Data-Driven Remaining Useful Life Estimation: Inference-Based Versus Direct Prediction

Ark Ifeanyi[1], Mattia Zanotelli[2], and Jamie Coble[3]

[1,2,3] *University of Tennessee, Knoxville, TN, 37996, USA*
aifeanyi@vols.utk.edu
mzanotel@vols.utk.edu
jamie@utk.edu

ABSTRACT

This paper explores the development and application of data-driven prognostic models for estimating the Remaining Useful Life (RUL) of Nuclear Power Plant (NPP) condensers experiencing tube fouling. Due to the unavailability of run-to-failure industry sensor data, we utilized simulated data generated by the Asherah Nuclear Power Plant Simulator (ANS), initially designed by the International Atomic Energy Agency (IAEA) and programmed in Simulink for cyber security simulations. ANS's adaptability allows it to simulate Pressurized Water Reactor (PWR) behaviors given a time series of operating conditions and to introduce degradation modules to mimic fouling effects. Our study compares two primary approaches applied to data generated by ANS: inference-based and direct prediction methods. The selected inference-based approach estimates the health state of the condenser using a pipeline formed by an Auto Associative Kernel Regressor and a Hidden Markov Model (HMM), which subsequently combines the state estimates with its parameters to predict the RUL. The direct prediction method employs a Gradient Boosting Regressor Decision Tree (GBRDT) to map input variables directly to RUL. Our findings demonstrate the efficacy and limitations of each method through the case study, providing valuable insights for the adoption of data-driven RUL estimation techniques in industrial and energy applications.

1. INTRODUCTION

In industrial and energy settings, the ability to predict the Remaining Useful Life (RUL) of critical components is essential for effective maintenance planning and operational efficiency. Accurate prognostics can lead to significant economic benefits by minimizing unplanned downtimes, optimizing maintenance schedules, and extending the service life of equipment (Elattar et al., 2016). These advantages have spurred considerable interest in data-driven methods for RUL estimation.

The motivation for exploring data-driven approaches stems from their potential to leverage vast amounts of operational data generated by modern industrial systems. Traditional physics-based models, while useful, often require extensive domain knowledge and may not fully capture the complex behaviors of advanced machinery (Cubillo et al., 2016). In contrast, data-driven methods can uncover hidden patterns and relationships within the data, providing robust and scalable solutions for RUL estimation (An et al., 2013).

In this context, two primary approaches have emerged: inference-based and direct prediction methods. Inference-based approaches involve a two-step pipeline where the health state of the investigated component or system is first estimated, and then the RUL is inferred from the estimated health state (Yu, 2015; Peng et al., 2019; Sankavaram et al., 2016). This method allows for a detailed understanding of the degradation process and can be particularly useful when there is a clear, interpretable path from an healthy state to failure. On the other hand, direct prediction methods involve mapping input variables directly to the available true RUL of components over time (Khelif et al., 2016). This approach simplifies the modeling process and can provide quick, accurate predictions without the need for intermediate steps. Both methods hold significant value for the industry. Inference-based methods offer detailed insights into the degradation mechanisms, which can be critical for diagnostic purposes and for improving design and operational strategies. Direct prediction methods, with their straightforward implementation and rapid results, are advantageous in scenarios where quick decision-making is paramount.

The primary contribution of this paper is to address a common question among practitioners: which data-driven prognostics approach should I adopt? This study offers a detailed comparison of two major data-driven methodologies, using a

nuclear power plant (NPP) condenser as a case study. By employing comprehensive quantitative and qualitative metrics, including uncertainty assessments, we draw meaningful conclusions about the efficacy of each approach. Through this comparison, we aim to shed light on the strengths and limitations of both inference-based and direct prediction methods, providing valuable insights for the implementation of data-driven remaining useful life (RUL) estimation techniques in industrial and energy sectors. The NPP condenser is critical to the power generation process, as it converts steam from the turbine back into water for reuse in the boiler. Ensuring the condenser's reliability and optimal performance is essential for the efficient operation of the power plant.

The rest of this paper is organized as follows: section 2 discusses relevant literature and underscores our contributions, section 3 details the data used, section 4 explains the methodology, section 5 presents the findings, and section 6 summarizes the key points and suggests future work.

2. LITERATURE REVIEW

Condenser tube fouling is the accumulation of debris on the surface of the condenser tubes. This process causes an increase in thermal resistivity and a slight reduction of the cross-sectional area of the flow (Ibrahim & Attia, 2015). As a consequence, the heat exchanger's effectiveness decreases, leading to plants' performance drops. Researchers have addressed this issue by developing methods that can detect fouling within condensers. For instance, the authors in (Sundar et al., 2020) outline several methods for detecting fouling in heat exchangers, incorporating both traditional and modern approaches such as empirical and analytical models, data-driven prediction approaches, and deep learning models. However, this article and many others do not explicitly cover prognostics and RUL prediction. Another research (Zhang et al., 2023) proposes a method for generating fouling data using a Digital Twin and performs online prognostics by combining particle filtering and differential modeling to reduce uncertainty. However, the employed method required defining an equation for fouling thermal resistance and did not explicitly consider NPPs. NPP condensers have slightly different requirements compared to thermal power plants like coal. Specifically, NPP condensers must manage a higher level of safety and regulatory compliance, and they often operate under different thermal and pressure conditions, necessitating our research.

Prognostics algorithms are designed to predict when a system or component will cease to function as intended by analyzing its deviation and degradation from normal operating conditions. Although the health state of an item generally degrades linearly with usage, predicting failures is challenging due to varying operational conditions, environments, and the complex nature of different parameters. To achieve this, three main modeling techniques are used: regression models, which model the degradation path to predict failure time; classification models, which predict if failure will occur within a specific time window; and survival models, which assess how the risk of failure changes over time. In this study, the regression approach is taken due to its applicability to the investigated system and available data. These regression strategies can be implemented using data-driven models that could fall into two broad categories: statistical, and machine learning-based models.

2.1. Statistical Models

a) General Path Models (GPMs): These models capture the degradation trend of a system over time using parametric regression. They work by fitting a probabilistic model to the degradation data, assuming that degradation follows a specific path influenced by both population and individual effects. (Coble & Hines, 2011) paper proposes a specific formulation of the General Path Model with dynamic Bayesian updating as one effects-based prognostic algorithm. The method is illustrated with an application to the prognostics challenge problem posed at PHM '08.

b) Stochastic Process Models (SPMs): These models are advantageous in situations where the accumulation of the degradation is random and requires a probabilistic approach. They model degradation as a stochastic process, accounting for randomness and environmental influences. Examples are Wiener processes (Wang, Ma, & Zhao, 2019), Gamma processes (K. Song & Cui, 2022), Gaussian processes (Valladares et al., 2022), Markovian-based models, which will be further revised later, and Filtering-based models (Cui, Li, Wang, Zhao, & Wang, 2022).

2.2. Machine Learning Models

a) Support Vector Machines (SVMs): Initially developed for classification, SVMs have been adapted to regression tasks (Support Vector Regression, SVR) and further employed for RUL prediction (Rezki & Rezgui, 2024). SVMs are effective in high-dimensional spaces and work well when the number of features exceeds the number of samples. However, they can struggle with noisy data.

b) Decision Trees (DTs) and Random Forests (RFs): DTs create a tree-like model to make predictions based on features, while RFs aggregate the predictions of multiple DTs to improve accuracy and robustness. These models are intuitive and easy to interpret but can become unstable with slight changes in data (Patil et al., 2018; Kundu, Darpe, & Kulkarni, 2020). RFs, in particular, are effective in handling large datasets with diverse features (Patil et al., 2018).

c) Back Propagation Neural Networks (BPNNs): BPNNs are a type of artificial neural network that uses gradient

descent to minimize prediction errors iteratively. They are flexible and powerful in capturing non-linear relationships between inputs and outputs. However, BPNNs can get stuck in local minima, limiting their performance (Ma, Yao, Liu, & Tang, 2022).

d) Convolutional Neural Networks (CNNs): CNNs are widely used for their ability to extract both local and global features from input data, such as time-series or image data. They consist of multiple layers, including convolutional, pooling, and fully connected layers, which allow them to learn complex patterns. CNNs are highly effective in feature extraction but require large datasets and significant computational resources (Xu, Li, Ming, & Chen, 2022; A. O. Ifeanyi, Coble, & Saxena, 2024).

e) Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs): RNNs are designed to capture sequential dependencies in data, making them ideal for time-series forecasting (Y. Song, Li, Peng, & Liu, 2018). LSTMs, a variant of RNNs, are particularly useful for modeling long-term dependencies and avoiding issues like gradient vanishing (Zhang, Xiong, He, & Liu, 2017). These models are effective for RUL prediction but can be challenging to train. Graph-based CNNs and RNNs have also been proposed for prognostic applications (A. Ifeanyi, 2024).

f) Bayesian Deep Learning: This approach extends deep learning into a probabilistic framework, where uncertainty in model predictions is quantified by placing prior distributions over the model's parameters. Bayesian deep learning is particularly valuable when dealing with limited data or when quantifying the uncertainty of predictions is critical, though it tends to be computationally expensive (Zhu, Chen, Peng, & Ye, 2022).

g) Transfer Learning: Transfer learning leverages knowledge from one domain (the source) and applies it to a different but related domain (the target). This is especially useful in RUL prediction when data in the target domain is scarce. Transfer learning can significantly improve model performance, but it requires the source and target domains to be sufficiently similar to avoid negative transfer (L. Song et al., 2024).

These methods are used in various industrial applications, such as rotating machinery, aircraft, power systems, and electronics, to predict equipment failure and optimize maintenance schedules (Wen, Rahman, Xu, & Tseng, 2022).

In this work, we want to compare two prognostic models, each belonging to one of the aforementioned categories: the hidden Markov models (inference-based and statistical) and the Gradient Boosting Regressor (direct prediction and machine learning-based). The first one can estimate the future health states and use these predictions to compute the RUL, while the second model directly predicts the RUL. Although these models were already applied for prognostics, no comparison between the results and applicability has been made. For instance, (Dong & He, 2007) presents a unified framework for these tasks using a segmental hidden semi-Markov model (HSMM), enhancing the accuracy and efficiency of failure prognostics by incorporating state duration modeling and a modified forward-backward algorithm for parameter estimation. Similarly, (Tobon-Mejia et al., 2010) propose a data-driven diagnostic and prognostic method based on Mixture of Gaussians Hidden Markov Models (MoG-HMM), which processes sensor data to model component degradation and predict remaining useful life (RUL). Another article discusses the development of a Hidden Markov Model with auto-correlated observations (HMM-AO) for predicting the remaining useful life (RUL) of manufacturing systems (Chen et al., 2019). This model, which incorporates previous observations to improve accuracy also proposes an improved maintenance policy based on RUL predictions and demonstrates its effectiveness through a case study. More recently, (Zhao, Shi, & Wang, 2020) developed a method for diagnosing and predicting bearing faults using an HMM with multiple features, including time domain, frequency domain, and wavelet packet decomposition, combined with PCA for dimensionality reduction. Another work, (Martins, Fonseca, Farinha, Reis, & Cardoso, 2021), involved developing an improved Degenerated Hidden Markov Model (DGHMM) with a quasi-power relation to better describe gradual performance degradation over time. It uses an improved genetic algorithm for parameter estimation, overcoming the limitations of the conventional EM algorithm. Additionally, a greed approximation algorithm based on the Viterbi algorithm is proposed to predict the residual life of the system. This method is validated with data from caterpillar hydraulic pumps, showing effectiveness in diagnosing system health and predicting residual life.

A Gradient Boosting Regressor is an ensemble learning technique used for regression tasks that builds a series of weak learners, in a sequential manner (Zemel & Pitassi, 2000). The process begins with an initial prediction, often the mean of the target values. The algorithm then calculates the residuals, which are the errors between the actual target values and the predictions made by the current model. A new model is fitted to these residuals with the goal of predicting them. The predictions from the new model are added to the previous predictions, and scaled by a learning rate to control the contribution of each model. This step updates the model to improve accuracy incrementally (Friedman, 2001). Common GBR models are decision tree-based where the described process is repeated for a predefined number of iterations or until the residuals are minimized, with each new tree helping to correct the errors made by the existing ensemble of trees. The final prediction is the sum of the initial prediction and the weighted contributions from all the trees (Friedman

& Meulman, 2003). Our research employs a decision tree-based GBR given its recorded success in performing complicated regression tasks (Huang et al., 2019) and in prognostics (Yang et al., 2020). Furthermore, it is particularly suitable for direct Remaining Useful Life (RUL) prediction for several reasons. It handles non-linearity effectively, which is crucial since RUL prediction often involves complex, non-linear relationships between the input variables and the remaining useful life. By focusing on the residuals in each step, the model continuously refines its predictions, leading to high accuracy. It also offers flexibility with different types of features and performs automatic feature selection, which is useful when dealing with a variety of sensor data (Huang et al., 2019). Additionally, parameters like learning rate and tree depth control overfitting, making the model robust for practical applications (Friedman & Meulman, 2003). Finally, Gradient Boosting can be scaled to handle large datasets, which is essential in industrial settings where data from multiple sensors over long periods is available (Huang et al., 2019).

Due to the unavailability of industry sensor data for our study, we utilized simulated data generated by the Asherah Nuclear Power Plant Simulator (ANS) model (Busquim e Silva et al., 2021). Originally designed by the International Atomic Energy Agency (IAEA) for cyber security simulations, ANS is versatile and can simulate Pressurized Water Reactor (PWR) behaviors and transients based on time series of operating conditions. This adaptability makes it an excellent tool for our research. Previous studies have also employed the Asherah model to generate data for prognostics using particle filtering (Xiao et al., 2023), underscoring its capability to provide reliable and relevant simulation data for advanced analytical purposes.

3. DATA

ANS (Busquim e Silva et al., 2021) was used to simulate condenser performance under different fouling conditions. We manipulated the final fouling thickness (2.5 mm, 3.0 mm, and 4.0 mm) in the condenser tubes to represent varying degradation rates. Since thicker fouling implies faster degradation, these values allowed us to respectively simulate scenarios with different degradation severities (low (NR_25), moderate (NR_30), and high (NR_40)). Although ANS can simulate immediate component repairs, none were performed in this study (denoted by "NR" for No Repair) to isolate the effects of progressive tube fouling on condenser performance.

To ensure that the simulated data reflects realistic operating conditions, several measures were taken. A consistent power profile, similar to plants with stable industrial demands, was used across all simulations. Each simulation lasted 10,000 seconds, with multiple runs at each fouling severity level to generate substantial data. Sensor data was collected at 10Hz, yielding 100,000 data points per simulation. To enhance data clarity, mean filtering with a window size of 100 (without overlap) reduced the data points to 1,000, retaining essential trends. The ANS model, initially for cybersecurity assessments, lacks inherent variability within simulations. To mimic realistic fluctuations from measurement errors and process variability, white noise was added at various points in the ANS model.

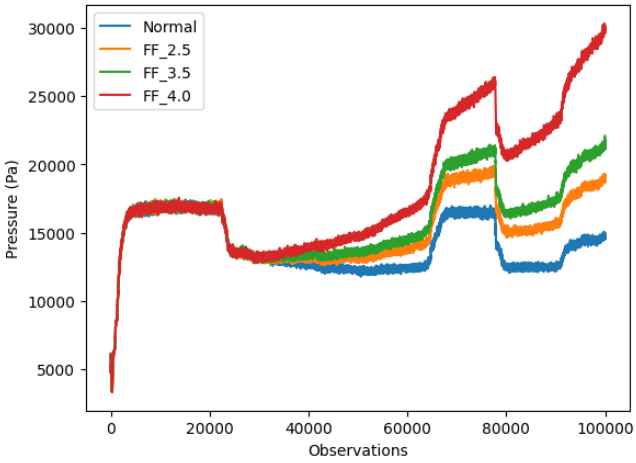

Figure 1. Pressure profile for different fouling rates

In this study, normal operation spans a complete 10,000-second sequence without degradation signs, while faulty operation involves degradation due to fouling. Degradation began at observation 30,000 (3000 seconds) in all tests. Figure 1 displays the unfiltered pressure profiles for components with different degradation rates, illustrating the potential of fault detection, diagnostics, and prognostics through raw signal analyses. Noticeable deviations occur around 30,000, with the fastest degradation showing the most significant departure from the normal profile.

During each simulation, four properties were measured namely input power (% full power), river temperature (K), condenser temperature (K), and condenser pressure (Pa). In other words, each simulation has an output array of shape $(1000, 4)$ after filtering. There are thirty simulation outputs in each degradation category but only two were used per degradation category for the inference-based method. However, all thirty samples were used for the direct RUL prediction method. The training strategy and the definition of a sample in each approach are elucidated in their respective subsections under section 4.

4. METHODOLOGY

4.1. Inference from Health

The implementation of the HMM considered in this work takes a single variable as input. Therefore, signals must be preprocessed to generate a parameter that is strictly corre-

lated to the degradation. We define a prognostics parameter as a composite measure derived from multiple sensed or inferred degradation indicators that characterize the system's progression toward failure. The signals must be combined to ensure that the parameter exhibits monotonicity, prognosability, and trendability. These qualities guarantee that the parameter is consistent as the degradation accumulates, has a predictable failure threshold, and shows a similar degradation pattern across different items of the same kind, facilitating accurate prediction of the remaining useful life (RUL) (Coble & Hines, 2009).

Therefore, the four output signals of the Asherah model are initially analyzed through an auto-associative kernel regression. Given a set of signal values collected from a degraded system, the AAKR can reconstruct a correspondent set of values as if the system was not degraded and compute the residuals between the two sets. This model does not require training but directly relies on data collected from the healthy condenser to detect faulty operations. A Sequential Probability Ratio Test (Wald & Wolfowitz, 1948) examines the generated residuals for fault detection. Once a fault is detected, a prognostic parameter is computed by subtracting the residual of the pump inlet temperature (river temperature) from the condenser temperature residual. The parameter can then be fed into the prognostic phase handled by a HMM.

The next section delineates the structure of a general Hidden Markov Model (HMM) and explains the specifics to make it Gaussian and Left-Right. Then, the algorithm to train it is briefly described. The method to estimate the health and to predict the Remaining Useful Life (RUL) of a test sequence is outlined in section 4.1.2.

4.1.1. The left-right HMM

The HMM (Rabiner, 1989) is an extension of a Markov model in which the states of the process are not directly measurable and can only be inferred through the analysis of the observed signal. The relationship between the signal values and the hidden states are modeled through conditioned probabilistic functions. In other words, the observed signal, which can be a sensor reading or an inferred degradation indicator, is a random variable whose probability density function (pdf) depends on the current hidden state of the underlying Markov model.

A standard HMM consists of a set of elements that are defined below:

- $Q = \{q_1, \ldots, q_i, \ldots, q_N\}$: a set of N hidden states,
- $A = \{a_{11}, \ldots, a_{ij}, \ldots, a_{NN}\}$: a transition probability matrix from any state i to any state j,
- $O = \{o_1, \ldots, o_t, \ldots, o_T\}$: a sequence of T observations (e.g., time series of prognostic parameter values),
- $B = \{b_1(o_t), ..., b_i(o_t), ..., b_N(o_t)$: a set of observation likelihoods (or emission probability densities) that can be discrete or continuous,
- $\pi = \{\pi_1, \pi_2, \ldots, \pi_N\}$: an initial probability distribution of the hidden states.

The HMMs inherit the main property of the Markov Models, known as the Markov assumption: $P(q_i \mid q_1, \ldots, q_{i-1}) = P(q_i \mid q_{i-1})$. Another assumption specific to HMMs is the output independence: $P(o_i \mid q_1, \ldots, q_N, o_1, \ldots, o_T) = P(o_i \mid q_1)$.

Since the analyzed industrial prognostic signal takes continuous values, the observation likelihoods are chosen to be continuous and Gaussian. Even though the likelihood function can be selected as any statistical distribution, we decided to select the Gaussian distribution due to its simplicity. Given the regular mathematical nature of the prognostic parameters, this choice still ensures reliable and efficient predictions. Furthermore, according to the nature of degradation processes, components cannot recover from a bad health condition to a better one over time without intervention. This property is translated by imposing: $a_{ij} = 0$ for $j < i$. The index of states numerically proceeds from left to right; that's why the employed model is called Left-Right HMM.

The training step is performed employing the Baum-Welch algorithm (B-W) (Rabiner, 1989) to train a single model λ. Given an observation sequence O and the initialization of the HMM parameters, it provides an estimate of the A and B that maximize the likelihood $L(O \mid \lambda)$ (Rabiner, 1989). A limitation of the B-W procedure is that it finds a parameter set for λ that is, in general, a local maximum for a given observation sequence O. It has been observed that the algorithm is sensitive to all the parameters' initialization (Sá et al., 2021); thus, the global maximum is reached only if the initialization is adequately performed.

4.1.2. Predicting RUL with HMMs

The prediction step employs the state-based RUL estimate method explained in (Chen et al., 2019). Briefly, given a new observation sequence $O_{\text{test}} = \{o_1, \ldots, o_t, \ldots, o_{T_{\text{test}}}\}$ and a trained HMM model λ, the probability for the system being in each discrete state at the current time is calculated. Then, a discrete probability distribution for $RUL(T)$ is calculated estimating the remaining number of time steps to reach the failure state q_N from the current time T. The point estimate for the RUL is then calculated by averaging the distribution. Since the model needs some data to compute reliable estimates of the current state probabilities and future states, in our implementation, the first prediction will be made when the test sequence has more than 10% (70) of observations after fault injection (700). Subsequently, the test sequence is updated with the true health parameter's values before the next prediction.

4.1.3. Initialization of model parameters

The parameters of Hidden Markov Models (HMMs) must be set or initialized before training. These operations must be done separately for each HMM (Zanotelli, Hines, & Coble, 2024). For a Left-to-Right Gaussian HMM (LR-GHMM), it's generally recommended to set the initial probabilities with $\pi_1 = 1$ and $\pi_i = 0$ for $i > 1$, as the state sequence typically begins in state 1 and ends in state N. Although initializing π to match a test sequence's starting values was attempted, it showed no significant improvement in predictions, as HMMs can quickly adjust to the correct hidden state after a few observations.

Initializing observation likelihoods is crucial for achieving a global likelihood maximum during training. While complex methods exist, for univariate and monotonic prognostics parameters, simpler initializations are effective. Here, means are evenly spaced between the upper and lower bounds of the training sequence, and standard deviations are set to half the distance between adjacent state means.

Transition probabilities were initialized as:

$$a_{i,i+1} = \frac{N}{T}, \quad \text{for } i = 1, \ldots, N-1, \tag{1}$$

where T is the number of training sequence observations. This approach assumes the number of transitions equals the number of hidden states.

Choosing the number of states in an HMM is challenging. The Bayes Information Criterion (BIC) was employed here and adapted to the current notations as:

$$\text{BIC} = k \cdot \ln(T_{train}) - 2 \cdot \ln(P(\boldsymbol{O_{train}} \mid \lambda)) \tag{2}$$

where k is the number of parameters estimated by the model. In the case of a LR-GHMM, k is the sum of the number of means (N), variances (N), and transition probabilities (N-1). The goal is to increasing the number of states until the BIC measure reaches a minimal plateau.

The final parameter is the length scale s of the Gaussian membership function, crucial in the aggregation procedure. For unsmoothed data, setting $s^2 = 2\sigma^2_{\text{noise}}$ is reasonable. When smoothing, dividing s by a factor between 5 and 10 can improve results.

4.2. Direct Prediction

The other approach this research takes to estimate RUL involves directly mapping it to two input variables: pressure and residual pressure. To put it differently, each sample sequence has shape $(1000, 2)$. The residual pressure was not directly measured from the simulation but was calculated as the pressure difference between the expected normal operation and the actual faulty operation of the component throughout the investigated period. This residual pressure served as the health index in this approach and was used to calculate the target variable (true RUL). Different parameter thresholds were employed to define failure in each degradation category. For the least degraded components, the threshold was set at 4,000 Pa, while for the moderate category, it was 6,500 Pa. Lastly, the most degraded condenser tubes had a threshold of 14,000 Pa. These thresholds marked failure points or end-of-life (EOL) to calculate the true RUL of the components in each category. The process involved assigning decreasing window values to the data points within a sample's sequence, starting with the highest value (1000) at the initial time step. Subsequently, the assigned window values were subtracted from EOL to derive the true RUL array. Any resulting negative values, occurring beyond the point of failure, were set to '0'. The true RUL before fault injection at 3000s is undefined.

This method leverages the relationship between the input variables and the condenser's remaining useful life. For this approach, a machine learning (ML) algorithm was selected because ML algorithms can effectively capture complex patterns and relationships in the data that traditional statistical methods might miss. The selected ML model, as alluded to in section 2, is a Gradient Boosting Regression Decision Tree (GBRDT). The applied Scikit learn GBRDT has 30 decision trees and an optimally selected learning rate of '0.2'. All other hyperparameters were set as the default (Prettenhofer & Louppe, 2014). Twenty-nine ('29') sequences reshaped to $(29000, 2)$ for proper GBRDT processing were used to train the model in each category, and one ('1') sample of shape $(1000, 2)$ was used to test. In the 'NR_25' category, one ('1') bad simulation that failed earlier than others was included to test the robustness of GBRDT. To add to that, we conducted thirty ('30') rounds of model initialization, training, and prediction, with a new random seed generated for the model each time. This approach introduces variability in the final parameters of the GBRDT after training, allowing us to further assess the robustness of the predictions.

4.3. Metrics

This paper employs two broad classes of metrics to analyze the performance of the different models and the effectiveness of the various approaches namely visual and numerical metrics. Since separate pipelines are compared, it is crucial to select metrics that are method-independent.

Two accuracy-based numerical metrics were resorted to in this paper. One focuses on the general error of predictions throughout the condenser's lifetime whereas the other accounts for how early or late the prediction is. Both errors are normalized by the variance of the true RUL values, so they can be used to compare methods that use different RUL scales. These two numerical metrics are the normalized sum of weighted errors (NSWE) adapted from (Saxena, Goebel,

et al., 2008), and normalized mean squared error (NMSE) adapted from (Saxena, Celaya, et al., 2008).

For NSWE, we first compute the difference d between the predicted RUL and the true RUL as shown in Eq. 3a. Next, we apply a piecewise function to compute the weight based on the sign of d, as described in Eq. 3b. If d is negative (early prediction), the weight is calculated using the formula $\exp\left(-\frac{d}{a1}\right)$, where $a1$ is set to 100. If d is non-negative (late prediction), the weight is calculated using $\exp\left(\frac{c \cdot d}{a1}\right)$, where c, the penalty severity parameter, is set to 2. Subsequently, we calculate the sum of weighted errors (swe) by summing the product of the absolute error and the corresponding weight for all data points, as shown in Eq. 3c. Finally, to account for variability in the true RUL values, we normalize the sum of weighted errors by dividing it by the variance of the true RUL values in Eq. 3d.

$$d_i = \text{predicted\_rul}_i - \text{true\_rul}_i \tag{3a}$$

$$\text{weight}_i = \begin{cases} \exp\left(-\frac{d}{a1}\right) & \text{if } d_i < 0 \\ \exp\left(\frac{c \cdot d}{a1}\right) & \text{if } d_i \geq 0 \end{cases} \tag{3b}$$

$$\text{swe} = \sum(|d_i| \times \text{weight}_i) \tag{3c}$$

$$\text{NSWE} = \frac{\text{swe}}{\text{var(true\_rul)}} \tag{3d}$$

$$NMSE = \frac{MSE}{\text{var(true\_rul)}} \tag{4}$$

Finally, MSE in Eq. (4) refers to the mean squared error, calculated as the average of the squared differences between the predicted and actual RUL values. Because of the normalization by variance, $NMSE$ has no units but $NSWE$ has the unit 'per_window' or $(/w)$.

In addition to the accuracy-based metrics, two computational metrics were utilized for comparison: model training time and model prediction time, both measured in seconds. The total CPU processing time was chosen as the primary metric because it accurately reflects the actual computational effort required for model training and prediction, independent of other tasks the computer might be handling or the system's configuration (Saxena, Celaya, et al., 2008). This focus ensures a fair comparison of computational efficiency across different environments. Measuring training time provides valuable insights into the computational demands of model development, revealing how efficiently a model can be trained. This is particularly important in real-world applications where models may need frequent retraining due to new incoming data. Shorter training times can lead to more agile and responsive systems. Prediction time, on the other hand, indicates the efficiency of the model in making predictions once trained. In scenarios requiring real-time or near-real-time predictions, such as predictive maintenance or fault detection in industrial settings, minimizing prediction time is crucial for timely decision-making and intervention. By evaluating both training and prediction times, we gain a comprehensive understanding of the model's operational performance, balancing the trade-offs between computational cost and predictive accuracy. This dual analysis helps in selecting models that not only provide accurate predictions but also operate within acceptable computational limits, ensuring practicality and efficiency in deployment.

To evaluate the models' ability to capture the overall trend and progression of the RUL, we examined the mean RUL prediction at each time step throughout the component's life (see Fig. 2). This comparison helps identify any systematic biases or discrepancies among the models. We also analyzed the spread of predictions at each time step to assess the models' consistency and variability (see Fig. 3). This analysis provides insights into the robustness and reliability of the models' predictions. To determine the reasonableness of the predictions, we plotted an alpha bound around the true RUL values, representing a +/- 10% range of the true value. This range, though arbitrary, serves as a threshold to evaluate if the predictions fall within an acceptable range. This qualitative assessment, which considers both the mean RUL estimation and the spread of predictions, offers a comprehensive understanding of the models' performance, complementing the numerical metrics.

5. RESULTS

This section presents the findings of our study using the methods and metrics described in section 4. For visual assessment, we focus on the final 7,000 seconds (700 data points) of operation following fault injection, as this period encompasses the defined degradation phase. The preprocessing for the inference-based prognostics necessitated excluding periods prior to fault detection to enhance performance. To maintain uniformity and facilitate ease of comparison, the direct prediction method results were also limited to this region. By concentrating on this critical phase, we ensure a clear and consistent comparison of the prediction profiles, highlighting the performance and reliability of both prognostic approaches under similar conditions.

5.1. Visual Assessment

In the NR_25 category, one bad training sequence was included in the training data for the GBRDT while the HMM was trained as described in section 4.1. As earlier mentioned, 30 predictions were made per prediction point after re-initializing and retraining the GBRDT. The means of these predictions are shown in Fig. 2a. The HMM inherently produces a distribution for each prediction point and the means of these distributions are reported in Fig. 2b. From Fig. 2, it is

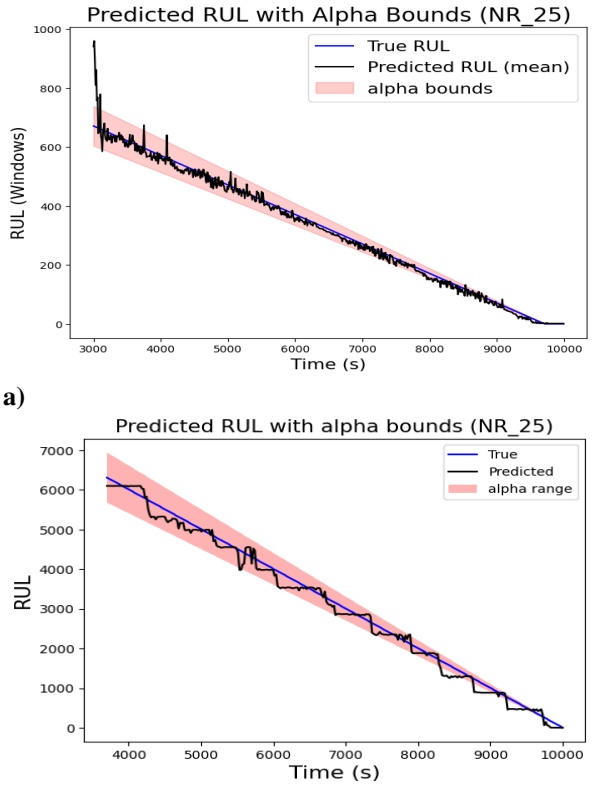

Figure 2. Mean predictions for final fouling 2.5mm. (a) GBRDT. (b) HMM.

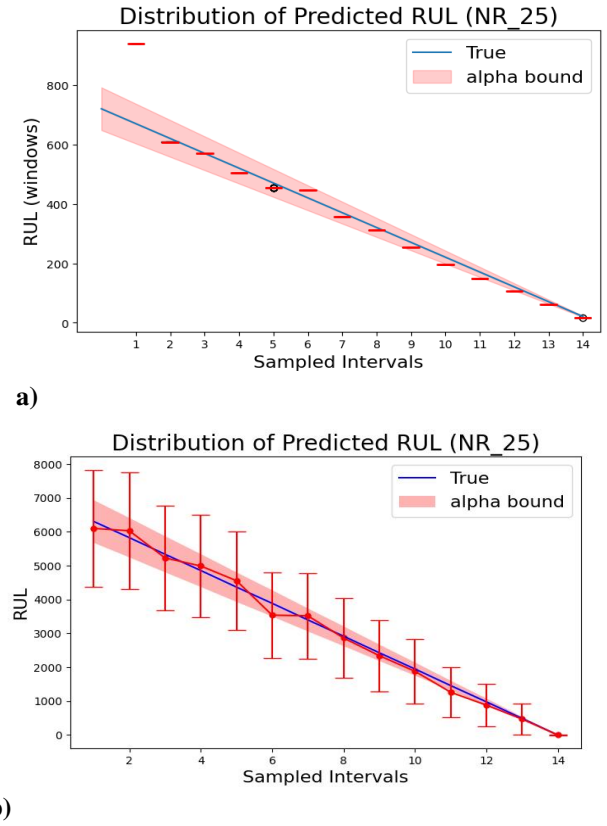

Figure 3. Distributed predictions for final fouling 2.5mm. (a) GBRDT. (b) HMM.

seen that GBRDT's predictions more closely follow the true RUL line indicating a better overall prediction in the NR_25 category. The shorter prediction period for the HMM (Fig. 2b) is because prediction starts around 3,700 seconds (observation 370) for reasons explained in section 4.1.2.

Analyzing the distribution of these NR_25 predictions at evenly spaced intervals (see Fig. 3), we observe that the whiskers of the box plots are not visible indicating that most GBRDT predictions are around the median values. This suggests a very stable and repeatable performance by the GBRDT. As seen, these median values mostly fall in the defined alpha range of the true RUL (see Fig. 3a) connoting low errors. The HMM model, on the other hand, displays a reasonably high prediction variability early on in the operational life of the condenser but the stability of predictions improves closer to the component's EOL (see Fig. 3b). The reduced variability close to the EOL is highly favorable since errors around this region have higher consequences. The median values of HMM's predictions also closely follow the true line and mostly fall within the alpha range.

For NR_30, similar results are seen in Fig. 4 as observed in the NR_25 category where GBRDT predicted RUL better than HMM. In addition, the GBRDT predictions appear less erratic, especially around the fault injection point (3000s). This could be because higher degradations make performance decay more profound and easier to identify. The distributions in Fig. 5 show the same trend of low spread closer to the EOL for HMM and high stability across time steps for GBRDT. The median values of the distributions also closely align with the true RUL line and fall comfortably within the alpha range.

For the third and most degraded component tested, a similar trend is seen as in the previous categories. Again, the GBRDT shows smaller deviations of predictions from the true values (Fig. 6) and promises higher repeatability because of smaller prediction variability than HMM (see Fig. 7) over the condenser's lifetime. like in the NR_30 category, the stable and relatively accurate prediction of GBRDT close to fault onset may be attributed to higher degradation. The reduced prediction spread of the HMM toward the EOL in all the degradation categories is expected because the model continuously takes more input to make predictions as it moves from fault point to EOL. The increased input likely aids in better pattern identification and reduces uncertainty.

Upon further investigation of the GBRDT, it was revealed that the decision trees predominantly relied on the pressure residual to predict RUL. Figure 8 shows that the monitored

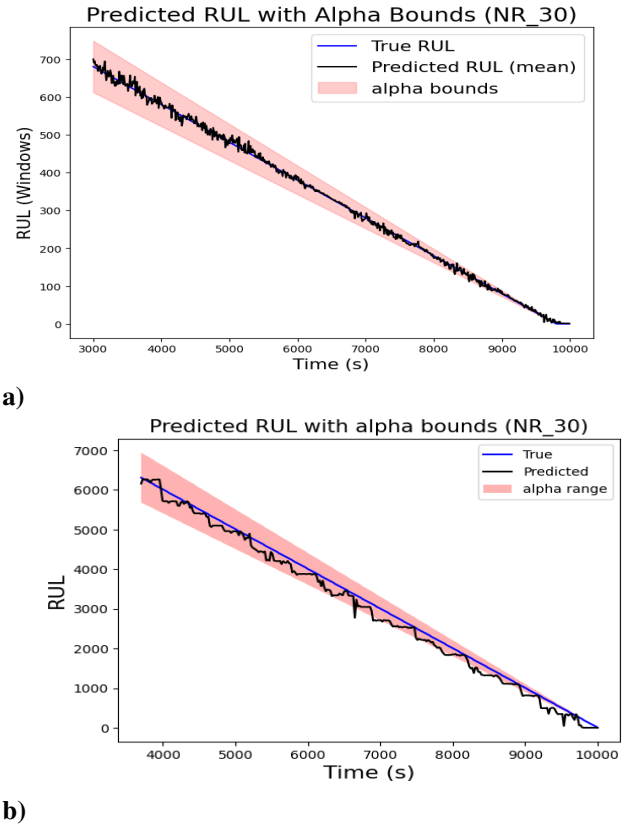

Figure 4. Mean predictions for final fouling 3.0mm. (a) GBRDT. (b) HMM.

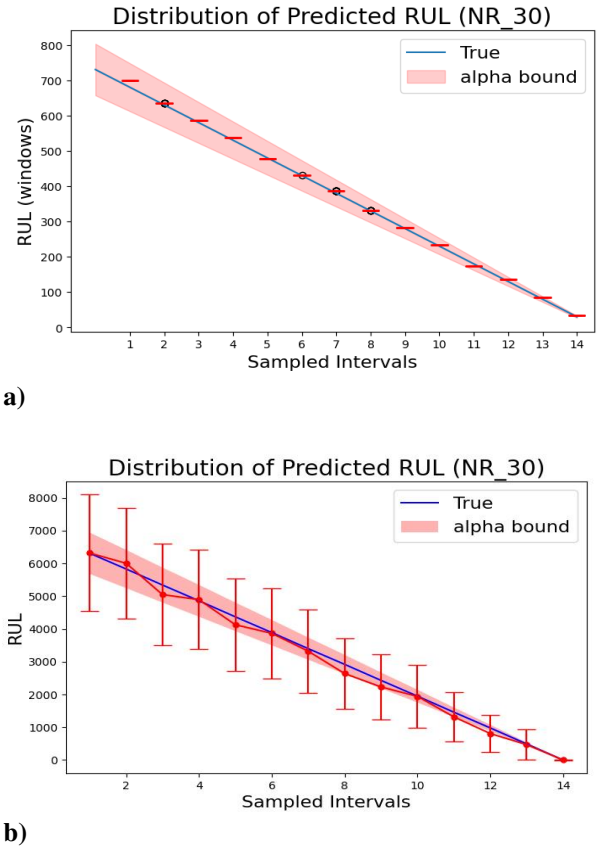

Figure 5. Distributed predictions for final fouling 3.0mm. (a) GBRDT. (b) HMM.

condenser pressure contributed only about 5% to the predictive capacity of the GBRDT, whereas the pressure residual accounted for approximately 95%. To explore the impact of this reliance, a reduced GBRDT was trained using only the most important feature, resulting in the mean prediction profile shown in Figure 9. As observed, the predictions deteriorated significantly, particularly toward the highly consequential EOL phase. This deterioration underscores the critical role of the raw pressure signal in providing essential context for accurate RUL estimation. Although it contributes only 5%, this signal is vital for capturing the nuances of the degradation process, highlighting the importance of incorporating multiple features for robust RUL prediction of the investigated condenser.

5.2. Numerical Assessment

From section 5.1, it is evident that predictive performances differ between the two approaches, with direct RUL prediction appearing superior. However, quantifying the performance gap can be challenging. In this section, we present our findings on measuring the performance gap between the two methods using the quantitative metrics discussed in section 4.3. For all selected performance measures (see Table 1), lower values indicate better performance.

Table 1. Performance Summary

Final Fouling 2.5 mm		
	GBRDT	HMM
NSWE (/w)	1.776	7.78
NMSE	0.00502	2.01
Training Time (s)	1.55	0.4
Prediction Time (s)	0.0015	197.4
Final Fouling 3.0 mm		
	GBRDT	HMM
NSWE (/w)	0.591	7.72
NMSE	0.00972	2.02
Training Time (s)	1.53	0.4
Prediction Time (s)	0.0016	197.1
Final Fouling 4.0 mm		
	GBRDT	HMM
NSWE (/w)	0.564	7.70
NMSE	0.0102	2.04
Training Time (s)	1.56	0.4
Prediction Time (s)	0.0014	198.1

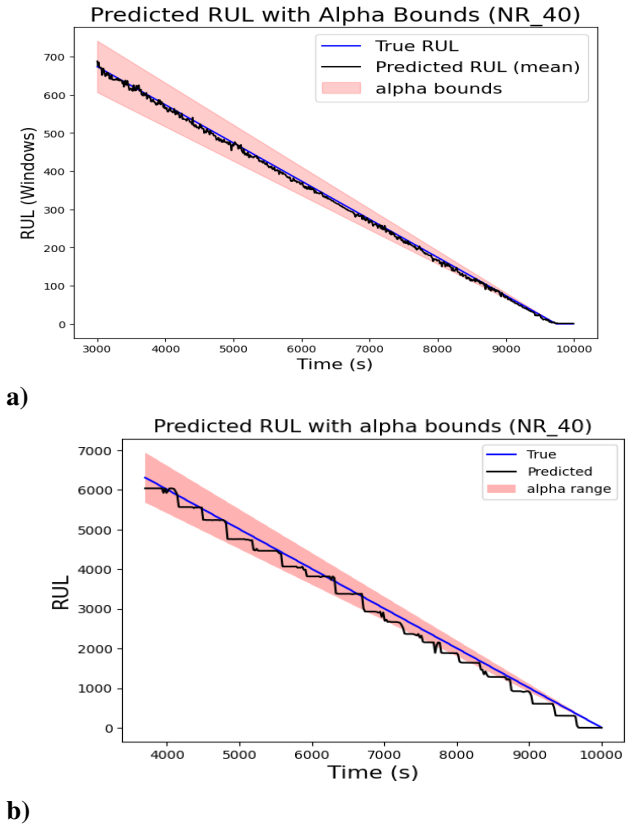

Figure 6. Mean predictions for final fouling 4.0mm. (a) GBRDT. (b) HMM.

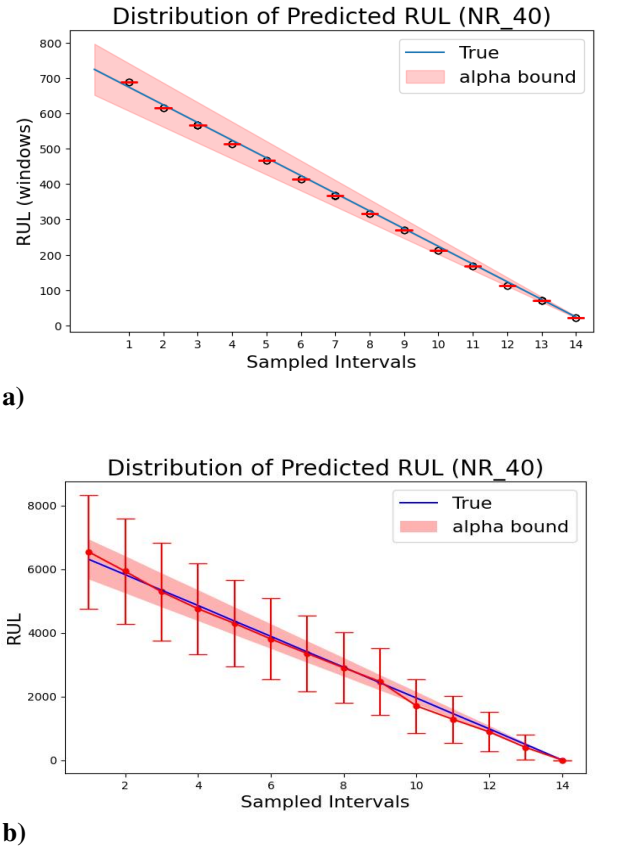

Figure 7. Distributed predictions for final fouling 4.0mm. (a) GBRDT. (b) HMM.

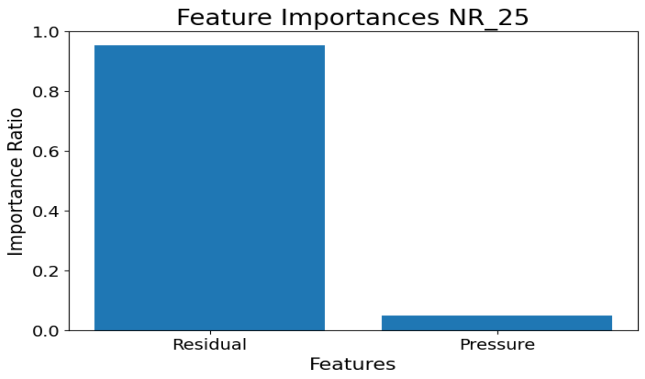

Figure 8. Importance of features (GBRDT)

Focusing on the computational metrics (Training Time and Prediction Time), each model's performance remained consistent across different degradation categories. This consistency is expected since computation largely depends on data size and model complexity, which were uniform across all categories. The training time of the HMM was much shorter than that of the GBRDT because the HMM was trained with a single input sequence of the health parameter, whereas the GBRDT processed twenty-nine (29) sequences with two (2) features during training. The GBRDT excelled in prediction

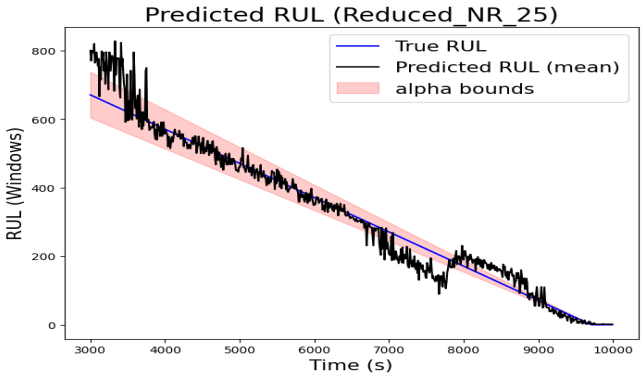

Figure 9. Reduced GBRDT mean predictions. Final fouling 2.5 mm.

speed, as it simultaneously predicted all time steps. In contrast, the HMM predicted one time step at a time, processing an updated set of inputs for each prediction, as described in section 4.1.2. About 700 time steps were predicted in both approaches meaning that the HMM took approximately 0.283 seconds per time step in the NR_40 category, for example,

which is still significantly slower than the GBRDT's 0.0014 seconds per time step.

Given these results, the HMM might be preferred in scenarios where rapid training is crucial, such as in applications requiring regular model retraining with new data. Conversely, the GBRDT's faster prediction speed makes it more suitable for real-time or near-real-time applications, where quick decision-making is essential. This dual analysis highlights the importance of choosing a model based on the specific computational requirements and operational constraints of the intended application.

In terms of accuracy, $NSWE$ and $NMSE$ underscore the superiority of GBRDT over HMM in all the tested categories. For the direct RUL prediction, the higher degradations (NR_30 and NR_40) show similar errors and improved accuracy compared to the low degradation especially when late predictions are punished ($NSWE$). For the inference-based method, similar errors are seen in all categories for both accuracy-based metrics. This is not surprising because the approach of feeding inputs in batches during prediction limits HMM's predictive performance, particularly in the early operational life of the condenser regardless of the degradation level.

5.3. Applicability and Limitations

As demonstrated by the results of our experiments in sections 5.1 and 5.2, the proposed data-driven approaches can be preferred under different real-world situations. In this section, we discuss the applicability and limitations of the compared specific methods.

5.3.1. Inference from Health Using HMMs

The HMM-based approach is particularly well-suited for systems where degradation is progressive and can be monitored through observable signals. Industries such as aerospace, automotive, and manufacturing, where equipment maintenance is critical, can benefit significantly from this method. The Left-Right HMM model is especially advantageous in scenarios where the progression of degradation is unidirectional and irreversible, such as wear and tear in mechanical systems or the gradual depletion of material in chemical processes both applicable in NPPs and other energy and industrial systems. The requirement for monotonicity, prognosability, and trendability in the prognostic parameters ensures that the model can consistently track and predict degradation across different units of the same equipment type, leading to reliable maintenance decisions.

However, the HMM approach also has some limitations. The assumption of output independence, while simplifying the model, may not always hold in real-world scenarios where multiple factors influence the system's state simultaneously. Additionally, the model's reliance on Gaussian distributions for the observation likelihoods may limit its effectiveness in environments where the degradation indicators cannot be modeled as a set of discrete states with Gaussian observation likelihoods. Another significant limitation is the sensitivity of the Baum-Welch algorithm to the initial parameter values. If the initialization is not carefully performed, the algorithm may converge to a local maximum, resulting in suboptimal predictions.

5.3.2. Direct Prediction Using GBRDT

The direct prediction method using GBRDT is highly applicable in situations where the relationship between input variables (such as pressure and residual pressure) and the RUL is complex and non-linear. The use of machine learning allows the model to capture intricate patterns in the data that traditional statistical methods might overlook. This approach is particularly useful in industries where real-time monitoring and rapid predictions are critical, such as in power generation or transportation. The flexibility of GBRDT in handling various types of input data, combined with its robustness in the presence of noisy data, makes it a powerful tool for predictive maintenance. The ability to incorporate multiple rounds of training and testing with different random seeds further enhances the model's reliability, ensuring that predictions are not overly sensitive to specific data configurations.

Despite its advantages, the GBRDT approach also faces some challenges in real-world applications. The method's reliance on accurately calculated residual pressure as a health index means that any errors in this calculation can significantly affect the model's performance. The model's complexity and the need for extensive computational resources for training can also be a drawback in some industrial environments, particularly where real-time predictions are required. Additionally, while GBRDT is effective in capturing complex relationships, it may not always provide interpretable results, making it difficult for operators to understand the underlying reasons for a particular prediction.

6. CONCLUSION

This study compared two data-driven prognostic models to estimate the RUL of NPP condensers with tube fouling. The second approach, GBRDT, demonstrated superior accuracy and lower prediction variability, particularly excelling in scenarios with higher degradation rates. Its ability to quickly and accurately map input variables to RUL makes it suitable for real-time applications.

While the HMM approach showed lower prediction accuracy, higher prediction variability, and longer prediction times, it provided valuable insights into the degradation mechanisms and offered more interpretability as well as quicker training for applications that require periodic updating of the model

with new data.

Our research highlights the importance of selecting the appropriate model based on the application's specific requirements and constraints. Future work should explore integrating these methods to leverage the strengths of both approaches and apply them to other critical components within NPPs and different industrial contexts.

ACKNOWLEDGMENT

This material is based upon work supported by the Department of Energy under Award Number DE-NE0009278.

REFERENCES

An, D., Choi, J. H., & Kim, N. H. (2013). Options for prognostics methods: A review of data-driven and physics-based prognostics. In *54th aiaa/asme/asce/ahs/asc structures, structural dynamics, and materials conference* (p. 1940).

Busquim e Silva, R., Piqueira, J., Cruz, J., & Marques, R. (2021). Cybersecurity assessment framework for digital interface between safety and security at nuclear power plants. *International Journal of Critical Infrastructure Protection*, *34*, 100453. doi: https://doi.org/10.1016/j.ijcip.2021.100453

Chen, Z., Li, Y., Xia, T., & Pan, E. (2019). Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy. *Reliability Engineering and System Safety*, *184*(C), 123-136. doi: 10.1016/j.ress.2017.09.00

Coble, J., & Hines, J. (2009, 01). Fusing data sources for optimal prognostic parameter selection. *Transactions of the American Nuclear Society*, *100*, 211-212.

Coble, J., & Hines, J. (2011, 01). Applying the general path model to estimation of remaining useful life. *International Journal of Prognostics and Health Management*, *2*, 2153-2648. doi: 10.36001/ijphm.2011.v2i1.1352

Cubillo, A., Perinpanayagam, S., & Esperon-Miguez, M. (2016). A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery. *Advances in Mechanical Engineering*, *8*(8), 1687814016664660.

Cui, L., Li, W., Wang, X., Zhao, D., & Wang, H. (2022). Comprehensive remaining useful life prediction for rolling element bearings based on time-varying particle filtering. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-10. doi: 10.1109/TIM.2022.3163167

Dong, M., & He, D. (2007). A segmental hidden semi-markov model (hsmm)-based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing*, *21*(5), 2248-2266. doi: https://doi.org/10.1016/j.ymssp.2006.10.001

Elattar, H. M., Elminir, H. K., & Riad, A. (2016). Prognostics: a literature review. *Complex & Intelligent Systems*, *2*(2), 125–154.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Friedman, J. H., & Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in medicine*, *22*(9), 1365–1381.

Huang, Y., Liu, Y., Li, C., & Wang, C. (2019). Gbrtvis: online analysis of gradient boosting regression tree. *Journal of Visualization*, *22*, 125–140.

Ibrahim, S. M., & Attia, S. I. (2015). The influence of condenser cooling seawater fouling on the thermal performance of a nuclear power plant. *Annals of Nuclear Energy*, *76*, 421-430. doi: https://doi.org/10.1016/j.anucene.2014.10.018

Ifeanyi, A. (2024). A graph neural network approach to system-level health index and remaining useful life estimation. In *16th annual conference of the prognostics and health management society (phmconf).*

Ifeanyi, A. O., Coble, J. B., & Saxena, A. (2024). A deep learning approach to within-bank fault detection and diagnostics of fine motion control rod drives. *International Journal of Prognostics and Health Management*, *15*(1).

Khelif, R., Chebel-Morello, B., Malinowski, S., Laajili, E., Fnaiech, F., & Zerhouni, N. (2016). Direct remaining useful life estimation based on support vector regression. *IEEE Transactions on industrial electronics*, *64*(3), 2276–2285.

Kundu, P., Darpe, A. K., & Kulkarni, M. S. (2020). An ensemble decision tree methodology for remaining useful life prediction of spur gears under natural pitting progression. *Structural Health Monitoring*, *19*(3), 854–872.

Ma, Y., Yao, M., Liu, H., & Tang, Z. (2022). State of health estimation and remaining useful life prediction for lithium-ion batteries by improved particle swarm optimization-back propagation neural network. *Journal of Energy Storage*, *52*, 104750.

Martins, A., Fonseca, I., Farinha, J. T., Reis, J., & Cardoso, A. J. M. (2021). Maintenance prediction through sensing using hidden markov models—a case study. *Applied Sciences*, *11*(16).

Patil, S., Patil, A., Handikherkar, V., Desai, S., Phalle, V. M., & Kazi, F. S. (2018). Remaining useful life (rul) prediction of rolling element bearing using random forest and gradient boosting technique. In *Asme international mechanical engineering congress and exposition* (Vol. 52187, p. V013T05A019).

Peng, W., Ye, Z.-S., & Chen, N. (2019). Bayesian deep-learning-based health prognostics toward prognostics uncertainty. *IEEE Transactions on Industrial Electronics*, *67*(3), 2283–2293.

Prettenhofer, P., & Louppe, G. (2014). Gradient boosted regression trees in scikit-learn. In *Pydata 2014*.

Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257-286. doi: 10.1109/5.18626

Rezki, N., & Rezgui, W. (2024). Enhancing turbofan engines reliability through comparative analysis of svr and ls-boost for remaining useful life estimation. In *2024 2nd international conference on electrical engineering and automatic control (iceeac)* (pp. 1–4).

Sankavaram, C., Kodali, A., Pattipati, K., Singh, S., Zhang, Y., & Salman, M. (2016). An inference-based prognostic framework for health management of automotive systems. *International Journal of Prognostics and Health Management*, *7*(2).

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., & Schwabacher, M. (2008). Metrics for evaluating performance of prognostic techniques. In *2008 international conference on prognostics and health management* (pp. 1–17).

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management* (pp. 1–9).

Song, K., & Cui, L. (2022). A common random effect induced bivariate gamma degradation process with application to remaining useful life prediction. *Reliability Engineering & System Safety*, *219*, 108200. doi: https://doi.org/10.1016/j.ress.2021.108200

Song, L., Gui, X., Du, J., Fan, Z., Li, M., & Guo, L. (2024). A novel transfer learning approach for state-of-health prediction of lithium-ion batteries in the absence of run to failure data. *IEEE Transactions on Instrumentation and Measurement*.

Song, Y., Li, L., Peng, Y., & Liu, D. (2018). Lithium-ion battery remaining useful life prediction based on grurnn. In *2018 12th international conference on reliability, maintainability, and safety (icrms)* (pp. 317–322).

Sundar, S., Rajagopal, M. C., Zhao, H., Kuntumalla, G., Meng, Y., Chang, H. C., ... Salapaka, S. (2020). Fouling modeling and prediction approach for heat exchangers using deep learning. *International Journal of Heat and Mass Transfer*, *159*, 120112. doi: https://doi.org/10.1016/j.ijheatmasstransfer.2020.120112

Sá, A., Andrade, A., Soares, A., Naves, E., & Nasuto, S. (2021, 01). On the initialization of parameters of hidden markov models.. doi: 10.37423/210203594

Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., & Tripot, G. (2010). A mixture of gaussians hidden markov model for failure diagnostic and prognostic. In *2010 ieee international conference on automation science and engineering* (p. 338-343). doi: 10.1109/COASE.2010.5584759

Valladares, H., Li, T., Zhu, L., El-Mounayri, H., Hashem, A. M., Abdel-Ghany, A. E., & Tovar, A. (2022). Gaussian process-based prognostics of lithium-ion batteries and design optimization of cathode active materials. *Journal of Power Sources*, *528*, 231026. doi: https://doi.org/10.1016/j.jpowsour.2022.231026

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 326–339.

Wang, H., Ma, X., & Zhao, Y. (2019). An improved wiener process model with adaptive drift and diffusion for online remaining useful life prediction. *Mechanical Systems and Signal Processing*, *127*, 370-387. doi: https://doi.org/10.1016/j.ymssp.2019.03.019

Wen, Y., Rahman, M. F., Xu, H., & Tseng, T.-L. B. (2022). Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective. *Measurement*, *187*, 110276.

Xiao, H., Hines, A., Zhang, F., Coble, J. B., & Hines, J. W. (2023). Prognostics and health management for maintenance-dependent processes. *Nuclear Technology*, *209*(3), 419–436.

Xu, X., Li, X., Ming, W., & Chen, M. (2022). A novel multi-scale cnn and attention mechanism method with multi-sensor signal for remaining useful life prediction. *Computers & Industrial Engineering*, *169*, 108204.

Yang, F., Wang, D., Xu, F., Huang, Z., & Tsui, K.-L. (2020). Lifespan prediction of lithium-ion batteries based on various extracted features and gradient boosting regression tree model. *Journal of Power Sources*, *476*, 228654.

Yu, J. (2015). Machine health prognostics using the bayesian-inference-based probabilistic indication and high-order particle filtering framework. *Journal of Sound and Vibration*, *358*, 97–110.

Zanotelli, M., Hines, J. W., & Coble, J. B. (2024). Combining similarity measures and left-right hidden markov models for prognostics of items subjected to perfect and imperfect maintenance. *Nuclear Science and Engineering*, 1–15.

Zemel, R., & Pitassi, T. (2000). A gradient-based boosting algorithm for regression problems. *Advances in neural information processing systems*, *13*.

Zhang, Y., Xiong, R., He, H., & Liu, Z. (2017). A lstm-rnn method for the lithuim-ion battery remaining useful life prediction. In *2017 prognostics and system health management conference (phm-harbin)* (pp. 1–4).

Zhang, Y., Yang, T., Zhou, H., Lyu, D., Zheng, W., & Li, X. (2023). A prognosis method for condenser fouling based on differential modeling. *Energies*, *16*(16).

Zhao, W., Shi, T., & Wang, L. (2020). Fault diagnosis and prognosis of bearing based on hidden markov model with multi-features. *Applied Mathematics and Nonlinear Sciences*, *5*(1), 71–84.

Zhu, R., Chen, Y., Peng, W., & Ye, Z.-S. (2022). Bayesian deep-learning for rul prediction: An active learning perspective. *Reliability Engineering & System Safety*, *228*, 108758.

BIOGRAPHIES

Ark O. Ifeanyi graduated from the University of Benin, Nigeria with a B.Eng. in Electrical and Electronic Engineering before obtaining an MSc. in Renewable Energy Systems Technology from Loughborough University, Leicestershire, UK. He is currently pursuing a Ph.D. in Energy Science and Engineering at the Bredesen Centre for Interdisciplinary Research and Graduate Education at the University of Tennessee. His current research is focused on the application of machine learning and artificial intelligence to the Prognostics and Health Management (PHM) of nuclear plants including small modular reactors (SMRs). Specifically, he is developing advanced machine learning algorithms to detect and diagnose faults in nuclear power plants, with the ultimate goal of improving the safety, reliability, and efficiency of nuclear energy production.

Mattia Zanotelli graduated from Politecnico di Milano, Italy, with a B.Eng. in Energy Engineering and an MSc. in Nuclear Engineering with a thesis on interpretable diagnostics for energy production plants. He joined the Nuclear Department at the University of Tennessee, Knoxville, in the Fall of 2021, and he is currently pursuing a Ph.D. in Nuclear Engineering. His research has supported two projects in nuclear plant PHM, first in prognostics approaches for imperfect maintenance and second in margin-based system-level prognostics.

Jamie B. Coble is an Associate Professor of Nuclear Engineering at the University of Tennessee, Knoxville (UTK). She earned her Ph.D. in Nuclear Engineering in 2010 from UTK. Prior to joining the faculty, she was a scientist in the Applied Physics group at Pacific Northwest National Laboratory. Her research interests lie mainly in applications of data analytics and machine learning in operations and maintenance of nuclear power plants. She is a member of American Nuclear Society and U.S. Women in Nuclear, senior member of IEEE, and fellow of the Prognostics and Health Management Society and of the International Society of Engineering Asset Management.