GROSS MEASUREMENT ERROR IDENTIFICATION USING THE GREY SYSTEM THEORY

Z. Wang [a], Y. Gao [b, *], S. Tse [b], P. Qin [c]

[a] Henan University of Science and Technology, Luoyang, 471039, China
[b] Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
[c] Xi'an Jiaotong University, Xi'an, 710049, China
* E-mail: meygao@ust.hk

Introduction

In a measurement process, measurement data may involve gross errors, which significantly exceed true values, when a measurement quantity is measured repeatedly without significant changes in measurement conditions. If the suspicious data remained during data processing, it will cause incorrect assessment on measurement accuracy due to the distorted measurement results have been used. Correct identification of gross measurement errors is an important issue to achieve reliable measurement results.

Gross errors may be reduced by using suitable measurement devices and under appropriate physical conditions. However, it would be quite difficult to avoid gross errors in measurement processes having large quantities of data.

Gross error identification has been based on statistics. A number of criteria have been used such as the 3σ criterion, the Chauvenet criterion, the Grubbs criterion, and the Dixon criterion [1]. The existing methods [2-10] have been based on a typical distribution such as the Gaussian distribution and require prior knowledge on the data that the measurement data conform to the Gauss distribution or conform approximately to it. In a practical measurement, it may be difficult to have a large quantity of data. Furthermore, the distribution may be found not conforming to the typical distribution. These may make the statistical methods [2-10] not applicable for gross error identification and removal.

In order to address the above issues to proceed with gross error identification and subsequently removal, a new method using the grey system theory is proposed. The advantages of the proposed new method is that the measurement data is not required to conform to a particular probability density distribution and the sampling size of the data does not need to be large. The principle of the gross error identification is presented and an identification criterion proposed. A case study is provided to demonstrate the effectiveness of the proposed grey system method. The proposed method should be a convenient and useful tool to identify gross errors in a precision measurement process.

Working Principle

The grey system method relies on the geometric features of the measurement data, instead of the probability density distribution and the sample size. In order to apply the grey system method, a measurement system is considered as a grey system and the measurement data is therefore a grey number within a particular interval [11].

To establish a complete grey area (Fig. 1), the distance between the data accumulation curve of the sorted data series, $x^{(0)}(k)$, k=1,2,…, n, and the reference line, noted as $\Delta(k)$, k=1,2,…n, is considered. In a measurement process, as the measurement proceeds, the number of measurement k would increase, and in the meantime the distance $\Delta(k)$ would increase, and would approach to the maximum distance Δ_{max}. After this point, $\Delta(k)$ would decrease and approaches zero at the last measurement point, where $k=n$.

To simplify the analysis, the medium point of the measurement number n is used as the turning point A to establish the 2-segment line, indicated as Line 2 in Fig. 1. The distance between the point A and the reference line would be h times of the maximum distance Δ_{max}. It can be seen that the 2-segment line (Line 2) is obtained by link up the origin of the coordinate, the turning point A and the end point of the data accumulation. The reference line (Line 3) and the 2-segment line (Line 2) would constitute the grey area that contains the data accumulation curve. It is considered that if there were no gross errors in the measurement data, the data accumulation curve would be restrained in the grey area, between Line 2 and Line 3. If the data accumulation curve cannot be restrained in the grey area, gross errors would exist. Therefore, if only random errors are involved, the data accumulation curve would be restrained within the grey area. Based on this, a criterion for the identification of gross errors can be established.

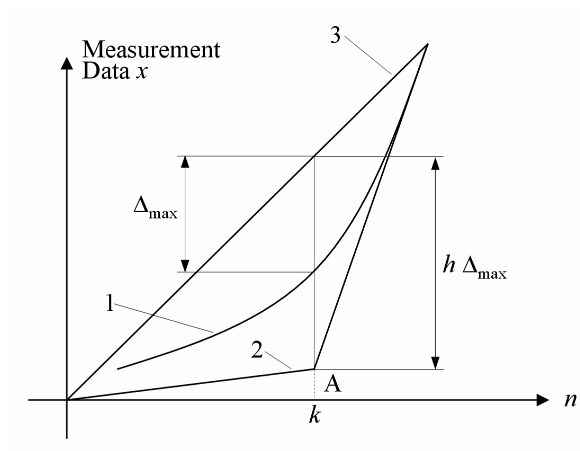

Fig. 1 Data processing using the grey system theory

Firstly, the original data series prior to the accumulation is sorted in an ascending order, noted as $x^{(0)}$. Therefore, $x^{(0)} = [\ x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\ ]$. Secondly, applying the 1-AGO algorithm to obtain an accumulated data series, noted as $x^{(1)}$ (Fig. 1). The turning point A of the 2-segment line is the pth measurement point and can be obtained as $p=n/2$ for an even n and $p=(n+1)/2$ for an odd n. The upper side of the grey area can be determined based on the reference line Line 3 (Fig. 1). The linear equation to describe the reference line is as

$$x^{(1)}_{\max}(k) = \frac{1}{n}x^{(1)}(n)k = [\frac{1}{n}\sum_{i=1}^{n} x^{(0)}(i)]k = \bar{x}k \quad , \tag{1}$$

where $k=1,2,\ldots,n,$ and $\bar{x}$ is the arithmetic mean of the measurement data. The lower side of the grey area can be determined based on the 2-segment line. The linear equation to describe the 2-segment line is as

$$x_{\min}^{(1)}(k)=\begin{cases}(\bar{x}-\frac{h\Delta_{\max}}{p})k, & 1\le k\le p\\ \bar{x}k-\frac{h\Delta_{\max}}{n-p}(n-k), & p<k\le n\end{cases} \quad (2)$$

Thirdly, the parameter *h* is used to define the lower side of the grey area. Through the analysis of many practical examples, *h*=3.75 seems appropriate. Finally, the criterion can be expressed mathematically as (a) if

$$x_{\min}^{(1)}(k)\le x^{(1)}(k)\le x_{\max}^{(1)}(k), \quad (3)$$

there would be no gross errors. This is due to the accumulated data points being within the grey area. Due to $x^{(1)}(i)<x^{(1)}(i+1)$, the above criterion could be simplified as $x^{(1)}(k)\le x_{\max}^{(1)}(k)$; (b) if

$$x^{(1)}(k)<x_{\min}^{(1)}(k), \quad (4)$$

there would be gross errors. This is due to the accumulated data points being outside the grey area (Fig. 1).

The data point identified to have gross errors should be removed and the remaining measurement data should be rearranged and re-assessed using the above criterion to identify and remove any suspicious data points. This process should be repeated until no gross error is identified in the data series.

Experimental Testing and Discussion

In the experimental testing of the errors of a measurement instrument, 20 repeated measurements were obtained [12]. The systematic errors were identified and compensated. The measurement data series, with the unit of mm, were obtained as

20.002, 20.000, 20.000, 20.001, 20.000, 19.998, 19.998, 20.000, 20.001, 19.998,
20.002, 20.002, 20.000, 20.004, 20.000, 20.002, 19.992, 19.998, 20.002, 19.998.

For the above measurement data, the turning point A of the lower side of the grey area was at p=10, the maximum distance $\Delta_{\max}$=0.0174mm. The arithmetic mean of the data $\bar{x}$=19.999mm. Therefore,

$$x_{\min}^{(1)}=\bar{x}-3.75\frac{\Delta_{\max}}{p}=19.999-3.75\frac{\Delta_{\max}}{7}=19.9934.$$

The first data point and the last one could have higher possibilities in having gross errors, after re-arranging the original sequence into an ascending order. For the first data point k=1, $x^{(0)}(1)$=19.992mm, and $x^{(1)}(1)$=19.992mm. $x^{(1)}(1)$=19.992mm was smaller than $x^{(1)}_{\min}$=19.9934mm. Therefore, it could be concluded that the data point was suspicious of having gross errors. For the last data point k=20, $x^{(0)}(20)$=20.004mm, $x^{(1)}(20\text{-}1)$=379.994, and

$$x_{\min}^{(1)} = (n-1)\bar{x} - 3.75\frac{\Delta_{\max}}{n-p}[n-(n-1)] = 379.9916\,.$$

It was greater than $x^{(1)}_{\min}$ =379.9916. Therefore, it was concluded that the last data point was not a suspicious one. Once the data point was identified to have gross errors, it should be removed. As a result, 19 data points remained in the data series. Using the proposed procedure, the remaining 19 data points were subject to the same identification process. It was found that no data point was suspicious of having gross errors.

If a statistics method is used, the data point 19.992mm was also treated as a suspicious one. This shows that the proposed method is more effective in the gross error identification. In general, the statistics methods [2-10] require that the data series conform to the normal distribution. The grey system method does not have such requirement. It is noted that the results from the use of the grey system criterion could be affected by the actual selection of the coefficient h. A systematic examination of the effect should be a topic of further investigation.

Conclusions

The grey system theory can be successfully applied in the identification of gross errors involved in a measurement process. In the proposed method, geometric aspects in the measurement data series plots are considered in the gross error identification criterion. The advantages of the new method are that the probability density distribution of the data series can be unknown and the sampling size small, thus suitable for both statistical and non-statistical measurement. The effectiveness of the proposed method is demonstrated in the experimental testing.

References

[1] Fei, Y., Error theory and data processing, China Mechanical Industry Press, 1995, Beijing.

[2] Bagajewicz, M. J., Gross error modeling and detection in plant linear dynamic reconciliation, Computers and Chemical Engineering, **22**(12), 1998, pp. 1789-809.

[3] Mangalvedekar, H. A., Gross error identification in current measurements, Electric Machines and Power Systems, **25**(8), 1997, pp. 907-16.

[4] Kim, I. W., Robust data reconciliation and gross error detection, the modified MIMT using NLP, Computers and Chemical Engineering, **21**(7), 1997, pp. 775-82.

[5] Sanchez, M., Gross measurement error detection/identification for an industrial ethylene reactor, Computers and Chemical Engineering, **20**, Pt. B, 1996, pp. S1559-64.

[6] Tong, H., Detecting persistent gross errors by sequential analysis of principal components, Computers and Chemical Engineering, **20**, Pt. A, 1996, pp. S733-8.

[7] Yang, Y., A study of gross error detection and data reconciliation in process industries, Computers and Chemical Engineering, **19**, 1995, pp. S217-22.

[8] Kaminski, W., The method of growing rigor for the adjustment of geodetic observations contaminated by gross errors, Manuscripta Geodaetica, **19**(2), 1994, pp. 55-61.

[9] Harikumar, P., A method to incorporate bounds in data reconciliation and gross error detection, Computers & Chemical Engineering, **17**(11), 1993, pp. 1121-8.

[10] Falcao, D. M., Linear programming state estimation: error analysis and gross error identification, IEEE Transactions on Power Systems, **3**(3), 1988, pp. 809-15.

[11] Lin, Y., and Wang, Z., Grey assessing the measurement uncertainty, Journal of Grey System, 2001, **14**(2), pp. 161-164.

[12] Zhu, J., Bin, H., and Wang, Z., Non-statistical distinguish of gross errors in measurement values, Journal of Huazhong University of Science and Technology, 2000, **28**(4), pp.17-19.