Remaining Useful Life Prediction of Underground Water Pipes Using Design Expert

M. Chikuruwo, M. Mutingi, S. Johns, H. Musiyarira, A. Ditima and T Garikayi.

***Abstract*—**Remaining usable life (RUL) of water pipes can be categorized as history based, data dependent and physics-centered prediction models. A few authors developed hybrid RUL prediction models which calls for precise and trustworthy data. The purpose of this research is to create a mathematical model that uses secondary data to forecast RUL of underground water pipes. Response surface methodology (RSM) was employed to build the model to estimate the pipe RUL. A coefficient of determination R^2 value of 0.9876 was obtained after the model underwent response surface optimization, indicating that most of the variation in the dependent variable can be explained in the model for practical reasons. For practical reasons, the soil type, groundwater-table location and operational factors were not considered as the data can be used regardless of the environmental conditions.

***Keywords*—** Coefficient of determination, Design of Experiments, Remaining useful life, Response Surface Methodology.

I. Introduction

Water pipes are essential components of a country's infrastructure since they are crucial for meeting humanity's most fundamental demand for water [1]. According to Tavakoli [2], the United States' water delivery system is thousands of miles long with pipes made of various materials, varying in size, age, and construction. These pipes are subjected to operational, structural, and physical stresses that eventually cause deterioration and failure. Given that degraded pipes are more vulnerable to leaks, pipe degradation is a fundamental factor in the detection of leaks. Since there is a time lapse between a pipe failure and its effects the failure are exceedingly difficult to diagnose [3]. The RUL of the mechanical system must be forecasted, specifically when the system is no longer performing reliably, due to the fact that most mechanical systems endure failure gradually ([4]). In accordance with [5], RUL forecasting is an approximation of the total amount of time that a system can still be used reliably, which is a crucial component of system health and the adoption of a robust maintenance strategy. RUL prediction models can be determined using the physics around the system, available data, and practical knowledge of the system. To create hybrid models, these models were combined by different authors.

Manuscript received 27 March 2023.

(M Chikuruwo is with the Harare Institute of technology, Zimbabwe
M Mutingi is with Namibia University of Technology, Namibia
H. Musiyarira, A. Ditima and T Garikayi, Namibia University of Technology, Namibia

The authors tested various combinations to create the best hybrid model. Forecasting the mechanical system's remaining useful life (RUL), particularly when it is no longer operating consistently, is necessary because most mechanical systems deteriorate gradually [4]. According to [5], RUL forecasting is an approximation of the total time that a system can still be reliably used, which is a very important aspect of system health and the adoption of a robust maintenance strategy. The physics surrounding the system, usage of the data at hand, and practical understanding of the system can all be used to determine RUL prediction models. Several writers integrated these concepts to create hybrid models. The authors tried different combinations in order to come up with the best hybrid model. Their conclusions were that a hybrid model can determine RUL more accurately as compared to using the models independently on the other hand, [6] proposed the use of genetic programming as a method to discover unconventional features that can determine the progression of RUL prediction. Their proposed methodology though, makes it difficult to interpret the physical meaning of the system degradation. [7] used Support Vector Regression-Particle Filter (SVR-PF) to forecast the remaining usable life of lithium-ion batteries. The prediction RUL model proposed provided the RUL quantity and updated the probability distribution of RUL to when the product is no longer usable. [8] developed a particle filter model with interacting multiple framework which used support vector regression so as to obtain multi-step-ahead approximation RUL of batteries.[9] suggested a prediction method that employed hot spot temperature method of prediction focusing on the functioning parameters of the transformer through the use of using a support that employs particle filter. In addition, an advanced particle filter (PF) approach was applied in the prediction of the RUL for bearings [9]. Particles at each iteration step were determined by a tunable significance density function, and a backpropagation (BP) neural network was applied to increase particle diversity before resampling.

A selection approach known as Adaptive First Predicting Time (FPT) was established by [10] for rolling elements bearings. The FPT was centered on 3σ interval in which particle filtering was applied to diminish random faults of the stochastic process. The process effectiveness was demonstrated through a replication and RUL prediction was done on four tests of bearing degradation. Furthermore, RUL prediction was done by, [11] though the use of a model-based method (a collective system) instead of a part of a machine. [12] used health indicator hinged on Recurrent neural network to forecast the RUL of bearings. [6], however, predicted the RUL of bearings by employing an integrated deep learning approach which combined both time and frequency domain features.

https://doi.org/10.17758/EARES14.EAP0723219

II. DESIGN

A. Data collection

Secondary data was collected from a publication which was documenting RUL prediction of underground pipes by Tavakoli in 2018. In this investigation soil type, groundwater-table location and operational factors were not considered which therefore made it practical to use the data regardless of environmental conditions. Table 1 shows data that was used to determine the most significant parameters influencing RUL of underground water pipes.

TABLE I: STEEL PIPE LEAKAGE DATA SAMPLES (TAVAKOLI, 2018, PAGE NO 186)

Mat	Age	Len	Dia	No	Wall	RUL
Steel	1	236	12	0	1	69
Steel	1	787.4	12	0	4	69
Steel	1	623	12	0	3	69
Steel	2	1,380	12	0	5	68
Steel	3	1,310	12	1	3	67
Steel	4	1,360	12	0	3	66
Steel	5	1,410	12	1	4	65
Steel	6	1,340	12	1	4	64
Steel	7	1,320	12	1	5	63
Steel	8	1,200	12	0	3	62
Steel	9	1,220	12	1	4	61
Steel	10	1,250	12	0	6	60
Steel	11	1,420	12	1	7	59
Steel	12	1,260	12	0	9	58
Steel	13	1,330	12	0	8	57
Steel	32	2,640	6	0	13	38
Steel	43	2,286.75	12	6	29	27
Steel	43	554.46	12	8	27	27
Steel	43	3,687.6	12	10	35	27
Steel	43	1,509	15	2	37	27
Steel	43	919	15	1	31	27
Steel	43	853	17	1	37	27
Steel	43	2,464	17	2	39	27

Where :Mat- Material, Age – Age in years, Len- Length in feet, Dia- Diameter in inches, No- Number of breaks, Wall – % Wall loss, RUL-Remaining useful life in years.

B. Design of Experiments Using Design Expert Software

Design expert 13 is a statistical software package that is used to perform design of experiments (DOE). DOE is a powerful tool for understanding and optimizing complex processes [13]. DOE assists in identifying very important variables and interactions that has an effect on RUL, allowing for more accurate and reliable predictions. For DOE using secondary data, RUL was defined as the response variable. The units and range of the RUL was set. The next step was to identify the input variables which, in this case, were the age, length and diameter of the pipe. These were the independent variables that affected the response variable. After defining the input variables, central composite design was identified to be used for the design of experiments. This design can identify the curvature of the response surface, and can provide insights into the optimal settings for the factors that maximize the RUL. Data was collected for each combination of input variables defined in the experimental design. The collected data was analyzed and significant main effects and interactions between input variables that affect RUL were identified. The software generated statistical models that predicted RUL based on the input variables. The process was optimized using the statistical models generated by the software and the best conditions for maximizing RUL were identified. DOE using Design Expert software is a valuable tool for predicting RUL using secondary data. By identifying the most important input variables and their interactions, the accuracy of RUL predictions can be improved and the maintenance process can be optimized for maximum efficiency and effectiveness.

C. Response surface methodology (RSM)

RSM was employed to come up with a model to predict RUL. RSM is a statistical procedure that is usually employed to model and optimize complex systems. RSM forecasts and optimizes the RUL of water pipes centered on pipe age, diameter and length. Experiments were designed to collect data on the system's performance over time and statistical tools were used to analyze the data and develop a framework that predicts the RUL of the system using experimental variables.

The model was initially validated by comparing its predictions with actual RUL data from the system. The model was used to optimize the RUL by identifying the most important factors and adjusting them accordingly. The most important parameters were displayed from a correlation table generated from initial variables and responses that were entered into design expert software. The tabulated results show a correlation of -0.346 between RUL and age of the pipe. This meant the older the pipe the less the number of useful years remain for the pipe. There was however, a 0.787 degree of correlation between age and number of breaks which indicates a high degree of positive correlation. Thus as the age increases the number of breaks also increases. A moderately high positive correlation existed between age and wall loss as shown by the correlation coefficient of 0.590. The table, presented a somewhat low negative correlation between length and RUL with a value of -0.434. This implied that there were greater chances of reduced RUL on longer pipes. A very low correlation of 0.185 existed between the pipe length and the breaks that occurred on the pipe. The correlation between length and wall loss was almost negligible as represented by the correlation coefficient of 0.038. Thus the rate at which wall loss took place was not affected by the length of the pipe. The correlation of 0.383 between RUL and diameter was such that the bigger the diameter, the higher the RUL. With larger diameter, the number of breaks decreased by -0.139. A negative correlation factor of -0.363 existed between diameter and wall loss. There was, however, no correlation at all between age, length and diameter.

III. MATHEMATICAL MODEL

In generating a mathematical model, design expert (DE) software identified the most probable equation for the model and the actual values had to be used in order to determine how accurate the prediction model was. Mode1 was the first model that was generated to predict RUL. In the model: A is the age, B

https://doi.org/10.17758/EARES14.EAP0723219

is the length and C is the diameter of the pipe.

$$RUL = 19.095 - 1.780A + 0.010B - 0.252C + 0.026A^2 \quad (1)$$

The predicted R2 of -4.9637, which is the coefficient of determination generated by model 1 indicated that the model was a poor fit. A larger R2 value indicates that the dependent variables variation can be explained by the independent variables in the model and is better at predicting the dependent variable figures founded on independent variables. The adequacy (Adeq) precision which measures how well the model can predict the response variable for new observations that were not used to build the model, was used in conjunction with R^2 and mean squared error, to approximate the performance of the model. The Adeq precision of 6.2494 was generated which indicated that the model could be useful in predicting new data. However, this model was discarded because of its poor R2 value as shown in Fig. 1.

Influential observations were produced using Cook's distance which is a statistical measure used to identify and generate influential points in regression analysis. Points with high Cook's values may have significant impact on the regression results.

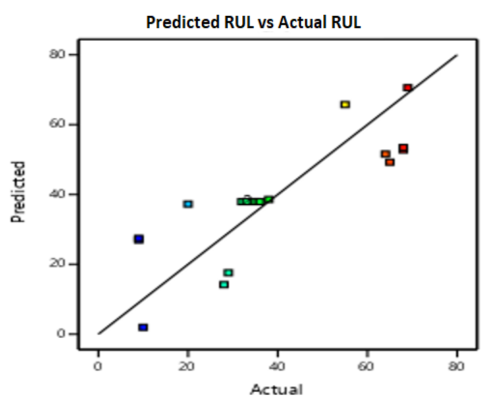


Fig. 1. Graph showing the predicted against the Actual RUL

In addition to Cook's distance, Design expert also provided leverage and studentized residuals, which helped in identifying influential points and other potential issues with the regression model. These diagnostic measures were used to refine the model and improve its predictive accuracy. In order to improve the model, observations number 2, 6, 9, 12, 18 and 20 which had Cook's distance that was greater than 0.2 (which is 4/20 and any point that is greater than 4/20 is an outlier), and the difference in fits (DFFITS) was either very low (-8.619) or very high (5.840). The removal of these influential observations helped in coming up with a better prediction model.

D. Reducing the Model

For a more significant model, model reduction was done by scrapping the results of the pipe aged 1, 2 and 61. The new values were entered into the Design Expert software for experimentation. Comparison with the first experiment were made and it was worth noting that the new model was also a quadratic model although a negative correlation of 0.979 was observed between age and RUL. This indicated age as the parameter with the highest influence on RUL of underground water pipes. A moderately high positive correlation existed between age and the number of breaks, and age and the wall loss with correlation coefficients of 0.552 and 0.819 respectively. The reduced model had an F-value of 102.80 which showed a significant model. There was only a 0.01 probability of such a high F-value occurring because of noise. An insignificant lack of fit of 0.13 was observed on the reduced model as compared to model 1. A probability of 0.978 of such a high value of lack of fit can transpire due to noise. The reduced model is represented by Model 2 where: A is the age, B is the length and C is the diameter of the pipe.

$$RUL = 72.768 - 1.3A + 0.002B - 1.14C - 0.001AC + 0.009A^2 + 0.056C^2 \quad (2)$$

IV. Results

From Table 2, R squared is 0.9893 which indicates that almost 99% of the total variation in RUL is explained by age, length and diameter of the pipe while the remaining percentage is due to other factors not included in the model. A predicted R2 of 0.9776 was generated from model 2 which is in sync with a value of 0.9797 generated for adjusted R2. There is a variation of smaller than 0.2 between the adjusted R2 and predicted R2 which indicating a good fitting model and is not overfitting the data. A 28.5577 Adeq precision value represents a high degree of accuracy in predicting new observations within the range of the design factors used in the experiment. It is calculated as the ratio of the range of predicted responses to the prediction error average.

TABLE II: FIT STATISTICS

Std. Dev.	1.99		**R^2**	0.9893
Mean	45.10		**Adjusted R^2**	0.9797
C.V. %	4.40		**Predicted R^2**	0.9776
			Adeq Precision	28.5577

The Adeq precision is greater than the average prediction error by 28.5577 times, which suggest that the model is highly accurate in predicting new observations within the range of the design factors. A favorable and a better fit was produced when the RUL predicted was related to the real RUL as indicated in Fig 2. The graph shows no significant deviation of the RUL predicted RUL from the real RUL.

https://doi.org/10.17758/EARES14.EAP0723219

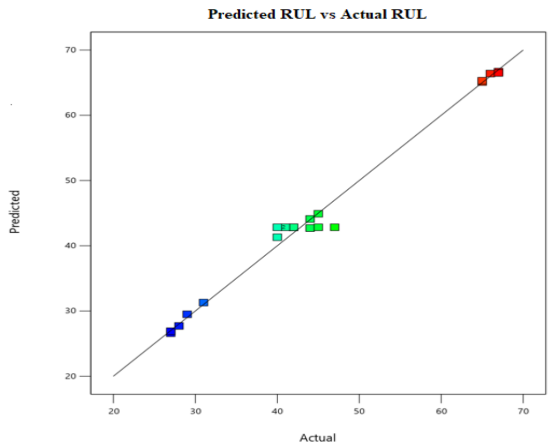


Fig. 2**:** Graph showing RUL predicted against the Real RUL of the cleaned up model

The relationship between age, length and the RUL of the pipe shown in Fig 3 which shows the response surface plot of age against length and length against RUL. From the surface plot, it is evident that age has very high influence on RUL as shown by the contour lines while the effect of length on RUL is almost negligible.

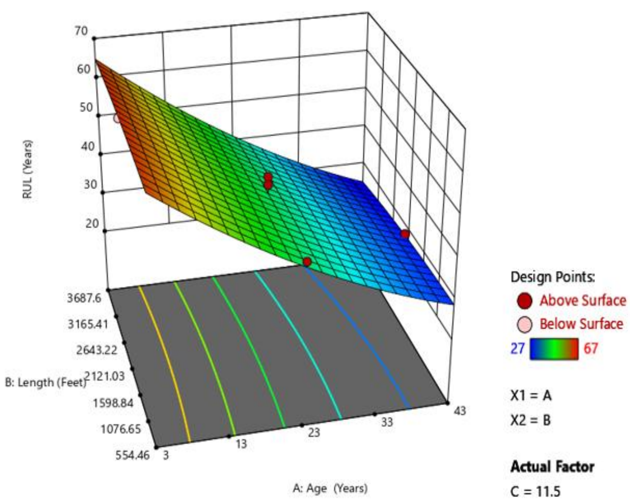


Fig. 3. Response Surface plot of RUL vs Age and Length.

Table 3 and 4 shows the points highlighted above the surface and below the surface of the response surface plot. The points highlight that even when the pipes are of the same length and have different ages the RUL will change significantly depending on the age.

TABLE III: RESULTS FROM ABOVE THE RESPONSE SURFACE PLOT.

Age	length	RUL
23	2121.03	47
23	2121.03	45
23	554.46	44
28	211.03	28

TABLE IV: RESULTS FROM BELOW THE SURFACE.

Age	length	RUL
23	3687.6	40
23	2121.03	41
23	2121.03	42
3	2121.03	65

A. *Comparing the model with Actual RUL*

RUL from the secondary data used was plotted against the RUL found from using the mathematical model and the results are shown graphically in fig 4, where the graphs have an almost similar trend.

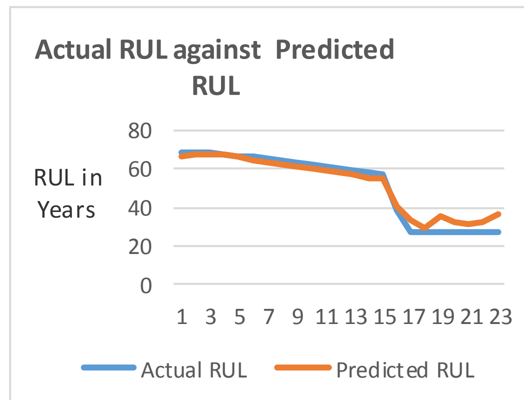


Fig. 4: Comparing the RUL predicted and the Real RUL

Fig 5 shows the trend line of the actual values generated from the scatter plot of the values of RUL predicted.

$$y = 0.7986x + 11.148 \quad (3)$$

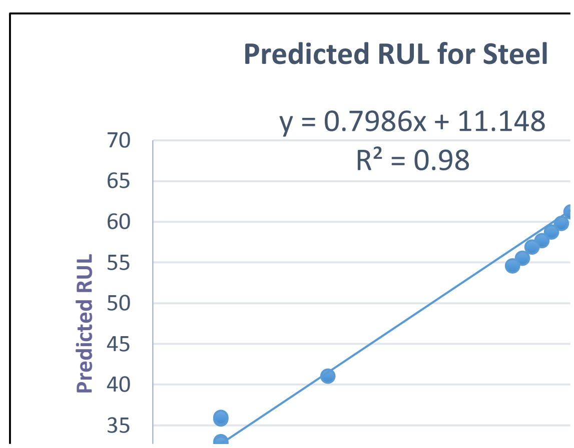


Fig. 5. Goodness of Fit for predicted RUL

From the graph, the predicted values of RUL do not deviate much from the trend line and a pattern can be seen from the set of data presented. This observation is supported by the trend line equation 3. The goodness of fit, R^2, of 0.9876 shows that the data fit the regression model very well. The X-axis shows the actual RUL and the Y-axis presents the predicted RUL. Actual RUL was evaluated using the secondary data provided and forecasted RUL was calculated using the mathematical model developed. The bulk of the figures shown are closer to y=0.7986x+11.148, and there is coefficient of determination of 99%. This very high coefficient of determination expresses that the developed mathematical model accurately foretold the pipe RUL of the pipeline and can be used reliably for extra investigation of the network.

V. Model validation

The model was validated using data on ductile iron from Tavakoli 2018. The pipe age of the data used for validation was from 10 years to 36 years with length ranging from 237 feet to 21120 feet.

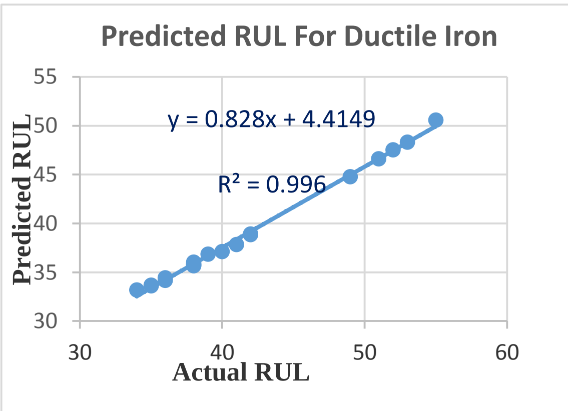


Fig. 6. Goodness of fit for prediction model on validation RUL

The diameter of the pipes used for validation was between 6 and 24 inches. Fig 6 shows the goodness of fit for the ductile iron actual RUL compared to the RUL predicted by the mathematical model developed. The X-axis shows the actual RUL while the Y-axis presents the predicted RUL.

VI. Conclusion

A mathematical RUL prediction model for underground water pipes was developed and validated. Age was seen to be the most significant parameter affecting RUL since a very high negative correlation exists between the age of the pipe and RUL. Response surface methodology further represented the results graphically to elaborate the relationship between age and RUL. The model created can be useful in forecasting RUL of pipes even when maintenance data is not available. Using the model enables a reliable health management system to be developed as scientific input will be used in determining the correct maintenance technique to be used.

VII. Recommendations

The accuracy of the model can be boosted by having more input parameters identified to have an effect on the water condition are available for modeling. Availability of primary data would help in coming up with a more accurate model as the soil type, groundwater-table location and operational factors would be considered for the prediction model.

References

[1]. D. Robert, D. Chan and J. K. P Rajeev, "Effects of operational loads on buried water pipes using field tests," *Tunnelling and Underground Space Technology,* vol. 124, no. 104463, 2022. https://doi.org/10.1016/j.tust.2022.104463

[2]. R. Tavakoli, A. Sharifara and M. Najaf, "Artificial Neural Networks and Adaptive Neuro-Fuzzy Models to Predict Remaining Useful Life of Water Pipelines," in *World Environmental and Water Resources Congress 2020:* , A. S. o. C. Engineers, Ed., Alexander, 2020 American Society of Civil Engineers , 2020. https://doi.org/10.1061/9780784482988.019

[3]. D. Thirka, E. Elwakil, H. Novoa and J. Delgado, "Artificial intelligence for the modeling of water pipes deterioration mechanisms," *Automation in Construction,* vol. 120, no. 103398, 2020. https://doi.org/10.1016/j.autcon.2020.103398

[4]. H. Dong, X. Jin, Y. Lou and C. Wang, "Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter," *Journal of Power Sources,* vol. 271, pp. 114-123, 2014. https://doi.org/10.1016/j.jpowsour.2014.07.176

[5]. L. Ren, J. Cui, Y. Sun and X. Cheng, "Multi-bearing Remaining Useful Life Collaborative Prediction: A Deep Learni Approach," *Elsevier,* pp. 1-17, 2017. https://doi.org/10.1016/j.jmsy.2017.02.013

[6]. X. Li, Q. Ding and J. Sun, "Remaining Useful Life Estimation in," *Elsevier Science,* pp. 1-28, 2017.

[7]. L. Guo, N. Li, F. Jia and Y. Lei, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing,* pp. 98-109, 2017. https://doi.org/10.1016/j.neucom.2017.02.045

[8]. S. Li, H. Fang and B. Shi, "Remaining useful life estimation of Lithium-ion battery based on interacting multiple model particle filter and support vector regression," *Reliability Engineering & System Safety,* vol. 210, no. 107542, 2021. https://doi.org/10.1016/j.ress.2021.107542

[9]. H. Liao, "State Dissatisfaction: Predicting the Occurrence of Interstate War," Asian Politics and Policy, pp. 2177-236, 2014. https://doi.org/10.1111/aspp.12106

[10]. Y. Qian and R. Yan, "Remaining Useful Life Prediction of Rolling Bearings Using an Enhanced Particle Filter," IEEE Transactions on Instrumentation and Measurement, pp. 2696-, 2015. https://doi.org/10.1109/TIM.2015.2427891

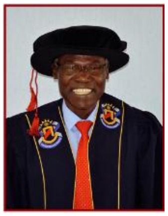
[11]. Y. Lei , N. Li, J. Lin, S. Radkowski and J. Dybala, "A Model-Based Method for Remaining Useful Life Prediction of Machinery," IEEE Transactions on Reliabil, pp. 1314-1326, 2016. https://doi.org/10.1109/TR.2016.2570568

[12]. L. Liao and F. Kottig, "Review of Hybrid Prognostics Approaches for Remaining Useful Life Prediction of Engineered Systems, and an Application to Battery Life Prediction," IEEE Transactions on Reliability, vol. 6, pp. 191-207, 2014. https://doi.org/10.1109/TR.2014.2299152

Eng. Mary Nyaradzayi Hughslar Chikuruwo is a lecturer and the current chairperson of Industrial and Manufacturing Engineering Department at Harare Institute of Technology. She is a holder of a Master of Engineering degree in Manufacturing Engineering and Operations Management (MEOM) from National University of Science and Technology, Zimbabwe, a Master of Technology degree in Design Engineering (DE) from Jawaharlal Nehru Technological University in India and a Bachelor of Technology degree in Industrial and ManufEng. acturing Engineering (IME) from Harare Institute of Technology, Zimbabwe. She is studying towards her PhD in Reliability Engineering with Namibia University of Science and Technology and her topic is “Development of an Intelligent Hybrid Prognostic and Health Management System for Underground Water Pipes”. She is passionate about sustainable engineering and is an advocate of engineering for sustainability in

Zimbabwe.

Prof. Michael Mutingi Associate Professor, Industrial Engineering, Namibia University of Science and Technology. DEng. Engineering Management, University of Johannesburg, South Africa (2015). MEng. Industrial Engineering, National University of Science and Technology, Zimbabwe (1998), BEng. Industrial Engineering, National University of Science and Technology, Zimbabwe (2004). Operations Management, Industry 4.0 Bio-inspired optimization algorithms in manufacturing and service systems. Systems thinking, system dynamics simulation.



Prof Samuel John studied Mechanical Engineering at the Ahmadu Bello University Zaria, in Nigeria. He obtained his Bachelor of Engineering in 1987 and Master of Science in mechanical engineering in 1992 in the same university. He then obtained his Doctor of Philosophy degree at the University of the Witwatersrand Johannesburg, South Africa in 2012. Professor John is also a theologian with a Master of Theology in integrated theology, which he obtained in 2019 at the South African Theological Seminary (SATS). Professor John joined NUST (formerly the Polytechnic of Namibia) on 1st August 2003 as a Senior Lecturer. He has served as Deputy HoD and HoD of the Department of Mechanical Engineering and was Dean of the Faculty of Engineering from 2013 to 2019. (SATS). Professor John joined NUST (formerly the Polytechnic of Namibia) on 1st August 2003 as a Senior Lecturer. He has served as Deputy HoD and HoD of the Department of Mechanical Engineering and was Dean of the Faculty of Engineering from 2013 to 2019.

Prof Harmony Musiyarira Executive Dean: PhD Process and Environmental Engineering (Witwatersrand University) 2011. M.Sc Manufacturing Systems and Operations Management (University of Zimbabwe) 2005. Bachelor of Science (Honours) in Metallurgical Engineering (University of Zimbabwe) 2001. President Elect of the international Society of Mining Professors (SOMP) (2021-2022). Member of the Australian Institution of Mining and Metallurgy (AusIMM). Member of the Southern African Institute of Mining and Metallurgy (SAIMM). Research Focus Areas Mine Closure and Rehabilitation and Cleaner Production and Circular economy

Dr. Eng. Talon Garikayi Currently the Pro-Vice Chancellor - Research, Innovation & Commercialisation. Holder of PhD in Mechatronics from Stellenbosch University South Africa specialising in Bionics and Artificial Intelligence and their applications in Biomedical Engineering. Master of Engineering from National University of Science and Technology Zimbabwe and Bachelor of engineering. (Hons) in Mechatronics. A Certified Global Engineer trained and groomed in German specializing in Design Thinking and System Engineering. He is also a Professional Member of ZIE and ECZ. He holds He has a strong bias towards firmware and hardware designs for IT embedded systems and has worked as a Robotics Engineer for Ford Motors, South Africa. As a Certified Automation Engineer, he has developed control systems, automated processes, optimised production lines, SCADA and HMI Systems for several manufacturing and processing engineering firms including Steelmakers, Unilever, Ford Motors, Greif, Kadoma Textiles, Glendale Spinners, Zimplats, Mega Pak, Canuad Metal Box and Cottco.

Mr. Allen Ditima is a holder of Bachelor of technology degree in Industrial and Manufacturing Engineering from Harare Institute of Technology in Zimbabwe. And currently study for a master of technology degree in Machine design at Harare Institute of Technology and a teaching assistant in the department of Industrial and Manufacturing engineering. His areas of research interests are design and systems engineering.