*Applied Mechanics and Materials* *Online: 2009-10-12*
*ISSN: 1662-7482, Vols. 16-19, pp 347-351*
*doi:10.4028/www.scientific.net/AMM.16-19.347*


# On the Flatness Uncertainty Estimation Based on Data Elimination

Meifa Huang[1,a], Qiaoyi Wang[1,b], Yanru Zhong[2,c],
Bing Kuang[1] and Xiangqian Li[1]

[1]School of Electromechanical Engineering, Guilin University of Electronic Technology, Guilin, P.R. China, 541004

[2]School of Computer and Control Engineering, Guilin University of Electronic Technology, Guilin, P.R. China, 541004

[a]hmhmf@guet.edu.cn, [b]wangqiaoyi@mails.guet.edu.cn, [c]rosezhong@guet.edu.cn

**Keywords:** Uncertainty of flatness, Data elimination, Least-square methods, Statistics

**Abstract.** Coordinate measuring machine (CMM) plays a more and more important role in the flatness evaluation. The pre-processing for eliminating error data is an important process to improve the quality of flatness evaluation. So, flatness uncertainty estimation becomes a key problem. To solve the problem, a flatness uncertainty estimation method based on data elimination is presented. Firstly, a method for excluding the error data based on statistical theory is expatiated with the consideration of probability theory and mathematical statistics. Secondly, details of calculating flatness uncertainty based on least-square method are analyzed. Then the measured 3-dimensional data are processed so as to put the excluding method into use. So the error data can be deleted and the flatness uncertainty can be decreased and can achieve a good evaluation quality. Finally, an example is given and the result shows that the method proposed in this paper is feasible.

## Introduction

The reliability and validity of the measured data have always been the bottleneck to our applications in large number of data processing. At present, CMM is the important instrument in flatness evaluation. Though CMM has a higher accuracy, some CMM lack the mature software to do the pre-job of the measured data. So on one hand, gross error can not be timely detected and deleted by the surveyors; On the other hand, verification requirements and remedial implement are also difficult to be put into effect [1-3].

The uncertainty theory is built based on the further error theory, and it indicates the changeable degree of uncertain error. That is to say the calculated error has been allowed in a certain scope, and uncertainty is used to describe the more subtle part of the error. Because gross error is out of the available evaluation range, it can be seen as the wrong data. In order to decrease the uncertainty, it is necessary to do the pre-job for eliminating error data. Excluding the gross error not only an effective way to get a available evaluating result, but also an essential method to realize the uncertainty management of the new generation of geometrical product specification and verification (GPS). Because the new GPS is established on metrology and most information is expressed in way of numerical form [1, 3, 4], so it needs a high precision to ensure the correctness of result.

In the test of the workpiece, flatness uncertainty has always been closely watched and many methods were adopted to evaluate it. Among the methods minimum zone method has a higher precision but has a higher requirement. In literature [1, 2], there are many examples about how to use it. To simplify the process, least-square method is adopted to calculate the flatness in literature [1, 4, 5]. Though the results above have been verified, the gross error was not considered in the process of the evaluation. So in complicated conditions, the result maybe not exact. To resolve the problem, a method considering the gross error based on least-square methods is proposed.

**Rules for Distinguishing and Eliminating the Gross Error**

There are many rules to eliminate the gross error [1, 2, 6, 7], such as 3$\delta$ criteria, T criteria and Grubbs criteria, etc. 3$\delta$ criteria has a higher requirement of measuring times while T guideline needs to establish a complex mathematical model. However Grubbs criteria has a low request to the measuring times and the model, and it can achieve a high accuracy in actual applications. From the actual experience of the measurement, we know that most measured data is subject to some kind of distribution. So the error data can be excluded by utilizing statistical criterion of the residuals. The basic idea of removing suspicious data is that a significant level '*a'* (or confidence level '1-*a*') is set at first, and one confidence limit should also be given. When confidence limit is higher than the limits of the error, it can be considered out of the scope of random error and should be removed. If measured data rows are $X_1, X_2, X_3, \cdots, X$n, then the criteria standard deviation can be expressed as:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}v_i^{\,2}} \quad . \tag{1}$$

Where $\bar{x} = \frac{1}{n}\sum_i x_i$ is the mean value of measured data. With the relevant knowledge in statistics, we know that the absolute value of reasonable data deviations should not exceed a multiple of *k*, so when $|v_i|> ks$, the data should be removed. In order to avoid the leakage and misuse of the deleting method, only the data having the biggest residual value should be removed. Meanwhile the rest data must be dealt with once again until there is no suspicious. In inequation $|v_i|> ks$, *k* may have different values and there are different criteria. Considering the new GPS standards, the functional requirements can play an important part in the contribution to get the value of *k*. At the same time in order to simple the eliminating process, the formula can be expressed as follows:

$$|\, v_i / s\,|_{\max} > k. \tag{2}$$

Here we have independent samples: $x_1, x_2, \cdots x_n$ which obey normal distribution. In order to use Grubbs criteria, a new statistics variable $g = (x_d - \bar{x})/s$ is constructed. The significance level $\alpha$ is usually selected 0.05 or 0.01, and then $G(\alpha, n)$ can be obtained by equation (3) [6, 7]:

$$P(|\, g\,| \geq G(\alpha, n)) = \alpha. \tag{3}$$

So the criterion is given: If $|\, x_d - \bar{x}\,| \geq G(\alpha, n)s$, data $x_d$ contain gross errors and should be removed; otherwise reserved. In our actual use the value of *k* in formula (2) is the substitution of $G(\alpha,n)$ and can be obtained according to Grubbs theory and the distribution rule of $G(\alpha,n)$. In accordance with the threshold criteria and use a consistent formula, *k* is gained from the following equation (4)[6, 7]:

$$k = \begin{cases} \ln(\mathrm{n} - 2.65)/2.31 + 1.305 + \rho & \mathrm{n} < 30 \\ \ln(\mathrm{n} - 3)/2.301 + 1.36 - \mathrm{n}/550 + \rho & \mathrm{n} \geq 30. \end{cases} \tag{4}$$

Where $\rho$ stands for the adjustment coefficient, and it can be calculated based on new GPS and functional requirements of the workpiece [4]. For example, the straightness of a machine tool (CA6410) guide surface in vertical direction is less than 0.018/1000 and only allowed to bump. So we can say that the guide surface can has a tiny convexity, so $\rho$ can be calculated according to its functional requirements and the correlative theory in new GPS. That is to say $\rho$ can express the partial information of the surface in a certain extent to make the evaluation more scientific and ensure the verification more scientific [1, 4, 9].

**Uncertainty Calculation of Flatness Based on Least-Squares Method [4, 8, 9, 10]**

Among the methods to estimate the flatness uncertainty, smallest area method is commonly used. This method only requests the square of error being a minimize value. Thus a balance among the equations can be established so as to prevent a certain extreme error obtaining dominant status. Here it is adopted to work with the elimination method to evaluate the flatness uncertainty.

Generally speaking, the measurement results $Y$ is a function of variables $X$:

$$Y = G(X_1, X_2 \cdots, X_i \cdots, X_{p+r}). \quad (5)$$

Where $p$ stands for the non-related amount of the variable with each other (coefficient $\rho$=0), and $r$ is the number of variables which have a strong correlation ($\rho$=1 or -1). In that case, the uncertainty of measurement results can be expressed as shown in equation (6):

$$u_c = \sqrt{u_r^2 + \sum_{i=1}^{p} (\frac{\partial Y}{\partial X_i} \times u_{Xi})^2}. \quad (6)$$

In order to get coefficient, the plane equation can be expressed as $z = ax+by+c$, Assuming that $(x_1,y_1,z_1)$ and $(x_2,y_2,z_2)$ are the two sampling points which have the largest and the smallest distance to the least-squares plane, then $\delta$ can be expressed as equation(7):

$$\delta = \frac{(z_1 - z_2) - a(x_1 - x_2) - b(y_1 - y_2)}{\sqrt{1 + a^2 + b^2}}. \quad (7)$$

From the actual measurement experience of CMM, only '$a$' and '$b$' are relevant. With the consideration of ISO / TS 14253-2, we can treat $\rho$ 1 or -1 in a conservative way. Here they are related enthusiastically, so $\rho$=1, and then $u_\delta^2$ is given in equation (8) [10]:

$$u_\delta^2 = (\frac{\partial \delta}{\partial x_1} u_{x_1})^2 + (\frac{\partial \delta}{\partial x_2} u_{x_2})^2 + (\frac{\partial \delta}{\partial y_1} u_{y_1})^2 + (\frac{\partial \delta}{\partial y_2} u_{y_2})^2 + (\frac{\partial \delta}{\partial z_1} u_{z_1})^2 + (\frac{\partial \delta}{\partial z_2} u_{z_2})^2 + (\frac{\partial \delta}{\partial a} u_a)^2 + (\frac{\partial \delta}{\partial b} u_b)^2 + 2\frac{\partial \delta}{\partial a}\frac{\partial \delta}{\partial b} \rho_{ab} u_{ab}. \quad (8)$$

Referring to mathematical statistics, we treat $a, b, c$ as a random vector to calculate $u_a$, $u_b$ and $u_{ab}$. Through times' fitting, the covariance matrix (9) can be got. So $u_a$, $u_b$, $u_{ab}$ and $u_\delta$ can be calculated.

$$Cov(r) = \begin{pmatrix} u_a^2 & u_{ab} & u_{ac} \\ u_{ba} & u_b^2 & u_{bc} \\ u_{ca} & u_{cb} & u_c^2 \end{pmatrix}. \quad (9)$$

**Process of Identifying and Removing Abnormal Data**

In order to facilitate the processing, we save data in file format. With the help of matlab7.0, the measured data are saved in file and stored with the '.txt'. The data $(x,y,z)$ measured by CMM have three parameters but there is always only one parameter in actual assessment, so the 3-dimension data should be turned into 1-dimension. To solve this problem, coordinates should be converted into distance between the point and the flat. Because $(x,y,z)$ obeys the normal distribution, so from the formula $d_i = \frac{z_i - ax_i - by_i - c}{\sqrt{1 + a^2 + b^2}}$, we can see that $d_i$ is the linear function about $(x,y,z)$. According to probability theory and mathematical statistics, in the process of transforming, the character of the value has not changed. That is to say $d_i$ also obeys the normal distribution. With powerful processing ability of computers, gross error can be eliminated quickly and reliably. So in one course of eliminating, no more than one outlier is deleted. Thus by way of deleting outlier value one time by one time, we can avoid skipping over or mistaken deleting as far as possibly. So the reasonable data got kept to a large extent and the data processing was effective. When the distance value $d[i]$ meets Grubbs criteria, this value is recorded as well as its related information, including sequence number and its location. From distance information the coordinate's information can be traced, and the eliminating work is completed. The remaining value can be automatically updated and get ready for the new kicking circle. After eliminating gross error, a backup storage document should be made to accomplish the processing job effectively.

### Verification Experiment

By using one CMM modeled MCMS-086, we measure a flat with a specification 260mm * 250mm. Assuming the plane is $z = ax+by+c$, and 60 points are adopted, which were divided into three groups. Each group has 20 data points. Due to the space, only one set of data was listed in table 1.

Table 1 Coordinate Data Values

| Item | *x* [mm] | *y* [mm] | *z* [mm] | Item | *x* [mm] | *y* [mm] | *z* [mm] |
|---|---|---|---|---|---|---|---|
| 1 | 12.999 | 10.100 | 2.114 | 11 | 12.999 | 136.100 | 2.114 |
| 2 | 66.999 | 10.099 | 2.115 | 12 | 67.000 | 136.099 | 2.112 |
| 3 | 120.999 | 10.100 | 2.107 | 13 | 120.999 | 136.100 | 2.117 |
| 4 | 174.999 | 10.099 | 2.107 | 14 | 175.001 | 136.099 | 2.107 |
| 5 | 228.999 | 10.100 | 2.107 | 15 | 228.999 | 136.100 | 2.112 |
| 6 | 229.001 | 73.099 | 2.110 | 16 | 229.002 | 199.099 | 2.106 |
| 7 | 175.002 | 73.099 | 2.116 | 17 | 175.001 | 199.099 | 2.122 |
| 8 | 121.001 | 73.099 | 2.113 | 18 | 121.000 | 199.099 | 2.123 |
| 9 | 67.002 | 73.099 | 2.117 | 19 | 67.001 | 199.096 | 2.117 |
| 10 | 13.001 | 73.099 | 2.113 | 20 | 19.002 | 199.089 | 2.110 |

Document is stored as table1.txt in matlab7.0 and by using [*c_x*1*,c_y*1*,c_z*1] = textread ('table1.txt', '% f% f% f') statement, we can read the contents of the document and sent these data to the default unit in matlab. Take the data in table 1 as the example, and then the distance value that the points from the least-squares fitting plane can be calculated and shown in table 2.

Table 2 Distance between Measured Data Points and Fitting Plane [mm]

| | | | | |
|---|---|---|---|---|
| 0.0006862 | 0.0011755 | 0.0000372 | 0.0024344 | 0.0009060 |
| 0.0007026 | 0.0015075 | 0.0020209 | 0.0002160 | 0.0012362 |
| 0.0023736 | 0.0015687 | 0.0010970 | 0.0006254 | 0.0012916 |
| 0.0025899 | 0.0008817 | 0.0006866 | 0.0001363 | 0.0041940 |

The distance values are also stored and marked $d[i]$. Setting $d_i = d[i]$, then the new statistics $g = (d_i - \bar{x})/s$ is constructed. Setting $\alpha = 0.01$, $G(\alpha,n)$ will be obtained by the Eq.3.

Because $n$<3, so $k = \ln(n-2.65)/2.31+1.305+\rho$. As a result of no-special functional requirements about the flat, $\rho$=0. That is to say, $k = \ln(n-2.65)/2.31+1.305$.

In the comparison, 0.0041940 was found the inordinate value which meets the deleting conditions. So we keep a record about it, and track it. Then we get data (11.002,198.999,1.020), which is the gross error data point, so we note it down and remove this point.

Through the analysis of least squares fitting method above, we know that the data were divided into three groups. In that case the correlation coefficient can be got by using the covariance matrix. By way of utilizing each set of data, appropriate parameter $a_1$, $b_1$, $c_1$, $a_2$, $b_2$, $c_2$, $a_3$, $b_3$, $c_3$, can be calculated separately. Then through the covariance matrix, we can get $u_a$, $u_b$ and $u_{ab}$.

Table 3 Comparison of Parameters before and after Eliminating

| Item | *a* | *b* | *c* | Flatness[mm] | Uncertainty[mm] |
|---|---|---|---|---|---|
| Before | -0.00009258 | 0.00002317 | 1.0231065 | 0.017026 | 0.005213 |
| After | -0.00009785 | 0.00002222 | 1.0226543 | 0.016835 | 0.004928 |

After obtained correlation coefficient respectively, we put the three sets of data together, then using the data combination, we can count *a, b* and *c*. So the flatness and uncertainty can be calculated by Eq.7 and Eq.8. Table 3 shows the parameters before and after deleting the gross error. From which we can know that after the elimination, the flatness and the uncertainty get improved.

In order to validate the correctness, stability and practicality of this method further more, another CMM of Croma6106 which has a higher precision is adopted. With the help of this CMM, we measured the same plan for three times. The flatness for the three measurements is 0.016908 mm, 0.016824 mm and 0.016833 mm. The uncertainty is 0.0049016 mm, 0.0048926 mm and 0.0048913 mm. As can be seen from the comparison, the data obtained after the eliminating can meet the requirements of precision and the result is accurate. What is more the method did not introduce some disadvantage to our evaluation. That is to say this method has certain stability and robusticity.

## Conclusions

Nowadays, in the evaluation of the flatness uncertainty, the gross error was rarely taken into consideration. In this paper a new scheme integrating the eliminating method and least square method is proposed to get rid of the gross error and calculate the uncertainty. In this process matlab 7.0 was used to realize the details of data processing. The analysis and the comparison results obtained in different way show that the scheme proposed in this paper is feasible. Of course the method also has application prospect, because it does not introduce disadvantage to our evaluation and it is easy to be integrated into CMM.

## Acknowledgements

This research is supported by the Natural Science Foundation of China (Grant No. 50865003), the Guangxi Natural Science Foundation of China (Grant No. 0640166) and the Ten, Hundred and Thousand Talents Project of New Century in Guangxi to Prof. Meifa Huang.

## References

[1] Z.Y. Wang, Z.M. Liu, X.T. Xia and L.Q. Zhu: *Measurement Error and Uncertainty Evaluation* (Science Press, China 2008) (In Chinese).

[2] G.X. Zhang: *Coordinate Measuring Machines* (Tianjin University Press, China 1999) (In Chinese).

[3] W. Choi, T.R. Kurfess and J. Cagan: Precision Engineering, Vol. 22 (1998), pp.153-163.

[4] X.Q. Jiang: *Theory and Applications of New-Generation Geometrical Product Specifications* (Higher education press, China 2007) (In Chinese).

[5] T. Weber, S. Motavalli, B. Fallahi and S.H. Cheraghi: Precision Engineering, Vol. 26 (2002): pp.269-278.

[6] Z. Sheng, S.Q. Xie and C.Y. Pan: *Probability Theory and Mathematical Statistics* (Higher Education Press, China 1989) (In Chinese).

[7] J.W. Arenberg: *Proceedings of ASPE Summer Topical Meeting on Uncertainty Analysis in Measurement and Design*. (Pennsylvania, USA 2004), Vol. 1, pp.40-45.

[8] Y.Q. Ni, ZH. Sun, M.Q. Li and M. Miao: Microcomputer Information. Vol.17 (2006), pp.178-180. (In Chinese)

[9] J. Yu: Journal of Test and Measurement. 2003, Vol. 17 (2003), pp.258-261. (In Chinese)

[10] ISO 14253-1 (1998): ISO, Geneva.