Unscented Kalman Filtering for Prognostics Under Varying Operational and Environmental Conditions

Luc S. Keizers[1,2], Richard Loendersloot[1], and Tiedo Tinga[1,2]

[1] *University of Twente, Enschede, 7522NB, the Netherlands*
l.s.keizers@utwente.nl,
r.loendersloot@utwente.nl,
t.tinga@utwente.nl

[2] *Netherlands Defence Academy, Den Helder, 1781AC, the Netherlands*

ABSTRACT

Prognostics gained a lot of research attention over the last decade, not the least due to the rise of data-driven prediction models. Also hybrid approaches are being developed that combine physics-based and data-driven models for better performance. However, limited attention is given to prognostics for varying operational and environmental conditions. In fact, varying operational and environmental conditions can significantly influence the remaining useful life of assets. A powerful hybrid tool for prognostics is Bayesian filtering, where a physical degradation model is updated based on real-time data. Although these types of filters are widely studied for prognostics, application for assets in varying conditions is rarely considered in literature. In this paper, it is proposed to apply an unscented Kalman filter for prognostics under varying operational conditions. Four scenarios are described in which a distinction is made between the level in which real-time and future loads are known and between short-term and long-term prognostics. The method is demonstrated on an artificial crack growth case study with frequently changing stress ranges in two different stress profiles. After this specific case, the generic application of the method is discussed. A positioning diagram is presented, indicating in which situations the proposed filter is useful and feasible. It is demonstrated that incorporation of physical knowledge can lead to highly accurate prognostics due to a degradation model in which uncertainty in model parameters is reduced. It is also demonstrated that in case of limited physical knowledge, data can compensate for missing physics to yield reasonable predictions.

1. INTRODUCTION

Maintenance of components or structures is required to keep them in a good health state. Maintenance strategies can be either corrective or preventive. In a corrective maintenance strategy, maintenance is performed after failure occurred. In a preventive maintenance strategy, maintenance is performed before failure such that unexpected downtime is prevented.

Two types of preventive maintenance are 1) *planned maintenance*, based on a predefined interval (Taheri et al., 2019) and 2) *condition-based maintenance* (CBM), based on the actual condition of a component or structure. Tiddens (2018) defined that a CBM strategy can be either measured (i.e. based on a real-time measured state) or calculated (i.e. based on a calculated state). A CBM strategy can be referred to as a predictive maintenance strategy when not only the current state is used, but also a prediction of the future state is taken into account (Tiddens, 2018). In the current paper, predictive maintenance refers to a maintenance strategy in which a prediction of future degradation is incorporated (calculated CBM). This prediction step is specified as the prognostic step.

The interest in predictive maintenance is expanding due to the trend towards Industry 4.0 (Montero Jimenez et al., 2020). This maintenance strategy leads to the highest availability of assets at the lowest costs because of the following reasons: 1) expensive, unexpected downtime can be reduced, 2) assets can be used until the end of their lifetime, preventing replacement of healthy components, 3) maintenance can be scheduled when it is most convenient and 4) it can lead to efficient warehousing. The prognostic step is crucial to achieve all benefits that predictive maintenance offers.

However, prognostic algorithms are adopted in industry to a limited extent yet. Although a significant amount of research papers has been published on development of algorithms, practical issues need to be solved (Tiddens et al., 2020). For

example, data-driven models gained a lot of research interest, but these models require representative training data that are often not available. This is especially problematic in cases where historical loading profiles are different from future loading profiles due to varying operational and environmental conditions.

Thus, varying operational and environmental conditions complicate the application of prognostic algorithms. However, also opportunities arise if varying conditions can be taken into account: if it is known how certain conditions affect degradation rates, operations can be adapted to influence the remaining useful life (RUL) of the asset. This can be used to delay maintenance activities to a more convenient moment by adapting the operational profile.

The aim of this paper is to develop a prognostic algorithm that can predict the RUL of a component under varying operational and environmental conditions. To solve practical issues that appear in purely data-driven or physics-based methods, combinations of data-driven and physics-based models are expected to be powerful. In this paper, it is proposed to utilize an unscented Kalman filter (UKF). It is applied to an artificial crack growth problem with varying loading profiles, where the effectiveness of the filter for short-term and long-term prognostics is demonstrated for different filter settings. Also, the benefit of including physical knowledge, in this case knowledge on the loading profiles, is shown.

The remainder of this paper is organized as follows. In section 2 a literature review describes prognostic methodologies from which Bayesian filtering is selected as a powerful method for further development. A specific Bayesian filter, the UKF, is explained in section 3. Section 4 introduces the artificial crack growth case with four different prognostic scenarios. The results of these scenarios are shown in section 5. In section 6 a positioning diagram is introduced to clarify when the approach can be used. A discussion is given in section 7 and the paper is concluded in section 8.

2. PROGNOSTIC APPROACHES

Three types of prognostic approaches exist: physics-based, data-driven and hybrid (Guo et al., 2020). This section first describes characteristics of physics-based and data-driven models. Thereafter, hybrid approaches are introduced. In addition, particular methods are described that consider varying operational and environmental conditions. The section ends with a discussion on the hybrid prognostic models to select the methodology adopted in this paper.

2.1. Physics-Based and Data-Driven Prognostics

On the one hand, physics-based models predict degradation using mathematical expressions of the physics-of-failure (PoF). On the other hand, data-driven models predict the RUL from measured parameters with statistical methods or machine learning algorithms (Elattar et al., 2016).

The deterministic nature of physics-based models makes them suitable for accurate long-term predictions (Cubillo et al., 2016). However, this requires a model that accurately describes the degradation process. Development of such a model requires expert knowledge and uncertainties arise in the input and model parameters.

In contrast, data-driven models are easier to implement: no physical model is required, but data unravels hidden patterns (Elattar et al., 2016). However, this requires large quantities of failure data that are not always available because failures are generally prevented (Pillai et al., 2016). Furthermore, the data sets need to include all possible degradations for all conditions (i.e. operational and environmental conditions) of the system to be accurate (Elattar et al., 2016). Consequently, implementing data-driven algorithms is not always feasible.

As both methods have their merits and limitations, a strong belief is that combining physics-based models and data-driven models can leverage the benefits of both. These combined methods are discussed in the following subsection.

2.2. Hybrid Prognostics

The aim of hybrid prognostics is to combine multiple prognostic methods for better performance. In literature, *hybrid prognostics* can relate to combinations of multiple data-driven models as well as to a combination of physics-based and data-driven models (Liao & Köttig, 2014). Due to the belief that especially physics-based and data-driven models can solve their mutual limitations, the focus in this paper is on models that combine physics with data.

In other disciplines, e.g. in smart manufacturing or chemical industries, these types of models are referred to as gray-box models (Yang et al., 2017; Zendehboudi et al., 2018), combining a white-box (physics-based) with a black-box (data-driven). These models can be configured in three ways: parallel, physics-to-data and data-to-physics, as visualized in Figure 1.

Parallel hybrid prognostics In a parallel configuration (Figure 1a), a data-driven model can be used to learn and correct the errors of a physical degradation model. Dourado & Viana (2020) include an artificial neural network (ANN) to reduce the error of a fatigue model of an air plane. This error reduction is accomplished by learning the ANN the effect of a corrosive environment on crack growth. Yucesan & Viana (2020) use a neural network (NN) to account for the effects of contamination of grease lubrication to bearing fatigue in a wind turbine. Another application of the parallel configuration is proposed by Goebel et al. (2006), applying

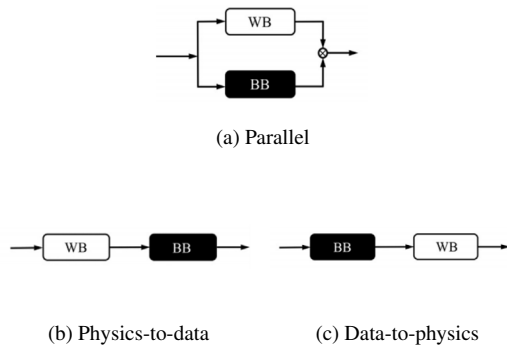(a) Parallel

(b) Physics-to-data   (c) Data-to-physics

Figure 1. Configurations of hybrid (gray-box) models (Zendehboudi et al., 2018) by combining white box (WB, physics) and black-box (BB, data) models

Dempster-Shafer regression to assign weights to estimated RULs of an empirical model and a physical model to obtain higher accuracy.

Physics-to-data hybrid prognostics An application of a physics-to-data configuration (Figure 1b) is the generation of artificial failure data to train a data-driven algorithm. For example, C-MAPPS is a software that creates run-to-failure data sets of turbofan engines (Frederick et al., 2007). Another application of a physics-to-data configuration is utilizing physical models to pre-process data before entering into a machine learning algorithm, which can make the algorithm less complex and more efficient (Rai & Sahu, 2020). For example, Chao et al. (2020) use physics-based performance models to improve a NN for prognostics of a turbofan engine, which is validated on the C-MAPSS dataset.

Data-to-Physics hybrid prognostics The majority of the developed hybrid methods are in the data-to-physics configuration (Figure 1c). For example, Kumar et al. (2008) developed a framework where data-driven methods detect anomalies in data and correlate them to the corresponding degradation mechanism. From a PoF database, the appropriate degradation model can then be selected to perform prognostics. Another data-to-physics approach is to update physical models with data to reduce the uncertainty of model parameters using Bayesian inference. For example, Zhao et al. (2013) update crack growth parameters of the Paris-Erdogan equation to predict crack propagation in a gear. Model parameters are updated from crack length measurements obtained by inspections.

Bayesian filters that apply Bayesian statistics in a recursive algorithm are widely studied. A distinction can be made between Kalman filters used for Gaussian types of noise and particle filters (PF) for arbitrary noise. This is explained in more detail in subsection 3.1. The PF in particular gained

a lot of research interest due to its broader applicability. A comprehensive review on the use of PF for prognostics is given by Jouin et al. (2016). A few examples of recursive Bayesian approaches in general are given below.

Recursive Bayesian filters require two models: 1) a state transition model that describes the degradation process by estimating a degradation parameter and 2) a measurement model that relates measurements to this estimated parameter. Because it is not always possible to estimate the required parameter directly or analytically, Baraldi et al. (2013) propose to use an ANN as measurement model in a PF. They applied it to predict the crack depth of a component.

A limitation of Bayesian filters is that they track the current state, but actually are not prognostic algorithms. This is due to the fact that no measurements are available in the future. Liao & Köttig (2016) propose a PF in which artificial measurements are created for the prognostic stage, using data-driven algorithms. The PF is used to predict the RUL of a battery. Another method to predict future degradation is to include parameter estimation in the recursive algorithm to be able to accurately extrapolate the latest stage into the future. This is similar to the Bayesian inference method as proposed by Zhao et al. (2013). As an example, Y. Wang et al. (2019) update parameters of the Paris-Erdogan equation in an extended Kalman filter (EKF) algorithm. These updated model parameters are used in a Monte-Carlo simulation to predict the RUL of fuselage panels of an aircraft.

2.3. Methods for Varying Operational and Environmental Conditions

Although a variety of prognostic approaches has been developed, the effect of varying operational and environmental conditions is not extensively researched yet. Frequently it is assumed that historical usage is representative for future usage or that degradation rates do not change. These can be valid assumptions for relatively stationary assets, but not for many moving assets.

Tinga et al. (2021) emphasize the effect of varying conditions on degradation and developed a new maintenance policy called *Functional Usage Profiles Based Maintenance*, completely based on operating profiles of assets. As an example, it is shown that degradation rates of wear pads of a military combat vehicle are highly dependent on the road and mission type of the asset. The approach shows promising results, but has no automated algorithm and the detailed assessment of expected future degradation is not included.

Other methods elaborate on transition probabilities between future operating profiles. Rezamand et al. (2021) propose an algorithm for wind turbine bearing prognostics that include multiple operating conditions. The method uses Kernel fuzzy C-means-based clustering on historical data to define a set

of operating conditions and their corresponding degradation rates. A Viterbi algorithm is utilized to predict future operating state transitions. Jain & Lad (2020) developed a prognostic approach for machine tool degradation in varying operating conditions in a deterministic scenario and in a probabilistic scenario. Uncertain future conditions are modelled by Markov chains. Li et al. (2019) propose a PF for varying conditions in which the effect of varying degradation rates and jumps in measured signals are considered. A set of degradation rates is defined for the state transition model and signal jumps are handled by the measurement model. The methods discussed in this paragraph are helpful if 1) future degradation rates are equal to historical degradation rates and 2) the sequence of occurrence of historical operating conditions is representative for the future.

Also methods are proposed that consider Bayesian filters with parameter updates and use them for assets in varying conditions. Zhao et al. (2015) perform Bayesian inference on parameters of the Paris-Erdogan equation for crack growth under varying loads, making use of a finite element model to predict the stresses. They stress the benefit of understanding the underlying degradation model. However, the method is limited to fully known future operating conditions. Y. Wang et al. (2021) utilize a PF to predict magnetic head wear of a hard disk drive including parameter updates in varying wear and stress conditions. Promising results are shown, but limited attention is given to the RUL predictions with uncertain future operating profiles. J. Wang et al. (2020) propose a PF for prognostics of wind turbine bearings when limited samples of historical data are available. They update model parameters to perform accurate prognostics. In this case, varying conditions are considered as stochastic process noise, but the relation between usage and degradation is neglected.

To conclude, some research is done in the field of prognostics and health management under varying conditions. However, further research is required when degradation rates and future operating conditions are uncertain.

2.4. Discussion on Prognostic Methods

The variety of developed hybrid algorithms described in this section show that combining physics-based and data-driven approaches can lead to successful solutions to predict the RUL. However, applicability depends on two aspects: 1) availability of representative data and 2) availability of a degradation model. A parallel approach works if sufficient data are available to train the data-driven model and a suitable physics-based model is available. For a physics-to-data model it is required to have a physical model that can properly describe all operating conditions and degradation signals of the component. Data-to-physics models require data that match the available degradation model.

As was mentioned, more research is needed that considers varying operating conditions. Because data-driven techniques are only valid if historical usage is representative for the future, involvement of physical models is expected to be helpful. However, a physical model can be hard to develop and may contain uncertainties.

From the discussed methods, Bayesian filtering is expected to be the most powerful tool to handle varying conditions and uncertainties. This is because of the following two reasons: first of all, performance of these methods does not depend on the quality and availability of large historical datasets. Instead of recognizing patterns in historical data and finding these patterns in newly observed data, as done in many data-driven methods, Bayesian filtering algorithms solely use newly observed data for the system under consideration to track its degradation state and extrapolate this to the future. Because the algorithm is independent of historical data of other systems, the newly observed degradation trend does not need to match historical degradation trends to yield accurate results. The second reason to select a Bayesian filtering approach is that data can compensate for missing information or can reduce uncertainties of the physical model. Therefore, relatively simple physical models can be used, even if model parameters are not exactly known.

In subsection 2.3 methods were discussed that consider varying operational conditions. It was concluded that these methods can work in specific scenarios. For example when transition probabilities and degradation rates can be fully defined or varying conditions can be handled as stochastic noise. The core contribution of this paper is to show how parameter updating of physical degradation models can be used to handle varying conditions in two situations: 1) to reduce uncertainties in predictions if loads are (partly) known and 2) to compensate for missing information if loads are unknown, in both cases for long-term as well as short-term prognostics.

In the next section Bayesian filtering is explained in more detail. The most studied Bayesian filter, the Kalman filter (Thrun et al., 2005), is described. Thereafter a modification of this filter, the unscented Kalman filter (UKF), is introduced. In contrast to the standard Kalman filter, the UKF can handle non-linear models. This filter will be utilized in this work for prognostics under varying operating conditions.

3. KALMAN FILTERING FOR PROGNOSTICS

This section first introduces the concept of Bayesian filtering techniques. Then the standard Kalman filter algorithm is described. Thereafter the UKF is illustrated which can, in contrast to the standard Kalman filter, handle non-linear models. Finally, the method to update model parameters is explained.

3.1. Introduction to Bayesian Filtering

In section 2 Bayesian filters were discussed. These filters can be explained from Bayes' rule. The core principle of Bayes' rule is to calculate a conditional probability of an event, given relevant information to this event. In the context of prognostics, the event can be defined as the degradation state. The relevant information is given by (noisy) measurements of the actual degradation state. In the form of probability density functions Bayes' rule can be written as follows (Arulampalam et al., 2002):

$$p(x_k|z_{1:k}) = \frac{1}{\int p(z_k|x_k)p(x_k|z_{1:k-1})dx}p(z_k|x_k)p(x_k|z_{1:k-1}) \tag{1}$$

with k the time steps, $p(x_k|z_{1:k})$ the posterior probability density of state x given measurements $z_{1:k}$, $p(z_k|x_k)$ the likelihood probability density, $p(x_k|z_{1:k-1})$ the prior probability density of state x and $\int p(z_k|x_k)p(x_k|z_{1:k-1})dx$ a normalization factor. The obtained posterior probability density of the state can then be extrapolated into the future up to a failure threshold, resulting in a probability density of the RUL.

Without assumptions on state transition models and process noises it is not always possible to solve equation 1 analytically (Elfring et al., 2021). Particle filters overcome this problem by representing the posterior distribution discretely based on Monte Carlo simulations. An explanation of this filtering approach is given by Gordon & Salmondand (1993). A disadvantage of the PF is that it is computationally expensive due to the large amount of calculations involved.

An analytical solution is found when assuming Gaussian noises and linear state transition models. These assumptions may lead to a decreased accuracy, but yield a faster and simpler algorithm. Equation 1 leads to the Kalman filter in this case. The normalization factor is still hard to solve, but because it is a constant it is known that the posterior is proportional to the numerator of equation 1. Although the original derivation of the Kalman filter (Kalman, 1960) does not start from Bayes' rule, the Kalman filter can be interpreted from the Bayesian point of view. An extensive derivation of the Kalman filter from Bayes' rule is given by Chen (2003). A brief description of the implementation of the filter is given in the following subsection.

3.2. Standard Kalman Filter

As already mentioned in section 2.2, the Kalman filter requires a state transition model and a measurement model. The state transition model describes how the system state vector x changes over time and the measurement model relates the estimated state variables to the obtained measurements z. For a standard Kalman filter for linear systems, these models are formulated as follows:

$$x_k = Fx_{k-1} + Gu_k + w_k \tag{2}$$
$$z_k = Hx_k + v_k \tag{3}$$

where F is a state transition matrix, G is a control input matrix, w is a vector with Gaussian process noises, H is the measurement matrix that predicts the measurement from the predicted state and v is a Gaussian measurement noise vector. The indices k represent the time step. In the remainder of this paper, G is neglected because the filter is used for a degradation model, which has no control inputs.

The algorithm consists of two steps: first, a prediction step is performed where a prior belief of the new system state and its covariance are calculated. Secondly, the update step is performed where the posterior estimate is calculated by incorporating the measurement. The calculations in these steps are given below (Saho & Masugi, 2015). To distinguish between prior (before updating) and posterior (after updating) estimates that appear in a single time step k, priors are indicated with a $(-)$ and posteriors are indicated with a $(+)$.

The prediction step is as follows:

$$\tilde{x}_k^- = F\tilde{x}_{k-1}^+ \tag{4}$$
$$P_k^- = FP_{k-1}^+F^T + Q \tag{5}$$

where F is the state transition matrix and Q the process noise matrix, representing F and w of equation 2 respectively. P is the error covariance matrix, representing the uncertainty of the estimated system state. T indicates a transposed matrix. It should be noted that in the first time step, no posterior $\tilde{x}_{k-1}^+$ and P_{k-1}^+ are available yet. In this first time step, the posteriors are replaced by specified initial beliefs of the system state.

After the prediction step, the update step is performed. This concerns two actions: First, the Kalman gain K is calculated. The Kalman gain weights the calculated state and the measurement to minimize the error in the posterior estimate. Then, the Kalman gain is used to calculate the posterior state estimate and the covariance matrix. The calculation of the Kalman gain is as follows:

$$K_k = P_k^-H^T(HP_k^-H^T + R)^{-1} \tag{6}$$

with H the measurement matrix and R the measurement noise matrix, representing H and v of equation 3 respectively. The Kalman gain is a value between 0 and 1. Note that a high measurement noise R, thus an inaccurate

measurement, leads to a low Kalman gain whereas a low measurement noise $\boldsymbol{R}$, thus an accurate measurement, leads to a high Kalman gain. Using the Kalman gain, the posterior system state vector is calculated. Lastly, the covariance matrix is calculated:

$$\tilde{\boldsymbol{x}}_k^+ = \tilde{\boldsymbol{x}}_k^- + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{H}\tilde{\boldsymbol{x}}_k^-) \tag{7}$$

$$\boldsymbol{P}_k^+ = \boldsymbol{P}_k^-(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}) \tag{8}$$

where $\boldsymbol{z}$ is the obtained measurement. It can be noted that the term $(\boldsymbol{z}_k - \boldsymbol{H}\tilde{\boldsymbol{x}}_k^-)$ is the residual between the obtained measurement and the estimation from the prediction step. As discussed, an inaccurate measurement leads to a lower Kalman gain, thus reducing the influence of the measurement on the posterior state estimate. An accurate measurement, with a high Kalman gain, has a high influence on the posterior state estimate.

The update step can only be performed if a new measurement is available. If this is not available, the new system state is predicted using the prediction step only. Future states are then estimated by extrapolation of the latest known measured state using the state transition model.

In subsection 3.1 it was discussed that two assumptions are made when using a standard Kalman filter. The first assumption is that the noises are additive Gaussian. The second assumption is that the state transition model is linear. This is due to the fact that a transformation of a Gaussian distribution through a non-linear function does not yield a proper Gaussian distribution. Because degradation models involved in prognostics are often non-linear, the following subsection introduces the UKF that can handle non-linear models.

3.3. Unscented Kalman Filter

Two variations of the standard Kalman filter to handle non-linear problems are the extended Kalman filter (EKF) and the unscented Kalman filter (UKF) (LaViola, 2003). The EKF is based on linearization of the model around the mean of the current state estimate. This requires a Jacobian matrix to be calculated each time step. The UKF is based on the idea that it is easier to approximate a Gaussian distribution than to approximate a non-linear function (LaViola, 2003). Because the UKF is easier to implement and in most cases performs better than the EKF (Labbe, 2015), in this work the UKF will be utilized. Because the model is non-linear, the linear matrix representations defined in equations 2 and 3 are rewritten as follows:

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}, \boldsymbol{u}_k) + \boldsymbol{w}_k \tag{9}$$

$$\boldsymbol{z}_k = \boldsymbol{h}(\boldsymbol{x}_k) + \boldsymbol{v}_k \tag{10}$$

where $\boldsymbol{f}$ contains the (non-linear) state transition equations, $\boldsymbol{u}$ the control input equations (which will be neglected) and $\boldsymbol{h}$ the measurement equations. $\boldsymbol{w}$ and $\boldsymbol{v}$ remain Gaussian process noise and measurement noise vectors respectively.

The core idea of the UKF is to select a set of sigma points that describe the most important statistical characteristics of the prior state (Julier, 1998). These sigma points are passed through the non-linear state transition function. Note that this is similar to the PF in which a large number of points is passed through the state transition function. However, in contrast to deriving an arbitrary probability density in the PF, the UKF approximates a new Gaussian with the transformed sigma points. This is called the unscented transform. The unscented transform is visualized in Figure 2. In this study, sigma points proposed by Van Der Merwe (2004) are used. These are considered to provide a good trade-off between performance and accuracy (Labbe, 2015). A more elaborate explanation of the algorithm can be found in Wan & Van Der Merwe (2000).

3.4. Joint State and Parameter Estimation

Kalman filters are in fact state tracking algorithms that estimate the current state accurately, utilizing an uncertain state transition model and noisy measurements. Since for prognostics not only the current health state, but also future degradation is of interest, accurate extrapolation from the latest known measured state is required. This can be done by proper estimation of the model parameters. In this paper, joint state and parameter estimation is considered. This means that model parameters are added to the system state vector $\boldsymbol{x}$:
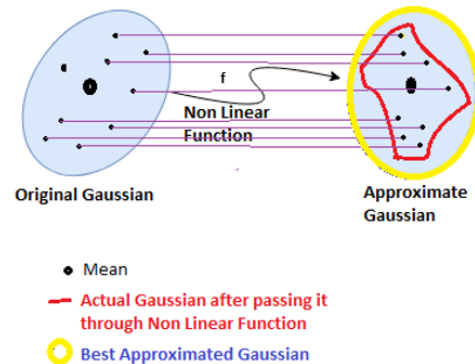

Figure 2. Schematic representation of the unscented transform (modified from Chadha (2018))

$$x = \begin{Bmatrix} s \\ \theta \end{Bmatrix} \tag{11}$$

with s the vector with the tracked states and θ the vector with the model parameters.

The algorithm can be divided in two stages: 1) the filtering stage in which the current state and model parameters are being estimated and 2) a prognostic stage where the latest state is extrapolated into the future, utilizing the obtained model parameters. In the next section, a case study is introduced to which a UKF is applied.

4. CASE STUDY

This section first introduces the case details. Then it is explained how the artificial dataset is generated and how the unscented Kalman filter is specified. The end of this section divides the case study into four scenarios. These four scenarios demonstrate the performance of the filter based on the knowledge of loads during the filtering and prognostic stage of the algorithm, and on the requirement for short-term or long-term prognostics.

4.1. Introduction to the Case

The case study concerns the hull of a naval vessel for which the owner wants to develop a predictive maintenance policy. Common failure modes for structural failures of the hull of a vessel are crack formation, buckling, indent and corrosion (Raju & Anandh, 2018). In this case, a criticality analysis has revealed that formation of fatigue cracks is the most critical failure mode. Causes of fatigue cracks can be for example wave loads (i.e. environmental conditions) or manoeuvring of the ship (i.e. operational conditions). The guidelines of Tiddens et al. (2018) reveal that for this case, a maintenance strategy that incorporates prognostics is suitable. After that, the prognostic ambition level (Tiddens et al., 2020) must be determined. For this case, predictions for a specific vessel are required, so one of the two individual ambition levels that involve prognostics of individual components (ambition level 4 or 5) is selected.

Wave loads and manoeuvring loads are highly varying. It is assumed that no representative historical degradation dataset is available that covers all operating conditions. Implementation of a physics-based degradation model is beneficial in this case (Tiddens et al., 2018). To incorporate a physics-based degradation model, the failure mechanism needs to be evaluated. In this case, the failure mechanism is fatigue. An appropriate degradation model for fatigue is the Paris-Erdogan equation (Paris & Erdogan, 1963), which is selected for the sake of simplicity. The relevant loads are stress ranges and the corresponding degradation measure is

crack length, which is assumed to be directly measurable.

In this situation a hybrid Bayesian filtering approach is a proper choice for a predictive maintenance strategy, as will be shown more generically in section 6. From this point on, the case can be further specified to end up in four specific scenarios. Subsection 4.4 describes these four different scenarios in more detail. But first it is explained how the dataset is generated and how the filter is specified.

4.2. Data Generation

As explained in subsection 4.1, the case study is based on the Paris-Erdogan equation (Paris & Erdogan, 1963). Therefore, the artificial crack growth curve is also generated according to this equation, which is defined as follows:

$$\frac{da}{dN} = C(\Delta K)^m \tag{12}$$

with a the crack length, dN the number of stress cycles, C and m material constants and ΔK the stress intensity factor, given by:

$$\Delta K = K_{max} - K_{min} \tag{13}$$

where K is calculated by:

$$K = F\sigma\sqrt{\pi a} \tag{14}$$

where σ is the stress and F is a geometry factor. For this case, it is assumed that the hull of the vessel can be described as an infinite plate with a crack centred in the middle. In this situation, F is 1 (Tinga, 2013).

Because crack growth is a random process and the crack growth cannot be negative, the curve created by equation 12 is multiplied with a log-normal noise $\omega \sim log - \mathcal{N}(0, 0.63^2)$. To simulate measurements, an additive noise $v \sim \mathcal{N}(0, 0.1^2)$ is added to the real crack length. This leads to the following equations for generation of the artificial crack length and measurements respectively:

$$a_k = a_{k-1} + C(\Delta\sigma\sqrt{\pi a_{k-1}})^m \omega_k \tag{15}$$

$$z_k = a_k + v_k \tag{16}$$

The initial crack length a is 1.14mm. The model parameter C is $2.68 \cdot 10^{-12}$ and m is 3.31, which are realistic values for 316L stainless steel (Wheatley et al., 1998). A zero mean stress is assumed such that the effect of a stress ratio is not present.

Varying conditions are included in two ways: first of all, the magnitudes of the stress ranges ($\Delta\sigma$) vary every 2500 stress

cycles between two levels. Secondly, there are two stress profiles defined. The magnitudes of the two alternating stress ranges are different in each stress profile. The vector with the stress ranges is defined as follows:

$$[\Delta\sigma] = [2500 \cdot [\Delta\sigma_1] + 2500 \cdot [\Delta\sigma_2]] \cdot n_1 \\ + [2500 \cdot [\Delta\sigma_3] + 2500 \cdot [\Delta\sigma_4]] \cdot n_2 \quad (17)$$

where $\Delta\sigma_1$ is 90MPa, $\Delta\sigma_2$ is 30MPa, $\Delta\sigma_3$ is 60MPa and $\Delta\sigma_4$ is 40MPa. n_1 and n_2 indicate how often each specific stress profile occurs. In this case, n_1 is 4 and n_2 is 7, such that a total of 55.000 stress cycles are simulated with a change of the stress profile at 20.000 stress cycles. The stress profiles are visualized in Figure 3.

A total of 200 measurements are simulated, yielding a measurement every 276th stress cycle. The generated crack growth curve and the obtained measurements using this specific stress sequence is visualized in Figure 4.

4.3. Algorithm Settings

In this subsection, the prognostic algorithm settings are described. First, the state vector is configured. Besides the crack length, also the model parameter m is updated. To simplify the model, model parameter C is assumed to be a known fixed value and is therefore not included in the system state vector. This results in the following vector:

$$x = \begin{Bmatrix} a \\ m \end{Bmatrix} \quad (18)$$

Process noise in the standard Kalman filter is included in the process noise matrix Q (see equation 5), but the random noise in this UKF is included in the state transition function. This is due to the fact that the process noise is not additive, but multiplicative. The model parameter m is not described by any mathematical function and is therefore defined to be a
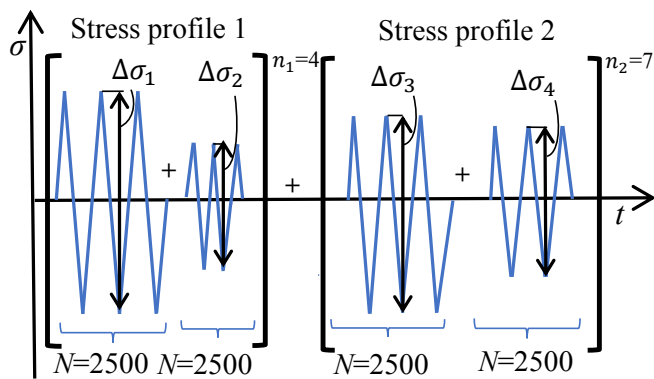

Figure 3. Visualization of the two stress profiles defined in equation 17

constant. This leads to the following state transition functions in the UKF:

$$\tilde{x}_1^k = \tilde{x}_1^{k-1} + (C(\Delta\sigma\sqrt{\pi\tilde{x}_1^{k-1}})^{\tilde{x}_2^{k-1}})\omega^k \quad (19)$$

$$\tilde{x}_2^k = \tilde{x}_2^{k-1} \quad (20)$$

It is assumed that the crack lengths can be measured directly. Therefore, the output of the measurement equation h is the first entry of the predicted system state vector.

The process noise matrix is defined as follows:

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & q_m \end{bmatrix} \quad (21)$$

where q_m represents noise in model parameter m. A high value of q_m means an uncertain model parameter, leading to a filter that quickly adapts the model to fit the measurements. Contrary, a lower value of q_m leads to a filter that adapts slower to these measurements. This also leads to wider and narrower uncertainty bounds around the parameter estimations respectively. Results of different settings of q_m will be discussed in section 5.

The measurement noise matrix contains a variance that matches the simulated measurement noise. Because the measurement equation only considers one value, the crack length, the measurement matrix is a 1×1 matrix:

$$R = \begin{bmatrix} \sigma_m^2 \end{bmatrix} \quad (22)$$

with $\sigma_m = 0.1$.

The initial error covariance matrix describes the expected uncertainty in the initial estimates of the crack length and model parameter m. This matrix is defined as follows:

$$P = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \quad (23)$$

the values of 0.1 are arbitrary selected. Higher values indicate a higher uncertainty on the initial states. That leads to wider uncertainty bounds in the initial iterations of the algorithms. The fact that P is a diagonal matrix suggests that the state a and parameter m are independent of each other. This is not true, as m directly affects the crack growth according to the Paris-Erdogan equation (12). However, the filter will update P such that suitable values for the covariance will be found.

The initial guess of the crack length and model parameter m slightly differ from the real values (which are 1.14mm and 3.31 respectively). By doing so, it can be demonstrated that the filter works even if the initial estimate of the state and its
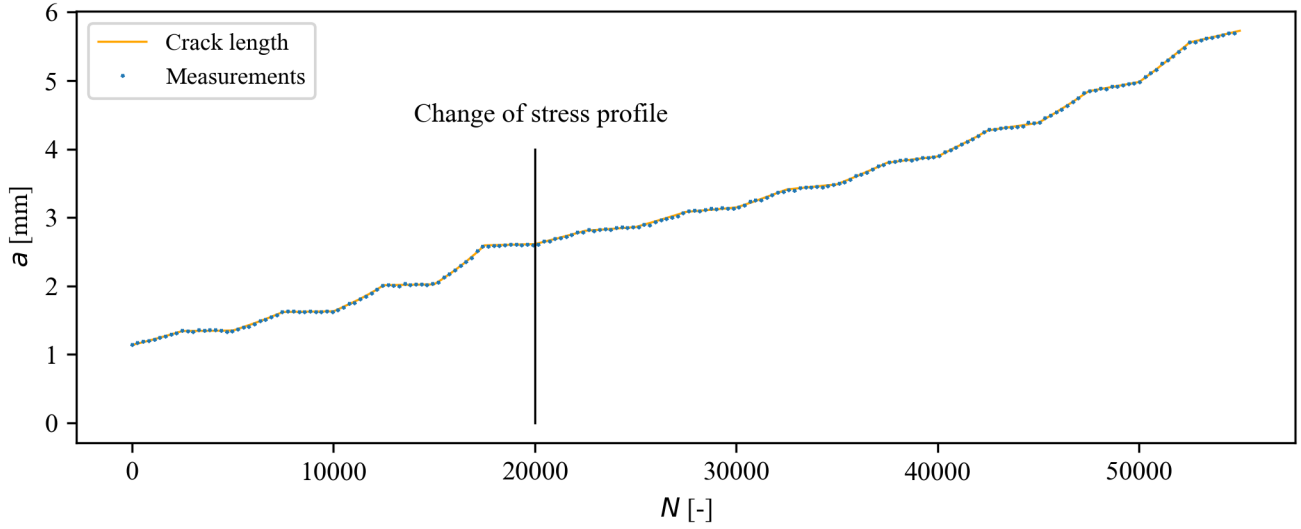
Figure 4. The generated dataset showing the evolution of crack length (a) versus the number of applied stress cycles (N). The dots indicate the generated measurement values (including noise).

transition model is inaccurate. The specified initial estimates are:

$$\boldsymbol{x}_{init} = \begin{Bmatrix} 1.0\text{mm} \\ 2.8 \end{Bmatrix} \tag{24}$$

To do a RUL prediction, the latest state can be extrapolated into the future. This is the prognostic stage of the algorithm. Then, the time difference between the moment the threshold is exceeded (N_{thr}) and the current time step (N_{pred}) can be calculated in terms of stress cycles:

$$\tilde{RUL} = \tilde{N}_{thr} - N_{pred} \tag{25}$$

The failure threshold is set to a crack length of 4.5mm.

4.4. Prognostic Scenarios

In this subsection it is explained how four different prognostics scenarios are defined for the artificial crack growth problem. These scenarios will be discussed in a more general context in section 6. In scenario 1 it is simulated that actual loads are being monitored and that future loads are known. In scenario 2 it is simulated that actual loads are being monitored, but future loads are only approximated. In scenario 3 it is simulated that no loads are being monitored and future loads are unknown, while long-term prognostics are desired. In scenario 4 it is simulated that no loads are monitored and future loads are unknown, while short-term prognostics are required.

In scenario 1 and 2, the actual stress ranges are included in the state transition equation (19) during the filtering stage of the algorithm, because it is assumed that the actual loads are being monitored. The difference between these two scenarios is that it is assumed in scenario 1 that all future stress ranges are known, whereas the future stress ranges are approximated in scenario 2. This means that in scenario 1 the actual future stress ranges are implemented in the prognostic stage of the algorithm. The load approximation in scenario 2 is simulated by randomly increasing the expected stress ranges in the second stress profile with 10% (i.e. $\Delta\sigma_3 = 66$MPa and $\Delta\sigma_4 = 44$MPa) in the prognostic stage. For both scenarios, process noise q_m is specified as $1 \cdot 10^{-7}$ (see equation 21).

In scenario 3 and 4, a fixed assumed stress range is implemented in equation 19 for both the filtering and the prognostic stage. A value of $\Delta\sigma = 60$MPa has been selected as a reasonable approximation of the occurring stress ranges. The difference between these two scenarios is that q_m is specified as $1 \cdot 10^{-7}$ in scenario 3, yielding a low adaptive filter whereas q_m is specified as $1 \cdot 10^{-5}$ in scenario 4, yielding a highly adaptive filter. Although these values are small in an absolute sense, the large amount of time steps makes it an appropriate choice to adapt model parameter m while preserving robustness of the filter.

An overview of the four scenarios is provided in Table 1.

5. RESULTS

This section shows the results of the filtering and prognostic stages of the four described scenarios. Each subsection contains three figures that demonstrate the performance of

Table 1. Overview of the prognostic scenarios

	Loads monitored?	Future loads known?	q_m
Scenario 1	Yes	Yes	$1 \cdot 10^{-7}$
Scenario 2	Yes	Partly	$1 \cdot 10^{-7}$
Scenario 3	No	No	$1 \cdot 10^{-7}$
Scenario 4	No	No	$1 \cdot 10^{-5}$

the filter: 1) an estimation of model parameter m during the filtering stage, 2) the filtering stage up till a specific number of stress cycles N and the prognostic stage from that point on and 3) a comparison of the estimated RUL at different moments compared to the actual RUL at those moments.

5.1. Scenario 1: Monitored Stress Ranges and Known Future Stress Ranges

In the first scenario the actually applied stress ranges are being monitored and known for future operations. When the loads are known, m is the only unknown parameter of the Paris-Erdogan equation (12). For this reason, m converges to its real value with reducing uncertainty. This is shown in Figure 5, where the evolvement of m during the filtering stage is plotted.

This scenario makes the filter highly effective, as it is 1) able to accurately track the current degradation trend and 2) can predict short-term and long-term crack growth. This is shown in Figure 6, which shows for three moments in time how the model tracks the actual crack propagation until a specific moment, and what the predicted crack growth from that moment into the future is. After the filter has converged to the real value of m, predictions can be done as soon as future loads can be estimated. It should be noticed that an error in model parameter m leads to an error in the crack growth prediction that propagates in the prognostic stage. A small error in model parameter m can lead to a large error on the long-term, as it shown in Figure 6a. However, in Figure 7 it can be observed that the estimated RUL quickly converges
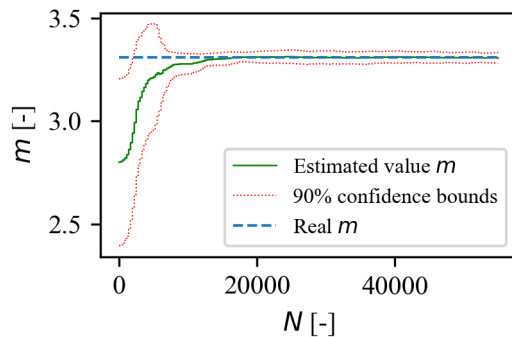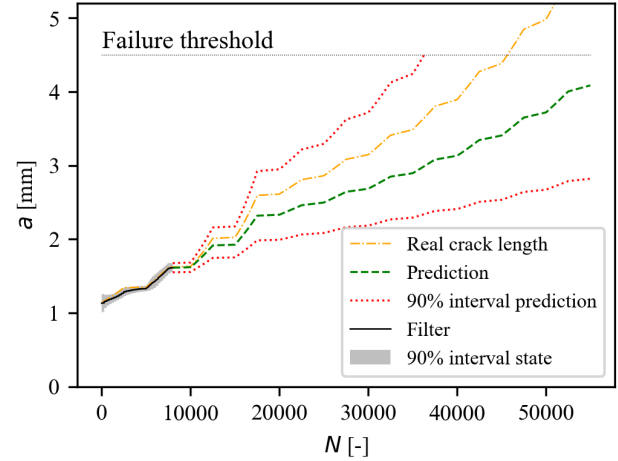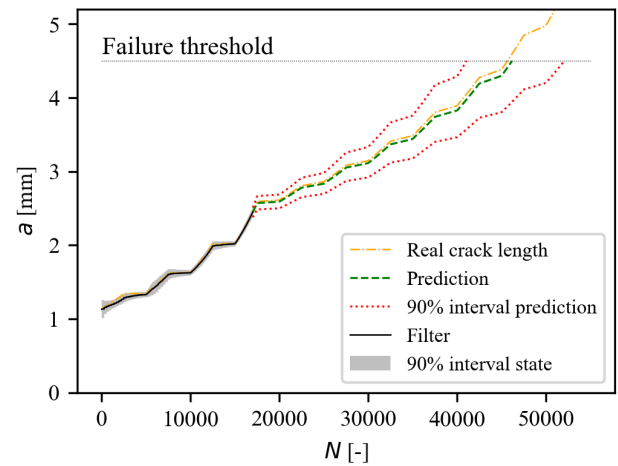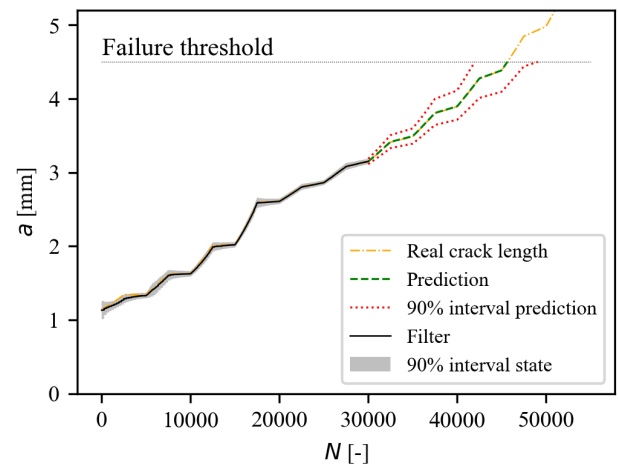

Figure 5. Estimated m during the filtering stage of the filter when loads are being monitored, compared to the actual value of m (scenario 1 and 2)

(a) Prediction after 8000 cycles

(b) Prediction after 17.000 cycles

(c) Prediction after 30.000 cycles

Figure 6. Crack growth predictions of the filter when loads are being monitored and future loads are known (scenario 1)

10

to the actual RUL. This figure shows in each time step what the estimated RUL is, compared to the actual RUL.

In practice, this is the optimal scenario for predictive maintenance. The filter predicts the RUL with the highest accuracy independent of the current operating conditions. However, in practice it is not realistic that the loads of an asset can be determined for its complete RUL beforehand. For example, for a naval vessel its operating profile can be quite certain for the upcoming week, but estimating the operating profile for a longer period of time comes with higher uncertainty. However, a highly accurate prediction is also only needed in the final stages of the life time (close to failure), when the actual maintenance must be executed. In that stage, the operating profile can probably be estimated quite accurately. Much earlier in the life time, when operating profiles are still uncertain, a more rough estimate of the RUL is sufficient to benefit from predictive maintenance. In the next subsection the performance of the filter is evaluated for the case in which future loads are not known exactly.

5.2. Scenario 2: Monitored Stress Ranges and Approximated Future Stress Ranges

In the second scenario the actually applied stress ranges are being monitored, but only an approximated value is available for future operations. This is simulated by randomly overestimating the stress ranges of 10% in the prognostic stage of the filter. Because the estimation of model parameter m only depends on the filtering stage, in which the stress ranges are being monitored, the estimation of m in scenario 2 is equal to the estimation of m in scenario 1 which is visualized in Figure 5.

Similar to scenario 1, the filter can be used to predict degradation paths as soon as a future stress profile can be approximated. Even if the approximated loads are not 100% accurate this 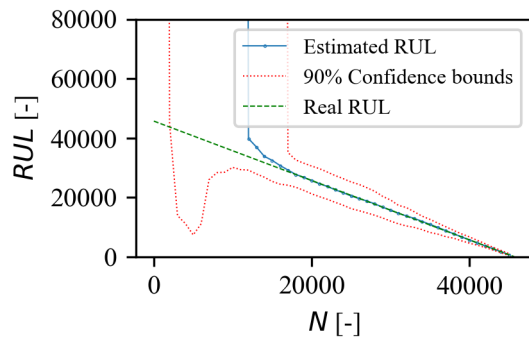is beneficial because the prognostic stage is still independent of the stress ranges observed in the filtering stage. A change in the stress profile such as the one observed at $N = 20.000$ cycles can still be included in the prognostic stage. When extrapolating a trend without understanding the underlying physics, this can generally not be done as the state then needs to be extrapolated from the latest known degradation trend. The predictions for the model of scenario 2 are shown in Figure 8.

Due to the overestimation of the stress ranges in the future, the RUL prediction is constantly underestimated after the model has been optimized in the filtering stage (which is not yet the case in Figure 8a). This underestimation of the RUL is visualized in Figure 9. It should be noticed that the uncertainty bounds only reflect the uncertainty in the model parameter and do not account for uncertainty in future loads. This limitation will be further discussed in section 7.

In practice, this scenario in which future loads can only be approximated is still useful for predictive maintenance. If the filtering stage correctly identified the model parameters of the degradation model, this model provides an insight in how future operations will affect the degradation rates. This can help to make decisions on how adapting the operational conditions of an asset can influence the RUL. Operations could for example be adapted to extend the RUL such that maintenance can be delayed to a more convenient moment in time.
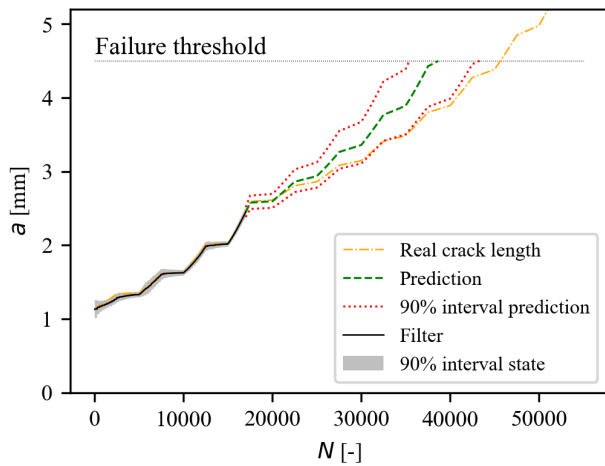
5.3. Scenario 3: Approximated Current and Future Stress Ranges With Low Adaptive Filter

In the third scenario the stress ranges are neither monitored, nor known for future operations. Model parameter m needs to compensate for the error between an assumed stress range and the actual stress ranges. For this reason, the estimated model parameter m does no longer represent the actual model parameter m of the material. The estimation of model parameter m during the filtering stage is shown in Figure 10. It clearly shows that m does not heavily fluctuate to compensate for the local stress ranges, but converges to a more or less constant value for each of the two stress profiles: up till 20.000 stress cycles the model adapts to fit the first stress profile (yielding $m \approx 3.5$) and beyond 20.000 stress cycles the filter adapts to the second stress profile (yielding $m \approx 3.2$). It can be observed that these values are above and below the actual value of m respectively.
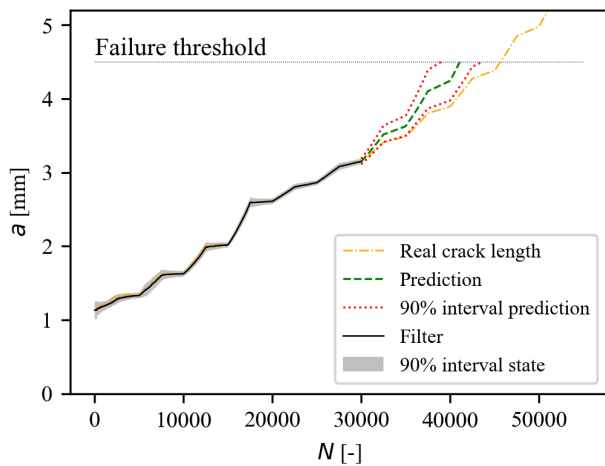
Due to its low adaptivity, this filter is less suitable to estimate the current degradation state with high accuracy. However, it clearly captures the global degradation trend over a longer period of time. This is shown in Figure 11. The prediction after 8000 cycles (Figure 11a) nicely follows the first part of the actual degradation curve. However, a problem arises when the stress profile changes: after 20.000 stress cycles the filter needs time to find a value of m that fits the new stress
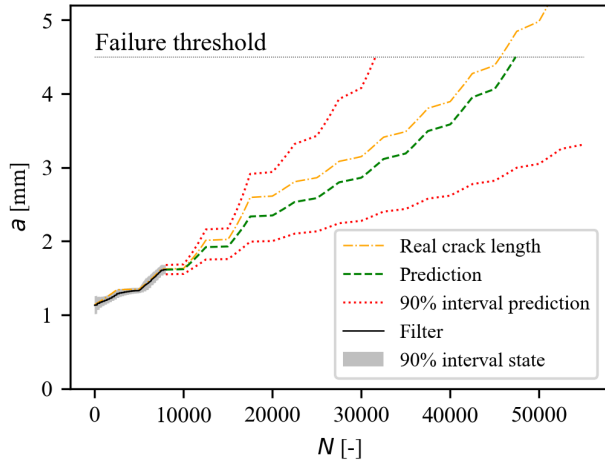
Figure 7. Estimated and calculated RUL in prognostic stage when loads are being monitored and future loads are known (scenario 1)

(a) Prediction after 8000 cycles

(b) Prediction after 17.000 cycles

(c) Prediction after 30.000 cycles

Figure 8. Crack growth predictions of the filter when loads are being monitored and future loads are approximated (scenario 2)
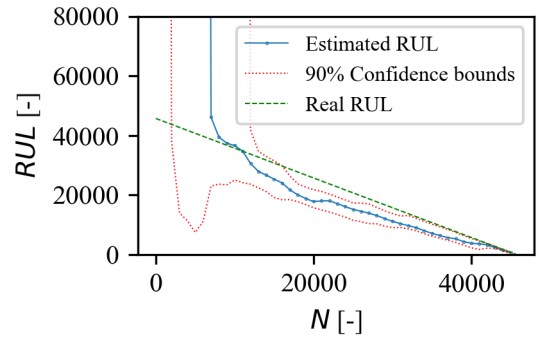

Figure 9. Estimated and calculated RUL in prognostic stage when stress ranges are monitored and approximated for future operations (scenario 2)

profile. From approximately 30.000 stress cycles, the new degradation trend is properly estimated.

The fact that a low adaptive filter can be used if the conditions remain roughly constant, is clarified in Figure 12. During the first stress profile (up till 20.000 cycles), the RUL is underestimated once it has stabilized. However, as soon as the change in the stress profile is detected by the filter, the estimated RUL converges to the actual RUL. It should be noted that the uncertainty bounds in this plot only represent uncertainties of the model parameter m during the filtering stage and do not reflect uncertainties in changing future stress profiles.

This reveals that for RUL predictions in frequently varying conditions, a low adaptive filter is accurate if the operational profile remains constant (i.e. the occurrence and magnitude of stress ranges remain similar). The low adaptivity slightly decreases the accuracy of the algorithm on the short-term, but the filter is much more capable in finding the degradation trend on the long-term. As was mentioned before, in the context of predictive maintenance it is of higher interest to
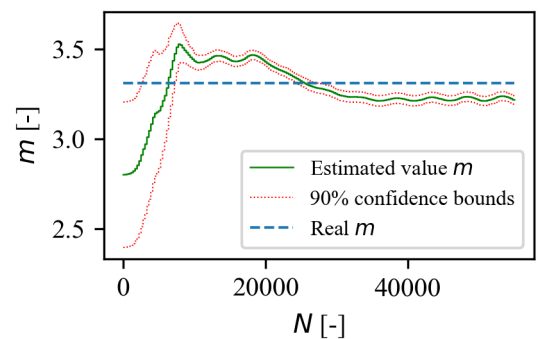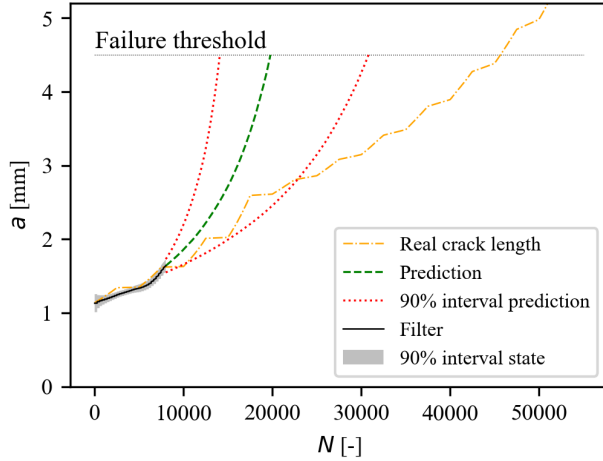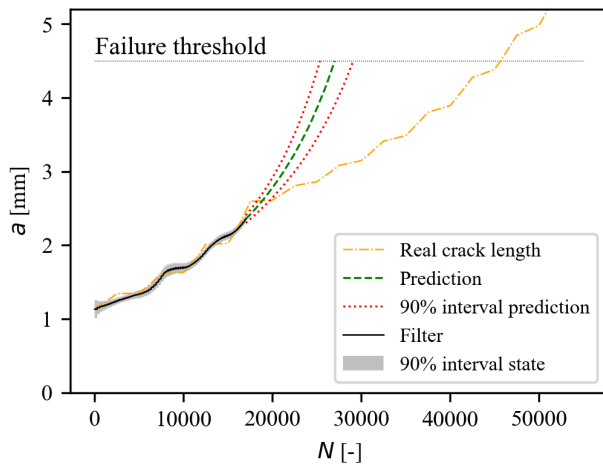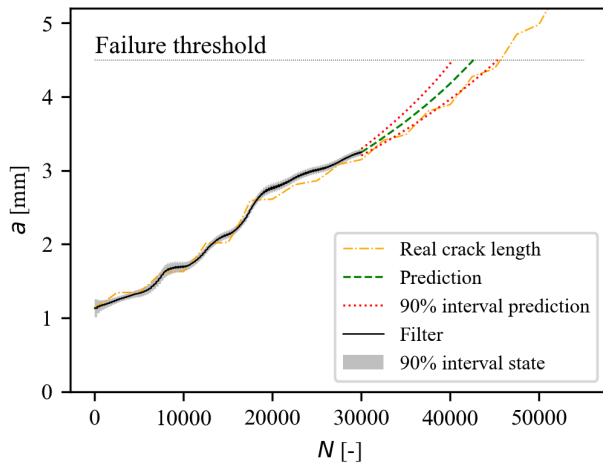

Figure 10. Estimated m during the filtering stage of the low adaptive filter, compared to the actual value of m (scenario 3)

12

(a) Prediction after 8000 cycles

(b) Prediction after 17.000 cycles

(c) Prediction after 30.000 cycles

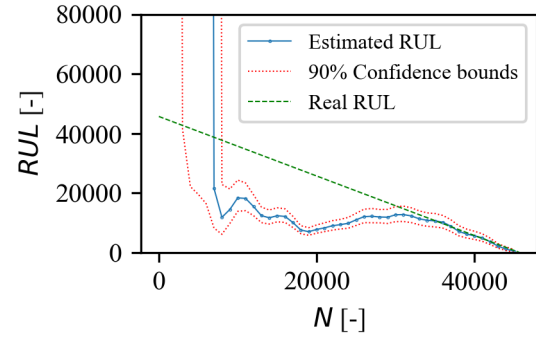Figure 11. Crack growth predictions of the low adaptive filter (scenario 3)

Figure 12. Estimated and calculated RUL in prognostic stage of the low adaptive filter (scenario 3)

approximate the degradation evolution over a longer period of time, than to accurately predict the degradation state on the short-term. In the example of the naval vessel, an estimation of the degradation trend can help to make a decision whether maintenance should be performed the earliest opportunity in the docks or if it can be postponed to a later moment. Furthermore, a prediction of the RUL is useful with respect to logistics to timely plan materials and workspace. However, it should be realized carefully that the predicted degradation trend can still be inaccurate if the stress profile significantly changes.

5.4. Scenario 4: Approximated Current and Future Stress Ranges With Highly Adaptive Filter

The fourth scenario describes the situation in which no stress ranges are being monitored nor known for future operations and a highly adaptive filter is used. This filter is similar to the filter used in scenario 3, but by specifying a higher process noise q_m model parameter m quickly adapts to fit the degradation trend. Similar to scenario 3, model parameter m does not represent the actual material parameter m, but is a parameter that compensates for the mismatch between the actual and the specified loads. In Figure 13 it is shown how model parameter m evolves during the filtering stage.

Predictions of this filter are shown in Figure 14. It can be observed that the highly adaptive filter tracks the local degradation state quite accurately (once it has initialized at ≈ 10.000 cycles). Thereby, the model parameter is adjusted continuously to fit the degradation trend corresponding to the current stress range, when a new measurement is available. However, it should be noted that the state estimation is less accurate compared to scenario 1 and 2. This is due to the fact that the accuracy of the prediction step in the filtering stage of the algorithm (see equation 4 and 5) is lower, as the actual loads are not incorporated.

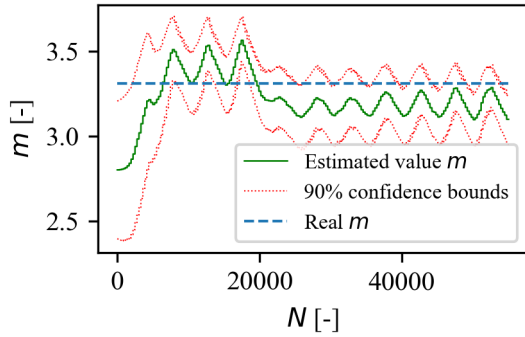Furthermore, Figure 14 shows that for long-term degradation

13

Figure 13. Estimated m during the filtering stage of the highly adaptive filter, compared to the actual value of m (scenario 4)

predictions, this filter is not suitable at all. Because the filter is continuously updating to fit the local stress range, the RUL prediction is highly dependent on the current operating conditions. The fact that the highly adaptive filter is not suitable for long-term prognostics, is visualized in Figure 15, clearly showing that the estimated RUL does not properly converge to the actual RUL.

In the perspective of predictive maintenance this is not optimal. The key information that is needed is a proper estimate of the time (in the far future) at which the crack reaches the critical value, i.e. the RUL. And especially that value is highly volatile when using the current filter, which means that it is unsuitable for decision making. For example, a naval vessel may need to visit a dock when a critical crack is present or may be on a military mission that cannot be interrupted on the short-term. In both situations, knowledge on the crack length in the next day, which can be accurately predicted by the present filter, is not very useful, but longer term predictions are needed. This makes scenario 4 the least effective scenario for predictive maintenance if loads vary frequently. This scenario can be used if loads do not vary in the (short) interval between the moment of prediction and the moment for which a state prediction is required.

6. POSITIONING OF THE PROPOSED APPROACH

Section 5 showed results of the UKF algorithm for varying conditions for a simulated crack growth problem. However, the algorithm is not limited to crack growth problems, but may also be applied for other cases with other failure mechanisms. The aim of this section is to properly position the methodology, clarifying when a Bayesian filtering approach can be used and which of the four scenarios is most suitable. In Figure 16 a positioning diagram is shown that visualizes this. All steps represented in the diagram are discussed in the remainder of this section.
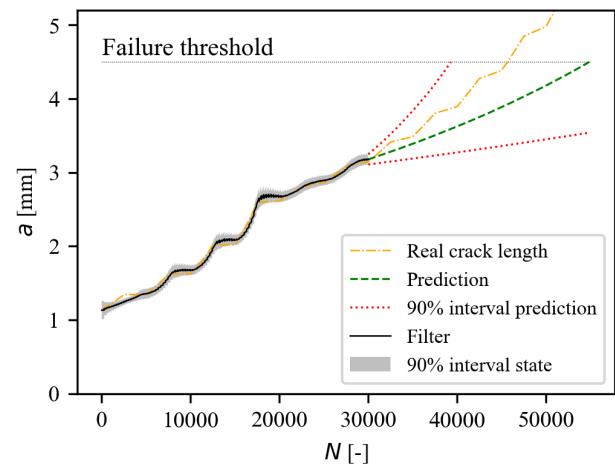
The starting point for the process is that an asset owner
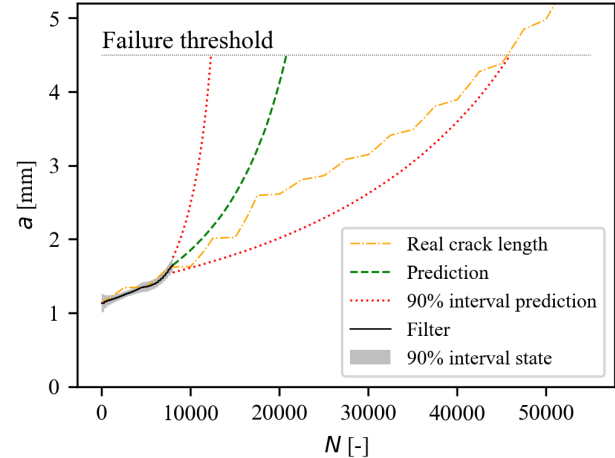
(a) Prediction after 8000 cycles
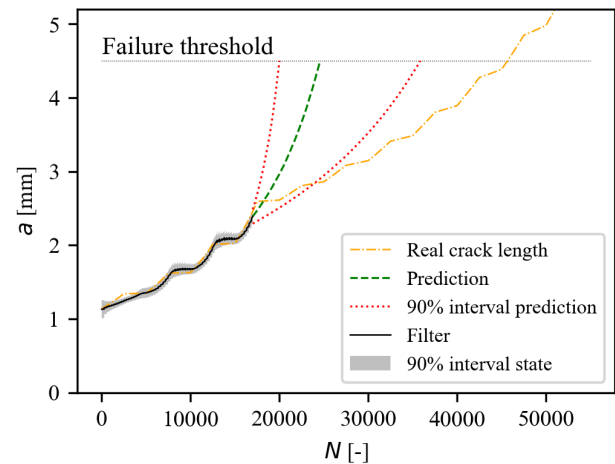
(b) Prediction after 17.000 cycles

(c) Prediction after 30.000 cycles

Figure 14. Crack growth predictions of the highly adaptive filter (scenario 4)
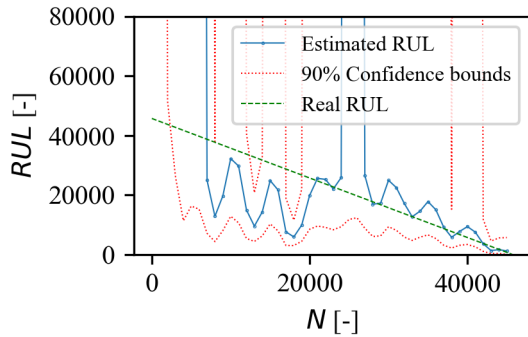
Figure 15. Estimated and calculated RUL in prognostic stage of the highly adaptive filter (scenario 4)

considers to apply a prognostic approach, and needs guidance in selecting the most suitable method. Tiddens et al. (2018) explained that besides criticality of components, also organizational, economic and technical feasibility need to be considered when selecting a new maintenance strategy. After that, selecting the most suitable prognostic method is dictated by the specific situation.

Three criteria have been derived to define for which situation the proposed hybrid Bayesian filtering approach is advantageous: 1) prognostics of a specific asset in specific environments is required. Otherwise, simpler solutions (e.g. based on reliability statistics) can be appropriate (Tiddens et al., 2020). 2) Operating conditions are varying. A large number of methodologies and algorithms for constant operating conditions is already available in literature (Cubillo et al., 2016; Tsui et al., 2015; Guo et al., 2020). 3) Representative degradation data in all possible operating conditions is unavailable. Otherwise, a purely data-driven approach could be adopted. Involvement of a physical model is very useful if future usage can be different from historical usage or if no useful data are available (Tiddens et al., 2020) which is a benefit of the hybrid filter.

The failure mechanism and its governing loads need to be identified if all criteria are fulfilled. This is required to define a suitable degradation model and the corresponding measure of degradation severity. Such a measure can be for example mass loss in case of a wear or corrosion problem or crack length in case of a fatigue problem. It is required to derive this measure either directly or indirectly. Analytical or data-driven measurement models as proposed by Baraldi et al. (2013) are indirect methods. Such a data-driven method requires training data, but it should be noticed that these data are of a diagnostic kind. For this reason, in contrast to data-driven prognostics, it is not required to have large representative datasets of historical degradation paths.

A hybrid Bayesian filtering approach is feasible if the

required degradation parameter can be obtained, which is the core of the present paper. The positioning diagram thus clearly indicates when the proposed UKF method is most suitable, and that in other situations alternative prognostic approaches should be selected. However, different variants of the UKF can be applied, which is shown in the lower part of Figure 16.

The next step is to determine whether loads are being monitored or not. This is important because it determines how the filter is going to be used. If loads are being monitored, the filter will optimize the degradation model by reducing uncertainties in model parameters as demonstrated in subsection 5.1 and 5.2. If loads are not being monitored, the filter will adapt the model parameters such that they compensate for the mismatch between a specified approximated load and the actual loads, as demonstrated in subsection 5.3 and 5.4. There are two ways to obtain actual loads: 1) by measuring them directly using sensors and 2) by estimating them from system usage, e.g. by using physical laws that relate usage of systems to loads.

Monitoring loads yields accurate state estimation during the filtering stage due to the accurate prediction stage of the UKF (equation 4 and 5). However, it should be noted that this does not automatically imply that loads in the prognostic stage are known as well. If future loads can be estimated, this leads to highly accurate prognostics due to the reduced uncertainty in the model parameters (scenario 1). If loads are not exactly known, they need to be approximated (scenario 2). However, if no reasonable approximation can be given, the degradation model is not able to provide an accurate RUL prediction. In this case, it can be a better option to use an adaptive filter which does not require the actual loads as input.

When loads are not being monitored, the accuracy of state estimation (in the filtering stage) depends on the filter settings. A filter that adapts slowly will provide a less accurate state estimation because it is less sensible for the most recent measurement (scenario 3), whereas a highly adaptive filter will quickly adapt to fit the latest known measurement (scenario 4), leading to a higher accuracy in the state estimation. Also the prognostic performance depends on the filter settings. For a long-term prediction, a low adaptive model can be applied. Such a model approximates the degradation trend, fitted on a specific loading profile which can consist of varying loads. This leads to reasonable prognostics if the loading profile remains constant. For a short-term prediction, a highly adaptive model can be adopted. As soon as a change in the operations is detected, e.g. a change in the loads, it immediately adapts to the new operation. In this case, it is assumed that the most recent loads are representative for the future loading.

7. DISCUSSION

Section 5 demonstrated how a hybrid Bayesian filtering approach, in this case an unscented Kalman filter, can be applied for prognostics under varying operational and environmental conditions. Four scenarios have been described in which the amount of physical knowledge and the filter settings varied. However, two major assumptions were done to generate this case study which need to be discussed in more detail to evaluate use of this approach in practice.

First of all, frequent measurements of the direct degradation severity (i.e. crack length) were assumed. Secondly, uncertainties in the case study were simplified and studied to a limited extend. These two assumptions are discussed in the following subsections.

7.1. Availability of Measurements

The assumption that the degradation severity is frequently and directly measured can be divided in two parts: 1) degradation severity is measured directly and 2) measurements are collected frequently. In practice, frequent direct measures of degradation severity are often not available.

To measure degradation severity directly, a measurement device that matches the governing failure mechanism at the specific degrading location is required. This can be expensive and complicated to implement. Another way of collecting data about the degradation severity is to perform inspections. However, then the location of degradation needs to be accessible. Furthermore, a manual inspection will be less feasible to perform frequently compared to a real-time
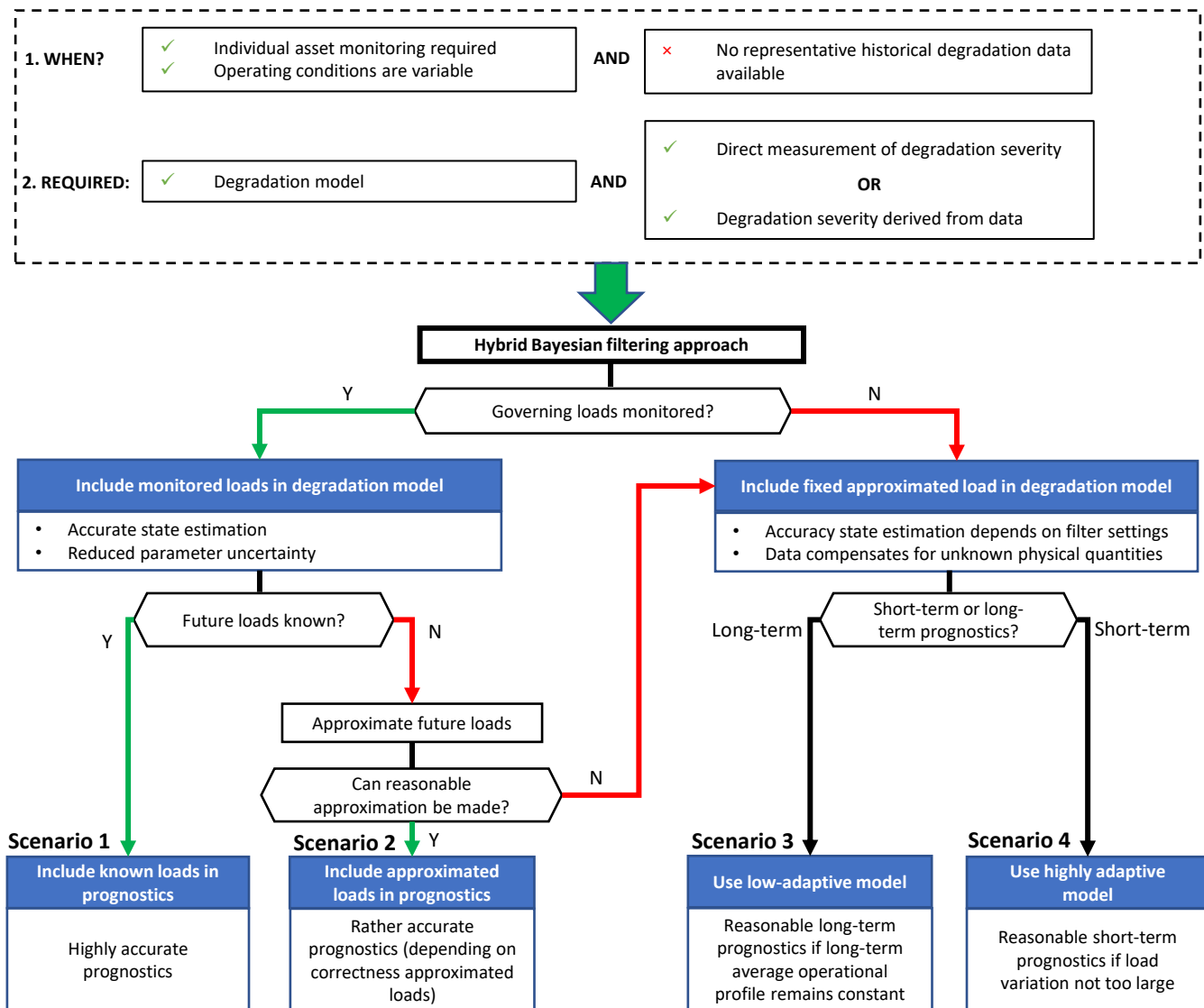

Figure 16. Positioning the prognostic maintenance strategy based on Bayesian filtering

16

monitoring device. A low number of measurements leads to less data points in an algorithm to optimize the degradation model or to adapt the model properly for long-term or short-term prognostics.

However, many assets collect data from e.g. SCADA systems. Also, assets can be equipped with cheaper and less complex sensors such as accelerometers which can be used to describe performance (Lee et al., 2014). Although these data can be used for fault detection or diagnostics or may even reveal degradation trends over time, they do not provide a direct measure of the degradation severity. This makes them hard to directly implement in a Bayesian filtering algorithm while preserving physical knowledge.

To benefit from Bayesian filtering algorithms in practice, it is therefore required to do additional research to either proper measurement devices for degradation severity, or to develop methodologies to infer these measures from indirect signals. This should be possible with proper physical models or development of digital twins. The research field of diagnostics should not only develop methods to diagnose what fault occurs, but also what the degradation severity of the specific fault is.

7.2. Uncertainty Quantification

In a prognostic problem, handling of uncertainties is unavoidable. In the case studies described in this paper, assumptions regarding uncertainty quantification are done in two aspects: 1) uncertainty in loads, 2) uncertainty in the degradation model. These will be discussed in this subsection.

First of all, in scenario 1 and 2 it was assumed that loads were monitored perfectly as no measurement error is included in the monitored stress ranges. Thereby, in scenario 1 all future loads were assumed to be perfectly known whereas in scenario 2 an overestimation of 10% of the actual loads was assumed in the second stress profile. In practice, a wide variety of loading profiles and uncertainties in loads within loading profiles can occur. Further research is required on how these uncertainties in loads can be handled properly, while preserving knowledge on how specific operating profiles affect the degradation rate. If this knowledge is not preserved, prognostics will be too generic such that the desired ambition level may not be achieved. Also the distinction between more certain short-term loads and less certain loads in the far future needs additional attention. Investigating a method that runs multiple models with different characteristics could be useful.

Secondly, the case study describes a problem in which only model parameter m is considered to be uncertain. In practice, also model parameter C will be uncertain. Furthermore, a relatively simple Paris-Erdogan equation is used, while in practice a more comprehensive degradation model can be

more appropriate. A more comprehensive degradation model can contain even more and more uncertain model parameters. Moreover, the UKF is based on approximation of Gaussian distributions, while the real world is not always Gaussian.

Besides uncertainty in the model parameter, a process noise was specified in the algorithm during the filtering and the prognostic stage. Because this process noise was also used to generate the artificial crack growth curve, its distribution was precisely known. In practice it is more complicated to assess the actual process noise. An approximation can be obtained from historical degradation data, but it was concluded that a Bayesian filter is especially useful when limited historical degradation data is available. This emphasizes the significance of understanding the underlying physics of the degradation process, as these may reveal sources of uncertainties.

From this discussion it can be concluded that further research is required on how to implement multiple (non-Gaussian) uncertainties in the methodology. Implementation of a particle filter is a logical next step, in which (non-Gaussian) uncertainties in both operating conditions and for multiple model parameters need to be addressed. A comparison between a PF and UKF can then be made to evaluate the trade-off between accuracy and computational effort of the algorithms.

8. CONCLUSION

This paper proposed to use an unscented Kalman filter for prognostics under varying operational and environmental conditions and provided guidance on the situations in which the proposed method is useful and feasible. The method is demonstrated on an artificial crack growth problem. It is shown that parameter estimation can help to predict degradation, even if the model input parameters are defined poorly.

Where a highly adaptive filter can track the degradation trend locally, a low adaptive filter is more suitable to predict the global degradation trend. To know in detail how operating conditions will affect the degradation rates, it is required to include physical knowledge in the model. In this paper, this was demonstrated by including actual loads in the Paris-Erdogan equation. It illustrated that uncertainty can be reduced by estimation of the model parameters and that predictions can be done if future loads can be estimated with sufficient accuracy.

A positioning diagram is introduced which helps to implement the approach for new case studies with arbitrary failure mechanisms. To make the methodology more applicable in practice, further research is required on obtaining sufficient real-time degradation data and on uncertainty quantification in a physics-based degradation model.

REFERENCES

Arulampalam, M., Maskell, S., N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174-188.

Baraldi, P., Compare, M., Sauco, S., & Zio, E. (2013). Ensemble neural network-based particle filtering for prognostics. *Mechanical Systems and Signal Processing*, *41*(1), 288 - 300.

Chadha, H. S. (2018). *The unscented kalman filter: Anything ekf can do i can do it better!* https://towardsdatascience.com/the-unscented-kalman-filter-anything-ekf-can-do-i-can-do-it-better-ce7c773cf88d. (Retrieved on 26-02-2021)

Chao, M. A., Kulkarni, C., Goebel, K., & Fink, O. (2020). Fusing physics-based and deep learning models for prognostics. , *arXiv:2003.00732*, 1-18. (Unpublished)

Chen, Z. (2003). Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics*, *182*(1), 1-69.

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), 1-21.

Dourado, A., & Viana, F. (2020). Physics-informed neural networks for bias compensation in corrosion-fatigue. In *Aiaa scitech 2020 forum* (p. 1-13). doi: https://doi.org/10.2514/6.2020-1149

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

Elfring, J., Torta, E., & van de Molengraft, R. (2021). Particle filters: A hands-on tutorial. *Sensors*, *21*(2), 1-28.

Frederick, D., DeCastro, J., & Litt, J. (2007). User's guide for the commercial modular aero-propulsion system simulation (c-mapss). *NASA Technical Manuscript*, *2007–215026*, 1-38.

Goebel, K., Eklund, N., & Bonanni, P. (2006). Fusing competing prediction algorithms for prognostics. In *2006 IEEE aerospace conference* (p. 1-10). doi: 10.1109/AERO.2006.1656116

Gordon, N., & Salmondand, D. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEEE Proceedings-F (Radar and Signal Processing)*, *140*(2), 107-113.

Guo, J., Li, Z., & Li, M. (2020). A review on prognostics methods for engineering systems. *IEEE Transactions on Reliability*, *69*(3), 1110-1129.

Jain, A. K., & Lad, B. K. (2020). Prognosticating ruls while exploiting the future characteristics of operating profiles. *Reliability Engineering System Safety*, *202*(C), 1-13.

Jouin, M., Gouriveau, R., Hissel, D., Péra, M.-C., & Zerhouni, N. (2016). Particle filter-based prognostics: Review, discussion and perspectives. *Mechanical Systems and Signal Processing*, *72-73*(5), 2-31.

Julier, S. J. (1998). Skewed approach to filtering. In *Signal and data processing of small targets 1998* (Vol. 3373, p. 271 - 282).

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35-45.

Kumar, S., Torres, M., Chan, Y. C., & Pecht, M. (2008). A hybrid prognostics methodology for electronic products. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)* (p. 3479-3485). doi: 10.1109/IJCNN.2008.4634294

Labbe, R. R. (2015). *Kalman and bayesian filters in python.* https://github.com/rlabbe/Kalman-and-Bayesian-Filters-in-Python. (Retrieved on 12-03-2021)

LaViola, J. J. (2003). A comparison of unscented and extended kalman filtering for estimating quaternion motion. In *Proceedings of the 2003 american control conference, 2003.* (Vol. 3, p. 2435-2440).

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. *Mechanical Systems and Signal Processing*, *42*(1), 314-334.

Li, N., Gebraeel, N., Lei, Y., Bian, L., & Si, X. (2019). Remaining useful life prediction of machinery under time-varying operating conditions based on a two-factor state-space model. *Reliability Engineering System Safety*, *186*(2), 88-100.

Liao, L., & Köttig, F. (2014). Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Transactions on Reliability*, *63*(1), 191-207.

Liao, L., & Köttig, F. (2016). A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. *Applied Soft Computing*, *44*(7), 191 - 199.

Montero Jimenez, J. J., Schwartz, S., Vingerhoeds, R., Grabot, B., & Salaün, M. (2020). Towards multi-model approaches to predictive maintenance: A systematic

literature survey on diagnostics and prognostics. *Journal of Manufacturing Systems*, *56*(7), 539-557.

Paris, P., & Erdogan, F. (1963). A critical analysis of crack propagation laws. *Journal of Fluids Engineering, Transactions of the ASME*, *85*(4), 528-533.

Pillai, P., Kaushik, A., Bhavikatti, S., Roy, A., & Kumar, V. (2016). A hybrid approach for fusing physics and data for failure prediction. *International Journal of Prognostics and Health Management*, *7*(4), 2153-2648.

Rai, R., & Sahu, C. (2020). Driven by data or derived through physics? a review of hybrid physics guided machine learning techniques with cyber-physical system (cps) focus. *IEEE Access*, *8*(4), 71050-71073.

Raju, M., & Anandh, A. (2018). A study on common ship structural failures. *International Journal of Mechanical Engineering and Technology*, *9*(7), 746-754.

Rezamand, M., Kordestani, M., Orchard, M. E., Carriveau, R., Ting, D. S.-K., & Saif, M. (2021). Improved remaining useful life estimation of wind turbine drivetrain bearings under varying operating conditions. *IEEE Transactions on Industrial Informatics*, *17*(3), 1742-1752.

Saho, K., & Masugi, M. (2015). Automatic parameter setting method for an accurate kalman filter tracker using an analytical steady-state performance index. *IEEE Access*, *3*(10), 1919-1930.

Taheri, E., Kolmanovsky, I. V., & Gusikhin, O. (2019). Survey of prognostics methods for condition-based maintenance in engineering systems. *ArXiv*, *abs/1912.02708*, 1-74. (Unpublished)

Thrun, S., Burgard, W., & Fox, D. (2005). *Probabilistic robotics (intelligent robotics and autonomous agents)*. The MIT Press.

Tiddens, W. (2018). *Setting sail towards predictive maintenance: developing tools to conquer difficulties in the implementation of maintenance analytics* (Unpublished doctoral dissertation). University of Twente.

Tiddens, W., Braaksma, A., & Tinga, T. (2018). Selecting suitable candidates for predictive maintenance. *International Journal of Prognostics and Health Management*, *9*(1), 1-14.

Tiddens, W., Braaksma, J., & Tinga, T. (2020). Exploring predictive maintenance applications in industry. *Journal of quality in maintenance engineering*, 1-18. (ahead-of-print) doi: https://doi.org/10.1108/JQME-05-2020-0029

Tinga, T. (2013). *Principles of loads and failure mechanisms. applications in maintenance, reliability and design*. Springer.

Tinga, T., Wubben, F., Tiddens, W. W., Wortmann, H., & Gaalman, G. (2021). Dynamic maintenance based on functional usage profiles. *Journal of Quality in Maintenance Engineering*, *27*(1), 21-42.

Tsui, K.-L., Chen, N., Zhou, Q., Hai, Y., & Wang, W. (2015). Prognostics and health management: A review on data driven approaches. *Mathematical Problems in Engineering*, *2015*(5), 1-17.

Van Der Merwe, R. (2004). *Sigma-point kalman filters for probabilistic inference in dynamic state-space models* (Unpublished doctoral dissertation). OGI School of Science & Engineering, Oregon Health Science University.

Wan, E., & Van Der Merwe, R. (2000). The unscented kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 adaptive systems for signal processing, communications, and control symposium* (Vol. Cat. No.00EX373, p. 153-158).

Wang, J., Liang, Y., Zheng, Y., Gao, R. X., & Zhang, F. (2020). An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples. *Renewable Energy*, *145*(1), 642-650.

Wang, Y., Gogu, C., Binaud, N., Bes, C., & Fu, J. (2019). A model-based prognostics method for fatigue crack growth in fuselage panels. *Chinese Journal of Aeronautics*, *32*(2), 396-408.

Wang, Y., Peng, Y., & Chow, T. W. S. (2021). Adaptive particle filter-based approach for rul prediction under uncertain varying stresses with application to hdd. *IEEE Transactions on Industrial Informatics*, *17*(9), 6272-6281.

Wheatley, G., Niefanger, R., Estrin, Y., & Hu, X. (1998). Fatigue crack growth in 316l stainless steel. *Key Engineering Materials*, *145-14*(2), 631-636.

Yang, Z., Eddy, D., Krishnamurty, S., Grosse, I., Denno, P., Lu, Y., & Witherell, P. (2017). Investigating grey-box modeling for predictive analytics in smart manufacturing. In *Asme 2017 international design engineering technical conferences and computers and information in engineering conference* (Vol. 2B, p. 1-10).

Yucesan, Y., & Viana, F. (2020). A physics-informed neural network for wind turbine main bearing fatigue. *International Journal of Prognostics and Health Management*, *11*(1), 1-17.

Zendehboudi, S., Rezaei, N., & Lohi, A. (2018). Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. *Applied Energy*, *228*(10), 2539 - 2566.

Zhao, F., Tian, Z., Bechhoefer, E., & Zeng, Y. (2015). An integrated prognostics method under time-varying operating conditions. *IEEE Transactions on Reliability*, *64*(2), 673-686.

Zhao, F., Tian, Z., & Zeng, Y. (2013). Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method. *IEEE Transactions on Reliability*, *62*(1), 146-159.

BIOGRAPHIES

Luc S. Keizers received his MSc degree in Mechanical Engineering at the University of Twente in 2020. He graduated on "Structural Fatigue Analysis using Flexible Multibody Dynamics" in the research chair of Applied Mechanics and Data Analysis. Shortly after, he started his PhD at the same university in the chair of Dynamics Based Maintenance in collaboration with the Netherlands Royal Navy. His research interest is in prognostics that combine his engineering background with data-driven models.

Richard Loendersloot received a MSc degree in Mechanical Engineering (2001) and a did his PhD research at the University of Twente, on thermoset resin flow processes through textile reinforcements during composite production process Resin Transfer Moulding and obtained his PhD degree in 2006. He worked in an engineering office on high-end FE simulations of a variety mechanical problems to return to the University of Twente in 2008 as assistant professor for Applied Mechanics. His research started to focus on vibration based structural health and condition monitoring, being addressed in both research and education. He became part of the research chair Dynamics Based Maintenance upon its initiation in 2012. His research covers a broad range of applications: from rail infra structure monitoring, to water mains condition inspection and aerospace health monitoring applications, using both structural dynamics and ultrasound methods. He is involved in a number of European and National funded research projects. He became associate professor in 2019, currently in charge of the daily lead of the Dynamics Based Maintenance group.

Tiedo Tinga is a full professor in dynamics based maintenance at the University of Twente since 2012 and full professor Life Cycle Management at the Netherlands Defence Academy since 2016. He received his PhD degree in mechanics of materials from Eindhoven University in 2009. He is chairing the smart maintenance knowledge center and leads a number of research projects on developing predictive maintenance concepts, mainly based on physics of failure models, but also following data-driven approaches.