Doctoral School on Engineering Sciences
Università Politecnica delle Marche

Extended summary

Modeling and Diagnosis of Complex Systems Dynamics by Data-Driven Approaches

Curriculum: Ingegneria Informatica, Gestionale e dell'Automazione

Author

Francesco Ferracuti

Tutor(s)

Prof. Giuseppe Orlando

Prof. Sauro Longhi

Date: 30-01-2014

Doctoral School on Engineering Sciences
Università Politecnica delle Marche

Abstract. Complex systems are found in almost all field of contemporary science, and are associated with a wide variety of financial, physical, biological, information and social systems. Although complex systems have many different properties, the most important are: dimensionality, uncertainty, nonlinearity and coupling between components. The procedures to obtain analytical models of complex systems cannot be implemented easily because the properties, which arise from these systems, make difficult the modeling. Since a mathematical model is a description of system behaviour, accurate modeling for a complex system is very difficult to achieve in practice. Furthermore, sometimes, it may even be impossible to describe the system by analytical equations. The present dissertation tries to address two issues regarding the modeling and diagnosis of complex systems. The first one deals with the issue of modeling a complex system, in the case the analytical model is not obtainable. The second one deals with the issue of diagnosing the system behaviour. Modeling and diagnosis of complex systems is addressed developing data-driven procedures, which are able to learn the complex system dynamics from data that are provided by installed sensors. A contribution to complex system dynamics modeling and diagnosis is given with particular attention to real systems, for this different applications are discussed. The first application deals with the issue of modeling and diagnose the defects and faults in a Quality Control scenario for electric motors. The second application deals with the issue of modeling and diagnose a complex industrial system as a paper mill plant. The third application deals with the issue of estimating the residual useful life of a turbofan engine and the last deals with the issue of modeling the Electroencephalography signals by data-driven algorithms in order to diagnose the user intentions in a Brain Computer Interface context.

Keywords. Fault Diagnosis, Electric Motors, Data-Driven, Brain Computer Interface, Complex Systems.

1 Problem statement and objectives

Complex systems consist of a large number of nonlinearly interacting components, often called agents, which displays collective behaviour that does not follow trivially from the behaviours of the individual parts. These systems are open, they interchange information with environment and constantly modify their internal activity structure and patterns in the self-organization process [1]. In many contexts such as industrial, natural and financial the development of a mathematical model of the process is a difficult task due to the complexity of the processes that are involved. Modeling may demand considerable engineering efforts and thus becomes not practical for complex processes. Modeling complex systems with physical models could be very difficult or not possible, on the other hand mathematical models that describe the complex systems dynamics could not be easily manipulated. Complex systems, of which analytical models are not available, could be modeled by data-driven approaches. Data-driven approaches, simply speaking, are based on data matrices, which usually contain measurements of physical process variables, and computational intelligence and machine learning algorithms [2]. In this thesis the modeling and diagnosis problems of complex systems dynamics are addressed by data-driven procedures for different case studies described in the following.

1.1 Modeling and Diagnosis of Electric Motors in a Quality Control Scenario

Electric motors are well known dynamical systems with accurate analytical models and extensive results in literature. Although these models, which are mainly nonlinear, describe accurately the dynamics, not all intrinsic dynamics are described by these models. Unmodeled dynamics can be vibrations, thermal drifts and faults. Mechanical faults such as misalignment, broken bars, gear and bearing defects are not simple to model by analytical way for this the modeling and the diagnosis, with Fault Detection and Isolation applications, is dealt with data-driven approaches. Data-Driven modeling and diagnosis solutions are proposed in order to model and diagnose mechanical faults, which could not be diagnosed by analytical models. Fault Detection and Diagnosis (FDD) algorithms could be used not only with the fault diagnosis purpose but also to improve the Quality Control (QC) of these machines. Two data-driven FDD algorithms based on Motor Current Signature Analysis (MCSA) and vibration signals in a QC scenario are proposed [3, 4, 5].

1.2 Modeling of Complex Systems with FDI and Prognosis Applications

A paper mill plant is a complex system, which consists of several coupling nonlinear systems (e.g. paper machine, stock preparation) and subprocesses (e.g, fun pump, tine unit, mixing unit, pope reel, pulping). Paper mill plant is often monitored by a lot of sensors, which measures many different variables that can be discrete, logic and acquired at different sampling frequency. One of the problems in complex industrial systems as a paper mill plant concerns the high number of monitored variables. This is a classic example of dimensionality problem in complex system. The Fault Detection and Isolation issue for a paper mill plant is faced by the Multi-Scale Principal Component Analysis (MSPCA) procedure [6, 7]. The second issue is related to a prognostic management of a turbofan engine: a data-driven algorithm is developed in order to estimate the residual useful life (RUL). A common turbofan consists of coupling nonlinear systems such as a compressor, combustor, and a turbine which drives the compressor. Turbofan engines is a complex system that requires adequate monitoring to ensure flight safety and timely maintenance [8].

1.3 Modeling and diagnosis of EEG signals in BCI Applications

The brain is assumed to be a classical example of a complex, self-organized system. As such, it exhibits hallmarks of nonlinearity, multistability, and non-diffusivity. These oscillations are produced by large ensembles of synchronized neuronal activity and the resulting electrophysiological signals in the different frequency bands are associated with different functional states (e.g. sleep, wake, perception and attention). Brain-Computer Interfaces (BCIs) are devices which translate the brain activity of the user into specific signals, which may be used for communicating or controlling external devices [9, 10] without the use of peripheral nerves and muscles [11]. BCIs represent an interesting option to people affected by neuromuscolar disorders, but whose brain activity is normal, such as in patients affected by Amyotrophic Lateral Sclerosis. In this context Electroencephalography (EEG) signals are monitored and modeled by data-driven algorithms in order to diagnose when the user focuses the attention and when he doesn't on auditory stimuli as common spoken words [12].

2 Research planning and activities

2.1 Data-Driven Modeling and Diagnosis of Induction Motor by Current Signals and Vibration Signals

In order to model the three-phase stator current signals, Principal Component Analysis (PCA [6]) and Kernel Density Estimation (KDE) are taken into account. PCA is used in data pre-processing to reduce the currents space in two dimensions. The Probability Density Function (PDF) of PCA-transformed signals is estimated by KDE, which is a non-parametric method useful to assess the data distribution [13]. PDFs are the models that can be used to identify each fault and defect. In the test, KDE with Gaussian kernel function is considered and the plug-in bandwidth selection procedure is applied [14]. Diagnosis has been carried out using the Kullback-Leibler (K-L) divergence [15], which measures the difference between two probability distributions. This divergence is used as a distance measure between classified statistic signatures obtained by KDE. By K-L divergence, the classification of each motor condition is performed. In the context of Machine Vibration Signature Analysis (MVSA), MSPCA is taken into account in order to model the vibration signals [6]. MSPCA transforms the process data information at different scales by Wavelet Transform (WT [16]). The information of each different scales is captured by PCA modeling. These models, which represent the process conditions, can be used to identify each fault and defect. In order to isolate the defects a KDE algorithm is used on the PCA residuals, and thresholds are computed for each sensor signal to determine if, for each wavelet matrix, the signals are involved in the defect or not. KDE method is widely recognized as a robust methodology to determine numerically the data PDF, in particular such estimation technique is introduced where Gaussian assumption is not recognized [17]. Fault diagnosis can be performed using the contribution plots because they represent the signatures of the rotating electrical machines conditions. In QC context, a supervised classificator, with input the PCA contributions, is used to diagnose each motor defect. The results show that the identified signatures by PCA contributions, are unique for each considered defect.

2.2 Data-Driven Modeling and Diagnosis of a Paper Mill Plant: FDI Application

MSPCA [6] is used to monitor some critical variables of the stock preparation of a paper mill plant in order to diagnose faults and malfunctions. MSPCA simultaneously extracts both, cross correlation across the sensors (PCA approach) and auto-correlation within a sensor (Wavelet approach). The advantage of MSPCA is validated on considered paper mill plant where several sensors are installed to control and monitor the automation system. Availability of many sensors provides valuable redundancy for fault detection and identification because sensors measurements are highly correlated under normal conditions [18]. Detection and diagnosis of stock preparation process of a paper mill plant is considered. Recursive MSPCA is applied for on-line fault detection and diagnosis: once a fault is detected a multi-scale fault identification is performed.

2.3 Data-Driven Modeling and Diagnosis of a Turbofan Engine: Prognosis Application

Residual life time of systems is a determinant factor for machinery and environment safety. The issue of estimate the RUL of turbofan engines is addressed [19]. A Hidden Markov Model (HMM [20]) is used to estimate the RUL of a turbofan engine, features are extracted by an Artificial Neural Network that is trained to identify faultless parameter of the turbofan in different flight conditions. Residuals are obtained at the end of each flight and a set of indexes are generated. The HMM uses these indexes and computes RUL estimation, as the number of remaining flights. These models give estimations on residual life and health-state by modeling observations (inputs) as probability density functions. Thus it is possible to define a model composed by a set of states that are described by PDF. This allows to use Bayesian inference algorithms for estimating the health conditions. Data are generated by the model simulator C-MAPSS (Commercial Modular Aero-Propulsion System Simulation) and are used in the Prognostics and Health Management competition (PHM'08 [21]).

2.4 Data-Driven Modeling of EEG Signals: BCI Application

An auditory BCI paradigm is proposed for systems based on P300 signals which are generated by auditory stimuli characterized by different sound typologies and locations. This paradigm is able to improve the classification scores with respect to the classical auditory BCI systems. A Head Related Transfer Function approach is chosen to virtualize auditory stimuli. When virtualized audio is used, the user has to focus the attention both on the type and location of the stimulus, thus generating P300 signals whose amplitude is higher than that generated without audio virtualization. The auditory algorithm processes the EEG signals in order to model, by data-driven algorithms, patterns that describe the user intentions, in particular if the user focuses the attention on auditory stimuli and when he doesn't. These patterns can be used as typical features in order to diagnoses the user intentions. Supervised classification is performed by Support Vector Machines in order to diagnose the user attention on a specific stimulus. The system has been validated with 14 users, who were asked to choose one among five common spoken words, previously virtualized and transmitted to stereophonic headphones. Classification results prove that the proposed auditory BCI system performed similarly to common visual BCI P300 systems, representing then an alternative to visual BCI for users with visual impairments [12].

3 Analysis and discussion of main results

3.1 Electric motors

The classification accuracy is considered as an index to evaluate the performance of the proposed algorithm. The simulation results, given in Table 1 show the faults diagnosis for broken rotor bars and connectors, while Table 2 shows the results related to real induction motors. In this last case, two different faults are tested: wrong rotor and cracked rotor. Wrong rotor refers to a non-compliant rotor, in particular a single phase rotor is assembled instead of a three phase rotor. In the first case study, by Monte Carlo simulations, all fault types are diagnosed with 100% accuracy with acquisition time above 0.4 s for each fault type, while below 0.4 s, the classification accuracy decreases as shown in Table 1.

Table 1. Classification accuracy in the case of finite element motor, changing n_{grid}, the points number in which the PDF is estimated, and the current signals acquisition time in steady-state. Label H means healthy motor, labels 1-10B mean broken bars with the relative number, labels 1-10C mean broken connectors with relative number.

n_{grid}	128 × 128		64 × 64		32 × 32	
Acquisition time (s)	0.3	0.15	0.3	0.15	0.3	0.15
	%					
H	100	100	100	100	100	100
1B	100	100	100	100	100	100
2B	100	100	100	100	100	100
3B	100	100	100	100	100	99.93
4B	100	100	100	100	100	99.89
5B	100	100	100	99.96	100	99.70
6B	100	100	100	100	100	99.87
7B	100	99.98	100	99.94	100	98.19
8B	99.82	95.34	99.89	95.97	99.55	91.41
9B	100	99.74	100	99.53	100	98.56
10B	100	99.43	100	99.32	99.99	99.43
1C	100	100	100	100	100	99.99
2C	99.98	96.89	99.88	95.99	99.56	91.01
3C	99.88	91.71	99.72	95.74	98.48	93.75
4C	99.79	97.61	99.84	98.10	99.93	96.87
5C	99.98	98.96	99.99	98.49	99.96	96.37
6C	100	100	100	100	100	99.98
7C	100	100	100	100	100	99.94
8C	100	100	100	100	100	99.88
9C	100	99.99	100	100	100	99.47
10C	100	99.89	100	99.77	100	96.95
Mean	99.97	99.03	99.97	99.18	99.88	98.15

In the second case study, by Monte Carlo simulations, all fault types are diagnosed with accuracy shown in Table 2. It can be noticed how the classification accuracy in the case of healthy motor is always 100%, therefore the algorithm is able to detect if the motors are healthy or if there are some faults or defects. In these tests, the algorithm never confuses the cases of healthy motors from those not healthy.

Table 2. Classification accuracy in the case of real motors, changing n_{grid}, the points number in which the PDF is estimated, and the current signals acquisition time in steady-state. Motor conditions are: healthy, motor with cracked rotor and motor with wrong rotor.

n_{grid}	128 × 128			64 × 64			32 × 32		
Acquisition time (s)	0.7	0.5	0.3	0.7	0.5	0.3	0.7	0.5	0.3
	%								
Healthy	100	100	100	100	100	100	100	100	100
Cracked rotor	98.82	95.08	77.08	99.00	94.74	86.54	98.29	94.02	81.45
Wrong Rotor	98.97	99.47	99.49	99.18	99.36	98.56	99.85	99.41	99.21
Mean	99.26	98.18	92.19	99.39	98.03	95.03	99.38	97.81	93.55

The following experiments show the application of the algorithm using the data acquired from accelerometer mounted on the kitchen hoods. In this case study two accelerometer sensor are placed by robotic arm on the top of the kitchen hoods, when in the production line the kitchen hoods reach the quality control bench. Motor rotation frequency is 50 Hz and sampling frequency is 20 kHz. Figures 1-2 show the contribution plot at approximation matrix A_7 of each sensor. In the production line, some kitchen hoods could be assembled with unbalanced impeller, this defect introduces higher energy consumption and more noise than faultless kitchen hoods. These figures show that the contributions for the first 11 kitchen hoods, which are faultless, are lower than the last 8, which are assembled with defective motors.

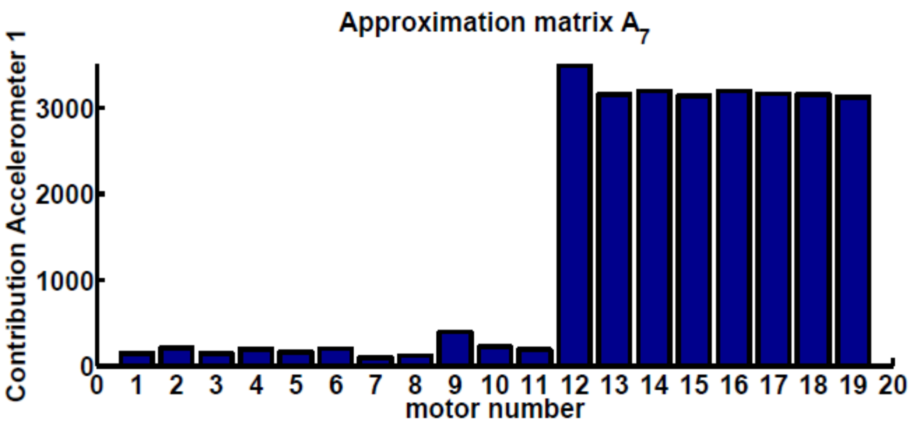

Figure 1. Contribution weights of A_7 scale matrix for the first accelerometer in the case of unbalanced impeller.

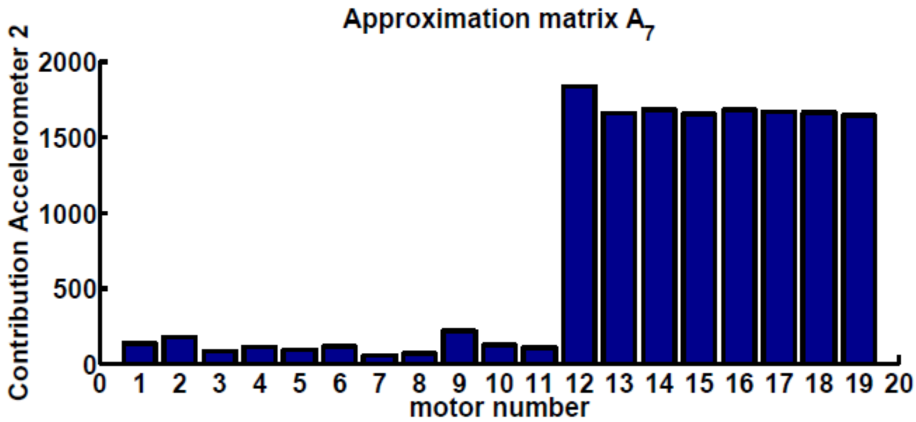

Figure 2. Contribution weights of A_7 scale matrix for the second accelerometer in the case of unbalanced impeller.

3.2 Paper Mill Plant

The experiment concerns with a real fault occurred in the stock preparation subprocess of paper mill during normal operation. The fault occurrence is identified on sensors 5 and 13, which measure the inlet pressure to the broke handling machine and the current of the motor broke deflacking. From sample 600 the motor is idling and is regarded as a soft fault, which is not considered as a motor fault but due to a malfunction of the machine as a jam. After such an interval it takes an abrupt fault identified in both sensors. As shown in Figure 3 and Table 3, the fault is detected by the SPE approximation matrix A_3 in the window 74 with a delay of 88 samples. Also the SPE of reconstructed PCA detected three faults at samples 3979, 4180 and 4481, that correspond to peaks in the sensor 5. The fault is identified in the sensors 5 and 13. It is possible to define the signature of the isolated fault by the SPE contribution at each level and identify its nature. The signature associated with the fault is unique and no other faults generated a similar signature in contribution plots.

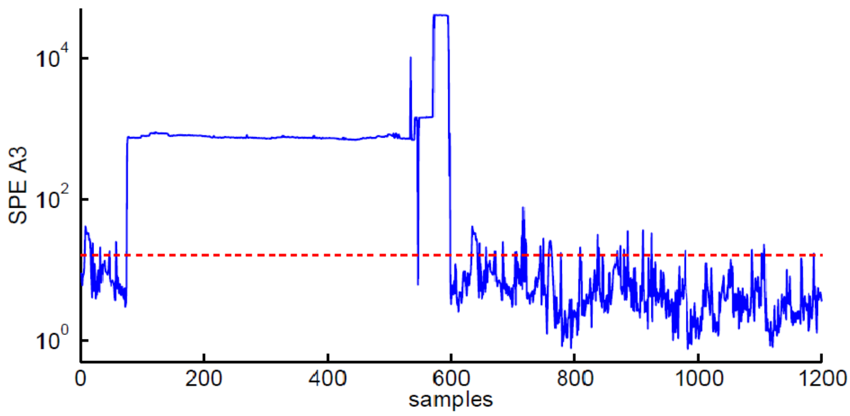

Figure 3. Real fault: SPE of approximation matrix A_3 in logarithmic scale.

Table 3. Fault diagnosis with real fault.

	A3	D1	D2	D3	Reconst. PCA
Fault	✓	✗	✗	✗	✓
	1218%	124%	120%	156%	345%
Sample	592-688				3979-4100 4180-4361 4481-4770

3.3 Turbofan Engine

The prognosis of the turbofan affected by the HPC fault is shown in Figure 4. The RUL converges to the real RUL after the fault occurrence. These simulations highlight important considerations, first the initial error of the RUL is derived from the different working point of the new data from the trained model. Second once the fault occurs, the RUL suddenly falls down, it shows the algorithm robustness and its ability to correct the predicted RUL in presence of faults. The simulation show that the developed procedure is able to track the true RUL even if perturbations occur.

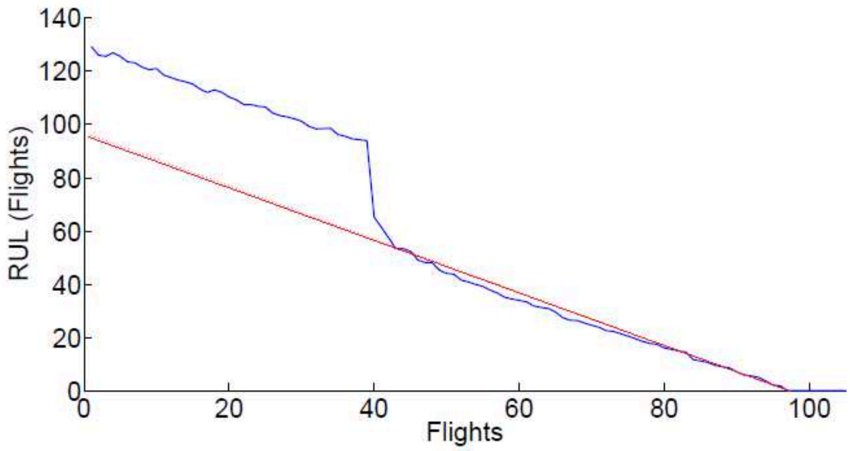

Figure 4. Turbofan estimated RUL in presence of LPT fault, solid blue line is the estimated RUL, dashed red line is true RUL.

3.4 Brain Computer Interface

Table 4 gives the classification, target and non-target accuracy for the BCI experiment when the SVM is required to perform a classification within a single run. In this case only one subject reaches 70% of target accuracy, while the remaining subjects scored a target accuracy below the 70% limit, which is assumed to be the minimal limit for useful BCI operations [22]. Please note that target accuracy being lower than non-target accuracy is considered normal: whenever the ratio between target and non-target words is small, the classificator tends to weight non-target words more than target ones.

Table 4. Classification accuracy, target accuracy and non-target accuracy for auditory stimuli (stimulus duration 1500 *ms*, ISI 250 *ms*) within a single run. Peak amplitude for the auditory condition is determined as the maximum amplitude in the range from 0 to 800 *ms*.

Participant	Classification accuracy (%)	Target accuracy (%)	Non-target accuracy (%)
1	76.8	42.0	85.5
2	78.8	44.0	86
3	84.0	51.6	90.7
4	82.9	48.4	90.0
5	84.0	52.0	90.4
6	78.0	33.9	86.8
7	86.2	58.1	92.0
8	82.5	47.3	89.5
9	78.9	36.7	87.3
10	87.3	61.3	92.7
11	82.5	47.3	89.5
12	80.2	40.6	88.1
13	90.6	71.0	94.7
14	85.1	54.8	91.3
Mean	82.5	49.1	89.6
SD	3.9	9.7	2.6

4 Conclusions

The first contribution of the thesis is the development of two data-driven diagnostic modules which can be applied to detect faults and defects, which cannot be described by analytical equations, of electric motors. The results show that the proposed data-driven diagnosis procedure is able to detect and diagnose different induction motor faults and defects. A possible future work is the extension of the algorithm to on-line FDD procedure in order to avoid one of the major drawback of the algorithm which concerns the data batch processing because it needs to acquire several current samples for the fault diagnosis procedure. The second one is based on vibration signals. Experiments on single-phase motors prove that the identified signature is unique for each defect. A possible future work is the improvement of fault diagnosis using the Wavelet Packet Transform (WPT), which is an extension of the classical wavelet analysis applied in MSPCA, and a filter that chooses the details and approximation scale matrices obtained by the WPT in a way to maximize the separation of the classes related to the motor conditions. A possible solution is to consider as filter the Common Spatial Pattern (CSP).

The second contribution is the development of two solutions for modeling and diagnose two different complex systems: a paper mill plant and a turbofan engine. These solutions are applied in order to monitor these complex systems in FDI and prognostic contexts. Possible future works for complex dynamics modeling should be mainly focused on the improvement of FDI and prognosis solutions in order to obtain models more robust to changes of operating conditions. Furthermore, since complex systems have a nonlinear behavior, seems natural to extend these algorithms by kernel models.

The last contribution is the development of an auditory BCI paradigm for systems based on P300 signals which are generated by auditory stimuli characterized by different sound typologies and locations. The auditory BCI paradigm diagnoses the user intention by auditory stimuli as common spoken words. In order to achieve this objective, EEG are modeled by data-driven approaches since it would not have been possible through analytical models. A possible future work could be the development of a hybrid modeling solution for EEG signals based on data-driven approaches and Kuramoto model [23], since this model is able to reproduce synchronization phenomena, which could be linked to P300 signal as recent findings indicate.

References

[1] A. Zecevic and D. Siljak, *Control of Complex Systems: Structural Constraints and Uncertainty*, ser. Communications and Control Engineering. Springer, 2010.
[2] C. Aldrich and L. Auret, *Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods*, ser. Advances in Computer Vision and Pattern Recognition. Springer, 2013.
[3] F. Ferracuti, A. Giantomassi, S. Iarlori, G. Ippoliti, and S. Longhi, *Induction motor fault detection and diagnosis using KDE and Kullback-Leibler divergence*, in Industrial Electronics Society, IECON 2013 - 39th Annual Conference of the IEEE, 2013, pp. 2923–2928.
[4] F. Ferracuti, A. Giantomassi, and S. Longhi, *MSPCA with KDE thresholding to support QC in electrical motors production line*, in Manufacturing Modelling, Management, and Control, vol. 7, no. 1, 2013, pp. 1542–1547.
[5] F. Ferracuti, A. Giantomassi, G. Ippoliti, and S. Longhi, *Multi-Scale PCA based fault diagnosis for rotating electrical machines*, in European Workshop on Advanced Control and Diagnosis, 8th ACD, Ferrara, Italy, 2010, pp. 296 – 301.
[6] B. R. Bakshi, *Multiscale PCA with application to multivariate statistical process monitoring*, AIChE Journal, vol. 44, pp. 1596–1610, 1998.
[7] F. Ferracuti, A. Giantomassi, S. Longhi, and N. Bergantino, *Multi-scale PCA based fault diagnosis on a paper mill plant*, in Emerging Technologies Factory Automation (ETFA), 2011 IEEE 16th Conference on, 2011, pp. 1–8.
[8] A. Giantomassi, F. Ferracuti, A. Benini, S. Longhi, G. Ippoliti, and A. Petrucci, *Hidden markov model for health estimation and prognosis of turbofan engines*, in ASME 2011 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, 2011, pp. 1–6.
[9] N. Birbaumer and L. Cohen, *Brain-computer interfaces: communication and restoration of movement in paralysis*, The Journal of physiology, vol. 579, no. 3, pp. 621–636, 2007.
[10] D. J. McFarland and J. R. Wolpaw, *Brain-computer interfaces for communication and control*, Communications of the ACM, vol. 54, no. 5, pp. 60–66, 2011.
[11] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, *Brain-computer interfaces for communication and control*, Clinical Neurophysiology, vol. 6, no. 113, 2002.
[12] F. Ferracuti, A. Freddi, S. Iarlori, S. Longhi, and P. Peretti, *Auditory paradigm for a P300 BCI system using spatial hearing*, in Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, 2013, pp. 871–876.
[13] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, *Kernel density estimation via diffusion*, Annals of Statistics, vol. 38, no. 5, pp. 2916–2957, 2010.
[14] M. P.Wand and M. C. Jones, *Multivariate plug-in bandwidth selection*, Computational Statistics, vol. 9, pp. 97–116, 1994.
[15] S. Kullback and R. A. Leibler, *On information and sufficiency*, Annals of Mathematical Statistics, vol. 22, pp. 49–86, 1951.
[16] S. Mallat, *A theory for multiresolution signal decomposition: the wavelet representation*, Pattern Analysis

and Machine Intelligence, vol. 11, pp. 674–693, 1989.

[17] P. E. Odiowei and Y. Cao, *Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations*, Industrial Informatics, IEEE Transactions on, vol. 6, no. 1, pp. 36 –45, feb. 2010.

[18] R. Dunia and S. Qin, *Joint diagnosis of process and sensor faults using principal component analysis*, Control Engineering Practice, vol. 6, pp. 457–469, 1998.

[19] F. Camci and R. B. Chinnam, *Health-state estimation and prognostics in machinery processes*, IEEE Transactions on Automation Science and Engineering, vol. 7, no. 3, pp. 581–597, 2010.

[20] L. R. Rabiner, *A tutorial on hidden markov models and selected applications in speech recognition*, Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, February 1989.

[21] A. Saxena, K. Goebel, D. Simon, and N. Eklund, *Damage propagation modeling for aircraft engine run-to-failure simulation*, in International Conference on Prognostics and Health Management PHM'08, Denver CO, USA, 2008.

[22] A. Kübler, N. Neumann, B. Wilhelm, T. Hinterberger, and N. Birbaumer, *Predictability of brain-computer communication*, Journal of Psychophysiology, vol. 18, no. 2, pp. 121–129, 2004.

[23] Y. Kuramoto, *Chemical oscillations, waves, and turbulence*, ser. Chemistry Series. Dover Publications, 2003, originally published: Springer Berlin, New York, Heidelberg, 1984.