



HAL
open science

The Survival Analysis for a Predictive Maintenance in Manufacturing

Bahrudin Hrnjica, Selver Softic

► **To cite this version:**

Bahrudin Hrnjica, Selver Softic. The Survival Analysis for a Predictive Maintenance in Manufacturing. IFIP International Conference on Advances in Production Management Systems (APMS), Sep 2021, Nantes, France. pp.78-85, 10.1007/978-3-030-85906-0_9 . hal-04022124

HAL Id: hal-04022124

<https://inria.hal.science/hal-04022124>

Submitted on 9 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

The Survival Analysis for A Predictive Maintenance in Manufacturing

Bahrudin Hrnjica¹[0000-0002-3142-1284] and Selver Softic²[0000-0003-2949-8239]

¹ University of Bihac, Bihac 77000, Bosnia and Herzegovina
bahrudin.hrnjica@unbi.ba

² IT and Business Informatics, CAMPUS 02 University of Applied Sciences, Graz 8010, Austria
selver.softic@campus02.at

Abstract. The Predictive Maintenance (PdM) as a tool for detection future failures in manufacturing has recognized as innovative and effective method. Different approaches for PdM have been developed in order to compromise availability of data and demanding needs for predictions. In this paper the Survival Analysis (SA) method was used for the probability estimation for the machine failure. The paper presents the use of the two most popular SA models Kaplan-Meier non-parametric and Cox proportional hazard models. The first model was used to estimate the probability of machine to survive certain amount of cycles time. The Cox proportional model was used to find out the most significant covariates in the observed data set. The analysis shown that use of SA in the PdM is a challenging task and can be used as additional tool for failure analysis. However, due to its foundation there are several limitations in the application of SA which in most cases are the availability of the right information in the data set.

Keywords: Survival Analysis, Predictive Maintenance, Machine Learning.

1 Introduction

Maintenance plays a vital role in a manufacturing and it can be defined as the set of activities in order to preserve the system in the functional state. The maintenance can be classified in different ways depending how it can be performed on the manufacturing system. Only small number of maintenance types can be performed without interrupting the manufacturing process. However, in most of cases the maintenance must be done only if the production process is shut down. When the maintenance requires production stopping it may lead for tension between the production and the maintenance departments in a way that the production needs quality and reliable maintenance without production interruption, or at least with minimum stopping time [1]. The worse case scenario may happen when the production stops due to the component failure. Such case should never happen in the production with well-planned and organized maintenance. Be able to create a perfect maintenance plan can cost more than planned by budget. It is obvious that good maintenance planning can offer nearly a perfect production with minimal production interruption. How to achieve such a maintenance plan

depend on adopted strategies and methods for a maintenance time calculation and prediction.

Instead of relying on planned maintenance entirely, the ability to predict possible failures in the manufacturing process has become largely popular in the last decades. By knowing the possible failure of the specific component before it occurs leads to multiple benefits in the production such as reduce maintenance costs, improve production quality and increase the productivity. Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) improved the maintenance process primarily in the ability to detect and predict errors before failure occurs. Such kind of maintenance is called Predictive Maintenance (PdM) and it requires the equipment to provide data from sensors monitoring the equipment as well as other operational data. Simply speaking, it is a technique to determine (predict) the failure of the machine component in the near future so that the component can be replaced based on the maintenance plan before it fails and stops the production process [2]. Different types of maintenance with ability to improve the production process are presented on Fig 1.

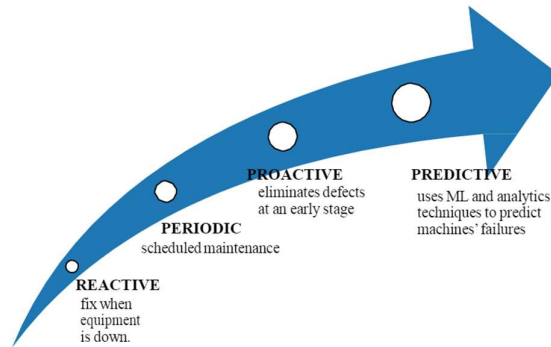


Fig. 1. Different type of maintenance

The PdM uses data collected from various sensors installed on the machines. The sensors are built into the Internet of Things (IoT) devices that send data in the cloud. Once the data are in the Cloud, different cloud solutions can use the data for processing and analysis. PdM can be implemented in the cloud solution as a part of Industry 4.0. Through the literature many different Cloud based PdM solutions have been implemented and published [3, 4, 5, 6].

1.1 PdM and SA Literature Overview

Without doubt, the PdM plays major role in the Industry 4.0. The Industry 4.0 key components Big Data, AI, ML, Cloud Computing and IoT are used in the PdM to predict the machine failure and related maintenance parameters. By using IoT and Cloud computing PdM reaches its full potential.

Combination of Big Data ecosystem orchestrated through the IoT for various PdM approaches in industrial IoT-based smart manufacturing can be found literature. In most case the IoT provide the data in real-time while the big data eco system provides the predictive analytics algorithms in order to dynamically manage preventive maintenance and failures. Beside predictive algorithms the ecosystem include numerous technologies including big data ingestion, integration, transformation, storage, analytics, and visualization in a real-time environment using various technologies such as the Data Lake, NoSQL database, Apache Spark, Apache Drill, Apache Hive, OPC Collector, and other techniques [6-10].

PdM based on Survival Analysis (SA) can be implemented on different level. In most case PdM based SA is used for the calculation of various survival curves and then incorporated into ML models. Survival curves are used in order to calculate the probability of defects derived from the Cox proportional-hazards model. Although in most cases SA is used in medical research of various diseases such as Alzheimer's, cancer, leukemia and similar [11-13], it is possible to find examples of successful application in production and related areas [14-16].

There are few examples where PdM based SA is used in Cloud based Solutions using AI, ML, Biga Data, IoT and other Industry 4.0 based technologies [15-18]. This paper presents an approach of using PdM based SA in the cloud-based solution.

2 Survival Analysis

Survival Analysis SA is a popular data analysis method first appeared in bio science and medicine science, and later spread to other scientific fields. SA tries to estimate the time to event data. The time to event T can be anything related to maintenance such as: duration of proper operation of the machine, frequency of the machine failure, or duration of the last maintenance etc. It is always positive function. In context of science the survival means probability [19]. In SA the survival function S is defined as function of time $S(t)$. Generally, the survival function $S(t)$ is defined as the probability for survival after time t of the random variable T :

$$S(t) = \Pr(T > t) \quad (1)$$

Obviously, $S(0) = 1$, which indicate the survival function is related to the lifetime distribution function. In SA the cumulative distribution function F represents the probability that the event variable occurs earlier than t .

$$F(t) = 1 - S(t) \quad (2)$$

The first derivative of cumulative density function defines the death density function (DDF) and can be expressed as:

$$f(t) = \frac{F(x)}{dt} \quad (3)$$

Hazard density function $h(t)$ represents the probability the event to be occurred in the next instant, given survival time t :

$$h(t) = \frac{f(t)}{s(t)} = -d \frac{[\ln(s)]}{dt} \quad (4)$$

3 Case Study

The Case Study of the paper is conducted from the NASA Turbofan Engine Degradation Simulation Data Set. A number of aircraft engines were monitored throughout usage history. Each engine was employed under different flight conditions while the 21 sensors recorded various states of the engine. Depending on the recorded sensor values the amount and rate of damage accumulation for each engine can be obtained. Data generation and specific meaning of the sensors are described in the literature [20].

The data was generated in the machine learning ready data sets, consisting of four different sets. Each set was generated by different operating conditions, fault mode and train/test size. In this paper the first data set FD001 was used which consisted of 100 different engines, one operating condition and one fault mode.

The data set used in the paper consisted of the following columns (covariates) separated by space:

- Engine identification number,
- The number of cycles which can be related to time,
- Three operational settings and
- 21 sensor readings.

There are two sets of FD001: the training and testing sets. For the privacy purpose the data sets are obfuscated of the column description except the engine id and the time which is represented as the number of cycles.

3.1 Data preparation for SA

In order to perform survival analysis on the presented data set the time to event and the indicator for censoring occurrence variables have to be defined. From the data set description file the train data set collect the information which monitored each engine till the failure time. This means that the training data set contains failure information for each engine, and it is the maximum cycle. However, the test set do not contain such information since the monitor history of engine usage is not completed. However, the third file of the data set contains such information and it is used for model test evaluation. The test set along the RUL file are used in the perdition scenario for other machine learning methods such that used in the literature [2].

As mentioned, the data set was generated so that each machine failed at certain cycle number thus the censored information is not provided. In order to provide the censored information, the data set is modified so that the maximum time cycle is defined before the engine is failed. The scenario used in the paper is defined so that the survival analysis analyzed the engines till certain number of cycles S . In such case the data set is

cutoff so that the engines which have failure time greater than S are censored. The several variants of S value are used in order to get the best possible model.

3.2 Kaplan-Meier non-parametric model

The Kaplan – Meier survival model [1] represent a non-parametric model that estimates the survival function from the lifetime data set. The model defines the survival function $S(t)$ which represent the probability that life is longer than time t . In our case the model estimates the probability that failure will not appear before 220 cycles:

$$\hat{S}(t) = \prod_{i, t_i \leq t=220} \left(1 - \frac{d_i}{n_i}\right), \quad (5)$$

where,

- t_i is a time at least one event happened,
- n_i the number of engines survived up to time t_i .

The following image shows the Kaplan-Meier probability function plot for the 100 engines for the maximum of 220-time cycles.

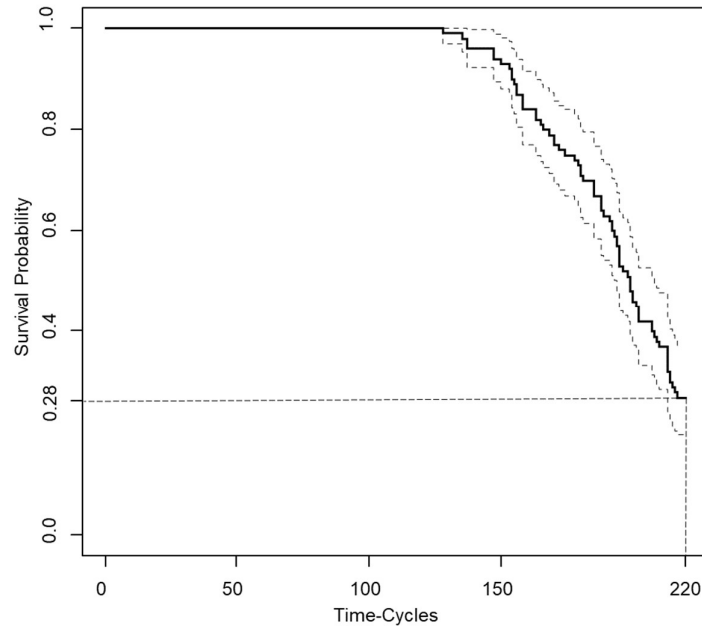


Fig. 2. Kaplan-Meier probability function plot

As can be seen all 100 engines have 100% probability of survival the first 128 cycles. Since the maximum cycle $S = 220$, one can see there is a 28% of probability that the engine will survive 220 cycles. The image also shows the confidence interval of the 95%.

3.3 Cox proportional-hazards model

The Cox proportional-hazards model [1] represents the regression model used for investigating the association between the survival time of the engine and predictor variables. In this paper the Cox proportional model was used in order to provided influence of several covariates (predictors) at rate of particular event. The model can be expressed based on the hazard function (4):

$$h(t) = h_0(t) \cdot \exp(b_1x_1 + b_2x_2 + \dots + b_kx_k), \quad (6)$$

where:

- $h(t)$ - hazard function estimated by the set of k covariates (x_1, x_2, \dots, x_k) ,
- b_1, b_2, \dots, b_k - regressors for measuring the influence of the covariates,
- $h_0(t)$ - baseline hazard which is related to the value of the hazard if all the x_i are equal to zero.

The Cox regression tests results shows several important indicators which defines the influence between the covariate and the hazard rates. The first indicator is statistical significance which defines the influence between each covariate on the hazard rate. From the data set there are more than 20 covariates potentially be included in the Cox model. However only few have influence in the model. The following tables shows covariates which have significant influence in the model:

Table 1. Statistical significance of the regressors of the the Cox proportional regression model.

Covariates	coef	exp(coef)	se(coef)	z	Pr(> z)	significance
Sensor4	8.986e-02	1.094e+00	2.711e-02	3.314	0.000919	***
Sensor7	-8.306e-01	4.358e-01	2.631e-01	-3.157	0.001592	**
Sensor15	2.673e+01	4.041e+11	6.812e+00	3.923	8.74e-05	***

Table 2. Statistical significance of the Cox proportional hazards ratios.

Cov.	exp(coef)	exp(-coef)	Low.0.95	upp.0.95	Concordance=0.773 (se=0.03)
Sensor4	1.094e+00	9.141e-01	1.037e+00	1.154e+00	l.h.r.t=91.27, 3df, p=<2e-16
Sensor7	4.358e-01	2.295e+00	2.602e-01	7.298e-01	W.test=44.61, 3 df, p=1e-09
Sensor15	4.041e+11	2.474e-12	6.430e+05	2.540e+17	S.test=56.62, df, p=3e-12

From the statistical significance (Table 2) one can see that the model is significant since all three p -values of the tests (likelihood, Wald and score) are far lower than 0.05. The tests also proved that regressors are significant.

From the Table 1 the p -value for Sensor4 is 0.000919, with hazard ration $HR = \exp(coef) = 1.094$ (Table 2) indicating the strong influence between the Sensor4

values and *increased* risk of failure. The similar statement can be defined for the Sensor15. The negative value of the Sensor7 indicates that the hazard ratio has strong influence between Sensor 7 and *decreased* risk of failure.

Comparison of the results with the results in the literature is hard to achieve due to the use of different data sets, tasks approaches and methods. Only qualitative analysis with limited comparison attributes can be established by several results in the literature[2, 21]. However, the comparison analysis should be our further step in the future research of this subject.

4 Conclusions, Limitations and Outlook

In this paper survival analysis method was applied for the predictive maintenance in order to estimate the probability of survival of 100 engines of the turbofan engine degradation data set. The application was carried out for the group of engines which were monitored throughout usage history. The engines were monitored by the 21 sensors and 3 different settings. During the time the amount of accumulated damage was estimated in form of number of cycles. In order to use such data set, the data transformation was performed in order to create SA compatible data set. The two most popular SA models Kaplan-Meier and Cox proportional models were created and tested. By using the first model we estimated the probability for the survival of each engine before certain amount of cycles has been reached. By using Cox proportional hazard model the influence of different sensor readings on hazard rates was estimated. General conclusion of the paper can be that the SA application in the PdM can be challenging task mostly because of leak of suitable data sets. SA for PdM can be used as additional tool for other classic PdM methods like Deep learning, Random Forest etc.

References

1. Budai G., Dekker R., Nicolai R.P. Maintenance and Production: A Review of Planning Models. In: Complex System Maintenance Handbook. Springer Series in Reliability Engineering. Springer, London, (2008). https://doi.org/10.1007/978-1-84800-011-7_13
2. Hrnjica B., Softic S. Explainable AI in Manufacturing: A Predictive Maintenance Case Study. In: Lalic B., Majstorovic V., Marjanovic U., von Cieminski G., Romero D. (eds) Advances in Production Management Systems. Towards Smart and Digital Manufacturing. APMS 2020. IFIP Advances in Information and Communication Technology, vol 592. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-57997-5_8
3. M. Paolanti, L. Romeo, A. Felicetti, A. Mancini, E. Frontoni and J. Loncarski, "Machine Learning approach for Predictive Maintenance in Industry 4.0," 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), Oulu, Finland, (2018), pp. 1-6. doi: 10.1109/MESA.2018.8449150
4. Hrnjica B., Mehr A.D. Energy Demand Forecasting Using Deep Learning. In: Al-Turjman F. (eds) Smart Cities Performability, Cogn., & Sec. EAI/Springer Innovations in Communication and Computing. Springer, https://doi.org/10.1007/978-3-030-14718-1_4, 2020.
5. Schmidt, B., Wang, L. Cloud-enhanced predictive maintenance. Int J Adv Manuf Technol 99, 5–13 (2018). <https://doi.org/10.1007/s00170-016-8983-8>

6. Annamalai, S., Udendhran R., & Vimal S. Cloud-Based Predictive Maintenance and Machine Monitoring for Intelligent Manufacturing for Automobile Industry. In Raj, P., & Koteswaran, S. (Ed.), *Novel Practices and Trends in Grid and Cloud Computing* (pp. 74-89). IGI Global, (2019). <http://doi:10.4018/978-1-5225-9023-1.ch006>
7. W. Yu, T. Dillon, F. Mostafa, W. Rahayu and Y. Liu, "A Global Manufacturing Big Data Ecosystem for Fault Detection in Predictive Maintenance," in *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 183-192, Jan. 2020, doi: 10.1109/TII.2019.2915846.
8. Wang, J., Zhang, L., Duan, L. et al. A new paradigm of cloud-based predictive maintenance for intelligent manufacturing. *J Intell Manuf* 28, 1125–1137 (2017). <https://doi.org/10.1007/s10845-015-1066-0>
9. March, S.T., Scudder, G.D. Predictive maintenance: strategic use of IT in manufacturing org. *Inf Syst Front* 21, 327–341 (2019). <https://doi.org/10.1007/s10796-017-9749-z>
10. He Y, Han X, Gu C, Chen Z. Cost-oriented predictive maint. based on mission reliability state for cyber man. systems. *Adv. in Mech. Eng.* doi:10.1177/1687814017751467, 2018
11. Mueller, Christoph, et al. "Associations of acetylcholinesterase inhibitor treatment with reduced mortality in Alzheimer's disease: a retrospective survival analysis." *Age and ageing* 47.1 (2018): 88-94. doi: 10.1093/ageing/afx098
12. Yao H, Xie M, Fu S, Guo J, Peng Y, Cai Z, Jiang Y, Zheng D, Wang Z. Survival analysis of patients with invasive extramammary Paget disease: implications of anatomic sites. *BMC Cancer*. doi: 10.1186/s12885-018-4257-1, 2018 .
13. Gnant, Michael, et al. "Adjuvant denosumab in early breast cancer: Disease-free survival analysis of 3,425 postmenopausal patients in the ABCSG-18 trial." (2018): 500-500.
14. de Almeida Costa, Mariana, Joaquim Pedro de Azevedo Peixoto Braga, and António Ramos Andrade. "A data-driven maintenance policy for railway wheelset based on survival analysis and Markov decision process." *Quality and Reliability Engineering International* 37.1 (2021): 176-198. <https://doi.org/10.1002/qre.2729>
15. Aydin, Olgun, and Seren Guldamlasioglu. "Using LSTM networks to predict engine condition on large scale data processing framework." 2017 4th International Conference on Electrical and Elec. Eng. (ICEEE). IEEE, (2017). doi: 10.1109/ICEEE2.2017.7935834.
16. Ramesh, P. G., et al. "Implementation of predictive maintenance systems in remotely located process plants under industry 4.0 scenario." *Advances in RAMS Engineering*. Springer, Cham, 2020. 293-326. https://doi.org/10.1007/978-3-030-36518-9_12
17. J. Wang, C. Li, S. Han, S. Sarkar and X. Zhou, "Predictive maintenance based on event-log analysis: A case study," in *IBM Journal of Research and Development*, vol. 61, no. 1, pp. 11:121-11:132, 1 Jan.-Feb. 2017, doi: 10.1147/JRD.2017.2648298.
18. Korvesis, Panagiotis, Stephane Besseau, and Michalis Vazirgiannis. "Predictive maintenance in aviation: Failure prediction from post-flight reports." 2018 IEEE 34th International Conference on Data Engineering (ICDE). IEEE, 2018.
19. Emmert-Streib, Frank, and Matthias Dehmer. "Introduction to survival analysis in practice." *Machine Learning and Knowledge Extraction* 1.3 (2019): 1013-1038.
20. A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation", in the *Proceedings of the 1st International Conference on Prognostics and Health Management (PHM08)*, Denver CO, 2008.
21. Lopes, Figueiredo, Sá Criticality evaluation to support maintenance management of manufacturing systems. *International Journal of Industrial Engineering and Management*, Vol. 11, Iss. 1, pp. 3-18. doi: 10.24867/IJEM-2020-1-248, 2020