CBM / PBM Operational Performance: Concepts & Applicable Formulas For Aligning Model Performance Metrics With Operational Ones At Failure Mode And Equipment Levels

Franck DESSERTENNE[1]

[1] *'Data Driven Services & Maintenance' Department*

Airbus Operations SAS, 316 route de Bayonne, 31060 TOULOUSE, FRANCE

franck.dessertenne@airbus.com

Abstract

Condition-Based Maintenance (CBM) and Predictive-Based Maintenance (PBM) have emerged as pivotal strategies in the aircraft industry, offering to revolutionize maintenance operations by optimizing schedules, reducing costs, and enhancing safety through Health Indicators (HIs). Realizing the full benefits of these approaches requires the ability to accurately measure the performance of both predictive models and the equipment itself.

However, widespread implementation faces challenges, as the concept of "performance" is often interpreted differently by various stakeholders. This paper addresses the limitations of conventional metrics like **MTBUR** (Mean Time Between Unscheduled Removals), which is no longer directly applicable in a CBM / PBM context without biasing performance calculations. The core ambiguity arises because proactive interventions are also "unscheduled," which wrongly penalizes the perceived success of a predictive program.

To resolve this, this paper advocates for a paradigm shift by proposing a new suite of operational metrics. The primary proposal is the adoption of **MTBRR (Mean Time Between Reactive Removals)**, used in conjunction with the classic **NFF (No Fault Found) rate**, to clearly distinguish true failures from non justified reactive removals.

These core metrics are then supplemented by a set of new indicators designed for a predictive context: **MTBPR** (Mean Time Between Predictive Removals), **NDF** (No Degradation Found) rate, and **MLR** (Mean Lifetime Reduction).

Furthermore, the paper delves into the crucial relationship between model and operational performance, demonstrating how standard data analytics metrics like **Recall** and **Precision** can be formally linked to these new equipment-level operational metrics. By establishing these connections, this work provides a generalized and robust framework that enables all stakeholders - including suppliers, Maintenance, Repair and Overhaul (MRO) providers, and operators - to define performance objectives, accurately monitor in-service performance, and foster clear, data-driven alignment.

1. Introduction

In the framework of maintenance operations, different maintenance strategies have been built following the evolution of technologies.

In particular CBM and PBM strategies, heavily reliant on the effective selection, monitoring, and analysis of Health Indicators, represent a significant advancement in maintenance practices. They move away from reactive and time-based approaches towards a more proactive and data-driven methodology, ultimately delivering substantial value to end users through improved equipment reliability, reduced costs, and enhanced operational efficiency. The ability to accurately assess the health of equipment and predict potential failures empowers organizations to make informed maintenance decisions at the right time.

The figures 1 and 2 illustrate a hierarchical view of the different maintenance strategies.

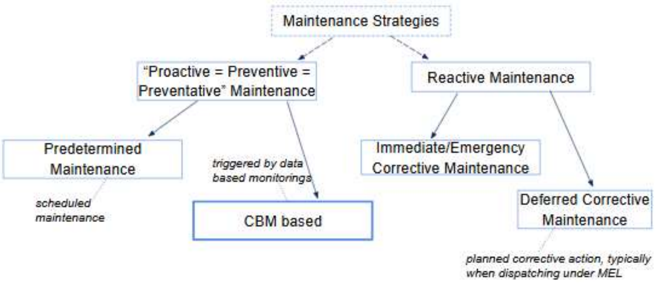

Figure 1. Hierarchy of maintenance strategies - part 1

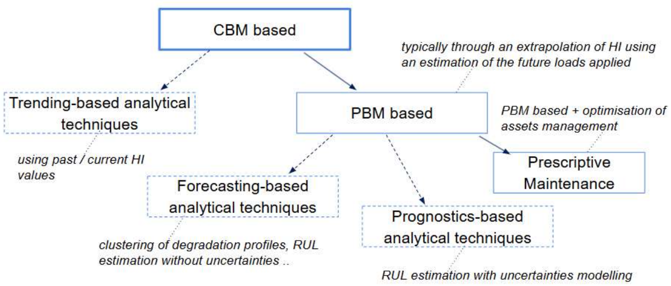

Figure 2. Hierarchy of maintenance strategies - part 2

In this work, we propose a modification of the way classical KPIs are estimated, such as the well known Mean Time Between Failures (MTBF). The introduction of new metrics is presented as key to take into account the new paradigm of 'Recognized Degradation Zone' of an equipment before reaching an actual failure. Focusing on their applicability at Failure Mode and equipment level, this paper allows to reconcile the conventional metrics of performance of predictive models like the Precision and Recall with operational metrics like the well-known MTBF. This way, a practical process is proposed to support without bias the in-service measurement of the operational performance.

In order to illustrate the concepts, and even if what follows is applicable regardless of the techniques used, we illustrate the scope of applicability with a prognostics approach based on Health Indicators.

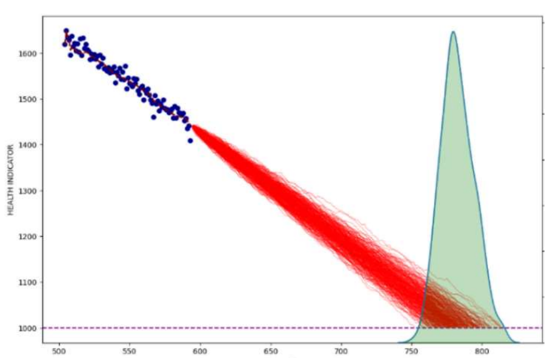

Figure 3. HI trajectories with uncertainties supporting the RUL PDF (Probability Density Function) building

This paper is organized as follows. The section 2 deals with the problem statement, showing that new concepts are needed with associated KPIs to be considered and officialized for the whole ecosystem of stakeholders. The section 3 addresses the impacts of such new concepts at Failure Mode and at equipment level, and the new set of formulas applicable when using a CBM / PBM approach. Section 4 then deals with the impacts on the in-service measurement process required to properly assess the performance through operational metrics. The conclusion eventually summarizes the most important formulas to consider as a take away and the perspective of next steps.

2. PROBLEM STATEMENT

In a CBM / PBM context, the formula below (traditionally and historically used within an unscheduled maintenance strategy) is no longer directly applicable without biasing the in-service MTBF computation (using measured NFF rate and MTBUR), and requires generalization.

$$\frac{1}{MTBF} = \frac{1-NFF}{MTBUR} \quad (1)$$

The ambiguity of the MTBUR metric stems from its origins in a traditional maintenance paradigm, which consisted of two distinct categories: Scheduled Maintenance (time or usage-based) and Unscheduled Maintenance (reactive work following a failure). In this model, "unscheduled" and "reactive" were effectively synonymous.

However, the adoption of CBM and PBM strategies disrupts this binary classification. Removals or repairs triggered by predictive alerts present a fundamental conflict with this model:

- They are not reactive, as they preempt functional failures.
- They are not scheduled in the traditional sense, as they do not appear on a long-term maintenance plan. Instead, these interventions are triggered by emerging data, transforming an unscheduled need into a planned and scheduled task, often on short notice.

This results in a critical ambiguity. Classifying a proactive removal from a CBM or PBM program as "unscheduled" inherently lowers the MTBUR metric. Consequently, the maintenance program is misleadingly penalized for its own success, creating the illusion of declining reliability while failures are actually being preempted.

To resolve this paradox, we introduce and utilize the **MTBRR** (Mean Time Between Reactive Removals). This metric eliminates the ambiguity by classifying removals based on their direct operational impact - a reactive removal or not - instead of their status as a scheduled or unscheduled calendar event.

Therefore without CBM / PBM, (1) is equivalent to:

$$\frac{1}{MTBF} = \frac{1-NFF}{MTBRR} \qquad (2)$$

2.1. Key concepts to be considered

Traditional reactive-based maintenance strategies operate on a binary model, typically defined with equipment suppliers, that recognizes only two operational states: 'Nominal' and 'Faulty'. The transition to the 'Fault Zone' is defined by a threshold corresponding to fault effects detectable by test benches and/or associated with in-service effects (e.g. cockpit alerts). Once an asset crosses this threshold and the fault is confirmed, a process of troubleshooting and replacement is initiated.

In the context of a CBM / PBM approach, based on a large usage of Health Indicators, a paradigm shift is required, involving new concepts and KPIs to be formalized.

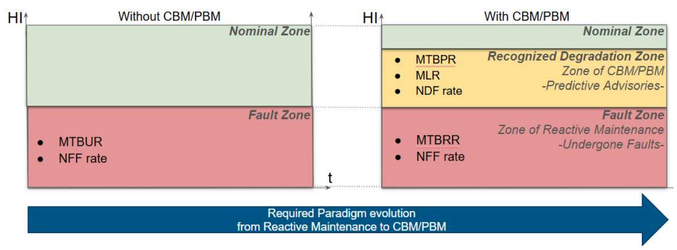

Figure 4. New concepts with corresponding operational metrics

The foundational step for implementing a CBM / PBM strategy is to evolve from the traditional binary health model (GREEN/RED) by formally defining a three-state model. This model introduces an intermediate AMBER zone, representing a zone of detectable degradation that is contractually recognized by the equipment supplier as justifying an early maintenance action before a fault manifests to the operator.

Of course the decomposition of the health indicator excursion into three zones - GREEN (Nominal zone), AMBER (Recognized Degradation Zone), and RED (Fault Zone) - requires unambiguous definitions for its boundaries.

The lower boundary (AMBER/RED), or Fault Threshold, is classically well-defined. It corresponds to a level of degradation (or functional loss with inability to operate properly) that triggers reconfigurations or redundancy loss with cockpit alarms or other 'Flight Deck Effects' and is determined through monitoring thresholds calibrated to reach specified objectives at aircraft level. A crucial aspect is that this threshold must be consistent with existing diagnostics systems, such as Maintenance Messages, to ensure a coherent health status across all monitoring platforms.

By contrast, the upper boundary (GREEN/AMBER) has historically lacked a formal definition. To make predictive maintenance viable, this boundary must also be unambiguously defined. It represents the precise level of degradation at which an early predictive maintenance action is both technically justified and contractually acceptable. Operationally, this is the point where a "weak signal" allows for the computation of a reliable Health Indicator (HI). Commercially, crossing this threshold is the key trigger for warranty applicability on a predictive removal, contractually obligating the supplier to endorse the cost of the early maintenance action during the warranty period.

Secondly, measuring operational performance in this new context requires a two-pronged approach. We must first adapt existing maintenance concepts and introduce clearer operational metrics like **MTBRR** and the associated **NFF** (No Fault Found) rate. Furthermore, it is essential to establish additional operational metrics that provide a common performance framework for the entire ecosystem. This includes stakeholders across the value chain, from equipment suppliers and system designers to MRO providers and data analytics specialists.

- The **MTBPR** (Mean Time Between Predictive Removals) is the operational performance metric for the AMBER zone, just as the **MTBRR** (Mean Time Between Reactive Removals) is for the RED one.
- The **MLR** (Mean Lifetime Reduction) formally quantifies the reduction in an asset's potential lifespan caused by a predictive removal.
- The **NDF** (No Degradation Found) rate is another key performance indicator for the AMBER zone, just as the NFF (No Fault Found) rate is for the RED one. The NDF rate represents the proportion of proactively removed assets for which no significant degradation has been confirmed during shop-level inspection.

These operational metrics are essential for performance assessment within the framework of a continuous improvement and feedback loop strategy.

From this baseline, the following steps need to be addressed:

- Make the link between the different set of metrics at Failure Mode and equipment levels
- Reconcile operational metrics with the classic ones used in data analytics because it is crucial to align

different teams and fields of expertise (project, service, operational experts, data analysts, data scientists…)

- Propose new ways of working to assess the in-service performance of "health ready" equipment (cf SAE JA6268) in the scope of a CBM / PBM strategy.

In particular, the measurement of the in-service performance (once a given monitoring algorithm is deployed) must be performed in an efficient way to ensure that there is no deviation from the defined objectives (e.g. detect the emergence of no new Failure Modes in particular when the machine is ageing, no new unexpected degradation profile, no new usage pattern …). This requires adapting the feedback loop process, the used set of metrics and the continuous in-service performance assessment taking into account the shop findings results.

The following figure compares key operational performance metrics across two distinct scenarios: a purely reactive maintenance strategy versus a purely predictive one.

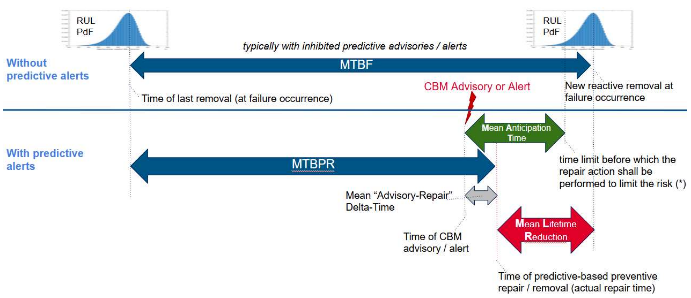

Figure 5. Links between operational metrics in a perfect predictive situation (Recall = Precision = 100%)

MTBF: Mean Time Between Failures

MTBPR: Mean Time Between Predictive Removals

MLR: Mean Lifetime Reduction

RUL: Remaining Useful Life

PDF: Probability Density Function

(*): time limit not to be exceeded, corresponding to the maximum accepted cumulative risk level. This is the time at which the Cumulative Distribution Function reaches the value '1 - targeted Recall'.

As shown in Figure (5), the MTBPR represents the mean time between removals triggered by CBM / PBM advisories or alerts. Consequently, this metric incorporates the time required by the operator to organize and perform the maintenance action (i.e. replacing the equipment).

The MLR represents the Mean Lifetime Reduction for an equipment lifetime due to the fact that replacing an equipment before a failure naturally reduces its operational usage time.

Besides, the NDF rate must be understood as corresponding to the wellknown "NFF" concept (which is applied to reactive removals), but when applied only to early removals due to a predictive-based monitoring.

Then, the couple (NDF, MTBPR) here proposed should play, for predictive removals, the same role as the couple (NFF, MTBRR) in a context of reactive removals.

'1 - NFF (rate)' is the Precision of a given reactive maintenance process, i.e. the percentage of removed equipment confirmed as faulty following a reactive removal.

'1 - NDF (rate)' is the Precision of a given predictive solution, i.e. the percentage of removed equipment confirmed as degraded for only predictive removals.

In the following chapters, the term 'Precision' will be used only in the context of predictive removals.

So we have: Precision = 1-NDF.

2.2. **Key End-to-End metrics**

In data analytics, especially in anomaly detection, the quality of detection, is usually measured using following main standard metrics, derived from the confusion matrix:

$$Recall \ = \ \frac{TP}{TP+FN} \qquad \text{(also called detection rate)}$$

$$Precision \ = \ \frac{TP}{TP+FP}$$

$$Specificity \ = \ \frac{TN}{TN+FP}$$

With a CBM / PBM strategy, i.e. considering the AMBER zone, following definitions allow to build also a confusion matrix:

TP (True Positives): The number of predictive advisories (triggered within a given validity window, that is not too early and not to too late) that result in an equipment removal, where the equipment's degradation is subsequently 'confirmed (e.g. in shop) and recognized' as sufficiently degraded to trigger an early removal.

FP (False Positives): The number of predictive advisories (triggered within a given validity window) that result in an equipment removal, where the equipment's degradation is NOT 'confirmed (e.g. in shop) and recognized'.

TN (True Negatives): The number of predictive points (e.g., flights, cycles) where the component remains in the nominal GREEN zone and without any predictive advisory triggered (within a given validity window).

FN (False Negatives): The number of operational faults (events in the RED zone) that occurred without being preceded (within a given validity window) by a predictive advisory.

Moreover establishing a reliable ground truth is a critical prerequisite for calculating these performance indicators. While several methods exist, two primary strategies are widely employed in the industry:

- Shop-based confirmation: Considered the gold standard, this direct strategy uses results from shop testing to physically confirm an equipment's state of degradation after removal.
- Data-driven validation: A practical alternative is a data-driven approach that analyzes the Health Indicator's (HI) behavior following a maintenance action. For instance, a True Positive is confirmed if the HI value promptly returns to the nominal zone post-replacement.

Also in data analytics, the Receiver Operating Characteristic (ROC) curve is a standard tool for evaluating model performance, typically plotting the True Positive Rate (i.e., Recall) against the False Positive Rate (1 - Specificity).

However, in aeronautics, the high reliability of systems (resulting in high MTBF values) leads to highly imbalanced datasets where faults are relatively "rare" events. In such cases, the ROC curve can be misleadingly optimistic. Therefore, the Precision-Recall curve, which plots Precision versus Recall, is often a more informative and suitable tool for visualization and model tuning. And this Precision metric is particularly important. A high Precision is key, as it signifies a low rate of unnecessary maintenance removals. Thus this metric is essential as it is directly linked to the additional costs and the maintenance burden resulting from inaccurate advisories (i.e. False Positives).

Besides, the introduction of new operational metrics also necessitates a revision of the formulas used within the feedback loop process. The classic formula (see (2)) cannot be applied to a mixed population of predictive and reactive removals without introducing significant bias.

3. IMPACT OF NEW APPLICABLE METRICS

Let's define **n**, the number of Failure Modes, and $\mathbf{n_p}$, the number of Failure Modes monitored by a predictive model.

$\mathbf{FM_i}$: the Failure Mode number <i> for a given equipment.

The following metrics are derived from the confusion matrix. This presupposes that a ground truth is established using a defined assessment method (e.g., 'Shop-Based Confirmation' or 'Data-Driven Validation').

$\mathbf{NDF_i}$**:** The proportion of predictive removals for a given FM_i that are subsequently confirmed as "No Degradation Found".

$\mathbf{NFF_i}$**:** The proportion of reactive removals for a given FM_i that are subsequently confirmed as "No Fault Found".

$\mathbf{Precision_i}$: the Precision of the predictive monitoring used to address a given FM_i.

$\mathbf{Recall_i}$: the Recall of the predictive monitoring used to address a given FM_i.

NDF: The proportion of predictive equipment removals that are subsequently confirmed as "No Degradation Found".

NFF: the proportion of reactive equipment removals with confirmed "No Fault Found".

Precision: The rate of predictive removals that correctly identify a degraded state. For a given piece of equipment, it is calculated as '1 - NDF rate'.

Recall: The proportion of all actual equipment faults that were successfully preempted by the predictive maintenance monitoring.

3.1. At Failure Mode level

Using a basic composition rule on failure rates at FM_i level, which is valid considering the principle of Mutually Exclusive and Collectively Exhaustive Events (MECE), we have :

λ_{FMi}= [rate of successful predictive removals] + [rate of successful reactive removals]

That we can express as follows:

$$\lambda_{FMi} = \lambda_{PREDICTIVE_{SUCCESSi}} + \lambda_{REACTIVE_{SUCCESSi}} \quad (3)$$

Using the following notation (practical to simplify the writing):

$$MTBPR_i^* = MTBPR_i + MLR_i$$

(2) becomes with a basic consideration on rates:

$$\frac{1}{MTBF_i} = \frac{1 - NDF_i}{MTBPR_i^*} + \frac{1 - NFF_i}{MTBRR_i} \quad (4)$$

Moreover by definition within the AMBER zone:

$$Precision_i = 1 - NDF_i \quad (5)$$

Besides, by definition, the Recall is the ratio of confirmed predictions versus actual faults.

Hence, introducing NDF_i, we have:

$$recall_i = \frac{\frac{1 - NDF_i}{MTBPR_i^*}}{\frac{1}{MTBF_i}} = \frac{(1-NDF_i).MTBF_i}{MTBPR_i^*} \quad (6)$$

Using (4), multiplying by $MTBPR_i^*$, then multiplying by $MTBF_i$, we have:

$$MTBPR_i^* = \frac{(1-NDF_i).MTBF_i}{1-\frac{MTBF_i}{MTBRR_i}(1-NFF_i)} \quad (7)$$

Injecting (7) in (6), we finally obtain:

$$recall_i = 1 - (1 - NFF_i).\frac{MTBF_i}{MTBRR_i} \quad (8)$$

3.2. At equipment level

3.2.1. Preamble

The formulas demonstrated in the previous section are scale-invariant and can therefore be applied at the equipment level. While this property may seem intuitive through the principle of composition, a formal demonstration is essential. Such a proof is provided in the following sections.

As a starting point for this demonstration, we will use an exponential reliability function as a good approximation. This model assumes a constant failure rate (λ), meaning that the risk of failure is entirely random, with no memory effect, no wear and tear, and no cumulative damage.

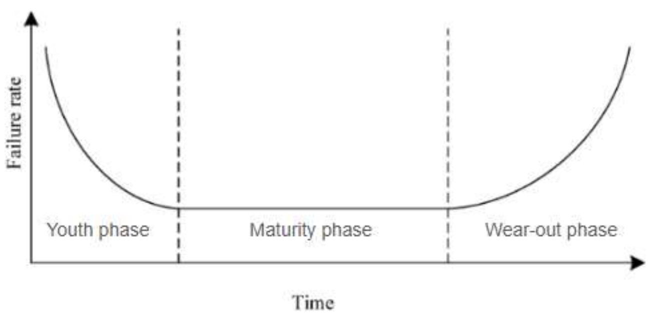

Figure 6. bathtub reliability curve

Engineers widely use this approximation in aeronautics because:

- The first non linear part of the curve is avoided through robustness tests ensuring to suppress youthful failures.
- The second non linear part of the curve is avoided using maximum safe-lifetime-limited parts.

With constant failure rates λ_i, the Reliability function $R_i(t)$ is expressed as:

$$R_i(t) = e^{-\lambda_i t} \quad (9)$$

If the FM_i are considered in series (that is, when any FM_i occurrence fails the equipment), the following laws apply:

$$R_{eqpt}(t) = \prod_i R_i(t)$$

This leads easily to:

$$\lambda_{eqpt} = \sum_{i=1}^{n} \lambda_i \Leftrightarrow \frac{1}{MTBF} = \sum_{i=1}^{n} \frac{1}{MTBF_i} \quad (10)$$

However, if the Failure Modes (FM_i) are configured in parallel, the calculation becomes more complex. A parallel configuration implies that the equipment fails only if all individual Failure Modes occur. The system's failure probability is therefore governed by the following principles:

$$F_{eqpt}(t) = \prod_i F_i(t)$$

$$R_i(t) = 1 - F_i(t)$$

where:

- $F_{eqpt}(t)$ is the failure probability of the equipment.
- $F_i(t)$ is the failure probability of the i-th Failure Mode.
- $R_i(t)$ is the reliability function of the i-th Failure Mode.

Therefore, calculating the overall system reliability entails analyzing the internal architecture of the equipment. And it is essential to consider the arrangement of components and Failure Modes, specifically whether they are configured in series or in parallel (i.e., with redundancies).

To connect these principles to the equipment level, the space of Failure Modes for a given piece of equipment is described as follows:

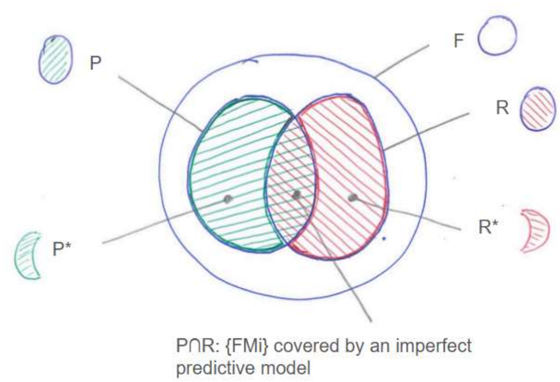

Figure 7. Sets of Failure Modes

Let the following sets define the space of all possible Failure Modes (FM_i) for the equipment under consideration:

- **F**: The complete set of all possible Failure Modes $\{FM_i\}$ for the equipment.
- **P**: The set of Failure Modes that are covered by a predictive monitoring function, therefore associated with predictive and reactive removals.
- **P***: The subset of P containing Failure Modes for which only predictive removals are performed (detection rate = 100%).
- **R**: The set of Failure Modes that can result in a failure leading to reactive removals. Such failures can originate from two sources:
  - Failure Modes that have no predictive monitoring coverage.
  - Failure Modes where the predictive monitoring is imperfect (detection rate < 100%).
- **R***: The subset of R associated with only reactive removals. This subset contains only the Failure Modes that are not covered by any predictive monitoring.
- **P∩R**: This intersection represents the set of Failure Modes covered by an imperfect predictive model. It corresponds to failures that are "missed" by the monitoring system, resulting in reactive removals.

Based on these definitions, the complete space of Failure Modes can be described as the union of the modes covered by predictive monitoring (P) and those that are not (R*):

$$F = P \cup R^*$$

3.2.2. Demonstration

The following demonstration will assume all Failure Modes (FM_i) are independent and configured in series. As previously established, the failure rate at the equipment level is therefore given by an additive law:

$$\lambda_{eqpt} = \sum_{i=1}^{n} \lambda_i \quad (11)$$

Then

$$\frac{1}{MTBF} = \sum_{i \in F} \frac{1}{MTBF_i} \quad (12)$$

Then

$$\frac{1}{MTBF} = \sum_{i \in P} \frac{1}{MTBF_i} + \sum_{i \in R^*} \frac{1}{MTBF_i} \quad (13)$$

$$\frac{1}{MTBF} = \sum_{i \in P} \left(\frac{1 - NDF_i}{MTBPR_i^*} + \frac{1 - NFF_i}{MTBRR_i} \right) + \sum_{i \in R^*} \frac{1 - NFF_i}{MTBRR_i} \quad (14)$$

Noting $MTBPR_i^* = MTBPR_i + MLR_i$

Hence:

$$\frac{1}{MTBF} = \sum_{i \in P} \frac{1 - NDF_i}{MTBPR^*_i} + \sum_{i \in F} \frac{1 - NFF_i}{MTBRR_i} \quad (15)$$

This can be also expressed as follows with appropriate definitions:

$$\frac{1}{MTBF} = \sum_{i \in P} \frac{1}{MTBSPR_i^*} + \sum_{i \in F} \frac{1}{MTBSRR_i} \quad (16)$$

Indeed let's consider these defined rates:

- $\frac{1}{MTBSPR_i^*}$: Rate of Successful Predictive Removals at FM_i level
- $\frac{1}{MTBSRR_i}$: Rate of Successful Reactive Removals at FM_i level

Considering the rules of addition regarding **independent** FM_i in **series**, we have at equipment level:

$$\frac{1}{MTBSPR^*} = \sum_{i \in P} \frac{1}{MTBSPR_i^*} \quad (17)$$

$$\frac{1}{MTBSRR} = \sum_{i \in F} \frac{1}{MTBSRR_i} \quad (18)$$

Hence we can express (15) as follows:

$$\frac{1}{MTBF} = \frac{1}{MTBSPR^*} + \frac{1}{MTBSRR} \quad (19)$$

This can also be expressed as a general law at equipment level:

$$\frac{1}{MTBF} = \frac{1 - NDF}{MTBPR^*} + \frac{1 - NFF}{MTBRR} \quad (20)$$

Using the following defined term at equipment level:

$$NDF = 1 - \frac{MTBPR^*}{MTBSPR^*} = 1 - \frac{\frac{1}{MTBSPR^*}}{\frac{1}{MTBPR^*}}$$

'NDF' being defined like this in order to have the following verified relation:

$$\frac{1}{MTBSPR^*} = \frac{1 - NDF}{MTBPR^*}$$

We can see that this 'NDF' actually corresponds to the right definition of the rate of 'No Degradation Found" at equipment level:

NDF = 1 **-** [% of success among the predictive removals]

NDF = [% of bad predictive removals]

Having the same approach for NFF, we can first define:

$$NFF = 1 - \frac{MTBRR}{MTBSRR} = 1 - \frac{\frac{1}{MTBSRR}}{\frac{1}{MTBRR}}$$

in order to have:

$$\frac{1}{MTBSRR} = \frac{1 - NFF}{MTBRR}$$

Again we can see that this "NFF" actually corresponds to the right definition of the rate of 'No Fault Found" at equipment level:

NFF = 1 - [% success among the reactive removals]

NFF = [% of bad reactive removals]

Thus, the terms NDF and NFF align with their standard definitions at the equipment level. However, a question remains: how do these indicators relate to those defined at the Failure Mode (FM_i) level? The answer is affirmative, and the relationship is demonstrated as follows:

Let's define normalisation coefficients respectively on F, P and R:

$$\alpha_i = \frac{\frac{1}{MTBF_i}}{\sum_{k \in F} \frac{1}{MTBF_k}} \quad over\ F \quad (21)$$

We have *obviously*: $\sum_{i \in F} \alpha_i = 1$

$$\beta_i = \frac{\frac{1}{MTBPR^*_i}}{\sum_{k \in P} \frac{1}{MTBPR^*_i}} \quad over\ P \quad (22)$$

We have obviously: $\sum_{i \in P} \beta_i = 1$

$$\gamma_i = \frac{\frac{1}{MTBRR_i}}{\sum_{k \in R} \frac{1}{MTBRR_k}} \quad over\ R \quad (23)$$

We have obviously: $\sum_{i \in R} \gamma_i = 1$

From the definition: $NDF = 1 - \frac{MTBPR^*}{MTBSPR^*}$

We can write it like this:

$$NDF = 1 - \frac{\sum_{i \in P} \frac{1}{MTBSPR^*_i}}{\sum_{k \in P} \frac{1}{MTBPR^*_k}} \quad (24)$$

$$NDF = \sum_{i \in P} \left(\frac{\frac{1}{MTBPR^*_i}}{\sum_{k \in P} \frac{1}{MTBPR^*_k}} \right) - \sum_{i \in P} \left(\frac{1}{\sum_{k \in P} \frac{1}{MTBPR^*_k} MTBSPR^*_i} \right) \quad (25)$$

$$NDF = \frac{1}{\sum_{k \in P} \frac{1}{MTBPR^*_k}} \sum_{i \in P} \frac{1}{MTBPR^*_i} - \frac{1}{\sum_{k \in P} \frac{1}{MTBPR^*_k}} \sum_{i \in P} \frac{1}{MTBSPR^*_i} \quad (26)$$

$$NDF = \sum_{i \in P} \left(\frac{\frac{1}{MTBPR^*_i}}{\sum_{k \in P} \frac{1}{MTBPR^*_k}} \left(1 - \frac{MTBPR^*_i}{MTBSPR^*_i} \right) \right) \quad (27)$$

Then, using following relations:

$$NDF_i = 1 - \frac{MTBPR^*_i}{MTBSPR^*_i}$$

$$NFF_i = 1 - \frac{MTBRR_i}{MTBSRR_i}$$

This leads to:

$$NDF = \sum_{i \in P} \left(\frac{\frac{1}{MTBPR^*_i}}{\sum_{k \in P} \frac{1}{MTBPR^*_k}} NDF_i \right) \quad (28)$$

$$NDF = \sum_{i \in P} \beta_i . NDF_i \quad (29)$$

This is logical, the equipment-level 'NDF rate' is computed as a weighted sum of the individual NDF_i values.

Using same demonstration principles for NFF, we have:

$$NFF = 1 - \frac{MTBRR}{MTBSRR} \quad (30)$$

$$NFF = \sum_{i \in R} \gamma_i . NFF_i \quad (31)$$

Furthermore, coming back to the raw definition, we have :

$$Precision = \frac{\%\ confirmed\ predictions}{\%\ predictions}$$

$$Precision = \frac{\frac{1 - NDF}{MTBPR^*}}{\frac{1}{MTBPR^*}} = 1 - \mathrm{NDF} \quad (32)$$

And considering that:

$$Precision_i = 1 - NDF_i$$

This easily leads to the following expression:

$$\mathrm{Precision} = \sum_{i \in P} \beta_i . Precision_i \quad (33)$$

The Recall at the equipment level is now considered.

We have by definition:

$$Recall = \frac{\%\ confirmed\ predictions}{\%\ actual\ failure}$$

$$Recall = \frac{\frac{1}{MTBSPR^*}}{\frac{1}{MTBF}} \quad (34)$$

$$Recall = \frac{\sum_{i \in P} \frac{1}{MTBSPR_i^*}}{\sum_{k \in F} \frac{1}{MTBF_k}} \quad (35)$$

$$Recall = \frac{\sum_{i \in F} \frac{1}{MTBSPR_i}}{\sum_{k \in F} \frac{1}{MTBF_k}} \quad (36)$$

Having $MTBSPR_i^* = \infty$ over R*

This leads to:

$$Recall = \sum_{i \in F} \left(\frac{\frac{1}{MTBSPR_i^*}}{\sum_{k \in F} \frac{1}{MTBF_k}} \right) \quad (37)$$

$$Recall = \sum_{i \in F} \left(\frac{\frac{1}{MTBF_i}}{\sum_{k \in F} \frac{1}{MTBF_k}} \times \frac{MTBF_i}{MTBSPR_i^*} \right) \quad (38)$$

$$\text{Recall} = \sum_{i \in F} \alpha_i \, . \, Recall_i \quad \text{with } \text{Recall}_i = 0 \text{ over R*} \quad (39)$$

$$\text{Recall} = \sum_{i \in P} \alpha_i \, . \, Recall_i \quad (40)$$

Besides, from (20) and based on same demonstration principles shown on §3.1, we obtain:

$$\text{Recall} = 1 - (1 - NFF) \, . \, \frac{MTBF}{MTBRR} \quad (41)$$

Furthermore, returning to the basic definition, the following laws also apply:

$$Recall = \frac{MTBF}{MTBSPR^*} = \frac{\frac{1 - NDF}{MTBPR^*}}{\frac{1}{MTBF}} \quad (42)$$

$$Recall = \frac{(1 - NDF) . MTBF}{MTBPR^*} \quad (43)$$

In addition, it is easy to demonstrate the following relations:

$$\text{MTBF} = \frac{1}{n} . \sum_{i \in F} \alpha_i \, . \, MTBF_i \quad (44)$$

$$\text{MTBPR}^* = \frac{1}{n_p} . \sum_{i \in P} \beta_i \, . \, MTBPR_i^* \quad (45)$$

$$\text{MLR} = \frac{1}{n_p} . \sum_{i \in P} \beta_i \, . \, MLR_i \quad (46)$$

$$\text{MTBPR} = \frac{1}{n_p} . \sum_{i \in P} \beta_i \, . \, MTBPR_i \quad (47)$$

3.3. **Consistency checks**

Without CBM / PBM, there are no predictive removals and MTBPR = ∞.

Thus, (20) reduces to (2), the historical law used in the context of a purely reactive maintenance strategy. This relationship confirms that the traditional approach is simply a specific case of the more comprehensive model presented here.

Besides (41) can be written as:

$$Recall = 1 - \frac{\frac{1 - NFF}{MTBRR}}{\frac{1}{MTBF}} \quad (48)$$

Which corresponds to:

$$Recall = 1 - \frac{\%\, confirmed\ reactive\ removals}{\%\, actual\ failures} \quad (49)$$

This is consistent with the definition.

Noting:

$$\rho = \frac{MTBF}{MTBRR} = \frac{\frac{1}{MTBRR}}{\frac{1}{MTBF}} = \frac{\%\, reactive\ removals}{\%\, actual\ failures}$$

We have :

$$\text{Recall} = 1 - \rho.(1\text{-NFF}) \quad (50)$$

With only successful predictive removals (i.e. without reactive removals), we have MTBRR = ∞.

Then ρ = 0, then Recall = 100%, which is consistent again.

4. IMPACTS ON THE MEASUREMENT PROCESS

4.1. **Equivalent expression of performance targets**

The equation (41) shows the equivalence in terms of performance targets expressed in a Purchase Technical Specification:

- When defining a minimum operational performance target expressed in the form:

  $$\frac{MTBRR}{MTBF} > coeff$$

  Using a coeff > 1 requires predictive monitoring functions

- When defining a "classic" set of performance targets (in data analytics) expressed in the form:

  Recall > value

4.2. **Formalisation of shopfinding results**

To compute performance indicators at the equipment level, it is necessary to understand the architectural relationship of the different Failure Modes (FM_i) involved.

By default, all FM_i should be provided as formalized, independent, and configured in series. If a more complex relationship exists (e.g., involving parallel configurations or dependencies), this architecture must be specified. This information is essential for calculating the system-level performance using the appropriate aggregation laws.

4.3. **A Methodology for In-Service Performance Evaluation**

To ensure an accurate and unbiased estimation of operational KPIs such as MTBF, the following in-service evaluation process must be applied. A core principle is to analyze distinct populations of equipment that share, on average, the same usage profiles and environmental conditions.

Step 1: Establish the Baseline MTBF

This requires a "control" population of equipment that is not subject to predictive maintenance removals (i.e., operating in a purely reactive or "shadow" mode). This ensures all inherent Failure Modes (FM_i) can manifest.

1. Estimate the Mean Time Between Reactive Removals ($MTBRR_R$): Measure the $MTBRR_R$ for this control population, where all removals are, by definition, reactive.
2. Determine the NFF_R for this population through a robust feedback loop with suppliers and MRO providers, using the defined ground truth assessment methods for fault confirmation (typically a Shop-Based Confirmation).
3. Compute the Baseline MTBF: Calculate the equipment's design-dependent MTBF using the classic formula:

$$MTBF = MTBRR_R / (1 - NFF_R)$$

Step 2: Evaluate the CBM / PBM Program

This requires a second representative population of equipment where predictive monitoring and removals are active.

4. Estimate the Mean Time Between Predictive Removals (MTBPR): Measure the MTBPR for this CBM / PBM-enabled population.
5. Estimate the No Degradation Found (NDF) rate: Determine the NDF rate for predictive removals through the feedback loop, using the defined ground truth assessment method for degradation confirmation.
6. Analyze the Reactive Sub-Population: Within this same CBM / PBM group, analyze the sub-population of assets that still fail reactively. Estimate their specific MTBRR and NFF rate. These values are expected to differ from the baseline $MTBRR_R$ and NFF_R due to the effect of the model's Recall (some failures are successfully prevented).
7. Compute the Mean Lifetime Reduction (MLR): Using the metrics from the CBM / PBM population, calculate the MLR by rearranging the generalized formula (19):
$$1 / MTBF = (1-NDF) / (MTBPR + MLR) + (1-NFF) / MTBRR$$

Step 3: Verify the CBM / PBM Program's Efficiency

8. Check the Efficiency Ratio: Ensure that the ratio of lifetime reduction to the theoretical mean lifetime remains within an acceptable target, typically in the range [1% - 5%].
$MLR/MTBF \leq Target\%$.

   In practice this target percentage should be defined based on factors such as the equipment's MTBF and price, the cost of an operational interrupt, and the relative costs of proactive versus reactive repairs. A high MLR/MTBF ratio can indicate that a predictive model is intervening too early, thus requiring improvement to its anticipation time.

<u>A Note on Statistical Validity:</u>

When calculating mean-time indicators like MTBRR and MTBPR, it is crucial to ensure statistical significance. If these metrics are based on only a few events, they may not be representative. In such situations, it is necessary to expand the measurement window or the size of the equipment population to gather more data.

5. CONCLUSION

This paper has addressed a fundamental challenge in evaluating the performance of modern maintenance strategies. We first established that, while deeply ingrained, the MTBUR metric is a legacy of a purely reactive paradigm and is ill-suited for a CBM / PBM context. Its use creates misleading assessments by penalizing successful proactive interventions. The logical and necessary first step for any organization adopting predictive maintenance is therefore to transition to unambiguous metrics that accurately reflect operational performance and foster a proactive culture.

To resolve this ambiguity, we introduced a new set of complementary operational metrics: **MTBRR, MTBPR, MLR, and NDF**. This comprehensive framework overcomes the inherent bias of traditional "unscheduled-based" measurements. Crucially, it provides a common language that reconciles the perspectives of the data analytics domain and the operational world. Key mathematical relationships were formalized to connect these different levels of analysis:

- **Linking Failure Mode to Equipment Level** (for independent, series-configured FM_i):

$$Recall = \sum_{i \in P} \alpha_i \, Recall_i \quad (51)$$

$$\text{Precision} = \sum_{i \in P} \beta_i \, Precision_i \quad (52)$$

$$NDF = \sum_{i \in P} \beta_i \, . \, NDF_i \quad (53)$$

$$NFF = \sum_{i \in R} \gamma_i \, . \, NFF_i \quad (54)$$

- **Linking Data Analytics and Operational Metrics at the equipment level:**

$$Recall = 1 - (1 - NFF)\frac{MTBF}{MTBRR} \quad (55)$$

This can be rearranged to generalize the classically used equation (1):

$$MTBRR = \frac{1-NFF}{1-Recall} MTBF \quad (56)$$

And finally, another useful expression for the 'Recall' metric was presented:

$$Recall = \frac{1-NDF}{MTBPR^{*}} MTBF \quad (57)$$

Furthermore, a detailed step-by-step measurement process was proposed to ensure that these metrics can be calculated consistently and without bias from in-service data. The framework and methodologies presented here now require experimental validation on a wide scale. Successful implementation will depend on acceptance by all stakeholders, particularly those involved in the feedback loop for in-service parts management. Adopting this rigorous approach will prevent misalignments between suppliers, integrator and operators viewpoints, leading to a more accurate and transparent assessment of operational performance in a predictive maintenance environment.

NOMENCLATURE

CBM	Condition Based Maintenance
E2E	End-to-End
EQPT	Equipment
FM	Failure Mode
FP	False Positive
FN	False Negative
HI	Health Indicator
MEL	Minimum Equipment List
MRO	Maintenance, Repair, and Overhaul
$MTBF$	Mean Time Between Failures
$MTBPR$	Mean Time Between Predictive Removals ("Predictive" term chosen to be PBM centric, but in fact we consider here all early CBM-based removals)
MLR	Mean Lifetime Reduction
$MTBPR^{*}$	Mean Time Between Predictive Removals added to the Mean Lifetime Reduction $MTBPR^{*}$ = MTBPR + MLR
MTBRR	Mean Time Between Reactive Removals
$MTBSPR$	Mean Time Between Successful Predictive Removals
$MTBSPR^{*}$	Mean Time Between Successful Predictive Removals added to the Mean Lifetime Reduction $MTBSPR^{*}$ = MTBSPR + MLR
$MTBUR$	Mean Time Between Unscheduled Removals
$MTBSRR$	Mean Time Between Successful Reactive Removals
NFF	No Fault Found, or rate of equipment with 'No Fault Found' when used in a formula

NDF	No Degradation Found, or rate of equipment with 'No Degradation Found' when used in a formula
PBM	Predictive Based Maintenance
PDF	Probability Density Function
RUL	Remaining Useful Life
TP	True Positive
TN	True Negative

REFERENCES

J. Sen Gupta, C. Trinquier, K. Medjaher and N. Zerhouni (2015). Continuous validation of the PHM function in aircraft industry". *First International Conference on Reliability Systems Engineering (ICRSE), Beijing, China, 2015, pp. 1-7, doi: 10.1109/ICRSE.2015.7366505.*

Li Jun, Xu Huibin, (2012). Reliability Analysis of Aircraft Equipment Based on FMECA Method

Joel J. Luna (2021). Metrics, Models, and Scenarios for Evaluating PHM Effects on Logistics Support Joel J. Luna

Mark Fitzpatrick, KS Robert Paasch. Analytical Method for the Prediction of Reliability and Maintainability and Maintainability Based Lifecycle Labor Costs

Carolin Wagner, Bernd Hellingrath (2020). Supporting the Implementation of Predictive Maintenance – a Process Reference Model. *International Conference on Industrial Engineering and Operations Management*

Salma Maataoui, Ghita Bencheikh, Maria Ezziani, Ghizlane Bencheikh, (2024). Data-Driven Predictive Analysis for Cutting Machine Failures: A Technical Report on Reliability Optimization

Dr. Jorge Luis Romeu (2006). Understanding Series and Parallel Systems Reliability

Pengfei Wei, Zhenzhou Lu, and Longfei Tian (2013). Addition Laws of Failure Probability and their Applications in Reliability Analysis of Structural System with Multiple Failure Modes

Mohammad Modarres, Reliability Engineering and Risk Analysis - A Practical Guide (3rd edition in 2016).

Christopher Teubert, Ahmad Ali Pohya, George Gorospe, (2023). An Analysis of Barriers Preventing the Widespread Adoption of Predictive and Prescriptive Maintenance in Aviation

Goebel, K.; and Rajamani, R. (2021). Policy, regulations and standards in prognostics and health management. International Journal of Prognostics and Health Management, vol. 12, no. 1

Teubert, C.; Jarvis, K.; Corbetta, M.; Kulkarni, C.; and Daigle, M.(2022). ProgPy Python Prognostic Packages v1.4. URL https://nasa.github.io/progpy.

Robert Meissner, Antonia Rahn, Kai Wicke (2021). Developing prescriptive maintenance strategies in the aviation industry based on a discrete-event simulation framework for post-prognostics decision making

Eleftheroglou, N. (2023): Prognostics: The Heart of Predictive Maintenance.

Abhinav Saxena, Jose Celaya, Edward Balaban, Kai Goebel, Bhaskar Saha, Sankalita Saha, Mark Schwabacher (2008). Metrics for Evaluating Performance of Prognostic Techniques. *International Conference on Prognostics and Health Management*

Abhinav Saxena , Jose Celaya , Bhaskar Saha , Sankalita Saha , and Kai Goebel (2010). Metrics for Offline Evaluation of Prognostic Performance. *International journal of prognostics and health management*

Carlos E. Torres (2024), Challenges in Implementing Condition-Based Maintenance. *Power-MI Blog*

Shuai Fu, Nicolas P. Avdelidis, Angelos Plastropoulos, Ip-Shing Fan (2023). Fusion and Comparison of Prognostic Models for Remaining Useful Life of Aircraft Systems

Shuai Fu, Nicolas P Avdelidis (2023). Prognostic and Health Management of Critical Aircraft Systems and Components: An Overview - PMC - PubMed Central

SAE (first edition 2018, last edition 2023). JA6268: Design & Run-Time Information Exchange for Health-Ready Components.

Biographies

Franck Dessertenne is a Technical Leader in Health Monitoring, Diagnostics, and Predictive Maintenance at Airbus Operations SAS. He earned a Master's Degree in Engineering from ISMRA–ENSICAEN and a Master's Degree in Materials Science & Radiation from the University of Caen (France), both in 1997.

Since joining Airbus in 2003, he has held several key roles. He began in software design and development for simulation products, including diagnostic tools, before becoming responsible for the technical integration of an airborne maintenance server on a military program. Subsequently, his work has focused on Condition-Based Maintenance (CBM) and Predictive-Based Maintenance (PBM). In this domain, he has served as a Health Engineer and Data Analyst, designing and implementing health monitoring and predictive solutions for various aircraft systems.

Currently, in a technical leadership role, he guides Research & Technology (R&T) activities in health monitoring, diagnostics, and prognostics. His research interests include reactive and predictive maintenance, advanced data analytics, artificial intelligence, and the application of machine learning to complex aerospace systems. This paper marks his first authored contribution to the PHM Society. He has also recently joined the SAE HM-1 committee.