Kraków, 05.09.2024 r.

Dr hab. inż. Adam Jabłoński, prof. AGH

Wydział Inżynierii Mechanicznej i Robotyki

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

Review of the doctoral thesis of M.Eng. Hamid Shiri

Modeling and analysis of long-term historical data of time-varying complex systems in the presence of impulsive noise for condition monitoring

The review was commissioned by the Rada Dyscypliny Naukowej Inżynieria Środowiska, Górnictwa i Energetyki Politechniki Wrocławskiej, 17.07.2024 r.

1 Subject of the doctoral thesis

Scientific research on the assessment of the technical condition of rotating machinery has been conducted for a long time due to the negative impact of excessive vibration levels on the wear of machine components. The concept of the "technical condition of an object" itself could be considered either solely as the level of vibrations measured at a given moment or in relation to vibration levels measured many times in the past, which makes it possible to predict the level of vibrations in the future. The scientific starting point of the doctoral dissertation is the so-called "bathtub curve" of the machine lifetime, the essence of which is the change in the value of the features of vibroacoustic signals over time for subsequent, successive phases of that curve. The subject of the doctoral dissertation are specific challenges in the analysis of data from machine parks, resulting from non-Gaussian signal distortions and the broadly understood non-stationarity of time series composed of scalar values of diagnostic indicators calculated from vibration signals, typical for heavy industry. The author stated that the main goal of the work is to develop an original method of data processing that is able to overcome the indicated challenges. In order to achieve this goal, the Candidate presented an original concept of modeling, segmentation and prediction of time series values. To verify the developed methods, the Author used his own synthetic data, two public databases from research test-rigs equipped with rotating machines and data from a wind turbine. In this way, the subject of the doctoral dissertation includes the extension of the scientific concept of the "bathtub curve" to more complicated cases than those described in the literature, as a result of which the work presents the possibility of additional conclusions about the dynamic state of the object in the future.

AJ

2 Contents of the doctoral thesis

The doctoral dissertation consists of 7 main chapters in the form of a single concise publication. The work contains one Appendix presenting specific probability density functions used in the analysis. The work contains an abstract, a detailed list of figures, tables and abbreviations used in the work. The last chapter is a list of 158 references. The work has 116 pages. The work provides a source of funding for scientific research - the ETN MOIRA GA 955681 project. The work is written in English, the level of which fully allows understanding its content.

In Chapter 1, the Candidate describes the development of the maintenance concept and provides a classification and a detailed description of the reactive, proactive and preventive approach. Then, with reference to the ISO-13374 standard, he describes the principle of operation and subsequent stages of assessing the technical condition of machines and devices. At the end of Chapter 1, the author refers to the main topic of the work, which is the analysis of relatively large data sets.

In Chapter 2, the Candidate formulates the main research problem, i.e. the negative impact of non-Gaussian noise and the non-stationary characteristics of time series on signal processing and formulates the main objectives of the work. The author states four questions regarding modeling, segmentation and prediction, which are discussed in the further part of the work. In subsection 2.5, the Candidate presents in details the scope of his own contribution to scientific publications, which directly refer to the presented work, from which it is concluded that the Candidate's contribution among the members of the research team is the largest.

In chapter 3, the Candidate describes the analyzed data sets.

In chapter 4, the Candidate presents the original concept of modeling and identification of model parameters for long-term time series of diagnostic indicators.

In chapter 5, the Candidate presents the original segmentation method, i.e. fragmentation of time series into three pre-defined phases of machine operation.

In chapter 6, the Candidate presents a probabilistic method for assessing the remaining useful life of the machine based on exponential degradation model with confidence intervals.

In chapter 7, the Candidate presents conclusions from the work performed and describes possibilities of using the developed method in analysis of time series in the practical condition monitoring of rotary machinery, as well as outlines the direction of further research work.

The structure of the dissertation is clear. From the editorial point of view, a certain difficulty for the reader is the absence of some abbreviations used in the work in the list of abbreviations included at the beginning of the work.

AJ

3 Evaluation of the doctoral thesis

The doctoral dissertation concerns the important issue of data analysis from a machine park in the detrimental regime of non-stationarity of signals, non-linearity of systems and the large presence of non-Gaussian noise. Each of the mentioned difficulties is analyzed separately in the work. The author correctly justifies the importance of long-term data analysis in decision-making processes, especially in the assessment of the technical condition of machines and in prediction of the remaining useful life. The Author also justifies the need for the undertaken topic by indicating the possible consequences of omitting the described issues in the data analysis, which include incorrect determination of parameters, incorrect classification of outlier observations and significant prediction errors.

In the chapter devoted to the literature review, the Candidate accurately presents existing solutions and indicates specific assumptions made by other authors regarding the nature of the stochastic part of vibration signals, which limit the possibilities of assessing the technical condition. Additionally, the review on modeling contains an important functional summary.

The next part of the dissertation concerns the description of synthetic data and real data that were used in the work. Important parameters of the data sets used are described correctly.

The rest of the dissertation, consisting of chapters 4-6, presents the Author's method of modeling stochastic processes for the purpose of automated analysis of time series, enabling reliable prediction. The presented algorithm consists of three stages. The first stage concerns modeling historical data series in order to extract their significant features. The second stage consists of appropriate segmentation (i.e. fragmentation) of time series, which indicates places on the time axis corresponding to the change in the characteristics of the series. The third stage includes prediction of subsequent values of a given series based on the model, available data, threshold value and the specified length of the prediction window. The most important part of the work in the field of modeling is subchapter 4.2.2, in which the Author presents algorithms for separation, normalization and calculation of model parameters for time series characterized by a non-Gaussian distribution. For each subsequent step of the algorithm, the Candidate refers to the previously discussed simpler case, indicating the specific modifications he has made, e.g. replacing the standard deviation with a scale parameter in the normalization process, which is suitable for asymmetric distributions, or introducing a robust version of the autocovariance function, which allows for less accurate model assumptions. The significance of the developed model is demonstrated by a comparative interpretation of Figures 4.8 and 4.9, illustrating underestimation and overestimation for Gaussian modeling and the absence of these defects to a significant degree for the model proposed by the Candidate.

AJ

Chapter 5 on segmentation algorithms is divided into two parts, which allows comparison two different approaches for determination of "changing points" of the change of time series features. The first approach concerns deterministic regressions on a historical set for a constant, linear and exponential function using various mathematical techniques, including Hidden Markov Model algorithms, to decipher hidden variables. In this part of the work, the Author provides detailed analysis of the results for each of the regression methods, for each successive data set, and thus shows differences (in the determination of the change points of features) not only in terms of range, but also by analyzing the changes between changing points. The presented results allow comparison of results for different techniques in detecting different stages of machine operation. The second approach, based on the use of the Kalman filter using the cost criterion determined from the combination of correlation and entropy of the time series, as shown in the doctoral thesis, is particularly attractive for non-Gaussian disturbances due to the uncertainty of estimation. In this approach, the author proposes a relatively computationally challenging adaptive-iterative-predictive method for determining time series segmentation intervals. The results presented in this chapter indicate the attractiveness of probability-based methods over deterministic techniques, which is consistent with the properties of stochastic processes. However, due to the iterative nature of the proposed method, it would be recommended to include data on the time of calculations performed in the dissertation, especially as a function of the number of time series points.

Chapter 6 presents the author's prediction algorithm for the remaining useful life of machinery, in which the parameters of the exponential function are time-varying (i.e. depend on the values of the predicted data), which enables the generation of prediction confidence intervals. For prediction, the Candidate proposes a state-space degradation algorithm using previously proposed techniques for determination of model parameters. A particularly important part of the work from the application point-of-view is subchapter 6.1.4, in which the Author discusses the principles of determination of hyperparameters of the prediction model. The described results correctly refer to the determined probability model, demonstrating the expected feature of convergence of estimation in the function of time for the studied cases.

The presented methods have potential in the analysis of time series, but some of their parts may cause difficulties. Firstly, the author assumes that a given time series is characterized by a constant number and a constant type of mathematical model of differentiable segments. Secondly, the estimation of the remaining working time depends on the provided threshold value, and not on the nature of the change in the value of the time series. Tertiary, subsequent modeling, segmentation and prediction algorithms require knowledge of the basic statistical properties of the distribution of values of a given time series, which in turn is a parametric operation itself (e.g., correct selection the number of class intervals that correctly reflect the essential features of the data).

AJ

The doctoral thesis under evaluation contains valuable results of both cognitive and practical nature, and thus contributes to the development of knowledge on the assessment of the technical condition of rotary machinery generating vibration signals with time-varying characteristics. The significant and original achievements included in the dissertation include:

- Extension of the object degradation model known from the literature to systems with time-varying statistical features of disturbances,
- Development of a concept for a two-stage analysis of time series based on probabilistic determination of model parameters adopted on the basis of physical features of the object under study,
- Development of a mathematical model of the time series of the diagnostic indicator containing parameters that enable appropriate modeling of the variability of features over time and providing initial numerical values for these parameters,
- Development of a methodology for determination of hidden dependencies of stochastic components of the time series,
- Development of an algorithm for normalization of stochastic components in the time series model,
- Development of a segmentation algorithm together with principles for limiting changes in model parameters during calculations,
- Development of a prediction algorithm based on determination of the probability of defined potential stages of object operation.

4 Critical remarks

1. The concept of "large data sets" appears frequently in the work, but the Author does not attempt to provide boundary values between small and large sets - which may lead to a certain arbitrariness and, consequently, to improper use of the developed tools.
2. The work does not contain a direct reference to the so-called "bathtub curve" described in the literature, which is a reference point for the assumption that there are specific moments in time that refer to the change in the dynamic state of the machine.
3. An important element of the work is subsection 2.7, which refers self-critically to the incompleteness of the method. Firstly, this subsection is located rather unfortunately in Chapter 2 within the description of the subject of the work and the review of the literature, which gave the Candidate the opportunity to just tackle listed problems. Secondly, subsection 2.7 is composed of several different issues - and not, as its title suggests - only "limitations" of the work. Point 1 of subsection 2.7 concerns the assumption made in the work. Point 2 concerns the limitation of the scope of scientific research. Point 3 concerns the limitation of the presented

AJ

method in practical application. Point 4 concerns the requirement to provide external variable values (e.g. by the machine operator). Individual points should be developed in the appropriate parts of the thesis.

4. On page 20, the Candidate indicates that based on the illustration of raw time histories, three stages of machine operation are distinguished in the context of its wear. It seems that based on this one example, in light of other solutions cited in the literature (e.g. 4 phases or 5 phases), the adoption of three phases in the model is arbitrary.
5. The block captions in Figure 4.1 suggest that the modeling includes the removal of a single component (constant or linear), while in the description, the Candidate uses the term "separation" of deterministic and stochastic components. Regardless of the adopted sliding window method, these are two different things.
6. In the description of the stochastic component modeling process (page 31), the Candidate distinguishes between the case of Gaussian noise and non-Gaussian noise, but in the analysis of real data there is no such distinction.
7. On page 31, the Candidate states that the time series test for normality (distribution) is performed visually, and next states that the Kolmogorov-Smirnov test for normality of distribution was used for synthetic data. What is the source of different approach? Why was this relatively simple test chosen?
8. In subsection 4.4.3, the Author states that the correctness of the modeling is verified by determining the line of two assumed quantiles from 100 "trajectories" determined on the basis of the model, but the Candidate does not state what exactly is the source of the random variable in the simulation, as they yield different results.
9. It follows from Figures 5.3 and 5.6 that for some estimation methods the dispersion of results exceeds 50 percent. How does the Candidate interpret such a large spread? Do the individual values in Figures 5.2 and 5.5 (relative to the ranges in Figures 5.3 and 5.6) refer to the first of a hundred estimates, or to the average value, for example?
10. How does the Candidate justify verifying the results for the broadband RMS indicator in Figure 5.8 (a) using the envelope signal values in Figure 5.8 (b)? For example, in the author's opinion, can this situation be reversed, i.e. if the indicator being tested were the envelope trend values (e.g. BPFO), would the author allow for its verification using the RMS indicator?

The comments and questions presented are of a discussion nature and do not affect the essential assessment of the Candidate's achievements described in the doctoral thesis.

AJ

5 Final remarks

The dissertation presents the general theoretical knowledge of the Candidate in the discipline of Environmental, Mining and Energy Engineering, and indicates Candidate's high skills in conducting own scientific work, in accordance with the applicable research methodology in the field. The subject of the dissertation, i.e. the developed method of modeling, segmentation and prediction of time series, is an original scientific solution and contribution to the problem of data analysis. The dissertation also presents examples of the application of the developed results in various branches of heavy industry. The Candidate has achieved the goal of the work, and the results of the data analysis with conclusions contained in the work confirmed the validity of undertaken topic, and prove the potential for application in the economic aspect. An additional confirmation of the required scientific quality of the dissertation is the presence of the Candidate's co-authored publications (largely reflected in the dissertation) in general publication databases (24 publications, 309 citations, h-index 9, Scopus database, as of 03.09.2024). Taking the above into account, I hereby declare that the submitted dissertation meets the conditions specified in *art. 187 ust. 1 i 2 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2023 r., poz. 742)* and thus I declare a positive conclusion to it, requesting to *Rada Dyscypliny Naukowej Inżynieria Środowiska, Górnictwa i Energetyki Politechniki Wrocławskiej* to allow it to a public defense.

Sincerely,

Dr hab. inż. Adam Jabłoński, prof. AGH