

## ANALYSIS OF DATA FROM MEASURING SENSORS FOR PREDICTION IN PRODUCTION PROCESS CONTROL SYSTEMS

Tomasz Rymarczyk<sup>1,2</sup>, Bartek Przysucha<sup>3</sup>, Marcin Kowalski<sup>2</sup>, Piotr Bednarczuk<sup>2</sup>

<sup>1</sup>Research and Development Centre, Netrix S.A., Lublin, Poland, <sup>2</sup>University of Economics and Innovation in Lublin, Lublin, Poland,

<sup>3</sup>Lublin University of Technology, Lublin, Poland

**Abstract.** The article presents a solution based on a cyber-physical system in which data collected from measuring sensors was analysed for prediction in the production process control system. The presented technology was based on intelligent sensors as part of the solution for Industry 4.0. The main purpose of the work is to reduce data and select the appropriate covariate to optimise modelling of defects using the Cox model for a specific mechanical system. The reliability of machines and devices in the production process is a condition for ensuring continuity of production. Predicting damage, especially its movement, gives the ability to monitor the current state of the machine. In a broader perspective, this enables streamlining the production process, service planning or control. This ensures production continuity and optimal performance. The presented model is a regressive survival analysis model that allows you to calculate the probability of failure occurring over a given period of time.

**Keywords:** Cox model, time to failure prediction, production control, intelligent platform

### ANALIZA DANYCH Z CZUJNIKÓW POMIAROWYCH DO PREDYKCJI W SYSTEMACH KONTROLI PROCESÓW PRODUKCYJNYCH

**Streszczenie.** Artykuł przedstawia rozwiązanie oparte na systemie cyber-fizycznym, w którym analizowano dane zbierane z czujników pomiarowych do predykcji w systemie kontroli procesów produkcyjnych. Przedstawiona technologia została oparta na inteligentnych czujnikach pomiarowych jako element rozwiązania dla Przemysłu 4.0. Głównym celem pracy jest redukcja danych i wybór odpowiedniego kowariantu w celu optymalizacji modelowania usterek za pomocą modelu Coxa dla konkretnego układu mechanicznego. Niezawodność pracy maszyn i urządzeń w procesie produkcyjnym jest warunkiem zapewnienia ciągłości produkcji. Przewidywanie uszkodzenia, a zwłaszcza jego momentu daje możliwość monitorowania bieżącego stanu maszyny. W szerszej perspektywie umożliwia to usprawnienie procesu produkcji, planowania serwisu, czy kontroli. Zapewnia to utrzymanie ciągłości produkcji i optymalnej jej wydajności. Przedstawiony model jest regresyjnym modelem analizy przeżycia, który pozwala na obliczanie prawdopodobieństwa wystąpienia awarii w określonym czasie.

**Słowa kluczowe:** model Coxa, predykcja uszkodzeń, sterowanie produkcją, inteligentna platforma

### Introduction

The article presents the results of research on the use of sensors for the analysis of technological processes using measuring devices [17]. Control of production processes and advanced automation and play an important role in industry. Production lines are an important element of production companies, where with the use of measuring and control systems it is possible to optimise technological processes that can provide high flexibility and quick adaptation of production processes, safety and efficiency at optimal costs. The presented concept consists in the use of cyber-physical systems and devices of the Internet of Things. The integration of business devices and processes gives greater opportunities and increases the efficiency of technological lines [16]. Modern production systems are based on the latest achievements in the field of information and measuring technologies (Fig. 1). There are many methods for solving optimisation problems [2, 5–12, 13–23]. Autonomy, optimisation and integration of analytical approaches is related to the operation of sensor networks, large amounts of data, analysis and interpretation of information, taking into account security aspects [1, 13–15].



Fig. 1. Industrial automation system

An important condition for ensuring continuity of production is the reliability of the equipment in the production process. Predicting damage time allows you to monitor the current status of the device and improve the planning process for inspections and service. In order to maintain production continuity and optimal production efficiency, the Cox model was used to determine the expected failure time. The presented solution is a regressive survival analysis model that allows you to calculate the probability of failure occurring at a given time or the average time of failure occurrence, etc. [3, 4, 24, 25].

### 1. Model of fault prediction

One of the most popular models for determining the expected failure time is the Cox model, which is a regressive model for survival analysis. It allows you to calculate the probability of failure occurring within a specified time. The main idea of the Cox model, the proportional threat is that devices age with time. The specific degradation process may depend on many factors that can be permanent.

In the Cox model, the proportional hazard is defined by the following terms:

The survival function is called the function given by:

$$S(x) = P(T \geq x) = 1 - F(x^-) \quad (1)$$

$F$  – lifetime distributor,  $f$  – lifetime density. Specifies the probability that the object will live longer than  $x$  time.

The hazard function is defined as the following relationship:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P[(t \leq T < t + \Delta t) | (T \geq t)]}{\Delta t} \quad (2)$$

Indicates the probability that a given object will live longer than the time  $t$ .

Cumulative risk – the cumulative hazard function is expressed by the formula:

$$H(t) = \int_0^t h(u) du \quad (3)$$

For time-dependent hazard function, covariants can be divided into time-dependent and constant.

Then the function is determined by

$$h(\mathbf{x}_i, t) = h_0(t) \exp \left[ \sum_{j=1}^{p_1} x_{ij} b_j + \sum_{j=1}^{p_2} x_{ij}(t) b_j \right] \quad (4)$$

$x_{ij}$  – time-fixed covariates,  $x_{ij}(t)$  – time-varying covariates.

A Cox proportional hazard model with time dependent variables is described:

$$h(t, \mathbb{x}_l(t), \beta) = h_0(t) \exp(\mathbb{x}_l(t)\beta) \quad (5)$$

Where  $\beta = (\beta_1, \beta_2, \dots, \beta_n)$  – model parameter vector (in the Cox proportional model it is assumed that this vector is constant in time),  $h_0(t)$  – baseline hazard function.

The hazard rate is given

$$HR(t, x_1, x_2) = \frac{h(t, \mathbb{x}_1(t), \beta)}{h(t, \mathbb{x}_2(t), \beta)} \quad (6)$$

In the event of a fault occurring at specific time intervals, when the failure times are approximate by the same amount, determining parameter estimates is complicated. It is helpful to use the algorithm proposed by Breslow:

$$l = \prod_{l=1}^m \frac{\exp(\mathbb{x}_{(l)+}\beta)}{[\sum_{j \in R(t_l)} \exp(\mathbb{x}_{(j)}\beta)]^{d_l}} \quad (7)$$

$d_l$  – number of objects with survival time  $t_{(l)}$ ,

$$\mathbb{x}_{(l)+} = \sum_{j \in D(t_{(l)})} \mathbb{x}_j \quad (8)$$

In the PHM model, the estimation of the basic hazard function takes place irrespective of the parameter model estimation. The most common form of primary threat is Weibull or exponential, Gompertz. The Weibull distribution is one of the most frequently used distributions in modelling the time of failure occurrence. He assumes that the intensity of damage is a monotonic variable. The Gompertz distribution is used when the population divides the cause of the fault into two parts. Devices or units at a young age are primarily prone to random events or diseases, while for advanced age natural resistance in people or parts wear in the case of machines decreases, which is included in the function in the form of parameters.

The Weibull distribution function looks like this:

$$h_0(t) = \frac{f(t)}{S(t)} \quad (9)$$

$$f(t) = \frac{\alpha}{\gamma} \left(\frac{t}{\gamma}\right)^{\alpha-1} \exp\left\{-\left(\frac{t}{\gamma}\right)^\alpha\right\} \quad (10)$$

$$S(t) = 1 - F(t) = \exp\left\{-\left(\frac{t}{\gamma}\right)^\alpha\right\} \quad (11)$$

where  $\alpha$  – shape parameter,  $\gamma$  – scale parameter.

The shape parameter tells us about the change in the probability of a fault occurring over time. For  $\gamma > 1$  the probability of occurrence of a fault increases with time, for  $\gamma = 1$  the probability of occurrence of a fault is constant, for  $\gamma < 1$  the probability of a fault decreases with time. The scale parameter is related to the time the fault occurred.

Therefore, the hazard function is ensured by:

$$h_0(t) = \frac{\alpha}{\gamma} \left(\frac{t}{\gamma}\right)^{\alpha-1} \quad (12)$$

After estimating the model parameters, you can test the statistical significance of the parameters. For this purpose, for example, the partial probability test or Wald test is used.

The partial likelihood ratio test statistics are provided by:

$$G = 2\{L(\hat{\beta}) - L(0)\} \quad (13)$$

$$L(0) = -\sum_{l=1}^p \ln(n_l) \quad (14)$$

where  $n_l$  is the number of objects in the risk set.

With the null hypothesis that the tested coefficient is equal to zero, it is statistics with distribution  $\chi^2$ .

Wald's Test:

This test assumes that the ratio of the estimated ratio to the standard error will have a normal distribution

$$z = \frac{\hat{\beta}}{SE(\hat{\beta})} \quad (15)$$

Partial probability tests are usually recommended.

The value of the standard deviation of the model parameters is estimated based on the inverse of Fisher's information matrix.

$$\mathbb{I}(\hat{\beta}) = -\frac{\partial^2 L(\beta)}{\partial \beta^2} \Big|_{\beta=\hat{\beta}} \quad (16)$$

On the other hand, the variance and standard deviation of the data are given in the formulas:

$$\widehat{Var}(\hat{\beta}) = \mathbb{I}(\hat{\beta})^{-1} \quad (17)$$

$$\widehat{SE}(\hat{\beta}) = \sqrt{\widehat{Var}(\hat{\beta})} \quad (18)$$

The model assumptions are tested by verifying four basic assumptions:

- Violation of the proportional threat determination,
- Appropriate functional form of accompanying variables,
- Remote observations,
- Influential observations.

Testing the correctness of the model is checked by testing the probability distribution of so-called Cox-Snell residues.

**Schoenfeld's residuals**

For the vector covariates  $\mathbb{x}_{(l)}$  calculated at time  $t_l$ ,  $k = 1, \dots, p$ . Schoenfeld's residuals are:

$$\hat{r}_k(\hat{\beta}) = \mathbb{x}_{(k)} - E(\mathbb{x}_{(k)} | R_k), \quad k = 1, \dots, p \quad (19)$$

where set  $R_k$  – a set of those objects that remain endangered until,  $t_k$ . In the event that Cox regression model assumptions are met rather than asymptotically  $E(\hat{r}_k) = 0$ . When fixing  $\hat{r}_k(\hat{\beta})$  instead of  $\beta$  we take  $\hat{\beta}$ .

**2. Results and data analysis**

For each covariant, the maximum percentage of significant models can be observed depending on the length of the time window. Both too short time windows and too long give lower percentages of significant models. Optimal values are obtained for a time window from 10 sec to 12 sec depending on the type of covariant. The least favourable case can be observed for the minimum, where the percentages of significant models decrease rapidly for all sample sizes. For the maximum, the percentage levels decrease in the least significant way, then for the average the situation is worse while for the median and for the minimum the worst. Based on this criterion, the optimal selection of a covariant in terms of choosing a time window is a maximum of the 10 seconds RMS time series.

An analysis of the selection of the covariant was also made due to the criterion of the percentage of statistically significant models depending on the sample size. The relationships between the covariant are presented in the graphs Fig 2–Fig 6.

In the selection of the covariant that gives the largest percentage of significant models depending on the time window length, the worst case is the covariant which is the minimum RMS value from the time window of the given length. From the graphs it can be seen that the larger the window size, the percentage of significant models is getting smaller just for the minimum with RMS, while the percentage of significant models does not differ significantly for other covariant.

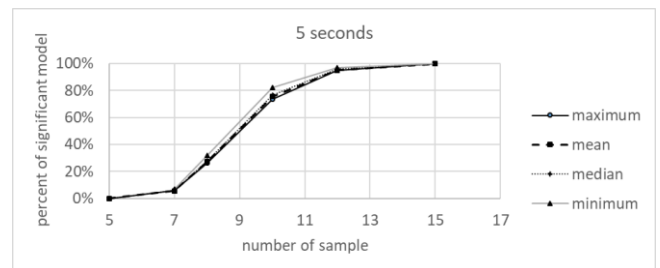


Fig. 2. Percentage of relevant models depending on the sample size for a 5-second time period

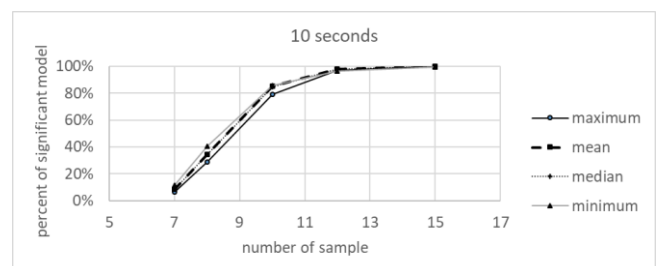


Fig. 3. Percentage of relevant models depending on the sample size for a 10-second time period

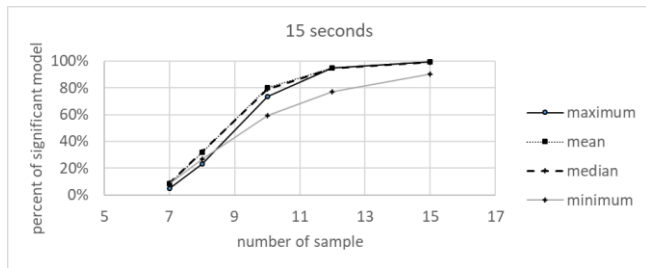


Fig. 4. Percentage of relevant models depending on the sample size for a 15-second time period

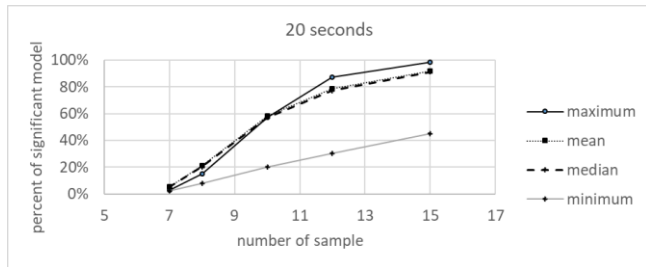


Fig. 5. Percentage of relevant models depending on the sample size for a 20-second time period

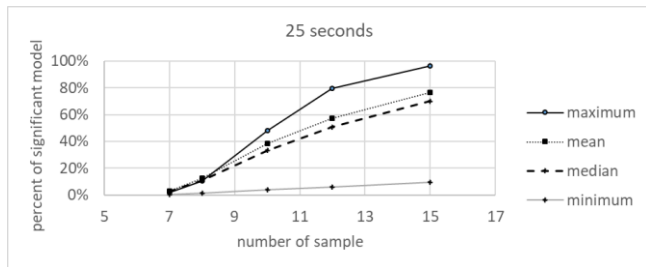


Fig. 6. Percentage of relevant models depending on the sample size for a 25-second time period

The quality of model fit depending on the covariant form was also analysed by examining the sum of standard parameter errors for the covariant. The results are shown in Fig. 7 to Fig 11.

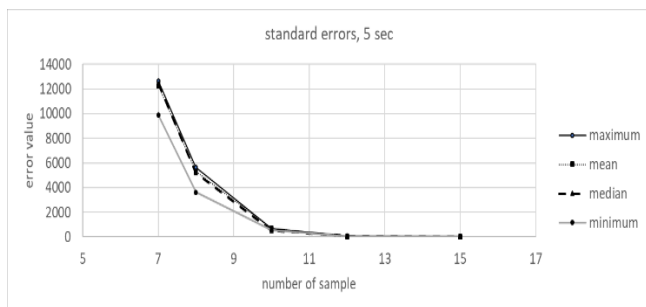


Fig. 7. Graph of the sum of standard errors of the parameter at the covariance in the Cox model depending on the sample size at the set time window length  $d = 5$  sec

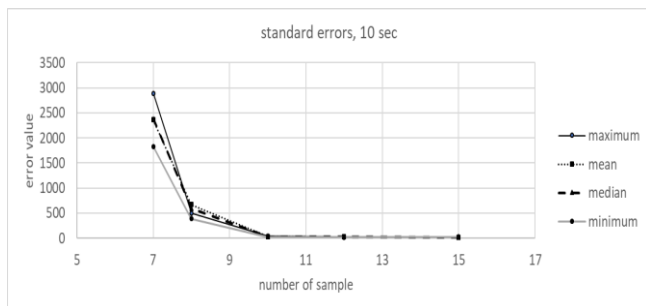


Fig. 8. Graph of the sum of standard errors of the parameter at the covariance in the Cox model depending on the sample size at the set time window length  $d = 10$  sec

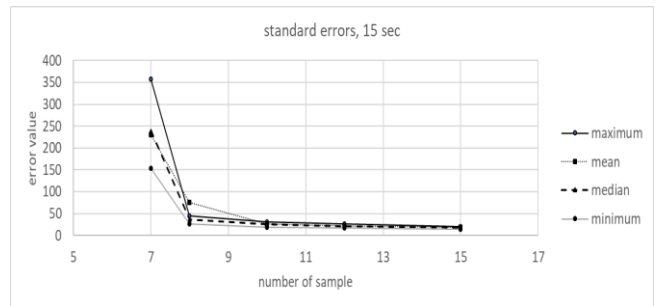


Fig. 9. Graph of the sum of standard errors of the parameter at the covariance in the Cox model depending on the sample size at the set time window length  $d=15$  sec

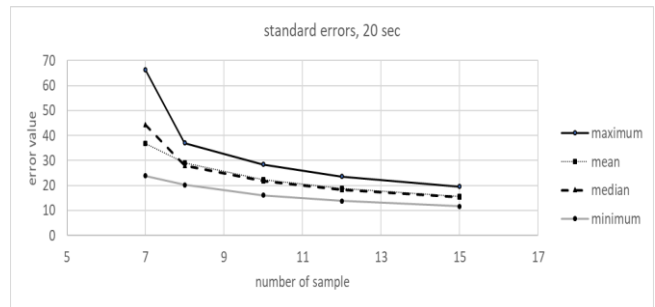


Fig. 10. Graph of the sum of standard errors of the parameter at the covariance in the Cox model depending on the sample size at the set time window length  $d=20$  sec

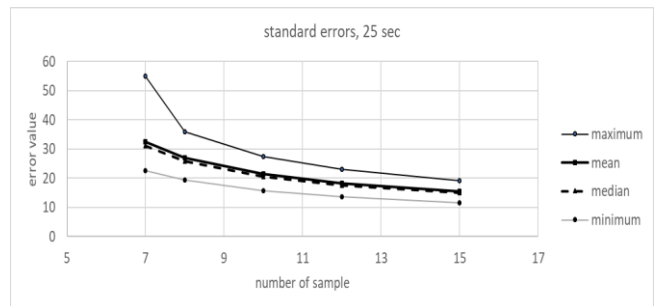


Fig. 11. Graph of the sum of standard errors of the parameter at the covariance in the Cox model depending on the sample size at the set time window length  $d=25$  sec

It can be seen from the graphs that the maximum RMS in a given time window generates the largest estimation errors. The smallest generates the minimum. The median and average generate error values between the maximum and minimum values.

### 3. Conclusion

The article presents the problem of choosing covariates in the Cox model and data reduction in order to optimise the quality of the obtained models. Data collected from measuring sensors for prediction in the production process control system. The presented technology was based on intelligent measuring sensors from an experiment consisting in introducing a defect in the mechanical shaft alignment system. Based on the received data, damage simulations were performed. Based on the generated time series data, the impact of the length of the data reduction time window on the quality of the model was examined. The maximum percentage of significant models can be observed depending on the time window length. For models where we care about the smallest estimation error, the most appropriate covariate will be the minimum RMS in the time interval. For 15-element samples, 99.9 percent of the models appear significant. For 5-element samples, no model turned out to be significant. For 7-element samples, significant models are beginning to appear, but this is only a few percent of all models. For models in which we are interested in maximizing the percentage of significant models, the best covariate will be the maximum RMS in the time window.

## References

- [1] Bergweiler S.: Intelligent Manufacturing based on Self-Monitoring Cyber-Physical Systems. UBICOMM 2015 The Ninth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, 2015.
- [2] Chen B., Abascal J., Soleimani M.: Electrical Resistance Tomography for Visualization of Moving Objects Using a Spatiotemporal Total Variation Regularization Algorithm. *Sensors* 18(2018), 1704.
- [3] Cox D., Snell E.: A general definition of residuals. *Journal of the Royal Statistical Society Series B (Methodological)* 30(1968), 248–275.
- [4] Deszyńska A.: Modele hazardów proporcjonalnych Coxa. *Matematyka stosowana* 13(54)/2011.
- [5] Dušek J., Hladký D., Mikulka J.: Electrical Impedance Tomography Methods and Algorithms Processed with a GPU. *PIERS Proceedings* 2017, 1710–1714.
- [6] Goetzke-Pala A., Hoła A., Sadowski L.: A non-destructive method of the evaluation of the moisture in saline brick walls using artificial neural networks. *Archives of Civil and Mechanical Engineering* 18(4)/2018, 1729–1742.
- [7] Grudzien K., Romanowski A., Chaniecki Z., Niedostatkiwicz M., Sankowski D.: Description of the silo flow and bulk solid pulsation detection using ECT. *Flow Measurement and Instrumentation* 21(3)/2010, 198–206.
- [8] Kozłowski E., Mazurkiewicz D., Kowalska B., et al.: Binary Linear Programming as a Decision-Making Aid for Water Intake Operators. 1st International Conference on Intelligent Systems in Production Engineering and Maintenance (ISPEM), Wrocław 2017.
- [9] Korzeniewska E., Walczak M., Rymaszewski J.: Elements of Elastic Electronics Created on Textile Substrate. *Proceedings of the 24th International Conference Mixed Design of Integrated Circuits and Systems – MIXDES 2017*, 2017, 447–454.
- [10] Kowalska A., Banasiak R., Romanowski A., Sankowski D.: Article 3D-Printed Multilayer Sensor Structure for Electrical Capacitance Tomography. *Sensors* 19(2019), 3416.
- [11] Kryszyn J., Smolik W.: Toolbox for 3d modelling and image reconstruction in electrical capacitance tomography. *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOŚ* 7(1)/2017, 137–145.
- [12] Kozłowski E., Mazurkiewicz D., Żabiński T., Prucnal S., Sęp J.: Assessment model of cutting tool condition for real-time supervision system. *Eksploatacja i Niezawodność – Maintenance and Reliability* 21(4)/2019, 679–685.
- [13] Monostori L. Cyber-physical production systems: Roots, expectations and R&D challenges. *Procedia CIRP* 17, 2014, 9–13.
- [14] Mosorov V., Grudzien K., Sankowski D.: Flow velocity measurement methods using electrical capacitance tomography. *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOŚ*, 7(1)/2017, 30–36.
- [15] Qian F., Xu G., Zhang L., Dong H.: Design of Hybrid NC Control System for Automatic Line. *International Journal of Hybrid Information Technology* 8(4)/2015, 185–192.
- [16] Repta D., Sacala I., Moisesescu M., Stanescu A.: Towards the development of a Cyber-Intelligent Enterprise System Architecture. 19th World Congress The International Federation of Automatic Control, Cape Town 2014.
- [17] Rymarczyk T., Przysucha, B.: Intelligent sensor platform for multi-source data analysis for monitoring and control of technological systems. *Applications of Electromagnetics in Modern Engineering and Medicine, PTZE 2019*, 171–175.
- [18] Rymarczyk T., Filipowicz S.F., Sikora J.: Level Set Method for Inverse Problem Solution In Electrical Impedance Tomography. *Journal Proceedings of the XII International Conference on Electrical Bioimpedance & V Electrical Impedance Tomography*, 2004, 519–522.
- [19] Rymarczyk T., Klosowski G.: Innovative methods of neural reconstruction for tomographic images in maintenance of tank industrial reactors. *Eksploatacja i Niezawodność – Maintenance and Reliability* 21(2)/2019, 261–267.
- [20] Rymarczyk T., Kozłowski E., Klosowski G., Niderla K.: Logistic Regression for Machine Learning in Process Tomography. *Sensors* 19(2019), 3400.
- [21] Rymarczyk T.: Characterization of the shape of unknown objects by inverse numerical methods. *Przegląd Elektrotechniczny* 88(7b)/2012, 138–140.
- [22] Rymarczyk T., Adamkiewicz P., Polakowski K., Sikora J.: Effective ultrasound and radio tomography imaging algorithm for two-dimensional problems. *Przegląd Elektrotechniczny* 94(6)/2018, 62–69.
- [23] Rymarczyk T., Szumowski K., Adamkiewicz P., Tchórzewski P., Sikora J.: Moisture Wall Inspection Using Electrical Tomography Measurements. *Przegląd Elektrotechniczny* 94/2018, 97–100.
- [24] Schoenfeld D.: Partial residuals for the proportional hazards regression model. *Biometrika* 69/1980, 239–241.
- [25] Xue Y., Schifano E. D.: Diagnostic for Cox model. *Communications for statistical Applications and Methods* 24(6)/2017, 583–604.

**Ph.D. Eng. Tomasz Rymarczyk**

e-mail: tomasz@rymarczyk.com

He is the director in Research and Development Centre in Netrix S.A. and the director of the Institute of Computer Science and Innovative Technologies in the University of Economics and Innovation, Lublin, Poland. He worked in many companies and institutes developing innovative projects and managing teams of employees. His research area focuses on the application of non-invasive imaging techniques, electrical tomography, image reconstruction, numerical modelling, image processing and analysis, process tomography, software engineering, knowledge engineering, artificial intelligence and computer measurement systems.

ORCID ID: 0000-0002-3524-9151

**Ph.D. Przysucha Bartosz**

e-mail: b.przysucha@pollub.pl

He is a doctor of mechanics AGH University of Science and Technology, and master of mathematic – Maria Curie-Skłodowska University in Lublin. He is a head of the Department of Quantitative Methods in Management in Lublin University of Technology. His research area focuses on acoustic, uncertainty measurement, statistical modelling and machine learning.

ORCID ID: 0000-0002-1117-8088

**Ph.D. Eng. Marcin Kowalski**

e-mail: marcin.kowalski@wsei.lublin.pl

He is the researcher in the Institute of Computer Science and Innovative Technologies in the University of Economics and Innovation, Lublin, Poland. He also works in IT department in a company of an energy sector. His research area focuses on the data bases, data analysis and methods of an optimisation of the SQL and PL/SQL code. He graduated in Electrotechnics at Technical University in Lublin. He has been post graduated in Data Bases and Data Analysis in Business.

ORCID ID: 0000-0002-1644-0612

**Ph.D. Eng. Piotr Bednarczuk**

e-mail: Piotr.Bednarczuk@wsei.lublin.pl

He is a doctor in the Institute of Computer Science at the University of Economics and Innovation in Lublin. He supports his scientific knowledge with professional practice gained in one of the leading IT companies, where he has been working for over 15 years, currently as the head of the database solutions department in the mobile systems department. His research area focuses on the software engineering web base systems, mobile-device systems and databases and data warehouses.

ORCID ID: 0000-0003-1933-7183

otrzymano/received: 30.10.2019

przyjęto do druku/accepted: 06.12.2019

