

ASDPS-based Geochemical Data Processing and Anomaly Identification

WANWU GUO

School of Computer and Information Science
Edith Cowan University
2 Bradford Street, Mount Lawley, WA 6050
AUSTRALIA

Abstract – Geochemical data processing aims firstly at reducing random and/or systematic errors resulted from the field survey and/or laboratory analysis, and secondly at identifying whether the data contain some useful information indicating the existence of mineral concentrations, oil fields, and pollution sources in the survey area. The first task is usually carried out by using various smoothing approaches. However, how to determine the ‘best’ outcome from many smoothing methods still depends on a geochemist’s personal judgment. The second task is achieved by making the data in reference to some geochemical benchmarks. In this paper, a statistical method is proposed to determine the likely ‘best’ outcome among many smoothed datasets, and then this ‘best’ fitted dataset is used to determine anomalies in reference to different geochemical benchmarks. The proposed statistical selector quantifies the determination of smoothing processing of geochemical data. The anomaly classifiers proposed can automatically identify and classify any geochemical anomalies contained in the data.

Key-Words: - Statistical model, ASDPS, Data matrix, Geochemical data processing, Anomaly classification

1 Introduction

In geochemical survey for energy/mineral exploration and land/soil quality assessment, sampling is normally carried out in different sized grids for different purposes [1][2][3][4]. Analysed geochemical data are commonly formed as matrixes, which represent values distributed in squared grids. The major tasks of geochemical data processing are firstly to reduce any random and/or systematic error generated in the field survey and/or laboratory analysis, and secondly to assess if the data contain some useful information indicating the existence of mineral deposits, oil fields, and pollution sources in the surveyed area.

The first task is usually achieved by using various smoothing approaches. A moving window varying from 2×2 to 9×9 may be used to smooth the data with either a locally best-fitted plane model or quadratic surface model. If a small window is used, some potential systematic errors may not be removed whereas if a large window is used, it is likely to artificially spread the elements to broader surrounding sub-areas, which may lead to the creation of some ‘false’ targets. Therefore, there has been some uncertainty in determining the ‘best’ outcome from many smoothed datasets in practice.

To do the second task, some geochemical references must be used as references. For example, each geochemical element has an average value in the Earth’s crust that is called the geochemical background (GCBG) of that element. For an element in an area to be considered having formed an anomaly, its value must be higher than its local geochemical anomaly threshold (GCAT), which is normally higher than its GCBG. Even if an element in an area is indeed higher than its GCAT, it may still not be considered having significance because the rocks/soils in that area may contain a higher background value of that element without economic value.

In this paper, a statistical model is proposed to determine the likely ‘best’ outcome among many smoothed datasets, and then this selected dataset is used to determine the potential anomalies in reference to different benchmarks. These approaches are applied to the processing and anomaly identification of Cu data collected in an area in northern China to prove their usefulness.

2 Methodology

Before determining the ‘best’ outcome among different smoothed datasets, we must first get these

processed results. As geochemical dataset is normally organized as a data matrix in a 2D array $A(m \times n)$, the parallel array processing system (ASDPS) proposed recently by Guo [5] can be used to perform smoothing for each geochemical element involved.

Assuming that the smoothed data arrays are available as $R_k(m \times n)$ ($k = 1, \dots, N$), we must first set selection criteria to determine which processed data array should be regarded as the 'best' outcome. The first factor to be considered is the window sizes used. Although in theory any window size can be used for smoothing processing, as mentioned in first section, windows with too small and too large sizes are actually not useful. Therefore, medium sized windows should be preferred in smoothing processing. This means that 3×3 and 5×5 windows are likely to produce reasonable results for smoothing processing.

With these two sized windows, either locally best-fitted plane model or quadratic surface model can be used to carry out smoothing. To evaluate the fitness of these results to their original data, a selector is designed based on the statistics of the residual array resulted from the original data array and its smoothed array, which can be expressed as

$$S_{\min}(ave, std) = \underset{i,j}{\text{MIN}} \left\{ \sum_{i,j}^{m,n} \text{abs}[R_k(i, j) - A(i, j)] \right\}$$

$$= \text{MIN} \sqrt{(ave)^2 + (std)^2} \quad (1)$$

where *ave* and *std* are the average and standard deviation of the residual array. This selector will

determine the smoothed array R_{\min} that has the closest fitness to the original data, which is regarded as the 'best' outcome for the smoothing processing. Data array R_{\min} is then used for anomaly identification.

Three different references are used for anomaly identification. The necessary condition for an element to form an anomaly is that its values must be higher than its GCBG, i.e.,

$$R_{BA}(i, j) = R_{\min}(i, j) - GCBG > 0 \quad (2)$$

This will result in a data array R_{BA} named as background anomaly (BA) of that element.

If GCAT is used, instead of GCBG, a data array R_{TA} called threshold anomaly (TA) is then produced as

$$R_{TA}(i, j) = R_{\min}(i, j) - GCAT > 0 \quad (3)$$

TA indicates the existence of potential anomalies in the sub-area where the anomaly is located.

In mineral exploration, usually a concentration of an element that is at least 2 times higher than its GCAT can be classified as a reliable anomaly. For this sake, the data array (R_{RA}) that represents such a reliable anomaly (RA) can be computed by

$$R_{RA}(i, j) = R_{\min}(i, j) - 2GCAT > 0 \quad (4)$$

This data smoothing and anomaly identification procedure is summarised in Figure 1.

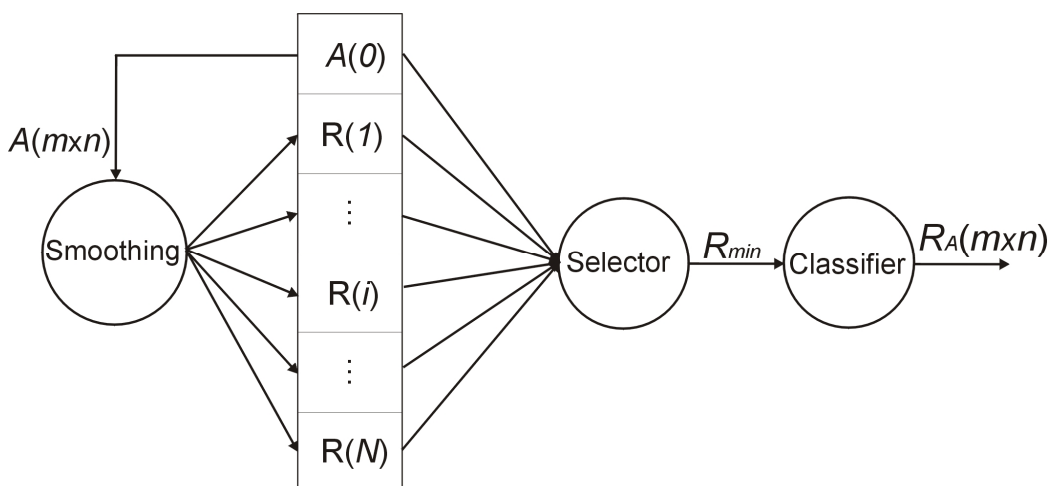


Figure 1. Procedure of geochemical data processing and anomaly identification

3 Application to Cu Data Processing and Anomaly Identification

The Cu data from a reconnaissance geochemical survey in northern China is shown in Figure 2. The GCBG of Cu element is 5.73 ppm and its GCAT is 11.49 ppm in the area. The rocks exposed in this area are predominately Jurassic sedimentary rocks and magmatic rocks. The Jurassic sedimentary rocks are mainly distributed in the central and western regions

mostly with Cu distribution below its GCBG. The magmatic rocks expose in the northeastern region and southeastern corner with Cu distribution over its local GCAT. These high Cu distributions imply the existence of potential Cu concentrations in these sub-areas. The method of data processing and classification proposed in the previous section can be used to identify and classify the Cu anomalies, if there is any, in the area.

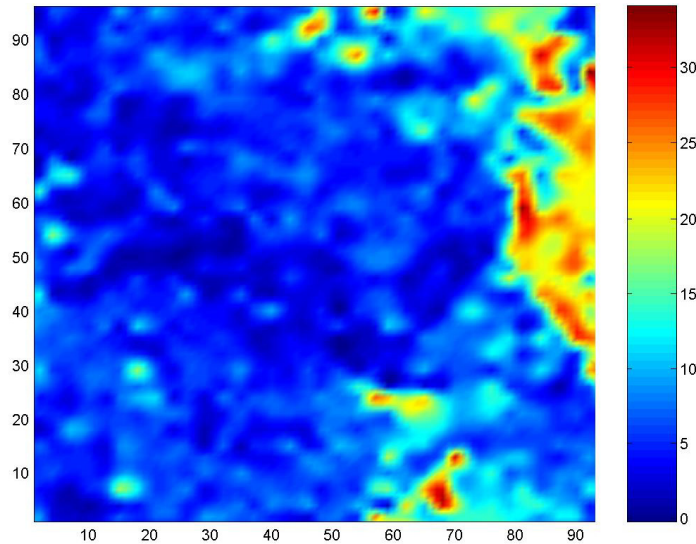


Figure 2. Image of original Cu data array

To smooth the original Cu data, we use the parallel processing system (ASDPS) proposed by Guo [5], in which each new array component is computed using a moving 3×3 or 5×5 window by

$$R_k(0,0) = \sum_{i=0}^N W_k(i)A(i) \quad (5)$$

where $A(i)$ is the circular average of all points on the i th circle, and $W_k(i)$ is the constant for all points in the i th circle of the k th smoothing method. The weight constants for the smoothing methods used to process Cu data array are given in Table 1 and their corresponding images are shown in Figure 3.

Visually there are no significant differences between the smoothed and original images, except the 5×5 plane model (Fig. 3c). With this smoothing, local Cu concentrations are weakened whereas the regional highs are expanded to the surrounding areas.

To find the best-fitted smoothing result among these four data arrays, these arrays are processed using the selector of Relation (1). The statistical results for all models are listed in Table 2. Obviously the 3×3 surface model produces the best result whereas the 5×5 plane model has the worst fit. Other two models also generate reasonably good outcomes.

Table 1. Weights used in Cu data smoothing using 3×3 and 5×5 moving windows

Result	Method	$W(0)$	$W(1)$	$W(2)$	$W(3)$	$W(4)$	$W(5)$
Fig. 3a	3×3 plane	0.1111	0.4444	0.4444	0	0	0
Fig. 3b	3×3 surface	0.5556	0.8888	0.4444	0	0	0
Fig. 3c	5×5 plane	0.0400	0.1600	0.1600	0.1600	0.3200	0.1600
Fig. 3d	5×5 surface	0.1543	0.5028	0.3886	0.1600	0.0914	-0.2972

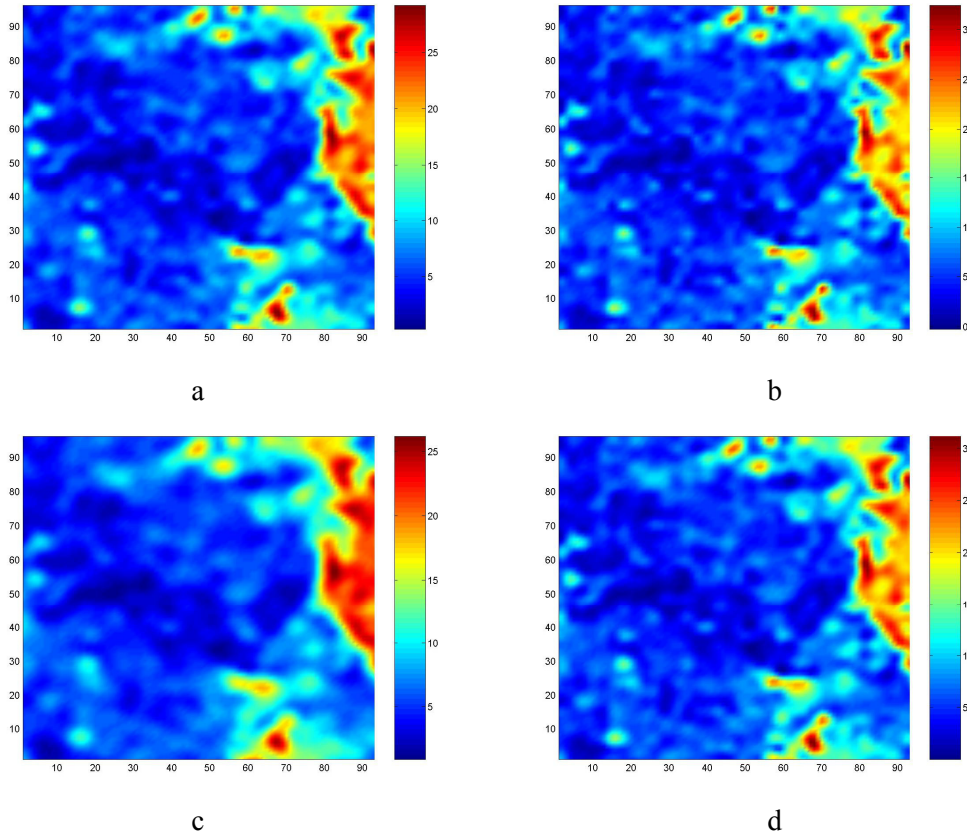


Figure 3. Images of smoothed Cu data
 a - 3×3 plane, b - 3×3 surface, c - 5×5 plane, d - 5×5 surface

Table 2. Statistical results of different smoothing models

Result	Method	<i>ave</i>	<i>std</i>	R_{min}
Fig. 3a	3×3 plane	0.3926	0.5596	0.6836
Fig. 3b	3×3 surface	0.0549	0.1049	0.1183
Fig. 3c	5×5 plane	0.9171	1.1290	1.4545
Fig. 3d	5×5 surface	0.2556	0.4340	0.5037

To identify Cu anomalies that may have been contained in this Cu data array, Relation (2) is used to determine the sub-areas where the Cu values in $R_{min}(m \times n)$ are higher than its GCBG. The resultant

BA image is shown in Figure 4. It is clear that the sub-areas where magmatic rocks expose have the potential in finding some useful Cu anomalies.

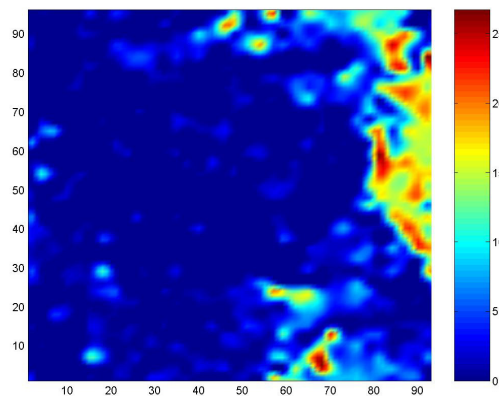


Figure 4. The background anomaly image of Cu element

By using Relation (3) to $R_{min}(m \times n)$, the threshold anomalies are identified associated with magmatic rocks, and most background anomalies in Jurassic rocks are eliminated as insignificance (Fig. 5).

By directly using Relation (4) to $R_{min}(m \times n)$, the resultant reliable anomalies are outlined in Figure 6. The reliable anomalies only appear in a few isolated

locations, which should be the primary targets for discovering potential copper deposits in this area. These anomalies are associated with magmatic rocks where copper may have been enriched locally to the industry standard as copper deposits.

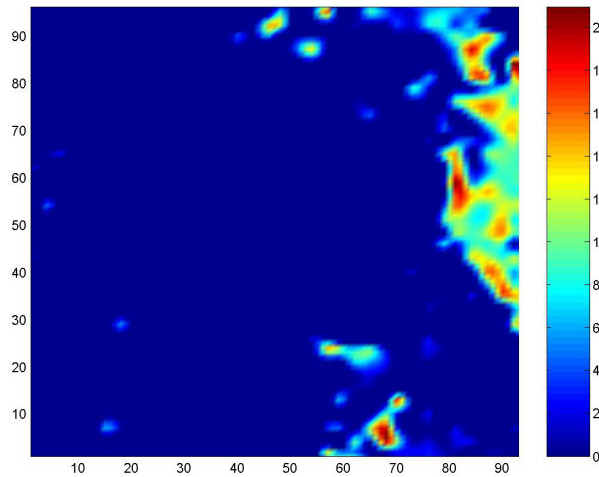


Figure 5. The threshold anomaly image of Cu element

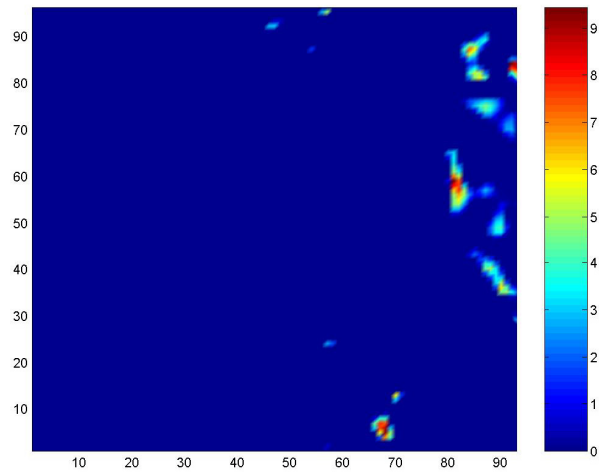


Figure 6. The reliable anomaly image of Cu element using Relation (4)

4 Conclusion

The example of Cu element demonstrates that the statistical model is useful for geochemical data processing and anomaly identification. The proposed statistical selector quantifies the determination of smoothing processing of geochemical data, which is the first time being used in geochemical data

processing. The anomaly identifiers proposed in this paper lead to automatic anomaly identification and classification in geochemical data interpretation, which will have a profound impact in geochemical exploration and survey in the future.

References:

- [1] A.H. Brownlow, *Geochemistry*, Prentice Hall, 1996.
- [2] G.J.S. Govett (ed.), *Handbook of Exploration Geochemistry, Volume 2, Statistics and Data Analysis in Geochemical Prospecting*, Elsevier, 1983.
- [3] A.A. Levinson, *Introduction to Exploration Geochemistry (2nd Edition)*, Applied Publishing Limited, 1980.
- [4] A.W. Rose, H.E. Hawkes and J.S. Webb, *Geochemistry in Mineral Exploration (2nd Edition)*, Academic Press, 1979.
- [5] W. Guo, "Adaptive spatial data processing system (ASDPS)", Knowledge-Based Intelligent Information & Engineering Systems, *Lecture Notes in Artificial Intelligence*, 3214, Springer-Verlag, 2004, pp. 432-439.