Published in final edited form as:

Int J Mach Tools Manuf. 2019 ; 143: 38–48. doi:10.1016/j.ijmachtools.2019.05.004.

Root-cause analysis of wear-induced error motion changes of machine tool linear axes

Gregory W. Vogl^{a,*}, N. Jordan Jameson^b, Andreas Archenti^c, Károly Szipka^c, M. Alkan Donmez^a

^aEngineering Laboratory, National Institute of Standards and Technology (NIST), 100 Bureau Drive, Gaithersburg, MD, 20899-8220, USA

^bDepartment of Mechanical Engineering, University of Maryland, College Park, MD, 20742, USA

^cDepartment of Production Engineering, KTH Royal Institute of Technology, Brinellvägen 68, 10044, Stockholm, Sweden

Abstract

Manufacturers need online methods that give up-to-date information of system capabilities to know and predict the performance of their machine tools. Use of an inertial measurement unit (IMU) is attractive for on-machine condition monitoring, so methods based on spatial filters were developed to determine rail wear conditions of linear guideways of a carriage from its IMU-based error motion. Rail wear-induced changes in translational and angular error motions as small as 1.5 μm and 3.0 microradians, respectively, could be resolved. A corresponding two-part root-cause analysis procedure was developed to determine the rail locations of error motion degradation as well as the most probable physical location of damage that causes the detected error motion changes. Another analysis method determined the root cause of non-localized damage along each rail. These approaches support the development of smart machine tools that provide actionable intelligence to manufacturers for early warnings of system degradation.

Keywords

Machine tool; Error; Diagnostics; Wear; Condition monitoring

*Corresponding author. National Institute of Standards and Technology (NIST), 100 Bureau Drive, Stop 8220, Gaithersburg, MD, 20899-8220, USA. gvogl@nist.gov (G.W. Vogl).

Publisher's Disclaimer: Disclaimer

NIST Disclaimer Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States. Approved for public release; distribution is unlimited.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijmachtools.2019.05.004>.

1. Introduction

Machine tools possess many geometric errors (e.g., there are 21 geometric errors for a three-axis machine tool) that affect the accuracy of machined parts [1]. Machine tool linear axes move the cutting tool and workpiece to their desired positions for component manufacturing [2], but over a machine tool's lifetime, various phenomena cause performance degradation, lowering accuracy of workpieces [3].

Yet proper assessment of axis performance degradation is often a manual, time-consuming, and potentially cost-prohibitive process. While direct methods for machine tool performance evaluation are well-established [1,4–8] and reliable for assessing position-dependent errors, such measurements typically interrupt production [9]. Recently, various types of embedded sensors have been used for in-process or process-intermittent metrology of machine tools [10–12] and dynamic monitoring of machine tool components [13]. Six-degree-of-freedom motion sensors exist within integrated circuit components [14], which facilitates use of an inertial measurement unit (IMU) for on-machine condition monitoring.

Previous work has shown the use of an IMU [15,16] that processes accelerometer and rate gyroscope data to detect changes in the translational and angular error motions due to axis degradation [16,17]. A small calibrated IMU can be permanently integrated within a machine tool; no setup change is required to measure the changing geometric performance. During pauses in production, an axis is commanded to move back and forth at various speeds [16], and the IMU data can be processed to yield metrics and the root causes of performance degradation.

Following the approach outlined in Fig. 1, this paper introduces a novel methodology based on spatial filters and data processing to determine root causes of performance degradation of linear guideways (consisting of rails and trucks) from the measured error motion of the carriage. The paper describes the IMU-based diagnostic methods that assess the locations, magnitudes, and root causes of such degradation for industrial applications.

2. Experimental setups with rail degradations

Two testbeds were designed to test various degradation patterns via defects intentionally imparted on rails of the linear axes. In each testbed, the carriage is guided by two rails using two trucks per rail.

2.1. Experiment for localized damage

The first testbed provides the case of localized damage of the guideway rails (e.g., due to spalling of a certain region). Fig. 2a shows the configuration of the setup including an IMU on the carriage of the linear axis system [16]. As seen in Fig. 2a, the linear axis consists of two rails, Rail 1 and Rail 2, upon which rides four trucks, Truck 1, Truck 2, etc. Each truck has two recirculating ball bearing arrays, one each for the inner and outer raceway of the rail. The interaction of the trucks and the rails causes nominally linear motion of the carriage that carries the IMU for collection of performance data. Fig. 2b shows an example of rail damage, in which the inner raceway of a rail was ground down to have wear with a nominal

depth of 50 μm (Fig. 2c). The length of the ‘localized’ damage region in the axis travel direction was increased incrementally from a state of no damage (Stage 1) to a state with a 75 mm-long damage zone (Stage 15), keeping the depth of damage approximately constant.

With the overall system explained, some details of the system can now be understood. Fig. 3a shows a zoomed-in front view of Rail 1 and Truck 2. Each rail has two raceway grooves upon which rides two recirculating ball loops from each truck, as denoted by the two dashed red circles in Fig. 3a. The balls of the trucks interact with the raceway grooves to provide constraints for a single degree-of-freedom linear motion. However, imperfections in this interaction create geometric error motions in six degrees of freedom. Therefore, as the balls and/or the raceway grooves deform (due to wear or other causes), the geometric performance of the carriage changes to reflect the damage.

In this study for the first testbed, Rail 1 is intentionally degraded to various levels, affecting the carriage motion. Fig. 3b shows a side view of Rail 1 and the carriage assembly. Rail 1 was degraded over a 10-cm-long measurement region, as seen in Fig. 3b. Thus, Truck 1 and Truck 2 interact with this region of the rail, changing the translational and angular error motions of the carriage. To mechanically simulate spalling, a handheld grinder was used to wear the surface of the raceway groove of Rail 1 seen in Fig. 4a. The damage zone length was increased incrementally by about 5.4 mm from its nominal state of no degradation, seen in Fig. 4a, to its final state of significant damage with a length of about 75 mm, seen in Fig. 4c. Fig. 4b shows the relative location and length of the damage zone in the raceway groove for each of the fifteen (15) stages.

For each stage of degradation, fifty (50) runs of IMU data were collected bidirectionally at slow (0.02 m/s), moderate (0.1 m/s), and fast (0.5 m/s) axis speeds [16]. Afterwards, ten (10) runs of laser-based reference data were collected bidirectionally at finite positions of travel, specifically every 1 mm between travel positions 1 mm and 321 mm. The laser-based commercial reference system measures straightness and angular error motions over the travel length of 322 mm with standard uncertainties of 0.7 μm and 3.0 μrad . Even though the IMU and reference data are neither collected simultaneously nor measure the same errors (the IMU measures inertial states, while the reference system measures differences via interferometry), the ‘reference’ system is treated as such without the availability of other data for a more accurate comparison. The IMU data (accelerometer and rate gyroscope data) for each run was processed according to an IMU-based method to yield the 6-DOF error motions [16]. Thus, six error motions are associated with each of fifty runs for all fifteen stages. These error motions were used for physics-based diagnostics described in the following sections.

2.2. Experiment for non-localized damage

The second testbed provides the case of non-localized damage of the guideway rails (e.g., due to deformation). Fig. 5a shows the configuration of the setup including high-precision mechanical pushers, which are placed under both guiding rails, to deform the rails in the vertical direction inducing degradation detected along the entire travel range. Fixturing bolts above the mechanical pushers were removed to enable maximum deformation. Fig. 5b shows an example schematic of rail damage. The magnitude of the damage was increased

incrementally by tightening the horizontal bolt of the mechanical pushers with a precision torque wrench, resulting in a repeatable vertical displacement. In total, five levels of damage were imparted (6 N m, 8 N m, 10 N m, 13 N m, and 16 N m), resulting in center rail deflections as large as 91 μm . Finally, fifty (50) runs of IMU data were collected at each stage of damage, similar to the data collection for the first testbed.

3. Contributors to measured error motions

The IMU-based methods should ultimately be used for diagnostics, prognostics, and health management of linear axes within production machines. Achieving these goals relies on using the IMU data to detect changes in error motions and determine the causes of these changes through diagnostic methods. However, the development of a root-cause analysis first requires an understanding of the general types of error motions, their sources, and their effects on the error motions.

3.1. Error motions

Like any ‘rigid’ body, the carriage experiences 6-DOF error motions to be determined by the IMU-based method [16]: one positioning error motion, two straightness error motions, and three angular error motions. For the experimental setup seen in Fig. 2a, any point on the carriage has three translational error motions as functions of X -axis position: a positioning error (E_{XX}), a horizontal straightness error (E_{YX}), and a vertical straightness error (E_{ZX}). Also, the carriage experiences three angular error motions as functions of X -axis position: a roll error (E_{AX}), a pitch error (E_{YX}), and a yaw error (E_{CX}). As shown in Table 1, the acceptable straightness error range is limited to 20 μm and the acceptable angular error range is limited to 60 μrad for many machine tool linear axes. For degradation tracking, it is then reasonable to desire the ability to track error motion changes on the order of several micrometers or microradians, before tolerances are exceeded.

3.2. Error motion sources and effects

Table 2 shows the physical sources of error motions in linear axes and the main errors affected by these physical sources. As seen in Table 2, the trucks and rails affect mainly the two straightness error motions and the three angular error motions, while the other components (lead screw, ball nut, and controlled motor) move the carriage and affect mainly the one positioning error motion. Technically, all components affect all error motions, but Table 2 focuses on the significant relationships between sources and error motions for root-cause diagnostics. For example, the positioning error motion is impacted by rail damage through angular error motions, but not significantly.

In general, each error motion is composed of repeatable and non-repeatable terms. The non-repeatable terms simply do not repeat from run to run among collected IMU data, e.g., due to thermal drift of the system or ball recirculation within trucks that causes an insignificant contribution on the average error motions. The repeatable components, on the other hand, are repeatable from run to run.

The methods described herein focus on diagnostics based on repeatable error motions. Hence, for each experiment, only the averages of the error motions are used for analysis,

based on the fifty IMU-based datasets and the ten laser-based datasets for each stage of degradation. Fifty IMU-based runs were sufficient for convergence of the results of the following diagnostic methods. Furthermore, as highlighted in Table 2, the rail mainly affects the straightness and angular error motions, but not the positioning error motion (E_{XX}). Because the sixth degree of freedom, the positioning error motion, is not significantly impacted by rail damage, the positioning error motion will not be included in the subsequent analysis.

4. Method to diagnose rail damage

The diagnostic method determines the positions and levels of rail damage from observed changes of the carriage motion, as outlined by example in Fig. 6. Changes in measured error motions (with respect to the healthy conditions) are processed to isolate only the damage-caused error motions and translate them to errors in each guide rail/truck interface location. The appropriate thresholds are then generated to assess the level and position of the damage along the guideway rail. The algorithms for diagnosing localized and non-localized damages are slightly different. The appropriate algorithm is currently selected manually; automating that step is left for future work. The details of the method are explained below for the case of localized degradation.

4.1. Primary method for localized rail damage

The use of the IMU to measure the translational and angular error components of linear axis motion and comparison with traditional laser-based measurements are described in Ref. [16]. Each of these error motions is plotted as a function of ‘axis position’, which refers to the distance travelled by the carriage relative to its starting position. An example is shown in Fig. 6a, where the pitch error motion corresponding to the undamaged (“healthy” Stage 1) guideway is compared against the increased (degraded) error motion corresponding to the guideway with localized damage.

The relative error motion is created via subtraction of the “healthy” response from the degraded response. For the cases of localized damage, ideally, the regions corresponding to undamaged locations will have no relative error motion. However, due to thermal distortions, axis motion non-repeatability, and sensor/instrumentation noise, the relative error motion is typically non-zero even at these undamaged regions. Hence, appropriate filters (‘offset functions’) are needed to isolate the relative error motion data associated with the damaged regions for further processing. The offset function is essentially the change in error motion that is not due to rail damage. The damaged region is identified by iteratively comparing, via the use of thresholding, the assumed damaged region to the rest of the data, which is assumed to be associated with the undamaged regions.

Fig. 6b shows the process to determine an initial offset function. The IMU-based method [16] integrates rate velocity, $\dot{\theta}(t)$, or acceleration, $\ddot{\theta}(t)$, as part of the process to yield relative angular displacement, $\Theta(t)$, or relative translational displacement, $\delta(t)$. The relative displacements are changes relative to the “healthy” error motions as functions of axis position, x . However, the integration of slowly-varying biases and drift causes approximately linear or quadratic offset functions:

$$\Theta(x) \approx \theta(x) + a_{\theta}x + b_{\theta} \quad (1a)$$

$$\Delta(x) \approx \delta(x) + a_{\delta}x^2 + b_{\delta}x + c_{\delta} \quad (1b)$$

The offset function is determined via fitting through the data that may belong to undamaged regions, as seen in Fig. 6b. This can be achieved by considering the response of a healthy linear axis to a commanded movement. The first derivative of $\Theta(x)$, the relative angular error response, yields $\Theta'(x) = \theta'(x) + a_{\theta}$ and the second derivative yields $\Theta''(x) = \theta''(x)$. The product of the first and second derivatives is: $\Theta'(x)\Theta''(x) = (\theta'(x) + a_{\theta})\theta''(x)$. Assuming $a_{\theta} \ll \theta'(x)$, then $\Theta'(x)\Theta''(x) \approx \theta'(x)\theta''(x)$. Ideally, for undamaged rails, $\theta(x)$ should equal zero due to no change in error motion, and $\Theta'(x)\Theta''(x) \approx \theta'(x)\theta''(x) \approx 0 \times 0 = 0$. Hence, for angular data, we look for locations where the product of the first and second derivatives is approximately zero. For the translational data, we look for locations where the product of second and third derivatives is approximately zero, since the offset is approximately quadratic as seen in Eq. (1b).

Fitting points through undamaged regions, as seen in Fig. 6b, depends on the definition of ‘localized’ damage used in the following method. Initially, a new guideway has no damage (zero degradation), but after a long period of cutting operations, the contact surfaces of the guideways can develop localized damage [19] from material fatigue, abrasion or adhesion between the trucks and rails [3] as well as material debris from machining processes [20]. Hence, the percentage of localized damage increases from zero as the damage progresses and potentially affects the error motions. The locations and magnitudes of this localized wear depends on factors such as the machining forces, positions of machined parts [19], and forces due to acceleration/deceleration of masses. Eventually, the localized damage may progress into non-localized damage once a certain percentage of the travel range is degraded. In the following algorithm, 70% was chosen as that threshold delineating the transition from localized damage to non-localized damage, leaving at least 30% of the travel range, which is undamaged, to be fitted for the scenario of ‘localized’ damage. Hence, in the first iteration shown in Fig. 6b and 30% of the total data with minimum values of the product of the first and second derivatives or the second and third derivatives is used to generate the initial offset function by a least-squares fit.

The offset function is either a fitted line, for angular error motions, or a fitted parabola, for straightness error motions. This function results in a set of data points that are outliers (points that are not in the undamaged regions), which will be removed in the next iteration. Also, it should be noted that before computing each derivative, the data is filtered using a Savitzky-Golay linear lowpass filter (cutoff wavelength of 20 mm) to mitigate the effects of noise in the data. The cutoff wavelength of 20 mm was chosen to be relatively small compared to the travel range, in order to keep larger defects, but yet relatively large to filter out higher frequencies related to noise and the recirculating truck bearing balls (nominal diameter equals 3.965 mm). At the same time, a sharp peak with a width smaller than 20 mm would be “smoothed” and could be detected if the peak is truly correlated with position.

The initial offset function of Fig. 6b is subtracted from the initial relative error motion data to yield an offset-corrected relative error motion, which is then filtered with a Savitzky-Golay linear lowpass filter (cutoff wavelength of 20 mm). The offset-corrected relative error motion is nominally zero, except at locations with degradation. Because the relative error motion is nominally zero at undamaged locations, a tolerance band is used to separate all undamaged and damaged locations. Each instance of the filtered relative error motion data within (or outside) the tolerance band is associated with an undamaged (or damaged) region. The half-width tolerance is computed as $T = T_0 + k \cdot SD\{\text{undamaged filtered data points}\}$. The base tolerance T_0 is selected as 1.5 μm or 3 μrad for straightness error motions (E_{YX} and E_{ZX}) or angular error motions (E_{AX} , E_{BX} and E_{CX}), respectively. Also, $SD\{\text{undamaged filtered data points}\}$ is the standard deviation (SD) statistic of the data points, and the multiplier k was set to a value of one. The base tolerances can be adjusted by machine users depending on needs, but as chosen here, they are more than ten times smaller compared to tolerances for typical machine performance [18]. Thus, the tolerance is composed of a user-defined base tolerance, which signifies the minimum desired deviation that constitutes “damage”, as well as an increase to account for statistical variations (the SD) in the data.

Because the undamaged locations are now defined based on a tolerance, a revised offset function is fit through the undamaged points (see Fig. 6d) and then removed from the relative error motion, leading to a new set of undamaged points and an updated tolerance. This process is repeated until the maximum change of offset function values (for every axis position) from iteration to iteration is less than $10^{-3} \mu\text{m}$ or $10^{-3} \mu\text{rad}$, usually within eight iterations, leading to the degradation error curve (relative error motion corrected by removing the offset function) shown in Fig. 6e. The localized degradation locations and values are now converged and able to be used for metrics and root-cause analysis.

4.2. Alternative method for localized rail damage

Another approach to determine the degradation error curve in a more robust, albeit computationally inefficient, manner is to use the algorithm shown in Fig. 6c. After identifying the set of points through which the offset function fit is created (see Fig. 6b), the set is refined in an iterative fashion, where for each iteration, 30% of the data points closest to the offset function are retained and then the offset function is re-fitted through those points. This process is repeated until the maximum change of offset function values (over the entire range of axis positions) from iteration to iteration is less than 0.1 μm or 0.1 μrad . The convergence criterion removes the outliers seen in Fig. 6b and ensures the identification of undamaged regions by the offset function (Fig. 6c), leading to a more robust initial set of points to the thresholding of Fig. 6d. In the example shown in Fig. 6c, nine iterations were needed to reach convergence. The algorithm works on the assumption that no more than 70% of the available travel range is degraded. In the experiment described herein, this algorithm captured many of the undamaged points, but not all of them; it yielded only an initial set of points that is a robust input to the final separation algorithm (Fig. 6d and e), which converges to less than $10^{-3} \mu\text{m}$ or $10^{-3} \mu\text{rad}$ within typically seven iterations.

4.3. Method for non-localized rail damage

In the case of the non-localized damage of the guideway rails (Fig. 3), since the performance over the entire travel range is affected, determination of the offset function is not possible. Therefore, the above described algorithm is modified. After calculation of the original relative error motion data (see Fig. 6b), data is processed using the same Savitzky-Golay linear lowpass filter, and only the base tolerance T_0 is used to determine the locations with significant damage.

5. Experimental results

5.1. Localized rail damage results

The localized-damage method outlined in Section 4 was applied to the data for the localized damage experiment (see Section 2.1). The resulting degradation error curves for various stages of damage (Stages 3, 7, 11, and 15) and detected axis positions and magnitudes of performance degradation are shown in Fig. 7. In this case, two performance degradation regions exist because two trucks (Truck 1 and Truck 2) interact with the same damaged section of the inner raceway of Rail 1 (Fig. 3b). The geometry of the damage affects this interaction. For example, the damage for Stage 15 was measured with a depth probe to have two distinct valleys along the direction of travel, which explains the two peaks between 0.1 m and 0.15 m for its degradation error curves in Fig. 7.

For verification and validation purposes, Fig. 8 shows the error motion changes determined from the laser-based data. As can be seen, the IMU-based results (see Fig. 7) closely match those from the laser-based commercial system (see Fig. 8) in both degradation magnitudes and locations of degraded regions. Hence, it is demonstrated that the IMU can provide a similar level of fault diagnosis as the laser-based systems, but at a much lower time expenditure and with greater setup flexibility. For a proper comparison, the laser-based results were used after transforming the error motions along the commercial system's measurement functional line to the IMU's measurement function line.

The performance degradation regions determined for fifteen stages of rail damage are summarized (in blue color) in Fig. 9. The dashed bounds in Fig. 9 denote the axis positions where engagement of Truck 1 or Truck 2 with the damage zone starts or ends, determined from dimensions and positions of trucks with respect to the actual damage zone on the rail. All identified degradation regions should be within the dashed bounds. However, as seen in Fig. 7b and d, there are degradation regions detected outside the bounds of engagement with the damage zones (i.e., the dashed lines). These can be considered as false positive and can be eliminated by increasing the base tolerance T_0 . The presence of false positives must be weighed by the user as part of the balance between detection sensitivity and the acceptable amount of rail damage.

After computing the degradation error curves, metrics can be used to track the performance degradation of linear axes. One metric, the "degradation area" is the integral of the absolute value of the relative error motions (Fig. 7) over only the degraded regions. Fig. 10 shows how the degradation area metric can be used as a reliable quantitative measure of rail damage. Other metrics were calculated for potential use as a measure of rail damage, as

shown in Table 3. The use of any metric depends on user needs. For example, perhaps the ‘Max’ metric could be more useful instead of the ‘Area’ metric for large-travel axes.

5.2. Non-localized rail damage results

The non-localized-damage analysis method was applied to the non-localized damage cases (Fig. 5). Fig. 11a shows that the change of vertical straightness error motion relative to that for zero torque (Stage 1) is significantly non-zero over most of the axis travel, as the input torque increases from 6 N m (Stage 2) to 16 N m (Stage 6). Degradation area metric values have a greater uncertainty for non-localized degradation cases, because ‘offsets’ remain in the method to diagnose rail wear (Section 4).

Nonetheless, Fig. 11b shows that the degradation area metric for this case has a similar pattern to that of Fig. 10 for localized degradation. The metric in Fig. 11b also increases after a critical stage (the second stage as indicated by the input torque of 6 N •m) to cause the rail deformation.

6. Root-cause analysis for localized rail wear

The trucks and rails typically affect the two straightness error motions and the three angular error motions, while the other components (lead screw, ball nut, and controlled motor) move the carriage and usually affect the one positioning error motion. A comprehensive root-cause analysis should not only separate the coupled influences of the trucks and rails on the straightness and angular error motions, but also determine the physical locations and causes of these error changes.

A method was developed to determine the root cause(s) of the relative error motions seen in Fig. 7. For localized rail damage cases, four possible physical causes exist: inner or outer raceway damage on Rail 1 or 2. The root-cause analysis should yield the specific locations of rail damage causing the changes in error motion. The expectation is that this rail-based root-cause analysis method is the first step towards a comprehensive root-cause analysis that includes the analysis of truck damage. Knowledge of the specific location of raceway damage could be vital as input for truck-based root-cause analysis, because of the mechanical coupling of trucks and rails.

6.1. Method to identify rail locations of damage

The first part of the method determines the physical locations of damage. Each error motion is sensitive to damage in varying degrees, resulting in little/asymmetric/many detected axis positions of damage in Fig. 9a/b/e. The method must distinguish the unique “common” and “uncommon” rail locations of damage, despite the varying sensitivities seen in Fig. 9. Fig. 12a shows the “common” and “uncommon” zones for the setup of Fig. 2a. All four trucks move within the “common” zone, but only two of the four trucks move within their respective “uncommon” zone. For example, whenever Truck 1 and Truck 4 are located at Point C (a position along a rail), which is within the common zone, Truck 2 and Truck 3 are in an uncommon zone. The ‘rail position’, henceforth, denotes the distance of a point on the rail from the rear-most end of the rail (see Fig. 2a).

Boolean logic is used to determine the rail locations of damage, as seen in the method outlined in Fig. 14. First, the method to detect degradation is performed according to Section 4. Then, the degradation error motions at the IMU are transformed to error motions at each nominal truck location via the use of rigid-body kinematics [21], as represented within Fig. 13. The transformation vector points from the location of the accelerometer sensor to the nominal center-rail location for each truck. Note that the angular error motions are the same for any point within a rigid body, so the rate gyroscope position is not involved in the transformation process.

Next, due to the varying sensitivities of detected degradation, the union of all five sets (one per error motion) of “common” degradation positions is determined for each of the four trucks. Note that the sixth degree of freedom (DOF), the positioning error motion (E_{XX}), is impacted by rail degradation through angular error motions, but not significantly enough to be included in the analysis. Then, the intersection of these four sets yields the “common” damaged rail locations. Furthermore, the method was applied to a testbed with four trucks, but could be applied to machine tools with more than four trucks. A similar process yields the unique “uncommon” damage locations for each stage of damage.

6.2. Example of method to identify rail locations of damage

While the method is applied sequentially for all stages of damage, Fig. 15 shows an example of the process of Fig. 14 where significant damage (Stage 12) exists. As seen in Fig. 15a, the five DOF error motion changes for each truck are combined and broken into two groups: rail locations of damage within the “common” physical zone (Fig. 15b) and damage locations within the “uncommon” zones (Fig. 15c). Then, the intersection of the four “common” sets yields a rather strict set, P^c , of the “common” rail locations of damage (Fig. 15d).

Alternatively, the union of the four “common” sets yields a rather loose set, P_{loose}^c , of possible common locations of damage (Fig. 15f). This “loose” set is used within the other branch of Boolean logic that determines the damage within the “uncommon” zones. As shown in Fig. 15e, the intersection of the “uncommon” locations is performed for each pair of trucks, the front trucks (Truck 1 and Truck 4) and the rear trucks (Truck 2 and Truck 3). Then, these truck-paired sets are compared to “shifted” versions of the loose set, P_{loose}^c , where the shifting is based on the center-to-center distance between the trucks, d . If any truck-paired points are not within the shifted loose sets, then those points are considered to be locations of damage in the “uncommon” zone. The result for the given example is that there are no “uncommon” locations of damage (Fig. 15g) because all truck-paired points are within the shifted loose sets. Because of the application of intersections and unions within Boolean set theory, this method gives priority to finding robust locations of damage within the “common” zone, such that an “uncommon” location of damage must be significant to be detected. Finally, the common positions of damage (P^c) and the uncommon positions of damage (P_{front}^{uc} and P_{rear}^{uc}) are combined to yield all rail locations of damage (Fig. 15h).

6.3. Rail locations of damage for first experiment

As seen in Fig. 16, the damage determination method of Fig. 14 was applied to the experiment of Section 2.1 to yield the detected rail positions of damage for all stages of

damage. The two error motion degradation regions seen in Fig. 7 were determined correctly to be caused by one “common” physical damage zone with no “uncommon” locations, for all stages of damage. Fig. 16 shows that the detected rail positions of significant damage mainly begin at Stage 8 and exist within 10 mm of the dashed lines in Fig. 16, which denote the true rail locations of the damage region (Fig. 4b). The degradation area metric of Fig. 10 also mainly began to increase at Stage 8, illustrating how the truck behavior does not change significantly until the ‘localized’ damage length is about 32 mm, increased from about 26 mm (Stage 7), which is approximately half of the 50 mm of “contact length” of the balls within a truck on the rail. Therefore, for the given unloaded case, a truck can “sense” the damage whenever most of its bearing balls in contact with a rail raceway are in contact with the ‘localized’ damage.

6.4. Method to identify most probable physical location of damage

The second part of the root-cause analysis method ranks the four damage modes (inner or outer raceway damage on Rail 1 or 2) to determine the most likely cause of damage at each location identified in the first part of the method, i.e., the locations shown in Fig. 16. First, when a given truck interacts with a damage zone, each of five error motions could change in a negative (N) or positive (P) manner. Table 4 shows the expected sign changes in error motions according to basic mechanics, whenever each truck moves over a damaged area on the inner or outer raceway. The error motion components of the truck will be compared to these expectations.

Because the identified positions in Fig. 16 are in the “common” zone of the rail, one of the four trucks could interact with damage on an inner or outer raceway to cause the detected error motion changes. Therefore, for each of the eight possible raceway states (four trucks with two raceway interactions per truck), each error motion is expected to change positively or negatively, as shown in Table 4. The method first assigns values to each error motion based on how the expected sign change matches the detected sign changes from the analysis of Section 4. If the detected change of error motion is greater in magnitude than the tolerance T , then the detected error motion is statistically significant and considered as a “match” or a “mismatch”, respectively, when the detected sign change matches or opposes the expected sign change seen in Table 4. For example, if the detected change of an error motion is $-10 \mu\text{m}$, but the expected sign change is positive, then despite the sufficient magnitude, the error motion is considered to be a “mismatch”. On the other hand, if the detected change of error motion is sufficiently small (e.g., $0.1 \mu\text{m}$) to be within the tolerance band, then the error motion is “inconclusive” at the truck position, irrespective of the detected sign change. Consequently, at each rail position, each of the eight possible raceway states has five values assigned as a label (“match”, “inconclusive”, or “mismatch”) for each error motion ($E_{YX}/E_{ZX}/E_{AX}/E_{BX}/E_{CX}$). Combinations of these five values results in a metric μ for each raceway state. The metric μ is calculated for each of the eight raceway states according to

$$\mu = \begin{cases} 0.2 \cdot M & , I = 0 \\ 0.2(M + 0.5^I) & , I > 0 \end{cases} \quad (2)$$

where M represents the number of “match” values at the given position, and I represents the number of “inconclusive” values at the given position. Hence, the metric ranges from zero (mismatch of all signs) to one (match of all signs).

Comparison of the eight metric values at each position of damage yields the most probable cause of damage out of four possibilities: inner or outer raceway on Rail 1 or Rail 2. The most likely root cause has the highest value of the metric. For example, Fig. 17 shows the most probable cause of damage (color-coded) for each identified location of damage (see Fig. 16) for the experiment of Section 2.1. The degradation was determined to be mainly caused by damage on the inner raceway of Rail 1, which is the correct diagnosis. In fact, most of the damage locations had a perfect metric value ($\mu = 1$).

6.5. Illustration of method to identify most probable physical location of damage

This metric-comparison process is illustrated in Fig. 18. Fig. 18a shows the most probable source of detected degradation for Stage 11, based on the maximum of eight values at each position: the four metric values for Rail 1 (Fig. 18b) and the four metric values for Rail 2 (Fig. 18c). As seen in Fig. 18a, most of the degradation locations had a perfect metric value ($\mu = 1$) for damage on the inner raceway of Rail 1, which is consequently identified as the most probable source of the detected degradation. On the other hand, as seen in Fig. 17, there are some rail positions for Stage 11 with the most probable cause of damage as the inner raceway on Rail 2.

In order to help explain this discrepancy, Table 5 and Table 6 (based on Table 4) compare the signs for the expected and measured error motion changes at Point F and Point G of Fig. 18, respectively. These sign comparisons are used in Eq. (2) to yield a metric μ for each of the eight possible raceway states related to the four possible spalling locations, as shown in the last columns of Tables 5 and 6. The most probable root cause of degradation is deemed to have the highest value of the metric: the inner raceway of Rail 1 ($\mu = 1$) at Point F, and the inner raceway of Rail 2 ($\mu = 0.9$) at Point G.

However, as seen in Table 6, the second highest possible location of damage at Point G is the inner raceway of Rail 1 ($\mu = 0.8$). The difference between the two metric values of 0.8 and 0.9 is because of a sign “mismatch” for an angular error (E_{AX}) in Table 6 that yields $\mu = 0.8$, in contrast to only a sign “inconsistency” that yields $\mu = 0.9$. Such a situation illustrates how uncertainty (no statistically significant error motion change) is valued slightly higher than incorrectness (a mismatched sign) in Eq. (2). Furthermore, this discrepancy occurs when the rail position is within 4 mm of 0.34 m, a limit of detected degradation seen in Fig. 18a. As both limits of detectability are approached, Fig. 18b shows how the highest metric tends to decrease, leading the way for slight discrepancies in this probability-based root-cause diagnostic method.

6.6. Summary of root-cause analysis for localized rail wear

Despite minor discrepancies, the two-part root-cause method can (1) identify the rail locations of degradation and (2) determine the most probable physical location of damage that causes the detected error motion changes. For the experiment of Fig. 2b, the degradation was correctly determined to be mainly caused by damage on the inner raceway of Rail 1 at

the known damage locations. In fact, because the root-cause method essentially ranks the physical behavior based on possibilities from basic mechanics, the method can be quite robust, e.g., for Point C and Point D in Fig. 12. Even though Point C and Point D are very close in rail position (see Fig. 18a), Point C corresponds to an axis position of 0.1358 m (Fig. 12c) while Point D corresponds to a different axis position of 0.2433 m (Fig. 12d). The physical reasoning for this is because Fig. 12c or Fig. 12d corresponds to whenever Truck 1 or Truck 2, respectively, is in contact with the damage on the inner raceway of Rail 1; the trucks cross the same “common” rail positions, but at different times. In other words, any “common” rail position (Point C or Point D) can be reached by two axis positions. As seen in Fig. 12c, whenever Truck 1 interacts with the damage, the truck moves as expected, leading to the highest metric value ($\mu = 1$) in Fig. 18b. On the other hand, as seen in Fig. 12d, whenever Truck 2 interacts with the damage, the truck does not move as expected; Truck 3 moves the most. This unexpected behavior leads to imperfect metrics ($\mu < 1$) for Truck 2. Nonetheless, Fig. 18b shows that a relatively high metric of $\mu = 0.8$ is still detected via Truck 2 on the inner raceway of Rail 1. Thus, the method determined the correct source of damage despite behavior that is physically real yet not completely expected.

7. Root-cause analysis for non-localized rail wear

Finally, Fig. 19 shows a method that was developed to determine the root cause of the non-localized degradation (Fig. 11a) for the nonlocalized damage experiment (see Section 2.2). First, the IMU data was transformed to the four truck locations (see Fig. 5) to obtain each truck’s Z-axis motion. Assuming a sufficient truck stiffness, the measured vertical deviation is approximately that of the rail. Next, because the motions of the two trucks of the rail overlap in a “common” physical zone of the rail (Fig. 19a), the vertical deviations should also overlap therein. In general, the two deviations match in this “common” zone to within a bias that was removed via a least-squares fit. The two vertical deviations (one per truck) were “stitched” via averaging to yield the vertical deviations along the rail (Fig. 19b). Then, the change of the vertical deviation was calculated via its difference from the initial deviations resulting from an experimentally applied torque of 0 N·m (Fig. 19c). Finally, the vertical deviations are normalized to be positive with respect to the assumed stiff machine base (Fig. 19d).

Fig. 20 shows the result of the stitching and normalization process for each rail (Rail 1 and Rail 2) and for each case of degradation. The carriage error motion seen in Fig. 11a is traced to the deformation of the rails shown in Fig. 20. Furthermore, the estimated vertical deviations at the rail centers follow the experimentally measured values (color-coded circles in Fig. 20) to within about 20 μm .

8. Conclusions

Manufacturers need online methods that give updated information of system capabilities to know and predict the performance of their machine tools. An inertial measurement unit (IMU) that is small and relatively inexpensive is attractive for on-machine condition monitoring. By measuring the changing geometric error motions of a carriage due to mechanical wear of linear guideways with the IMU, data can be processed to yield metrics

and the root causes of performance degradation. Data analytics can then be performed to optimize production and lead to preventative maintenance of manufacturing assets.

An iterative method based on spatial filters and thresholding was developed to determine performance degradation for localized damage, under the assumption that no more than 70% of the available travel range is damaged. It was shown that rail wear-induced changes in translational and angular error motions as small as 1.5 μm and 3.0 μrad , respectively, could be resolved. A “degradation area” metric was also shown to be of value for tracking the performance degradation of linear axes. A modified version of the method is used for cases of non-localized damage.

A two-part root-cause analysis was developed to determine the locations of damage on the guideway rails. The first part of the method determines the physical locations corresponding to the degradation, while the second part ranks the probability of various damage locations to determine the most likely root cause of damage at each rail position identified in the first part of the method. Most damaged rail locations were identified with the correct root cause (damage on the inner raceway of Rail 1). Another root-cause analysis method based on transformations and data “stitching” was developed to determine the non-localized rail deviations for each rail, correctly identifying the mechanical deformation of the rails.

These analytical methods show the potential use of IMU-based data to inform machine tool users of the locations, magnitudes, and root causes of performance degradation. Improvements and further testing could enable differentiation for cases with simultaneous localized and non-localized degradations. Specifically, a more comprehensive root-cause analysis could be developed to distinguish various sources of changes of error motions from one another, including changes due to friction and thermal deformation. Additional data such as motor currents, rotor and linear encoder readings, and component temperatures could complement the current approach to achieve a more detailed root-cause analysis. Furthermore, the proposed root-cause methodology can be extended via similar mechanics-based logic to unique linear axes with various numbers of trucks and rails. The monitoring and diagnostic methods support the development of smart machine tools that can diagnose themselves and provide actionable intelligence to manufacturers.

Acknowledgement

The authors thank Brian Pries, Travis Shatzley, Dan Falvey, and Jay Brandenburg of the Fabrication Technology Group (NIST) for their outstanding contributions with the experimental setup.

References

- [1]. Xiang ST, Altintas Y, Modeling and compensation of volumetric errors for five-Axis machine tools, *Int. J. Mach. Tool Manuf* 101 (2016) 65–78.
- [2]. Altintas Y, Verl A, Brecher C, Uriarte L, Pritschow G, Machine tool feed drives, *CIRP Ann. - Manuf. Technol* 60 (2) (2011) 779–796.
- [3]. Li Y, Wang X, Lin J, Shi S, A wavelet bicoherence-based quadratic nonlinearity feature for translational Axis condition monitoring, *Sensors* 14 (2) (2014) 2071–2088. [PubMed: 24473281]
- [4]. International Organization for Standardization, ISO 230–1 - Test Code for Machine Tools – Part 1: Geometric Accuracy of Machines Operating under No-Load or Quasi-Static Conditions, (2012).

- [5]. Tang H, Duan JA, Zhao QC, A systematic approach on analyzing the relationship between straightness & angular errors and guideway surface in precise linear stage, *Int. J. Mach. Tool Manuf* 120 (2017) 12–19.
- [6]. Wan A, Song LB, Xu J, Liu SL, Chen K, Calibration and compensation of machine tool volumetric error using a laser tracker, *Int. J. Mach. Tool Manuf* 124 (2018) 126–133.
- [7]. Chen JX, Lin SW, Zhou XL, A comprehensive error analysis method for the geometric error of multi-Axis machine tool, *Int. J. Mach. Tool Manuf* 106 (2016) 56–66.
- [8]. He ZY, Fu JZ, Zhang XM, Shen HY, A uniform expression model for volumetric errors of machine tools, *Int. J. Mach. Tool Manuf* 100 (2016) 93–104.
- [9]. Khan AW, Chen W, Calibration of CNC milling machine by direct method. 2008, *Int. Conf. Opt. Instrum. Technol.: Optoelectron. Meas. Technol. Appl* 7160 (2008) 716010.
- [10]. Smith KS, Hocken RJ, *Dynamic Metrology Methods and Systems*. United States Patent and Trademark Office: US8401691 B2, (2013).
- [11]. Sato R, Nagaoka K, Sato T, Machine Motion Trajectory Measuring Device, Numerically Controlled Machine Tool, and Machine Motion Trajectory Measuring Method, United States Patent and Trademark Office, 2015 US9144869 B2.
- [12]. Sato R, Nagaoka K, Motion trajectory measurement of NC machine tools using accelerometers, *Int. J. Autom. Technol* 5 (3) (2011) 387–394.
- [13]. Feng GH, Pan YL, Investigation of ball screw preload variation based on dynamic modeling of a preload adjustable feed-drive system and spectrum analysis of ball-nuts sensed vibration signals, *Int. J. Mach. Tool Manuf* 52 (1) (2012) 85–96.
- [14]. InvenSense Incorporated, MPU-6050 Six-Axis (Gyro + Accelerometer) MEMS MotionTracking™ Device, (2016) <https://www.invensense.com/products/motion-tracking/6-axis/mpu-6050/>.
- [15]. Vogl GW, Weiss BA, Donmez MA, A sensor-based method for diagnostics of machine tool linear axes, *Annu. Conf. Prognostics Health Manag. Soc* (2015) 1–10.
- [16]. Vogl GW, Donmez MA, Archenti A, Diagnostics for geometric performance of machine tool linear axes, *CIRP Ann. - Manuf. Technol* 65 (1) (2016) 377–380. [PubMed: 28603291]
- [17]. Vogl GW, Pavel R, Archenti A, Winnard TJ, Mennu MM, Weiss BA, et al., Identification of machine tool geometric performance using on-machine inertial measurements, 6th International Conference on Virtual Machining Process Technology, VMPT, 2017, pp. 1–6.
- [18]. International Organization for Standardization, ISO 10791–2 - Test Conditions for Machining Centres – Part 2: Geometric Tests for Machines with Vertical Spindle or Universal Heads with Vertical Primary Rotary Axis (Vertical Z-Axis), (2001).
- [19]. Fan KC, Chen HM, Kuo TH, Prediction of machining accuracy degradation of machine tools, *Precis. Eng.-J, Int. Soc. Precis. Eng. Nanotechnol* 36 (2) (2012) 288–298.
- [20]. Uhlmann E, Geisert C, Hohwieler E, Monitoring of slowly progressing deterioration of computer numerical control machine axes, *Proc. IME B J. Eng. Manufact* 222 (10) (2008) 1213–1219.
- [21]. Meirovitch L, *Introduction to Dynamics and Control*, Wiley, New York, NY, 1985, pp. 31–62.

Fig. 1.
Schematic of a root-cause analysis for condition monitoring of error motions of machine tool linear axes.

Fig. 2.

(a) Photorealistic view of rails and carriage assembly of a one-axis testbed, (b) example of localized rail damage and (c) its depth profile measured along the rail cross section using a depth probe with a 3-mm-diameter sphere end.

Fig. 3.

(a) Zoomed-in front view and (b) side view of rails and carriage assembly.

Fig. 4.

(a) Rail with no damage, (b) schematic of relative location and length of damage zone in raceway groove for each stage, and (c) rail with about 75 mm of 'spalling'. Damage zones in (b) were measured within a 10-cm-long measurement region indicated by a ruler in (c).

Fig. 5.

(a) Testbed for low-spatial frequency changes in straightness and example of (b) non-localized rail damage.

Fig. 6.

(a) Mean pitch error motion (rotation about Y -axis, E_{BX}) for 50 runs at the twelfth stage of degradation, and the average of 50 runs at the first stage (or "healthy" stage) of damage; (b) relative pitch error motion along with the initial set of points used to fit the offset function; (c) relative pitch error motion along with the offset function, and a set of data points used to determine the offset function; (d) relative pitch error motion along with the offset function, a set of "undamaged" data points used to determine the offset function, and the damage thresholds used to classify points as damaged or undamaged; and (e) degradation error curve (offset-corrected relative error motion that is nominally zero, except at locations with degradation) after removal of offset function and convergence, along with filtered response and final damage tolerance.

Fig. 7.

Relative error motions for Stages 3, 7, 11, and 15 of the rail-damage experiment of Section 2.1 as measured by the IMU. The regions with bold lines indicate detected degradation.

Fig. 8.

Relative error motions for Stages 3, 7, 11, and 15 of the rail-damage experiment of Section 2.1 as measured by the laser-based reference system. The regions with bold lines indicate detected degradation.

Fig. 9.

Detected damaged axis positions for (a) E_{YX} , (b) E_{ZX} , (c) E_{AX} , (d) E_{BX} , and (e) E_{CX} , based on IMU data processed with the damage detection algorithm. The dashed lines denote axis positions where engagement of Truck 1 or Truck 2 with the damage zone starts and ends.

Fig. 10.

Area under the damaged regions for every rail-damage stage based on IMU measurements.

Fig. 11.

(a) Mean vertical straightness error motion change for each rail-degradation stage and (b) “damage area” metric versus torque.

Fig. 12.

(a) Illustration of “common” and “uncommon” rail positions over which trucks move, and (b–e) various schematic representations of the changes of rigid-body motion of the carriage with Trucks 1, 2, 3, and 4 at rail locations B, C, D, and E for Stage 11 of the localized-damage experiment. The different rail positions for Truck 1 and Truck 2 are denoted with square and circular symbols, respectively.

Fig. 13.
Schematic of transformation of error motions from the accelerometer within the IMU to nominal truck positions.

Fig. 14.

Flowchart of process to determine the rail positions of damage.

Fig. 15.

Example of process to determine the rail positions of damage.

Fig. 16.

Detected rail positions (solid black) of damage for 530-mm long rails as a function of degradation stage. In reality, the rail damage exists only within physical bounds (see dashed lines) for each stage.

Fig. 17.

Most probable source of detected damage as a function of rail position using data collected by the IMU. Only the inner raceway of Rail 1 truly has 'spalling' that exists within physical bounds (see dashed lines) for each degradation stage.

Fig. 18.

(a) Most probable source of detected damage as a function of rail position for Stage 11, based on the metric values for (b) Rail 1 and (c) Rail 2. Point F is at Rail Position = 0.2724 m and Point G is at Rail Position = 0.3014 m.

Fig. 19.
Method to estimate non-localized vertical deviations of the rails.

Fig. 20.

Estimated vertical deviations of (a) Rail 1 and (b) Rail 2 for the non-localized rail damage experiment. Experimental data points are color-coded circles that correspond with the IMU-based results.

Table 1

Tolerances for linear axis errors of vertical machining centers.

Error	Tolerance ^a
Straightness	20 μm
Angular	60 μrad

^aFor axes capable of up to 1 m of travel, according to ISO 10791–2 [18].

NIST Author Manuscript

NIST Author Manuscript

NIST Author Manuscript

Table 2

Error sources and error motions in typical linear axis.

Main Error Motions Affected by Source			
Error Source	Positioning	Straightness	Angular
Truck	-	Yes	Yes
Rail	-	Yes	Yes
Lead Screw	Yes	-	-
Ball Nut	Yes	-	-
Controlled Motor	Yes	-	-

Table 3

Computed health metrics.

Metric	Description
Length	Total track length of damage regions
Area	Integral of the absolute value of the offset-corrected error curve over the locations in the damage regions
Mean	Average absolute value of the error at damage regions
Peak-to-peak	Maximum value of error curve minus minimum value of error curve within damage regions
Max	Maximum absolute value of error curve within damage regions

Table 4

Expected sign change of the carriage's error motion when a truck moves over a theoretical spalling region on the inner (outer) raceway of its rail. Each expected sign change is either positive (P) or negative (N).

	E_{YX}	E_{ZX}	E_{AX}	E_{BX}	E_{CX}
Rail 1, Truck 1	N (P)	N (P)	P (N)	P (N)	N (P)
Rail 1, Truck 2	N (P)	N (P)	P (N)	N (P)	P (N)
Rail 2, Truck 3	P (N)	N (P)	N (P)	N (P)	N (P)
Rail 2, Truck 4	P (N)	N (P)	N (P)	P (N)	P (N)

Table 5

Comparison of each carriage error motion sign change with the expected sign change when the truck moves over a theoretical spalling region on an inner or outer raceway of its rail at Point F in Fig. 18. The table cell text is the expected sign change, and the table cells are color-coded as green, red, or yellow if the measured and expected signs match, do not match, or are inconclusive, respectively.

	E_{YX}	E_{ZX}	E_{AX}	E_{BX}	E_{CX}	μ
Rail 1, Inner, Truck 1	N	N	P	P	N	1
Rail 1, Outer, Truck 1	P	P	N	N	P	0
Rail 1, Inner, Truck 2	N	N	P	N	P	0.25
Rail 1, Outer, Truck 2	P	P	N	P	N	0.45
Rail 2, Inner, Truck 3	P	N	N	N	N	0.45
Rail 2, Outer, Truck 3	N	P	P	P	P	0.25
Rail 2, Inner, Truck 4	P	N	N	P	P	0.4
Rail 2, Outer, Truck 4	N	P	P	N	N	0.6

Table 6

Comparison of each carriage error motion sign change with the expected sign change when the truck moves over a theoretical spalling region on an inner or outer raceway of its rail at Point G in Fig. 18. The table cell text is the expected sign change, and the table cells are color-coded as green, red, or yellow if the measured and expected signs match, do not match, or are inconclusive, respectively.

	E_{YX}	E_{ZX}	E_{AX}	E_{BX}	E_{CX}	μ
Rail 1, Inner, Truck 1	N	N	P	P	N	0.5
Rail 1, Outer, Truck 1	P	P	N	N	P	0.5
Rail 1, Inner, Truck 2	N	N	P	N	P	0.8
Rail 1, Outer, Truck 2	P	P	N	P	N	0.2
Rail 2, Inner, Truck 3	P	N	N	N	N	0.6
Rail 2, Outer, Truck 3	N	P	P	P	P	0.4
Rail 2, Inner, Truck 4	P	N	N	P	P	0.9
Rail 2, Outer, Truck 4	N	P	P	N	N	0.1