Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-1 (2016) 385–390

Training data selection criteria for detecting failures in industrial robots

Sathish V [*,***] **Srini Ramaswamy** [**] **Sachit Butail** [***]

[*] *ABB India Development Center, Bangalore, India (e-mail: sathish.v@in.abb.com)*
[**] *ABB Inc, Wickliffe, USA (e-mail: srini@ieee.org)*
[***] *Indraprastha Institute of Information Technology Delhi (IIIT-Delhi), New Delhi, India (e-mail: sbutail@iiitd.ac.in)*

Abstract: We study the effect of source and type of training data on detecting failures in industrial robots using Principal Component Analysis (PCA). Specifically, using field data across multiple robots performing different tasks, we compare two scenarios: first, where training data obtained from a single robot is used to evaluate multiple robots (one-to-many), and second, where each robot is evaluated on the basis of its own training data (one-to-one). We further investigate if the data preprocessing prior to running PCA affects the ability to detect and predict failures. To reduce task dependence of the raw signal, we preprocess the same by computing the absolute difference between successive measurements and compare the results with a PCA model that is built using raw signal alone and another that is built from a combined signal having both raw measurements and their absolute difference. We quantify effectiveness of detecting failures in terms of three measures: coefficient of variation of the Q-residual obtained by projecting the test data on the PCA model, number of samples above a data-driven confidence threshold, and lead time, measured as the number of days prior to failure when the residual error rises above a given threshold. Specifically, we show that while both one-to-one and one-to-many training sources are valid for detecting failures, signal preprocessing has a significant influence. Our results show that coefficient of variation of the Q-residual from a PCA model built using absolute difference between measurements serves as a robust descriptor for predicting and detecting failure in robots in the one-to-many training scenario. With the same signal, when using number of samples above threshold, we find that one-to-one training source is able to detect failure in robots. Finally, with lead time, we find that one-to-one training scenario with absolute difference as signal type can be used to raise warning as early as nineteen days before failure.

Keywords: Industrial Robots, Failure Detection, Principal Component Analysis.

1. INTRODUCTION

In a remote service set-up for industrial robots, information such as events and measurements are sent to a remote database for hundreds of robots (Blanc (2009)). This information is further processed to prepare advisory reports that assist field engineers. The value of a remote service support can be enhanced by improving the capability of predicting and detecting failures across multiple robots.

Failure detection can be improved by modeling the robot operation subject to different loads (Bittencourt (2012)). However, given the variability in the type of tasks and loads borne by industrial robots, and the difficulty in accurately modeling a failure within a robot operation, data-driven approaches provide an alternative black-box type approach to detect failures (Samhouri et al. (2009); Datta et al. (2007)). Data-driven methods do not require prior knowledge about the process and rely entirely on collected data (Lin et al. (2010). Such algorithms learn the nominal behavior of the system from the actual data itself and are validated on the basis of accurate classification of new unseen data called the test data. Anomalous behavior is then detected in the form of error in a model fit (Schwabacher and Goebel (2007); Luhui and Jie (2010); Singh et al. (2011); Bittencourt et al. (2014)).

Industrial robots are typically instrumented with many sensors, monitoring and reporting the health of the robot in terms of multiple measurements including torque loads, temperature, arm speeds, and position. With more than a hundred such measurements reported on an hourly basis, it is not immediately clear which measurement is important for which type of failure.

Principal component analysis (PCA) is a statistical learning approach that helps to reduce the dimensionality of the input dataset under the assumption that the data points are linearly separated. It has been used in the past effectively for dimensionality reduction (Howley et al. (2006); Lin et al. (2010); Pettersson (2005)) and fault detection in equipment and processes (Ahmed et al. (2012); Pan et al. (2013); Mnassri et al. (2008); Wise and Gallagher (1996); Penha and Hines (2001); Yu (2011)). However, PCA effectiveness is evaluated on the basis of model fit which traditionally entails using data from the same source

2405-8963 © 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.03.084

that has to be tested (Howley et al. (2006); Ahmed et al. (2012)).

Here, we first investigate if failure detection across multiple robots is significantly affected by the source of training data, that is, a single reference robot for all tests (one-to-many) or nominal data from the same robot (one-to-one) for all tests (Hypothesis 1). Instead of using measurement data as is, which is likely a function of the specific task performed by the robot, we further investigate if signal pre-processing can make detection of failure task invariant. For example, assuming that a failure is accompanied by large changes in measurement values, measurement change over time may be used to represent normal operation across multiple robots. Accordingly, we test if the effectiveness of the data-driven approach depends on three types of signal input used to build the PCA model: raw signal, absolute difference between two successive raw signal values, and a combination of both (Hypothesis 2).

We test hypotheses 1 and 2 on the basis of the descriptors on the residual error obtained after projecting the test data on the PCA model. In particular, we use multiple descriptors such as coefficient of variation of Q-residuals, number of test samples above Q-residual threshold limit, and lead time (in days), defined as the remaining time in days left in the test data set when the first sample above threshold is detected, to quantify the effectiveness of the failure detection and prediction approach using PCA. The contributions of this paper are: (i) we collect data from industrial robots for analysis of fault-detection strategies, (ii) we describe and test alternative data selection criteria to reduce collection time and task dependence, and (iii) we perform experimental analyses to demonstrate the effectiveness of signal processing on a PCA based approach for failure detection.

This paper is organized as follows: Section 2 gives a background on different failures in industrial robots and training strategies for PCA model development. Data collection, preprocessing and experimental analysis are described in Section 3. Section 4 provides the results followed by discussion, conclusion and a summary of ongoing work in Section 5.

2. BACKGROUND

2.1 Failures in Industrial robots

Failures in industrial robots can be classified as mechanical based, sensor related, and actuator system related (Fantuzzi et al. (2003)). Different reasons have been attributed to each of these failures. For example, gearbox and brake failures occur mainly due to wear (Bittencourt et al. (2012)), and motor failure can be due to a short circuit (Fantuzzi et al. (2003)). Because the pattern of change of data and the time of change can vary widely across all failures (Bittencourt et al. (2012)), combining all the failures into a single failure-detection framework is difficult when a single model is used. Instead, an approach that uses all reported measurements and events may be able to capture robot operation for failure detection across multiple tasks and operating environment.

2.2 Training data selection and preprocessing for PCA

Principal Component Analysis is a multivariate analysis technique (Pettersson (2005)) that helps to reduce the number of features to a small number of principal components without losing important information (Luhui and Jie (2010), Pan-Ning Tan (2013)). PCA has been widely used in other fields such as face recognition (Givens et al. (2003)), fault detection in semiconductor processes (Yu (2011)), and anomaly detection in by monitoring sensor data in air plane systems (Schimert (2008)).

Given n instances of m different measurements, the PCA algorithm eigen-decomposes the input data $\mathcal{X} \in \mathbb{R}^{n \times m}$ into the principal components, where the number of principal components are selected such that they account for maximum variation in data (Penha and Hines (2001)). The accuracy with which data can be projected onto the principal components represents the degree of model fit and can be represented using Q-residual (Chen et al. (2004)). Q-residual is calculated as the difference between a sample and its projection into the principle components retained in the model (Yu (2011)). We use a 95% confidence level to set the threshold on the Q-residual beyond which a model is declared not fit (Devore (2002)).

The effectiveness of PCA to classify new data accurately depends on the size and extent of the training data (Givens et al. (2003)). At the same time, a generic PCA model that is built using a large volume of training data may give rise to higher incidence of false negatives in a failure-detection framework (Li et al. (2000)). PCA model is traditionally developed with training data from the same system where the test is to be performed (Ahmed et al. (2012); Lin et al. (2010)). To overcome variability and underfitting due to new data, Li et. al. (Li et al. (2000)) have used a recursive PCA algorithm that is periodically revised to recalculate the thresholds for the Q-residuals. This approach helps to reduce false alarms because of slow and normal changes in the industrial processes over period such as equipment aging and sensor drifting.

3. DATA COLLECTION, PRE-PROCESSING AND EXPERIMENTAL ANALYSIS

3.1 Data collection

Time-series data collected from industrial robots can be classified into two categories: measurements, which in turn can be mechanical (e.g. speed, torque) and non mechanical (e.g. CPU temperature, fan speed), and events, which are generated using rule-based criteria on measurements, reported or otherwise. An example of an event is the over-temperature event, generated because temperature of the main computer unit is too high. We obtained data from seven different instances of four types of failures and three instances of normal operation from ten different robots. The seven failure instances consist of two gearbox failures, two motor failures, two brake failures and one computer unit failure. All robots have six axes performing different tasks within the same application domain. For each robot, measurements and event frequency was extracted by querying the remote servicing database using a custom SQL script. Fault types, wherever a failure instance was found, was reported in the comment log of the service engineer. Robot type and application were confirmed using remote service data [1].

3.2 Preprocessing of data

Data from the three normally operating robots was further split so that six instances of normal operation were available for testing. Specifically, two datasets of 20 days each were extracted from each normal robot so that they were furthest apart in time. Each training and test dataset comprised 60 days of operation time that was further split into first 40 days of training data followed by 20 days of test data. The last day of the 60-day dataset (20th day in the test data) was the day of failure. To avoid including measurements and events very close to failure in time, the last five days of test data (including the day of failure if it occurred) was removed from the test data. This approach also ensured predictive capability of a model for up to at least five days (one work week) before failure. While there was some variability in the number of measurements and events reported by each robot, a total of 104 measurements and 13 events were found to be common across all robots and were therefore used for training and testing. The events were reported in the form of number of observations per day.

The measurements and events were further preprocessed to provide three different types of inputs as training data for PCA: i) raw signal, ii) absolute difference between two successive raw signal values and iii) combination of both (i) and (ii). Some axis-specific measurements such as position and torque were available in the form of a frequency distribution. To compare such values we used a probabilistic measure called Kullback Leibler (KL) distance. While preserving the difference in such measurements, this approach also reduced the total number of measurements. Given two probability distributions of measurement X, $p(x)$ and $q(x)$, KL distance $D(p||q)$, is defined as (Cover and Thomas (2012),Dasu et al. (2006))

$$D(p||q) = \sum_{x \in \chi} p(x) \times \log \frac{p(x)}{q(x)} \tag{1}$$

In the case of raw signal, KL distance was computed between the current measurement and the first measurement. For absolute difference, KL distance was computed between successive measurements. Linear trends in measurements were removed prior to training by detrending the signal. The number of final measurements and events for each dataset were 117 for raw and absolute difference signal, and 234 for the combined signal.

3.3 Test setup

A three-way analysis is designed to test the effect of training source and training signal type and robot condition on the ability to isolate failures using the defined descriptors. The training source was set as one-to-many (OTM), or one-to-one (OTO) (Figure 1); the training signal type was set to raw, absolute difference, and combined; and the robot condition was selected as failure or normal. To avoid bias due to training data from particular robot in the OTM scenario, PCA models were trained using 3 failure and 3 normal robots to create 42 test cases. Table 1 provides a list of test scenarios. The number of Principal Components were chosen in such a way that selected components explains 90% variability in the training data (Fig. 2).

In addition to testing hypotheses 1 and 2 on the 15-day test data, we varied the number of test samples to 10, 12 and, 13 days to identify a higher lead time for predicting failure.

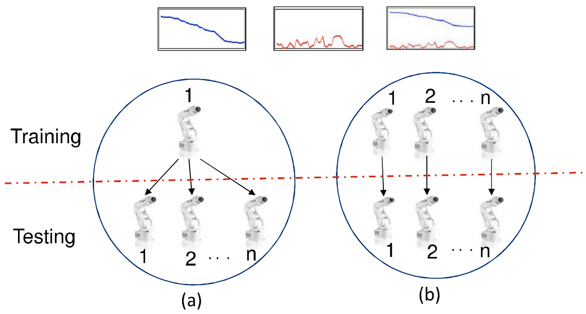

Fig. 1. Test Conditions. (a) One-to-many scenario: PCA model built using training data from a single robot. (b) One-to-one scenario: PCA model is built for each robot using its own training data, shown as dashed line. The data for training and testing the model are: raw signal, absolute difference between successive signal values and combination of both raw and absolute difference. Robot image is obtained from ABB product catalogue.

Table 1. Experimental Design

Training Type	Signal Type	Robot Condition	No of Cases
OTM	Raw	Failure	42
OTM	Raw	Normal	36
OTM	Difference	Failure	42
OTM	Difference	Normal	36
OTM	Combined	Failure	42
OTM	Combined	Normal	36
OTO	Raw	Failure	7
OTO	Raw	Normal	6
OTO	Difference	Failure	7
OTO	Difference	Normal	6
OTO	Combined	Failure	7
OTO	Combined	Normal	6

We compared different scenarios using the following descriptors based on Q-residual: coefficient of variation (Devore (2002)), defined as the ratio of standard deviation to the mean; number of test samples above a threshold limit; and lead time defined as the number of days prior to failure when the model fit error first rises above threshold. Note that since we ignore the last five days of the test data the actual lead time adds five days to the value obtained.

We used analysis of variance (ANOVA) to compare the different scenarios. Given groups of experiments for each independent variable, ANOVA measures the source of variation in the data and compares the relative sizes (Montgomery and Runger (2010)). The sources of variation can

[1] For confidentiality reasons ABB specific information such as robot type, location, and measurements are not disclosed.

be within-group or between-groups. The ANOVA comparison reports the probability of rejection of the null hypothesis (p-value), namely that the groups are not significantly different, and the F-statistic which is the ratio of variation between the group to within the group, we evaluate the significance. A large F value indicates that there is more difference between groups compared to within groups (Montgomery and Runger (2010)). If the p-value is less than 0.05, then the group means are considered to be significantly different.

We assume an interaction model for the following reasons: (i) signal type and training source will likely be impacted by the condition of the robot that is tested; (ii) it is also likely that the impact of the signal type (raw or difference) will be affected by the training source (OTO or OTM); and (iii) the absolute difference signal captures variation in data and is therefore more representative of the condition of the robot (normal or failure), where as raw signal is more representative of the task.

Significance level is set to $p < 0.05$. If an interaction effect is found, post-hoc comparisons are made using Tukey's honest significant difference (hsd) test. All statistics were performed using MATLAB (R2014a, Mathworks, Natick, MA).

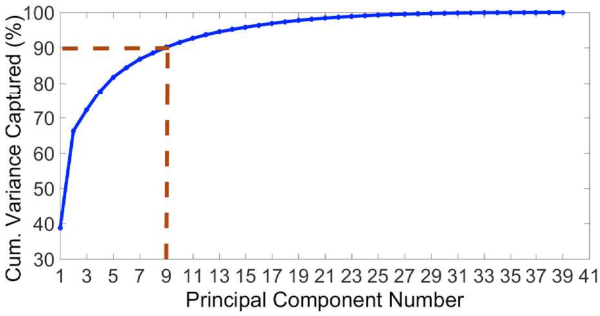

Fig. 2. Selection of Principal Components: number of PCs selected which explains 90% variability in the data .

4. RESULTS

ANOVA comparison of coefficient of variation (COV) of Q-residual across different scenarios shows a combined interaction effect between signal type and robot condition. Specifically, results show that while there was no interaction effect found between training source × signal type × robot condition ($p = 0.789$), training source × robot condition was significant ($p = 0.033$); there was no interaction between training source × signal type ($p = 0.661$) and signal type × robot condition ($p = 0.123$). Averaging across all other variables (main effects), the coefficient of variation from both signal type and robot condition was found to be significantly different (p<0.01), but not for training source (p=0.823). Post-hoc comparisons between different robot conditions across signal types show that the normal and failure robots were significantly different when absolute difference was used as the signal type and training source as OTM (Figure 3). With fewer days of test data, the significant difference between normal and failure robots was lost for 12 days and less.

ANOVA comparison of the number of test samples above threshold limit shows that the interaction effects as well as main effects are significant ($p < 0.01$). Post-hoc analysis for all pairs shows that number of samples above threshold is significantly more for failure robots than normal robots when absolute difference is used as signal type with OTO as the training source (Figure 4). This result of significant difference between normal and failure robot continued to hold even when the test-data was reduced to 10 days.

ANOVA results on lead time in days shows that the interaction effects are significant for all combinations and main effects ($p < 0.01$). Post-hoc comparisons across all pairs of conditions show that when signal type is absolute difference, the lead time is significantly more for a failure robot with OTO training source (Figure 5). This result of significant difference between normal and failure robot continued to hold even when the test-data was reduced to 10 days.

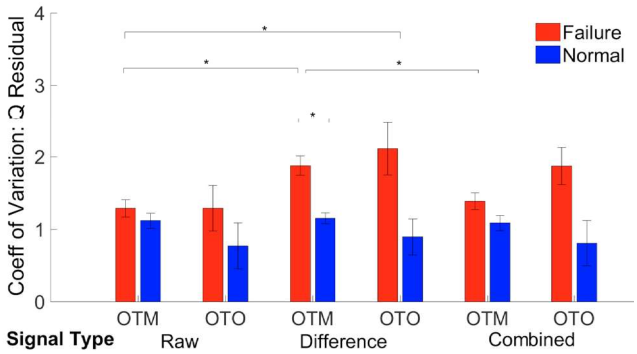

Fig. 3. Coefficient of variation of Q-residual for each robot condition. * indicates significantly different conditions in post-hoc comparisons. Error bars denote standard error of the mean.

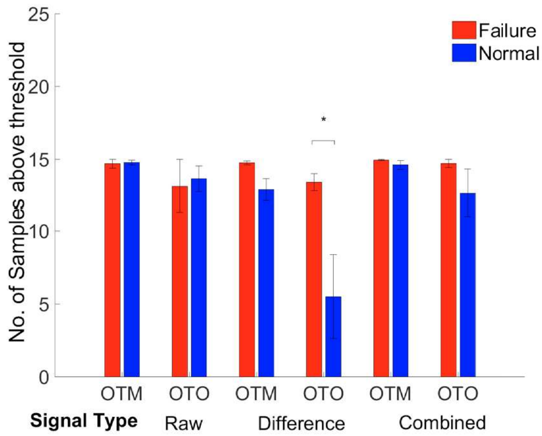

Fig. 4. Number of samples above threshold for each robot condition. * indicates significantly different conditions in post-hoc comparisons. Error bars denote standard error of the mean.

5. DISCUSSION AND CONCLUSION

Our results show that the failure in robots can be detected using coefficient of variation of Q-residual while using signal type as absolute difference and training source as OTM. Using lead time and number of samples above

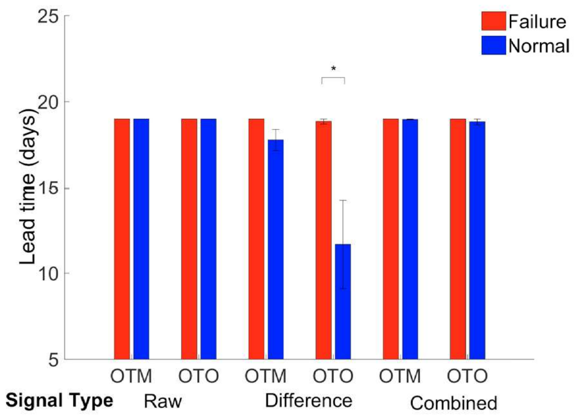

Fig. 5. Lead time (measured as number of days prior to failure) for each robot condition. * indicates significantly different conditions in post-hoc comparisons. Error bars denote standard error of the mean.

threshold as the descriptor, the failure can be detected if training source is OTO and signal type is absolute difference.

Contrary to our expectations in hypothesis 1, training source was not found to significantly affect COV across different conditions. However, the combination of robot condition and signal type was significantly different across other two descriptors. Further analysis showed that it is possible to use a single reference robot to generate training data for testing multiple scenarios if absolute difference is used as the signal type for training the PCA model. The OTO strategy is task and robot dependent and is therefore expected to reflect any variability present in different robots. On the other hand, the OTM strategy along with the absolute difference as the signal type produces a single model to test against and therefore likely dilutes any variability creating a more robust failure detection approach.

In agreement with hypothesis 2, signal type affects the ability to detect a failure in robots. In particular, all descriptors were significantly different for different signal types (main effect). For signal type as absolute difference, all descriptors were able to differentiate between a failure and a normal robot at least seven days before failure. The absolute difference is a measure of change in robot operation and is expected to be similar across different tasks as long as the robot operates normally without any sudden changes. The sampling rate of once per day possibly further reduced any anomalous changes due to task stop and start operations that occurred within a single day.

The significant difference in lead time for detecting failure as the time when the first data point is detected above threshold is an expected result for OTO. At the same time, the high variability in Q-residual itself implies that it is difficult to use a single threshold that can be used across all scenarios. In this respect, the lead time can be used as a warning to further process the data over the next few days for computing COV and number of samples above threshold. This can help in reducing the number of false positives. Between training sources, we find that while OTO finds significant difference between failure and normal robots when using number of samples, OTM is able to register a similar change with COV. In an industrial setting, a combination of these training strategies can be adopted. Using OTM approach, effort on training model for every robot can be avoided and help to identify the robots that may require closer inspection.

Future studies will focus on collecting more data to highlight significant differences and contributing factors to different types of failures and identifying the correlation between events and measurements in time domain. In our analysis, we have considered measurements that are periodic as well as event driven. In ongoing work we seek to extend the current approach to study the effect of measurement type in prediction capability.

ACKNOWLEDGEMENTS

This work was supported by ABB, India Development Center (IDC). Authors are thankful to Mr. Rajeshwar Datta (ABB India) for discussions related to remote service operations.

REFERENCES

Ahmed, M., Baqqar, M., Gu, F., and Ball, A. (2012). Fault detection and diagnosis using principal component analysis of vibration data from a reciprocating compressor. In *Proceedings of 2012 International Conference on Control (CONTROL), UKACC*, 461–466.

Bittencourt, A., Saarinen, K., Sander Tavallaey, S., et al. (2012). A data-driven method for monitoring systems that operate repetitively–applications to wear monitoring in an industrial robot joint. In *Fault Detection, Supervision and Safety of Technical Processes*, volume 8, 198–203.

Bittencourt, A.C. (2012). *On Modeling and Diagnosis of Friction and Wear in Industrial Robots.* Ph.D. thesis, Linkping University, Linkping.

Bittencourt, A.C., Saarinen, K., Sander-Tavallaey, S., Gunnarsson, S., and Norrlöf, M. (2014). A data-driven approach to diagnostics of repetitive processes in the distribution domain–applications to gearbox diagnostics in industrial robots and rotating machines. *Mechatronics*, 24(8), 1032–1041.

Blanc, D. (2009). Service from afar ABBs remote service concept is revolutionizing the robotics industry. In *ABB Review 1.*

Chen, Q., Kruger, U., Meronk, M., and Leung, A. (2004). Synthesis of T^2 and Q statistics for process monitoring. *Control Engineering Practice*, 12(6), 745–755.

Cover, T.M. and Thomas, J.A. (2012). *Elements of information theory.* John Wiley & Sons.

Dasu, T., Krishnan, S., Venkatasubramanian, S., and Yi, K. (2006). An information-theoretic approach to detecting changes in multi-dimensional data streams. In *In Proc. Symp. on the Interface of Statistics, Computing Science, and Applications.* Citeseer.

Datta, A., Mavroidis, C., Krishnasamy, J., and Hosek, M. (2007). Neural network based fault diagnostics of industrial robots using wavelet multi-resolution analysis. In *Proceedings of the IEEE American Control Conference*, 1858–1863.

Devore, J. (2002). *Probability and Statistics for Engineering and the Sciences.* Thomson Learning.

Fantuzzi, C., Secchi, C., and Visioli, A. (2003). On the fault detection and isolation of industrial robot manipulators. In *7th International IFAC Symposium on Robot Control*.

Givens, G., Beveridge, J.R., Draper, B.A., and Bolme, D. (2003). A statistical assessment of subject factors in the PCA recognition of human faces. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, volume 8, 96–96. IEEE.

Howley, T., Madden, M.G., OConnell, M.L., and Ryder, A.G. (2006). The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data. *Knowledge-Based Systems*, 19(5), 363–370.

Li, W., Yue, H.H., Valle-Cervantes, S., and Qin, S.J. (2000). Recursive PCA for adaptive process monitoring. *Journal of process control*, 10(5), 471–486.

Lin, L., Ma, J., Ye, X., and Xu, X. (2010). Mechanical fault prediction based on principal component analysis. In *Proceedings of the IEEE International Conference on Information and Automation (ICIA)*, 2258–2262.

Luhui, L. and Jie, M. (2010). Fault prediction based on data-driven technique. In *Proceedings of the 2010 International Conference on Electrical and Control Engineering*, 997–1001. IEEE Computer Society.

Mnassri, B., El Adel, E., and Ouladsine, M. (2008). Fault localization using principal component analysis based on a new contribution to the squared prediction error. In *Proceedings of IEEE Mediterranean Conference on Control and Automation,*, 65–70.

Montgomery, D.C. and Runger, G.C. (2010). *Applied statistics and probability for engineers*. John Wiley & Sons.

Pan, D., Liu, Z., Zhang, L., Liu, Y., and Zuo, M.J. (2013). A study on applications of principal component analysis and kernel principal component analysis for gearbox fault diagnosis. In *Proceedings of the International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)*, 1917–1922. IEEE.

Pan-Ning Tan, Vipin Kumar, M.S. (2013). *Introduction to Data Mining*. Pearson.

Penha, R.M. and Hines, J.W. (2001). Using principal component analysis modeling to monitor temperature sensors in a nuclear research reactor. In *Proceedings of the maintenance and reliability conference (MARCON)*, 6–9.

Pettersson, O. (2005). Execution monitoring in robotics: A survey. *Robotics and Autonomous Systems*, 53(2), 73–88. doi:10.1016/j.robot.2005.09.004.

Samhouri, M., Al-Ghandoor, A., Ali, S.A., Hinti, I., and Massad, W. (2009). An intelligent machine condition monitoring system using time-based analysis: neuro-fuzzy versus neural network. *Jordan Journal of Mechanical and Industrial Engineering*, 3(4), 294–305.

Schimert, J. (2008). Data-driven fault detection based on process monitoring using dimension reduction techniques. In *Aerospace Conference*, 1–12. IEEE.

Schwabacher, M. and Goebel, K. (2007). A survey of artificial intelligence for prognostics. In *AAAI fall symposium*, 107–114.

Singh, S., Pinion, C., and Subramania, H.S. (2011). Data-driven framework for detecting anomalies in field failure data. In *Aerospace Conference*, 1–14. IEEE.

Wise, B.M. and Gallagher, N.B. (1996). The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control*, 6(6), 329–348.

Yu, J. (2011). Fault detection using principal components-based gaussian mixture model for semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, 24(3), 432–444.