Méthode de comparaison de performances de fiabilité de différents matériels roulants

Methodology to compare different rolling stock reliability performances

MILLIOT Frédéric
Senior RAM Expert

*ASLTO*MSAINT OUEN sur SEINE
frederic.milliot@alstomgroup.com

RÉSUMÉ/SUMMARY

This paper presents a method to compare different rolling stocks performances eradicating the bias introduced by different mission profiles (mileage and speed). Once the scope of study is defined (can be at train level, at subsystem level or lower), the starting point is to perform Reliability Growth models for each product, on a defined set of field data. Alstom uses the Crow-AMSAA model which has a good ability to fit data, especially for the end of the observation. Then, it is possible to plot Reliability Growth Curves for each product by using a single and referenced mileage profile defined by the study. It is also necessary to suppress the speed influence on the failure intensities (generally measured in km^{-1}) ; to do that a correction is done "as if" all the products had the same referenced commercial speed. Finally, this comparison is the introduction to a qualitative analysis to determine the best practices and best solutions in the reliability perspective.

Cette publication présente une méthode de comparaison de performances de fiabilité de différents matériels roulants qui supprime le biais introduit par des profils de mission kilométriques et de vitesses qui sont nécessairement différents. Une fois le périmètre de l'étude établi, son point de départ est de réaliser des modèles de croissance de fiabilité sur des données de terrain pour chacun des produits sélectionnés. Alstom recommande le modèle de Crow-AMSAA pour sa capacité à ajuster les données, spécialement les derniers points d'observation. Dès lors, il devient possible de tracer les courbes de croissance de fiabilité sur un profil de mission kilométrique unique dit de référence. De manière analogue, on doit corriger l'influence des vitesses moyennes qui influent sur l'intensité de fiabilité (mesurée en km^{-1}) ; la comparaison se déroule « comme si » les flottes de trains avaient la même vitesse moyenne. Au final, cette comparaison vise à lancer une analyse de causes qualitative devant déterminer les bonnes pratiques et les solutions pour garantir la fiabilité.

MOTS-CLEFS/KEY WORDS —

Comparison, Reliability Growth, mission profile, Crow-AMSAA, and lessons learnt.

I. INTRODUCTION

In the Railway industry, reliability is a key performance and become more and more challenging. The customer targets become more and more stringent and intermediate objectives can also be defined. The customers want the best reliability as soon as possible after the manufacturer's delivery. In this context, the rolling stock is one of the most critical systems in terms of reliability. The reason is mainly due to a high complexity and the integration of multiple functions. The financial risk for a company like Alstom is important. Indeed, the consequences of service perturbation are huge: potential penalties, fees, the cost of spare parts, maintenance workload, cash payment milestone or other contractual clauses.

So, the management of reliability performance is critical. One way to manage this risk is to predict the performance and, by anticipation, take actions to optimize or improve the reliability. Practically there are 2 possible ways to perform this prediction

- a theoretical one which is a combination of figures from supplier's studies which is not addressed in this paper
- another one use Return of EXperience performances from previous projects.

Ideally, both approaches are combined since the tender stage of the project. The first approach provides a steady state prediction.

The second one could provide better confidence in the result but introduce bias due to the specific conditions of the projects. So at the company level, it is necessary to be able to capitalize this REX and get the best estimations from that. One use case is to use the REX to predict performance for a new projects since the tender stage to estimate the risk of non-compliance. Another use case is to compare on field project performances and identify quantitatively which one is the best/worst product in a family; this is the topic of this article. Once this identification is done, the good and bad practices as well as the good and the bad solutions can bring lesson learnt to improve the future products.

The main challenge faced to perform this comparison is that every product is slightly different from each other, but, in addition, the customer mission profiles can also vary (speeds, mileage, power on time, time in service,…), as well as the size of the fleet (can be between 10 to 50 trains), the commercial service introduction planning of the train are different, the observed fleet mileages are different (between 3 and 15 million kilometers) and the various climates must be considered. So many different factors can influence the comparison.

The goal of this method is to eradicate at least the bias introduced by the mission profile, that is to say the difference of mileage run by the fleets and the average commercial speeds. The main idea is to use the Reliability Growth Models and the referenced conditions of mileage and speed to simulate performances as if all products were operated in the same condition. In this way, the comparison is becoming possible and correct. This paper presents a methodology breakdown in 5 steps.

II. Revue de literature

[CROW] AMSAA Technical report n° 138 Reliability Analysis for complex, repairable systems, Larry Crow 75

[IEC] IEC 61710:2013 Power law model – Goodness-of-fit tests and estimation methods

[DUANE] "Learning Curve Approach To Reliability Monitoring," Duane, J.T. 64

[TAN] MATHEMATICAL ASPECTS OF RELIABILITY GROWTH ANALYSIS, Tananko, 2019

III. Methodology Overview

The proposed methodology is structured into five distinct steps, as illustrated in Figure 1:

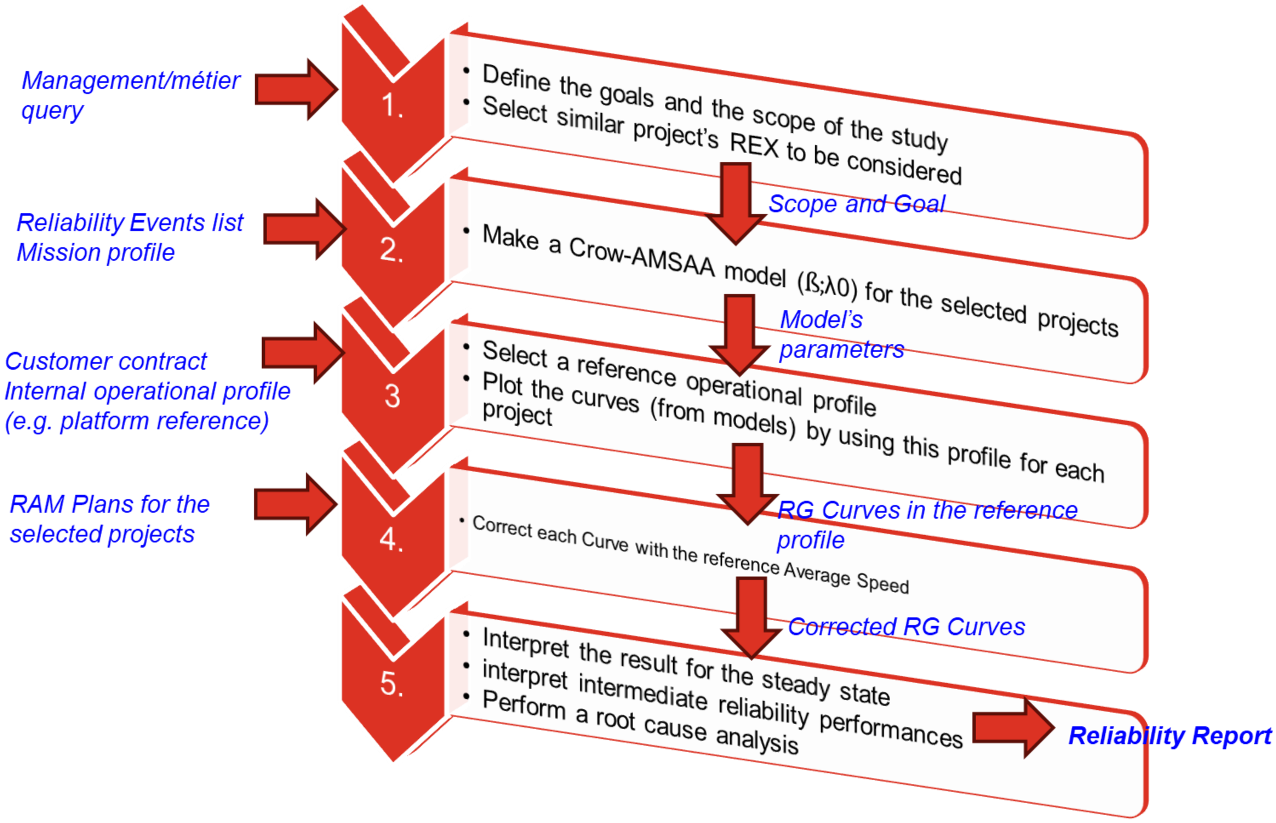

Fig. 1. Methodology Diagram

IV. DEFINE THE SCOPE AND THE GOAL OF THE ANALYSIS

The initial step involves defining the objective of the analysis. This goal may originate from various sources such as a management directive, a tender requirement, or internally from pure Reliability, Availability or Maintenance concerns.

Possible objectives include: assessing whether the previous generation of metros performs better than the new one, determining the benchmark product within a certain range to offer customers, or demonstrating that a product meets or exceeds a benchmark standard. Once the objective is established, we identify the relevant projects for methodological development. This involves collecting raw data and key mission profile characteristics from the warranty phase, during which each incident's service impact is evaluated by both the customer and the manufacturer. The current method is not modifying the content of those data, the only action is to group all Service Affecting Failure (SAF) in a global one. Then we can plot the reliability in term failure intensity vs observed mileage.

In the railway industry the Mean Distance Between Failure or the Failure intensity are used to measure the reliability. As defined in [IEC] like:

$$FI\ \ (\mathrm{t}) = \frac{d(E(N(t)))}{dt} \quad (1)$$

with N(t) the number of events and E(N(t)) the mean value of events.

We cannot use directly this formula, we have to transform it into a mileage base:

$$FI\ \ (\mathrm{km}) = \frac{d(E(N(km)))}{dkm} \quad (2)$$

Thus the estimator is:

$$FI_{estimator}(\mathrm{km}) = N(km)/km \ \ (3)$$

This is what is usually measured in the projects.

For illustration, consider a study aimed at comparing two metro fleets.

Metro 1 comprises a fleet of 23 trains observed over a 3 million km cumulated distance. By plotting the Service Affecting Failure intensity vs mileage it comes fig 2:

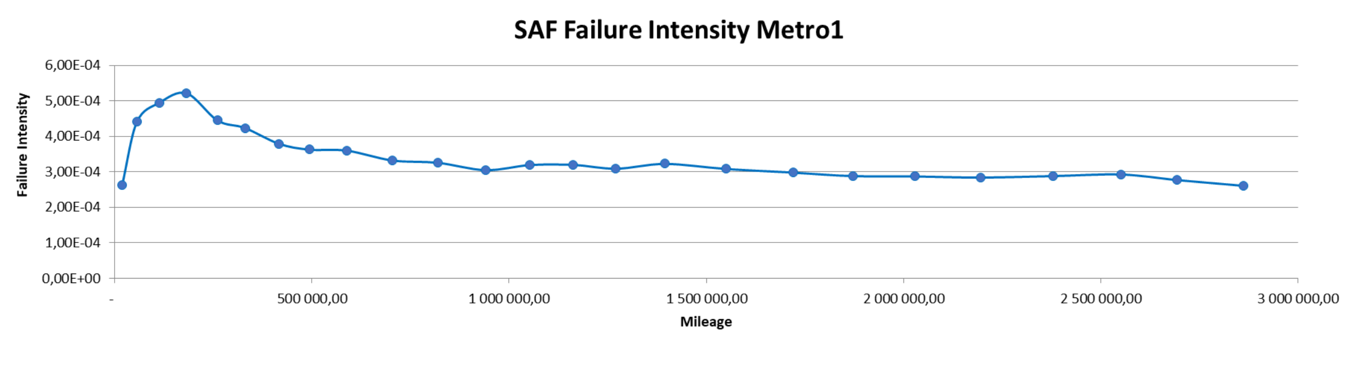

Fig. 2. Measured SAF Failure Intesnsity for Metro 1

Metro 2 consists of a fleet of 50 trains, observed over 16 million km cumulated distance see fig 3:

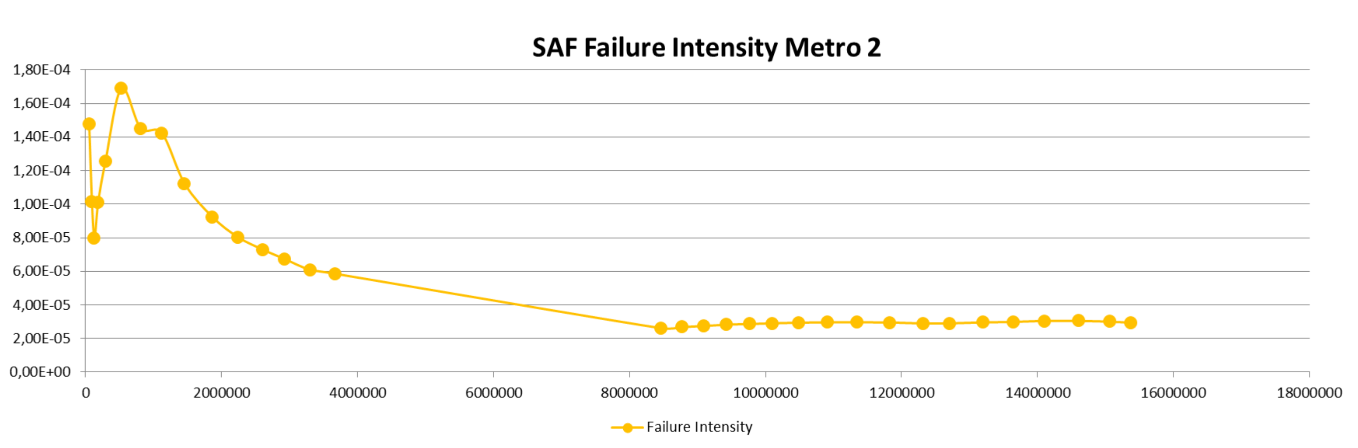

Fig. 3. Measured SAF Failure Intesnsity for Metro 2

Reliability Growth (RG) curves are highly useful for evaluating reliability over time, addressing questions such as: 'Is the product reaching its optimal performance? What are its intermediate performance milestones (e.g., after X months or Y kilometers)? What is its initial reliability?' The mean value may not accurately represent the best performance achievable, highlighting the importance of RG curves.

For the objective of this analysis—comparing Metro 1 and 2—we attempt a direct comparison by overlaying both curves on the same graph, aligning them along the same mileage on the X-axis, as shown in: Fig 4:

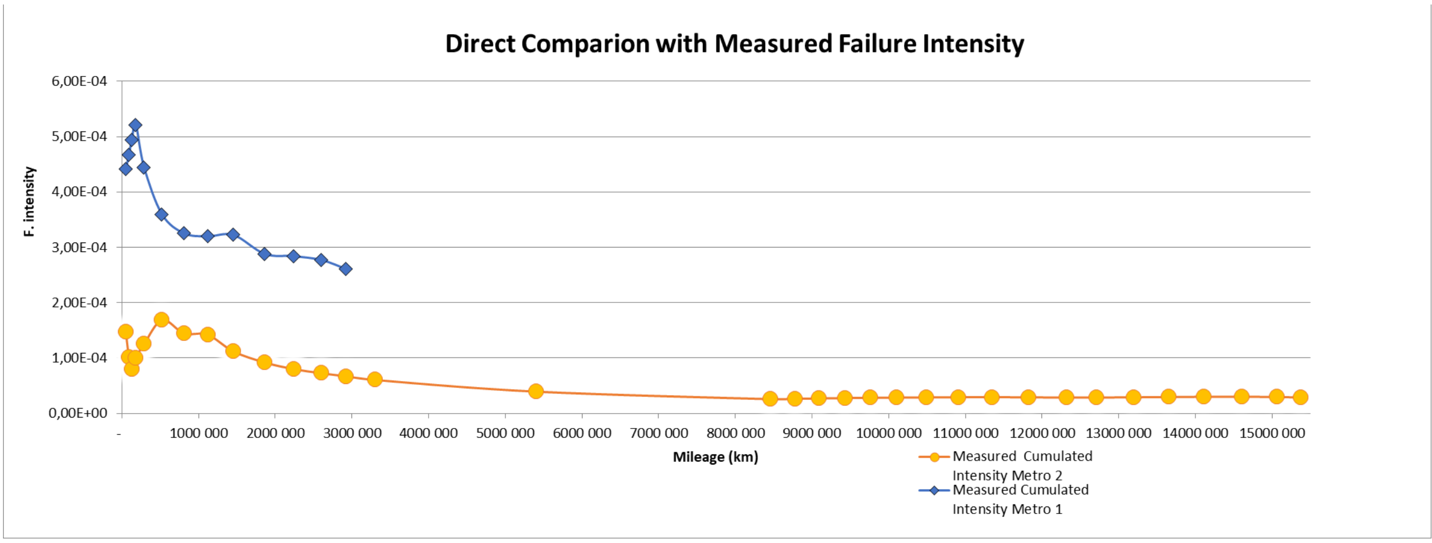

Fig. 4. Direct Performance Comparison with measured Failure Intensities

Initially, it seems that Metro 2 is significantly more reliable than Metro 1, with Metro 1's failure intensity reaching only 26% of that of Metro 2. However, this quick assessment warrants scrutiny. The observation periods differ significantly: Metro 1's data is capped at 3 million kilometers (marking the end of its contractual warranty period), compared to 16 million kilometers for Metro 2. This discrepancy raises questions: How would Metro 1 perform over an observation period equivalent to Metro 2's? Has Metro 1 achieved its optimal performance? Furthermore, the quantity of observations differs between the two fleets, resulting in lower confidence in the reliability assessment for Metro 1 compared to Metro 2. Such a 'direct comparison' approach, while once considered state-of-the-art, is precisely what the methodology introduced in this article aims to refine.

.

V. Reliability Growth Modelling

The subsequent step in the methodology involves calculating a Reliability Growth (RG) model for each product, offering several advantages. Firstly, the model affords a 50% confidence level for any given distance. Secondly, it effectively filters the randomness of the observations. Actually, the number of observed events N(t) can be considered as a random variable with a given spread. RG models are able to predict the average value of N(t).

There are two main models:

- Duane described in [DUANE]
- Crow-AMSAA described in [CROW]

The difference between those and the assumptions are detailed in [TAN].

In a few words and put in the Railway scope of application Duane model postulates that:

$$Ln(N/Cumulated\ km) = Ln(\lambda) - \alpha * Ln(Cumulated\ km) \qquad (4)$$

- α is the shape parameter
- λ is the scale parameter

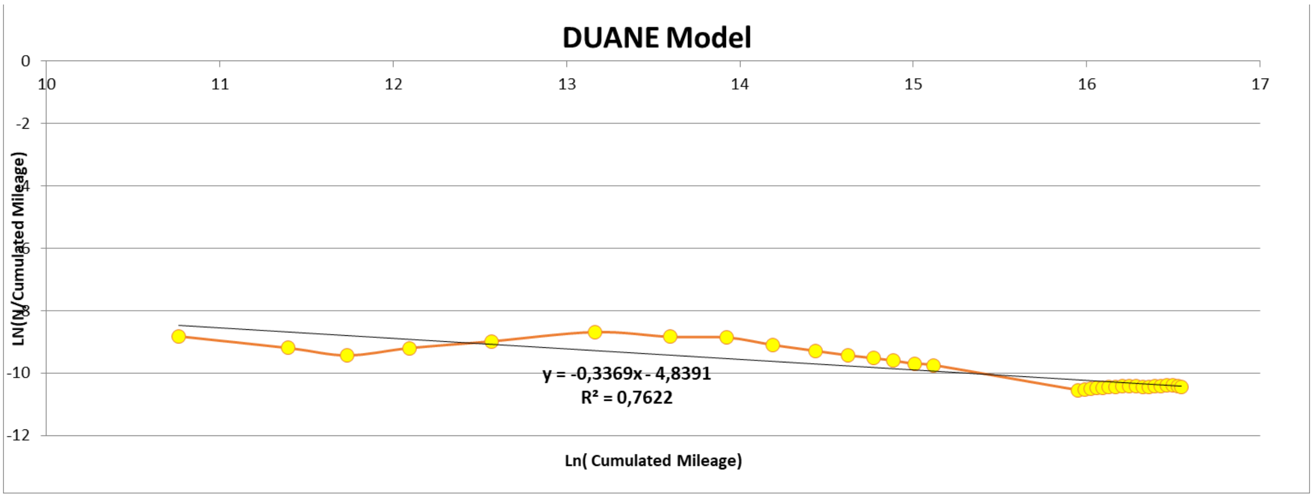

Fig. 5. Duane Parameters estimation

By plotting $Ln(N/Cumulated\ km\ \ )$ on y axle and $Ln(Cumulated\ km)$ on x axle the parameter's estimation is straightforward.

$\alpha = -Slope$ (5)

$\lambda = Exp(\text{Y intercept for x} = 1)$ (6)

This can be done with special Reliability Tools or with a spreadsheet with a linear regression. The modelling error can be evaluate with the correlation factor R^2 taken from this linear regression.

Dr. Larry Crow converted the Duane model into a stochastic model using a Weibull distribution as failure intensity function [CROW].

$$Instantaneous\ Failure\ Intensity_{Fleet}(km) = \lambda_0 * \beta * (km)^{\beta-1} \qquad (7)$$

- β is the shape parameter
- λ_0 is the scale parameter

in this model the Total Number of events is no more deterministic (as in Duane) and is a random variable.

The model parameters (β, λ_0) are rendered independent of the original mileage, enabling the derivation of an analytical formula for failure intensity (1) from [IEC]

$$Cumulated\ Failure\ Intensity_{Fleet}(Cumulated\ km) = \lambda_0 * (Cumulated\ km)^{\beta-1} \qquad (8)$$

The parameters estimation uses the Maximum Likelihood which improve the fitting capabilities compare to the Duane estimation. In this case the calculation is more sophisticated. The algorithm is provided in [IEC] time data for group of relevant failures. Each month, one counts the relevant events for the current month. In addition the actual mileage run by the fleet is collected. This can be done with special Reliability Tools or with a spreadsheet with optimisation capabilities.

This methodology advocates for the use of the Crow-AMSAA model. With the type of data we analyze, it has been observed that this model possesses a superior capability to accurately fit data towards the end of the observation period compared to the Duane model. In figure 6 we present an example of fitting capabilities comparing Duane and Crow-AMSAA for the same set of data. We calculate the errors at each km:

$$Error\ (km) = \left| \frac{Ncumulated(km)}{km} - Model(km) \right| \qquad (9)$$

The Duane model, due to its estimation technique, generates an average prediction that considers the entire history of the process. Given that reliability experience (REX) curves may exhibit irregularities at the start — attributable to gradual train deliveries and the establishment of the customer's operations — the Duane model can introduce discrepancies. The end of the observation is the most important for the manufacturer because it represents the optimal performances of the product for which the customer set a target. With the Duane model the weight of the beginning and the end of the observation are the same, which leads to pessimistic results, most of the time.

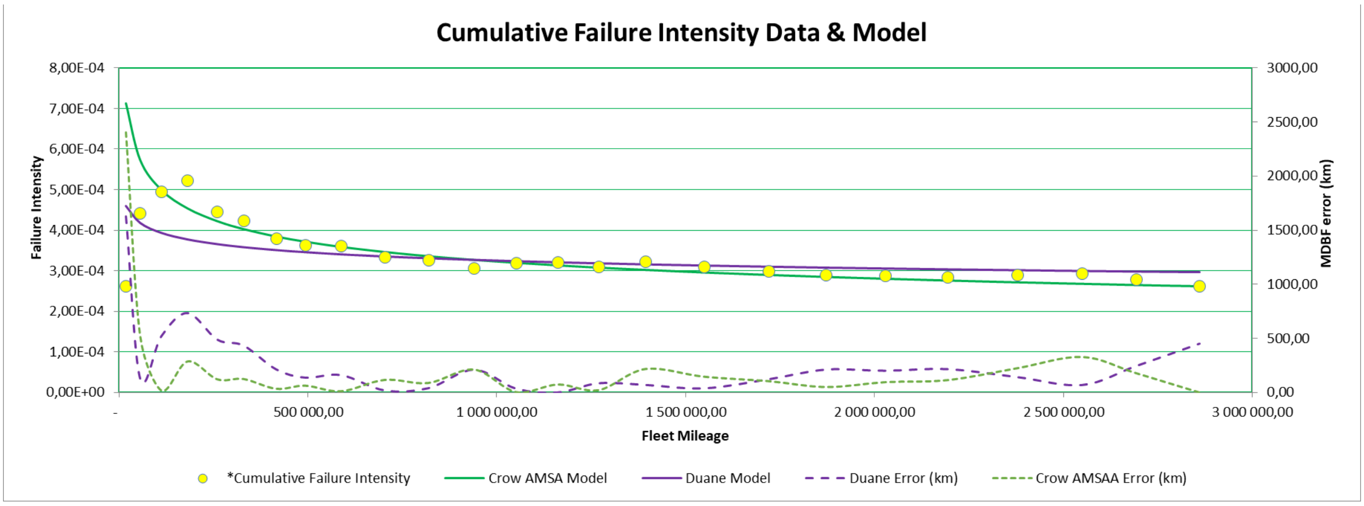

Fig. 6. Metro 1 : Example of Fitting Performances with Crow-AMSAA and Duane algorithms

VI. ASSESS THE RELIABILITY WITH A REFERENCE PROFILE

Once (βi,λ_{0i}) parameters are estimated with (i) the number of projects selected as similar and relevant for the scope of the study (2 in our example), it becomes possible to calculate or plot dedicate curves with any fleet mileage (different from the original ones by using the one for a new tender for example) or to have a direct analytical estimation (1). Indeed (βi,λ_{0i}) are independent from the original fleet mileage. They represent the product reliability and also the company's efforts to reach that level of performance. The more effort is done the more βi is close to 0.

So the projection done in another operational profile assume that the initial reliability λ_{0i} have the same order of magnitude and that the company's effort βi will be same. Those conditions are explicitly mentioned to the tender team. In case of the fleet mileage is too low (for example if the steady state is not reached or with censored data) or the representativeness of the data is not ensured, we can also calculate the confidence bounds. In the illustration it is not necessary.

To make a comparison we can choose the same mission profile for two metros. Thus we can select a controlled and Reference profile. For this we can use a typical metro profile (10 000 km/month for example) or select according to a specific customer hypothesis which is interesting for new tender. In our example let' us assume that the expected optimal reliability will be reached after 7 Millions km.

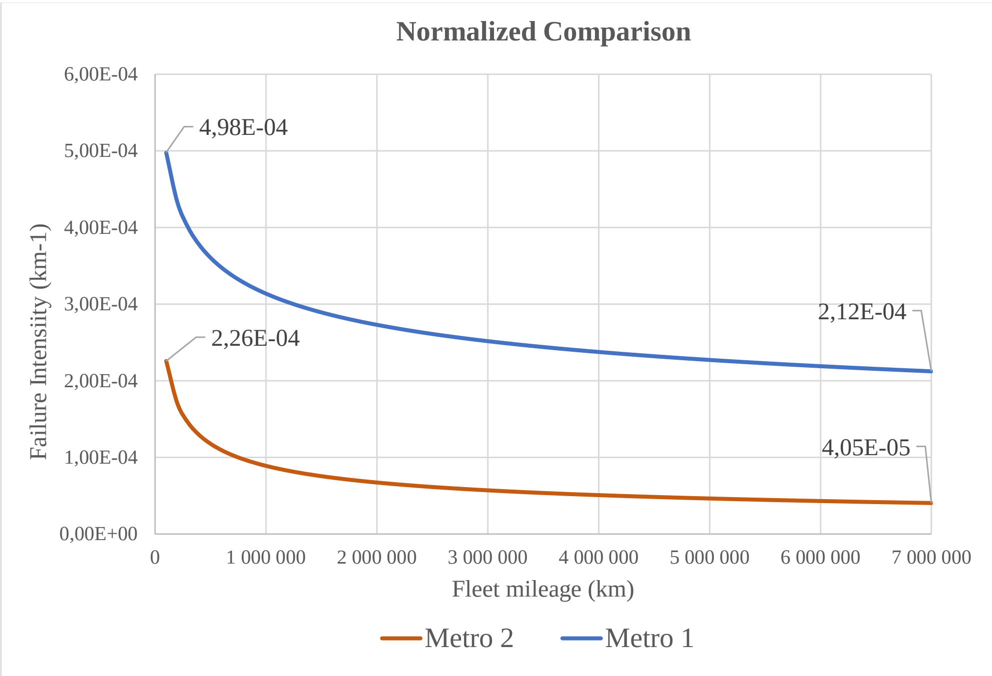

Fig. 7. Performance comparison with a reference mileage profile

These curves confirm that Metro 2 outperforms Metro 1 over a distance of 7 million km, despite their respective reliability growths. Notably, Metro 1 exhibits superior initial performance and reaches its steady state more rapidly than Metro 2. Utilizing these curves to evaluate both optimal and intermediate performances is particularly useful, as these metrics can significantly impact contractual agreements with customers. When assessing optimal performance and the timeline to achieve it, it becomes clear that the average failure rate is not an appropriate KPI for either metro.

Furthermore, the direct comparison indicates the smallest performance gap between Metro 2 and Metro 1 is 26%, implying that Metro 2's performance does not exceed 26% of Metro 1's. However, under this standardized mission profile, the gap narrows to 19%. This suggests that when evaluated under reference conditions, the discrepancy is greater than initially observed through direct comparison.

VII. AVERAGE SPEED CORRECTION

With the previous step we have compared the performances as if both fleet were running in a same mileage condition. Similarly it is possible to use the time in operation or powered up time as a reference x axis. Another option is to make a focus on a given subsystem. If the method is applied with mileage as basis time can be used also depending on the need.

Indeed in the railway industry and for the rolling stock, the reliability target is usually allocated by the customer in term of failure intensity. The target is a fixed value for a given time T (for example end of the fleet delivery + x months). This can be formalized in (10).

$$Failure\ intensity\ Target_{Fleet}(T) = \frac{Total\ Number\ of\ failures\ observed\ within\ the\ Fleet(T)}{Cumulated\ Fleet\ Mileage\ (T)} \quad (10)$$

The underlying assumption is that the failure intensity is constant in time which is not the case on the real field conditions. This is a simplification for the specification. In project condition this target is check against field estimation. From (10) one can link the cumulated fleet mileage with the average speed and T cumulated operating time:

$$Failure\ intensity\ Target_{Fleet}(T) = \frac{Total\ Number\ of\ failures\ observed\ within\ the\ Fleet}{Average\ Speed_{Fleet} * \ T} \quad (11)$$

This simple calculation shows that the Failure intensity depends on the total mileage and then also to the average speed. The Average speed can be considered as a critical factor in fleet operations, varying widely among customers due to differences in timetables, track lengths, number of stations, and other factors. As indicated by (11), it significantly influences failure intensity. Therefore, it become possible to correct prediction considering the varying average speeds.

Let’s take an example; assume that we want to predict the performance for a new product for a tender during a T period with an average speed of 35 km/h given by the customer in the contract. However, the best similar project with REX had an 27 km/h average speed. By having a higher average speed and knowing that failure occurrences are mainly time base, we can deduce that the performance of the new project will be higher than the old one for the same timeframe. So the problem is what would be the performance of the new project considering the “new” average speed is at 35 km/h and knowing that the REX is at 27 km/h?

Analogous to adjusting for fleet mileage, we recommend calculating the failure intensity as if both fleets operated at a uniform average speed—a reference average speed. This approach is formalized in equation (5).

$$Failure\ intensity\ Corrected\ _{Fleet}(T) = \frac{Total\ Number\ of\ failures\ observed\ within\ the\ Fleet}{Average\ Speed_{Nomalized} * T} \quad (12)$$

Calculated with (11):

$$Failure\ intensity\ Corrected\ _{Fleet}(T) = \frac{Average\ Speed}{Average\ Speed_{Nomalized}} * Failure\ intensity_{Fleet}\ (T) \quad (13)$$

For our example for tender it comes:

$$Failure\ intensity\ Corrected\ _{Fleet}(T) = \frac{27\ (REX)}{35(Target)} * Model\ Rex\ _{Fleet}\ (T) \quad (14)$$

In this case we apply the target speed of a new project to a given REX, but it not the only application for your proposal. the original need is to compare respective performances. Let's apply this correction to the first comparison example, which is particularly notable due to the significant difference in speeds. Metro 2 operates on a long track with a substantial distance between two stations, unlike Metro 1. Consequently, Metro 2 has an average speed of 47 km/h, while Metro 1's average speed is only 28 km/h. We will conduct an analysis assuming an average speed of 35 km/h for both metros (35 km/h is the average speed set by the platform as key assumption). For each project (i) we plot (15)

$$Model\ Corrected\ _{Fleet\ i}\ (T) = \frac{Average\ Speed(i)}{35} * \mathrm{Model}_{Fleet}\mathrm{i}\ (T) \quad (15)$$

This correction results are in the following graph:

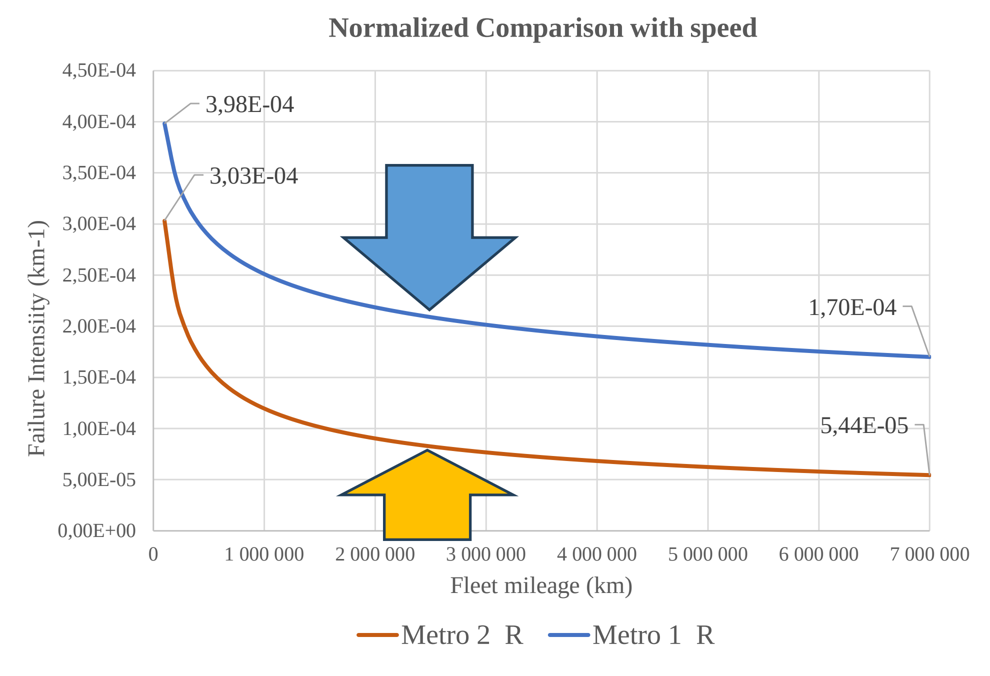

Fig. 8. Comparion with corrected an reference Average Speed

By doing this correction the gap between Metro 1 and 2 is reduced and goes to 32% : Metro 1 reaches at best 32% of Metro 2 Failure intensity. Mathematically it is formalized in (16)

$$Cumulated\ Failure\ Intensity\ Corrected_{Fleet\ i}\ (Cumulated\ km) = \lambda 0iRex * \frac{Average\ Speed(i)}{35} * (Cumulated\ km)^{\beta i-1} \quad (16)$$

In other words the speed correction change the scale parameter $\lambda 0$ (17):

$$\lambda 0i\ Corrected = \lambda 0i\ Rex * \frac{Average\ Speed\ (i)}{Average\ Speed_{Nomalized}} \quad (17)$$

The shape parameter ß is not affected by the speed considering that it rely on the process performance to detect and fix the systematic issues in the product.

VIII. Results

This method is allowing the RAM Engineer to get the Rolling Stock performances in the same conditions for mileage and speed.

Then, the remaining gaps are due to certain customer sensitivity (customers are more or less demanding in term off performances), or the climate, or the product itself: the way it was designed and manufactured but also the operational or maintenance practices.

This is the starting point of a qualitative analysis which is looking for the root causes of those differences. This analysis shall be done with the warranty teams of each project to understand their difficulties, their organization, their good practices and lesson learnt. A multi project analysis is able to find common weaknesses and strengths and push a platform evolutions.

IX. Discussion and open points

It should be understood that this comparison does not determine which product is superior in absolute terms. Rather, it aims to correct for biases introduced by differing mission profiles. The key assumption is that the failure occurrence is mainly time based which is actually the case for almost all embedded devices (except for bogie). We can also face an issue in term of lack of data representativeness, or censored data, of, for example, a reduce period of observation. In this case both parameters might be affected that why we recommend to use confidence bounds.

Another limitation for this method does not account for other significant factors such as product architecture, customer expectations, climatic conditions, or variations in manufacturing and maintenance practices. A critical aspect is product architecture; for instance, metros can range from 2-car to 9-car configurations, each with differing numbers of subsystems like traction, braking systems, auxiliary power supplies, and air conditioning units. Comparing products with significant architectural differences—such as between a 2-car and a 9-car metro—might misleadingly suggest the larger is less reliable due to its complexity. This methodology attempts to mitigate such bias during the initial scope definition and selection process. However, there are instances where a direct comparison is challenging due to the lack of comparable field data. In our analysis, comparing a 5-car to a 6-car metro assumes a level of comparability that may not exist in practice, potentially leading to inaccurate conclusions about reliability. Therefore, an ideal next step involves refining the methodology to better estimate and adjust for architectural biases.

.

X. Conclusion

The methodology offers a structured approach to comparing the reliability performances of various rolling stocks, acknowledging that they operate under diverse conditions, such as differing numbers of trains, total mileage, and operating times. In the railway industry, identical mission profiles for two distinct fleets are highly improbable due to variations in infrastructure and schedules. Direct comparisons, therefore, lack the robustness needed for accurate gap measurement. Our proposed method involves comparing performance within a standardized mission profile, necessitating Reliability Growth (RG) modelling to simulate conditions of equal mileage and speed for all subjects. The preference for the Crow AMSAA model is due to its effectiveness in fitting data towards the observation's end, providing insights into the potential optimal performance against set contractual targets. This methodology lays the groundwork for a qualitative analysis (not covered here) aimed at identifying best practices and revealing specific products' strengths and weaknesses by first quantifying performance before exploring underlying causes.

REMERCIEMENTS/THANKS

Special thanks to E. Cozzolino, B. Vidal, and P. Moran from the Metro platform for challenging the requirements and supporting me in developing the proof of concept. I would also like to express my gratitude to R. Roland, F. Seigneuret, P. Cracco and J Barbotin for their support in RAMS Metier and to the all Experts from the Alstom RAM Core Competency Network A. Peronne, M. Chenavier, A. Ragot and R. Tiennot for their valuable feedbacks, detailed reviews and corrections.