Unscented Kalman Filter with Gaussian Process Degradation Model for Bearing Fault Prognosis

Christoph Anger B.Sc.[1], Dipl.-Ing. Robert Schrader[1], and Prof. Dr.-Ing. Uwe Klingauf[1]

[1] *Institute of Flight Systems and Automatic Control, Darmstadt, 64287, Germany*
anger@fsr.tu-darmstadt.de
schrader@fsr.tu-darmstadt.de
klingauf@fsr.tu-darmstadt.de

ABSTRACT

The degradation of rolling-element bearings is mainly stochastic due to unforeseeable influences like short term overstraining, which hampers the prediction of the remaining useful lifetime. This stochastic behaviour is hardly describable with parametric degradation models, as it has been done in the past. Therefore, the two prognostic concepts presented and examined in this paper introduce a nonparametric approach by the application of a dynamic Gaussian Process (GP). The GP offers the opportunity to reproduce a damage course according to a set of training data and thereby also estimates the uncertainties of this approach by means of the GP's covariance. The training data is generated by a stochastic degradation model that simulates the aforementioned highly stochastic degradation of a bearing fault. For prediction and state estimation of the feature, the trained dynamic GP is combined with the Unscented Kalman Filter (UKF) and evaluated in the context of a case study. Since this prognostic approach has shown drawbacks during the evaluation, a multiple model approach based on GP-UKF is introduced and evaluated. It is shown that this combination offers an increased prognostic performance for bearing fault prediction.

1. INTRODUCTION

Forecasting the remaining useful lifetime (RUL) of line-replaceable units (LRUs) with a high accuracy is one of the main issues in aviation to avoid unnecessary maintenance cycles and, therefore, to reduce aircraft life cycle costs. One component of those LRUs can be rolling-element bearings, whose RUL is of great interest and are therefore in the centre of this enquiry.

Rolling-element bearings ensure the functionality of rotating assembly parts in case of varying loading and frequency. During their life cycle, bearings degrade in two different ways

according to the amount, duration and the nature of loading and other influences. Those are e.g. contaminants or constructive incertitudes that can affect the tribological system of a bearing. Overstraining and solid friction caused by a high cyclic stress and a lack of lubricant, respectively, leads to a rapid degradation of the bearing, as in case of calculated strains like wear and tear or fatigue, the course of damage increases continuously. A degradation process of a real bearing results from both kinds of strains and, therefore, has a strong stochastic character (Sturm, 1986).

To simulate this behaviour, several approaches of degradation models (DMs) have been developed in the past. Most of them base on the Paris-Erdogan law that describes a relation between crack growth rate and effective stresses in the examined material. By adjusting this law to existing test results of real bearings, several enhancements were formulated and evaluated, as Choi et al. did in (Choi & Liu, 2006b) and (Choi & Liu, 2006a), respectively. Other DMs are based on the Lundberg-Palmgren model that describes the correlation between the probability of survival and among others the maximum shearing stress. Yu et al. refined this approach by adding a more precise geometrical description of the contact surface in (Yu & Harris, 2001).

All aforementioned models describe the degradation to depend on the application of external load difference, as a non-loaded bearing would not degrade at all. In reality, the degradation is also a function of the current degradation, since detached particles can lead to solid friction. The DM in this paper that is applied to generate reliable degradation courses considers both the degradation rate due to loading and due to the state of degradation itself.

Most of these DMs are used as prognostic models (PMs) in combination with state estimation. Usually particle filters based on the aforementioned Paris-Erdogan model are implemented, as done in (Orchard & Vachtsevanos, 2009). Other prognostic concepts are based on the Archard wear equation. Daigle et al. presented a model-based prognostic approach by estimating the RUL of a pneumatic valve with a fixed-lag par-

ticle filter (Daigle & Goebel, 2010). The appropriated model relates the current degradation to the wear of material based on the Archard equation.

Besides a DM that accounts for the current degradation state, Orsagh et al. presented a prognosis approach of a rolling-element bearing (Orsagh, Sheldon, & Klenke, 2003). By measuring several features like e.g. the oil debris of the bearing or the vibration signal, they predicted the RUL depending on the measured fused features by correlation with the current state of degradation. The RUL was then forecast according to the applied PM.

The prognostic concept at hand attempts another approach, as it is not based on a physical model. Therefore, a dynamic Gaussian Process (GP) model is trained on a degradation process and combined with the Unscented Kalman Filter (UKF) for state estimation. Ko et al. analysed this dynamic prognostic model in (Ko, Klein, Fox, & Haehnel, 2007) by tracking an autonomous micro-blimp. Additionally, the expected benefits of the GP-UKF concept in combination with a multiple model approach is examined.

This paper is divided into four parts. In Section 2 the appropriated DM of a rolling-element bearing is presented. The prognostic approach with a short introduction in the two components UKF and GP and the multiple model approach are described in Section 3 and in Section 4 the two concepts are tested and evaluated in the context of a case study.

2. BEARING FAULT DEGRADATION

As the objective of this paper is to forecast the degradation of a bearing, a feature has to be identified that directly corresponds to the current state of degradation. One variable is the surface of pitting A, i.e. excavation of macroscopic particles caused by material fatigue, either in rolling-elements or in the inner- or outer-race. As one pitting does not immediately lead to the failure of a bearing, its functionality remains. However, this fault produces single impacts due to the geometrical irregularity to the assembly group directly contacting the bearing. Depending on the location of the fault, these impacts appear with certain frequencies as a function of the rotation speed Ω of the shaft, the number of rolling elements n and geometric magnitudes, summarised by Antoni et al. in (Antoni, 2007) and depicted in Table 1.

Inner-race fault	$\frac{n}{2}\Omega(1 + \frac{d}{D}\cos\theta)$
Outer-race fault	$\frac{n}{2}\Omega(1 - \frac{d}{D}\cos\theta)$
Rolling-element fault	$\frac{D\Omega}{d}(1 - (\frac{d}{D}\cos\theta)^2)$
Cage fault	$\frac{\Omega}{2}(1 - \frac{d}{D}\cos\theta)$
Inner-race modulation	Ω
Cage modulation	$\frac{\Omega}{2}(1 - \frac{d}{D}\cos\theta)$

Table 1. Ω = speed of shaft; d = bearing roller diameter; D = pitch circle diameter; n = number of rolling elements; θ = contact angle
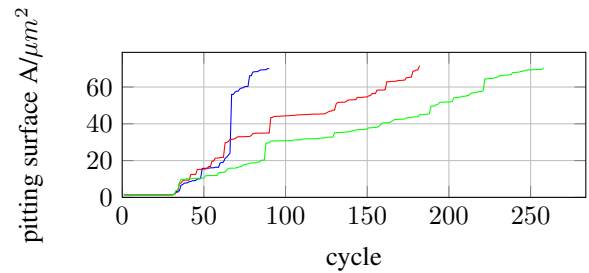
These impacts produce a structure-borne noise and by using an acoustic emission sensor, the frequency and the amplitude of the impacts generated by the rotating fault can be detected. Herewith the location of the fault (according to the frequency and Table 1) and the degree of degradation can be determined as the acceleration amplitude is assumed to correlate with the pitting surface and, therefore, the current condition of the bearing.

The degradation of a real rolling-element bearing results from two different courses of damage, as described in the introduction: a continuously rising damage caused by material fatigue that is interrupted by abrupt steps as a result of overstraining. Both effects can be mathematically described in the appropriated DM by the following differential equation of the degradation, i.e. pitting surface A
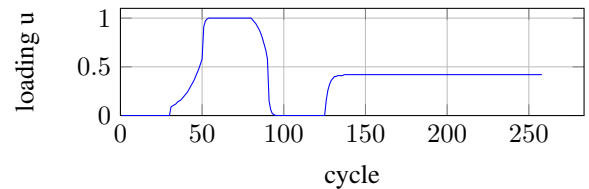
$$\Delta A_i = k_A \cdot A_{i-1} + k_u \cdot \Delta u_{i-1}, \qquad (1)$$

where Δu_i is the external loading difference of the bearing during one cycle and $k_A \sim \mathcal{W}(\lambda'(A_{i-1}), k'(A_{i-1}))$ is a factor drawn from a Weibull distribution, whose scaling parameter k' and shape parameter λ' are expected to be functions of the previous degradation A_{i-1}. The product $k_A \cdot A$ represents the influence of the degradation on the transition rate. Analogously, $k_u \cdot \Delta u$ stands for the increased degradation rate caused by loading, as $k_u \sim \mathcal{E}(\mu(A_{i-1}))$ is drawn from an exponential distribution, whereas the mean μ is also a function of A_{i-1}. Both coefficients k_A and k_u realise the stochastic character of the degradation. Therefore, the current degradation in cycle i can be calculated as

$$A_i = A_{i-1} + \Delta A_i. \qquad (2)$$

(a)

(b)

Figure 1. (a) three different degradation courses as the result of Equation (1), (b) applied normalised load spectrum

In Figure 1, three different damage courses generated by the DM and the applied normalised loading are depicted. Figure 1a clearly shows the stochastic character of a real degradation, as the RUL of all courses differ strongly and the mainly continuous course is interrupted by steps in case of high strain. The correlation between the applied loading in Figure 1b and the degradation rate is obvious, as the load difference between cycle 100 and 125 is zero and the degradation in this range is quite flat. Thus, the applied DM is assumed to reproduce the damage course of a faulty bearing for the use of this paper instead of real test rig measurements.

3. PROGNOSTIC APPROACH

The applied prognostic concepts are introduced in this section. The UKF is used for state estimation and prediction of the degradation. Instead of a parametric model, the UKF is founded on a trained dynamic GP. The basics of both prognostic tools are presented in the following subsections. Subsection 3.1 is based on (Ko et al., 2007), (Ko & Fox, 2011) and (Rasmussen & Williams, 2006). In Subsection 3.3, a multiple model approach that promises an increased prognostic performance is explained.

3.1. Dynamic Gaussian Process for Fault Degradation

The GP offers the feasibility of learning regression function from sample data without any parametric model. Rasmussen et al. describe the GP in (Rasmussen & Williams, 2006) as defining a Gaussian distribution over a function. In other words, the GP establishes a function f out of a given training data set $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ...(\mathbf{x}_n, y_n)\}$ according to a given noisy process

$$\mathbf{y} = f(\mathbf{X}) + \epsilon, \tag{3}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$ is a $n \times m$ input matrix with m the number of inputs and n the length of the single input vector $\mathbf{x}_i$. $\mathbf{y}$ is a n-dimensional vector of scalar outputs and ϵ represents a noise term, which is drawn from a Gaussian distribution $\mathcal{N}(0, \sigma^2)$.

A Gaussian distribution is basically defined by its mean μ and covariance Σ. Therefore, the GP defines a zero-mean joint Gaussian distribution over the given outputs $\mathbf{y}$ of the training data D, as follows

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{0}; \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2\mathbf{I}). \tag{4}$$

The covariance of this joint distribution consists of the kernel matrix $\mathbf{K}(\mathbf{X}, \mathbf{X})$ that represents the deviation of the inputs among each other and the term $\sigma_n^2\mathbf{I}$ for the Gaussian noise caused by ϵ. The entries of $\mathbf{K}$ are the kernel functions $k(\mathbf{x}_i, \mathbf{x}_j)$, where the squared exponential

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)\mathbf{W}(\mathbf{x}_i - \mathbf{x}_j)^T). \tag{5}$$

is a standard kernel function. Here, σ_f^2 is the signal variance and $\mathbf{W}$ is a diagonal matrix that contains the distance measure

of every input.

To calculate the mean GP_μ and the covariance GP_Σ out of a given test input $\mathbf{x}_*$ and test output y_* w.r.t. the training data D, the following expression can be applied

$$GP_\mu(\mathbf{x}_*, D) = \mathbf{k}_*^T[\mathbf{K} + \sigma_n^2\mathbf{I}]^{-1}\mathbf{y} \tag{6}$$

for the mean and

$$GP_\Sigma(\mathbf{x}_*, D) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T[\mathbf{K} + \sigma_n^2\mathbf{I}]^{-1}\mathbf{k}_* \tag{7}$$

for the covariance. Here, the compact form $\mathbf{K}(\mathbf{X}, \mathbf{X}) = \mathbf{K}$ and $\mathbf{k}_*$ the covariance function between the test input $\mathbf{x}_*$ and the training input vector $\mathbf{X}$ is used. Obviously, the mean prediction in Equation (6) is a linear combination of the training output $\mathbf{y}$ and the correlation between test and training input. The covariance is the difference of the covariance function w.r.t. the test inputs and the information from the observation $\mathbf{k}(\mathbf{x}_*, \mathbf{x}_*)$.

The GP possesses three so-called hyperparameters $\theta = [\mathbf{W} \quad \sigma_f \quad \sigma_n]$ from the kernel function and the process noise. Optimal hyperparameters θ can be found by maximising the log likelihood

$$\theta_{max} = \arg\max_\theta \{\log(p(\mathbf{y}|\mathbf{X}, \theta))\}. \tag{8}$$

Considering a stochastic dynamic degradation process, Equation (3) can be written as

$$r_{k+1} = r_k + \Delta r_k + \epsilon_k. \tag{9}$$

Therefore, the state transition Δr_k is trained with the GP. The generated training data set $D_r = \{X, X'\}$ consists of the inputs $X = [(r_1, \Delta u_1), (r_2, \Delta u_2), ..., (r_n, \Delta u_n)]$ and the state transition $X' = [\Delta r_1, \Delta r_2, ..., \Delta r_n]$, which is calculated as

$$\Delta r_k = r_k - r_{k-1} \tag{10}$$

or w.r.t. the mean of the dynamic GP of Equation (6)

$$r_k = r_{k-1} + GP_\mu(u_{k-1}, r_{k-1}, D_r) \tag{11}$$

and the covariance $GP_\Sigma(u_{k-1}, r_{k-1}, D_r)$, both fully describe the Gaussian distribution of the GP. The additional benefit generated by this approach is the time invariance caused by the transition from a static to the dynamic system and the ability to capture different kinds of degradation processes without physical knowledge of the actual process.

3.2. Combining GP and Unscented Kalman Filter

In case of a nonlinear dynamic system, the application of the UKF is the appropriate choice, because it estimates the state of nonlinear systems by means of observation $\mathbf{z}$ and system inputs $\mathbf{u}$. As the presented prognostic approach intends to omit a physical degradation model, the Extended Kalman filter is also inapplicable, since an analytic model is required due to the linearisation step.

In general, a nonlinear dynamic system in kth time step can be described as

$$\mathbf{x}_k = G(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \epsilon_k \tag{12}$$

with the state transition function G, the n-dimensional state vector $\mathbf{x}$, the input vector $\mathbf{u}$ and an additive Gaussian noise term ϵ drawn from a zero-mean Gaussian distribution $\epsilon \sim \mathcal{N}(0, Q_k)$ with the process noise Q_k as covariance.

An analogue description of the observation $\mathbf{z}_k$ can be formulated as

$$\mathbf{z}_k = H(\mathbf{x}_k) + \delta_k. \tag{13}$$

Here, H relates the state to the observation and δ is also an additive noise term $\delta \sim \mathcal{N}(0, R_k)$, where R_k is the measure noise.

Through the scaled unscented transformation by Julier et al. (Julier, 2002) sigma points $\chi^{[i]}$ are defined according to the covariance Σ and the mean μ of the previous time step

$$
\begin{aligned}
\chi^{[0]} &= \mu \\
\chi^{[i]} &= \mu + (\sqrt{(n+\lambda)\Sigma})_i \quad \text{for } i = 1, ..., n \\
\chi^{[i]} &= \mu - (\sqrt{(n+\lambda)\Sigma})_{i-n} \text{for } i = n+1, ..., 2n,
\end{aligned} \tag{14}
$$

where λ is a scaling parameter that, in case of the scaled unscented transformation, is defined as

$$\lambda = \alpha'^2(n+\kappa) - n. \tag{15}$$

Here, α' and κ are further scaling parameters to determine the spread of the sigma points. These sigma points according to the standard UKF are transformed depending on function G to generate a new distribution with mean and covariance.

As the applied UKF contains the dynamic GP, this state transition function G is replaced by the Gaussian predictive distribution of Equation (6) and thereby defines a new set of sigma points

$$\bar{\chi}_k^{[i]} = GP_\mu(\chi^{[i]}, D). \tag{16}$$

Similarly, the process noise Q_K is defined by Equation (7). With this information, a priori mean and covariance can be generated by

$$
\begin{aligned}
\tilde{\mu} &= \sum_{i=0}^{2n} w_m^{[i]} \bar{\chi}_k^{[i]} \\
\tilde{\Sigma} &= \sum_{i=0}^{2n} w_c^{[i]} (\bar{\chi}_k^{[i]} - \mu')(\bar{\chi}_k^{[i]} - \mu')^T + \\
&\quad + GP_\Sigma(\mathbf{x}_*, D)
\end{aligned} \tag{17}
$$

with weights w_m and w_c set up in (Julier, 2002).

The whole applied GP-UKF algorithm is depicted in Table 2. In comparison to Equation (16), the new sigma points in line 3 are generated by $\bar{\chi}_k^{[i]} = \chi_{k-1} + GP_\mu(u_{k-1}, \chi_{k-1}^{[i]}, D_G)$, as the applied GP is trained according to Equation (9). The prediction of the mean and covariance in time step k described in Equation (14) to (17) takes places from line 1 to 5. A priori estimation is corrected according to the measured observation $\mathbf{z}_k$ from line 7 to 13. This correction step proceeds similarly to the prediction.

In line 6 the transformed sigma points of line 3 $\bar{\chi}_k$ are used as observation $\hat{\mathcal{Z}}_k^{[i]}$. Line 8 is comparable to line 5 and in line

1:	Inputs $\mu_{k-1}, \Sigma_{k-1}, u_{k-1}, z_k, R_k$
2:	$\chi_{k-1} = (\mu_{k-1}, \mu_{k-1} + \gamma\sqrt{\Sigma_{k-1}}, \mu_{k-1} - \gamma\sqrt{\Sigma_{k-1}})$
3:	for $i = 0, ..., 2n$:
	$\bar{\chi}_k^{[i]} = \chi_{k-1} + GP_\mu(u_{k-1}, \chi_{k-1}^{[i]}, D_g)$
	$Q_k = GP_\Sigma(u_{k-1}, \mu_{k-1}, D_g)$
4:	$\tilde{\mu}_k = \sum_{i=0}^{2n} w_m^{[i]} \bar{\chi}_k^{[i]}$
5:	$\tilde{\Sigma}_k = \sum_{i=0}^{2n} w_c^{[i]} (\bar{\chi}_k^{[i]} - \hat{\mu}_{k-1})(\bar{\chi}_k^{[i]} - \hat{\mu}_k)^T + Q_k$
6:	$\hat{\mathcal{Z}}_k^{[i]} = \bar{\chi}_k^{[i]}$
7:	$\hat{z}_k = \sum_{i=0}^{2n} w_m^{[i]} \hat{\mathcal{Z}}_k^{[i]}$
8:	$S_k = \sum_{i=0}^{2n} w_c^{[i]} (\hat{\mathcal{Z}}_k^{[i]} - \hat{z}_k)(\hat{\mathcal{Z}}_k^{[i]} - \hat{z}_k)^T + R_k$
9:	$\hat{\Sigma}_k^{x,z} = \sum_{i=0}^{2n} w_c^{[i]} (\bar{\chi}_k^{[i]} - \tilde{\mu}_k)(\hat{\mathcal{Z}}_k^{[i]} - \hat{z}_k)^T$
10:	$K_k = \hat{\Sigma}_k^{x,z} S_k^{-1}$
11:	$\mu_k = \tilde{\mu}_k + K_k(z_k - \hat{z}_k)$
12:	$\Sigma_k = \tilde{\Sigma}_k - K_k S_k K_k^T$
13:	Outputs μ_k, Σ_k

Table 2. Applied GP-UKF Algorithm

9 the cross-covariance of prediction and observation is determined. Depending on both, the Kalman gain K_k is generated in line 10 and based on this, the new mean and covariance in time step k are defined in line 11 and 12, respectively.

3.3. Multiple Model Approach

Selecting one model to predict the RUL of bearing faults ignores the uncertainty due to the stochastic nature of the degradation process. To take the uncertainties into account, more prognostic models (PMs) are needed to improve the prediction. A Bayesian formalism is used to combine the knowledge of a set $\mathcal{M}$ of PMs, by weighting each model to be the correct one, as demanded by Li et al. in (Li & Jilkov, 2003). Therefore, the Interacting Multiple Model (IMM) estimator, which bases on the Autonomous Multiple Model (AMM), is applied. In contrary to latter, the IMM belongs to the group of cooperating multiple model approaches, since every model $m^i \in \mathcal{M}$ interacts with the other. Thus, the multiple model filters are reinitialised during every time step k according to information of the previous time step.

Consider Equations (12) and (13) with one PM. Then the extension to the multiple model approach follows as

$$
\begin{aligned}
\mathbf{x}_k &= G(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, m^i) + \epsilon_k \\
\mathbf{z}_k &= H(\mathbf{x}_k, m^i) + \delta_k
\end{aligned} \tag{18}
$$

according to (Schaab, 2011).

The first steps of the IMM algorithm consist of a reinitialising step with a calculation of a mode probability of every ith model

$$
\begin{aligned}
\mu_{k|k-1}^{(i)} &= P(m_k^{(i)}|y_{1:k}) \text{ for } i = 1, ..., n_z \\
&= \sum_{j=1}^{n_z} h_{ij} \mu_{k-1}^{(j)}
\end{aligned} \tag{19}
$$

with the entries $h_{ij} = P\{m_k = m^j | m_{k-1} = m^i\}$ of the transition matrix $\mathbf{H}$ according to Markov. The application of

4

the transition matrix $\mathbf{H}$ prevents the prognostic approach of insisting on one model, as it offers the possibility of a change from model i to j during every time step. Therefore, the transition matrix $\mathbf{H}$ describes a Markov chain, whereas $\mathbf{H}$ is assumed to be time invariant.

By using the information of the previous time step and $\mu_{k|k-1}^{(i)}$, a weighting factor according to

$$
\begin{aligned}
\mu_{k-1}^{j|i} &= P(m_{k-1}^{(i)}|y_{1:k-1}, m_k^{(i)}) \\
&= \frac{h_{j|i}\mu_{k-1}^{(j)}}{\mu_{k|k-1}^{(i)}}
\end{aligned}
\tag{20}
$$

is calculated. Herewith an individual reinitialising value for every filter

$$
\begin{aligned}
\bar{x}_{k-1|k-1}^{(i)} &= E[x_{k-1}|y_{1:k-1}, m_k^{(i)}] \\
&= \sum_{j=1}^{n_z} \hat{x}_{k-1|k-1}^{(j)} \mu_{k-1}^{j|i}
\end{aligned}
\tag{21}
$$

and similarly a covariance $\bar{P}_{k-1|k-1}^{(i)}$ (s. Table 3) is computed. After the reinitialising of the models, these initial values are provided to the applied filters, which are in case of the appropriated prognostic approach GP-UKFs with different PMs.

According to the likelihood $L_k^{(i)}$, which depends on the residuum $e_k^{(i)} = z_k^{(i)} - \hat{z}_k^{(i)}$ and indicates the probability that i is the correct model, the state probability of model i is calculated as

$$
\mu_k^{(i)} = \frac{\mu_{k|k-1}^{(i)} L_k^{(i)}}{\sum_{j=1}^{n_z} \mu_{k|k-1}^{(j)} L_k^{(j)}}.
\tag{22}
$$

Finally, the results of the single i filters are fused to the state $\hat{x}_{k|k}$ and covariance estimation $P_{k|k}$ by means of the minimum mean squared error (MMSE) weighted with the state probability of Equation (22)

$$
\begin{aligned}
\hat{x}_{k|k} &= \sum_{i=1}^{n_z} \hat{x}_{k|k}^{(i)} \mu_k^i \\
P_{k|k} &= \sum_{i=1}^{n_z} [P_{k|k}^{(i)} + (\hat{x}_{k|k} - \hat{x}_{k|k}^{(i)})(\hat{x}_{k|k} - \hat{x}_{k|k}^{(i)})^T] \mu_k^{(i)}.
\end{aligned}
\tag{23}
$$

The entire algorithm is depicted in Table 3.

4. CASE STUDY

The previously defined prognostic concepts are tested and evaluated for the case of a degrading rolling-element bearing according to the DM of Section 2. Since the training data of the GP are computer-generated, the appropriated vibration model (VM) is defined and described. Afterwards, the results and problems of the GP-UKF approach and the IMM prognostic approach are presented.

1:	Inputs $\mu_{k-1}^{(i)}, \hat{x}_{k-1|k-1}, P_{k-1|k-1}$
2:	$\mu_{k|k-1}^{(i)} = \sum_{j=1}^{n_z} h_{ij}\mu_{k-1}^{(j)}$, for $i = 1, ..., n_z$
3:	$\mu_{k-1}^{j|i} = \frac{h_{j|i}\mu_{k-1}^{(j)}}{\mu_{k|k-1}^{(i)}}$
4:	$\bar{x}_{k-1|k-1}^{(i)} = \sum_{j=1}^{n_z} \hat{x}_{k-1|k-1}^{(j)}\mu_{k-1}^{j|i}$
5:	$\bar{P}_{k-1|k-1}^{(i)} = \sum_{j=1}^{n_z} [P_{k-1|k-1}^{(j)} + (\bar{x}_{k-1|k-1}^{(i)} - \hat{x}_{k-1|k-1}^{(j)}) (\bar{x}_{k-1|k-1}^{(i)} - \hat{x}_{k-1|k-1}^{(j)})^T]\mu_{k-1}^{j|i}$
6:	s. Table 2, Inputs: $\bar{x}_{k-1|k-1}^{(i)}, \bar{P}_{k-1|k-1}^{(i)}, y_k, R^{(i)}$ Outputs: $e_k^{(i)} = z_k^{(i)} - \hat{z}_k^{(i)}, \hat{x}_{k|k}^{(i)}, P_{k|k}^{(i)}$
7:	$L_k^{(i)} = P(e_k^{(i)}|m_k^{(i)}, y_{1:k}) = \mathcal{N}(e_k^{(i)}; 0, S_k^{(i)})$
8:	$\mu_k^{(i)} = \frac{\mu_{k|k-1}^{(i)} L_k^{(i)}}{\sum_{j=1}^{n_z} \mu_{k|k-1}^{(j)} L_k^{(j)}}$
9:	$\hat{x}_{k|k} = \sum_{i=1}^{n_z} \hat{x}_{k|k}^{(i)}\mu_k^i$
10:	$P_{k|k} = \sum_{i=1}^{n_z} [P_{k|k}^{(i)} + (\hat{x}_{k|k} - \hat{x}_{k|k}^{(i)})(\hat{x}_{k|k} - \hat{x}_{k|k}^{(i)})^T]\mu_k^{(i)}$
11:	Outputs $\mu_k^{(i)}, \hat{x}_{k|k}, P_{k|k}$

Table 3. IMM Algorithm

4.1. Simulation of Structure-borne Noise

The aim of this subsection is to generate a vibration signal of a faulty bearing, as it could be measured in reality. Therefore, a combination of a later on set up VM and a DM is required, since the latter creates a monotonically rising acceleration amplitude of impulses, as described in Section 2, which are modulated by the VM. By this means, a vibration signal is generated and can be evaluated in frequency range to detect the state of degradation.

The impulses in the area of the bearing can be measured by an acoustic emission sensor. Antoni et al. determined in (Antoni, 2007) that this measured vibration signal consists of several modulations of the initial impulses and can be summarised in a VM in time domain as

$$
x(t) = \sum_{i=-\infty}^{+\infty} h(t - iT - \tau_i)\, q(iT)\, A_i + n(t).
\tag{24}
$$

Here, τ_i and A_i represent the uncertainties of the measured signal in the arrival time and the amplitude of the ith impact, respectively, as e.g. the penetration of the rolling-element into the pitting of an inner-race is a stochastic process. The transmission behaviour of the surrounding machine parts up to the acoustic emission sensor is considered by the impulse response function $h(t)$ with the inter-arrival time iT of two consecutive impulses. The amplitude modulation $q(t)$ is caused by the cage frequency and $n(t)$ represents the additive background noise.

The applied VM is based on Equation (24). For a more realistic signal, it is assumed that there is another additive impulse sequence caused by other mechanisms in the LRU. For example a rotating systems with a rotating-frequency is also measured by the acoustic emission sensor and its amplitudes outclass those of the fault. The resulting vibration signal $x(t)$ in time range and its power spectral density (PSD) is depicted
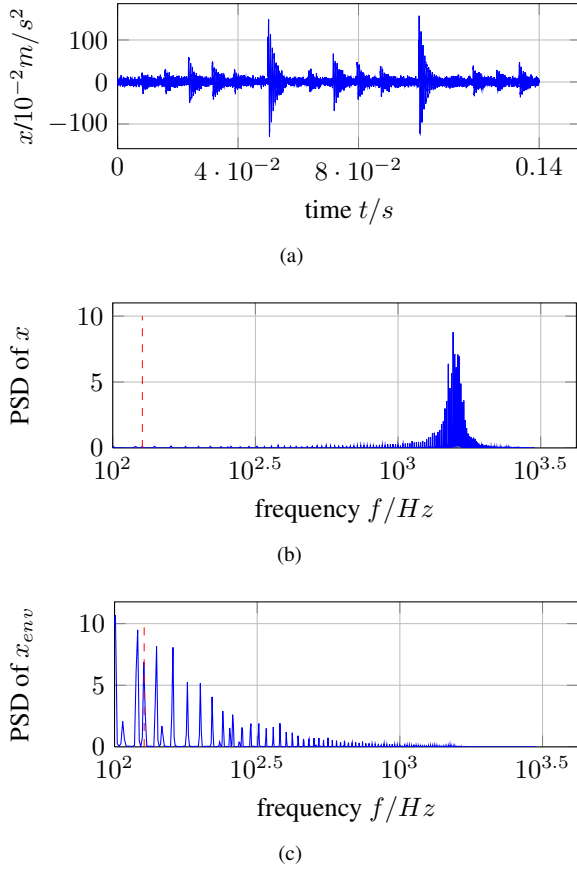
Figure 2. (a) vibration signal $x(t)$ in time range, (b) PSD of $x(t)$ with the marked fault frequency $f_f = 127\,Hz$, (c) PSD of the envelope of $x(t)$
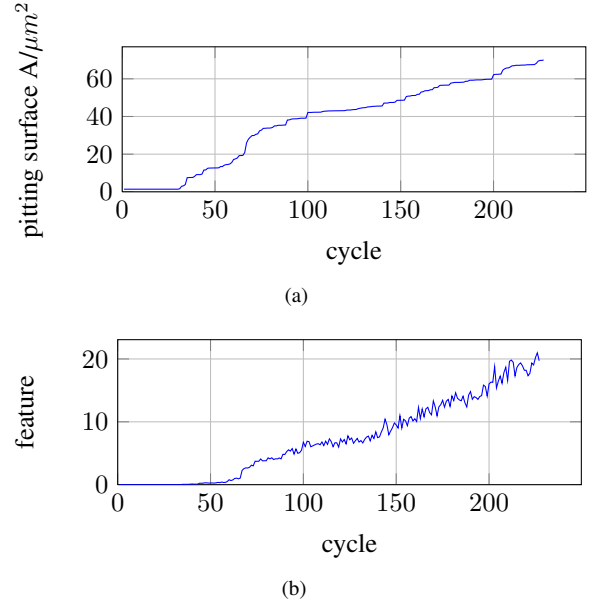

Figure 3. simulated degradation of a rolling-element bearing (a) real degradation as the result of Equation (1), (b) measured normed degradation of sampling the PSD generated by the previously set up VM at the fault frequency

in Figures 2a and 2b, respectively. In Figure 2b, there is a mark at the fault frequency $f_f = 127\,Hz$, as a fault was assumed to be localised at the inner-race.

The vibration signal in time range is dominated by the background noise and the impact sequence of other mechanisms with a frequency of $f_o = 20\,Hz$. In Figure 2b the assumed system behaviour of a second-order lag element with an eigenfrequency of $f_{SB} \approx 1600\,Hz$ representing the path between the bearing and the sensor is clearly visible in contrast to the impulses caused by the fault, which are almost overshadowed by the background noise.

By the application of an envelope x_{env} of the original vibration signal, the influence of the system behaviour is reduced, as depicted in Figure 2c. Besides the impulse sequence and its modes, the amplitude of the fault frequency can be scanned. The amplitude of the PSD at the fault frequency is related to the amplitude given by the DM and therefore is an appropriate feature for the prognostic process, as it determines the current state of degradation. In addition the sidebands caused by the cage modulation $q(t)$ at the frequency $f = f_f \pm f_c$ with an expected cage frequency $f_c = 20\,Hz$ get visible.

The comparison of the real degradation of the applied DM in Equation (1) and the feature r_{feat} is depicted in Figure 3, whereas the applied loading is given in Figure 1b. The measured degradation is quite noisy due to the background noise that dominates the PSD of the vibration signal in lower frequency range. It indicates a different course compared to the real degradation due to the frequency analysis, but as it also denotes the monotonically rising character, the measured amplitude directly correlates with the pitting surface in Figure 3a, i.e. the current degradation.

The prognostic range is set within the normed feature boundary $r_{feat} \in [0.001, 1]$, which is related to a pitting surface range of $A \in [5\mu m^2, 70\mu m^2]$. So this measured degradation course is applied for training and testing the GP-UKF prognostic concepts.

4.2. Applied Performance Metrics

To analyse the prognostic performance, several performance metrics have to be applied. Those metrics can be one single analytical characteristic for the entire prediction or a graphical depiction of every prediction step. Selected performance metrics are summarised by Saxena et al. in (Saxena et al., 2008), whereas a few of these are used for evaluation of the appropriated prognostic concept in the following sections.

Some notations of the metrics domain are given in the following glossary:

UUT Unit under test
EOL End of life

6

EOP	End of prediction - predicted failure feature crossed threshold
i	Time index
l	Number of UUT index
P	Time index of the first prediction
L	Total number of predictions
λ	Normed time range of the entire prediction
$r^l(i)$	Estimation of RUL at time step t_i for the lth UUT
$r^l_*(i)$	Real RUL at time step t_i

In the following subsections the applied performance metrics are defined.

4.2.1 Error

The error $\Delta^l(i)$ indicates the difference between the predicted RUL and the true RUL in time step i

$$\Delta^l(i) = r^l_*(i) - r^l(i) \qquad (25)$$

The error is one of the basic accuracy indicators and is, therefore, included directly or indirectly in most of the selected metrics.

4.2.2 Average Bias

By averaging the error w.r.t. the entire prediction range, the average bias AB of lth UUT is defined as

$$AB_l = \frac{\sum_{i=P}^{EOP} \Delta^l(i)}{EOP_l - P_l + 1} \qquad (26)$$

Thus, the perfect score of AB_l is zero.

4.2.3 Mean absolute percentage error

The Mean Absolute Percentage Error (MAPE) can be written as

$$MAPE(i) = \frac{1}{L} \sum_{i=1}^{L} \left| \frac{100\Delta^l(i)}{r^l_*(i)} \right|. \qquad (27)$$

As it contains the error w.r.t. the actual RUL, derivations in early states of prediction are not as weighted as those near the EOP.

4.2.4 Mean squared error

One most commonly used metric is the Mean Squared Error MSE, since it averages the squared error w.r.t. the number of predictions L

$$MSE(i) = \frac{1}{L} \sum_{i=1}^{L} \Delta^l(i)^2 \qquad (28)$$

An advantage in comparison to the average bias is that the MSE considers both negative and positive errors, as the average bias decreases at the appearance of positive and negative derivations within one prediction.

4.2.5 Prognostic horizon

The Prognostic horizon PH describes the difference between the EOP and the current time step i

$$PH(i) = EOP - i, \qquad (29)$$

whereas the PH can be dictated to fulfill certain specifications. Those are e.g. to remain within a given constant error bound depending on an accuracy value α, i.e.

$$[1 - \alpha] \cdot r^l_* \leq r^l(t) \leq [1 + \alpha] \cdot r^l_*, \qquad (30)$$

comparable to the metrics in the following last subsection. Throughout the whole expectations the accuracy value is $\alpha = 0.05$.

4.2.6 α - λ Performance

Similarly to the PH, the α - λ Performance describes the time span, when the predicted RUL remains within a given error bound. In comparison to the PH, the bound decreases linearly according to

$$[1 - \alpha] \cdot r^l_*(t) \leq r^l(t) \leq [1 + \alpha] \cdot r^l_*(t). \qquad (31)$$

Like the MAPE of Equation (27), this metric favours early predictions at $\lambda \approx 0$ and tightens the demands for predictions near EOP ($\lambda \approx 1$). Throughout the whole expectations the accuracy value α of the α - λ accuracy is $\alpha = 0.20$.

4.3. Prognostic Results of GP-UKF Approach

The aim of this section is to evaluate the GP-UKF approach. In Figure 4, four different test trials are depicted, whereas all courses were generated by the DM in Equation (1). However, the difference especially between trial 1 and 4 in terms of RUL at the beginning of the observation is obvious. Therefore, trial 2 has a similar course to trial 3 with nearly the same life cycle.

In Figure 4b, the corresponding features scanned at the fault frequency $f_f = 127\,Hz$ in the PSD are given, which show a slightly different character in comparison to the real degradation. The noise increases at higher degradation and is filtered by a low-pass filter with regard to the GP training. In Figure 5, the results of the GP training using trial 3 is shown, whereas the estimated training and the real degradation is overlaid by the estimated test degradation.

The results of both the state estimation and the prediction of data set 2 and 3 using the UKF are presented in Figure 6. To compare the state estimation performance, the real feature course is also plotted. The first prediction started at cycle 2, whereas the following predictions began every 10 cycles later. The state estimation of data set 2 matches the real course recognizably. But also the predictions show a high accuracy with an initial error $\Delta^{(2)}(i = 2) \approx 2$ cycles. In sum, all predictions represent the behaviour of the damage course of trial 2. When the GP-UKF is trained and tested with data set 3, the prediction performance becomes slightly worse, as the error
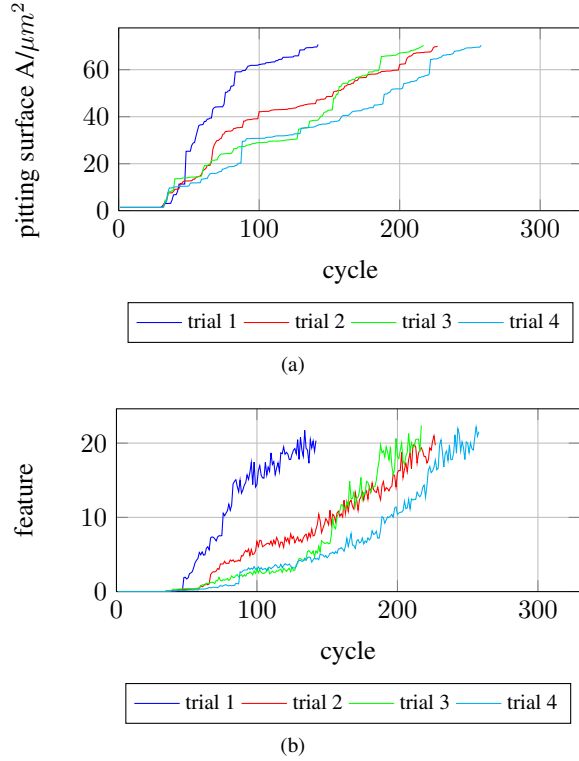
Figure 4. four applied data sets, (a) real degradation, (b) feature

of the forecast RUL of early predictions is about 20 cycles lower than the real RUL. However, later predictions ($\approx$ 6th) match the real degradation with an accuracy allowing the prediction of the RUL.

The same aforementioned behaviour is depicted in several performance metrics, summarised in Table 4. The two similar metrics PH and α-λ accuracy are given as fractions of the normed prediction range λ to indicate the time range, when the predictions fulfill the specifications until the EOP. Additionally, the RUL is also normed to allow comparison of the four trials with different RUL.

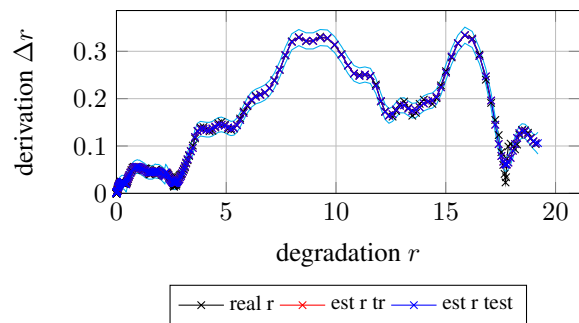In general, the predictions of all four trials show a high ac-
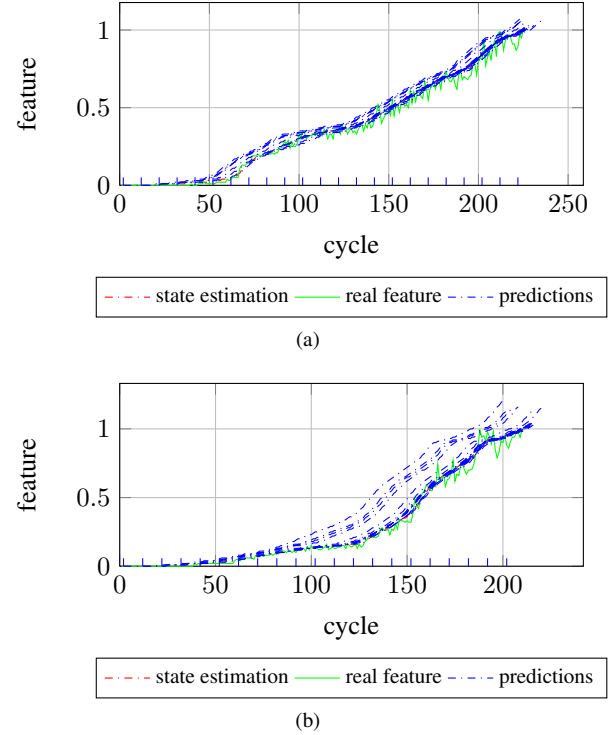

Figure 5. training of the GP with data set 3

Figure 6. the state estimation and prediction of (a) data set 2 and (b) data set 3 in comparison to the real degradation course

Performance Metric	AB	MSE	MAPE	α - λ	PH
Data Set 1	−0.21	21.79	19.52	1	0.86
Data Set 2	0	10.87	13.11	1	1
Data Set 3	−3.48	78.71	24.55	1	0.81
Data Set 4	−2.28	56.44	23.86	1	0.84

Table 4. performance metrics of the four trials tested with themselves

curacy, since every trial remains within the given α-λ error bound during the entire prediction range and also satisfies the tighter specification of the PH after 20% of the normed prediction range λ, as displayed in Figure 7. Therefore, every trial converges to the actual RUL with only slight deviations at 0.3 λ. Additionally, all predictions indicate a rather conservative character, since the AB of all trials is mainly negative. In sum, the selected metrics correspond with the graphical results of Figure 6. The appropriated GP-UKF prognostic concept offers a high accuracy for long turn prediction of a rolling-element bearing, in case of the degradation following the model the filter is trained with.

4.4. Generalisation of the Prognostic Approach

Now the prognostic results of a GP-UKF that is tested with a degradation course, which differs from the training set, are discussed as it occurs in real applications. Figure 8a and 8b
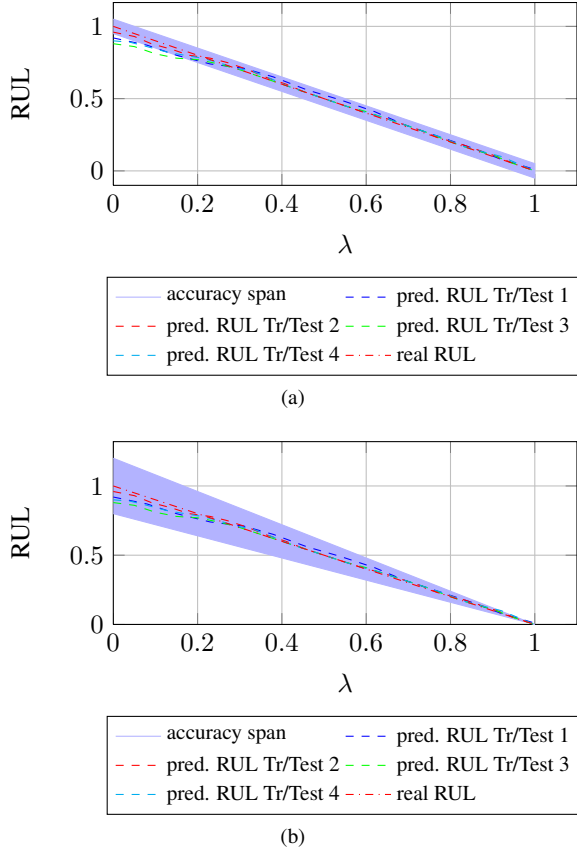
Figure 7. (a) prognostic horizon of the four given trials at $\alpha = 5\%$, (b) α-λ accuracy at $\alpha = 20\%$

show the results of the state estimation and prediction of trial 1 and 4, respectively, when the prognostic model is trained with the data of trial 2. Additionally, the real degradation is plotted. The state estimation of both sets is satisfactory, since there are only slight deviations over the whole prognostic range. The predictions generally indicate the course of the training data set 2 with a progressive degradation at the beginning and a flat degradation rate at the end of the life cycle, whereas both characteristics differ from the tested sets. Therefore, the forecast degradations are not as convincing as in comparison to Section 4.3, according to expectations.

Performance Metric	AB	MSE	MAPE	α - λ	PH
Data Tr 2 Test 1	44.64	3522.93	242.65	0	0.07
Data Tr 2 Test 4	8.64	544.8	116.60	0	0.04
Data Tr 3 Test 1	22.79	1514.79	151.34	0	0.071
Data Tr 3 Test 4	−28.12	1172.28	149.59	0	0.28

Table 5. performance metrics of different test data sets, when GP is trained with data set 2

The performance is analysed again by means of the metrics in Table 5. Additionally, the prognosis performance in case of training data set 3 is depicted. Compared to the results of
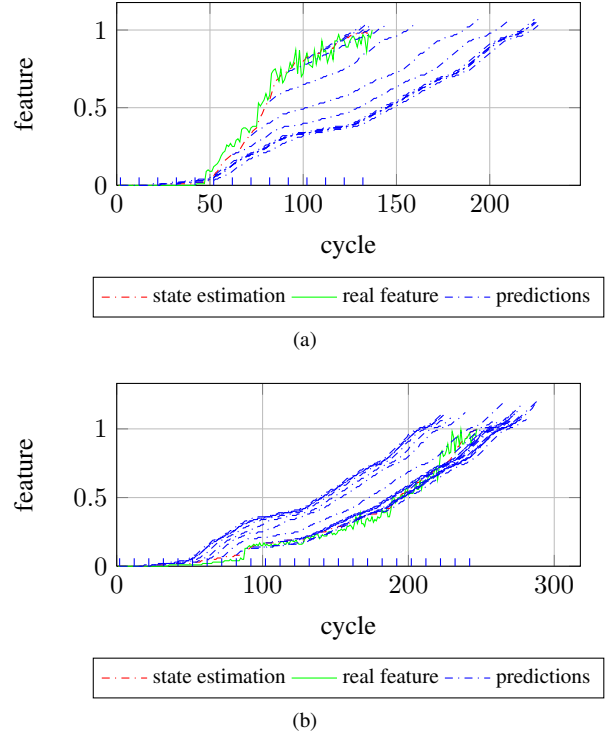

Figure 8. the state estimation and prediction of (a) data set 1 and (b) data set 4 with the GP trained with data set 2

the previous section, all metrics increased considerably due to both aforementioned causes of different degradation behaviour of test and training sets and slightly inaccurate state estimation. Hereby the predictions from cycle 60 to 90 (corresponds to $\lambda \approx 0.5$) of test data set 1 and at the end of data set 4 are not able to predict the real damage course. These deviations are also depicted in the graphical metrics in Figure 9, whereas neither the specifications of α-λ accuracy nor of the PH are fulfilled satisfactory.

In comparison to training data set 3, the predictions with a dynamic GP trained with data set 2 show beneficial prognostic results in case of test data set 4 according to Table 5, whereas w.r.t. test data set 1 the third data set is advantageous. Therefore, a combination of both training data sets through a multiple model approach is assumed to exhibit benefits in comparison to the GP-UKF with only one set of training data.

4.5. Improvements by means of Multiple Model Approach

The prediction performance of an IMM approach with two different GP-UKFs is discussed in this section. The two models are the GP-UKF trained with trial 2 and 3, respectively, as the combination of both is supposed to indicate the benefits of the IMM approach.

In Figure 10, the state estimation and prediction results of test set 1 and 4 are depicted. The real degradation of both test cases is estimated very accurately with only a slight devi-
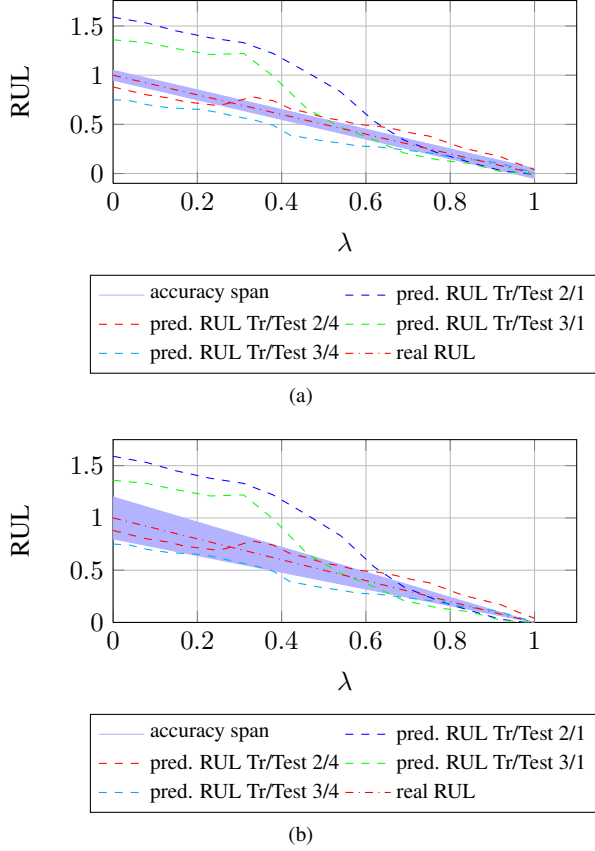
(a)

(b)

Figure 9. (a) prognostic horizon of the two tested data sets 3 and 4 with training data set 2 ($\alpha = 5\%$), (b) α-λ accuracy at $\alpha = 20\%$
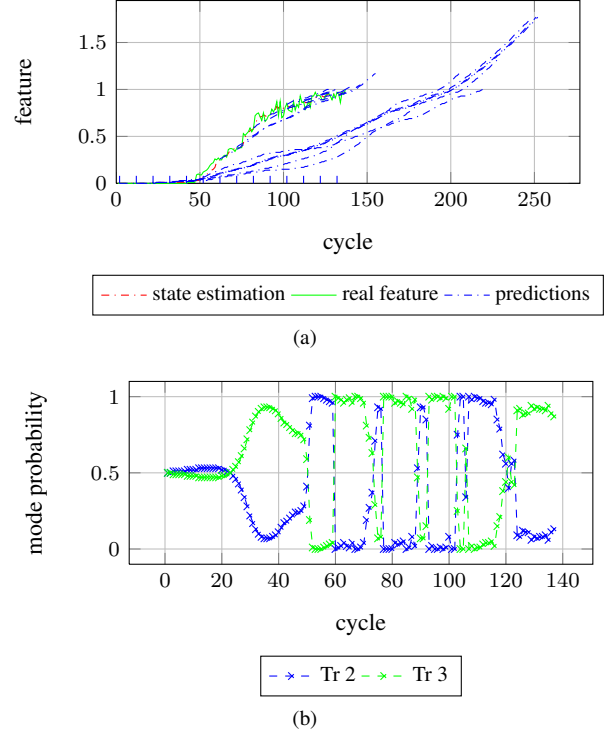

(a)

(b)

Figure 10. prediction results of training sets 2 and 3 tested with (a) trial 1 and (b) mode probability of training sets 2 and 3 during prediction of trial 1

Performance Metric	AB	MSE	MAPE	α - λ	PH
Data Tr 2,3 Test 1	25	1841.14	160.54	0	0.07
Data Tr 2,3 Test 4	-7	895.32	132.99	0.04	0.12

Table 6. performance metrics of IMM with training set 2 and 3 testing set 1 and 4

ation in Figure 10a between cycles 50 and 60. At first sight, the prediction results show a similar performance comparable to the GP-UKFs of Section 4.4. Especially the first predictions determine the RUL rather inaccurately, since the error $|\Delta(i = 2)|$ amounts in both cases about 60 cycles. As described in the previous section, both test sets differ from the applied training sets, which causes poor reproduction of the damage course. The mode probability of trial 1 in Figure 10b indicates the same reason, since especially at the beginning of the prediction neither of both training sets replicate the real degradation to satisfaction and therefore the model probabilities $\mu_k^{(1)}$ and $\mu_k^{(2)}$ are about 0.5. In comparison to Figure 8a, the later predictions of the MM approach in Figure 10a indicate a more accurate RUL estimation due to the domination of training set 3 that reproduces the test set 1 more precisely. In Table 6, the performance metrics of the IMM approach are shown. Due to the inaccurate forecasts at the beginning of the prognosis, the metrics values are comparable to Table 5.

To identify the advantages of the IMM approach, the net diagrams in Figure 12 give an overview of the collected results. They show the metrics normalised to the major value within a test trial. Since most metrics describe an inaccurate RUL

estimation with large values, the time range, when the predictions fulfill the specifications of the α-λ and the PH error bound, is exchanged for the time range those specifications are not met, i.e. $PH_{net} = 1 - PH$.

The diagrams show the advantage of the IMM approach, as all measured performance metrics lie between the metrics of the GP-UKFs trained with one data set. That means this approach increases the robustness of a bearing's RUL prediction, as it provides the possibility of discarding inaccurate models depending on the mode probability.

However, since the MM approach consists of the two PMs, which both differ from the test sets, the prognostic performance is still hardly able to outperform the results of both GP-UKFs with reference to Table 5 and Table 6. Especially the α - λ accuracy and PH of trial 1 (Figure 12a) of both GP-UKFs indicate the same behaviour and, thus, an IMM approach based on those models is not able to increase this performance.

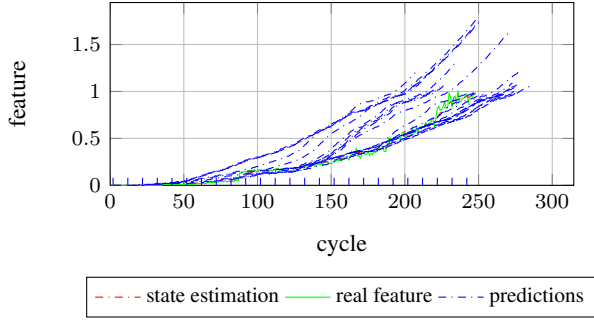The great benefit of the increased robustness is assumed to

Figure 11. prediction results of training trials 2 and 3 tested with trial 4

raise by including more models with different degradation courses. Especially a progressive degradation rate at the beginning of the prediction range in case of forecasting test set 1 is expected to be beneficial concerning the prognosis performance.
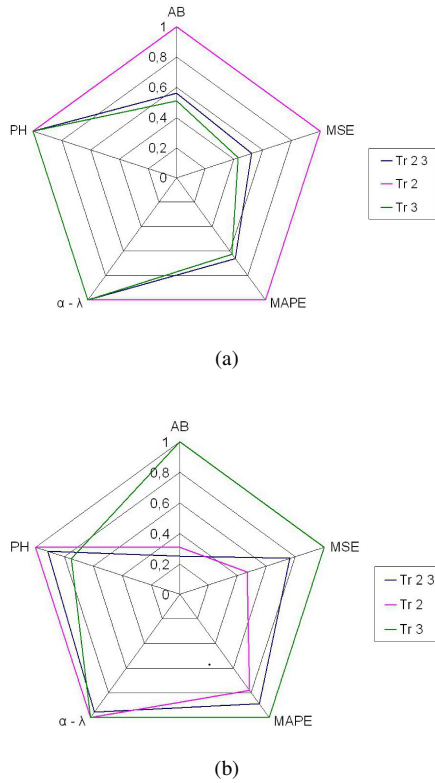

(a)

(b)

Figure 12. comparison of the single performance metrics of (a) test data 1 and (b) test data 4

5. CONCLUSION

Two prognostic concepts based on the GP-UKF approach to predict the RUL of a rolling-element bearing were examined in the context of a case study. The results showed that a dynamic GP in combination with an UKF estimates the RUL

of a bearing very accurately, when the applied training data is equal to the trial data. If the training data differs from the trial data, the GP-UKF is not able to forecast the degradation precisely, but mainly insists on the characteristics of the trained damage course. To solve this problem, an IMM approach based on two different GP-UKF models has been evaluated. It was assumed that the IMM algorithm, restoring several prognostic models, is more likely to forecast a damage course of an unknown trial. The results proved these expectations, since the robustness of the predictions was highly increased by the approach.

By incorporating more prognostic models into the IMM approach, which should mainly differ from the applied GP-UKFs, this approach is expected to even outperform the prognostic results of a single GP-UKF. This will be in the focus of further research.

NOMENCLATURE

Symbols

A	Pitting surface
ΔA_i	Increase of surface during cycle i
Δu_i	Loading difference during cycle i
k_A, k_u	Coefficients of applied degradation model
λ', k'	Shape/scaling parameter of Weibull distribution
μ'	Expected value of Exponential distribution
D	Data set
$\mathbf{x}_n$	Inputs of GP
$\mathbf{X}$	Input matrix
y_n	Outputs of GP
$\mathbf{y}$	Output matrix
ϵ	Noise term
μ	Mean of model
Σ	Covariance of model
$\mathbf{K}$	Kernel matrix
σ_n	Noise term
$k(\mathbf{x}_i, \mathbf{x}_j)$	Kernel function
σ_f	Signal variance of Kernel function
$\mathbf{W}$	Distance measure weighting matrix
GP_μ	Mean of GP
GP_Σ	Covariance of GP
$\mathbf{x}_*$	Test input
θ	Hyperparameters of GP
r_k	Degradation at kth time step
Δr_k	Degradation rate at kth time step
X'	Degradation rate matrix
D_r	Training data set
G	Transition function
Q_k	Process noise
H	Observation function
$\mathbf{z}_k$	Observation at kth time step
R_k	Measure noise

δ_k	Noise Term
$\chi^{[i]}$	Sigma points
α', κ	Scaling parameters of UKF
$\tilde{\mu}, \tilde{\Sigma}$	A priori mean/covariance
w_m, w_c	Weights of mean/covariance
$\hat{\mathcal{Z}}_k^{[i]}$	Observation
K_k	Kalman gain
$\mathcal{M}$	Prognostic model set
m^i	Prognostic model i
$\mu_{k|k-1}^{(i)}$	Mode probability
h_{ij}	Entries of transition matrix $\mathbf{H}$
$\bar{x}_{k-1|k-1}^{(i)}$	Reinitialising state of IMM
$\bar{P}_{k-1|k-1}^{(i)}$	Reinitialising covariance of IMM
$L_k^{(i)}$	Likelihood of model i
$e_k^{(i)}$	Residuum
$\mu_k^{(i)}$	State probability of model i
$x(t)$	Total vibration signal
$h(t)$	Impulse response
τ_i, A_i	Uncertainties of arriving impulse response
$q(t)$	Amplitude modulation
$n(t)$	Background noise
f_f	Fault frequency
f_o	Frequency of other mechanisms
f_{SB}	Eigenfrequency of system behaviour
x_{env}	Envelope of $x(t)$
f_c	Cage frequency
r_{feat}	Degradation of feature
$\Delta^l(i)$	Error of RUL prediction
AB_l	Average Bias of lth UUT
$MAPE(i)$	Mean absolute percentage error
$MSE(i)$	Mean squared error
$PH(i)$	Prognostic horizon
α	Accuracy value

Shortcuts

RUL	Remaining Useful Lifetime
LRU	Line-replaceable Unit
DM	Degradation Model
PM	Prognostic Model
GP	Gaussian Process
UKF	Unscented Kalman Filter
IMM	Interacting Multiple Model
VM	Vibration Model
PSD	Power Spectral Density
Tr	Training

(See also Glossary in Section 4.2)

REFERENCES

Antoni, J. (2007). Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *Journal of Sound and vibration*, *304*(3-5), 497–529.

Choi, Y., & Liu, C. R. (2006a). Rolling contact fatigue life of finish hard machined surfaces - Part 1. Model development. *Wear*, *261*(5-6), 485–491.

Choi, Y., & Liu, C. R. (2006b). Rolling contact fatigue life of finish hard machined surfaces - Part 2. Experimental verification. *Wear*, *261*(5-6), 492–499.

Daigle, M., & Goebel, K. (2010). Model-Based Prognostics under Limited Sensing.

Julier, S. (2002). The scaled unscented transformation. In *American Control Conference, 2002. Proceedings of the 2002* (Vol. 6, pp. 4555–4559).

Ko, J., & Fox, D. (2011). Learning GP-BayesFilters via Gaussian process latent variable models. *Autonomous Robots*, *30*(1), 3–23.

Ko, J., Klein, D., Fox, D., & Haehnel, D. (2007). GP-UKF: Unscented Kalman filters with Gaussian process prediction and observation models. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on* (pp. 1901–1907).

Li, X., & Jilkov, V. (2003). A survey of maneuvering target tracking—Part V: Multiple-model methods. In *Proc. SPIE Conf. on Signal and Data Processing of Small Targets* (pp. 559–581).

Orchard, M. E., & Vachtsevanos, G. J. (2009). A particle-filtering approach for on-line fault diagnosis and failure prognosis. *Transactions of the Institute of Measurement and Control*, *31*(3-4), 221–246.

Orsagh, R., Sheldon, J., & Klenke, C. (2003). Prognostics/diagnostics for gas turbine engine bearings. In *Proceedings of IEEE Aerospace Conference.*

Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian Processes for Machine Learning.

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., et al. (2008). Metrics for evaluating performance of prognostic techniques. In *Prognostics and Health Management, 2008. PHM 2008. International Conference on* (pp. 1–17).

Schaab, J. (2011). *Trusted health assessment of dynamic systems based on hybrid joint estimation* (Als Ms. gedr. ed.). Düsseldorf: VDI-Verl.

Sturm, A. (1986). *Wälzlagerdiagnose an Maschinen und Anlagen*. Köln: TÜV Rheinland.

Yu, W. K., & Harris, T. A. (2001). A New Stress-Based Fatigue Life Model for Ball Bearings. *Tribology Transactions*, *44*(1), 11–18.