

Assoc. Prof. Eng. Rafał Stanisławski, Professor of the University  
Chair of Computer Science  
Faculty of Electrical Engineering, Automatic Control and Informatics  
Opole University of Technology  
e-mail: r.stanislawski@po.edu.pl

Opole, 2021-04-23,

*/stamp:/* OFFICE  
of the AEE Scientific Council  
Received on 7.05.2021  
Registered with reference number -  
Signature */illegible signature/*

## REVIEW OF A DOCTORAL THESIS

**Thesis title:**

**Statistical reasoning analysis of fault occurrences in industrial applications**

**Author of the thesis: Tian Cong, M. Sc. Eng.**

**Thesis supervisor: Assoc. Prof. Eng. Jerzy Baranowski**

**Auxiliary supervisor: James R. Ottewill, PhD**

This review has been prepared on the order of Assoc. Prof. Eng. Ryszard Sroka, Professor of the University, President of the Automatics, Electronics and Electrical Engineering Scientific Council of the AGH University of Science and Technology in Krakow, of 15 February 2021.

### 1. Formal assessment of the thesis

The doctoral thesis submitted for assessment consists of a total of 113 pages divided into seven chapters, references including 238 items, and two annexes containing definitions of terms used in process condition monitoring systems and list of publications co-authored by the PhD Student. The thesis also includes summaries in Polish and English. The dissertation has been written in English.

Chapter one introduces the problems discussed in the thesis. The research problem has been formulated, the purpose of the study has been characterized, and specific tasks have been defined to be accomplished in the area of Process Condition Monitoring (PCM) together with the reasons for their undertaking. In addition, in chapter one, the structure of the dissertation in question has been explained.

Chapter two introduces the problems of PCM. The basic tasks for PCM systems have been presented, and individual processes related to these systems have been characterized. The basic classification of the tools for the modeling/identification of the systems has been presented, the methods used in the monitoring of *single mode processes* and *multimode processes* have been discussed, and *decision-making processes* have been described. The chapter also characterizes the capabilities of the PCM systems operating in the industrial environment and problems and challenges faced by these systems.

Chapter three is devoted to an original fault detection algorithm based on a binary classifier, the so-called BaFFle (*Binary Classifier for Fault Detection*). In this chapter, firstly, the possibility of adaptation of various fault detection algorithms has been discussed. A one-dimensional and multi-dimensional method of Shewhart control chart has been presented, and tools implementing this method in the multimode process monitoring task have been proposed. Next, the Principal Component Analysis (PCA) method has been analyzed, and selected multi-dimensional probability density distribution estimators have been implemented. Based on the above mentioned elements, the Author has developed a BaFFle algorithm recommended for single mode processes.

Introduction of the Dirichlet process - Gaussian Mixture Model (DP-GMM), and its implementation in the clustering task have been presented in chapter four. In the first place, selected probability distributions, i.e. Bernoulli distribution, binomial distribution, multinomial distribution, Beta distribution and generalized Dirichlet distribution have been presented, and relationships between the selected distributions in the context of the Bayes' theorem have been presented. Next, various versions of Gaussian Mixture Model (GMM) have been presented, including finite and infinite implementations of the model, as well as the Dirichlet model. Furthermore, the relationships between DP-GMM distribution and inverse-Wishart distribution have been analyzed. The chapter also includes the analysis of the methods of calculation implementation of the finite and infinite versions of the GMM and DP-GMM models. The final part of the chapter analyzes in detail the problems of selection of DP-GMM model parameters and their effect on modeling accuracy. It should be noted that the theoretical considerations have been precisely supported by simulation research. Chapter four is concluded with the presentation of the implementation of the models built as part of the diagnostic tool operation.

Chapter five is devoted to the detection of anomalies in multimode systems. To differentiate between

multimode process conditions and anomaly detection, the Candidate employs a specific extension of Kalman filter, enabling the estimation of process conditions, parameters and disturbances, called FKF (*Field Kalman Filter*). The FKF filter has been implemented to Multivariate Autoregressive State-Space (MARSS) model. Finally, in the chapter, an algorithm has been developed for the implementation of the above mentioned tool in the constructed diagnostic system, and extensive simulation research of the introduced method is conducted.

Chapter six has practical nature. The research of the diagnostic system developed as part of the dissertation has been conducted on a public access PRONTO dataset which is based on measurements made in a multiphase flow facility. The set includes data from heterogeneous sources, including process signal measurements (high frequency ultrasonic flow and pressure measurements), alarm records, operation logs and video recordings. The set includes data collected in various operating conditions, including fault conditions. The discussed chapter analyzes the methods based on binary classifier (discussed in chapter three) and the methods implementing Kalman filter (chapter five). The chapter is concluded with the efficiency analysis of individual methods.

Chapter seven contains the summary and determines planned directions of further research.

The layout of the thesis is correct. The content of the dissertation has been logically divided into individual chapters. The only doubt is raised by the order of the first two chapters. Firstly, in chapter one, the description of the study contributions has been presented, and in chapter two, the introduction to the PCM problems has been presented. In the opinion of the reviewer, the reverse chapter order would be more logical, because in the current structure, the reader is acquainted with the purposes of the study first, and learns the background in the form of the PCM problems later. The layout of individual chapters does not raise any doubts. In addition, each substantive chapter is concluded with conclusions, which additionally facilitates the reading of the dissertation.

It is worth to particularly highlight the very good editorial side of the thesis. The chapters, subchapters, sections, headers, footers, etc., are maintained in the same convention, which makes the thesis look very good. The figures are carefully prepared, with attention paid to aesthetic and precise graphical representation of the research results and flowcharts. The language used in the thesis is clear and precise which makes the thesis read quite well. The only doubt in this regard is raised by the PhD Student's use of a very large number of abbreviations which have not been listed anywhere. The thesis is also very well prepared editorially; few typographical errors are pointed out later in the review.

## **2. Substantive assessment of the thesis**

For a long time, the detection of faults/anomalies and diagnostics in industrial processes have been one of the most important aspects of the control and automatic adjustment systems reliability improvement. Over the past decades, a very large number of papers have been devoted to this field, presenting diverse approaches to these topics. Some of the more popular and effective methodologies used in this field are the solutions based on process observers, systems based on identification tools and statistical approaches employing the Bayes' theorem. Among the above mentioned methods, the approaches based on observers are usually effectively applied in combination with stochastic systems described in state-space. Most diagnostic methods for stochastic systems consider models with Gaussian interferences employing Kalman filters. However, in the case of numerous industrial processes, the solutions based on stochastic signal probability density distribution instead of instantaneous values have proved to be more effective. In this respect, the effective use of statistical information has become an approach offering new possibilities. This has led to an intensive development of this method class during the last dozen or so years. In view of the above, the Author is involved in an important and current field of study, offering the possibility to obtain significant results, both in the theoretical and practical perspective. It should also be noted that the subject matter of the thesis is clearly located in the field of Automatics, Electronics and Electrical Engineering in the discipline of engineering and technical sciences.

The Author of the thesis has set herself an ambitious and important goal to construct a system enabling the monitoring of processes of varying complexity. Achieving the goal required developing separate algorithms for relatively simple single mode processes as well as for more complex multimode processes. The content of the thesis, described in the previous section of the review, results directly from the goal set, and focuses on the development followed by the analysis of the constructed tool.

For the research of single mode processes, the Author has proposed a new, interesting algorithm for anomaly detection with binary classifier (called BaFFle). The method, based on PCA analysis and two alternative methods of probability density distribution estimation, seems quite simple. However, an interesting idea is to introduce into it a mechanism of adaptive change of signal level limits during system operation. The results presented in the practical chapter show that the algorithm for adaptive change of signal limit significantly

improves the effectiveness of anomaly detection. This method is particularly effective in combination with Kernel Density Estimation (KDE).

For the detection of multimode process anomalies, the Author has employed a clustering algorithm based on the Dirichlet process - Gaussian Mixture Model (DP-GMM). The result in the form of the final grouping algorithm implemented in the anomaly detection system I assess absolutely positively. However, it should be pointed out that the above-mentioned implementation has been preceded by an in-depth theoretical and simulation analysis of the method used. In the opinion of the reviewer, the detailed and critical analysis of the DP-GMM method, constituting the subchapters 4.4 to 4.6, is an important contribution by the PhD Student.

The most advanced anomaly detection system in multimode processes has been presented in chapter five. The original concept, presented by the PhD Student, is based on the process condition model based on autoregressive state-space model and on an estimator constituting a generalization of Kalman filter (called FKF). In the study, the Author has constructed theoretical foundations of the system, prepared operating algorithms and implemented them. The simulation research as well as practical research based on the data from a multiphase flow facility have shown high effectiveness of the algorithm. In the opinion of the reviewer, this original solution constitutes the most important achievement of the reviewed thesis. The great significance of the achievement in question is demonstrated by the PhD Student's articles published in prestigious scientific journals.

Finally, the solutions presented in the doctoral thesis have led to the development of a system enabling the monitoring of processes of varying complexity. Therefore, the purpose of the study has been achieved.

To sum up the substantive assessment, it should be noted that the Candidate's main contribution in the development of the field of Automatics, Electronics and Electrical Engineering has been presented in chapters three, four and five. However, in the opinion of the reviewer, the results presented in chapter five are particularly important. In the practical approach, chapter six is also important, verifying the developed methods on the basis of real data.

### **3. Analysis of sources, thesis position, the significance of the Author's results, the ability to present results**

The motivation to undertake the thesis subject matter has resulted from a carefully performed analysis of the publications on the subject, comprising 238 items. With extensive analysis, the current state of knowledge on the issues related to fault detection methods based on statistical inference used in the study has been correctly reflected.

It should be noted that the thesis is distinctive in relation to the state of knowledge as represented in the world literature. It is particularly worth to highlight the PhD Student's co-authorship of six valuable articles in the field of the reviewed thesis, three of which have been published in prestigious journals, i.e. *EURASIP Journal on Advances in Signal Processing*, *Journal of Process Control*, *IEEE Transactions in Control System Technology*.

The Author has the ability to present results correctly, convincingly and accurately. Both, the thesis and the articles are characterized by concise and clear language as well as precision. The PhD Student's orderliness in individual chapters of the thesis should also be emphasized. The Author begins each substantive chapter with basic elements, and then gradually conducts the discussion of further elements of the developed methods. It greatly facilitates the reading of the thesis and the interpretation of the results.

### **4. Major shortcomings of the thesis, weaknesses, comments and questions**

It should be emphasized that the reviewed doctoral thesis is written very precisely and presents a number of new, interesting and important issues on a scientific level. The substantive level of the thesis required strong involvement on the part of the reviewer in the reading and verification of the presented results. However, despite the absence of significant flaws and shortcomings, it is worth pointing out some, mostly minor, deficiencies and ambiguities. In addition, during familiarizing with the thesis, general questions came to the reviewer's mind. They have been divided into two groups: a) debatable comments and questions, and b) minor typographical and stylistic shortcomings.

Debatable comments and questions:

- There is a large number of abbreviations used in the thesis (several dozen in total). The Author precedes each abbreviation with full name when used for the first time, but such large number of them makes it significantly more difficult to read the thesis. The introduction of a synthetic description of the abbreviations at the beginning of the thesis would considerably facilitate the reading.

- Some parts, particularly in chapter two, are of very general nature. Apart from a very high level of generality, which in the opinion of the reviewer is unnecessary, the chapter contains elements remotely associated with the rest of the thesis, e.g. subchapter 2.2 Methodologies of process modelling.
- The algorithm presenting the operation of the BaFFle algorithm in chapter three (Fig. 3.3) does not show the signal density distribution estimation process. In the opinion of the reviewer, this process is quite important in the entire algorithm and should be included.
- The adaptation of parameter changes in the BaFFle method ((3.23) and (3.24) equations) employs the method of Least Mean Squares. Would the implementation of the method of Adaptive Least Squares in this location improve the operation of the algorithm?
- The method examined in chapter five is based on linear model. However, in industrial practice non-linear processes are typically encountered. In what manner does the PhD Student consider to expand the proposed method to include non-linear processes?

#### Minor typographical and stylistic shortcomings

- The table of contents of the thesis does not include the references.
- In the Table 6.3 no units of the presented results are identified. To understand the results presented in the table, it is necessary to read the Author's comment.
- Page 24, lines 11-12: The graphic layout of the page gives the impression that there is something missing between the lines 11 and 12.
- Page 54, line 12: Space missing between words.

It should be emphasized that the shortcomings and comments presented above are mostly of a disputable character and in no way affect the very positive assessment of the study.

#### 5. Review summary and final conclusion

In the opinion of the reviewer, the reviewed doctoral thesis constitutes an original solution to an important research problem, and demonstrates the Candidate's great general theoretical and applicable acquaintance with the field of Automatics, Electronics and Electrical Engineering as well as her skills necessary to engage in independent research. Therefore, on the basis of Art. 179 of the Act of 3 July 2018 Regulations implementing the Act - Law on higher education and science (Journal of Laws 2018 item 1669), I state that **this thesis satisfies with excess** the conditions specified in Art. 13 par. 1 and par. 2 of the Act of 14 March 2003 on academic degrees and title and degrees and title in the arts (Journal of Laws 2003 item 595, as amended).

Having regard to the originality of the solution to the research problem presented in the thesis, the Candidate's specialist knowledge in the field of Automatics, Electronics and Electrical Engineering, and her skills necessary to engage in independent research, **I propose to admit the doctoral thesis by Tian Cong, M. Sc. Eng., to public defense**

*/illegible signature/*