ROC-based Business Case Analysis for Predictive Maintenance – Applications in Aircraft Engine Monitoring

Lily Koops

Bauhaus Luftfahrt e.V., Willy-Messerschmitt-Str. 1, 82024 Taufkirchen, Germany
lily.koops@bauhaus-luftfahrt.net

ABSTRACT

Predictive maintenance approaches leveraging integrated knowledge, fleet-wide data and machine-learning techniques allow for earlier warnings on impeding failures and for higher accuracy in remaining useful life predictions compared with traditional maintenance strategies. However, in case relative to correctly predicted maintenance needs, missed detections or false alarms occur too often, benefits can be outweighed by follow-up costs due to unexpected damage or unnecessary inspections. For business case evaluation, we demonstrate the value of a general approach to cost-benefit analysis based on the Receiver Operating Characteristics (ROC) curve. It allows for deducing application-specific requirements on prediction quality for achieving a net benefit and for comparing and optimizing failure prediction algorithms regarding cost-efficiency. As example of use, the approach is applied within aircraft engine maintenance to assess potentials for reducing unscheduled engine removals by more accurate prediction of turbine blade failures. Based on realistic, literature-based assumptions on various costs, failure probability and algorithm performance, maximal cost-saving potentials of up to 17 Mio $ are found per mature-run, widebody engine and per mean-time between removals. The machine-learning based fusion of a pure physics-of-failure model with relevant data, e.g. pertaining to environment and inspection, is shown to allow for an up to 42% higher cost benefit, demonstrating the value of data for predictive maintenance purposes. Generalizations of the presented approach, e.g. to cost-optimize engine workscope planning or other system maintenance, are discussed.

1. INTRODUCTION

One of the general expectations of Prognostics and Health Management is the translation of raw data related to the health state of engineering systems into actionable information to facilitate rapid and informed maintenance decision making.

Within the last decades, several new methods for prognostics and for assessing the performance of predictions for health management have been developed (Saxena, Sankararaman, & Goebel, 2014). Lately, e.g. in aeronautics, approaches are investigated that leverage integrated knowledge and fleet-wide data of various sources to reduce uncertainties related to modeling of systems, and the impacts of usage profiles, operating environments and Maintenance, Repair and Overhaul (MRO) actions such as on asset health state, degradation rate and performance. Capturing complex relationships in the data that may be difficult to describe using physics, allows for earlier warnings on impeding failures and for higher accuracy in remaining useful life predictions (Saxena et al., 2014; Wagner, Saalmann, & Hellingrath, 2016).

Yet, not all use cases are economically favorable for approaches relying on predictive analytics. Qualification criteria of business problems involve sufficient high-quality data and business needs, the latter arising from a significant influence of particular failure modes of components or (sub) systems of an asset on reliability, availability, installation or maintenance effort or operational costs. Moreover, importantly, for achieving a net average benefit compared to conventional strategies like preventive maintenance, application-specific requirements on predictive algorithm performance as well as on maximally allowed overhead costs such as for development, implementation and maintenance of the predictive analytics solution arise that have to be met for economic viability.

As key performance measure of predictive algorithms, the Receiver Operating Characteristics (ROC) curve, discussed in more detail in the next section, indicates all possible combinations of relative occurrences of various kinds of correct and incorrect predictions (Metz, 2018). It can be directly linked to cost-benefit analysis of diagnostics- / prognostics-based decision-making, allowing for determining the optimal compromise among various kinds of prediction errors and finally, for business case identification. While ROC curves provide a common basis to medical decision making (Metz, 2018), and in recent years have been increasingly adopted in the machine learning and data mining

research communities, their natural relation to cost-benefit analysis is not commonly exploited for business case evaluation with regards to predictive maintenance potentials in engineering disciplines like Prognostics and Health Management.

It is the aim of this paper to highlight the value of the ROC-based approach in a) evaluating, comparing and optimizing predictive algorithm performance and b) for business case analysis specifically for aeronautical applications such as for optimizing engine maintenance expenses.

In section 2, some general conclusions are drawn resulting from accounting for costs of the predictive maintenance solution itself as well as for those resulting from decisions / actions taken on its basis including the negative effects of false alarms (e.g. inspection maintenance event, delays) and missed detections (leading e.g. to failures, cascading effects, delays / cancellations). Section 3 concerns with a meaningful application case of ROC-based cost-benefit analysis for determining cost saving potentials of various failure prediction algorithms regarding unscheduled engine removals. Finally, in section 4 we conclude and provide an outlook on future work in section 5.

Besides giving a practical guide for assessing the economic value of research approaches in failure prediction, this study is meant to give directions for industry decision making.

2. ROC-CURVE AND COST-BENEFIT ANALYSIS

As mentioned in the introduction, for a binary classification problem (true / false), there are four potential outcomes

1. Predicted as 'true', actual value is 'true', i.e. a 'True Positive' *TP*

2. Predicted as 'true', actual value is 'false', i.e. a 'False Positive' *FP*

3. Predicted as 'false', actual value is 'false', i.e. a 'True Negative' *TN*

4. Predicted as 'false', actual value is 'true', i.e. a 'False Negative' *FN*

Actual Positives *P* and Negatives *N* are hence respectively given by the sum of *TP* and *FN* as well as *TN* and *FP*. This leads to the definition of

True Positive Rate $TPR = \frac{TP}{P}$, (1)

False Positive Rate $FPR = \frac{FP}{N}$,

True Negative Rate $TNR = 1 - FPR$,

False Negative Rate $FNR = 1 - TPR$

Note that of the four rates only two are independent such that all rates may be expressed in terms of *TPR* and *FPR*. Accordingly, when applied to failure prediction, for instance

within a predefined time-window, this is associated with the following outcomes

1. Correct failure prediction: avoiding a potential unscheduled maintenance event and possible contingency damage costs due to cascading effects, allowing for timely planning of necessary MRO actions with estimated RUL as latest due date

2. False alarm: an impending failure is indicated, even though no failure is impending or it is reported early leading to unnecessary inspection costs / potential labor and logistic costs associated with component or system replacement for testing and resulting costs associated with Aircraft On Ground (AOG)

3. Correct prediction of normal operation, no positive or negative cost impact

4. Missed detection / failure: impending failure is not predicted or predicted late. In practice, the same consequence as a failure not covered by the prediction system leading to a potential unscheduled event, possible contingency damage costs, potential labor and logistic costs associated with component or system replacement and AOG-related costs

Typically, a failure prediction algorithm would associate each prediction with some instance probability or score (btw. 0 and 1) (Metz, 1978). Regarding positive (negative) predictions, the closer the score is to 1 (0), the higher is the algorithm's confidence in this classification result. In order to produce a discrete classifier output, it becomes evident that the operator has the freedom of choosing a decision threshold above / below which the prediction is rated as positive (i.e. failure, to the right of the threshold) / negative (i.e. normal operation, to the left of the threshold) (cf. Figure 1). Taking besides the two choices presented in the figure, all possible combinations of relative occurrences of correct / incorrect outcomes of the prediction (that are evaluated as such by later inspection of actual conditions of monitored components or systems) allows for the construction of the so-called Receiver Operating Characteristics (ROC) curve. Here, each point on the ROC curve corresponds to a different choice of classification threshold and subsequent evaluation of the relative occurrences of incorrect and correct predictions. Here, the ROC-curve is a key algorithm-specific performance measure (cf. Figure 2a)) (Metz, 1978) that enables evaluating the trade-off between *FPR* and *TPR*.

A measure for the algorithm's discriminability between positive and negative instances is given by the Area Under Curve (*AUC*). Minimal *AUC* corresponds to that of a random classification (i.e. $AUC = 0.5$), and maximal *AUC* to that of a perfect classifier (i.e. $AUC = 1$), allowing for an operating
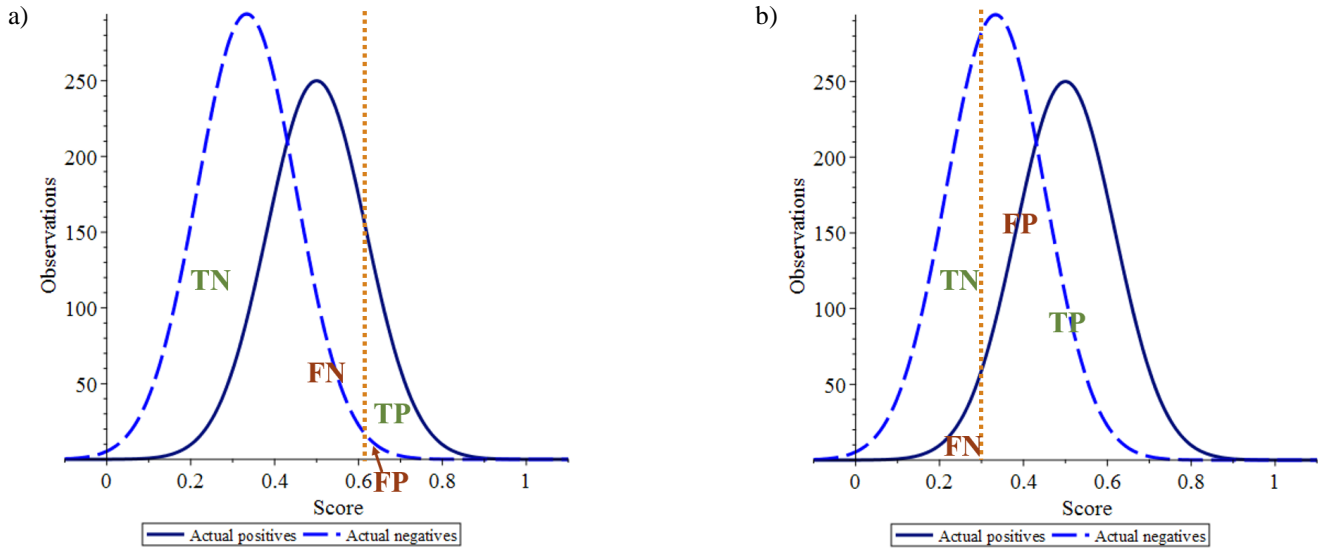
Figure 1. Distributions of actual positives (solid) and negatives (dashed) as a function of score with a) a rather strict decision threshold (dotted) leading to fairly low *FPR* and *TPR* and b) a rather lax decision threshold leading to fairly large *FPR* and *TPR*.

point on the ROC curve only associated with benefits ($TPR = 1$) and no penalties ($FPR = 0$) (cf. Figure 2a)) (Metz, 1978).

Considering that the ROC curve captures all possible combinations of correct / incorrect predictions, each of which is associated with specific actions / follow-up costs, the natural connection between the ROC-curve and cost-benefit analysis becomes apparent. The task is to find the optimal operating point on a ROC-curve that is associated with the best cost-benefit balance, promoting this choice to an application-specific business decision.

As concerns the average costs C for all possible prediction outcomes, these are given by (Metz, 1978)

$$C = C_0 + C_{TP} \cdot p(TP) + C_{FP} \cdot p(FP) \\ + C_{FN} \cdot p(FN) + C_{TN} \cdot p(TN), \qquad (2)$$

where C_0 summarizes overhead costs associated with the predictive maintenance solution (e.g. development, implementation and maintenance costs) and the other summands correspond to the average costs of each type of the four possible predictions, i.e. respectively, the costs of the prediction consequence, multiplied by the probability that this prediction occurs. Since true negative predictions (i.e. predictions of normal operation) are not associated with specific actions / follow-up costs ($C_{TN} = 0$), the last term vanishes. Furthermore, for instance $p(TP)$ corresponds to the occurrence probability of the failure mode p_{fail} multiplied by the probability that an actual failure will be predicted as such (i.e. *TPR*) such that $p(TP) = p_{fail} \cdot TPR$. Similarly, it is $p(FP) = p_{no} \cdot FPR$, $p(FN) = p_{fail} \cdot (1 - TPR)$ and

$p(TN) = p_{no} \cdot (1 - FPR)$, where p_{no} denotes the probability for normal operation.

The maximally achievable net benefit NB_{max} due to failure prediction results from the difference in conventional costs C_{ref} (without the specific failure prediction algorithm) and C_{min}, the minimal value of the cost function C,

$$NB_{max} = C_{ref} - C_{min} \qquad (3)$$

Accordingly, from analyzing the net benefit e.g. as a function of prediction error (*FPR*), the optimal operating point can be found as that leading to the maximally achievable value of the net benefit, NB_{max}.

In Figure 2b), an exemplary calculation of the relative net benefit as a function of prediction error (*FPR*) is shown, which assumes fixed follow-up costs for all possible prediction outcomes according to Eq. (2), the same prediction quality (cf. ROC-curve in Figure 2a)), but varying occurrence probability of a failure mode. Compared with the benefit due to correct failure prediction (*TP*), fairly high cost penalties are assumed for *FPs* and *FNs*. The reference is chosen such that

$$C_{ref} = C_{fail} \cdot p_{fail,} \qquad (4)$$

where $C_{fail} = C_{FN}$ and $p_{fail} = (p(TP) + p(FN))$. This case may be interpreted as typical for corrective maintenance, where failed equipment is typically only restored after a damage has occurred. Accordingly, the associated costs correspond to those arising from missed detections, C_{FN}, i.e. from unexpected damage for a system with failure prediction. This case is chosen for simplicity, since here a direct relation
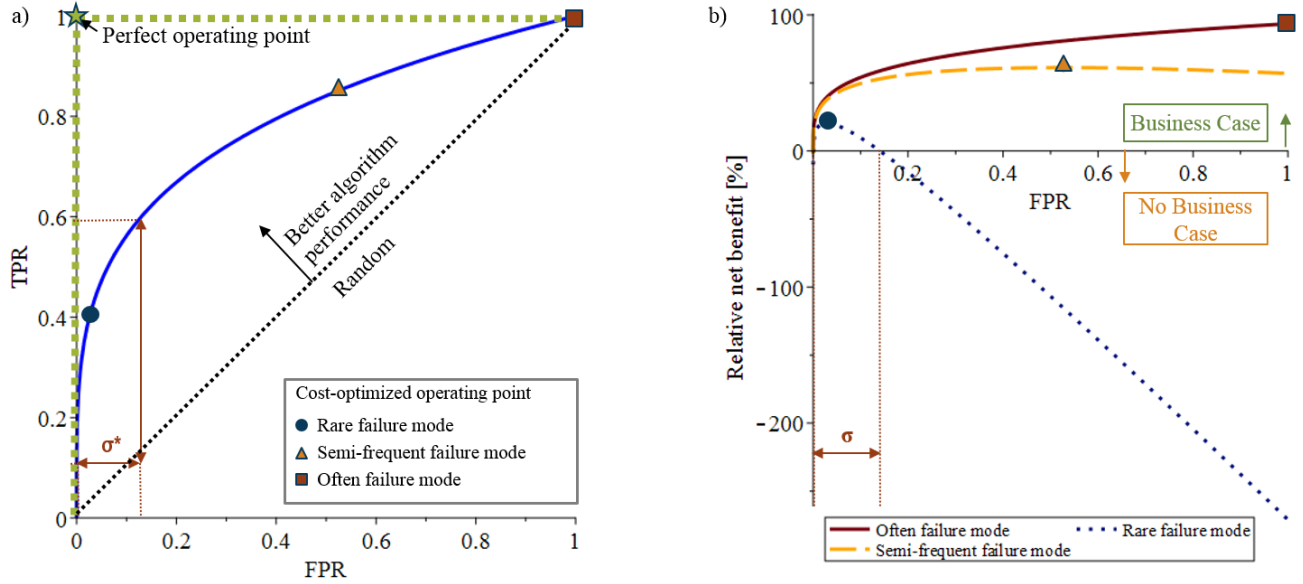
Figure 2 a) The ROC curve can be directly related to b) cost-benefit analysis of predictive analytics for specific business problems allowing for algorithm performance evaluation, optimization and finally, for identification of business cases with net relative benefit.

exists between the parameters specifying the costs with failure prediction and without (i.e., the reference).

It becomes apparent from Figure 2 that the net benefit achievable with the same ROC-curve is application specific: while for a semi-frequent and for an often-occurring failure mode, a business case is given for failure prediction irrespective of the prediction quality, for a rare failure mode, σ denotes the economically viable range of *FPR*. Furthermore, it becomes apparent that the choice of operating point is a business decision. Its optimal value minimizes costs due to prediction errors and maximizes benefits resulting from correct predictions e.g. of impending failures of components or (sub-)systems. For decreasing failure occurrence probability, the optimal operating point on the ROC curve (solution to Eq. (3)) moves to lower values of *FPR* and *TPR* (cf. Figure 2). This means here, a strict decision threshold is most favorable, while respectively a medium and lax threshold are best for medium and high failure occurrence probability. This is related with the fact that in case a failure mode is rare, almost all positive predictions will be false positive. Due to the assumed large cost penalties arising from false positive predictions, this implies that NB_{max} is associated with smaller values of *FPR* and hence also of *TPR* for lower failure event rate. Since in comparison, costs benefits due to correctly predicted failures occur less often and furthermore, lower values of *TPR* tend to enhance the negative effect of *FNs* (cf. Eq. (2)), the achievable overall relative benefit decreases with failure occurrences probability.

Some (further) general conclusions from ROC-based cost-benefit analysis may be drawn:

- In particular for rather poor prediction quality, the

choice of a strict / lax threshold is not only beneficial for (Metz, 1978) a rare / often failure mode, but also for

- $C_{FP} \gg C_{FN} - C_{TP}$ / $C_{FP} \ll C_{FN} - C_{TP}$, i.e., e.g. if failure prediction is of little benefit, but false alarms are very costly / if costs of actual failure resulting from a missed detection (e.g. due to cascading effects) are much larger than costs for timely maintenance / repair before actual failure occurs

- If conventionally, costs for unscheduled maintenance are comparatively high / low, allowable prediction error / overhead costs are comparatively high / low as well.

- Overall costs may increase, despite of good algorithm performance, if overhead costs C_0 are too high

3. APPLICATION EXAMPLE: ENGINE MAINTENANCE

Over its service lifetime, the majority of an aircraft's maintenance exposure arises from three main areas: airframe, engine and components. Making up a significant contribution of about 30-40% of the total maintenance expenses, expenditures arising from the engine exhibit an important impact on the market value of the whole aircraft at any given time (Ackert, 2011).

Regarding engine maintenance practices in aeronautics, there has been a shift in industry from fixed maintenance intervals towards engine on-condition monitoring. The aim is to remove engines only when internal components reach their individual life limits, or performance monitoring indicates

operation outside of parameter values suggested by manufacturers (Ackert, 2011).

In support of this paradigm, the ability has been improving of accurately predicting the time to failure (or Remaining Useful Life (RUL)) of various components. In particular, this enables engine removal from service for repair and/or refurbishment before secondary damage may result from failed parts. The further development of this capability is in particular in demand for failures of hot section components such as turbine blades, nozzles, rotor or combustor components that can induce high economical penalties arising both from turbine downtimes and from potential cascading effects inducing high down-stream damage costs (Pillai, Kaushik, Bhavikatti, Roy & Kumar, 2016).

Various factors such as operating conditions, specific material and manufacturing characteristics or environmental conditions can significantly influence the lifetime of components and are partly difficult to incorporate in a physics framework. Correspondingly, Pillai et al. (2016) firstly exploited a physics-based damage accumulation model based on Computational Fluid Dynamics (CFD) and Finite Element (FE) simulations that translates turbine operation data into the probability of failure of the considered components (by comparing estimated damage with a damage threshold expected to lead to a failure). Then, they fused the latter with data e.g. on manufacturing, geography and environment as well as customer and inspection information by means of machine learning techniques. This hybrid approach has been shown to allow for significantly improving predictive capability in failure detection of turbine blades e.g. regarding creep-driven cracking. This is manifested in a 60% increase in AUC of the respective ROC-curves for failure prediction of the hybrid physics-/data-based compared to the pure physics-based approach (Pillai et al., 2016).

In the following, it is demonstrated that ROC-based cost-benefit analysis can be applied to evaluate cost reduction potentials of such predictive approaches with regards to Unscheduled Engine Removal (UER) in dependence on the achievable failure prediction quality (as measured by the corresponding ROC-curve, cf. section 2). Possible extension of the approach to further optimize engine workscope planning e.g. with regards to maximizing time-on wing or minimizing the number of shop visits will be discussed in section 0.

3.1. Potential Analysis for Reducing Unscheduled Engine Maintenance Costs: ROC-based Approach

With the aim of assessing cost reduction potentials regarding UER based on failure prediction and the ROC-based approach outlined in section 2, in the next section the stage will be set for deriving quantitative results in section 3.1.2.

3.1.1. Setting the stage for ROC-based Cost-benefit Analysis

In general, the Shop Visit Rate (SVR) of an engine may be broken into the scheduled removal rate (e.g. resulting from expiry of Life-Limited Parts (LLPs), performance deterioration and service bulletin compliance) and unscheduled removal rate. The latter measures the number of times unexpected engine anomalies or failures require engine removal for repair or refurbishment before normal maintenance intervals are reached (Ackert, 2012). This causes a shop maintenance event with associated Shop Visit Costs *SVCs* and the necessity of installing an airworthy (new or repaired) engine.

The reciprocal of the total SVR is the engine's Mean-Time-Between Removals *MTBR*, another important reliability metric (Ackert, 2015).

As discussed in section 2, unexpected maintenance is more expensive than scheduled MRO actions based on knowledge of an impending failure. Besides potential Contingency Damage Costs *CDC*, Logistic Costs *LC* increase, if the engine needs to be replaced outside of the base owing to an in-service failure (Batalha, 2012). Here, one can discriminate two cases associated with decreasing occurrence probabilities, but increasing severity of economical penalties:

- On-ground occurrence or detection of engine failure with probability p_g and logistic costs LC_g
- In-flight occurrence of failure with probability $p_f < p_g$ and logistic costs $LC_f > LC_g$ due to potential engine In Flight Shut-Down (IFSD) that may cause the necessity to replace the engine in an alternate airport (Batalha, 2012).

Adding to this, in general, unexpected AOG is associated with a contribution loss *CL* (revenue-variable costs) (Batalha, 2012).

Assuming the UER event E_{UER} to happen with a certain event rate λ_{UER} per 1000 FH that is (roughly) constant, but completely at random, gives a Poisson process with probability p_{UER} growing as a function of time t (Batalha, 2012):

$$p_{UER} = \lambda_{UER} t e^{-\lambda_{UER} t} \qquad (5)$$

With conditional probability $p_f = p(E_{UER} | E_f)$ the engine failure is an in-flight event E_f. For a (roughly) constant in-flight occurrence rate λ_f this yields (Batalha, 2012)

$$p_f = \frac{\lambda_f t e^{-\lambda_f t}}{p_{UER}} \qquad (6)$$

In Figure 3, a schematic representation of all possible events with / without failure prediction with the associated follow-up costs and occurrence probabilities is given (cf. section 2

a)

b)

```
                                    ┌─────────────────────┐
                                    │   True Negative     │
                                    │     No costs        │
                                    │ (1-FPR)·(1-p_UER)   │
                                    └─────────────────────┘

        ┌──────────────────┐        ┌─────────────────────┐         ┌──────────────────────┐
        │ In-flight event E_f │     │   True Positive     │         │ In-flight event E_f  │
        │ SVC_f (CDC_f) LC_f, │     │        SVC          │         │ SVC_f (CDC_f) CL_f   │
        │        CL_f         │     │     TPR·p_UER       │         │   (1-TPR)·p_f        │
        │  p_f=p(E_f│E_UER)   │     └─────────────────────┘         │  p_f=p(E_f│E_UER)    │
        └──────────────────┘                                        └──────────────────────┘
```
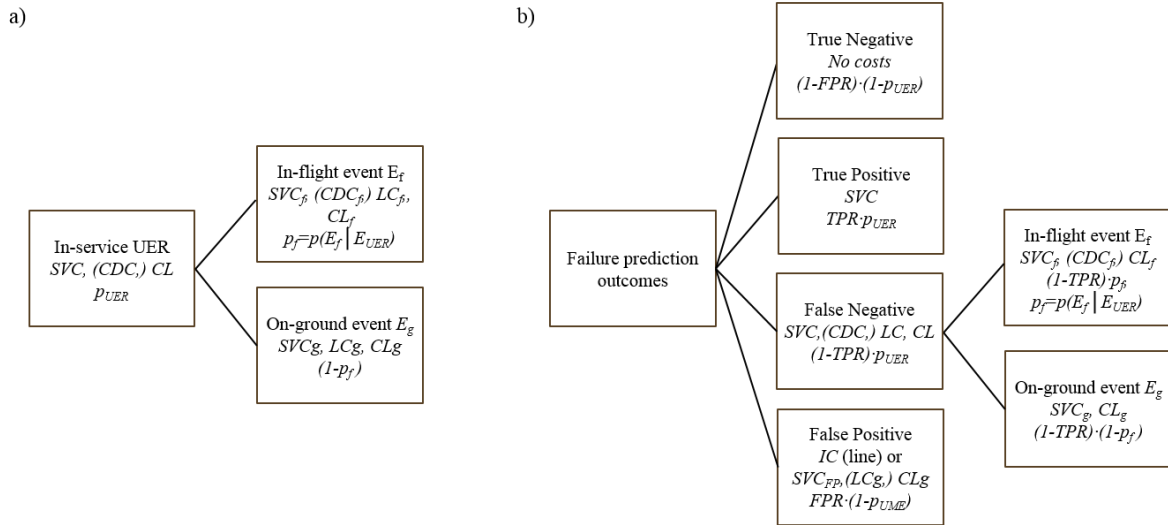
Figure 3. Schematic representation of all possibilities of events with (potential) associated follow-up costs and occurrence probabilities for a) the reference and b) the failure prediction scenario.

for the discussion of occurrence probabilities of all possible correct / incorrect predictions). Note that with failure prediction, the only source of UERs is provided by FN predictions. These, however, occur with a factor of $(1 - TPR)^{-1}$ lower probability than in the reference scenario. As concerns *FP* failure predictions, the severity of cost-penalties strongly depends on whether line inspections with comparatively low Inspection Costs *ICs* suffice for revealing them (e.g. excluding safety-critical crack-growth by borescope inspections) or whether the necessity of engine removal (potentially outside of the base with additional logistic costs depending on the determined RUL) and subsequent shop visit arises.

Engine removal causes are generally highly dependent on type of operation. In the following, a focus is placed on widebody engines and events caused by High Pressure Turbine (HPT) components such as HPT stage 1 and stage 2 blades. Creep is an important type of time-dependent degradation mechanism of turbine blades, while not being the only one (Pillai et al., 2016). Given the exploratory nature of this work and for the purpose of concreteness, the corresponding achievable prediction quality of creep-induced cracking of turbine blades presented by Pillai et al. (2016) is taken as representative for the prediction of HPT blade failures in the following. Refinements to this approach are left for future work.

Overhead costs for failure prediction (cf. section 2) are neglected in this study. Clearly, for economic viability, an upper limit will be given by the achievable net benefit of failure prediction.

Furthermore, as a reference, mature-run, widebody engines are considered, where the SVR can be taken to be stabilized (Ackert, 2012). Based on these assumptions, literature results elaborated on in the following will be used for specifying the corresponding typical UER rates and related follow-up costs necessary to quantitatively perform a ROC-based cost-benefit analysis as outlined in section 2. A collection of representative, literature-based values for various cost factors mentioned in Figure 3 can be found in Table 1.

Table 1. Various sources of costs with values based on Batalha (2012) for in-service engine removal due to on-ground / in-flight failure.

Costs	Estimates [k $]
$SVC_{g/f}$ Shop visit costs after on-ground / in-flight failure	12-120
Potential contingency damage costs CDC_f due to in-flight failure	500
Potential (line) inspection costs IC due to a false alarm (FP)	0.2
SVC_{FP} potential SVCs due to a false alarm (FP)	2
$LC_{g/f}$ Logistic costs to replace an engine outside the base due to on-ground / in-flight failure (e.g. at alternate airport due to IFSD)	100 / 250
$CL_{g/f}$ loss of contribution (revenue – variable costs) during AOG time due to on-ground / in-flight engine failure	266 / 372

Here, a typical cost split into 60-70% material costs, 20-30% labor costs and 10-20% repair costs (Ackert, 2015) has been assumed to estimate the respective *SVCs* for turbine failures as well as *ICs* potentially associated with *FPs*. Furthermore, the estimate of $CL_{g/f}$ assumes a contribution loss of 14 FH per day and respectively 5 and 7 days AOG for engine replacement after on-ground and in-flight failure (Batalha, 2012).

6

Typically, for the considered failure mode, the failure rate would grow with the number of flight hours since the component has come into service or since last repair. However, as will be demonstrated in section 3.1.2, the net average benefit depends only mildly on variations in the event rate for sensible parameter ranges.

Typical event rates for UER caused by turbine blades in different stages after last shop visit (respectively, after a couple of years in operation as well as soon after coming into service / after repair) are extracted from Ackert (2012) for a widebody engine. Respectively, they take values of 3.6 10^{-3} and 7.0 10^{-4} per 1000 FH for HPT stage 1 blades as well as 1.2 10^{-3} and 3.0 10^{-4}, per 1000 FH for HPT stage 2 blades. In total, this respectively amounts to 17 % as well as 6% of the total UER rate, which has a typical value of 0.026 per 1000 FH for a mature-run, widebody engine (Ackert, 2012). The total SVR amounts to 0.032 per 1000 FH, corresponding to a MTBR of 31,250 FH. Furthermore, the total IFSD rate is about 5 10^{-3} per 1000 FH (Batalha, 2012).

Based on the parameter presented in this section, a sensitivity analysis will be performed in the next section, in order to single out the most important contributions influencing the achievable net average benefit resulting from the prediction of turbine failures.

3.1.2. Results: ROC-based Cost-Benefit Analysis

This section demonstrates the value of ROC-based cost-benefit analysis for assessing cost-reduction potentials regarding unscheduled engine removals (UER) achievable by means of failure prediction algorithms. In general, their influence on the relative benefit in relation to the reference without UER failure prediction depends on

- Event rates and corresponding failure probabilities as analyzed by varying event rates for mature-run, widebody engines within sensible limits (cf. the last section) and by considering the time dependence according to Eqs. (5) – (6)
- Various costs associated with all possible cases with and without failure prediction (cf. Figure 3) as analyzed by a variation of those costs within sensible limits
- Relative occurrence of all possible correct and incorrect predictions as analyzed by varying the operating point on the considered ROC curves
- Prediction quality as analyzed by considering three different ROC curves (cf. Figure 4), two of which correspond to those achievable for creep-induced turbine blade failures deduced in Pillai et. al (2016) (hybrid approach fusing physic and data (LASSO model) as well as pure physics-based approach, respectively). The better of the two is approximated by a continuous function (referred to as ROC 1) that is considered first for cost-benefit analysis.

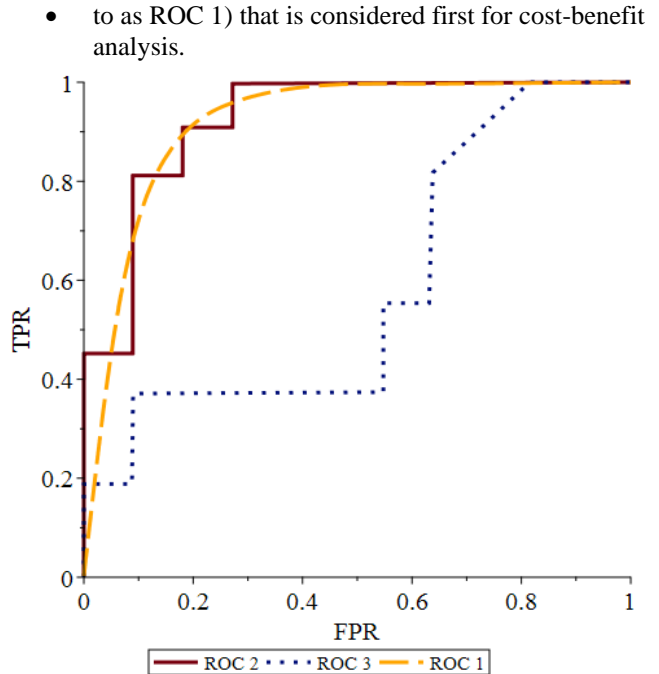- to as ROC 1) that is considered first for cost-benefit analysis.

Figure 4: Considered ROC curves for failure prediction, ROC2 and ROC3 respectively corresponding to a hybrid physics-/data-based as well as pure physics-based approach (Pillai et al., 2016).

Key results of the analysis are subsequently presented. These emerge from using Eqs. (2) – (3), the costs specified according to Figure 3, the failure occurrence probabilities according to Eqs. (4) – (5) as well as realistic (ranges of) parameters discussed in section 3.1.1.

In Figure 5, for various cost scenarios, the relative net benefit achievable by the prediction of turbine blade failures (with performance according to ROC 1) compared with the reference (no UER failure prediction) is shown as a function of *FPR* (i.e. representing all possible operating points on the ROC-curve).

Here, a fixed failure rate of 17% of the total unscheduled events rate is assumed. For a widebody aircraft, this event rate is typical for events caused by HPT 1st and 2nd stage turbine blades a couple of years after last shop visit (cf. the discussion in the last section). As demonstrated by Eqs. (5) – (6), for fixed event rate the probability of occurrence of an unscheduled event grows with flight hours, e.g. since last shop visit. In Figure 5, this effect is taken into account by integrating over flight time with $T_{min} = 0$ and $T_{max} = MTBR$. This approach allows determining the optimal operating point on the ROC curve (i.e. the most cost-efficient relative occurrence of *TP* and *FP* for ROC 1) for the respective scenarios considered. It turns out to exhibit little dependence on time for the considered scenarios. Thereafter, the influence of growing failure probability with time will be
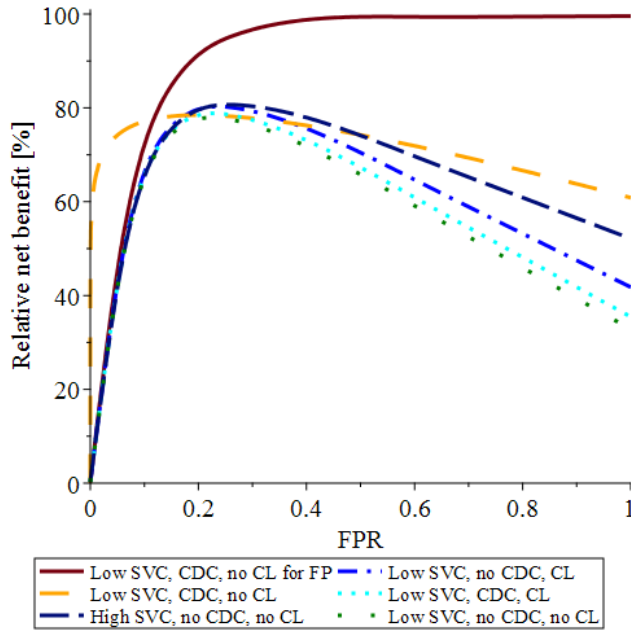
Figure 5. Impact of various cost terms on relative net benefit achievable per widebody engine per MTBRs.

analyzed for the respective optimal operating points on the ROC-curve that are roughly constant with time. One observes from Figure 5 that for all cost scenarios a business case for failure prediction emerges, since the relative net benefit is positive irrespective of the choice of the decision threshold (i.e. for all possible combinations of *FPR* and *TPR*). Furthermore the achievable relative net benefit increases both with shop visit costs *SVC* and contingency damage costs *CDC*, since both tend to increase reference costs and hence make failure predictions more valuable.

Yet, if false positive predictions are assumed to always require an engine removal, then for fixed prediction quality, the relative net benefit will decrease with the resulting AOG costs.

In contrast, if line inspections would suffice to detect false positive failure predictions without requiring engine removals, then the achievable relative net benefit would be superior to all other cases, reaching values as high as almost 100% for $FPR = 1$ and $TPR = 1$. Since in this case, false positive predictions are not associated with large follow-up costs, here, the penalties associated with incorrect predictions mainly arise from missed detections (i.e. false negatives). However, their relative occurrence probability decreases with increasing TPR. Hence, since choosing an operating point on the ROC curve with $FPR = 1$ allows for high $TPR = 1$, this explains why a large relative benefit is achievable compared to the reference. The respective maximally achievable relative net benefit for the six considered scenarios is summarized in Table 2 together with the corresponding optimal operating point on the ROC curve. While some of the scenarios where chosen in order to demonstrate the influence

of the various costs terms, scenarios 3), 5) and 6) are considered as meaningful options that will be further pursued in the following.

As mentioned before, the failure rate for the considered failure mode would typical increase with flight time since last shop visit. In the following, this effect is demonstrated to have little to hardly any impact on the achievable relative net benefit for the meaningful cost scenarios.

Table 2. Optimization of relative net benefit (cf. Figure 5 and Table 1).

Scenario	Optimized relative net benefit [%]	Optimized FPR, TPR
1) Low SVC, no CDC, no CL	78.00	0.23, 0.94
2) Low SVC, CDC, no CL	78.46	0.20, 0.91
3) Low SVC, CDC, CL	78.85	0.23, 0.94
4) High SVC, no CDC, no CL	80.68	0.26, 0.95
5) Low SVC, no CDC, CL	80.28	0.24, 0.95
6) Low SVC, CDC, no CL for FP	99.6	1.00, 1.00

For this purpose, the influence of varying the failure rate (i.e. occurrence probability of the considered failure mode) is analyzed respectively for cost scenarios 3) and 6) (cf. Table 2), while apart from the failure probability all other parameters are kept fixed.

Typical fractions of the total unscheduled event rate in different stages after last shop visit are considered (cf. the discussion in the last section). Furthermore, as an upper limit on the specific event rate of the considered failure mode, the total unscheduled event rate is taken accounting for all possible failure causes. Moreover, as above, two extreme cases regarding the costs associated with false alarms are considered: firstly assuming the necessity of engine removal and subsequent shop visit (i.e. scenario 3)), secondly assuming that all false positive failure predictions are discovered e.g. by borescope inspections, while the aircraft is on ground (i.e. scenario 6)).

In general, follow-up costs associated with false positive predictions grow in proportion to $(1-p_{UER})$, *FPR* and the corresponding cost penalty (cf. Eq. (2)).

From Figure 6, it becomes evident that in the first case, the maximally achievable relative net benefit in relation to the reference varies only slightly with failure occurrence probability in the full range of conceivable probabilities, taking values between maximally 78 % (lowest assumed event probability) and 82 % (highest assumed probability).
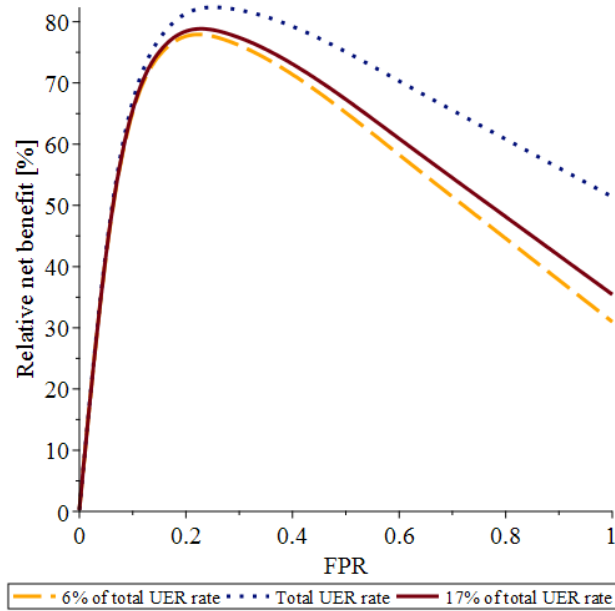
Figure 6. Impact of event rate on relative net benefit within MTBRs for scenario 3) as specified by Table 1 – Table 2.

Figure 7. Impact of event rate on relative net benefit within MTBRs for scenario 6) as specified by Table 1 – Table 2.

Furthermore, the optimal operating point on the ROC curve moves to lower values of *FPR* and *TPR* for decreasing failure occurrence probability. The reason is that if a failure mode is rare, almost all positive predictions will be false positive. Owing to the assumed large cost penalties arising from false positive predictions causing engine removal and associated downtime, this implies that the tolerable *FPR*-rate decreases with the failure event rate at the dispense of also decreasing *TPR*. In total, the positive effect of true positive predictions is less frequently coming into play for reduced event rates and furthermore, lower values of *TPR* tend to enhance the negative effect of missed detections (cf. Eq. (2)) such that the overall relative benefit decreases in comparison to that of larger event rates.

In the second case, the situation is quite different. Here varying the failure event rate has hardly any impact on the achievable relative benefit and for the whole spectrum of chosen event rates an operating point on the ROC curve with $FPR = 1$ and $TPR = 1$ is optimal. Clearly, this is due to the fact that false alarms lead to fairly negligible cost penalties such that a high *FPR* can be tolerated, correspondingly leading to a high *TPR* that minimizes the occurrence of missed detections (i.e. *FNs*) and their cost penalties and optimizes the achievable benefit due to true positive predictions. Thereby, almost independent of event rate, a net relative benefit of 100% is achievable.

In the following, the event rate is kept fixed at a typical value of 17% of the total UER rate (Ackert, 2011), while the failure prediction performance is varied. For this purpose, all ROC curves in Figure 4 are considered. Here, the second and third ROC curves correspond to the best (hybrid physics- and data-
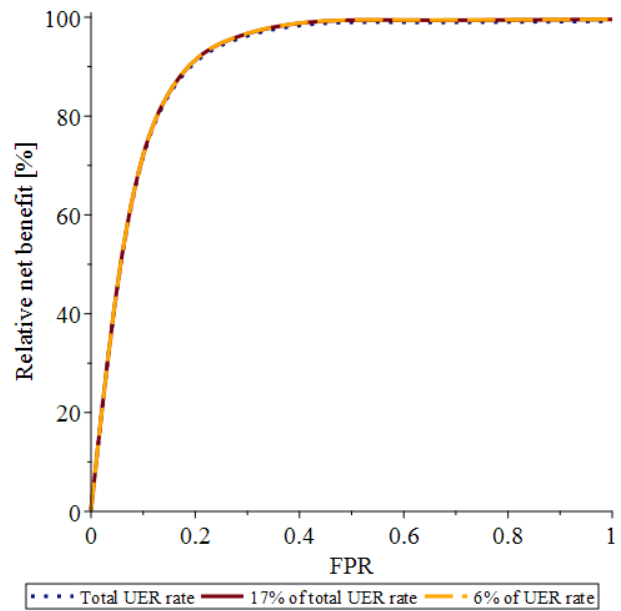
based) and the worst (pure physics-based) cases respectively found in Pillai et al. (2016). Note that also for these ROC curves, it has been verified that the relative net benefit depends only mildly on deviations from the assumed typical event rate for HPT blade failures within sensible limits.

For the meaningful scenarios 3), 5) and 6) (cf. Table 2), a comparison of the achievable (relative) net benefit for all three ROC curves as a function of prediction error (*FPR*) can be found in Figure 8 –Figure 10. In Figure 8a) – Figure 10a), similarly to the approach taken in the calculation for Figure 5, it is assumed that $T_{min} = 0$ and $T_{max} = MTBR$ for the time integration yielding the cumulative relative net benefit in dependence on the prediction error (*FPR*).

In Figure 8b) – Figure 10b), for the optimal operating point of the respective ROC curves, the increase of the net benefit with time T in units of 1000 FH is plotted by integrating from $T_{min} = 0$ up to T, with $T_{max} = MTBR$.

In particular, this calculation would turn out useful for optimizing shop visit intervals as it demonstrates the time dependence of the reduction potential of unscheduled engine removals for the considered failure mode. In addition, the respective optimal operating points and resulting values for the maximally achievable (relative) net benefit are summarized in Table 3.

It becomes apparent that for cost scenarios 3) and 5) the results are qualitatively similar. First of all, the prediction performance as measured by the respective ROC curves, has a sizeable impact on the achievable net benefit. While the first and the second ROC curve for both cost scenarios allow for relative net benefit around 80% with somewhat larger values
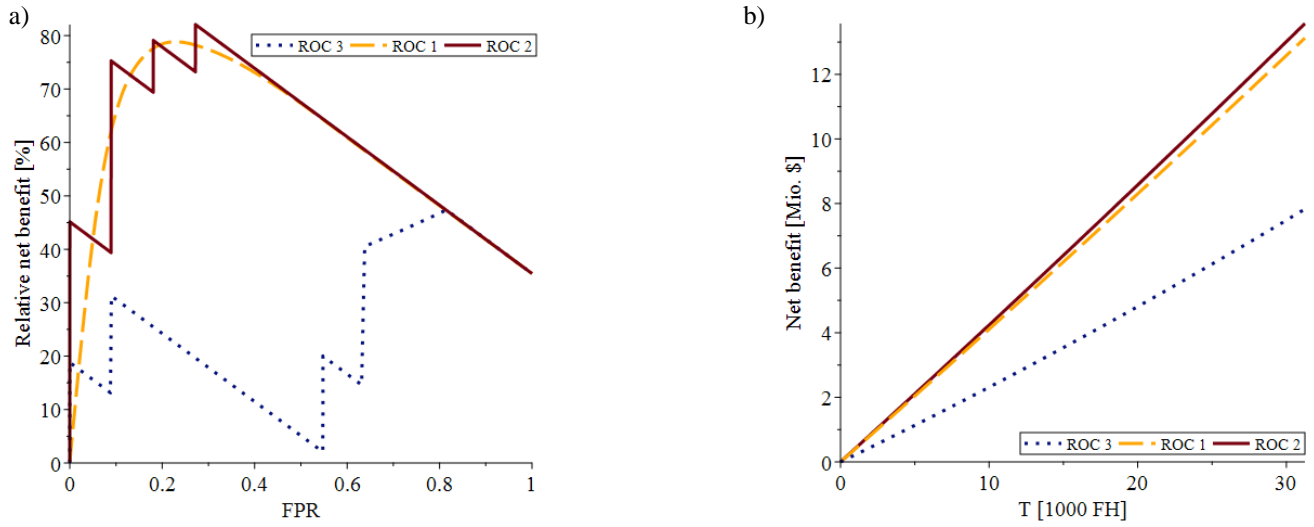
Figure 8. Cost scenario 3), a) relative net benefit as a function of FPR and b) net benefit as a function of T.
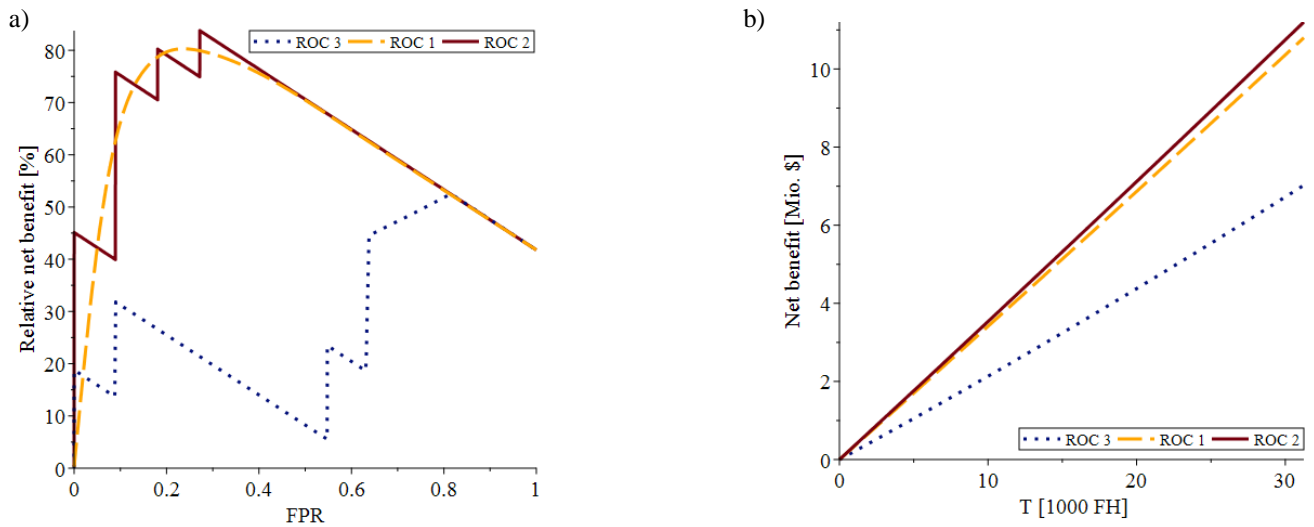

Figure 9. Cost scenario 5), a) relative net benefit as a function of *FPR* and b) net benefit as a function of T.
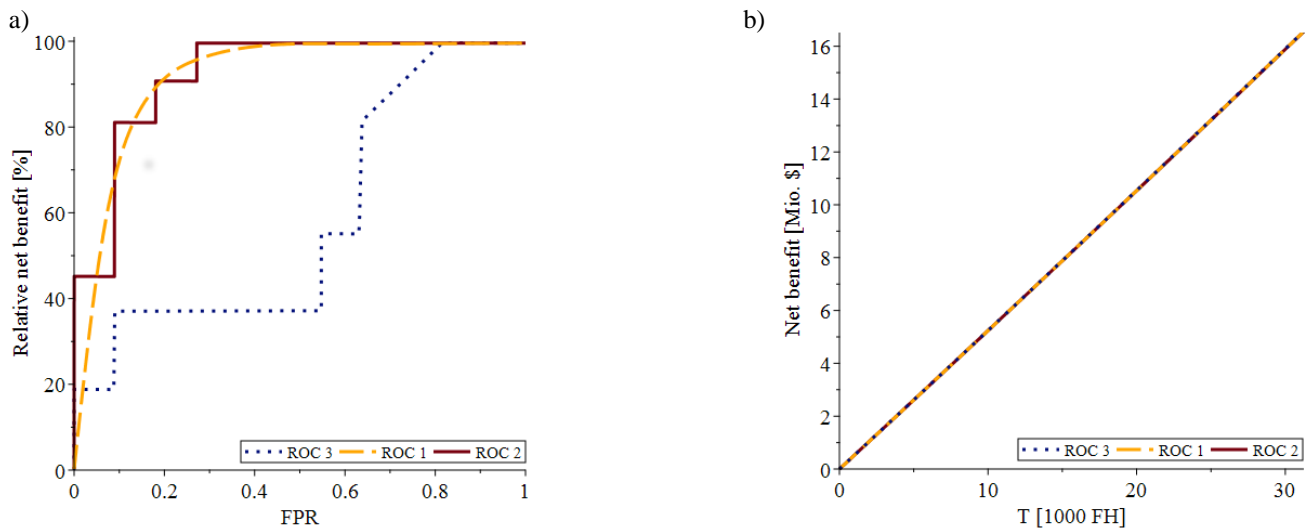

Figure 10. Cost scenario 6), a) relative net benefit as a function of *FPR* and b) net benefit as a function of T.

Table 3. Optimized results for costs scenarios 3), 5) and 6).

Cost scenario	Prediction performance	Costs reference [Mio. $]	Optimized relative net benefit [%]	Max. net benefit [Mio. $]
3)	ROC 1	16.65	78.90	13.13
	ROC 2		80.54	13.57
	ROC 3		47.04	7.83
5)	ROC 1	13.45	80.29	10.80
	ROC 2		83.29	11.20
	ROC 3		52.22	7.02
6)	ROC 1	16.65	99.50	16.56
	ROC 2		99.50	16.56
	ROC 3		99.50	16.56

in scenario 5) compared to scenario 3), the third ROC curve (corresponding to the pure physics-based approach for failure prediction) only yields about 47% and 52%, respectively.

As typical for a ROC curve with poor classification performance, there furthermore exists no cost optimum at low / high values of FPR / TPR as in the other two cases with good ROC performance.

Instead, the best operating point on ROC 3 lies at $FPR = 0.82$ and $TPR = 1$ for all cost scenarios. This corresponds to the case where the benefit due to true positive predictions is maximized and the negative cost impact of missed detections proportional to $(1 - TPR)$ is minimized, but traded for a fairly high cost penalty due to large false alarm rate (i.e. FPR). For ROC 2, there exists a cost optimum at significantly smaller FPR of 0.28 in all scenarios, corresponding to a TPR of close to 1, leading to a factor of about 1.7 / 1.5 (scenario 3) / 5)) higher optimal relative net benefit compared with that resulting from ROC 3.

The corresponding cost saving potentials per MTBRs are with about 13.57 and 11.20 Mio. $ for the hybrid physics-/data-based approach (corresponding to ROC 2) respectively about 5.75 Mio. $ (cost scenario 3)) and 4.18 Mio $ (cost scenario 5)) higher than the pure physics-based approach, impressively demonstrating the value of data for predictive maintenance purposes.

Note that while the relative net savings in scenario 3) are somewhat lower than that in scenario 5), the net benefit is slightly higher. The reason is that the assumed large contingency damage costs in this case drive up the reference costs arising from UERs from 13.45 Mio $ (scenario 5)) to 16.65 Mio. (scenario 3)) within MTBRs such that the relative net benefit in scenario 3) is somewhat lower than in scenario 5) (cf. Table 3). The situation is quite different for scenario 6) (cf .Figure 10). Here, due to the fact that false alarms are assumed to lead to comparatively low cost penalties, for all three ROC curves, the optimal operating point lies at $FPR = 1$ and $TPR = 1$. Accordingly, fairly irrespective of prediction

performance, a relative net benefit of about 100% can be achieved corresponding to a cost saving potential within MTBRs per widebody engine of 16.56 Mio. $ in all cases (cf. Table 3).

4. CONCLUSIONS

In this study, the value of ROC-based cost-benefit analysis for identifying and optimizing cost saving potentials associated with predictive maintenance applications was first generally discussed. The approach was thereafter applied to a representative use case within aircraft engine maintenance: potentials for reducing unscheduled engine removals by (more) accurate failure prediction were quantitatively assessed from an operator's perspective, for realistic, literature-based ranges of costs, failure occurrence probabilities and algorithm performances. A focus was placed on events caused by turbine blade failures to make contact with literature results on the achievable prediction quality, considering both a pure physics-of-failure-based approach and a hybrid physics-/data-based one with superior prediction performance (Pillai et al., 2016).

As a key result of this analysis, for sensible parameter ranges, the more accurate prediction of turbine blade failures was generally found to allow for significant cost savings. These extend up to roughly 17 Mio. $ per widebody engine and per MTBRs (i.e. a relative net benefit of 100% compared with current practice), neglecting any overhead costs for the failure prediction system itself. The highest parameter influence on the achievable net benefit was identified to stem from false alarms. Associated cost penalties can grow from comparatively low to high, in case for the identification of false positive failure predictions line (borecope) inspections do not suffice, but an engine removal and subsequent shop visit are required. If follow-up costs of false alarms are low, the net benefit of failure prediction is maximized (about 17 Mio. $ per widebody engine and per MTBRs, cf. above) and turns out to be independent of prediction quality. However, for all other considered cost scenarios, the hybrid physics-/data-based approach yielded a significantly higher net benefit than the pure physics-based one of up to about 42%. Accordingly, as a further key result of this study, the additional use of relevant data e.g. on environment and inspections, pertaining to factors that are not easily modeled using physics principles, was found to be worth up to about 6 Mio. $ per widebody engine and per MTBRs. This quantitative result impressively demonstrates the value of data for predictive maintenance purposes.

5. OUTLOOK

While in this study, a focus was placed on the impact of turbine failure prediction on unscheduled engine removal, the approach may be extended to include that associated with other failure modes and to select the corresponding most cost-efficient prediction algorithms. The results may be exploited

in order to further optimize engine workscope planning e.g. with regards to maximizing time-on wing or minimizing the number of shop visits based on more accurate failure predictions. For instance, shop visit costs increase with time on wing due to deteriorating engine condition (Ackert, 2011). Yet, performing the shop visit at a later point in time e.g. may result in discounted cash flow savings (Batalha, 2012). Furthermore, in case a significant fraction of scheduled tasks may be eliminated by means of reliable failure prediction algorithms, the total workload and potentially also the downtime due to maintenance checks could be reduced. Here, a ROC-based cost-benefit approach allows for optimally trading all involved cost factors for optimized workscope planning in dependence on failure prediction quality.

Finally, the approach may also be applied to assess cost-saving potentials from failure prediction for aircraft systems other than the engine as investigated in on-going work.

REFERENCES

Ackert, S. (2011). *Engine Maintenance Concepts for Financiers: Elements of Turbofan Shop Maintenance Costs.* Tech. rep. San Francisco, CA: Aircraft Monitor. http://www.aircraftmonitor.com/uploads/1/5/9/9/15993320/engine_mx_concepts_for_financiers___v2.pdf

Ackert, S. (2012). *Basics of Aircraft Maintenance Reserve Development and Management.* Tech. rep. San Francisco, CA: Aircraft Monitor. http://www.iata.org/whatwedo/workgroups/Documents/Paperless%20Supply%20Chain/Basics-AC-MR.pdf

Ackert, S. (2015). *Keeping Score: Analysis of an Engine's Shop Visit Rate.* Tech. rep. San Francisco, CA: Aircraft Monitor. http://www.aircraftmonitor.com/uploads/1/5/9/9/15993320/keeping_score_analysis_of_an_engine%E2%80%99s_svr____v1.pdf

Batalha, E. da C. P., (2012). *Aircraft Engines Maintenance Costs and Reliability: An Appraisal of the Decision Process to Remove an Engine for a Shop Visit Aiming at Minimum Maintenance Unit Cost.* Master Thesis. The Universidade Nova de Lisboa, Lisbon, Portugal.

Metz, C. E.. (1978). Basic Principles of ROC Analysis. *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283-298. doi.org/10.1016/S0001-2998(78)80014-2

Pillai, P., Kaushik, A., Bhavikatti, S., Roy, A. & Kumar, V. (2016). A Hybrid Approach for Fusing Physics and Data for Failure Prediction. *International Journal of Prognostics and Health Management*, pp. 2153–2648.

Saxena, A., Sankararaman, & Goebel, K. F. (2014). Performance Evaluation for Fleet-based and Unit-based Prognostic Methods. *Proceedings of European Conference of the Prognostics and Health Management Society*, July 8-10, Nantes, France.

Wagner, C., Saalmann, P., & Hellingrath, B. (2016). An Overview of Useful Data and Analyzing Techniques for Improved Multivariate Diagnostics and Prognostics in Condition-Based Maintenance. *Proceedings of Annual Conference of the Prognostics and Health Management Society*, October 3-6, Denver, Colorado, US.