machines

MDPI

Article

# The Thermal Error Estimation of the Machine Tool Spindle Based on Machine Learning

**Yu-Cheng Chiu , Po-Hsun Wang and Yuh-Chung Hu ***

Department of Mechanical and Electromechanical Engineering, National ILan University, Yilan 26041, Taiwan; b0722061@ms.niu.edu.tw (Y.-C.C.); b0722097@ms.niu.edu.tw (P.-H.W.)
* Correspondence: ychu@niu.edu.tw; Tel.: +886-3-9317-469

**Abstract:** Thermal error is one of the main sources of machining error of machine tools. Being a key component of the machine tool, the spindle will generate a lot of heat in the machining process and thereby result in a thermal error of itself. Real-time measurement of thermal error will interrupt the machining process. Therefore, this paper presents a machine learning model to estimate the thermal error of the spindle from its feature temperature points. The authors adopt random forests and Gaussian process regression to model the thermal error of the spindle and Pearson correlation coefficients to select the feature temperature points. The result shows that random forests collocating with Pearson correlation coefficients is an efficient and accurate method for the thermal error modeling of the spindle. Its accuracy reaches to 90.49% based on only four feature temperature points—two points at the bearings and two points at the inner housing—and the spindle speed. If the accuracy requirement is not very onerous, one can select just the temperature points of the bearings, because the installation of temperature sensors at these positions is acceptable for the spindle or machine tool manufacture, while the other positions may interfere with the cooling pipeline of the spindle.

**Keywords:** Gaussian process regression; machine learning; machine tool spindle; Pearson correlation coefficient; random forest; thermal error



**Citation:** Chiu, Y.-C.; Wang, P.-H.; Hu, Y.-C. The Thermal Error Estimation of the Machine Tool Spindle Based on Machine Learning. *Machines* **2021**, *9*, 184. https://doi.org/10.3390/machines9090184

Academic Editor: Francisco J. G. Silva

Received: 18 July 2021
Accepted: 27 August 2021
Published: 30 August 2021



## 1. Introduction

The machining errors of machine tools mainly include geometrical errors, thermal errors, errors caused by cutting-force, fixture-dependent errors, etc. [1–4]. Bryan [1], Ramesh [2], and Li et al. [3] mentioned that thermal errors account for 40% to 70% of the total machining errors. The heat sources in the machining process of machine tools include two categories: internal heat sources, and external heat sources [2,4]. Internal heat sources include the heat produced by cutting, the heat induced by the friction of bearings, spindle, gearbox, and motion guides; the heat generated in the motor; and the heating or cooling effects produced by the cooling system. The external heat sources include ambient temperature changes, solar radiant heat, and human body radiant heat. Being a key component of the machine tool, the spindle will generate a lot of heat during the machining process and thereby result in a thermal deformation/error of itself. The main heat sources of the spindle are the friction of bearings and the heat generated by the motor. Takabi [5] mentioned that the heat-generation of the bearings is affected by the bearing type, the torque and preload applied on the bearing, the lubrication, etc. The heat transfer of the spindle involves complex physical phenomena such as conduction, convection, and radiation because the spindle is composed of many components of different materials and it rotates at high speeds during machining process.

The design and manufacture of machine tools cannot completely solve the aforementioned diverse machining errors appearing in machining process [1,2,4]. A feasible solution is error compensation in the machining process. For compensation of thermal errors during the machining process, the machine tool should be capable of knowing the real-time thermal errors. However, the thermal errors of the machine tool cannot be measured at any time during the machining process, because this will interrupt the machining process and thereby reduce the production efficiency. A feasible way to realize the thermal error compensation in the machining process is by estimating the thermal error from the temperature data of the machine tool, because measuring temperature is much easier than measuring thermal deformation, and this will not interrupt the machining process. However, the relationship between the temperature field of the machine tool and its thermal deformation/error must be known. This problem has to be solved prior to thermal error compensation.

The thermal error of the spindle accounts for the majority of the thermal errors of the machine tool, because the spindle will generate a lot of heat in the machining process. This heat mainly comes from the friction of bearings, the cutting heat of the tool, and the heat generation of the motor, etc., which results in the thermal error of the spindle and thereby changes the relative position of the cutting tool and the workpiece. Therefore, this paper intends to establish a method to estimate in real time the thermal error of the spindle in the machining process based on the spindle's temperature data and rotational speed. To achieve this goal, one has to solve two problems: one is the temperature field of the spindle under diverse machining parameters, and the other is the relationship between the temperature field of the spindle and its thermal error.

In the absence of cooling oil/water lines, the heat convection between the shaft of the spindle and the atmosphere around the shaft can be regarded as the heat dissipation mechanism of the shaft. Kendoush [6] derived an analytical solution of forced heat convection between a rotating shaft and the surrounding atmosphere, which was useful to evaluate the heat dissipation of the spindle. Industry and academia commonly use the finite element method (FEM) or the finite difference method (FDM) to analyze the complex heat transfer of the spindle. FEM and FDM can analyze the temperature field of the spindle at the transient or steady state and then use the temperature distribution to calculate the thermal deformation/error of the spindle [7–11]. More efficient is the thermal network model (TNM), which lumps the temperature field of the spindle into several feature temperature points called nodes, and these nodes are networked by the lumped parameters of thermal resistances, thermal capacitances, and heat sources. The lumped parameters are determined based on the theory of heat transfer and empirical formula and the measured nodes temperatures [12–15].

As mentioned, heat transfer of the spindle is very complex; furthermore, its heat generation depends on the bearing type, the torque and preload applied to the bearing, the lubrication, the diverse machining parameters, etc. Therefore, an analytical solution is not a feasible approach to obtain the temperature field of the spindle, while FEM and FDM are too time-consuming to meet the real-time requirements in the machining process. The lumped parameters in TNM are determined based on the theory of heat transfer and empirical formulas, which means that TNM only applies to certain idealized assumptions. Therefore, TNM cannot meet the diverse machining parameters in practice. Measuring temperature is much easier than measuring thermal deformation, and, furthermore, the latter will interrupt the machining process. Based on the aforementioned considerations, the authors intend to adopt a data-driven approach to establish a machine learning model for estimating the thermal error of the spindle from its several feature temperature points and rotational speeds.

The selection, including the location and number, of the feature temperature points will dramatically affect the accuracy and efficiency of the spindle's thermal error model based on machine learning. Yan and Yang [16] adopted grey correlation theory (GCT) to determine the location and number of feature temperature points, which reduced the number of temperature points from 16 to 4 and improved the accuracy of the thermal error model. Lo et al. [17] adopted the Mallows's Cp in Statistics to group the temperature points and remove the less relevant temperature points. Yuan and Ni [18] adopted the coefficient of correlation (CC) to group the temperature points, and then, in each group, selected the point most correlative to the thermal error to conduct thermal error modeling. Han et al. [19] adopted fuzzy c-means (FCM) cluster analysis to group the temperature points, and then in each group selected the point closest to the center point to conduct thermal error modeling. Yang et al. [20] adopted k-harmonic means clustering (KHM) to reduce the number of feature temperature points for the thermal error modeling of machine tools. FCM and KHM have similar clustering effects, but KHM has faster calculation speed than FCM. However, different initial partitions will lead to different final clusters of KHM, and thereby lead to different selections of feature temperature points. This is the disadvantage of KHM. Krulewich [21] adopted the Gaussian integration model (GIM) to select the temperature points and model the thermal error of a machine tool.

After determining the feature temperature points, it is necessary to find the relationship between the feature temperature points and the thermal error of the spindle. Referring to the published literature [3], the thermal error modeling methods include the least square method, multi-variable regression, grey system, neural network, support vector machine, hybrid model, etc. Li and Zhao [22] adopted the least square method (LSM) to find the polynomial function relationship between the axial deformation of the spindle and its feature temperature points. Ruijun et al. [23] adopted the multiple linear regression (MLR) model to model the relationship between the multiple temperature points of the spindle and its thermal error and found that MLR is more accurate than the back propagation neural network (BPNN) model and the radial basis function neural network (RBFNN) model, but the disadvantage is worse reliability. Yang and Ni [24] adopted the system estimation (SE) method to model the dynamic thermal error of the spindle based on the thermo-elastic theory and pseudo-hysteresis. The dynamic model is more reliable than the previous static model. Creighton et al. [25] adopted linear regression (LR) to determine the thermal error model of the spindle that could predicts the thermal error from the rotational speed and steady-state temperature of the spindle. Chen et al. [26] adopted the multiple variable regression (MVR) method and the neural network (NN) for thermal error modeling. Both MVR and NN have their own systematic and computerized processing methods to automatically find the non-linear and interactive terms in different temperature variables. Li [27] adopted the AutoRegressive method to model the thermal error of the spindle, which used the thermal error at the previous moment of the spindle and its rotational speed to estimate the current thermal error. Pahk and Lee [28] adopted three methods, namely LR, NN, and SE, to obtain the model between the spindle's temperature increase and the thermal error and compared the performances of the three methods. Shi et al. [29] adopted fuzzy c-means clustering and correlation analysis to select feature temperature points and used the Bayesian neural network to establish a thermal error model. However, the disadvantage of the Bayesian neural network is that it requires a lot of calculations, and its modeling is poor with high-dimensional data.

In summary, a thermal-error model that can accurately estimate the thermal error from real-time temperature data is required to compensate in real-time for the thermal error in the machining process. According to the above literature survey, the methods of thermal error modeling can be roughly divided into three categories: FEM, MVR, and machine learning. The accuracy of FEM depends on whether the heat source, the coefficients of heat transfer, and the boundary conditions are clearly defined [30,31]. However, the spindle is composed of many components of different materials, and its heat sources and boundary conditions are highly dependent on assembly and machining parameters. Thereby, it is difficult to define commonly applicable heat sources, coefficients of heat transfer, and boundary conditions for any types of machining parameters and conditions. Thus, FEM is not suitable for real-time thermal-error compensation in machining processes. MVR is commonly adopted to find the mathematical relationship of multiple inputs and a single output, such as between the temperature or speed of the spindle and its thermal error [12,32]. However, the MVR model would lack robustness if the data samples were small and dirty. Lei et al. [33] revealed that random forest is suitable for poor data samples and used it to model the thermal error of the spindle of a boring machine. Machine learning is very suitable for a system with multiple inputs and multiple outputs, especially suitable for a system that is difficult to describe in terms of mathematical expression.

In order to adapt to the changing machining conditions and environment, the thermal error model must be verified and updated with the continuous use of the machine tool. Therefore, the authors decided to adopt machine learning to set up the thermal error model of a spindle. This paper will adopt two methods, namely Gaussian process regression (GPR) and Random Forest (RF), to model the thermal error of the spindle and compare the performance of these two methods. RF is suitable for small and dirty data [33], and GPR can give a reliable estimate of their own uncertainty, namely the probability distribution [34]. The inputs of the thermal error model are the feature temperatures of the spindle and its rotational speed, while the output is the thermal error of the spindle. Then, the authors adopted the Pearson correlation coefficient (PCC) method to reduce the number of the feature temperature points to search for a more efficient and economic thermal error model for the spindle. GPR [35,36] has high robustness and accuracy and is easy to implement. It can adjust the hyperparameters by maximizing the marginal likelihood and by accurately optimizing them according to the value of the hyperparameters. Furthermore, the tradeoff between penalty and data-fit in GPR is automatic, that is, it is needless to set the weighting parameter through other external methods. RF [37,38] is very suitable for processing high-dimensional data and can handle both discrete and continuous data. It is not necessary to normalize the data and easy to adjust hyperparameters, and has less of an over-fitting problem.

## 2. Methodology

In the following subsections, the authors describe the spindle running-in setup for the experiment, the method of feature temperature measurement, the method of thermal error measurement, and the machine learning methods for the thermal error model of the spindle.

### 2.1. Spindle Running-in Setup

This paper used a spindle running-in system (Figure 1a) to demonstrate the present methodology [15]. The spindle, a belt-driven spindle manufactured by the Precision Machinery Research Development Center (PMC) in Taiwan, is screwed on the running-in platform through the flange. The power source of the spindle is a 7.5 kW motor, and the motor controller controls its rotational speed. The personal computer is responsible for collecting the measurement data and controlling the motor speed, namely the spindle speed. The authors used ten Pt1000 resistance temperature detectors (RTDs) (Figure 1b) to acquire the temperature data of the spindle, and two Keyence LK-H055 laser displacement sensors (Figure 1c) to measure the axial displacement of the flange and the shaft end. Table 1 lists the specification of the sensors.

**Table 1.** The specifications of the sensors.

| Specification of the Temperature Sensor Probe | | Specification of the Laser Displacement Meter | |
|---|---|---|---|
| Sensor type | PT1000 (A class) | Sensor type | Keyence LK-055 |
| Accuracy (°C) | $\pm(0.15 \pm 0.002\,|t|)$ | Reference distance (mm) | 50 |
| Measurement range (°C) | −50 to 300 | Measurement range (mm) | ±10 |
| Excited current limit (mA) | ≤5 | Light source | Red semiconductor laser |
| Thermal response (s) | ≤0.3 @ air | Wave length (nm) | 655 |
| Package material | Stainless steel 304 | Light spot diameter (μm) | 50 × 2000 |
| Protection level | IP 65 | Linearity | ±0.02% |
| | | Repeatability (μm) | 0.025 |
| | | Sampling time (μs) | 20 |
| | | Package material | Al |
| | | Protection level | IP 67 |

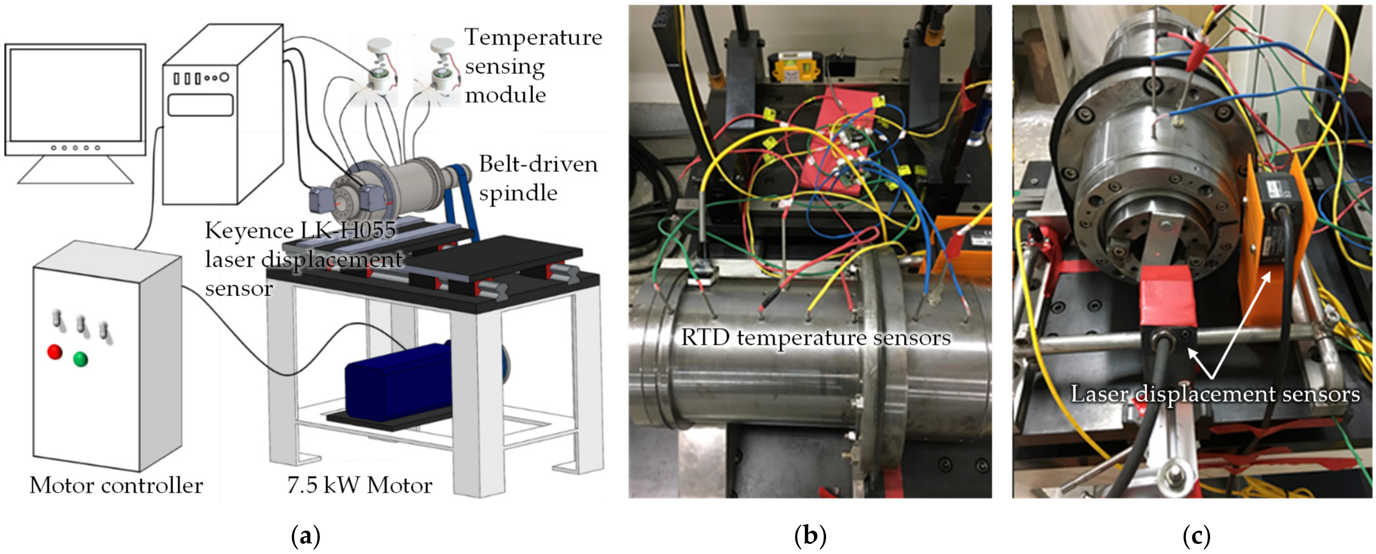


**Figure 1.** The spindle running-in experiment setup: (**a**) the scheme of the spindle running-in setup [15], copyright authorization has been obtained; (**b**) the side-view photograph of the spindle and the temperature measurement RTD; (**c**) the front-view photograph of the front end of the spindle and the laser displacement sensors.

### *2.2. Feature Temperature Measurement*

The authors used Pt1000 RTDs to measure the temperature of the spindle. For details of the architecture of the sensor and circuit, one can refer to Hu's (one of the authors of this paper) published literatures [12,32]. The number and position of the temperature points have an influence on the accuracy of the thermal error modeling [39]. It is better to select the positions where the temperature is highly correlated to the thermal error, and these positions are called the feature temperature points. There are some guideline for the selection of the feature temperature points [40]: the temperature points should be near the heat source; the feature temperature points should be able to describe the temperature field; and the temperature points should be closely related to the thermal error. Based on the aforementioned literatures, the authors selected ten feature temperature points (Figure 2): front bearings A and B; rear bearings C and D; front and rear outer housings; front and rear inner housing; mid inner housing; and ambient. These feature temperature points were measured by the use of ten Pt1000 RTDs.

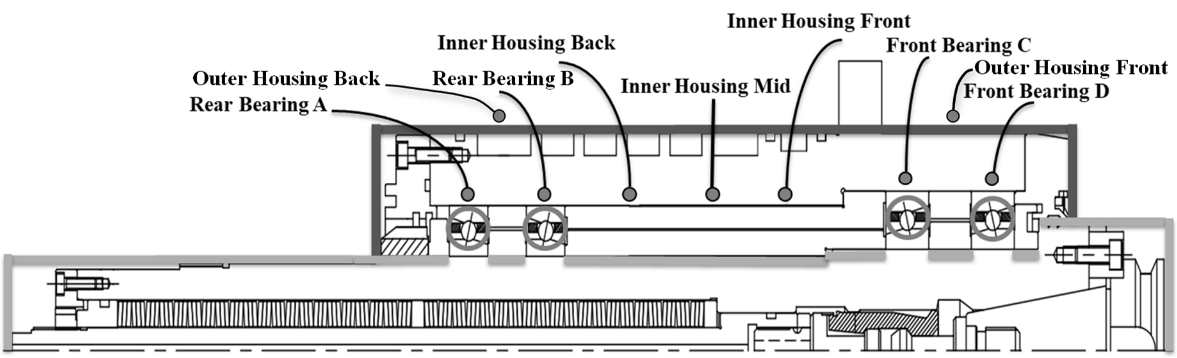


**Figure 2.** The feature temperature points [15], copyright authorization has been obtained.

The authors collected the temperature data under the two conditions of constant speed and variable speed. The constant speed conditions were 5000, 6000, 7000, and 8000 rpm; the variable speed conditions were divided into four stages of 5000, 6000, 7000, and 8000 rpm; and speeds were increased step-by-step from low to high. The data samples were collected at night because of no sunshine. The experiments were carried out in a laboratory with an air conditioner, and the ambient temperature and humidity were controlled within 25 ± 1 °C and 55–65%, respectively. Each constant speed and variable speed condition performed 3 runs on different days. Each constant speed condition lasted 6.5 h (Figure 3); while each speed stage of the variable speed condition lasted 1.6 h and thus 6.4 h in total (Figure 4). Figure 3 shows that the temperature initially increased rapidly and then reached stable conditions. The temperatures of the rear bearings were higher than the other temperature points; this is because the rear bearings were near the driving belt. The temperatures of the front bearings were lower than the other temperature points; this is because the diameter and volume of the spindle near the front bearings were larger and thus had a larger heat capacity.

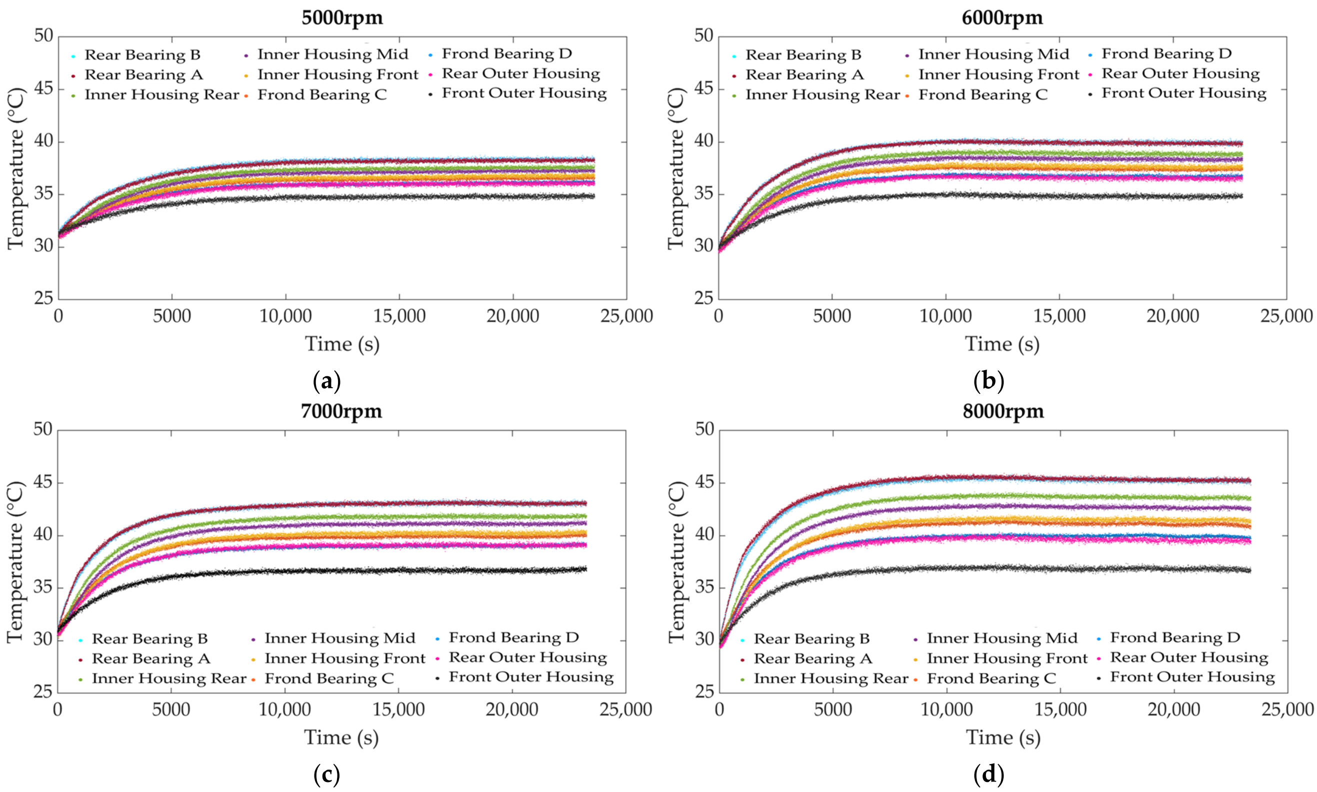


**Figure 3.** The demonstration of the feature temperatures measurement results under constant speed: (**a**) 5000 rpm; (**b**) 6000 rpm; (**c**) 7000 rpm; (**d**) 8000 rpm.

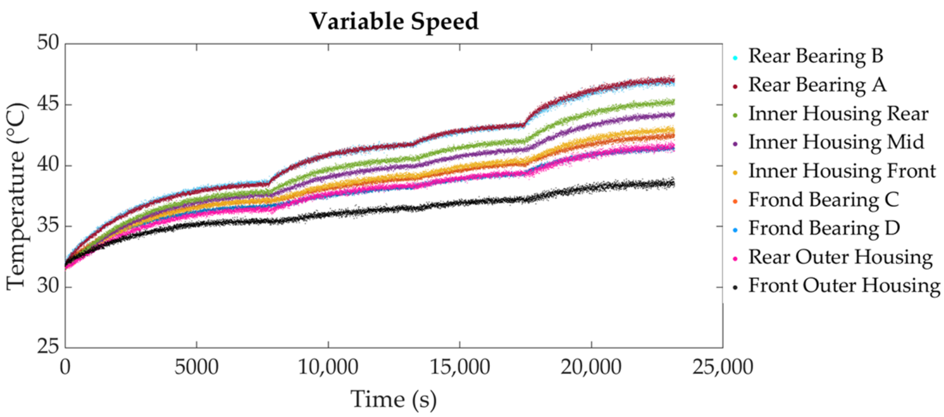


**Figure 4.** The demonstration of the feature temperatures measurement results under variable speed: 5000→6000→7000→8000 rpm.

*2.3. Thermal Error Measurement*

The spindle was screwed to the ring bracket of the running-in platform through its flange. Ideally, the flange is a fixed support of the spindle. However, the tensile force of the belt will exert a downward lateral force on the spindle and cause the rear end of the spindle to tilt downward slightly (Figure 5) because the ring bracket is not perfectly rigid. Therefore, the authors used two Keyence LK-H055 laser displacement sensors (Figure 1c) to measure the axial displacement of the flange and the shaft's front end. Thus, subtracting the displacement of the flange from that of the shaft's front end gave the thermal deformation/error of the spindle. Figure 6 shows the thermal error under the constant speed and the variable speed. The speeds of the constant speed experiments were, respectively, 5000, 6000, 7000, and 8000 rpm; while the speeds of the variable speed experiment were divided into four stages, namely 5000, 6000, 7000, and 8000 rpm, and were sped up from low to high. This figure shows that the thermal error initially increased rapidly and then reached stable conditions. This phenomenon is consistent with that of the temperature (Figures 3 and 4).

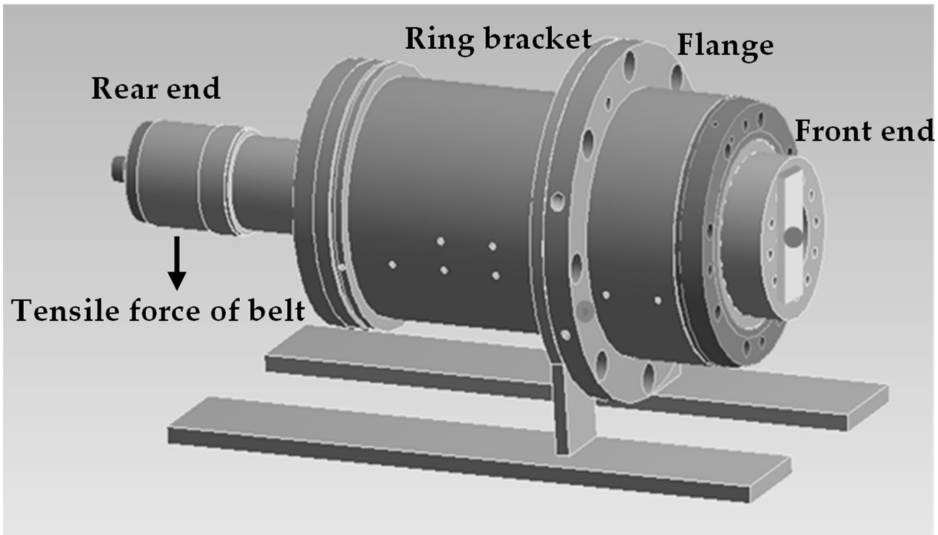


**Figure 5.** The schematic diagram of the spindle screwed to the ring bracket of the running-in platform through its flange.

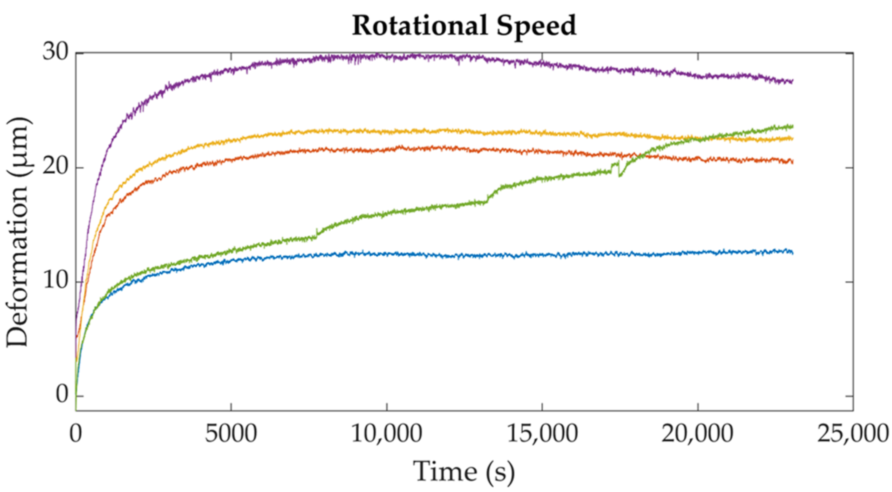


**Figure 6.** The demonstration of the thermal error measurement results under constant speed and variable speed.

### 2.4. Training Data and Validation Data

Data samples were divided into training data group and validation data group. Each rotation speed mode performed 3 runs on different days. Therefore, there were 3 training data sets for each rotation speed mode (Table 2). Similarly, in the validation data group, there were 3 validation data sets for each rotation speed mode. The validation data group was collected on different days from the training data group and did not participate in the model training. Each data set contained the time series data of the ten feature temperature points of the spindle and its rotational speed and thermal error, similar to Figures 3, 4 and 6. In the training process of the thermal error model, the temperature and rotation speed data in the training data sets were the input, while the thermal error data were the output. The temperature and rotation speed data in the validation data sets were to be input to the thermal error model to predict the thermal error and then to compare the predicted thermal error with the thermal error data in the validation data sets to validate the thermal error model.

**Table 2.** The list of the training and testing data sets.

| Data Group | Data Set | Speed (rpm) |
|---|---|---|
| Training group | $T_{5000,1}$, $T_{5000,2}$, $T_{5000,3}$ | 5000 |
| | $T_{6000,1}$, $T_{6000,2}$, $T_{6000,3}$ | 6000 |
| | $T_{7000,1}$, $T_{7000,2}$, $T_{7000,3}$ | 7000 |
| | $T_{8000,1}$, $T_{8000,2}$, $T_{8000,3}$ | 8000 |
| | $T_{var,1}$, $T_{var,2}$, $T_{var,3}$ | 5000→6000→7000→8000 |
| Validation group | $V_{5000,1}$, $V_{5000,2}$, $V_{5000,3}$ | 5000 |
| | $V_{6000,1}$, $V_{6000,2}$, $V_{6000,3}$ | 6000 |
| | $V_{7000,1}$, $V_{7000,2}$, $V_{7000,3}$ | 7000 |
| | $V_{8000,1}$, $V_{8000,2}$, $V_{8000,3}$ | 8000 |
| | $V_{var,1}$, $V_{var,2}$, $V_{var,3}$ | 5000→6000→7000→8000 |

### 2.5. Thermal Error Modeling by Machine Learning

In the following sub-subsections, the authors briefly describe Gaussian process regression (GPR) and random forest (RF) adopted in the thermal error modeling of the spindle.

#### 2.5.1. Gaussian Process Regression (GPR)

GPR was conducted to find a function that can describe the relationship between the independent variables (inputs), $\boldsymbol{x}$, and the dependent variables (outputs), $\boldsymbol{y}$, for a given data set, and that can predict the corresponding dependent variables based on the new independent variables. For the thermal error model of the spindle, the inputs, $\boldsymbol{x}$, are the time series data of its feature temperature points and the output, $\boldsymbol{y}$, is the time series data of its thermal error. It does not need to choose the form of the fitting function to fit the relationship between the independent variable and the dependent variable. For a given data set, there will be many potential functions that can be used to fit the relationship between the input and output data. GPR will assign a probability value to each potential function and uses the mean value of the distribution of the probability values to represent the most possible feature of the data set. A Gaussian process is composed of the mean function and the kernel function. Through modifying the mean function and the kernel function, one can obtain a good machine learning model to map the input and output data. For details of GPR, one can refer to [41].

The authors adopted the MATLAB built-in GPR to model the thermal error of the spindle. The procedure is as follows:

1. Use the "Import Data" to import the training data sets and the testing data sets;
2. Choose the "Machine Learning and Deep Learning" variety in the App toolbar, and then use the "Regression Learner";
3. Choose a training data set, set the thermal error data to response, set the temperature data to predictors, and set "validation" to prevent the model from overfitting;

4. Choose the Exponential of Gaussian Process Regression, one can use "Advanced" to adjust the training parameters;
5. Click "Export Model", input testing data set to the model to get a response, get mean error and accuracy by calculating the true value, and predicted value based on the testing data set.

#### 2.5.2. Random Forests (RF)

RF is a method of machine learning proposed by Leo Breiman [42] in 2001. RF is composed of many trees, called decision trees, and there is no correlation between each decision tree. The decision trees classify the data in the training data sets layer by layer until they can no longer be classified. There are many types of decision trees, such as classification trees (CT), regression trees (RT), and classification and regression trees (CART) [43]. CT classifies the data into multiple classes; therefore, each node may split into many branches, and the leaf/output of CT is a class. RT classifies the data based on the threshold values; therefore, each node may split into many branches, and the leaf/output of RT is a number. This paper is to find the relationship between the feature temperature points of the spindle and its thermal error; therefore, it is a regression problem and the decision tree is RT. RF is composed of many CARTs. CART is a combination of CT and RT, but the difference is that its node will only split into two branches, that is, it is a binary tree. Its advantage is less prone to overfitting. This paper uses RF to model the thermal error of the spindle; the output of RF is the mean value of all CARTs. For a given training data set, T = $\{(x_1 y_1,), (x_2,y_2), (x_3,y_3), \ldots, (x_n,y_n)\}$, where $\boldsymbol{x}$ is an 11-dimensional vector because, in addition to the spindle speed, there are 10 feature temperature points of the spindle, and $\boldsymbol{y}$ is the thermal error. The goal of regression is to find a model to satisfy all the elements in the training data set, so that the mean square error is the smallest. The details of RF can refer to [42,43]. The authors used Python programming, and the source code is listed in Appendix A.

## 3. Results

This section shows the modeling results of GPR and RF. In addition, since ten feature temperature points is too many and some temperature points are unacceptable in the industry, the authors here used Pearson correlation coefficient (PCC) to reduce the number of the feature temperature points.

### *3.1. Thermal Error Model Based on Gaussian Process Regression*

The GPR randomly selected 75% of the training data set to train the thermal error model of the spindle and selected the remaining 25% of the training data set to self-validate the thermal error model. The values of root mean square error (*RMSE*) and R-squared were the two indices to check the accuracy of the thermal error model. *RMSE* means the deviation between the predicted value and the actual value, and the smaller the better. R-squared means the proportion of variance of the dependent variable explained by the independent variables in a regression model, and the closer it is to 1 the better. Figure 7 shows the self-validation results of the thermal error model trained by GPR. R-squared approached 1 and the average *RMSE* was about 0.22 µm. In addition, the thermal error model had to be validated externally by new data sets, namely the validation data sets listed in Table 2. Figure 8 shows the external validation results of the thermal error model trained by GPR. The average *RMSE* was about 2.21 µm, and the average accuracy was about 87.22%.

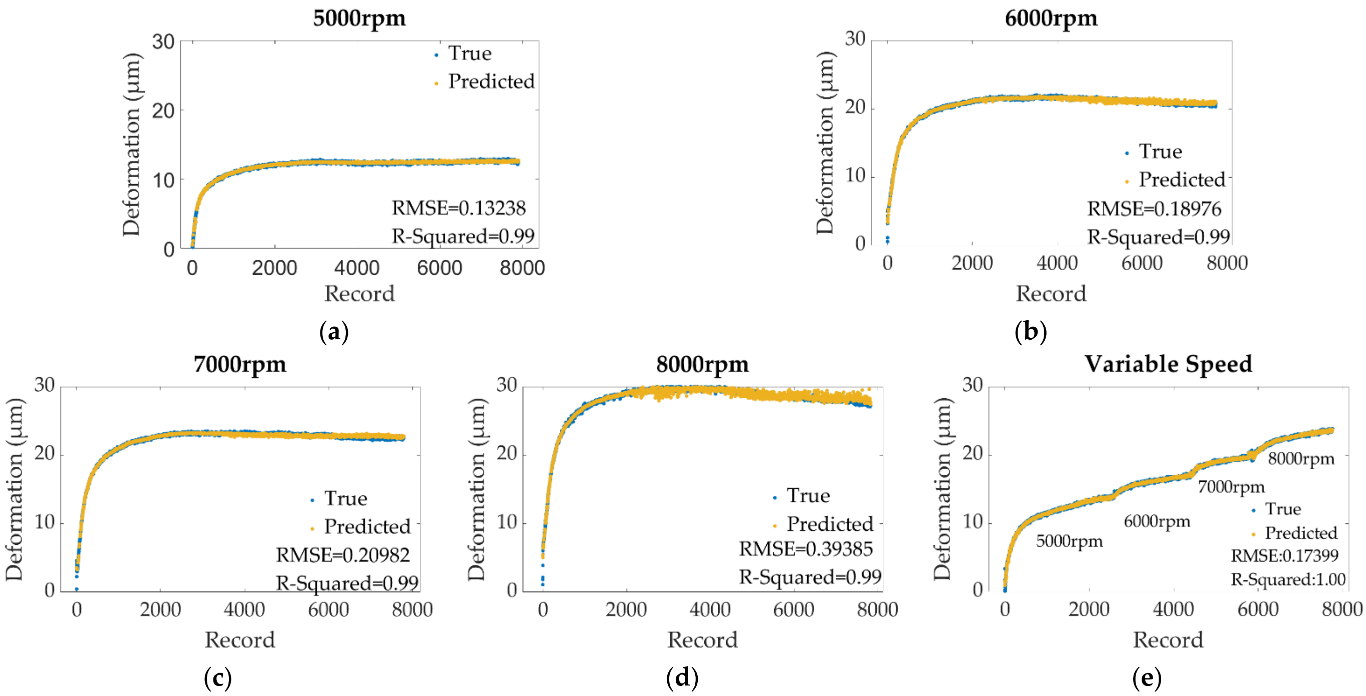


**Figure 7.** The self-validation of the thermal error model trained by Gaussian process regression under the rotational speed of (**a**) 5000 rpm; (**b**) 6000 rpm; (**c**) 7000 rpm; (**d**) 8000 rpm; (**e**) variable speed.

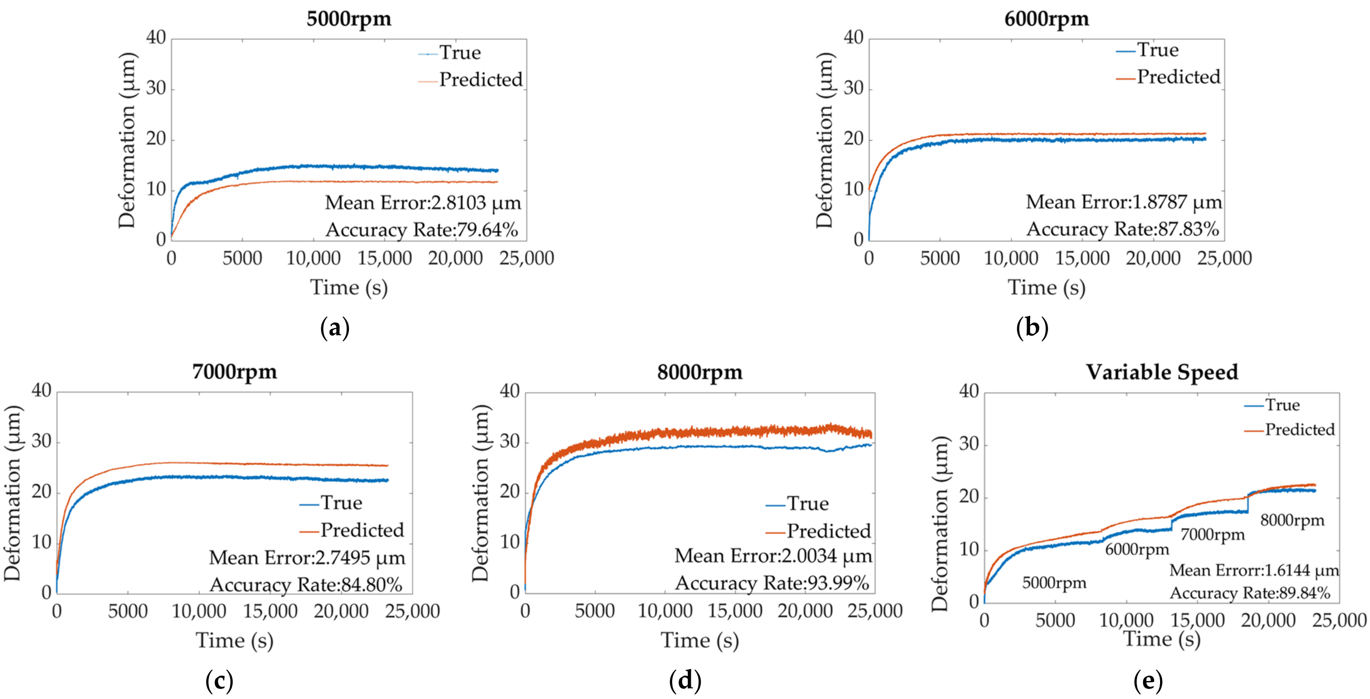


**Figure 8.** The external validation of the thermal error model trained by Gaussian process regression under the rotational speed of (**a**) 5000 rpm; (**b**) 6000 rpm; (**c**) 7000 rpm; (**d**) 8000 rpm; (**e**) variable speed.

### *3.2. Thermal Error Model Based on Random Forest*

The RF randomly selected 80% of the training data set to train the thermal error model of the spindle and the remaining 20% of the training data set to self-validate the thermal error model. Figure 9 shows the self-validation results of the thermal error model trained by RF. The R-squared approached 1, and the average *RMSE* was about 0.22 μm. In addition, the thermal error model had to be validated externally by new data sets, namely the validation data sets listed in Table 2. Figure 10 shows the external validation results of the thermal error model trained by RF. The average *RMSE* was about 1.85 μm, and the average accuracy was about 89.7%. The thermal error model of the spindle obtained by RF was slightly better than that obtained by GPR. Though the deviation of *RMSE* between RF and GPR seemed to be only about 0.36 μm, a deviation of 0.36 μm is important for ultra-precision machining such as the machining of the cavity of contact lens.

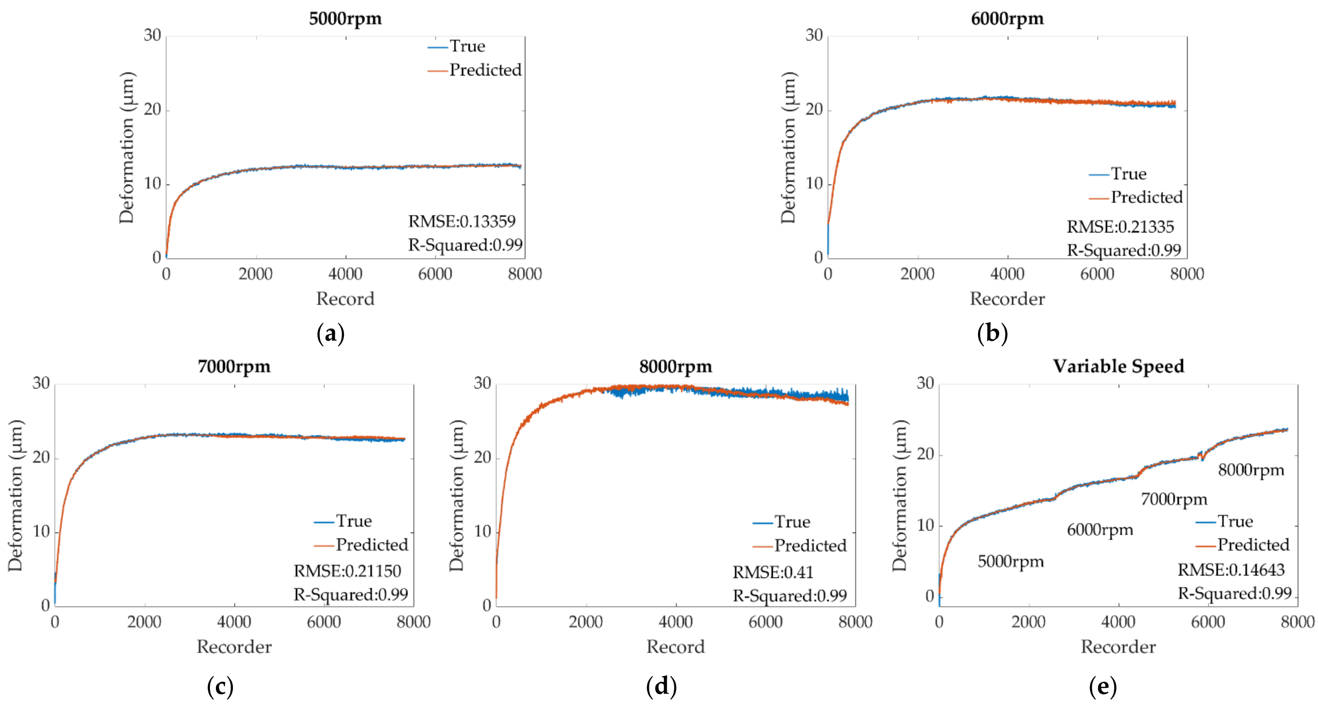


**Figure 9.** The self-validation of the thermal error model trained by random forests under the rotational speed of (**a**) 5000 rpm; (**b**) 6000 rpm; (**c**) 7000 rpm; (**d**) 8000 rpm; (**e**) variable speed.

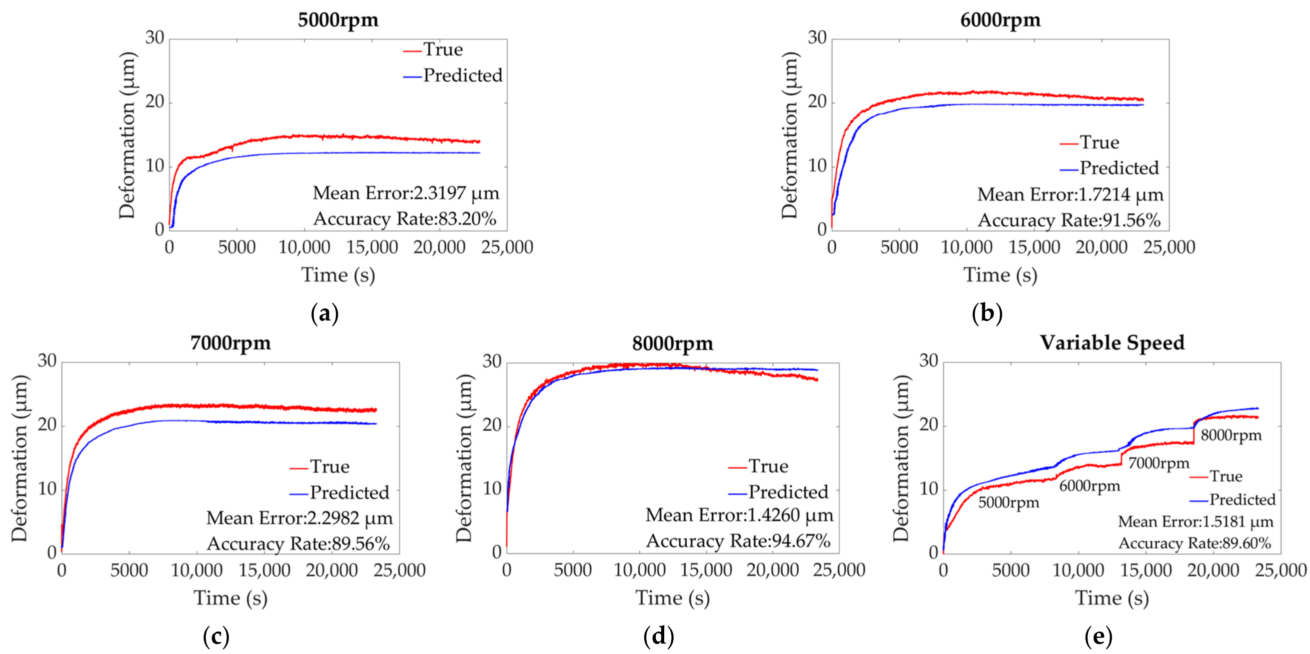


**Figure 10.** The external validation of the thermal error model trained by random forests under the rotational speed of (**a**) 5000 rpm; (**b**) 6000 rpm; (**c**) 7000 rpm; (**d**) 8000 rpm; (**e**) variable speed.

### 3.3. Feature Temperature Points Selection by Pearson Correlation Coefficient (PCC)

Since ten feature temperature points are too many, and some temperature points are unacceptable in the industry, the authors here used Pearson correlation coefficient (PCC) to reduce the number of feature temperature points. PCC is used to measure the linear correlation between independent and dependent variables. Its value is between the range $[-1, +1]$, where $-1$ means absolutely negative correlation, +1 means absolutely positive correlation, and 0 means no correlation. Figure 11 visualizes the average PCC of each feature temperature point under the speeds of the running-in test in the order from high to low. The highest correlation points were the rear bearings.

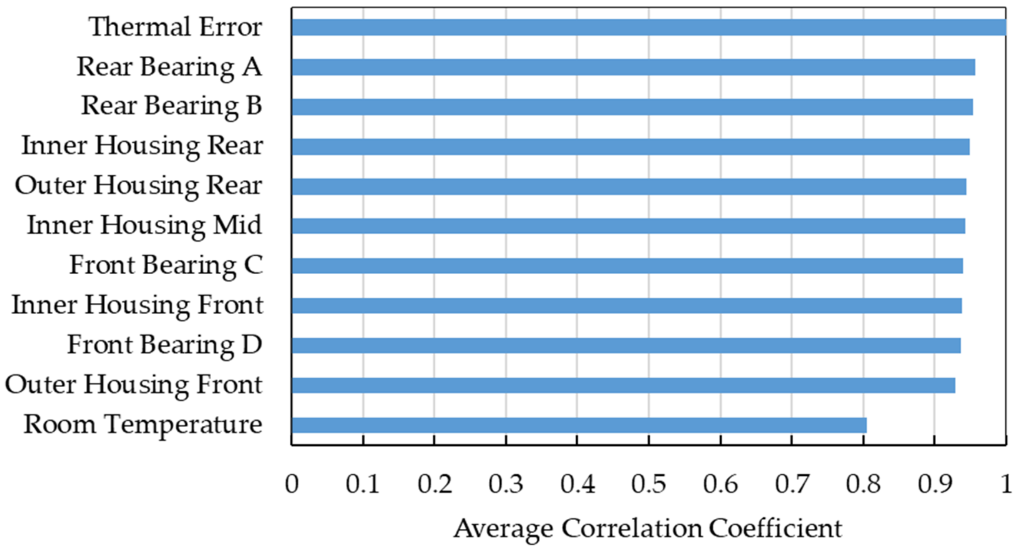


**Figure 11.** The visualization of the average correlation coefficients between the feature temperature points and the thermal error.

There are many indices to justify the prediction performance of thermal error model, such as mean error (*ME*), root mean square error (*RMSE*), and accuracy, whose definitions are,

$$ME = \frac{\sum_{i=1}^{n} |y_i - y_i'|}{n}, \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - y_i'\right)^2}{n}}, \tag{2}$$

$$Accuracy = \left(1 - \frac{\sum_{i=1}^{n} |y_i - y_i'|}{\sum_{i=1}^{n} |y_i|}\right) \times 100\%, \tag{3}$$

where $y_i$ and $y_i'$ are the measured thermal error and the predicted thermal error, respectively, *ME* and *RMSE* are the mean error and root mean square error of the predicted thermal error with respect to measured thermal error, and *Accuracy* is the deviation percentage of the predicted thermal error with respect to the measured thermal error. In addition, the coefficient of determination in statistics ($R^2$) is a common measure of how well the output data are replicated by the model. Its value ranges from 0 to 1, and the closer to 1, the better the model. $R^2$ is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - y_i'\right)^2}{\sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2}, \tag{4}$$

where $\overline{y}$ is the mean value of the measured thermal error.

In order to compare the influence of the feature temperature point selection on the thermal error prediction of the model, the authors selected the feature temperature points from Figure 11 in the order of the correlation coefficient from high to low and re-trained the thermal error model under the variable speed condition. Tables 3 and 4 list their external validation results obtained by GPR and RF, respectively. The best result of GPR was obtained by selecting nine feature temperature points (excluding the room temperature), and the accuracy was improved to be 89.71%; while that of RF was only four (including rear bearings A and B, inner housing rear, and outer housing rear), and the accuracy was improved to be 90.49%, respectively.

**Table 3.** The influence of the number of the feature temperature points on the external validation result of the thermal error model trained through GPR under variable speed conditions.

| No. F. Pt. [1] | *ME* [2] (μm) | *RMSE* [3] (μm) | $R^2$ * | *Accuracy* |
|---|---|---|---|---|
| 2 | 1.6423 | 1.7481 | 0.9884 | 88.92% |
| 3 | 1.6128 | 1.6860 | 0.9888 | 89.09% |
| 4 | 1.5960 | 1.7098 | 0.9888 | 89.19% |
| 5 | 1.5595 | 1.6805 | 0.9889 | 89.45% |
| 6 | 1.5769 | 1.6759 | 0.9894 | 89.38% |
| 7 | 1.5458 | 1.6557 | 0.9899 | 89.57% |
| 8 | 1.5374 | 1.6734 | 0.9906 | 89.64% |
| 9 | 1.5145 | 1.6539 | 0.9905 | 89.71% |
| 10 | 1.5160 | 1.6573 | 0.9910 | 89.62% |

[1] The number of feature temperature points selected from Figure 11 in the order of the correlation coefficient from high to low; [2] mean error of the machine learning model; [3] root mean square error of the machine learning model; * R-squared value.

**Table 4.** The influence of the number of the feature temperature points on the external validation result of the thermal error model trained through RF under variable speed conditions.

| No. F. Pt. [1] | *ME* [2] (μm) | *RMSE* [3] (μm) | $R^2$ * | *Accuracy* |
|---|---|---|---|---|
| 2 | 1.5728 | 1.6633 | 0.9914 | 89.41% |
| 3 | 1.5730 | 1.6656 | 0.9909 | 89.41% |
| 4 | 1.5273 | 1.6402 | 0.9905 | 90.49% |
| 5 | 1.5167 | 1.6474 | 0.9901 | 89.61% |
| 6 | 1.5125 | 1.6417 | 0.9902 | 89.62% |
| 7 | 1.5140 | 1.6495 | 0.9898 | 89.61% |
| 8 | 1.5129 | 1.6482 | 0.9899 | 89.62% |
| 9 | 1.5130 | 1.6490 | 0.9898 | 89.62% |
| 10 | 1.5181 | 1.6506 | 0.9898 | 89.60% |

[1] The number of feature temperature points selected from Figure 11 in the order of the correlation coefficient from high to low; [2] mean error of the machine learning model; [3] root mean square error of the machine learning model; * R-squared value.

## 4. Discussion

Based on the results adopting 10 feature temperature points, RF is slightly better than GPR for the thermal error modeling of the spindle. Though the deviation of *RMSE* of RF and GPR seems to be only about 0.36 μm, the deviation of 0.36 μm is important for ultra-precision machining, such as the machining of the cavity of a contact lens. Collocating with PCC, RF can reduce to only four feature temperature points that are closest to the rear end of the spindle; while GPR still requires nine feature temperature points, excluding room temperature, to approach the accuracy of RF. Therefore, RF collocating with PCC is an efficient and accurate method for the thermal error modeling of the spindle. If the accuracy requirement is not very difficult, one can choose just the temperature points nearing the bearings, because the installation of temperature sensors at these positions is acceptable for the spindle or machine tool manufacture, while the other positions may interfere with the cooling pipeline of the spindle. The external validations shown in Figures 8e and 10e reveal that the thermal error prediction for variable speed conditions has larger deviations than those of constant speed conditions. This is because under variable speed conditions, the system has not yet reached a thermally steady state. When the system is in a thermally steady state, the relationship between the thermal error and the temperature is close to linear. However, when the system does not reach a thermally steady state, the relationship between the thermal error and the temperature is close to nonlinear. Fortunately, in practical machining processes, the spindle speed will not change frequently and even keep constant in the whole machining process but just change the depth of the cut and the feed rate. Otherwise, the model training has to adopt nonlinear regression.

Table 5 compares the prediction performance of this paper with the other literatures using machine learning modeling. Note that the prediction performance indices of this paper shown in the table are under the variable spindle speed condition, while most of the other literatures are under specific constant speed conditions based on their own experiments. For specific constant spindle speed conditions, most literatures could show not bad results, as does this paper. However, the performance of this paper under variable spindle speed conditions is even better than those of the other literatures under constant spindle speed conditions.

**Table 5.** Comparison of this paper with published literatures. Note that the prediction performance indices of this paper are under the variable speed condition, while most of the other literatures are under specific constant speed conditions based on their own experiments.

| Paper | M. L. Model * | F. Pt. Selection * | No. F. Pt. * | *ME* * | *RMSE* * | $R^2$ * | *Accuracy* * |
|---|---|---|---|---|---|---|---|
| This paper | RF ** | PCC *** | 4 | 1.5273 | 1.6402 | 0.9905 | 90.49% |
| | GPR ** | PCC | 4 | 1.5960 | 1.7098 | 0.9888 | 89.19% |
| [16] | DCM ** | IGM *** | 4 | 3.62 | - | - | - |
| [17] | Mallows' Cp | CC *** | 4 | - | - | 0.982 | 89% |
| [19] | MLR ** | FCM *** | 4 | 1.8 | - | - | - |
| [20] | MLR | KHM *** | 3 | - | 6.9690 | 0.9356 | 90.86% |
| [21] | GI ** | - | 9 | - | 2.4928 | - | 93% |
| [22] | LS ** | SRCC *** | 11 | - | - | - | 85% |
| [23] | RBF ** | CC | 5 | - | 2.4 | - | 75% |
| [29] | BNN ** | FCM | 3 | 1.741 | 1.998 | 0.807 | 74.1% |

* M. L. Model: machine learning model; F. Pt.: feature temperature point; No. F. Pt.: number of feature temperature point; ** RF: random forest; GPR: Gaussian process regression; DCM: direct criterion method; MLR: multiple linear regression; GI: Gaussian integration; LS: least square method; RBF: radial basis function network model; BNN: Bayesian neural network; *** PCC: Pearson correlation coefficient; IGM: indirect grouping method; CC: correlation coefficient; FCM: fuzzy C-means; KHM: k-harmonic means clustering; SRCC: Spearman's rank correlation coefficients.

## 5. Conclusions

Though the thermal error model for estimating the thermal error of the spindle from its feature temperature points is a regression problem, random forests is a better method than Gaussian process regression. Furthermore, random forests collocating with Pearson correlation coefficients provides an accurate and efficient method for the thermal error modeling of the spindle. The result shows that random forests collocating with Pearson correlation coefficients is an efficient and accurate method for the thermal error modeling of the spindle. The temperatures at the bearings and the inner housing nearing the bearings are the most important for the thermal error modeling of the spindle. If the accuracy requirement is not very onerous, one can select just the temperature points of the bearings, because the installation of temperature sensors at these positions is acceptable for the spindle or machine tool manufacture, while the other positions may interfere with the cooling pipeline of the spindle. Most of the published literatures showed the performance of their model only under specific constant spindle speed conditions. For specific constant spindle speed conditions, most literatures using the other machine learning models could show not bad results, and so does this paper. However, the performance of this paper under variable spindle speed conditions is even better than those of the other literatures under constant spindle speed conditions. Note that the experiment is on a running-in platform of the spindle, which means no cutting load. If the spindle is to conduct machining, the front bearings should be the important feature temperature points, because the front bearings are closest to the cutting heat. This is worth investigating in the near future.

**Author Contributions:** Conceptualization, Y.-C.H.; methodology, Y.-C.C. and P.-H.W.; software, Y.-C.C. and P.-H.W.; validation, Y.-C.H.; formal analysis, Y.-C.H.; investigation, Y.-C.C. and P.-H.W.; resources, Y.-C.H.; data curation, Y.-C.C. and P.-H.W.; writing—original draft preparation, Y.-C.C. and P.-H.W.; writing—review and editing, Y.-C.H.; visualization, Y.-C.C. and P.-H.W.; supervision, Y.-C.H.; project administration, Y.-C.H.; funding acquisition, Y.-C.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Technology, Taiwan, grant numbers MOST 110-2221-E-197-029 and MOST 110-2622-E-197-009.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Appendix A

```
#Import data set
import pandas as pd
data = pd.read_csv(“Data_random_1.csv”)
#Remove time and deformation
data2 = pd.read_csv(“Data_random_1.csv”, usecols = [1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 13])
#Import data split module and drawing module
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
%matplotlib inline
import numpy as np
np.random.seed(42)
#Set the input and output of the model
x = data2
y = data.Deformation
#Divide training data and testing data (training data account for 80%)
x_train, x_test, y_train, y_test = train_test_split (x, y, test_size = 0.2)
#Use grid-searching to fine suitable hyperparameters
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import RandomizedSearchCV
from scipy.stats import randint
param_distribs = {
‘n_estimators’: randint(low = 0, high = 500),
‘max_depth’: randint(low = 1, high = 15),
‘min_samples_split’: randint(low = 2, high = 8),
‘min_samples_leaf’: randint(low = 1, high = 8),
‘max_features’: randint(low = 1, high = 9),
}
forest_reg = RandomForestRegressor(random_state = 42)
rnd_search = RandomizedSearchCV(forest_reg, param_distributions = param_distribs,n_iter
= 10, cv = 5, scoring = ‘neg_root_mean_squared_error’, random_state = 42)
rnd_search.fit(x_train, y_train)
cvres = rnd_search.cv_results_
for mean_score, params in zip(cvres[“mean_test_score”], cvres[“params”]):
print(np.sqrt(-mean_score), params)
rnd_search.best_params_
#Substitute the hyperparameters into the modeling
from sklearn.ensemble import RandomForestRegressor
```

```
model = RandomForestRegressor(n_estimators = 199,criterion = ‘mse’, max_depth = 11, min_samples_split = 4, min_samples_leaf = 4, max_features = 5, random_state = 42)
predict = model.fit(x_train,y_train).predict(x_test)
#Use learning curve to judge whether the model is over-fitting or not
from sklearn.metrics import mean_squared_error
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor(n_estimators = 199,criterion = ‘mse’, max_depth = 11, min_samples_split = 4, min_samples_leaf = 4, max_features = 5, random_state = 42)
def plot_learning_curves(model, x, y):
X_train, X_val, y_train, y_val = train_test_split(x, y, test_size = 0.2, random_state = 10)
train_errors, val_errors = [], []
for m in range(1, len(X_train)):
model.fit(X_train[:m], y_train[:m])
y_train_predict = model.predict(X_train[:m])
y_val_predict = model.predict(X_val)
train_errors.append(mean_squared_error(y_train[:m], y_train_predict))
val_errors.append(mean_squared_error(y_val, y_val_predict))
plt.plot(np.sqrt(train_errors), “r-”, linewidth = 2, label = “train”)
plt.plot(np.sqrt(val_errors), “b-”, linewidth = 3, label = “val”)
plt.legend(loc = “upper right”, fontsize = 14)
plt.xlabel(“Training set size”, fontsize = 14)
plt.ylabel(“RMSE”, fontsize = 14)
#Calculate the mean absolute error of the model
errors = abs(predict—y_test)
print(‘Mean Absolute Error:’, round(np.mean(errors), 2), ‘degrees.’)
#Calculate the root mean square error of the model
from sklearn.metrics import mean_squared_error
from math import sqrt
rmse = sqrt(mean_squared_error(y_test,predict))
print(rmse)
#Calculate the score of the model to judge whether over-fitting or not
print(“Traing Score:%f”%model.score(x_train,y_train))
print(“Testing Score:%f”%model.score(x_test,y_test))
#Import the validation data set to the thermal error model
data3= pd.read_csv(“801rpm.csv”)
data4= pd.read_csv(“801rpm.csv”, usecols = [1, 2, 3, 4, 5, 6, 7, 8, 9, 11])
x2 = data4
y2 = data3.Deformation
#Predict the outputs from the inputs of the validation data set through the thermal error model
predict_df = []
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor(n_estimators = 199,criterion = ‘mse’, max_depth = 11, min_samples_split = 4, min_samples_leaf = 4, max_features = 5, random_state = 42)
model.fit(x_train,y_train)
for m in range(0, len(x2)):
X_new = [x2.iloc[m,:]]
predict_df.append(model.predict(X_new))
#Export the prediction results to CSV file
def OutputVar():
Result = ‘C://Users/User/Desktop/Varresponse.csv’
```

```
df_SAMPLE = pd.DataFrame.from_dict(predict_df)
df_SAMPLE.to_csv(Result, index = False)
print('Success' + Result)
OutputVar()
```

## References


1. Bryan, J. International status of thermal error research. *CIRP Ann.* **1990**, *39*, 645–656. [CrossRef]
2. Ramesh, R.; Mannan, M.A.; Poo, A.N. Error compensation in machine tools—A review: Part I: Geometric, cutting-force induced and fixture-dependent errors. *Int. J. Mach. Tools Manuf.* **2000**, *40*, 1235–1256. [CrossRef]
3. Li, Y.; Yu, M.; Bai, Y.; Hou, Z.; Wu, W. A review of thermal error modeling methods for machine tools. *Appl. Sci.* **2021**, *11*, 5216. [CrossRef]
4. Postlethwaite, S.R.; Allen, J.P.; Ford, D.G. Machine tool thermal error reduction—An appraisal. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **1999**, *213*, 1–9. [CrossRef]
5. Takabi, J.; Khonsari, M.M. Experimental testing and thermal analysis of ball bearings. *Tribol. Int.* **2013**, *60*, 93–103. [CrossRef]
6. Kendoush, A.A. An approximate solution of the convective heat transfer from an isothermal rotating cylinder. *Int. J. Heat Fluid Flow* **1996**, *17*, 439–441. [CrossRef]
7. Jędrzejewski, J.; Kaczmarek, J.; Kowal, Z.; Winiarski, Z. Numerical optimization of thermal behavior of machine tools. *CIRP Ann.* **1990**, *39*, 379–382. [CrossRef]
8. Chen, D.; Bonis, M.; Zhang, F.; Dong, S. Thermal error of a hydrostatic spindle. *Precis. Eng.* **2011**, *35*, 512–520. [CrossRef]
9. Feng, W.; Li, Z.; Gu, Q.; Yang, J. Thermally induced positioning error modelling and compensation based on thermal characteristic analysis. *Int. J. Mach. Tools Manuf.* **2015**, *93*, 26–36. [CrossRef]
10. Venugopal, R.; Barash, M.; Shaw, M.C. Thermal effects on the accuracy of numerically controlled machine tools. *CIRP Ann.* **1986**, *35*, 255–258. [CrossRef]
11. Mayr, J.; Ess, M.; Weikert, S.; Wegener, K.J.P.L. Compensation of Thermal Effects on Machine Tools Using a FDEM Simulation Approach. In Proceedings of the 9th Lamdamap Conference, Brunel University, Uxbridge, UK, 30 June–2 July 2009; ISBN 1861941188. Available online: https://www.iwf.mavt.ethz.ch/ConfiguratorJM/publications/Compensati_129500865499587/O1.5_Mayr.pdf (accessed on 9 July 2021).
12. Lo, Y.-C.; Hu, Y.-C.; Chang, P.-Z. Parameter estimation of the thermal network model of a machine tool spindle by self-made Bluetooth temperature sensor module. *Sensors* **2018**, *18*, 656. [CrossRef]
13. Huang, D.; Hong, J.; Zhang, J.; Wu, D.; Li, C. Thermal resistance network for solving temperature field in spindle system. *J. Xi'an Jiaotong Univ.* **2012**, *46*, 63–66.
14. Liu, Z.; Pan, M.; Zhang, A.; Zhao, Y.; Yang, Y.; Ma, C. Thermal characteristic analysis of high-speed motorized spindle system based on thermal contact resistance and thermal-conduction resistance. *Int. J. Adv. Manuf. Technol.* **2015**, *76*, 1913–1926. [CrossRef]
15. Wu, C.-H. Thermal Deformation Prediction of a Machine Tool Spindle by Thermal Network Model. Master's Thesis, National Taiwan University, Taipei, Taiwan, 1 July 2019.
16. Yan, J.; Yang, J.G. Application of synthetic grey correlation theory on thermal point optimization for machine tool thermal error compensation. *Int. J. Adv. Manuf. Technol.* **2009**, *43*, 1124–1132. [CrossRef]
17. Lo, C.-H.; Yuan, J.; Ni, J. Optimal temperature variable selection by grouping approach for thermal error modeling and compensation. *Int. J. Mach. Tools Manuf.* **1999**, *39*, 1383–1396. [CrossRef]
18. Yuan, J.; Ni, J. The real-time error compensation technique for CNC machining systems. *Mechatronics* **1998**, *8*, 359–380. [CrossRef]
19. Han, J.; Wang, L.; Cheng, N.; Wang, H. Thermal error modeling of machine tool based on fuzzy c-means cluster analysis. In Proceedings of the 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, Harbin, China, 12–14 August 2011. [CrossRef]
20. Yang, L.; Zhao, J.; Ji, S.; Liang, F. The selection of temperature-sensitivity points based on K-harmonic means clustering and thermal positioning error modeling of machine tools. *Int. J. Adv. Manuf. Technol.* **2019**, *100*, 2333–2348. [CrossRef]
21. Krulewich, D.A. Temperature integration model and measurement point selection for thermally induced machine tool errors. *Mechatronics* **1998**, *8*, 395–412. [CrossRef]
22. Li, Y.; Zhao, W. Axial thermal error compensation method for the spindle of a precision horizontal machining center. In Proceedings of the 2012 IEEE International Conference on Mechatronics and Automation, Chengdu, China, 5–8 August 2012. [CrossRef]
23. Ruijun, L.; Wenhua, Y.; Zhang, H.H.; Qifan, Y. The thermal error optimization models for CNC machine tools. *The International J. Adv. Manuf. Technol.* **2012**, *63*, 1167–1176. [CrossRef]
24. Yang, H.; Ni, J. Dynamic modeling for machine tool thermal error compensation. *J. Manuf. Sci. Eng.* **2003**, *125*, 245–254. [CrossRef]
25. Creighton, E.; Honegger, A.; Tulsian, A.; Mukhopadhyay, D. Analysis of thermal errors in a high-speed micro-milling spindle. *Int. J. Mach. Tools Manuf.* **2010**, *50*, 386–393. [CrossRef]
26. Chen, J.S. Thermal Error Modeling for Volumetric Error Compensation. *ASME Winter Annu. Meet.* **1992**, *PED-55*, 113–125.
27. Li, S.; Zhang, Y.; Zhang, G. A study of pre-compensation for thermal errors of NC machine tools. *Int. J. Mach. Tools Manuf.* **1997**, *37*, 1715–1719. [CrossRef]

28. Pahk, H.; Lee, S.W. Thermal error measurement and real time compensation system for the CNC machine tools incorporating the spindle thermal error and the feed axis thermal error. *Int. J. Adv. Manuf. Technol.* **2002**, *20*, 487–494. [CrossRef]
29. Shi, H.; Jiang, C.; Yan, Z.; Tao, T.; Mei, X. Bayesian neural network–based thermal error modeling of feed drive system of CNC machine tool. *Int. J. Adv. Manuf. Technol.* **2020**, *108*, 3031–3044. [CrossRef]
30. Haitao, Z.; Jianguo, Y.; Jinhua, S. Simulation of thermal behavior of a CNC machine tool spindle. *Int. J. Mach. Tools Manuf.* **2007**, *47*, 1003–1010. [CrossRef]
31. Wang, Y.-C.; Kao, M.-C.; Chang, C.-P. Investigation on the spindle thermal displacement and its compensation of precision cutter grinders. *Measurement* **2011**, *44*, 1183–1187. [CrossRef]
32. Hu, Y.-C.; Chen, P.-J.; Chang, P.-Z. Thermal-feature system identification for a machine tool spindle. *Sensors* **2019**, *19*, 1209. [CrossRef] [PubMed]
33. Lei, M.; Jiang, G.; Yang, J.; Mei, X.; Xia, P.; Zhao, L. Thermal error modeling with dirty and small training sample for the motorized spindle of a precision boring machine. *Int. J. Adv. Manuf. Technol.* **2017**, *93*, 571–586. [CrossRef]
34. Towards Data Science. Available online: https://towardsdatascience.com/an-intuitive-guide-to-gaussian-processes-ec2f0b45c71d (accessed on 5 August 2021).
35. Rasmussen, C.E. Gaussian Processes in Machine Learning. In *Advanced Lectures on Machine Learning*; Lecture Notes in Computer Science; Bousquet, O., von Luxburg, U., Rätsch, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 63–71. [CrossRef]
36. Singh, A.; Nagar, J.; Sharma, S.; Kotiyal, V. A Gaussian process regression approach to predict the k-barrier coverage probability for intrusion detection in wireless sensor networks. *Expert Syst. Appl.* **2021**, *172*, 114603. [CrossRef]
37. Biau, G.; Scornet, E. A Random Forest Guided Tour. *Test* **2016**, *25*, 197–227. [CrossRef]
38. Louppe, G. Understanding Random Forest—From Theory to Practice. Ph.D. Thesis, University of Liège, Liège, Belgium, 28 July 2014. [CrossRef]
39. Guo, Q.; Yang, J.; Wu, H. Application of ACO-BPN to thermal error modeling of NC machine tool. *Int. J. Adv. Manuf. Technol.* **2010**, *50*, 667–675. [CrossRef]
40. Dexian, Z.; Xiaolian, L.; Hanming, S.; Ri-Yao, C. Identification of position of key thermal susceptible points for thermal error compensation of machine tool by neural network. In Proceedings of the International Conference on Intelligent Manufacturing, Wuhan, China, 28 August 1995. [CrossRef]
41. Gramacy, R.B. *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*, 1st ed.; CRC Press: Lake Wanaka, New Zealand, 2020; pp. 143–222.
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
43. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*, 1st ed.; CRC press: Boca Raton, FL, USA, 1984.