Dynamic Error Analysis of CMM Based on Variance Analysis and Improved PLSR

Zhang Mei*[1], Cheng Fang[2], Li Guihua[3]
[1,3]College of Electrical Engineering and Automation, Anhui University, Hefei, 230601, China
[2]School of Mechanical & Aerospace Engineering, Nanyang Technological University, 639798, Singapore
*Corresponding author, e-mail: hfren@126.com[1], chengf@artc.a-star.edu.sg[2], guihuali1@sina.com[3]

Abstract

It is difficult to build an accurate model to predict the dynamic error of CMM by analyzing error sources. An innovative modeling method based on Variance Analysis and Improved Partial Least-square regression (IPLSR) is proposed to avoid analyzing the interaction of error sources and to overcome the multi-collinearity of Ordinary Least-square regression (OLSR). Among many impact factors the most influential parameters are selected as the independents of the model, by means of variance analysis. The proposed modeling method IPLSR can not only avoid the analysis of the error sources and the interactions, but can also solve the problem of multi-collinearity in OLSR. From experimental data the expository capability of this IPLSR model can be calculated as 85.624 percent, and the mean square error is 0.94µm. As comparison, the mean square values of conventional PLSR and OLSR are 1.04µm and 1.39µm, respectively. So IPLSR has higher predicting precision and better expository capability.

Keywords: dynamic error, partial least-squares regression (PLSR), variance analysis, multi-collinearity

1. Introduction

In the field of dynamic error modeling for Coordinate Measuring Machine (CMM), many efforts have been made on the analysis of error sources [1-3]. It has notable theoretical significance especially for designing a measuring machine, but for practical measurement its application is limited because the modeling accuracy is notably influenced by the interactions between different error sources. Weekers and Schellekens [4] used 8 position sensors to detect the deformation and acceleration changing of 8 main articulating points on a bridge type CMM. With these parameters a dynamic error model of the probe was established. Dong [5] directly measured the angular error of the main connecting mechanism during the movement to model the probing errors. Zhang Yi and Liu Jizhu[6], Wei Jinwen and Chen Yanling [7] focused on the deformation of the crossbeam. By analyzing the crossbeam deformation with acceleration, constant speed and deceleration with ANSYS, a common model of beam deformation was established under any load. But this method is only useful for error compensation in one direction.

It's known that the influence of each error source will be finally reflected in the measured values, namely, (x, y, z) coordinates. On the other hand, the Direct Computer Control（DCC）parameters, which are effective during the whole process of measurement, are easy for controlling and sampling. So if these parameters are seen as the independent variables, the complicated source analysis for dynamic errors can be avoided. So it's a novel research to study the influential power of each independent variable as well as the elimination of the multi-collinearity in between [8-9].

Partial Least-squares (PLS) Regression [10-13] has attracted many researchers' interests these years. Recently some improved PLS algorithms [14-15] are proposed and have some successful application in different fields. Similar with PCR (Principal Component Regression), PLS is also effective for reducing the dimensions and eliminating the multi-collinearity. The method of PCR, however, has not acceptable fitting for dependents, because it only concentrates on the principle components of the independents, but it's irrelative with the dependents. As comparison, PLS starts from the dependents to find a linear combination of the independents which have the most influential powers. Therefore it has better predictive capability than PCR. Besides, In the cases that the sample size is smaller than the quantity of

variables, it still has acceptable expository capability [16-19]. Yang Hongtao and Liu Yong[8] used the hybrid modeling method about PLS regression and the support vector machine, and Zhang Mei[9] applied 3B spline-orthogonal projection PLS regression model to analyze CMM dynamic error. The results show that the PLS can be used to build the high-precision model of CMM's dynamic errors. The authors' group has employed improved PCR in the field of CMM dynamic errors. By analyzing the relationship between (x, y, z) coordinates and the DCC parameters, an accurate model with clear physical meanings can be established.

2. Research Method

Assuming there is a single dependent Y, a set of independents $x_1 \cdots x_p$, and n sample points are acquired. With the n-dimensional dependent vector and the p-dimensional independent vector, an n×p observing matrix can be configured as: $X = [x_1 \cdots x_p]_{n \times p}$. Then the PLS algorithm[10] can be described as follow.

In the observing matrix X a component t_1, a linear combination of $(x_1 \cdots x_p)$, is extracted, which should to the largest extent include the mutation information and has most correlation with Y [11]. So t_1 includes most information of X and has a good expository capability for Y. Then PLS regression of X on t_1 and that of Y on t_1 can be worked out, respectively. If the regression equation has reached the required accuracy the operation stops; otherwise the residual information in X should be extracted for the next operation. This iterative process should be repeated until the required accuracy is achieved. Finally if k components are extracted from X: $t_1 \ldots t_k$, the regressive operations of Y on $t_1 \ldots t_k$ should be done. Then the regression model can be expressed in form of $Y = f(x_1 \cdots x_p)$.

2.1. Modeling Process

According to the references [10-11], the modeling process can be summarized as below:

Step 1: Standardization

The observing matrix X is standardized as $E_0 = (E_{01} \cdots E_{0p})_{n \times p}$; The single dependent vector Y is standardized as $F_0 = (F_{01})_{n \times 1}$.

Step 2: Components extraction.

k components can be extracted as Equation (1):

$$t_k = E_{k-1} w_k \tag{1}$$

Where $w_k = \dfrac{E_k F_0}{\|E_{k-1} F_0\|}, E_k = E_{k-1} - t_k p_k, p_k = \dfrac{E_{k-1}^T t_k}{\|t_k\|^2}$, and, $E_1 \cdots E_k$ are the residual error matrixes after the standardization of independents.

For the kth components, the coefficients of the fitting equation can be determined by iterative operation, expressed by:

$$r_k = \frac{F_{k-1}^T t_k}{\|t_k\|^2} \tag{2}$$

Where $F_k = F_{k-1} - t_k r_k, F_1 \cdots F_k$ are the residual error vectors after the standardization of the dependent.

Then the kth fitting equation can be expressed as:

$$\hat{y}* = r_1 t_1 + r_2 t_2 + \cdots + r_k t_k \tag{3}$$

Step 3: Number of components.

The best compromise should be made to determine the number of components. The extracted components should have enough expository capability to the system, but the modeling reliability cannot be decreased by over fitting. Currently the method called Cross Validation (CV) is widely used to determine the number of components for LS. The process of CV can be described as follow:

The ith sample is removed from the sample data set. With the rest sample data, a regression equation is worked out on the k PLS components. When the ith sample is taken into this regression equation, the fit value of the ith sample can be worked out, namely, $\hat{y}_{k(-i)}$. For any i = 1, 2, …, n , the above operation is repeated. Then the PRESS (Prediction Residual Error Sum of Squares) can be calculated:

$$PRESS(k) = \sum_{i=1}^{n}(y_i - \hat{y}_{k(-i)})^2 \tag{4}$$

With all the sample data, another regression equation on k components can be derived. Assuming $\hat{y}_{ki}$ is the calculated from all these sample data, with the same operation in the above paragraph, the sum of squares can be defined as:

$$SS(k) = \sum_{i=1}^{n}(y_i - \hat{y}_{ki})^2 \tag{5}$$

CV can be defined as:

$$Q_k^2 = 1 - \frac{PRESS(k)}{SS(k-1)} \tag{6}$$

Only if $Q^2{}_k \geq 0.0975$, the model quality can be improved by increasing the number of PLS components, t_k .

Step 4: This regression equation with the optimized components is then deducted to that with original variables to analyze the direction and extent, to which the dependents are influenced by the independents [10-14].

2.2. Improvement of PLS Model (IPLS)

In the process of PLS regression, the fundamental principle of extracting t_k is to make the covariance of the dependent Y to acquire the maximal value. The covariance of Y can be defined as $Cov(t_k, Y)$. Normally t_k has the best expository capability to the dependents. But there's still a problem: It's seen from equation (7) that a big value of $Cov(t_k, Y)$ will not necessarily result in a big value of $\rho(t_k, Y)$, the correlation coefficient of t_k and Y. In some cases, therefore, a big value of $Var(t_k)$, the variance of t_k , may cause wrong selection of the component t_k .

$$Cov^2(t_k, Y) = \rho^2(t_k, Y)Var(t_k)Var(Y) \tag{7}$$

To solve this problem, Cheng and Wu [15] proposed an improved algorithm of PLS: Firstly the orthogonal matrix of Y is worked out and named as B, which is composed by the eigenvectors $b_1, \cdots b_{p-1}$ corresponding to the zero eigenvalues of X^TYY^TX . Then the eigenvalues and eigenvectors of B^TX^TXB are calculated and named as $\lambda_1, \cdots \lambda_{p-1}$ and $\alpha_1, \cdots \alpha_{p-1}$, respectively. Among the eigenvalues the largest s values $\lambda_1, \cdots \lambda_s$ are extracted and the

corresponding eigenvectors are selected to form a matrix *A*. The determination of *s* should make the value of $\sum_{i=1}^{s}\lambda_i \Big/ \sum_{i=1}^{p-1}\lambda_i$ close to 99%.

A new orthogonal matrix of *Y* is configured as $U = XBA$. So the projection of *X* in the direction that is orthogonal to *U* can be expressed by:

$$\left(I_p - P_U\right)X = X - U\left(U^TU\right)^{-1}U^TX = X\left(I_p - BA\left(U^TU\right)^{-1}U^TX\right) \equiv XD \tag{8}$$

Where I_P is an identity matrix, P_UX is a projection which is the *X* on the *U*. The process of projection helps to eliminate the information with unobvious relativity of *Y*. The operation of PLS with *XD* shows the improvement of conventional PLS modeling, which is named as IPLS in this paper:

$$Y_{IPLS} = XD\beta_{PLS} = X\beta_{IPLS} \tag{9}$$

This IPLS model has the best expository capability to the dependents and can improve the predicting accuracy.

2.3. Acquisition of Experimental Data

Yang etc [1-3], [8] proposed the experimental scheme by driving the probe running in the measuring space freely, without touching or practical measurement. So only the positioning errors of different positions are sampled. In this study the practical errors are acquired by touching the specimen at different positions and with different DCC parameters, such as positioning velocity v_1, touching velocity v_2 and approaching distance a. This process of the experiment corresponds to the definition of dynamic errors and includes the consideration of all the main error sources, such as mechanical structure, guild way, environment, and most important, the probing errors. The composite errors, therefore, are acquired in the proposed experimental process. Figure 1 is the principle diagram of dynamic error collection experiment.

Figure 1. Principle Diagram of Dynamic Error Collection Experiment

A moving bridge CMM MC850 (equipped with the probe RenishawTP20, stylus length: 20mm, tip ball diameter: 4mm) is used to testify the proposed modeling method. The experiment for error sampling is arranged as following.

Table 1. Data Collection Plan of Dynamic Measuring Error

Independents	Values	Number of groups	Number of data
x/(mm)	0,150,300,450,600,750		
y/(mm)	150,300,450,550	72	
z/(mm)	-581,-473,-324		3456
zv_1/(mm/s)	20,60,100		
a/(mm)	1,2,5,8	48	
v_2/(mm/s)	2,4,6,8		

There are 6 independents: spatial coordinates x(mm), y(mm), z(mm), and DCC parameters: positioning velocity v_1(mm/s), approaching distance a(mm), and contacting velocity v_2(mm/s). The dependent is the composite spatial dynamic error e(µm). Different values of each independent are used and in every individual experiment only one variable is changed. Table 1 shows the combinations of the independents. There are totally 3456 combinations of the variables.

In order to testify the proposed method, only 5% of these 3456 groups data, about 173 points, were randomly selected. For each point the measurements were repeated five times and then the mean values are worked out. So147 data were used for the model evaluation, while the rest 26 data were used for modeling.

3. Results and Analysis
3.1. Analysis of Experimental Data

In 1923 R.A.Fisher proposed Analysis of Variance (ANOVA), which is used to determine the influential factors of a certain variable and the intersections among these factors. This method is widely used in biology and agriculture, but still seldom used in the field of mechanical engineering. In order to determine the influencing variables of dynamic errors and their interactions (expressed in form of products in following sections), the method of variance analysis is employed in this study. The experiment is repeated by 5 times because it's needed to distinguish the interaction of the influential factors against the random errors. The analysis can be done by the software SPSS16.0. The results are recorded in Table 2.

Table 2. Variance Analysis the Impact Factors of Measuring Error

Factor	Type III Sum of Squares	Degree of Freedom	Mean Square	F Value	P Value
C	290.49	1	290.49	61.91	0.00
x	335.75	1	335.75	71.55	0.00
y	100.92	1	100.92	21.51	0.00
z	530.50	1	530.50	113.06	0.00
v_1	13.86	1	13.86	2.95	0.08
a	28.77	2	14.39	3.07	0.05
v_2	22.89	2	11.44	2.44	0.09
v_1*a	30.37	2	15.19	3.24	0.04
$a*v_2$	101.81	4	25.45	5.42	0.00
$x*a$	40.20	2	20.10	4.28	0.00
$y*a$	4.58	2	2.29	0.49	0.62
$z*a$	26.22	2	13.11	2.79	0.08
v_1*v_2	1.78	2	0.89	0.19	0.83
$x*v_1$	132.40	1	132.40	28.22	0.00
$y*v_1$	14.69	1	14.69	3.13	0.07
$z*v_1$	13.16	1	13.16	2.81	0.09
$x*v_2$	27.37	2	13.68	2.92	0.06
$y*v_2$	31.90	2	15.95	3.40	0.04
$z*v_2$	35.20	2	17.60	3.75	0.03
$x*y$	135.16	1	135.16	28.80	0.00
$x*z$	15.05	1	15.05	3.21	0.07
$y*z$	35.07	1	35.07	7.47	0.01
e	525.53	112	4.69		
Sum	2493.65	147			

Note: The product of two variables expressed the interaction between the variables

The data in Table 2 show that except the interaction between coordinate value y and approaching distance a (expressed by y*a), and the interaction between positioning velocity and contacting velocity (expressed by v_1*v_2), all the other factors have significant influence on the dynamic errors at the level of 10%. Because all the factors affect the dynamic errors to different extents, they are all used as the independents for the model of dynamic error prediction. Besides, it should be considered that the coordinates (x, y, z), positioning velocity, approaching distance and contacting velocity may affect the measurement errors in form of variable nonlinearity.

3.2. IPLS Modeling for Dynamic Errors

It's known from experience that the measurement errors of a CMM have a nonlinear relationship with the selected independents. In practice the influence of errors can be synthesized in form of sum:

$$e=\beta_0+\beta_1 x+\beta_2 y+\beta_3 z+\beta_4 a+\beta_5 v_1+\beta_6 v_2+\beta_7 x^2+\beta_8 y^2+\beta_9 z^2+\beta_{10} a^2+\beta_{11} v_1^2+\beta_{12} v_2^2+\beta_{13} xy+\beta_{14} xz$$
$$+\beta_{15} xa+\beta_{16} xv_1+\beta_{17} xv_2+\beta_{18} yz+\beta_{19} yv_1+\beta_{20} yv_2+\beta_{21} az+\beta_{22} zv_1+\beta_{23} zv_2+\beta_{24} av_1+\beta_{25} av_2+\varepsilon \tag{10}$$

Where β_i (i=0~25) is the parameter that needs estimation and ε is the item of random error. The productions of variables respect the interactions in between.

With Ordinary Least Square (OLS), the largest value of VIF (Variance Inflation Factor) is as high as 625.47, which means serious multi-collinearity exists. To overcome the above limitation, IPLS is employed to eliminate the dynamic error expressed by equation (10). The whole process can be divided into two steps: ① The orthogonal projection of independents matrix is worked out by MATLAB. ② PLS regression is worked out by SICAM-P. The result is shown in Table 3.

Table 3. Parameters of the model Fitting Effect

	Component	R^2	R^2(cum)	Q^2	Q^2(cum)
	Comp1	0.64473	0.64473	0.59605	0.59605
e/μm	Comp2	0.21151	0.85624	0.20168	0.79773
	Comp3	0.07101	0.92725	0.07116	0.86889

The analysis in section 2.1 shows that only if $Q^2 \geq 0.0975$ can the modeling quality be improved by increasing the components of PLS. When two components are extracted, the expository capability of the model is $R^2=0.85624$ and the CV of the dynamic errors is $Q^2=0.79773$, which means the model has good precision.
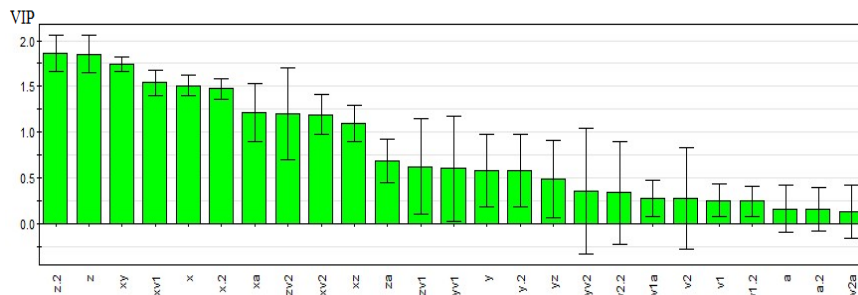

Figure 2. Variable Importance

It's seen from Figure 2 (Variable Importance (VIP)) that among the independents the coordinate x and z have the most influence upon the dynamic errors; the influences of v_1, y, a are weaker and v_2 is the weakest. This phenomenon can be explained by the practical conditions: for a bridge-type machine that is driven on one side, a closer position to the driving side will cause larger errors (x axis). A higher position will also cause larger errors (z axis). But along the driving side (y axis) the position has less influence. All the other factors have no notable influence on the dynamic errors.

Figure 3 lists the regression coefficients of the regression equation for the standardized data, which have no items of constants. It's seen that the items z, v_2, z^2, v_2^2, z*a, z*v_1, z*v_2 and

y*z have negative effects, which means the smaller they are, the bigger error they will cause. The other items, however, have the positive effects. This conclusion provides the instruction for optimizing the parameters combination.
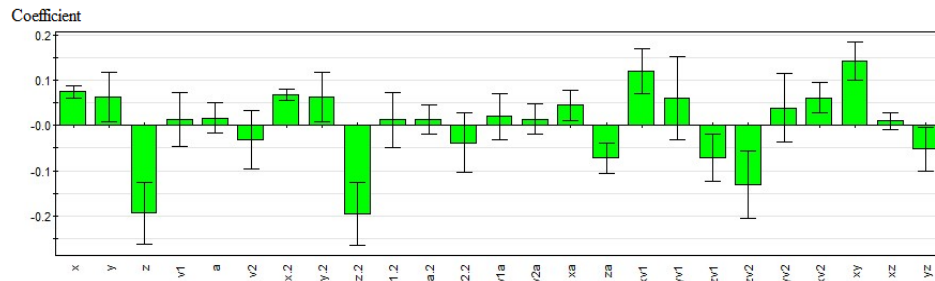

Figure 3. Coefficient Plot

3.3. Analysis of IPLS Predicting Effect

26 sets of data among all 173 are be selected and taken into the fitting function, the predicted mean square error (MSE) of IPLS regression equation, which can be evaluate the predicting effects, is calculated to be 0.94µm.As comparison, the predicted MSE of OLS and PLS are calculated to be 1.39µm and 1.04µm, respectively. Figure 4 shows the prediction accuracy of IPLS and PLS. This results show that IPLS has the better predicting effect.
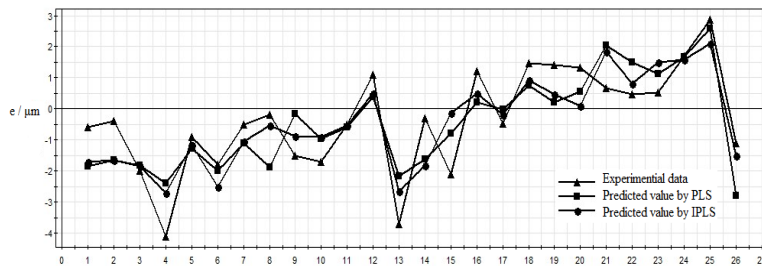

Figure 4. IPLS & PLS Observed vs. Predicted Plot

4. Conclusion

The error sources of CMM are very complicated and have uncertain interactions. So it is difficult to establish an accurate model to predict the dynamic errors by analyzing error sources. In this paper an improved modeling method based on PLS regression is proposed to avoid analyzing the interaction of error sources and to overcome the multi-collinearity of OLS regression. The results show that the proposed method IPLS has better performance of predicting and better explicability, compared with OLS and PLS.

References

[1] Fei Yetai, Zhao Jing, Wang Hongtao, Ma Xiushui. A Review of Research on Dynamic Errors of Coordinate Measuring Machines. *Chinese Journal of Scientific Instrument*. 2004; 25(4): 773-776.

[2] Yang Hongtao. Research on Error Model Building and Error Correcting Technique of Coordinate Measuring Machines. PhD Thesis. China: *Hefei University of Technology*. 2007.

[3] Yang Hongtao, Fei Yetai. Research on characteristic parameter optimization of fast probing CMM. *China Mechanical Engineering*. 2008; 19(20): 2403-2406.

[4] CWG Weekers, PHJ Schellekens. Compensation for dynamic errors of coordinate measuring machines. *Measurement*. 1997; 20(3): 197-209.

[5] Dong Chensong, Mu Yuhai, Zhang Guoxiong. Assessing the Dynamic Characteristics of CMMS with a Laser Interferometer. *Journal of Tian Jin University*. 1998; 31(5): 621-626.

[6] Zhang Yi, Liu Jizhu. Research on the CMM Compensation Errors Models Based on Finite Element Simulations. *Manufacturing Technology and Machine Tool*. 2009; 6: 78-80.

[7] Wei Jinwen, Chen Yanling. The geometric dynamic errors of CMMs in fast scanning-probing. *Measurement*. 2011; 44(3): 511–517.

[8] Yang Hongtao, Liu Yong, Fei Yetai, Chen Xiaohuai. Hybrid modeling method for CMM dynamic error. *Chinese Journal of Scientific Instrument*. 2010; 31(8): 1861-1866.

[9] Zhang Mei, FeiYetai. Hybrid Modeling of CMM Dynamic Error Based on Improved Partial Least Squares. *Nanotechnology and Precision Engineering*. 2012; 10(6): 525-530.

[10] Wang Huiwen, Wu Zaibin, Meng Jie. *Editors.* Partial Least-Squares Regression Linear and Nonlinear Methods. China: National Defense Industry Press. 2006.

[11] Wayawo-Mandata Augustin, Liu Lan. The Transmission Multicast and The Control of QoS For IPv 6 Using The Infrastructure MPLS*. International Journal of Information and Network Security (IJINS).* 2012; 1(1): 9-27.

[12] Mohammad Yanuar Hariyawan. Comparison Analysis of Recovery Mechanism at MPLS Network. *International Journal of Electrical and Computer Engineering (IJECE).* 2011; 1(2): 151-160.

[13] Zeng Xueqiang, Li Guozheng. An examination of a classification model with partial least square based dimension reduction. *Journal of Shandong University (Engineering Science).* 2010; 40(5): 41-47.

[14] Yang Maolong, Wang Yuanfang, Sun Quansen, Xia Deshen. Improved Partial Least Squares and Feature Extraction. *Computer Engineering and Applications.* 2011; 47(1): 179-181.

[15] Cheng Bo, Wu Xizhi. A modification of the PLS method. Advances in Mathematics, 1999; 28(4): 365.

[16] Chang Yingjie, Lu Xianzhong, Wang Shilong. Study on the Linearization of Analyzer for Engine Exhaust Based on Partial Least Squares. *Chinese Journal of Mechanical Engineering*. 2011; 47(10):76-81.

[17] Wu Xiaohua, Chen Dezhao. Recent Development of Non-linear Partrital Least Squares in Chemometrics. *Chinese Journal of Analytical Chemistry.* 2004; 32(4): 534-540.

[18] Nurmaini Siti, Tutuko Bambang. A new classification technique in mobile robot navigation. *TELKOMNIKA Telecommunication Computing Electronics and Control.* 2011; 9(3): 453-464.

[19] Khairudin Mohammad,Mohamed Zaharuddin,Husain Abdul Rashid.Dynamic model and robust control of flexible link robot manipulator. *TELKOMNIKA Telecommunication Computing Electronics and Control.* 2011; 9(2): 279-286.