CIRP Annals - Manufacturing Technology 70 (2021) 345–348

Contents lists available at ScienceDirect

CIRP Annals - Manufacturing Technology

journal homepage: https://www.editorialmanager.com/CIRP/default.aspx

Influence of bearing ball recirculation on error motions of linear axes

Gregory W. Vogl[a,*], Kyle F. Shreve[b], M. Alkan Donmez (2)[a]

[a] *Engineering Laboratory, National Institute of Standards and Technology (NIST), 100 Bureau Drive, Gaithersburg, Maryland 20899-8220, USA*
(Official contribution of the National Institute of Standards and Technology; not subject to copyright in the United States.)
[b] *Department of Mechanical Engineering, University of Maryland, College Park, Maryland 20742, USA*

ARTICLE INFO

Article history:
Available online 10 June 2021

Keywords:
Machine tool
Error
Diagnostics

ABSTRACT

For positioning systems utilizing linear guides and trucks with recirculating balls, a method is presented that uses the measured total error motions and the measured phase of ball loops within trucks to determine the influence of each ball loop on the error motions. The influence of ball recirculation on the error motions is estimated a priori via a least-squares solution based on data collected from a multitude of motion tests in which varying phases were measured by sensors integrated into the trucks. This method enables real-time estimation of performance degradations and identification of their sources.

1. Introduction

Most modern machine tools and linear positioning systems use linear motion guideways, consisting of a rail and multiple carriages (trucks) with recirculating (looping) rolling elements, for precision motion. Manufacturing relies on linear positioning systems [1], and achieving high-precision and robust production requires accurate knowledge of error motions as well as sources of emerging faults. Thus, methods that enable higher precision and condition-based maintenance via the monitoring of error motions could impact various manufacturing applications.

The geometric accuracy and structural health of linear motion systems are affected by the interaction between the guideways and the moving elements in the trucks. At any given position of a truck, a subset of recirculating balls in a ball loop contacts a section of a guideway. As the truck moves back and forth to the same position, the subset of recirculating balls contacting the guideway changes, because the entire ball loop has shifted relative to the truck due to micromechanics. This ball loop shift can be thought of as a phase change. The combination of guideway/active balls (in contact with the guideway) at any axis position influences the error motions [2] due to local imperfections of both the guideways and the active balls, and the use of multiple trucks increases the apparent non-repeatability of the error motions. Furthermore, error motions change with machine usage, since abrasion and adhesion between rolling elements causes material fatigue, pitting, cracking, and wear. If not properly mitigated, these faults will grow to affect the quality of parts produced, leading to parts becoming out of tolerance and/or machine failure [3].

To enable increased accuracy and proactive maintenance, manufacturers need automated methods for diagnosing machine tool linear axes without halting production. A new method is proposed to identify the contributions of each truck to the overall error motion, which provides information that can be used to diagnose the condition of linear axis components [4,5] needed for intelligent sensor-based systems of manufacturing processes [6]. This paper introduces a sensor-based method and demonstrates its experimental validation.

2. Method

The new method utilizes the phases of recirculating ball loops, measured by in-situ sensors, to estimate the effect of recirculation on error motions in real time. Fig. 1 shows a schematic of the new method applied to a system with two rails and four trucks. As the linear axis moves back and forth to the same axis position (x), the balls inside the trucks recirculate and the phase of each loop changes, which is represented in the figure by the changing positions of the black balls between two states (blue and red). Also, the change of balls in contact with the rails at the same position causes a given

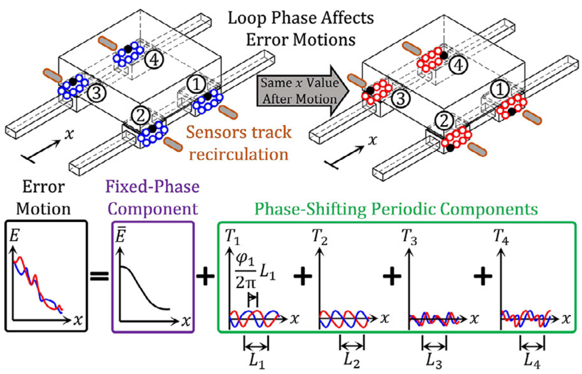

Fig. 1. Schematic of the method to determine and monitor influence of bearing ball recirculation in ball loops of four moving trucks (Truck 1 to Truck 4) on geometric error motions of a linear positioning system.

* Corresponding author.
E-mail address: gvogl@nist.gov (G.W. Vogl).

https://doi.org/10.1016/j.cirp.2021.04.078
0007-8506/Published by Elsevier Ltd on behalf of CIRP.

error motion, $E(x)$, to change its state from the “blue” state to the “red” state in the error plot in Fig. 1.

This method considers an error motion, as a function of nominal position, to be the summation of the influences of the interactions of each ball loop with the guideway rails. By decomposing the error function into aperiodic (fixed-phase) and periodic (phase-shifting) components, the position-based error function can be expressed in terms of the phases of the ball loops. Fig. 1 shows that the error motion for the j^{th} state, $E_j(x)$, is assumed to be the sum of the fixed-phase component, $\overline{E}(x)$, due to the average interaction of all balls and guideways, and the phase-shifting components, $T_{ij}(x, \varphi_{ij})$, for each i^{th} ball loop and j^{th} state due to recirculation:

$$E_j(x) = \overline{E}(x) + \sum_{i=1}^{N_{\text{BL}}} T_{ij}(x, \varphi_{ij}) \tag{1}$$

where N_{BL} is the total number of ball loops interacting with guideway rails in the linear positioning system. Each phase-shifting component is a periodic function $\overline{T}_i(x)$ with period L_i that is shifted in space with phase φ_{ij} due to recirculation; that is,

$$T_{ij}(x, \varphi_{ij}) = \overline{T}_i\left(x - \frac{\varphi_{ij}}{2\pi} L_i\right) \tag{2}$$

The method assumes that the influences of each ball loop on an error motion is periodic with the same period and phase as the ball loop itself. The phase of $\overline{T}_i(x)$ for $T_{ij}(x, \varphi_{ij})$ is associated with the location each specific ball contacts the rail, which requires tracking of the ball movements. This was accomplished by replacing a few of the original balls with marker balls (see Fig. 2b) and observing the locations of these markers through a machined slot during motion with proximity sensors (see Fig. 2a). Fig. 3 shows the proximity signal peaks (denoted A through E) corresponding to the selected arrangement of the markers within the ball loop (Fig. 2b) and the distances between these peaks, as ratios (α, β, γ, δ, and ε) of the signal period, L_i. Six marker balls were used which balances the need for minimizing their influence on the motion accuracy while enabling visual identification of every unique ball through a slot (see Fig. 2). For industrial implementation, sensor-integrated trucks with only one marker ball could be used without any slot.

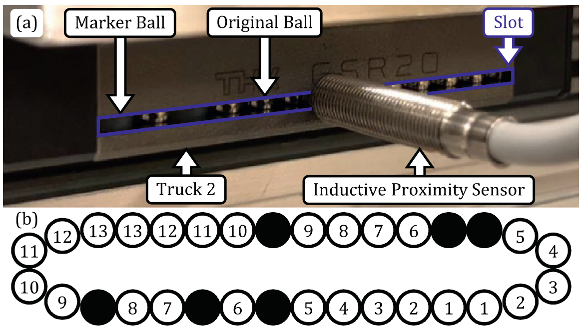

Fig. 2. (a) View of truck with integrated inductive proximity sensor, and (b) example ball loop pattern with marker balls (dark filled circles) and a numbering scheme of original balls (see Section 4).

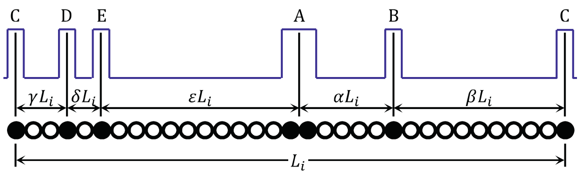

Fig. 3. Diagram of the distances between example proximity signal peak centers as a function of the ball loop signal period, L_i.

Due to spaces between the balls and the friction characteristics, the period and the phase are not constant. The method uses the measured periods and phases of the periodic ball-loop signals, determined from the proximity data, to solve for the error functions ($\overline{E}(x)$ and every $\overline{T}_i(x)$) that comprise the right-hand side of Eq. (1). Any measured error motion on the left-hand side of Eq. (1) is decomposed into average aperiodic and recirculation-induced periodic components as described in Section 3.

2.1. Ball loop period and phase

The signal period L_i and phase φ_{ij} of each error function $\overline{T}_i(x)$ of Eq. (2) are determined for each j^{th} state before solving for all error functions ($\overline{E}(x)$ and every $\overline{T}_i(x)$). The ratios (α, β, γ, δ, and ε) are known based on the marker ball locations within the ball loop. Towards this end, Fig. 4 shows a diagram comparing the theoretical distances and experimental distances between the peak centers for the i^{th} ball loop and j^{th} state. The differences between the theoretical and experimental peak-to-peak distances are minimized in a least-squares approach to yield

$$L_{ij} = \left(\mathbf{r}^T\mathbf{r}\right)^{-1}\left(\mathbf{r}^T\mathbf{d}\right) \tag{3}$$

where $\mathbf{r}$ is the vector of sequential theoretical peak-to-peak ratios, e.g., $[\gamma;\ \delta;\ \varepsilon;\ \alpha;\ \beta;\ \gamma]$ in Fig. 3, and $\mathbf{d}$ is the vector of sequential experimental peak-to-peak distances, $[d_1;\ d_2;\ \cdots\ d_{n-1}]$, for the $(n-1)$ distances corresponding to the n peaks. Once L_{ij} is known from Eq. (3), the signal period L_i for the i^{th} ball loop is set as the mean of its L_{ij} values. Since, at any given state, the starting point for period calculation can be arbitrary, we assign it a variable y_0 (see Fig. 4) and calculate it by minimizing the sum of the squares of the residuals between the vector of theoretical locations, $\mathbf{y} = [y_1;\ y_2;\ \cdots\ y_n]$, and the vector of experimental locations, $\mathbf{x} = [x_1;\ x_2;\ \cdots\ x_n]$, for the n peaks. The least-squares solution for y_0 is the solution of

$$\text{mean}(\mathbf{y}) = \text{mean}(\mathbf{x}) \tag{4}$$

where mean($\mathbf{v}$) is the arithmetic mean of a general vector $\mathbf{v}$. Finally, the phase, φ_{ij}, of the ball loop is defined as

$$\varphi_{ij} = \text{wrap}(2\pi\ y_{\text{A},\ ij}/L_i) \tag{5}$$

where $y_{\text{A},}\ ij$ is the theoretical position of an index marker (Peak A in Fig. 4) for the i^{th} ball loop and j^{th} state that is closest to the origin ($x = 0$), and wrap() is the function that wraps the phase to within $[0, 2\pi]$.

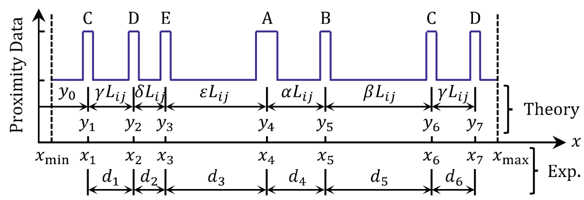

Fig. 4. Diagram with theoretical and experimental signal peak centers.

3. Problem Formulation and Least-Squares Solution

All error functions ($\overline{E}(x)$ and every $\overline{T}_i(x)$) are determined at specific locations and interpolated to locations in between. Each type of error motion (e.g., straightness and angular error motions) is measured with a relevant instrument (e.g., laser interferometer) at $k = 1, 2, \ \ldots$ N equally-spaced locations, $x_k \in [x_{\min}, x_{\max}]$, where Δx is the nominal distance between adjacent positions. To reduce the number of solvable variables, the variables representing a signal are placed at positions with an interval spacing of $\nu\Delta x$, where ν is a positive integer. Next, Eqs. (1) and (2) are approximated in matrix form as

$$\mathbf{E}_j = \overline{\mathbf{E}} + \sum_{i=1}^{N_{\text{BL}}} \mathrm{P}_{i,j}(L_i, \varphi_{ij})\ \overline{\mathbf{T}}_i \tag{6}$$

where $\mathbf{E}_j$ is the vector of measured errors, $\overline{\mathbf{E}}$ is the aperiodic error vector with a length of $\tau_0 = 1 + \text{ceil}\left((x_{\max} - x_{\min})/\nu\Delta x\right)$, $\overline{\mathbf{T}}_i$ is the vector of periodic errors with a length of $\tau_i = 1 + \text{ceil}(L_i/\nu\Delta x)$, in which ceil() is the ceiling function, and $\mathrm{P}_{i,j}(L_i, \varphi_{ij})$ is a matrix that accounts for linear interpolation, the periodicity of $\overline{T}_i(x)$, and phase shifts. Lastly, the mean of every $\overline{\mathbf{T}}_i$ is set to zero for a unique solution. Thus, four rows are added to Eq. (6) for all error motion states ($j = 1, 2, \ldots R$) to yield the final system of equations as

$$\begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \vdots \\ \mathbf{E}_R \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \mathrm{P}_0 & \mathrm{P}_{1,1} & \mathrm{P}_{2,1} & \mathrm{P}_{3,1} & \mathrm{P}_{4,1} \\ \mathrm{P}_0 & \mathrm{P}_{1,2} & \mathrm{P}_{2,2} & \mathrm{P}_{3,2} & \mathrm{P}_{4,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathrm{P}_0 & \mathrm{P}_{1,R} & \mathrm{P}_{2,R} & \mathrm{P}_{3,R} & \mathrm{P}_{4,R} \\ \Omega_0 & I_1 & \Omega_2 & \Omega_3 & \Omega_4 \\ \Omega_0 & \Omega_1 & I_2 & \Omega_3 & \Omega_4 \\ \Omega_0 & \Omega_1 & \Omega_2 & I_3 & \Omega_4 \\ \Omega_0 & \Omega_1 & \Omega_2 & \Omega_3 & I_4 \end{bmatrix} \begin{bmatrix} \overline{\mathbf{E}} \\ \overline{\mathbf{T}}_1 \\ \overline{\mathbf{T}}_2 \\ \overline{\mathbf{T}}_3 \\ \overline{\mathbf{T}}_4 \end{bmatrix} \quad (7)$$

where Ω_i is a matrix of zeros and I_i is a matrix of ones, both with a size of $1 \times \tau_i$. The least-squares solution of Eq. (7) yields $\overline{\mathbf{E}}$ and every $\overline{\mathbf{T}}_i$ and can be applied for the two straightness and three angular error motions. Eq. (6) applies for constant ball-loop periods and captures lower frequencies via linear interpolation.

4. Experimental Setup

Each of the four trucks is instrumented with an inductive proximity sensor (see Fig. 2a) to monitor the outer loop containing 32 balls of 4 mm diameter with the pattern of Fig. 2b. At any given time, about 13 balls in each loop contact the guideway. Hence, about 104 balls (13 balls per loop $\times$ 2 loops per truck $\times$ 4 trucks) are in contact with the two rails at any time. However, a preload was applied within the system such that the net forces between the outer ball loops and the rails dominate over the forces for the inner ball loops. Therefore, this study only concerns the four outer ball loops that significantly influence the error motions; therefore, $N_{\mathrm{BL}} = 4$ in Eq. (1).

In this experiment, the balls are incrementally replaced with marginally larger balls to illustrate how the method can monitor the bearing-ball influences on error motions for diagnostic purposes. First, with the original set of balls (Condition 0), pitch error motion data is collected using a laser-based measuring instrument and the four proximity sensors as the axis moves back and forth over its entire range of travel (0.35 m) for a total of 90 runs ($R = 90$). In practice, this type of data collection would be performed once, or as desired, for updating the phase-based error model, Eq. (7), for health monitoring purposes.

To change the condition, two balls in Truck 2, labeled "1" in Fig. 2b, were replaced with balls with diameters about 12 µm larger, and 90 runs of data were collected for that condition (Condition 1). Next, two more balls labeled "2" in Fig. 2b were replaced in the same manner and 90 runs of data were collected for that condition (Condition 2). This process of replacing balls with subsequent data collection was repeated (Conditions 3 to 13) according to Fig. 2b until each original ball in Truck 2 was replaced (Condition 13).

For each motion between each end of the axis travel, the time-sampled proximity data were processed as functions of position. Fig. 5 shows an example of proximity data, binarized as either 0 of 1 via simple thresholding. The binary data will be analysed to yield the signal period and phase for each outer ball loop signal for every state. The signal periods and phases are used to solve for all error functions ($\overline{E}(x)$ and every $\overline{T}_i(x)$).

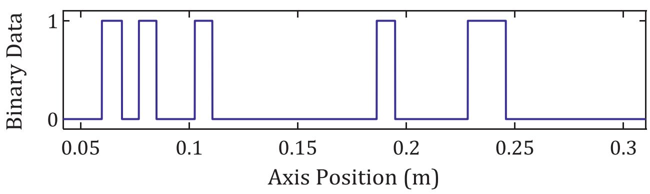

Fig. 5. Example of binarized proximity data for a single axis travel.

The ball loop period, L_{ij}, for each state is calculated according to Eq. (3). The period L_{ij} corresponds mainly to the numbers of balls per loop. Due to the kinematics of rolling, the signal period for the binary proximity data, as well as every error function $\overline{T}_i(x)$, should be approximately 256 mm, which is twice the physical ball loop length of 128 mm (= 32 $\times$ 4 mm). A gap exists within each ball loop, to allow for motion of balls within the truck, and was measured to be between 2 mm and 5 mm in length, which increases the signal period by roughly 4 mm to 10 mm.

Table 1 lists the mean and standard deviation of the ball loop periods for all states (every speed, direction, run, and condition). The means range from 264.1 mm to 269.1 mm, which is a difference of 5 mm and is reasonable given the known gap and preload differences that affect the micromechanics of rolling. Also, given that the standard deviations of the ball loop periods are less than 2 mm (see Table 1) and the error motion data has a spacing of $\Delta x = 1$ mm, ν is set to equal 2, which yields a point spacing of 2 mm for $\overline{\mathbf{E}}$ and every $\overline{\mathbf{T}}_i$ and is small enough to identify details in errors due to relatively small physical features.

Table 1
Mean and standard deviation (std) of the distribution of estimated ball loop period for all motions.

	Truck 1	Truck 2	Truck 3	Truck 4
mean(L_{ij}) (mm)	269.12	265.38	264.05	267.81
std(L_{ij}) (mm)	1.52	1.13	0.99	0.77

Similarly, the ball loop phase for each state is calculated according to Eq. (5). Fig. 6 shows an example of the phase of the ball loops for the initial condition (Condition 0). Over the 90 runs, the phases of the ball loops changed by typically more than 2π, meaning that each ball loop performed more than one full loop rotation over time for a fixed axis location. The trends are nominally linear with unique slopes due to rolling micromechanics that are difficult to predict but are observable via the truck sensors.

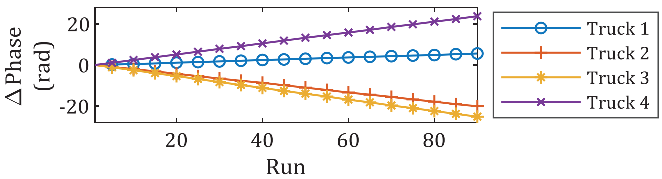

Fig. 6. Change in phase of the ball loops for Condition 0.

5. Results

Fig. 7 shows the five error functions for the initial pitch error motion, E_{BX} (Condition 0). The aperiodic component, $\overline{E}(x)$, has a range of about 200 µrad, but $\overline{T}_i(x)$ has a range of 5 µrad that is about three times smaller than those for the other periodic components ($\overline{T}_2(x)$, $\overline{T}_3(x)$, and $\overline{T}_4(x)$). Hence, the new method reveals that Truck 1 has a potentially low preload.

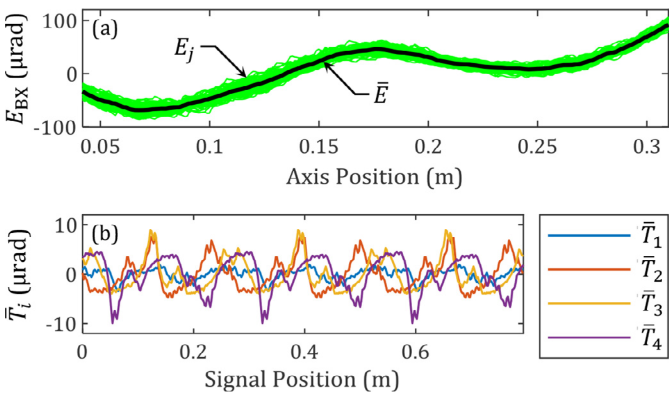

Fig. 7. (a) Fixed-phase error function (black curve), $\overline{E}(x)$, with measured pitch error motion for each j^{th} state (green curves) and (b) periodic error functions for pitch error motion for Condition 0.

The new method also reveals how the error functions change as the balls in Truck 2 are replaced (Conditions 1 to 13). Fig. 8 shows how $\overline{T}_1(x)$, $\overline{T}_3(x)$, and $\overline{T}_4(x)$ change from their initial values (Condition 0) with high-frequency changes of only about 1 µrad. $\overline{E}(x)$ and $\overline{T}_2(x)$ also have high-frequency terms but change significantly at low frequencies. Fig. 8c shows that $\overline{T}_2(x)$ changes by more than 20 µrad at Condition 5, mainly near the locations of the replaced balls. However, $\overline{T}_2(x)$ then changes back to its initial state at the final state, i.e., $\Delta\overline{T}_2(x) \approx 0$ at Condition 13. Simultaneously, $\overline{E}(x)$ changes increasingly with ball replacements (see Fig. 8a). The behaviors of $\overline{E}(x)$ and $\overline{T}_2(x)$ are related; as the balls in Truck 2 are replaced, the truck preloads change, which changes $\overline{E}(x)$, but once all the original balls in Truck 2 are replaced (Condition 13), $\overline{T}_2(x)$ returns to its initial values because the replaced balls are similar to each other, just like the original balls.

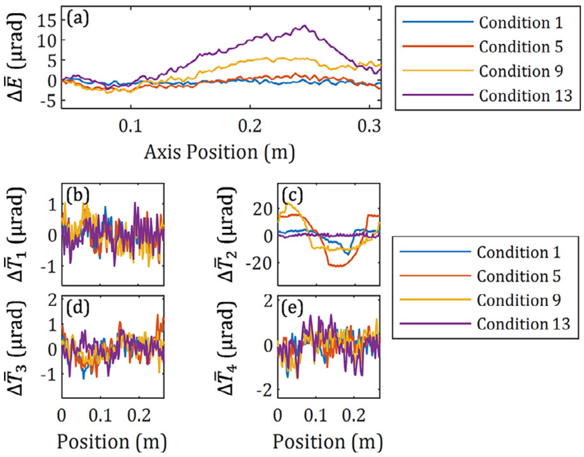

Fig. 8. Change in (a) fixed-phase error function and (b)-(e) periodic error functions for pitch error motion for various conditions.

Conventionally, the mean measured error motion (mean of the repeated error motion measurements) would be used for error motion analyses. However, as the Truck 2 condition is changed, the mean measured pitch error motion becomes a less accurate representation of the actual error motion. For example, Fig. 9 shows how the modeled pitch error motion (red curve) from Eqs. (1) and (2) for six replaced balls (Condition 3) accurately tracks the measured error motion (green curve) from state to state, in contrast to the mean measured error motion (blue curve).

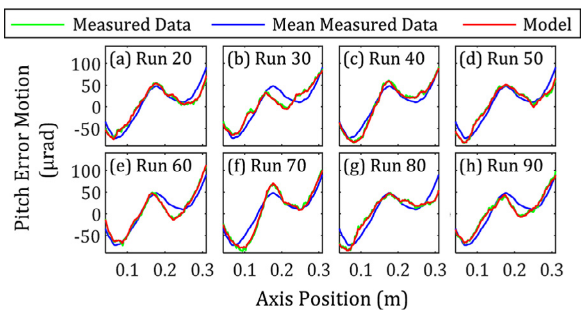

Fig. 9. The measured pitch error motions, mean measured pitch error motion, and modeled pitch error motion for various runs for Condition 3.

Fig. 10 shows that the new method estimates error motions, based on the model from Eqs. (1) and (2), within a standard error of less than 2 µm or 5 µrad, independent of condition. Thus, the new method is able to reduce what would have been perceived as poor repeatability of the mean measured error motion.

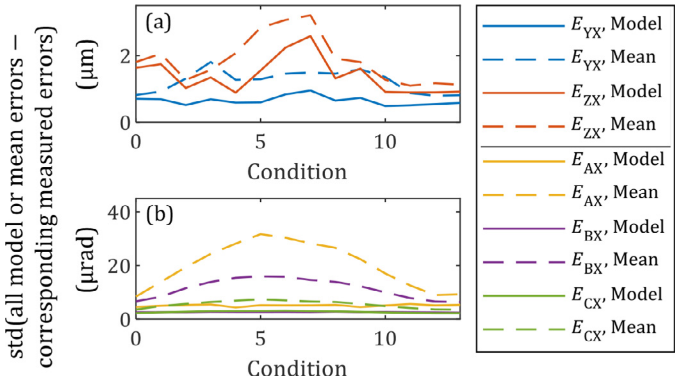

Fig. 10. The standard deviations of all differences between the modeled error motions or mean measured error motions and the measured error motions for (a) straightness error motions or (b) angular error motions.

6. Conclusions

A new method is proposed that uses the measured total error motion data and the measured phase of ball loops within trucks to determine the influence of each ball loop on the error motions. Each truck was instrumented with marker balls and sensors to measure the phases of the ball loops during motion. The method assumes that the influences of each ball loop on an error motion is periodic with the same period and phase as the ball loop itself. The new method solves for the fixed-phase (aperiodic) error component, due to the average interaction of all balls and guideways, and a phase-shifting (periodic) error component for every ball loop. The modeled pitch error motions accurately tracked the measured pitch error motions as balls were replaced in a truck. Most of the phase-shifting error components were initially similar in range, and hence their influences on the pitch error motion were similar, yet the model was able to isolate the physical changes occurring in only one truck.

Results revealed the potential of the new method for health monitoring of linear positioning systems and the influence of bearing balls on surface finish. Once the contributions of each truck on error motions are identified, the performance of each truck can be tracked and maintenance can be planned. The periodicities within the assumed model and the ability to diagnose simultaneous changes in multiple trucks will further be investigated in the next phase of this study. The same method could also be applied to ball screw health monitoring, in which the ball nut is instrumented with a sensor and a marker ball.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Altintas Y, Verl A, Brecher C, Uriarte L, Pritschow G (2011) Machine Tool Feed Drives. *CIRP Ann Manuf Technol* 60(2):779–796.
[2] International Organization for Standardization (ISO). *ISO 230-1 - test code for machine tools – part 1: geometric accuracy of machines operating under no-load or quasi-static conditions*, .
[3] Uhlmann E, Geisert C, Hohwieler E (2008) Monitoring of slowly progressing deterioration of computer numerical control machine axes. *Proc Inst Mech Eng Part B-J Eng Manuf* 222(10):1213–1219.
[4] Vogl GW, Galfond BC, Jameson NJ (2019) Bearing metrics for health monitoring of machine tool linear axes. *2019 manufacturing science and engineering conference (MSEC 2019)*, ASME, . MSEC2019-2911.
[5] Vogl GW, Jameson NJ, Archenti A, Szipka K, Donmez MA (2019) Root-cause analysis of wear-induced error motion changes of machine tool linear axes. *Int J Mach Tools Manuf* 143:38–48.
[6] Teti R, Jemielniak K, O'Donnell G, Dornfeld D (2010) Advanced monitoring of machining operations. *CIRP Ann Manuf Technol* 59(2):717–739.