Early life failures and services of industrial asset fleets

Arinan Dourado, Felipe A.C. Viana*

University of Central Florida, Orlando, FL 12309, United States

ARTICLE INFO

Keywords:
Fleet management
Fleet unreliability
Bayesian networks
Reliability
Prognosis

ABSTRACT

In the service market targeting fleets of industrial assets (e.g., aircraft, jet engines, wind turbines, etc.), original equipment manufacturers and service providers compete with one another through offers covering day-to-day service as well as major maintenance and repairs over. Since decision-making is highly guided by reliability models, it is safe to say that services profitability depends on the ability to understand the complex stochastic interactions between operating conditions and component capability. Unfortunately, factors such as aggressive mission mixes introduced by operators, exposure to a harsh environment, inadequate maintenance, and problems with mass production can lead to large discrepancies between predicted and observed useful lives. This paper focuses on the quantification of the infant mortality impact on fleets of industrial assets. A numerical experiment is used to study how the number of failure observations and fleet size impacts the modeling of fleet reliability. Dynamic Bayesian networks implementing physics-based models are used to model fleet unreliability considering the effects of bad batch of materials. The results demonstrate that material capability, penetration of bad batch of material in the fleet, and fleet size drastically influence the model accuracy. Therefore, small fleet operators, which naturally observe a low number of failures, have to deal with larger uncertainties when quantifying infant mortality. This negatively impacts their ability to allocate resources such as inventory, labor, and account for the loss of productivity while servicing their fleet. With large fleet operators, on the other hand, large number of failure observations can cause high financial burden. Nevertheless, it also allows for reduced uncertainty in building/updating the reliability models, which can help their ability to forecast future failures and make provisions for service and maintenance. Finally, the results also show that measures such as re-commissioning of the fleet and inspection campaigns can mitigate the effects of fleet-wide early life problems.

1. Introduction

On April 1, 2011, a Boeing 737–300 operating in the Southwest Airlines flight 812 suffered rapid depressurization while cruising, leading to an emergency landing. The National Transportation Safety Board (NTSB) investigation pointed out that the incident was caused by structural failure of the fuselage skin. Evidence of manufacturing errors in the joining fuselage crown skin panels was found (see ref. [1]) accelerating metal fatigue on the panels. This was the second incident with a Boeing 737–300 in a period of two years in which a fuselage skin manufacturing issue lead to an unexpected decompression during a flight (the first was Southwest Airlines flight 2294 in 2009). These incidents are an example of how manufacturing problems can significantly reduce useful life and lead to unexpected failures in large fleets of engineering assets (e.g. airplanes, jet engines, wind turbines, etc.).

In general, unexpected premature asset failures is an issue referred to as infant mortality failures [2]. This issue is usually identified by higher than expected failure rates in early life (short period after deployment). It is a major concern among original equipment manufacturers (OEM) and operators of industrial assets. Early life failures (a) increase the total cost of ownership due to an increase in costs of maintenance, warranty, services, etc., and (b) can reduce asset performance and availability. In addition, they can impose difficulties in meeting compliance and regulations standards (as hardware degradation can be a lead cause of safety standard infringements, elevation of noise and emission levels, etc.).

Understanding how the complex interactions between operating conditions and component capability define useful life is a key issue in the services market (ability to comprehend hardware degradation and predict remaining useful life). This gives operators the chance to make decisions that directly impact their financial outcomes. For instance, Pattabhiraman et al. [3] showed by means of model updating that scheduled interval-based maintenance can be safely avoided depending on model predictions, which directly impact the cost of ownership. A similar outcome is also described by Ling et al. in [4], where...
information gain theory is used to evaluate the usefulness of aircraft component inspection which helps to decide whether the inspection is worthwhile or not. Haddad et al. in [5] and [6] discussed a cost-benefit-risk approach to manage the actions to be taken following a prognostic model. The discussion included important aspects such as overall maintenance (cost of unscheduled maintenance, collateral damage during repair, fault isolation), shortening of remaining useful life, spare parts management, etc.

As can be noticed, prognosis models are at the core of fleet management wherein the main goal is to correctly forecast failure events in order to avoid unexpected asset failure and unscheduled system maintenances. Usually this is achieved through models derived considering features of condition-based maintenance [7,8,9], in which the current degradation state of the asset is inferred based on diagnosis information and used to estimate the asset remaining useful life (RUL); or predictive maintenance [10,11,12,13], whereas physics-based degradation models [14,15,16,17] are used to forecast assets RUL and recommend possible maintenance interventions.

While an extensive review and discussion about data-driven versus physics-based modeling is not the goal of the current contribution, we believe that most practitioners would agree that:

- Data-driven models are flexible and can be easily tuned and reused. As a matter of fact, the growth in computational power has triggered interest in machine learning techniques such as support vector machines [18,19,20,21], neural networks [22,23,24] and dynamic Bayesian networks [25,26]. However, data-driven approaches present drawbacks such as the requirement of significant data sets of failure data (unavailable in some applications) and lack of clear physical interpretation of the asset degradation process (making it difficult to implement risk mitigation and remaining useful life extension).

- Physics-based models are specific to failure modes and help understanding root cause analysis as well as implementing mitigation plans. An in-depth discussion on physics-based approaches can be found on [27,28].

In this paper, we forecast remaining useful life through a prognosis model composed of two dynamic Bayesian networks: an asset level network that implements component stress-life curves employing a tunable lognormal distribution, and a fleet-level network that accounts for the integration of Palmgren-Miner’s rule. Nonetheless, readers still interested in a more deep discussion regarding data-driven and physics-based prognosis models are referred to [27,28].

When using physics-based approaches, one has to focus on quantifying uncertainty in model form, model parameters, and data. Jiang et al.[29] addressed the issue of bias correction with the systematic error being corrected by using a statistical model for the bias term (e.g., a Gaussian process) calibrated with actual experimental data, or through high-fidelity simulations. Pecherstorfer et al.[30] reviewed strategies for handling multi-fidelity models when performing computationally intensive uncertainty quantification. Asher et al.[31] and Coppeet al. [32] discussed how to calibrate important parameters in fatigue crack growth applications (initial flaw size and crack growth parameters). In real-world applications, it is also very common that models are updated while data is gathered for a particular instantiation (asset of interest). Li et al.[26] discussed the use of dynamic Bayesian networks for model updating with observed data (including loads). The updated model was used in the diagnosis and prognosis of an aircraft digital twin.

When the physics of failure is well understood and hardware degradation is linked with the operation, the physics-based prognosis can be highly effective. Unfortunately, unaccounted factors such as manufacturing problems can lead to unexpected forms of asset degradation that might not manifest in performance degradation or might not be easily captured in an inspection. In these cases, it becomes challenging to build a physics-based prognosis model and discrepancies between predicted and observed failures could prevent the models to be used. This is a very important issue in industrial applications where a large fleet of assets (hundreds to thousands of units) are constantly monitored [33,34,35].

Additionally, in a modern fleet management analysis, identifying the root problem is just the first stage. After diagnosing an issue (e.g. bad batch of materials, contamination, loss in material capability, etc.) a prognosis model should be able to provide some sort of risk mitigation analysis. The derived model should be able to take into account the impact in forecasted failures of mitigation approaches such as asset derating (i.e. recommissioning), inspection campaigns (see ref. [4]) or any other approach that can offer a compromise solution between the losses in performance as an increase in reliability. Recommissioning through derating (assigning to missions at reduced load level) is usually accompanied by a penalty in performance.

This work aims at presenting a probabilistic framework able to identify and characterize early life problems in fleets of assets. We also provide an assessment of mitigation measures (derating and inspection) on a fleet level. We will present the analysis of an infant mortality problem by means of a numerical experiment, focusing on answering the following fundamental question: **how do fleet size and the number of failures interact with each other when characterizing an infant mortality problem?** We answer this question using prognosis, uncertainty quantification, reliability and fleet management. We use a dynamic Bayesian network in which some of the nodes are physics-based as a way to forecast the remaining useful life through the progression of hardware distress by fusing design, manufacturing, and services information. The proposed framework is composed of dynamic Bayesian networks used to model fleet unreliability considering the effects of manufacturing problems (due to bad batch of materials). The proposed Bayesian models use two model parameters: material capability degradation, model as a shift in the material nominal stress-life curve (SN curve); and fleet penetration level, i.e. the percentage of units in the fleet plagued by the material capability degradation. Additionally, the derived fleet model is used to evaluate possible mitigation approaches such as fleet recommissioning (i.e. derating) and inspection/repair campaigns. One advantage of the approach is that we can leverage loads and environment information without the need of direct damage measurement. Although we point out that our approach offers only a damage estimate (which is only confirmed with inspection).

The remaining of the paper is organized as follows. Section 2 describes the case study that will illustrate the issue of infant mortality in fleet reliability. Section 3 presents and discusses the numerical results. Finally, Section 4 closes the paper recapitulating salient points and presenting concluding remarks.

2. Case study: infant mortality in a fleet of assets due to a bad batch of materials

We use a numerical experiment to study how fleet size and number of failures impact the characterization of infant mortality in fleets of assets. This numerical experiment aim at emulating scenarios in which fleets of industrial assets are plagued with a bad batch of materials. In other words, the material contaminated during manufacturing yielding in a significant loss in its capability. We consider a hypothetical component to be made out of the AI 2024-T3 alloy \(^1\) subjected to alternating loads and assume that initiation cycles dominate fatigue life. This hypothetical component is mission-critical (its failure does not affect directly the safety of asset operation). Due to the nature of the considered manufacturing problem, components made with pristine material and

\(^1\) AI 2024-T3 is commonly used in aircraft fuselage, flaps, trim tabs, servo tabs and control surfaces.
components made with degraded material would be indistinguishable in operation other than by the unexpected failures. Hence, repairment of plagued components is not possible, although the components could be replaced. Therefore, the numerical experiment focuses on fleets of different sizes plagued with contaminated components that can not be detected based solely on operational performance (i.e., yielding unexpected early life failures).

We use readily available S-N curves commonly found in material handbooks [36] to model low cycle fatigue life at different average and alternating stress levels. Then, we mimic problems with manufacturing (bad batch of materials) by degrading the S-N curves. We designed two missions and two mission mixes to emulate variation due to customer behavior. Finally, we simulate different fleet sizes to understand how failure observations affect overall fleet reliability and detection of emerging issues.

2.1. Damage accumulation at the component level

We used the readily available S-N curves illustrated in Fig. 1 and the following suggested fatigue life model as a function of equivalent stress [36]:

\[
N_f \sim \log\text{Normal}(\mu_{N_f}, \sigma_{N_f}),
\]

\[
\mu_{N_f} = \delta_1 \log(S_{max} + \delta_2) + \delta_3,
\]

\[
S_{eq} = (S_{mean}(1 - (2(S_{mean} - S_{max}))^\delta))^{\delta},\text{ and}
\]

\[
\delta_1 = -3.33, \delta_2 = -12.3, \delta_3 = 9.2, \delta_4 = 0.68, \delta_5 = \sigma_{N_f} = 0.89,
\]

for \(S_{eq} + \delta_2 > 0\) \hspace{1cm} (1)

where:

- \(N_f\) is the fatigue life,
- \(\mu_{N_f}\) and \(\sigma_{N_f}\) are the parameters of the fatigue life lognormal distribution,
- \(S_{eq}\) is the equivalent stress of a given load cycle,
- \(S_{mean}\) and \(S_{max}\) are the mean and maximum stress of a given load cycle, and
- \(\delta_1\) to \(\delta_5\) are model parameters.

Damage is accumulated following Palmgren-Miner's rule:

\[
D = \sum \Delta d_i = \sum \frac{n_i}{N_f^{(i)}},
\]

where:

- \(D\) is the damage accumulated throughout the life of the component
- \(\Delta d_i\) is the damage accumulated by running \(n_i\) cycles at the \(i\)th load level
- \(n_i\) is the number of cycles run at the \(i\)th load level (uniquely defined by mean and maximum stress).

- \(N_f^{(i)}\) is the fatigue life at the \(i\)th load level, and
- the threshold for end of life is \(d_{TH} = 1\).

In many research studies, fatigue lives of metal materials are generally assumed to follow either lognormal or Weibull distributions [37,38]. In this contribution we assume fatigue life follows a lognormal distribution. Since fatigue life \(N_f^{(i)}\) follows a lognormal distribution, \(D\) is a random variable with no closed-form expression for its probability density function. However, considering that all \(\Delta d_i\) have the same variance, then, \(D\) can be approximated by:

\[
D \sim \log\text{Normal}(\mu_D, \sigma_D),
\]

\[
\sigma_D^2 = \ln\left(\mu_D^2 - 1\right)\left(\sum \sigma_{N_f}^2 + 1\right),
\]

\[
\mu_D = \ln\left(\sum e^{\sigma_{N_f}^2}\right) + \frac{\sigma_D^2}{2},\text{ and}
\]

\[
\Delta d_i \sim \log\text{Normal}\left(-\mu_{D, i} + \ln(n_i), \sigma_{D, i}\right).\]

(3)

Since damage is accumulated after each mission, for a given component, the number of missions to failure (MTF) is a random variable with cumulative density function defined by

\[
F_{MTF}(m) = \Pr[M_{TF} \leq m] = \Pr[D^{(m)} \geq 1],
\]

where \(D^{(m)}\) is the damage accumulated up to \(m\) missions of component \(i\). This implies that \(i\)th component reliability \(R_i(m)\) (probability to perform below the threshold) and unreliability \(Q_i(m)\) (probability to perform above the threshold) at mission \(m\) are given by:

\[
R_i(m) = 1 - \Pr[D^{(m)} \geq 1],\text{ and}
\]

\[
Q_i(m) = 1 - R_i(m) = \Pr[D^{(m)} \geq 1].
\]

(5)

We designed the two missions shown in Fig. 2. At the end of each mission, the accumulated damage is distributed around \(6.55 \times 10^5\) and around \(2.63 \times 10^4\) and around \(6.55 \times 10^5\) for mission #1 and #2, respectively. The 50th percentile of fatigue life is approximately 3,800 and 15,260 missions for different sizes plagued with contaminated components that can not be detected based solely on operational performance (i.e., yielding un-expected early life failures).

\[
LH^{(m)} = \begin{cases} 
\text{Mission 1, if } M_{IDX} = 1 \\
\text{Mission 2, if } M_{IDX} = 2 
\end{cases},\text{ and}
\]

\[
M_{IDX} \sim \text{Bernoulli}(k = M_{IDX}, p = \theta_k).
\]

(6)

where:

- \(LH^{(m)}\) is the load history for mission \(m\),
- \(M_{IDX}\) is an index that defines which mission to assign, and
- \(\theta_k\) is a Bernoulli distribution parameter.

Fig. 1. S-N curves for the Al 2024-T3 alloy (adapted from [36]).

Fig. 2. Alternating stress levels (\(S_{Min}\) to \(S_{Max}\)) for the two designed missions.
Table 1
Mission mix formulation. Every asset in the fleet is expected to operate at or between, or even alternating between, aggressive and mild mission mixes.

| Mission 1 | 50% | 15% |
| Mission 2 | 50% | 85% |
| 50th percentile $N_f$ | 9230 | 13360 |

-$\theta_0$ is a calibration parameter.

In this contribution, we considered two mission mixes as detailed in Table 1, in which mission #1 and mission #2 bound the life distributions for any mission mix. This way, when variations due to both loads in the form of mission mixes and material capability (spread in S-N curves) are considered, the distribution of fatigue life is as illustrated by Fig. 3.

We emulate degradation in material capability by shifting S-N curves to the left (i.e., for the same stress level, the material has a shorter fatigue life as compared to the nominal material). For simplicity, we can consider that the shift on the S-N curves affects its mean prediction and preserves the variance (i.e., a shift on the mean value of the damage $D$ lognormal distribution given in Eq. (3)). In this contribution, the degradation shift is modeled through the introduction of the model parameter $\theta_i$, as illustrated by Eq. (7)

$$D_{\text{Debit}} \sim \text{logNormal}((1 - \theta_i)C, \sigma)$$

where $\theta_i$ is a calibration parameter (defining the degradation in material capability).

Fig. 4 illustrates the effects of considered material capability degradation on the fatigue life distribution for the aggressive mission mix. At the highest degradation considered (20%), the median of missions to failure can be reduced from 9230 to 860 missions.

Fig. 3. Fatigue life distribution in terms of missions to failure considering both load (mission mix) and material capability variations (spread in SN curve).

2.2. Fleet commissioning, reliability, and failure observations

Large fleets of assets are usually commissioned over a period of time (as production follows a backlog of orders, commissioning ramps up for a while before it starts to decline). Commissioning schedule determines the number of units running (and as a consequence, it impacts the number of failure observations). In this study, we arbitrarily model commissioning time through a truncated Gaussian distribution (as illustrated by Fig. 5), in which assumed commissioning time after product launch $T_{\text{Comm}} \sim N(\mu = 4.5, \sigma = 2.625)$ and $0 \leq T_{\text{Comm}} \leq 10$. With the considered assumptions the expected fleet size in the 10th year is 10,000 units. In real life, this distribution is first estimated based on market analysis and can be updated as units are sold and commissioned.

Different commissioning time makes the units across the fleet to have different accumulated service lives (and damage, as a consequence).

After commissioning, we assume that each unit runs one mission per day. Integration of asset unreliability, for pristine and material with degradation in capability, up to fleet reliability (i.e., the complementary set of the accumulated number of failures in the fleet $(1 - \frac{N_f}{N_{\text{fleet}}})$) is straightforward:

$$Q_{\text{Pristine}}^{(i)} = \frac{\sum_{t=1}^{t(t_{\text{comm}})} N_f^{(i)}}{N_{\text{fleet}}},$$

$$Q_{\text{Debit}}^{(i)} = \frac{\sum_{t=1}^{t(t_{\text{comm}})} N_f^{(i)}}{N_{\text{fleet}}}.$$

where:

- $Q_{\text{Pristine}}^{(i)}$ and $Q_{\text{Debit}}^{(i)}$ are the fleet unreliability due to material with the nominal capability and material with certain capability degradation, respectively, both at time $t$. $Q_{\text{Debit}}^{(i)}$ is not related to failures yielded by materials with a specific capability degradation (i.e. to a specific $\theta_i$), but rather accounts for all observed failures due to the material with a non-nominal capability (i.e. degraded material).

- $Q_{\text{Pristine}}^{(i)}$ and $Q_{\text{Debit}}^{(i)}$ are the unreliability of component $i$ assuming it is made of pristine material and material with certain capability degradation, respectively, both at time $t$ greater than the component commissioning time ($t_{\text{Comm}}^{(i)}$). $Q_{\text{Pristine}}^{(i)}$ and $Q_{\text{Debit}}^{(i)}$ only start being computed if the unit is already commissioned (otherwise both are null).

- $N_{\text{fleet}}$ is the final fleet size.

Fig. 4. Fatigue life distribution for aggressive mission mix and different levels of material capability.

Fig. 5. Commissioned units over time.
As a consequence, at time \( t \), fleet unreliability \( Q_{\text{Fleet}}^{(t)} \) is:

\[
Q_{\text{Fleet}}^{(t)} = \delta_t \times Q_{\text{Fleet}}^{(t)} + (1 - \delta_t) \times Q_{\text{Prior}}^{(t)}
\]

where \( \delta_t \) is a model parameter that defines the penetration of units in the fleet (in terms of a fraction of the fleet) made of a material with a certain degradation level in capability.

With fleet unreliability we can predict the number of failures at any year after product launch by using the binomial distribution to model the number of failures:

\[
f_{\text{Fail}}(n = N_{\text{Fleet}}; p = Q_{\text{Fleet}}^{(t)}, N = N_{\text{Fleet}}) = \binom{N_{\text{Fleet}}}{N_{\text{Fail}}}(1 - Q_{\text{Fleet}}^{(t)})^{N_{\text{Fleet}} - N_{\text{Fail}}} \]

where:

- \( f_{\text{Fail}}(.) \) is the probability density function of the binomial distribution,
- \( N_{\text{Fail}} \) is the accumulated number of observed failures up to year \( y \),
- \( Q_{\text{Fleet}}^{(t)} \) is the fleet unreliability at year \( y \), and
- \( N_{\text{Fleet}} \) is the fleet size at year \( y \).

Conversely, given a number of observed failures, we can estimate the model parameters through the Bayes rule:

\[
f_{\text{Prior}}(\theta|N_{\text{Fail}}, N_{\text{Prior}}) \propto f_{\text{Prior}}(N_{\text{Fail}}|Q_{\text{Prior}}, N_{\text{Fleet}}) f_{\text{Fail}}(\theta) \] (11)

where \( Q_{\text{Prior}} \) is a function of \( \theta \).

### 2.3. Fleet management

Our physics-based prognosis model for fleet management is composed of two Bayesian networks, one for the asset reliability and another one for the fleet unreliability. Fig. 6 shows the asset-specific dynamic Bayesian network that relates material properties and loads with damage accumulation, where:

- \( LH \) corresponds to load history,
- \( S_{eq} \) is the equivalent stress of a given load cycle,
- \( \mu_{\text{eq}} \) and \( \delta_{\text{eq}} \) are parameters of the fatigue life lognormal distribution,
- \( \Delta d^{(t)} \) is damage accumulated after running through \( LH^{(t)} \),
- \( D_i^{(t)} \) is the damage accumulated by component \( i \) up to \( t \),
- \( \theta_{1,4} \) and \( \theta_5 \) are parameters defining material properties,
- \( \theta_6 \) is the parameter defining the mission mix,
- \( \theta_7 \) is the parameter defining degradation in material capability.

Superscripts \((t - 2), (t - 1), \) and \((t)\) indicate the timestamps in which inference/estimation is performed. Load histories for missions 1 and 2 are defined in Fig. 2.

As formulated, the general case for the asset-specific model yields in a Bayesian network with seven model parameters, \( \theta_1 \) to \( \theta_7 \), as illustrated by Fig. 6. Albeit, the problem treatment may significantly vary from the general case to a simplified model, depending on the application some of the model parameters can be set to a given value yielding in a reduced number of model parameters to be estimated. This is the case for the numerical application addressed in Section 3. We will consider material properties defining the nominal S-N curves to be known; hence, parameters \( \theta_1 \) to \( \theta_2 \) are set to the values given by [36], as shown in Eq. (1). In the mentioned numerical application we also consider the mission mixes to be fixed throughout asset life, thus, \( \theta_3 \) will also be fixed at the values shown in Table 1. Therefore, with respect to the asset-specific model, for the considered application, only \( \theta_7 \), which defines the degradation in material capability, will be calibrated with failure observations (as will be discussed in Section 3). Again, it worth highlighting that the problem treatment and uncertainty predictions presented for this simplified model, may not hold for the general case.

Fig. 7 shows the fleet dynamic Bayesian networks, where \( Q_{\text{Fleet}}^{(t)}\mu_{\text{eq}} \) and \( Q_{\text{Fleet}}^{(t)}\delta_{\text{eq}} \) are vectors of damage accumulated up to \( t \) for the \( N_{\text{Fleet}} \) assets in the fleet with and without degradation in material capabilities, respectively, \( Q_{\text{Fleet}}^{(t)}\mu_{\text{eq}} \) and \( Q_{\text{Fleet}}^{(t)}\delta_{\text{eq}} \) are vectors of unreliability at time \( t \) for the \( N_{\text{Fleet}} \) assets in the fleet with and without degradation in material capabilities, respectively. \( \theta_t \) is one more calibration parameter related to the fraction of the fleet made of material with a certain degradation level in capability, i.e., the fleet penetration of material with low capability, that will also be calibrated with failure observations. In the numerical application discussed in Section 3, both the simplified asset-specific model and the fleet dynamic Bayesian network model will be used to make inference about \( \theta_7 \) and \( \theta_8 \), as well as estimate and forecast fleet unreliability \( Q_{\text{Fleet}}^{(t)} \).

The number of failures at time \( t, N_{\text{Fail}}^{(t)} \), can be modeled through a binomial distribution where \( N_{\text{Fail}} \) (fleet size) is the number of trials and \( Q_{\text{Fleet}}^{(t)} \) is the probability associated with each trial:

\[
N_{\text{Fail}}^{(t)} \sim \text{Bin}(Q_{\text{Fleet}}^{(t)}, N_{\text{Fleet}}). \] (12)

This model can be applied in simple estimation at current time or even in forecast (when \( Q_{\text{Fleet}}^{(t)} \) and \( N_{\text{Fleet}} \) can be forecasted).
In this contribution, we study the effects of fleet sizes in the ability to forecast the number of failures and its implication to fleet management. From Eqs. (11) and (12), it is expected that inference performed with data from small fleets will result in large uncertainty about the model parameters. This is a direct consequence of the reduced number of observations (in the binomial distribution, estimation of small probabilities is accompanied by large uncertainties when the number of observations is small, see [39] for further details). This is problematic as the model parameters are then used to estimate and forecast the number of future failures. Large uncertainty in the number of future failures drives conservativeness in the way operators manage their fleets. On the other hand, operators of large fleets, large services and maintenance companies, and original equipment manufacturers tend to observe a large number of failures and should be able to benefit from it in terms of uncertainty quantification. Regardless of the fleet size, effective fleet management asks for a continuous model update as new information is made available throughout service lives (including re-visiting the assumptions about model form, failure modes, etc.).

The estimated number of failures can be used to assess the “risk” associated with the forecast. One very straightforward measure is the uncertainty about the forecasted number of failures (i.e., companies have to be prepared to absorb that variation from a financial perspective). There are a number of ways to quantify variation in the number of failures. One can simply use the standard deviation, which might not be convenient given the asymmetric nature of the $N_{fail}$ estimator. Alternatively, the difference between 97.5 and 2.5 percentiles of the forecasted number of failures can also be used. Operators of small fleets can use this range to support the decision to either self-perform or buy a contractual service agreement from a third party company. Operator of small fleets tend to have difficulties in absorbing large variations in the forecasted number of failures due to liability associated with it (both in terms of inventory, labor, etc., as well as in terms of loss of revenue). This can make operator of small fleets overzealous and perform excessive inspection and services in the hope to prevent costly maintenance or catch serious problems when units are still under manufacturer warranty (minimizing the impact of unscheduled removals and cost of repairs/replacements). For large fleet operators, the problem shifts from unexpected downtime to an excessive number of costly maintenance and contractual obligations regarding availability and reliability.

We compare results from a small and a large fleet to mimic the operator of a small fleet vs large service provider dynamic. For this portion of the case study, we assume that the large service provider is more likely to provide an unbiased and accurate estimation of the number of failures. If that is the case, the difference in the forecasted number of failures can help us judge whether self-performing is a good decision or not. Mathematically

$$F_{Exp} = N_{Small}^{Large} - N_{Small}^{Small}$$

where $N_{Fail}^{Small}$ and $N_{Fail}^{Large}$ are the estimated/forecasted number of failures on the small fleet coming from the large service provider and the small fleet operator models, respectively.

The fleet exposure index ($F_{Exp}$) becomes an indicator of whether the small fleet operator is likely to save or lose money by self-performing services and maintenance:

- $F_{Exp} > 0$: the operator of a small fleet over predict failures, which drives the allocation of more resources than needed. In other words, the behavior is conservative and it translates in savings due to avoided unscheduled maintenance, reduced downtime, etc.
- $F_{Exp} < 0$: the operator of a small fleet under predicts failures, which drive allocation of fewer resources than needed (operator loses money due to unscheduled maintenance, downtime, etc).

Again, assuming that we can use $N_{Fail}^{Large}$ as an estimator of $N_{Fail}^{Small}$ (i.e. the service provider estimates for the failure observations in the small fleet is close to the true number of failures), it is expected that inference performed with data from small fleets will result in large uncertainty about the model parameters (see [39], for further details). Although $N_{Fail}^{Large}$ can be obtained in this numerical example (through Eq. (12) since fleet reliability for known loads can be obtained at any point in time), we avoid using it as it is not available in real life though.

In this contribution we also use the calibrated Bayesian networks to evaluate another two mitigation approaches: fleet recommissioning (i.e. derating the operation of individual assets) and inspection campaigns. Fleet recommissioning is performed by changing assets from an aggressive mission mix to a mild mix. Its goal is to mitigate failures by reducing mission stress levels. Although, in practice, reducing stress...
levels can imply reduced performance (e.g., derating the machine). An inspection campaign targets improving fleet reliability by identifying and repairing units with problems.

Inspection is modeled in the Bayesian network employing a two-parameter Palmberg model [40] that represents the probability of detection (POD) curve as:

\[
POD = \frac{D^n}{1 + D^n}, \quad 0 \leq D < \infty
\]

(14)

where POD is the probability of detecting an asset with damage level \(D\). \(D^*\) is a hyperparameter that measures the quality of the detection technique, by imposing POD equals 50% when \(D = \) is equal to \(D^*\). The other hyperparameter \(n\) addresses the detection uncertainty since increasing its value yields in steeper POD curves. Note that we use the Palmberg model only to demonstrate the Bayesian network ability to handle inspection campaigns, and discussions regarding its hyperparameters tuning are not in the scope of this contribution. Here, we arbitrarily considered inspection characterized by the probability of detection with \(D^* = 0.05\) and \(n = 0.25\). It is worth mentioning that in practical applications a technique-specific POD model can be used to replace Eq. (14), or at least \(D^*\) and \(n\). Additionally, in our formulation, \(D = 1\) characterizes failure undoubtedly. It is important to point out that our proposed inspection campaign assumes that all failed units would be removed and inspection is carried only in units with a damage level lower than 1.

Due to the nature of the considered numerical experiment and for simplicity we opt to focus our analysis on asset derating (i.e. fleet recommissioning) and fleet inspection. It is important to point out that these are not the only mitigation measures and depending on the nature of the addressed problem other measures could be considered (e.g. burn-in, as usually applied in the electronics industry). Also, the proposed numerical experiment is an academic simplification, in a real-life application many other aspects (e.g. asset decommissioning) would influence observed failures and the performance of the proposed mitigation measures.

3. Results and discussions

3.1. Numerical example

In order to evaluate the effect of fleet size in the number of failure observations, we defined two distinct fleets:

- a large fleet of 10,000 units: emulating an original equipment manufacturer or a large service provider, and
- a small fleet of 1,000 units: emulating a small fleet operator. These units come from the larger 10,000 unit fleet, which also means that the large fleet operator has visibility into what happens with this small fleet.

Both fleets are plagued with a material degradation level of 15%. However, to make things more interesting, we distributed the failures across the fleet such that the small fleet operator has a penetration of 20% of units plagued with material of inferior capability (i.e., 200 out of 1,000 units are plagued), while the larger fleet has an overall 10% penetration (i.e., 1,000 out of 10,000 units are plagued). Thus, in this example we considered that model parameter \(\theta_7\) and \(\theta_8\) are constant in time (as they represent debit and fraction of the fleet plagued). In real life this assumption may not hold if there are changes in manufacturing process over time (affecting the level of material degradation, \(\theta_7\), and the proportion of plagued assets in the fleet, \(\theta_8\)).

The implications in fatigue life distribution are shown in Fig. 8, in which load history is exclusively coming from the aggressive mission mix. There is a considerable shift in fatigue life if the entire fleet is plagued with a material of poor capability. Nevertheless, the 10% and 20% penetration levels for the large and small fleets, respectively, the effect mostly manifested in the lower tail of fatigue life distribution.

As discussed in Section 2.2, commissioning has an effect on fleet unreliability as the fleet grows bigger with asynchronous aging. As an illustration, Fig. 9 shows a comparison between fleet unreliability over time with and without the effect of commissioning. The drastic reduction in unreliability values results in a delay in rising failures observations. Most industrial engineering assets (the focus of this paper) are commissioned over a period of time. In the remainder of this section, we will discuss the results following the commissioning detailed in Section 2.2.

With the fleet unreliability over time, we can forecast the number of failures. Fig. 10 highlights the contribution of each subpopulation by material type (pristine and with degradation in capability) in the resulting failure observations. Besides the obvious penetration of material with poor capability (10% versus 20% for the large and small fleet, respectively), commissioning also affects the relative contribution of each material to the number of failures. Early on, most failures come from components plagued with material with poor capability. Over time, the unreliability for pristine material increases (see Fig. 9), and the relative contribution of each population starts to change. Around the 3rd year after deployment, at least for the large fleet, failures are dominantly coming from components made out of pristine material (although the contribution from the subpopulation with plagued material is still substantial).

3.2. Bayesian networks calibration

At the 3rd year after deployment, we assume the following number of failure observations:

- Small fleet: 17 failures (lower tail of the predicted number of failures).
- Large fleet: 127 failures (roughly 50th percentile of the predicted number of failures). Obviously, the small fleet failures are contained in this set.

We use these failure observations and fleet unreliability (from known load histories) to calibrate:

- \(\theta_7\): degradation in material capability with a uniform prior between 1% to 30%, and
- \(\theta_8\): penetration of units with poor material capability in the fleet with a uniform prior between 0.01% and 20%.

Figs. 11 and 12 detail the calibration results with regards to both model parameters and the estimated number of failures for the small and large fleets, respectively. Even for the small fleet operator, there is considerable uncertainty reduction and failure estimates are much improved as compared with the uniform priors (see Table 2, in which for instance we can observe the reduction in the 95% prediction interval, from an interval range of (8,22) in the small fleet prior, to (9,20) in its updated posterior).

As we mentioned before, the large fleet operator has full visibility into what happens with the small fleet. In this numerical example, the relative number of failures with respect to the fleet size can be used to map the posterior distribution of the number of failures at the large fleet into the small fleet, as illustrated by Fig. 13 and summarized in Table 2. When compared to Fig. 11-(c), the uncertainty in Fig. 13 is much smaller and clearly attributed to the much richer information available at the large fleet level.

The models with updated model parameters can be used to forecast the number of failures over time. Fig. 14 shows how these forecasted values look like for the small fleet. In the 3rd year, both models are unbiased. The uncertainty in the posterior distribution of model
parameters for the small fleet model, Fig. 11-(b), is larger than the one for the large fleet model, Fig. 12-(b). The result is the larger and ever-increasing uncertainty that the small fleet model exhibits when compared to the large fleet model.

### 3.3. Mitigation assessment

Once the infant mortality issue is quantified, operators undergo a number of mitigation actions to reduce costs associated with unscheduled maintenance, asset unavailability, etc. (as discussed in Section 2.3 Fleet management). We first look at fleet recommissioning (changing the mission mix from aggressive to mild). This approach can be costly, as mild mission mixes are usually associated with some loss in performance or productivity. Fig. 15 illustrates the estimated/forecasted fleet unreliability and forecasted number of failures over time for the small fleet. Fig. 15-(a) shows that after the entire small fleet is recommissioned from the aggressive to the mild mission mix, the estimated fleet unreliability with the updated model falls between the estimates for the entirely pristine fleet and the actual fleet composition both operating at the aggressive mission mix. This means that although there is a significant improvement in unreliability, the levels are still above design intent. Interestingly, the distribution in forecasted fleet unreliability might still be useful for estimating the number of failures. Fig. 15-(b) shows the forecasted number failures coming out of the unreliability estimates of Fig. 15-(a). The aggressive mission mix and pristine material represent the design intent. The estimated penetration and material degradation represent the forecasts if the fleet keeps operating at the aggressive mission mix. Visual comparison between the two cases makes it clear that the number of failures could be potentially much larger than what was intended. Recommissioning the fleet knocks down the number of failures and makes the prediction interval overlap with the one from the design intent.

As expected, Fig. 16 shows that recommissioning as a mitigation measure is much more effective at the large fleet level. Coincidently,
Fig. 11. Calibration results for the small fleet. Failure observations in the 3rd year after deployment and uniform priors feed the Bayesian update. (a) The prior distribution of the number of failures. (b) Posterior distribution of model parameters. (c) Posterior distribution of the number of failures.

Fig. 12. Calibration results for the large fleet. Failure observations in the 3rd year after deployment and uniform priors feed the Bayesian update. (a) The prior distribution of the number of failures. (b) Posterior distribution of model parameters. (c) Posterior distribution of the number of failures.
As shown in Fig. 19, this is an option, as the operator of the small fleet does not know the outcomes of the large fleet operator model (and model form, assumptions, etc. also tend to be unknown). Nevertheless, we can study that in this synthetic example. Fig. 17 shows the forecasted $F_{\text{Exp}}$ index as defined in Eq. (13) before and after fleet recommissioning for the small fleet. Continuous and dotted lines represent the median and 95% prediction intervals, respectively. The operator of the small fleet is likely to lose money by self-performing services and maintenance when $F_{\text{Exp}}$ is negative (since unbiased predictions from the operator of the large fleet tend to be larger than the ones from the operator of the small fleet). Conversely, the operator saves money by self-performing when $F_{\text{Exp}}$ is positive. Although the $F_{\text{Exp}}$ median is relatively small up to the 5th or 8th year, depending on recommissioning, the uncertainty about it tends to be large and continuously increasing.

Another way of looking at the risk associated with self-performing maintenance is through the probability of reward and loss. With the $F_{\text{Exp}}$ index defined by Eq. (13), there are three things to keep in mind: (1) when $F_{\text{Exp}}$ is positive, the operator of the small fleet saves money by self-performing services and maintenance, (2) conversely, when $F_{\text{Exp}}$ is negative loss of money is more likely; and finally, (3) unbiased number of failure estimates imply that the expected value of $F_{\text{Exp}}$ is zero.

With that in mind, Fig. 18-(a) shows that self-performing is reasonable in the short term (obviously, the operator attitude towards unaccounted failures determines how long that option remains reasonable). In order to illustrate this idea, consider that an arbitrary threshold of $P_{\text{Prob}}[F_{\text{Exp}} \geq 0] \geq 0.4$ is imposed (in practical applications this threshold would be imposed by the small fleet operator depending on its willingness to handle unexpected failures). Then, the operator of the small fleet could sustain the aggressive mission mix until almost the end of the 4th year (without having to buy a services and maintenance contract). If the operator decides to recommission the fleet at the third year; then, with the $P_{\text{Prob}}[F_{\text{Exp}} \geq 0] \geq 0.4$ threshold, self-performing is reasonable until the 7th year.

Now, let us assume that the operator of the small fleet is willing to accept underestimating failure of 10 units. Then, Fig. 18-(b) shows the probability that the operator will have to pay for the extra 10 units (unplanned failures). If again an arbitrary threshold of $P_{\text{Prob}}[F_{\text{Exp}} \leq -10] \leq 0.2$ is imposed, then the operator of the small fleet could sustain the aggressive mission mix until the middle of the 4th year (without having to buy a services and maintenance contract). Switching to a mild mission mix early on, extends that window to the middle of the 5th year.

Finally, we evaluate inspection as a mitigation measure. Our case study focuses on failures due to a manufacturing problem. Therefore, it is hard to identify the problem at an asset level purely by monitoring operation (i.e., through sensors and performance). Although a massive fleet-wide inspection can be considered, fleet size and associated downtime would make it prohibitively expensive. Here, we considered inspection characterized by the probability of detection (POD) defined in Eq. (14), with $D^* = 0.05$ and $n = 0.25$. As shown in Fig. 19, this is representative of a relatively good inspection technique, as there are high probabilities of detections for small damage levels.

We estimated the impact of inspection campaigns covering three different percentages of the small fleet: 25%, 50%, and 100%. The damage distributions come from the nominal and shifted SN curves. All considered campaigns were performed in the 3rd year when the small fleet had 256 commissioned units. Thus, in each campaign, an equivalent of 64, 128 and 256 units went through the inspection routine. Damage detection is defined by a Bernoulli trial where the probability $p$ is given by the probability of detection. We also defined replacement according to an arbitrary damage threshold of 0.25. In other words, if an inspected unit has accumulated damage greater than 0.25 (assessed as a result of the inspection), then the unit is replaced otherwise no action is taken. When replacement is performed, it is assumed that a component with pristine material is put in place (i.e., no new manufacturing defects are introduced after the replacement of units). This essentially keeps the number of units in the fleet the same.

### Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>Percentile</th>
<th>2.5</th>
<th>50</th>
<th>97.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design intent</td>
<td>3</td>
<td>8</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Small fleet prior</td>
<td>8</td>
<td>10</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Small fleet posterior</td>
<td>9</td>
<td>14</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Large fleet posterior</td>
<td>14</td>
<td>17</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

* Number of failures considering only pristine material.

![Fig. 13.](image-url) Poster distribution of the number of failures in the 3rd year after deployment for the small fleet as estimated by the large fleet operator.

![Fig. 14.](image-url) The forecasted number of failures for the small fleet as estimated by both small and large fleet models (error bars represent the 95% prediction intervals).

the fleet unreliability after recommissioning converges to design intent. Obviously, it comes at the cost of a mild mission mix. In real life, even though the reliability levels would be back to design intent, it might not be practical to recommission the fleet due to performance losses. The final compromise between reliability and performance is a decision that is highly industry dependent (and discussion is outside the scope of this paper). Fig. 16-(a) shows uncertainty levels in fleet unreliability after recommissioning are much smaller than those shown in Fig. 15-(b). This has direct implications in the forecasted number of failures, as illustrated by Fig. 16-(b), to the point that there is a good overlap between estimated and intended error bars.

Besides recommissioning, the small fleet operator can also consider contracting out services and maintenance from a large service provider as a way to reduce financial exposure due to the upcoming high number of failures. In real life, it is difficult to forecast the costs associated with such an option, as the operator of the small fleet does not know the
(however the replaced units have zero damage and pristine material).

For illustration purposes, a simplified flowchart of the inspection routine is presented in Fig. 20. In real life, the damage threshold could be linked to the unit lifetime and nominal degradation rate. The results of this part of the study highlight the potential benefits of an inspection campaign in the overall forecasted number of failures. This way, operators could assess whether it pays to invest in an inspection.

Table 3 shows the 95% prediction intervals for the number of replaced units as estimated by the Bayesian model of the small fleet (predicted) and considering the actual values of material degradation and fleet penetration (actual). While the predicted values account for uncertainty in the Bayesian model as well as in the probability of detection, the actual estimates only account for uncertainty in the probability of detection. Fig. 21 illustrates the effects of the inspection/replacement campaign on the Bayesian model forecasted number of failures. For comparison purposes the error bars of the fleet recommissioning of the entire fleet in the third year were also plotted. Despite the high levels of uncertainty it can be noticed that partial or complete inspection makes the forecasted number of failures to overlap with design intent, as shown in Fig. 21. Another feature from the inspection campaign forecast, is the increase in the overall uncertainty as the years accumulate, as shown in Fig. 21 (a) and (b). This is a direct result of the fleet splitting since the uncertainty of each distinct group accumulates in the model forecast.

Also, from Fig. 21, it can be inferred that the inspection campaigns, either partial or complete, have a drastic impact in the number of failures in the short term (especially in the first years after the campaign) with the overall fleet failures returning to high levels as time passes. Hence, from the Bayesian model failure forecast, it seems that in the short term the inspection campaign would be more efficient than fleet recommissioning. However, in the long run, as failures start to accumulate both approaches would lead to similar results. It is important to point out that a more thorough investigation taken into

Fig. 15. Small fleet recommissioning. Recommissioning curves show the 50th percentile and the 95% prediction interval and error bars represent the 95% prediction intervals. (a) Small fleet unreliability after recommissioning; (b) Small fleet forecasted number of failures.

Fig. 16. Large fleet recommissioning. Recommissioning curves show the 50th percentile and the 95% prediction interval and error bars represent the 95% prediction intervals. (a) Large fleet unreliability after recommissioning; (b) Large fleet failure observations intervals.

Fig. 17. Small fleet $\text{FExp}$ index (defined by Eq. (13)) forecast before and after recommissioning.
Fig. 18. Self-performing reward and loss probabilities for the small fleet. (a) \( \text{Prob}\{F_{\text{Exp}} \geq 0\} \). (b) \( \text{Prob}\{F_{\text{Exp}} \leq -10\} \).

Fig. 19. Probability of detection used in this study.

Table 3
Predicted intervals for the number of replaced units as a function of the percentage of fleet inspected (inspection performed when operator of the small fleet has 256 commissioned units).

<table>
<thead>
<tr>
<th>Percentage of fleet inspected</th>
<th>Percentiles</th>
<th>2.5</th>
<th>50</th>
<th>97.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>Predicted</td>
<td>2</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Actual</td>
<td>6</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>50%</td>
<td>Predicted</td>
<td>4</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Actual</td>
<td>15</td>
<td>21</td>
<td>29</td>
</tr>
<tr>
<td>100%</td>
<td>Predicted</td>
<td>7</td>
<td>30</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Actual</td>
<td>31</td>
<td>43</td>
<td>53</td>
</tr>
</tbody>
</table>

Fig. 20. Simplified flowchart of the proposed inspection routine for the fleets Bayesian models.
account financial aspects should be performed to verify this observation.

4. Summary and closing remarks

In this work, we studied early life failures as applied to fleet management. Depending on the scale of the problem, early failures can have a significant impact on the safety, availability, and operational profit of industrial equipment. We used Bayesian networks to model fleet reliability as well as to characterize and quantify a manufacturing-related problem (material capability degradation). We designed a simple numerical experiment where:

- degradation in material capability is used to characterize infant mortality, and
- fleet commissioning is a function of time.

We have studied:

- The effect of degradation in material capability: results confirmed that it dramatically reduces fleet reliability.
- The effect of fleet commissioning over time: results highlighted how it masks the increase in fleet unreliability.
- Fleet size and number of failures interaction and its effects on the infant mortality characterization: in the considered numerical experiment, the proposed framework was able to characterize the percent of the fleet plagued by poor material capability. Nevertheless, results illustrated how the fleet size can hinder the characterization of the problem, especially for small fleet operators.
- Bayesian models for assessing possible mitigation approaches: the proposed framework was able to evaluate two distinct mitigation approaches, derating assets (i.e. recommissioning them at reduced load levels) and inspection. Presented results indicate that the proposed model could be used as a tool for fleet management and aid operators to decide on a proper action course. For example, an inspection campaign could mitigate the problem by screening out units that could potentially fail (even if this leads to temporary disruption of machine usage). Alternatively, recommissioning mitigates the loads and therefore slows the failure rate down (at the cost of performance loss).

The results motivate us to extended the study in several aspects. For example, in order to have a better understanding of the impact that early failure in fleets of assets, we suggest extending the study and include, among other factors:

- Improved physics of failure models: not only by separating the cycles spent in initiation and propagation, but also improving the stress models to account for geometry and boundary conditions.
- Other sources of uncertainty, such as:
  - field inspection: qualitative inspection (distress ranking) instead of damage measurements
  - service level (repair versus replace failed units).

Declaration of Competing Interest

None.
Acknowledgments

This work was supported by the University of Central Florida (UCF). Nevertheless, any view, opinion, findings and conclusions or recommendations expressed in this material are those of the authors alone. Therefore, UCF does not accept any liability in regard thereto.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ress.2020.107225,

References