Comparison of Data-driven Prognostics Models: A Process Perspective

Li, Rui; Verhagen, Wim; Curran, Richard

DOI
10.3850/978-981-11-2724-3_0503-cd

Publication date
2019

Document Version
Final published version

Published in
Proceedings of the 29th European Safety and Reliability Conference, ESREL 2019

Citation (APA)
Li, R., Verhagen, W., & Curran, R. (2019). Comparison of Data-driven Prognostics Models: A Process Perspective. In M. Beer, & E. Zio (Eds.), *Proceedings of the 29th European Safety and Reliability Conference, ESREL 2019* (pp. 1133-1140). Article 503 (Proceedings of the 29th European Safety and Reliability Conference, ESREL 2019). Research Publishing. https://doi.org/10.3850/978-981-11-2724-3_0503-cd

Comparison of Data-driven Prognostics Models: A Process Perspective

Rui Li, Wim J.C. Verhagen & Richard Curran

Aerospace Engineering, Delft University of Technology, Netherlands. E-mail: R.Li-4@tudelft.nl W.J.C.Verhagen@tudelft.nl, R.Curran@tudelft.nl

Remaining useful life (RUL) prediction is crucial for the implementation of Prognostics and Health Management (PHM) systems, enabling application of predictive maintenance strategies for critical systems (e.g. in aviation, power, railway). Existing literature addresses aspects of data-driven prognostic approaches, with a predominant focus on introducing and testing various novel prediction techniques which are purposed towards improving prediction accuracy performance. However, a relative lack of research can be identified when considering a comparative evaluation of competing for data-driven approaches. In particular, the contributing process elements and characteristics of data-driven prognostics methods are typically not compared in detail. To overcome these drawbacks, this paper aims to evaluate the underlying technical processes for statistical and artificial neural networks (ANN) methods for prognostics. A case study is conducted to implement both approaches on the PHM08 Challenge Data Set for comparison. This research comprehensively compares the statistical and ANN prognostic methods in a systematic manner, covering and comparing their respective technical processes, and evaluates the results with respect to prediction accuracy.

Keywords: Remaining useful life (RUL), Prognostics and Health Management (PHM), Data-Driven Prognostics, Statistical Prognostic, Artificial Neural Network (ANN).

1. Introduction

Prognostic and Health Management (PHM) approaches are designed to conduct maintenance before system failure, via assessing system condition including operating environments and estimating the risk or Remaining Useful Life (RUL) in a real-time way, based on historical trajectory data in Zhao et al. (2017). RUL estimation can improve maintenance schedules to avoid catastrophic failures and consequently save resultant costs for industries (e.g. civil aerospace, automobile, and manufacturing) in Tsui et al. (2014).

A considerable amount of research considers the development of prognostics for different components or systems. For example, Chen et al. (2011) present a review on the RUL prediction of aircraft engines, in which existing RUL estimation approaches are discussed. These are categorized as model-based, data-driven, and hybrid methods, and their characteristics are comprehensively reviewed. Particularly, the data-driven methods do not involve a priori knowledge of the physical behavior or models information of the system, but are instead relying on a collection of operational, environmental or failure/repair data for health prediction Baptista et al. (2017)). Data-driven prognostic approaches encompass a group of statistical models, e.g. Gaussian process regression (Pan et al. (2016)), support/relevance vector machine (Chen es al. (2018); Leahy et al.(2018)), gamma/wiener process (Susto et al. (2018); Zhang et al. (2018)), etc. Data-driven methods also include artificial neural network (ANN) models which learn the mapping between feature vectors and the associated RUL values. Examples include feed-forward neural networks (Ahmadzadeh and Lundberg (2013)), recurrent neural networks (RNN) (Gugulothu et al. (2017)), and convolution neural networks (CNN) (Guo et al. (2016); Zhao, and Li (2016)).

While several reviews provide excellent overviews of RUL prediction methods, these are typically covered from a holistic perspective based on data-driven methods. Yet, little research is available for comparative evaluation of data-driven prognostic methods, in particular considering statistical versus ANN models. The contributing process elements and characteristics of data-driven prognostics are typically not compared in detail. To overcome this shortcoming, this paper evaluates the prognostic technical processes for statistical and ANN

Proceedings of the 29th European Safety and Reliability Conference.
Edited by Michael Beer and Enrico Zio

Published by Research Publishing, Singapore.
ISBN: 978-981-11-2724-3; doi:10.3850/978-981-11-2724-3_0503-cd

models. Furthermore, specific process techniques are discussed and compared comprehensively.

The remainder of this article is structured as follows: In Section 2, a comparative breakdown of the prognostic processes underlying statistical and ANN models is introduced. Section 3 presents the experimental case study, including a discussion of the results. Finally, conclusions and further research are addressed in section 4.

2. Data-Driven Prognostic Models – Process Representation

2.1. Prognostic Process

Data-driven prognostics utilize feature extraction from observation data to assess the system performance and enable prediction of degradation trends in a consistent operating environment. It relies completely on the analysis of data obtained from sensors and exploits operational or performance related signals that can indicate the health of the monitored system Bailey et al. (2015). A generic representation of technical processes for prognostics using statistical and ANN models is illustrated in Fig. 1. Model-specific technical elements are emphasized using the color-coded representation; these elements are described in more detail below.

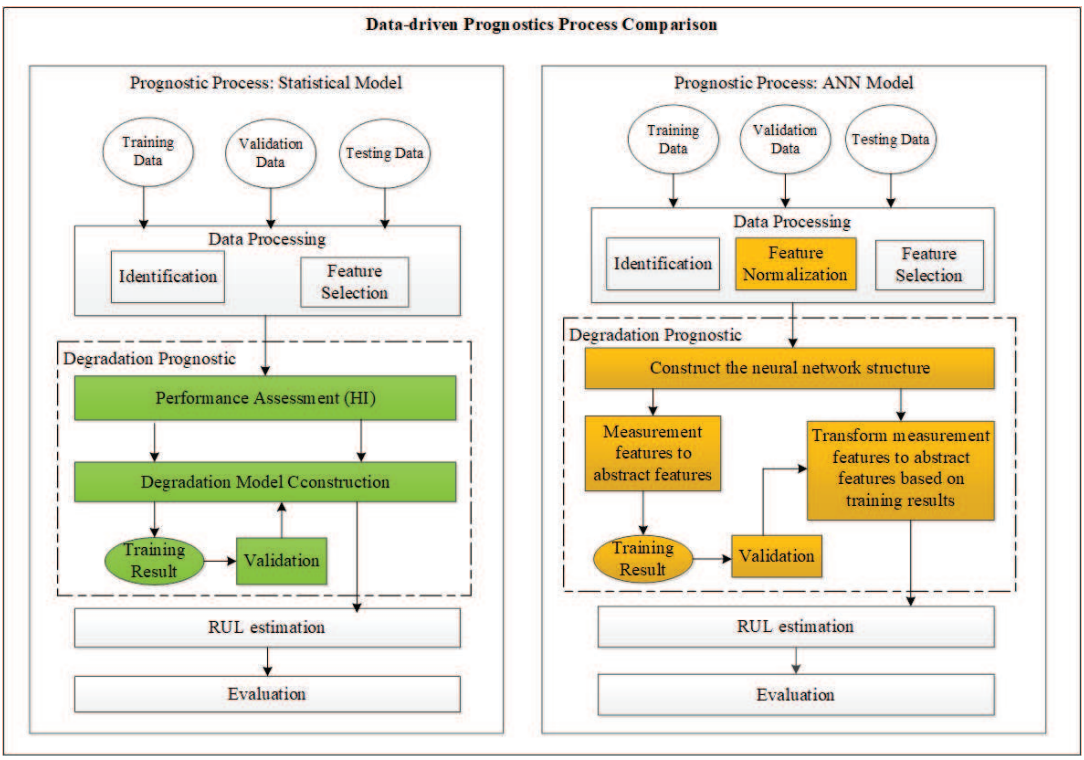

Fig.1. Data-Driven Prognostic Process Comparison

(i) Statistical Model

The statistical models for prognostics rely on available historical observational data and statistical techniques, e.g. statistical principles, stochastic processes, and mathematical models regression, for nonlinear system prediction. In Fig. 1, all the observation data are assumed to be acquired from measurement sensors. Subsequently, system performance is assessed via health indicators (HIs) constructed from the training data with some necessary data processing (Li et al. (2018)). Afterward, the degradation lifetime of the system is described by statistical models representing varying degradation trends. The constructed degradation model must be validated by validation data before the testing stage. Thus, it is able to predict the degradation lifetime after the last observation cycle of testing data in accordance with the trained degradation models. Then, the RULs of testing instances are estimated with a pre-specified failure threshold, enabling accuracy evaluation.

(ii) ANN Model

The ANN methods use training instances to obtain the desired outputs (e.g. RULs). The key is to collect useful features from the observation data. As shown in Fig.1, the ANN models perform data processing on the observation data, including options for feature identification, normalization, and selection. Next, a category of neural network (e.g. RNN, CNN) with specific configurations (e.g. layers, neurons) is set up to extract predictive characteristics from the measurement data using artificial learning. In other words, neural networks have the capability to artificially learn the relationship between the observation data and the target RULs via iteration with the objective to reduce the errors. Validation data is used to validate the hyper-parameters settings of the neural network for the specific case. In the testing stage, the measurement observations (testing data) are transformed to abstract some features under the learned neural network with the aim to estimate corresponding RULs for testing instances. Finally, the estimated RULs are evaluated using suitable accuracy metrics.

(iii) Comparison and Highlights

Refined views of the technical processes governing statistical and ANN models are highlighted in Fig. 1, which includes a comparative layer. The characteristics of each model, as well as the comparison of similar and different viewpoints, have been briefly discussed in the preceding subsections. First of all, similarities can be summarized as follows:

- **Datasets:** Both models require measurement observations, as represented in training, validation, and testing data.
- **RUL estimation and evaluation:** The essence of prognostics is to estimate RUL based on training and validation results for testing instances, as well as evaluate the estimated RULs through appropriate metrics.

On the other hand, critical differences regarding technical elements, as highlighted in Fig. 1, can be identified as follows:

- **Data processing:** Data processing is performed to transform raw data into an understandable format that can be consumed by an automated filtering process. Proper data processing is able to improve prognostic performance. Alternatively, a model implementation based on poorly partitioned data could produce misleading outcomes. The operations of feature identification and selection are commonly used in both model categories for identifying the regimes corresponding to the operation settings, as well as selecting the sensors with useful features for RUL estimation. However, the normalization processing is generally utilized in ANN models, but rarely in statistical models. Thus, this research mainly focuses on the process of normalization for comparison in Section 2.2.
- **Degradation Prognostic:** Statistical models stress the process of feature redefining and assessment techniques (e.g. through principle component analysis (PCA) or regression), and the subsequent process of developing a statistical representation for varying degradation models, as highlighted in green in Fig. 1. In contrast, the construction of a neural network (e.g. layers, neurons), feature abstraction and techniques for artificial learning (e.g. CNN, RNN) play a significant role in the technical process of ANN, as highlighted in orange in Fig. 1.

2.2. Process-specific approaches

(i) Statistical based Model

Principal component analysis (PCA) technique provides the capability to eliminate those variables that contribute the least to data variance, while makes it possible to remove those sensors and operational settings which have a weak relationship with the RUL (Son et al. (2013). As a result, this paper applies the PCA technique to construct the HIs as performance assessment for statistical-based models.

In practice, the observation data on the health of a system are noisy due to the presence of different sensors or measurement errors. For such observations, Son et.al (2016) consider a noisy observed degradation dataset. They use Gibbs' sampling technique to approximate the degradation states, and then regress the degradations into a gamma process under the approximated data for RUL estimation. In this case, the RUL is defined as the time between the observation time and the failure time with respect to the noisy observation. The detailed implementation of this prognostic algorithm is described in Son et.al (2016).

(ii) ANN Model

Generally, the first step of data processing is to identify the operational regimes in all trajectories via the analysis of the operational parameters in observation data. The number of regimes can be obtained by finding the number of clusters in the operational settings. Taking the example of the PHM08 challenge dataset: three operational settings are concentrated in six different clusters, pointing out six operating regimes in Ramasso and Saxena (2014). A normalization method (e.g. min-max, z-score, and k-means) can carry out adjustments by returning raw values into a common scale, to accordingly increase the efficiency of sensor selection and useful feature extraction for prognostics. In this paper, the collected measurements data are normalized using the min-max normalization method, where a range from -1 to 1 is considered (Li et al. (2017)). In addition, sensor selection is mostly relevant to the application with the aims to mitigate the unnecessary redundancy while maximizing the relevance in the sensors subset. Bektas et al. (2018) propose three factors to identify and evaluate key sensor data, using measures for monotonicity, prognostic-ability, and trend-ability. In the remainder of this paper, the sensor selection results from Bektas et al. (2018) are used directly for a simple case study.

ANN models are able to learn complex non-linear relationships by training multi-layer networks, which is a useful characteristic when considering RUL prediction of complex systems. When considering convolutional neural networks (CNN), high-level abstract features can be successfully extracted by the CNN architecture with the defined objective functions, so that it is able to estimate the associated RULs based on the learned representations. More specifically, the convolutional layers convolve multiple filters with raw input data and generate features, and the following pooling layers extract the most significant local features afterward in Li et al. (2017). In this sense, the process of constructing the neural network structure and abstracting features are discussed alongside.

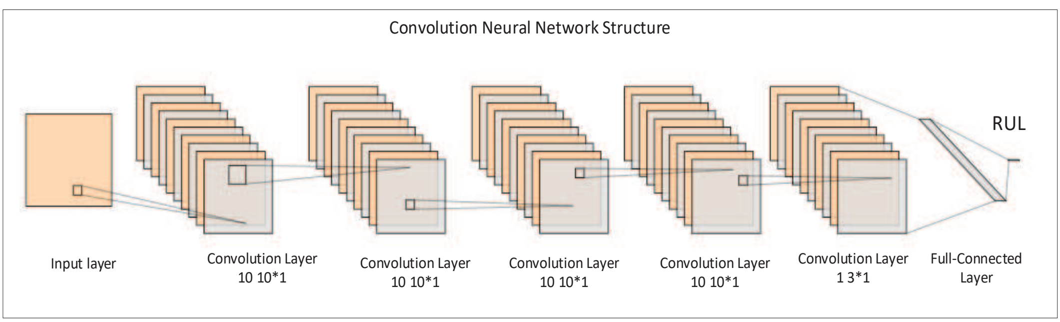

Fig. 2. Convolution Neural Network Structure (Li, Ding, and Sun 2017)

For instance, Fig. 2 illustrates the structure of a convolution neural network with a specific configuration, which provides the capability to carry out the problem of RUL estimation for a complex system, as described in the paper by Li et al. (2017). This neural network starts with 2-dimensional data as the input data in the input layer, and then it connects with 4 convolution layers of 10 kernels filters in each layer. Another convolution layer with 1 filter combines the previous feature maps to be a unique one. The feature map is flattened and connected with a fully-connected layer. Finally, one neuron is constructed at the end of this CNN network for RUL estimation.

(iii) Comparison and Highlights

Section 2.1 has introduced the general prognostic process and summarized the different viewpoints of the statistical and ANN models. This paper will implement a set of prognostic approaches to compare the highlighted technical factors, as defined in Table 1

Table 1 Comparison Matrix

Comparison items		**Methods**
Data Process		Compare the ANN prognostic models (e.g. CNN and basic ANN) with normalization and the same approaches without normalization of the observation data.
Degradation Prognostic	Statistical model	Compare the approach of similarity-based with regression and similarity-based with PCA for performance assessment.
		Compare the degradation formulation of Wiener process with PCA, similarity-based with PCA and gamma process with PCA for degradation model construction.
	ANN model	Compare the CNN prognostic approach with a basic NN prognostic approach for neural network structure and artificial learning techniques.

3. Experimental Case Study

3.1. Experimental Data

The PHM08 Challenge Dataset consists of multivariate time series that are collected from 218 (218 training instances and 218 testing instances) identical and independent instances of a turbofan engine. Each instance consists of 3 operational settings that have a substantial effect on the performance, and 21 sensors measurements in Ramasso and Saxena (2014). The case study only implements the training dataset to simplify the comparison. The first 150 instances are used for training, while the last 68 instances are applied for testing. The experimental sensors data are selected as #2, #3, #4, #7, #11, #12, #15 for application. In addition, the evaluation metrics (e.g. PHM'08 estimation metric, root squared error (RSE), mean squared error (MSE) are reflecting accuracy, meaning that lower scores indicate better performance (Ramasso and Saxena (2014)).

3.2. Implementation

3.2.1. *Statistical Model*

This case study applies the method of gamma process with PCA techniques for RUL estimation of the experimental data. Firstly, the health indicators are constructed through PCA for each training instance. A vector of HIs corresponds to an instance degradation phenomenon with noisy measurement. Gibbs technique can reduce the noise using the approximated degradation state instead of the relevant vector of HIs, to describe system degradation trends (Son et al., (2016)). As an illustration, Fig. 3 presents the observations degradation of some training instances and their Gibbs iteration degradation trends.

A Gamma process with noise is applied to estimate the degradation trend and predict the RULs for testing instances. Consequently, the estimated RULs are evaluated respectively with the results of PHM08 Score =162.1089, RSE=103.8192, and MSE=232.6176, as shown in Table2.

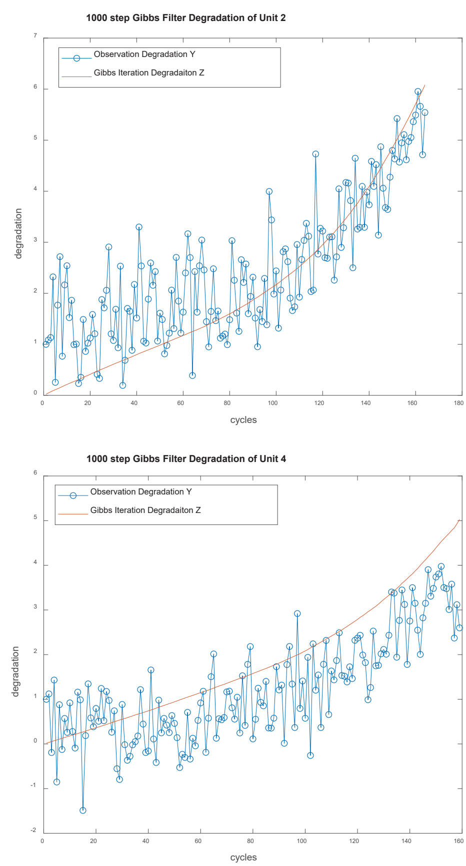

Fig. 3. Examples of Degradation Trend with Gibbs Filter

3.2.2. *ANN Model*

Regarding the ANN model, a CNN prognostic approach performing back-propagation learning in a configured convolution neural network structure is adopted, as shown in Fig. 2. Raw collected data of selected sensors with normalization is directly used as input to the proposed network, and no prior expertise on prognostics and signal processing is required, that facilitates the application of the proposed method. The measurement data is normalized using the min-max method to map in the range of [-1,1]. Fig. 4 expresses the comparison of the raw sensor data and the normalized sensor data of PHM08 Challenge Dataset.

All the training and testing instances data are prepared via sampling operation with a time window size of 15 cycles. Subsequence, the CNN learns the significant useful features of all training samples through a number of convolution filters. The operation of artificial learning can reduce the errors between the estimated RULs and labeled RULs.

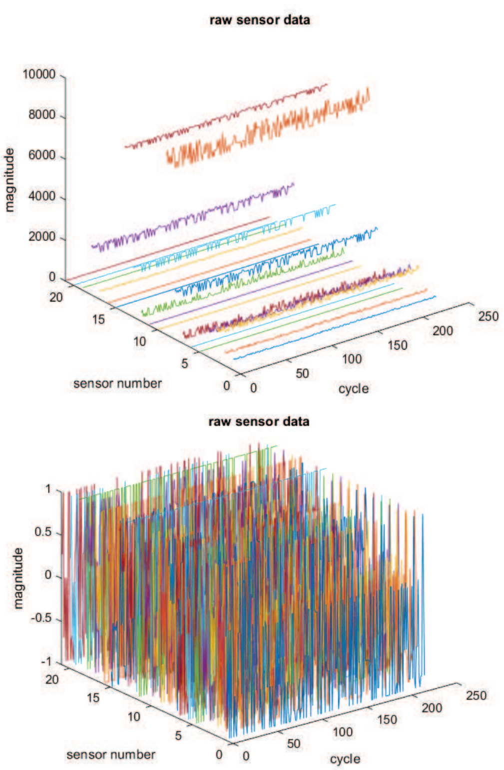

Fig. 4. Normalization of PHM08 Challenge Data

To optimize, the samples are randomly divided into multiple mini-batches with each batch containing 512 samples as the input in each epoch (Li et al. (2017)). The weights in each layer are optimized based on each mini-batch though back-propagation learning. The total number of epochs of artificial learning are set as 250 to obtain a stable convergence iteration result. All the parameters setting have been validated in the research, therefore it is assumed that the parameters are also suitable for this case study due to the same characteristics of the dataset. Finally, the testing samples are fed into the trained network for the RUL estimations, and the evaluation results are PHM08 Score= 219.8198, RSE=107.7404, and MSE=170.7059 as given in Table 3.

3.3. ***Result and Discussion***

This section analyses the evaluation results to support the comparison of statistical and ANN models as list Table 1. On one hand, Table 2 expresses that the gamma process improves the evaluation score, when comparing with the Wiener process and similarity-based prognostics methods, while using the same health indicators as found via PCA. The results reflect the efficiency of the gamma process for RUL estimation on the PHM08 Challenge Dataset. Moreover, it also shows that the linear regression technique provides better performance compared with PCA technical for HI construction on such experimental data (Li et al. (2018)).

Table 2 Evaluations of Statistical Models

Approach	*PHM08 Evaluation*	*RSE*	*MSE*
Similarity-based with Regression	231.0338	155.1644	354.0588
Wiener Process with PCA	190.2898	133.9813	263.9853
Similarity-based with PCA	305.0142	179.8027	475.4265
Gamma-Process with PCA	**162.1089**	**103.8192**	**232.6176**

To sum up, the case study demonstrates that the statistical model mainly depends on the technical process of performance assessment, and degradation model construction, involving the regressed formulation for the varying degradation models. These critical technical factors for the statistical prognostic model have been highlighted in Fig.1. Indeed, the proper techniques applied in these specific processes can provide the improvement of efficiency and accuracy of RUL estimation.

On the other hand, the CNN approach has better performance when compared with the basic neural network structure, due to the convolution operation to abstract the features in an effective way, as shown in Table 3. Further, the results also reveal that the experimental case with normalization processing before artificial neural learning can improve the accuracy of RUL estimation in this case. It means that the normalization operation can effectively extract the useful features from observations data to improve prognostic performance.

In conclusion, it indicates that the proper data processing (e.g. normalization) and the structure of the neural network (e.g. CNN) prove a significant contribution for RUL estimation problem due to the improvement of extracting useful features from measurements. These key technical processes for ANN prognostic models have been highlighted in Fig. 1.

Table 3 Evaluations of ANN models

Approach	PHM08 Evaluation	RSE	MSE
Basic NN with normalization	293.6918	130.4301	250.1765
Basic NN without normalization	1077	223.8772	737.0735
CNN without normalization	338.9281	139.0396	284.2941
CNN with normalization	**219.8198**	**107.7404**	**170.7059**

4. Conclusion

This paper presents a process-oriented comparison of data-driven prognostics, considering statistical and ANN models for RUL estimation in particular. Technical process representations for statistical and ANN models are proposed as illustrated in Fig. 1. Furthermore, the processes are compared through discussion of theoretical aspects and through application, covering the specific technical processes (e.g. data process, degradation prognostic) and evaluating the results based on common accuracy criteria.

Statistical models depend on the technical process of performance assessment, and degradation model construction regarding the regressed formulation for the varying degradation models. ANN models rely on proper data processing (e.g. normalization) and the structure and underlying technique(s) of the neural network (e.g. CNN) to improve estimation performance.

Data-driven approaches can provide excellent estimation results for prognostics. However, there are some limitations. For example, ANNs generally require a large quantity of training data, which are difficult to capture in industrial applications. Therefore, future research requires a comprehensive identification of the advantages and disadvantages of data-driven approaches with specific models with an eye towards industry application.

Acknowledgments

This research is supported by China Scholarship Council (CSC), [grant number 201609110122].

References

Ahmadzadeh, Farzaneh, and Jan Lundberg. (2013). Remaining Useful Life Prediction of Grinding Mill Liners Using an Artificial Neural Network. *Minerals Engineering* 53. Elsevier Ltd: 1–8.

Babu, Giduthuri Sateesh, Peilin Zhao, and Xiao Li Li. (2016). Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9642: 214–28.

Bailey, Chris, Thamo Sutharssan, Chunyan Yin, and Stoyan Stoyanov. (2015). Prognostic and Health Management for Engineering Systems: A Review of the Data-Driven Approach and Algorithms. *The Journal of Engineering*, 1–8.

Baptista, Marcia, Ivo P. de Medeiros, Joao P. Malere, Cairo Nascimento, Helmut Prendinger, and Elsa M.P. Henriques. (2017). Comparative Case Study of Life Usage and Data-Driven Prognostics Techniques Using Aircraft Fault Messages. *Computers in Industry* 86. Elsevier

B.V.: 1–14.

Bektas, Oguz, Jeffrey A Jones, Shankar Sankararaman, Indranil Roychoudhury, and Kai Goebel. (2018). A Neural Network Filtering Approach for Similarity-Based Remaining Useful Life Estimation. *The International Journal of Advanced Manufacturing Technology*. The International Journal of Advanced Manufacturing Technology.

Chen, Xiongzi, Jinsong Yu, Diyin Tang, and Yingxun Wang. (2011). Remaining Useful Life Prognostic Estimation for Aircraft Subsystems or Components: A Review. *Proceedings - IEEE 2011 10th International Conference on Electronic Measurement and Instruments, ICEMI 2011* 2: 94–98.

Chen, Zhongzhe, Shuchen Cao, and Zijian Mao. (2018). Remaining Useful Life Estimation of Aircraft Engines Using a Modified Similarity and Supporting Vector Machine (SVM) Approach. *Energies* 11 (1).

Gugulothu, Narendhar, Vishnu TV, Pankaj Malhotra, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff. (2017). Predicting Remaining Useful Life Using Time Series Embeddings Based on Recurrent Neural Networks. *Journal of Learning Disabilities* 12 (6): 408–15.

Guo, Xiaojie, Liang Chen, and Changqing Shen. (2016). Hierarchical Adaptive Deep Convolution Neural Network and Its Application to Bearing Fault Diagnosis. *Measurement: Journal of the International Measurement Confederation* 93. Elsevier Ltd: 490–502.

Leahy, Kevin, R Lily Hu, Ioannis C Konstantakopoulos, Costas J Spanos, Alice M Agogino, and Dominic T.J. O'Sullivan. (2018). Diagnosing and Predicting Wind Turbine Faults from SCADA Data Using Support Vector Machines. *International Journal of Prognostics and Health Management* 9 (1): 1–11.

Li, Rui, Wim J C Verhagen, and Richard Curran. (2018). A Comparative Study of Data-Driven Prognostic Approaches : Stochastic and Statistical Models. *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 1–8.

Li, Xiang, Qian Ding, and Jian-Qiao Sun. (2017). Remaining Useful Life Estimation in Prognostics Using Deep Convolution Neural Networks. *Reliability Engineering & System Safety* 172 (June 2017). Elsevier Ltd: 1–11.

Pan, Donghui, Jia Bao Liu, and Jinde Cao. (2016). Remaining Useful Life Estimation Using an Inverse Gaussian Degradation Model. *Neurocomputing* 185. Elsevier: 64–72.

Ramasso, Emmanuel, and Abhinav Saxena. (2014). Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets. *International Journal of Prognostics and Health Management*, no. ISSN2153-2648: 1–15.

Son, Khanh Le, Mitra Fouladirad, and Anne Barros. (2016). Remaining Useful Lifetime Estimation and Noisy Gamma Deterioration Process. *Reliability Engineering and System Safety* 149. Elsevier: 76–87.

Son, Khanh Le, Mitra Fouladirad, Anne Barros, Eric Levrat, and Benoît Iung. (2013). Remaining Useful Life Estimation Based on Stochastic Deterioration Models: A Comparative Study. *Reliability Engineering and System Safety* 112: 165–75.

Susto, Gian Antonio, Andrea Schirru, Simone Pampuri, Alessandro Beghi, and Giuseppe De Nicolao. (2018). A Hidden-Gamma Model-Based Filtering and Prediction Approach for Monotonic Health Factors in Manufacturing. *Control Engineering Practice* 74 (February). Elsevier Ltd: 84–94.

Tsui, Kwok L, Nan Chen, Qiang Zhou, Yizhen Hai, and Wenbin Wang. (2014). Prognostics and Health Management : A Review on Data Driven Approaches. *Mathematical Problems in Engineering., Hindawi Publishing Corporation* 2015.

Zhang, Zhengxin, Xiaosheng Si, Changhua Hu, and Yaguo Lei. (2018). Degradation Data Analysis and Remaining Useful Life Estimation: A Review on Wiener-Process-Based Methods. *European Journal of Operational Research* 271 (3). Elsevier B.V.: 775–96.

Zhao, Zeqi, Bin Liang, Xueqian Wang, and Weining Lu. (2017). Remaining Useful Life Prediction of Aircraft Engine Based on Degradation Pattern Learning. *Reliability Engineering & System Safety* 164 (457). Elsevier Ltd: 74–83.