Error Analysis of Distribution Curve

ZHANG Xue-mei*, LI Dong, MA Qing-hua and HAO Jing-yuan

Energy & Chemical Engineering Research Center, Xi'an Siyuan University, No. 28, Shuiyan Road, Dongjiao District, Xi'an, Shannxi, China

E-mail: 476333285@qq.com

Abstract. 11 sets of microscopical data from one same briquette samples have been converted into the Pseudo Strength Index PSI curves, each curve is consisted of 37 points. PSI curves have been used to illustrate the procedures of analyzing and calculating the error. This study shown that even in each PSI curve, there always are same points with the point error greater than 3s, that it is, however, impossible to determine which curve is an “abnormal curve” based on the characteristic (number, size and location) of abnormal points. Because Euclid distance converts the error of a curve into the error of a point, it must first calculate the Euclid distance for the series sets of PSI curves. All 11 Euclid distances is less than the distance limit error of 3s. Therefore, there are no abnormal curves in 11 PSI curves evaluated.

1. Introduction

In order to select scientific measurement methods to study material characteristics, one must be able to calculate the error generated by the experimental process and to cull the data of large errors. There is a detailed mathematical formula, such as (arithmetic mean, average error, standard error, limit error, outlier, etc.) to deal with the point error, i.e. a same physical quantity in the same situations measured repeatedly n times [1, 2]. But in many cases, the objective needed to be treated is not a value, but a curve or even a surface. How to carry out the similar error analysis? For example, based on the national standard of the People's Republic of China “method of determining microscopically the reflectance of vitrinite in coal”, GB/T6948-2008[3], a reflectance curve of vitrinite in coal could be obtained. As for the relationship between the best active inert ratio graph and coal's metamorphic degree, Amosov-Shapiro Theory [4, 5] converts the reflectance curve of vitrinite in coal into the strength index SI. Some researchers have introduced the Pseudo Strength Index PSI. In this paper, the error calculation and analysis of the distribution curve is discussed.

2. Point error calculation

At the same conditions, one physical property has been measured n times with the results of $x_1$, $x_2$, $x_3$, …… $x_n$, the arithmetic mean, error, average absolute error, and mean square root error $\sigma$ could be calculated. According to Gaussian error theory, the probability of error in any of the measurements in the range of $3\sigma$ is 99.7%. $3\sigma$ is generally called the limit error. If there is an error of a measurement is more than $3\sigma$ in these repeat n times measurements, then this measurement must be considered as an outlier and be, therefore, culled.
2.1. The specular group reflectivity measurement of coal

The reflectance of vitrinite in coal starts at 0.3%, increasing by 0.05%, such as 0.3%, 0.35%, .... up to 2.1%. A total of 37 elements from start to finish can be output from a reflectance of vitrinite in coal. The strength index SI is defined as:

$$SI = \frac{a_1x_1 + a_2x_2 + ... + a_{37}x_{37}}{\sum_{i=1}^{37} x_i}$$

(1)

In formula, the value of the underlying does represent a certain reflectivity, such as $x_1$ is the content of the active component at 0.3% of vitrinite reflectivity, $x_2$ is the content of the active component at 0.35% of vitrinite reflectivity, ....; $x_{37}$ is the content of the active component at 2.1% of the end vitrinite reflectivity, $a_i$ is the strength index of various vitrinite [6].

2.2. Pseudo strength index distribution curve and error calculation

Pseudo Strength Index Distribution PSI [7-11] has been developed in order to solve the problem of how to mix most suitable coking coals. PSI is a curve containing 37-points, not SI, a single value:

$$PSI_l = \frac{a_1x_1}{\sum_{i=1}^{37} x_i}$$

(2)

The line is made up of points, so you can calculate the error of each point on the line. However, you still cannot determine the abnormal curve by the amount and/or size of the anomaly point (the error exceeds 3σ) of the curve.

The Minkovsky distance is a standard pattern of date sets to measure the similarity between two sets [12,13]. Euclidean Distance is a special form of the Minkovsky Distance:

$$dist(X, \bar{X}) = \left( \sum_{i=1}^{37} (x_i - \bar{x_i})^2 \right)^{1/2} = d$$

(3)

$X$ is a measurement of a data set, $\bar{X}$ is the arithmetic average data set of repeatedly measures $m$ times. All 37 points of the arithmetic average curve are calculated in order by the arithmetic average of the points, as shown in the following equation:

$$\bar{X} = \{ \bar{x_1}, \bar{x_2}, ..., \bar{x_{37}} \}$$

(4)

Using Euclid distance to make the similarity of the line into a common “error calculation and determination of the value of the repeated determinatio”.

3. Results and discussions

A briquette coal product from northern Shaanxi province, has been shattered and processed into total 11 samples. According to the national standard “method of determining microscopically the reflectance of vitrinite in coal”, using the MSP9000C to do the coal petrographic analysis, 11 reflectivity test results were reported. According to Equation 2, 11 Pseudo Strength Index Distribution PSI have been obtained, one of them is presented as Figure 1. In general, the larger m, the closer the arithmetic average curve $\bar{X}$ is to the real curve. But it cannot be answered from the 11 PSI distribution curves, without further mathematic analysis, regarding:

1) Are there any suspicious outlier curves amount those 11 PSI distribution curves?
2) If yes, which one is the suspicious outlier curve? And why?
3) If not, how close is the arithmetic average curve $\bar{X}$ to the real curve? And why?
3.1. Point error calculation
Because each PSI curve has 37 points, the point error calculation has been carried out. The results show that in each PSI curve, there always are same points with the error over 3σ, which are defined as “abnormal points”. However, the abnormal points of each of PSI curves are different in number, size and location. It cannot determine the suspect outlier curve based on the characteristic (number, size and location) of abnormal points.

3.2. Curve error calculation
The Euclid distance of all 11 PSI curves can be calculated by eq. 3. With 11 Euclidean distances, the Euclid distance error calculation has been carried out. Table 1 lists the test serial number, Euclid distance, the absolute value of the distance error, and the distance limit error, respectively.

| PSI test # | Euclid Distance d | Distance Error Absolute | Distance limit error 3σ |
|------------|-------------------|-------------------------|-------------------------|
| 1          | 32.033            | 28.137                  | 87.337                  |
| 2          | 33.057            | 27.113                  | 87.337                  |
| 3          | 41.589            | 18.581                  | 87.337                  |
| 4          | 45.730            | 14.440                  | 87.337                  |
| 5          | 43.170            | 17.000                  | 87.337                  |
| 6          | 58.715            | 21.455                  | 87.337                  |
| 7          | 68.716            | 8.546                   | 87.337                  |
| 8          | 129.494           | 69.324                  | 87.337                  |
| 9          | 67.512            | 7.342                   | 87.337                  |
| 10         | 62.238            | 2.068                   | 87.337                  |
| 11         | 99.614            | 39.444                  | 87.337                  |

Based on the results of table 1, it is easy to determine:
1) The smaller the distance average error, the greater the reliability of the measurement.

![Figure 1. One of the 11 PSI distribution curves](image)
2) The relative size of the distance limit error of $3\sigma$ can be used to determine the abnormal curve. From Table 1, the absolute value of the error of all 11 Euclid distances is less than the distance limit error of $3\sigma$.

3) Therefore, there are no abnormal curves in all 11 PSI curves.

4. Conclusions
When the series sets of data to be processed is not a value but a curve, such as the PSI curve, it is impossible to determine which curve is an “abnormal” based on the characteristic (number, size and location) of abnormal points.

Instead of it must first calculate the Euclid distance for the series sets of curves, such as in the PSI curves. The Euclid distance turns the error calculation of curves into the point error calculation.

Acknowledgments
I am very grateful to Professor Li Dong, the director of our department, for his great help in my work. At the same time, I also thank my colleagues Ma Qinghua and Hao Jingyuan very much. In the process of the experiment, two colleagues took the initiative to help me successfully complete the experiment.

5. References
[1] Du Y C, Sun Y and Wang B, The important influence of error in the experiment 2019 J. Guangdong Cheml Indus. 46 157-158.
[2] Liu D H, Study on detection and error of electronic components 2019 J. Applications of IC. 36 70-71.
[3] The national standard of the People's Republic of China “method of determining microscopically the reflectance of vitrinite in coal”, GB/T6948-2008.
[4] Shapiro N, Gray R J and G R, Recent development in coal petrography: Blast furnace, coke oven and raw materials 1961 J. Iron and Steel Division. 20 89-112.
[5] Shapiro N and Gray R J, The use of coal petrography in coke making 1964 J. Journal of institute of fuel. 37 234-242.
[6] Chen Q R 2009 Clean coal in China (Xu Zhou: China University of mining and Technology Press) p 1038.
[7] Zhang X M, Li D and Xia H, Evaluation on blended coal for coking by using fuzzy sets measurement 2014 J. Shanxi Sci and Technol. 4 67-68.
[8] Zhang X M, Liu X G,Wang M H et al, Study on the prediction of coke performance by fuzzy set coal blending method 2018 J. Goal Quality Technol. 6 15-18.
[9] Li D and Hao J Y, Prediction of coke properties based on fuzzy set coal blending 2017 J. Coal Sci and Technol. 45 213-217.
[10] Xia H, Li D and Zhang X M, Coal blending calculation with fuzzy set measurement method 2014 J. Shanxi Sci and Technol. 29 66-68.
[11] Li D, Li H R and Mu Q Y, The feasibility of coking coal blending evaluation with the fuzzy set measurement method 2014 J. Shanxi Sci and Technol. 29 57-59.
[12] Zhang G L, Zhang H and Kong Q 2011 Foundation and application of fuzzy mathematics(Bei Jing: Chemical Industry Press) p 22-25.
[13] Jiang Z J 2003 Fuzzy mathematics course(Bei Jing:National Defense Industry Press )p 18-23.