

Fault Detection Based on Optimal Transport Theory

Bingsen Wang

Energy Department, Politecnico di Milano, Italy. E-mail: bingsen.wang@polimi.it

Piero Baraldi

Energy Department, Politecnico di Milano, Italy. E-mail: piero.baraldi@polimi.it

Xuefei Lu

Predictive Analytics, University of Edinburgh, UK. E-mail: xuefei.lu@ed.ac.uk

Enrico Zio

Energy Department, Politecnico di Milano, Italy.

MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France.

Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea.

Aramis Srl, Milano, Italy. E-mail: enrico.zio@polimi.it

Most of the existing methods for fault detection are residual-based, i.e., they reconstruct the expected values of the signals in normal condition by using large amounts of data collected in the past and require to formulate hypotheses on the distributions. Since in many industrial applications the available data do not cover all the possible operating conditions and data distributions are unknown, their performance can be unsatisfactory. In this work, we propose a data-driven fault detection method based on Optimal Transport (OT). The Wasserstein distance between the distribution of the signals measured under the current conditions and a baseline distribution derived from the signals measured under normal conditions is used as abnormality score, and the OT solution is computed using the Cumulative Distribution Transform (CDT). The proposed method is verified considering a real bearing dataset. The performance of the detection is evaluated in terms of missed and false alarm rates, and compared to that of other traditional fault detection methods.

Keywords: Fault detection, data-driven, optimal transport, Wasserstein distance, abnormality score, cumulative distribution transform.

1. Introduction

Fault detection methods are classified as model-based and data-driven (Schwabacher and Goebel (2007); Zio (2009)). Model-based methods use first principles and physical laws to describe the physical phenomenon of interest. For example, Sekhar (2004) builds a model of the behaviour of a rotor using the finite element method and successfully applies it to fault detection. However, the practical application of model-based methods is limited by the difficulty of developing accurate mathematical models of the behaviour of complex systems (Wan and Ni (2019)).

On the other hand, data-driven fault detection approaches rely on the use of normal-condition (healthy) data (Qin (2009)). In particular, residual-based methods reconstruct the expected values of the signals in normal conditions and use the residuals, i.e., the difference between the real measurements and the reconstructed signals, to identify the normal/abnormal status of the

plant. Examples of residual-based methods include Auto-Associative Kernel Regression (AAKR) (Yu et al. (2018); Di Maio et al. (2013); Baraldi et al. (2015)), Principal Component Analysis (PCA) (Li and Wen (2014)), One-class Support Vector Machine (OC-SVM) (Yan et al. (2017); Liu et al. (2017)), and Artificial Neural Networks (ANN) (Bangalore et al. (2017)). A well-fitted empirical model providing accurate signal reconstructions plays an essential role in the above procedure. However, its training may require a large amount of healthy data collected under various operating conditions (Odgaard et al. (2008)). Besides, different choices of the reconstruction model may yield different detection results (Yang et al. (2019)). Eventually, the detection of an abnormal condition is performed considering whether the obtained residuals exceed a threshold or by statistical tests. For example, Di Maio et al. (2013) use the Sequential Probability Ratio Test (SPRT) on the residuals obtained from an AAKR model;

Mujica et al. (2011) apply T^2 and Q -statistics of the PCA residuals to detect damages in structures; Huang et al. (2017) establish a statistical hypothesis model in the residual subspace of PCA transform, to detect and isolate sensor faults based on the Bayesian and the generalized likelihood ratio test (GLRT). Notice that, although these methods assume a certain distribution of the residuals, most distributions of real-world data may be a priori unknown or may not actually follow the assumed distributions (Seo (2006)).

Another challenge of fault detection is signal pre-processing (Jardine et al. (2006)). The aim is to extract the information useful for enabling the detection. Various pre-processing techniques, such as Fast Fourier Transform (FFT) (Wang et al. (2018)), Continuous Wavelet Transform (Widodo and Yang (2008); Wang et al. (2011); Kankar et al. (2011)), Mathematical Morphology (Raj and Murali (2012)), have been applied to raw signals before fault detection (Caesarendra and Tjahjowidodo (2017)). The quality of the selected features strongly impacts the detection result, but unfortunately there is no universal rule for choosing the optimal pre-processing algorithm.

Recently, transport-related methods have been employed in various applications (Kolouri et al. (2017)), such as signal and image processing (Li et al. (2013); Reich (2013)), computer vision (Rabin et al. (2014); Ferradans et al. (2013)), machine learning and statistics (Montavon et al. (2015); Gretton et al. (2012)). Commonly used optimal transport distances include Wasserstein distance (or Kantorovich distance) (Cuturi and Doucet (2014)) and Earth Mover's distance (EMD) (Pele and Werman (2009)). Wasserstein distance has been shown as a promising statistic for the nonparametric two-sample test (Ramdas et al. (2017)). Ng et al. (2015) study the bearing diagnosis problem using EMD combined with dynamical system reconstruction. Kammammettu and Li (2019) use a PCA scheme combined with the Kantorovich distance (KD) for fault detection in the process industry.

The objective of this work is to develop a fault detection method with the following characteristics: 1) able to directly deal with raw signals, avoiding signal reconstruction and feature extraction; 2) be distribution free, i.e., not requiring to formulate hypotheses on the distributions of the data.

The basic idea behind the proposed approach is to generate an abnormality score, which quantifies the dissimilarity between the probability distributions of the currently monitored and healthy data. To this aim, we consider the Wasserstein distance as abnormality score for fault detection and the Cumulative

Distribution Transform (CDT) (Park et al. (2018)) is used to find the univariate Optimal Transport (OT) solution.

The proposed method has been applied to a real bearing dataset and compared with two other fault detection methods of literature: a Z -test based method (Panda and Khilar (2012); Gelso et al. (2008)) and a PCA-based method for signal reconstruction, combined with the Q -statistic for residual analysis (Garc a-Alvarez (2009)). The detection performance of the developed method is evaluated in terms of missed and false alarm rates.

The rest of this paper is organized as follows. Section 2 introduces the basic theory of OT and Wasserstein distance. Section 3 presents the proposed method of fault detection. Section 4 shows the application of the proposed method to a bearing dataset. Finally, the conclusions of the work are presented in Section 5.

2. Optimal Transport Theory and Wasserstein Distance

In general, optimal transport aims at realizing the push-forward of one distribution onto another one, with minimal cost. One of its most promising applications is the measure of the difference between two samples of data collected from different sources (Courty et al. (2017)).

Let \mathcal{X}, \mathcal{Y} be two supports over which signals are defined and μ and ν two probability measures defined on \mathcal{X} and \mathcal{Y} , respectively. The probability density functions of μ and ν exist, and are indicated by I_0 and I_1 , i.e., $d\mu(x) = I_0(x)dx$ and $d\nu(x) = I_1(x)dx$. The aim of optimal transport is to find the map $f: \mathcal{X} \rightarrow \mathcal{Y}$ that pushes μ onto ν by minimizing the objective function:

$$M(\mu, \nu) = \inf_{f \in \Pi(\mu, \nu)} \int_{\mathcal{X}} c(x, f(x)) I_0(x) dx \quad (1)$$

where $f(\cdot)$ is the pushforward of measure μ , $\Pi(\mu, \nu)$ the set of all possible maps, $c(\cdot, \cdot)$ is the cost function, which can be regarded as the energy required to move a unit mass $I_0(x)dx$ from x to $f(x)$. For any measurable subset $A \subset \mathcal{Y}$, f is characterized by:

$$\int_{f^{-1}(A)} d\mu(x) = \int_A d\nu(x) \quad (2)$$

If the cost function of Eq. (1) is defined by:

$$c(x, f(x)) = |x - f(x)|^p \quad (p \geq 1) \quad (3)$$

the p -Wasserstein distance for probability measures μ and ν is obtained:

$$W_p(\mu, \nu) = \left(\inf_{f \in \Pi(\mu, \nu)} \int_{\mathcal{X}} |x - f(x)|^p I_0(x) dx \right)^{\frac{1}{p}} \quad (4)$$

In the case $p = 2$, the space of probability measures is endowed with a structure of a Riemannian manifold (Villani (2003)), in which the 2-Wasserstein distance W_2 is the shortest path connecting any two probability measures.

When two Gaussian probability measures $u \sim \mathcal{M}(\mu_1, \Sigma_1)$ and $v \sim \mathcal{N}(\mu_2, \Sigma_2)$, defined on \mathbb{R}^d with mean vectors μ_1 and μ_2 , and covariance matrices Σ_1 and Σ_2 , are considered, the 2-Wasserstein distance between u and v has the closed-form expression (Takatsu (2011)):

$$W_2(u, v) = \left\{ \|\mu_1 - \mu_2\|_2^2 + \text{Tr} \left(\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2})^{1/2} \right) \right\}^{1/2} \quad (5)$$

3. Fault Detection Method

We assume to have available a set of N^σ training data $\{x_i^\sigma\}_{i=1}^{N^\sigma}$ collected in normal conditions, where x_i^σ refers to the value of signal X^σ at time i . These data are used to estimate the reference probability distribution I_0 . To this aim, we consider time windows made by L consecutive measurements $w_0^{m_\sigma} = \{x_i^\sigma\}_{i=m_\sigma}^{m_\sigma+L}$ with $m_\sigma = 1, 2, \dots, N^\sigma - L$, and we estimate the signal probability density function (PDF) $I_0(x)$ in the time window, using the kernel density estimation method (Silverman (1986)). The objective of the fault detection is to detect whether an abnormality has occurred in a time window of L consecutive test data, $w_{test} = \{x_i^{test}\}_{i=1}^L$.

This is done by estimating the PDF of the test data, I_1 . According to Eq. (2), the mapping f

from the training data to the test data is uniquely defined. The Cumulative Distribution Transform (CDT) (Park et al. (2018)) of I_1 (denoted as $\hat{I}_1: \mathcal{X} \rightarrow \mathbb{R}$) with respect to I_0 is:

$$\hat{I}_1(x) := (f(x) - x) \sqrt{I_0(x)} \quad (6)$$

where $f(\cdot)$ is the mapping function which minimizes the objective function in Eq. (1).

For a discrete-time univariate signal X , the L^2 norm of \hat{I}_1 is:

$$\|\hat{I}_1\|_{L^2} = \left(\int_{\mathcal{X}} |\hat{I}_1(x)|^2 dx \right)^{1/2} = \left(\int_{\mathcal{X}} (f(x) - x)^2 I_0(x) dx \right)^{1/2} \quad (7)$$

Since the mapping f is uniquely defined for nonzero probability densities, $\|\hat{I}_1\|_{L^2}^2$ is equal to the minimal energy applied for the optimal transportation from I_0 to I_1 , when a quadratic cost $c(x, f(x)) = (f(x) - x)^2$ is used (Eq. (1)). Notice that $\|\hat{I}_1\|_{L^2}^2$ is equal to the square of the 2-Wasserstein distance, W_2^2 (Eq. (4)). Therefore, $\|\hat{I}_1\|_{L^2}^2$ represents a measure of dissimilarity between the distributions I_0 of the training data and I_1 of the test data.

According to this interpretation, we define the *abnormality score* of the test data $w_{test} = \{x_i^{test}\}_{i=1}^L$ as:

$$Q^{OT}\{w_{test}\} = \|\hat{I}_1\|_{L^2}^2 \quad (8)$$

Then, if Q^{OT} is greater than a certain threshold h , a fault is detected. The threshold h is set empirically using a fraction of the training data.

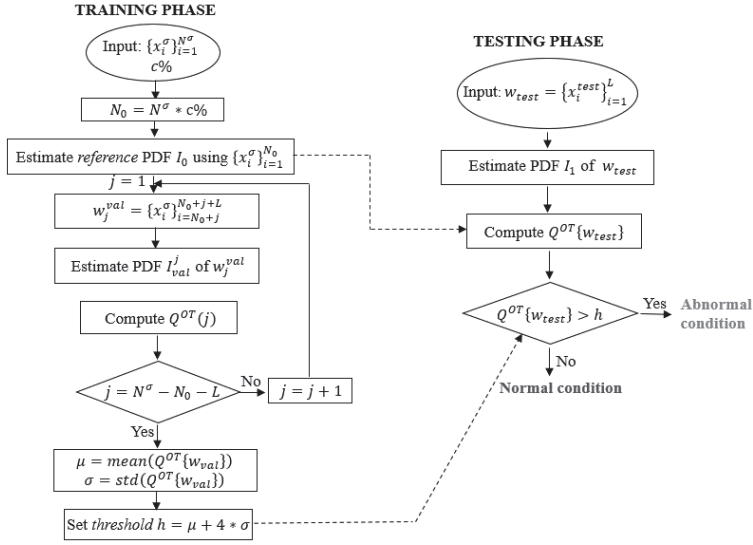


Fig. 1. The flowchart of the proposed fault detection method.

In practice, two sets of normal condition data are obtained from $\{x_i^\sigma\}_{i=1}^{N^\sigma}$. The first (training set) is used to estimate the reference PDF I_0 . A moving window of length L observations is applied to the second (validation set) and the PDF I_{val}^j of each data segment w_j^{val} , $j = 1, 2, \dots, N^\sigma - N_0 - L$ with N_0 indicating the $c\%$ of N^σ of the healthy data, is estimated. The abnormality score $Q^{OT}(j)$ associated to the j -th time window is computed by applying Eq. (7), and the threshold h is set as: $h = \mu + 4 * \sigma$, where μ and σ are the mean and standard deviation of $Q^{OT}(j)$, respectively. Fig. 1 shows the flowchart of the method.

4. Case Study

Bearings are one of the most widely used components in machines, and bearing faults are one of the most common reasons for machine breakdown (Ng et al. (2015)). In this Section, the proposed fault detection method is applied considering a bearing dataset provided by the Case Western Reserve University (CWRU) Bearing Data Center (<http://csegroups.case.edu/bearingdatacenter/home>). It contains test data collected from normal and faulty bearings.

The fault detection results are compared to those obtained by applying the Z-test (Panda and Khilar (2012); Gelso et al. (2008)), and a method which combines PCA for signal reconstruction

and the Q -statistic for residual analysis (Garca-Alvarez (2009)).

The fault detection performance is evaluated considering:

- the false alarm rate (FAR):

$$FAR = \frac{N_{normal}^M}{N_{normal}} \quad (9)$$

where N_{normal} is the total number of test windows containing patterns in normal condition and N_{normal}^M the number of normal condition patterns misclassified as abnormal condition.

- the missed alarm rate (MAR):

$$MAR = \frac{N_{abnormal}^M}{N_{abnormal}} \quad (10)$$

where $N_{abnormal}$ is the total number of test windows containing patterns in abnormal condition and $N_{abnormal}^M$ the number of abnormal condition patterns misclassified as normal condition.

- the Receiver Operating Characteristic (ROC) curve (Carter et al. (2016)), which reports FAR on the x -axes and $1-MAR$ on the y -axes, and is obtained by varying the thresholds h used for the detection. The Area Under the Curve (AUC) can be used to assess the performance of a fault detection model: larger the AUC , better the fault detection.

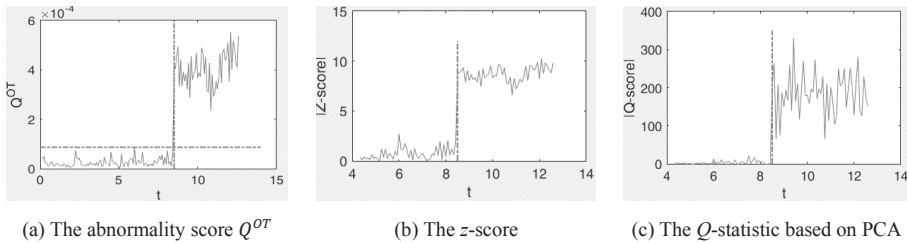


Fig. 3. Evolution of the abnormality score when the proposed method, the Z-test and the Q -statistic methods are tested with the fault of Fig. 2. The true onset of the abnormal condition is at time $t = 8.5s$.

The training set contains 5.2×10^4 healthy observations of the amplitude of the acceleration measured by a sensor at the drive end of the motor housing. The data was collected at a frequency of 12,000 observations per second. The first 2000 observations are used to estimate the reference PDF, I_0 , and the remaining observations to estimate the threshold h . Other 10^5 observations are used for testing the method performance. They include 50% of healthy data and 50% of abnormal condition data, obtained by artificially adding a bias or a noise to healthy data. A window of length $L = 1000$ observations has been used.

Fig. 2 shows an example of simulated abnormal condition, which has been obtained by adding a bias of magnitude $c = 0.02$, corresponding to a mild fault according to Zhang et al. (2019), to the signal evolution in normal condition from time $t = 8.5s$.

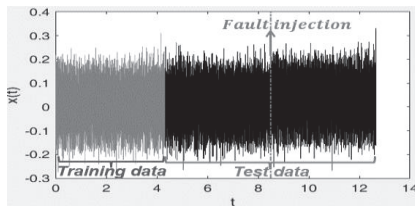


Fig. 2. Healthy bearing data corrupted with a bias $c=0.02$.

The time evolution of the abnormality score, Q^{OT} , is shown in Fig. 3, which also compares the results of the proposed method with that of the fault detection approaches based on Z-test and Q -statistic. Notice that all the abnormality scores

are able to promptly detect the occurrence of the abnormal condition.

The horizontal dash line in Fig. 3(a) indicates the threshold $h = \mu + 4 * \sigma$. With respect to the z-score (Fig. 3 (b)), we use the mean μ_0 and the population variance σ^2 estimated from the training data of Fig. 2. The statistical features considered by the PCA model, which have been selected by trial-and-error, are mean, maximum, 3rd moment, mean frequency, frequency Kurtosis and max power.

Fig. 4 shows another example of simulated abnormal condition, obtained by adding a noise of magnitude $c * e$, where $e \sim N(0,1)$, $c = 0.05$, to the original signal from time $t = 8.5s$.

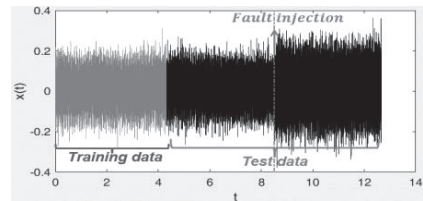


Fig. 4. Healthy bearing data corrupted with a noise $c * e$, $e \sim N(0,1)$, $c = 0.05$.

The time evolution of the abnormality score, Q^{OT} , is shown in Fig. 5, which also compares the results of the proposed method with those of the proposed approaches based on Z-test and Q -statistic. Notice that only the abnormality score Q^{OT} and the Q -score are able to promptly detect the occurrence of the abnormal condition.

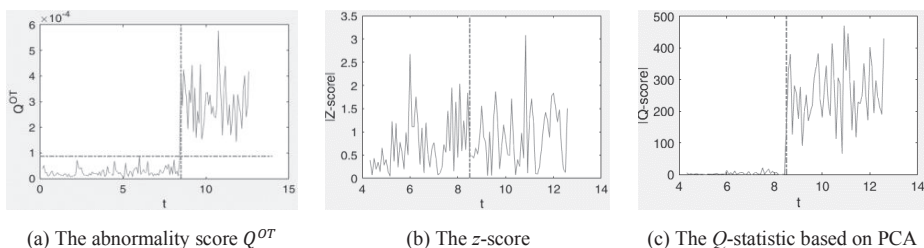


Fig. 5. Evolution of the abnormality score when the proposed method, the Z-test and the Q -statistic methods are tested with the fault of Fig. 4. True onset of the abnormal condition is at time $t = 8.5s$.

Considering the performance of the three methods on the two types of simulated abnormal conditions, it is possible to conclude that: 1) the proposed method is able to detect different types of abnormal conditions (e.g. bias and noise); 2) the proposed method can identify the time of the first fault injection, without delay; 3) the Z -test can easily recognize the change of the mean of the observations, but fails to detect the change of the variance; 4) the performances of the proposed method and the Q -statistic are similar; however, the latter requires a preliminary step of feature selection.

If we assume a Gaussian distribution of the data, the 2-Wasserstein distance can be computed using Eq. (5). Fig. 6 shows the comparison of the time evolution of the abnormality scores Q^{OT} obtained by the analytical solution and by CDT. Notice that when we consider the two abnormal conditions of bias (Fig. 2) and noise (Fig. 4), CDT (Fig. 6 (b), (d)) provides almost the same results of the closed-form solution (Fig. 6 (a), (c)), without any assumption on the distribution of the measured signal.

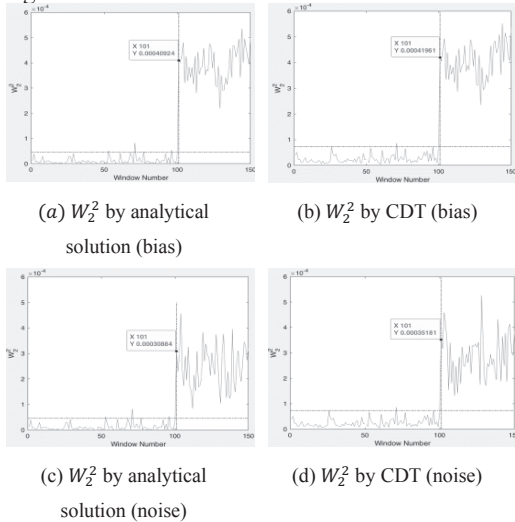


Fig. 6. Evolution of W_2^2 obtained by the analytical solution and by CDT, with the faults of Fig. 2 (top) and Fig. 4 (bottom).

The influence of the window length of observation, L , on the performance of the proposed method is now investigated. Fig. 7 shows the evolution of the abnormality score when L is set equal to 2000, 1000, 500 and 200 observations, respectively.

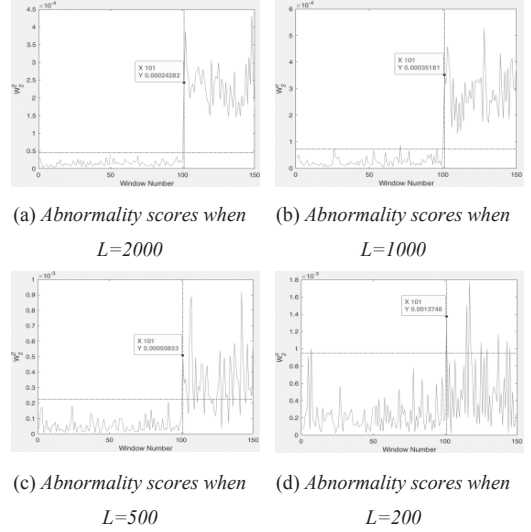


Fig. 7. Evolution of the abnormality score when the proposed method is tested with the fault of Fig. 4 and various observation window lengths.

Table 1. FAR , MAR and AUC of the detection results with various observation window lengths.

Window length L	FAR	MAR	AUC
2000	0	0	1.0
1000	2%	0	1.0
500	0	36%	0.977
200	0	86%	0.737

FAR , MAR and AUC of the detection results with different observation window length are reported in Table 1. There is a trade-off between the window length and time-to-detection: although larger window lengths require to collect more observations for estimating I_j and $Q^{OT}(j)$, and therefore cause delayed detections, they allow obtaining better performance in terms of FAR , MAR and AUC .

5. Conclusions

This work has addressed the problem of fault detection in industrial systems. A novel method based on optimal transport theory has been proposed. The abnormality score has been built using the Wasserstein distance. It differs from state-of-the-art methods for fault detection since it does not require: 1) the use of signal reconstruction methods; 2) to make hypotheses on the signal distributions. The method has been verified considering the detection of abnormal

conditions in bearings. Future work will include the verification of the proposed method with real abnormal conditions and its extension to the case of monitoring a system with multiple sensors.

Acknowledgement

Bingsen Wang gratefully acknowledges the financial support from the China Scholarship Council (No. 201801810028). The participation of Piero Baraldi and Enrico Zio has been funded by “Smart maintenance of industrial plants and civil structures by 4.0 monitoring technologies and prognostic approaches – mac4pro”, sponsored by the call BRIC-2018 of the National Institute for Insurance against Accidents at Work – INAIL.

References

- Baraldi, P., Di Maio, F., Turati, P., and Zio, E. (2015). Robust signal reconstruction for condition monitoring of industrial components via a modified Auto Associative Kernel Regression method. *Mechanical Systems and Signal Processing*, 60, 29-44.
- Bangalore, P., Letzgus, S., Karlsson, D., and Patriksson, M. (2017). An artificial neural network-based condition monitoring method for wind turbines, with application to the monitoring of the gearbox. *Wind Energy*, 20(8), 1421-1438.
- Caesarendra, W., and Tjahjowidodo, T. (2017). A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Machines*, 5(4), 21.
- Courty, N., Flamary, R., Habrard, A., and Rakotomamonjy, A. (2017). Joint distribution optimal transportation for domain adaptation. In *Advances in Neural Information Processing Systems*, 3730-3739.
- Carter, J. V., Pan, J., Rai, S. N., and Galandiuk, S. (2016). ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery*, 159(6), 1638-1645.
- Cuturi, M., and Doucet, A. (2014). Fast computation of Wasserstein barycenters. In *Proceedings of the 31st International Conference on Machine Learning*.
- Di Maio, F., Baraldi, P., Zio, E., and Seraoui, R. (2013). Fault detection in nuclear power plants components by a combination of statistical methods. *IEEE Transactions on Reliability*, 62(4), 833-845.
- Ferradans, S., Xia, G. S., Peyré, G., and Aujol, J. F. (2013, June). Static and dynamic texture mixing using optimal transport. In *International Conference on Scale Space and Variational Methods in Computer Vision*, 137-148.
- Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1), 723-773.
- Gelso, E. R., Biswas, G., Castillo, S., and Armengol, J. (2008, September). A comparison of two methods for fault detection: a statistical decision, and an interval-based approach. In *19th international workshop on principles of diagnosis dx*, 261-268.
- García-Alvarez, D. (2009, January). Fault detection using principal component analysis (PCA) in a wastewater treatment plant (WWTP). In *Proceedings of the International Student's Scientific Conference*.
- Huang, H. B., Yi, T. H., and Li, H. N. (2017). Sensor fault diagnosis for structural health monitoring based on statistical hypothesis test and missing variable approach. *Journal of Aerospace Engineering*, 30(2), B4015003.
- Jardine, A. K., Lin, D., and Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7), 1483-1510.
- Kankar, P. K., Sharma, S. C., and Harsha, S. P. (2011). Fault diagnosis of ball bearings using continuous wavelet transform. *Applied Soft Computing*, 11(2), 2300-2312.
- Kolouri, S., Park, S. R., Thorpe, M., Slepcev, D., and Rohde, G. K. (2017). Optimal mass transport: Signal processing and machine-learning applications. *IEEE signal processing magazine*, 34(4), 43-59.
- Kammammettu, S., and Li, Z. (2019). Change point and fault detection using Kantorovich distance. *Journal of Process Control*, 80, 41-59.
- Li, S., and Wen, J. (2014). A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform. *Energy and Buildings*, 68, 63-71.
- Liu, J., Li, Y. F., and Zio, E. (2017). A SVM framework for fault detection of the braking system in a high speed train. *Mechanical Systems and Signal Processing*, 87, 401-409.
- Li, P., Wang, Q., and Zhang, L. (2013). A novel earth mover's distance methodology for image matching with gaussian mixture models. In *Proceedings of the IEEE International Conference on Computer Vision*, 1689-1696.
- Mujica, L. E., Rodellar, J., Fernandez, A., and Güemes, A. (2011). Q-statistic and T2-statistic PCA-based measures for damage assessment in structures. *Structural Health Monitoring*, 10(5), 539-553.
- Montavon, G., Müller, K. R., and Cuturi, M. (2015). Wasserstein training of Boltzmann machines. *arXiv preprint arXiv:1507.01972*.
- Ng, S. S., Cabrera, J., Peter, W. T., Chen, A. H., and Tsui, K. L. (2015). Distance-based analysis of dynamical systems reconstructed from vibrations for bearing diagnostics. *Nonlinear Dynamics*, 80(1-2), 147-165.

- Odgaard, P. F., Lin, B., and Jorgensen, S. B. (2008). Observer and data-driven-model-based fault detection in power plant coal mills. *IEEE Transactions on Energy Conversion*, 23(2), 659-668.
- Park, S. R., Kolouri, S., Kundu, S., and Rohde, G. K. (2018). The cumulative distribution transform and linear pattern classification. *Applied and Computational Harmonic Analysis*, 45(3), 616-641.
- Panda, M., and Khilar, P. M. (2012). Distributed soft fault detection algorithm in wireless sensor networks using statistical test. In *2012 2nd IEEE International Conference on Parallel, Distributed and Grid Computing*, 195-198.
- Pele, O., and Werman, M. (2009). Fast and robust earth mover's distances. In *2009 IEEE 12th International Conference on Computer Vision*, 460-467.
- Qin, S. J. (2009). Data-driven fault detection and diagnosis for complex industrial processes. *IFAC Proceedings Volumes*, 42(8), 1115-1125.
- Raj, A. S., and Murali, N. (2012). Early classification of bearing faults using morphological operators and fuzzy inference. *IEEE Transactions on Industrial Electronics*, 60(2), 567-574.
- Reich, S. (2013). A nonparametric ensemble transform method for Bayesian inference. *SIAM Journal on Scientific Computing*, 35(4), A2013-A2024.
- Rabin, J., Ferradans, S., and Papadakis, N. (2014, October). Adaptive color transfer with relaxed optimal transport. In *2014 IEEE International Conference on Image Processing (ICIP)*, 4852-4856.
- Ramdas, A., Trillos, N. G., and Cuturi, M. (2017). On wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19(2), 47.
- Schwabacher, M. and Goebel, K. (2007, November). A Survey of Artificial Intelligence for Prognostics. In *AAAI Fall Symposium: Artificial Intelligence for Prognostics*, 108-115.
- Sekhar, A. S. (2004). Model-based identification of two cracks in a rotor system. *Mechanical Systems and Signal Processing*, 18(4), 977-983.
- Seo, S. (2006). A review and comparison of methods for detecting outliers in univariate data sets (Doctoral dissertation, University of Pittsburgh).
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC press.
- Sheriff, M. Z., Mansouri, M., Karim, M. N., Nounou, H., and Nounou, M. (2017). Fault detection using multiscale PCA-based moving window GLRT. *Journal of Process Control*, 54, 47-64.
- Takatsu, A. (2011). Wasserstein geometry of Gaussian measures. *Osaka Journal of Mathematics*, 48(4), 1005-1026.
- Villani, C. (2003). *Topics in optimal transportation* (No. 58). American Mathematical Soc..
- Wan, H. P., and Ni, Y. Q. (2019). Bayesian multi-task learning methodology for reconstruction of structural health monitoring data. *Structural Health Monitoring*, 18(4), 1282-1309.
- Wang, B., Tang, D., Yue, Q., Zhou, J., and Deonauth, N. (2018). Study on nonlinear dynamic characteristics inherent in offshore jacket platform using long-term monitored response of ice-structure interaction. *Applied Ocean Research*, 72, 23-38.
- Widodo, A., and Yang, B. S. (2008). Wavelet support vector machine for induction machine fault diagnosis based on transient current signal. *Expert Systems with Applications*, 35(1-2), 307-316.
- Wang, S., Huang, W., and Zhu, Z. K. (2011). Transient modeling and parameter identification based on wavelet and correlation filtering for rotating machine fault diagnosis. *Mechanical systems and signal processing*, 25(4), 1299-1320.
- Yu, J., Yoo, J., Jang, J., Park, J. H., and Kim, S. (2018). A novel hybrid of auto-associative kernel regression and dynamic independent component analysis for fault detection in nonlinear multimode processes. *Journal of Process Control*, 68, 129-144.
- Yan, K., Ji, Z., and Shen, W. (2017). Online fault detection methods for chillers combining extended kalman filter and recursive one-class SVM. *Neurocomputing*, 228, 205-212.
- Yang, C., Liu, J., Zeng, Y., and Xie, G. (2019). Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model. *Renewable Energy*, 133, 433-441.
- Zio, E. (2009). *Computational methods for reliability and risk analysis* (Vol. 14). World Scientific Publishing Company.