

A brief assessment of outliers and malfunctions detecting techniques with an application on lubricant condition monitoring

G Grebenișan^{1,3}, S Bogdan¹, N Salem², D C Negrău¹

¹University of Oradea, Romania

²Zarqa University, Jordan

³grebe@uoradea.ro

Abstract. Condition monitoring and machine status classification are of great practical importance in the manufacturing industry as it provides online updates on the state of the machine, avoiding loss of production and minimizing the probability of generating catastrophic damage to the machine. In this paper, the classification of conditions is based on the processing of information using wavelets based on the results of the monitoring and the data collected during such an action, measuring the characteristics of the lubricating oil over some time sufficient to produce a time series of results. In this paper, the classification system is tested and validated using observation sequences based on the maximum wavelet distribution obtained from the collected signals, monitoring the state of the lubricating oil, to define and diagnose singularities in time series.

1. Introduction

In recent years, the rapid development of industrial automation has prompted the need for smarter and more reliable processing systems.

To minimize the losses caused by production interruption and the high cost due to machine and plant faults, it is necessary to monitor the status of the on-line machine using an effective condition monitoring system to provide timely information for faulty decision making, of any type, and diagnosing these defects promptly. In general, monitoring conditions involves observing machine condition using periodic dynamic response measurements, such as vibration signals obtained from multiple transducers, or measurable features for lubricating oils. Vibration measurements are taken from the machine usually contain lots of useful information but also noise components that should be removed from the signal before the information is used for machine condition classification and maintenance planning. Measurements made for oils refer to detections of the measurable components of the matrix of physical, chemical, electrical, properties. All these measurements, relating to the properties of the oils, are based on the use of performance sensors, [1].

2. Detecting and diagnosing the malfunction

When a process error occurs, it should be detected as soon as possible. The fault and malfunction detection system must indicate that something is wrong in the process. After detection, fault diagnosis is performed, the fault is isolated, and the cause of the malfunction is detected. Typically, the techniques used to detect and diagnose defects are divided into two broad categories: estimation methods and pattern recognition methods [2].

2.1 Methods of estimation

Estimation methods require mathematical models of the processes, not very complicated, to represent the real process satisfactorily, and solving the mathematical model is not excessively time-consuming. Detecting malfunctions based on state variables also involves the risk of a significant number of non-measurable variables that must be estimated. For directly estimation, a dynamic process model is linearized around an operating point. Estimation can be done using different methods depending on how stochastic the model is, then the residue evaluation, i.e., the differences between estimated and actual measured variables. This approach, based on mathematical models, requires relatively accurate knowledge of the linearized model parameters, with estimation errors generating model errors and validation errors. The detection of defects based on parameter estimation requires the existence of a known mathematical relationship, to be able to estimate their behavior by the physical parameters of the process. This requires specific knowledge of measurable process parameters as well as their variability. Because not all process parameters are directly measurable, their changes are calculated using the estimated process parameter. This is why the relationship between model parameters and process coefficients must be unique and, preferably, known precisely, a very rarely fulfilled condition.

2.2 Methods of pattern recognition.

These methods do not require the mathematical models of the process, the creation of neural network architecture, an algorithm for fault detection and diagnosis. The idea is that the process is classified according to measured data. From spatial representation, this classification is, in fact, a transposition of the measurement space into the decision-making area. The development of an algorithm for pattern recognition and pattern classification can usually be focused on three steps: measurement collection, extraction of features, and classification. As an example of the development of these algorithms, in the case of a real process, in a first step, the measured data is collected, then a characteristic vector is calculated, the redundant data is removed by the extraction, respectively the situations in which the data are missing (missing data), creating, at this extraction stage, the prerequisites for generating decision areas. In the last step, the characteristic vector is classified into one or more classes. These classes depend on the purpose of the algorithm created. If the target algorithm refers to error detection and diagnosis, the classes could be, for example, regular operation, number of A-type faults, number of B-type faults, etc. Any neural network or pattern recognition pattern or pattern classification algorithm performs the classification and recognition of patterns based on a complex overlapping operation, as is evident, of the measurement space in the decision making space. A human being has an amazing ability to recognize patterns, and often uses a very complex logic in pattern recognition and classification, but can not often define laws and rules for doing these operations. When a classification is done with neural networks, the entire mapping from the measurement space to the decision space is done at the same time, and the classification scheme is learned through examples extracted from the classified data collection. Thus, it is evident that methods of model recognition and classification do not require mathematical, analytical models but need representative data for training.

An outlier can be defined as a "data point" in a series of times that is very important difference from the rest of the data points. Outliers are important observations that affect data analysis and should, therefore, be treated with caution. There are different types of exceptional values that can occur in a series of times: a) -additive outlier (AO) is a measurement error at time T , $1 \leq T \leq N$, due to factors outside the system (e.g., o machine malfunction or a human error in data recording could be called outlier additions, an outlier additive does not affect the trend of a process); b) - Another type of outlier is innovative outlier (IO), which is caused by a certain change in a process/system. The main difference between an AO and an IO is that it indicates the beginning of a new trend in the process, which may eventually return to normal. Sometimes, changes in process/ system characteristics may involve a permanent change of process/system status (a change from stationary to non-stationary). The two main aspects associated with outliers are the detection of outliers, and the decision to be taken after the outliers have been detected. Detecting outliers involves identifying the occurrence time T , which may not be known, as well as recognizing the outlier type.

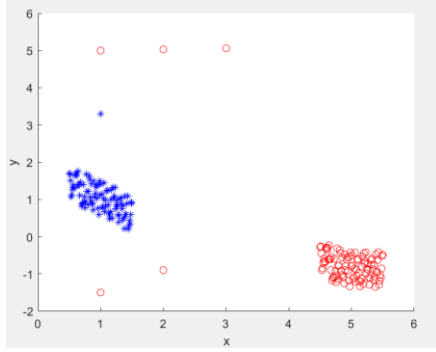


Figure 1- Normal regions and anomalies data

Figure 1 illustrates the dispersion of some anomalies in a simple set of data represented in a two-dimensional plot. Two distinct regions are distinguished, in which the data are grouped, each representing the normal behavior of a system or systems, while scattered, discordant data sufficiently distanced by the two typical regions are singularities, or anomalies of this data set, [4]. It is essential that these anomalies are not treated as "noises". Noises are generally defined as obstacles in the work of analyzing phenomena and processes, and should be treated as components of the mathematical model of the process, [5]. Considering a simple linear model that expresses the relationship between the time series of inputs, X_t , and the time series of outputs, Y_t , in the form:

$$Y_t = v(B)X_t + N_t \quad (1)$$

where, $v(B) = v_0 + v_1B + v_2B^2 + \dots$ is the transfer function, expressed as polynomial,

$$v(B) = \frac{\Omega(B)}{\delta(B)} = \delta^{-1}(B)\Omega(B) \text{ respectively, represented as the ratio of mathematical operators, and}$$

which denotes the dynamic relation between outputs and inputs, where B is a backward shift operator (a set of time z_t observable with the components: z_1, z_2, \dots, z_N then $Bz_t = z_{t-1}, B^m z_t = z_{t-m}$), while N_t represents the filtered noise component superimposed over the input signal, a_t (converts) the z_t component as follows:

$$z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t \quad (2)$$

where μ is a parameter that generally expresses the "level" of z_t , and a_t , is a sequence consisting of randomly weighted components, after the operator $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$, called the filter transfer function (it is different, as a definition and mathematical model of anomalies, singularities, etc.: $N_t = \psi(B)a_t$, so that analysis of a process or a system behavior requires the determination of both the transfer function $v(B)$, and the filter transfer function, $\psi(B)$, [5], [6], [7]. An iterative numerical procedure is presented in [7], starting from the analysis of two classes of outliers generated by dynamic models of exceptional interventions at unknown times: innovational outlier (IO) and additive outlier (AO) Starting from a stochastic process model, x_t , following an autoregressive-integrated-moving average (ARIMA) model (possibly with the characteristics: p - a positive integer indicating the degree of the nonseasonal autoregressive polynomial, d -a nonnegative integer indicating the degree of nonseasonal integration in the linear time series, q -a positive integer indicating the degree of the known nonseasonal moving average polynomial), [8], [9], [10].

In [6], it is analyzed the case where the moment T of the occurrence and the unfolding of the exceptional intervention in the model of the process, and finally the appearance of the outliers, is known,

and in [7] there is presented a survey including a practical procedure for the development of the analysis a stochastic process in which the time of intervention is unknown, it can be estimated statistically, or using other techniques, [10].

3. Experimental setup

The experiment was conducted in several phases, being part of a Research Program, supported by private companies in Romania and Italy. A first test phase of an experimental model that took place at Mecoil (Italy), in the company's lab, on a demo system with simulated parameters. The elements that were tested (physicochemical parameters) were evaluated in the first stage, on a single level. Data processing and computational work have yielded positive results in gaining the necessary data points to advance in the actual monitoring of a physical system. The overcoming of this early phase has enabled the decision to install the hydraulic system containing sensors and transducers sets, designed and physically installed in the industrial installations (pilot systems) at EMSIL Techtrans SA Oradea-Romania, as well as in four other locations in Italy, to the partners Mecoil [11], [12], [13].

4. Results and discussions

Status monitoring and diagnosis, respectively, predicting the estimation of some output parameters for a technical system can be framed in the "extrapolation" chapter as an immediate mathematical operation [14], [15]. Of course, the time series for which mathematical operations of this kind can be applied are subject to errors due to the statistical nature of these mathematical classes. The example prepares an LSTM network for two case studies: one to predict oil moisture and the other to predict the oil temperature using neural network analysis [13] based on the retrieved records by the sensors located in the hydraulic system designed to simulate the operation of an oiling plant in an automatic plant, figure 2. Prediction is understood as an estimate of the values of a time function based on values of a time series, values that can be, or can not be, affected by random errors. For example, a prediction problem could be expressed as follows: Given a series of time, $S(t)$, which consists of a set of values, and a random set of disturbing signals assimilated to a set of noise, $Z(t)$, it is proposed to estimate a future value, a prediction, therefore, $P(t + \tau)$, where τ is a positive constant, the prediction is also a continuous function of time. This case study shows how to predict data from time series using long short term memory (LSTM) network, [9]. To predict the values of the next steps of a sequence, a sequence regression LSTM network will be used, where the responses are the training sequences with values changed over a time step, figure 3. This means that at each step of the input sequence, the LSTM network learns to predict the output value at the next time step. To forecast the multiples of time values in the future, the `predictedAndUpdateState` function in Matlab was used to predict the time steps one by one, and to update the LSTM network state for each prediction, figures 4 and 5. This experiment uses the data set collected within the National Program PN II project, ERA MANUNET: NR 13081221 / 13.08.2013. The example prepares an LSTM network for two case studies: one to predict oil moisture and the other to predict the oil temperature using neural network analysis [13] based on the retrieved records by the sensors located in the hydraulic system designed to simulate the operation of an oiling plant in an automatic plant. The data set is partitioned as follows: 90% of the data volume will be used for network training, and the remaining 10% of the collected data set is used to test the network. The working procedure states that the response values of the system are the network training data sequence. An LSTM regression network was used for which it was specified to have an architecture of 200 hidden units, and train for 50 epochs. This will prevent the gradients from exploding the gradient threshold. The initial value of learning rate was set to 0.005 and provided to drop the learning rate after 25 epochs to be multiplied by a factor of 0.2. The network is initialized when the predictive training data is made. Next should be the first prediction using the last time step of the training response. The network function will be useful when prediction and outputs of the network are updated sequentially using predicted data. The phase of training will be monitored permanently by the calculated root-mean-square error (RMSE), which is the measure of errors induced by the network, figure 6.

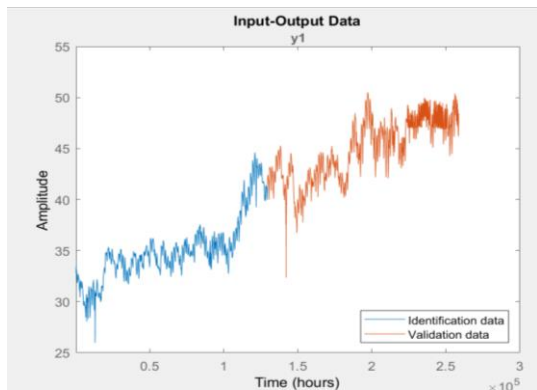


Figure 2 Separating the data into identification and a validation segment(time's scale in seconds, instead of hours)

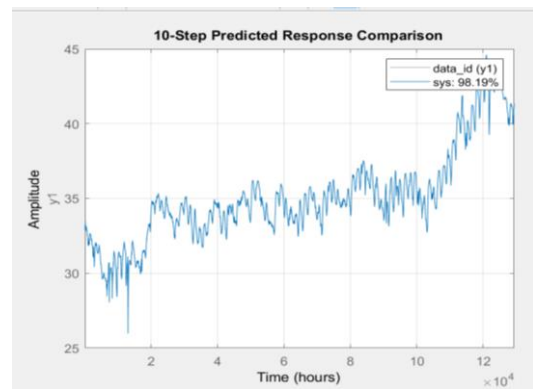


Figure 3 Using the 10 steps ahead predictor for the identification data and the independent validation data(time's scale in seconds, instead of hours)

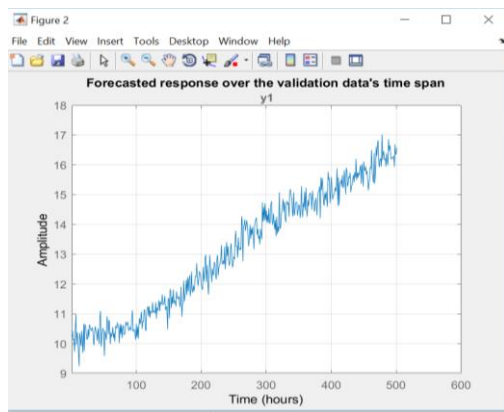


Figure4 The forecasting procedure uses the y_i , $i=1:n$, measured data recorded, to compute the model state at time step n

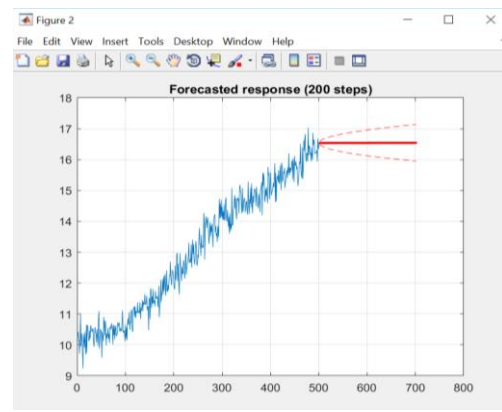


Figure 5 The forecasting results after 200 steps, shows not significant variance

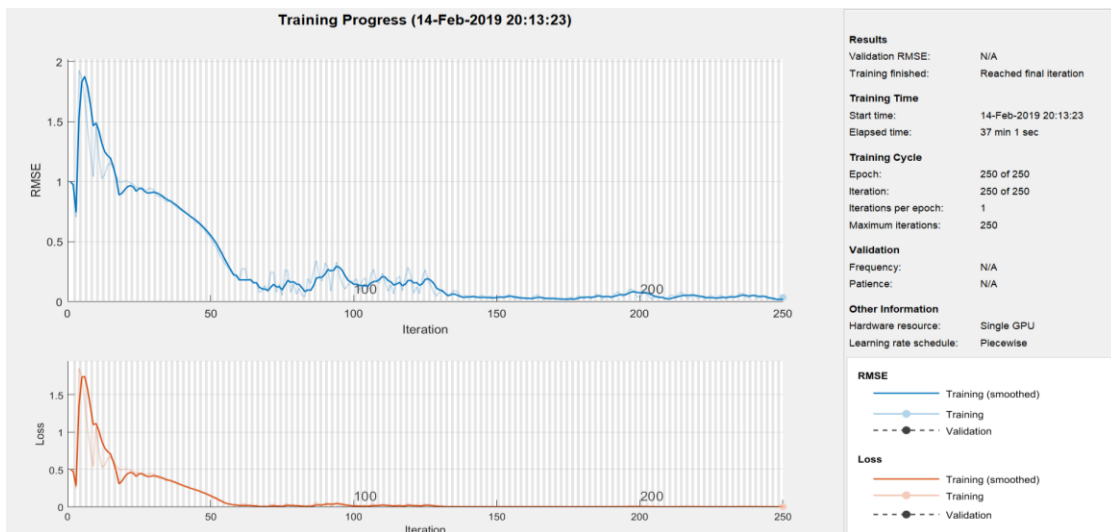


Figure 6 The plot of the training time series with the forecasted values, in terms of RMSE (root-mean-square-error) and LOSSE parameter

Acknowledgment

This work is supported by Machine Intelligence for Diagnosis Automation (MIDA), Research Program, National Program PN II, ERA MANUNET: NR 13081221/13.08.2013

References

- [1] Sorsa T., Koivo H. N., and Koivisto H. 1991 *Neural Networks in Process Fault Diagnosis*, IEEE Transactions On Systems, Man, And Cybernetics, **Vol. 21**, no. 4, DOI: 10.1109/21.108299
- [2] Teng H. S., Chen K. and Lu S. C-Y 1990 *Adaptive Real-time Anomaly Detection Using Inductively Generated Sequential Patterns* Conference Proceedings on Research in Security and Privacy IEEE Xplore DOI: 10.1109/RISP.1990.63857
- [3] Chandola V., Banerjee A., and Kumar V. 2009 *Anomaly detection: A survey*. ACM Comput. Surv., 58 pages. <http://doi.acm.org/10.1145/1541880.1541882>
- [4] Hodge V. J. and Austin J. 2004 *A Survey of Outlier Detection Methodologies*, Artificial Intelligence Review **22**: pp. 85–126, Kluwer Academic Publishers, doi: 10.1023/B: AIRE.0000045502.10941.a9
- [5] Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M. 2015 *Time Series Analysis: Forecasting and Control*, Fifth Edition, John Wiley & Sons.
- [6] Box G. E. P. & Tiao G. C. 1975 *Intervention Analysis with Applications to Economic and Environmental Problems*, Journal of the American Statistical Association, **70:349**, pp. 70-79, DOI: [10.1080/01621459.1975.10480264](https://doi.org/10.1080/01621459.1975.10480264).
- [7] Chang I., Tiao G. C., and Chen C. 1998 *Estimation of Time Series Parameters in the Presence of Outliers*, *Technometrics* **30**, no. 2 193-204. doi:10.2307/1270165.
- [8] Abraham B., Box G. E. P. 1979 *Bayesian Analysis of Some Outlier Problems in Time Series* *Biometrika* **66**(2):229-236, DOI: 10.1093/biomet/66.2.229
- [9] Statistics and Machine Learning Toolbox™ User's Guide, https://www.mathworks.com/help/releases/R2018a/pdf_doc/stats/stats.pdf, Accessed in February 2019.
- [10] Rousseeuw P. J., Leroy A. M. 1987 *Robust Regression and Outlier Detection*, John Wiley & Sons.
- [11] Adriani G., Campatelli M., Paoli M., 2013 *Monitoil® - Online sensors for efficient and cost-saving oil conditions' monitoring*, Mecoil Diagnosi Meccaniche s.r.l., Florence, Italy, http://www.mecoil.net/wp-content/uploads/2013/04/20120914-Articolo-colonne_Monitoil_eng.pdf
- [12] Grebenișan G., Salem N., and Bogdan S. 2017 *The lubricants parameters monitoring and data collecting*, MATEC Web of Conferences, **Volume 126**, 2017, <https://doi.org/10.1051/mateconf/201818403008>.
- [13] Grebenișan G., Salem N., and Bogdan S. 2018 *An approach of classification and parameters estimation using the neural network, for lubricant degradation diagnosis*, MATEC Web of Conferences, **Volume 184** <https://doi.org/10.1051/mateconf/201818403009>.
- [14] Wiener N. 1949 *Extrapolation, Interpolation, and Smoothing Of Stationary Time Series With Engineering Applications* The M.I.T. Press Massachusetts Institute Of Technology Cambridge, Massachusetts
- [15] Zadeh L. A. and Ragazzini J. R. 1950 *An Extension of Wiener's Theory of Prediction* IEEE Xplore, Journal of Applied Physics **21**(7) pp. 645 – 655, DOI:10.1063/1.1699725