TMAS2001 - Degradation modelling

Jørn Vatn
Email: jorn.vatn@ntnu.no

Updated 2019-11-11

Introduction

Background

According to EN133006 *condition-based maintenance* is preventive maintenance which include assessment of physical conditions, analysis and possible ensuing maintenance actions. *Predictive maintenance* is condition-based maintenance carried out following a forecast derived from repeated analysis or known characteristics and evaluation of the significant parameters of the degradation of the item. That is, predictive maintenance is a subset of condition-based maintenance but where we explicitly need to understand the degradation of the item.

The term 'forecast' is often replaced with the term 'prognostics'. To support maintenance decisions it is desirable to know the point of time when a failure will occur. Due to the random nature of the degradation, we are not able to give one number for the point of time when the failure will occur. However, in many situations we are able to give a prediction with an *uncertainty interval* for the time to failure. The remaining useful life (RUL) of an item is the point of time where the item is not "usable" any more. RUL will depend on the state and age of an item at the current time. "Remaining" therefore means "from now on". Further we recognize that the RUL is a stochastic variable (random quantity). Note that it is not always easy to say when an item is not "usable" any more. This has to be defined. In some cases RUL points to the breakdown of the item, whereas in other cases RUL points to the point of time when the item is "shut down" by protective systems.

Degradation

Degradation is the act or process of degrading, where degrading means to impair in respect to some physical property. Wear, fatigue, erosion and corrosion are all degradation mechanisms.

To measure degradation is not always easy. In some situations we have direct measures of degradation i terms of for example size of fatigue cracks.

Crack size could therefore be a degradation measure. Still for this measure we have challenges because it might be hard to correctly measure the crack size, further the direction of the crack size and where the crack is situated is crucial for when a breakage will occur.

When it comes to rotating equipment it is even harder to measure degradation. We may measure degradation indirectly by for example the vibration level. But the vibration level is often not a very precise measure to forecast a coming failure.

In this presentation we use the term ‘degradation’ in a rather loose manner, independent on the challenges we will encounter. In some presentations the term ‘health indicator’ is used rather than degradation to emphasize that we are not always measuring degradation directly.

Anomaly detection, diagnostics and prognostics

A distinction is made between the terms ’anomaly detection’, ‘diagnostics’ and ‘prognostics’. All these terms are essential in predictive maintenance.

Anomaly detection

Anomaly detection is the process to distinguish normal behaviour from abnormal behaviour. In some situations we rather use the term ‘early warning detection’. For anomaly detection there are generally three approaches that are reported as promising:

1. First principle approaches where physical laws are used to represent the normal behaviour of the system. Several related measurements are linked together. Under normal operation we expect to observe some correlation between these measurements. In case of an anomaly we often observe a different pattern in the correlation, and this is used as basis for anomaly detection. For example a sump pump in a hydro power station is used to drain the sump regularly. When the level in the sump exceeds a predefined level, the pump is started and runs until a lower set-point is reached. When the pump runs the level is decreasing. By measuring the level over time, we indirectly measure the pump capacity. Upon degradation of the pump, we expect that it takes longer time to empty the sump. However, since the level in the sump is influenced both by the pump capacity and the inflow into the sump the time to empty the sump is not a precise measure. But if we in addition measure the inflow into the sump when the pump is not running, we can easily calculate the pumping capacity by comparing two curves. In other situations we may apply laws from thermodynamics to establish such “first principle” laws.

2. In recent years machine learning (ML) approaches replace the "physical based" models for correlation with models that are purely obtained by data-driven approaches. Several cases studies have been performed that show promising results. Machine learning is attractive because we do not need to specify in detail the physical laws. What we do is to present the data for the "machine". For example we might measure temperature, pressure, flow, rotation speed etc in a multi-stage compressor system. Under normal operation we assume that there is a relation between these measurements but we are not really able to specify. Then we present for example 70% of the available data to the "machine" which now will search for "correlation" among the measurements. Example of "algorithms" used in machine learning is "artificial neural network", "deep learning networks", and "support vector machines". Different algorithms have different "performance" for different data sets. When we have "trained" the model on the 70% training set, we test the model on the remaining 30% of the data.

3. Signal processing. For rotating equipment the standard approach for anomaly detection has been vibration analysis. For a rotating equipment we may measure distance, velocity and acceleration with a higher sampling frequency than the rotation frequency. These measurements can then be analysed in various ways. Both time domain and frequency domain analysis (FFF = fast Fourier analysis) is used. The tutorials on the vibration rig in the RAMS lab should indicate aspects of such anomaly detection

Diagnostics

In predictive maintenance diagnosis usually relates to search for root causes behind symptoms observed. For rotating machinery very often analyse the data stream in the frequency domain is the preferred approach. FFT of the data can identify typical "frequencies" and these are mapped to various underlying root causes. For static equipment and structures we usually have less "high resolution" information and hence it is more difficult to use the data to set the correct diagnosis. In recent years also machine learning has demonstrated success when it comes to diagnostics. When machine learning is used for anomaly detection we may process large amount of data rather automated because the period of "normal" behaviour is assumed to be known, and hence we know what is "normal". In diagnostics machine learning is more challenging because we manually need to code, or more precisely, *label* the data. For example when we detect a "potential anomaly", we manually label the situation, for example this is an outer ring failure of the bearing, a ball failure etc. Since manual labelling is required, machine learning requires significant analysis work to supervise the "training".

Diagnostics is crucial in order to get meaningful result out of prognostics. The various degradation mechanisms might show quite different degradation pattern, and hence a precise diagnose will give better prognoses.

Prognostics

Prognostics is the process of establishing models for the degradation when we look ahead.

Prognostics and residual useful life prediction comprise two main challenges; (i) how to describe in the time domain the development of some degradation measure used to monitor the condition of a system, and (ii) what is the failure threshold, i.e., at what "value" measured by the chosen degradation measure will result in a failure.

For the first problem we may use both ML methods and more classical models either physic based models with random loads and stochastic processes. A limitation of ML methods is usually that we have very few observations where the system is run to failure, hence in general we have less trust in ML methods wrt shed light on what happen in late life of the component. To cope with this last phase physical models and stochastic processes together with stochastic modelling of the failure threshold is considered more appropriate.

The remaining part of this memo will elaborate on some relevant models used in prognostics.

PF-model

The so-called PF-model is a popular model used in predictive maintenance. Figure 1 shows a principal sketch of the failure progression (degradation level) of a component. Up to point of time P there is no indication of a failure. But then starts failure progression until the failure progression exceeds the failure limit at point of time F. The point of time P is often referred to as a *potential failure* whereas the point of time F is a real failure. The time interval between the points P and F is denoted the PF-interval. The PF-interval is treated as a stochastic variable. Given that we are at time P in 1, the PF-interval is corresponding to the RUL. More formally let T denote the length of the PF-interval. In the modelling it is now crucial to assess the probability density function, or the cumulative distribution function, $F_T(t)$. In Bane NOR the PF-interval for rail cracks were assessed and used for determination of the frequency of ultrasonic inspection of the rails. A general model with $\mathrm{E}(T) = 5$ years and $\mathrm{SD}(T) = 3$ years were used.

The PF-model as presented above will not be able to take current condition and future load into account. A reasonable simple extension of the model used in the previous section will be derived. The starting point is the failure rate function for the PF-interval, $z(t)$. If se stick to the Weibull distribution

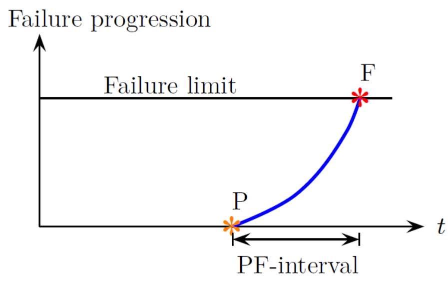

Figure 1: PF-model

the failure rate function is given by $z(t) = \alpha\lambda^{\alpha} t^{\alpha-1}$. We observe that $z(t)$ does not contain neither the current state nor the future loads. The so-called Cox-proportional hazard model is often used to incorporate the current state in the failure rate function. Let $\boldsymbol{y}$ be the vector of current relevant state information for the component, for example temperature, vibration level and so on. Next let Let $\overline{\boldsymbol{x}(t)}$ be the vector of average loads in the time period $[0,t)$. The failure rate function may be written on the form:

$$z(t|\boldsymbol{y},\overline{\boldsymbol{x}(t)}) = z_0(t)e^{\boldsymbol{\beta}_1\boldsymbol{y}}e^{\boldsymbol{\beta}_2\overline{\boldsymbol{x}(t)}} \tag{1}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are regression coefficient vectors established by for example statistical analysis of data. Statistical data analysis is beyond the scope of this lecture. $z_0(t)$ is a baseline failure rate function, typically on the form $z_0(t) = \alpha\lambda^{\alpha} t^{\alpha-1}$

Now assume that the parameters α, λ, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are all estimated by analysis of statistical data. Further assume that the current component state, $\boldsymbol{y}$, is known and that we have an estimate of future load $\overline{\boldsymbol{x}(t)}$.

It may now be shown that the cumulative distribution function is given by:

$$F_T(t|\boldsymbol{y},\overline{\boldsymbol{x}(t)}) = 1 - \exp\left(-\int_0^t z(u|\boldsymbol{y},\overline{\boldsymbol{x}(u)})\,du\right) \tag{2}$$

When we know the cumulative distribution function it is rather straight forward to optimize for example the optimal time for replacement/corrective maintenance. We might also use this model to determine if relaxing on the future loads will be a strategy when there is a long time to wait for the "maintenance window". For example in offshore wind maintenance the weather window might be closed for several months in the winter, and an operational strategy would be to shut down a damaged wind turbine when wind speed exceeds for example $10 m/s$.

Markov chain

Another popular degradation model is the Markov chain model. Figure 2 illustrates the situation.

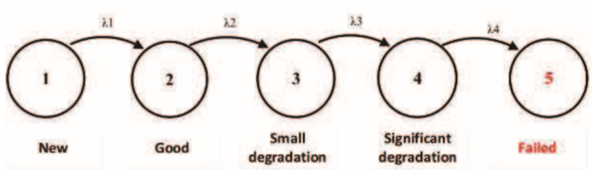

Figure 2: Markov chain model for emergency shutdown valve

A component is assumed to have five states, where state 1 corresponds to a new component, whereas the other states represents more and more degradation. State 5 is the fault state. In this model the λ's represent transition from one state to another. The times to jump from one state to another is in the Markov chain model exponentially distributed, that is the reason for the name 'Markov' after a Russian mathematician. Compared to the PF-model, we may say that the Markov chain model is a refined model where we distinguish between various states. A potential failure corresponds to the second state in the Markov model, whereas the failure state corresponds to the fifth state in the Markov model. The mathematical analysis simplifies if we can assume exponentially distributed transition times. In a more general situation we may model the transition times similarly to equation (1). So-called phase type distributions may then be used to overcome the non-Markovian behaviour of the intermediate transitions. The mathematical framework required is beyond the scope for this lecture. The usage of such a model could be multiple:

1. Determination of inspection intervals. Here the inspection interval may vary between states.

2. Determination of repair strategies, i.e., in which state is a repair required, and is it always required to repair to an "as good as new" condition

3. How long can we wait before we repair?

4. Can we relax on the operational load in order to postpone the repair activity, i.e., if we are able to link the transition rates to the operational load mathematically?

There are two main reasons for working with a discrete model for degradation. Firstly the available data will in many cases only provide discretized state information, This is the case in many oil and gas companies as well in hydro power stations. Further a Markov chain model is often more attractive to work with than the stochastic processes where the state (degradation) variable is continuous. In the next section we will, however, introduce a stochastic process where the state variable is continuous.

Continuous state stochastic processes

Figure 3 shows how may imagine that degradation evolves in time. Degradation means a change of condition (or health) from a good state to a worse state. The degradation level, here denoted $D(t)$, is a stochastic process, and the future degradation level is uncertain and can therefore represented by a stochastic variable (random quantity). This is indicated by one path up to the current time, and then we indicate different paths that may be followed. If there is a threshold, or failure limit L, the hitting time (i.e. the time the stochastic project reaches/hits the failure limit L) will also be a stochastic variable. The remaining useful life (RUL) is defined as the time from now until the failure limit is hit.

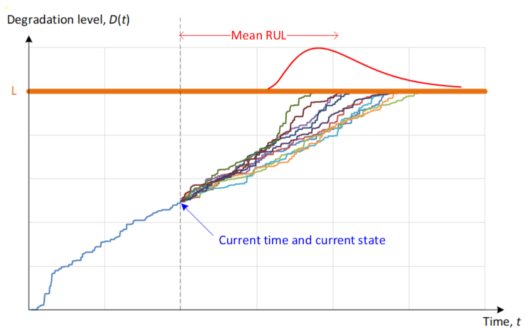

Figure 3: Continuous degradation over time

The model has similarities with the Markov chain model, but here we assume that degradation is continuous rather than following a discrete path.

Several mathematical models exist for describing the degradation. In these models we assume that it is the *increments* in the degradation level from day to day that represent the randomness in the process. We let $\Delta D(t, \Delta t)$ denote the degradation in a small time interval from t to $t+\Delta t$. This can often be understood in terms of loads that cause damage to an item and hence an increase in the degradation level. Here the increments are typically stochastic independent for different points of times

The gamma and Wiener processes are classical stochastic processes used for degradation in the maintenance models. For the stationary gamma process we assume that increments are independent and gamma distributed. For the stationary Wiener process we assume that the increments are independent and normally distributed. Thus:

$$D(t+\Delta t) = D(t) + \Delta D(t, \Delta t) \tag{3}$$

Physical arguments may be used to argue that the increments depend on the current degradation level, cf. the Paris law for fatigue cracks. A relevant stochastic process to consider in this situation is a Geometric Brownian Motion (GBM):

$$\Delta D(t, \Delta t) = \mu D(t)\Delta t + \sigma D(t)\Delta W(\Delta t) \tag{4}$$

Where $\Delta W(\Delta t)$ is normally distributed with zero mean and standard deviation Δt. μ represents the "percentage drift" and σ represent "percentage volatility".

Given that the degradation level at time t is observed to be $D(t)$ and we know the parameters μ and σ, uncertainty intervals can be obtained for the degradation level at the future time $t+s$ rather easily. To find uncertainty intervals for the time to hit the failure threshold by means of analytical methods is more complicated, but rather easy if we apply Monte Carlo methods.

The GBM model will result in an exponentially increasing degradation. This may be reasonably from physical arguments as discussed for fatigue cracks. Case studies have also indicated that the exponentially trajectory is changing at some critical values. A so-called twin-exponential model has been proposed for this situation.

In the twin-exponential model, it is proposed to use a two-stage process by combining two GBMs with different parameters. Before the first prediction time (FPT), which is the time when the process changes from slow to fast development, GBM has parameters μ_1 and σ_1. After the FPT, GBM parameters are changed to μ_2 and σ_2. The transition occurs at time $t = \phi$. The parameters $\mu_1, \sigma_1, \mu_2, \sigma_2$ and ϕ (=FPT) are unknown in the model and needs to be estimated.

Figure 4 shows the degradation level (health indicator) for pump bearing data:

The health indicator here is the so-called root mean square (RMS) of the wavelet transform of the raw vibration data. The health indicator is established by a so-called feature extraction procedure. Feature extraction is beyond the scope of this lecture.

In a study performed in the MonitorX project within the hydro power industry, this model were tested. In addition to model the degradation with the twin-exponential model, also the failure threshold L was modelled by a stochastic variable.

Around Christmas time we plan to use our vibration rig in the RAMS lab to conduct experiments. Bearings will be exposed to extreme loads (force and/or particles enforced inside the bearing) to accelerate the degradation process. The twin-exponential model is one out of many model we will test for RUL prediction with our own generated data.

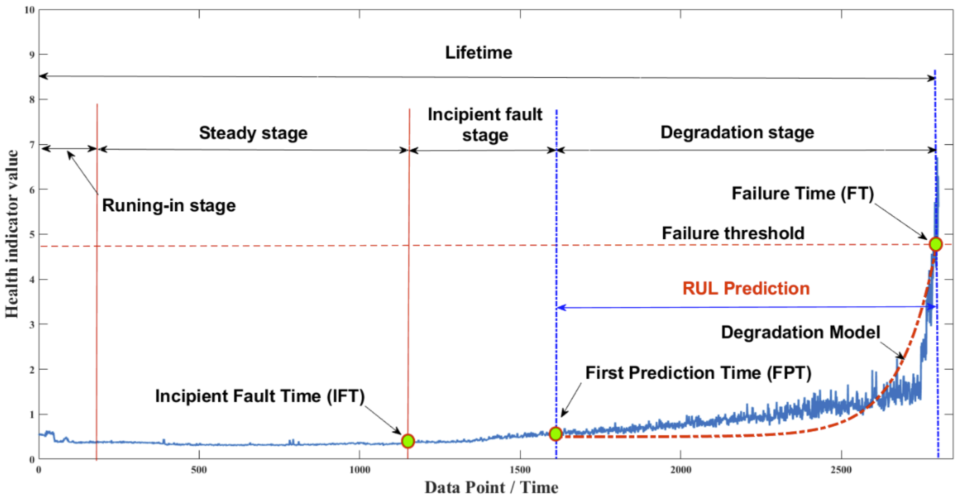

Figure 4: Twin-exponential model