Int J Adv Manuf Technol (2009) 42:168–179
DOI 10.1007/s00170-008-1570-x

ORIGINAL ARTICLE

Thermal-error modeling for complex physical systems: the-state-of-arts review

J. W. Li · W. J. Zhang · G. S. Yang · S. D. Tu · X. B. Chen

Received: 15 July 2007 / Accepted: 14 May 2008 / Published online: 17 June 2008

Abstract Thermal errors are very important to modern systems including both physical and biological systems. It is well known that temperature rise causes thermal expansion of an object and may induce internal stresses in the object when it is subject to constraints. It is further known that friction is a passive source of heat causing temperature rise in many systems. In this paper we present a critical review of the work, published in the last decade, towards modeling of thermal errors due to the friction-induced heat in complex physical systems. We first develop a framework of the issues so that we can place each significant work in a proper "location" in the framework. To model a phenomena, we consider that there are two general schools, namely principle-based (or white-box) and empirical-based (or black-box). We further classify the principle-based model into the analytical model and numerical model and the empirical-based model into the static model and dynamic model. We discuss the key studies reported in the literature by examining the issues that the studies have addressed and the modeling techniques or methods that the studies employed. As a result, we conclude several new research divisions such as (1) modeling of the whole physical process of thermal errors (i.e., from the heat source, especially due to friction, to the thermal deformation, and from the deformation to friction, and to the heat further—a closed-loop process), (2) studies of reliability and resiliency of the thermal-error management along with accuracy in prediction of thermal errors, and (3) studies of an integrated principle-based and empirical-based modeling and management approach to thermal errors.

Keywords Friction · Temperature · Thermal error · Modeling · Deformation · Heat

J. W. Li · W. J. Zhang · S. D. Tu
School of Mechanical & Power Engineering,
East China University of Science and Technology,
130 Meilong Road,
Shanghai 200237, China

J. W. Li · W. J. Zhang (✉) · G. S. Yang · X. B. Chen
Department of Mechanical Engineering,
University of Saskatchewan,
Saskatoon, SK S7N5A9, Canada
e-mail: Chris.Zhang@Usask.ca

G. S. Yang
Information Engineering College,
Central University for Nationalities,
Beijing 100081, China

1 Introduction

Any system stays in a thermal changing environment. When the temperature of a component changes, properties of the component may change, which is generally called thermal effect. In machine or motion systems, change in the deformation of a component due to a change in the temperature of the component is of most interest. Thus, in this paper, the thermal effect refers to change in deformation due to a change in temperature. Furthermore, the change in deformation can be function-positive (e.g., shape memory alloy actuator) or function-negative (e.g., work-piece distortion in a machining operation); in the latter case, the deformation due to a change in temperature is also called error—namely thermal error. This paper is interested in thermal error, especially modeling of thermal errors.

Figure 1 shows a typical example of the machine tool system [1]. From this figure, it can be seen that there are several pairs of components that perform a relative contact motion, namely (1) ball screw, (2) rolling bearing, (3) motor, and (4) tool–chip interface. There are frictions in the contact surfaces (e.g., the contact between the tool and chip, the

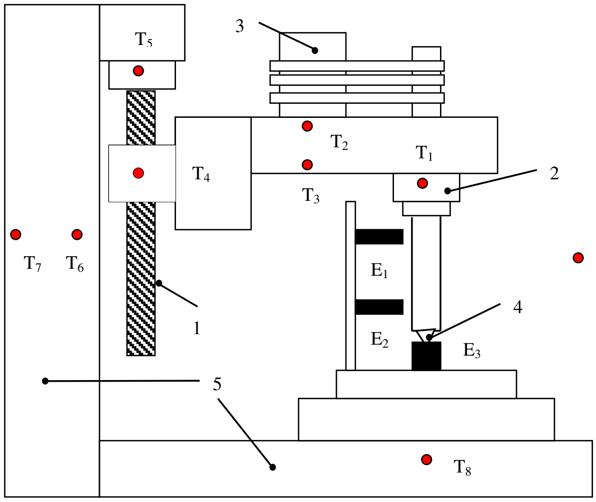


Fig. 1 The general configuration of a vertical machining centre [1] (*1* ball screw, *2* rolling bearing, *3* motor, *4* tool, *5* main structure)

contact between the ball and screw bar), and these frictions can generate significant heat in the machine and its environment (e.g., there may be significant heat generated over the tool–chip interface). It is well known that heat will transfer based on the three mechanisms, i.e., heat conduction, heat convection, and heat radiation. The effects of heat transfer over the components are such that the temperature of these components changes, which cause the components deform, thus departing away from their original or normal geometrical shapes, i.e., errors. Furthermore, such deformation may lead to a significant change in the interface between two components, leading to a change in the friction behavior and friction-induced heat; the heat further causes "additional" deformation of the components. For example, in machine tool systems, the deformation in the screw bar can affect its interaction with the ball, which further changes the friction behavior leading to a degraded performance of the ball screw mechanism.

Our previous work [2] shows that even in a system with simple structures, the thermal errors might be complex, see Fig. 2 which shows a piezoelectric stick-slip (PZT-SS) actuator system. In this system, a piezoelectric (PZT) actuator pushes the stage. The relation between the stage and end-effector is such that when the PZT actuator extends, there is a sufficient friction between the stage and end-effector such that they move together towards right. While the PZT actuator returns to its original state, i.e., moving towards left, there is a sliding motion between the stage and end-effector—in particular the end-effector tends to remain due to its inertia and the stage with the PZT actuator moves back to left. Temperature rises continuously due to the friction between the stage and end-effector. The change in temperature further causes both the stage and end-effector to deform, and this deformation can in turn affect the friction between the stage and the end-effector. Eventually, the altered friction further affects the stick-slip motion, resulting in the poor accuracy and repeatability of the actuator system and even the failure in motion in some cases [2]. A finding was also made with the PZT-SS actuator system that the PZT actuator can also generate heat which contributes to temperature rise on the interface between the stage and end-effector [2].

From the above two examples, the coupling relationship between the friction and structure deformation can be further illustrated in Fig. 3 in a more generalized manner. With this understanding, a classification of thermal-error modeling issues can be defined. In particular, Issue 1: modeling of the friction–temperature relation; Issue 2: modeling of the temperature-deformation relation; Issue 3: modeling of simultaneous interactions among friction, heat, temperature, and deformation. In Table 1, we list some of the studies reported in the literature in the context of these issues and two schools.

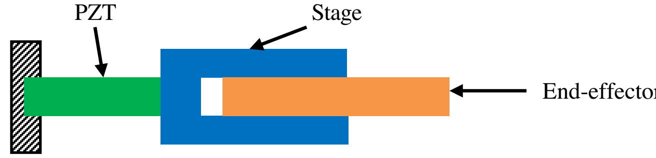
It is interesting to observe from Table 1 that (1) on Issue 2, principle-based approach is only suitable for the system with simple structures; (2) there appears no work on Issue 1 and Issue 3 with the empirical-based approach; (3) there appears no work on modeling the coupling relationship between friction and deformation—namely Issue 3.

The remainder of this paper is organized as follows: principle-based models are discussed in Section 2. Empirical-based models are discussed in Section 3. Challenges of the principle-based models are elaborated in Section 4. Challenges of the empirical-based models are presented in Section 5. Conclusions and future work are discussed in Section 6.

2 Principle-based models

According to Fig. 3, the entire process from heat generation to deformation consists of two steps: (1) from heat generation to temperature field, and (2) from temperature field to deformation (i.e., thermal error). For (1), the principle is available which governs the relationship between friction and heat, for example the widely used law for the friction-induced heat is $Q = \tau v$, where Q denotes the heat, τ the

Fig. 2 PZT-SS actuator



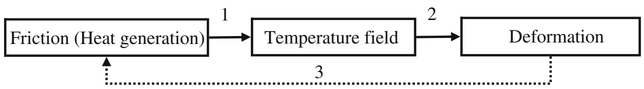


Fig. 3 Coupling relationship among friction, temperature field and deformation

shearing stiffness, and ν the relative velocity [3, 4]. Further, the principle from heat to temperature field is available of three heat transfer mechanisms (i.e., conduction, convection, and radiation). For (2), the principle is available; for example, $\Delta L = \alpha L \Delta T$, where α denotes thermal expansion coefficient, L the length of a sample, and ΔT the temperature change.

The heat generated due to friction is partially dissipated into the environment, while a part of heat will remain in two objects that are interacting. Particularly, the heat remains in the objects can be calculated by using the heat partition ratio [4]: the heat of $(1-B)\ Q$ remains in body 1 and the heat of BQ remains in body 2 provided that there is no heat loses, where $0<B<1$. In principle-based models, the relationship between thermal errors and heat generated are described by a system of non-linear differential equations. The solution to such equations can be obtained by analytical methods as well as numerical methods. Therefore, principle-based models are further classified into analytical models and numerical models.

2.1 Analytical models

The analytical model is often used to establish the relationship between heat generation and temperature field. Komanduri and Hou [5] proposed an analytical model for temperature distribution due to frictional heat source at the tool–chip interface in the metal-cutting process. The model is based on the Jaeger's heat source model, e.g., the chip is considered as a moving-band heat source and the tool is regarded as a stationary square heat source, while Komanduri and Hou's model takes into account the effect of additional boundaries along with a superimposed non-uniform distribution of heat intensity. Karpat et al. [6] also proposed an analytical model to study temperature field at tool–chip interface in the cutting process. Compared with Komanduri and Hou's model, Karpat's heat source model considers mechanical shearing in the primary zone as well. Chen et al. [7] proposed an analytical model to study the interface temperature rise in polishing a polycrystalline diamond (PCD) surface, in which the heat generated is taken as the product of the friction force and the relative sliding velocity between the PCD asperities and the metal disk surface. The Jaeger's moving heat source analysis is then applied to determine the fractions of heat flux flowing into the PCD asperities and their counterpart in contact sliding and to give rise to the average temperature rise. Kuo and Lin [8] proposed an analytical model to calculate the temperature rise in surface grinding. They reported that the difference between the calculated and experimental value are always less than 4%.

2.2 Numerical models

Principle-based models may not render to an analytical solution when the geometric boundary condition is complex. Numerical solutions to the thermal-error models are thus found. Newton–Raphson method was recommended in a thermal model for a roller bearing in [3]. Wang et al. [9] proposed an interesting method to evaluate heat generation and temperature field in high-speed ball bearing. The temperature field which is a two-dimensional model was obtained using APDL language of ANSYS. A finite element method (FEM) was employed by Sukaylo et al. [10] to simulate the thermal deformation in hard turning. Their experimental results showed that the deviation between the simulated and experimental values is within 10%. Zhao et al. [11] used the finite element (FE) model to simulate the spindle thermal deformation of a CNC turning center and to calculate its thermal error. They showed that the spindle thermal errors calculated from their model well agree with the experimental results.

Quite naturally, numerical solutions to thermal-error models tend to be time-consuming due to a large number of calculations. For example, today's finite element models may require meshes of thousands of nodes and several minutes to hours of calculation even with high-speed computers. However, for many application problems in manufacturing industries, there is not much demand on real-time thermal-error management at the time scale of, say, a few milliseconds.

Table 1 Overview of the studies in the literature in terms of the issues and schools

	Principle-based		Empirical-based	
	Literatures	Applications	Literatures	Applications
Issue 1	Refs. [3–9, 46] Refs. [47–50]	Rolling bears, gears, tool–chip interface, gears, etc.	–	
Issue 2	Refs. [10, 11] Refs. [50]	Only for simple structure	Refs. [1, 14–43, 51, 52]	Machine tools, etc.
Issue 3	–		–	

The numerical solution to principle-based thermal-error models is readily acceptable.

In the optimal design of physical structures in light of thermal-error management, there is a need to repeatedly mesh the system and compute the result. In this occasion, despite the off-line activity in nature, computational overhead can be intolerable. Techniques are available to develop semi-analytic solutions to the thermal-error models. The general idea of the semi-analytical solution is to choose certain parameters in the models and then to determine them by measurements or numerical calculations of behaviors of the system at a few time spans [12, 13].

3 Empirical-based models

The empirical-based model is based on the assumption that thermal errors can be considered as a function of some critical discrete temperature points on the machine. There are two theories to support this assumption: (1) the thermal deformation is associated with the temperature field and (2) a few critical temperature points on the machine can be employed to sufficiently predict the temperature field of the entire machine [14, 15]. Back to Fig. 1, the measured temperatures at a set of locations on the machine are denoted by $T_1, T_2, \cdots T_n$ (n is the total number of thermal sensors), and the measured displacements at a set of locations are denoted by E_1, $E_2, \cdots E_m$ (m is the total number of displacement sensors) [16]. Two sets of locations may not be the same.

A measurement procedure can be described with the machine tool system as an example. Initially, the coordinates of the machine tool and temperatures around the machine are measured as a reference under a machine cold-start condition by displacement sensors and thermal sensors, respectively. The machine is then run under air or real cutting conditions to warm-up to make the machine temperature field change. Continually interrupt the running in a series of time periods to record the sensor readings of coordinates and temperatures. Thermal errors and temperature changes can be obtained by the discrepancy between the on-going measurements (displacement and temperature) and the reference. A relationship between the thermal errors and temperature changes can be established based on these measured data using a proper method.

There are extensive studies on the method of establishing this relationship (i.e., empirical-based model). The empirical-based model can be further classified into two categories: static and dynamic. Static models are established on the basis of information at one isolated point of time or present time, while in dynamic models, not only information at the present time but also information at previous times is used. In the following, the static model and dynamic model are discussed.

3.1 Static models

Static models are based on the quasi-static assumption that thermal errors vary slowly in time and are only related to the structure of the machine tool. Static models are obtained off-line by capturing the empirical relationships between the inputs (e.g., temperature, spindle speed, etc.) and the thermal errors at present time. There are two well-known methods of the static model: multivariable regression analysis (MRA) and artificial neural network (ANN).

3.1.1 MRA models

MRA models take the following form of regression equations [17, 18]:

$$y_1 = a_{11}T_1 + a_{12}T_2 + \cdots + a_{1n}T_n + b_1 \tag{1}$$

where y_1 is a thermal-error component, $a_{11}, a_{12}, \cdots, a_{1n}$ are coefficients for temperature, $T_1, T_2, \cdots T_n$ are temperature inputs, b_1 is the constant for the thermal-error model. Eq. (1) can be extended to m thermal-error components in a matrix form as follows:

$$Y = AT \tag{2}$$

where,

$$Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \end{bmatrix}^T,$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix},$$

$$T = [\,T_1\, T_2 \cdots T_n\,]^T$$

Coefficient matrix A can be obtained by the least square method. Thus, the relationship between thermal errors and temperatures is established.

Chen et al. [16] proposed a MRA model to compensate thermal errors of a horizontal machining centre. Their experimental results showed that the spindle thermal growth or error was reduced from 196 to 8 µm. The MRA model was employed by Yang et al. [19] to form an error synthesis model which combines both the geometric and thermal errors of the NC twin-spindle lathe. Their experimental results indicated that the size variations of the work-piece could be reduced from 60 to 14 µm. The MRA model proposed by Pahk [18] was applied to compensate the thermal errors caused by the spindle and feed axis of a CNC machining centre. Their experimental results showed that the machine tool accuracy could be improved four to five times. Tseng et al. [20] proposed a non-linear MRA model for predicting thermal errors of a high-precision computer numerical control (CNC) lathe. Their experimental results showed that 40% of thermal errors can be reduced by their linear MRA model and 60% of thermal errors can be reduced by their non-linear model.

3.1.2 ANN models

ANN techniques can easily be used to correlate thermal errors and temperatures in a multiple-input and multiple-output form. As such, many thermal errors can be modeled with only one ANN model and many previous experiments [21, 22] suggested that ANN models can significantly compensate thermal errors since the interaction effects among outputs are well considered in one ANN model. Generally, in these ANN models, temperature measurements are used as inputs and thermal errors as outputs. The general concept of ANN model [17, 21–23] is illustrated in Fig. 4. In Fig. 4, $T_1, T_2, \cdots T_n$ denote n temperature inputs, $E_1, E_2, \cdots E_m$ denote m thermal errors, and W_{ij}, W_{jk} are the weights of the network.

An ANN is composed of the artificial neurons which process input signals and the network which connects the artificial neurons. The artificial neurons receive and process a set of input signals, and then present the outputs to other neurons [17]. Usually, a multiple-layer feedforward network is used to connect the artificial neurons, as shown in Fig. 4. In thermal-error modeling, the input layer is used to receive the temperature measurements and the output layer is used to store the machine tool thermal errors. Layers between the input layer and output layer are hidden layers which are to perform feature extraction and noise suppression between the temperature inputs and thermal errors. When temperature inputs are entered, the outputs of the network are calculated and compared with the corresponding target thermal error. The weights are then adjusted according to the discrepancy between the network output and the target which is fed back through the network. Such a training process does not stop until the error of the entire training set reaches an acceptable level. A thermal-error model can thus be obtained.

Chen et al. [21] employed an ANN model with 15 input nodes, 15 hidden nodes, and six output nodes to compensate thermal errors caused by the spindle and lead-screws on a vertical machining center. The network is trained by a training set with 540 training pairs and a new cutting condition which is not within the collected training pairs was used to test the prediction accuracy of the ANN model. Experimental results showed that 70–90% thermal errors can be reduced after compensation.

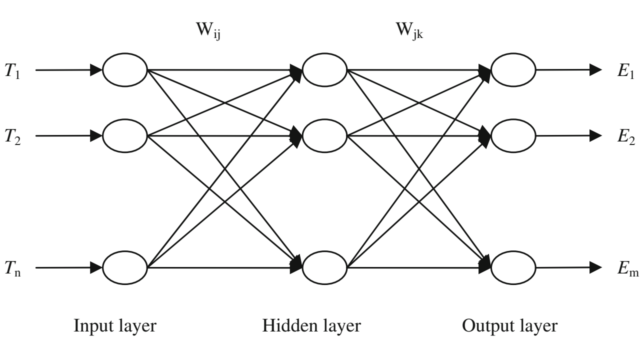


Fig. 4 The three-layer feedforward neural network

Various modified ANN models were proposed in the past decade. A cerebellar model articulation controller (CMAC) neural network was proposed by Yang et al. [24] for modeling spindle drift errors. CMAC can accept input vectors (temperature inputs) and produce output vectors (computed thermal errors) through the active association cell vector which was with weights of the network. Input vectors were mapped onto locations in the random address tables which were then mapped by hashing onto a much smaller set address table. This mapping scheme has the advantage of providing automatic interpolation between input vectors (that is, similar inputs produce similar outputs) [24]. Experiments on both a horizontal machining center and a CNC turning center showed that CMAC model had better robustness to temperature sensor locations and speed of learning than that of MRA model and multiple-layer feedforward network (MFN) model. A neural network model based on fuzzy artificial resonance theory (fuzzy ART-map) was proposed by Srinivasa [25], and then used to predict and compensate the tool point errors of a three-axis machining center [26]. The fuzzy ART-map was used to correlate the relationship between multivariable inputs and outputs through continually classifying groups of input and output data according to output of a fuzzy logic equation called vigilance. The network was then trained by using these classified data. When the network was used for thermal-error prediction, inputs are first compared with previous classified input data, and then the selected group of data was used as input to find corresponding outputs. It was reported that the prediction accuracy of the model was limited by the similarity of the training set to that of the future inputs [26]. Therefore, ART-map model can be used in a relatively stable working condition and it is necessary to establish several ART-map models for usage in different machining conditions.

A radial basis function (RBF) neural network was proposed by Li [27, 28]. The RBF neural network has a similar form to a multi-layer perceptron (MLP) and they both have a multi-layer, feedforward network. However, the difference is that the hidden units in the RBF neural network have the "radial basis function" which is a statistical transformation based on Gaussian distribution and these hidden units with basis function make the process of training the RBF network easier than that of MLP. The test results showed that RBF model had good prediction accuracy and 85% of thermal errors can be compensated.

Wang [29] proposed a hierarchy-genetic-algorithm (HGA) trained neural network to map the temperature change to thermal drift of machine tool. In another paper of Wang [30], he proposed a model combining GM (1,*m*) scheme and an adaptive network-based fuzzy inference system (ANFIS). A hybrid learning algorithm including forward pass and backward pass was developed to fast identify parameters of the model. Steepest descent (SD) and the least square method were used in backward pass and forward pass, respectively. Experimental results showed that thermal error could be less than 9.2 µm under the real cutting conditions. Tseng et al. [31] proposed a neural-fuzzy model to improve the accuracy of a CNC machining center. The results showed that thermal errors of the CNC machining center can be reduced from 80 to 3 µm. And comparing with the multiple regression analysis (MRA) model, the accuracy of the CNC machining center using the neural-fuzzy thermal-error model can be improved from 10 to 3 µm. Ramesh et al. [32] presented a hybrid support vector machines (SVM)-Bayesian network (BN) model. BN model was designed for classifying the error into groups depending on different operating conditions and SVM model was used for mapping temperature and thermal errors.

3.2 Dynamic models

Philosophically, dynamic models have a better prediction accuracy and robustness than that of static models [33, 34]. The reason is that the mapping from temperatures to thermal errors in the static model is considered as one-to-one in steady state. However, the fact is that temperature field of machines is not always in a steady state. Studies [33, 34] have shown that when the temperature field changes quickly, the static model is difficult to describe the non-unique relationship between the temperatures and thermal errors because the thermal deformation not only depends on the current temperature but also the previous thermal states. As such, dynamic models have better capability of capturing the dynamic behavior (e.g., non-linearity of the thermal-elastic process) of thermal deformation. Dynamic models are classified into common dynamic models and adaptive dynamic models in this paper.

3.2.1 Common dynamic models

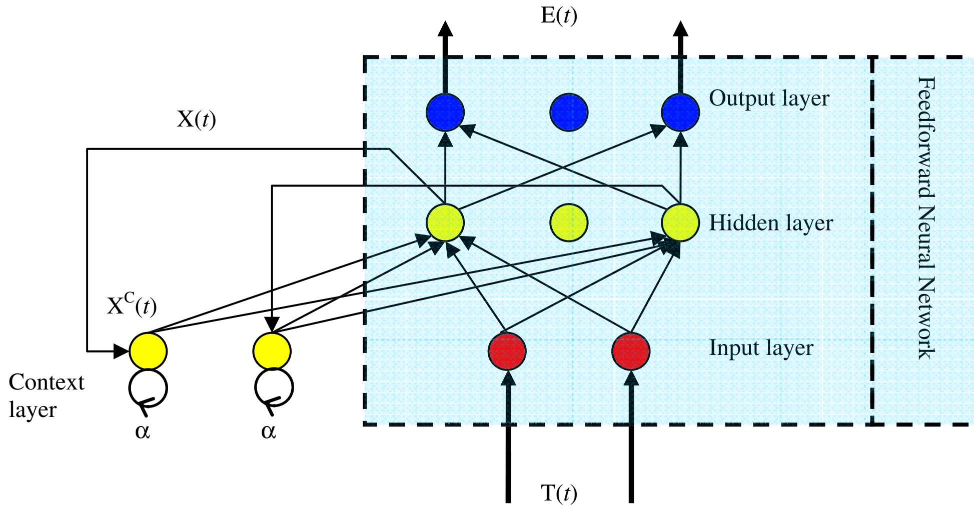
A recurrent neural network (RNN) model [33] (see Fig. 5) is taken as an example to illustrate the working principle of dynamic models in order to better distinguish the difference from the static model which is shown in Fig. 4. In Fig. 5, $T(t)$ are temperature measurements, $E(t)$ are thermal deformation measurements, $X(t)$ are hidden unites, $X^c(t)$ are context units, α $(0<\alpha<1)$ is the coefficient for self-connection. As shown in Fig. 5, a RNN model is established by adding a context layer into the feedforward neural network. Some previous information of the hidden layer can then be memorized by the context layer and saved there as inputs with current temperature inputs together for the current training step of the network. The RNN model can be described by non-linear, state space, difference equation (3).

$$\begin{aligned}
&X(t) = F\left\{W^{xc}X^c(t), W^{xT}T(t)\right\},\\
&X^c(t) = \alpha X^c(t-1) + X(t-1) = X(t-1) + \alpha X(t-2)\\
&\qquad + \alpha^2 X(t-3) + \cdots,\\
&E(t) = W^{Ex}X(t)
\end{aligned} \tag{3}$$

where W^{xc}, W^{xT}, W^{Ex} are weight matrixes, and $F(\cdot)$ is a non-linear vector function.

By integrating the output difference $E(t)$, an integrated recurrent neural network (IRNN) model was obtained by Yang and Ni [33]. The same data from spindle thermal-error experiments were used to compare the performance of the IRNN, RNN, multi-layer feedforward neural network (MFN) and MRA models. Test results showed that all these models had good accuracy within modeling data. However, dynamic models (IRNN, RNN) had better accuracy than that of static

Fig. 5 Recurrent neural network (RNN) [33]



models (MFN, MRA) under new working conditions which suggests dynamic models have better robustness as well.

Wang et al. [14] proposed a thermal-error model based on the grey system theory which is a kind of dynamic model described by a differential equation, as shown in equation (4).

$$\frac{d^n X_1^{(\xi)}(k)}{dt^n} + a_1 \frac{d^{n-1} X_1^{(\xi)}(k)}{dt^{n-1}} + \cdots + a_n X_1^{(\xi)}(k) = b_1 X_2^{(\xi)} + b_2 X_3^{(\xi)}(k) + \cdots + b_{m-1} X_m^{(\xi)}(k) \quad (4)$$

where n is the order of the differential equation, m is the number of the types of the data, and ξ is the number of transformations. Eq. (4) is called a grey dynamic model, commonly denoted as GM (n, m). The coefficients of the model are estimated by the least square method. Usually, X_1 is chosen as thermal error of machine tools, and $X_2, X_3, \cdots, X_m$ as temperature measurements.

The dynamic characteristic of GM model results from a data process method called the accumulated generating operation (AGO). In fact, much of the irregular data obtained is difficult to be expressed directly by a mathematical equation. In the AGO, a set of known non-negative series

$$\left\{X_i^{(0)}(k)\right\}, i = 1, 2, \ldots, m, k = 1, 2, \ldots, n$$

can be transformed as follows:

$$x_i^{(\xi)}(k) = \sum_{k=1}^{n} X_i^{(\xi-1)}(k) \quad (5)$$

By using the AGO transformation, all of the previous data can be used to reduce the randomness of the data and to easily find the trend of the new series [14]. Usually, GM (1,m) is employed to model thermal errors of machine tools, which has the following form:

$$\frac{dX_1^{(1)}(k)}{dt} + a_1 X_1^{(1)} = b_1 X_2^{(1)}(k) + b_2 X_3^{(1)}(k) + \cdots + b_{m-1} X_m^{(1)}(k) \quad (6)$$

The coefficients of the model are denoted as follows:

$$\hat{a} = [a_1, b_1, b_2, \cdots, b_{m-1}]^T$$

which can be obtained by the least square method as follows:

$$\hat{a} = \left[B^{\mathrm{T}} \cdot B\right]^{-1} \cdot B^{\mathrm{T}} \cdot Y_k$$

where

$$B = \begin{bmatrix} -\frac{1}{2}\left(X_1^{(1)}(2) + X_1^{(1)}\right) & X_2^{(1)}(2) & \cdots & X_m^{(1)}(2) \\ -\frac{1}{2}\left(X_1^{(1)}(3) + X_1^{(1)}(2)\right) & X_2^{(1)}(3) & \cdots & X_m^{(1)}(3) \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{2}\left(X_1^{(1)}(n) + X_1^{(1)}(n-1)\right) & X_2^{(1)}(n) & \cdots & X_m^{(1)}(n) \end{bmatrix}$$

$$Y_k = \left[X_1^{(0)}(2), X_1^{(0)}(3), \ldots, X_1^{(0)}(n)\right]$$

GM(1, 4) model was used in Wang's study and experimental results showed that average 75% of the thermal drift of the machining centre can be compensated. Total data GM(1,1) model, new information GM(1,1) model and metabolic GM (1,1) model for thermal-error compensation were studied by Li et al. [35]. Their difference as opposed to Wang's model is that they are modeled by different data sequence and details can be referred in [35]. Experimental results showed that both new information GM(1,1) model and metabolic GM (1,1) model had a better prediction accuracy than that of total data GM(1,1) since the old data used in total dada GM(1,1) could not well reflect the system behavior compared with the former models.

Lin et al. [36] proposed a complex multivariable regression analysis model to predict spindle thermal errors, which is essentially a dynamic model since previous thermal errors and previous spindle speed were as inputs to predict current thermal errors in the model. The complex MRA model is essentially a set of different common MRA model established according to different spindle rotational speed and a suitable MRA model can be selected from them to predict thermal errors according to relatively close spindle rotational speed under real machining conditions. The complex MRA model can be written into the following equation:

$$e(t) = a_0(t) + \sum_{i=1}^{n} a_i(t) e(t-i) + \sum_{i=0}^{m} b_i(t-i) N^2(t-i) \quad (7)$$

Where, i is unit timeslot, $e(t-i)$ is the spindle thermal error when it is in the time slot of $(t-i)$, $N^2(t-i)$ is the square of spindle rotational speed when it is in the time slot of $(t-i)$, $a_i(t)$ is the regression coefficient related to the parameters of spindle thermal error in the time slot of (t), $b_i(t-i)$ is the regression coefficient related to the parameters of spindle rotational speed in the time of $(t-i)$, n, m is the number of $e(t-i)$, and the number of $N^2(t-i)$ parameters deciding how many previous measurements are used to predict thermal error. The selection of n and m can be based on the contribution of the previous thermal error and spindle speed to the current thermal error. Experiments indicated the variable of $a_i(t)$ and $b_i(t-i)$ whose weight percentage is less than 5% could be neglected. The complex MRA model is suitable to use when the spindle speed varies frequently in a large range and experimental results showed that the spindle thermal error could be reduced from 62 to 4.62 μm when the spindle rotational speed varied from 4,000 to 10,000 rpm.

3.2.2 Adaptive dynamic models

The non-adaptive models could not have good prediction accuracy for two cases [37]: (1) for the small-batch production, where the manufacturing processes change frequently, and (2) for the long-term application, where some parameters of machine tools will vary slowly. The key reason for this observation is that non-adaptive models are all off-line pre-established and then used to on-line predict thermal errors, they cannot be updated according to continuous changes of manufacturing conditions. The nature of this problem is that real machining conditions are not "identical" to the experimental conditions used for deriving the model. To overcome these shortcomings, adaptive dynamic models are proposed.

Barakat et al. [38] proposed an adaptive model for compensation of quasi-static errors of a coordinate measuring machine (CMM). First, the CMM errors were modeled at a constant reference thermal state. The effect of ambient temperature change was then incorporated through non-linear regression analysis. The model thus can update itself according to ambient temperature change. However, the change of parameters of the machine due to different operating conditions cannot be accounted for with this model.

Yang et al. [37] proposed an adaptive model which can be refined according to continuous changes of operation status. The general concept is shown in Fig. 6. The dynamic model, which is also called integrated time-series (ITS) model, was pre-established off-line and used to predict thermal errors on-line according to measurement data under real machining conditions. A recursively updated model was integrated in the entire thermal-error compensation system and it was used to update ITS model in a period of time according to the measurement data under current machining conditions. When the model is not updating, at time instants $t \neq N \cdot T_S \cdot k$, $(k = 1, 2, 3, \ldots)$, the latest previous updated ITS will be used to predict thermal errors according to input data.

The approach proposed by Wang et al. [14] may also be considered as an adaptive model. A GM model based on grey system theory is established on-line by using the measurement data of the first 30 min after the machine have been started. The experimental results showed that on average about 90% of thermal error can be compensated for using the on-line GM model, while about 75% of the error can be compensated by using the off-line pre-established GM model. The similar approach with Wang's can be found in earlier research [25]. A fuzzy ARTMAP was trained on-line using the data collected in the first 6 h after the machine was started. However, it cannot be employed in industry due to its too-long-training time.

The difference between Wang's model and Yang's model is as follows: (1) Wang's model is established by using the data of only a short time after the machine starts. As such, Wang's model is more suitably employed in a short term and relatively stable working conditions, while Yang's model can be updated over a period and thus it is suitably used in long-term working. (2) Yang's updating process is obtained by recursively interrupting machining operations for data acquisition; while once Wang's model is established, it does not need to recursively interrupt the machining process. (3) Wang's model is established by using the data from air cutting conditions after machine starts a short time, while Yang's model is obtained by using the data from real cutting conditions.

4 Challenges of the principle-based models

The heat source model and the heat transfer model are vital to any principle-based models. Heat source models are used to calculating how much heat is generated. Heat transfer models are used to study how the heat flux flows in a system. The most challenging part of building a principle-based model is to accurately build a heat source model and a heat transfer model.

The challenge of building heat source models for a complex physical system is that every single heat source model is dependent on a specific friction model which is usually complex. The friction model is complex due to the following reasons: (1) friction force is much more complex than the common model such as $F_r = \mu \times N$ (F_r is the friction force, μ is called friction coefficient which is a constant, and N is the force applied on the interface between two objects in consideration). In fact, the friction force is a function of velocity, load, displacement, temperature, and the properties of the interface (e.g., roughness, hardness, material, etc.), many of which are time dependent (e.g., velocity, displace-

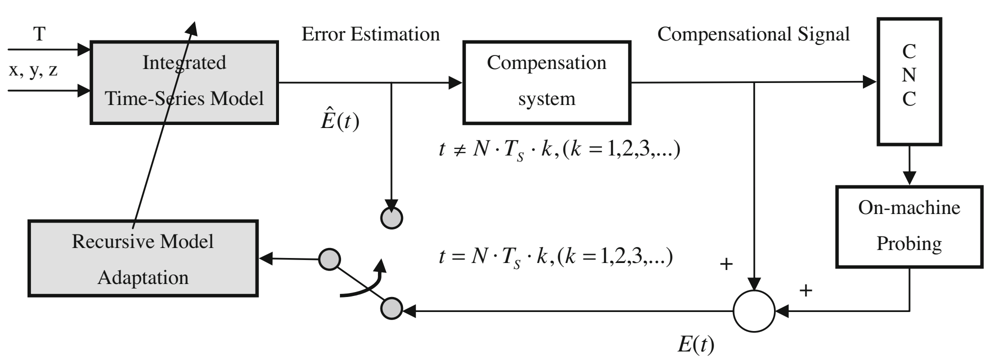


Fig. 6 Recursive model adaptation with multi-sample horizon (N is sampling horizon and T_S is sampling interval) [37]

ment, temperature, etc.) [4]. (2) In a friction model, say for the rolling bearing system, the motion behavior of two bodies that are interacting is a combination of rolling, sliding, and spinning motions, which is much more complex than is indicated by pure rolling we usually assume. (3) The interaction or contact model cannot be represented by a simple analytical expression in some cases, e.g., rolling element-raceway. The contact model usually depends on the micro-geometry of the contacting bodies and the properties of the lubricant or oxidation (if any). The microcontacts may deform plastically as well as elastically with the result that the microgeometry varies with time.

There are usually a number of heat sources in a complex physical system. For example, in a machine tool (see Fig. 1), there are five main heat sources. In each of them, e.g., in a rolling bearing, all of the following factors contribute to the heat generation and they are [4]: (a) elastic hysteresis in rolling, (b) sliding in rolling element-raceway contacts due to a geometry of contacting surfaces, (c) sliding due to the deformation of contacting elements, (d) sliding between the cage and rolling elements and, for a land-riding cage, sliding between the cage and bearing rings, (e) viscous drag of the lubricant on the rolling elements and cage, (f) sliding between roller ends and inner and/or outer ring flanges, and (g) seal friction.

Building heat transfer models along with the principle-based model is a challenge as well. There exists three fundamental heat transfer mechanisms, (1) heat conduction, which is the conduction of heat within solid structures, (2) heat convection, which is the convection of heat from solid structures to fluids in motion (or apparently at rest), and (3) heat radiation, which is the radiation of heat between masses separated by space. Heat convection is the most difficult one among the three mechanisms. For example, heat convection from a surface can generally be described by the equation of $H_v = h_v S(T_1 - T_2)$, in which H is heat flow; S is area normal to heat flow; T is temperature; h_v is the film coefficient of heat transfer [4]. h_v is a function of surface temperatures, surface dimensions and attitude in the case of dry friction. In the case of lubricated friction, fluid viscosity and density, fluid thermal conductivity, and fluid velocity adjacent to the interacting surface need to be considered for h_v as well. Many of these properties are more temperature dependent, which makes a heat transfer model much more complex.

5 Challenges of the empirical-based models

To train an empirical-based model requires a large number of data-pairs of temperature and thermal error along with the patterns of working conditions. With the current data acquisition technology, the process of getting these data can last about tens of hours. Optimal data acquisition and processing for high efficiency, accuracy, and robustness is the most important challenge in building an empirical-based model. This challenge can be further divided into six specific challenges: (1) thermal-sensor selection, (2) system variables selection, (3) real cutting and air cutting, (4) thermal-error measurements, (5) interruption or non-interruption during measurements, and (6) safety and reliability.

5.1 Thermal-sensor selection

Studies have shown that a proper selection of thermal sensors and their locations is vital to the prediction accuracy and robustness of models. A poor selection of location and use of too little a number of thermal sensors will result in poor prediction accuracy. However, too large a number of thermal sensors may lead to poor robustness of a model because there may be too much noise in the model (notice: each sensor brings noise to the model while bringing useful information). Experimental results and industrial practices have shown that the location and number of thermal sensors can be optimized by engineering judgment, multiple regression analysis, grey system theory, etc [29]. Yang et al. [39] proposed a model combing MRA and grouping approach to compensate thermal errors of an INDEX-G200 turning center. With their approval, the number of temperature variables was reduced and the best combination of temperature variables was selected by the grouping approach. GM (1,m) model of the grey system theory is applied to minimize the numbers of temperature and select the suitable sensors in [28, 29, 39, 40]. A simple free-expansion-based distortion model combined with thermal imaging techniques was proposed to predict the spindle thermal error [41]. A high specification thermal imaging camera was used to record a sequence of pictures of temperature distributions across the whole machine structure. These pictures were then analyzed by MATLAB to find out the structural elements significantly affected by temperature change. According to the results of such an analysis, thermocouples were installed at the proper location to monitor thermal status of the machine. Chen et al. proposed a stepwise regression analysis procedure for selecting temperature variables both in their MRA model [17] and ANN model [42]. In Yang's model [43], the number of thermal sensors was reduced from 16 to four by using thermal-error mode analysis.

5.2 System variables selection

In most of thermal-error models of machine tools, temperatures of critical discrete points are usually as inputs to predict thermal errors. However, spindle speed, feed speed, and other parameters of the machine may also need to be taken into account, because they are responsible for major heat generation—a source of thermal errors. In Lin's model [36], previous thermal errors and previous spindle speed were as inputs in a dynamic MRA model. In Yang's model [43], the produced-part dimensions and temperature rise were used as inputs. By using these parameters, a modeling process becomes easier since a large number of thermal sensors are avoided. However, studies showed that use of spindle speed and feed speed only as inputs of the model makes it difficult to describe the phenomenon of time-lag between thermal deformation and temperature field [43].

5.3 Real cutting and air cutting

Chen [22, 42] studied the effect of cutting conditions on the thermal errors, e.g., real cutting, air cutting, application of coolants, spindle speed, feed speed, cutting paths, workpiece materials, etc. The experimental results showed that the prediction accuracy of a model established using the data from air cutting conditions is unacceptable when the model is used under real cutting conditions. He thus proposed a hybrid ANN model developed using both the data form air cutting and real cutting conditions.

5.4 Thermal-error measurements

Traditional thermal-error (e.g., displacement or angular) measurement procedures are based on the simultaneous measurement of a machine tool by external measurement devices only when the machine stops. The accuracy of such measurements is improved by arithmetic average obtained by repeating measurement many times using the same sensor, which means that the number of sensors is generally equal to the number of degree of freedom (DOF). Repeating measurements definitely is too time-consuming for practical applications. Redundant measurement method was proposed to overcome this shortcoming. The redundant measurement is based on the idea that the same effect of improved measurement accuracy can be realized by the repetition in space by simultaneous use of more sensors [44], which means the number of measured variables is greater than the number of DOF, e.g., the number of DOF plus one, two or more. The work of Valasek et al. [45] showed that the accuracy of the redundant method was significantly better than that of the non-redundant method.

5.5 Interruption or non-interruption during measurements

During empirical-based modeling, the machine system has to be stopped by human interruption in order to measure thermal errors. While in redundant calibration (or called self-calibration) [44], no external measurement devices are needed since each drive is equipped with a position sensor. If the method of redundant calibration can be used in the future thermal-error modeling process, the empirical-based modeling process will be automatically and continuously done by computers, and no interruption is needed, which makes online real-time model updating practical.

5.6 Safety and reliability

The safety and reliability of a system should not be compromised in the empirical-based models. This is a serious problem since the thermal-error compensation models are incorporated in the controller of the system. The thermal-sensor failure or random noises are likely to generate a wrong control signal, leading to a disaster result (e.g., causing machine tool break or endangering people's lives in medical devices). In 1996, Yang et al. [24] developed a sensor failure detection algorithm. The concept is that once the reading of some sensor is far beyond a tolerable range, it will be diagnosed as a defective sensor and the sensor readings will be replaced by other related sensors mounted on the same machine. This approach needs to establish the relationship with other sensors for each sensor, which is time-consuming and the related sensors should be carefully selected because not all sensors are related with each other. However, almost no other work on safety and reliability due to the use of thermal-error model was found in the last decade.

6 Conclusions and future directions

In this paper, various techniques for modeling thermal errors of complex physical systems in the last decade were reviewed and discussed. The main conclusions can be drawn from these discussions.

(1) Principle-based models are more suitable to find the relation between the friction-induced heat and temperature field. The empirical-based models are more suitable for building the relation between the temperature change and deformation.
(2) The coupling effect between thermal deformation and friction is not well considered in either principle-based models or empirical-based models. On a general note, the coupling effect can be significant on the performance of micro-motion systems, which is found from our previous work on a PZT-SS actuator [2] (see Fig. 2).
(3) A potential compromise of safety and reliability due to a potential failure of thermal sensors has not been sufficiently studied in the literature.

The state of arts of the thermal-error modeling for complex physical systems can be advanced further along the following directions.

First, the coupling effect between thermal deformation and friction must be considered, especially for the micro-motion systems. One effort may be taken on understanding of the friction behavior for a macro system which performs micro-motions. It is noted that the macro system implies a larger inertia which inherently hinders the realization of small increments in movement (i.e., small resolutions of motions) as well as demands high-end measurement instruments, while the micro-motion inherently hinders the relative movement between two contacting objects or blurring the relative motion (i.e., a physical phenomenon called "stick") as well as demands low-end measurement instruments. Furthermore, when the relative motion between two contacting objects gets to the nano-scale, the so-called size effect may challenge the contemporary friction models which are under an "implicit" assumption that the relative motion is at the macro level. Another important effort may be taken to develop a procedure to establish a "simultaneous" calculation of friction variables and deformation variables as they are, in essence, expressed at the same time, the coupling effect as we called. The general idea to develop this procedure is to integrate contact mechanics for friction and heat transfer for deformation by footing on the asperities (micro features) on the surfaces of two contacting objects.

Second, there is a promise on study of integrating principle-based models and empirical-based models. Both models have advantages and disadvantages; an integrated one may be complementing to each other, resulting in a more reliable, robust, and resilient model as well as thermal-error system. For instance, when sensors fail, principle-based models may be put in place to "derive" thermal errors; as such, a thermal-error compensation system may still work. Another scenario is that the "derived" knowledge from the principle-based models can be used as a reference to screen out noises coming from the sensor measurements—a way to improve the empirical-based models for more accurate prediction of thermal errors.

Acknowledgements The corresponding author would like to thank a partial financial support from NSERC in a Discovery grant and East China University of Science and Technology (ECUST) in a Chair Professorship Award.

References

1. Chen J-S (1996) Neural network-based modeling and error compensation of thermally-induced spindle errors. Int J Adv Manuf Technol 12:303–308
2. Li JW, Yang GS, Zhang WJ, Tu SD, Chen XB (2008) Thermal effect on piezoelectric stick-slip actuator systems. Rev Sci Instrum 79:046108
3. Harris TA (2001) Rolling bearing analysis. Wiley, New Jersey
4. Harris TA, Kotzalas MN (2007) Advanced concepts of bearing technology. CRC, Boca Raton
5. Komanduri R, Hou ZB (2001) Thermal modeling of the metal cutting process—part II: temperature rise distribution due to frictional heat source at the tool–chip interface. Int J Mech Sci 43:57–88
6. Karpat Y, Özel T (2006) Predictive analytical and thermal modeling of orthogonal cutting process—part I: predictions of tool forces, stresses, and temperature distributions. J Manuf Sci Eng 128:435–443
7. Chen Y, Zhang LC, Arsecularatne JA, Montross C (2006) Polishing of polycrystalline diamond by the technique of dynamic friction, part 1: prediction of the interface temperature rise. Int J Mach Tools Manuf 46:580–587
8. Kuo WL, Lin JF (2006) General temperature rise solution for a moving plane heat source problem in surface grinding. Int J Adv Manuf Technol 31:268–277
9. Wang LQ, Chen GC, Gu L, Zheng DZ (2007) Operating temperature in high-speed ball bearing, Proc. IMechE Vol. 221 Part C: J. Mechanical Engineering Science 353–359
10. Sukaylo VA, Kaldos A, Krukovsky G, Lierath F, Emmer T, Pieper H, Kundrak J, Bana V (2004) Development and verification of a computer model for thermal distortions in hard turning. J Mater Process Technol 155–156:1821–1827
11. Haitao Z, Jianguo Y, Jinhua S (2007) Simulation of thermal behavior of a CNC machine tool spindle. Int J Mach Tools Manuf 47:1003–1010
12. Bwolin NK, Chen XB, Zhang WJ (2007) A method for modeling time-dependant mechanical properties of tissue scaffolds. ICAM 2007. Harbin, China, August
13. Chen XB, Schoenau G, Zhang WJ (2005) Modeling and control of time-pressure dispensing processes for Surface Mount Technology. IEEE/ASME Trans Mechatron 10(3):326–334, June
14. Wang Y, Zhang G, Moon KS, Sutherland JW (1998) Compensation for the thermal error of a multi-axis machining center. J Mater Process Technol 75:45–53
15. Venugopal R, Barash M (1986) Thermal effects on the accuracy of numerically controlled machine tools. Ann CIRP 35(1):255–258
16. Srinivasa N, Ziegert JC, Mize CD (1996) Spindle thermal drift measurement using the laser ball bar. Precis Eng 18:118–128
17. Chen JS, Yuan J, Ni J (1996) Thermal error modeling for real-time error compensation. Int J Adv Manuf Technol 12:266–275
18. Pahk HJ, Lee SW (2002) Thermal error measurement and real time compensation system for CNC machine tools incorporating the spindle thermal error and the feed axis thermal error. Int J Adv Manuf Technol 20:487–494
19. Yang JG, Ren YQ, Du ZC (2002) An application of real-time error compensation on an NC twin-spindle lathe. J Mater Process Technol 129:474–479
20. Tseng PC, Ho JL (2002) A study of high-precision CNC lathe thermal errors and compensation. Int J Adv Manuf Technol 19:850–858
21. Chen JS, Chiou G (1995) Quick testing and modeling of thermally-induced errors of CNC machine tools. Int J Mach Tools Manufact 35(7):1063–1074
22. Chen J-S (1996) A study of thermally induced machine tool errors in real cutting conditions. Int J Mach Tools Manufact 36(12):1401–1411
23. Chen J-S, Ling CC (1996) Improving the machine accuracy through machine tool metrology and error correction. Int J Adv Manuf Technol 11:198–205
24. Yang S, Yuan J, Ni J (1996) Improvement of thermal error modeling and compensation on machine tools by CMAC neural network. Int J Mach Tools Manufact 36(4):527–537
25. Srinivasa N, Ziegert JC (1994) An application of fuzzy ARTMAP neural network to real-time learning and prediction of time-variant machine tool error maps. IEEE 1725–1730
26. Mize CD, Ziegert JC (2000) Neural network thermal error compensation of a machining center. Precis Eng 24:338–346
27. Li X (2001) Real-time prediction of workpiece errors for a CNC turning centre, part 1. Measurement and identification. Int J Adv Manuf Technol 17:649–653
28. Li X (2001) Real-time prediction of workpiece errors for a CNC turning centre, part 2. Modeling and estimation of thermally induced errors. Int J Adv Manuf Technol 17:654–658

29. Wang K-C 2006 Thermal error modeling of a machining center using grey system theory and HGA-trained neural network. IEEE
30. Wang K-C (2006) Thermal error modeling of a machining center using grey system theory and adaptive network-based fuzzy inference system. IEEE
31. Tseng PC, Chen SL (2002) The neural-fuzzy thermal error compensation controller on CNC machining center. JSME Int J 45 (2):470–478
32. Ramesh R, Mannan MA, Poo AN, Keerthi SS (2003) Thermal error measurement and modeling in machine tools. Part II. Hybrid Bayesian network-support vector machine model. Int J Mach Tools Manuf 43:405–419
33. Yang H, Ni J (2005) Dynamic neural network modeling for nonlinear, nonstationary machine tool thermally induced error. Int J Mach Tools Manuf 45:455–465
34. Yang H, Ni J (2003) Dynamic modeling for machine tool thermal error compensation. J Manuf Sci Eng 125:245–254
35. Li Y, Yang J, Zhang H, Tong H (2006) Application of grey system model to thermal error modeling on machine tools. International Federation for Information Processing 207:511–518
36. Lin Z-C, Chang J-S (2006) The building of spindle thermal displacement model of high speed machine center. Int J Adv Manuf Technol
37. Yang H, Ni J (2005) Adaptive model estimation of machine-tool thermal errors based on recursive dynamic modeling strategy. Int J Mach Tools Manuf 45:1–11
38. Barakat NA, Spence AD, Elbestawi MA (2000) Adaptive compensation of quasi-static errors for an intrinsic machine. Int J Mach Tools Manuf 40:2267–2291
39. Yang JG, Ren YQ, Liu GL, Zhao HT, Dou XL, Chen WZ, He SW (2005) Testing, variable selecting and modeling of thermal errors on an INDEX-G200 turning center. Int J Adv Manuf Technol 26:814–818
40. Li YX, Yang JG, Gelvis T, Li YY (2008) Optimization of measuring points for machine tool thermal error based on grey system theory. Int J Adv Manuf Technol 35:745–750
41. Postlethwaite SR, Allen JP, Ford DG (1998) The use of thermal imaging, temperature and distortion models for machine tool thermal error reduction. Proc Instn Mech Engrs 212:671–679
42. Chen J-S (1997) Fast calibration and modeling of thermally-induced machine tool errors in real machining. Int J Mach Tools Manufact 37(2):159–169
43. Yang J, Yuan J, Ni J (1999) Thermal error mode analysis and robust modeling for error compensation on a CNC turning center. Int J Mach Tools Manuf 39:1367–1381
44. Valášek M, Šika Z, Štembera J (2004) PKM Calibration by redundant measurements. In: Neugebauer R (ed) Parallel kinematic machines in research and practice. Wissenschaftliche scripten, Zwickau, pp 739–744
45. Valášek M, Šika Z, Hamrle V (2007) Dexterity to calibrability of parallel kinematical structures. In Proceedings of 12th World Congress in Mechanism and Machine Science [CD-ROM]. Comité Français pour la Promotion de la Science des Mécanismes et des Machines, Besançon, pp 1–6
46. Horejš O (2007) Thermo-mechanical model of ball screw with non-steady heat sources, Thermal Issues in Emerging Technologies, ThETA 1, Cairo, Egypt, Jan 3–6th:133–137
47. Stachowiak GW, Batchelor AW (2006) Engineering tribology. Elsevier Butterworth-Heinemann, Oxford, ISBN-13: 9780750678360
48. Stolarski TA (1999) Tribology in machine design. Elsevier Butterworth-Heinemann, Oxford, ISBN-13: 9780750636230
49. Lee KY, Kang JW (2003) Boundary element analysis for thermo-mechanical interface stress in the thin-layered media under sliding contact. Eng Anal Bound Elem 27:583–587
50. Wang W-Z, Hu Y-Z, Wang H, Liu Y-C (2006) Effect of thermo-elastic deformation on lubricated point contacts. Proceedings of the Institution of Mechanical Engineers, Part J (Journal of Engineering Tribology), v 220, n J6, Sept., p 523–34
51. Postlethwaite SR, Allen JP, Ford DG (1999) Machine tool thermal error reduction—an appraisal. Pro Instn Mech Engrs 213:1–9
52. Yuan J, Ni J (1998) The real-time error compensation technique for CNC machining systems. Mechatronics 8:359–380