Thermal Expansion Error Prediction of High-speed Electric Spindle

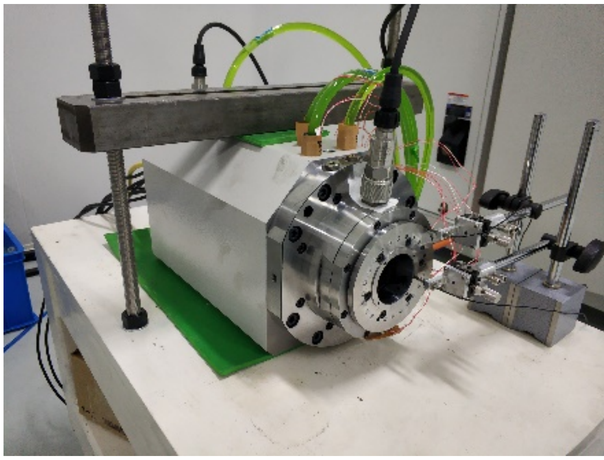
Ruoheng Du

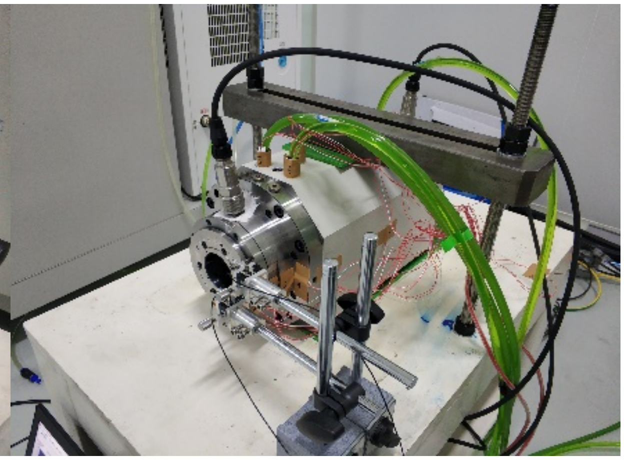
I. Introduction

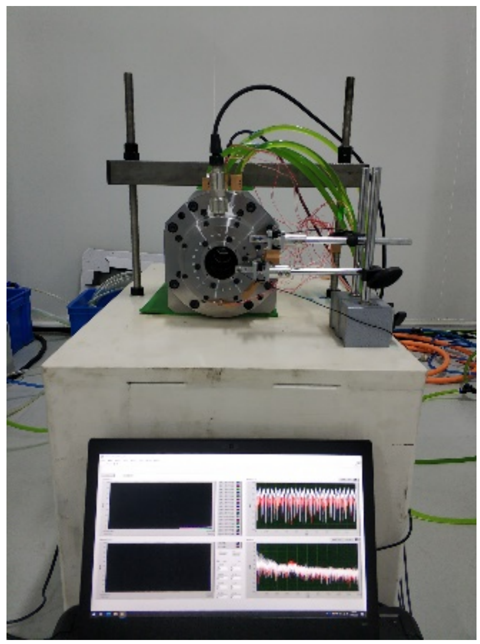
High-speed electric spindles play a critical role in the machining industry, enabling precise and efficient operations. However, the challenge lies in managing their thermal expansion error, which can lead to machining errors and decreased accuracy (Bryan, 1990). To maintain optimal precision in manufacturing processes, it is vital to monitor and forecast the thermal expansion error of these spindles. Modern manufacturing might experience a revolution as a result of the improvements in the detection and compensation of these errors, which would increase output, reduce waste, and improve product quality.

Hence, it becomes imperative to closely monitor and predict thermal expansion errors in high-speed electric spindles to uphold machining precision. Several factors, including spindle design, material properties, environmental conditions, and operating parameters, influence these errors (Li et al., 2019). In this report, it investigates whether the thermal expansion error (i.e., delta L in the following analysis) in the central shaft of electric spindles at the present moment and the current temperature of the surrounding environment as well as the current temperatures at various spindle components could estimate the thermal expansion error at the future moment with some degree of accuracy.

The experiment took place at the School of Mechanical Engineering, Shanghai Jiao Tong University, and the following images illustrate the details of the conducted study.

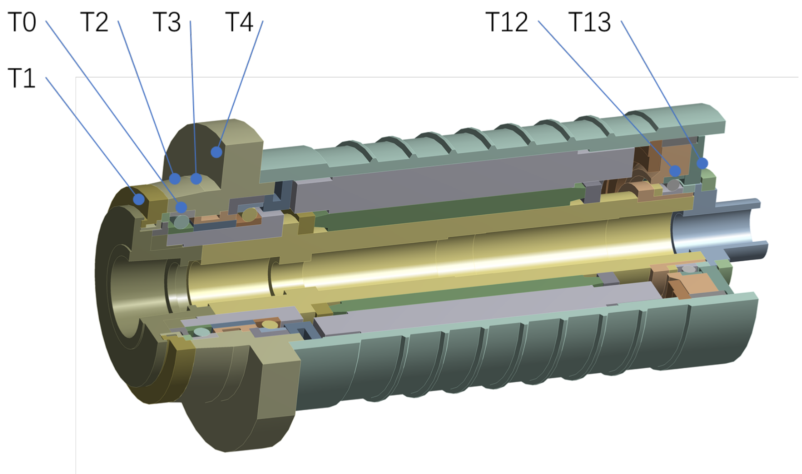






II. Methodology

The experiment comprises a total of 16 temperature sensors, with 7 of them distributed along the central shaft of the electric spindle. The diagram below illustrates the placement of those sensors.



Here, T0 is the temperature measured inside the front bearing, T1 is the temperature of the front bearing housing end face, T2 and T3 represent the temperature of two front bearing housing side, T4 is the temperature of the front bearing housing flange surface, T12 is the temperature inside rear bearing, and T13 is the temperature outside rear bearing. Another 7 temperature sensors are distributed separately around the water inflow and outflow of the front bearing, the water inflow and outflow of the electric motor, and the front, middle and rear parts of the cooling jacket. Besides, the remaining 2 sensors detects the temperature of the worktable (T14) and the surrounding environment (T15).

The following analysis is based on data from a sample of 4682 set of thermal expansion error of the central shaft of the electric spindle and the temperatures, which are collected from the experiment. The data are collected at one-minute intervals, representing the relative change in thermal expansion error per minute under various temperature settings. Besides, the unit of the target variable is micrometer (μm) and all temperatures are measured in Celsius (°C).

Here is the descriptive statistics of the variables.

Descriptive Statistics: delta L, T0, T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, T12, T13, T14, T15

Statistics

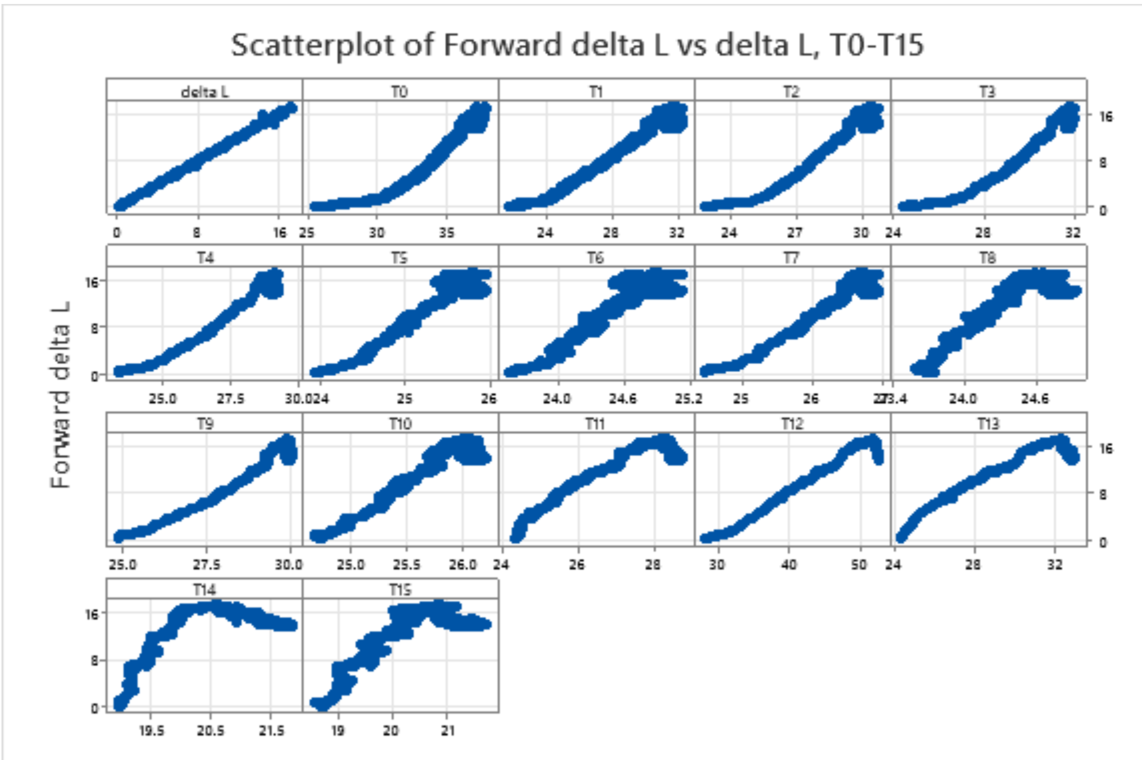
Variable	Total Count	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
delta L	4682	14.050	0.0510	3.490	-0.0088	13.845	15.331	16.105	17.372
T0	4682	36.623	0.0259	1.773	25.361	36.952	37.238	37.432	37.958
T1	4682	30.784	0.0265	1.815	21.765	30.985	31.417	31.704	32.325
T2	4682	29.723	0.0177	1.214	22.864	29.808	30.129	30.358	30.801
T3	4682	31.222	0.0170	1.161	24.433	31.470	31.630	31.754	32.063
T4	4682	28.479	0.0149	1.020	23.326	28.591	28.814	29.006	29.251
T5	4682	25.556	0.00519	0.355	23.916	25.512	25.662	25.769	25.977
T6	4682	24.720	0.00398	0.272	23.552	24.643	24.792	24.901	25.124
T7	4682	26.536	0.00651	0.445	24.452	26.549	26.672	26.779	26.958
T8	4682	24.536	0.00403	0.276	23.585	24.477	24.587	24.737	24.932
T9	4682	29.490	0.0145	0.992	24.823	29.688	29.868	29.951	30.075
T10	4682	25.899	0.00439	0.301	24.662	25.886	25.985	26.084	26.228
T11	4682	27.835	0.0161	1.099	24.366	27.785	28.299	28.497	28.773
T12	4682	49.659	0.0758	5.186	28.130	49.758	52.195	52.436	52.648
T13	4682	31.419	0.0321	2.196	24.286	31.394	32.389	32.728	32.984
T14	4682	20.688	0.0115	0.788	18.951	20.087	20.819	21.336	21.848
T15	4682	20.619	0.0102	0.697	18.550	20.312	20.691	21.170	21.763

As we can see, delta L has a mean of 14.050µm and a standard deviation of 3.490µm, which means that on average, the thermal expansion error of the central shaft of the electric spindle is 14.050µm. For other predicting variables (i.e., the detected values of the temperature sensors), the mean values from T0 to T15 are 36.5623µm, 30.784µm, 29.723µm, 31.222µm, 28.479µm, 25.556µm, 24.720µm, 26.536µm, 24.536µm, 29.490µm, 25.899µm, 27.835µm, 49.659µm, 31.419µm, 20.688µm, and 20.619µm, respectively. The standard deviation values are 1.773µm, 1.815µm, 1.214µm, 1.161µm, 1.020µm, 0.355µm, 0.272µm, 0.445µm, 0.276µm, 0.992µm, 0.301µm, 1.099µm, 5.186µm, 2.196µm, 0.788µm, and 0.697µm, respectively.

Taking into consideration the characteristics of the data presented above, this report will continue with an evaluation into the effectiveness of both multiple linear regression and deep learning techniques in the ensuing analysis, and compare how well these techniques individually do at predicting thermal errors.

III. Results

a) Multiple Linear Regression



As can be seen from the scatterplots, there does appear to be a positive linear relationship between almost each of all the 17 predicting variables and the target variable forward delta L, although these relationships do not appear to be absolutely linear and the possibility of a more complicated regression with deep learning techniques will be discussed later.

As a result, the relationship between delta L and temperatures of the environment and different parts of the electric spindle can be expressed by a multiple regression model:

$$forward\ delta\ L = \beta_o + \beta_1 * T0 + \beta_2 * T1 + \beta_3 * T2 + \beta_4 * T3 + \beta_5 * T4 + \beta_6 * T5 + \beta_7 * T6 + \beta_8 * T7 + \beta_9 * T8 + \beta_{10} * T9 + \beta_{11} * T10 + \beta_{12} * T11 + \beta_{13} * T12 + \beta_{14} * T13 + \beta_{15} * T14 + \beta_{16} * T15 + \beta_{17} * delta\ L + random\ error.$$

The below is the least squares regression with all the seventeen predicting variables.

THERMAL EXPANSION ERROR

Regression Analysis: Forward delta L versus T0, T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, T12, T13, T14, T15, delta L

Regression Equation

Forward delta L = 3.79 + 0.0171 T0 - 0.0009 T1 + 0.0128 T2 + 0.0258 T3 - 0.0146 T4 - 0.1414 T5 + 0.139 T6 - 0.0003 T7 - 0.0726 T8 - 0.0013 T9 - 0.0674 T10 - 0.0622 T11 + 0.0144 T12 + 0.1554 T13 - 0.2457 T14 + 0.0374 T15 + 0.93215 delta L

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	3.79	1.22	(1.41, 6.17)	3.12	0.002	
T0	0.0171	0.0136	(-0.0096, 0.0438)	1.25	0.210	175.21
T1	-0.0009	0.0131	(-0.0267, 0.0248)	-0.07	0.944	170.51
T2	0.0128	0.0309	(-0.0478, 0.0734)	0.41	0.679	422.46
T3	0.0258	0.0450	(-0.0624, 0.1140)	0.57	0.567	817.70
T4	-0.0146	0.0709	(-0.1536, 0.1244)	-0.21	0.837	1567.10
T5	-0.1414	0.0939	(-0.3255, 0.0427)	-1.51	0.132	333.55
T6	0.139	0.104	(-0.064, 0.342)	1.34	0.180	238.53
T7	-0.0003	0.0836	(-0.1642, 0.1635)	-0.00	0.997	415.27
T8	-0.0726	0.0531	(-0.1767, 0.0316)	-1.37	0.172	64.26
T9	-0.0013	0.0457	(-0.0910, 0.0884)	-0.03	0.977	616.91
T10	-0.0674	0.0741	(-0.2127, 0.0780)	-0.91	0.363	148.86
T11	-0.0622	0.0570	(-0.1739, 0.0496)	-1.09	0.275	1175.96
T12	0.0144	0.0156	(-0.0161, 0.0449)	0.92	0.356	1951.98
T13	0.1554	0.0435	(0.0702, 0.2407)	3.57	0.000	2734.12
T14	-0.2457	0.0232	(-0.2912, -0.2002)	-10.58	0.000	100.25
T15	0.0374	0.0144	(0.0092, 0.0656)	2.60	0.009	30.09
delta L	0.93215	0.00532	(0.92172, 0.94258)	175.19	0.000	103.38

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AICc	BIC
0.124949	99.87%	99.87%	73.4149	99.87%	-6168.62	-6046.20

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	17	56734.8	99.87%	56734.8	3337.34	213763.04	0.000
T0	1	50711.4	89.27%	0.0	0.02	1.57	0.210
T1	1	75.4	0.13%	0.0	0.00	0.00	0.944
T2	1	303.5	0.53%	0.0	0.00	0.17	0.679
T3	1	18.9	0.03%	0.0	0.01	0.33	0.567
T4	1	47.4	0.08%	0.0	0.00	0.04	0.837
T5	1	9.6	0.02%	0.0	0.04	2.27	0.132
T6	1	406.7	0.72%	0.0	0.03	1.80	0.180
T7	1	25.1	0.04%	0.0	0.00	0.00	0.997
T8	1	434.9	0.77%	0.0	0.03	1.86	0.172
T9	1	560.7	0.99%	0.0	0.00	0.00	0.977
T10	1	797.6	1.40%	0.0	0.01	0.83	0.363
T11	1	384.3	0.68%	0.0	0.02	1.19	0.275
T12	1	68.6	0.12%	0.0	0.01	0.85	0.356
T13	1	659.7	1.16%	0.2	0.20	12.78	0.000
T14	1	1724.7	3.04%	1.7	1.75	111.98	0.000
T15	1	27.3	0.05%	0.1	0.11	6.76	0.009
delta L	1	479.1	0.84%	479.1	479.15	30690.24	0.000
Error	4664	72.8	0.13%	72.8	0.02		
Total	4681	56807.6	100.00%				

Overall, the F-statistic of this regression has a p-value of 0.000, which is statistically significant, and this regression is quite strong as the R-squared value and adjusted R-squared value are both 99.87%. In other words, T0, T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, T12, T13, T14, T15, and delta L are able to account well for the observed variability in forward delta L. The constant coefficient indicates that when all the temperature variables equal zero Celsius and the current thermal expansion error is zero, the estimated expected forward delta L is 3.79µm (that is, the thermal expansion error of the central shaft of the electric spindle in the next moment is expected to be 3.79µm). But this point does not have any practical interpretation because none of the predicting variables take on values close to zero, and this should not be discussed.

As for T0, T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, T12, T13, T14, T15, and delta L coefficients, in general, for one specific temperature coefficient, holding all other coefficients constant, 1 °C change in the temperature of this coefficient is associated with a certain micrometer estimated expected increase or decrease in the variation of the thermal expansion error. Take T15 coefficient as an example. It indicates that holding other predicting variables constant (i.e., holding T0, T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, T12, T13, T14, and delta L constant), 1 °C change in the temperature of the environment is associated with an estimated expected 0.0374μm change in the forward delta L (i.e., the thermal expansion error of the central shaft of the electric spindle in the future moment). The standard error of the estimate of 0.124949μm says that this model could be used to predict delta L within ±0.249898μm, roughly 95% of the time.

However, not all the coefficients are statistically significant. As for the T15 coefficient, with the null hypothesis of $\beta_{16} = 0$ and the alternative hypothesis of $\beta_{16} \neq 0$, the t–statistic for the slope is:

$$t = \frac{\widehat{\beta_{16}} - Hypothesized\ Value}{\delta_{\widehat{\beta_{16}}}} = \frac{0.0374 - 0}{0.0144} = 2.60.$$

This is statistically significant compared with the critical value of 2.58. And the p-value of this coefficient is 0.009, which is greater than 0.01. So, we are 99% confident that the null hypothesis should be rejected and the alternative hypothesis of $\beta_{16} \neq 0$ should be accepted. However, T6 coefficient is not statistically significant, and its t–statistic for the slope is:

$$t = \frac{\widehat{\beta_7} - Hypothesized\ Value}{\delta_{\widehat{\beta_7}}} = \frac{0.139 - 0}{0.104} = 1.34.$$

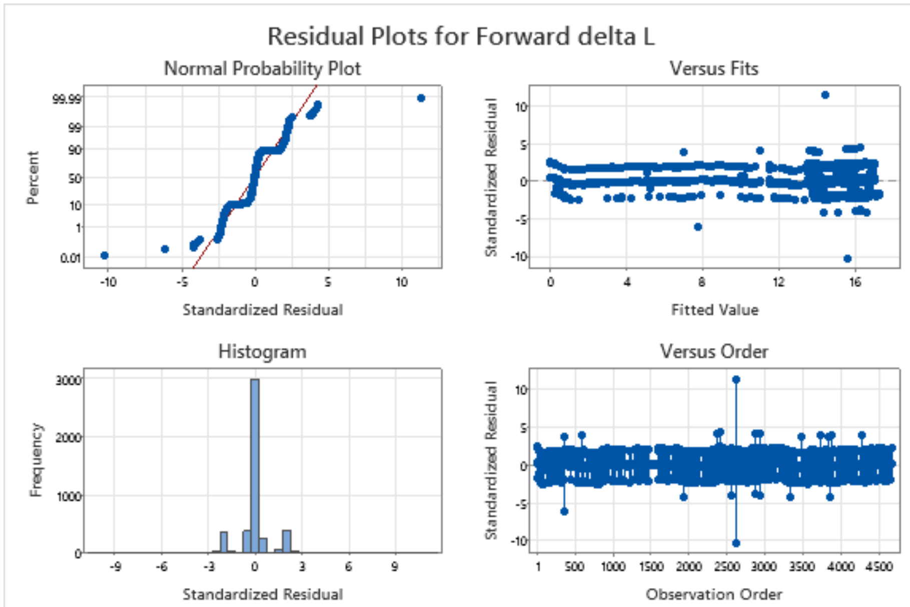
The t-statistics value of 1.34 is less than 1.645, and p-value is 0.180, which is greater than 0.10. So, we are 90% confident that the null hypothesis of $\beta_7 = 0$ could be accepted.

Besides, multicollinearity does seem to be a problem in this regression, as shown by the VIF value of all the variables above. The general guideline for VIF value is that

$$\max\left(10, \frac{1}{1 - R^2_{model}}\right) = max\left(10, \frac{1}{1 - 99.87\%}\right) = \max(10, 769.23) = 769.23.$$

So, some of the VIF values above, such as T3 coefficient, are problematic because they are greater than 769.23.

Next, the “four in one” plot will be analyzed to see if the assumptions of least squares regression hold true.



There’s insufficient evidence to say that the first assumption of zero expected value is violated and the error has a non-zero expected value because there does not exist specific subgroups with residuals that systematically appear to be deviated from the zero line. As for the second assumption of homoscedasticity, the points are slightly more concentrated in the right half of the plot of standardized residuals versus fitted values, which does not exhibit a lack of pattern, and therefore, indicates nonconstant variance and violation of this assumption. The third assumption suggests that the errors should be uncorrelated with each other, and this is violated because forward delta L is not only very much influenced by the temperature variables in this model, but also affected by the current delta L. However, the fourth assumption, that errors are normally distributed, is slightly violated. As we can see from the normal probability plot, some points on the top-right and bottom-left deviates from the line, indicating non-normality.

Given the analysis above, a model selection is performed to select relevant predicting variables more systematically, and thus, improve the model.

Response is Forward delta L

Vars	R-Sq	R-Sq (adj)	PRESS	R-Sq (pred)	Mallows Cp	S	AICc	BIC	Cond No	L	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
1	99.9	99.9	75.4	99.9	144.5	0.12684	-6044.309	-6024.960	1.000	X																
1	89.3	89.3	6115.6	89.2	385795.7	1.1413	14528.717	14548.066	1.000		X															
2	99.9	99.9	75.3	99.9	138.3	0.12674	-6050.231	-6024.433	35.336	X	X															
2	99.9	99.9	75.3	99.9	140.3	0.12677	-6048.320	-6022.522	30.085	X				X												
3	99.9	99.9	74.1	99.9	61.8	0.12572	-6125.217	-6092.973	515.294	X													X		X	
3	99.9	99.9	74.2	99.9	67.0	0.12579	-6120.093	-6087.849	683.517	X														X	X	
4	99.9	99.9	73.5	99.9	22.5	0.12518	-6164.225	-6125.534	1073.988	X	X													X	X	
4	99.9	99.9	73.7	99.9	34.4	0.12534	-6152.435	-6113.744	1004.151	X													X	X	X	
5	99.9	99.9	73.2	99.9	7.6	0.12497	-6179.170	-6134.033	3937.104	X	X											X		X	X	
5	99.9	99.9	73.4	99.9	16.2	0.12509	-6170.514	-6125.378	1544.570	X	X										X			X	X	
6	99.9	99.9	73.3	99.9	7.7	0.12496	-6179.067	-6127.486	4876.495	X	X								X			X		X	X	
6	99.9	99.9	73.3	99.9	8.2	0.12497	-6178.537	-6126.956	5098.943	X	X										X	X		X	X	
7	99.9	99.9	73.2	99.9	4.3	0.12490	-6182.444	-6124.419	6346.394	X	X										X		X	X	X	X
7	99.9	99.9	73.2	99.9	5.6	0.12492	-6181.123	-6123.098	5648.477	X	X								X			X		X	X	X
8	99.9	99.9	73.2	99.9	4.2	0.12489	-6182.493	-6118.025	12381.989	X	X										X	X	X	X	X	X
8	99.9	99.9	73.2	99.9	5.3	0.12490	-6181.463	-6116.995	7337.202	X	X							X			X		X	X	X	X
9	99.9	99.9	73.2	99.9	5.9	0.12489	-6180.835	-6109.925	20012.436	X	X								X		X	X	X	X	X	X
9	99.9	99.9	73.2	99.9	5.9	0.12489	-6180.776	-6109.866	14971.732	X	X		X								X	X	X	X	X	X
10	99.9	99.9	73.2	99.9	5.7	0.12488	-6181.016	-6103.665	16708.786	X	X					X	X		X			X	X	X	X	X
10	99.9	99.9	73.2	99.9	6.2	0.12488	-6180.539	-6103.188	8960.366	X	X			X		X	X		X			X		X	X	X
11	99.9	99.9	73.2	99.9	7.1	0.12488	-6179.635	-6095.843	28722.916	X	X					X	X		X		X	X	X	X	X	X
11	99.9	99.9	73.2	99.9	7.3	0.12489	-6179.388	-6095.597	20950.492	X	X		X			X	X		X			X	X	X	X	X
12	99.9	99.9	73.3	99.9	8.3	0.12489	-6178.365	-6088.134	35438.532	X	X			X		X	X		X		X	X	X	X	X	X
12	99.9	99.9	73.3	99.9	8.4	0.12489	-6178.306	-6088.075	32286.603	X	X		X			X	X		X		X	X	X	X	X	X
13	99.9	99.9	73.3	99.9	10.2	0.12490	-6176.525	-6079.856	38736.495	X	X		X	X		X	X		X		X	X	X	X	X	X
13	99.9	99.9	73.3	99.9	10.2	0.12490	-6176.499	-6079.830	39429.197	X	X			X		X	X		X	X	X	X	X	X	X	X
14	99.9	99.9	73.3	99.9	12.0	0.12491	-6174.659	-6071.552	42884.164	X	X		X	X	X	X	X		X		X	X	X	X	X	X
14	99.9	99.9	73.3	99.9	12.1	0.12491	-6174.606	-6071.499	42676.476	X	X		X	X		X	X		X	X	X	X	X	X	X	X
15	99.9	99.9	73.4	99.9	14.0	0.12492	-6172.650	-6063.106	54540.621	X	X	X	X	X	X	X	X		X		X	X	X	X	X	X
15	99.9	99.9	73.4	99.9	14.0	0.12492	-6172.646	-6063.102	47855.798	X	X		X	X	X	X	X	X	X		X	X	X	X	X	X
16	99.9	99.9	73.4	99.9	16.0	0.12494	-6170.636	-6054.656	68348.547	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X
16	99.9	99.9	73.4	99.9	16.0	0.12494	-6170.635	-6054.655	58552.343	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X
17	99.9	99.9	73.4	99.9	18.0	0.12495	-6168.620	-6046.205	78415.830	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Based on the above model selection outputs by Minitab and the selection criteria of high R^2_{adj}, small Mallows C_p, high R^2_{pred}, small AIC_c, and small BIC, the highlighted row is the regression model with seven predicting variables that has the best performance among all. It takes delta L, T0, T10, T12, T13, T14, and T15 as the predicting variables, and the below is the regression output.

Regression Analysis: Forward delta L versus delta L, T0, T10, T12, T13, T14, T15

Regression Equation

Forward delta L = 3.694 + 0.02321 T0 - 0.1495 T10 + 0.02211 T12 + 0.1035 T13 - 0.2436 T14 + 0.0456 T15 + 0.93447 delta L

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	3.694	0.774	(2.176, 5.212)	4.77	0.000	
T0	0.02321	0.00798	(0.00756, 0.03885)	2.91	0.004	60.10
T10	-0.1495	0.0342	(-0.2165, -0.0825)	-4.38	0.000	31.64
T12	0.02211	0.00875	(0.00496, 0.03927)	2.53	0.012	617.90
T13	0.1035	0.0185	(0.0673, 0.1397)	5.60	0.000	493.70
T14	-0.2436	0.0210	(-0.2848, -0.2023)	-11.58	0.000	82.36
T15	0.0456	0.0115	(0.0231, 0.0681)	3.97	0.000	19.19
delta L	0.93447	0.00520	(0.92427, 0.94466)	179.65	0.000	98.87

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AICc	BIC
0.124900	99.87%	99.87%	73.1975	99.87%	-6182.44	-6124.42

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	7	56734.7	99.87%	56734.7	8104.95	519551.41	0.000
T0	1	50711.4	89.27%	0.1	0.13	8.46	0.004
T10	1	128.1	0.23%	0.3	0.30	19.14	0.000
T12	1	933.6	1.64%	0.1	0.10	6.39	0.012
T13	1	56.2	0.10%	0.5	0.49	31.42	0.000
T14	1	4348.9	7.66%	2.1	2.09	134.10	0.000
T15	1	53.0	0.09%	0.2	0.25	15.77	0.000
delta L	1	503.5	0.89%	503.5	503.46	32273.58	0.000
Error	4674	72.9	0.13%	72.9	0.02		
Total	4681	56807.6	100.00%				

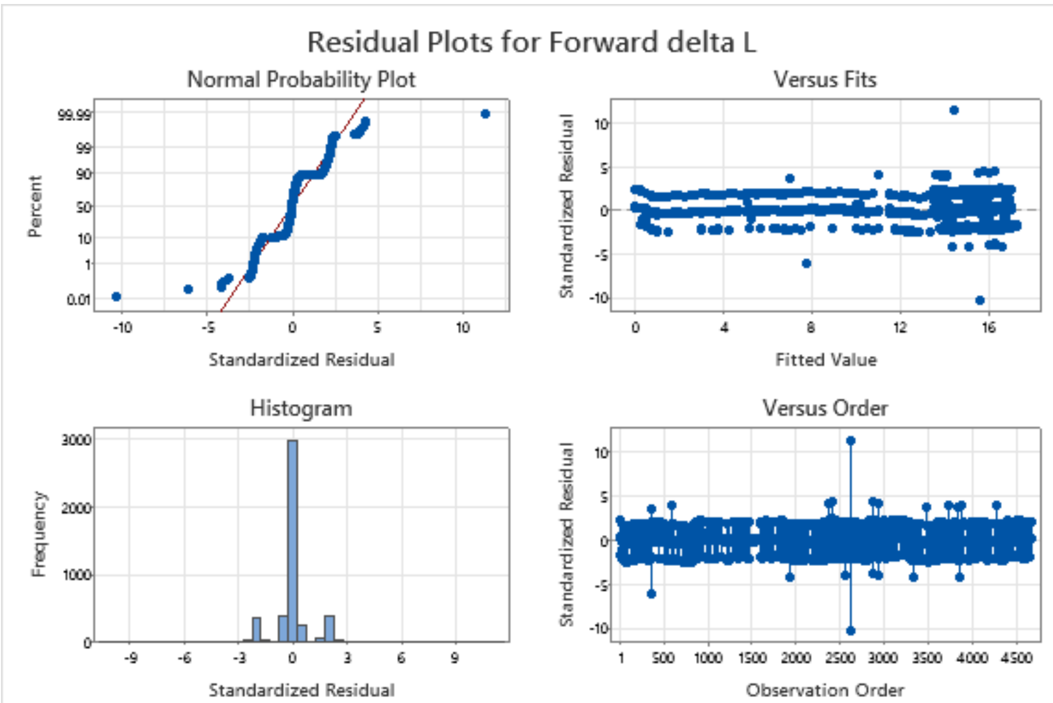
This regression is still strong as the R-squared value and adjusted R-squared value are the same as before, which is 99.87%. In other words, delta L, T0, T10, T12, T13, T14, and T15 are able to account for the observed variability in forward delta L well. This time, all the seven coefficients are statistically significant as all the absolute values of the t-statistics are greater than 1.96, which is the critical value of the 5% significance level, and all the p-values are less than 0.05.

Besides, multicollinearity is not a problem in this regression. The new general guideline for VIF value is that

$$\max\left(10, \frac{1}{1-R^2_{model}}\right) = max\left(10, \frac{1}{1-99.87\%}\right) = \max(10, 769.23) = 769.23.$$

And as shown above, the VIF values for all predicting variables are less than 769.23.

The "four in one" plot of the new regression is shown below.



This "four in one" plot is almost the same as the one with seventeen predicting variables before. The second assumption of homoscedasticity is still somewhat violated as the points are still more concentrated in the right half of the plot of standardized residuals versus fitted values, indicating nonconstant variance. And the existence of delta L in predicting variables still fails to fulfill the third assumption. Besides, the fourth assumption of the normal distribution of errors is still slightly violated by top-right and bottom-left points in the normal probability plot.

To take a further look at the current two models, a F-test is needed to examine whether restricting ten predicting variables is meaningful. This model, $forward\ delta\ L\ =\ \beta_o + \beta_1 *$

$T0 + \beta_2 * T10 + \beta_3 * T12 + \beta_4 * T13 + \beta_5 * T14 + \beta_6 * T15 + \beta_7 * delta\ L +$ $random\ error$, is a simpler model (i.e., restricted model), which is a special case of the previous full model (i.e., unrestricted model), $forward\ delta\ L = \beta_o + \beta_1 * T0 + \beta_2 * T1 + \beta_3 * T2 + \beta_4 * T3 + \beta_5 * T4 + \beta_6 * T5 + \beta_7 * T6 + \beta_8 * T7 + \beta_9 * T8 + \beta_{10} * T9 + \beta_{11} * T10 + \beta_{12} * T11 + \beta_{13} * T12 + \beta_{14} * T13 + \beta_{15} * T14 + \beta_{16} * T15 + \beta_{17} * delta\ L +$ $random\ error$, with seventeen predicting variables. The F-test can be formulated as a hypothesis testing problem of

$$H_0: \beta_{2_{full}} = \cdots = \beta_{10_{full}} = \beta_{12_{full}} = 0$$

versus

$$H_a: at\ least\ one\ of\ \beta_{2_{full}}, \dots, \beta_{10_{full}}, \beta_{12_{full}} \neq 0\ .$$

The F-statistics can be calculated as

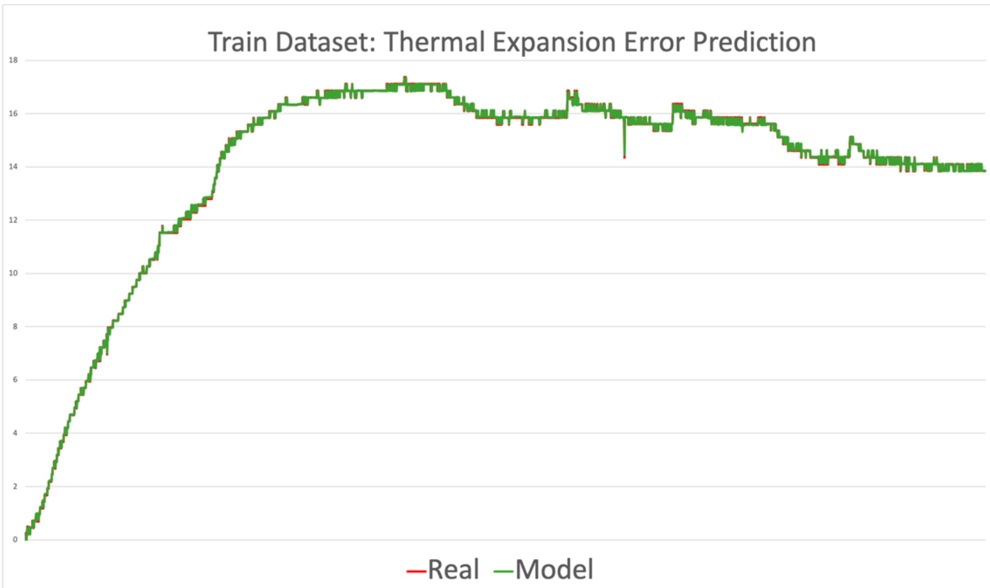
$$F = \frac{\frac{R^2_{full} - R^2_{subset}}{10}}{\frac{1 - R^2_{full}}{4682 - 17 - 1}} = \frac{\frac{99.87\% - 99.87\%}{10}}{\frac{1 - 99.87\%}{4682 - 17 - 1}} = 0$$

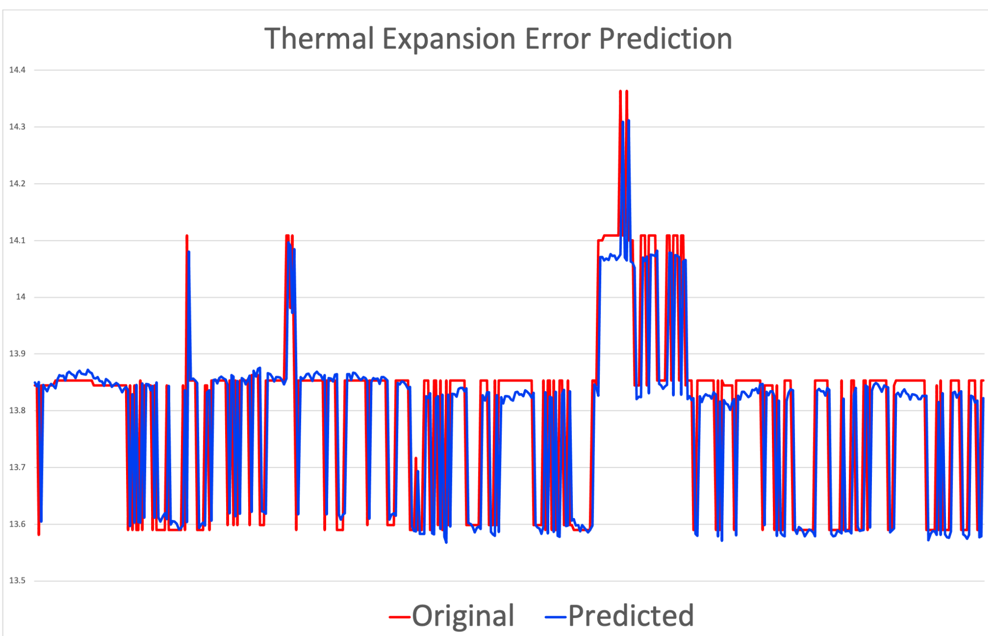
on (10, 4664) degree of freedom, which is smaller than the critical value of 1.83 for significance level of 5%. This result fails to reject the null hypothesis.

Furthermore, the calculation of the root mean square error (RMSE) for the original dataset and the predicted values may provide a tool to assess the regression models' ability to predict future values.

Number of Predicting Variables	Train RMSE	Test RMSE
7	0.124	0.133
17	0.124	0.133

Just to clarify, the dataset has been split into two parts here, which correspond to the training set and the testing set in the subsequent deep learning model, in order to preserve a standard comparing criterion. The first 90% of the data belongs to the training set, and the last 10% belongs to the testing set. As we can see from the table, there's not much difference between the two regression models with different numbers of the predicting variables. Both could reach a train RMSE of 0.124 and a test RMSE of 0.133. And below are the comparison charts.

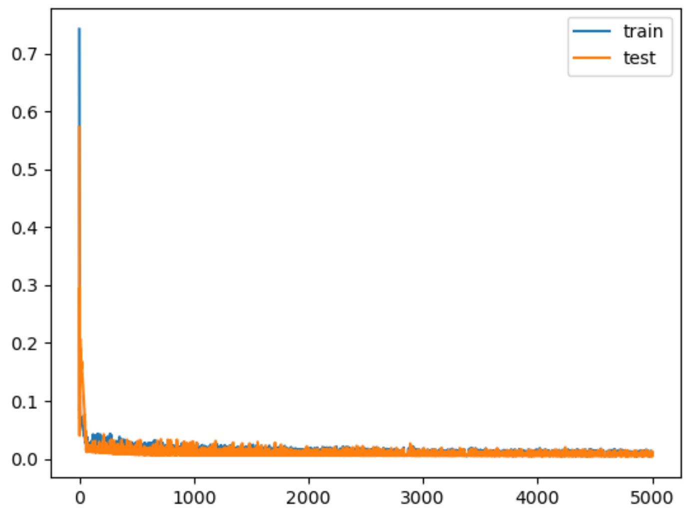


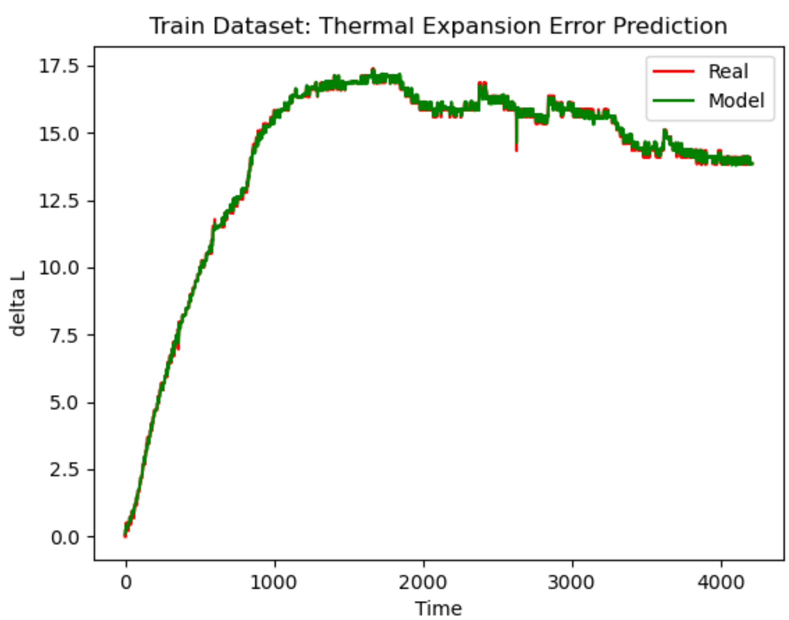


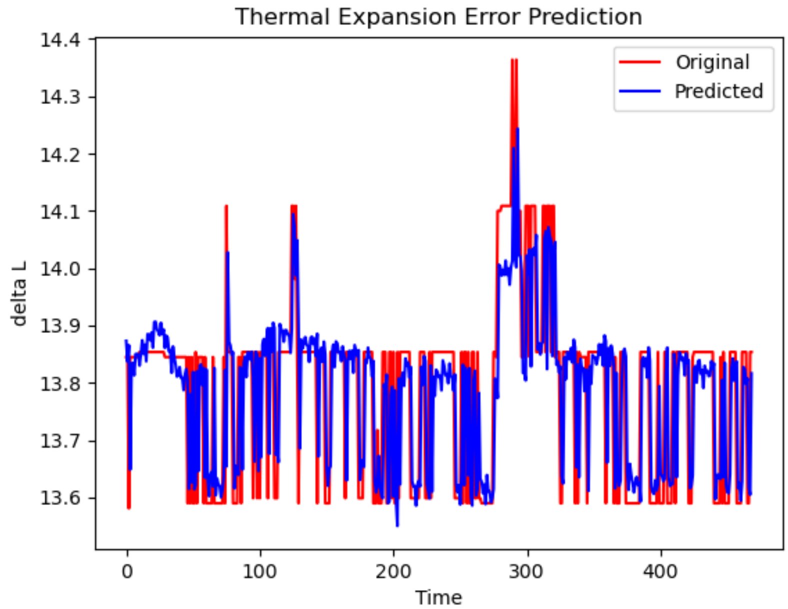
b) Long Short-Term Memory

In the area of deep learning, LSTM is a type of recurrent neural network. It possesses the ability to gather and retain information across extended sequences, allowing it to adeptly comprehend connections and structures within data featuring time gaps.

This report employs a Python-incorporated Keras function. This model consists of 64 hidden layers, a batch size of 512, and underwent a total of 5000 epochs. It has a train RMSE of 0.126 and a test RMSE of 0.127. The graph of the loss function and the result are presented below.







c) Comparison

This report conducts a comparison between the multiple linear regression model and the LSTM model by evaluating their respective train RMSE values and test RMSE values.

Techniques	Train RMSE	Test RMSE
Multiple Linear Regression	0.124	0.133
Long Short-Term Memory	0.126	0.127

As shown above, despite having a train RMSE that is marginally higher than that of the multiple linear regression model, the LSTM model's test RMSE value of 0.127 is noticeably lower. This suggests that the LSTM model might be able to predict data more accurately in both training and testing sets, particularly when it comes to the last 10% of the data. In other words, the multiple linear regression model might not be able to predict the future thermal expansion error as well as the LSTM model. The LSTM model therefore performs better in this case.

IV. Conclusion

To summarize, investigating how to predict the forward thermal expansion error of the central shaft of the electric spindle through the current thermal expansion error and the temperatures of the environment and different parts of the electric spindle is crucial for maintaining high precision in machining processes. In this report, with a train RMSE of 0.126 and a test RMSE of 0.127, the LSTM network has a better performance than the multiple linear regression. By studying this, researchers can develop effective compensation methods, optimize spindle design and operating parameters, thus improving machining accuracy.

Works Cited

Bryan, J. (1990). International status of Thermal Error Research (1990). *CIRP Annals, 39*(2), 645–656. https://doi.org/10.1016/s0007-8506(07)63001-7

Li, B., Tian, X., & Zhang, M. (2019). Thermal error modeling of machine tool spindle based on the improved algorithm optimized BP Neural Network. *The International Journal of Advanced Manufacturing Technology, 105*(1-4), 1497–1505. https://doi.org/10.1007/s00170-019-04375-w