Data imputation of missing values from marine systems sensor data. Evaluation, visualisation, and sensor failure detection

C Velasco-Gallego and **I Lazakis**, University of Strathclyde, UK

SUMMARY

To enable Condition-Based maintenance, sensors need to be installed, and thus Internet of Ships (IoS) needs to be implemented. IoS presents several challenges, an example of which is the imputation of missing values. A data assessment imputation framework that is utilised to assess the accuracy of any imputation model is presented. To complement this study, a real-time imputation tool is proposed based on an open-source stack. A case study on a total of 4 machinery systems parameters obtained from sensors installed on a cargo vessel is presented to highlight the implementation of this framework. The multivariate imputation technique is performed by applying Kernel Ridge Regression (KRR). As the explanatory variables may also contain missing values, GA-ARIMA is utilised as the univariate imputation technique. The case study results demonstrate the applicability of the suggested framework in the case of marine machinery systems.

Keywords – data imputation, machine learning, marine machinery systems, condition-based maintenance (CBM), data monitoring, imputation assessment.

NOMENCLATURE

AdaBoost	Adaptive Boosting
ARIMA	Autoregressive Integrated Moving Average
CBM	Condition-Based Maintenance
DAIF	Data Assessment Imputation Framework
DTR	Decision Tree Regression
EWMA	Exponentially Weighted Moving Average
GA	Genetic Algorithm
k-NN	k-Nearest Neighbors
KRR	Kernel Ridge Regression
MAR	Missing at Random
MCAR	Missing Completely at Random
MedAE	Median Absolute Error
MICE	Multiple Imputation by Chained Equations
MNAR	Missing Not at Random
MSE	Mean Squared Error
MSLE	Mean Squared Log Error
NN	Neural Network
OEM	Original Equipment Manufacturers
PLS	Partial Least Squares
RBF	Radial Basis Function
RMSE	Root Mean Square Error
STL	Seasonal and Trend decomposition using Loess
SVR	Support Vector Regression
VAR	Vector autoregression

1. INTRODUCTION

In 2019 alone a total of 2904 casualties and 49 fatalities occurred in relation to maritime transportation. Of all the causes of accidents to ships, 14% refers to damage to ship equipment [1]. Between 75% and 96% of marine accidents are the result of human error [2]. In consideration of climate change concerns, CO_2 emissions are expected to increase between 50% and 250% by 2050 if no measure is implemented [3]. Hence, there is no doubt whatsoever about the improvement threshold that can be implemented within the maritime industry.

In relation to maintenance, the maritime industry is currently considering state-of-the-art maintenance and inspection processes, an example of which is Condition-Based Maintenance (CBM). This is a strategy hinged on the condition monitoring of the asssets [4][5][6][7][8]. To enable this strategy, sensors need to be installed, and thus Internet of Ships (IoS) needs to be implemented. IoS presents several challenges, an example of which is the imputation of missing values [9][10][11].

Data imputation is a compelling pre-processing step, the aim of which is the estimation of identified missing values to avoid under-utilisation of data. Hence, if missing values are not tackled, the results obtained from applying data analysis may be unreliable and inaccurate, which could lead to bias in further steps due to the utilisation of poor models implemented in decision-making processes [12].

Although there are more than 10,000 publications about data imputation in Scopus database, only two publications refer to the maritime industry. This indicates a lack of analysis and formalisation of data imputation in this industrial sector. For this reason, a Data Assessment Imputation Framework (DAIF) is developed to assess the accuracy of any imputation model. Thus, the selection of the imputation approach is not biased by human decisions and its selection is purely objective based on the characteristics and contexts of the data.

In addition, a real-time imputation tool is presented based on an open-source stack, so that any organisation can implement this framework to monitor sensor data, and thus prevent sensor failure. The application of both the DAIF and the real-time imputation tool is highlighted through a case study, in which missing values from the main engine power parameter are imputed.

The following paragraphs are structured as follows. Section 2 presents data imputation studies performed within the maritime industry. Section 3 describes the proposed methodology. Section 4 reflects on the result obtained after implementing the proposed methodology through a case study. Lastly, in Section 5 the conclusions are presented.

2. LITERATURE REVIEW

A total of two articles were identified from Clarivate Analytics Web of Science and Scopus. [14] presented a hybrid imputation method by combining *k*-Nearest Neighbors (*k*-NN) and Multiple Imputation by Chained Equations (MICE) models. A total of three metrics (Absolute Percentage Error (APE), Mean Absolute Percentage Error (MAPE), and the standard deviation of the error) were estimated to compare the proposed framework with *k*-NN and MICE algorithms. A case study based on the imputation of missing values from time-series data collected from a total of 8 sensors coupled to the turbocharger and to the main engine of a chemical tanker was implemented to demonstrate that the proposed methodology outperforms *k*-NN and MICE methods.

[13] performed a comparative study that examined a total of 20 widely implemented machine learning and time series forecasting algorithms (mean imputation, Seasonal and Trend using Loess (STL) decomposition, exponential smoothing, Autoregressive Integrated Moving Average (ARIMA) models, linear regression models (Partial Least Squares (PLS) regression, Least Absolute Shrinkage and Selection Operator (LASSO) regression, Ridge regression, and ElasticNet regression, *k*-NN, Support Vector Regression (SVR) with linear and Radial Basis Function (RBF) kernels, Neural Networks (NNs) with 1, 2, and 3 hidden layers, Vector autoregression (VAR), Decision Tree Regression (DTR), and ensemble methods (bagged trees with SVR, bagged trees with *k*-NN, random forests, and Adaptive Boosting (AdaBoost)). A case study was also implemented on a total of 7 machinery parameters obtained from sensors installed on a cargo vessel to assess their performance. It was concluded that ARIMA outperformed the remaining models.

Although both articles presented new methodologies to impute missing values from sensor data of marine machinery, there are yet several challenges to be tackled within the maritime industry. An example of which is the development of a generic framework to assess the performance of any imputation model, novelty that is introduced in this paper. A list of the novelties presented in this paper in relation to data imputation in the maritime industry is described hereunder.

- The development of a data assessment imputation framework to evaluate the accuracy of any imputation model.
- The implementation of Exponentially Weighted Moving Average (EWMA) as a denoising method in the pre-processing step to enhance the performance of the data imputation model.
- The application of the Kernel Ridge Regression (KRR) as a multivariate data imputation model.
- The employment of GA-ARIMA as a univariate data imputation model.
- The proposal of a real-time imputation tool based on an open-source stack.

3. METHODOLOGY

Having explored the novelties presented in this conference paper, this section presents the proposed framework, which is graphically represented in Figure 1.

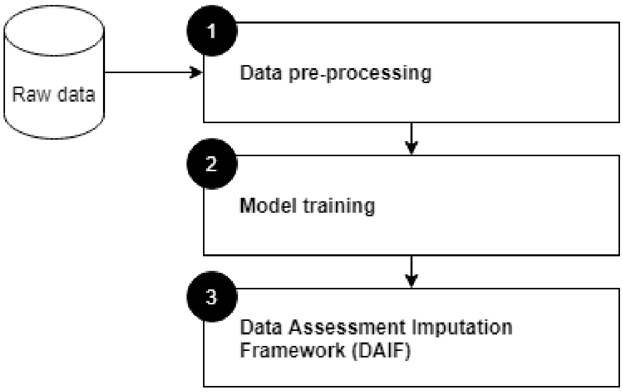


Figure 1. Graphical representation of the proposed methodology.

3.1 DATA PRE-PROCESSING

Data pre-processing is essential to be implemented prior to model training due to the characteristics of the data, as the raw data may not only include steady operational states but also both manoeuvring and transient states of machinery, which need to be excluded from the analysis. To define proper steady operational states Original Equipment Manufacturers (OEMs) of the systems being analysed are consulted.

Subsequently, standardisation is applied so that features can contribute equally when distance-based algorithms are implemented. Correlation analysis is also performed by the estimation of both the Pearson's correlation coefficient and Spearman's rank correlation coefficient to identify linear and non-linear relationships between features. Feature extraction has been performed prior to all these steps and only four parameters of the main engine system of a cargo vessel are included in this study (main engine power, main engine rotational speed, main engine fuel flow rate, and scavenging air pressure of the scavenge air receiver). It is assumed that all occurrences of the sample refer to normal operational conditions.

In addition, as time series data are susceptible to containing high noise, Exponential Weighted Moving Average (EWMA) is implemented as a denoising technique.

3.2 MODEL TRAINING

The model implemented to impute missing values in this study is the Kernel Ridge Regression (KRR), which combines Ridge regression with the kernel tricks. Ridge regression is a biased model, as it adds a penalty to regularise the parameter estimates, and thus apply a trade-off between the bias and the variance to reduce the Mean Squared Error (MSE). Ridge regression considers the addition of a second-order penalty on the parameter estimates. Among all the kernel functions that encompass linear and non-linear functions of the predictors, the Radial Basis Function (RBF) is considered in this study. Hence, two main hyperparameters need to be selected optimally (regularisation strength, α, and gamma parameter for the RBF kernel, γ). To that end, a Genetic Algorithm (GA) is applied, which is a variety of evolutionary algorithms inspired by the process of natural selection.

The selection of these two hyperparameters is essential as an inadequate selection of them may yield either under-fitting or over-fitting of the KRR model, and thus lead to inaccurate imputations. Therefore, time series cross-validation is implemented to select the hyperparameters that best conduct the learning process.

3.3 DATA ASSESSMENT IMPUTATION FRAMEWORK (DAIF)

In general, a total of three different missingness mechanisms have been identified to encompass underlying causes of missing data.

The first mechanism, Missing Completely at Random (MCAR), involves those situations in which the missingness is independent of the data. An example of which is a random failure produced in the fuel flowmeter [14]. To simulate this situation, the following procedure is reproduced:

- n samples with different missing ratios (r_1, r_2, ..., r_m) are generated. Each sample contain values missing completely at random.
- Univariate imputation is performed to impute missing values from predictors' instances. The method applied is GA-ARIMA. GA-ARIMA is a variation of Autoregressive Integrated Moving Average (ARIMA) models, in which GA is utilised to estimate the coefficients.
- Multivariate imputation is performed as indicated in the preceding section to impute missing values from the response variable.

The second mechanism is Missing at Random (RAM), in which the missingness hinges on another feature. For instance, if a component of a main engine fails, the operating condition of dependent components may be altered. This case is simulated by applying the strategy expressed hereunder.

- n large gaps of m size are genereted at random. As a considerable amount of successive occurrences are missing for each feature, it is highly probable that instances contain more than one feature, the occurrences of which are missing.
- Analogous to the MCAR mechanism, univariate imputation is implemented, and thus missing values from predictors' instances are imputed. The technique applied to perform the univariate imputation is GA-ARIMA.
- Multivariate imputation is performed as indicated in the preceding section to impute missing values from the response variable.

The third and final mechanism is established as Missing Not at Random (MNAR) and refers to those scenarios in which the missingness is related to the feature itself. An example of which is the failure of a component, which is unable to record operating data as it is malfunctioning. However, as indicated in the preceding sections, some assumptions have been made to conduct this study, an example of which is the absence of fault data. Thus, this mechanism is out of the scope of this paper.

To evaluate the accuracy of the imputations performed in this section of the methodology, a total of five metrics can be estimated (Root Mean Square Error (RMSE), MSE, max error, Median Absolute Error (MedAE), Mean Squared Log Error (MSLE)). Once the selected metric has been estimated for all the imputation scenarios, the results are pooled to obtain a unique value that summarises the imputation performance of the model being tested.

4. RESULTS

To highlight the implementation of the methodology described above, a case study is presented. To that end, data have been collected from a DMD-MAN B&W 6S50MC-C main propulsion engine of a cargo vessel. Specifically, the main engine power, the main engine rotational speed, the main engine fuel flow rate, and the scavenging air pressure of the scavenge air receiver parameters have been considered for this case study. Time series plots of all four parameters can be observed in Figure 2.

The samples contain a total of 1,200 records, which refer to the steady operational states of machinery. As observed in Figure 2, a total of four steady states are identified in the samples. For instance, if the main engine power is considered, there is a large steady state that initiates at the first instant and persists around 200 minutes where the values are stabilised around 4,200 kW. Subsequently, there is an adjustment where the main engine power decreases to approximately 3,800 kW. This state remains for roughly 700 minutes, and then suddenly an abrupt increase is perceived. This is when the maximum value of the time series is achieved, which presents a value greater than 4,500 kW. This state remains for 200 minutes. Finally, the last abrupt change occurs. This is when the minimum value of the time series is achieved, which presents a value lower than 3,800 kW. Analogous patterns are observed in the remaining parameters being analysed. All the adjustments observed are applied due to the contractual agreements between the charterer and the shipowner in relation to the vessel speed and the fuel oil consumption per day.

Summary statistics of each time series can also be consulted by analysing the descriptive statistics in Table 1. Correlation analysis has also been performed and the resulting Pearson's and Spearman's correlation matrices can be consulted in Tables 2 and 3, respectively. As indicated, all the parameters are linearly correlated.

To finalise the data pre-processing step, denoising has been applied by implementing EWMA to remove the high noise that time-series data contain. The results of which can also be perceived in Figure 2.

The second step of the methodology is model training, in which KRR model is trained to be used as a multivariate imputation model. 5-fold time-series cross-validation technique is applied to prevent either under-fitting or over-fitting. As two main hyperparameters need to be estimated (α and γ), an optimisation algorithm needs to be implemented. For this study, GA is applied. The sets considered to perform GA are expressed hereunder.

$\alpha \in [10^{-10}, \ldots, 1]$
$\gamma \in [10^{-10}, \ldots, 10^{-3}]$

Due to the extent limitations of the present conference paper, only the main engine power is considered for assessing the imputation of missing values, and thus the three remaining parameters (main engine rotational speed, main engine fuel flow rate, and scavenging air pressure of the scavenge air receiver) are only utilised as explanatory variables. However, the presented framework can be used to assess the imputation performance for any parameter.

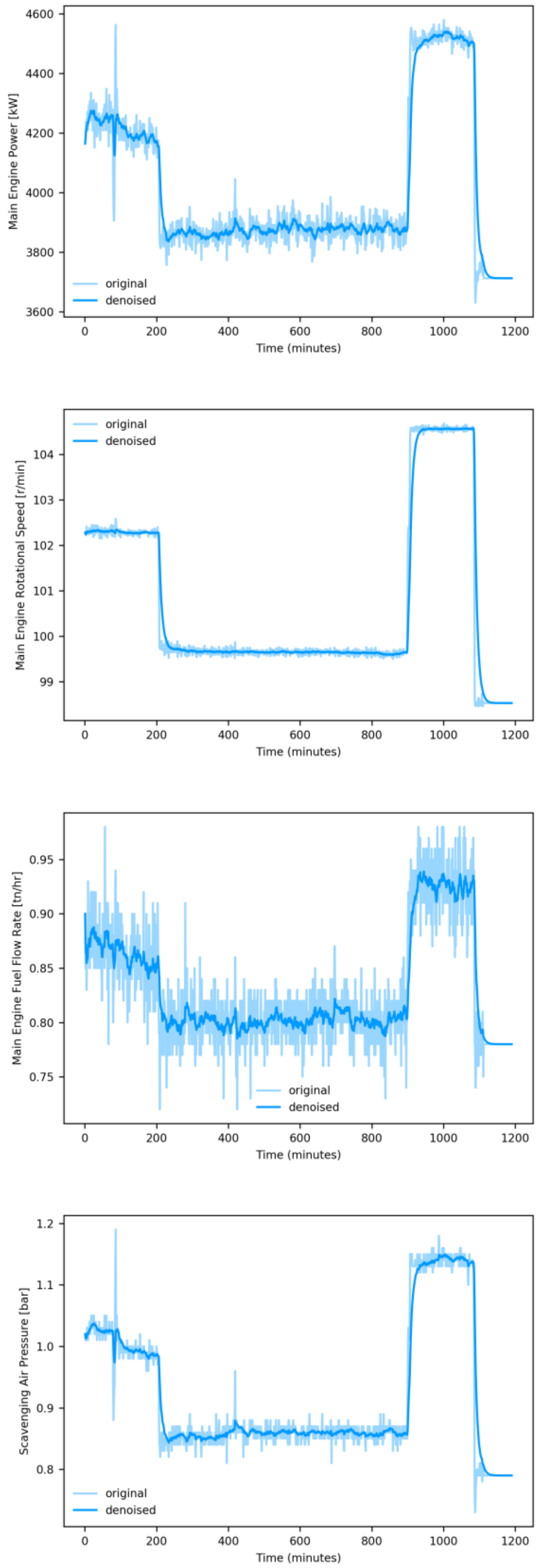


Figure 2. Time series plot of (a) the main engine power, (b) the main engine rotational speed, (c) the main engine fuel flow rate, and (d) the scavenging air pressure of the scavenge air receiver.

To assess the imputation performance of KRR, the data assessment imputation framework is implemented. The ratios considered to generate the MCAR samples are 0.05, 0.15, 0.20, 0.25, and 0.3. In relation to the MAR samples, a total of 5 large gaps, the size which is between 50 and 150 occurrences, are generated at random for each analysed sample. These values are selected based on the characteristics of missingness and contexts perceived in the data collected, as well as suggestions made by other academics who have analysed marine machinery data.

To compare the imputation performance of KRR, GA-ARIMA is also implemented. The criteria utilised to select this model for the comparative study is due to its implementation as a univariate imputation technique to impute missing values from predictors' instances. In addition, ARIMA models presented the best results in the case study performed in [13]. For each model two variations are considered: 1) denoising is implemented in the pre-processing step by the implementation of EWMA, and 2) no denoising is implemented.

MCAR results for each fold can be observed in Table 4. Table 5 presents both large gaps and final results of each implemented model.

As it can be perceived, and as deduced in [13], ARIMA models outperform multivariate imputation techniques when MCAR context is considered. For all MCAR scenarios, ARIMA achieved a RMSE no greater than 30 kW, with the exception of the cases introduced in fold 5, in which the maximum error (406.08 kW, ratio 0.20) is achieved. This increment can be perceived for all four approaches, which indicates that these substantive alterations are related with the characteristics of the time series. As identified in Figure 2, there are a total of four steady states, the last one being the shortest one. This steady state is not considered in the training set of any of the models, and thus the predictions cannot be constructed adequately. This is an indicator of how important it is to prevent failures that lead to the collection of either incorrect or missing values, as this may yield to biased imputations if analogous preceding instances have not been perceived.

Another consideration that can be distinguished is the enhancement of the imputation when denoising is applied. For instance, when this pre-processing step is implemented prior to the imputation of missing values by the utilisation of GA-ARIMA a percentage of improvement of approximately 50% can be observed. This indicates that denoising may need to be applied when a model sensitive to noise is being performed and the time-series data being analysed contain high noise.

In relation to MAR results, KRR leads to better results than GA-ARIMA, as GA-ARIMA cannot predict unexpected events. Thus, GA-ARIMA is unable to detect when an adjustment is introduced due to either the contractual agreements between the charterer and the shipowner or weather conditions. For this reason, multivariate imputation techniques are recommended when the instances with missing values are perceived and when unexpected events occur. This also highlights the importance of preventing failures that lead to the collection of either incorrect or missing values, as biased imputations may be obtained if the unexpected events are not recorded.

Therefore, as indicated in the preceding paragraphs, there is no unique model that outperforms the remaining imputation techniques for all characteristics and contexts. Hence, the implementation of comparative studies and assessment frameworks to determine the best model to impute missing values based on the scenario presented is of paramount importance, as under-utilisation of data can then be prevented. Furthermore, the frameworks applied can be part of a holistic predictive framework, in which diagnosis and prognosis can be performed to assess the current and future health of machinery to assist instant decision-making processes. Thus, maintenance and inspection tasks, crew management, and spare parts stocks can be optimised.

To make profit of all this potential, the proposed frameworks are combined with data warehouse and data dashboard solutions. The implementation of data dashboards is considered one of the three critical aspects (data warehouse, data dashboard, and training development) that need to be implemented to advance a company towards data maturity. Data warehouse is the core component of business intelligence, utilised for data analysis and reporting. Data dashboard is an information management tool that supports a specific insight based on the extraction, analysis, measurement, and monitoring of data. As data is being democratised, there is also a necessity to adapt the culture of the company to orient decision-making strategies toward data-driven approaches. Thus, the development of new skills in the personnel is of paramount importance to guarantee the success of this transition.

Accordingly, to finalise the results section, an example of a data dashboard developed based on the open-source stack is presented in Figure 4. As observed, instances of the four parameters analysed in this study are graphically represented by the implementation of Grafana software (https://grafana.com). These visualisations can be utilised by the personnel to identify missing values, and thus implement inspection activities if needed. Alarms can also be set so that a message is sent to the personnel when a certain amount of missing values is achieved per hour. Hence, sensor failure can be detected at an early stage or even prevented, guaranteeing the quality of data required to perform data-driven decision-making strategies.

Table 1. Descriptive statistics of the monitored features.

	Power [kW]	Speed [rpm]	Fuel Flow Rate [tn/hr]	Scav. Air Press. [bar]
Count	1300	1300	1300	1300
Mean	4415.28	104.19	0.87	1.09
Std.	41.94	0.10	0.02	0.03
Min.	4313.49	103.92	0.81	1.02
25%	4387.015	104.12	0.86	1.07
50%	4408.425	104.19	0.87	1.09
75%	4433.1675	104.27	0.88	1.11
Max.	4566.69	104.51	0.93	1.19

Table 2. Pearson's correlation matrix.

	Power [kW]	Speed [rpm]	Fuel Flow Rate [tn/hr]	Scav. Air Press. [bar]
Power [kW]	-	0.99	0.90	0.99
Speed [rpm]	0.99	-	0.90	0.99
Fuel Flow Rate [tn/hr]	0.90	0.90	-	0.91
Scav. Air Press. [bar]	0.99	0.99	0.91	-

Table 3. Spearman's correlation matrix.

	Power [kW]	Speed [rpm]	Fuel Flow Rate [tn/hr]	Scav. Air Press. [bar]
Power [kW]	-	0.80	0.76	0.90
Speed [rpm]	0.80	-	0.74	0.82
Fuel Flow Rate [tn/hr]	0.76	0.74	-	0.78
Scav. Air Press. [bar]	0.90	0.82	0.78	-

Table 4. a) MCAR results fold 1.

	0.05	0.15	0.20	0.25	0.30
GA-ARIMA (EWMA)	21.17	13.69	16.55	20.01	20.19
GA-ARIMA	38.29	41.15	41.18	38.35	42.34
KRR (EWMA)	245.06	320.97	291.54	309.98	290.15
KRR	272.49	343.85	339.18	320.08	326.70

Table 4. b) MCAR results fold 2.

	0.05	0.15	0.20	0.25	0.30
GA-ARIMA (EWMA)	12.127	12.070	11.732	14.015	14.825
GA-ARIMA	45.468	48.617	53.193	53.739	54.790
KRR (EWMA)	365.439	372.451	368.937	373.311	373.738
KRR	393.447	386.104	393.150	394.653	384.923

Table 4. c) MCAR results fold 3.

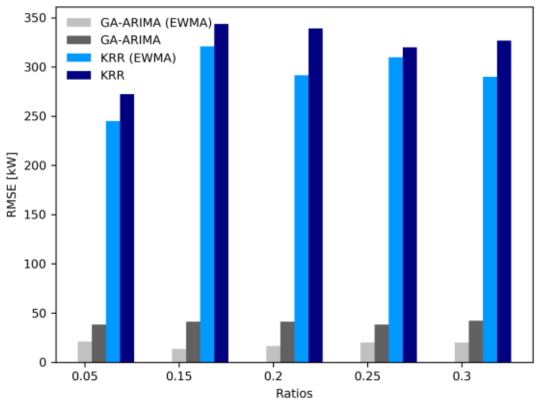
	0.05	0.15	0.20	0.25	0.30
GA-ARIMA (EWMA)	21.17	16.97	16.85	26.43	23.91
GA-ARIMA	47.63	45.04	55.38	45.90	46.48
KRR (EWMA)	365.08	365.27	368.05	366.35	365.03
KRR	340.91	326.56	342.13	321.13	324.75

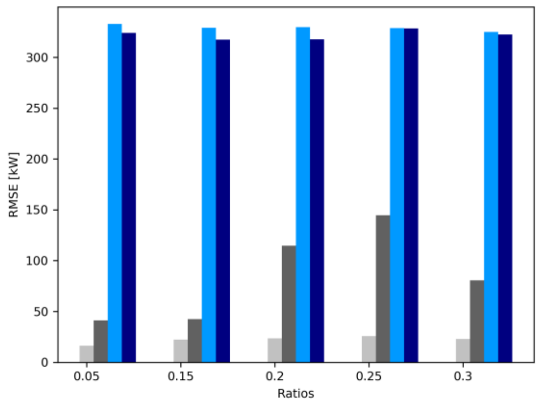
Table 4. d) MCAR results fold 4.

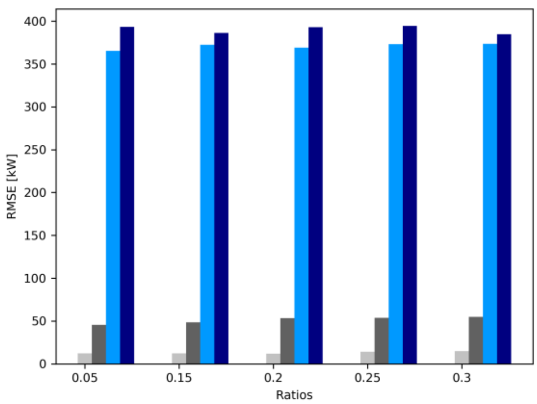
	0.05	0.15	0.20	0.25	0.30
GA-ARIMA (EWMA)	16.36	22.23	23.53	25.64	23.00
GA-ARIMA	41.13	42.29	114.52	144.51	80.47
KRR (EWMA)	332.94	329.03	329.55	328.72	325.09
KRR	323.93	317.42	317.78	328.38	322.54

Table 4. e) MCAR results fold 5.

	0.05	**0.15**	**0.20**	**0.25**	**0.30**
GA-ARIMA (EWMA)	347.88	370.19	406.08	402.20	389.98
GA-ARIMA	575.34	608.99	568.87	567.17	603.60
KRR (EWMA)	295.18	289.66	297.84	295.90	322.83
KRR	391.58	409.42	422.26	399.49	427.91







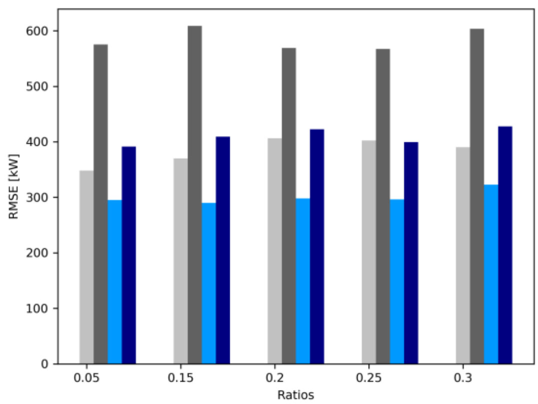


Figure 3. MCAR results plot of (a) fold 1, (b) fold 2, (c) fold 3, (d) fold 4, and (e) fold 5.

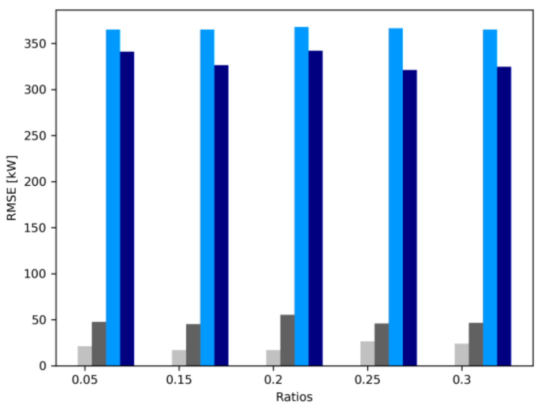


Table 5. DAIF results.

	MCAR	MAR	Final
GA-ARIMA (EWMA)	91.67	257.99	174.83
GA-ARIMA	161.634	279.41	220.52
KRR (EWMA)	331.524	234.24	282.88
KRR	354.83	201.26	278.05

Figure 4. Data dashboard of four main engine parameters in Grafana.

5. CONCLUSIONS

The maritime industry is currently considering state-of-the-art maintenance and inspection processes, an example of which is CBM. To enable this maintenance strategy, sensors need to be installed, and thus IoS needs to be implemented. IoS presents several challenges, including the imputation of missing values.

As previously stated, while there are more than 10,000 publications about data imputation in Scopus database, only two of these refer to the maritime industry. This indicates a lack of analysis and formalisation of data imputation in this industrial sector. To contribute to the development of a data imputation framework in the maritime industry, a total of five novelties were proposed: 1) the development of a data assessment imputation framework to evaluate the accuracy of any imputation model, 2) the implementation of EWMA as a denoising method in the pre-processing step to enhance the performance of the data imputation model, 3) the application of KRR as a multivariate imputation model, 4) the employment of GA-ARIMA as a univariate data imputation model, and 5) the proposal of a real-time imputation tool based on an open-source stack.

In relation to these novelties, the main conclusions are expressed hereunder.

- There is not a unique model that outperforms the remaining imputation techniques for all possible characteristics and contexts described in the maritime industry. Hence, the implementation of comparative studies and assessment frameworks is of paramount importance.
- KRR leads to better results when large gaps of missing values need to be imputed, as GA-ARIMA cannot predict unexpected events.
- GA-ARIMA outperforms KRR when MCAR context is considered.
- The enhancement in the performance that can be produced when denoising is applied to time-series data that contain high noise and when a model sensitive to noise is implemented.
- The importance of preventing failures that lead to the collection of either incorrect or missing values, as this may yield to biased imputations if analogous preceding intances have not been perceived and unexpected events are not recorded.
- The influence of an effective data dashboard on the prevention of sensor failure.

Although some novelties are presented in this paper, further research needs to be addressed due to the importance of this pre-processing step. For this reason, the authors of this paper suggest the following work guidelines:

- A comprehensive comparative methodology of deep learning models as imputation techniques.
- The development of a software that provides data imputation tool to be included in a holistic predicteve framework to assist real-time data-driven decision-making strategies.
- The analysis of optimisation algorithms for tuning hyperparameters.
- The study of denoising alogirithms.

6. REFERENCES

1. EUROPEAN MARITIME SAFETY AGENCY (EMSA), 'Preliminary annual overview of marine casualties and incidents', 2020.

2. ELLEFSEN, A. L. ÆSØY, V., USHAKOV S., ZHANG H., 'A Comprehensive Survey of Prognostics and Health Management Based on Deep Learning for Autonomous Ships', *IEEE Transactions on reliability*, 2019.

3. INTERNATIONAL MARITIME ORGANIZATION (IMO), 'Greenhouse Gas Emissions', 2020.

4. LAZAKIS I., DIKIS K., MICHALA A.L., THEOTOKATOS G., 'Advanced ship systems condition monitoring for enhanced inspections, maintenance and decision making in ship operations', *Transportation Research Procedia*, 2016.

5. LAZAKIS I., GKEREKOS C., THEOTOKATOS G., 'Investigating an SVM-driven, one-class approach to estimating ship systems condition', *Ships and Offshore Structures*, 2018.

6. LAZAKIS I., RAPTODIMOS Y., VARELAS T., 'Predicting ship machinery system condition through analytical reliability tools and artificial neural networks', *Ocean Engineering*, 2018.

7. RAPTODIMOS Y., LAZAKIS I., 'Using artificial neural network self-organising map for data clustering of marine engine condition monitoring applications', *Ships and Offshore Structures*, 2018.

8. RAPTODIMOS Y., LAZAKIS I., 'Application of NARX neural network for predicting marine engine performance parameters', *Ships and Offshore Structures*, 2019.

9. BALAKRISHNAN S.M., SANGAIA A.K., 'Chapter 6 – aspect oriented modelling of missing data imputation for Internet of Things (IoT) based healthcare infrastructure', *Intelligent Data-Centric*, 2018.

10. IZONIN I., KRYVINSKA N., TKACHENKO R., ZUB K., 'An approach towards missing data recovery within IoT smart systems', *Procedia Computer Science*, 2019.

11. NOOR N.M., ABDULLAH M.M.A.B, YAHAYA A.S., RAMLI N.A., 'Comparison of linear interpolation method and mean method to replace missing values in environmental data set', *Mater. Sci*., 2014.

12. FEKADE B., MAKSYMYUK T., KYRYK M., JO M., 'Probabilistic recovery of incomplete sensed data in IoT', *IEEE Internet of Things Journal*, 2018.

13. VELASCO-GALLEGO, C., LAZAKIS, I., 'Real-time data-driven missing data imputation for short-term sensor data of marine systems. A comparative study', *Ocean Engineering*, 2020.

14. CHELIOTIS M., GKEREKOS, C., LAZAKIS, I., THEOTOKATOS, G., 'A novel condition and performance hybrid imputation method for energy efficient operations of marine systems', *Ocean Engineering*, 2019.

7. AUTHORS BIOGRAPHY

Mr. Christian Velasco-Gallego is a PhD student in the Department of Naval Architecture, Ocean, and Marine Engineering at the University of Strathclyde. He has experience from working in multinationals as a technical consultant and supply chain specialist. His research interests include predictive maintenance and supply chain management.

Dr. Iraklis Lazakis, CEng, MRINA, MSNAME, PhD is a Reader at the Department of Naval Architecture, Ocean, and Marine Engineering at the University of Strathclyde. He has extensive academic experience from working in UK, EU, and international projects. His research interests include ship systems condition monitoring, systems maintenance and reliability, ship operations, risk analysis tools and methodologies, system criticality assessment, shipyard manufacturing and productivity, and asset management.