ISSN : 0974 - 7435 Volume 10 Issue 20

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(20), 2014 [12365-12370]

Gross error detection method based on wavelet theory of mining spatial data

Chen Ling-Xia*, Zhang Jun-Li
College of Tourism and Resource Environment, Xianyang Normal University, Shanxi, Xianyang 712000, (CHINA)
E-mail : alingyaoxin@sina.com

ABSTRACT

Some gross errors in mining spatial data may occur during the process of data collection owning to the natural or human factors. The existence of gross error will affect the result of measuring, so it is of great necessity to explore an effective detection method to find out and eliminate the errors. The paper aims to observe the detection of the gross error in mining spatial data by using the multi-resolution capability of wavelet analysis, and as well to make an analysis of the influence of different wavelet function and decomposition upon the gross error by the case study of mine drilling data, and therefore it finally confirms the use of db2 wavelet decomposition into four layers for gross error detection. Meanwhile, it accurately pinpoints the existing gross error which should be eliminated combining with the spatial distribution characteristics of the mine drilling data. It is proved to be practical by applying the way of wavelet analysis to detecting mining spatial data and which is of great value to solve the deficiency of traditional gross error detection.

KEYWORDS

Wavelet theory; Mining spatial data; Gross error detection.

INSTRUCTION

Gross error, also can be regarded as negligence error or abnormal error. Usually called a bad or outliers from a gross error of the observation value, which showed big difference between the observation results and true value, that need a certain kind of detection method to find it and get rid of it, for a influence results. Actually in data collection, the natural and human environment always made some gross errors. So far the detection methods are getting more, and the classic detection methods are simple, but need normal data, meanwhile, the calculation are more complicated without strict theory; although the data detection method is better, there are still complex faults; Even the calculation of regression analysis is relatively simple, practical, but the precision is not high. Because of the low pass filter of wavelet analysis functions, the original signal could be decomposed, the use of performance in different scales could be separated the gross error from noise. By the wavelet transform of the multi-scale analysis, in the high frequency part can detect the normal signal transient abnormal phenomenon and show its composition. These gross error detection data of wavelet theory which applicated from GPS dynamic monitoring[1,2], gas monitoring[3], earthquake[4] and foundation pit[5], make better effect. So the wavelet theory should be more applied in mine spatial data for gross error detection.

THE DECOMPOSTION AND RECONSTRUCTION OF WAVELET

Due to multi-resolution, the wavelet analysis is easier to implement also with better effect, using the algorithm of Mallat multiresolution decomposition[6,7], analysis of signal may be coarse almost in position. Mallat algorithm of the basic idea is to put the wavelet base which expressed as a series of high pass and low pass filter group. Generalization, S is original signal, the wavelet coefficient (cA_j) the output of the low-pass branch, branch of wavelet detail coefficients (cD_j) qualcomm output, as show in Figure 1.

The characteristics of the wavelet decomposition is the largest scale, under the limit of iteration. The largest scales related to the signal length and the length of wavelet base. Refactoring, contrary to decompose, from the bottom to 2 times factor to start up to synthesis of original signal, A_j means reconstruction of wavelet signals, D_j means wavelet details, as shown in Figure 2.

$F(x)$ is a data model, use the Mallat algorithm, the signal is decomposed into different frequency components:

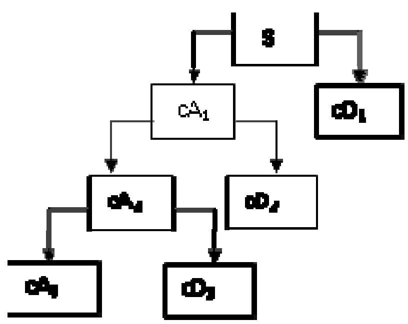

Figure 1 : The wavelet decomposition

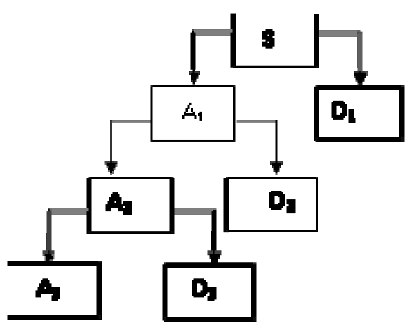

Figure 2 : The wavelet reconstruction

$$f(x) = A_j f(x) + D_j f(x) \tag{1}$$

Among them, $A_j f(x)$ is the frequency of the signal over the composition of 2^{-j}, $D_j f(x)$ is the frequency between the components of 2^{-j} and 2^{-j+1}. The matrix form of the wavelet decomposition as the following:

$$C_{j+1} = HC_j, D_{j+1} = GC_j (i = 1,2,\cdots, J) \quad (2)$$

Type: J --- The layers of wavelet decomposition;
H --- Low pass filter,
G --- High-pass filter,

After the Mallat decomposition, new filtering sequence $\tilde{C}_{j+1}$ and $\tilde{D}_{j+1}$ be formed, Mallat algorithm reconstruction.

$$\tilde{C}_j = H^* \tilde{C}_{j+1} + G^* \tilde{D}_{j+1} (j = J, J-1,\cdots,1) \quad (3)$$

Among them, the conjugated H* and G* were H and G

GROSS ERROR DETECTION OF MINING SPATIAL DATA BASED ON WAVELET THEORY

The distribution of underground space object has a certain continuity, in the presence of geological structures (folds, faults,) the location of the discontinuous phenomenon. If using wavelet analysis, data trend is low frequency part of these objects, and mutation is reflected in the high frequency signal. So detection of gross error on mining spatial data the basic idea is: First, the wavelet decomposition of the original signal; then the wavelet coefficients of the high frequency part of the decomposed after threshold processing; finally the signal reconstruction to eliminate the gross error.

The steps of outlier detection based on wavelet decomposition and reconfiguration is:

(1) The wavelet decomposition. The key of wavelet decomposition is on the selection of wavelet functions and decomposition layers, according to the characteristics of data, the choice of wavelet function, decomposition level must be on the analysis of different decomposition level data, to determine the decomposition level N. By the wavelet decomposing mutation was found in the high frequency part of the position, and then according to the actual situation, to determine whether it is true or not due to significant changes in the environment of normal value; otherwise, would be identified as gross error[8].

(2) the threshold value of high frequency coefficient quantification. If the rejection of gross errors is the choice of correction method, it can be to each layer of high frequency coefficient of N layer first, processing. Threshold processing is for a threshold processing exceeds the threshold value; on the other hand, is not processing. This threshold can be the default threshold and soft threshold and compulsion is the high frequency coefficient of 0[9].

(3) The wavelet reconstruction. To reconstruct the last layer of the low frequency coefficient after wavelet decomposition and threshold de-noising treatment after the layers of high frequency coefficient, coarse sent out the signal.

THE EXPERIMENT AND RESULT ANALYSIS

The use of wavelet analysis for gross error detection, need to select the type of wavelet function and scale. Used to obtain the data of 56 drilling grouting elevation coordinates, for example, the use of Matlab software by wavelet tools, first analysis of different wavelet functions and scale significance in gross error detection, and then determine the gross error detection measure adopted by the wavelet functions and scale size (decomposed layers). The original signal images of mining spatial data as shown in Figure 3.

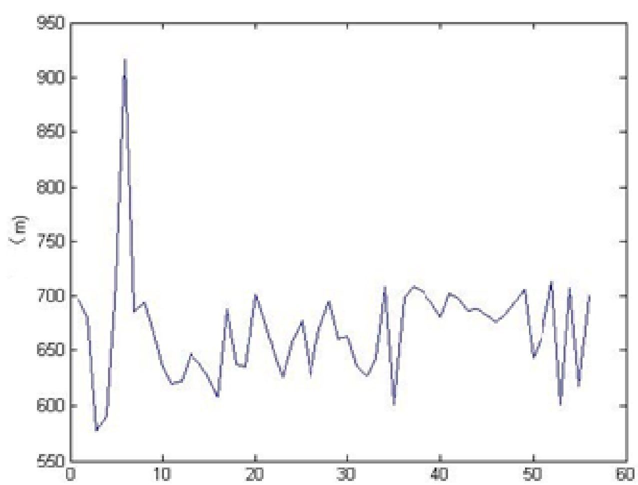

Figure 3 : Original signal

From the original signal graph, it can be seen that there exists in peak signal, this value is the gross error detection method, must be adopted to judge outliers. By using the method of wavelet theory for gross error detection and rejection. The gross error detection of the wavelet theory, there are two important problems need to be identified, one is the choice of wavelet function, and the other is a scale selection.

The influence of different wavelet functions for gross error detection

Different regularity of different wavelet functions, regular higher smoothness make better function. Effect of regular stability reconstruction wavelet coefficients, so for the different wavelet functions, the regularity of different, different signal decomposition and reconstruction effect. Considering the characteristics of mine data, select the regularity and vanishing moments better dbN wavelet function as the basis function. The following use of crude difference signal is decomposed into 4 layers DB2, db3, DB4, db5 respectively, as show in Figure 4, Figure 5, Figure 6 and Figure 7.

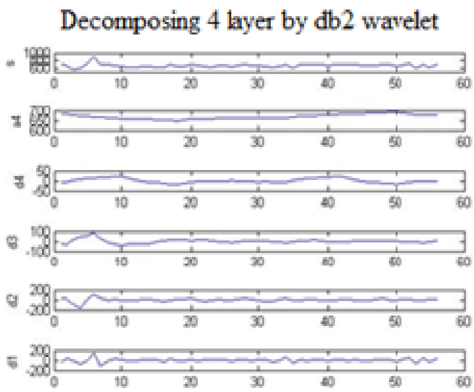

Figure 4: Db2 wavelet

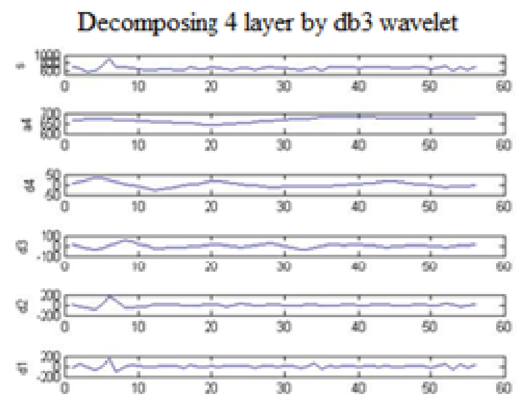

Figure 5 : Db3 wavelet

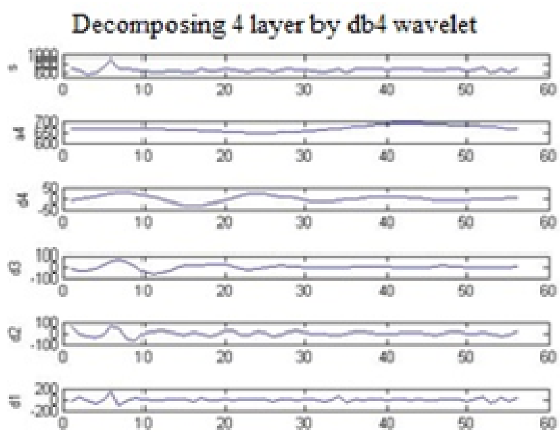

Figure 6 : Db4 wavelet

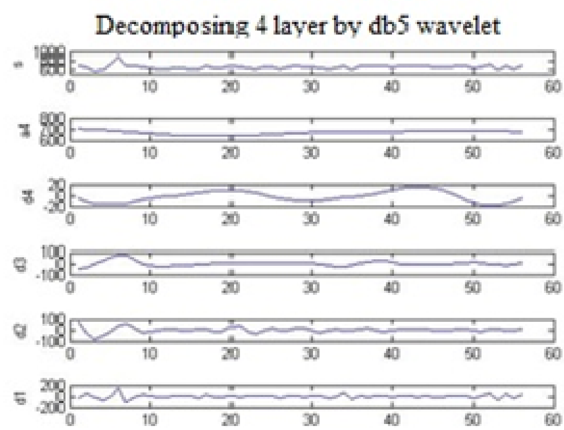

Figure 7 : Db5 wavelet

From Figure 4 to Figure 7 db2, db3, db3 and db5 wavelet, in d1, d2, d3, the high frequency part of point mutation, but the db2, db3, db3, but in these wavelet low resolution of the fourth floor. In db3, db3 wavelet low frequency part and some outstanding high frequency part. In db2 four layers of decomposition, high-frequency layer 1 and layer 2 and layer 3 mutations part is consistent, and very obvious. Other wavelet mutation have reflected, but not as good as the db2 mutation point is remarkable. Considering various wavelet function in the performance of the low frequency part and high frequency part, select the db2 wavelet for gross error identification

The influence of different decomposition level of gross error detection

Choose the right db2 regularity wavelet including gross error signal respectively for 2 layer, 3, 5 and 6 layer decomposition, the results are shown in Figure 8 and Figure 9, Figure 10 and Figure 11.

From the Figure 8 and Figure 9, Figure 10 and Figure 11 learn that gross error in the high frequency part is reflected, but for three layers of decomposition, in the second and third layer mutation point obviously; Decomposition is four layer on layer 1 and layer 2 and layer 3 mutation point clear; In the fifth and sixth floors down in the third layer mutation point is remarkable. In different decomposition layers, the third layer of the singular point is very obvious, can accurate positioning. Comprehensive consideration, the choice of db2 4 layer wavelet decomposition to locate the gross error.

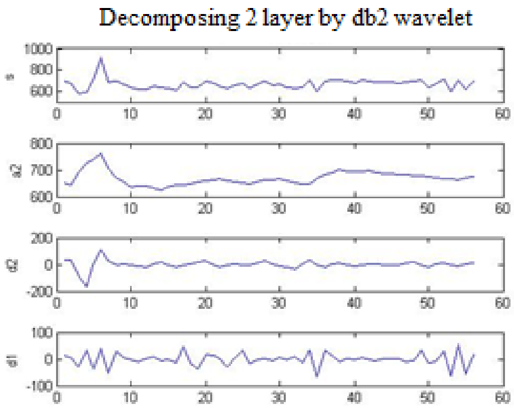

Figure 8 : Decomposing 2 layer

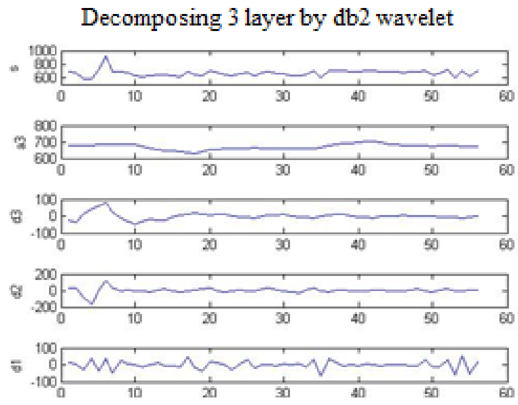

Figure 9 : Decomposing 3 layer

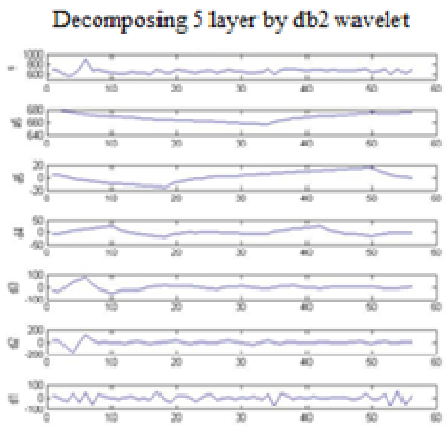

Figure 10 : Decomposing 5 layer

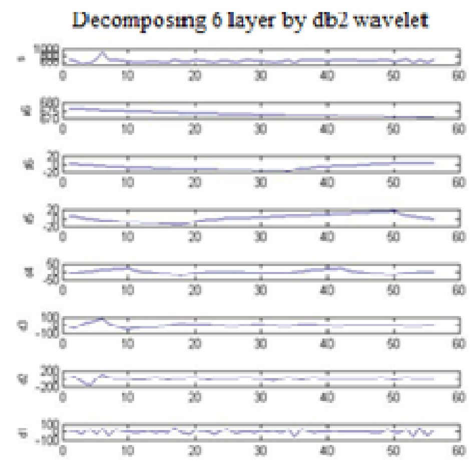

Figure 11 : Decomposing 6 layer

The gross error detection

From Figure 4, can be precise the positioning of singular point, which be found in the sixth data values change is bigger, show that the data may be a gross error.

The gross error elimination

As the sixth point data in the signal, and It is the borehole, CH4-1 in Figure 12, the fault is shown in black and point said, drilling can be seen from the graph, there is no geological structure, and its surrounding after comprehensive analysis of data, determine the abnormal data values, should choose out.

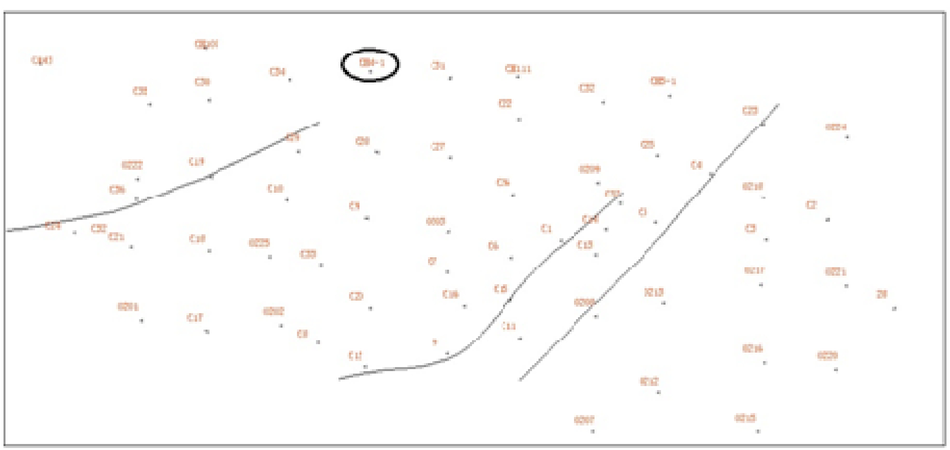

Figure 12 : The gross error elimination graph

CONCLUSIONS

Drilling data mining analysis of borehole data based on whether is gross error, by using the multiresolution of wavelet analysis, to study the different layer number of wavelet function and wavelet decomposition in the gross error detection.

With the experimental analysis, make sure to use db2 4 layer wavelet decomposition, determine the borehole data is gross error, and pinpoint the existence of gross error data. In combination with the practical situation of data distribution, the gross error should be removed. If data anomalies are reasonable, should be adopted based on the theory of wavelet threshold.

ACKNOWLEDGEMENT

This study was founded by Xianyang Normal College special fund (No.13XSYK031), The authors also gratefully acknowledge the help of anybody for providing valuable comments on the manuscript.

REFERENCES

[1] Huang Lian-Ying; Tthe Analysis of Deformation Characteristics to GPS Dynamic Monitoring Data Based on Wavelet Analysis. J.Changchun Inst.Tech.(Nat.Sci.Edi), **13(3)**, 15-20 **(2012)**.

[2] Chen Xiang-Yang; Application of Wavelet Analysis in Gross Error Detection of GPS Dynamic Monitoring Data. Journal of Nantong Vocational & Technical Shipping College, **12(3)**, 47-50 **(2013)**.

[3] Deng Gan-Bo, Zou Yun-Long; Denoising Processing of Gas Monitoring Data Based on The Theory of Wavelet. Echnological Development of Enterprise, **31(11)**, 172-173 **(2012)**.

[4] Shi Yu-Cheng, Zou Li-Hua, Li Shu; Optimizing to Earthquake Wave Based on Wavelet Packet. Journal of Lanzhou University (Natural Sciences), **46**, 132-137 **(2010)**.

[5] Wu Hong, He Yue-Guang, Xiong Sha, Ren Yun-Zhi; The Processing of Foundation Pit Monitoring Date Based on Wavelet De-noising Method. Beijing Surveying and Mapping, **1**, 7-10 **(2013)**.

[6] Wang Jian, Gao Jing-Yang, Sun Xiang-Zhong, et al; Wavelet Thresholding of the GPS Single Epoch Deformation Signal. Science of Surveying and Mapping, **29(1)**, 24-25 **(2004)**.

[7] Wu Dong-Hui, Tian Lin-Ya, Zhang Jin-Hua; Application of Wavelet Time-series in Settlement Monitoring for Urban Subway. Science of Surveying and Mapping, **38(2)**, 150-151 **(2013)**.

[8] Ji Zhong-Yan, Tang Huai; Gross Error Detection of GPS Observational Data Based on First Order Differential and Wavelet Analysis. Science Technology and Engineering, **13(27)**, 8206-8210 **(2013)**.

[9] Li Xi-Pan; Study on Process of GPS Dynaminc Deformation Data Based on Wavelet Analysis and Forecast Model, Master.s Thesis, Hebei University of Engineering, Hebei **(2009)**.