

Warped Gaussian processes for predicting the degradation of aerospace structures

Structural Health Monitoring

2023, Vol. 22(4) 2531–2546

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14759217221127252

journals.sagepub.com/home/shm



Simon Pfingstl^{1,2} , Christian Braun¹, Amir Nasrollahi², Fu-Kuo Chang² and Markus Zimmermann¹

Abstract

Gaussian processes (GPs) can be used to predict future states of a system with credible intervals when considering multiple previous trajectories for training. For example, predicting the degradation of mechanical structures is one application in which they have shown their usefulness. In modeling the system output as a GP, the output is presumed to be normally distributed—assuming the predictions to be defined from negative to positive infinity. However, this assumption does not hold in many applications as, for example, crack lengths and damage indices can only assume positive values. Moreover, several degradation trajectories for training are rare in real-world applications, and the current state of a monitored system, which is used to update the prediction, can often be not directly measured. This paper presents an approach that utilizes warped GPs for treating data that is not normally distributed while considering multiple degradation trajectories for training. The approach is successfully applied to two different crack propagation examples: first, an analytically computed pre-cracked infinite plate, and second, two equally manufactured aluminum structures that resemble a lower section of a wing. For the investigated aerospace structures, we use finite element (FE) simulations to generate multiple degradation trajectories for training. To estimate their hidden degradation states, we infer the current crack length from strain measurements by using Bayesian inference. The results show that the approach of warped GPs provides more accurate predictions than using standard ones for non-normally distributed data, as is the case for crack growth problems. The approach enables quick training of warped GPs while considering multiple training trajectories. Additionally, the crack lengths estimated from strain measurements agree well with the visually inspected ones. Ultimately, the presented approach enables estimating the current and future degradation states with credible intervals that can be used to improve maintenance scheduling.

Keywords

Warped Gaussian processes, fatigue crack growth, aerospace structures, structural health monitoring, prognostics and health management

Introduction

Mechanical structures are usually used under changing load conditions and must withstand fatigue loads. Fixed inspection intervals ensure safe operation but are often too conservative, resulting in unnecessary system downtimes and increased operational costs. In order to overcome these drawbacks, new approaches consider variable inspection intervals. These include structural health monitoring (SHM) and prognostics and health management (PHM). While SHM monitors the current state of a system by processing sensor data, PHM extends this by predicting the state of the system and thus forecasting future damages.¹ This leads to benefits

such as improved operational reliability and reduced operating costs.²

¹TUM School of Engineering and Design, Laboratory for Product Development and Lightweight Design, Technical University of Munich, Garching, Germany

²Department of Aeronautics and Astronautics, Structures and Composites Laboratory, Stanford University, Stanford, CA, USA

Corresponding author:

Simon Pfingstl, TUM School of Engineering and Design, Laboratory for Product Development and Lightweight Design, Technical University of Munich, Boltzmannstrasse 15, Garching 85748, Germany.
Email: simon.pfingstl@tum.de

Several machine learning methods can be used to predict the degradation of structures. Since recurrent neural networks (NNs) are explicitly suitable for dealing with time-series data, they are widely used for prognostics.³ For example, Do et al.⁴ utilizes recurrent NNs to predict the crack growth in structures and Malhi et al.⁵ to estimate the degradation of bearings. Moreover, Heimes⁶ and Peng et al.⁷ used them to forecast the time to failure of turbines. However, one drawback of recurrent NNs is that they do not provide any information on their predictions' uncertainties by default, which is especially important for fatigue problems.

Support vector machines are another method that researchers frequently use. Many authors^{8–13} utilized support vector machines to predict the degradation of bearings. Furthermore, Fumeo et al.¹⁴ used them to estimate the remaining useful life of railway transportation systems. As recurrent NNs, support vector machines do not provide estimates about the uncertainty associated with the prediction by default. Yet, the fatigue life of structures, especially at the beginning of usage, is highly uncertain. Using a prediction method that does not account for uncertainties would suggest a false certainty about the fatigue life.

By contrast, Gaussian process regression (GPR) is a data-driven approach for PHM¹⁵ which has the advantage of modeling the uncertainty of the predicted states. This is particularly important in fatigue mechanics since fatigue-induced damage is highly uncertain.¹⁶ Even though there exist methods for leveraging recurrent NNs and support vector machines to predict their outputs with credible intervals (e.g. mixture density networks or relevance vector machines), Gaussian processes (GPs) describe distributions of trajectories rather than points and thus enable predicting the entire degradation process instead of merely the fatigue life. This allows us to estimate the probability of failure for each future time point. A GP is defined by a mean and covariance function^{17,18} that greatly influences prediction accuracy. In order to improve the selection of these functions, Pfingstl and Zimmermann¹⁹ introduced an approach to infer them from previously collected trajectories. Yet, a major disadvantage of GPs is that they presume the data to be normally distributed, thus assuming the predictions to be defined from negative to positive infinity.

Contribution

In many applications, presuming normally distributed data is not valid as, for example, crack lengths and damage indices can only assume positive values. Moreover, training data such as degradation trajectories of aircraft structures is often rare in real-world

applications, and the state of the system, for example, the crack length, is hidden. In order to improve and utilize GPs for PHM problems, we (1) transform the data before training the GP, (2) incorporate prior knowledge by using analytical equations and results from finite element (FE) analyses, and (3) infer the hidden state of the system from sensor data by applying Bayesian inference. The proposed approach is used to predict the crack growth in an infinite plate and in two similar aerospace structures resembling a lower part of an aircraft wing. Even though it is currently common practice in the aviation industry to repair a crack as soon as it is detected, in the future, with a fully operating SHM system, this might become too expensive. Predicting the crack growth makes it possible to wait until either multiple parts require repair or structural integrity can no longer be assured. Current requirements are primarily based on NDI/visual inspections where the aircraft is already grounded, and the location can be accessed rather easily once the engineer has found the crack. This might change when reliable SHM systems are demonstrated. The proposed approach is not limited to predicting crack lengths. Predicting other damage indices is also possible.

Definition of terms

In the course of this work, we distinguish between *previous* and *current* data. In this paper, previous data consists of several trajectories describing the degradation of a mechanical system. These trajectories can be gathered by executing simulations or experiments. By contrast, current data represents the data of the currently monitored mechanical system for which degradation is to be predicted. Furthermore, we use the term *degradation* to describe the fatigue process in mechanical systems. In this paper, we use this term particularly for propagating fatigue cracks.

Gaussian processes

Over the last decade, many researchers have frequently been using GPR for nonlinear regression, see Reference,^{20–32} as the model provides not only the predicted function value itself but also its credible interval. A GP is fully defined by its mean and covariance functions, $m_{y, \theta}(x)$ and $k_{y, \theta}(x, x')$, with

$$y \sim \mathcal{GP}(m_{y, \theta}(x), k_{y, \theta}(x, x')), \quad (1)$$

where usually predefined mean and covariance functions with some free parameters θ are used. These parameters are optimized by minimizing the negative log-likelihood (NLL) in order to fit the model to the training data.

Basis functions

In order to use GPs for PHM, an approach based on several previously collected trajectories was recently published by Pfingstl and Zimmermann.¹⁹ By using p linearly independent basis functions $\Phi_y(x) = [\phi_y^{(1)}, \dots, \phi_y^{(p)}]^\top$ with $k \in 1, \dots, p$ to fit all m previously collected trajectories $\{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ with $j \in 1, \dots, m$

$$f_y^{(j)}(\mathbf{x}^{(j)}) = \mathbf{y}^{(j)} = \Phi_y(\mathbf{x}^{(j)})^\top \boldsymbol{\beta}^{(j)}, \quad (2)$$

the GP can be expressed as

$$m_y(x) = \Phi_y(x)^\top \hat{\boldsymbol{\mu}}_\beta \quad (3)$$

$$k_y(x, x') = \Phi_y(x)^\top \hat{\boldsymbol{\Sigma}}_\beta \Phi_y(x') \quad (4)$$

$$f_y(x) = y \sim \mathcal{GP}(m_y(x), k_y(x, x')), \quad (5)$$

where $\hat{\boldsymbol{\mu}}_\beta$ and $\hat{\boldsymbol{\Sigma}}_\beta$ are the sample mean vector and the sample covariance matrix of the fitting coefficients $\boldsymbol{\beta}^{(j)}$, respectively. The fitting coefficients $\boldsymbol{\beta}^{(j)}$ can be determined by applying linear regression. Therefore, no free parameters $\boldsymbol{\theta}$ of the mean and covariance function have to be optimized by minimizing the NLL. This turns the, in general, non-convex optimization problem into a convex one and thus significantly reduces the training time, see Reference¹⁹ for more details.

The approach first tries to describe all possible functions by a GP. Second, after determining the mean and covariance function, the GP can be updated by computing the conditional distribution based on currently measured data, that is, data describing the current state of the monitored system. When the mean function $m_y(x)$ and the covariance function $k_y(x, x')$ are defined, and we have measured the data $\mathbf{y}^{(+)}$ (e.g., crack lengths) at the locations $\mathbf{x}^{(+)}$ (e.g., numbers of cycles), we can compute the conditional distribution (e.g., over the crack length) at the location x (e.g., a future number of cycles) by

$$f(x) | \mathbf{x}^{(+)}, \mathbf{y}^{(+)} \sim \mathcal{GP}(m_{y|+}(x), k_{y|+}(x, x')), \quad (6)$$

with

$$m_{y|+}(x) = m_y(x) + k_y(x, \mathbf{x}^{(+)}) (k_y(\mathbf{x}^{(+)}, \mathbf{x}^{(+)}) + \sigma_y^2 \mathbf{I})^{-1} (\mathbf{y}^{(+)} - m_y(\mathbf{x}^{(+)})) \quad (7)$$

and

$$k_{y|+}(x, x') = k_y(x, x') - k_y(x, \mathbf{x}^{(+)}) (k_y(\mathbf{x}^{(+)}, \mathbf{x}^{(+)}) + \sigma_y^2 \mathbf{I})^{-1} k_y(\mathbf{x}^{(+)}, x')^\top, \quad (8)$$

where σ_y is the observation error. The conditional distribution described by the conditional mean $m_{y|+}(x)$ and variance $k_{y|+}(x, x)$ is then our prediction at x . We can, of course, evaluate the conditional distribution at

multiple locations x , allowing us to rather predict the entire future degradation trajectory than just a single point in future. By splitting the approach into a training and an updating process, the method enables predicting the degradation trajectory rapidly since only the conditional distribution (Equations (7) and (8)) must be evaluated.

Warped GPs

By modeling the system output with a GP, the output is assumed to be normally distributed—presuming it to be defined from negative to positive infinity. This assumption, however, is often not valid as, for example, crack lengths, damage indices, sunlight time, etc., are only defined in the positive domain. For a better treatment of data that is not normally distributed, Snelson et al.³³ introduced warped GPs. The idea is to transform the observed space y onto a latent space z by a so-called warping function $\psi_\theta(y) = z$ which usually has some free parameters θ . The GP is then modeled in the latent space z as

$$\psi_\theta(y) = z \sim \mathcal{GP}(m_{z,\theta}(x), k_{z,\theta}(x, x')), \quad (9)$$

where $m_{z,\theta}(x)$ and $k_{z,\theta}(x, x')$ are the mean and covariance function in the latent space z , respectively. For predicting the function values of the observed space y , the predictions of z have to be transformed by the inverse warping function $\psi_\theta^{-1}(z)$.

Some researchers have already used warped GPs, for example, to predict the power supplies of wind turbines. In contrast to wind speeds, power supplies cannot be assumed to be normally distributed due to the nonlinear correlation between wind speed and power.³⁴ Therefore, the authors of^{35–37} utilized warped GPs to predict the power supplies of wind turbines. They used a sum of *tanh* as their warping function and proved the approach’s usefulness on real data. Moreover, Mateo-Sanchis et al.³⁸ applied warped GPs to oceanic content data. They predicted the oceanic chlorophyll content from multispectral data and concluded that warped GPs outperform standard GPs. Again, a sum of *tanh* as the warping function was used.

However, if the inverse warping function is not available in closed form, which is the case in the before-mentioned papers, additional complexity arises from numerical approximations.³⁹ Therefore, one can use, for example, the Box–Cox transformation⁴⁰

$$\psi_\theta(y) = \begin{cases} \frac{y^{\theta_1} - 1}{\theta_1}, & \text{if } \theta_1 \neq 0 \\ \log y, & \text{if } \theta_1 = 0. \end{cases} \quad (10)$$

as the warping function. For example, Rios and Tobar³⁹ utilized a slightly adjusted type of the Box–

Cox transformation function and showed its effectiveness on real data, enforcing their predicted yearly sunspot numbers to be strictly positive. A similar application of warped GPs was presented by Gonçalves et al.⁴¹ The authors estimated future sunspot numbers and enforced their predictions to be positive by using an integrated softplus function in their warping function.

Even though these studies show that warped GPs are able to deal with non-normally distributed data, researchers tend to use them rarely. This might be because the warping function introduces additional parameters to the modeling task. These parameters must be optimized in addition to the mean and covariance function parameters. Therefore, one entailed problem is the arising computational complexity by determining not only the mean and covariance function parameters but also the ones of the warping function. Another problem is that after introducing a warping function, an optimizer might find, in fact, a different solution. But the result is not necessarily better, as minimizing the NLL is, in general, a non-convex optimization problem. In order to remove the mean and covariance function parameters, we can derive these functions from previously gathered trajectories. In this way, we significantly reduce the computational complexity and additionally integrate prior knowledge into warped GPs.

Approach

In order to exploit the advantages of warped GPs and the use of basis functions for inferring the GP model quickly from previously collected data, both approaches are combined in the following. For doing so, it is assumed that the warped realizations $f_z^{(j)}(x)$ can be approximated by a linear combination of p linearly independent basis functions $\Phi_z(x) = [\phi_z^{(1)}, \dots, \phi_z^{(p)}]^\top$ in the latent space z . The warped trajectories $\{\mathbf{x}^{(j)}, \psi_\theta(\mathbf{y}^{(j)}) = \mathbf{z}^{(j)}\}$ can therefore be represented by a weighted sum of basis functions as

$$f_z^{(j)}(\mathbf{x}^{(j)}) = \mathbf{z}^{(j)} = \psi_\theta(\mathbf{y}^{(j)}) = \Phi_z(\mathbf{x}^{(j)})^\top \beta^{(j)} \quad (11)$$

and the GP in the latent space as

$$m_z(x) = \Phi_z(x)^\top \hat{\mu}_\beta \quad (12)$$

$$k_z(x, x') = \Phi_z(x)^\top \hat{\Sigma}_\beta \Phi_z(x') \quad (13)$$

$$f_z(x) = z \sim \mathcal{GP}(m_z(x), k_z(x, x')). \quad (14)$$

Note that the GP model is independent of any free parameter θ . However, by introducing a warping function, free parameters, which need to be

optimized, are integrated in the formulation. In order to determine the free parameters, the NLL in the observed space p_y is minimized. If the warping function is strictly monotonic, the optimization problem can be stated as

$$\begin{aligned} \theta^* &= \underset{\theta}{\operatorname{argmin}} - \sum_{j=1}^m \log p_y(\mathbf{y}^{(j)}) \\ &= \underset{\theta}{\operatorname{argmin}} - \sum_{j=1}^m \log \left(p_z(\mathbf{z}^{(j)}) \left| \det \left(\operatorname{diag} \frac{d\psi_\theta(\mathbf{z}^{(j)})}{d\mathbf{y}} \right) \right| \right) \\ &= \underset{\theta}{\operatorname{argmin}} - \sum_{j=1}^m \left(\log p_z(\psi_\theta(\mathbf{y}^{(j)})) + \sum_{i=1}^{n_j} \log \left| \frac{d\psi_\theta(\mathbf{y}^{(i,j)})}{d\mathbf{y}} \right| \right) \end{aligned} \quad (15)$$

with the given probability density function in the latent space

$$\begin{aligned} p_z(\mathbf{z}^{(j)}) &= \frac{1}{\sqrt{(2\pi)^{n_j} \det \mathbf{K}_z}} \\ &\exp - \frac{1}{2} (\psi_\theta(\mathbf{y}^{(j)}) - m_z(\mathbf{x}^{(j)}))^\top \mathbf{K}_z^{-1} \\ &(\psi_\theta(\mathbf{y}^{(j)}) - m_z(\mathbf{x}^{(j)})), \end{aligned} \quad (16)$$

where n_j is the number of data points of trajectory j , and

$$\mathbf{K}_z = k_z(\mathbf{x}^{(j)}, \mathbf{x}^{(j)}) + \operatorname{diag} \sigma_z^2(\mathbf{y}^{(j)}). \quad (17)$$

The gradient with respect to the parameter θ_q is given by

$$\begin{aligned} \frac{\partial}{\partial \theta_q} - \log p_y(\mathbf{y}^{(j)}) &= \frac{1}{2} \left(\left(\frac{\partial \psi_\theta(\mathbf{y}^{(j)})}{\partial \theta_q} \right)^\top \mathbf{K}_z^{-1} (\psi_\theta(\mathbf{y}^{(j)}) - m_z(\mathbf{x}^{(j)})) \right. \\ &\quad \left. + (\psi_\theta(\mathbf{y}^{(j)}) - m_z(\mathbf{x}^{(j)}))^\top \mathbf{K}_z^{-1} \left(\frac{\partial \psi_\theta(\mathbf{y}^{(j)})}{\partial \theta_q} \right) \right) \\ &\quad - \sum_{i=1}^{n_j} \left(\left(\frac{d\psi_\theta(\mathbf{y}^{(i,j)})}{d\mathbf{y}} \right)^{-1} \frac{\partial}{\partial \theta_q} \left(\frac{\partial \psi_\theta(\mathbf{y}^{(i,j)})}{d\mathbf{y}} \right) \right). \end{aligned} \quad (18)$$

By transforming the observed data into the latent space, the observation error is warped too. A constant observation error in the observed space σ_y can be approximated in the latent space by

$$\sigma_z(y) = \sigma_y \left| \frac{d\psi_\theta(y)}{d\mathbf{y}} \right|. \quad (19)$$

This approximation is particularly accurate if the observation error σ_y and the second derivative of the warping function at y are rather small. Considering this nonconstant observation error in the latent space, a weighted least squares regression is applied to determine the weights of the basis functions

$$\hat{\boldsymbol{\beta}}^{(j)} = (\boldsymbol{\Phi}_z^{(j)\top} \mathbf{W} \boldsymbol{\Phi}_z^{(j)})^{-1} \boldsymbol{\Phi}_z^{(j)\top} \mathbf{W} \mathbf{z}^{(j)}, \quad (20)$$

with

$$\mathbf{W} = \text{diag} \left(\frac{d\psi_\theta(\mathbf{y}^{(j)})}{d\mathbf{y}} \right)^{-2} \quad (21)$$

and

$$\boldsymbol{\Phi}_z^{(j)} = \boldsymbol{\Phi}_z(\mathbf{x}^{(j)}) = [\phi_z^{(1)}(\mathbf{x}^{(j)}), \dots, \phi_z^{(p)}(\mathbf{x}^{(j)})]. \quad (22)$$

Now, the mean and covariance function can be determined with Equations (20)–(22) (fitting coefficients) and 12–13 (mean and covariance function) within every iteration step of the warping function parameter optimization. The result of the presented approach is an optimized warping function and a determined GP in the latent space. The observation error σ_y can be approximated by

$$\hat{\sigma}_y = \sqrt{\frac{1}{m} \sum_{j=1}^m \frac{1}{n_j - p} \sum_{i=1}^{n_j} \left(\psi_\theta^{-1}(\boldsymbol{\Phi}_z^{(j)} \hat{\boldsymbol{\beta}}^{(j)}) - \mathbf{y}^{(j)} \right)^2}. \quad (23)$$

Afterwards, the conditional distribution in the latent space can be computed by substituting $m_y(x)$, $k_y(x, x')$, $\mathbf{y}^{(+)}$ and $\sigma_y^2 \mathbf{I}$ for $m_z(x)$, $k_z(x, x')$, $\mathbf{z}^{(+)}$ and $\text{diag } \sigma_z^2(\mathbf{y}^{(j)})$ in Equations (7) and (8).

Application to an infinite plate

As a first example, we examine the proposed method on an academic example, a pre-cracked infinite plate, for which the underlying formulas and solutions are known.

Data generation

Considering an infinite plate with a centered crack, the range of the stress intensity factor (SIF) ΔK_I can be expressed by

$$\Delta K_I(a) = \Delta \sigma_\infty \sqrt{\pi a}, \quad (24)$$

where a is the crack length and $\Delta \sigma_\infty$ the remotely applied stress range. With Paris' law

$$da/dN = C \Delta K_I(a)^\alpha, \quad (25)$$

where C and α are material parameters, the crack growth rate da/dN can be quantified. With the initial condition $N_0 = 0$ and an initial crack length a_0 , the differential equation in Equation (25) can be analytically solved by

$$a(N) = \left(\frac{(2 - \alpha) C \Delta \sigma_\infty^\alpha \pi^{\frac{\alpha}{2}}}{2} N + a_0^{\frac{2-\alpha}{2}} \right)^{\frac{2}{2-\alpha}}, \quad (26)$$

where N is the number of cycles. For this example, 50 trajectories are simulated by setting $\Delta \sigma_\infty = 48.26$ MPa, $a_0 = 9 \times 10^{-3}$ m, $\alpha = 2.9$, and sampling C from a normal distribution with $\mu_C = 8.7096 \times 10^{-11}$ and $\sigma_C = 1.519 \times 10^{-11}$ (C with $[da/dN] = \text{m/cycle}$ and $[\Delta K_I] = \text{MPa} \sqrt{\text{m}}$) for each trajectory. Often, researchers assume C to follow a log-normal distribution. A log-normal distribution approaches a normal distribution for $\frac{\mu_C}{\sigma_C} \rightarrow \infty$. In our case, $\frac{\mu_C}{\sigma_C} = 5.73$ which approximately leads to a normal distribution. This is why we can assume C to be normally distributed. In order to represent the measurement noise, an observation error $\sigma_y \sim \mathcal{N}(0, 0.16 \times 10^{-6} \text{m}^2)$ is added to the computed crack lengths. Figure 1(a) displays the crack growth trajectories already split into training and test data.

Training of GPs

First, a GP without the use of a warping function is trained on 35 trajectories of the simulation data representing previously collected data. A set of polynomial basis functions $\boldsymbol{\phi}_y(x)$ with orders 0 to 5 is chosen, resulting in a polynomial mean and covariance function of order 5, see Equations (2)–(4). The determined GP with its mean function and symmetric 95% credible region is shown in Figure 1(b). The figure reveals that the credible region assumes negative values, which is nonphysical since crack lengths can only be zero or positive.

In the second case, the warping function in Equation (10) is considered. The free parameter θ_1 is optimized with respect to Equation (15). In this case, two polynomial basis functions in the latent space $\boldsymbol{\phi}_z(x)$ with orders 0 and 1 are chosen since the inner part of Equation (26) represents a GP with straight lines. This is because the parameters α , $\Delta \sigma_\infty$ and a_0 are constants, N is the variable, and C is normally distributed. The optimization of the warping function should therefore lead to the inverse of the outer exponent

$$\theta_1^* = \frac{2 - \alpha}{2} = -0.45 \quad (27)$$

such that the latent space is modeled by a GP with straight lines. In this example, the optimizer leads to a value of $\hat{\theta}_1 = -0.4489$ with a relative error of 0.25%, which is a close approximation of the optimal solution. The difference results from approximating the observation error in the latent space with Equation (19). Considering the trajectories without the added noise leads to the analytical solution.

The resulting GP in the latent space and the corresponding warped GP, which is mapped to the observed space by the inverse warping function, are shown in Figures 1(c) and (d), respectively. Figure 1(d) shows

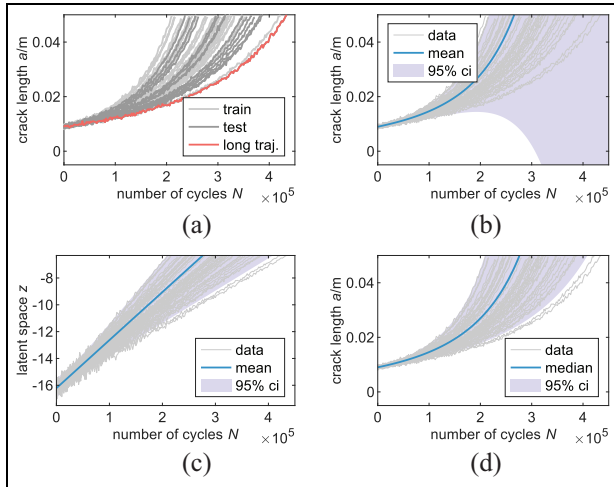


Figure 1. (a) Crack growth trajectories, (b) GP in the observed space, (c) warped GP in the latent space and (d) in the observed space.
GP: Gaussian process.

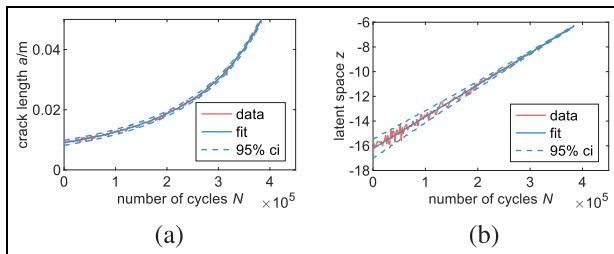


Figure 2. Observation error in (a) observed space and (b) latent space.

that the credible region of the warped GP assumes only positive values. Figure 1(c) additionally shows that the combination of the warping function and the choice of polynomial basis functions with orders 0 and 1 leads to the desired solution. During the optimization process, the optimizer tries, on the one hand, to warp the data such that the trajectories follow a normal distribution in the latent space. On the other hand, the optimization tries to straighten the simulated data in the latent space to fit the presumed basis functions.

Figure 2 shows another effect related to the transformation, which we mentioned at the end of the Approach section. The constant observation error σ_y , see Figure 2(a), is warped too and varies over N in the latent space, see Figure 2(b). For higher cycle numbers, the error σ_z decreases. The proposed approach takes this into account by applying weighted least squares

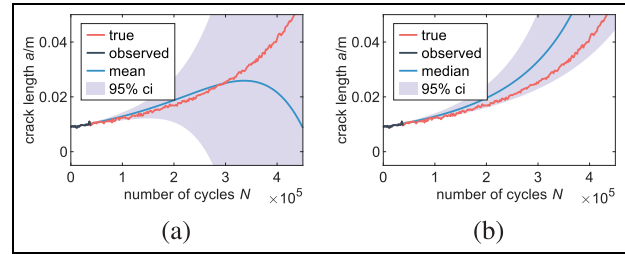


Figure 3. Crack growth prediction for the infinite plate at cycle number 38,000 of (a) GP and (b) warped GP.
GP: Gaussian process.

regression and approximating the warped observation error σ_z according to Equation (19).

Condition GPs on current crack length data

In the next step, we gradually condition the GPs on currently observed data to update the prediction for an unseen trajectory according to Equations (6)–(8). This is done for the entire test set, even though only one test trajectory is depicted here. Figure 3 shows the GP's prediction of the longest test trajectory (indicated in red) conditioned on currently observed data. Since the monitoring process is continuous, not all measurements are available from the beginning. The currently available ones are shown in black. In both cases (standard and warped), the credible regions narrow down with more available current data. Figure 3(a) reveals that the standard GP initially updates the mean function so that it becomes negative. This is because the current data has a relatively flat trend which, in combination with the assumed basis functions, results in such predictions. By comparison, the warped GP in Figure 3(b) shows strictly positive predictions. The warped GP is conditioned in the latent space, and the results are afterwards mapped to the observed space by the inverse of the warping function, leading to strictly positive predictions.

In order to compare the two models, the mean negative log-likelihood (MNLL), the mean absolute error (MAE), and the root mean squared error (RMSE) of the test set are quantified. The crack lengths predicted at the trajectory's last number of cycles for each conditioning step are compared to the latest realized crack length for all metrics. The results are listed in Table 1. For all metrics, the warped GP performs better than the standard GP as it predicts strictly positive crack lengths and resembles the optimal solution (GP with straight lines in the latent space with $\theta_1^* = -0.45$) closely. The advantage of warped GPs is particularly apparent for long trajectories. One reason for this is

Table 1. Evaluation results for the infinite plate of the standard and warped GP.

Model	MNLL		MAE		RMSE	
GP	-3.6		14.7 mm		22.2 mm	
Warped GP	-4.7	-32.2%	6.8 mm	-53.6%	6.8 mm	-69.4%

GP: Gaussian process; MNLL: mean negative log-likelihood; MAE: mean absolute error; RMSE: root mean squared error.

that long trajectories have a relatively small slope at the beginning.

As we have seen, utilizing the Box-Cox transformation for warping our data leads to approximations close to the analytical solution for the infinite plate example. However, using this exact type of transformation function results in imaginary numbers if θ_1 is not an integer and y is negative. Additionally, if θ_1 is an even number, the warping function becomes non-monotonic over $y \in \mathbb{R}$, violating our assumption.

As presented earlier, we first compute our predictions in the latent space and second map them by the inverse warping function to the observed space. Since the inverse of the Box-Cox transformation reads

$$\psi_{\theta}^{-1}(z) = y = (\theta_1 z + 1)^{\frac{1}{\theta_1}} = \sqrt[\theta_1]{\theta_1 z + 1}, \quad (28)$$

we must also ensure that $z \geq -\frac{1}{\theta_1}$ for $\theta_1 > 0$ and $z \leq -\frac{1}{\theta_1}$ for $\theta_1 < 0$ (both for $\theta_1 \neq 0$ and $\theta_1 \neq 1$). In the infinite plate example, we do not encounter any of the before-mentioned problems as our data is strictly positive ($y > 0$) and our predictions in the latent space comply with the requirement $z \leq -\frac{1}{\theta_1}$.

These constraints limit the applicability of this specific type of Box-Cox transformation. Since the Box-Cox transformation can be useful even in situations where no power transformation can produce normality exactly,⁴² it is worth adjusting it. Bickel and Doksum⁴³ presented the modified Box-Cox transformation function

$$\psi(y) = z = \frac{\text{sgn}(y)|y|^{\theta_1} - 1}{\theta_1} \text{ for } \theta_1 > 0, \quad (29)$$

which was also used by Rios and Tobar.³⁹ This modification leads to strictly monotonic warping functions over $y \in \mathbb{R}$ also if θ_1 becomes an even number. Yet, for $\theta_1 < 0$, the modified Box-Cox transformation function results in a discontinuity at $y = 0$. Therefore, we must constrain θ_1 to be positive, ensuring our warping function complies with the assumptions made for our approach.

Ultimately, the infinite plate example reveals the great predictive capabilities of warped GPs. Since the GP is modeled in a latent space, the warping approach is more flexible than standard GPs. The approach

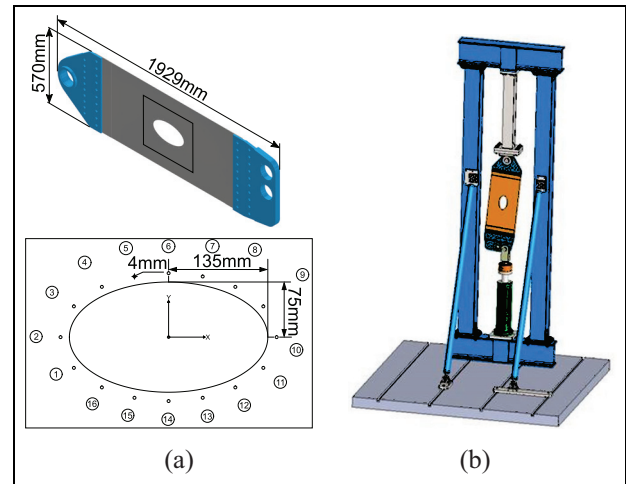


Figure 4. (a) Specimen and (b) CAD of the test rig. CAD: computer-aided design

enables the prediction of non-normally distributed trajectories, which is especially useful for PHM problems due to their non-negative nature. By contrast, since standard GPs rely on a normal distribution, the possible functions of their conditional distribution can become negative. Yet, one drawback of warped GPs is that the parameters of the warping function need to be optimized, leading to greater computational effort.

Application to an aerospace structure

The method proposed in the Approach section is also applied to predict the crack growth in an aluminum panel that resembles a lower section of a civil aircraft wing. The structure is made of the aluminum alloy Al 2024-T351 with Young's modulus of $E = 72,000$ MPa and a Poisson's ratio of $\nu = 0.34$. It is 1,920 mm long and 570 mm wide and has an elliptical armhole in its center with a length of 135 mm and a width of 75 mm. Around the armhole are 16 holes with a diameter of 4 mm. The specimen is shown in Figure 4(a).

The armhole in the center of the structure is usually covered by a lid that is fixed on the smaller holes around it. The armhole allows reaching into the wing to inspect the inner part of the structure with an endoscope. This aerospace structure is prone to fatigue

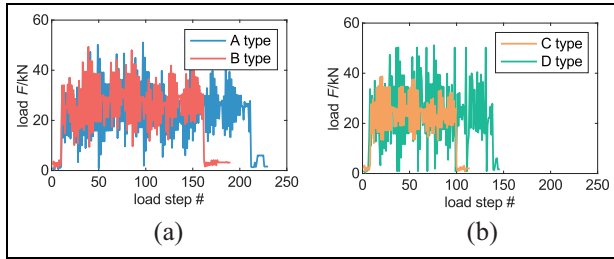


Figure 5. Loading program with flight types (a) A and B and (b) C and D.

cracks, and it is used to showcase the developed method on warped GPs.

The present section is divided into five parts: First, we describe the experimental setup of the executed fatigue tests. Second, the simulation of multiple crack growth trajectories for generating training data is explained. Third, the generated trajectories are used for training a warped GP according to the proposed method. Fourth, we show how to estimate the current state of the structure, that is, the current crack length, based on measured strain data. And fifth, future crack lengths are predicted by using the trained warped GP and the estimated current crack lengths.

Experimental setup

During the fatigue test of the aerospace structure, the load was applied by a hydraulic cylinder at an angle in order to represent the shear stresses in a wing resulting from twisting, see Figure 4(b). The loading program is based on four different flight types, A, B, C, and D, to realistically simulate the forces acting on an aircraft wing, see Figure 5.

They consist of 230, 190, 114, and 146 load steps, respectively. The flight types are ordered randomly for the first 100 flights and repeated consecutively afterwards. The occurrences of the different flight types, A, B, C, and D, in the first 100 flights are 55, 15, 20, and 10 times, respectively.

Two equally manufactured specimens, P02T01 and P03T01, were tested with the same loading program. Several sensors were attached to the panel to monitor the structures. They were predominantly positioned according to the method described in Reference.^{44,45} According to this method, the change of strain regions due to different possible cracks are evaluated with FE analyses. Based on a certain requirement for the change of strain that can be detected by the applied sensors, the positions that satisfy this requirement for every possible crack are determined. Therefore, the method ensures that the deployed sensors are able to detect the occurring cracks. In total, two single strain gauges and

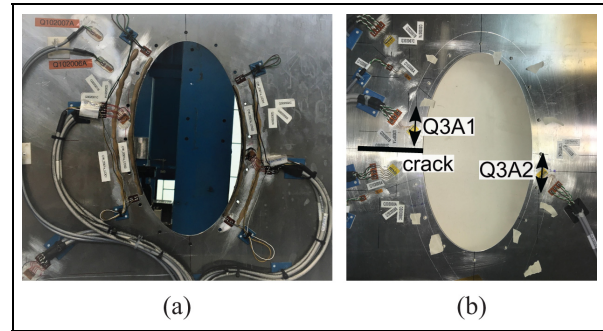


Figure 6. Sensors of (a) P02T01 and (b) P03T01 with the two investigated strain directions Q3A1 and Q3A2.

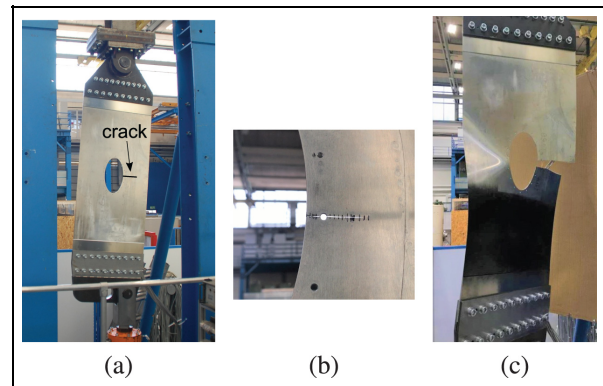


Figure 7. Fatigue process of aerospace structure with (a) intact structure, (b) crack growth, and (c) final fracture.

three strain rosettes were attached to specimen P02T01 and two single strain gauges and four strain rosettes to specimen P03T01, see Figures 6(a) and (b). For each flight, the strains were measured at every load step.

The experiments were run until final fracture of the structure, see Figure 7. During the experiment, the structure was regularly inspected by test engineers in order to detect cracks and measure their lengths.

Generation of training data

The method proposed in the Approach section enables integrating prior knowledge from previously collected degradation trajectories into GPs. As large structural fatigue tests are usually carried out only once, previously measured degradation trajectories are missing. However, due to analytical equations and FE analyses, much of the degradation behavior is understood a priori. In order to integrate this knowledge into GPs, degradation trajectories are gathered by conducting virtual simulations. First, the fatigue life and second, the crack growth are computed to simulate the

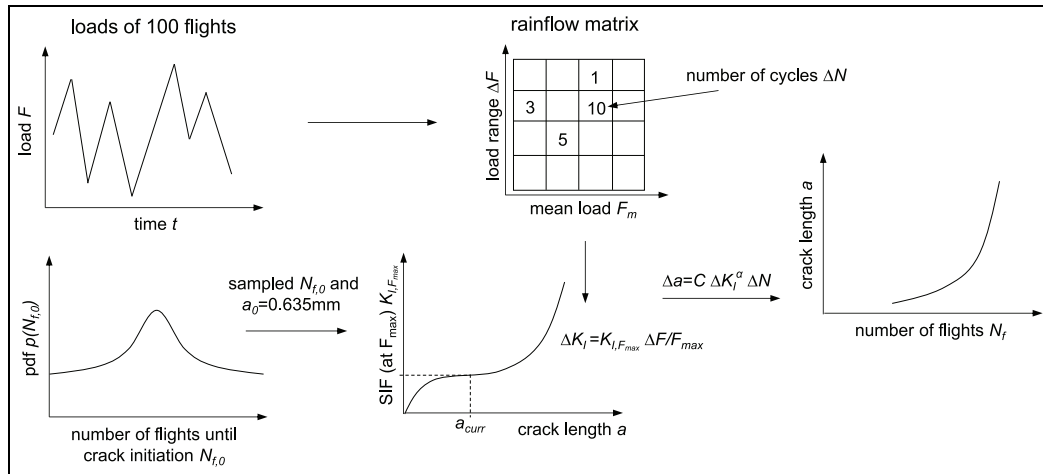


Figure 8. Schematic representation on how the crack growth in the aerospace structure is simulated where F_{max} is the maximum force in the loading program.

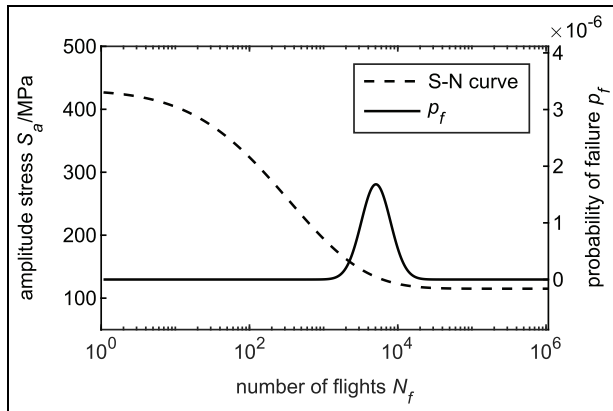


Figure 9. S–N curve and log10 normal distribution for the present loading program.

structure’s degradation. The entire simulation process is schematically shown in Figure 8.

Fatigue life. First, an FE analysis is carried out to quantify the stresses in the structure. Two local hot spots are found at the small holes 5 and 6 on the side toward the armhole. In the following, we assume the crack to start at hole 6. However, the methodology can also be extended to consider cracks starting from multiple spots by modeling a mixture of GPs.

By applying rainflow counting and the Haigh diagram to the computed stress at hole 6, the stresses for the entire loading program can be mapped to

amplitude stress blocks with a constant stress ratio of $R = -1.0$. Using Miner’s linear damage accumulation rule and the 50% S–N curve corresponding to the structure’s material, the median number of flights after which a crack will occur can be determined. Note that we denote the number of flights as N_f . In order to quantify the uncertainty of the crack initiation, a material corresponding scatter parameter $s = 0.197$ of the S–N curve from Reference⁴⁶ is used. The resulting log10 normal distribution and the S–N curve are shown in Figure 9. Now, the number of flights for an initial crack $N_{f,0}$ can be sampled from this distribution.

Crack growth. Second, the crack growth in the structure is evaluated. To quantify the relationship between the SIF and the crack length a , multiple crack computations are evaluated using the extended finite element method (XFEM).^{47–49} The crack is assumed to first propagate toward the armhole (crack length a_1) and then toward the edge of the structure (crack length a_2), see Figure 10.

In total, 382 XFEM computations (static analyses), see Figure 11(a), with different crack lengths are evaluated to quantify the relationship between a and K_I . Figure 11(b) reveals a scatter of the simulated data mainly toward longer crack lengths ($a > 30\text{mm}$). The scatter results from numerical errors due to larger elements further away from the armhole. As we do not want to consider these errors, we model only the general relationship and not the scatter/distribution of the data points. Two separate NNs, for a_1 and a_2 , with

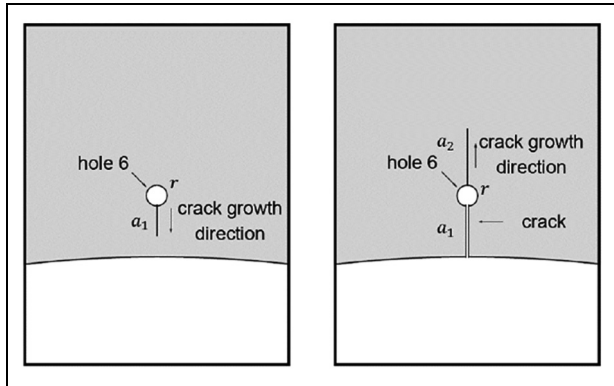


Figure 10. Assumption for crack propagation where crack 1 first propagates toward the armhole (left) and crack 2 to the edge of the specimen afterwards (right).

two and four neurons (one hidden layer) are trained to map the crack length onto the SIF $K_I(a)$, see Figure 11(b).

In this paper, we define the total crack length a as

$$a = \begin{cases} a_1, & \text{if } a < 10 \text{ mm} \\ a_1 + 2r + a_2, & \text{if } a > 10 \text{ mm}. \end{cases} \quad (30)$$

As Pfingstl et al.,⁴⁴ an initial crack length of $a_0 = 0.635$ mm which is, according to Ryschkewitsch,⁵⁰ the smallest crack length detectable by eddy current testing, is assumed. The crack growth can then be computed by the Paris law of Equation (25) with

$$\Delta K_I = K_{I, \max} - K_{I, \min}. \quad (31)$$

In order to reduce computational time, the crack growth computations are simplified by applying rain-flow counting⁵¹ to the load steps of 100 flights to compute the load ranges and using 1/100 of their frequencies for each flight. Furthermore, we assume $K_{I, \min} = 0$ and that the crack length is constant during

one flight. According to Virkler et al.,⁵² crack growth is subject to great uncertainties. Therefore, C is assumed to be a random variable that is normally distributed with $\mu_C = 8.7096 \times 10^{-11}$ and $\sigma_C = 6.5680 \times 10^{-12}$ (determined from the crack growth data published by Virkler et al.⁵² with C with $[da/dN] = m/\text{cycle}$ and $[\Delta K_I] = \text{MPa}\sqrt{m}$), and the material parameter α is set to $\alpha = 2.9$ according to Spencer et al.⁵³ Moreover, the error of the load range is varied according to a normal distribution with $\sigma_{dF} = 5.0\%$ for each trajectory. For computing different crack growth trajectories, a set of parameters ($N_{f,0}$, C , dF) is sampled for each trajectory.

Figure 12 shows the computed degradation trajectories. Note that the step in each trajectory results from the two different cracks, a_1 and a_2 , and the added diameter of the hole after crack 1 reaches the armhole. We can also see the huge scatter of the initial starting points of the trajectories that results from the S–N curve distribution shown in Figure 9. Furthermore, the curvatures of the trajectories vary due to different values for C and ΔF . Figure 12 also shows that no observation error is artificially added to the simulation data.

Training of GP

A warped GP can now be trained on this simulation data set according to the method proposed in the Approach section. By using simulation data for training, however, one has to be sure that the simulations are correct. When there are doubts about whether the simulations represent the experiment well, they should be verified based on experiments before using them as the training set.

For this example, the modified Box–Cox transformation of Equation (28) is used as the warping function and a polynomial of degrees 0 and 1 (intersection and slope) as basis functions. Therefore, the optimizer tries again to achieve not only a normal distribution in the latent space but also straight lines. The optimized

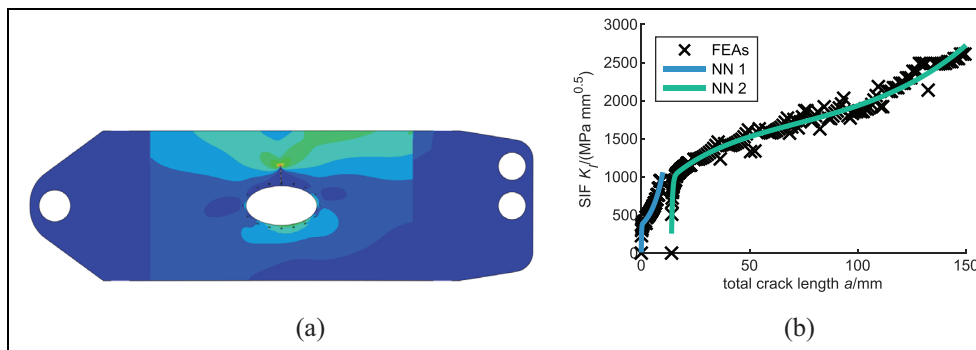


Figure 11. (a) XFEM computation and (b) trained NN. XFEM: extended finite element method; NN: neural network.

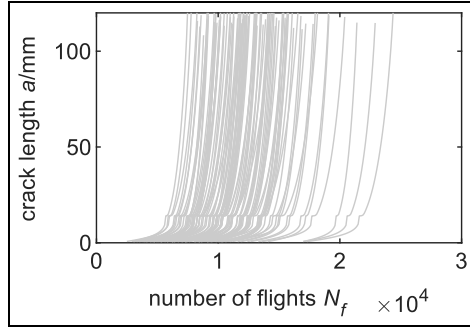


Figure 12. Computed degradation trajectories.

solution is $\hat{\theta}_1 = 4.46 \times 10^{-8}$ which is close to a log transformation. And for a log transformation, the distribution in the observed space is completely defined on the positive domain. Thus, the predicted crack lengths will stay positive after transferring the values from the latent space to the observed space. Moreover, by using polynomial basis functions of degrees 0 and 1, we ensure that the resulting predictions are strictly monotonic in the observed space. Since no observation error is present in the simulation data, a non-weighted least squares regression is applied to determine the basis functions' weights. The modeling error in the latent space $\sigma_{z,m}$ is assumed to be constant and is approximated by taking the square root of the average of all squared residuals in the latent space.

Figures 13(a) and (b) show the training data and the determined warped GP in the observed and the latent space. It can be seen that the trajectories are almost straight lines in the latent space. Moreover, the mean function and the credible region in the observed space assume only positive values which is in agreement with the physics. Figure 13(a) depicts the prediction before any data of the monitored structure is available.

Estimation of current states

In order to update the trained GP, the current state of the system has to be observed. As the present GP is based on crack growth data, the current crack length must be determined. In the present study, the applied strain gauges are used to determine the current crack length. With Bayes law, the crack length a can be inferred from the measured strain ϵ_{SG} by

$$p(a|\epsilon_{SG}) = \frac{p(\epsilon_{SG}|a)p(a)}{\int_{a_0}^{a_c} p(\epsilon_{SG}|a)p(a)da}, \quad (32)$$

where $p(a)$ is the prior distribution and $p(\epsilon_{SG}|a)$ the likelihood. For measuring multiple strains and assuming them to be measured independently, the probability

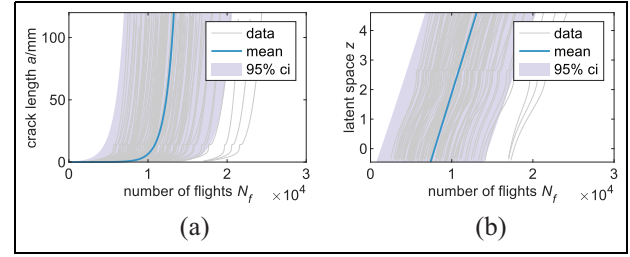


Figure 13. Trained GP in (a) the observed space and (b) the latent space.

GP: Gaussian process.

density function of ϵ_{SG} for s applied strain gauges given a becomes

$$p(\epsilon_{SG}|a) = p(\epsilon_{SG}^{(1)}, \dots, \epsilon_{SG}^{(s)}|a) = \prod_{l=1}^s p(\epsilon_{SG}^{(l)}|a). \quad (33)$$

Additionally assuming the prior $p(a)$ to be uniformly distributed within a_0 and a_c , the probability density function of a given all strain measurements becomes

$$p(a|\epsilon_{SG}) = \frac{\prod_{l=1}^s p(\epsilon_{SG}^{(l)}|a)}{\int_{a_0}^{a_c} \prod_{l=1}^s p(\epsilon_{SG}^{(l)}|a)da}. \quad (34)$$

In order to be able to use Equation (34), the likelihoods $p(\epsilon_{SG}^{(l)}|a)$ with $l \in 1, \dots, s$ that incorporate how the measured data ϵ_{SG} depends on the current crack length a have to be known. Therefore, the strains of all XFEM computations (see the subsection Generation of training data) are evaluated for each sensor position to quantify the relationship between the crack lengths and the strains. Then, the results are used to fit an NN for each strain gauge position. Since we only consider strictly monotonic strain gauges, in total, seven NNs $\epsilon_{NN}^{(l)}(a)$ are trained. By assuming a normally distributed measurement error of $\sigma_\epsilon = 200 \mu\text{m/m}$, the likelihoods $p(\epsilon_{SG}^{(l)}|a)$ are completely defined and can be evaluated.

In order to cancel out the bias term which might emerge from the difference between the FE analysis and the real measurement, only the relative change of strain due to a crack

$$\Delta \epsilon_{NN, \text{rel}}^{(l)}(a) = \frac{\epsilon_{NN}^{(l)}(a) - \epsilon_{NN}^{(l)}(a_0 = 0)}{\epsilon_{NN}^{(l)}(a_0 = 0)} \quad (35)$$

and

$$\Delta \epsilon_{SG, \text{rel}}^{(l)} = \frac{\epsilon_{SG}^{(l)} - \epsilon_{SG}^{(l)}(a_0 = 0)}{\epsilon_{SG}^{(l)}(a_0 = 0)} \quad (36)$$

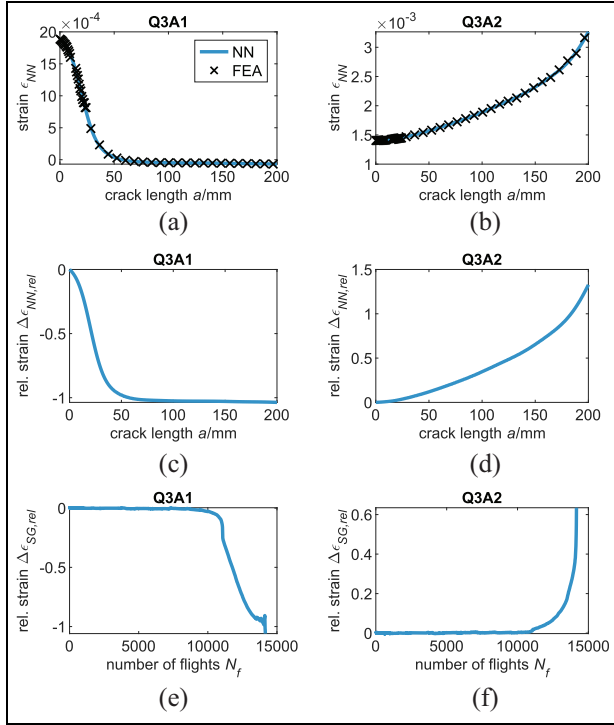


Figure 14. Trained NNs on absolute strains for positions (a) Q3A1 and (b) Q3A2. NNs representing the relative change of strains for (c) Q3A1 and (d) Q3A2. Measured change of strains over the number of flights for (e) Q3A1 and (f) Q3A2. NNs: neural networks.

are considered. Since settlement effects happen at the beginning of each test, the measurement of flight 500 is used as $\epsilon_{SG}(a_0 = 0)$.

Figures 14(a) and (b) show the absolute strains for two strain gauge positions computed by FE analyses (black crosses). Additionally, the blue lines indicate the trained NNs. As mentioned before, we do not use the absolute values since there might be a bias term between the FE analysis and the measurements. Therefore, we compute the relative change of strains shown in Figures 14(c) and (d). Moreover, the change of strains for two different sensor positions (Q3A1 and Q3A2) over the number of flights are displayed in Figures 14(e) and (f). By using the strain's relative change, all lines begin at zero, canceling the bias term. The figures also show that if the crack grows, the strain becomes smaller at the position of sensor Q3A1 and larger at Q3A2. Furthermore, Q3A1 indicates a very sensitive behavior for small cracks and Q3A2 for larger ones. This can be explained by looking at Figure 6(b). The figure shows that the crack occurs close to position Q3A1 and cuts the load path so that smaller

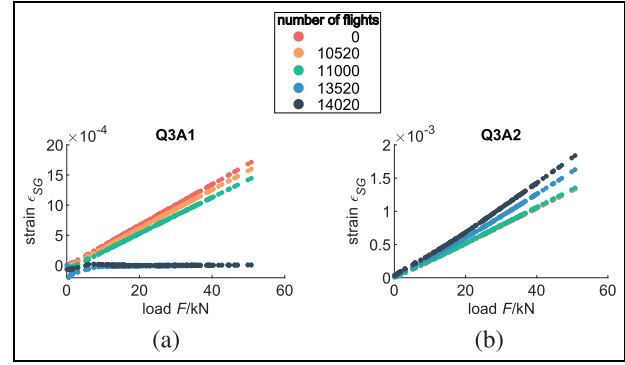


Figure 15. Measured strains for different flights at position (a) Q3A1 and (b) Q3A2. The color indicates a certain number of flights.

strains are measured. By contrast, position Q3A2 lies on the opposite side of the crack. If the crack starts growing, no significant change of the strain is measured. However, once the crack is long enough, the load path is shifted to the side of position Q3A2, increasing the measured strain.

The relative change of strain for every flight is computed by the relative change of the strain–force slope shown in Figure 15. While the change of the slope for position Q3A1 can be well distinguished for smaller numbers of flights (smaller cracks), position Q3A2 is better for indicating the crack lengths at larger numbers of flights (larger cracks), see Figure 15(a) and (b), respectively.

By using the relative change of strain instead of the absolute values, the likelihood becomes

$$p(\Delta \epsilon_{SG,rel}^{(l)} | a) = \frac{1}{\sqrt{2\pi \left(\frac{\sigma_\epsilon}{\epsilon_{NN}^{(l)}(a_0)} \right)^2}} \exp - \frac{\left(\Delta \epsilon_{SG,rel}^{(l)} - \Delta \epsilon_{NN,rel}^{(l)}(a) \right)^2}{2 \left(\frac{\sigma_\epsilon}{\epsilon_{NN}^{(l)}(a_0)} \right)^2}, \quad (37)$$

which is used instead of $p(\epsilon_{SG}^{(l)} | a)$. The current crack length is determined by

$$\hat{a} = \arg \max p(a | \epsilon_{SG}) \quad (38)$$

and its variance by

$$\sigma_a^2 = \int_{a_0}^{a_c} (a - \mu_a)^2 p(a | \epsilon_{SG}) da. \quad (39)$$

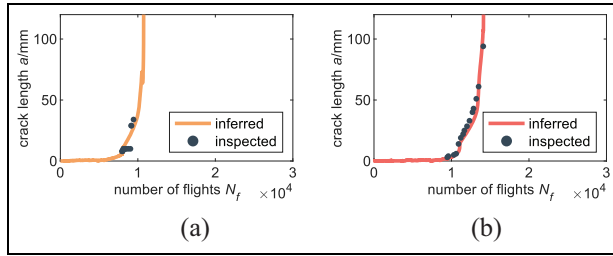


Figure 16. Inferred crack lengths and visual inspections for specimen (a) P02T01 and (b) P03T01.

$$\mu_a = \int_{a_0}^{a_c} ap(a|\epsilon_{SG})da \quad (40)$$

Figure 16(a) and (b) show the crack lengths inferred from the measured strains and the corresponding crack lengths visually inspected by test engineers for both specimens. It can be seen that the inferred crack lengths for P03T01 closely match the inspected ones ($R^2 = 0.926$), whereas the match of the P02T01 trajectory is not so close ($R^2 = 0.656$). A big step in the inspected data of specimen P02T01 can be seen. It is likely that the test engineers did not detect the crack on the outer part of the small hole straightaway.

As the P03T01 specimen was tested after P02T01, the test engineers were already familiar with the type of structure and were, therefore, able to measure the crack lengths more accurately. Still, both crack growth behaviors are very similar to the inspected ones, and are close to the simulations in terms of locations, numbers, and crack growth rate.

Prediction of future states

Although Figures 16(a) and (b) show the entire trajectories of the inferred crack lengths, during the test, the crack length is only partially known, that is, from $N_f = 0$ up to the current number of flights. After each new flight, the current crack length and its uncertainty can be determined and used to compute the conditional GP leading to an updated prediction. For doing so, all inferred crack lengths \hat{a} and their estimated observation errors $\hat{\sigma}_a$ up to the current flight cycle are transformed to the latent space using Equations (10) and (19), respectively. In this case, the total squared error in the latent space σ_z^2 is assumed to be the sum of the squared modeling error $\sigma_{z,m}^2$ and the transformed squared observation error $\sigma_{z,a}^2$ ($\sigma_z^2 = \sigma_{z,a}^2 + \sigma_{z,m}^2$). Furthermore, crack sizes smaller than the initial crack length of 0.635 mm are ignored. Then, the conditional GP is computed in the latent space using Equations (6)–(8).

The updated prediction is transformed to the observed space using the inverse warping function.

Figures 17(a) to (h) show the updated predictions for specimens P02T01 and P03T01 at different time states. Initially, the GP's prediction is entirely based on the knowledge gained from analytical equations and FE analyses. Once a crack length greater than 0.635 mm is inferred from the strain data, the mean function starts to change and the credible region narrows down, leading to a more accurate and precise prediction. Since the GP is defined by a set of polynomial basis functions with orders 0 and 1, the step due to the two different crack regimes is not apparent in the prediction. As in the infinite plate example, the predictions are strictly positive again, which complies with the physics.

Discussion

Warped GPs

As demonstrated for an infinite plate and an aerospace structure, the proposed approach on warped GPs can even handle data that is not normally distributed. In both cases, introducing a warping function leads to predictions that assume strictly positive values. This is in agreement with the physics since crack lengths can only be positive. The approach reproduces the analytical solution for problems without an observation error and leads to a close approximation ($<0.5\%$) for cases where an observation error is present. However, by using warped GPs, free parameters θ are introduced. These need to be optimized by minimizing the NLL, which is, in general, a non-convex optimization problem requiring increased computational effort. In contrast to existing warped GP approaches, no GP parameters of the mean and covariance function need to be optimized by minimizing the NLL since the GP is determined by solving a linear regression problem. By modeling the GP with a weighted sum of basis functions therefore allows integrating prior knowledge quickly into GPs.

Simulations

Prior knowledge in the form of degradation trajectories is often rare, especially if the mechanical system is large and expensive. Using analytical and FE-based simulations can produce valuable information that can be incorporated into GPs by the presented approach, see section Application to an aerospace structure. By splitting the approach into two parts, (1) training a GP on previous data and (2) computing the conditional distribution for updating the prediction based on currently monitored data, allows rapid predictions that may be

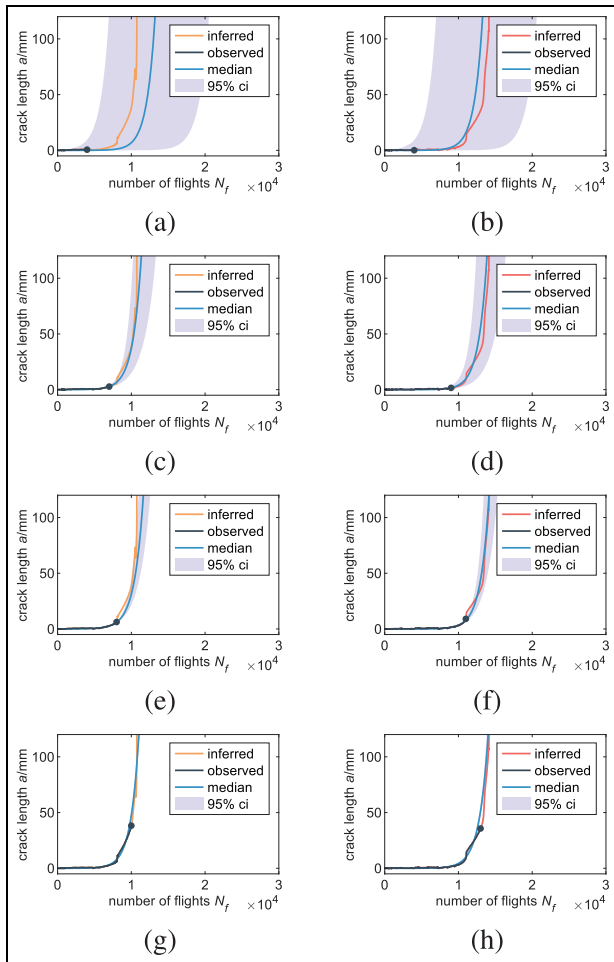


Figure 17. GP predictions for specimen P02T01 at number of flights equal to (a) 4000, (c) 7000, (e) 8000, and (g) 10,000, and for specimen P03T01 at (b) 4000, (d) 9000, (f) 11,000, and (h) 13,000 flights.

GP: Gaussian process.

evaluated online. Nevertheless, incorrectly executed simulations resulting, for example, from the use of incorrect parameters, can lead to weak predictions. Therefore, the simulation and its parameters, as well as their uncertainties, should be well known. To avoid overconfidence, it is better to assume overly large variances than variances that are too small since there might be sources of uncertainties that are not known in advance.

Hidden state of the system

After the GP is defined by its mean function and covariance function, it can be conditioned on current data. The current state of the system, however, is often hidden. Therefore, the present paper shows how to infer the crack length from strain data. Based on the

coefficient of determination, the resulting crack lengths using Bayesian inference match the crack lengths measured during inspections with $R^2 = 0.656$ and $R^2 = 0.926$ accuracy for the first and second aerospace specimens, respectively. The approach enables continuous monitoring of the crack length and its uncertainty. By using this information, the GP's predictions can be continuously updated.

Conclusion

The present paper proposes a PHM algorithm that is based on GPs. It is successfully applied to an infinite plate and an aluminum aerospace structure in order to predict crack growth. The established model predicts not only the crack length for every future time step but also its credible intervals. By describing the crack length as a random variable, different credible regions, for example, a 99% or 99.999% region, can be computed, which allows the user to ensure different levels of safety.

The proposed algorithm is based on warped GPs. Using warped GPs reduced the MAE by 53.6% and the MNLL by 32.2% for the infinite plate example investigated in this paper. It also leads to strictly positive predictions for crack lengths, which is in accordance with the physics. Furthermore, the proposed approach can quickly integrate prior knowledge in the form of several previously generated degradation trajectories. Since prior knowledge is often only available in terms of analytical equations and FE analyses, the present paper shows how to generate them for an aerospace structure prone to fatigue cracks. In general, the simulations must be trustworthy and might be verified before using them. Therefore, the approach is currently limited to crack growth predictions for isotropic materials. In future research work, it would be interesting to apply a similar approach to composite structures.

In order to update the GP's prediction, the current crack lengths are inferred from strain data, which agree well with the visually inspected ones. Ultimately, the estimated and predicted crack lengths can be used to compute the probability of failure and therefore to better schedule maintenance tasks.

Acknowledgements

We thank Prof. Dr. Kochenderfer from the Department of Aeronautics and Astronautics of Stanford University for the valuable discussion.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Federal Ministry for Economic Affairs and Energy based on a decision by the German Bundestag in the national aeronautics program LuFo V as a part of the research project Strubatex.

ORCID iD

Simon Pfingstl  <https://orcid.org/0000-0002-7282-889X>

References

1. Tinga T and Loendersloot R. Aligning PHM, SHM and CBM by understanding the physical system failure behaviour. In: Bregon A and Daigle M (eds.) *Proceedings of the European Conference of the Prognostics and Health Management Society*. Nantes: PHM Society, 2014, pp. 162–171.
2. Kim NH, An D and Choi JH. *Prognostics and health management of engineering systems: an introduction*. Switzerland: Springer International Publishing, 2017.
3. Lei Y, Li N, Guo L, et al. Machinery health prognostics: a systematic review from data acquisition to RUL prediction. *Mech Sys Signal Process* 2018; 104: 799–834.
4. Do DT, Lee J and Nguyen-Xuan H. Fast evaluation of crack growth path using time series forecasting. *Eng Fract Mech* 2019; 218: 106567.
5. Malhi A, Yan R and Gao R. Prognosis of defect propagation based on recurrent neural networks. *IEEE Trans Instrum Meas* 2011; 60(3): 703–711.
6. Heimes FO. Recurrent neural networks for remaining useful life estimation. In: *International conference on prognostics and health management*, Denver, CO, USA, 06–09 October 2008, pp. 59–64. IEEE.
7. Peng Y, Wang H, Wang J, et al. *A modified echo state network based remaining useful life estimation approach*. In: *IEEE conference on prognostics and health management*, Denver, CO, USA, 18–21 June 2012, pp. 1–7. IEEE
8. Benkedjough T, Medjaher K, Zerhouni N, et al. Remaining useful life estimation based on nonlinear feature reduction and support vector regression. *Eng Appl Artif Intell* 2013; 26(7): 1751–1760.
9. Carino JA, Zurita D, Delgado M, et al. Remaining useful life estimation of ball bearings by means of monotonic score calibration. In: *IEEE international conference on industrial technology*, Seville, Spain, 17–19 March 2015, pp. 1752–1758. IEEE
10. Dong S and Luo T. Bearing degradation process prediction based on the PCA and optimized ls-svm model. *Measurement* 2013; 46(9): 3143–3152.
11. Liu J and Zio E. An adaptive online learning approach for support vector regression: online-SVR-FID. *Mech Sys Signal Process* 2016; 76: 796–809.
12. Soualhi A, Medjaher K and Zerhouni N. Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. *IEEE Trans Instrum Meas* 2014; 64(1): 52–62.
13. Widodo A and Yang BS. Machine health prognostics using survival probability and support vector machine. *Expert Syst Appl* 2011; 38(7): 8430–8437.
14. Fumeo E, Oneto L and Anguita D. Condition based maintenance in railway transportation systems based on big data streaming analysis. *Procedia Comput Sci* 2015; 53: 437–446.
15. An D, Kim NH and Choi JH. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliab Eng Syst Saf* 2015; 133: 223–236.
16. Gobbato M, Kosmatka JB and Conte JP. A recursive bayesian approach for fatigue damage prognosis: an experimental validation at the reliability component level. *Mech Sys Signal Process* 2014; 45(2): 448–467.
17. Rasmussen CE and Williams CKI. Gaussian processes for machine learning. In: Dietterich T, Bishop C, Heckerman D, et al. (eds.) *Adaptive computation and machine learning*, 3rd ed. Cambridge, Mass: MIT Press, 2008: 13.
18. Williams C and Rasmussen C. Gaussian processes for regression. In: Touretzky D, Mozer MC and Hasselmo M (eds.) *Advances in neural information processing systems 8*. Cambridge, MA, USA: MIT Press, 1996, pp. 514–520.
19. Pfingstl S and Zimmermann M. On integrating prior knowledge into Gaussian processes for prognostic health monitoring. *Mech Sys Signal Process* 2022; 171: 108917.
20. Avendano-Valencia LD, Chatzi EN and Tcherniak D. Gaussian process models for mitigation of operational variability in the structural health monitoring of wind turbines. *Mech Syst Signal Process* 2020; 142: 106686.
21. Aye S and Heyns P. An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission. *Mech Syst Signal Process* 2017; 84: 485–498.
22. Chen J, Yuan S and Wang H. On-line updating Gaussian process measurement model for crack prognosis using the particle filter. *Mech Syst Signal Process* 2020; 140: 106646.
23. Gentile R and Galasso C. Gaussian process regression for seismic fragility assessment of building portfolios. *Struct Saf* 2020; 87: 101980.
24. Hong S and Zhou Z. *Remaining useful life prognosis of bearing based on gauss process regression*. In: *2012 5th International conference on BioMedical engineering and informatics*, Chongqing, China, 16–18 October 2012, pp. 1575–1579. IEEE
25. Kong D, Chen Y and Li N. Gaussian process regression for tool wear prediction. *Mech Sys Signal Process* 2018; 104: 556–574.
26. Kwon D, Azarian MH and Pecht M. Remaining-life prediction of solder joints using RF impedance analysis and Gaussian process regression. *IEEE Trans Compon Packag Manuf Technol* 2015; 5(11): 1602–1609.
27. Li G, Datta S, Chattopadhyay A, et al. An online-offline prognosis model for fatigue life prediction under biaxial cyclic loading with overloads. *Fatigue Fract Eng Mater Struct* 2019; 42(5): 1175–1190.
28. Li M, Sadoughi M, Hu Z, et al. A hybrid Gaussian process model for system reliability analysis. *Reliab Eng Syst Saf* 2020; 197: 106816.

29. Liu D, Pang J, Zhou J, et al. Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression. *Microelectron Reliab* 2013; 53(6): 832–839.
30. Mohanty S, Chattopadhyay A, Peralta P, et al. Bayesian statistic based multivariate Gaussian process approach for offline/online fatigue crack growth prediction. *Exp Mech* 2011; 51(6): 833–843.
31. Su G, Peng L and Hu L. A Gaussian process-based dynamic surrogate model for complex engineering structural reliability analysis. *Struct Saf* 2017; 68: 97–109.
32. Yu J. State of health prediction of lithium-ion batteries: Multiscale logic regression and Gaussian process regression ensemble. *Reliab Eng Syst Safe* 2018; 174: 82–95.
33. Snelson E, Ghahramani Z and Rasmussen C (2004) Warped Gaussian processes. In: Thrun S, Saul L and Schölkopf B (eds.) *Advances in neural information processing systems*. Cambridge, Massachusetts: MIT Press.
34. Lange M. On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors. *J Sol Energy Eng* 2005; 127(2): 177–184.
35. Kou P, Gao F and Guan X. Sparse online warped Gaussian process for wind power probabilistic forecasting. *Appl Energy* 2013; 108: 410–428.
36. Kou P, Gao F, Guan X, et al. Prediction intervals for wind power forecasting: using sparse warped Gaussian process. In: *IEEE Power and Energy Society General Meeting*. San Diego, CA, USA, 22–26 July 2012. IEEE
37. Kou P, Liang D, Gao F, et al. Probabilistic wind power forecasting with online model selection and warped Gaussian process. *Energy Convers Manag* 2014; 84: 649–663.
38. Mateo-Sanchis A, Muñoz-Marí J, Pérez-Suay A, et al. Warped Gaussian processes in remote sensing parameter estimation and causal inference. *IEEE Geosci Remote Sens Lett* 2018; 15(11): 1647–1651.
39. Rios G and Tobar F. Learning non-Gaussian time series using the box-cox Gaussian process. In: *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, 08–13 July 2018, pp. 1–8. IEEE.
40. Box GE and Cox DR. An analysis of transformations. *J R Stat Soc Series B Stat Methodol* 1964; 26(2): 211–243.
41. Gonçalves ÍG, Echer E and Frigo E. Sunspot cycle prediction using warped Gaussian process regression. *Adv Space Res* 2020; 65(1): 677–683.
42. Draper NR and Cox DR. On distributions and their transformation to normality. *J R Stat Soc Series B Stat Methodol* 1969; 31(3): 472–476.
43. Bickel PJ and Doksum KA. An analysis of transformations revisited. *J Am Stat Assoc* 1981; 76(374): 296–311.
44. Pfingstl S, Steiner M, Tusch O, et al. Crack detection zones: computation and validation. *Sensors* 2020; 20(9): 2568.
45. Pfingstl S and Zimmermann M. Strain-based structural health monitoring: computing regions for critical crack detection. In: *12th International workshop on structural health monitoring: Enabling Intelligent Life-cycle Health Management for Industry Internet of Things (IIOT)* (eds F-K Chang and F Kopsaftopoulos), Stanford, CA, 2019, pp.132–139. Lancaster, PA: DEStech Publishing Inc, 2019.
46. Haibach E. *Betriebsfestigkeit*. Berlin, Heidelberg: Springer, 2006.
47. Belytschko T and Black T. Elastic crack growth in finite elements with minimal remeshing. *Int J Numer Methods Eng* 1999; 45(5): 601–620.
48. Belytschko T, Mœ̈les N, Usui S, et al. (2001) Arbitrary discontinuities in finite elements. *Int J Numer Methods Eng* 2001; 50(4): 993–1013.
49. Mœ̈les N, Dolbow J and Belytschko T. A finite element method for crack growth without remeshing. *Int J Numer Methods Eng* 1999; 46(1): 131–150.
50. Ryschkewitsch M. *Nondestructive evaluation requirements for fracture critical metallic components. NASA Technical Standard (NASA Technical Standard NASASTD-5009)*. Washington, DC: National Aeronautics and Space Administration, 2008.
51. Clormann U and Seeger T. Rainflow-hcm. ein zählverfahren für betriebsfestigkeitsnachweise auf werkstoffmechanischer grundlage. *Stahlbau, Der* 1986; 55(3): 65–71.
52. Virkler DA, Hillberry B and Goel P. The statistical nature of fatigue crack propagation. *J Eng Mater Tech* 1979; 101(2): 148–153.
53. Spencer B, Tang J and Artley M. Stochastic approach to modeling fatigue crack growth. *AIAA J* 1989; 27(11): 1628–1635.