



**HAL**  
open science

## Combining Approaches of Brownian Motion and Similarity Principle to Improve the Remaining Useful Life Prediction

Dima El Jamal, Mohammed Al-Kharaz, Bouchra Ananou, Guillaume Graton,  
Mustapha Ouladsine, Jacques Pinaton

► **To cite this version:**

Dima El Jamal, Mohammed Al-Kharaz, Bouchra Ananou, Guillaume Graton, Mustapha Ouladsine, et al.. Combining Approaches of Brownian Motion and Similarity Principle to Improve the Remaining Useful Life Prediction. 2021 IEEE International Conference on Prognostics and Health Management (ICPHM), Jun 2021, Detroit (Romulus) (virtual), France. pp.1-7, 10.1109/ICPHM51084.2021.9486507. hal-03325720

**HAL Id: hal-03325720**

**<https://hal.science/hal-03325720v1>**

Submitted on 25 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Combining Approaches of Brownian Motion and Similarity Principle to Improve the Remaining Useful Life Prediction

Dima EL JAMAL

Aix Marseille Univ, Univ de Toulon  
CNRS, LIS (UMR 7020)  
Avenue Escadrille Normandie-Niemen,  
F-13397 Marseille Cedex 20, France  
Email: dima.el-jamal@lis-lab.fr

Mohammed AL-KHARAZ

Aix Marseille Univ, Univ de Toulon  
CNRS, LIS (UMR 7020),  
Avenue Escadrille Normandie-Niemen,  
F-13397 Marseille Cedex 20, France

Bouchra ANANOU

Aix Marseille Univ, Univ de Toulon  
CNRS, LIS (UMR 7020),  
Avenue Escadrille Normandie-Niemen,  
F-13397 Marseille Cedex 20, France

Guillaume GRATON

Ecole Centrale Marseille,  
Technopôle de Château-Gombert,  
38 rue Frédéric Joliot-Curie,  
F-13451 Marseille Cedex 13, France

Mustapha OULADSINE

Aix Marseille Univ, Univ de Toulon  
CNRS, LIS (UMR 7020),  
Avenue Escadrille Normandie-Niemen,  
F-13397 Marseille Cedex 20, France

Jacques PINATON

STMicroelectronics Rousset,  
190 avenue Celestin Coq,  
ZI - Rousset-Peynier,  
F-13106 Rousset, France

**Abstract**—This paper proposes a data-driven framework for Remaining Useful Life (RUL) prediction, based on the Brownian Motion model (BM) and the similarity principle, for an operating system given its health indicator. It addresses the issues of noisy and limited run-to-failure (R2F) data. The Percentile filtering is used to extract, from the R2F data, 100 monotonic profiles used as references in the modeling and the RUL prediction. Then, the similarity is computed between these references and the Health Indicator (HI) of the operating system. Fitting the most similar reference into the BM improves the RUL prediction. A numerical application using simulated data justifies the accuracy of this approach.

## I. INTRODUCTION

Nowadays, Condition-Based Maintenance (CBM) [1] is a maintenance strategy that the manufacturing industry tries to adopt more and more in order to reduce the life cycle cost of the equipment units and to extend their availability. CBM uses real-time information to optimize the maintenance moment by restoring the functional properties of the equipment units. It is based on the current health monitoring of the equipment units, thus it is important to add prognosis tools to predict the future state and to anticipate the maintenance.

Fault prognosis is one of the major tasks in CBM. It estimates the RUL of the equipment units based on the condition monitoring information.

Generally, prognosis approaches can be classified into three main categories depending on the type of the used information. These categories [2], [3] are defined as physical model-based approaches, data-driven approaches and fusion-based approaches.

The physical model-based approaches [4] use an explicit mathematical model to represent a dynamical system's degradation. Data-driven approaches are based on condition mon-

itoring data for analyzing and predicting the current and the future health conditions. Fusion-based approaches [5] are a combination of these two approaches. In the real case, understanding and modeling the degradation of complex systems are difficult especially when taking into account all possible degradation effects. Thus, the physical model-based approaches are complicated to use even though they give the most accurate RUL estimation. On the other hand, data-driven approaches do not require understanding the physical behavior and the degradation process which makes them faster to deploy compared to the other approaches. Hence, for these reasons, this category of approaches is adopted in this paper.

Data-driven approaches include mainly statistical and Artificial Intelligence (AI) methods. The statistical methods estimate the RUL based on the expected health trend under known operating conditions (e.g. Wiener Process [6], Gamma Process [7], autoregressive model [8], Hidden Markov / Semi-Markov models [9]). The AI methods estimate the RUL by mimicking human brain structure using several methods like neural networks (e.g. Long short-term memory [10]), support vector machine [11], etc.

The Brownian Motion [12] is one of the statistical methods widely used to model degradation processes. In [13], it is used with an adaptive drift parameter to model the HI of the operating device and to predict its future state for RUL prediction, based on its past degradation information and R2F indicator. In [14], the BM is introduced in the similarity-based approach. This category of approach employs all the R2F profiles generated in the same operating conditions of the operating device to predict its RUL [15]. These R2F profiles similar to the operating device's HI provide additional contributions to the prediction. Thus, in our work we combine

the strengths of these two methods by using the adaptive BM model proposed in [13] that introduces the references's information and the operating device's HI after a similarity computation as in [14]. However, the problem is that the R2F data are sometimes limited as the devices are not allowed to be used until failure for finance and safety issues. Hence, the objective of this paper is to deal with the lack of R2F data. The proposed solution consists in generating references for the RUL prediction in the case of the availability of just one noisy R2F indicator. This can be done using the Percentile filtering suggested in [16]. Then, the similarity principle is introduced in the BM model in a new way in order to predict the RUL.

The remainder of this paper is structured as follows: the proposed approach is detailed in section II. Section III presents the approach's application on simulated data and a comparison with a non-similarity-based approach [13]. Finally, section IV concludes the work and presents some perspectives.

## II. APPROACH FRAMEWORK

The proposed approach is illustrated in Fig.1. Given the R2F indicator of a system, the objective is to predict the RUL of an operating device sharing similar operating conditions. Generally, health indicators are highly noisy, hence a pre-processing step is required. The filtering method proposed in [16] and based on the notion of percentile is applied for this purpose. This filter extracts from the R2F indicator, 100 monotonic profiles that cover it and are considered as references in this study. Similarly, the HI of the operating device is also filtered but only the 50<sup>th</sup> extracted profile is used for RUL prediction as it is the median profile. At each prediction time, the similarity between the references and the 50<sup>th</sup> extracted profile from the operating device's HI is calculated and the most similar reference to it is chosen and modeled by BM as in [13]. Its parameters are used to predict the RUL of the operating device. The different parts of this approach are detailed in the following subsections.

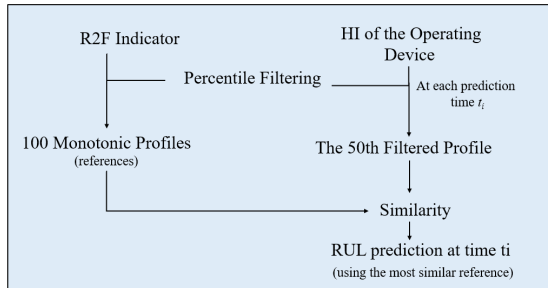


Fig. 1: Framework of the proposed approach

### A. HI filtering

The Percentile filtering is used since it minimizes information loss and uncertainty in the data compared to other existing methods, and it is suitable for finding a solution to our issues [16]. This method is briefly described in the following steps.

Let  $\mathcal{Y}_t$  be the raw HI that measures the degradation over time  $t$ ,  $V_N$  and  $V_F$  be respectively the normal and the failure threshold values set by the experts.

- First, the interval  $[V_N, V_F]$  is divided into several sub-intervals with step  $s$ , where  $s$  is chosen in such a way to make a compromise between the preservation of information and the complexity of calculation:

$$s = \min \left[ \frac{\text{Median}(|\mathcal{Y}_t - \mathcal{Y}_{t-1}|)}{10}, 0.01 \right] \quad (1)$$

where,  $\text{Median}(|\mathcal{Y}_t - \mathcal{Y}_{t-1}|)$  is the median value of the variability  $(\mathcal{Y}_t - \mathcal{Y}_{t-1})$ .

- The splitting values  $v$  of  $[V_N, V_F]$  are summarized in  $\mathcal{V} = \{V_N, V_{N+s}, V_{N+2s}, \dots, V_F\}$ . A set  $T_v = \{t_1, t_2, \dots, t_n\}$ , with  $n$  the number of its elements, is constructed for each value  $v$ , by selecting the time  $t_i$  verifying the following condition (see Fig.2):

$$(\mathcal{Y}_{t_i} < v) \ \& \ (\mathcal{Y}_{t_{i+1}} \geq v) \quad (2)$$

This selection is based on the fact that the HI is noisy and it can cross the value  $v$  many times.

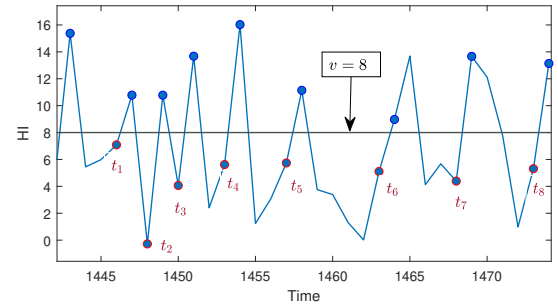


Fig. 2:  $\mathcal{Y}_t$  passes through a value  $v$  many times

- In the following, for each  $v \in \mathcal{V}$ , the goal is to calculate the percentiles denoted as  $t_v^p$  for  $p \in \{1, 2, \dots, 100\}$  of the set  $T_v$ . They are used next to construct the monotonic profiles  $(t_v^p, v)$  relative to each  $p$ . These percentiles are computed in two steps:

- 1) After placing the values  $t_i$  of the set  $T_v$  in ascending order, the  $n$  corresponding percentiles to these values are computed according to relative position calculation as:  $p = 100(j - 0.5)/n$ ,  $j \in \{1, 2, \dots, n\}$ .
- 2) The value of the rest  $(100 - n)$  percentiles is calculated as follows:

$$t_v^p = \begin{cases} t_1 & \text{for } p < 100(1 - 0.5)/n \\ t_n & \text{for } p > 100(n - 0.5)/n \\ \text{LI} & \text{for } (1 - 0.5) < \frac{np}{100} < (n - 0.5) \end{cases} \quad (3)$$

where LI stands for Linear Interpolation: given  $(p_1, t_v^{p_1})$  and  $(p_2, t_v^{p_2})$  two pairs of a percentile and its corresponding value respectively, the percentiles  $p_i$  between  $p_1$  and  $p_2$  (with  $p_1 < p_2$ ) are computed by LI as follows:

$$t_v^{p_i} = t_v^{p_1} + (p_i - p_1) \frac{t_v^{p_2} - t_v^{p_1}}{p_2 - p_1} \quad (4)$$

- 100 profiles are generated, each one corresponding to a percentile  $p$  (see Fig. 3). These profiles  $Y_t^p$  are constructed by assimilating to each splitting value  $v$  of  $V$ , the time  $t_v^p$  corresponding to the percentile  $p$  of  $T_v$ . The  $t_v^p$  that breaks the monotony of the profile is omitted.

### B. HI modeling

The Brownian Motion with an adaptive drift parameter is used to model the HI as in [13] and [17]. The evolution of the filtered HI is modeled iteratively over time. The strength of this method relies on the introduction of the current and the historical degradation information in the model to forecast the future trend of the health indicator. This is done by updating the drift parameter whenever a new observation is available. The degradation model is given as follows:

$$y_{t_i} = y_{t_{i-1}} + \mu_{t_{i-1}}(t_i - t_{i-1}) + \sigma \varepsilon_{t_{i-1}, t_i} \quad (5)$$

where,  $y_{t_i}$  is the degradation information at time  $t_i$ ,  $\mu_{t_{i-1}}$  is the updated drift parameter at  $t_{i-1}$  after observing  $y_{t_{i-1}}$ ,  $\sigma$  is a constant and  $\sigma \varepsilon_{t_{i-1}, t_i}$  is the error such that  $\varepsilon_{t_{i-1}, t_i} \sim \mathcal{N}(0, t_i - t_{i-1})$  by the BM. The variance  $t_i - t_{i-1}$  makes sense as the random quantity should somehow increases when  $t_i - t_{i-1}$  increases.

The Kalman Filter (KF) [18] is a recursive procedure used in this case to estimate and update the drift parameter. It computes the optimal estimator of  $\mu_t$  at time  $t$ , based on the available observations up to  $y_t$  and including it. In order to apply the KF, the system and the observation equations are given respectively in the following:

$$\mu_{t_i} = \mu_{t_{i-1}} + \nu \quad (6)$$

$$y_{t_i} - y_{t_{i-1}} = \mu_{t_{i-1}}(t_i - t_{i-1}) + \sigma \varepsilon_{t_{i-1}, t_i} \quad (7)$$

where,  $\nu$  is the system error normally distributed such as  $\nu \sim \mathcal{N}(0, Q)$ .

Under the normality assumption, the initial  $\mu_{t_0}$  has a normal distribution with mean  $\hat{\mu}_{t_0}$  and variance  $P_0$ . Thus,  $\mu_{t>0}$  is a linear combination of two random variables, both with normal distributions, hence it is itself normally distributed and its mean (8) and variance (9) are updated by KF as follows:

$$\hat{\mu}_{t_i} = \hat{\mu}_{t_{i-1}} + P_{i|i-1}(t_i - t_{i-1})F_i^{-1}(y_{t_i} - y_{t_{i-1}} - \hat{\mu}_{t_{i-1}}(t_i - t_{i-1})) \quad (8)$$

$$P_i = P_{i|i-1} - P_{i|i-1}(t_i - t_{i-1})^2 F_i^{-1} P_{i|i-1} \quad (9)$$

where,

$$\begin{cases} P_{i|i-1} &= P_{i-1} + Q \\ F_i &= (t_i - t_{i-1})^2 P_{i|i-1} + \sigma^2(t_i - t_{i-1}) \end{cases} \quad (10)$$

Thus, the model of the future behavior of the HI after observing  $y_{t_i}$  is:

$$y_t = y_{t_i} + \hat{\mu}_{t_i}(t - t_{i-1}) + \sigma \varepsilon_{t_{i-1}, t_i} \quad (11)$$

where  $\hat{\mu}_{t_i}$  is a function of  $(y_1, \dots, y_{t_i})$  for  $t_1 < t_2 < \dots < t_i \leq t$ .

### C. Similarity

In this subsection, the key elements of the similarity principle are presented. Let  $Y_{op}$  and  $Y_{rk}$  be respectively the health indicator of the operating device and the reference  $k$  considered both in the same operating conditions. Thus, the references can improve the device's RUL prediction.

As in [14] and [15], the similarity between the operating device's HI and a reference is measured over a time interval  $I$  by a similarity function  $S$ . For the similarity computation,  $(N + 1)$  consecutive monitoring points are considered. In particular, the last  $(N + 1)$  observations of the operating device's HI are chosen as they reflect the current state of the device. Thus, the time interval can be represented as  $I = (N + 1)\Delta t$ , where  $\Delta t$  is the condition monitoring time interval.

In this paper,  $S$  is the Euclidean distance and  $S_{op \leftrightarrow k}(t_i, N)$  denotes the similarity calculated between the operating device and the reference  $k$  at the time  $t_i$  for a determined duration  $I$ :

$$S_{op \leftrightarrow k}(t_i, N) = \sum_{n=0}^N [Y_{op}(t_i - n\Delta t) - Y_{rk}(t_i - n\Delta t)]^2 \quad (12)$$

Note that the following condition should be satisfied:

$$N + 1 \leq \min(N_{op}, N_{rk}) \quad (13)$$

where  $N_{op}$  and  $N_{rk}$  are respectively the number of condition monitoring points of  $Y_{op}$  and  $Y_{rk}$ .

At each prediction time  $t_i$ , the most similar reference to  $Y_{op}$  is determined and used as a support for the RUL prediction. It is chosen in such a way to minimize the similarity function  $S_{op \leftrightarrow k}(t_i, N)$  and is denoted as  $r_{sim}$ :

$$r_{sim} = \arg \min_{1 < k < N_T} S_{op \leftrightarrow k}(t_i, N) \quad (14)$$

where,  $N_T$  is the total number of references.

### D. RUL prediction

The objective of our work is to predict accurately the time required for the system to fail. It is done by forecasting the path of the degradation indicator  $Y_{op}$  up to the failure threshold. For the RUL prediction, only the profile  $Y_{op}^{50}$  of the operating device ( $p = 50$ ) is considered after filtering as it is the median of the generated profiles. At each prediction time, the degradation information is available until  $t_i$  ( $Y_{op}^{50}(t_1 \rightarrow t_i)$ ) and when the profile  $Y_{op}^{50}$  exceeds the normal operating threshold  $V_N$ , the prognosis is triggered.

In order to have an accurate prediction, the degradation information of both  $Y_{op}^{50}$  and  $Y_{r_{sim}}$  are used in the BM model [13]: the behavior of  $Y_{r_{sim}}$  is taken into account in order to forecast  $Y_{op}^{50}$  until failure. This is done by predicting the future evolution of the drift parameter of  $Y_{op}^{50}$  based on the evolution of  $Y_{r_{sim}}$  as detailed just below. At each prediction time  $t_i$ :

- The corresponding time  $t_c$  on which  $Y_{r_{sim}}$  has almost the same degradation state as  $Y_{op}^{50}$  is determined as :

$$Y_{r_{sim}}(t_c) \leq Y_{op}^{50}(t_i) \leq Y_{r_{sim}}(t_{c+1}) \quad (15)$$

- The drift parameter  $\hat{\mu}_{op}^{50}$  of  $Y_{op}^{50}$  and  $\hat{\mu}_{r_{sim}}^{sim}$  of  $Y_{r_{sim}}$  are calculated using the KF (equations 8-10).
- The residual is computed as:

$$r_i = \hat{\mu}_{t_i}^{op} - \hat{\mu}_{t_i}^{sim} \quad (16)$$

- The predicted values of the drift parameter  $\hat{\mu}^{op}$  to consider for  $t > t_i$  are:

$$P_{t_i} = \{\hat{\mu}_{t_c}^{sim} + |r_i|, \hat{\mu}_{t_c+1}^{sim} + |r_i|, \dots, \hat{\mu}_{t_{Max}}^{sim} + |r_i|\} \quad (17)$$

where  $t_{Max}$  is the time of the last monitoring point of  $Y_{r_{sim}}$ . The absolute value of  $r_i$  is taken in order to keep the index monotone and subsequently to avoid its divergence from the failure threshold  $V_F$ .

- The expected value of  $Y_{op}^{50}$  is predicted iteratively over  $t$  as:

$$\diamond \text{ for } 0 \leq k < t_{Max} - t_c$$

$$\begin{aligned} \mathbb{E}[Y_{op}^{50}(t_{i+k+1})] &= \mathbb{E}[Y_{op}^{50}(t_{i+k}) + \hat{\mu}_{t_c+k}^{sim} + |r_i| + \sigma\epsilon_{0,1}] \\ &= \mathbb{E}[Y_{op}^{50}(t_{i+k})] + \hat{\mu}_{t_c+k}^{sim} + |r_i| \end{aligned} \quad (18)$$

$$\diamond \text{ for } [\diamond]k \geq t_{Max} - t_c$$

$$\begin{aligned} \mathbb{E}[Y_{op}^{50}(t_{i+k+1})] &= \mathbb{E}[Y_{op}^{50}(t_{i+k}) + \hat{\mu}_{t_{Max}}^{sim} + |r_i| + \sigma\epsilon_{0,1}] \\ &= \mathbb{E}[Y_{op}^{50}(t_{i+k})] + \hat{\mu}_{t_{Max}}^{sim} + |r_i| \end{aligned} \quad (19)$$

- The predicted RUL at  $t_i$  is equal to  $\hat{k}\Delta t$  unit of time, such that:

$$\hat{k} = \inf\{k : E(Y_{op}^{50}(t_{i+k})) \geq V_F\} \quad (20)$$

### III. APPLICATION

The proposed approach is tested on simulated data on *Matlab*. Its evaluation is then performed using metrics dedicated to fault prognosis.

In order to highlight the effectiveness of our approach, it is compared to a non-similarity-based approach [13]. For that, the RUL is computed in two ways:

- With similarity (approach presented in this paper): the RUL is computed by picking at each prediction time  $t_i$ , the suitable reference to  $Y_{op}^{50}$  from the entire available collection.
- Without similarity (non-similarity-based approach): the RUL is predicted using the same reference during all the prediction times  $t_i$  as in [13]. For this computation, the 50<sup>th</sup> profile of the operating device and the reference  $Y_{r_{50}}$  are used.

The two approaches are compared on two different scenarios: in the first one, the HI of the operating device and the reference present almost the same trend, but their tendencies are different in the second one.

In the remainder of this section, the given data and the Percentile filtering are presented in (III-A), the RUL prediction based on the similarity principle and the BM is detailed in (III-B). Finally, the approaches's evaluation is given in (III-C).

#### A. Data & Filtering

The R2F indicator and its 100 extracted profiles by the Percentile filtering are displayed in Fig.3. The normal and the failure threshold are predefined as  $V_N = -10$  and  $V_F = 28$  respectively. As shown, this indicator (in blue) is noisy. Thus, the filtering is required as the quality of the HI affects the RUL prediction. The extracted profiles are monotonic and each one corresponds to a percentile  $p$ . They represent the useful information drowned in the noise of the HI. These extracted profiles are the references in the two cases.

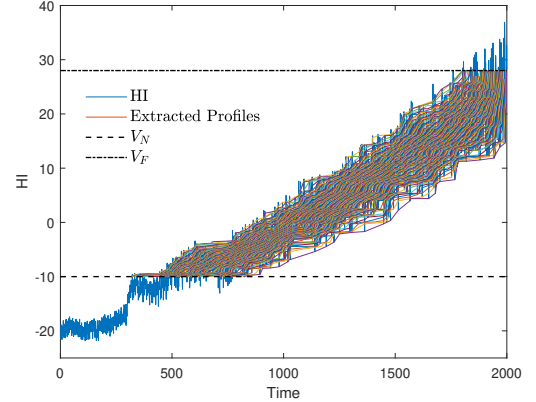


Fig. 3: Profiles extracted from the R2F indicator

In Fig.4, the health indicator of the operating device (in red) and the previous R2F indicator (in blue) are illustrated for the two scenarios.

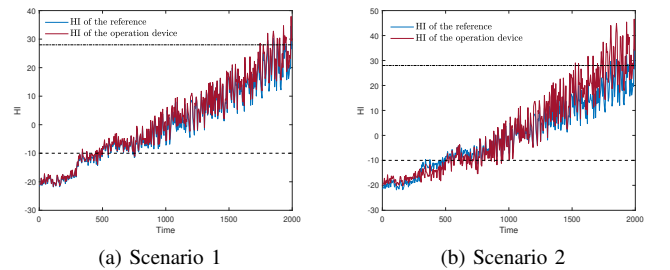


Fig. 4: Health indicators of the operating device and the reference

These indicators are then filtered by the Percentile method, and as mentioned previously, only  $Y_{op}^{50}$  of the operating device is used. This profile, in addition to the references, are illustrated in Fig.5.

For the first scenario, the profiles  $Y_{op}^{50}$  and  $Y_{r_{50}}$  have almost the same degradation dynamics but a notable difference is especially shown in the second scenario. In fact, the degradation of  $Y_{op}^{50}$  is slow at first in comparison with the reference  $Y_{r_{50}}$ , and becomes faster after the 1000<sup>th</sup> instant. Therefore, this scenario is considered to show the impact of the similarity on RUL prediction.

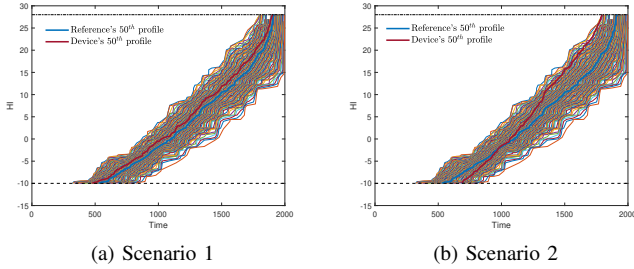


Fig. 5:  $Y_{op}^{50}$ ,  $Y_{r50}$  and the rest extracted profiles from the R2F indicator

### B. RUL Prediction

When the profile  $Y_{op}^{50}$  exceeds the phase of normal functioning ( $V_N = -10$ ), RUL prediction is launched. Thus, the prediction starts at  $T_{init} = 500$  in the first scenario, and at  $T_{init} = 600$  in the second one (Fig.5). In this application,  $N$  is set to 50 and  $\Delta t$  to 1 unit of time.

The real failure time denoted as  $T_{end}$  is the time when the 50<sup>th</sup> filtered profile of the operating system reaches the failure threshold as:

$$T_{end} = \inf\{t_i : Y_{op}^{50}(t_i) \geq V_F\} \quad (21)$$

In the first scenario,  $T_{end} = 1907$ , and the real RUL at each prediction time  $t_i$  is  $RUL_{real}(t_i) = T_{end} - t_i$ . This RUL is computed in order to evaluate our predictions. Idem for the second case where  $T_{end} = 1807$ .

The comparison between the two predicted RUL (with and without similarity) is performed for the two scenarios and illustrated respectively in Fig.6 and Fig.7.

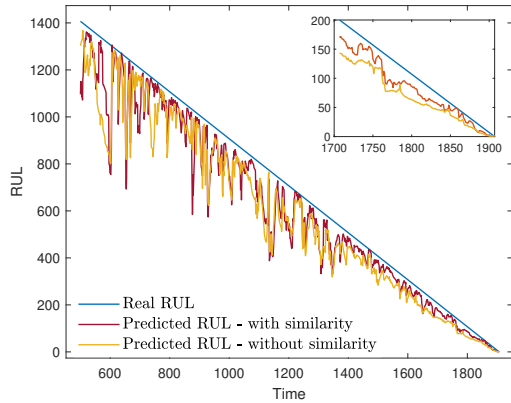


Fig. 6: Predicted RUL (with and without similarity) of the operating device and a zoom on the last 200 times - Scenario1

The result shows that the proposed approach outperforms the non-similarity-based approach as the difference between the real and the predicted RUL is more important when fixing the reference than in our approach. It can also be noticed that the RUL prediction is improved during the time as more data become available. This is especially notable on the last 200

prediction times of our approach as shown on the zoom in the figures.

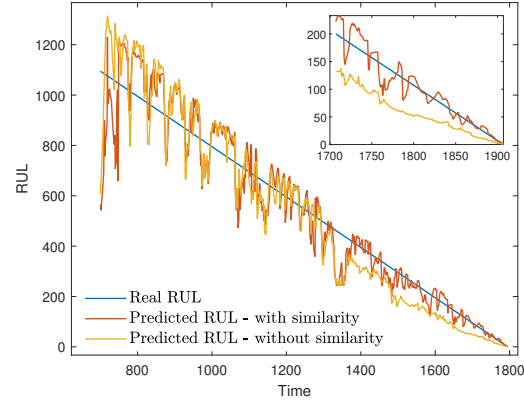


Fig. 7: Predicted RUL (with and without similarity) of the operating device and a zoom on the last 200 times - Scenario2

The improvement of the RUL prediction by the proposed approach is explained by the fact that using the most similar reference to  $Y_{op}^{50}$  at each prediction time  $t_i$  is more effective especially when the two HI do not have exactly the same trend. In Fig.8, the similar references chosen by our approach during the prediction are presented.

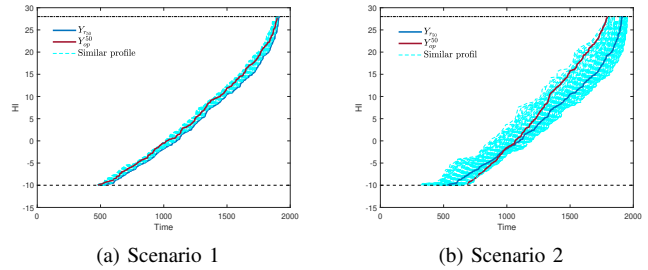


Fig. 8: The similar profiles to  $Y_{op}^{50}$  extracted from the R2F indicator

For the first scenario, these references belong to the set  $p = \{27, \dots, 48\}$  and are close to  $Y_{r50}$  and  $Y_{op}^{50}$  since the two HI have the same trend all along the time interval. However, in the second scenario, a greater variety of similar profiles is chosen: 67 profiles are particularly picked by the proposed approach between  $p = 1$  and  $p = 75$  following the same evolution of  $Y_{op}^{50}$ . Thus, using only  $Y_{r50}$  as the reference for the RUL prediction is not sufficient specially when  $Y_{r50}$  and  $Y_{op}^{50}$  have different trends.

The introduction of the similarity principle in the RUL prediction is suitable for cases like this scenario where the degradation information of the operating device and the R2F data are not exactly the same as in the first one (ideal scenario). In addition, this approach is more robust and more suitable for real cases as it improves RUL prediction when R2F data are limited.



### C. Evaluation

To evaluate numerically the two presented approaches, the Root Mean Square Error (RMSE) and the  $(\alpha - \lambda)$  accuracy are computed. These metrics [19] can only be used when the real RUL is available. They evaluate the prognosis's results from different aspects and allow us to compare these approaches.

1) RMSE is defined by:

$$RMSE = \sqrt{\frac{\sum_{t_i=T_{init}}^{T_{end}} (\text{RUL}_{real}(t_i) - \text{RUL}_{pred}(t_i))^2}{T_{end} - T_{init}}} \quad (22)$$

This metric is computed for the two approaches throughout the prediction interval and over the last 200 predictions time (see Table I).

TABLE I: The result of the metric RMSE

Prediction interval	Scenario	Non-similarity-based approach	Proposed approach
	Scenario 1	141.22	121.85
	Scenario 2	104.23	102.29
Last 200 predictions	Scenario 1	37.20	22.76
	Scenario 2	48.67	19.59

It can be noticed that the prediction error is smaller when the similarity principle is introduced in the BM model to predict the RUL.

2)  $(\alpha - \lambda)$  accuracy is a binary metric that determines whether a prediction result falls or not within the accuracy zone at a specific time  $t_\lambda$ , where  $\lambda \in [0, 1]$ . The accuracy zone is defined with respect to a percentage  $\alpha$  of accuracy related to the real RUL. This zone becomes smaller as time increases by reflecting that the prediction accuracy increases as more data are available. In this application,  $\alpha$  is fixed at 20% for precision.

The time  $t_\lambda$  is computed between the beginning time of the prediction ( $\lambda = 0$ ) and the real failure time ( $\lambda = 1$ ) as in the following:

$$t_\lambda = T_{init} + \lambda(T_{end} - T_{init}) \quad (23)$$

It can be noticed as shown in Fig.9 and Fig.10, that the predicted RUL corresponding to  $t_\lambda$  for  $\lambda \in [0, 0.1, 0.2, \dots, 1]$  falls more in the accuracy zone in our approach compared to the non-similarity-based approach. This result is also valid for  $\lambda$  calculated in a step of 1/100 between 0 and 1. When computing the RUL at each of these  $t_\lambda$ , we found that 70 predictions in our approach fall in the accuracy zone compared to only 53 in the non-similarity-based approach. Similarly, for the second scenario where 29 more predictions in the proposed approach belong to the accuracy zone especially in the last 200 prediction times as seen in the figures.

According to these metrics, the proposed approach outperforms the non similarity-based approach confirming the graphic results.

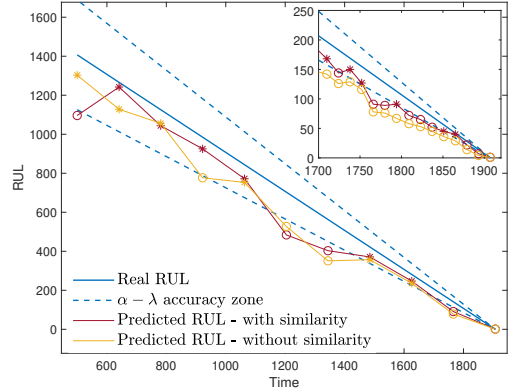


Fig. 9: The result of the metric  $(\alpha - \lambda)$  accuracy - Scenario1

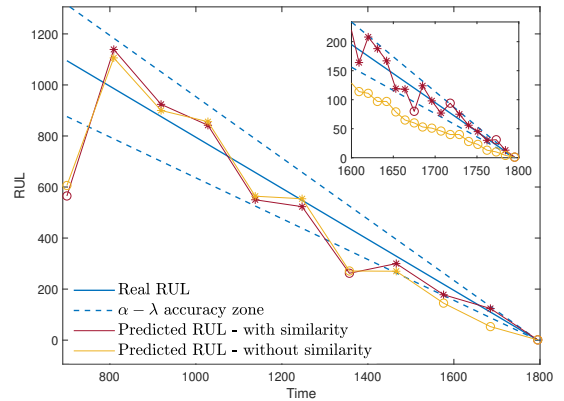


Fig. 10: The result of the metric  $(\alpha - \lambda)$  accuracy - Scenario2

## IV. CONCLUSION

This paper deals with the problem of RUL prediction when R2F data are limited. It proposes a new approach that combines the Brownian Motion model and the similarity principle after a pre-processing step of the health indicators by the Percentile filtering. An implementation with experiments on simulated data is carried out. The obtained results show the effectiveness of our approach compared to existing methods particularly to the non-similarity-based approaches.

In the future, we intend to extend this work in at least two directions: testing another method for HI modeling and applying the approach to real data.

## REFERENCES

- [1] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [2] B. Mrugalska, "Remaining useful life as prognostic approach: a review," in *International Conference on Human Systems Engineering and Design: Future Trends and Applications*. Springer, 2018, pp. 689–695.
- [3] K. Javed, R. Gouriveau, and N. Zerhouni, "State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels," *Mechanical Systems and Signal Processing*, vol. 94, pp. 214–236, 2017.

- [4] J. Luo, K. R. Pattipati, L. Qiao, and S. Chigusa, "Model-based prognostic techniques applied to a suspension system," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 38, no. 5, pp. 1156–1168, 2008.
- [5] L. Liao and F. Köttig, "A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction," *Applied Soft Computing*, vol. 44, pp. 191–199, 2016.
- [6] X. Wang, C. Hu, X. Si, Z. Pang, and Z. Ren, "An adaptive remaining useful life estimation approach for newly developed system based on nonlinear degradation model," *IEEE Access*, vol. 7, pp. 82 162–82 173, 2019.
- [7] Z. Wu, Z. Wang, C. Qian, B. Sun, Y. Ren, Q. Feng, and D. Yang, "Online prognostication of remaining useful life for random discharge lithium-ion batteries using a gamma process model," in *2019 20th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*. IEEE, 2019, pp. 1–6.
- [8] S. Ahsan and T. A. Lemma, "Remaining useful life prediction of gas turbine engine using autoregressive model," in *MATEC Web of Conferences*, vol. 131. EDP Sciences, 2017, p. 04014.
- [9] L. Fan, S. Wang, H. Duan, and H. Ran, "Fatigue crack fault diagnosis and prognosis based on hidden semi-markov model," *The Journal of Engineering*, vol. 2019, no. 13, pp. 406–410, 2019.
- [10] Q. Huang, Z. Kang, Y. Zhang, and D. Yan, "Tool remaining useful life prediction based on edge data processing and lstm recurrent neural network," in *2020 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE, 2020, pp. 1–5.
- [11] H.-Z. Huang, H.-K. Wang, Y.-F. Li, L. Zhang, and Z. Liu, "Support vector machine based estimation of remaining useful life: current research status and future trends," *Journal of Mechanical Science and Technology*, vol. 29, no. 1, pp. 151–163, 2015.
- [12] G. A. Whitmore and F. Schenkelberg, "Modelling accelerated degradation data using wiener diffusion with a time scale transformation," *Lifetime data analysis*, vol. 3, no. 1, pp. 27–45, 1997.
- [13] T.-B.-L. Nguyen, M. Djeziri, B. Ananou, M. Ouladsine, and J. Pinaton, "Fault prognosis for discrete manufacturing processes," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 8066 – 8072, 2014, 19th IFAC World Congress.
- [14] H. Zhang, C. Hu, X. Kong, and W. Zhang, "A model for residual life prediction based on brownian motion in framework of similarity," *Asian Journal of Control*, vol. 18, no. 4, pp. 1406–1416, 2016.
- [15] M.-Y. You and G. Meng, "A framework of similarity-based residual life prediction approaches using degradation histories with failure, preventive maintenance, and suspension events," *IEEE Transactions on Reliability*, vol. 62, no. 1, pp. 127–135, 2013.
- [16] T. L. Nguyen, M. Djeziri, B. Ananou, M. Ouladsine, and J. Pinaton, "Fault prognosis for batch production based on percentile measure and gamma process: Application to semiconductor manufacturing," *Journal of Process Control*, vol. 48, pp. 72–80, 2016.
- [17] W. Wang, M. Carr, W. Xu, and K. Kobbacy, "A model for residual life prediction based on brownian motion with an adaptive drift," *Microelectronics Reliability*, vol. 51, no. 2, pp. 285–293, 2011.
- [18] A. C. Harvey, *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.
- [19] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to rul prediction," *Mechanical systems and signal processing*, vol. 104, pp. 799–834, 2018.