

# Distribution of Anomaly Elements in Soil Heavy Metals Based on Singular Value Decomposition and Joint Sparse Model

Luo Qiu<sup>a, b\*</sup>, Gui Zhang<sup>b</sup>, Deqing Yu<sup>c</sup>, Qiming Xiong<sup>b</sup>

<sup>a</sup>School of Geosciences and Info-Physics, Central South University, Changsha 410012, China

<sup>b</sup>Central South University of Forestry and Technology, Changsha 410004, China

<sup>c</sup>Hunan Province Remote Sensing Center, Changsha 410007, China

bear218512@csu.edu.cn

Heavy metal enrichment in the soil affects health of animal, plant, and human. Therefore, the situation of heavy metal pollution and distributing rule is worth to research. This article investigated and assessed the status of pollution of heavy metals such as Cd, Hg, Pb, Cr, Ni, Ti, and Au in the study area. The singular value decomposition and joint sparse model have been successfully applied to map the distribution of elements such as Ni, Ti and Au in the research area. And the algorithm was found to be efficient, accurate and consistent with Nemerow index data and existing methods. This study lays a theoretical and instrumental foundation for future research using remote sensing analysis in a variety of areas, including geochemistry and human health.

## 1. Introduction

Soil, as a carrier of crops, its environmental quality directly affects the safety of agricultural products, and has a far-reaching impact on human health. With the rapid economic development, a large number of pollutants have entered the soil environment, and heavy metals are one of the important pollutants. There are many heavy metal elements, and it mainly refers to elements with significant biological toxicity such as Hg, Cd, Pb, Cr, and metalloids. It also includes certain toxic heavy metals such as Zn, Cu, Co, Ni, Sn, etc. Currently, the most compelling ones are Hg, Cd, Pb, As, Cr, and others.

In recent years, a program of cooperation in several political districts has resulted in expanded domestic research regarding environmental geochemistry. The research is primarily conducted using conventional methods and adopts remote-sensing and other auxiliary methods for a comprehensive analysis of the regions under study (Yu et al., 2015; Kemkin and Kemkina, 2015; Lin, 2003). The most relevant previous works related to this paper are: Ali, etc. (2015) who designed a geographic environment information and acquisition system to study geochemical flows and developed a sedimentation analysis model for the detection of gold, alkali and other metals in the north of Pakistan; Cong et al., (2013). who studied the relationship between spectral characteristics and distribution of gold deposits in Zhulazhaga region to perform a composition analysis of geochemical analysis data; Zhang, etc. who proposed a multipurpose remote sensing geochemical analysis method to guide field geochemical research; and Zhao, etc. who analyzed the vegetation in a copper ore region in Jiangcheng, Yunnan Province using remote sensing data for the prediction of the mineral field. The literatures above study the geochemical environment from different perspectives, all involving remote sensing image analysis methods.

As this technique evolves, it is critical to develop and field test new algorithms which can reveal previously imperceptible elements accurately. In this paper, a remote sensing method based on singular value decomposition, joint sparse model and field geochemical analysis is used to study both the Chang-Zhu-Tan urban agglomeration area and the remote sensing algorithms used.

## 2. Materials and methods

### 2.1 Study area

Chang-zhu-tan urban agglomeration is located at 26°3'N - 28°40'N, longitude 110°53'E - 114°15'E. It has about

2100 km<sup>2</sup> in terms of area, as shown in Figure 1, including three large dense-population areas which totally have 300 km<sup>2</sup> in terms of area and named as Changsha, Zhuzhou and Xiangtan. The three main parts of Chang-zhu-tan urban agglomeration are in the same physical geographical system, which belongs to the belt of hill and valley in downstream part of Xiangjiang River. Their physiognomy with small relief and abundant hydrographic net influences the diffusion of elements.

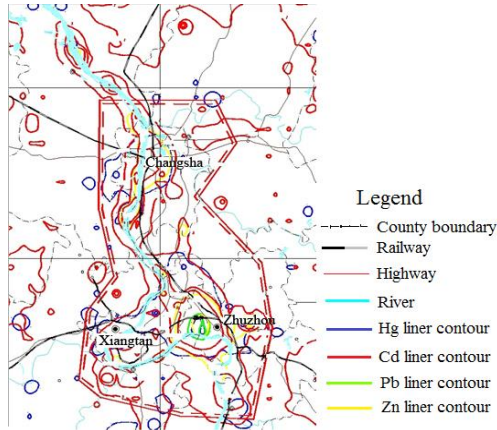


Figure 1: study area

## 2.2 Remote sensing geochemical analysis process

In geochemistry applications, the key procedures of remote sensing analysis mainly involve collection of samples, abnormality verification and abnormality analysis, etc. Apart from the use of general remote sensing methods (such as image collection, inlaying, revision, confirmation of training samples, categorization surveillance and interpretation), special treatment is also required in the flow chart, which is shown as Figure 2.

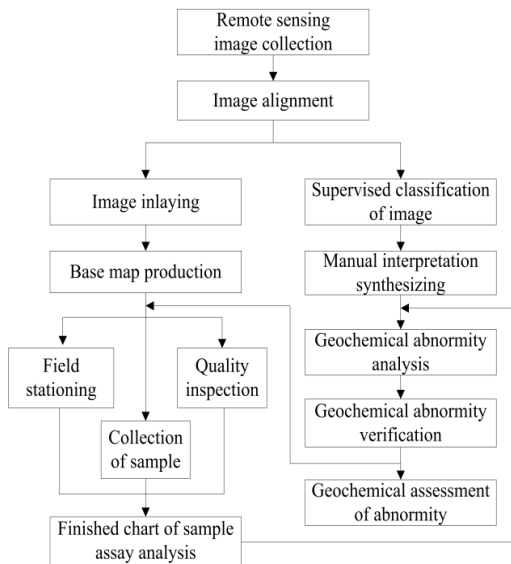


Figure 2: Geochemistry remote sensing analysis process

## 2.3 Joint sparse reconstruction of remote sensing

Remote sensing image analysis mainly performs sparse reconstruction operation based on matching tracing mode and matching tracking algorithm, and is essentially a type of greedy optimization algorithm, which finds the best matching solution through the iterative method. A significant feature in the matching pursuit method is that the algorithm is simple and efficient, so it has strong practicability when the accuracy requirement is not too high. For the above sparse restructuring, the greedy tracing optimization is used to achieve iterative optimization. Equations (1) and (2) describe the optimization model.

$$a = \arg \min \|a\|_0, \text{ s. t. } Px - AaP_2^2 < \varepsilon \quad (1)$$

$$a = \arg \min \|x - Aa\|_2^2, \text{ s. t. } PaP_0 < s \quad (2)$$

The commonly used greedy optimization approaches are OMP algorithm, MP algorithm and their various improved variations. To facilitate representation and analysis, it is assumed that all the atoms in dictionary D are standard (in which the  $l_2$  norm value is 1).

The pixel element (pixel) of remote-sensing high-spectral figure in adjacent space can be approximately demonstrated by constructing sparse linear combination of dictionary atoms. The proportion and weight of these atoms are diverse from each other, that means, the same atoms are representing the adjacent pixels, but the coefficients of adjacent pixels are different.

In the combined sparse analytical model of remote sensing images, it is usually assumed that the high-spectral adjacent pixels  $x_i$  and  $x_j$  can be composed of similar materials. The dictionary representation form of  $x_i$  is:

$$x_i = Da_i = a_i, a, + a_{i,var} a_{var} + L + a_{i,\lambda_k}, a_{\lambda_k} \quad (3)$$

In formula (1), the size of dictionary D is the matrix of  $M \times N$ .  $D_k = (\lambda_1, \lambda_2, L, \lambda_k)$  is the sparse base of coefficient  $a_i$ . Assume that  $x_i$  and  $x_j$  are composed of similar terrain feature elements, then  $x_j$  can be represented based on the linear combination of the same training sample  $\{a_k\}_{k \in A_k}$  and different training coefficients  $\{a_{j,k}\}_{k \in A_k}$ :

$$x_j = Da_j = a_{j,\lambda_1} a_{\lambda_1} + a_{j,Az} a_{var} + L + a_{j,\lambda_k}, a_{\lambda_k} \quad (4)$$

Under this assumption, the process can be extended to the adjacent pixels. Assume that  $N_\varepsilon$  contains T different pixels, which are represented based on the same training samples and the linear combination of their coefficients. If  $X = [x_1 x_2 L x_T]$  is a coefficient matrix with a size of  $M \times T$ , then  $\{x_t\}_{t=1, L, T} \in N_\varepsilon$  in the formula is the adjacent space pixel in the high-spectral remote sensing image. Then X can be represented as:

$$X = [x_1 x_2 L x_T] = [Da_1, Da_2, L, Da_T] = D[a_1, a_2, L, a_T] = DS \quad (5)$$

The sparse vector with form of  $\{a_t\}_{t=1, y} \cdot T$  is based on the same training samples and linear combination of their coefficients, i.e., it has the same form of  $D_k$ . The sparse matrix with form  $S \in R^{D \times T}$  has only a K-row non-zero matrix, i.e., the number of sparse matrix S composed of  $a_t$  is  $D_k$ . Assuming that training dictionary D is known, then matrix S can be represented and solved based on the following combined sparse restructuring form:

$$\text{Minimize } P S P_{row,0}, \text{ Subject to: } DS = X \quad (6)$$

In formula (6),  $P S P_{row,0}$  is the number of non-zero rows in the restructuring sparse matrix S, which can represent the heterogeneity characteristic of S. Then there is a need to solve  $\hat{S} = [\hat{a}_1, \hat{a}_2, L, \hat{a}_T]$ , while the sparse restructuring matrix with a size of  $M \times T$  only has a few non-zero rows. Based on the analysis of real data, formula (6) can be rewritten as (7) and (8) considering error:

$$\hat{S} = \arg \min P S P_{row,0}, \text{ subject to } PDS - XP_F \leq \sigma \quad (7)$$

$$\hat{S} = \arg \min \|DS - X\|_F, \text{ subject to } \|S\|_{row,0} \leq K_0 \quad (8)$$

In the above formula,  $\|\cdot\|_F$  is the norm of Frobenius. The combined sparse representation of the remote sensing image is essentially a problem of solving NP-Hard, which can be accomplished by constructing a new greedy algorithm. This algorithm is explained in short for it is not the major research focus concerning the paper.

## 2.4 Restructuring dictionary representation

Field and Olshausen were the first to study the construction of training dictionaries from data samples. Their method uses a construction dictionary that is over-complete and redundant. Selecting irrelevant natural image pixel block and taking it as training set, thus design line-direction partial filter on the basis of dictionary learning. There are many additional methods of dictionary learning, such as isolated component learning algorithm, and sparse prior learning dictionary construction. Dictionary learning problems can be represented and solved based on the sparse constraint form:

$$D = \arg \min \|X - Da\|_2^2, \text{ s. t. } \begin{cases} \forall i, Pa_i P_0 \leq k \\ \forall j, Pa_j P_2 = 1 \end{cases} \quad (9)$$

In formula (9),  $X = \{x_i\}$  is the sample figure set,  $a_i$  is the column vector of matrix a, sparse representation coefficient of image sample block, and standard atom existing in dictionary A. The difference between the above coefficient representation models and antecedent models is that the above models further refine the optimization objective from D to a, therefore further increasing operation of the optimization process. The problem shown in

formula (9) is non-convex and it can be transformed into a convex optimization problem by making one variable constant. Thus, in the actual computation, the universal practice is to assume the value of one variable is fixated, then estimate other variables. The commonly used algorithms are MOD (Method of Optimal Directions), general PCA (Generalized PCA), singular value decomposition (K-SVD), etc. Here we choose K-SVD algorithm for the establishment of the over-complete dictionary.

## 2.5 Sparse representation dictionary training

When dictionary parameters are updated, updating of optimization procedure evolution is pushed forward continuously by combining the optimization course with the sparse model representation. Here we use the greedy tracing mode to provide the training update of the constructed dictionary.

K-SVD decomposition algorithm is evolved from the K mean cluster, and it will degrade into the K mean cluster analysis algorithm if all the signals are only permitted to be approximated by only one standard atom. The following objective optimization problems in the above restructuring procedure of the dictionary need to be solved:

$$\min\{PY - DX P_F^2\} \text{ s. t. } \forall i, P X_i P_0 \leq T_0 \quad (10)$$

In formula (10),  $y \in R^n$  is the training sample,  $D \in R^{n \times K}$  is the dictionary and  $x \in R^K$  is the sparse vector.  $Y$  is the training set.  $X$  is the sparse vector.  $T_0$  is the upper limit amount of the non-zero sparse vectors.

K-SVD algorithm is essentially the updating process of iterative optimization and its dictionary updating is done column by column. Sparse vector updating is made along with that of the dictionary. If the dictionary to be updated is  $k$ -column coefficient  $d_k$ , assuming  $X$  and  $D$  are fixed, then:

$$PY - DX P_F^2 = \left\| Y - \sum_{j=1}^K d_j X_T^j \right\|_F^2 = \left\| Y \sum_{j=1}^K d_j X_T^j - d_j X_T^k \right\|_F^2 = P E_k d_k X_T^k P_F^2 \quad (11)$$

In formula (11),  $X_T^k$  is  $k$ -th row of  $X$ . First decompose  $DX$  into two parts:  $\sum_{j \neq k} d_j X_T^j + d_k X_T^k$  and  $E_k$ , which stands for the remaining  $N$  samples and error matrix of the original sample after removal of atomic  $k$ . Then  $d_k$  and  $X_T^k$  are obtained using SVD algorithm. In solving formula (11) above for the atom  $d_k$ , updating does not carry out sparse constraint representation, then the solving of  $X_T^k$  has errors. Therefore, the first step is to define set form:

$$\omega = \{i | 1 \leq i \leq K, x_T^k(i) \neq 0\} \quad (12)$$

$\omega$  is the index construction set of all the samples  $\{y_i\}$  of  $d_k$ , and define  $\Omega_k$  to make matrix of  $N \times |\omega_k|$  form. And the values are 1 at position  $(\omega_k(i), i)$  of the matrix, other positions are 0.  $X_R^k = X_R^k \Omega_k$ ,  $Y_R^k = Y_T^k \Omega_k$  and  $E_R^k = E_k \Omega_k$  are all shrinkage results after zero input is eliminated. Then the problem of formula (12) can be converted into:

$$P E_k \Omega_k - d_k X_T^k \Omega_k P_F^2 = P E_R^k - d_k X_R^k P_F^2 \quad (13)$$

Formula (13) can be solved using K-SVD decomposition. The K-SVD algorithm can train the dictionary by applying the flexible optimization algorithm.

## 3. Results and discussion

### 3.1 Analysis of abnormal remote sensing image

Figure 3 shows that the algorithm in this paper can get an accurate abnormal distribution image of Ni and Ti, and it can quickly and efficiently map and contrast the abnormal element distribution region with the normal distribution region. This greatly reduces surveying and mapping costs. Remote sensing image tones correlate with concentrations of geochemical pollution elements, so it can be preliminarily concluded that the above abnormalities are caused by human factors. The image tone within the interpretation borderline in figure 3(a) is basically uniform, while in the geochemically abnormal element images as shown in figure 3(b) and (c) the tones vary dramatically, with red areas having abnormally high element content and blue areas having abnormally low element content. Our analysis found out about 20 elements having higher content, such as Ni, Cu, Co, Cr, MgO, etc., and about 10 elements having lower content, such as Ba, Ti, Rb, etc. The variations in concentrations of these elements are highly consistent.

Figure 4 shows the remote sensing analysis image of Au, As and Sb distributions in the Changsha area in Chang-Zhu-Tan urban agglomeration as an example, by comparing (a) and (b) in Figure 4, it can be seen that the algorithm can also identify abnormalities in Au, As, and Sb accurately. Abnormality amounts in Figure 4(a) are indicated by contour lines, while in the remote sensing image analysis in Figure 4(b), the element concentration is shown by the denseness of red mark points. The result is highly consistent with the distribution results of measured contour lines in Figure 4 (a).

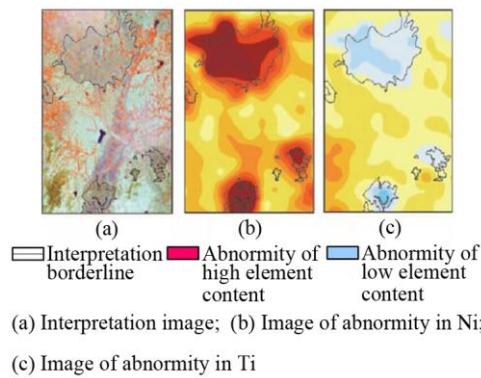


Figure 3: Remote sensing geochemical abnormality somewhere in Zhuzhou

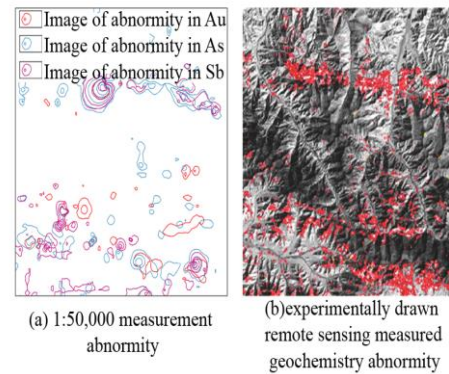


Figure 4: Remote sensing geochemical abnormality somewhere in Changsha

### 3.2 Nemerow index

In order to validate the model, basic pollution conditions of the Chang-Zhu-Tan region were determined using the Nemerow index for urban pollution.

$$I_{ij} = \sqrt{[(1/n \sum p_{ij})^2 + P_{ij\max}^2]/2} \quad (14)$$

In formula (14),  $I_{ij}$  is the Nemerow index, which is mainly used to evaluate the combined environment pollution index.  $P_{ij}$  is the standard pollution index which is equal to the measured single element value divided by the standard evaluation value.  $P_{ij\max}$  is the maximum value of Nemerow index in the pollution elements.  $n$  is the element quantity of pollution media.

Here we use the Nemerow evaluation index to analyze the medium based on national standards. The grading strategy, according to the Guide for Eco Region Geochemical Analysis, is shown in Table 1, while Table 2 shows the results of Nemerow overall evaluation in the Chang-Zhu-Tan region.

Table 2 shows that pollution level in Changsha is lower than Xiangtan and Zhuzhou, and the number of severe pollution samples in the two cities accounts for more than 70% of the total samples. In Zhuzhou, the number of severe pollution samples accounts for nearly 50% of the total samples. The marshaling sequence of the general pollution in the three cities in terms of severity is: Zhuzhou > Xiangtan > Changsha.

Table 1: Grading of Nemerow evaluation

<1.0	1.0~2.0	2.0~3.0	3.0~6.0	>6.0
Clear area	Mild pollution	Moderate pollution	Heavy pollution	Severe pollution

Table 2: Results of Nemerow overall evaluation in Chang-Zhu-Tan region

City	Changsha		Xiangtan		Zhuzhou	
	Sample	Proportion	Sample	Proportion	Sample	Proportion
Environmental quality						
Clear area	0	0.0	0	0.0	0	0.0
Mild pollution	25	8.5	10	5.6	10	5.5
Moderate pollution	123	41.7	38	21.2	41	22.7
Heavy pollution	94	31.8	67	37.4	41	22.7
Severe pollution	53	17.9	64	35.8	89	49.2

### 3.3 Comprehensive pollution distribution of Chang-Zhu-Tan urban agglomeration

To confirm the validity of the algorithm used in this study, the distribution contrast results are shown in Figure 5 along with the results of a contrasting algorithm developed by Liu Jiying et al. The analysis results shown in Figure 5(a) and 5(b) indicate similar color/pollution concentration distributions calculated by the two algorithms. Both algorithms calculate the same general pollution severity marshaling sequence for the three cities as Zhuzhou > Xiangtan > Changsha, which is consistent with the Nemerow results provided in section 5.2. However, the algorithm used in this paper appears to have better accuracy in terms of resolution of detail and contrast between adjacent areas with different pollution levels. The convergence curves of the objective functions in

Figure 6 show that the algorithm in this paper offers a performance advantage over that of Liu Jiying in terms of convergence speed and convergence accuracy.

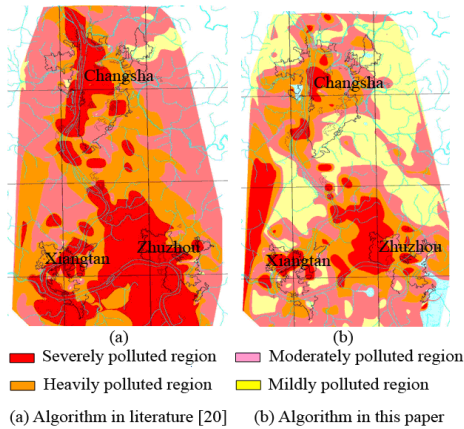


Figure 5: Analytic result of remote sensing image of Chang-Zhu-Tan urban agglomeration

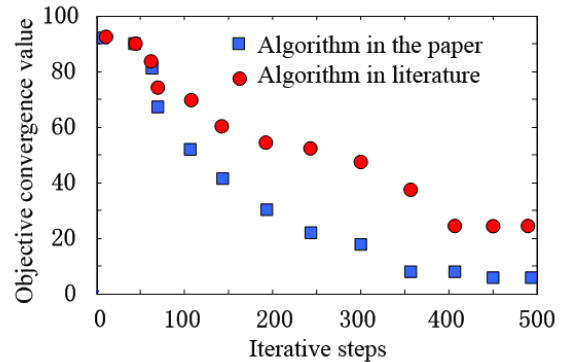


Figure 6: convergence curve (Liu et al., 2010)

#### 4. Conclusion

This paper provides a practical, accurate remote sensing algorithm, which is an improvement on previous work in the field of geochemical environmental remote sensing. The algorithm uses singular value decomposition and joint sparse modeling techniques. A test of the algorithm found that it provides a map of the abnormal geochemical pollutants in the Chang-Zhu-Tan urban agglomeration area that is consistent with other remote sensing algorithms as well as Nemerow index data. This study lays a theoretical and instrumental foundation for future research using remote sensing analysis in a variety of areas, including geochemistry and human health.

#### Acknowledgements

This paper is based on the key research project in Hunan, which is "Natural disasters data based on satellite remote sensing monitoring interpretation technology research"(Item number: 2016SK2088), and the project of "eco-geochemical investigation in Dongting Lake District of Hunan Province". Thanks to Professor Gui Zhang and Yu Deqing for providing remote sensing data and guidance, and my junior fellow apprentice for inspiration in the written process. Here, I wish to express my heartfelt thanks.

#### Reference

- Ali L., Moon C.J., Williamson B.J., Shah M.T., Khattak S.A., 2015, A GIS-based stream sediment geochemical model for gold and base metal exploration in remote areas of northern Pakistan, *Arabian Journal of Geosciences*, 8(7), 5081-5093, DOI: 10.1007/s12517-014-1531-7
- Cong L.J., Cen K., Yu X.Y., 2013, Relationship between spectral characteristics and geochemical composition of Zhulazhaga gold deposit, *Journal of Central South University (Science and Technology)*, 44(1), 266-273.
- Kemkin I.V., Kemkina R.A., 2015, Geochemical evidence of an oceanic provenance of cherts in accretionary complexes in the Sikhote Alin, *Geochemistry International*, 53(8), 700-712, DOI: 10.1134/S0016702915080029
- Lin N., 2003, Study on the Metallogenic Prediction Models Based on Remote Sensing Geology and Geochemical Information: A Case Study of Lalingzaohuo Region in Qinghai Province, *Jilin University*, 26-38.
- Liu J.Y., Zhu J.B., Yan F.X., Zhang Z.H., 2010, Design of remote sensing imaging system based on compressive sensing, *Systems Engineering and Electronics*, 32(8), 1618-1623, DOI: 10.3969/j.issn.1001-506X.2010.08.14
- Yu D.S., Liang D.L., Lei L.M., Zhang R., Sun X.F., Lin Z.Q., 2015, Selenium geochemical distribution in the environment and predicted human daily dietary intake in northeastern Qinghai, China. *Environmental Science and Pollution Research*, 22(15), 11224-11235, DOI: 10.1007/s11356-015-4310-4