Bayesian Networks for Remaining Useful Life Prediction

Erik Hostens, Kerem Eryilmaz, Merijn Vangilbergen, and Ted Ooijevaar

Flanders Make, Gaston Geenslaan 8, 3001 Heverlee, Belgium
erik.hostens@flandersmake.be
kerem.eryilmaz@flandersmake.be
merijn.vangilbergen@flandersmake.be
ted.ooijevaar@flandersmake.be

Abstract

Remaining useful life (RUL) prediction is a critical task in the field of condition-based maintenance. It is important to perform RUL prediction in a statistical sound way. However, it is not straightforward to properly combine multiple information sources about an asset, such as available statistics, measurements, derived features, and prior knowledge in the form of mathematical models and relations, including their uncertainties. Bayesian networks (BNs) are a means of graphically representing all statistical information in a comprehensible way and allow for correctly combining all information. BNs allow for inference in all directions, thereby not merely providing a RUL prediction with explicit uncertainty, but select the most informative features, diagnose which degradation mechanism is manifest if multiple mechanisms exist, provide decision support in the form of optimal condition-based maintenance points when combined with a cost model. BNs also explicitly quantify the model uncertainty arising from the scarcity of the training data. We illustrate these benefits on two real-world industrial examples: solenoids and bearings. We also provide a method to correctly include the effect of changing operating conditions.

1. Introduction

Condition-based maintenance (CBM) has gained a strong interest from the industry in recent years, both driven by the market-driven necessity of ever-increasing efficiency and sustainability of industrial systems, and the opportunity opened up by the fast growing industrial digitization and sensorization. For an extensive recent literature survey, we refer to (Quatrini, Costantino, Di Gravio, & Patriarca, 2020). The prediction of remaining useful life (RUL), which is the time at which an industrial asset will have been degraded to such extent that it can no longer perform its intended function, plays an important role in CBM to schedule maintenance, optimize operating efficiency, and above all avoid unplanned downtime.

Given the intrinsic randomness of the drivers of degradation leading up to the ultimate failure of the asset, a proper statistical treatment of these phenomena is required (Sankararaman, 2015). Indeed, because of the costly consequence of an unanticipated failure, even when its probability is small, the expected cost may become significant and requires early action. Many performance metrics for RUL prediction focus on the deviation of mean prediction from the ground truth, but the decision support for maintenance actions should rather focus on the tail of the prediction distribution.

In this paper, we aim to present a generic and systematic method using *Bayesian networks* (BNs) to incorporate all available knowledge and data in the RUL prediction. When many factors contribute to this prediction, it is a challenge to manage these relations and correctly calculate the overall statistics. BNs, although in essence nothing more than a representation of the statistics, offer a comprehensible approach. The explicit modeling and quantification of RUL prediction uncertainty has been extensively studied in literature, typically focused on a particular industrial asset, such as (Mishra, Martinsson, Rantatalo, & Goebel, 2018) for batteries and (Prakash, Narasimhan, & Pandey, 2019) for bearings.

A Bayesian network is a graphical representation of a joint distribution of a set of variables (Pearl, 1988). The joint distribution is factorized into root probabilities and conditional dependencies, which are graphically represented by a directed acyclic graph. BNs are ideal for taking an event that occurred and assessing the probability that any one of several possible known causes was the contributing factor. For example, a BN could represent the probabilistic relationships between failure mechanisms and their manifestations in the sensor data. Given sensor data, the network can be used to compute the probabilities of the presence of various failure mechanisms. This is analogous to how medical doctors need to diagnose a patient showing symptoms of disease, see for example Fig. 1. In summary, BNs are a means for graphically representing complex statistical relations that become otherwise intractable.

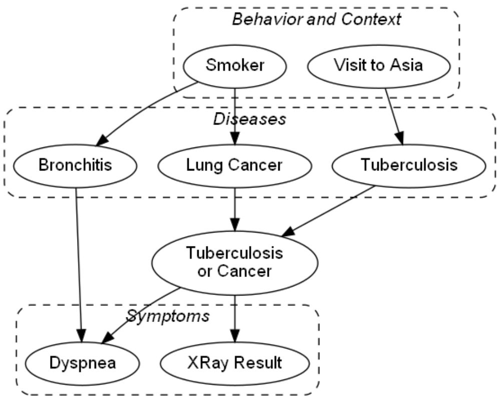

Figure 1. An example of a BN in medicine, analogous to condition-based maintenance. The intuitive (causal) understanding of how stochastic variables relate is well captured: diseases (failure mechanisms) are caused by behavior and context (settings and operating conditions) and become manifest in observed symptoms (measurements or features derived thereof). In this particular example, smoking increases the probability of bronchitis or lung cancer, whereas if a patient recently visited Asia, there is a higher probability of having contracted tuberculosis. The symptoms alone cannot distinguish tuberculosis from cancer, therefore an auxiliary variable "Tuberculosis or Cancer" is used to make this explicit.

The purpose of this paper is to show the advantages of using BNs for RUL prediction: (i) BNs naturally combine all prior knowledge, in the form of models and statistics, and data, and as such maximally exploit the available information; (ii) distributions of unknown variables can be inferred from observed variables, in all possible directions, depending on the application (parameter estimation, diagnostics, prognostics, decision support); (iii) explicit model uncertainty, which can be used to assess the (in)sufficiency of the available training data; (iv) BNs are easily extended with more variables that affect RUL and its assessment, such as operating conditions. We will illustrate how BNs are used on two examples of assets widely used in industry, a solenoid-operated valve (SOV) and a bearing, on which we conducted accelerated life testing.

This paper is organized as follows. In Section 2, we introduce the two industrial assets, SOVs and bearings, on which we validated the BN method for RUL prediction. In Section 3, we explain the methodology of BNs for RUL prediction and show how they provide the aforementioned advantages. In particular, we explain (i) how the model is built, trained and used, (ii) how a cost model can be integrated in the BN to provide *decision support for maintenance*, (iii) how *model uncertainty* is taken into account in the RUL prediction, and (iv) how the BN is adjusted to incorporate *operating conditions*. In Section 4, we illustrate the methodology by the application on the two industrial assets in four case studies, corresponding to the topics (i)-(iv). Finally, we formulate some conclusions and future directions of research in Section 5.

2. The application cases

2.1. Solenoid-operated valves

We have conducted our research on a historic dataset of accelerated life tests (ALT) on a set of 3/2-way normally closed alternating current powered solenoid operated valves (SOV). The SOVs were subjected to on-off switch cycles until failure or end-of-life (EOL), defined as the moment that the solenoid's magnetic force is insufficient to overcome the friction and move the plunger. This moment is observed both in the current signal, as the solenoid then behaves as a fixed nonlinear inductance, and in the thermal mass flow detecting leakage measured at the outlet ports and blow-off holes of the valves. The experimental dataset has been used before in (Tod et al., 2019; Mazaev, Ompusunggu, Tod, Crevecoeur, & Van Hoecke, 2020), where full details can be found.

In previous work, we have defined a number of features on the current signal and quantitatively assessed their feature performance for health monitoring quantitatively. For a detailed description of these features, we refer to (Ompusunggu & Hostens, 2021, 2023). For our purposes, it suffices to know that some features can be extracted from the current signal that contain information on the state of health of the SOV. Without loss of generality, we do not distinguish between direct measurement or derived features, we call them both *measurements*. Fig. 2 shows an important measurement *time-to-hit*, defined as the time between start of induced current in the solenoid at the beginning of the cycle and the plunger hitting the end of the shaft stopping its movement. In the induced increasing current signal, this stopping of the plunger is seen as a small dip in the first period. The time-to-hit increases when the SOV degrades, as it is related to the friction between plunger and shaft and therefore indicative of health. Fig. 3 shows the evolution of time-to-hit for 10 SOVs as a function of the number of past on-off cycles, together with the corresponding EOLs, except for two solenoids that did not fail before the end of the ALT. Note the strong increase of time-to-hit approaching the EOL.

2.2. Bearings

The SOVs were all tested under the same operating conditions. In order to illustrate the BN methodology for RUL prediction under *varying operating conditions*, we reused datasets of ALTs that we have conducted on bearings (Geurts, Eryilmaz, & Ooijevaar, 2023). These were generated using the Flanders Make Smart Maintenance living lab, an open test and development platform that aims to support the adoption

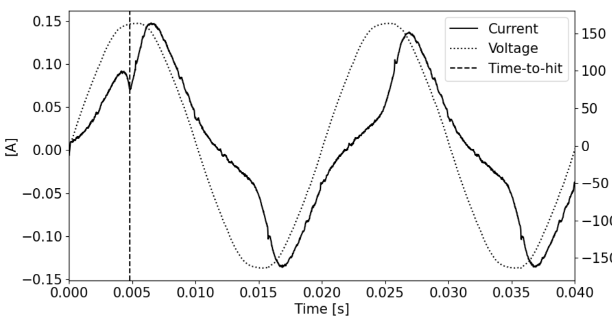

Figure 2. Definition of time-to-hit measurement in the SOV's alternating current signal.

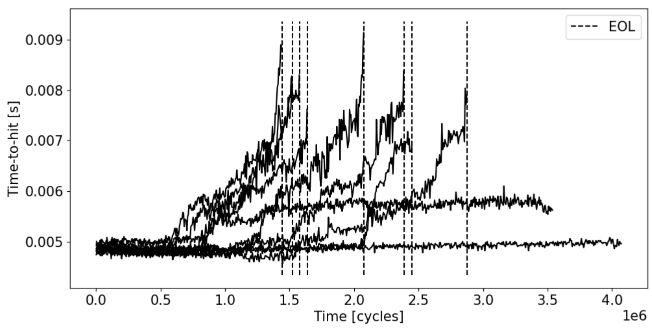

Figure 3. Evolution of time-to-hit for 10 SOVs. Two solenoids did not reach their EOL during the ALT.

of condition monitoring technologies in industry (Ooijevaar, Di, et al., 2019). Details of the setup and the tests can be found in the given references, we restrict ourselves here to the information required for a self-contained comprehension of this paper.

Before the start of each ALT, a small initial indentation was created in the bearing inner race in a repeatable manner. This serves as a local stress riser emulating a local plastic deformation caused by, for instance, a contamination particle. The EOL of a bearing is defined as the moment where the measured vibrations exceed a peak-to-peak acceleration of about 200 m/s^2. One set of bearings was subjected to stationary operating conditions, being a radial load of 9 kN and a rotary speed of 2000 rpm, another set of bearings to the same radial load but a varying speed going from 1000 to 2000 rpm in a cyclic saw-tooth pattern with a period of about 10 minutes, as shown in Fig. 4. Note how the acceleration depends on the speed, and how it increases exponentially with time near the EOL, similarly as the time-to-hit for the SOVs.

3. Methodology

3.1. Building the BN

The BN defines the joint distribution of all considered random variables X_i as a product of the individual density functions, conditional on their parent variables $\mathrm{pa}(i)$, i.e. the variables

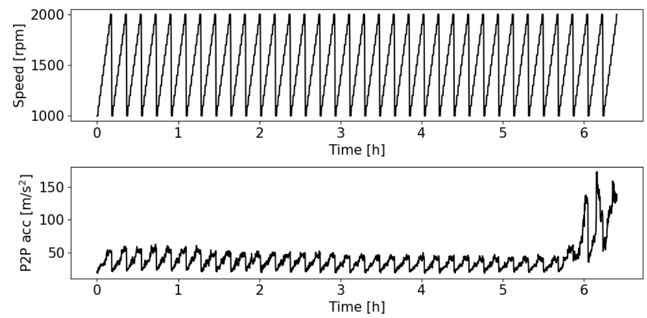

Figure 4. Speed and peak-to-peak acceleration of a bearing subjected to varying rotary speed.

that point to variable X_i in the graph representation:

$$p(X) = \prod_i p\left(X_i | X_{\mathrm{pa}(i)}\right). \tag{1}$$

We show how BNs naturally combine all available information for RUL prediction. The information we consider comprises:

- lifetime statistics,
- a degradation model,
- measurements revealing the underlying level of degradation.

A lot of research has been spent to each of these elements of information, either generic or specific to the considered asset. It is the sole purpose of this paper to show how these are *combined*, so we make a few simple assumptions, that sufficiently fit our example.

- In the following, we will refer to *time* not in a literal sense, but rather expressed in a unit that naturally relates to the usage of an asset. For solenoids, it is the number of on-off cycles; for bearings, it is the number of rotations. Similarly, lifetime and RUL are expressed in the same unit.
- For lifetime statistics, we assume a Weibull distribution, that corresponds to a failure rate that is proportional to a power of time (Jiang & Murthy, 2011). On top of that, in cases where the asset has not yet failed, we know that lifetime is greater than the current time.
- As a degradation model, we assume a hidden dimensionless degradation state, where the rate of degradation is proportional to the level of degradation itself. This simple first order dynamics boils down to an exponentially increasing degradation, or equivalently an exponentially decreasing health, which intuitively corresponds to the well-known P-F curve (Nowlan & Heap, 1978). The relation between health and measurements is a function that we will preferably describe with only a few parameters so as to keep complexity low, but the method allows any function fitting algorithm, including neural net-

works. Here we will adopt simple linear relations with normally distributed random noise, motivated by the exponential decay of health and the measurement evolution plots shown in Figs. 3-4.

These relations are captured in the generic BN structure of Fig. 5. Some of the variables are shown in rectangles, to in-

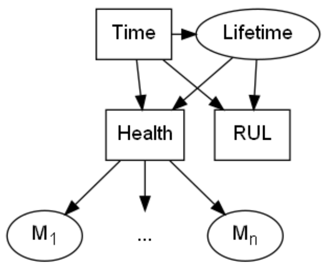

Figure 5. The generic BN structure for RUL prediction: time is given, and is a lower bound for the lifetime. Their difference is the RUL, and also defines the hidden health state. Measurements $M_1, \ldots, M_n$ are function of health.

dicate that they are deterministic, either always observed or a deterministic function of their parent variables. This simple model is captured in the following explicit relations:

$$\begin{aligned} L &\sim \text{Weibull}\left(k, \lambda\right), \\ L &> T, \\ \text{RUL} &= L - T, \\ H &= 1 - \exp\left(\frac{T-L}{D}\right), \\ M_i &\sim \mathcal{N}\left(M_{i0}(1-H) + M_{i1}H, \sigma_i^2\right), \end{aligned} \tag{2}$$

where L denotes the lifetime, T time, and H health. The latter starts very close to 1 (at $T = 0$) and ends on 0 (at EOL, or $T = L$). This BN model leverages on expert knowledge captured in simple relations between the variables and is therefore capable of describing those relations using only a few parameters, as opposed to using e.g. neural network models that easily have hundreds of free parameters. The free parameters are in this case: the Weibull distribution shape k and scale λ, the degradation time D, which approximately corresponds to the time between onset of degradation and end-of-life, and for each measurement M_i the spread σ_i and the linear coefficients M_{i0}, M_{i1}, which are the mean measurement values at start and at EOL, respectively. One could wonder why health H is made a deterministic variable, and not a random variable. We motivate this by the fact that by defining lifetime L and the measurements M_i as random variables, all *real* stochasticity is already captured. Indeed, as H is never observed directly, it can be considered merely as an auxiliary variable linking the M_i with L and T. Its actual value is of no importance, unless it would have an impact on the (observed) performance, but here we only consider the EOL and the prediction thereof.

3.2. Training and prediction

The purpose of *training* the BN model is to fit the parameter values to the data from the ALTs. We can then use this model to *predict* RUL for new data of another asset. To this end, we have used PyMC, a probabilistic programming library for Python that allows users to build Bayesian models and fit them using *Markov chain Monte Carlo (MCMC)* methods (Patil, Huard, & Fonnesbeck, 2010). Essential to PyMC is that there is no distinction between parameters and variables, there are only (random) variables, including the parameters. This enables *Bayesian hierarchical modeling*, a type of Bayesian modeling where information is available on different levels (Allenby & Rossi, 2006). In our case, we assume a single set of parameters k, λ, M_{i0}, M_{i1}, D, σ_i for the entire population, but we have a different L for each asset, and H, RUL, M_i are different for each asset and each time T.

MCMC is used for both training and prediction, they only differ in which variables are observed and which not. This is graphically explained in Fig. 6, showing all variables including the parameters (but only one measurement, to not overload the picture), and marking their being observed as gray shading. In training, the parameters are unknown and are fit-

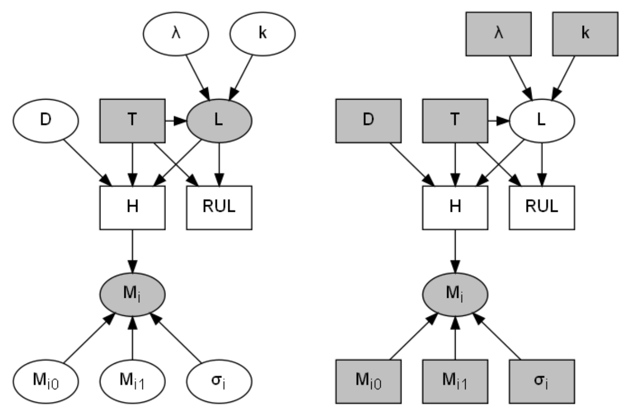

Figure 6. The BNs for training (left) and prediction (right). Gray shading represents known/observed variables, whereas white represents unknown/unobserved variables.

ted to the data on lifetime and measurements. In prediction, the parameters are known, but the unknown lifetime distribution is to be inferred from the lifetime statistics and the measurements.

The MCMC algorithm updates the distributions of all unobserved random variables given their prior distributions and the data, through the likelihood of the data. The prior distributions, also called prior *beliefs*, quantify the uncertainty in the prior knowledge *before* data is acquired. It is based on previous data campaigns or *physical* and *statistical* knowledge of the asset's behavior and degradation. For instance, the shape parameter k is related to the trend of the failure rate, typically going up ($k > 1$) as the asset ages. If little is known beforehand, prior beliefs should be chosen sufficiently wide, so-called *weakly informed priors*. The updating of prior beliefs into posterior distributions when new data comes in is the central paradigm of Bayesian statistics. Therefore, it is important to note that the BN for prediction in Fig. 6 is an oversimplification: the parameters do not become exactly known by the training, but if their posterior distributions become sufficiently narrow, their low remaining uncertainties can be ignored in their contribution to the total uncertainty of the RUL prediction.

Another important nuance to make about Fig. 6 is the fact that, as we saw in Section 2, some of the assets' EOL is never observed, simply because their ALT is stopped early. So not all L-nodes in the left graph (training) of Fig. 6 should be gray shaded (observed). Such *censoring* is quite common in statistical analysis of survival data in medicine (Kalbfleisch & Prentice, 2011). Because some of the lifetimes L are not observed, their values cannot be directly used to infer the Weibull parameters. This is exemplary for why the BN framework is powerful: although L is not observed, the uncertainty on its unknown value can nonetheless be significantly reduced through its relation with the measurements and the time during which it did not fail. Therefore this information still contributes to the fitting of the Weibull parameters. The BN truly leverages on all the available information combined.

3.3. Decision support for maintenance scheduling

We explain how the BN for prediction, shown in Fig. 6, is adjusted to provide decision support for maintenance scheduling. We only consider asset replacement as the maintenance action, but an analogous reasoning can be followed for other maintenance actions. The BN prediction yields a probability distribution of RUL. This is more useful than a single RUL expected value, because it allows to better balance the risk of unanticipated failure with the economic loss of early replacement. The BN allows the integration of a cost model and evaluate the prediction of cost given the replacement scheduling strategy. We illustrate this with a simple cost model. We assume the asset's cost C_A in the normal situation. This is for instance the sum of the costs of purchase and installation, the latter coinciding with the scheduled replacement of its used predecessor. If the asset fails before its scheduled replacement, there will be an extra *cost of failure* C_F. This cost is very dependent on the application: it can be very high for high impact failures, such as significant production loss or damage to other equipment, but it can also be low or even zero. In that case, the asset should only be replaced after EOL.

Another parameter we assume in this example is the *prediction horizon* T_{PH}, defined as the time required to schedule the replacement up front, for instance because it takes some time to send a maintenance engineer to the asset's remote location. If it is decided at time T to replace the asset, the actual replacement can take place at time $T+\Delta T$, where $\Delta T \geq T_{\mathrm{PH}}$. This is schematically depicted in Fig. 7.

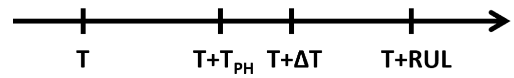

Figure 7. Replacement is scheduled at time T, and takes place at time $T+\Delta T$, which is $T+T_{\mathrm{PH}}$ at the earliest. $T+\Delta T$ should precede $T+\mathrm{RUL}$ to avoid the cost of failure.

Let $T+\Delta T$ be the asset's *replacement time* in the future, decided at the *scheduling time* T hence depending on the RUL prediction at time T. The resulting total cost depends on the actual RUL:

$$\begin{aligned} \text{total cost} &= C_A && \text{if RUL} \geq \Delta T, \text{ or} \\ &= C_A + C_F && \text{if RUL} < \Delta T. \end{aligned} \tag{3}$$

To balance the extra cost C_F with the cost of early replacement, where more assets are used in the long run, we have to evaluate the *cost per used time unit* C_T:

$$\begin{aligned} C_T &= \frac{C_A}{T+\Delta T} && \text{if RUL} \geq \Delta T, \text{ or} \\ &= \frac{C_A + C_F}{T+\mathrm{RUL}} && \text{if RUL} < \Delta T. \end{aligned} \tag{4}$$

Note that C_T is a deterministic function of other variables. We include it in the BN for prediction, shown in Fig. 8. Alongside the RUL prediction at scheduling time T, we can use this model to calculate the distribution of the cost $C_T(\Delta T)$ corresponding to the replacement time $T+\Delta T$, for multiple values of ΔT. Replacement should then be scheduled as soon as the

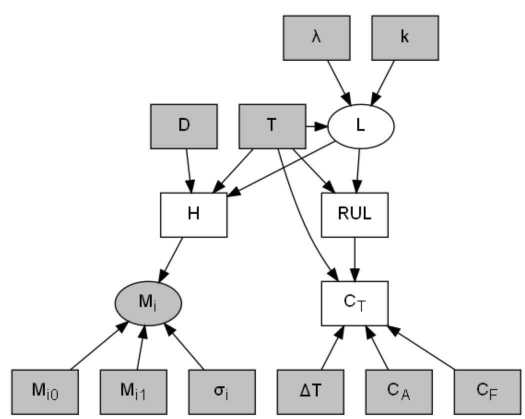

Figure 8. The BN for prediction, including the simple cost model.

expected value of this cost $E[C_T(\Delta T)]$ reaches a minimum for $\Delta T = T_{\mathrm{PH}}$, or if:

$$E[C_T(\Delta T)] > E[C_T(T_{\mathrm{PH}})], \quad \forall \Delta T : \Delta T > T_{\mathrm{PH}}. \tag{5}$$

3.4. Model uncertainty

As we mentioned in Section 3.2, the training of the BN will result in a posterior distribution of all non-observed random variables, including the model parameters. If these parameter posteriors are sufficiently narrow, we may consider the remaining uncertainty insignificant and select the means of the posteriors as fixed known parameters for the prediction BN, as was shown in Fig. 6. However, if training data is scarce, the remaining posterior uncertainty cannot be ignored and should be included in the prediction. This is achieved by making the parameter nodes in the prediction BN stochastic and unobserved, and using the posterior distribution after training as its prior. It is important to note that, in many cases, the parameter distributions will be mutually *dependent* after training. Therefore, one should use a single joint prior distribution for the parameters in the prediction BN.

MCMC does not output explicit posterior distributions, but a sample thereof, due to the way it works. To include it in the prediction, there are two options: either combine training and prediction in one MCMC run, or approximate the posterior parameter distribution. The former option is the most correct, since in this way we are combining all information at once, both of the past ALTs and the running one. However, this requires a lot of calculations and all data need to be kept, so this approach may become cumbersome. The latter option is most practical, since there is only one run of MCMC involving the training data, after which they are not longer needed. For approximating the posterior joint distribution of the parameters, in most cases a multivariate normal distribution is suited, motivated by the fact that, if the model is well designed, it is expected that the parameter estimates will converge. To this end, PyMC also supports *automatic differentiation variational inference (ADVI)* as an alternative to MCMC followed by the approximation of the posterior distribution from the sample. ADVI turns this around by up front assuming a parameterized approximation of the posterior distribution and reformulating its calculation as an optimization problem (Kucukelbir, Tran, Ranganath, Gelman, & Blei, 2017).

3.5. Varying operating conditions

In our original BN model of Section 3.1, we assumed the asset's degradation as the sole driver of further degradation. This works fine if other influences do not have a significant contribution to degradation. However, in most cases, the *operating conditions* (OC) do have a strong impact on degradation, and should be taken into account. Secondly, the OC also influence the measurements. This is clearly seen in Fig. 4 for the bearings. This influence further complicates the analysis, as the measurements serve as indicators for degradation, so we need to distinguish whether changes in the measurement are resulting either by changed OC, or by degrading health, or both.

Inquiring the effect of OC on degradation is particularly difficult, since the OC consist of multiple variables that often have a combined effect where one OC variable strengthens or weakens the effect of another. In such cases, on the one hand a detailed understanding is needed of how the asset's health evolves under given OC, in the form of engineering laws or physical models. On the other hand, a sufficient amount of ALT data is required to validate and quantify these models. However, ALT data are typically scarcely available because they are costly to generate. Again, maximally leveraging on all available knowledge and data is key. We show how the original BN for RUL prediction is adjusted such as to account for OC.

Let us first address the simpler case of *stationary* operating conditions. When the OC are stationary over the entire lifetime, even when they are different for different assets, training and using a BN for prediction of RUL is not a lot more complicated than before. Given sufficient ALT data for each OC, one can simply retrain another BN for each different OC. Of course it is more useful to leverage on knowledge of how OC affect lifetime, ideally in the form of a relation with the OC adding few free parameters, such as the empirical basic rating life model for bearings of (ISO281, 2007). Such knowledge is integrated in the BN by adding a relation from the OC to the lifetime and the measurements. The generic BN for RUL prediction, previously shown in Fig. 5, is then updated to the BN structure of Fig. 9.

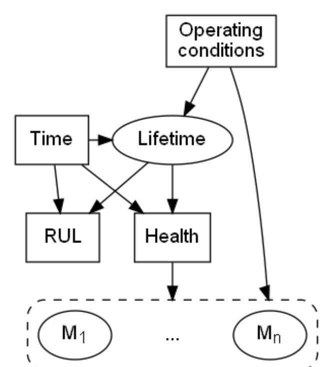

Figure 9. The generic BN structure for RUL prediction under stationary operating conditions.

For *varying* operating conditions, the relation between OC and lifetime is more complicated. One could interpret the relation between OC and lifetime in Fig. 9 as the *aggregated effect* of OC on lifetime, but as such the combined effect of the OC with health is overlooked. For instance, a higher load might have a larger damaging effect if the asset was already in a degraded state. Therefore, the OC effect should be aggregated in such a way that it takes that aspect into account, which it does not in the BN structure of Fig. 9.

We resolve this issue by assuming *nominal* parameters for a single nominal OC defined up front. We relate all other OC to this nominal OC, and express lifetime and RUL as their equivalent lifetime and RUL under the nominal OC. To deal with varying OC, we locally *compress* and *stretch* time into an equivalent time under the nominal OC. This idea is shown in the BN structure of Fig. 10. The variables Health, Lifetime and RUL are all expressed in the *Equivalent Time* corresponding to nominal OC. For the nominal OC, Equivalent Time and Time progress at the same rate. Note that Equivalent Time is defined as a stochastic variable (ellipse), because the relation between OC, Time and Equivalent Time might be uncertain. This BN structure also includes the immediate effect of OC on the measurements. The RUL that this model predicts is expressed as the equivalent RUL under nominal OC. Decision support should take the *expected future* OC into account for predicting the actual RUL. Note that such BN can also be used to recommend to change the future OC, if the application allows it, in order to delay potential failure.

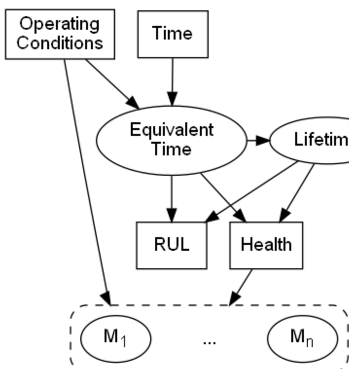

Figure 10. The generic BN structure for RUL prediction under varying operating conditions.

4. Application on case studies

In this section, we illustrate the methodology by the application on either SOVs or bearings in four case studies, corresponding to the topics explained in the previous section: (i) training and prediction, (ii) decision support for maintenance, (iii) model uncertainty, and (iv) varying operating conditions.

4.1. BNs for RUL prediction on SOV

We have trained the BN model of Fig. 6 on the ALT data of the 10 SOVs, whose time-to-hit measurement evolution was shown in Fig. 3, with 2 ALTs censored. For simplicity, we have only incorporated the time-to-hit measurement. Including other measurements would only reduce the prediction uncertainty, although not significantly since time-to-hit is the most informative on the hidden health. The BN automatically weighs the measurement contributions according to the amount of information they provide on the health. As such, it is an implicit form of *feature extraction*. We then used the fitted model parameters to predict the RUL of an SOV not used in the training, over its entire lifetime. The prediction yields a distribution of RUL, the evolution of which is shown in Fig. 11, compared to the RUL ground truth. Note the sud-

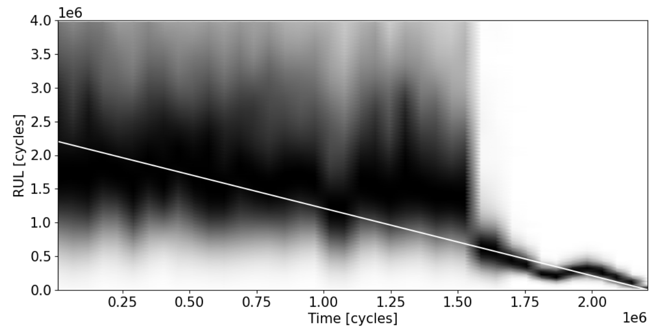

Figure 11. RUL prediction for one SOV (probability density in grayscale) compared to the ground truth RUL (white line).

den decrease of uncertainty around 1.5 million cycles. This decrease is due to the fact that at that point, the time-to-hit measurement starts increasing, thus providing crucial information on the imminent EOL. Before, the measurement reveals little on the SOV's health, so the prediction is mainly based on the Weibull statistics, truncated at the current cycle. This transition is naturally taken care of by the BN because it combines all information sources available and automatically weighs their uncertainties in the statistical posterior, as opposed to an explicit switching such as the one used in (Geurts et al., 2023).

The RUL prediction shown in Fig. 11 is based on the last measured time-to-hit at the present cycle. However, the time-to-hit measurement itself displays stochastic fluctuations, as can be clearly seen in Fig. 3. Therefore, it makes sense to include the *full history* of the measurement in the RUL prediction, so that this inherent stochasticity is filtered. Yet it is important to note that the fluctuations are not white noise, rather colored noise, which means the measurements are correlated over time. M_i should then no longer be defined as separate univariate normal random variables like in Eq. 2, but as a single multivariate normal random vector with the same mean and the measurement autocovariance as covariance. The resulting prediction ($5\% - 95\%$ quantiles near the EOL) is shown in Fig. 12, for both the last measurement only and the full history prediction, illustrating the advantage of the latter: it is more consistent and accurate compared to the true RUL.

4.2. BNs for maintenance decision support on SOV

For the SOV example of Fig. 11, we show a detail of the RUL prediction approaching EOL and the corresponding expected relative cost $E[C_T(T_{\mathrm{PH}})]/C_A$ in Fig. 13, for arbitrary cost model parameters $C_F/C_A = 10$ and $T_{\mathrm{PH}} = 2e4$. Note that

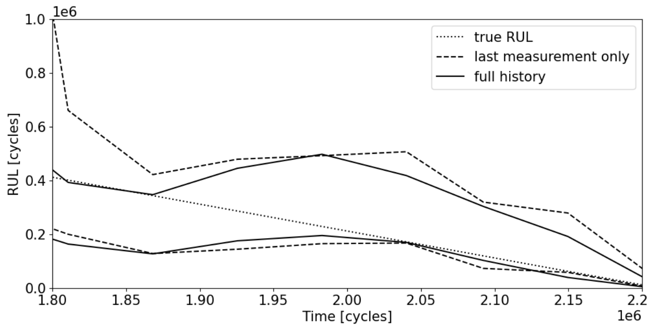

Figure 12. $5\%-95\%$ quantiles for RUL predictions from only the last measurement or from the full measurement history.

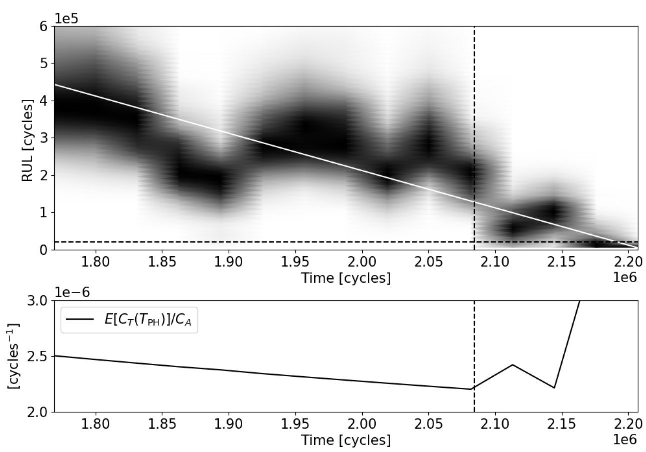

Figure 13. Top: RUL prediction. The dashed horizontal line marks the prediction horizon T_{PH}. Bottom: expected relative cost of replacement time $T + T_{\mathrm{PH}}$ at scheduling time T. The dashed vertical line marks the optimal scheduling time T^*.

at the optimal scheduling time T^*, the RUL prediction distribution still has the most part *above* the T_{PH} line. The optimal maintenance scheduling strategy therefore involves probing the *tail* of the RUL prediction distribution, which emphasizes the importance of correctly calculating this distribution.

4.3. Model uncertainty in BNs on SOV

We have redone the prediction of the SOV of Fig. 11, now using 40 SOVs in the training set instead of only 10. Both are compared in Fig. 14 through their $5\% - 95\%$ quantiles. In the RUL predictions, we have now included the posterior parameter uncertainty after training. Clearly, more training data results in a more accurate RUL prediction. It can be seen that this effect is most manifest in the healthy phase of the SOV, where the RUL prediction is mainly based on the EOL statistics and not on the measurement. This is to be expected from a statistical perspective. However, approaching the EOL, where the prediction accuracy is more important for optimal maintenance scheduling, both predictions become very close. This illustrates the power of BNs for RUL prediction, as in this case it suffices to have a training set of only 10 ALTs, two of which are censored, and a simple degradation and measurement model.

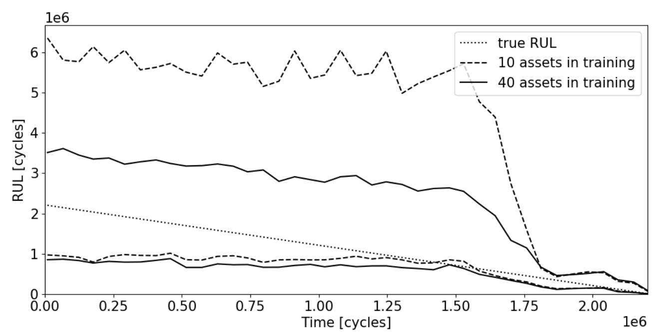

Figure 14. $5\%-95\%$ quantiles for RUL predictions including parameter uncertainty, where the BN parameters are trained either on 10, or on 40 SOVs.

4.4. BNs for varying OC on bearings

The bearing ALT dataset introduced in Section 2.2 is insufficiently rich to validate the proposed BN for RUL prediction under varying OC of Fig. 10. Indeed, because of the very uniform saw-tooth pattern of speed (Fig. 4) in the varying speed ALT, its long-term influence on degradation effectively corresponds to a stationary OC, albeit different from the stationary speed ALT. As a consequence, we have only two different long-term aggregated OC. We therefore use a combination of the BN structures of Figs. 9-10 for stationary OC and varying OC, respectively: we assume the long-term effect of OC on lifetime as equivalent to stationary, but the immediate effect of OC on the P2P measurement as varying. The resulting BN for training, now including the parameters, is shown in Fig. 15. There are two extra variables SET and

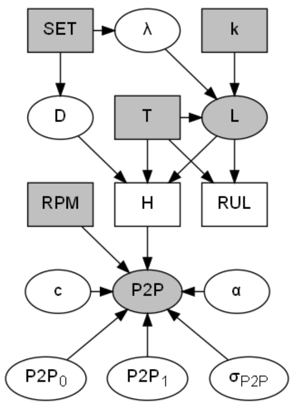

Figure 15. The BN for training of bearing RUL prediction, either under stationary or under varying RPM.

RPM, that relate to the aggregated effect of speed on lifetime and to the immediate effect of speed on the P2P acceleration measurement, respectively. The binary variable SET defines to which dataset the bearing belongs, either subjected to stationary speed ALT or to varying speed ALT. This variable essentially selects either one of two values for the parameters D and λ. The shape parameter k was fixed up front to a value of 1.3 building on historical knowledge on bearing fatigue lifetime statistics (NSWC, 2011). Extending Eq. 2 defining the relations between all variables, we define the distribution of the P2P measurement as:

$$\mathrm{P2P_{nom}} \sim \mathcal{N}\left(\mathrm{P2P}_0(1-H)+\mathrm{P2P}_1 H, \sigma^2_{\mathrm{P2P}}\right), \quad (6)$$

$$\mathrm{P2P} = c\,\mathrm{P2P_{nom}}\,\mathrm{RPM}^{\alpha}. \quad (7)$$

The expression in Eq. 7 with parameters c and α was established through a qualitative inspection of P2P data, both in healthy and degrading state as shown in Fig. 4, by comparing P2P values to the corresponding nominal P2P values around the nearest time where the speed is 2000 rpm. All 7 parameters of the BN model are simultaneously fitted to the training data.

We have trained this BN model on a set of 48 ALT, of which 7 were subjected to the varying speed profile. Bearings to validate the resulting RUL prediction were left out of the training data. As a benchmark, we also trained and validated the original model of Fig. 6 on the same data. An example of the resulting RUL prediction for both models on the same varying speed bearing is shown in Fig. 16. The RUL prediction with the original model is clearly disturbed by the varying conditions, emphasizing the need for including them. The same problem was manifest in the work of (Geurts et al., 2023).

A single asset's prediction may illustrate the added value, yet a proper comparison should be built on adequate RUL prediction performance metrics. A thorough overview and analysis of metrics is given in (Saxena, Celaya, Saha, Saha, & Goebel, 2010). To keep things simple, we have compared the benchmark model and the varying OC model by the log-likelihood evaluated at various relative locations in the lifetime and averaged over the varying speed bearings, as shown in Fig. 17. The log-likelihood is a straightforward generalization of mean squared error (MSE) that also covers the uncertainty of the prediction. The evaluation at multiple relative locations in the lifetime, similar to the alpha-lambda performance metric, addresses the application-specific prediction horizon as explained in Section 3.3.

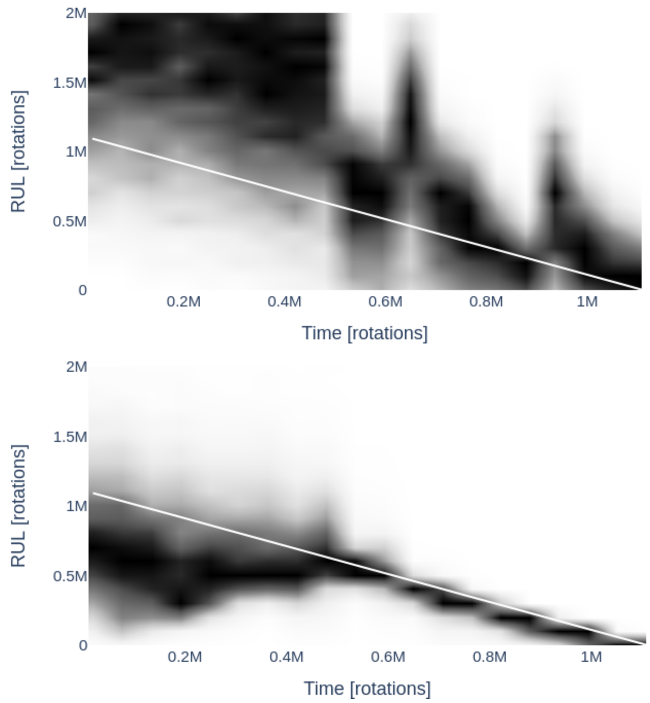

Figure 16. The RUL prediction for a varying speed bearing, both for the benchmark model (top) and the new model that includes the effect of varying OC (bottom), compared to the true RUL (white line).

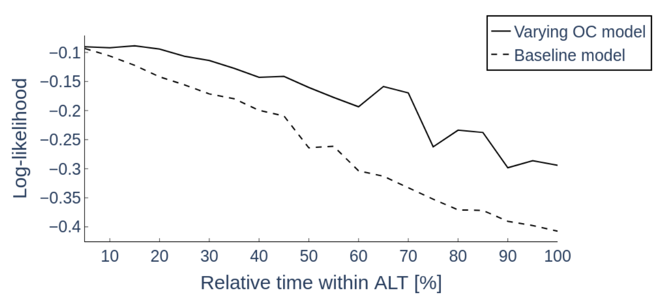

Figure 17. The average log-likelihood for the benchmark model and the new model, as function of relative time within the ALT.

5. Conclusion

We have shown a method for building, training and using Bayesian networks for RUL prediction. Next to the advantage of its comprehensibility, even when many factors contribute to the prediction, we have focused on the extension of RUL prediction with decision support for maintenance and the explicit inclusion of model uncertainty arising from the scarcity of training data. We have shown how the BN is adjusted to allow for RUL prediction under varying operating conditions.

This work is part of a larger study on the application of BNs for CBM and for maintaining quality in industry. We see the following open challenges and future research topics:

- Our current ALT datasets on SOVs and bearings do not allow for a proper validation of the generic method of Section 3.5. To this end, we are currently conducting a new ALT data campaign on SOVs under varying operation conditions.
- The inclusion of model uncertainty and its propagation to the RUL prediction is still lacking a quantified decision support for further data campaigns and design-of-experiments (DoE). We will investigate a practical method to assess the need for more training data and DoE, for instance through a criterion on the trend of a suitable performance metric such as the average leave-one-out log-probability.
- Investigate more complex degradation mechanisms, arising from multiple root causes that have different degradation dynamics.
- Instead of focusing on RUL which assumes the asset's quality as a binary variable and the EOL as a specific moment in time, we will shift towards the prognostics of a more nuanced application-oriented quality condition and corresponding decision support, such as condition-aware control.

Acknowledgment

This work has been carried out within the framework of Flanders Make's Strategic Basic Research project QUASIMO (Quality via a System Intelligence Methodology). Flanders Make is the Flemish strategic research centre for the manufacturing industry.

References

Allenby, G. M., & Rossi, P. E. (2006). Hierarchical bayes models. *The handbook of marketing research: Uses, misuses, and future advances*, 418–440.

Geurts, K., Eryilmaz, K., & Ooijevaar, T. (2023). A sequential hybrid method for full lifetime remaining useful life prediction of bearings in rotating machinery. In *Annual conference of the phm society* (Vol. 15).

ISO281. (2007). *Rolling bearings-dynamic load ratings and rating life.* ISO: International Organization for Standardization.

Jiang, R., & Murthy, D. (2011). A study of weibull shape parameter: Properties and significance. *Reliability Engineering & System Safety*, *96*(12), 1619–1626.

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The statistical analysis of failure time data*. John Wiley & Sons.

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. (2017). Automatic differentiation variational inference. *Journal of machine learning research*, *18*(14), 1–45.

Mazaev, G., Ompusunggu, A. P., Tod, G., Crevecoeur, G., & Van Hoecke, S. (2020). Data-driven prognostics of alternating current solenoid valves. In *2020 prognostics and health management conference (phm-besançon)* (pp. 109–115).

Mishra, M., Martinsson, J., Rantatalo, M., & Goebel, K. (2018). Bayesian hierarchical model-based prognostics for lithium-ion batteries. *Reliability Engineering & System Safety*, *172*, 25–35.

Nowlan, F. S., & Heap, H. F. (1978). Reliability-centered maintenance.

NSWC. (2011). *Handbook of reliability prediction procedures for mechanical equipment*. Naval Surface Warfare Center West Bethesda, MD.

Ompusunggu, A. P., & Hostens, E. (2021). Physics-inspired feature engineering for condition monitoring of alternating current-powered solenoid-operated valves. In *International conference on maintenance, condition monitoring and diagnostics* (pp. 139–151).

Ompusunggu, A. P., & Hostens, E. (2023). Quantitative evaluation of electric features for health monitoring and assessment of ac-powered solenoid operated valves. *IFAC-PapersOnLine*, *56*(2), 3725–3731.

Ooijevaar, P. K., Ted and, Di, Y., et al. (2019). Smart machine maintenance enabled by a condition monitoring living lab. In *8th ifac symposium on mechatronic systems (mechatronics 2019) and the 11th ifac symposium on nonlinear control systems (nolcos 2019)* (Vol. 52). Elsevier.

Patil, A., Huard, D., & Fonnesbeck, C. J. (2010). Pymc: Bayesian stochastic modelling in python. *Journal of statistical software*, *35*(4), 1.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann.

Prakash, G., Narasimhan, S., & Pandey, M. D. (2019). A probabilistic approach to remaining useful life prediction of rolling element bearings. *Structural health monitoring*, *18*(2), 466–485.

Quatrini, E., Costantino, F., Di Gravio, G., & Patriarca, R. (2020). Condition-based maintenance—an extensive literature review. *Machines*, *8*(2), 31.

Sankararaman, S. (2015). Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction. *Mechanical Systems and Signal Processing*, *52-53*, 228-247.

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and health management*, *1*(1), 4–23.

Tod, G., Mazaev, G., Eryilmaz, K., Ompusunggu, A. P., Hostens, E., & Van Hoecke, S. (2019). A convolutional neural network aided physical model improvement for ac solenoid valves diagnosis. In *2019 prognostics and system health management conference (phm-paris)* (pp. 223–227).