Evaluation of time series forecasting approaches for the reliable crack length prediction of riveted aluminium plates given insufficient data

Osarenren Kennedy Aimiyekagbon[1], Amelie Bender[2], and Walter Sextro[3]

[1,2,3] *Chair of Dynamics and Mechatronics, Paderborn University, 33098 Paderborn, Germany*
osarenren.aimiyekagbon@uni-paderborn.de
amelie.bender@uni-paderborn.de
walter.sextro@uni-paderborn.de

ABSTRACT

In all fields, the significance of a reliable and accurate predictive model is almost unquantifiable. With deep domain knowledge, models derived from first principles typically outperforms other models in terms of reliability and accuracy. When it may become a cumbersome or an unachievable task to build or validate such models of complex (non-linear) systems, machine learning techniques are employed to build predictive models. However, the accuracy of such techniques is not only dependent on the hyper-parameters of the chosen algorithm, but also on the amount and quality of data. This paper investigates the application of classical time series forecasting approaches for the reliable prognostics of technical systems, where black box machine learning techniques might not successfully be employed given insufficient amount of data and where first principles models are infeasible due to lack of domain specific data. Forecasting by analogy, forecasting by analytical function fitting, an exponential smoothing forecasting method and the long short-term memory (LSTM) are evaluated and compared against the ground truth data. As a case study, the methods are applied to predict future crack lengths of riveted aluminium plates under cyclic loading. The performance of the predictive models is evaluated based on error metrics leading to a proposal of when to apply which forecasting approach.

1. INTRODUCTION

Where a crack size with a dimension of $1/32$ Inch (≈ 0.79 mm) is negligible in an eggshell according to Hendron (Hendron, 1963), the same crack size can result in catastrophic failure of other technical systems such as a rocket. Such failures are not only economically devastating but could lead to injury and loss of lives. To avert sudden failures of increasingly complex technical components or systems, and thereby improve the reliability and availability of such systems through their life-cycle, prognostics and health management (PHM) has evolved as an enabling engineering discipline. Hereinafter, a system refers to a single component or a system of components. PHM generally encompasses a diagnostics and prognostics module, where the former is carried out a posteriori, that is after a fault or failure as occurred, to detect, isolate and identify system fault(s) or failure. Given current condition monitoring data, the prognostics module involves the future state prediction of an health indicator or condition(s) relating to the degradation of such system. The prognostics module also facilitates the estimation of the remaining useful life (RUL) of technical systems when the health indicator or degradation-related condition(s) is propagated to the a priori known end of life (EOL). Over the years, several diagnostics and prognostics techniques have been developed and published in scholarly literature, ranging from physics-of-failure (PoF)-based methods to application of artificial intelligence methods. The goal is to improve accuracy and efficiency, while striving for application-independent solutions and domain expert independence. With deep domain knowledge, PoF models derived from first principles are typically reliable and accurate. When it may become a cumbersome or an unachievable task to build or validate such models of complex (non-linear) systems, artificial intelligence methods, such as machine learning techniques are employed to derive models from condition monitoring data acquired over time. Under the assumption that the data reflects several fault modes and possibly failure, degradation trends are learned from data with these techniques. However, the accuracy of such techniques is not only dependent on the hyper-parameters of the chosen algorithm, but also on the amount and quality of data. When neither a physical model nor sufficient amount of run-to-failure data is available, several methods that lay between these two extrema are

adopted to tackle these challenges. They leverage the advantages of both approaches (Elattar, Elminir, & Riad, 2016; Atamuradov, Medjaher, Dersin, Lamoureux, & Zerhouni, 2017).

The following literature review gives an overview of PoF-based through artificial intelligence based prognostics methods, while highlighting the data source, data size and application domain.

Most PoF-based methods are built upon fatigue crack growth equations such as the Paris–Erdogan law (Paris & Erdogan, 1963) which are incorporated in Bayesian-based algorithms such as widely employed Kalman or particle filter for prognostics. For example, Ray and Tangirala (Ray & Tangirala, 1996) employed the Kalman filter with the Newman model for micro-cracks to predict the RUL of aluminium specimen under varying amplitude loading based on the statistical Virkler fatigue crack growth data with a sample size of 68 specimens (Virkler, Hillberry, & Goel, 1979). A major limitation of the application of such crack growth equations is that the parameters are often geometry-dependent, material-dependent and experimentally acquired with sufficient specimens.

In cases where the underlying physics or failure mechanism may be unknown or difficult to model, analytical functions can be fitted to acquired run-to-failure data or to features extracted from such data. To estimate the RUL of the gearbox of an SH-60 helicopter, Engel et al. (Engel, Gilmartin, Bongort, & Hess, 2000) applied a polynomial function to model the temporal evolution of normalized kurtosis, root mean square (RMS), variance and the proportional energy feature value of the time synchronous average residue signal. The accelerometer data set consisted of 36 recordings of a total duration of 548 minutes from healthy to developed gear tooth crack condition.

A general path model as proposed by Lu and Meeker (Lu & Meeker, 1993) or variations thereof can be employed to derive a general model when the parameters of the analytical functions may vary for each degradation trajectory.

The artificial neural network (ANN) or its variations are one of the most applied machine learning techniques in all domains to model more complex non-linear relations or multivariate features. For fault classification and prognostics of a bearing with an inner race crack, Wang and Vachtsevanos (P. Wang & Vachtsevanos, 2001) utilized (dynamic) wavelet neural networks. Input data for the network were signal peak values and the maximum power spectral densities (PSDs) of windowed triaxial accelerometer data with 100 data points each. At the time, with eight hidden neurons and two output neurons, the model training took several hours to complete.

Javed et al.(Javed, Gouriveau, Zemouri, & Zerhouni, 2011, 2012) performed a comparative analysis of features as input to an adaptive neuro-fuzzy inference system (ANFIS) for long term prediction and estimation of the RUL based on the CMAPPS data set (Saxena, Goebel, Simon, & Eklund, 2008). 40 data sets each with 8 out of 21 features were selected from the multivariate data set for model training. In the test phase, predictions were performed over different horizons with five data sets.

There is founded research and scholarly literature in the PHM-field for the estimation of the current health state or the estimation of the RUL of technical systems from condition monitoring data. However, there is little scholarly literature pertaining to the $h-$step, that is, one-step- or multi-step ahead prediction of an health indicator or degradation-related condition, while particularly considering condition monitoring data with very small sample size and little or no system-specific or domain information. Here, h denotes the prediction horizon. Nowadays, condition monitoring data or specifically run-to-failure data can be scarce for safety-critical systems or prototype systems, which poses challenges while developing diagnostics and prognostics modules.

This paper attempts to highlight some possibilities in tackling $h-$step ahead prediction tasks based on past and current recorded data. The prediction task is formulated as a time series forecasting problem which opens up more possibilities. Time series forecasting methods are a group of manifold methods that aim at predicting future times based on measurements detected in the past (Hyndman & Athanasopoulos, 2018). In this paper, classical time series forecasting approaches are employed for the reliable prognostics of technical systems, where black box machine learning techniques might not successfully be employed given insufficient amount of data and where first principles models are infeasible due to lack of domain specific data.

Considering the general PHM-steps (Elattar et al., 2016; Atamuradov et al., 2017), the next section begins by briefly describing the experiment involved in generating the case study data. Given that the raw data are inherently noisy, the data are thereafter preprocessed. In the following subsection, characteristic features are extracted in the time-, frequency-, and time-frequency domain from the preprocessed data. The extracted and selected features are then mapped to crack lengths by three carefully selected machine learning techniques, to conclude this subsection. The estimated crack lengths are propagated in the future via four forecasting techniques in the penultimate subsection. Finally, performance evaluation metrics are laid out and employed to evaluate the presented forecasting techniques.

2. METHODOLOGY

The methods described in this paper are exemplarily demonstrated on the data set provided by the PHM-society at the conference's 2019 data challenge (PHM Society, 2019).

2.1. Data Description

The data set consists of wave signals from lead zirconate titanate (PZT)-based piezeoelectric sensors mounted on riveted aluminium specimens under cyclic loading as exemplarily shown in Figure 2. Experiments were obtained with a hydraulic material testing machine working at 5 Hz at room temperature. Tension-tension fatigue test with constant amplitude as well as low-high block loading were studied as shown in Figure 1. In order to reduce the statistical uncertainty, measurements were acquired twice.

Edges of rivet holes are critical regions of high stress concentration that promote fatigue crack initiation and propagation which subsequently leads to failure (Harris, Piascik, & Newman Jr, 1999). For the non-destructive monitoring of the evolution of the edge cracks, a piezoelectric actuator-sensor pair was installed at a distance of 161 mm on opposite sides each of the the rivet holes along the lap joints as exemplarily illustrated in Figure 2. The actuator was excited with a 200 kHz Hanning-windowed sine wave signal which propagated through the rivet holes and was sensed by the sensor. The idea was that as crack develops along the wave propagation path, the received signal will reflect this change, thus acting as an health indicator for diagnostics and prognostics. Eight specimens, subsequently referred to as $T1$ through $T8$ were investigated. $T1 - T7$ were tested under constant amplitude, while $T8$ was tested under low-high fatigue loading. EOL of the specimen was declared when an undisclosed critical crack length was reached. During testing, crack length was also measured intermittently with an optical microscope. In the algorithm training phase, measured crack lengths were only available for specimens $T1$ through $T6$ as depicted in Figure 3. As can be seen in this figure, the cycle at crack initiation and the cycle at critical crack length are not necessarily the same for all data sets. Data sets were recorded sporadically, that is not periodically and continuously and only actuator-sensor pairs that registered a crack were made available. Actuator-sensor data sets, loading profile, as well as measured crack lengths were available up to EOL for 6 specimens, that is $T1$ through $T6$. Actuator-sensor data sets up to half of the EOL as well as loading profile were also made available for $T7$ and $T8$. The wave signal was 200 μs long and consists of 4000 data points.

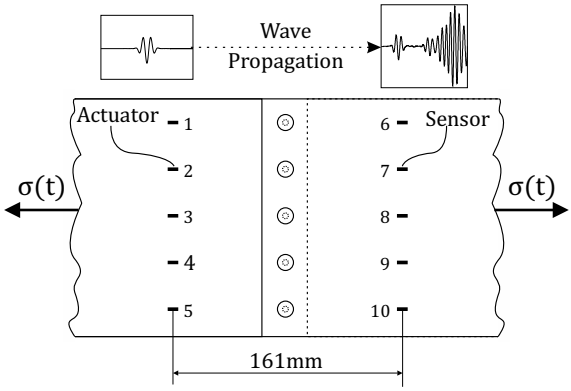


Figure 2. Section of the aluminum specimen with actuator-sensor placement

The task of the data challenge was two-fold. Firstly, **crack length estimation** for the specimens $T7$ and $T8$ given the aforementioned data sets. Secondly, **crack length prediction** for both specimens, where only number of cycles for the prediction horizon was given.

2.2. Crack Length Estimation

This section gives a breakdown of the process followed to obtain crack length estimates for the provided specimens and exemplarily for specimen $T4$.

2.2.1. Data Preprocessing

As with most sensor measurements, the provided wave signal measurements were inherently noisy and hence could not be used in the raw form to extract useful information pertaining to the condition of the system. Figure 4(a) depicts the noisy raw sensor signals exemplarily for specimen $T4$. The raw signals were filtered with the $morlet$ wavelet family of the scaled continuous wavelet transform (CWT), given that the

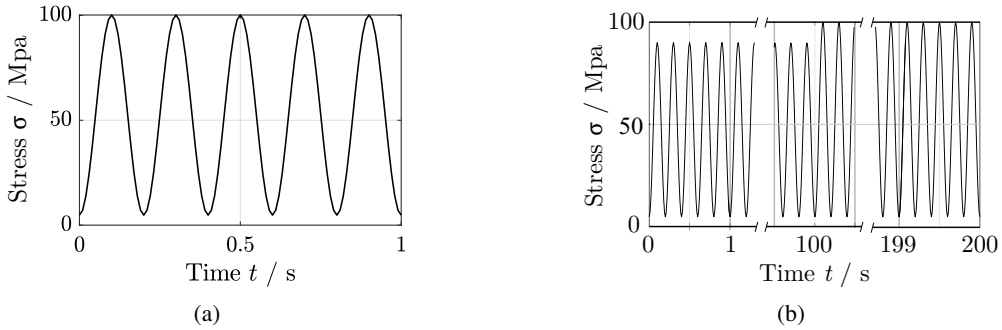


Figure 1: (a) constant amplitude loading and (b) low-high block loading

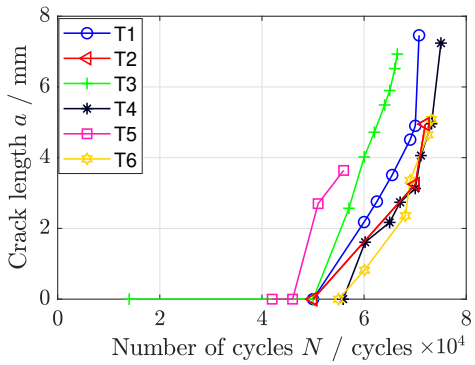


Figure 3: Actual measured crack lengths and respective cycles for specimens $T1 - T6$

shape of the mother wavelet is consistent with the shape of the excitation. The extracted scales correspond to the signal frequency range 160kHz - 235kHz. Apart from signal denoising, other preprocessing steps included outlier removal and mean normalization of the sensor signal.

2.2.2. Feature extraction and selection

The presence of a crack in a structure generally affects the vibration properties of the structure, that is modal and/or structural properties. The structural properties include the mass, stiffness and the damping. The modal properties include the natural frequencies and the mode shapes (Kindova-Petrova, 2014). To analyse these influences, a total of 24 characteristic features were extracted in the time-, frequency- and time-frequency domain from the preprocessed wave signals.

Time domain

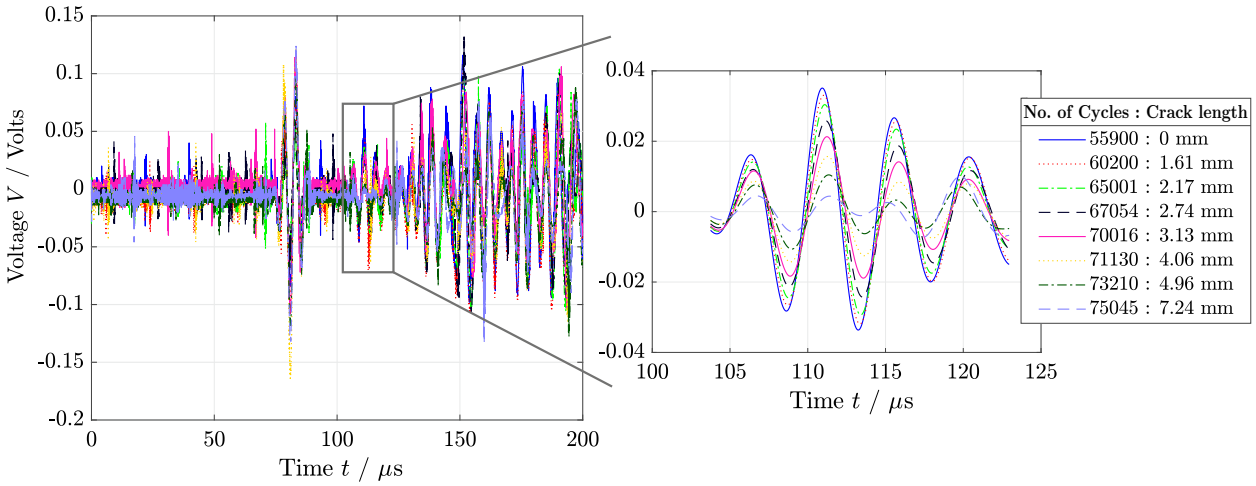
Statistical features were extracted in the time domain of the 200 μs sensor wave signal. A majority of the extracted features from the preprocessed 200 μs sensor wave signal did not show distinctive monotonic trends reflecting crack initiation and propagation. Considering the wave propagation velocity and subsequently the time of flight of the wave, only a section of the signal was thereafter considered. Concretely only about 20 μs time signal segment as exemplarily depicted in Figure 4(b) was then employed to extract the undermentioned features. Some of the extracted features are briefly described below.

Peak value: as evident from Figure 4(b), the $peak$ value decreases as the crack size increases. Two phenomena can explain the decrease in maximum amplitude of the signal, namely, the attenuation effect of crack development on the wave propagation and scattering and the stiffness reduction as a result of crack growth, which implies higher damping.

Peak-to-peak: also known as peak-to-valley value is defined as the difference between the maximum and minimum amplitude value. For a symmetric wave signal with a zero mean, it is twice the peak value. It was considered because the analysed wave was not symmetric for all provided specimens, particularly as crack developed.

The ***RMS*** has often been employed for diagnosis of various technical systems such as a gearbox (Večeř, Kreidl, & Šmíd, 2005). It is said to describe the energy content of a signal. Its value decreased as the crack size increases, also due to the effect of crack development on the wave propagation and scattering.

Relative time at peak value: to capture the phase shift, the relative time at peak value, that is the time difference between the actuator and sensor signal was evaluated. As can be seen in Figure 4(b), there is a slight shift of the maximum amplitude to the right as the crack size increases, which implies a



(a) (b)

Figure 4: Specimen $T4$ (a) raw wave signal (b) analysed preprocessed time segment

retardation of the wave propagation velocity through the cracked aluminium material and thus leading to an increase in the time of flight. However, this feature was not robust because this trend was not clear and consistent in all considered specimens.

Statistical moments: the variance, skewness and kurtosis are the second, third and fourth centralized moment of a signal respectively. They are measures to describe distributions. The skewness describes the shape, while the kurtosis describes the peakedness or flatness of a distribution. These measures and particularly the kurtosis has been found to be a useful condition indicator (Večeř et al., 2005).

Mean absolute deviation (MAD): as the name implies, it gives the absolute deviation around the mean of a signal. It was found to pronounce the similarity characteristics between specimens as highlighted in Table 1. The values within the table are normalized in the range $[-1, 1]$.

Frequency domain

In the frequency domain, the presence or appearance of high resonant frequencies as crack develops as well as possible change in amplitude and frequency shifts were analysed. The analysis of high resonant frequencies did not yield any discernible feature. This is possibly attributable to sensor characteristics, such as measurement frequency range. There was noticeable decrease in amplitude in certain frequencies. However, to reduce possible feature redundancy, this feature was not further considered given that it has been considered in the time domain.

Time-frequency domain

In the time-frequency domain, the $bior3.7$ mother wavelet of the wavelet packet decomposition (WPD) was employed to decompose the signal up to three levels. The wavelet energy was thereafter obtained from the wavelet coefficients.

2.2.3. Mapping features to crack length

The mapping of features to crack length is considered here as a supervised machine learning task. That is, machine learning algorithms are fed the aforementioned features as inputs and actual crack lengths as targets in the training phase to learn the underlying relationship between features and crack lengths. This learned model is employed in the test and validation phase to output crack lengths given features as inputs. The training, test and validation phase implies data partitioning. Thus, the extracted features and actual crack lengths of the specimens $T1, T3$ and $T4$ were utilised for algorithm training and testing. The extracted features and actual crack lengths of the specimens $T2, T5$ and $T6$ were used for algorithm validation, given that $T2$ and $T5$ only had three to four data points, with two non-zero crack length values as evident from Figure 3. The inputs were min-max normalized, such that the inputs fall within the range $[-1, 1]$. Three algorithms were employed for this task because they have successfully found application to other problems (Kimotho, 2017), each with its strengths. These algorithms are briefly described below:

Extreme Learning Machines (ELM)

The ELM is a learning algorithm for a special class of neural networks, that is the single hidden layer feed-forward neural network. Only the output weights are adjusted while the input weights and hidden nodes are randomly generated. Thus eliminating the need for exhaustive search of the optimal amount of neurons, which makes the ELM extremely fast. Given data samples, the number of hidden nodes $\widetilde{N}$ and activation function $g(x)$. The weight matrix $\boldsymbol{\beta}$ connecting the $\widetilde{N}$ hidden nodes to the output nodes, is calculated while minimizing the error: $\min_{\beta} \|\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{T}\|$, where $\boldsymbol{H}$ is the hidden layer output matrix and $\|\cdot\|$ is the Frobenius norm. The optimal solution is achieved by $\boldsymbol{\beta} = \boldsymbol{H}^{\dagger}\boldsymbol{T}$, where $\boldsymbol{H}^{\dagger}$ is the Moore-Penrose generalized inverse of $\boldsymbol{H}$ (Huang, Huang, Song, & You, 2015). The activation function $g(x)$ employed here is the radial basis function (RBF) because it was found to yield the best performance according to Kimotho (Kimotho, 2017). Other activation functions include sigmoid function and hard limit function (Huang et al., 2015).

The hyper-parameters initialised during the training of the algorithm were the free parameter of the RBF, and the regularization coefficient, which regularizes the output weights to improve generalization.

Support Vector Regression (SVR)

SVR is an adaptation of the Support Vector Machine (SVM), which was originally developed for classification problems, to solve regression problems (V. Vapnik, Golowich, & Smola, 1997; V. N. Vapnik, 1998). The underlying theory of SVR is finding the weights w and bias b that minimizes the so-called risk with an ε-insensitive loss function, which accepts deviations within an ε-bound and penalises error otherwise. SVMs adopts the structural risk minimization principle, that is the simultaneous minimization of the empirical risk as well the so-called Vapnik–Cervonenkis (VC)-dimension, leading to better generalization performance (L. Wang, 2005). Kernel functions are typically employed to map possibly non-linear inputs to higher dimensional feature space (V. N. Vapnik, 1998). The RBF kernel and LIBSVM MATLAB® toolbox as implemented by Chang and Lin (Chang & Lin, 2011) was adopted in this paper. Other kernel functions are a linear kernel, a polynomial kernel of degree d and a sigmoid kernel, etc. each with kernel-specific parameters.

The LIBSVM MATLAB® toolbox requires a set of hyper-

parameters, such as the type of SVM, kernel type, kernel parameters and cost C. There are two SVM types implemented for regression, the `epsilon-SVR` (ε-SVR) and the `nu-SVR` (ν-SVR). The ε-SVR implementation utilises an ε-parameter, which is the aforementioned error tolerance ε. On the contrary, in the ν-SVR implementation, the ν-parameter not only places a bound on the admissible errors and support vectors, but also finds optimal ε (Chang & Lin, 2011). Its value ranges from zero to one, that is $\nu \in (0, 1]$.

Random Forest (RF)

A RF as introduced by Leo Breiman (Breiman, 2001) is an ensemble of decision trees, each consisting of an independent and identically distributed (i.i.d.) random vector, for classification and regression. Each tree of the ensemble is constructed from a bootstrap sample (Efron & Tibshirani, 1991), that is a random sample drawn with replacement from the original data, and each non-leaf node of the tree is split by randomly selecting a subset of the features as best split. The trees are recursively grown till a predefined number of data points per leaf or a maximum number of leaf nodes is reached. The predictions of each tree are averaged to obtain a global prediction (Breiman, 2001; Cutler, Cutler, & Stevens, 2012).

The MATLAB® implementation of the random forest[1] by Jaiantilal (Jaiantilal, 2010) was adopted in this paper. The hyper-parameters includes but are not limited to the number of random samples m_{try} and the number of trees n_{tree}. Also adapted was the $nodesize$, which gives a lower bound on the amount of terminal nodes. During training, min-max normalization had a rather negative impact on the predictions, so the inputs were not normalized contrary to the previous techniques.

The hyper-parameters for all described techniques were tuned with the differential evolution and particle swarm optimization successively (Kimotho, 2017). After training the aforementioned algorithms, the validated models were then employed as an ensemble to estimate the crack lengths for specimens $T7$ and $T8$. The results of the estimation appended with the results of the next task can be found exemplarily in Figure 5.

2.3. Crack Length Prediction

After estimating the crack lengths for specimens $T7$ and $T8$, the next task is the crack length prediction for both specimens where only number of cycles for the prediction horizon was given. Several techniques ranging from naive to sophisticated methods are employed in comparison for this task. They are presented in the following subsections.

[1]Originally implemented by Leo Breiman and Adele Cutler in Fortran (Breiman & Cutler, 2004) and ported in R by Andy Liaw and Matthew Wiener (Liaw, Wiener, et al., 2002)

2.3.1. Forecasting by Analogy

Forecasting by analogy or forecasting through similarity measures implies that a new specimen, that is, $T7$ or $T8$ is similar to some extent or follows the same crack length growth trend as previous specimens, that is, $T1 - T6$. This forecasting approach finds application where there is little data sets and data points and where there is not a common model that covers the historical data sets. Several characteristics were compared such as the extracted features as well as the sampling interval, which suggests same crack propagation rate. For example, it was found that the specimens $T7$ share similar characteristics with $T4$ not only because their sampling intervals were similar but also because several features and most especially the value of MAD were similar as seen in Table 1. Figure 5 depicts the result of $T4$ to forecast the crack lengths for $T7$ for the prediction horizon under the assumption of the same rate of change. The result could be minimally improved by considering the mean of $T1$ and $T4$.

2.3.2. Forecasting by Analytical Function Fitting

Several analytical functions such as linear, polynomial, exponential, etc. were fitted to historical data to describe the crack growth trend. An analytical function that relates the cycles to crack length can be evaluated for arbitrary points. For the prediction horizon, the analytical function can then be extrapolated beyond available data points, that is cycles outside the fitted range. Given the variability in the presented historical data sets, a general path model that describes all the crack growth trends could not be adopted. However, with the already highlighted similarity between $T7$ and $T4$, the crack lengths for $T7$ could be predicted with the five parameter exponential model as in equation 1 fitted with $T4$.

$$f(cycle) = \alpha + \beta \cdot \mathrm{e}^{(\gamma \cdot cycle)} + \delta \cdot \mathrm{e}^{(\kappa \cdot cycle)} \qquad (1)$$

The parameters $\alpha, \beta, \gamma, \delta$ and κ were found through optimization by minimizing the sum of squared deviations. The result of the application of this forecasting method can be found in Figure 6.

2.3.3. Exponential Smoothing

The exponential smoothing as the name suggests is a smoothing function. It has mostly found application in PHM-related literature as a data smoothing or noise filtering technique. Beyond noise filtering, it is a well-established time series forecasting approach. The underlying principle of the exponential smoothing forecasting approach is that newer data points of a time series are more relevant than older points of the series, thus more weights are assigned to newer points (Hyndman & Athanasopoulos, 2018). There are three main types of exponential smoothing methods, namely single, double and triple exponential smoothing method, depending on the time series components been model-

Table 1. Mean absolute deviation (MAD) over all specimens

T1		T2		T3		T4		T5		T6		T7		T8	
Cycle	**MAD**	**Cycle**	**MAD**	**Cycle**	**MAD**	**Cycle**	**MAD**	**Cycle**	**MAD**	**Cycle**	**MAD**	**Cycle**	**MAD**	**Cycle**	**MAD**
50000	n/a	50000	0.00	14000	n/a	55900	0.88	42000	-0.60	55000	0.93	36001	0.89	40000	1.00
60000	0.30	70033	-0.58	50000	0.65	60200	0.77	46000	n/a	60078	0.71	40167	0.79	50000	0.93
62500	0.33	72000	-0.94	57038	0.61	65001	0.57	51000	-0.82	68091	0.59	44054	0.54	70000	0.68
65500	0.50			60035	0.57	67054	0.30	56000	-1.00	69018	0.30	47022	0.37	74883	0.69
69025	-0.04			62017	0.34	70016	0.02			72516	0.20	49026		76931	0.59
70026	-0.20			64019	0.07	71130	-0.33			73211	0.04	51030		89237	
70766	-0.61			65029	-0.09	73210	-0.61					53019		92315	
				66012	-0.32	75045	-0.75					55031		96475	
				66510	-0.44									98492	
														100774	

led. The single exponential smoothing method applies a single smoothing parameter. It is best employed when the data does not have a trend. The double exponential smoothing method, also known as Holt's linear method, is an extension of the foregoing method to account for multiplicative or additive trend, which can be damped. This method was further extended by Holt and Winter to account for seasonality, leading to the triple exponential smoothing method. The exponential smoothing is also coined ETS, from the initials of the three time series components that is Error or residual component, Trend and Seasonal component (Hyndman & Athanasopoulos, 2018).

As with most time series forecasting method, exponential smoothing is formulated for forecasting evenly spaced time series. The presented data did not possess a constant sampling interval as previously mentioned. Several techniques exist for dealing with unevenly spaced time series data. However, the transformation from unevenly- to evenly-spaced time series via interpolation was employed here. Several interpolation schemes such as cubic and linear scheme were evaluated. Linear interpolation was chosen because there was not a noticeable advantage of any other scheme over this scheme for the short time series with 4 or 5 data points for $T7$ and $T8$ respectively. The Holt Winters exponential smoothing method as implemented in the time series analysis Python package (`statsmodels.tsa`) was employed in this paper. The function input parameters were found through grid search as proposed by Brownlee (Brownlee, 2018). The hyper-parameters smoothing level α, smoothing slope β, and smoothing seasonal γ all in the interval $(0, 1]$ were found via the Limited-memory Broyden-Fletcher-Goldfarb-Shanno optimization algorithm (L-BFGS) in the scipy optimization Python package (`scipy.optimize.minimize`) while minimizing the mean squared logarithmic error. Figure 7 depicts the results obtained by applying this forecasting approach.

2.3.4. Long Short-Term Memory

Long short-term memory (LSTM) is a widely employed type of recurrent neural network (RNN) for processing sequential data and particularly for time series forecasting. It was developed to overcome the drawback of the traditional RNN, that is, its inability to handle long time dependencies. The network topology consists of an input layer, a hidden and an output layer. As opposed to the traditional neurons, the hidden layers contain so-called memory cells and corresponding input and output gates. Long- or short-term information are stored in memory cells. The input and output gates determine what information is stored or removed from the memory cells and controls the access to and from connected memory cells (Hochreiter & Schmidhuber, 1997).

According to Brownlee (Brownlee, 2018), there are four classes of LSTM models for time series forecasting, depending on the number of time dependent variables and the forecasting task. They are univariate, multivariate, multi-step and multivariate multi-step LSTM models. The task at hand is a multi-step ahead prediction task, thus the multi-step LSTM model and specifically the Stacked LSTM was used here. The linearly interpolated data was split in train and test data sets and formatted accordingly as input and output for the network. The employed Python's Keras sequential LSTM model was optimized with the adaptive moment estimation (Adam) optimizer while minimizing the mean squared error. The results obtained by applying this forecasting approach can be found in Figure 8.

2.4. Performance Evaluation

There are several measures to evaluate the performance of predictive methods. For the 2019 data challenge (PHM Society, 2019), the cumulative penalty score was formulated to evaluate the performance of the presented methods. It is the sum of the products of the time, asym-

metric and monotonicity penalty functions, which are briefly described below.

Time penalty function: because of the dire consequences that a critical crack length might have, prediction error incurred at the end of life are heavily penalized than error incurred at the initial stages of crack growth. It is mathematically expressed as:

$$T(i) = \alpha + \beta \cdot \tilde{a}_i, \tag{2}$$

where $i = 1, \cdots, n$ and n is the cardinality of the set of all predictions. $\alpha = 2$ and $\beta = 10$ are arbitrary constants and $\tilde{a}$ is the normalized crack length and it is given as:

$$\tilde{a} = \frac{\text{actual crack length, } a}{\text{critical crack length, } a_{crit.}}. \tag{3}$$

Although the predictions are not explicitly considered in the formulation, the time penalty function produces multiplicative weights that have impact on the cumulative penalty score.

Asymmetric penalty function: crack length underestimation implies a conservative estimate. For safety critical systems, such conservative estimates could lead to catastrophic events. Thus, this function penalizes crack length underestimation more than overestimation. It is formulated as:

$$A(i) = \begin{cases} exp\left\{\dfrac{|\tilde{a}_i^p - \tilde{a}_i|}{\gamma}\right\} - 1, & \text{if } (\tilde{a}_i^p - \tilde{a}_i) \geq 0 \\ exp\left\{\dfrac{|\tilde{a}_i^p - \tilde{a}_i|}{\delta}\right\} - 1, & \text{otherwise} \end{cases} \tag{4}$$

where $\tilde{a}_i$ and $\tilde{a}_i^p$ are actual and predicted crack lengths normalized as in equation 3, $\gamma > \delta > 0$. $\gamma = 0.5$ and $\delta = 0.2$ are arbitrary constants. $i = 1, \cdots, n$ and n is the cardinality of the set of all predictions.

Monotonicity penalty function: given that degradation typically follows a monotonic trend, this function penalizes cases which deviates from this trend. It is given as:

$$M(i) = \begin{cases} 1 + \omega \cdot (|\tilde{a}_i^p - \tilde{a}_{i-1}^p|), & \text{if } (\tilde{a}_i^p - \tilde{a}_{i-1}^p) < 0 \\ 1, & \text{otherwise} \end{cases} \tag{5}$$

where $\tilde{a}^p$ is the predicted crack length normalized as in equation 3, $\omega = 10$ is an arbitrary constant. $i = 1, \cdots, n$ and n is the cardinality of the set of all predictions.

The cumulative penalty score CPS with the range $[0, \infty)$ then becomes:

$$CPS = \sum_{i=1}^{n} T(i) \cdot A(i) \cdot M(i). \tag{6}$$

Given that the components of the cumulative penalty score all have arbitrary constants that impose the weights on the score, it is difficult to interpret the results and also to apply this metric to other problems. Thus, among other known measures, the mean absolute error (MAE) metric was also employed to evaluate the performance of the methods. MAE is a metric employed to evaluate the average absolute deviation of the predicted from the actual crack length over the prediction horizon. It is calculated as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |a_i^p - a_i|, \tag{7}$$

where a_i and a_i^p are actual and predicted crack lengths respectively. $i = 1, \cdots, n$ and n is the cardinality of the set of all predictions. Table 2 shows the results of the performance evaluation based on the cumulative penalty score and the MAE for the presented forecasting methods.

2.5. Results and Discussion

As can be deducted from Table 2, the forecasting method via analytical function fitting outperforms the other methods for $T7$. However, this method is deterministic and not easily transferable for new unseen cases as in the case for $T8$. It should be adopted if viable historical data sets, consisting possible degradation modes and trends exist. Forecasting by analogy and specifically the rate of change method is the first-choice forecasting method when presented with very little data sets. It is computationally inexpensive, and it produces reasonable results and for $T8$ even the best cumulative penalty score. Exponential smoothing method builds on past, possibly evenly spaced data points and produces conservative predictions. The training of the LSTM method was time consuming and as a machine learning technique, it could be computationally expensive. However, it outperforms the ot-

Table 2. Performance of the proposed forecasting approaches based on CPS and MAE for specimens $T7$ and $T8$.

		$T7$	$T8$
Forecast by Analogy / rate of change	CPS	9.58	25.76
	MAE	0.48	0.76
Forecast by Analytical Function Fitting	CPS	5.50	147.55
	MAE	0.28	0.94
Exponential Smoothing	CPS	16.87	35.93
	MAE	0.34	0.62
LSTM	CPS	9.73	27.58
	MAE	0.46	0.58

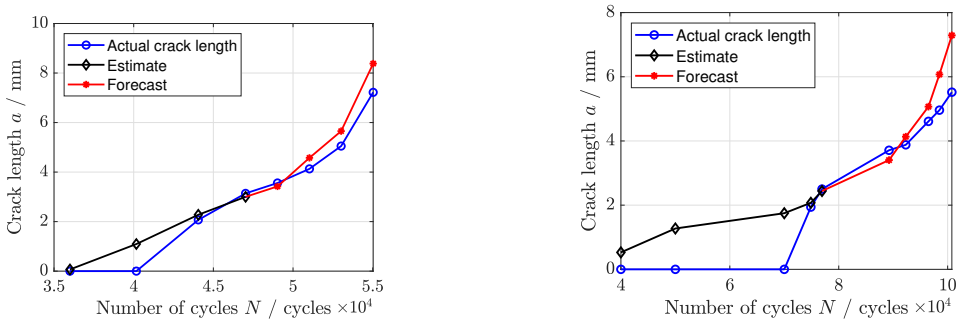


Figure 5: Results obtained via forecasting by analogy for (a) $T7$ and (b) $T8$

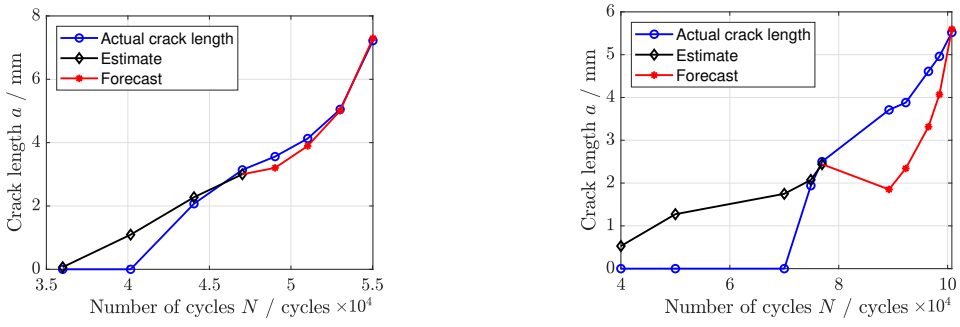


Figure 6: Results obtained via analytical function fitting for (a) $T7$ and (b) $T8$

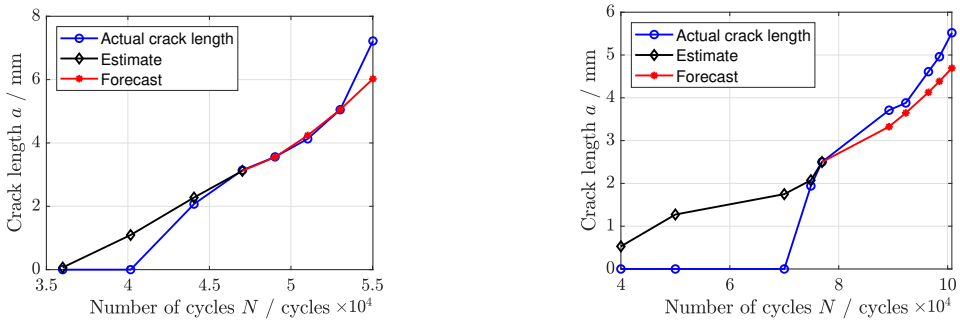


Figure 7: Results obtained with Holt Winters exponential smoothing method for (a) $T7$ and (b) $T8$

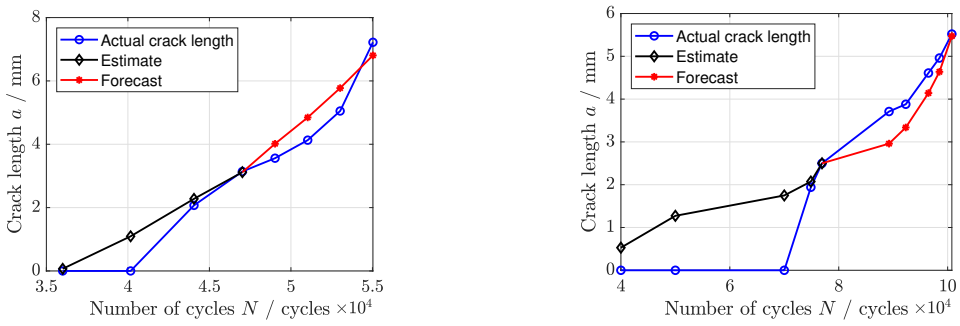


Figure 8: Results obtained with long short-term memory LSTM for (a) $T7$ and (b) $T8$

her methods for $T8$ while considering the MAE. As can clearly be seen, the presented methods underperform for $T8$ in comparison to $T7$. This is attributable to the loading condition variation, that is, the low-high amplitude loading situation as opposed to the constant amplitude loading for the presented test data sets $T1$ through $T6$ and also specifically for $T7$. In summary, the choice of the method depends on the goal to be achieved, the amount and quality of data and the amount of variability.

3. CONCLUSION

Several time series forecasting methods were presented in this paper. As a case study, these methods were successfully employed to predict the crack lengths of riveted aluminium specimens given limited historical data and lack of domain specific data. The case study data set was provided by the PHM-society at the conference's 2019 data challenge (PHM Society, 2019), where the tasks were to estimate and predict the crack lengths for the specimens $T7$ and $T8$ based on specimens $T1$ through $T6$.

From the presented results in Section 2.5 and specifically Table 2, it can be concluded that forecasting by analogy and via rate of change method outperforms the other presented methods, while considering the combined CPS for both specimens $T7$ and $T8$. Although only evaluated on the case study data set, the presented methods are easily transferable to other prediction problems.

As a future outlook, the presented exponential model can be incorporated in Bayesian-based algorithms such as widely employed Kalman or particle filter (Bender & Sextro, 2018) to account for uncertainties and model variations, provided more knowledge about the data generating processes is available. Further case studies also based on real-world data sets are required and planned to further evaluate the application of the presented methods for reliable prognostics of technical systems.

REFERENCES

Atamuradov, V., Medjaher, K., Dersin, P., Lamoureux, B., & Zerhouni, N. (2017). Prognostics and health management for maintenance practitioners-review, implementation and tools evaluation. *International Journal of Prognostics and Health Management*, *8*(060), 1–31.

Bender, A., & Sextro, W. (2018). A particle filtering approach for temperature based prognostics. *Safety and Reliability–Safe Societies in a Changing World*, 1025–1033.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Breiman, L., & Cutler, A. (2004). *Random forests - fortran implementation.* Retrieved from `https://www.stat.berkeley.edu/~breiman/RandomForests/`

Brownlee, J. (2018). *Deep learning for time series forecasting: Predict the future with mlps, cnns and lstms in python*. Machine Learning Mastery.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*, 27:1–27:27. (Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`)

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In *Ensemble machine learning* (pp. 157–175). Springer.

Efron, B., & Tibshirani, R. (1991). Statistical data analysis in the computer age. *Science*, 390–395.

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

Engel, S. J., Gilmartin, B. J., Bongort, K., & Hess, A. (2000, March). Prognostics, the real issues involved with predicting life remaining. In *2000 ieee aerospace conference. proceedings (cat. no.00th8484)* (Vol. 6, p. 457-469 vol.6). doi: 10.1109/AERO.2000.877920

Harris, C. E., Piascik, R. S., & Newman Jr, J. C. (1999). A practical engineering approach to predicting fatigue crack growth in riveted lap joints.

Hendron, J. (1963). Nondestructive testing of high-strength steel rocket motor cases. In *Symposium on recent developments in nondestructive testing of missiles and rockets.*

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Huang, G., Huang, G.-B., Song, S., & You, K. (2015). Trends in extreme learning machines: A review. *Neural Networks*, *61*, 32–48.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts: Melbourne, Australia. Retrieved from `OTexts.com/fpp2` (Retrieved on: 05.04.2020)

Jaiantilal, A. (2010). *Random forest (regression, classification and clustering) implementation for matlab (and standalone).* Retrieved from `https://code.google.com/archive/p/randomforest-matlab/`

Javed, K., Gouriveau, R., Zemouri, R., & Zerhouni, N. (2011). Improving data-driven prognostics by assessing predictability of features..

Javed, K., Gouriveau, R., Zemouri, R., & Zerhouni, N. (2012). Features selection procedure for prognostics: An approach based on predictability. *IFAC Proceedings Volumes*, *45*(20), 25 - 30. (8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes) doi: https://doi.org/10.3182/20120829-3-MX-2028.00165

Kimotho, J. K. (2017). *Development and performance evaluation of prognostic approaches for technical systems* (Vol. 4). Shaker Verlag.

Kindova-Petrova, D. (2014). Vibration-based methods for detecting a crack in a simply supported beam. *Journal of Theoretical and Applied Mechanics*, *44*(4), 69–82.

Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. *R news*, *2*(3), 18–22.

Lu, C. J., & Meeker, W. O. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, *35*(2), 161-174. doi: 10.1080/00401706.1993.10485038

Paris, P., & Erdogan, F. (1963, 12). A Critical Analysis of Crack Propagation Laws. *Journal of Basic Engineer-*

ing, *85*(4), 528-533. doi: 10.1115/1.3656900

PHM Society. (2019). *Phm data challenge.* 2019 PHM Conference Data Challenge. Retrieved from https://www.phmdata.org/2019datachallenge/ (Retrieved on: 31.03.2020)

Ray, A., & Tangirala, S. (1996). Stochastic modeling of fatigue crack dynamics for on-line failure prognostics. *IEEE Transactions on Control Systems Technology*, *4*(4), 443–451.

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).

Vapnik, V., Golowich, S. E., & Smola, A. J. (1997). Support vector method for function approximation, regression estimation and signal processing. In *Advances in neural information processing systems* (pp. 281–287).

Vapnik, V. N. (1998). *Statistical learning theory* (Vol. 1). Wiley New York.

Večeř, P., Kreidl, M., & Šmíd, R. (2005). Condition indicators for gearbox condition monitoring systems. *Acta Polytechnica*, *45*(6).

Virkler, D. A., Hillberry, B. M., & Goel, P. K. (1979, 04). The Statistical Nature of Fatigue Crack Propagation. *Journal of Engineering Materials and Technology*, *101*(2), 148-153. doi: 10.1115/1.3443666

Wang, L. (2005). *Support vector machines: Theory and applications* (Vol. 177). Heidelberg, Germany: Springer Berlin.

Wang, P., & Vachtsevanos, G. (2001). Fault prognostics using dynamic wavelet neural networks. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *15*(4), 349–365. doi: 10.1017/S0890060401154089

BIOGRAPHIES

Osarenren Kennedy Aimiyekagbon holds a bachelor's degree in engineering informatics with a major in mechanical engineering and a master's degree in mechanical engineering with a major in computer science both from Paderborn University. Since 2018, he is with the Chair of Dynamics and Mechatronics, Paderborn University. His research focuses on prognostic and health management of mechatronic systems.

Amelie Bender studied mechanical engineering at RWTH Aachen University and Newcastle University, Australia. Since 2015, she is with the Chair of Dynamics and Mechatronics, Paderborn University. Her research focuses on prognostic methods, especially PoF-based methods.

Walter Sextro studied mechanical engineering at the Leibniz University of Hanover and at the Imperial College in London. After his studies, he was development engineer at Baker Hughes Inteq in Celle, Germany and Houston, Texas. He was awarded the academic degree Dr.-Ing. as a research assistant at the University of Hanover in 1997. Afterward he habilitated in the domain of mechanics under the topic Dynamical contact problems with friction: Models, Methods, Experiments and Applications. From 2004 till 2009 he was professor for mechanical engineering at the Technical University of Graz, Austria. Since March 2009 he is professor for mechanical engineering and head of the Chair of Dynamics and Mechatronics, Paderborn University.