Oil condition monitoring and predicting actions using an Artificial Intelligence technique: Principal Components Analysis algorithm

N. Salem

Zarqa University, Jordan

nsalem@zu.edu.jo

Abstract. Like other lubricants, oils play a crucial role in providing the space needed to separate moving parts. In addition, the lubricants ensure the suspension and transport of contaminants, neutralize corrosive acids, protect surfaces likely to wear, ensure heat dissipation, and provide the performance and increasing performance characteristics of industrial equipment and beyond. During use, lubricating oils undergo several chemical transformations due to oxidation at high temperatures (regime temperatures) due to degradation and contamination, water, ethylene glycol, coolants, waste residues. Finally, oils reach a lifetime due to a wide variety of degradation mechanisms, which lead to increased oxidation and nitration, base depletion, acid build-up, water contamination, cooling fluids, and viscosity changes. The rather complex nature of lubricants, along with the distinct variety of industrial equipment, especially the latest generation, equipped with high-performance techniques and artificial intelligence, make it very difficult, if not impossible, to predict all possibilities of generating defects. This study was intended to show how to expect action-time series data using Artificial Intelligence techniques on a set of data collected using direct/ indirect sensors and computational determinations based on empirical relationships. The algorithm Principal Components Analysis (PCA) has been approached to predict the values of the next steps of any sequence by support vector machines (SVM) models. The PCA approach was considered a favourable experiment. The answers obtained characterize and equate the training sequences with values changed by a step of the time. This means that the data structure learns to predict the next step's output value at each stage of the input sequence.

1. Introduction

Typically, an oil change is based on criteria, often based on empirical determinations, calculations that refer to hours of operation, etc. Based on sound analysis, there are fewer criteria based on chemical composition and existing lubricants resulting from tests or online records. Generally, oils are tested in specialized laboratories, the tests mainly being time-consuming and energy-intensive. Due to the tests often performed by laboratories, and operations have, for this reason, disadvantages related to the date of acquisition instantaneity, only basic degradation features coverage. Therefore, using laboratory tests to monitor the condition of oils is not a practical solution, nor is it generally applicable. It is the time factor that requires that oil state analyses be almost simultaneous, if not concurrent with sampling time. Only in this way do oil analyses reflect the equipment's actual condition.

Electrochemical analytical techniques (electrochemical impedance spectroscopy (EIS), micro acoustic viscosity), chemical (pH measurement, thin-film contaminant monitor), physical (the kinematic

viscosity, ultrasonic techniques, thermo-conductivity, ferrography, optical (optical absorption), and other types of analysis, especially those combining different kinds of scenarios, are best suited for data collection using high-performance sensors, [1].

2. Experimental setup

The experiment was conducted in several phases, being part of a Research Program supported by private companies from Romania and Italy, in collaboration with researchers and academics staff, [1], [2], [3]. The data set was collected for six months, continuously validating the data in several 258646 instances for 19 operating parameters. A first test phase of an experimental model took place at a private Italian company on a simulated demo system in the company's lab.

The tested elements (i.e., physicochemical parameters) were evaluated in the first stage on a single level. As data was collected, based on predefined scenarios, "event/ error modules" were interleaved at a higher level to test the system's response speed [3]. All the data was then processed and interpolated. At that time, the "Artificial Intelligence" contribution became essential to evaluate the infinitesimal transformations (those that can escape the vigilance of a human operator) of multiparametric elements. The system has been tested to detect between the "n" elements that need to be considered simultaneously and to evaluate in real-time the evolution and trend of the monitored parameters, being "taught" to detect singularities, outliers, essential anomalies, or possible failures, [19].

Data processing and computational work have yielded positive results to obtain the necessary data points for advancement in effectively monitoring a correctly identified mechanical system. The overcoming of this early phase has allowed the decision to install the hydraulic system containing sensor sets and transducers designed and physically established in industrial installations (called pilot systems) at a private company in Romania, as well as in four other locations in Italy, to partners of the Research Program, named above, [1], [2], [3].

The data set contains 19 parameters, not all independent, whose values have been collected over six months or determined empirically. From this set, the data collected for three status parameters (which determines the degree of degradation of the lubricating oil) and a bunch of values corresponding to the collection of target values were selected, the latter being used as a comparison element, training, learning, testing, and validation, respectively. The number of time steps is estimated at 258646 items. Outliers detection techniques have replaced missing or erroneous values by comparison with the closest mean value set. The values with which these abnormalities have been replaced complies with the condition that they are interpolated linearly over a strictly determined range in the valid set of values.

This experiment uses the data set collected within the Romanian National Program PN II project, ERA MANUNET: NR 13081221. These experiment models and analyzes at the same time the behaviour of a mechanical system, studying its behaviour based on data collected from applied sensors the hydraulic system of automatic machine lubrication. Computational determinations have been made for a small set of parameters to estimate the degradation state of the lubricating oil (A, B, C), respectively, for the complete set of data (are used all nineteen independent parameters, or not, collected during the experiment). The latter analysis required an excessive computational effort in the necessary resources, involving parallel computing techniques on an i7 computer with eight kernels. For this approach, it was also necessary to process the data provided by the last data column, which contains calculated estimates of the qualitative degradation of the oil (empirical determination, [4-11]) so that the data column reflects the correct classification of the qualitative estimates as well as the values collected from the system sensors for the other parameters of the lubricating oil. In this sense, the last data column in the data set matrix contains binary values (0 and 1) for the two possible states of the oil in the plant: the logical value "1" corresponds to a degraded oil and needs to be changed, respectively the logical value "0", represents a suitable quality oil, and can be used in the lubrication system. The data set is partitioned as follows: 70% of the data volume will be used for artificial network training, and 15% of the collected data set is used for network testing, and 15% of the dataset is used to validate the results. The working procedure states that the response values of the system are the network training data sequence. Carefully prevent situations of divergence of the forecast; it will be necessary to standardize the training data are so

organized that it has zero mean and unit variance. Also, the data set used for the test will be standardized in the same way as the training data. The long-term memory (LSTM) regression network was used, which has been specified to have an architecture of 10 hidden units and train for 100 epochs. This will prevent the gradients from exploding the gradient threshold. The initial learning rate was defined to 0.005 and provided to drop the learning rate after 25 epochs 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 helpful when the forecast and outputs of the network are updated sequentially using predicted data. The training phase is monitored permanently by the calculated root-mean-square error (RMSE), which measures errors induced by the network. This was the preparatory stage of the work

3. Principal Components Analysis algorithm and application

The Principal Components Analysis (PCA) is a very commonly used method of extracting statistical features [4], [12-16]. Fortunately, in variable data sets, varying groups evolve concurrently. An explanation for this would be that several variables could follow the same evolutionary principle that governs the behaviour of a mechanical system. There are only a few such driving forces in technical systems, but plenty of instrumentation allows one to measure dozens of system variables. When this happens, the immediate danger would be to create a redundancy of information. In this instance, the subject appears hard to be solved because of the surplus of information. Still, it can be simplified by replacing the variable set with a new set of variables, generally smaller in number and size. Thus, considering a collection of measured data (a matrix with m lines and n columns), $\boldsymbol{X_{mxn}}$, where each column is a single sample (or an instance) of the data set (ie $\vec{\boldsymbol{X}}$), and either $\boldsymbol{Y_{kxn}}$ another matrix, which is a linear transformation $\boldsymbol{P}$ of the first matrix. In conclusion, $\boldsymbol{X}$ is the original set of recorded data, and $\boldsymbol{Y}$ is a re-representation of this dataset. Then, expressing the change of the base $\boldsymbol{X}$, with the transformation $\boldsymbol{P}$, in $\boldsymbol{Y}$, results [4], [13-18]:

$$P*X=Y \tag{1}$$

where $\boldsymbol{P}$ is a matrix with $\boldsymbol{k}$-rows and $\boldsymbol{m}$ columns, $\boldsymbol{P_{kxm}}$And the elements of this matrix (the vectors $\boldsymbol{p_1, p_2, \ldots, p_m}$) are the set of vectors in the new base. Thus, by applying the transformation matrix $\boldsymbol{P}$ to the matrix $\boldsymbol{X}$, then the matrix $\boldsymbol{Y}$ is formally obtained:

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ \vdots & & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{km} \end{bmatrix} * \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ \vdots & & \ddots & \vdots \\ y_{k1} & y_{k2} & \cdots & y_{kn} \end{bmatrix} \tag{2}$$

alternatively, in a more simplified form:

$$\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{bmatrix} * \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} p_1 * x_1 & p_1 * x_2 & \cdots & p_1 * x_n \\ p_2 * x_1 & p_2 * x_2 & \cdots & p_2 * x_n \\ \vdots & \ddots & \cdots & \vdots \\ p_m * x_1 & p_m * x_2 & \cdots & p_m * x_n \end{bmatrix} \tag{3}$$

It should remark that each coefficient of $\boldsymbol{y_i}$ is a dot product of $\boldsymbol{x_i}$ with the corresponding row in $\boldsymbol{P}$:

$$y_i = \begin{bmatrix} p_1 * x_i \\ p_2 * x_i \\ \vdots \\ p_m * x_i \end{bmatrix} \tag{4}$$

In other words, the coefficient of j^{th} of y_i is a projection on the j^{th} line of the matrix $\boldsymbol{P}$. The matrix $\boldsymbol{P}$ rows are a new vector base indeed for representing the columns of $\boldsymbol{X}$. Hence, the elements of $\boldsymbol{Y}$ are linear combinations of those vectors. This is why the PCA is a method that projects a set of data into a new coordinate system by determining its eigenvectors and eigen-matrix values. This involves calculating a covariance matrix of a data set to minimize redundancy and maximize the variance. Mathematically, the PCA is defined as a linear orthogonal transformation and assumes that all the base vectors are an orthogonal matrix [4], [13], [16], [17]. Therefore, the PCA is targeted towards finding variations and coefficients of a data set by computing and determining its eigenvalues and eigenvectors of the covariance matrix. The PCA is calculated by determining the eigenvectors and the eigenvalues of the covariance matrix. The covariance matrix evaluates how much each dimension differs from the relative median reported to the differences between them. The covariance of two random variables (dimensions) represents their tendency to vary together:

$$\mathrm{cov}(X, Y) = E[E[X] - X] \cdot E[E[Y] - Y] \tag{5}$$

where $\boldsymbol{E}[\boldsymbol{X}]$ moreover, $\boldsymbol{E}[\boldsymbol{Y}]$ are (expected) values of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively.
For a sample of values

$$\mathrm{cov}(X, Y) = \sum_{i=1}^{N} \frac{(x_i - \mu_x)(y_i - \bar{y})}{N}$$

$$\text{where:} \tag{6}$$

$$\bar{x} = \frac{1}{N}\sum_{1}^{N} X; \quad \bar{y} = \frac{1}{N}\sum_{1}^{N} Y$$

$\bar{x}$ and $\bar{y}$ is the means of X and Y, respectively.

Equation (1), from a geometric point of view, represents rotation and stretch, which transforms $\boldsymbol{X}$ into $\boldsymbol{Y}$.

The first principal component is a single axis in space. When the projection is constructed, each observation on that axis, then the resulting values form a new variable. The variation of this variable is the maximum of all possible options of the first axis.

The second principal component is another axis in space, perpendicular to the first one. The projection of the observations on this axis generates a new variable. The variance of this variable is the maximum of all possible options of this second axis. The complete set of principal components is as large as the original set of variables. However, the sum of variations in the first few principal components is usually expected to exceed 80% of the total variance of the original data.

In MATLAB®, owned by The MathWorks Inc., one may use the *pca* MATLAB function to find the principal components. To use the *pca*, the measured data one wants to analyze should be matrix organized. For example, suppose you do not have the actual data, but you have a sample for which you know the correlation matrix or the data covariance matrix. In that case, you can use the *pcacov* MATLAB function to perform a principal component analysis.

A very intuitive method for calculations and PCA visualization is presented in detail [16], [17]. In addition, the authors have implemented an interactive PCA view, and a text viewer called Text Variation Explorer is developed. It allows the user to interactively study the outcome of the PCA and provides a better understanding of the process.

An exciting and practically experienced perspective in exploring medical data is presented in [18], [19]. Since the primary purpose of the PCA is to extract new uncorrelated characteristics, it is logical to introduce a correlation-based criterion with the possibility of defining a threshold value. Such a criterion is Kaiser-Meyer-Olkin (KMO), a benchmark that explains both the total correlation and the partial correlation:

$$KMO = \frac{\sum_i \sum_j r_{ij}^2}{\sum_i \sum_j r_{ij}^2 + \sum_i \sum_j a_{ij}^2} \quad (7)$$

in which: $r_{ij} = r(x^{(i)}, x^{(j)})$ it is an element of the correlation matrix $\boldsymbol{R}$, and $\boldsymbol{a_{ij}}$ is an element of partial matrix correlation A: $a_{ij,x^{(i,j)}} = \frac{-R_{ij}}{\sqrt{R_{ii} R_{jj}}}$ is a partial correlation coefficient for $\boldsymbol{x^{(i)}}$ and $\boldsymbol{x^{(j)}}$ without considering the effect of all of the other variables, and the characteristic $\boldsymbol{X^{(i,j)}}$ is fixed and $\boldsymbol{R_{kl}}$ is the algebraic complement for $\boldsymbol{r_{kl}}$ from the determinant of the correlation matrix $\boldsymbol{R}$. It is easy to note that if two characteristics have a common correlation factor with other characteristics, it results that partial correlation $\boldsymbol{a_{ij}}$, is small, indicating a corresponding variation. Thus, if $\boldsymbol{a_{ij}}$ approaches zero, then KMO approaches the unit value, while if $\boldsymbol{a_{ij}}$ approaches the unit value (indicating that the variables do not have a common correlation factor) then KMO approaches zero value. The author recommends using PCA if KMO> 0.5.

Since the experiment results contain data on many variables (19 parameters), it was chosen to evaluate these data using the Principal Components Analysis model, being retained for the study of three principal components in a 3D graphic representation. The names and meaning of the symbols of the collected parameters are further presented for this analysis in the Classification Learning approach and the Artificial Intelligence techniques of this paper Table 1:

Table 1- Parameters' definitions and symbols' significance

Crt. nr.	Parameter Symbol	• Significance of parameters
1	icm_rh	• Relative humidity of the oil (%)
2	icm_flow	• Oil flow (ml/min)
3	icm_temp	• Temperature of the sensor (°C)
4	icm_iso4	• Disaggregated ISO Cleanliness Code 4406:1999
5	icm_iso6	• Disaggregated ISO Cleanliness Code 4406:1999
6	icm_iso14	• Disaggregated ISO Cleanliness Code 4406:1999
7	icm_pc4	• Particle counter 4µ
8	icm_pc6	• Particle counter 6µ
9	icm_pc14	• Particle counter 14µ
10	fps_vcst	• Kinematic viscosity (cSt)
11	fps_v	• Dynamic viscosity (cPoise)
12	fps_density	• Density (g/cm^3)
13	fps_dielectric	• Dielectric
14	fps_temp	• Oil temperature (°C)
15	oh_temp	• The temperature of the sensor (°C)
16	oh_parama	• Param A used for oil degradation calculation
17	oh_paramb	• Param B used for oil degradation calculation
18	oh_paramc	• Param C used for oil degradation calculation
19	oh_od	• Oil degradation

Figure 1 shows the relative distribution of sensor data collected or analytically determined:

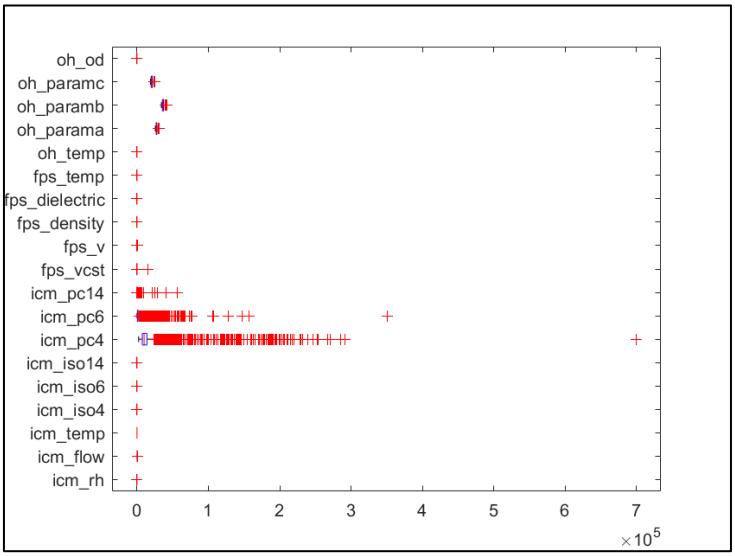

Figure 1- The distribution of the" AllOilParametersMonitor" data file

The Principal Components are represented bi-dimensionally in figure 2, also in this figure are indexed, several outliers, using the interactive features of digital graphical interfaces. Thus, you can easily see the contour of the data volume on two main components and the "scattered" set of independent data (outliers) in the isolated, but concentrated, area of the principal components.

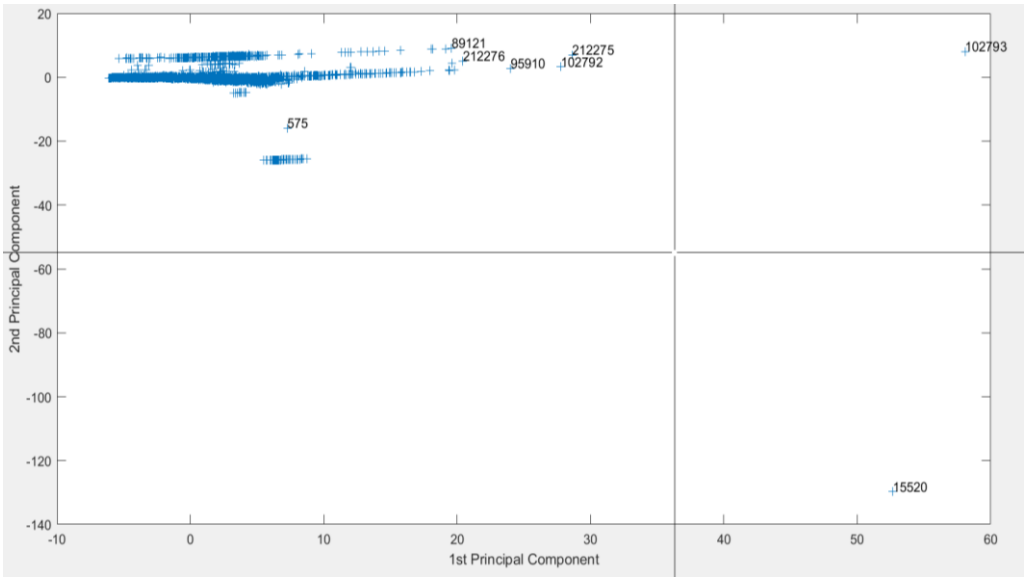

Figure 2- The Principal Components plot, with outliers outstanding

In figure 2, it can be easy to observe the volume of data falling within the "abnormalities" or "outliers" category, data to be eliminated, or for which special conditions should be imposed when processing the results.

The two-dimensional approach of the Principal Components, for all variables, offers the possibility of processing the data with the appropriate tools for the study, figure 3. It can be seen from figure 3 that the collected data is grouped quite uniformly over a relatively narrow area and that some parameters have distributed values in well-defined ranges. Anomalies being concentrated in isolated areas allow conclusions about the occurrence of anomalies and confirm the existence of "outliers" values.

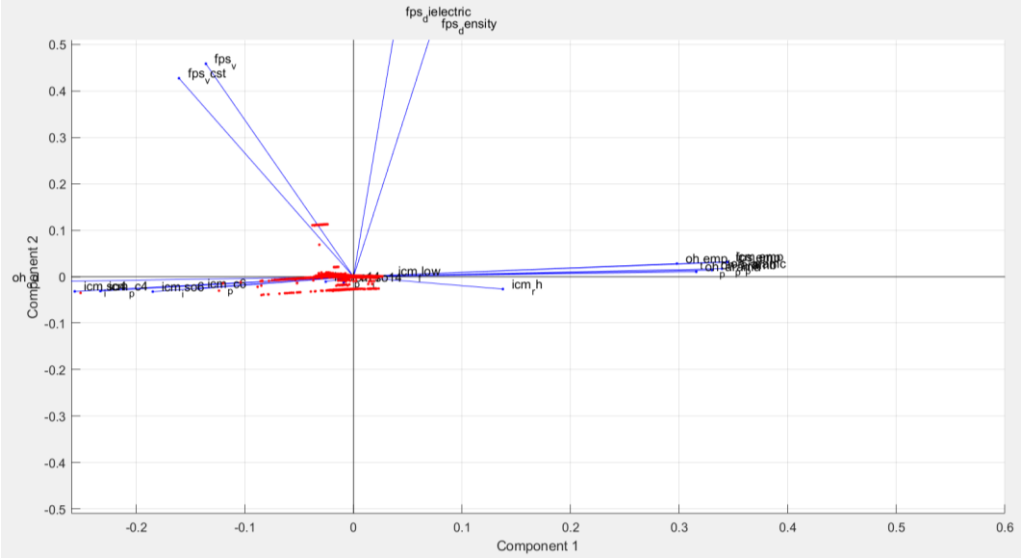

Figure 3- The two Principal Components plot, with all 19 parameters outstanding

Highlighting outliers is particularly well worth the 3D Graphic representation of Principal Components, using as bases the three main components, analytically determined, figure 4.

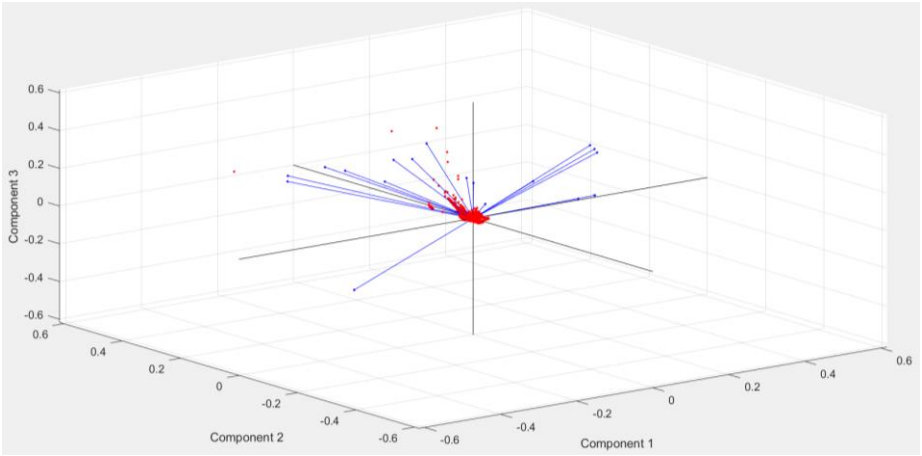

Figure 4-Three Principal Components, 3D plot representation

The variance of the three Principal Components is graphically displayed in figure 5. Again, observation is required: the figure puts into perspective the variance of the first ten components, which fall within the 95% threshold of the total variation. It is to be noticed that the first principal component, the most important one, represents a variance of almost 40%, the other components having much more comparable weights, their variance being almost insignificant, the first three components covering two-thirds of the total variance. This is why we can make the decision dimensionally reduced components, that is, consider only three of them, most obviously critical.

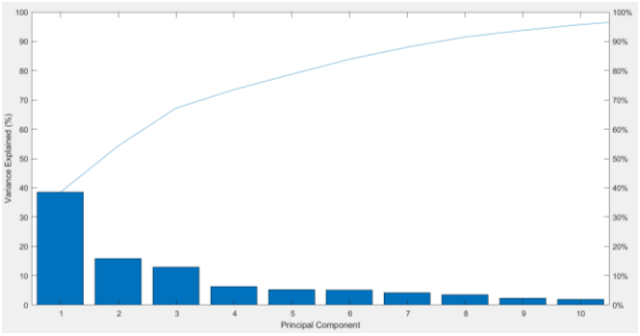

Figure 5- The amount of variance accounted for each component

4. Conclusions

Condition monitoring and diagnosis and predicting the estimation of some output parameters for a technical system, respectively, can be framed in the "extrapolation" chapter as a mathematical evaluation, testing, and validation operation [13], [15]. Doubtless, 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. Therefore, a 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. Thus, 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, $\boldsymbol{P(t+\tau)}$, where τ is a positive constant, the forecast is also a continuous function of time.

The success of a classification system depends significantly on the effectiveness of the extracted observation sequence to represent a particular condition or condition of the machine. Therefore, significant efforts have been made in research works, developing various techniques for extracting features and monitoring the machinery and installations. Characteristics extraction algorithms can be grouped, according to the basic methods used, into three main categories: the Principal Components Analysis, which is based on the identification of the axes on which data is most variable, the learning approach (Neural Networks, Pattern Classification, which have the advantages of superior learning, noise suppression, and parallel computational skills) and signal processing (Fourier transform, Hidden Markov Models or wavelet, which allow the classification based on modulus maxima distribution).

The Principal Components Analysis is a rigorous quantitative mathematical procedure for making such a simplification. The PCA is a stable and efficient method for finding an algorithm for structuring a multidimensional data set. The PCA is based on orthogonal transformations that convert a group of multidimensional values into linear-uncorrelated variables, called core components. The main disadvantage of the PCA approach is that the procedure and the outcome are often challenging to understand. The link between input and output can be confusing, a slight change of inputs can generate an utterly different result, and the user can often ask if the PCA does the right thing. The PCA generates a new set of variables, called main components, and each principal component is a linear combination of the initial variables. The fundamental property of these main components is that they are orthogonal, so there is no redundant information. The quantifiable results of this algorithm are that all the main components as a whole form an orthogonal basis for the data space.

References

[1] G. Adriani, M. Campatelli, M. Paoli, -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

[2] Grebenişan G., Salem N., and Bogdan S., *The lubricants parameters monitoring and data collecting,* MATEC Web of Conferences, **Volume** 126, 2017, https://doi.org/10.1051/matecconf/201818403008.
[3] Grebenişan G., Salem N., and Bogdan S. -*An approach of classification and parameters estimation, using the neural network, for lubricant degradation diagnosis*, MATEC Web of Conferences, **Volume** 184, 2018, https://doi.org/10.1051/matecconf/201818403009.
[4] Statistics and Machine Learning Toolbox™ User's Guide, https://www.mathworks.com/help/releases/R2018a/pdf_doc/stats/stats.pdf, Accessed in February 2019.
[5] Lihui Wang and Robert X. Gao (Eds.) Condition Monitoring and Control for Intelligent Manufacturing, 2006, Springer-Verlag.
[6] I.T. Jolliffe, Principal Component Analysis, Second Edition, Springer Series Statistics, 2010
[7] Chandola, V., Banerjee, A., and Kumar, V. 2009. Anomaly detection: A survey. ACM Comput. Surv., 2009, 58 pages. http://doi.acm.org/10.1145/1541880.1541882
[8] Victoria J. Hodge and Jim Austin, A Survey of Outlier Detection Methodologies, Artificial Intelligence Review 22: 85–126, 2004, Kluwer Academic Publishers, DOI: 10.1023/B:AIRE.0000045502.10941.a9
[9] B.C. Sharma and O.P. Gandhi, Performance evaluation and analysis of lubricating oil using parameter profile approach, Industrial Lubrication and Tribology 60(3):131-137, DOI: 10.1108/00368790810871057
[10] Junda Zhu, David He, Eric Bechhoefer, Survey of Lubrication Oil Condition Monitoring, Diagnostics, and Prognostics Techniques and Systems, Journal of Chemical Science and Technology July 2013, Vol. 2 Iss. 3, PP. 100-115
[11] N.K. Myshkin and L.V. Markova, Online Condition Monitoring in Industrial Lubrication and Tribology, Applied Condition Monitoring 8, Springer International Publishing AG, 2018, DOI 10.1007/978-3-319-61134-1_1
[12] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung, Time Series Analysis: Forecasting and Control, Fifth Edition, 2015, John Wiley & Sons.
[13] Jon Shlens, A Tutorial on Principal Component Analysis, 2014, International Journal of Remote Sensing 51(2).
[14] Parinya Sanguansat ed., Principal Component Analysis – Engineering Applications, Rijeka, 2012, InTechOpen, DOI: 10.5772/2693.
[15] Dong Hyun Jeong, Caroline Ziemkiewicz, William Ribarsky and Remco Chang, Understanding Principal Component Analysis Using a Visual Analytics Tool, *Charlotte visualization center UNC Charlotte*, 2009.
[16] Harri Siirtola, Tanja Säily, Terttu Nevalainen, Interactive Principal Component Analysis, 2017 21st International Conference Information Visualisation (IV), IEEE Xplore, **DOI:** 10.1109/iV.2017.39.
[17] Mykola Pechenizkiy, Alexey Tsymbal, Seppo Puuronen, PCA-based Feature Transformation for Classification: Issues in Medical Diagnostics, 2004, Proceedings of the IEEE Symposium on Computer-Based Medical Systems 17:535- 540, IEEE Xplore, DOI: 10.1109/CBMS.2004.1311770
[18] J. Edward Jackson, A user's guide to principal components, 1991, John Wiley & Sons Inc., DOI:10.1002/0471725331, New York.
[19] Bernhard Schölkopf, Alex J. Smola, Klaus-Robert Müller, Nonlinear Component Analysis as a Kernel Eigenvalue Problem, Neural Computation 10(5):1299-1319, DOI: 10.1162/089976698300017467, 1998