machines

MDPI

*Article*

# Deep Learning to Directly Predict Compensation Values of Thermally Induced Volumetric Errors

**Huy Vu Ngoc *, J. R. R. Mayer and Elie Bitar-Nehme**

Mechanical Engineering Department, École Polytechnique de Montréal, P.O. Box 6079, Station Downtown, Montréal, QC H3C 3A7, Canada
* Correspondence: huy.vu-ngoc@polymtl.ca

**Abstract:** The activities of the rotary axes of a five-axis machine tool generate heat causing temperature changes within the machine that contribute to tool center point (TCP) deviations. Real time prediction of these thermally induced volumetric errors (TVEs) at different positions within the workspace may be used for their compensation. A Stacked Long Short Term Memories (SLSTMs) model is proposed to find the relationship between the TVEs for different axis command positions and power consumptions of the rotary axes, machine's linear and rotary axis positions. In addition, a Stacked Gated Recurrent Units (SGRUs) model is also used to predict some cases, which are the best and the worst predictions of SLSTMs to know the abilities of their predictions. Training data come from a long motion activity experiment lasting 132 h (528 measuring cycles). Adaptive moment with decoupled weight decay (AdamW) optimizer is used to strengthen the models and increase the quality of prediction. Multistep ahead prediction in the testing phase is applied to seven positions not used for training in the long activity sequence and 31 positions in a different short activity sequence of the rotary axes lasting a total of 40 h (160 cycles) to test the ability of the trained model. The testing phase with SLSTMs yields fittings between the predicted values and measured data (without using the measured values as targets) from 69.2% to 98.8%. SGRUs show performance similar to SLSTMs with no clear winner.

**Keywords:** volumetric errors; deep learning; machine tool; thermal errors



**Citation:** Ngoc, H.V.; Mayer, J.R.R.; Bitar-Nehme, E. Deep Learning to Directly Predict Compensation Values of Thermally Induced Volumetric Errors. *Machines* **2023**, *11*, 496. https://doi.org/10.3390/machines11040496

Academic Editor: Gianni Campatelli

Received: 14 March 2023
Revised: 12 April 2023
Accepted: 18 April 2023
Published: 20 April 2023



## 1. Introduction

Machine tool accuracy may degrade significantly due to thermal effects resulting partly from heat generated by the drives and the motions of rotary axes [1]. One way to reduce these errors is real time compensation to cancel TCP deviation at an arbitrary point in the work space by adjusting its command position [2]. Mayr et al. stated that precise predictions of thermal errors through models are fundamental for effective real time compensation [3]. The higher the predictability of the thermal error models, the more accurate the compensation will be.

Brecher et al. developed first and second order transfer functions between the thermal volumetric errors (TVEs) of a three-axis machine tool as outputs and the internal control data (rotational speeds and motor currents) and environment temperature as inputs [4]. Horejs et al. compensated thermal displacements in the y and z directions of the TCP through thermal transfer functions that were achieved using Matlab's System Identification Toolbox. The inputs are the temperatures from inside the main spindle, ram and base of the horizontal milling machine, ambient temperature and spindle speed and the outputs are the thermal displacements [5]. Liu at al. modelled thermally induced volumetric errors based on the 21 geometric error components and nine thermal drift errors for milling and boring machine using rigid body kinematics. Only volumetric error in the z direction identified as the main error was compensated with Programmable Logic Controller and the machine's computer numerical controller (CNC) [6]. Mayr et al. found compensation

values for individual thermal errors of rotary axis and main spindle from a gray box model using cooling power as input [7]. Yu et al. investigated the dependence of volumetric errors on spindle speed and temperature at pre-defined points through polynomial regression model. The model with closest predictability to the measured outputs is chosen [8]. Bitar-Nehme and Mayer developed first order transfer functions with delay terms to predict individual machine geometric errors as output from power consumptions as input. The machine table thermal expansion was also predicted [1]. Baum et al. modeled thermally induced volumetric error from the 21 geometric errors components for a three-axis machine tool through rigid body kinematics. Error components were measured with integral deformation sensors in the training process and the R-test based procedure in the testing process [9]. Liu et al. established a mechanism-based error model to identify three thermally induced error terms of linear axes X, Y, and Z as functions of time and positions. With spindle system, they modelled relationship between two angular thermal errors yaw and pitch, two thermally translational errors in x and y directions, and spindle elongation with temperatures of critical key locations by multivariate linear regression analysis [10].

The above models show their ability to predict the TVEs (data sequence) of the machine tools. However, the volume and variety of data become bigger; the model has more clues to predict outputs better, but the calculation cost and internal parameters increase. In some cases are over the capability of manual analysis [11,12]. Elaborate training of the model is expected to be able to guarantee reliable results in the long term [13]. Deep learning technique enables learning complex relations and is a new approach for these problems. LSTM and GRU networks have attracted attention in recent time and Liang et al. proved that they have a better quality than Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) for processing sequence data and avoiding overfitting problem [14]. Ngoc et al. [11] applied LSTM to predict multi step ahead data of thermally induced geometric errors, which define the errors between and within individual machine axes, from activity cycles involving the two rotary axes B and C of a five-axis machine tool. Others [15–18] presented the efficiency of LSTM networks for modelling the thermal behaviors of a machine tool. Refs. [15,16] modeled the thermal elongation of spindle as a function of rotational speed. Ref. [17] used thermal elongation of spindle in the z direction as output and temperature at key locations as inputs. Ref. [18] found the relationship between thermal errors of ball screws and temperature of the feed drive system. While in [11] thermally induced geometric errors (TGEs) were predicted using SLSTM, the current paper applies SLSTM and SGRU to directly predict the TVEs without the use of a rigid body kinematic model. This avoids having to develop the machine kinematic model and may be able to model error sources that a kinematic model may not consider. Calculating them from predicted TGEs does not consider errors that cannot be explained with the estimated geometric errors [1,19]. TVEs directly relate to compensation values of the thermal errors of paths to move the tool to target positions in the working space of the machine tool. These predictions are more difficult than doing TGEs because the predictors need to realize not only the change of thermal behavior of machine, but also the positions we want to move the tool to. Therefore, the input data include not only the power consumptions of axes B and C, but also the machine linear and rotary axis positions. Increasing the number of inputs also increases the calculation cost. With a multivariate and multi-layer network, it may be trapped within multiple local minima; Sagheer and Kotb suggest having a suitable optimizer to facilitate the training of deep learning models [20]. Stochastic optimization is successfully applied in Deep learning [21] and Adam [22] is a popular optimizer. Adaptive moment with decoupled weight decay (AdamW) optimizer [23] is a modification of Adam with flexible learning rate, decoupling the weight decay from the gradient update, is expected to make the networks converge faster and improve accuracy. The proposed scheme is presented in six sections: Section 1 reviews previous works. Section 2 describes the experimental process. Section 3 presents the neural network configurations of the SLSTMs and SGRUs models. Section 4 shows the training and predicting processes. The results and discussion are in Section 5 and the conclusion follows in Section 6.

## 2. Experiment Process

Motion sequences for machine axes were designed to warm up the machine tool without any machining [1,24], while measuring the TVEs, or TCP displacements between a contactless R-test device (sensor nest) [25] and ball artefacts on a Mitsui-Seiki HU40T horizontal machining center with wCBXfZY(S)t topology as described in [1]. Figure 1 illustrates the experimental setup. Four separate artefact balls are fixed on stems with different lengths and a scale bar made of Invar with two balls separated by a calibrated distance on the workpiece side, while the sensor nest is mounted in the spindle on tool side.

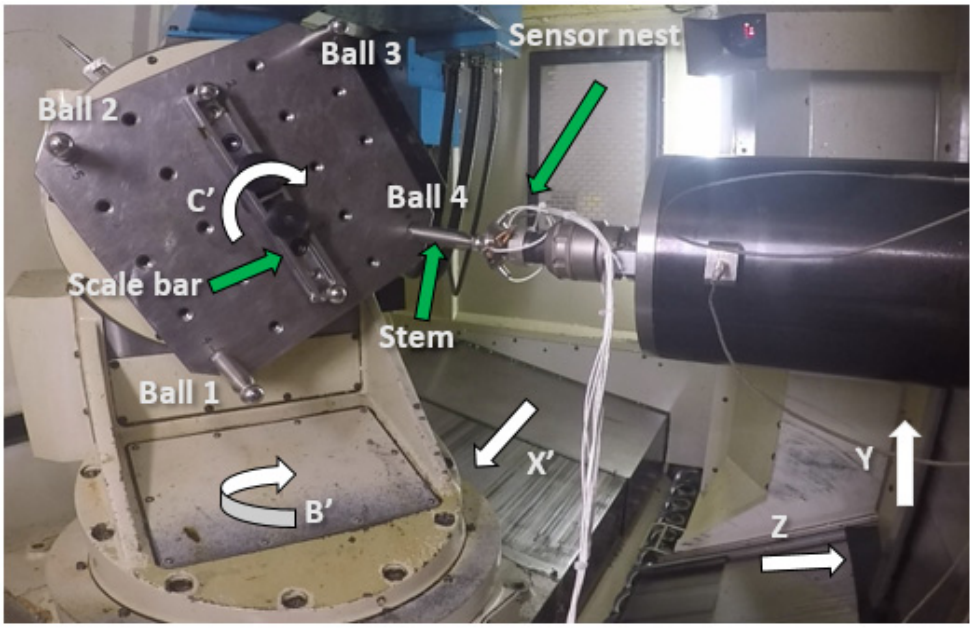


**Figure 1.** Experimental setup.

Heating (rotary axes' motions at different speeds) and cooling (machine stopped) cycles are shown in Figure 2 (on the left). Figure 2 (on the right) illustrates ambient temperature controlled between 22.1 and 23.5 Celsius degrees in 132 h (528 cycles) training process.

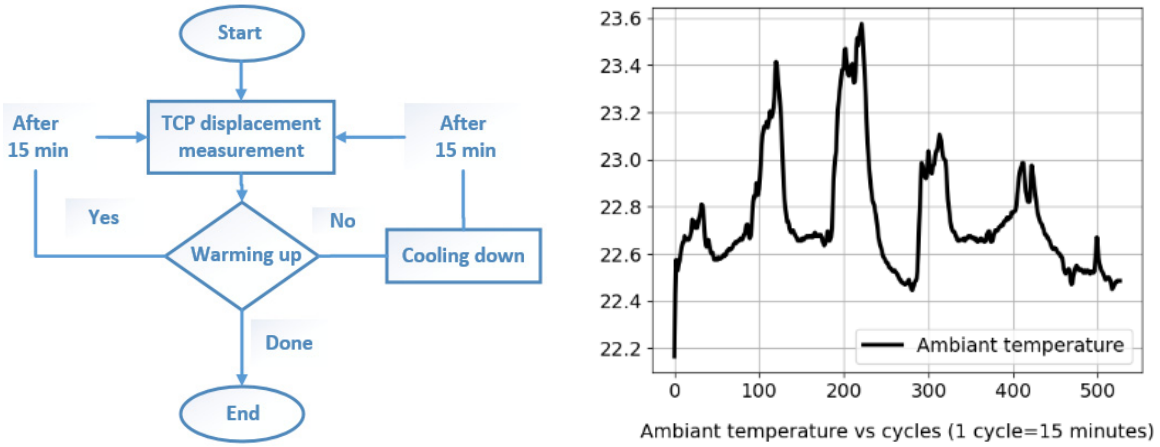


**Figure 2.** Heating and cooling processes and ambient temperature.

The activity sequences are implemented using a machine tool G-code. TCP positions are measured at 15 min intervals for both training and testing processes. TVEs, estimated by the Scale And Master Ball Artefact (SAMBA) method [26] in both processes in the x,

y, and z directions, are the subtractions of the initial value from all subsequent values to avoid the effect of machine quasi-static geometric errors as given by Equation (1).

$$\mathbf{TVE} = \mathbf{VE} - \mathbf{VE}_0 \quad (1)$$

Figure 3 on the left shows the long process of activity with different measured powers of the B and C rotary axes lasting 132 h. In this process, each axis is exercised separately to capture their individual effect on the behavior of the machine tool. Figure 3 on the right shows the short process lasting 40 h with individual and simultaneous motions of rotary axes, which was run after some days and with similar starting condition to the long process. The short process is different from the long process and includes individual axis motions as well as combined axis motions as would be likely during a machining process.

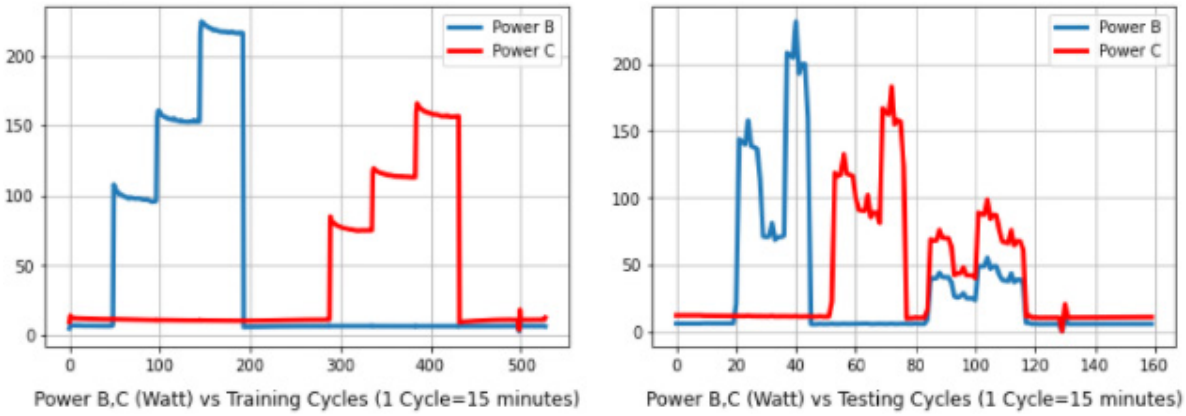


**Figure 3.** Long (**left**) and short (**right**) cycling processes [11].

## 3. Stacked LSTMs and GRUs

An LSTM unit has three gates: an input gate ($i_t$), an output gate ($0_t$) and a forget gate ($f_t$) as shown in Figure 4 (on the left).

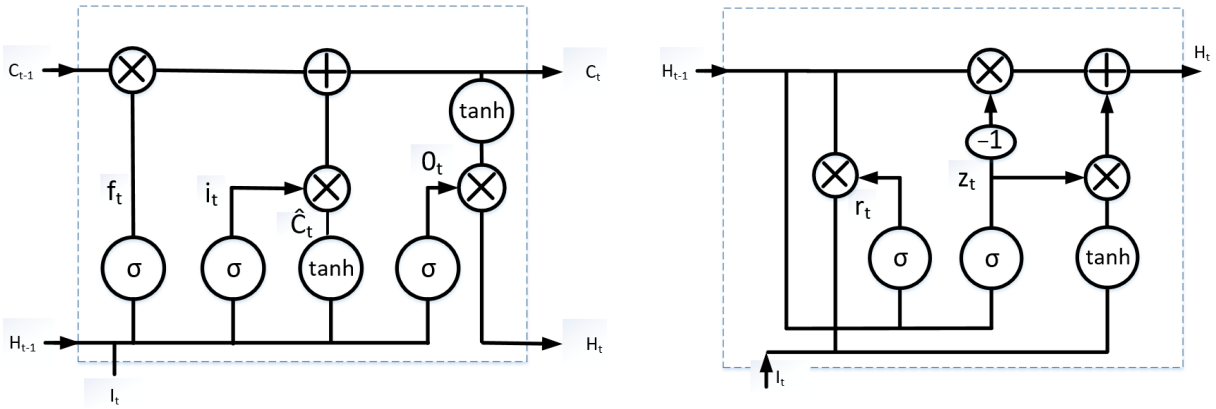


**Figure 4.** LSTM and GRU unit.

The function of the LSTM unit is to remove or add information to the cell state $C_t$ that is carefully regulated by these three gates. $\hat{C}_t$ represents new information that can be applied to cell state $C_t$. $H_t$ is the hidden state; σ and tan h are sigmoid and tangent activation functions, respectively.

Gated recurrent unit (GRU) [27] is also working by a gating mechanism. GRU unit contains a reset gate $r_t$ and an update gate $z_t$. GRU does not have an output gate $O_t$ and long-term memory unit $\hat{C}_t$ as the LSTM unit. Both of them have the ability to eliminate the vanishing gradient problem by keeping the relevant information and passing it to the next time steps of the network.

SLSTM/SGRUs in Figure 5 consists of stacked layers of LSTMs/GRUs and the output layer makes a prediction [20]. The SLSTMs/SGRUs run when the first LSTM/GRU layer takes the input sequence and every LSTM/GRU layer feeds hidden states to the next LSTM/GRU in the stack. SLSTMs/SGRUs calculates only one error signal at the final output layer, and then back propagates it through all previous layers [20]. This structure is programmed with Pytorch library in Python using torch.nn.LSTM/GRU.

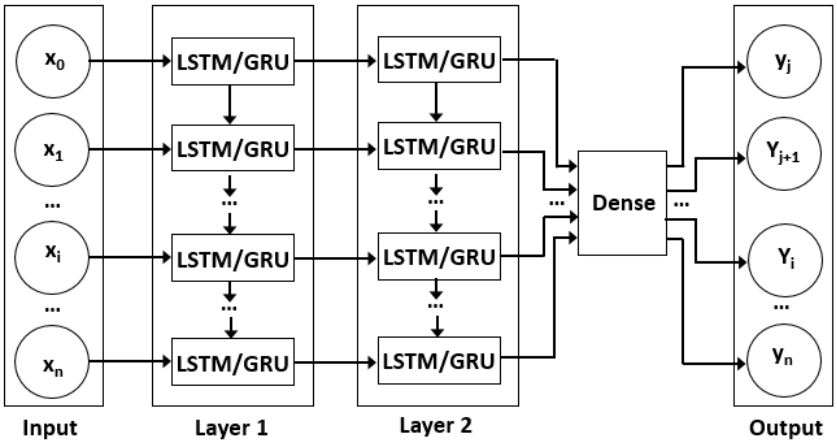


**Figure 5.** Stacked LSTMs and GRUs structure.

## 4. Training and Testing Processes

The machine schematic in Figure 6 shows TVEs measured at four artefact balls. The workpiece and tool branches are illustrated in green and black colors, respectively.

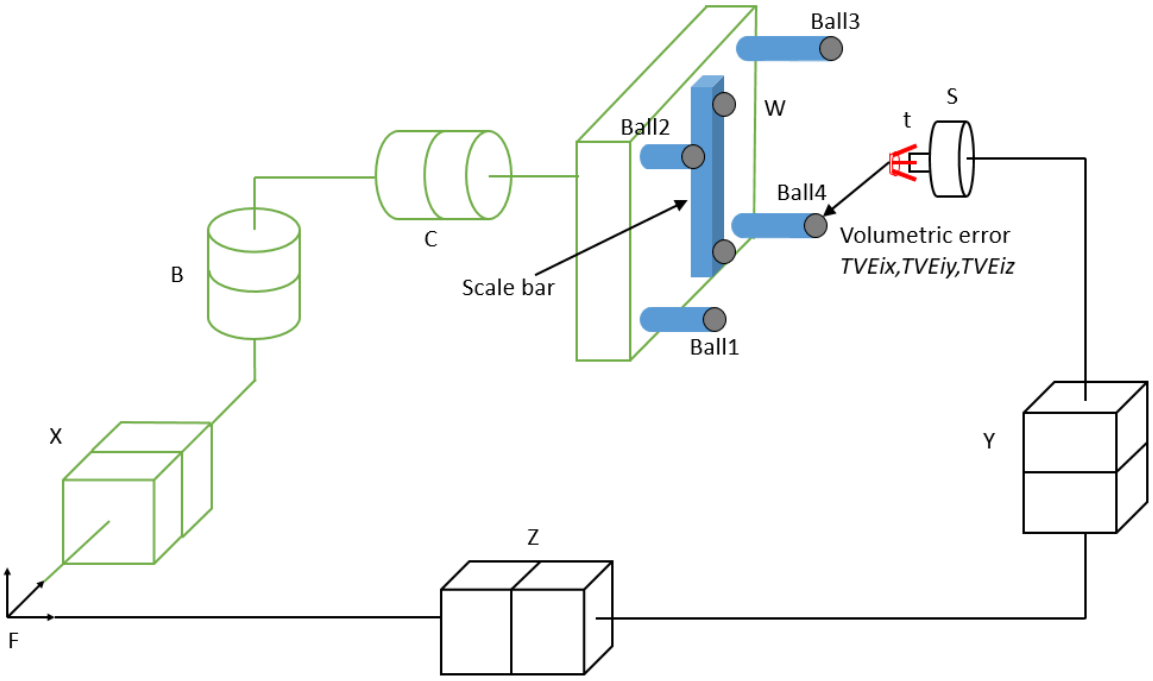


**Figure 6.** Machine schematic with thermally induced volumetric errors.

The TVEs have three positioning error components as described in Equation (2). These error components may change with activities of the B- and C-axis, positions, and indexations.

$$\mathbf{TVE_i} = [\mathrm{TVE_{ix}}\ \mathrm{TVE_{iy}}\ \mathrm{TVE_{iz}}]^{\mathbf{T}} \text{ with } i = 1{:}N \quad (2)$$

where N is the number of measured positions.

In the training phase, the SLSTMs/SGRUs models are built using long process data, as illustrated in Figure 7, and TVE*s* in 24 positions of the artefact balls 1, 2, and 3, shown in Table 1, are used as targets. The power consumptions of the axes B and C, the 24 linear axis positions used to move the sensor nest to the ball artefacts and the nine (B, C) indexations, or rotary axis positions, are used as inputs.

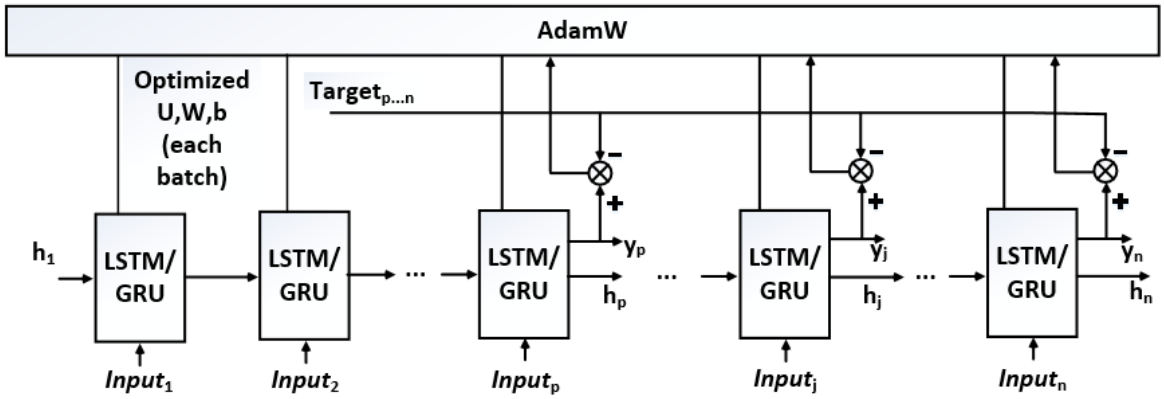


**Figure 7.** Flow chart of the training process.

**Table 1.** 31 positions with nine different indexations of balls 1, 2, 3, and 4.

| Positions of Artefact Balls | | | | Indexations A,B,C (Degrees) | | |
|---|---|---|---|---|---|---|
| $P_{1,1,1}$ | $P_{1,2,2}$ | $P_{1,3,3}$ | | 0 | −90 | 135 |
| $P_{2,4,1}$ | $P_{2,5,4}$ | $P_{2,6,3}$ | | 0 | −90 | −45 |
| $P_{3,7,3}$ | $P_{3,8,4}$ | $P_{3,9,1}$ | $P_{3,10,2}$ | 0 | −30 | −60 |
| $P_{4,11,1}$ | $P_{4,12,2}$ | $P_{4,13,3}$ | $P_{4,14,4}$ | 0 | 0 | 180 |
| $P_{5,15,2}$ | $P_{5,16,3}$ | $P_{5,17,4}$ | $P_{5,18,1}$ | 0 | 0 | 0 |
| $P_{6,19,3}$ | $P_{6,20,4}$ | $P_{6,21,1}$ | $P_{6,22,2}$ | 0 | 20 | 15 |
| $P_{7,23,2}$ | $P_{7,24,3}$ | $P_{7,25,4}$ | | 0 | 90 | 45 |
| $P_{8,26,3}$ | $P_{8,27,2}$ | $P_{8,28,1}$ | | 0 | 90 | −45 |
| $P_{9,29,4}$ | $P_{9,30,1}$ | $P_{9,31,2}$ | | 0 | 90 | −135 |

Where: $P_{q,i,k}$ with $q$ = indexation (1 to 9); $i$ = linear axis position (1 to 31); $k$ = artefact (1 to 4).

The testing phase will consider two cases: long test data consisting of TVE*s* at seven positions of artefact ball 4 as well as the short test data of all 31 positions at all four artefact balls.

The 31 linear axis positions are shown in Figure 8. These positions are pre-determined by calculations based on the indexations of B and C axes and the artefact's positions in the machine frame. They are used for the G-code to move the sensor nest to the four artefact balls.

SLSTMs/SGRUs [28] are capable of capturing longer patterns in sequential data. AdamW optimizer is used to generate the optimized weights and biases based on the outputs of the SLSTMs\SGRUs and the measured TVEs at 24 different positions. In Figure 7, $\boldsymbol{Input}_j = \left[PowerB_j, PowerC_j, x_j, y_j, z_j, b_j, c_j\right]^T$ with j = 1 : n; p is the number of inputs in one sample = 8; n is the number of training data = 528 × 24 = 12,672, corresponding to 24 position measurements or observations at each of the 528 measurement cycles;

$\text{Target}_j$ is the desired $y_j$ of SLSTMs/SGRUs equal to the measured $\mathbf{TVE}_j$; $[y_p, \ldots, y_n]$ = trained $[\text{TVE}_p, \ldots, \text{TVE}_n]$; **U**, **W** and **b** are weight and bias that need to be updated after each batch (number of training samples) in the training process as Equations (3)–(5) [29]

$$\mathbf{W}^{new} = (1 - \lambda_w)\mathbf{W}^{old} - \frac{\eta_w}{\sqrt{\hat{v}_{wt} + \epsilon}}\hat{m}_{wt} \quad (3)$$

$$\mathbf{U}^{new} = (1 - \lambda_u)\mathbf{U}^{old} - \frac{\eta_u}{\sqrt{\hat{v}_{ut} + \epsilon}}\hat{m}_{ut} \quad (4)$$

$$\mathbf{b}^{new} = (1 - \lambda_b)\mathbf{b}^{old} - \frac{\eta_b}{\sqrt{\hat{v}_{bt} + \epsilon}}\hat{m}_{bt} \quad (5)$$

where $\eta$ is the learning rate, $\epsilon = 10^{-8}$, $\hat{m}$ and $\hat{v}$ are bias-corrected first and second moment estimates and $\lambda$ is the rate of the weight decay per step.

The diagram in Figure 9 shows the testing procedure applying multistep predictions with the trained SLSTMs/SGRUs and inputs without using measured TVEs as targets. In this phase, to match the initial conditions, eight first states with values of zeros are added to the data sequence, as Ngoc et al. conducted this in [11].

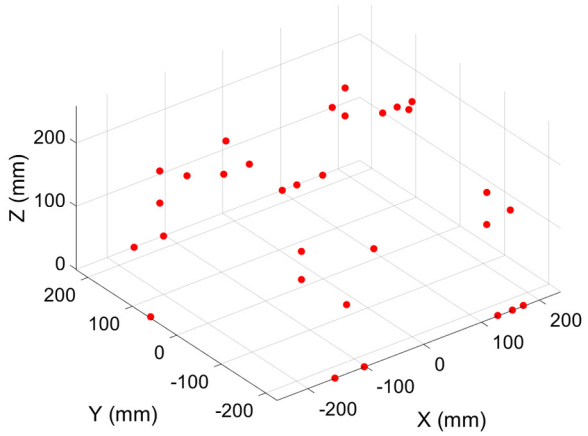


**Figure 8.** 31 linear axis positions at four artefact balls in the machine frame.

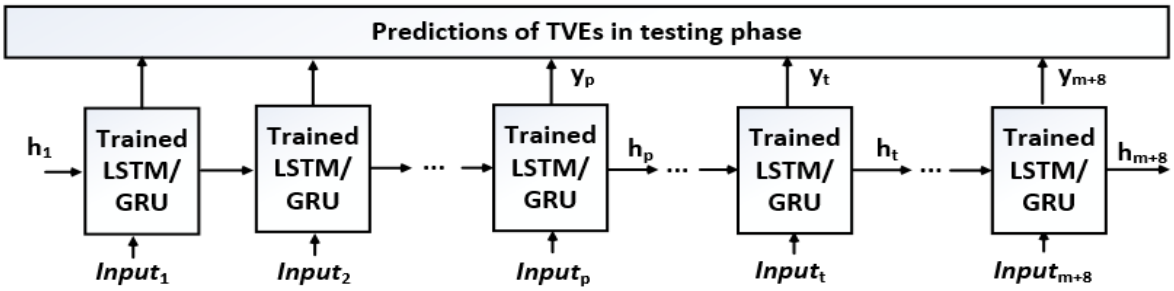


**Figure 9.** Prediction process of thermally induced volumetric errors.

In Figure 9: $\boldsymbol{Input_t} = [PowerB_t, PowerC_t, x_t, y_t, z_t, b_t, c_t]^T$, t = 1 to m + 8, p = 8, m = 528 and 160 cycles in the long and short processes, respectively; and $[y_p \ldots y_{m+8}]$ = *Predicted* $[\text{TVE}_p \ldots \text{TVE}_{m+8}]$.

## 5. Results and Discussions

In this section, the prediction ability of the trained SLSTMs for all TVEs are shown before some the best and worst cases are chosen to compare with the SGRUs'. In the training phase, three Stacked LSTMs are used to build the models for TVEx, TVEy and

TVEz separately. Data of the thermal behaviors of the machine in 24 positions (31 positions of Table 1 less positions 5, 8, 14, 17, 20, 25, and 29 of ball 4) were put in series. Figure 10 shows the measured and predicted thermal volumetric errors (TVEs) for each of the 24 balls' positions where for each position, 528 measuring and prediction cycles are conducted. The results for each position are concatenated horizontally, so the timeline of the abscissa repeats for every position (Pi) for compactness of presentation.

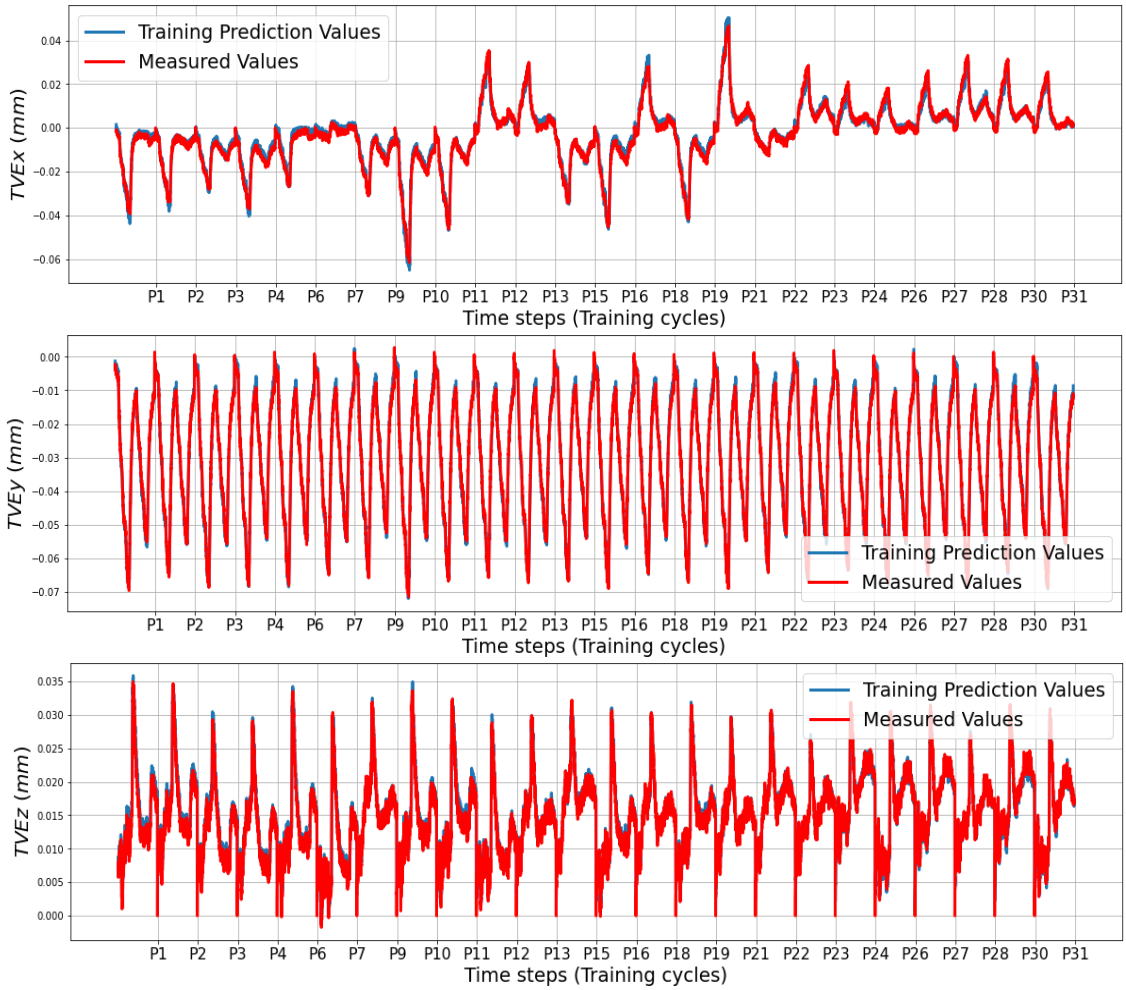


**Figure 10.** Training TVEx, TVEy, and TVEz from 24 positions in the long training process. For example, between P2 and P3 are the 528 cycle measured and predicted thermal volumetric errors at position P3 (1 cycle = 15 min).

The B axis has a greater impact on volumetric errors (TVEs) than the C axis. In fact, B axis activity dominates almost all of the TVEx. However, the influence of C axis activity relative to the B axis increases for TVEy and TVEz.

In the testing phase, three sets of volumetric errors were considered to test the ability of the trained model: data of seven positions at ball 4 in the long process (different positions in workspace but same process with training data), and data of 31 positions in the short process which is divided in two sets: the first set with 24 positions at balls 1, 2, 3 (same positions but different process) and the second set with seven positions of ball 4 (different positions and process). Due to the limitation of space, some of the best and the worst predictions of the TVEs in each set are presented. Multistep predictions (528 steps (cycles)) for positions P14 and P20 in the long process are computed and illustrated from Figures 11–13. In the short process, 160 prediction steps are shown from Figures 14–19 for positions P4, P5, P9, and P25.

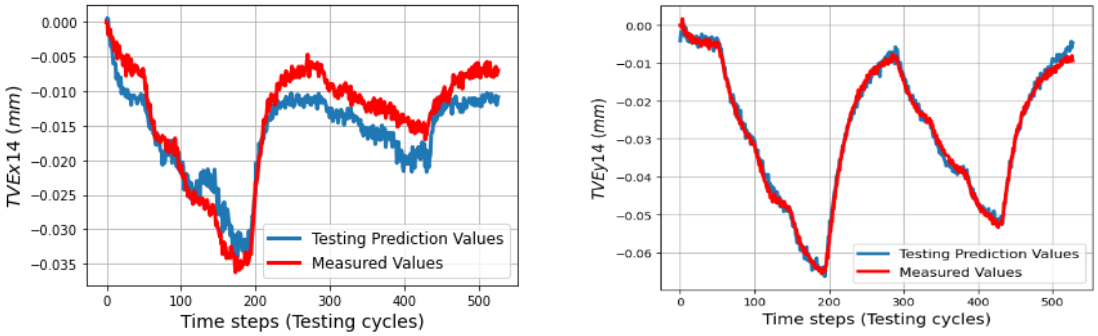


**Figure 11.** Prediction of TVEx14 and TVEy14 in the long process (LP) (1 cycle = 15 min).

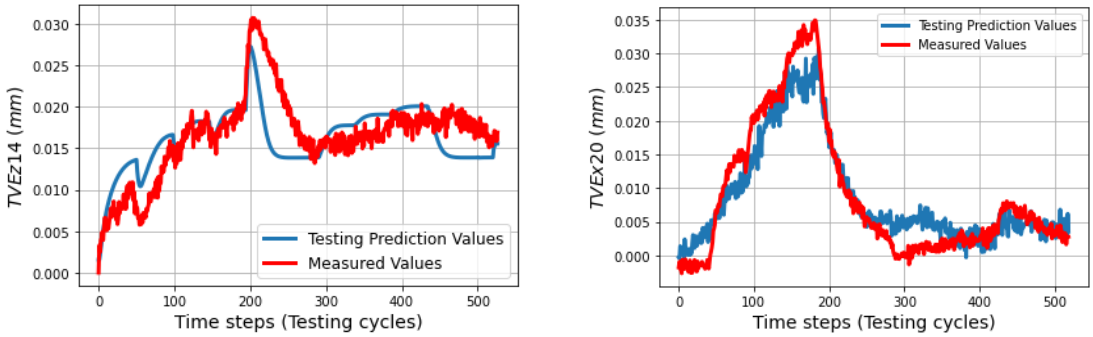


**Figure 12.** Prediction of TVEz14 and TVEx20 in the long process (LP) (1 cycle = 15 min).

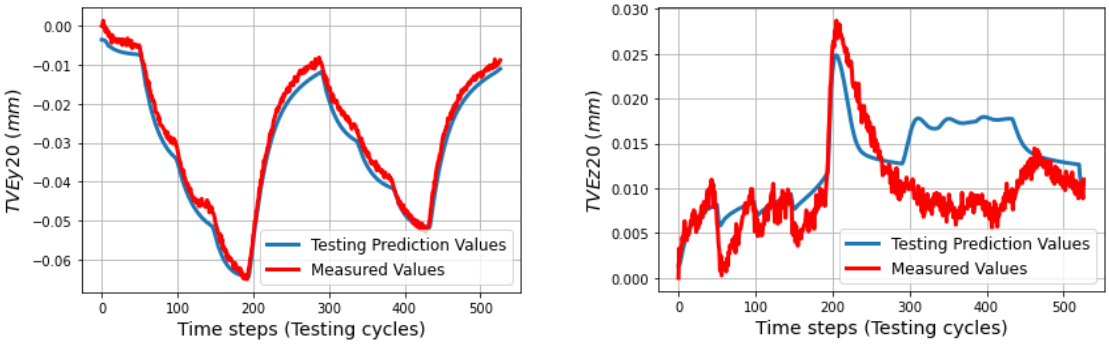


**Figure 13.** Prediction of TVEy20 and TVEz20 in the long process (LP) (1 cycle = 15 min).

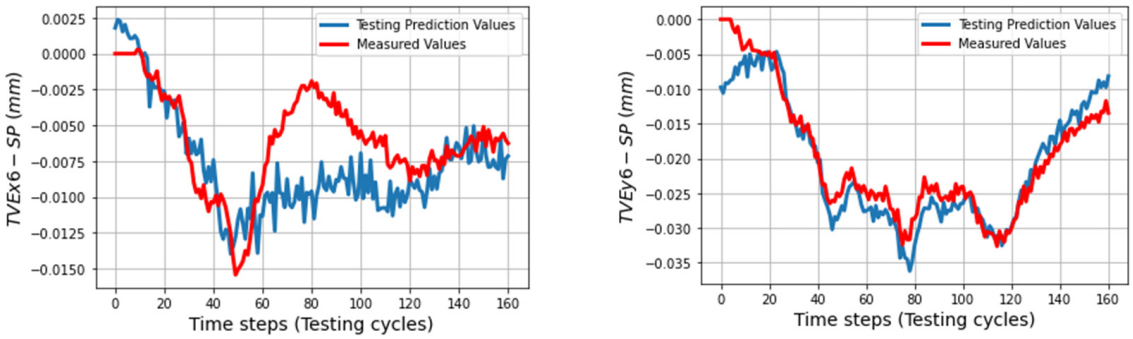


**Figure 14.** Prediction of TVEx6 and TVEy6 in the short process (SP) (1 cycle = 15 min).

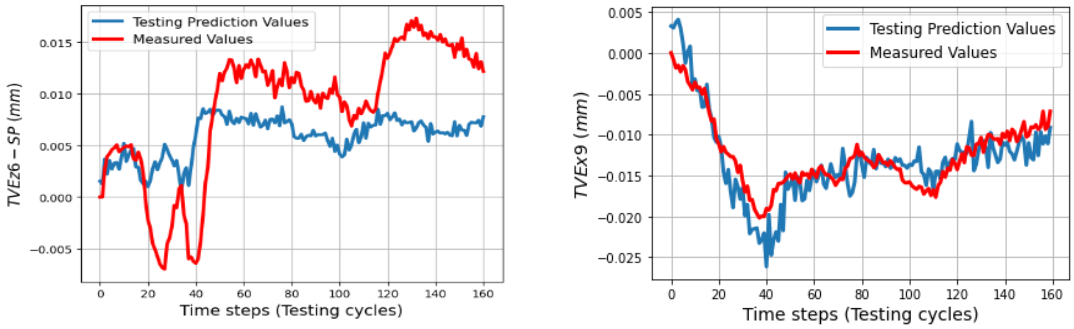


**Figure 15.** Prediction of TVEz6 and TVEx9 in the short process (SP) (1 cycle = 15 min).

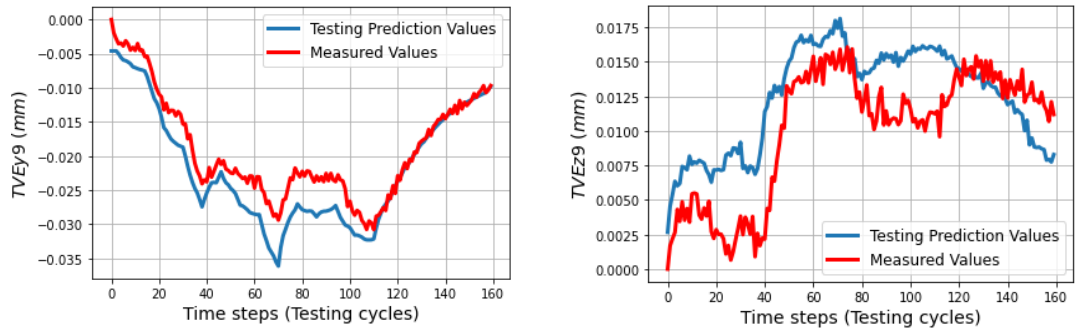


**Figure 16.** Prediction of TVEy9 and TVEz9 in the short process (SP) (1 cycle = 15 min).

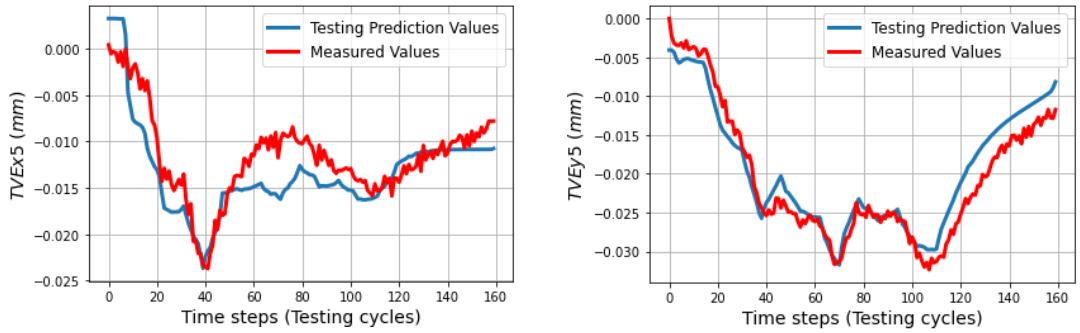


**Figure 17.** Prediction of TVEx5 and TVEy5 in the short process (SP) (1 cycle = 15 min).

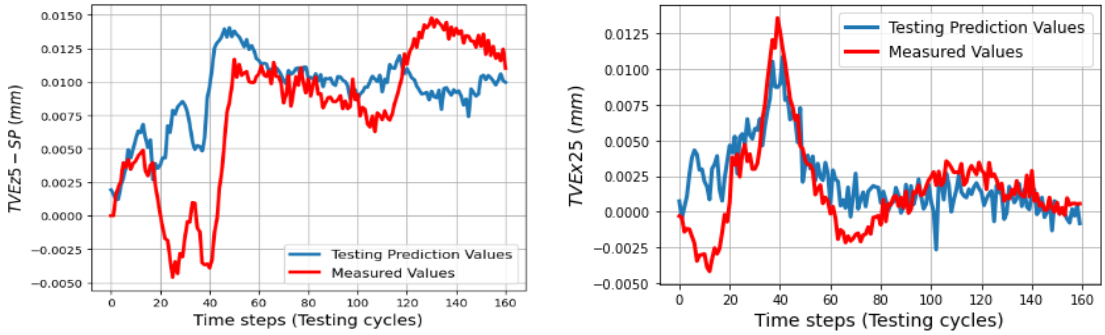


**Figure 18.** Prediction of TVEz5 and TVEx25 in the short process (SP) (1 cycle = 15 min).

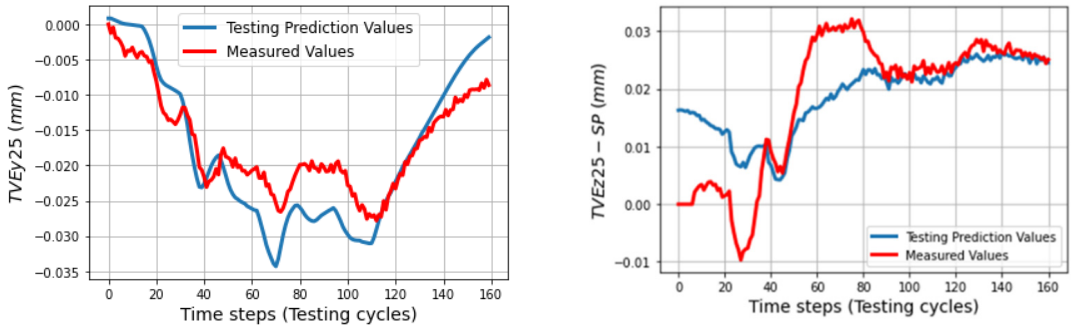


**Figure 19.** Prediction of TVEy25 and TVEz25 in the short process (SP) (1 cycle = 15 min).

The trained model appears to have difficulty predicting TVEz20 from cycles 300 to 450, as its parameters tend to rely heavily on the patterns observed in the training data and do not adjust well to changes in TVEz20 during the testing process. As a result, the model's predictions for TVEz20 during this period are likely to be inaccurate. Conversely, TVEz14 tends to be better predicted during this time. The model is able to adapt more easily to changes in TVEz14.

The Root Mean Square Errors (RMSEs) and fittings of predictions compared with measured values are calculated by Equations (6) and (7) and are shown from Figures 20–24.

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{n} \left(\text{Prediction values}_j - \text{Measured values}_j\right)^2}{n}} \tag{6}$$

$$\text{fitting}(\%) = \left(1 - \frac{\text{RMSE}}{\left|\text{Measured values}_{jmax} - \text{Measured values}_{jmin}\right|}\right) * 100 \tag{7}$$

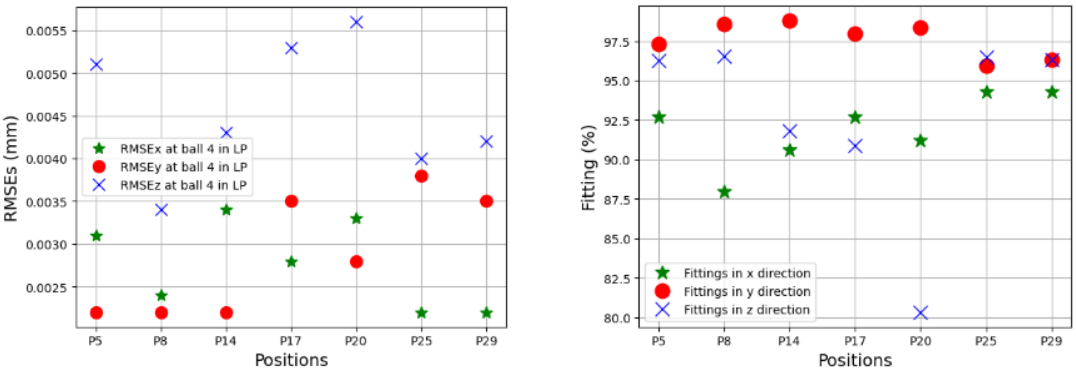


**Figure 20.** RMSEs and fittings of seven positions at ball 4 in the long process (LP).

The changes of TVEs depend on the inputs: powers of rotary axes, durations of activities, positions, and indexations. The trained models have capability to predict TVEs, respectively illustrated from Figure 11 to Figures 24 and 25 presents a set of TVEs in 33 linear axis positions with directional components of TVEx, TVEy, and TVEz at one cycle in the testing process. Predictions of TVEs of seven positions at ball 4 in the long process can reach very high fitting with RMSE = 2.2 µm (fitting = 98.8%) for TVEy14, while the lowest proportion of fitting in this process is for TVEz20 with RMSE = 5.6 µm (fitting = 80.3%).

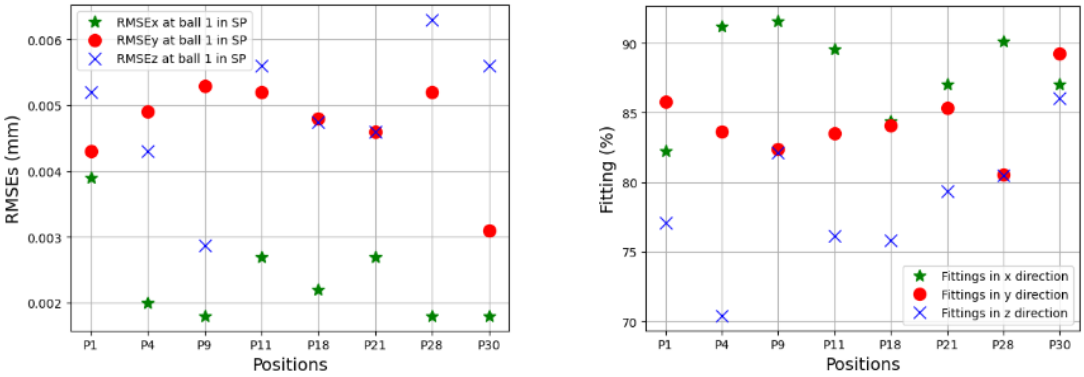


**Figure 21.** RMSEs and fittings of eight positions at ball 1 in the short process (SP).

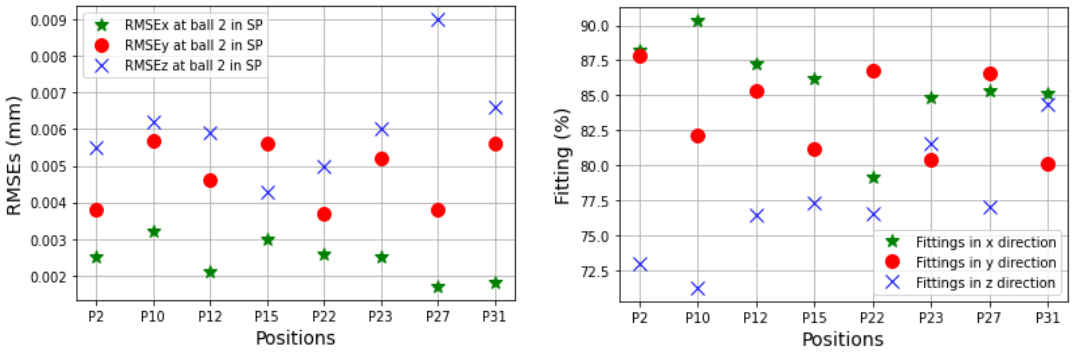


**Figure 22.** RMSEs and fittings of eight positions at ball 2 in the short process (SP).

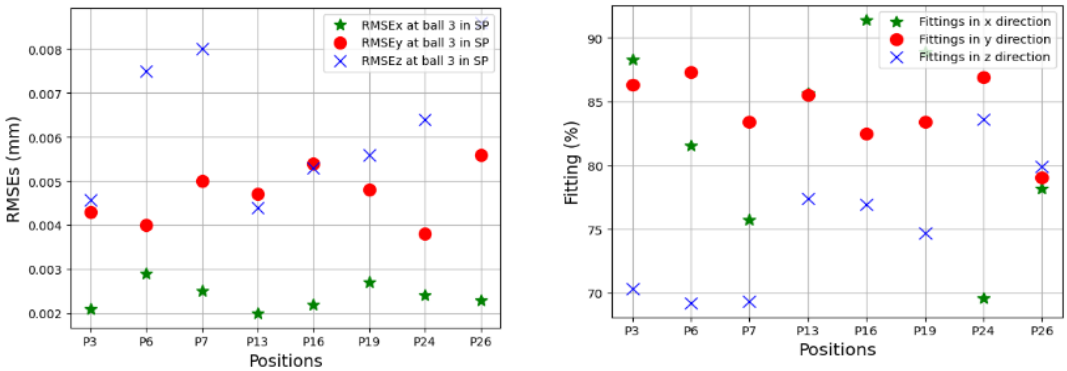


**Figure 23.** RMSEs and fittings of eight positions at ball 3 in the short process (SP).

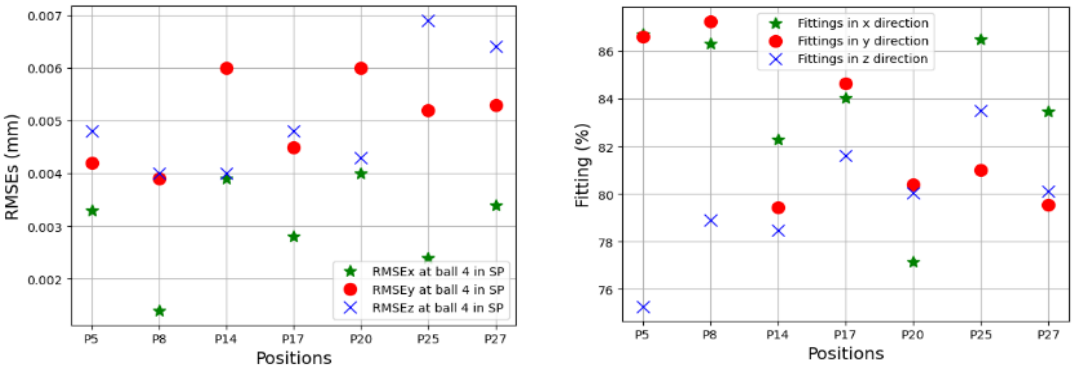


**Figure 24.** RMSEs and fittings of seven positions at ball 4 in the short process (SP).

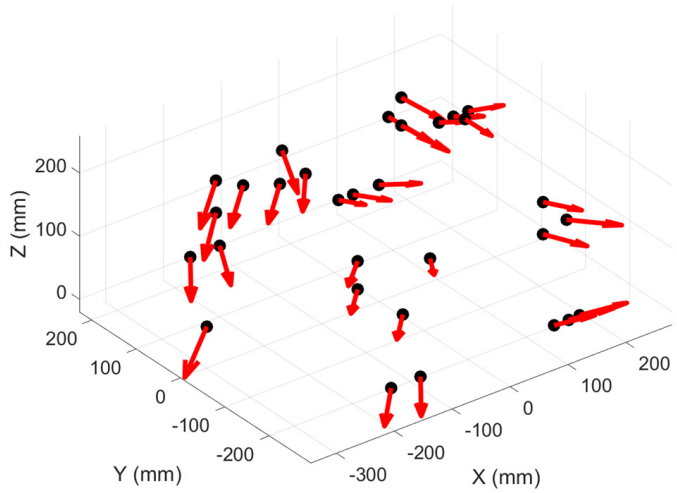


**Figure 25.** 31 linear axis positions and an example set of TVEs.

The trained models gave less quality when they predict TVEs at 7 positions at ball 4 in the short testing process as shown by RMSEs from 3.3 μm (86.7% fitting) for TVEx5 to 4.8 μm (75.27% fitting) for TVEz5. The predictions are good for positions at ball 1, 2, and 3 in the short testing process with the highest fitting 91.6% (1.8 μm) for TVEx9, and the worst case in this process with 69.2% (7.5 μm) for TVEz6. The results also indicate that the model gives the highest capability of prediction for TVE*y* and the worst for TVE*z*. These situations happened in [11], with geometric errors relating to y and z components. There are some over responses of the trained model in the beginning of the short process for all TVEs, but the predictions gradually improve to closely track the measured values for the rest of the process. Figures 20–24 reveal an interesting finding: for the positions located at the artifact ball four, TVEy tends to have a higher fitting than TVEx and TVEz. On the other hand, for the positions located at the artifact balls one, two, and three, TVEx tends to dominate the others.

The SLSTMs model showed its ability to predict thermally induced volumetric errors of the machine tool based on the motor powers of the rotary axes. GRU unit was invented after LSTMs with a simpler structure for time series application. The question is: which one is better for data sequences? To answer this question, this work continues by taking some cases including the best and worst predictions of SLSTMs and uses SGRUs to train model and predict these thermally induced volumetric errors with similar structure, learning rate, and weight decay.

SLSTMs predict TVEy14, TVEx4 better than SGRUs as shown in Figures 26–29 with RMSEs = 2.2 μm, 2 μm and 2.4 μm, 4.4 μm, respectively. For TVEz5, 6, SGRUs performs better with RMSEs = 3.2 μm and 6.6 μm, while SLSTMs have RMSEs turnoff 4.8 μm and 7.5 μm. Both the SLSTMs and SGRUs proved that they have potential to predict the sequence data. The results in this section partly prove the ability of SLSTMs and SGRUs to predict TVEs at the 40 h process with different activities and inputs.

This study has several limitations that should be taken into account. Firstly, the heat sources are limited to the rotary axes, which may affect the quality of predictions of TVEz. To improve the accuracy of the predictions, it may be beneficial to include more information on the activities of the linear axes and temperatures as inputs. Additionally, the ambient temperature should not be overlooked as it can directly affect the change in temperature of the axes following a sequence of activities.

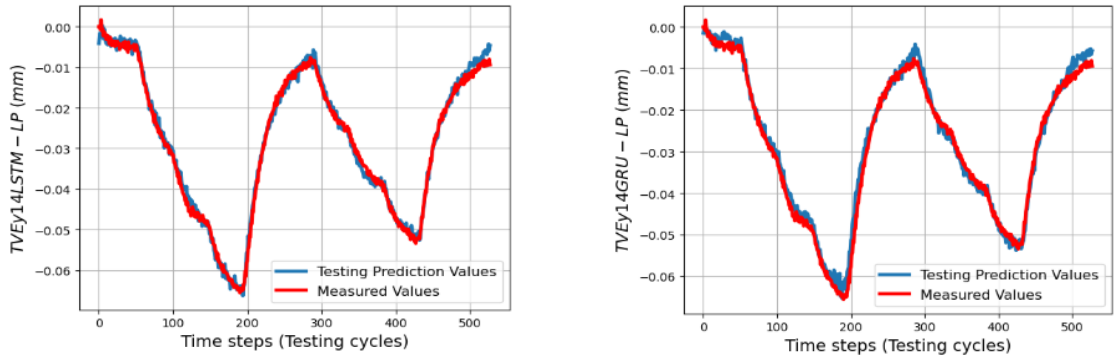


**Figure 26.** Prediction of TVEy14 by SLSTMs and SGRUs in the long process (LP) (1 cycle = 15 min).

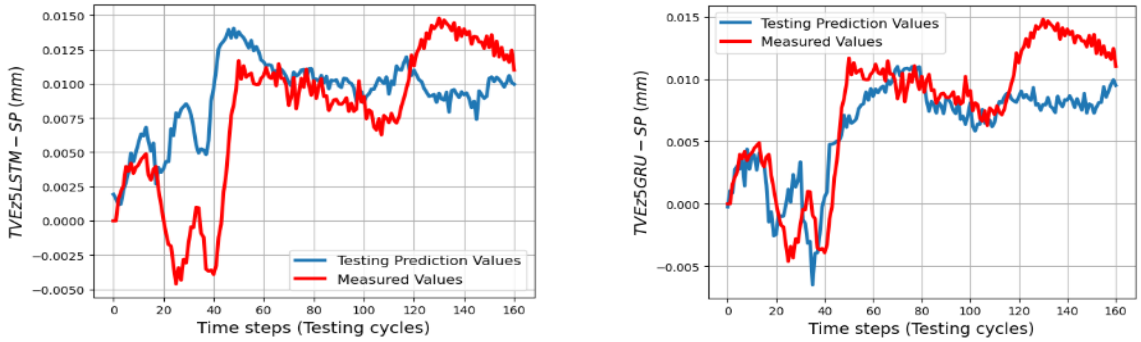


**Figure 27.** Prediction of TVEz5 by SLSTMs and SGRUs in the short process (SP) (1 cycle = 15 min).

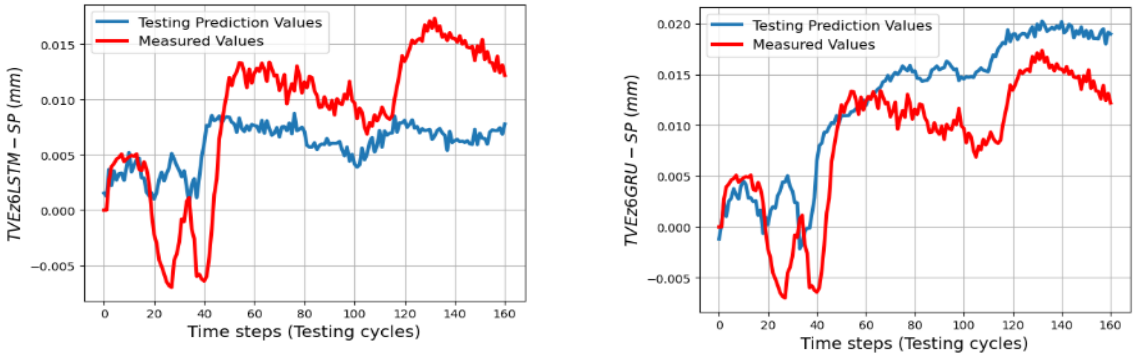


**Figure 28.** Prediction of TVEz6 by SLSTMs and SGRUs in the short process (SP) (1 cycle = 15 min).

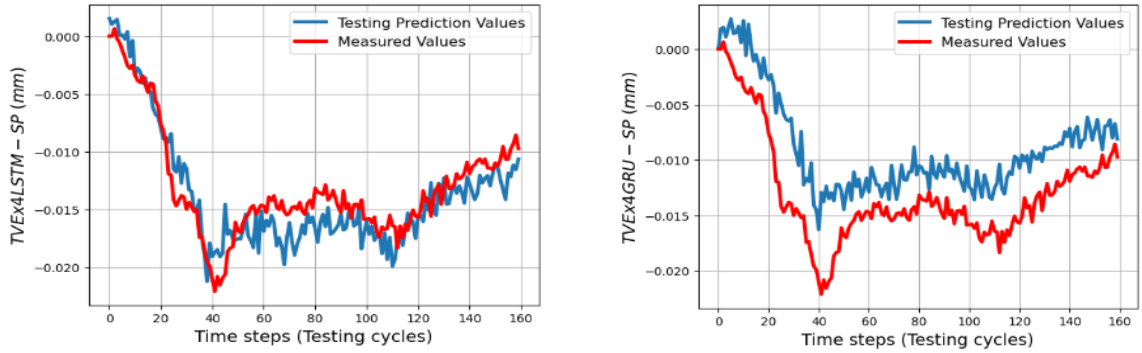


**Figure 29.** Prediction of TVEx4 by SLSTMs and SGRUs in the short process (SP) (1 cycle = 15 min).

## 6. Conclusions

In this study, an approach using SLSTM/SGRU for directly predicting thermally induced volumetric errors without the use of a kinematic model with geometric error parameters was proposed. SLSTMs\SGRUs+AdamW are used for modelling and predicting the sequential data. The training and testing data sets used different positions and different B- and C-axis exercise cycles. The SLSTMs model's best prediction in the long process reached RMSE of 2.2 µm (98.8% fitting to the experimental measurement), while in a short testing process, it achieved 1.8 µm (91.6%). The worst case predicted by the model in processes is 7.5 µm (69.2% fitting). This work gives a potential solution in practice to predict the thermal volumetric errors at different positions in the workspace directly over an extended period of 40 h without the need for remeasuring the thermal errors, which allows sufficient time for the machining of most workpieces.

A comparison between two popular deep learning models for sequential data, SLSTMs and SGRUs, is carried out for some outstanding TVEs. The purpose is to determine which one is better for this application. Their capabilities were similar.

Future work can focus on exploring the heat sources associated with the activities of five-axis machine tools, as well as examining the robustness of deep learning models when faced with variations in ambient temperature.

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

**Funding:** This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery, Grant number RGPIN-2016-06418.

**Institutional Review Board Statement:** I declare on behalf of my co-authors that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Available from the corresponding author on reasonable request.

**Acknowledgments:** Elie Bitar-Nehme, technicians Guy Gironne and Vincent Mayer are acknowledged for experimental data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bitar-Nehme, E.; Mayer, J.R.R. Modelling and compensation of dominant thermally induced geometric errors using rotary axes' power consumption. *CIRP Ann.* **2018**, *67*, 547–550. [CrossRef]
2. Ibaraki, S.; Knapp, W. Indirect measurement of volumetric accuracy for three-axis and five-axis machine tools: A review. *Int. J. Autom. Technol.* **2012**, *6*, 110–124. [CrossRef]
3. Mayr, J.; Jedrzejewski, J.; Uhlmann, E.; Donmez, M.A.; Knapp, W.; Härtig, F.; Wendt, K.; Moriwaki, T.; Shore, P.; Schmitt, R. Thermal issues in machine tools. *CIRP Ann.* **2012**, *61*, 771–791. [CrossRef]
4. Brecher, C.; Hirsch, P.; Weck, M. Compensation of thermo-elastic machine tool deformation based on control internal data. *CIRP Ann.* **2004**, *53*, 299–304. [CrossRef]
5. Horejs, O.; Mares, M.; Kohut, P.; Barta, P.; Hornych, J. Compensation of machine tool thermal errors based on transfer functions. *MM Sci. J.* **2010**, *3*, 162–165. [CrossRef]
6. Liu, Y.; Lu, Y.; Gao, D.; Hao, Z. Thermally induced volumetric error modeling based on thermal drift and its compensation in *Z*-axis. *Int. J. Adv. Manuf. Technol.* **2013**, *69*, 2735–2745. [CrossRef]
7. Mayr, J.; Egeter, M.; Weikert, S.; Wegener, K. Thermal error compensation of rotary axes and main spindles using cooling power as input parameter. *J. Manuf. Syst.* **2015**, *37*, 542–549. [CrossRef]
8. Yu, B.-F.; Liu, K.; Li, K.-Y. Application of Multiple Regressions to Thermal Error Compensation Technology—Experiment on Workpiece Spindle of Lathe. *Int. J. Autom. Smart Technol.* **2016**, *6*, 103–110.

9. Baum, C.; Brecher, C.; Klatte, M.; Lee, T.H.; Tzanetos, F. Thermally induced volumetric error compensation by means of integral deformation sensors. *Procedia CIRP* **2018**, *72*, 1148–1153. [CrossRef]
10. Liu, J.; Ma, C.; Wang, S. Data-driven thermally-induced error compensation method of high-speed and precision five-axis machine tools. *Mech. Syst. Signal Process.* **2020**, *138*, 106538. [CrossRef]
11. Ngoc, H.V.; Mayer, J.R.R.; Bitar-Nehme, E. Deep learning LSTM for predicting thermally induced geometric errors using rotary axes' powers as input parameters. *CIRP J. Manuf. Sci. Technol.* **2022**, *37*, 70–80. [CrossRef]
12. Provost, F.; Fawcett, T. Data science and its relationship to big data and data-driven decision making. *Big Data* **2013**, *1*, 51–59. [CrossRef]
13. Dahlem, P.; Emonts, D.; Sanders, M.P.; Schmitt, R.H. A review on enabling technologies for resilient and traceable on-machine measurements. *J. Mach. Eng.* **2020**, *20*, 5–17. [CrossRef]
14. .Liang, Y.; Li, W.; Lou, P.; Hu, J. Thermal error prediction for heavy-duty CNC machines enabled by long short-term memory networks and fog-cloud architecture. *J. Manuf. Syst.* **2020**, *62*, 950–963. [CrossRef]
15. Liu, J.; Ma, C.; Gui, H.; Wang, S. Thermally-induced error compensation of spindle system based on long short term memory neural networks. *Appl. Soft Comput.* **2021**, *102*, 107094. [CrossRef]
16. Liu, P.-L.; Du, Z.-C.; Li, H.-M.; Deng, M.; Feng, X.-B.; Yang, J.-G. Thermal error modeling based on BiLSTM deep learning for CNC machine tool. *Adv. Manuf.* **2021**, *9*, 235–249. [CrossRef]
17. Yu-Chi, L.; Kun-Ying, L.; Yao-Cheng, T. Spindle Thermal Error Prediction Based on LSTM Deep Learning for a CNC Machine Tool. *Appl. Sci.* **2021**, *11*, 5444.
18. Gao, X.; Guo, Y.; Hanson, D.A.; Liu, Z.; Wang, M.; Zan, T. Thermal Error Prediction of Ball Screws Based on PSO-LSTM. *Int. J. Adv. Manuf. Technol.* **2021**, *116*, 1721–1735. [CrossRef]
19. Mchichi, N.A.; Mayer, J. Axis location errors and error motions calibration for a five-axis machine tool using the SAMBA method. *Procedia CIRP* **2014**, *14*, 305–310. [CrossRef]
20. Sagheer, A.; Kotb, M. Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems. *Sci. Rep.* **2019**, *9*, 19308. [CrossRef]
21. Curtis, F.E.; Scheinberg, K. Adaptive Stochastic Optimization: A Framework for Analyzing Stochastic Optimization Algorithms. *IEEE Signal Process. Mag.* **2020**, *37*, 32–42. [CrossRef]
22. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
23. Gugger, S.; Howard, J. AdamW and Super-Convergence Is Now the Fastest Way to Train Neural Nets. Available online: https:/www.fast.ai/2018/07/02 (accessed on 19 July 2018).
24. Bitar-Nehme, E.; Mayer, J.R.R. Thermal volumetric effects under axes cycling using an invar R-test device and reference length. *Int. J. Mach. Tools Manuf.* **2016**, *105*, 14–22. [CrossRef]
25. Weikert, S. R-test, a new device for accuracy measurements on five axis machine tools. *CIRP Ann. Manuf. Technol.* **2004**, *53*, 429–432. [CrossRef]
26. Mayer, J.R.R. Five-axis machine tool calibration by probing a scale enriched reconfigurable uncalibrated master balls artefact. *CIRP Ann.* **2012**, *61*, 515–518. [CrossRef]
27. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
28. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
29. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.