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

doi: 10.3850/978-981-18-8071-1_P592-cd

Datasets envelope impact on marine engines prognostics and health management models accuracy

Jaehan Jeon

Maritime Safety Research Centre, Department of Naval Architecture, Ocean, and Marine Engineering, University of Strathclyde, Glasgow, G4 0LZ, United Kingdom. E-mail: jaehan.jeon@strath.ac.uk

Gerasimos Theotokatos

Maritime Safety Research Centre, Department of Naval Architecture, Ocean, and Marine Engineering, University of Strathclyde, Glasgow, G4 0LZ, United Kingdom. E-mail: gerasimos.theotokatos@strath.ac.uk

The machinery health management system required for developing maritime autonomous surface ships can be realised by employing prognostics and health management (PHM) methods. Pertinent PHM models are typically trained by using datasets corresponding to limited operating conditions and are subsequently employed to analyse a wide envelope of conditions. This study employs a PHM model that consists of a Deep Neural Network (DNN) submodel and an Auto-Regressive Integrated Moving Average (ARIMA) submodel for predicting the health indicator of a marine four-stroke engine. In specific, this study aims to quantify the accuracy of this PHM model predictions. The PHM model is developed by employing limited datasets and subsequently validated by employing extended datasets. The extended datasets reflect practical operating conditions including ambient temperature variations, stochastic degradation trends, several engine loads, and multiple simultaneous degradations. The results demonstrate that, when the testing dataset is employed, the PHM model predicts the engine exhaust valve health indicator for future time slices with high accuracy of R-squared values of 0.998. However, the model accuracy deteriorated reaching R-squared values of 0.707 when validation datasets representing extended operating envelope are used. This study's results emphasise that the PHM model accuracy is affected by the available datasets for training, necessitating the generation of trustworthy datasets and scientific methods for developing trustworthy PHM models.

Keywords: Marine Engine, Exhaust Valve, Degradation, Prognostics and Health Management, DNN, ARIMA, Health Indicator, Operating Envelope

1. Introduction

Maritime Autonomous Surface Ships (MASS) are expected to address current challenges of the waterborne transportation. However, it is widely acknowledged that commercial MASSs' operation is a medium- to long-term target. The machinery health monitoring and management without a human in the loop must be developed to enable autonomous shipping. An increasing number of research studies employed Prognostics and Health Management (PHM) with data-driven approaches for assessing ship machinery health conditions Zhang et al. (2022). However, acquiring appropriate datasets to develop the PHM model is a persisting challenge Saxena et al. (2008).

Pertinent maritime PHM studies typically acquire datasets from shipboard measurements, experimental data, released public data, or simulations Lei et al. (2018). Cheliotis et al. (2020) employed shipboard measurements to develop an early detection model of growing faults. However, because the measurement data represented only healthy operating conditions, the fault detection model employed simulated datasets with linear degradation to the engine manufacturer's fault limit. Han et al. (2021) collected marine diesel engine run-to-fail data from laboratory experiments and developed prognostic models. The experiments assumed two operating profiles, however, they characterised only limited conditions in a controlled environment. Oikonomou et al. (2019) employed the datasets reported in Coraddu et al. (2016); Cipollini et al. (2018) to develop real-time condition monitoring system of ship machinery. This labelled dataset represented different conditions of a ship propulsion plant, however, it replicated fouling degradations with discretised coefficients in limited ranges of 5% and 2.5% Coraddu et al. (2016). Coraddu et al. (2021) employed a marine dual fuel engine model considering air cooler degradation to generate datasets, which were used to test fault detection algorithms. However, the degradation was approximated by varying the model input parameters.

All the preceding studies achieved high accuracy for the developed PHM models. However, in practical situations, these models are expected to be employed for assessing the machinery health conditions under a wide operating envelope with several disturbances. The results of the tests in the limited envelope cannot guarantee the accuracy of the trained model in the extended operating envelope. Validation tests in a wide envelope of operating conditions are required to benchmark the PHM model's accuracy. This study aims at quantifying the accuracy of a trained PHM model by employing extended datasets that replicate probable operating conditions.

2. Framework

The framework for quantifying the accuracy of the trained PHM model in extended envelopes is illustrated in the flowchart shown in Figure 1. The simulation-based data generation approach creates limited datasets and extended datasets corresponding to different operating envelopes. The PHM model is trained by using a limited dataset that reflects a controlled environment. The trained PHM model is validated with the extended datasets representing a wide envelope of operating conditions encountered in the machinery lifetime. The ultimate output is the quantification of the trained PHM model accuracy considering the extended envelope.

3. Method

3.1. *Simulation-based Data Generation*

The datasets used in this study are generated using the digital twin of a marine four-stroke engine. The digital twin is customised based on the previous author's studies Stoumpos et al. (2018,

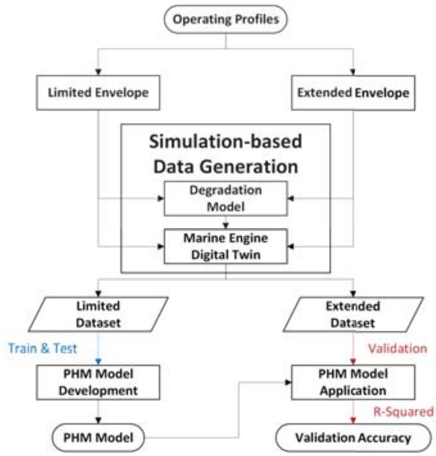

Fig. 1. Framework for quantifying the validation accuracy of the trained PHM model

2020), and the trustworthiness of the DT and the generated datasets are addressed in the previous author's study Jeon and Theotokatos (2023). Each dataset contains 500 samples; each sample consists of 8 features (the engine component running hours, and the following engine performance parameters: engine power, turbocharger shaft speed, exhaust gas temperature upstream turbine, maximum cylinder pressure, charge air pressure and temperature, as well as fuel oil consumption).

3.2. *Degradation Sub-Model*

The study employs degradation sub-models by combining an empirical degradation model and a stochastic process model. The empirical model is developed using the degradation mechanism and verified through laboratory experiments. However, it is unable to represent environmental uncertainty Rui et al. (2020). To simulate degradation by considering both physical behaviour and uncertainty, this study integrates the empirical model with the stochastic process model.

For the empirical model, the valve recession model reported in Lewis and Dwyer-Joyce (2001) is employed. The model was derived from laboratory experiments and describes the valve wear based on Eq. (1) and Eq. (2). The original model considered both the sliding and impact wear, however, this study considers only the impact wear since the effect of sliding wear is negligible on the valves with high closing velocities, such as the large marine engines Slatter et al. (2006).

$$V = K\ N\ e^n \left(\frac{A_i}{A}\right)^j \tag{1}$$

V denotes the wear volume, K is the impact wear constant, N is the number of cycles, e is the energy on impact, n is the impact wear constant, A_i is the valve's initial contact area, A is the contact area after N cycles, and j is the wear constant.

$$r = \left(\sqrt{\frac{V}{\pi R_i cos\theta_s sin\theta_s} + w_i^2} - w_i\right) sin\theta_s \tag{2}$$

r denotes the valve recession, R_i is the initial seat insert radius, θ_s is the seat insert seating face angle, and w_i is the initial seat insert seating face width.

The stochastic wear sub-model is based on the gamma process model that employs Eq. (3) van Noortwijk (2009). The gamma process is irreversible and continuous Rui et al. (2020); Shahraki et al. (2017), hence it can represent the wear behaviour.

$$Ga(x|v,u) = \frac{u^v}{\Gamma(v)} x^{v-1} exp\{-ux\} I_{(0,\infty)}(x) \tag{3}$$

Ga denotes the gamma process probability density function. $I_A(x) = 1$ for $x \in A$ and $I_A(x) = 0$ for $x \notin A$, and $\Gamma(a) = \int_{z=0}^{\infty} z^{a-1} e^{-z} dz$ for $a > 0$. The shape parameter $v(t)$ is a non-decreasing and right continuous function of time t.

3.3. *PHM Model*

The employed PHM model predicts the Health Indicator (HI) using as input the engine performance parameters and engine component(s) running hours. The HI ranges between 0 and 1. The HI value 1 denotes the engine component(s) healthy condition, whereas 0 denotes faulty conditions.

The PHM model consists of a Deep Neural Network (DNN) sub-model and an Auto-Regressive Integrated Moving Average (ARIMA) sub-model, as presented in Figure 2. The DNN model estimates engine component(s) HI at past time slices by using historical data as input. The DNN can reduce the number of neurons compared to the shallow neural network to achieve the same accuracy Gökgöz and Filiz (2018). The ARIMA submodel, which is a standard time series forecasting method, predicts the engine component(s) HI at future time slices. The ARIMA submodel combines the autoregressive process and the moving average process to predict near-future data by using time series Siami-Namini et al. (2018).

3.4. *Error Metric*

The accuracy of the PHM model is evaluated using the R-squared (R^2) according to Eq. (4). R^2 exhibits high interpretability compared to other metrics including the mean square error and the mean absolute error Chicco et al. (2021).

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(\hat{Y}_i - Y_i)^2}{\sum_{i=1}^{m}(\bar{Y}_i - Y_i)^2} \tag{4}$$

Where m is the number of samples, Y is the actual value, $\hat{Y}$ is the predicted value, and $\bar{Y}$ is the mean value of actual values.

4. Case Studies

This study considers the degradation of the engine cylinder valve recession. The impact of the dataset's characteristics on the PHM model accuracy is investigated by considering different operating envelopes. The PHM model for training and testing employs a limited operating envelope that includes a degradation trajectory with weak stochasticity, fixed engine load at 75%, and single-component degradation. The trained PHM model is subsequently applied to the three validation cases that consider practical operating conditions, such as degradation trajectories with strong stochasticity, engine load variations, and simultaneous degradation of multiple components. The employed operating envelopes for each case are listed in table 1, whereas the employed health indicator time variations are illustrated in Figure 3.

Fig. 2. PHM Model structure

Table 1. Operating envelopes for case studies

Dataset	Stochasticity	Load (%)	Degradation Component
Training	Weak	75	EV
Testing	Weak	75	EV
Validation 1	Strong	75	EV
Validation 2	Strong	70 - 85	EV
Validation 3	Strong	75	EV & IV

Note: EV: Exhaust Valve, IV: Intake Valve

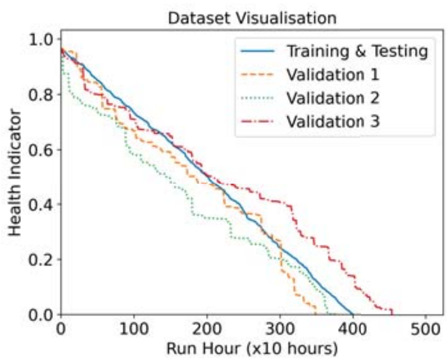

Fig. 3. Dataset Visualisation

4.1. *Testing Dataset - Limited operating envelope*

To quantify the trained PHM model accuracy, the testing dataset in the limited operating envelope was employed. The trained PHM model showed exceptional accuracy with R-squared values of 0.998. The PHM model results are shown in Figure 4.

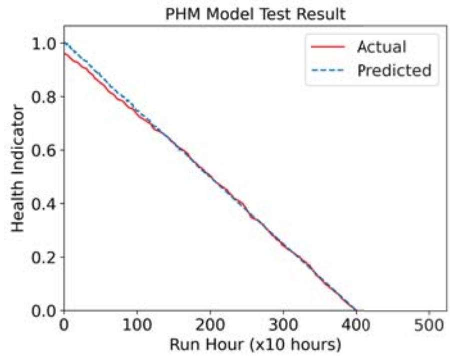

Fig. 4. Test Result for Prediction Model

4.2. *Validation 1 - Strong Stochasticity*

The first validation test was performed to quantify the PHM model accuracy with strong stochasticity datasets. The actual operations are usually exposed to a more stochastic environment compared to laboratory conditions. The operational uncertainties including the quality of the spare parts, environmental conditions, and ship operating mode increase the stochasticity of the component's degradation. The accuracy of the trained PHM model decreased to R-squared values of 0.754 under the strong stochasticity of the degradation pattern. The first validation test results are presented in Figure 5.

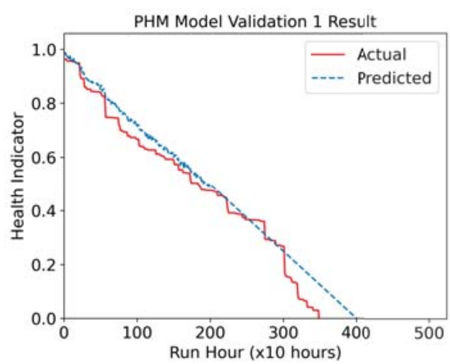

Fig. 5. Validation 1 Result for the trained PHM Model

4.3. *Validation 2 - Load Variations*

The second validation test was performed to quantify the PHM model accuracy considering datasets corresponding to engine load variations. The engine loads typically vary in actual operation corresponding to the ship voyage schedule and required power demand. The accuracy of the trained PHM model was reduced to the R-squared values of 0.855 obtained under the degradation patterns with varying engine loads. The second validation test results are shown in Figure 6.

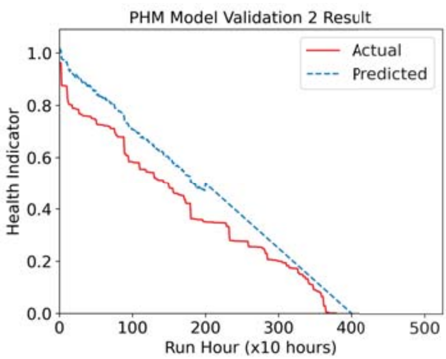

Fig. 6. Validation 2 Result for the trained PHM Model

4.4. *Validation 3 - Multiple Components Degradation*

The third validation test was performed to quantify the PHM model accuracy under multiple components' degradations. Several PHM studies considered limited components into account when training a PHM model, therefore the other components' degradations impact its accuracy. The intake valve was employed as the additional component that was not considered during the training phase. The accuracy of the trained PHM model for predicting the HI of the exhaust valve at future time slices decreased, as the R-squared reduced to 0.707. The third validation test results are shown in Figure 7.

5. Conclusions

This study quantified the accuracy of the trained PHM model for predicting the health indicator

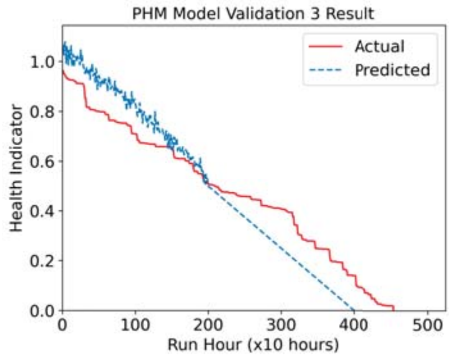

Fig. 7. Validation 3 Result for the trained PHM Model

of a marine engine exhaust valve by employing four different operating envelopes. The validation results verify the impact of the used operating envelope on the accuracy of the PHM model.

The trained PHM model exhibited remarkable accuracy on the testing dataset, which deteriorated when using the extended datasets representing a wide envelope of operating conditions. The accuracy deterioration was the most critical when the dataset considered the simultaneous degradations of multiple components from R-squared values of 0.998 to 0.707.

The study emphasises the significance of operating envelopes in datasets, however, it is challenging to change operating envelopes of measurement datasets like historical data and experimental data due to physical constraints. The simulation-based data generation approach is recommended for future studies to acquire datasets in a wide operating envelope. The study employed only three disturbances for the operating envelopes, however, practical operations have a wider range of operating disturbances. Future study is advised to consider appropriate operating disturbances corresponding to the purpose of a PHM model.

Acknowledgement

The study was partially supported by the AUTOSHIP project funded by the European Union's Horizon 2020 research and innovation programme (agreement No. 815012). The authors greatly acknowledge the funding from DNV AS and RCCL for the MSRC establishment and operation. The opinions expressed herein are those of the authors and should not be construed to reflect the views of EU, DNV AS, and RCCL.

References

Cheliotis, M., I. Lazakis, and G. Theotokatos (2020). Machine learning and data-driven fault detection for ship systems operations. *Ocean Engineering 216*, 107968.

Chicco, D., M. J. Warrens, and G. Jurman (2021). The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science 7*.

Cipollini, F., L. Oneto, A. Coraddu, A. J. Murphy, and D. Anguita (2018). Condition-based maintenance of naval propulsion systems: Data analysis with minimal feedback. *Reliability Engineering System Safety 177*, 12–23.

Coraddu, A., L. Oneto, A. Ghio, S. Savio, D. Anguita, and M. Figari (2016). Machine learning approaches for improving condition-based maintenance of naval propulsion plants. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment 230*, 136–153.

Coraddu, A., L. Oneto, D. Ilardi, S. Stoumpos, and G. Theotokatos (2021). Marine dual fuel engines monitoring in the wild through weakly supervised data analytics. *Engineering Applications of Artificial Intelligence 100*, 104179.

Gökgöz, F. and F. Filiz (2018). Electricity price forecasting: A comparative analysis with shallow-ann and dnn. *International Journal of Energy and Power Engineering 12*(6), 421–425.

Han, P., A. L. Ellefsen, G. Li, V. Æsøy, and H. Zhang (2021). Fault prognostics using lstm networks: Application to marine diesel engine. *IEEE Sensors Journal 21*.

Jeon, J. and G. Theotokatos (2023). A framework for assessing digital twins trustworthiness–application to marine engines. submitted.

Lei, Y., N. Li, L. Guo, N. Li, T. Yan, and J. Lin (2018). Machinery health prognostics: A systematic review from data acquisition to rul prediction. *Mechanical Systems and Signal Processing 104*, 799–834.

Lewis, R. and R. S. Dwyer-Joyce (2001). Design tools for prediction of valve recession and solving valve/seat failure problems. *SAE Transactions*, 1868–1877.

Oikonomou, S., I. Lazakis, and G. Papadakis (2019). An innovative machine learning system for real time condition monitoring of ship machinery. pp. 753–768. Springer.

Rui, K., G. Wenjun, and C. Yunxia (2020). Model-driven degradation modeling approaches: Investigation and review. *Chinese Journal of Aeronautics 33*, 1137–1153.

Saxena, A., K. Goebel, D. Simon, and N. Eklund (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. pp. 1–9. IEEE.

Shahraki, A. F., O. P. Yadav, and H. Liao (2017). A review on degradation modelling and its engineering applications. *International Journal of Performability Engineering 13*, 299.

Siami-Namini, S., N. Tavakoli, and A. S. Namin (2018). A comparison of arima and lstm in forecasting time series. In *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, pp. 1394–1401. IEEE.

Slatter, T., R. Lewis, and R. S. Dwyer-Joyce (2006). Valve recession modelling.

Stoumpos, S., G. Theotokatos, E. Boulougouris, D. Vassalos, I. Lazakis, and G. Livanos (2018). Marine dual fuel engine modelling and parametric investigation of engine settings effect on performance-emissions trade-offs. *Ocean Engineering 157*, 376–386.

Stoumpos, S., G. Theotokatos, C. Mavrelos, and E. Boulougouris (2020). Towards marine dual fuel engines digital twins—integrated modelling of thermodynamic processes and control system functions. *Journal of Marine Science and Engineering 8*, 200.

van Noortwijk, J. M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety 94*(1), 2–21.

Zhang, P., Z. Gao, L. Cao, F. Dong, Y. Zou, K. Wang, Y. Zhang, and P. Sun (2022). Marine systems and equipment prognostics and health management: A systematic review from health condition monitoring to maintenance strategy. *Machines 10*, 72.