A method of thermal error prediction modeling for CNC machine tool spindle system based on linear correlation

Wenjie Cao · Haolin Li · Qiang Li

Received: 13 April 2021 / Accepted: 30 September 2021 / Published online: 9 October 2021 © The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract
In order to improve the machining accuracy of the thermal error prediction model of CNC machine tools, a new method for calculating the position of the measuring points optimal combination researched on linear correlation is proposed, according to the thermal–mechanical finite element analysis (FEA) model of spindle system established after analyzing the thermal characteristics of heat source temperature field of CNC machine tool spindle system. Based on the correlation analysis (CA) of the finite element model of heat source temperature field of CNC machine tool spindle system, combined with the concept of mutual information (MI), this method measures the information of the measurement point variables including the thermal error variables and uses principal component analysis (PCA) to eliminate the collinearity effect within measuring point variables. By using multilinear regression (MR), The thermal error prediction model (CAMI-PCAMR) is established. The accuracy of the prediction model is verified by comparing the actual measurement thermal error with the predicted thermal error through the experimental measurement and analysis of the thermal error of the CNC end grinder test machine tool system. That the axial prediction accuracy of this method can reach 1.099 \( \mu m \), and the prediction radial accuracy can reach 1.28 \( \mu m \) under the variable ambient condition, so as to provide parameters and theoretical guidance for embedding temperature sensors in the machine tool to compensate thermal error in the design stage. And the experimental results also show that the CAMI-PCAMR method is superior to the gray correlation and fuzzy clustering(FC-GCA) modeling method.

Keywords Spindle system, Thermal error, Error modeling, Linear correlation, MI, PCA, MR

1 Introduction
The error of machine tool caused by heating accounts for 70% of the total machining error. Among these heat sources, the heat generated by the spindle is considered to be the main heat source that directly affects the workpiece accuracy. The thermal elongation of the spindle end will cause the relative position deviation of each axis, tool inclination, and machining difficulties [1, 2].

The way of compensation after the event has certain limitations and high cost, and the result of reducing thermal error by optimizing machine tool structure is unsatisfactory [3]. Therefore, it is of great significance to study the components of the thermal error of machine tools, so as to establish an accurate and reliable prediction model. At the same time, it is combined with the compensation method to embed sensors in the design stage of machine tools for effective real-time compensation, which can effectively improve the machining accuracy of machine tools.

In order to solve the above problems, the research of thermal error prediction mainly includes the following three parts: acquisition of thermal model of CNC machine tool, optimization of measuring points, and establishment of the thermal error prediction model.

Building an accurate thermal test platform and obtaining the thermal data model of the machine tool is the key to establish the thermal error prediction model. The modeling of thermal error requires to be based on the accurate data of the machine tool temperature field and deformation field [4].

At present, the finite element analysis method is more constantly used to study the thermal characteristics of the
spindle system. For example, the thermal network method based on finite difference method can be used to establish the calculation models of thermal resistance, power loss, and convective heat transfer, so as to further obtain the temperature field distribution of the spindle system [5].

Uhlmann et al. [6] proposed a 3D finite element prediction model of thermal characteristics of motorized spindle under complex boundary conditions such as heat source, convection, and contact between motorized spindle components; Abdulshahed et al. [7] developed an intelligent compensation system based on data collected by a thermal imaging camera. The system uses a scheme based on grey model and fuzzy c-means clustering method to identify key temperature points in different groups in thermal images; Xinyuan Wei et al. [8] established the 2D thermal error diagram compensation model of the whole workbench, and used the fuzzy clustering of grey correlation algorithm to select the temperature sensitive variables for modeling.

The research on the optimization of measuring points directly determines the accuracy of the prediction model. Temperature parameters are the only input variables to the CNC machine tool thermal deformation prediction model. Therefore, it is necessary to find some measuring points on the machine tool through certain methods, which can perform the influence of thermal temperature field on the thermal error to the maximum extent. Through the mathematical modeling method, the temperature measurement point data can be used to establish the functional relationship compensation model for the prediction of thermal error [9].

To find these temperature measuring points, Xu Jun et al. [10] combined RBF neural network with a genetic algorithm to discuss the optimal selection of measuring points, but the genetic algorithm has a long training time and slow iteration speed, and its identification parameters have no practical physical significance, which needs further discussion. Lee et al. [11] used the method of correlation coefficient combined with linear regression to study the temperature-sensitive points, and took the minimum residual sum of squares as the selection basis of temperature variables, and obtained a better prediction model. Xie Fei et al. [12, 13] used the method of fuzzy clustering and grey correlation degree to optimize temperature variables.

For the establishment of the prediction model, MIZE et al. [14] established a prediction model based on an art map neural network, used ASCII data to calculate and train the network with different network forms, and obtained good results, but the network method model is more complex and easy to fall into local optimum. Wang Jianchen et al. [15] used grey theory and BP model to model, and Miao Enming et al. [16] adopted the principal component regression algorithm for modeling, which significantly reduced the influence of temperature-sensitive point changes on the robustness and accuracy of the model. However, the selected combination of measuring points cannot guarantee the maximization of machine tool information points, and it is easy to fall into the local optimum when data was missing, resulting in poor results [17, 18]. Therefore, the accuracy of the model needs to be further studied.

According to the shortcomings of the existing thermal error modeling methods, a new thermal error prediction modeling for CNC machine tool spindle system based on linear correlation is proposed to improve the prediction accuracy, which mainly includes the following three parts:

1. In this paper, the temperature field and deformation field of machine tool are obtained based on the accurate finite element model of CNC machine tool.

2. Integrating with the concept of mutual information (MI), the information of measuring point variables including thermal error variables is measured, and the combination of measuring points is optimized. To eliminate the collinearity of the measured point variables, the measured point variables were optimized by principal component analysis (PCA). Different from the modeling method from the perspective of thermal sensitivity, the temperature measuring points are selected through correlation analysis (CA), and only the linear correlation points under working conditions are studied.

3. Finally, the prediction model (CAMI-PCAMR) is established by using multiple linear regression (MLR) model method.

The effectiveness of the proposed CAMI-PCAMR method is verified by comparing and analyzing the thermal error measurement experiments of the CNC machine tools experimental platform.

The framework of this paper is about to be conducted as follows: The first section makes a brief introduction to this paper. In Section 2, the CAMI-PCAMR method to obtain the optimal linear combination measuring point position selecting method and establishing the prediction model is put forward. In Section 3, based on the thermal structure of the machine tool, the thermo-mechanical coupling FEA model of the machine tool spindle system is established, and the optimization implementation method of temperature measuring points based on the modeling theory is elaborated in detail. In Section 4, the validity and reliability of the thermal error model are verified under actual working conditions, and the modeling accuracy is compared with other traditional modeling methods. Finally, the conclusions are drawn in Section 5.
2 CAMI-PCAMR modeling theory

2.1 The linear screening of heat-sensitive nodes

In order to ensure the maximum sensitivity and the principle of being closer to linearity as much as possible, and improve the optimization efficiency of the measuring points, the CA of the temperature and deformation data of the data points is carried out first, and the points that have less influence on the thermal deformation are screened out.

Through the analysis of the thermal–mechanical coupling FEA model of the machine tool spindle system, the temperature matrix composed of the temperature changes of the n-th node with time is expressed as $T_i(j)(i = 1, 2, \ldots, n; j = 1, 2, \ldots, m)$, and the matrix composed of axial deformation and radial deformation of the grinding wheel spindle with time is expressed as $D_j(j = 1, 2, \ldots, m)$. According to the CA theory [19], the correlation coefficient between the node temperature of the grinding wheel spindle and the axial and radial deformation of the grinding wheel spindle can be expressed as follows:

$$r_i = \left(\frac{1}{m} \sum_{j=1}^{m} (T_i(j) - \bar{T}_i)(D_j - \bar{D})\right) / \sqrt{\left(\frac{1}{m} \sum_{j=1}^{m} (T_i(j) - \bar{T}_i)^2\right) \left(\frac{1}{m} \sum_{j=1}^{m} (D_j - \bar{D})^2\right)}$$

where $T_i$ is the average value of $T_i(j)(j = 1, 2, \ldots, m)$; $\bar{D}$ is the average value of $D_j$, and $m$ is the time substep.

According to Eq. (1), the correlation coefficient between the thermal deformation and the temperature field is calculated. If $0 \leq r_i < 0.3$, it means that the temperature of the node is weakly related to the deformation; if $0.3 \leq r_i < 0.5$, it means that the temperature of the node has a low correlation with the deformation; if $0.5 \leq r_i < 0.8$, it means that the temperature of the node is significantly related to the deformation; if $0.8 \leq r_i < 1$, the temperature of the node is highly correlated with the deformation. For the node variables of $r_i < 0.5$, they are excluded because of the small correlation between temperature and deformation.

In order to ensure the minimum correlation coupling of the selected temperature measuring points, so as to meet the principle of uncorrelation as far as possible [19], and at the same time meet the principle of minimum point distribution, so as to obtain more accurate measuring points. Therefore, the MI theory is introduced to analyze the temperature variables between the data of each point.

The MI between variables can measure the information of another random variable contained in the random variable, which can describe the amount of information provided by each temperature measurement point for the thermal deformation of the CNC machine tool [20].

After the node temperature $T_i(j)$ and deformation $D_j$ are renumbered and analyzed, $T = \{T_1(t), T_2(t), \ldots, T_i(t)\}$ and deformation $D(t)$ are obtained. The range $[T_{min}, T_{max}]$ of each node temperature value is divided into $N$ equal parts, and the range $[D_{min}, D_{max}]$ of machine tool axial and radial deformation is divided into $M$ equal parts.

According to the probability distribution statistics, the probability $P_T(j)$ and $P_D(k)$ of $T_i(t)$ and $D(t)$ falling into each cell are calculated, where $i = 1, 2, \ldots, n; j = 1, 2, \ldots, N; k = 1, 2, \ldots, M$.

In order to ensure that the combination of measuring points meet the information maximum principle at the same time and to improve the optimization efficiency of the measurement points, the information content of the temperature and deformation data of the points is analyzed according to the entropy function $H(T_i)$ in the information theory. The uncertainty of the change of the temperature value at the position of the temperature measuring point of the machine tool is expressed by the following formula:

$$H(T_i) = -\sum_{j=1}^{N} P_T(j) \log_2 P_T(j)$$

(2)

The uncertainty of machine tool deformation can be described by the entropy function $H(D)$:

$$H(D) = -\sum_{k=1}^{N} P_D(k) \log_2 P_D(k)$$

(3)

Then, the MI provided by the temperature $T_i(t)$ of the measuring points for the deformation of the CNC machine tool is a follows:

$$I(D;T_i) = \sum_{k=1}^{M} \sum_{j=1}^{N} P(D_kT_i(j)) \log_2 \frac{P(D_kT_i(j))}{P(T_i(j))P(D_k)}$$

(4)

The temperature nodes area with the large amount of MI are selected. After the temperature nodes in the area sorted by mutual information, a total of 20 axial and radial temperature measuring points with the largest amount of information are found.

2.2 The optimal combination of temperature-measuring points

On the premise of keeping the loss of data information as small as possible, the high-dimensional temperature variable space is recombined. According to the principle of principal factor, the variable value of each measuring point in the thermal error prediction modeling needs to have a strong correlation with its thermal error variable value.

Principal component analysis (PCA) is used to recombine a large number of variable data with a certain correlation into a non-correlated comprehensive variables to replace the original variables, so as to simplify the data and reveal the relationship between variables [18, 21].

© Springer
For the node temperature $X = [x_1, x_2, \cdots, x_p]^T$ after the screening, the $n$ groups of data are expressed as $x_i = [x_{i1}, x_{i2}, \cdots, x_{ip}]^T, i = 1, 2, \cdots, n$.

The calculation steps of principal components are as follows:

(1) The covariance $\sigma_{ij}$ and the sample correlation coefficient $r_{ij}$ are obtained from the variable data, which expressed as follows:

$$
\begin{align*}
\sigma_{ij} &= \frac{1}{n-1} \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \\
\end{align*}
$$

(5)

$$
\begin{align*}
\bar{x}_i &= \frac{1}{n} \sum_{k=1}^{n} x_{ki} \\
\end{align*}
$$

(6)

In which:

In which:

$$
\begin{align*}
\bar{x}_i &= \frac{1}{n} \sum_{k=1}^{n} x_{ki} \\
\end{align*}
$$

(6)

The covariance matrix $V$ and correlation matrix $R$ of samples are obtained as follows:

$$
\begin{align*}
V &= \text{diag}(\sigma_{ij}) \\
R &= \text{corr}(\sigma_{ij}) \\
\end{align*}
$$

(7)

(8)

(2) Solve $p$ eigenvalues of $R$, denoted as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.

(3) The number of principal components is selected by using the variable contribution rate of principal components, and the contribution rate of the $i$ th principal component is represented by $\lambda_i / \sum_{j=1}^{p} \lambda_j$; that is, the degree of the original data information retained by the $i$ principal component.

(4) Calculate the unit eigenvector $l_i, i = 1, 2, \cdots, m$ corresponding to the first $m$ eigenvalues.

(5) Calculate the principal component $y_i = l_i^T X$ of $X$. Then, the principal component variable is obtained by analysis.

$$
\begin{align*}
Y &= [y_1, y_2, \cdots, y_m]^T \\
\end{align*}
$$

(9)

Finally, according to the theory of MR, the linear coefficients of temperature measuring points are determined by using the least square method, and the principal axis thermal error prediction model of axial and radial errors with respect to principal components is established by linear fitting, which is convenient and quick with less computation.

According to the CA theory, combining the advantages of PCA and MI theory, the CAMI-PCAMR method for calculating the position of the measuring points optimal combination proposed in this paper follows the principle including nearest linearity principle, mutual irrelevance principle, minimum distribution principle, maximum temperature-sensitivity principle, and principal factor principle, which can obtain better thermal error modeling of MR theory.

### 3 Temperature measuring points identification and prediction model establishment

In order to discuss the performance of the CAMI-PCAMR modeling method, this paper takes the spindle system of CNC machine tools as an example to explain the effectiveness of the proposed method.

#### 3.1 Thermal error analysis of spindle system

Firstly, through the thermal test experiment of the machine tool, the temperature field and displacement data are obtained by the thermal imager shooting and displacement measurement. Or through the thermal characteristic analysis of the machine tool, using the spindle system of the machine tool, and according to the data obtained from the actual working conditions, setting the boundary conditions of the thermal coupling finite element model with enough accuracy to obtain the data, this method is adopted in this paper.

The analysis shows that the temperature load in the CNC machine tool spindle system in the process of rotation is mainly caused by the front bearing, rear bearing, and end bearing, which causes the axial and radial deformation of the grinding wheel and affects the machining accuracy.

Through the transient analysis in ANSYS, the analysis process is divided into 20 sub-steps, each of which lasts for 6 min. The temperature distribution nephogram of the spindle system is captured as shown in Fig. 1. The local maximum temperature is 45.58°C, which occurred at the rear bearing. The transient thermal deformation of the spindle system can be obtained by adding boundary conditions and loading the thermal load of the transient temperature to the transient-mechanical coupling analysis. Figure. 2 a and b respectively show the radial and axial thermal deformation nephogram of the spindle working for 4h under the working condition. It can be seen from Fig. 2 a that the maximum axial deformation of the spindle system appears on the upper surface of the grinding wheel, and it can be seen from Fig. 2 b that the maximum radial deformation appears at the rear bearing. At the same time, it can be concluded from the analysis that the thermal deformation of the spindle is little affected by the thermal deformation of other parts of the CNC machine tool which can be ignored.
3.2 Screening of thermal temperature nodes

According to the study of linear measuring points of the spindle system of the CNC machine tool, the correlation analysis is made on all nodes of the model, and the correlation coefficient between temperature and thermal deformation components can be known from Table 1.

There are different degrees of correlation between temperature and thermal deformation components. If $r_i > 0$, the temperature of the joint is positively correlated with the deformation. If $r_i < 0$, it shows that the temperature of the node is negatively correlated with the deformation. The closer the correlation coefficient value $r_i$ is to $-1$ or $+1$, the stronger the correlation coefficient is, while the closer to 0, the weaker the correlation coefficient is. If $r_i = 1$ or $-1$, it means that the temperature of the node is linearly related to the deformation. According to the needs, the temperature nodes with $-0.5 < r_i < 0.5$ which are weak linear correlations are eliminated.

Then, based on the principle of mutual irrelevance and the principle of minimum distribution points, the maximum amount of information contained in each measuring point is guaranteed, and the mutual information between temperature and deformation of model nodes is calculated, and the node distribution table with larger mutual information of temperature and deformation is obtained, as shown in Table 2.

The analysis shows that the nodes with large mutual information are mainly distributed in several small areas, that is, the key temperature-sensitive areas. From each area with large mutual information, the nodes with large information are selected as temperature measuring points, and 20 axial and radial temperature measuring points are found, as shown in Table 2.

3.3 Extraction of thermal measuring points

Based on the principle of mutual uncorrelation, PCA was used to eliminate the collinearity between the larger temperature measurement points of $I(D; T_i)$, and the temperature variables of node sub-steps were extracted as sample data, and the measurement points were reanalyzed and numbered as $T_1 \sim T_{20}$.

Firstly, the Bartlett test was used to test the variables of the measurement points:

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} r_{ij} \cdot 1,2, \ldots, k}$$

The value of KMO sampling suitability is 0.768 > 0.6. It is verified that the mutual information screening is effective, the data variables have little relevance and can be classified, and the significance of bartlett sphere is 0.001 < 0.05, which is suitable for PCA.

Through PCA, Caesar normalized maximum variance method is used to rotate iteration until convergence, and the extracted values of common factor variance of variables are all greater than 0.5. It can be seen from Table 3 that the main factors 1 and 2 reflect the variable $T_6, T_{15}$, namely nodes 41,903 and 48,433, and main factors 3,
4, and 5 reflect variable $T_{19}$, $T_5$, and $T_{17}$, namely nodes 49,125, 39,527, and 49,006.

The factor $F_1, F_2, F_3, F_4$, and $F_5$ in Table 3 are determined by node $T_6, T_{15}, T_{19}, T_5$, and $T_{17}$. And the position of node $T_6, T_{15}, T_{19}, T_5$, and $T_{17}$ is shown in Fig. 3.

Fig. 2 Thermal error distribution of spindle system. a Axial thermal error nephogram of spindle, b Radial thermal error nephogram of spindle

Table 2 The distribution of measuring points with large amount of information in key temperature-sensitive areas

| Heat-sensitive area         | Node number of temperature measuring point with larger $I(D; T_i)$ |
|----------------------------|---------------------------------------------------------------------|
| $D_1$                      | Front bearing area                                                  |
| $D_2$                      | Back bearing area                                                   |
| End bearing area           |                                                                     |

Table 3 The composition score coefficient matrix

| Factor $F_1$ | Factor $F_2$ | Factor $F_3$ | Factor $F_4$ | Factor $F_5$ |
|--------------|--------------|--------------|--------------|--------------|
| $T_6$        | 0.709        | 0.705        | 0.000        | 0.000        | 0.000        |
| $T_{15}$     | -0.705       | 0.709        | 0.000        | 0.000        | 0.000        |
| $T_{19}$     | 0.000        | 0.000        | -0.987       | 0.163        | -0.011       |
| $T_5$        | 0.000        | 0.000        | -0.163       | -0.980       | 0.113        |
| $T_{17}$     | 0.000        | 0.000        | -0.008       | -0.114       | -0.994       |

The expression of component $F_1, F_2, F_3, F_4$, and $F_5$ is as follows:

$F_1 = 0.709 \times T_6 - 0.705 \times T_{15}$

$F_2 = 0.705 \times T_6 - 0.709 \times T_{15}$

$F_3 = -0.987 \times T_{19} - 0.163 \times T_5 - 0.008 \times T_{17}$

$F_4 = 0.163 \times T_{19} - 0.980 \times T_5 - 0.114 \times T_{17}$

$F_5 = -0.011 \times T_{19} + 0.113 \times T_5 - 0.994 \times T_{17}$

The expression of component $F_1, F_2, F_3, F_4$, and $F_5$ is as follows:
3.4 The establishment and analysis of thermal error prediction model

The temperature variables of the sub-step and the thermal deformation variables of the spindle obtained from the optimized measuring points are taken as the characteristic sample variables, and the principal components extracted and combined are analyzed by multiple linear regression. In this way, the prediction model of axial and radial thermal deformation error of the spindle system can be established.

Firstly, the variables of the prediction model are evaluated. It can be seen from Table 4 that the Durbin-Watson coefficients of the axial deformation and the radial deformation are 1.515 and 1.966, respectively, which indicates that the principal component analysis is effective and the sample independence is good. And the independent variable can represent the axial deformation error of 98.4% and the radial deformation error of 96.4% because the $R^2$ is more than 30%, which is acceptable.

It can be seen from Table 5 that the variance inflation factor (VIF) value which is a measure of the severity of complex collinearity in multiple linear regression model of measuring points $F_1, F_2, F_3, F_4$, and $F_5$ is less than 5, which indicates that there is no multicollinearity, and the mutual information analysis effect is good. $F_1, F_3$ positively affects the axial error, and $F_2, F_4, F_5$ negatively affects the axial error; $F_1, F_2, F_3$ positively affects axial error, and $F_4, F_5$ negatively affects the axial error.

Therefore, the spindle thermal error model can be established by multiple linear regression analysis:

$$
E_1 = [1, F_1, F_2, F_3, F_4, F_5] \begin{bmatrix}
-38.20 \\
-167.56 \\
16.73 \\
-258.21 \\
217.82 \\
193.23
\end{bmatrix}
$$

$$
E_2 = [1, F_1, F_2, F_3, F_4, F_5] \begin{bmatrix}
-8.66 \\
-182.20 \\
5.77 \\
-246.83 \\
230.46 \\
193.26
\end{bmatrix}
$$

Table 4 The evaluation of model samples

| Model          | $R^2$  | Durbin Watson coefficient |
|----------------|--------|---------------------------|
| Axial deformation | 98.4%  | 1.515                     |
| Radial deformation | 96.4%  | 1.966                     |

According to the regression model evaluation and analysis of $R^2$ value, the axial deformation and radial deformation are 98.4% and 96.4%, respectively, which indicates that the results are very good.

Figure 4 shows the comparison between the axial predicted $E_1$ and radial predicted $E_2$ of the thermal error model and the actual value.

It can be seen from Fig. 5 that the error prediction model conforms to the normal distribution. The maximum deviation between the axial error prediction value and the actual value of the model is $0.8411 \mu m$, and the maximum deviation between the radial error prediction value and the actual value is about $0.7295 \mu m$. So, it can be seen that the prediction modeling and analysis data based on MR are reliable and accurate, and can also achieve good results.
4 Verification and example of thermal error of spindle system

4.1 The experimental device

In order to verify the thermal sensitivity characteristics of the machine tool spindle system, relevant tests were carried out on the experimental platform of the actual CNC machine tool. As shown in Fig. 6, the rated speed of the spindle system of the grinder is 1488r/min, and it takes over 4h to achieve thermal stability from starting work.

4.2 Test data collection and analysis modeling

According to the method shown in Fig. 7 of the flow chart, the position and regression model of the thermal sensitive temperature measuring point combination of the CNC spindle system are used, and the sensors are arranged at the measuring point location.

In order to measure the thermal deformation of the spindle system of CNC machine tools and the temperature of temperature-sensitive measuring points, the temperature-displacement synchronous measuring system is selected to collect the temperature field and thermal deformation of the CNC machine tool spindle. The acquisition system can collect 600 corresponding temperature data per minute. The PT100 temperature sensor is used to measure the temperature change of the machine tool spindle, and the collected data is recorded as $T_1$, $T_2$, $T_3$, $T_4$, and $T_5$. The eddy current displacement sensor is used to measure the axial thermal deformation and radial thermal deformation of the spindle system, which are placed near the workpiece position and recorded as $D_1$, $D_2$, respectively.

The arrangement of temperature sensors and displacement sensors is shown in Fig. 8.
To reduce the influence of ambient temperature on the test results, the ambient temperature (room temperature) fluctuated in the range of 19~20 °C during the test, which met the requirements of the thermal characteristic measurement standard.

The collected temperature and deformation data are processed to remove the invalid samples and invalid features, and the abnormal values are removed according to the Laida criterion [22]:

\[ P(|x - \mu| > 3\sigma) \leq 0.003 \]  

(11)

In which:

\[
\sigma = \left( \frac{1}{n-1} \sum_{i=1}^{n} v_i^2 \right)^{1/2} = \left( \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2 / n \right) / (n-1) \]  

(12)

The data whose average value of sample data is greater than 3\(\sigma\) is regarded as outliers, and the data is completed and fitted by the least square method, so as to make the data as close to the real value as possible, and ensure the continuity and smoothness, and also reduce the gross error caused by experimental operation and instrument influence.

Figure 9 shows the temperature variation of five measuring points with time after processing.

Under the working condition that the rotating speed of the spindle is 1088 r/min, the points with great correlation between the temperature measurement data and the axial thermal deformation and radial thermal deformation are calculated based on the CA algorithm, and the temperature measurement points are selected by the PCA, and a MR model is established to predict the axial and radial thermal deformation of the spindle system under the working condition, as shown in Figs. 10a and b.

According to the comparative analysis between the predicted values of the model and the measured values, it can be seen that the axial maximum absolute residual error is 1.991 \(\mu m\), the average residual error is 0.623 \(\mu m\), the radial maximum absolute residual error is 1.59 \(\mu m\), and the average residual error is 0.551 \(\mu m\). Besides, because the experimental platform is limited by its cannot being disassembled. The errors of the result can be reduced to a smaller range. if the sensor position is chosen closer to the selected measuring point, that is, if the sensor is buried in the design stage.

### 4.3 The comparison between prediction models’ performance

In order to further verify the accuracy and superiority of CAMI-PCAMR (CP) prediction model, the results of CP prediction model are compared with those of another two methods [11]: the method of correlation coefficient combined with linear regression (CL) and method [16], a commonly used modeling method; Fuzzy clustering prediction model based on gray correlation (FG), whose result and comparison of residual have shown in Fig. 11.

According to the comparative analysis among the predicted values of three models, it can be seen that axial maximum absolute residual error using the FG method is 2.469 \(\mu m\), the average residual error is 0.915 \(\mu m\), the radial maximum absolute residual error is 1.926 \(\mu m\) and the average residual error is 0.767 \(\mu m\), and the axial maximum absolute residual error using the CL method is 2.651 \(\mu m\), and the average residual error is 0.908 \(\mu m\).
It is obviously shown from the comparison results that the relative residual error of CP thermal error prediction model is smaller than that of FG prediction method and CL prediction method. It can be seen that the CNC machine tool thermal error modeling method which combines MI to optimize measuring point and build RM model based on PCA has higher prediction accuracy and is superior to FG prediction method and CL prediction method.

5 Conclusion

In this paper, a high-precision prediction model for linear thermal error under different conditions is proposed, which combines the advantages of the PCA method and MI theory, follows five principles, and verifies the effectiveness of the proposed research method by analyzing and comparing it with the thermal error measurement experiment on the experimental platform of CNC machine tools. Different from the modeling method from the perspective of thermal
sensitivity traditionally, the temperature measuring points in this paper are selected through correlation analysis, and only the linear correlation points under working conditions are studied. It provides a new idea from the perspective of correlation for researchers to work on the thermal error prediction model. And it makes up for the problem that the regression method in the prediction model modeling is not good enough due to the multiplicity of variables, which improved the establishment method of the regression model.

From practical significance, the prediction method CAMI-PCAMR with axial accuracy of 1.099 μm and radial accuracy of 1.28 μm, which is superior to the CL and the FG modeling method, lays a foundation for the work of thermal error compensation theory and can be used as a reference and guidance.

1. For linear and nonlinear thermal errors in machine tool spindle, model compensation is more useful for linear thermal errors. The study provides a basis for this conclusion.
2. This study confirmed that the CAMI-PCAMR method can effectively eliminate the temperature-insensitive points on thermal error.
3. MI theory can eliminate multi-collinearity between temperature data, optimize thermal variables involved in principal component analysis, and improve accuracy. It has overcome the problem that the accuracy of the regression model cannot be more accurate due to the multi-collinearity of dependent variables. It is proved that the mutual information theory can theoretically eliminate the multicollinearity between temperature data and optimize the variables involved in the principal component analysis.
4. The optimized PCA dimension reduction can effectively identify the optimal temperature measuring point combination of MR. It is proposed that the PCA dimensionality reduction after MI optimization can effectively identify the most optimal combination of temperature-measuring points in multiple linear regression, which contributes to the problem of measuring point optimization.
5. It is proposed that the CAMI-PCAMR modeling method can be generally used to predict the linear thermal error of the machine tool spindle system theoretically. However, due to the limitation of research conditions, it has not been verified on other machine tools, which needs to be further confirmed and discussed.
6. Since the influence of thermal hysteresis and thermal drift during the operation of the studied CNC machine tool is too small to be ignored, relevant problems have not been considered in this study. Subsequent scholars can do more research from this point of view. It is recommended that the follow-up researchers can do more work from this perspective for supplement and further research.

Acknowledgements The authors pay great appreciation to the anonymous referees and editor for their valuable comments and suggestions.

Author contributions Cao performed the experiment and the data analyses, and wrote the manuscript; Li Haolin gave constructive suggestions on the ideas of the research and provided experimental platforms and instruments; Li Qiang helped perform the analysis and made great contributions in discussion. All authors read and approved the final manuscript.

Availability of data and materials All data generated or analyzed during this study are included in this published article.

Declarations

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to Participate Not applicable.

Competing interests The authors declare no competing interests.

References

1. Mayr J, Jedrzejewski J, Uhlmann E et al (2012) Thermal issues in machine tools. CIRP Ann Manuf Technol 61(2):771–791
2. Krulewich DA (1998) Temperature integration model and measurement point selection for thermally induced machine tool errors. Mechatronics 8:395412 Technology 69(9):2593–2603
3. Yang S, Yuan J, Ni J (1996) Accuracy enhancement of a horizontal machining center by real-time error compensation [J]. J Manuf Syst 15(2):113–124
4. Liu Q, Yan J, Pham DT, Zhou Z, Xu W, Wei Q, Ji C (2016) Identification and optimal selection of temperature-sensitive measuring points of thermal error compensation on a heavy-duty machine tool [J]. The International Journal of Advanced Manufacturing Technology 85(1–4):345–353
5. Liu CH H, Luo GJ, He W et al (2010) Steady state analysis of a spindle system based on thermal net-work [J]. Chin J Mech Eng 21(6):631–635
6. Uhlmann E, Hu J (2012) Thermal modelling of a high speed motor spindle [J]. Prorelia CIRP 1:313–318
7. Abdulshahed AM, Longstaff AP, Fletcher S, Myers A (2014) Thermal error modelling of machine tools based on ANFIS with fuzzy c-means clustering using a thermal imaging camera. Appl Math Model 39(7):1837–1852
8. Wei XY, Miao EM, Liu H, Liu SL, Chen SX (2019) Two-dimensional thermal error compensation modeling for worktable of CNC machine tools [J]. Int J Adv Manuf Technol 101:501–509
9. Miao E, Gong Y, Niu P, Ji C, Chen H (2013) Robustness of thermal error compensation modeling models of CNC machine tools [J]. Int J Adv Manuf Technol 69(9):2593–2603
10. Xu J, Zhou YH, AJ JY, Lei LY, Fu XS, Qin YH. (2019) Application of genetic neural network in optimal layout of water quality monitoring points [J]. Environ Eng 2019(6):177–183
11. Lee J, Yang S (2002) Statistical optimization and assessment of a thermal error model for CNC machine tools[J]. Int J Mach Tools Manuf 42(1):147–155
12. Xie F, Wang L, Yin M, Tan F, Yin GF (2019) Optimization method of temperature measurement points for thermal error of CNC machine tools[J]. Modular Mach Tools Autom Process Technol 2019(6):45–49
13. Zhang L, Chen GH, Zhao DZ (2020) Optimization method for temperature measuring points of machine tool spindle based on fuzzy clustering and grey theory [J]. Mach Tools Hydraul 48(22):85–89
14. Mize CD, Ziegert JC (2000) Neural network thermal error compensation of a machining center[J]. Precis Eng 24(4):338–346
15. Wang JC, Lin SQ, Shen YX, Xie CX, Deng XL. (2019). Research on thermal error measuring point optimization and modeling technology of CNC machine tool spindle [J]. Aeronautical manufacturing technology, 2019, 6:41–46,59
16. Miao EM, Liu Y, Gao Z, Han LH (2016) Variability of temperature sensitive points of CNC machine tools and its influence [J]. Chin Mech Eng 27(3):285–322
17. Zhang WW, Wang ZG, Liu XR, Li G, Wan pinlei, Wang Wei. (2017). Research on temperature measurement point optimization and thermal error prediction method of CNC machine tools.
18. Yang ZQ, LV SY, Bo JD. (2020) Optimization of thermal sensitive points of machine tool spindle system based on gra-pca [J]. Mach Tools Hydraul 48(23):93–98
19. Shi H, Mei XS, Yang J, Tao T. (2012). Thermal sensitivity analysis and Experimental Research on motorized spindle of CNC machine tool [J]. Huazhong University of science and Technology (NATURAL SCIENCE EDITION), 2012, 40 (S2): 26–30
20. Thermal analysis and thermal error compensation of high speed machining center (2014) [D]. Guo Songlu. Zhengzhou University
21. Li HL, Y XJ (2010) Optimization design method of thermal error temperature measurement point for CNC machine tool spindle system [J]. Chin Mech Eng 21(7):804–808
22. Li ZW, Lei XY, Gao L (2014) Analysis on pretreatment method of track irregularity detection data [J]. J Railw Sci Eng 11:43–47

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.