Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P113-cd

Degradation Process Modeling based on Reliability Test and Machine Learning Regression

Quoc Tiep La

Faculty of Military Technology, University of Defense, the Czech Republic. E-mail: quoctiep.la @unob.cz

Zdenek Vintr,

Faculty of Military Technology, University of Defense, the Czech Republic. E-mail: zdenek.vintr @unob.cz

David Valis

Faculty of Military Technology, University of Defense, the Czech Republic. E-mail: david.valis@unob.cz

Anh Dung Hoang

Military Institute of Mechanical Engineering, Vietnam. E-mail: hoanganhdung1982@gmail.com

A degradation process is the deterioration of an object's internal and external properties, resulting in a decline in performance quality and ability to meet design and operational requirements. Modelling the degradation process has received significant attention from reliability and statistical scientists. Many methods have been proposed and developed to model the degradation process and also to predict and estimate reliability measures. The regression methods are essential methods in machine learning, and it has proved the huge potential for modelling the deterioration process of the objects. In essence, machine learning regression is a concept that represents a series of methods based on i) supervised learning and on collecting data from actual object operations or on ii) reliability tests in laboratories. This paper presents general knowledge of machine-learning regression and uses them to model the degradation process of light-emitting diodes (LEDs) based on the data obtained from reliability tests in laboratories in two specific cases of input data. Afterwards, the paper compares the performance of these methods and assesses the suitability and effectiveness of these methods for modelling the degradation process of LEDs.

Keywords: Light-emitting diode, degradation process, machine learning regression, reliability testing, support vector regression, Gaussian process regression, k-nearest neighbors, random forest.

1. Introduction

Technical systems have become an integral part of human society to solve demands in the industry, research and all aspects of human life. The designed systems with purposes could replace humans in implementing some challenging, complex tasks, so the systems need operation with high performance, high accuracy and high reliability. For this reason, reliability research, Prognosis and Management of Health (PHM) of technical equipment and systems are essential besides researching, designing and developing products, and become a science field which has attracted the attention of many researchers. Researchers in this field have to deal with many existing problems such as research on failure characteristics, degradation process, Remaining Useful Life (RUL), maintenance and repair decisions, designing failure prevention methods, etc.

Degradation is a process in which quality, lifetime and abilities to address design and operation requirements of equipment and systems degrade over time under the affection of internal and external conditions. Understanding the degradation process trend is an essential aim of engineers and researchers to make accurate predictions of reliability and lifespan. However, the developments of the degradation are much more complex and much more challenging to describe and predict, especially for systems with many elements and subsystems. In recent decades, many approaches to modelling the

process have been proposed and proven their performances in solving the issue.

Many authors used physics-based models, which are based on a deep understanding of research systems/subsystems to model their degradation process (Samareh-Mousavi and Taheri-Behrooz, 2020; Volkov, et al., 2015; Fukuda and Morita, 2017; Tsai et al., 2022). Some other authors used data-driven approaches, which are based on observed data from experiments in the laboratories or the actual operation of objects for their studies (Sharifzadeh et al., 2019; Lim et al., 2022; Zhang, et al., 2018; Ma, et al., 2019).

Regression methods, which a class of data-driven approaches, are the simplest, most straightforward methods to model and predict degradation. The aim of methods is to attempt the relationship between dependent variables (output variables) and one or more independent variables (input variables) to catch the development of the degradation trend.

As data-driven methods, the regression methods also have inherent drawbacks, which depend on the quality and quantity of data collected (Xiao-Shen, et al., 2011). For proper outcomes the time for collecting data shall typically be long period of time, therefore updated ability of the model is not guaranteed. In addition, traditional regression methods do not have high accuracy, depending on the ability of the operator as many test steps must be performed before giving the final result.

The evolution of sensor technology, data acquisition technology, and semi-natural simulation technology has made it possible to perform reliability tests in the laboratory with high accuracy, performance and effectiveness. Obtained data can be efficiently processed using advanced algorithms such as machine learning, deep learning, etc. Therefore, using regression machine learning based on data from reliability tests can be an effective approach to model degradation processes (Sharifzadeh et al., 2019; Lim et al., 2022).

In this paper, the authors focus on studying machine learning regression to model the degradation process of light-emitting diodes (LEDs), which are a type of equipment with many applications in engineering and industry based on data collected from reliability testing in the laboratory. Simultaneously, the paper attempts to assess the performance of the methods for this object.

2. Machine Learning Regression

Machine Learning Regression (MLR) is a class of data-driven regression methods. It is a subfield of machine learning that involves predicting a continuous target variable based on one or more input features. It is a supervised learning technique that involves learning a mapping function between input and target features, so that given a new set of input features, the model can predict the corresponding target value. MLR methods can be divided into static MLR methods such as Linear Regression, Decision Trees, Neural Networks Regression (NNR), etc., and dynamic MLR methods such as Support Vector Regression (SVR), Gaussian Process Regression (GPR), and Random Forest Regression (RFR). In this section, the paper presents some selected MLR methods to modelling the degradation process of LEDs.

2.1. *Support Vector Regression*

SVR is a type of regression analysis in machine learning that uses support vectors to modelling and making predictions. The goal of the method is to find a function that best fits the data and has the smallest error that is not limited to a linear function but can be non-linear.

SVR uses support vectors to model non-linear functions and makes predictions, intending to find the optimal hyperplane to maximize the amplitude between the positive layer (observed data) and the negative layer (error). SVR is a flexible and powerful regression algorithm that can handle non-linear and multi-dimensional data, it is robust to outliers, and can achieve good generalization performance. Some authors (Lu and Wang, 2020; Qian et al., 2013; Sharifzadeh et al., 2019) used this method to model the degradation of technical objects.

In SVR, a fixed number of mappings is used to map input $\mathbf{x}$ to m-dimensional feature space, and then a linear model is built in this feature space. The linear model $f(\mathbf{x},w)$ can be given by:

$$f(\mathbf{x},w)=\sum_{j=1}^{m} w_j g_j(\mathbf{x})+b \tag{1}$$

where $g_j(\mathbf{x})$ represents a set of non-linear transformations, b is bias and w_j are model parameters. The estimation quality is measured by loss function $L(y, f(\mathbf{x},w))$ or ε-intensive loss function. proposed by Vapnik and Chervonenkis (1974). The aim of SVR is to minimize the

function as follows (Yang et al., 2016):

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \quad (2)$$

where C is penalty factor, $\xi_i, \xi_i^*, i = 1,..,n$ are slack variables and ε is radius of a "tube" around the value of y to measure deviations of training samples outside the tube.

The estimation accuracy of SVR depends on the setting of parameters C, ε and the kernel function parameters. SVR training involves finding an optimal hyperplane in a high-dimensional feature space that maximizes the margin between the predicted values and the actual target values. The training and adjust parameter process can be summarized in the following steps:

- Data is processed by normalizing or scaling the input features and target variable.
- A kernel function is selected to transform the input variables into a higher-dimensional space where the separation of data points becomes easier.
- SVR initializes the model by defining a hyperplane with an initial set of parameters.
- SVR training involves solving an optimization problem to find the optimal hyperplane that maximizes the margin while minimizing the errors between the predicted and actual values.
- SVR adjusts its parameters through an iterative optimization process.
- Once the training process converges, the trained SVR model is evaluated using evaluation metrics.

The process of parameter adjustment in SVR can be computationally intensive, especially for large datasets or complex problems.

2.2. *Gaussian Process Regression*

Gaussian Process Regression (GPR) is a non-parametric regression analysis method in machine learning based on Gaussian process to construct model and make prediction. Based on a set of observed data points, GPR uses Bayesian inference to make predictions about the values of the target variable at new, unobserved locations.

The predictions are made by computing the posterior distribution over the target variable given the observed data and the choice of mean and covariance functions. GPR is a robust and flexible regression algorithm that can adapt to many types of data, provides estimates of uncertainty, and is based on a Bayesian framework, making it suitable for many applications. Jamei et al. (2021), Baraldi et al. (2015); Lim et al., 2022 and Sharifzadeh et al. (2019) used GPR to model and predict degradation process in their studies.

Training and adjusting parameters in GPR involve the following steps:

- Data is processed by normalizing or scaling the input features and target variable.
- In GPR, the choice of the kernel function, also known as the covariance function or similarity function, is crucial. The kernel determines the shape and characteristics of the Gaussian process
- GPR initializes the model by defining a Gaussian process prior, which represents the belief about the underlying function before seeing any data.
- GPR updates the model parameters by maximizing the likelihood of the observed data. The likelihood quantifies the probability of the observed target values given the input variables and the model's parameters.
- GPR employs an optimization algorithm to find the optimal values for the hyperparameters of the kernel function.
- Once the training process is completed, the trained GPR model can be used to make predictions on new input data.

In summary, GPR trains and adjusts its parameters by optimizing the likelihood of the observed data, updating the hyperparameters of the kernel function.

2.3. *Random Forest Regression*

Random Forest Regression (RFR) is a type of ensemble machine learning technique that constructs multiple decision trees and combines their predictions to make the final prediction. RFR is a robust, flexible regression algorithm that can process complex relationships between input features and output variables. This method has the ability to extend plus to help reduce the variance of prediction and to improve performance generalizing, making it a useful technique for solving regression problems. However, like SVR the RFR regression may require more data and computation time compared to other dynamic regression methods, especially when dealing with high-dimensional time series data. Jamei et al.

(2021) and Lu and Wang (2020) used and developed RFR method to model degradation process of selected mechanical and semiconductor items.

The training and parameter adjustment process involves the following steps:

- Preprocess data by handling missing values, encoding categorical variables, and scaling or normalizing the input features.
- RFR creates an ensemble of decision trees. The number of trees is an important parameter to consider.
- RFR builds each decision tree in the ensemble using a random subset of the training data. This process is called bootstrapping or bagging.
- RFR further introduces randomness by selecting a random subset of features at each split point of a decision tree. This process is called feature subsampling.
- Each decision tree in the RF is trained on the bootstrapped subset of data and the randomly selected subset of features. The decision trees are built recursively by selecting the best feature and split point at each node based on a given criterion. The process continues until a stopping criterion is met.
- After training all the decision trees, the predictions from each tree are aggregated to obtain the final prediction.
- RFR offers additional parameters to adjust. These include the maximum depth of the decision trees, minimum samples required to split an internal node, maximum number of leaf nodes, and more.
- Once the RFR model is trained, it can be evaluated on validation or test data using appropriate metrics.

By adjusting the parameters and controlling randomness, Random Forest Regression can effectively balance bias and variance and provide accurate predictions.

2.4. *K-Nearest Neighbors*

K-Nearest Neighbors (KNN) is a simple, instance-based, lazy learning algorithm used for both classification and regression tasks. It is a non-parametric method that stores all available cases and classifies new cases based on a similarity measure.

In KNN, the output is a class membership, an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. The idea is that similar instances tend to belong to the same class, so the algorithm classifies an instance based on its k nearest neighbors. KNN has a simple and intuitive learning process, and it is computationally efficient for small datasets. However, it can be computationally expensive when dealing with large datasets and multidimensional data. In addition, the choice of the number of nearest neighbors k, and the distance metric used for computing similarity are important parameters that have to be carefully selected for each specific problem.

KNN regression does not involve explicit parameter adjustment or training in the traditional sense, as it does not learn a model with adjustable parameters. Instead, it uses the training data as the model itself. The KNN algorithm can be considered:

- The parameter k represents the number of nearest neighbors used to make predictions.
- KNN regression relies on a distance metric to determine the similarity between instances in the feature space.
- Before applying KNN regression, it is often beneficial to preprocess and scale the input features.
- The choice of weighting scheme depends on the problem, and popular options include inverse distance weighting or kernel-based weighting.

2.5. Performance assessment

In order to accurately assess the effectiveness of the regression methods, metrics have been proposed to quantitatively evaluate the effectiveness. Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2), and Adjusted R-squared are the principal metrics which can be used to compare different regression models and choose the best one for a given problem (Plevris et al., 2022). In the study, Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Correlation Coefficient (R) are chosen to assess performance of regression methods are given by (Lewis, 1982):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{3}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{4}$$

$$R = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (5)$$

where $\hat{y}$ are estimation values of output variable, y are true values of output variables.

3. Reliability Testing Details and Data processing

The LEDs voltage degradation datasets used in this paper were obtained from Cu (2019). The authors used 10 LEDs 10W to conduct ADTs under conditions with a current of 1050 mA and temperature of 60 °C (in 1080 hours), 70 °C (in 840 h), 80 °C (in 480 h) and 90 °C (in 240 h).

All LEDs were placed in climate chamber to keep experiment conditions during performing. The Voltage data from each LED was systematically recorded every 10 minutes. In our work we have only used the data of LEDs in the first period of the experiment (under the temperature of 60 °C in 1080 h) with acceleration factor A_F = 5.6 to be the observed data for the study. The dataset included 5439 observed points over time (corresponding with 1080 test hours), representing the changes in LEDs voltage.

The purpose of study is only compare capacity of some MLR methods in modeling degradation process of LEDs. Thus, we chose two cases to access: (i) access the performance of methods in Time Series (data of LED1 is chosen), (ii) access the performance of methods in Time Series and use data of another LED as adding input (using data of LED2 is as an adding input).

Graph of Voltage Degradation Trend of LED1 and LED2 are shown in Fig.1.

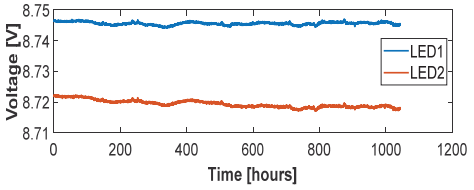

Fig. 1. The Voltage Degradation Trend of LEDs.

To preparing data, correlation coefficient between data of LEDs is accessed. The correlation coefficient between data of chosen LEDs is 0.95. It means that the data of LED2 can be used to train data LED1.

To assess the performance of each method, the observed dataset is divided into three sub-datasets, namely "Data Training" with 3107 points, "Data Test" with 1331 points, "Data validation" with 1001 points.

4. Results and discussion

In this section, the GPR, SVR, KNN, RFR models are used as methods to perform regression analysis of LED degradation based on observed dataset from ADTs in Cu (2019). The study presents the following two cases as shown in Section.3. The regression results using Machine Learning Regression Methods are shown in Fig. 2 to Fig. 6 and Tables.1, 2 present metrics to assess the performance of all used methods.

Three graphs are shown in each figure (Fig.2 to Fig.6). In Fig.3 to Fig.6, the 1st graph is a comparison graph between measured and regression data in the Training Stage, Test Stage and Validation Stage in all cases. The 2nd and 3rd graphs are the Error Distribution of the models in case 1 and case 2. According to the simulation results, in two cases, all methods have relatively high prediction performance, indicated by low RMSE, MAPE and high correlative coefficient R (see Table.1, 2 and Fig.2-6), and the performance of these methods in case 1 seem to be better than case 2 in Training Stage and Test Stage.

However, in the model validation phase, some differences begin to emerge. In particular, the values of the correlation coefficients for all the methods used have dropped significantly. The performance of these methods in case 2 is much better than in case 1 (see Table.1, 2).

In case 1, the correlation coefficient of these methods is small, indicating that the correlation between the target variables and the regression results is not good. It shows that these methods are inappropriate in modelling the degradation of LEDs in only Time Series.

In case 2, the methods have good performance in all stages, except for SVR in Validation Stage (See Table.2). Therefore, they have the ability to model the degradation process if using more input instead of only on Time Series.

In addition, RF, GPR and KNN are more suitable for solving the problems in chosen cases and can be used to model the degradation process of LED relying on observed data from reliability testing. However, they depend on observed datasets and chosen independent variables to describe the dependent variable.

Table 1. Comparison of performance metrics of four methods in case 1.

Statistical criteria		GPR	SVR	KNN	RF
Training stage	*RMSE*	0.00019	0.00040	0.00023	**0.00013**
	MAPE	0.0018	0.0040	0.0021	**0.0013**
	R	0.98147	0.95820	0.97901	**0.99202**
Testing stage	*RMSE*	0.00015	0.00031	0.00018	**0.00013**
	MAPE	0.0011	0.0024	0.0013	**0.0010**
	R	**0.96679**	0.93711	0.94870	0.95155
Validation Stage	*RMSE*	0.00028	**0.0011**	0.00022	0.00022
	MAPE	0.0018	0.0088	**0.0015**	**0.0015**
	R	**-0.31111**	**-0.21488**	**1.79e-29**	**1.79e-29**

Table 2. Comparison of performance metrics of four methods in case 2

Statistical criteria		GPR	SVR	KNN	RF
Training stage	*RMSE*	0.00023	0.00046	0.00026	**0.00015**
	MAPE	0.002	0.0042	0.0022	**0.0014**
	R	0.98836	0.95933	0.98515	**0.99445**
Testing stage	*RMSE*	**0.00031**	0.00053	0.00041	0.00038
	MAPE	**0.0027**	0.0047	0.0035	0.0032
	R	0.98993	0.96408	0.98578	**0.99134**
Validation Stage	*RMSE*	0.00095	**0.00012**	0.00038	0.00043
	MAPE	0.0070	0.0094	**0.0024**	0.0028
	R	0.7724	**-0.20339**	0.81342	**0.86301**

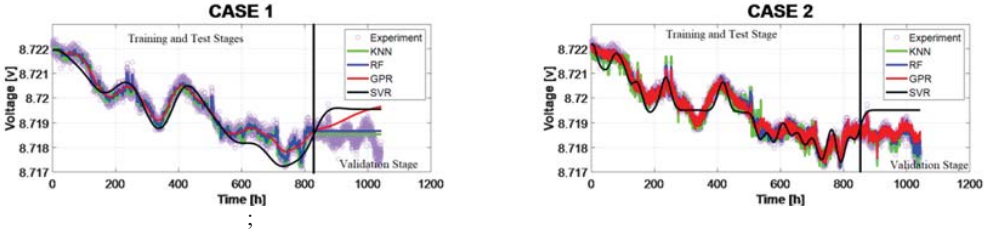

Fig.2. Compare regression results using SVR, GPR, KNN and RF in two cases

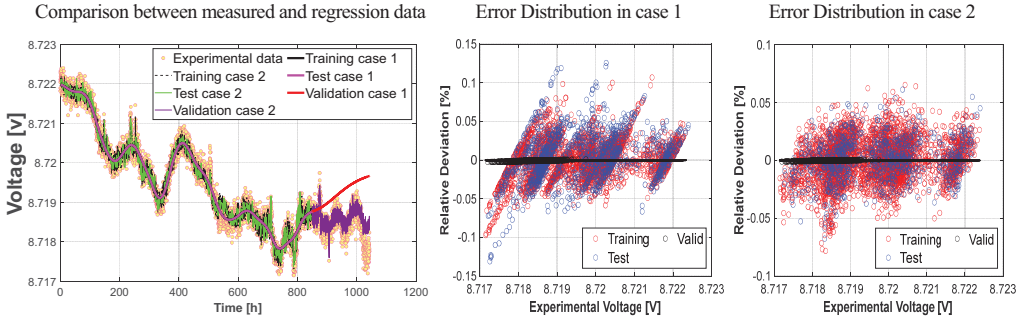

Fig. 3. Regression results of the LED Voltage degradation process using GPR.

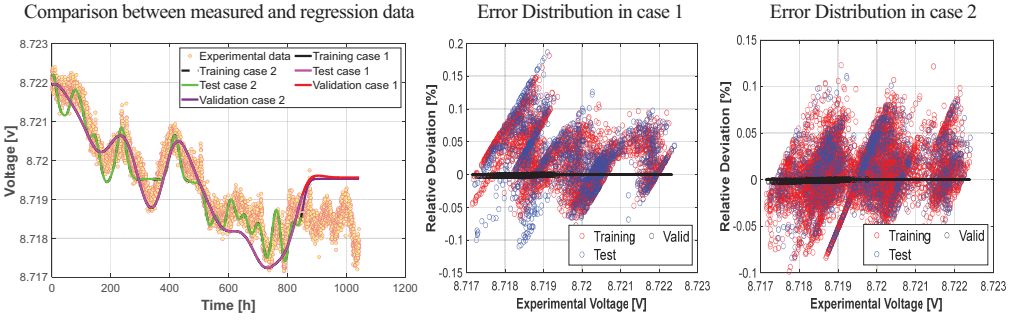

Fig. 4. Regression results of the LED Voltage degradation process using SVR.

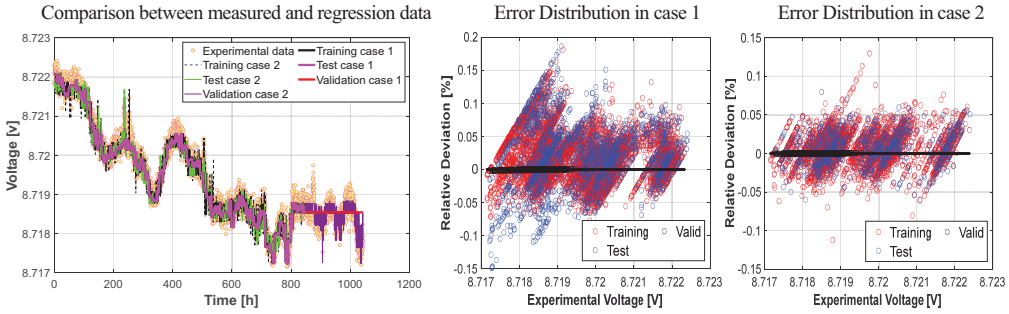

Fig. 5. Regression results of the LED Voltage degradation process using KN.

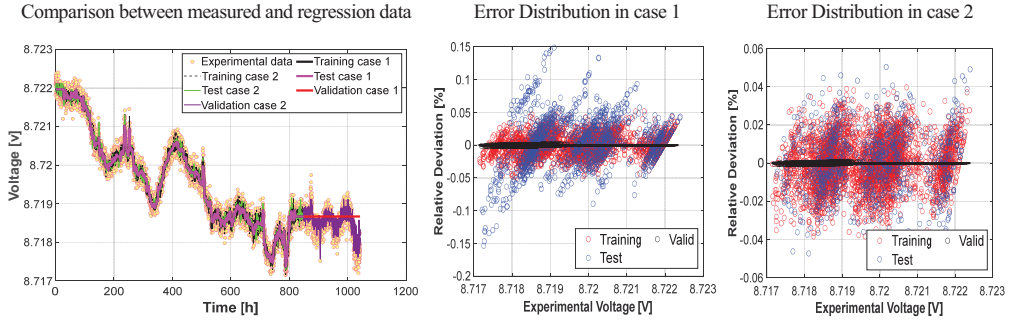

Figure 6. Regression results of the LED Voltage degradation process using RF.

5. Conclusion

The paper has attempted to present a summary of the machine-learning regression methods and applied these methods to model the degradation of light-emitting diodes based on observed data from reliability testing in the laboratory. According to the results represented in Section 4, some methods show the ability and potential to model the degradation processes of LEDs.

Since we have not assessed all cases of each method in this paper therefore choosing a method which is the most suitable to solve this problem is not correct because of not enough evidence. Thus, some problems need to continue to be discussed: (i) consider the complete set of cases with different regression methodologies instead of selected cases for comprehensive and precise assessment; (ii) compare the performance of these methods with other methods; (iii) combine regression methods with other methods to model the degradation process of LEDs instead of using only regression methods; (iv) use models to predict the degradation trend and estimate Remaining Useful Life of systems/subsystems in the future.

Acknowledgement

This paper has been prepared with the support of the Ministry of Defence of the Czech Republic – Project for the Development of the Organization DZRO VAROPS and Specific Research Project SV22-202.

References

Samareh-Mousavi, S.S. and Taheri-Behrooz, F. (2020). A novel creep-fatigue stiffness degradation model for composite materials. *Composite Structures, 237*, 11955.

Volkov, I.A. et al. (2015). Assessment of the service life of structural steels by using degradation models with allowance for fatigue and creep of the material. *Journal of Applied Mechanics and Technical Physics*, *56(6)*, pp. 995-1006.

Fukuda, K and Morita, T. (2017). Physical model of adhesive wear in early stage of sliding. *Wear*, *376(B)*, pp. 1528-1533.

Lu, C. and Wang, S (2020). Performance Degradation Prediction Based on a Gaussian Mixture Model and Optimized Support Vector Regression for an Aviation Piston Pump. *Sensors*, *20(14)*, 3854.

Qian, Y., Hu, S. and Yan, R. (2013). Bearing performance degradation evaluation using recurrence quantification analysis and auto-

regression model. In *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC),* pp. 1713-1716. IEEE.

Sharifzadeha, M., Sikinioti-Lock, A. and Shah, N. (2019). Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. *Renewable and Sustainable Energy Reviews, 108*, pp. 513-538.

Lim, S.L.H, Duong, P.L.T., Park, H. and Raghavan, N. (2022). Expedient validation of LED reliability with anomaly detection through multi-output Gaussian process regression. *Microelectronics Reliability, 138,* 114624.

Baraldi, P., Mangili, F. and Zio, E. (2015). A prognostics approach to nuclear component degradation modeling based on Gaussian Process Regression. *Progress in Nuclear Energy, 78,* pp.141 – 154.

Jamei, M., Ahmadianfar, I., Olumegbon, I.A., Karbasi, M. and Asadi, A. (2021). On the assessment of specific heat capacity of nanofluids for solar energy applications: Application of Gaussian process regression (GPR) approach, *Journal of Energy Storage, 33,* 102067.

Cu, X.P. (2019). Accelerated reliability testing of electronic parts of combat vehicles. Dissertation Thesis. University of Defense in Brno.

Rasmussen, C.E. and Williams, C.K. (2005). *Gaussian Processes for Machine Learning.* MIT Press.

Lewis, C.D. (1982). *Industrial and Business Forecasting Methods: a Practical Guide to Exponential Smoothing and Curve Fitting.* Butterworth-Heinemann.

Tsai, Y.C., Leburton, J.P. and Bayram, C. (2022). Quenching of the Efficiency Droop in Cubic Phase InGaAlN Light-Emitting Diodes. *IEEE Transactions on Electron Devices*, *69,* pp. 3240-3245.

Wang F.K., Huang, C.Y, Mamo, T. and Cheng, X.B. (2020). Ensemble model for the degradation prediction of proton exchange membrane fuel cell stacks. *Quality and Reliability Engineering International, 37*, pp. 34-46.

Zhang, Z. Si, X, Hu, C. and Lei, Y. (2018). Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods, *European Journal of Operational Research 271(3)*, pp.775-796.

Ma, Z. Wang, S. Liao, H. and Zhang, C. (2019). Engineering-driven performance degradation analysis of hydraulic piston pump based on the Inverse Gaussian process. *Quality and Reliability Engineering International, 35(7),* pp. 2278-2296.

Yang, C. Liu, J. Zeng, Y. and Xie, G. (2016). Prediction of components degradation using support vector regression with optimized parameters. *Energy Procedia, 127*, pp. 284–290

Plevris V., Solorzano, G., Bakas, N. and Seghier, M.B. (2022). Investigation of performance metrics in regression analysis and machine learning-based prediction models. In *The 8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2022),* pp. 1-25. Scipedia.

Vapnik, V.N. and Chervonenkis, A.Y. (1974). *Theory of pattern recognition*, Nauka.

Xiao-Shen, S, Wenbin, W, Chang-Hua, H, Dong-Hua, Z (2011). Remaining useful life estimation-A review on the statistical data driven approaches. *European Journal of Operational Research*, 213(1), pp. 1-14.