An ensemble of models for integrating dependent sources of information for the prognosis of the remaining useful life of Proton Exchange Membrane Fuel Cells

Dacheng Zhang, Piero Baraldi, Catherine Cadet, Nadia Yousfi-Steiner, Christophe Bérenguer, Enrico Zio

An ensemble of models for integrating dependent sources of information for the prognosis of the remaining useful life of Proton Exchange Membrane Fuel Cells

D. Zhang[1,2], P. Baraldi[3], C. Cadet[2], N. Yousfi-Steiner[4,5,6], C. Bérenguer[2,*], E. Zio[3,7]

Abstract

This paper presents a prognostic approach based on an ensemble of two degradation indicators for the prediction of the Remaining Useful Life (RUL) of a Proton Exchange Membrane Fuel Cell (PEMFC) stack. When the fuel cell stack experiences variable operating conditions, degradation indicators, such as the stack voltage and the stack State Of Health (SOH), are not able to individually provide precise and robust RUL predictions. The stack voltage does not directly measure the component degradation, as it is only related to degradation symptoms, which are significantly affected by operating conditions. The SOH provides aging information but it can only be measured at low frequency in industrial applications. The objective of this work is to combine the two indicators, leveraging their strengths and overcoming their drawbacks. Two different physics-based models are used to this aim: the first model receives a signal directly observable and related to the stack voltage, which can be frequently and easily measured; the second model is fed by periodic measurements from the physical characterization of the stack, which gives reliable information on the SOH evolution. The prognostic procedure is implemented using Particle Filtering (PF), and the outcomes of the two prognostic filters are aggregated to obtain the ensemble predictions. The ensemble-based approach employs a local aggregation technique that combines the outcomes of two prognostic models by assigning to each model a weight and a bias correction, which are obtained considering the individual models' local performances. The dependence between the two indicators is also accounted for, by dependent Gamma processes. The results obtained show that the accuracy of the RUL predictions obtained by the proposed ensemble-based method outperforms that obtained by the individual models.

Keywords: Proton Exchange Membrane Fuel Cell (PEMFC), Prognostics, Remaining Useful Life (RUL), Particle Filtering (PF), Ensemble, Dependent processes

*Corresponding author

Email address: christophe.berenguer@grenoble-inp.fr (C. Bérenguer)

[1]Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, 650500, China

[2]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

[3]Energy Department, Politecnico di Milano, Via la Masa 34, 20156 Milano, Italy

[4]FEMTO-ST, CNRS / Univ. Bourgogne Franche-Comté, Belfort, France

[5]FCLAB, CNRS / Univ. Bourgogne Franche-Comté, Belfort, France

[6]LABEX ACTION, CNRS, Belfort, France

[7]Chair on System Science and the Energetic Challenge, Foundation Electricité de France (EDF), Centrale Supélec, Univ. Paris Saclay, 92290 Chatenay-Malabry, France

1. Introduction

Proton Exchange Membrane Fuel Cell (PEMFC) has been considered as one of the most promising technologies for both stationary and transportation applications. However, it is not yet ready for large-scale industrial deployment and commercialization because of its limited durability [1]. Prognostics and Health Management (PHM) approaches offer solutions to estimate the State Of Health (SOH) of fuel cell stacks and to predict their Remaining Useful Life (RUL). Prognostic results can be helpful in making decisions on maintenance scheduling and control strategy to properly operate components and systems [2, 3].

The degradation of a PEMFC stack can be assessed from different measurements. The stack voltage (or power) is the most commonly used indicator in literature [4, 5, 6, 7]. Since this indicator can be continuously monitored in PEMFC applications, it is a good candidate for online prognostic purposes. However, the coexistence of reversible and irreversible degradation phenomena significantly limits its accuracy [6]. Degradation information can also be obtained by estimating the SOH using characterization measurements, such as Electrochemical Impedance Spectroscopy (EIS) [8] and polarization curves [9, 10]. Although the SOH typically provides an accurate estimation of the PEMFC degradation, it can only be measured parsimoniously given that this characterization is intrusive to the stack performance and may introduce additional complex degradation phenomena [6]. The contribution developed in this work aims at proposing a way of using jointly such two complementary deterioration indicators to elaborate a RUL prognosis result.

Prognostic approaches are typically distinguished into model-based, data-driven and hybrid approaches [11, 12, 13]. Model-based approaches use mathematical equations to describe degradation phenomena and predict failure times. They can be applied only when knowledge about failure mechanisms, material properties and external loading is available [5]. These approaches are used in specific applications, where an accurate analytic description of the system behavior has been developed. However, the necessary knowledge is not always available or mature, and for a real system, it is difficult (or even impossible) to obtain a degradation model in analytic form to integrate the degradation phenomena. Data-driven approaches such as those based on neural networks [14] and other meta-modeling techniques, do not require the availability of analytical models of the system behavior. They are often considered as "black boxes" because the behavior of the system is directly learned from historical data. They are usually easy to implement, as they do not explicitly model the links between internal phenomena and external observations. Data-driven approaches are, therefore, flexible to different problems but impose a high cost of data collection [15, 16]. One of the main limitations of data-driven approaches lies in the requirement of training data, *i.e.* historical data used to infer correlations, establish patterns and evaluate data trends leading to failure. Hybrid approaches are the combination of the two previous types. They are based on physical equations, whose parameters change over time and are estimated by data-driven learning [7, 17, 18]. The advantage of combining data-driven approaches (such as degradation trend regression or Artificial Neural Networks) with physical model-based approaches is that the data limitation and lack of knowledge can be mitigated. In [19], RUL predictions are carried out using statistical degradation model obtained built on real degradation tests by a Bond-graph technique. The prognostics problem is formulated as the joint state-parameter estimation problem within a Particle Filtering framework where estimations of state of health (SOH) is obtained in probabilistic terms. In [7], an innovative robust prediction algorithm for performance degradation of PEMFC is proposed based on the combination of a degradation trend model and a Nonlinear Auto Regressive Neural Network (ANRNN) model. A recent review on PHM for PEMFC is presented in [12]. Al-

though several important research works have already been reported on PHM and RUL prognosis of PEMFC systems, this field is still at the early stage and needs further development. One of the main obstacles identified is the lack of available test and failure data. Another problem is that the aging and failure mechanisms of the PEMFC are not yet entirely clear, because of the influence of the different operating conditions.

In this work, we develop a hybrid approach based on an ensemble of models, which uses prediction outcomes provided by different degradation models fed by different deterioration measurements and properly combines them to provide the prognostic results. Ensemble of models have shown promising results for the prognostics of industrial systems [20, 21, 22, 23, 24, 25]. For example in [23], an ensemble approach based on a semi-Markov model and a fuzzy similarity model has been developed for the predictions of the RUL of a heterogeneous fleet of aluminum electrolytic capacitors used in electric vehicle power trains. The ensemble output is obtained by local aggregation of the outcomes of two prognostic models, assigning to each model a weight and a bias correction, which are estimated considering the models' local performances, *i.e.* the inaccuracies in predicting the RUL of validation patterns similar to the one under assessment. However, this kind of approach requires a sufficient amount of historical run-to-failure data, and it cannot be applied to PEMFC, due to the high cost of run-to-failure testing. Authors in [22] presented an ensemble degradation model that integrates two regression models: an exponential model and a polynomial model for RUL prediction of Lithium-ion batteries. The ensemble model is a more accurate parametric model than each of the above two models because it takes into account on-board applications, and global and local regression characteristics. To be effective, ensemble-based systems and algorithms need diversity among the models, which can be achieved in different ways, [26]:

- Different training sets boosting, bagging, which is not the approach followed here, in particular because we work with time series;
- Different types of mathematical processing models;
- Same mathematical processing models but with different set of input features ; this is the case of the work presented in this paper.

Our work proposes an ensemble prognosis approach formed on the basis of different and dependent deterioration data sources, which makes it original when compared to existing ensemble prognosis approaches using rather a single data source processed through different models. More precisely, here, our proposition is to use an ensemble approach to take the best benefit from different measurements gathered at different levels in the monitored system, with a different quality and available with a different periodicity (namely here a stack voltage measurement and a SOH estimation), in order to improve the quality and the performance of the prognosis of the overall PEMFC system.

Particle Filtering (PF), which has been developed recently among applications for PHM [27], is applied so that the RUL uncertainties are also obtained.

The remainder of this paper is organized as follows. In Section 2, the problem formulation and the models for RUL prognosis are presented. Then, the Particle Filtering-based prognostic approach and the criteria for performance evaluation are presented in Section 3. The Ensemble-of-models approach to RUL prediction is explained in Section 4. In Section 5, we pay particular attention to data simulation and the process dependencies, and two different procedures of data generation are explained. Finally, the results of a numerical experimentation are reported in Section 6 showing the good performance of the proposed ensemble approach for RUL prognosis.

2. Problem Formulation

2.1. Fuel Cell Degradation

For fuel cell systems, just like for any other technological system, performance degradation is unavoidable but can be minimized by proper operation and maintenance, based on a comprehensive understanding of degradation mechanisms. PEMFC degrades due to calendar aging, which can occur even under constant optimal conditions, start and stop cycles and inadequate operating conditions such as temperature, pressure and poor water management. In a fuel cell, the aging process reduces the component performance and modifies its material physical properties. Figure 1 shows the evolution of the stack voltage degradation. The decreasing trend represents the irreversible degradation, whereas the voltage jumps represent the reversible behavior caused by operating conditions modification.

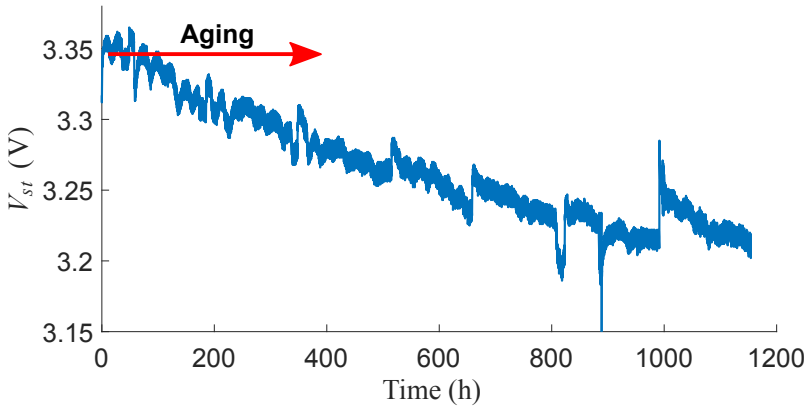

Figure 1: Voltage degradation with aging (under constant current density) [28].

During the lifetime of a PEMFC, its "health" and performance gradually deteriorate, due to irreversible physical and chemical changes, which take place with usage and with aging, until the moment the stack is no longer usable. The State Of Health (SOH) (*e.g.* from its Begin Of Life (BOL) status of full performance to its End Of Life (EOL) status of functional failure, i.e. its performance does not meet the desired operational standard) provides an indication (not an absolute measurement) of the performance which can be expected from the PEMFC in its current condition and of the amount of lifetime already spent by the component.

Any parameter significantly changing with age, such as cell impedance, can be used for indicating the SOH of the cell. These parameter changes are typically identified by performing characterization measurements such as polarization curves. The polarization curve describes the working performance of PEMFC. The variations of internal parameters, including physical and empirical ones, have great impact on the polarization characteristic. Figure 2 shows the variation of the polarization curves under aging. In this work, physical and empirical parameters are used to predict the performance of the fuel cell. We will consider two indicators of the PEMFC degradation: the stack voltage and the SOH. The stack voltage V_{st} can be measured at a high frequency ($\approx 0.6s$) and the SOH are characterized every week in practice.

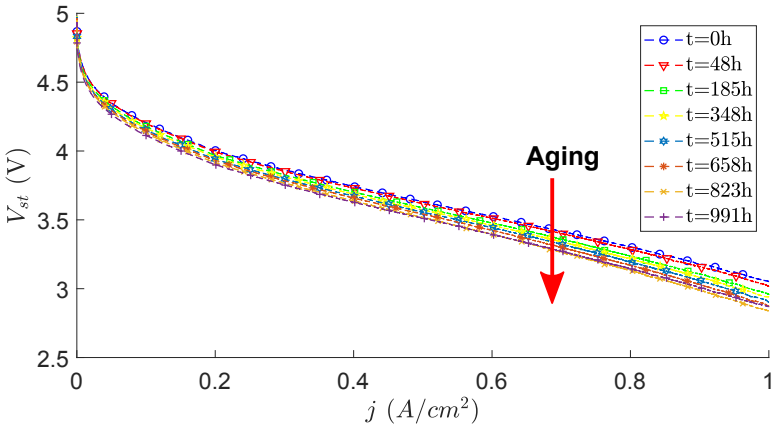

Figure 2: Polarization curves during aging [28].

2.2. Models Description

2.2.1. Voltage Model

Being electrochemical cells, fuel cells obey to thermodynamic and kinetic laws. The static voltage of a fuel cell stack, depicted in Figure 2, is given by [9]:

$$V_{st} = n \cdot (E - V_{ohm} - V_{act} - V_{trans}) \tag{1}$$

where V_{st} is the stack voltage, n is the number of cells in the stack, E is the open circuit voltage (OCV), V_{act} is the activation polarization, V_{ohm} represents the ohmic losses (due to the electrical resistance of individual components and their contact), and V_{trans} is the concentration polarization (due to mass transport limitation). For a stack operating at a current density j [29]:

$$V_{st} = n \cdot \left(E - r \cdot j - A \cdot ln(\frac{j}{j_0}) - m_1 \cdot exp(m_2 \cdot j) \right) \tag{2}$$

where r is the internal resistance, j is the operating current density, A is the Tafel coefficient, j_0 is the exchange current density, m_1 and m_2 are the mass-transfer constants. Considering different current current density values, a static polarization curve is obtained.

2.2.2. SOH Degradation Model

A limitation of the stack voltage is that it does not allow separating the effect of the load variation, which causes current density variations, from that of the stack inner degradation, which influences the OCV [30] and the global resistance parameters [10, 19]. Since physical laws describing the effects of the degradation on E and r are not known, in this work we adapt linear equations for simplicity of illustration and without loss of generality of the proposed approach. The changes in the two parameters are coupled by variable $\gamma(t)$, which reflects the SOH degradation:

$$\begin{aligned} r(t) &= r_0(1 + \gamma(t)) \\ E(t) &= E_0(1 - \gamma(t)) \end{aligned} \tag{3}$$

where r_0 and E_0 are the initial values of r and E. Since it has been proven in [10, 19] that the SOH indicator $\gamma(t)$ can be estimated from polarization curves, in this work we assume the

availability of the procedure which returns the SOH degradation estimation $\gamma(t)$ from characterization measurements of the PEMFC stack. Thus, $\gamma(t)$ can be taken as an input for our prognostic procedure.

2.2.3. Prognostic Models for RUL Prediction

Two stochastic state transition models are used for describing the SOH deterioration $\gamma(t)$ and the stack voltage degradation $V_{st}(t)$.

2.3. Problem Statement

When the fuel cell stack experiences variable operating conditions, a single degradation indicator is not able to provide a precise and robust RUL prediction. The stack voltage does not directly measure the component degradation but it is only related to degradation symptoms, which are significantly affected by operating conditions. The SOH provides aging information but it can only be measured at low frequency in industrial applications.

In this work, we consider prognostics based on two different measurements of the stack degradation:

- An external signal, such as the stack voltage, which is easily accessible and frequently measured, but of "poor quality", *i.e.* its measurement is affected by significant noise.
- A signal which provides an internal characterization of the component, such as the stack SOH, which is seldom measured due to the complexity and cost of the measurement procedure that requires to take the fuel cell stack out of service for the measurements.

The objective is to combine the predicted RUL outcomes based on the two signals.

3. Particle Filtering-based RUL Prognosis

3.1. RUL Prognosis

The RUL is the time remaining from the current moment and the moment when the system is considered failed. As been depicted in Figure 3, degradation measurements are used to train the prognostic model during the learning phase until the prediction time t_λ. Then the learned behavior of this degradation path is used to predict the future evolution with time. The End Of Life (EOL) is the time when the estimated degradation state reaches the failure threshold where the RUL can be computed. The performance of the prognostic model is typically evaluated by comparing the RUL prediction with ground truth RUL (see Appendix 1).

3.2. Particle Filtering

Particle Filtering (PF) relies on state-space description of the system evolution and observation with possibly non-linear and non Gaussian features [31]. It is a recursive state estimation techniques based on a Bayesian approach [32].

For our purposes, the degradation dynamics and its observations are assumed to be governed by a discrete-time state transition model:

$$x_k = f_k(x_{k-1}, \omega_{k-1}, \Theta_{k-1}) \tag{4a}$$

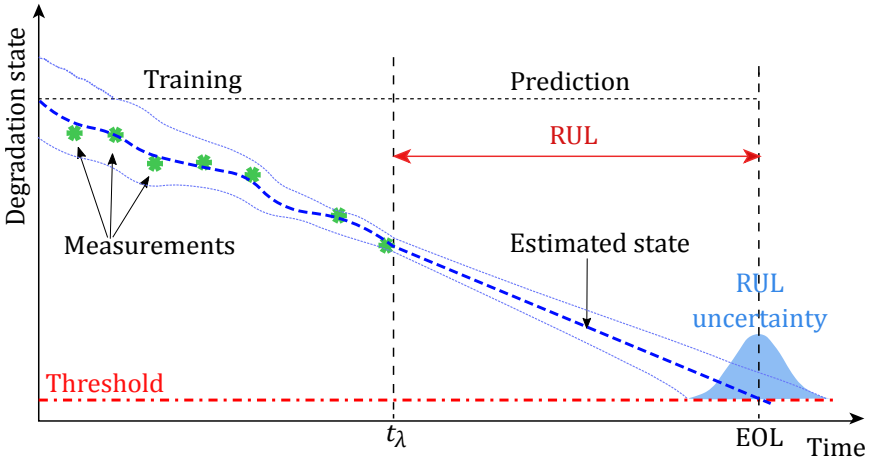

Figure 3: Degradation estimation and RUL prediction. t_λ is the prediction time.

$$z_k = h_k(x_k, \nu_k) \tag{4b}$$

where k is the time index, x is the system state, z is the measurement, f is the degradation model (state transition function), ω is the system noise, Θ is the vector of model parameters ($\Theta = [\theta_1, \theta_2, ...]$), h is the measurement model and ν is the measurement noise. Both the process noise ω_k and the observation noise ν_k are assumed to be sampled from a zero-mean Gaussian distribution, i.e. $\omega_k \sim \mathcal{N}(0, \sigma_{\omega_k}^2)$ and $\nu_k \sim \mathcal{N}(0, \sigma_{\nu_k}^2)$.

The PF algorithm is summarized in Algorithm 1. The approximation of the probability distribution of the system state is based on sampled particles and associated weights. Bayesian updating is processed sequentially by propagating particles carrying probabilistic information on the unknown states and model parameters. The probabilistic model for the particles propagation relies on the state transition model (4a) and the probability distribution of the process noise ω_k:

1. Propagate $i = 1, \ldots, n$ particles representing the system state probability density function (PDF) from x_{k-1} to x_k by the state transition model described in Equation (4a) (*Algorithm 1, line 5*).

2. For each particle, estimate the associated weight by calculating its likelihood given an online measurement z_k (*Algorithm 1, line 6*). This gives the corresponding weight of each particle (assuming the measurement noise $\nu_k \sim \mathcal{N}(0, \sigma_{\nu_k}^2)$ is normally distributed):

$$\mathcal{L}(z_k | x_k^i, \sigma_{\nu_k}^i) = \frac{1}{\sqrt{2\pi}\sigma_{\nu_k}^i} exp[-\frac{1}{2}(\frac{z_k - x_k^i}{\sigma_{\nu_k}^i})^2] \tag{5}$$

3. Perform resampling [33] to remove the particles with small weights relative to a given weight limit and replicated those with large weights (*Algorithm 1, line 10 to 17*).

4. The posterior PDF built using resampling in step (3) is used as the prior for the following iteration.

Algorithm 1 Particle Filtering

Initialize $x_0^i, \sigma_{\omega 0}^i, \sigma_{\nu 0}^i$ and Θ_0^i // drawn from initial uniform distributions
Time step $k = 1$
while $x_k^i > FT$ and $k \leq k_p$
  for $i = 1, \ldots, n$
  // Importance sampling:
    Draw particles $x_k^i \sim p(x_k^i | x_{k-1}^i, \sigma_{\omega k-1}^i, \Theta_{k-1}^i)$ using Equation (4a)
    Assign weight $w_k^i = \mathcal{L}(z_k | x_k^i, \sigma_{\nu k}^i)$ using Equation (5)
  end for
  Normalize weight $w_k^i = w_k^i / \sum_{i=1}^{n} w_k^i$
  Calculate the cumulative sum of normalized weights:
  $\{Q_k^i\}_{i=1}^n = Cumsum\left(\{w_k^i\}_{i=1}^n\right)$
  for $i = 1, \ldots, n$
  // Resampling (Multinomial):
    $j = 1$
    Draw a random value $u^i \sim \mathcal{U}(0, 1]$
    while $Q_k^j < u^i$
      $j = j + 1$
    end while
    Update state $x_k^i = x_k^j$
    Update noises $\sigma_{\omega k}^i = \sigma_{\omega k}^j, \sigma_{\nu k}^i = \sigma_{\nu k}^j$
    Update parameters $\Theta_k^i = \Theta_k^j$
  end for
  $k = k + 1$
end while

The process is performed until no measurement is available (prediction time $t_\lambda = k_p \cdot \Delta t$ reached).

For the RUL prediction, the posterior PDFs of the state and model parameters, given the observation sequence up to time t_λ, are used to estimate the future evolution of the particles. The RUL PDF can be obtained when the particles representing the system state reach the preset failure threshold, as illustrated in Figure 3. The prognostic procedure is summarized in Algorithm 2.

Algorithm 2 RUL prediction

$k = k_p$ // Start from the prediction time
for $i = 1, \ldots, n$ // For each particle
  Use model parameters estimated at time t_λ (from Algorithm 1) : $\Theta_k^i, \sigma_{\omega_k}^i$
  while $x_k^i > FT$
    $k = k + 1$
    Propagate particles $x_k^i = f(x_{k-1}^i, \sigma_{\omega}{}_{k-1}^i, \Theta_{k-1}^i)$ using Equation (4a)
  end while
  Estimate $\widehat{RUL}_k^i = (k - k_p) \cdot \Delta t$
end for

4. Ensemble-based Prognostic Approach

4.1. Prognostic Models for RUL Prediction

For the PF-based estimation stage, the following simplified state models are used for the SOH deterioration $\gamma(t)$ and the stack voltage degradation trend $V_{st}(t)$:

- $\gamma(t)$ represents the SOH degradation. It assumes values from 0 (healthy) to 100% (failed), following a linear model:
$$\gamma(t+1) = c^{(1)}(t) \cdot \gamma(t) \tag{6}$$
where $c^{(1)}(t)$ is the time-dependent SOH degradation model parameter [10].

- $V_{st}(t)$ represents a symptom of the stack degradation, which, according to Equation (6), follows a linear trend:
$$V_{st}(t+1) = c^{(2)}(t) \cdot V_{st}(t) \tag{7}$$
where $c^{(2)}(t)$ is the time-dependent voltage degradation parameter

The two linear degradation models of Equation (6) and (7) are used in two different particle filtering algorithms to provide the RUL predictions $\widehat{RUL}_t^{(m)}$ (m=1, 2), respectively. Model 1 in Equation (6) uses measurements of good quality, but not frequently acquired, whereas Model 2 in Equation (7) uses measurements that are regularly available, whose quality can be poor due to higher measurements noise and lower correlation with the true health states. The objective is, then, to combine the individual estimates $\widehat{RUL}^{(1)}$ and $\widehat{RUL}^{(2)}$, taking into account their "local" qualities.

4.2. Ensemble of Models

Fusing the outputs of an ensemble of diverse prognostic models can improve overall prediction accuracy [34]. Local aggregation dynamically assigns weights to each model according to its local performance, typically evaluated on the available historical patterns [25]. For prognostics, local aggregation requires the computation of the local performances of the individual models on a set of run-to-failure degradation trajectories.

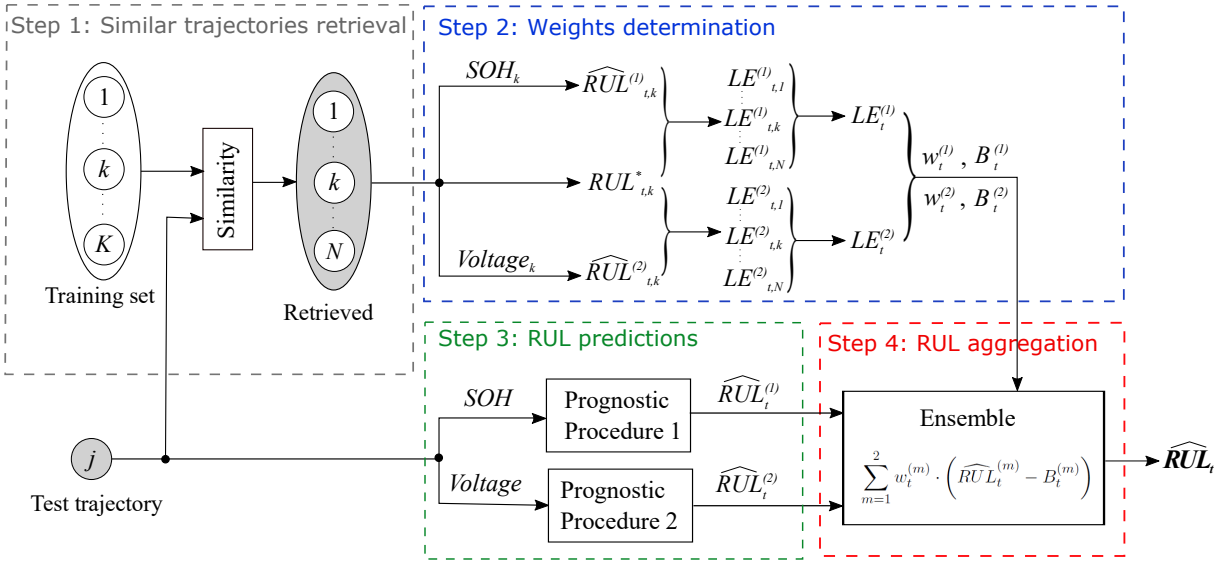

Figure 4: Scheme of the proposed prognostic approach.

Figure 4 presents the scheme of the ensemble-based prognostic approach proposed in this work. As mentioned previously, we assume the availability of the measurements of the signals $x^{(1)} = \gamma$ and $x^{(2)} = V_{st}$ collected during the life of K identical fuel cell stacks:

$$\{x_k^{train}\}_{k=1}^{K} = \{(x_k^{(1),train}, x_k^{(2),train})\}_{k=1}^{K} \tag{8}$$

These run-to-failure trajectories form a training set, which is also used within the ensemble approach for the aggregation of the individual model outcomes. The local fusion approach for the aggregation of the individual model outcomes is based on the following steps, illustrated in Figure 4:

Step 1 : Similar trajectories retrieval - For the test trajectory x^{test}, evaluate the most similar trajectories in the training set by computing the point-wise Euclidean distance on a time window of length L. The lower the Euclidean distance, the greater is the similiarity between the trajectories. For each model m, identify among the training trajectories the most similar to the test trajectory x^{test}, by finding the ones with the minimum Euclidean Distance (ED) [25]:

$$d_t^{(m)} = min\{ED\left(x_{(t-L):t,k}^{(m),train}, x_{(t-L):t}^{(m),test}\right)\}_{k=1}^{K} \tag{9}$$

where $d_t^{(m)}$ is the minimum ED of the test trajectory for the m^{th} model at time t, on a time window L. Then, select the N nearest (*i.e.* most similar) trajectories among the K training trajectories of each measurements for future analysis.

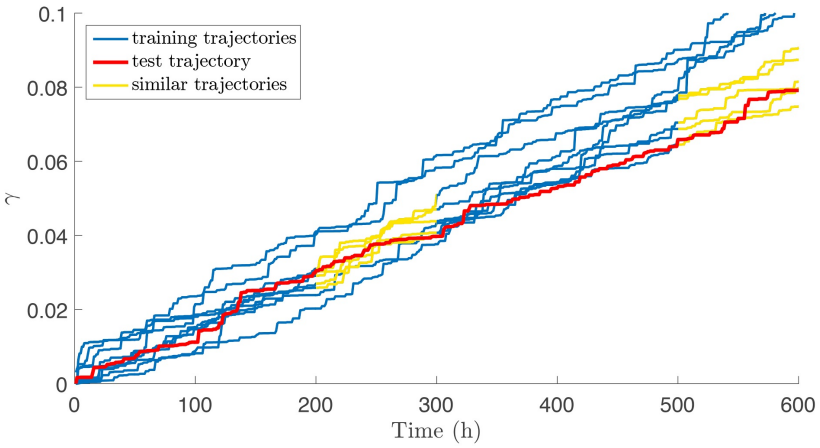

Figure 5: Selection of similar trajectories for RUL prediction assessment.

Figure 5 illustrates the similar training trajectories retrieval for one test trajectory. For example, five nearest trajectories in the training set are selected regarding their ED to the test trajectory, which is computed during a time window of 100 hours. Note that at different prediction time, different trajectories are considered as the nearest ones. Therefore, the similar trajectories are selected locally.

Step 2 : Local weights determination - The local weight assigned to each model of the ensemble is computed based on its local performance in terms of RUL prediction accuracy on the N selected training trajectories [20]. Consider the average local error $LE_t^{(m)}$ on the RUL prediction defined at time step t for the m^{th} model as:

$$LE_t^{(m)} = \frac{1}{N}\sum_{k=1}^{N} |RUL_{t,k}^{*} - \widehat{RUL}_{t,k}^{(m)}| \tag{10}$$

where $RUL_{t,k}^{*}$ is the corresponding ground truth RUL and $\widehat{RUL}_{t,k}^{(m)}$ is the estimated RUL of the k^{th} trajectory predicted by the m^{th} model. The larger the prediction error on the RUL predicted by a model, the poorer the performance of prediction of this model, at least locally (i.e. at the considered time step t). Thus, the local error $LE_t^{(m)}$ provides information about the (poor) performance of the m^{th} model in the reconstruction of the patterns of the training set which are closer to the test pattern. The weight associated to each model in the ensemble should then be a decreasing function of its local error, and we use the inverse of the local error to compute the weight [20]:

$$w_t^{(m)} = \frac{1/LE_t^{(m)}}{\sum_{m=1}^{M}\left(1/LE_t^{(m)}\right)} \tag{11}$$

where M is the number of models (in our case $M = 2$). The local weights $w^{(m)}$ are non-negative and sum to 1. Note that the weights are "local" in the sense that the RUL estimation

$RUL_t^{(m)}$ is evaluated at different time steps dynamically. Before aggregating the RUL predictions with their corresponding weights, a bias correction $B_t^{(m)}$ of the i^{th} model is subtracted:

$$B_t^{(m)} = \frac{1}{N}\sum_{k=1}^{N}\left(RUL_{t,k}^{*} - \widehat{RUL}_{t,k}^{(m)}\right) \tag{12}$$

This quantity represents the accuracy of the RUL predictions obtained by each m^{th} model on the N selected training trajectories. The reason of introducing the bias correction is that at the early prediction stage, due to insufficient available observations, the prognostic algorithm usually provides predictions characterized by large variability. Exploiting the historical data, the average variation can be learned from the training trajectories and used as an offset.

Step 3 : RUL predictions for the different models - Predict the RULs for the test trajectory using the PF method described in Section 3, based on the M models.

Step 4 : Ensemble RUL aggregation - Aggregate RUL predictions based on the individual models and weighted based on prognostic performances:

$$\widehat{RUL}_t = \sum_{m=1}^{M} w_t^{(m)} \cdot \left(\widehat{RUL}_t^{(m)} - B_t^{(m)}\right) \tag{13}$$

where $\widehat{RUL}_t^{(m)}, m = 1, 2, \ldots, M$, is the predicted RUL of the test trajectory x^{test} and $B_t^{(m)}$ is the bias correction evaluated on all N training trajectories.

The ensemble approach allows obtaining the PDF density of the predicted RUL. Various mathematical methods and approaches for combining probability distributions are discussed in [35]. Among them, in this work, we consider the Linear Opinion Pool (LOP), which is a common method for weighted linear combination of the experts probabilities [36, 37, 38] and it is easily understood and calculated:

$$p(\widehat{RUL}_t) = \sum_{m=1}^{M} w_t^{(m)} \cdot p\left(\widehat{RUL}_t^{(m)}\right) \tag{14}$$

where $p(\widehat{RUL}_t)$ represents the merged probability distribution, and $p(\widehat{RUL}_t^{(m)})$, represents the RUL distributions predicted by the M particle filters.

5. Data Generation

Because of the lack of real data available to test and validate the proposed approach, we have to resort to simulated data. Under this situation, it is necessary to clearly show how the chosen procedure to generate the simulated data allows to mimic reasonably the reality and how their generation can be controlled to test different aspects, i.e. here the variability of the considered measurements, and their dependency. Hence, we aim at simulating a realistic evolution of the signals $\gamma(t)$ and $V_{st}(t)$, properly accounting for temporal and stack-to-stack variability, and also the dependence between the two signals. By "realistic", we mean that both signals should be

correlated, but not fully equivalent nor exchangeable with respect to the degradation information they carry. Given the unavailability of real data describing the degradation of a fleet of similar PEMFC stacks, the degradation trajectories are generated by applying the physics-based models of Equation (2) described in Section 2.2. This procedure allows obtaining the SOH and the voltage degradation paths of similar stacks, realistically taking into account their variaibility and dependence by resorting also to stochastic processes, and in particular here to Gamma processes. The simulated degradation trajectories are then divided into a training set made by K trajectories and a test set made by J trajectories.

5.1. Gamma Process

An homogeneous Gamma process is a stochastic process with independent, non-negative increments following a Gamma distribution. If $(X_t)_{t\geq 0}$ is a Gamma process, then:

$$\Delta X_{t-s} = X_t - X_s \sim \mathcal{G}(\alpha.(t-s), \beta) \quad \text{for all} \quad 0 \geq t \geq s \tag{15}$$

where $X_0 = 0$, with probability equal to 1, ΔX_t are independent, $\mathcal{G}(\alpha, \beta)$ denotes the Gamma distribution with shape parameter α and scale parameter β, with the following probability density function

$$f(x; \alpha, \beta) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \tag{16}$$

Over a time interval t, the average degradation rate (slope) is $\overline{x} = \alpha \cdot \beta \cdot t$, the process variance $Var = \alpha \cdot \beta^2 \cdot t$.

The Gamma process is suitable to model gradual damage monotonically accumulating over time in a sequence of tiny increments, such as wear, fatigue, corrosion, crack growth, degrading health index, etc [39]. Thus it is used here for simulating the irreversible degrading SOH of the PEMFC stack. The choice of α and β allows one to set different values for the deterioration average and variance, and hence to model various degradation behaviors, from almost-deterministic to very chaotic. Given the degradation measurements, both parameters can be estimated using classical statistical methods, such as maximum likelihood method, moment method, Bayesian statistics method, *etc.*

Using a stochastic process-based (specifically here a Gamma process-based) degradation models make it possible to take both the temporal variability and the item-to-item variability into account [40]. Several kinds of stochastic processes could have been used ; one advantage of using a Gamma process for degradation modeling is that the required mathematical calculations are relatively straightforward. The RUL can be, thus, obtained in an analytic form if necessary.

5.2. Signal Simulation

5.2.1. SOH Simulation

The degradation path of $\gamma(t)$ is generated by a Gamma process, which accounts for the randomness of the degradation process. The failure threshold FT_γ, here set to the value of 0.15, is obtained by estimating the internal resistance from EIS characterization [8]. Figure 6 shows one simulated degradation path. The average Gamma process $\overline{\gamma}$ classifies the type of fuel cell stack, and the variation from stack to stack is represented by drawing different realizations from $\overline{\gamma}$. The average End Of Life ($\overline{EOL}$) can be found at the time point when $\overline{\gamma}$ crosses the threshold FT_γ:

$$\overline{EOL} = \frac{FT_\gamma}{\alpha \cdot \beta} \tag{17}$$

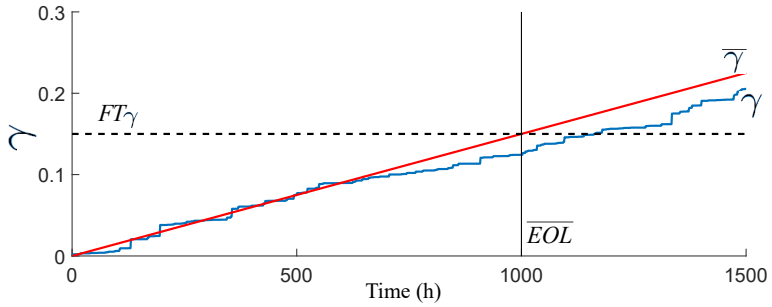

Figure 6: Simulated average SOH degradation $\overline{\gamma}$ and one realization of signal γ representing one stack.

5.2.2. Stack Voltage Simulation

The degradation pattern of PEM fuel cells degradation in stack power is not linear; the decreasing trends are not monotonic. According to Equation (2), the stack voltage is influenced by the loading current density j, which is here simulated by a Markov process [41]. It is used here to simulate the operating conditions during stack usage (Figure 7). In order to be as realistic as possible, the current profile is simulated to its practical value around 0.7mA/cm with 5% random variation in a test bench [28].

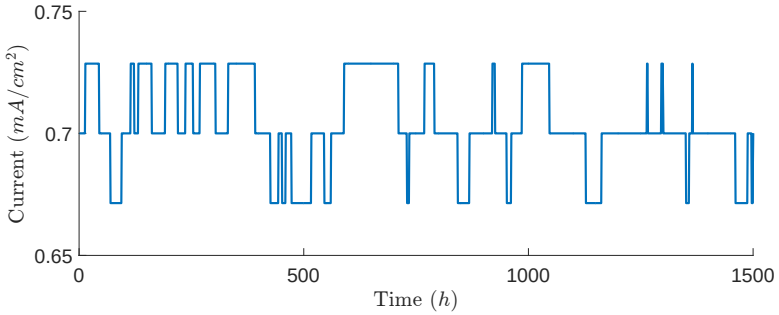

Figure 7: Loading current density j of one stack.

From a given γ, the degradation path of V_{st} is simulated using Equation (2) where the failure threshold $FT_{V_{st}}$ is obtained by substituting $t = \overline{EOL}$ (Figure 8). Note that this failure threshold is deduced from the failure time $\overline{EOL}$.

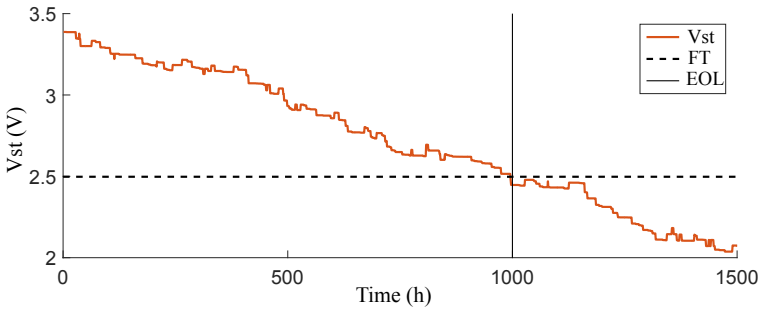

Figure 8: Voltage state V_{st} of one stack.

5.2.3. Observation

The two γ and V_{st} trajectories simulated above are considered as the ground truth. Since measurements revealed by sensors are affected by noises, we randomly sample their values by adding to the ground truth states zero-mean Gaussian noises:

$$\gamma_{meas} = \gamma + \mathcal{N}(0, \sigma_{\gamma}^2) \tag{18}$$

$$V_{stmeas} = V_{st} + \mathcal{N}(0, \sigma_{V_{st}}^2) \tag{19}$$

where γ and V_{st} are the system true states, γ_{meas} and V_{stmeas} are the measurement readings, σ_{γ} and $\sigma_{V_{st}}$ are the standard deviations of those two types of measurements, respectively. Note that $\sigma_{\gamma} < \sigma_{V_{st}}$ given that the SOH measurements γ_{meas} is more precise than the voltage measurements V_{st}.

5.2.4. Data Availability

As mentioned in Section 2.2, the stack voltage can be measured more frequently than the SOH degradation. Thus, the measurement data of SOH degradation are constrained such that they are available only every 100 hours, whereas the measurement data of stack voltage are available every hour. The γ and V_{st} measurements for one single stack are shown in Figure 9. For computational convenience, ten-time steps between two successive measurements are considered for the measurements of V_{st}.

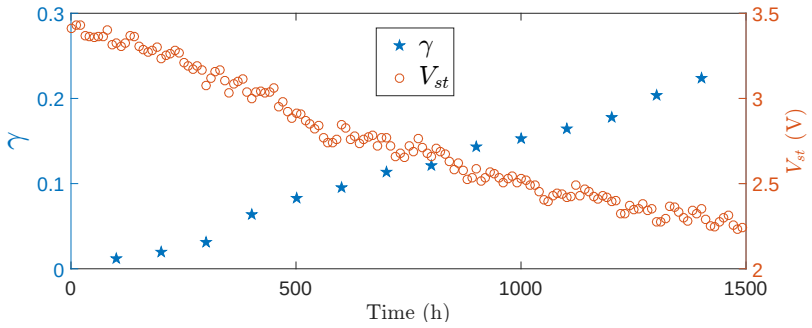

Figure 9: SOH degradation γ and voltage V_{st} measurements of one stack.

5.3. Degradation Simulation Procedure : Introducing Dependence between γ and V_{st}

The algorithm developed in 5.2 allows the simulation of synthetic deterioration data with temporal and item-to-item variability and with measurements uncertainty. However, it does not take into account the dependence between γ and V_{st}. It can be seen from Equations (2) and (3) that $V_{st}(t)$ is a symptom of the degradation $\gamma(t)$, the deterioration levels of the two indicators are correlated. To properly produce realistic simulations of the degradation trajectories that model the different sources of variability, randomness, and dependence between the signals, it is necessary to introduce dependence in the proposed simulation scheme and to control the level of dependence between both deterioration indicators. Two approaches have been proposed to this aim :

- Approach 1 : the two indicators V_{st} and γ are generated from the same realization of a Gamma degradation process, with different additive noises ;
- Approach 2 : the two indicators V_{st} and γ are simulated from two different degradation processes, dependent by construction. To this aim, a bivariate dependent Gamma process is constructed by trivariate reduction in the case of bivariate Gamma random vectors [42].

The technical details of these two approaches and the complete algorithms for synthetic data simulation are given in Appendix 2.

6. RUL Prognosis Results & Performance Evaluation

Considering that in real industrial applications we expect to have available a limited number of PEMFC stacks degradation trajectories. We simulate 100 trajectories of which we use each type of measurement: $K = 50$ for training and $J = 50$ for testing. By performing a sensitivity analysis (See Appendix 3) regarding the prediction accuracy and the computation time, we have set the number of nearest trajectories in the training set to $N = 5$ and the time window for the similarity calculation to $L = 100$ hours.

6.1. RUL Prognosis for Data Simulation Approach 1

The variance of the degradation process, $\alpha\beta^2$, depends on the choice of the Gamma process parameters α and β. It stands for the similarity in degradation behavior of identical PEMFC stacks. As being discussed in the presentation of the simulation procedure (see Appendix 2), the objective of introducing the variance is to represent stack-to-stack variability around the average behavior. Figure 10 shows three examples of degradation paths with different levels of variance: 1) low variance ($\alpha = 0.6, \beta = 2.5e{-}4$); 2) medium variance ($\alpha = 0.1, \beta = 1.5e{-}3$); 3) high variance ($\alpha = 0.03, \beta = 5.0e{-}3$).

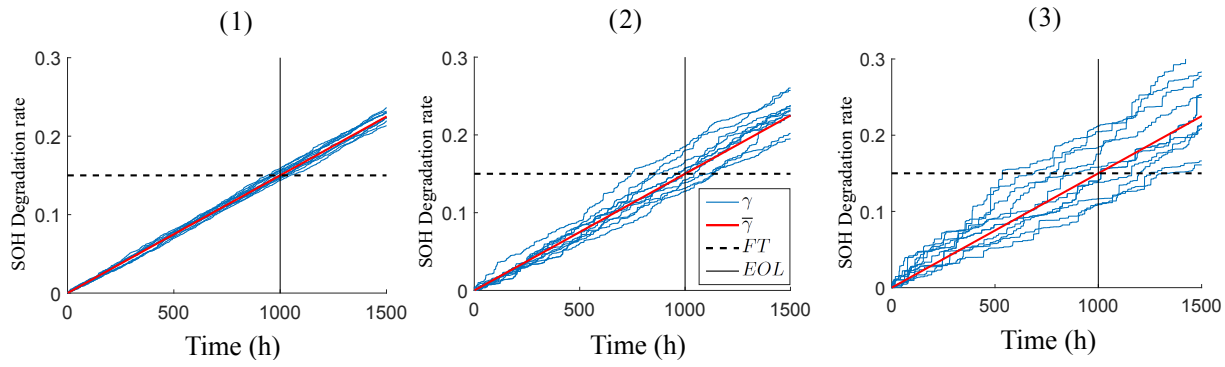

Figure 10: Data simulation with three levels of variance: (1) Low variance; (2) Medium variance; (3) High variance.

The RUL predictions for all the degradation trajectories are carried out by the Particle Filtering-based approach described in Section 3. For each trajectory in the test set ($J = 50$ trajectories), the RUL predictions are made every 100 time steps with Model 1 and every 10 time steps with Model 2. The RUL predictions based on Model 1 are less frequent than the ones based on Model 2, because the measurements that feed Model 2 are intermittently taken. Thus, to have a fair comparison between the two models, the missing predictions of Model 1 are reconstructed by linear

interpolation. The simulation is carried out with the data dependence generation of Approach 1 and Gamma process with medium variance. Figure 11 shows the Local Error (LE) at different prediction time steps obtained for a single test trajectory (№40).

6.1.1. RUL Aggregation

Figure 12 shows the corresponding weights which are dynamically assigned to the two models according to their local error evaluated at each time step. Notice that:

- Model 1 weights are larger at the beginning of the component life compared to that of Model 2. This can be justified by the fact that Model 1 is fed by more precise SOH measurements and it is not influenced by loading current variations.
- Model 2 weights are larger than those of Model 1 after approximately 600 hours. This can be justified by the fact that Model 2 is trained by using more data. Thus, its prediction performance is improved much faster than the one of Model 1, especially near the end of life when Model 1 is no longer updated due to lack of new incoming measurements.

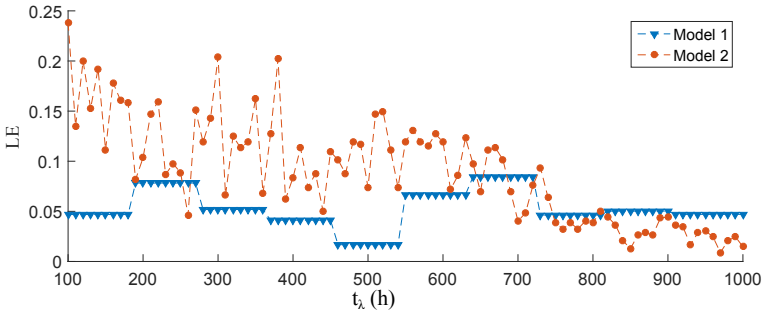

Figure 11: Local error evaluated over 50 training trajectories for test trajectory №40.

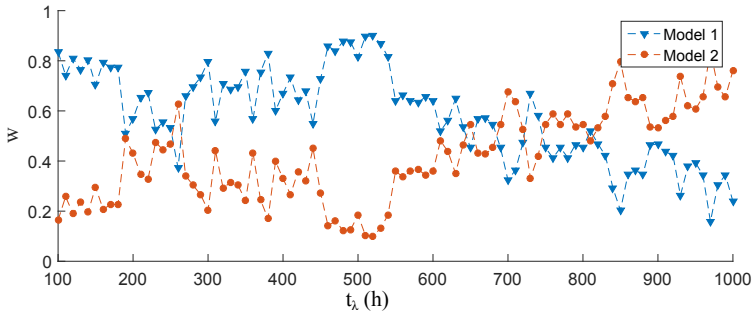

Figure 12: Weight assigned to each model for test trajectory №40.

The RUL predictions based on both models are aggregated according to Equation (13). Figure 13 shows the RUL predictions and the aggregation for one test trajectory. The ensemble RUL predictions take advantage of the complementary behaviors of individual models. Indeed, the analysis of Figure 13 suggests that:

- The predictions provided by the two models are comparable, even if Model 1 [8] provides more accurate RUL predictions at the early life stages of the stack №40, Model 2 provides more accurate predictions when this stack approaches the EOL.
- The ensemble of the two models allows obtaining more accurate predictions throughout the RUL predictions of stack №40 than each individual model.

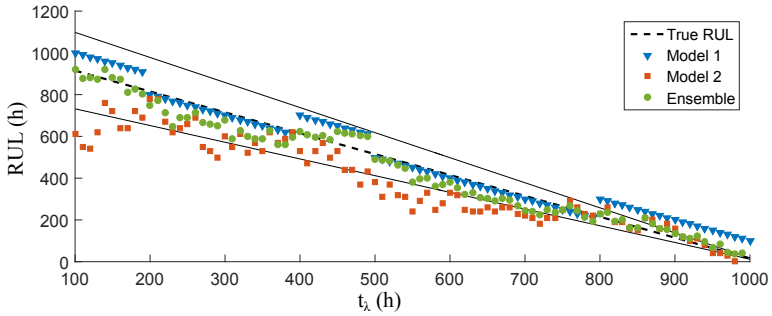

Figure 13: RUL predictions aggregation for trajectory №40.

Figure 14 provides a global view of the average local error for all 50 test trajectories. Since each trajectory (stack) has different EOL, we normalized the time index considering the EOL ratio $\lambda_j = \dfrac{t_\lambda}{EOL_j}$. Globally, Model 1 prediction errors are lower at earlier life stages, whereas Model 2 errors gradually decrease thanks to the updating by sufficient incoming measurements and finally becomes lower.

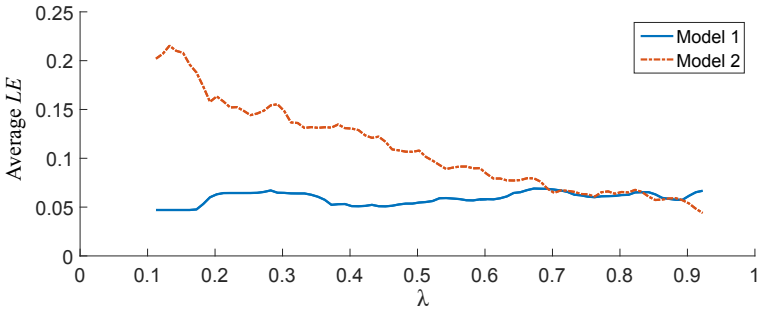

Figure 14: Average local error from 50 test trajectories.

Figure 15 shows the average prediction error for each test trajectory. It is the average local error of each trajectory along the entire prediction horizon (from 100 hours to 1000 hours): we can see that the ensemble gives the smallest prediction error for almost all the test trajectories.

[8] Here "Model 1" stands for "the prognostic approach based on Model 1", "Model 2" for "the prognostic approach based on Model 2", and "Ensemble" for "the prognostic approach based on the ensemble of models". This simplification is to avoid the wordy expression, and is used in the rest of the paper.

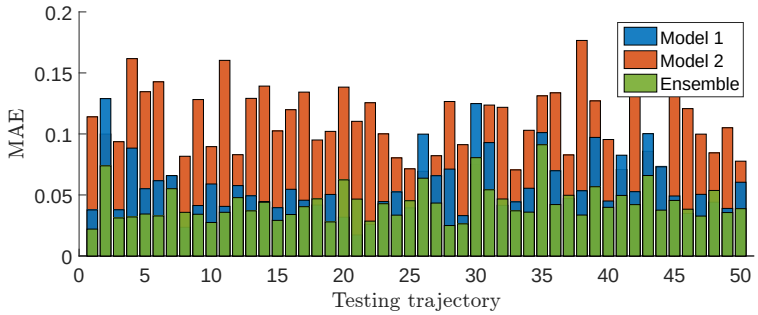

Figure 15: Mean absolute error for 50 test trajectories.

6.1.2. RUL Uncertainty Aggregation

Figure 16 shows the 25^{th} and 75^{th} percentiles of the RUL PDF provided by the Ensemble, which is obtained by merging the RUL PDFs of Model 1 and Model 2 according to Equation (14), for trajectory №40. By aggregating the two PDFs, we obtain not only the RUL but also the uncertainty of the predictions, which is very important for post-prognosis maintenance decision-making. As expected, the prediction becomes closer to the ground truth RUL and the uncertainties (PDFs) of the ensemble become smaller when approaching the end of life. Figure 17 depicts the RUL uncertainty at different life stages: at the early prediction time of 200 hours (Figure 17.1), at half-life of 500 hours (Figure 17.2) and near the EOL of 800 hours (Figure 17.3). All three models become less spread and centered to the true RUL accuracy zone when the prediction time approaches the end of life. The less spread distribution indicates that the RUL predictions become more accurate and more precise when more observations become available, which meets our expectation.

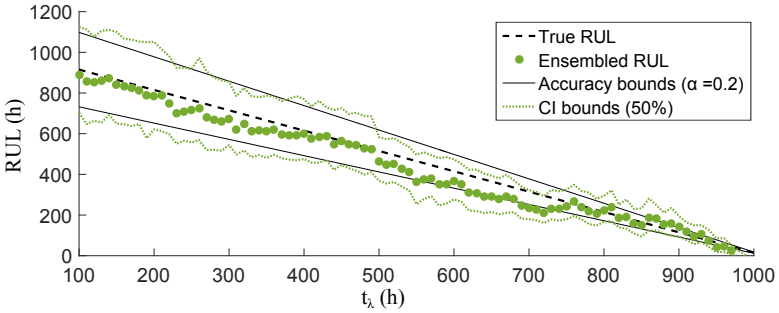

Figure 16: Aggregated RUL predictions with uncertainty for trajectory №40 with accuracy and confidence interval (CI) bounds.

6.1.3. Prognostic Performance Evaluation

The quality of the RUL predictions of the individual models and the ensemble are evaluated using the prognostic performance metrics in Table 1, which reports the average performance over $J = 50$ test trajectories and all t_λ time steps.

The metrics used for performance evaluation in this work consist of the accuracy index Acc, the α-λ accuracy αAc, the steadiness index Std, the risk index Rsk, the precision index Prc and the coverage index Cvg, as explained in Appendix 1. The values in Table 1 suggest that:

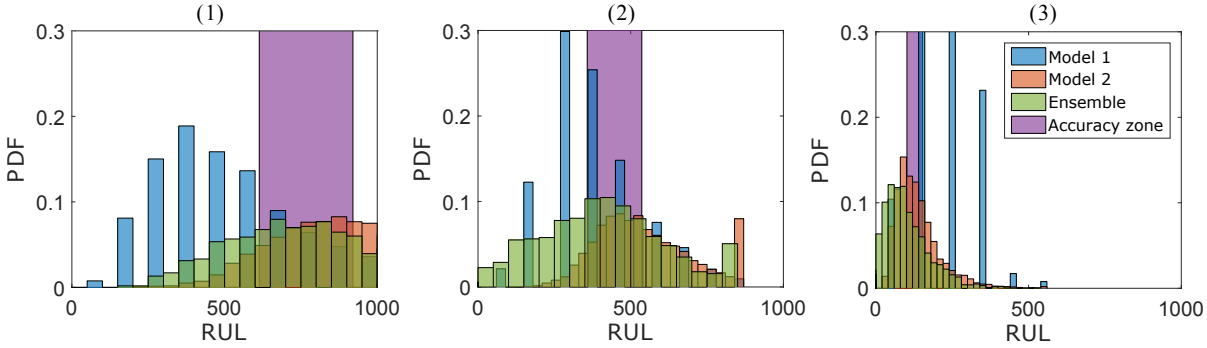

Figure 17: Histogram of aggregated RUL uncertainties for trajectory №40 at different prediction time steps: (1) $t_\lambda = 200$ hours; (2) $t_\lambda = 500$ hours; (3) $t_\lambda = 800$ hours.

Table 1: Prognostic performance metrics (Approach 1, medium variance)

Average Performance	Model 1	Model 2	Ensemble	
			Point	PDF
Acc	0.52	0.12	0.55	**0.56**
αAc	0.31	0.24	**0.54**	0.49
Std	0.16	0.14	**0.07**	**0.07**
Rsk	**0.34**	0.49	0.44	0.36
Prc	**0.25**	0.27	–	0.32
Cvg	0.48	0.37	–	**0.74**

- The Ensemble shows better performance than any individual model with respect to *Acc*, αAc, *Std* and *Cvg* indexes.
- Model 1 shows better performance in *Rsk* index, which means that the RUL predictions based on Model 1 are early notifications. This does not mean that all early predictions are good predictions: an early notification which is too far from the true failure time leads to unnecessary maintenance, which incurs extra cost. The *Rsk* performance needs to be considered jointly to the accuracy indexes (*Acc* and αAc). The *Rsk* of the Ensemble is between Model 1 and Model 2, with respect to both point values and uncertainty.
- The *Prc* index of the Ensemble is the weakest, whereas its *Cvg* index is the strongest. This is due to the fact that the PDFs of the Ensemble merges Models 1 and 2 PDFs. The spread of its distribution is, thus, broader than the individual models, but it provides a larger coverage.

Above all, in this example of Approach 1 with medium variance, we can conclude that the Ensemble-based approach globally provides the best prognostic performance.

6.2. RUL Prognosis for Data Simulation Approach 2

Similarly to what has been done for the data simulation Approach 1, three different levels of process variance are simulated. Furthermore, for each level of variance, seven different levels of processes dependence between $\gamma(t)$ and V_{st} are considered to represent the underneath correlation between the two signals.

6.2.1. Parameters Used for the Simulated Examples

The parameters used for the generation of the simulated examples of dependent Gamma processes are reported in Table 2. The correlation coefficient ρ ($0 \leq \rho \leq \rho_{max} = \frac{min(\alpha_1,\alpha_2)}{\sqrt{\alpha_1 \alpha_2}}$) indicates the dependence level of the two final degradation processes after the trivariate reduction.

Table 2: Parameters used for the simulated examples ($\rho_{max} = 0.9128$)

ρ	α_1	α_2	β	a_1	a_2	a_3
0	0.60	0.50	4.00	0.60	0.50	0
$10\%\rho_{max}$	0.60	0.50	4.00	0.55	0.45	0.05
$25\%\rho_{max}$	0.60	0.50	4.00	0.47	0.38	0.13
$50\%\rho_{max}$	0.60	0.50	4.00	0.35	0.25	0.25
$75\%\rho_{max}$	0.60	0.50	4.00	0.22	0.13	0.38
$90\%\rho_{max}$	0.60	0.50	4.00	0.15	0.05	0.45
ρ_{max}	0.60	0.50	4.00	0.10	0	0.50

6.2.2. Prognostic Performance Evaluation

Figures 18 and 19 show the performance improvements of the Ensemble with respect to the two individual models, considering different dependence scenarios, for point values and uncertainty,

respectively. The improvements of performance metrics are computed in terms of percentage increased in the metrics' values, for example:

$$\begin{aligned} Acc(gain) &= \frac{Acc_{Ensemble} - max\left(Acc_{Model1}, Acc_{Model2}\right)}{max\left(Acc_{Model1}, Acc_{Model2}\right)} \\ Std(gain) &= \frac{min\left(Std_{Model1}, Std_{Model2}\right) - Std_{Ensemble}}{min\left(Std_{Model1}, Std_{Model2}\right)} \end{aligned} \tag{20}$$

The values indicate the improvements of the Ensemble with respect to the best between Model 1 and Model 2. Gains above 0 indicate that the Ensemble performance is more satifactory than that of the individual models.

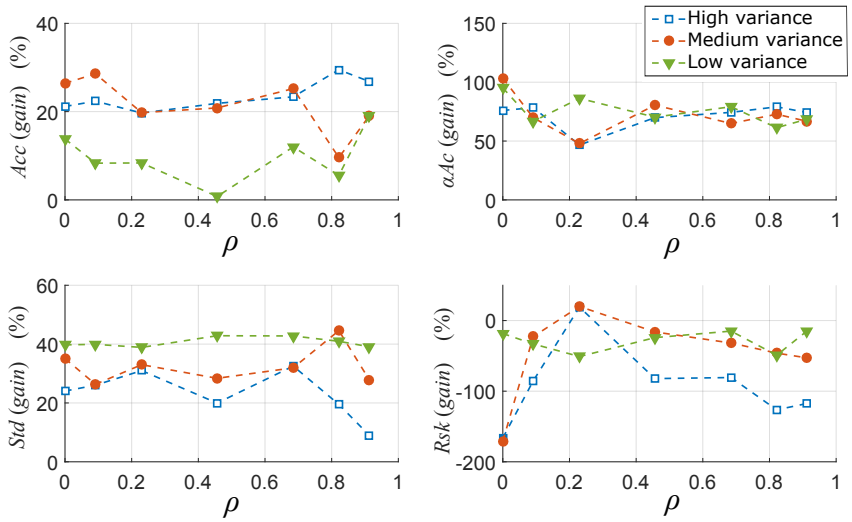

Figure 18: Point values aggregation: prognostic performance gain vs. dependence.

With respect to the point values, notice that:

- Considering the *Acc*, αAc and *Std* metrics, the Ensemble always outperforms any of the individual models. Therefore, we can conclude that the Ensemble is more accurate than Models 1 and 2. The larger the process variance, the larger the gain in performance for the Ensemble approach.
- Similar to the case of data simulation Approach 1, the *Rsk* index of the Ensemble tends to decrease, which means that the Ensemble provides RUL predictions exceeding the ground truth RUL, even though they are located in the accuracy zone.

For the uncertainties aggregation, the analysis of Figure 19 indicates that:

- Considering the *Acc*, αAc and *Std* metrics, the Ensemble model with any process dependence outperforms any of the individual models. The gains of *Std* are nearly the same as the one with point values.
- The *Rsk* index for low variance processes is sometimes improved. It can also be noticed that this index is better than the one with point values aggregation, which indicates that with complete information (uncertainties), some "risky" predictions can be avoided.
- Not surprisingly, the *Prc* index of the Ensemble is the weakest and, on the other hand, its *Cvg* index is the strongest. This can be explained by the fact that the RUL PDF obtained by

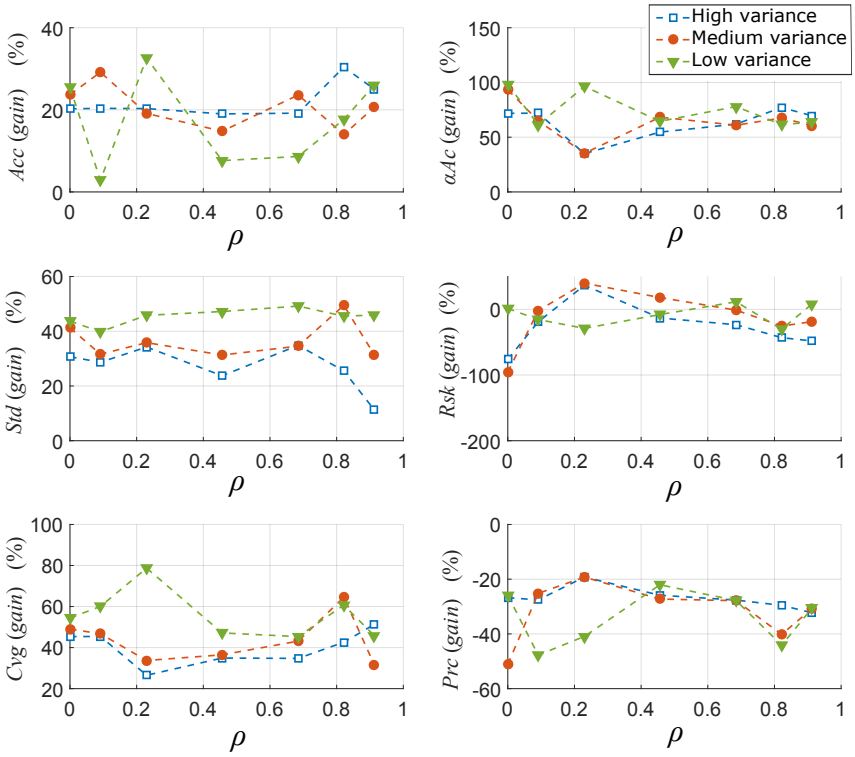

Figure 19: Uncertainties aggregation: prognostic performance gain vs. dependence.

the ensemble approach results from the merging of of both RUL PDFs obtained with Model 1 and Model 2. The spread of the distribution is, thus, broader than those of the individual models, which provides for a larger coverage.

Hence, the ensemble models approach can largely improve the prognostic performance for degradation processes with different variances and different dependencies.

7. Conclusion

In this work, the coexistence of two different sources of information on the degradation of a component, characterized by different levels of accuracy and acquisition rates, has been considered. We developed an ensemble approach that combines the RUL predictions from two different sources of information on the system deterioration, gathered at different levels. The RUL predictions of both models are dynamically aggregated according to their local weights estimated considering the prognostic performance evaluated on a set of historical data. The method has been applied to the prediction of the RUL of simulated PEMFC stack SOH and voltage degradation signals. The results show that the prediction accuracy is improved.

This work can be extended in several directions:

- In our approach, the weights are computed on the basis of offline historical run-to-fail data and a significant amount of historical data can thus be required, which makes it impossible to validate the proposed approach on a real PEFMC test-bench because of the cost of obtaining enough run-to-fail data. To improve the applicability of this approach it can be worthwhile to investigate other solutions for the weights definition, that are less dependent on the quantity of historical data.

- As already explained, ensemble approaches aims at taking advantage of the diversity in the data, processing methods and models. In this work, we have mainly investigated the interest of using deterioration data from different sources, with different characteristics, to improve the prognosis performance. It could be interesting to couple this use of different data sources with different ways of modeling the deterioration evolution (both physics-based and data-driven) and different ways of processing the data using these different models. Such an increased diversity could be a way to devise a more performant ensemble approach for prognosis.

Acknowledgment

This work is partly supported by InnoEnergy and Labex ACTION (ANR-11-LABX-0001-01). We thank Dr. Sameer Al-Dahidi for sharing the pearls of wisdom with us during the course of this research.

References

References

[1] J. Wang, Barriers of scaling-up fuel cells: Cost, durability and reliability, Energy 80 (2015) 509–521. doi:10.1016/j.energy.2014.12.007.

[2] M. Jouin, R. Gouriveau, D. Hissel, M. C. Péra, N. Zerhouni, Prognostics and Health Management of PEMFC - State of the art and remaining challenges, Int. J. Hydrogen Energy 38 (35) (2013) 15307–15317. doi:10.1016/j.ijhydene.2013.09.051.
[3] K. Javed, R. Gouriveau, 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, Mech. Syst. Signal Process. 94 (2017) 214–236. doi:10.1016/j.ymssp.2017.01.050.
[4] J. K. Kimotho, T. Meyer, W. Sextro, PEM fuel cell prognostics using particle filter with model parameter adaptation, in: 2014 Int. Conf. Progn. Heal. Manag., 2014, pp. 1–6. doi:10.1109/ICPHM.2014.7036406.
[5] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, N. Zerhouni, Degradations analysis and aging modeling for health assessment and prognostics of PEMFC, Reliab. Eng. Syst. Saf. 148 (2016) 78–95. doi:10.1016/j.ress.2015.12.003.
[6] D. Zhang, C. Cadet, C. Bérenguer, N. Yousfi-Steiner, Some Improvements of Particle Filtering Based Prognosis for PEM Fuel Cells, IFAC-PapersOnLine 49 (28) (2016) 162–167. doi:10.1016/j.ifacol.2016.11.028.
[7] D. Zhou, F. Gao, E. Breaz, A. Ravey, A. Miraoui, Degradation prediction of PEM fuel cell using a moving window based hybrid prognostic approach, Energy 138 (2017) 1175–1186. doi:10.1016/j.energy.2017.07.096.
[8] T. Kim, H. Kim, J. Ha, K. Kim, J. Youn, J. Jung, B. D. Youn, A degenerated equivalent circuit model and hybrid prediction for state-of-health (SOH) of PEM fuel cell, in: 2014 Int. Conf. Progn. Heal. Manag. PHM 2014, 2014, pp. 1–7. doi:10.1109/ICPHM.2014.7036407.
[9] E. Lechartier, E. Laffly, M.-C. Péra, R. Gouriveau, D. Hissel, N. Zerhouni, Proton exchange membrane fuel cell behavioral model suitable for prognostics, Int. J. Hydrogen Energy 40 (26) (2015) 8384–8397. doi:10.1016/j.ijhydene.2015.04.099.
[10] M. Bressel, M. Hilairet, D. Hissel, B. Ould Bouamama, Remaining useful life prediction and uncertainty quantification of proton exchange membrane fuel cell under variable load, IEEE Trans. Ind. Electron. 63 (4) (2016) 2569–2577. doi:10.1109/TIE.2016.2519328.
[11] M. S. Kan, A. C. Tan, J. Mathew, A review on prognostic techniques for non-stationary and non-linear rotating systems, Mech. Syst. Signal Process. 62 (2015) 1–20. doi:10.1016/j.ymssp.2015.02.016.
[12] T. Sutharssan, D. Montalvao, Y. K. Chen, W.-C. Wang, C. Pisac, H. Elemara, A review on prognostics and health monitoring of proton exchange membrane fuel cell, Renew. Sustain. Energy Rev. 75 (2017) 440–450. doi:10.1016/j.rser.2016.11.009.
[13] M. Tahan, E. Tsoutsanis, M. Muhammad, Z. A. Abdul Karim, Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review, Appl. Energy 198 (2017) 122–144. doi:10.1016/j.apenergy.2017.04.048.
[14] K. Javed, R. Gouriveau, N. Zerhouni, D. Hissel, Prognostics of Proton Exchange Membrane Fuel Cells stack using an ensemble of constraints based connectionist networks, J. Power Sources 324 (2016) 745–757. doi:10.1016/j.jpowsour.2016.05.092.
[15] X. S. Si, W. Wang, C. H. Hu, D. H. Zhou, Remaining useful life estimation - A review on the statistical data driven approaches, Eur. J. Oper. Res. 213 (1) (2011) 1–14. doi:10.1016/j.ejor.2010.11.018.
[16] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, W. Wang, Prognostics and health management: A review on data driven approaches, Math. Probl. Eng. 2015 (2015) 1–17. doi:10.1155/2015/793161.
[17] L. Liao, F. Köttig, Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction, IEEE Trans. Reliab. 63 (1) (2014) 191–207. doi:10.1109/TR.2014.2299152.
[18] L. Liao, F. Köttig, A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction, Appl. Soft Comput. 44 (2016) 191–199. doi:10.1016/j.asoc.2016.03.013.
[19] M. S. Jha, G. Dauphin-Tanguy, B. Ould-Bouamama, Particle filter based hybrid prognostics for health monitoring of uncertain systems in bond graph framework, Mech. Syst. Signal Process. 75 (2016) 301–329. doi:10.1016/j.ymssp.2016.01.010.
[20] P. Baraldi, A. Cammi, F. Mangili, E. E. Zio, Local fusion of an ensemble of models for the reconstruction of faulty signals, IEEE Trans. Nucl. Sci. 57 (2) (2010) 793–806. doi:10.1109/TNS.2010.2042968.
[21] P. Baraldi, M. Compare, S. Sauco, E. Zio, Ensemble neural network-based particle filtering for prognostics, Mech. Syst. Signal Process. 41 (1-2) (2013) 288–300. doi:10.1016/j.ymssp.2013.07.010.
[22] Y. Xing, E. W. M. Ma, K. L. Tsui, M. Pecht, An ensemble model for predicting the remaining useful performance of lithium-ion batteries, Microelectron. Reliab. 53 (6) (2013) 811 – 820. doi:10.1016/j.microrel.2012.12.003.
[23] S. Al-Dahidi, F. Di Maio, P. Baraldi, E. Zio, Remaining useful life estimation in heterogeneous fleets working under variable operating conditions, Reliab. Eng. Syst. Saf. 156 (2016) 109–124. doi:10.1016/j.ress.2016.07.019.

[24] S. Al-Dahidi, F. Di Maio, P. Baraldi, E. Zio, A locally adaptive ensemble approach for data-driven prognostics of heterogeneous fleets, Proc. Inst. Mech. Eng. Part O J. Risk Reliab. 231 (4) (2017) 1–14. doi:10.1177/1748006X17693519.
[25] M. Rigamonti, P. Baraldi, E. Zio, I. Roychoudhury, K. Goebel, S. Poll, Ensemble of optimized echo state networks for remaining useful life prediction, Neurocomputing 0 (2017) 1–18. doi:10.1016/j.neucom.2017.11.062.
[26] R. Polikar, Ensemble Based Systems in Decision Making, IEEE Circuits and Systems Magazine (2006) 21–45.
[27] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, N. Zerhouni, Particle filter-based prognostics: Review, discussion and perspectives, Mech. Syst. Signal Process. 72-73 (2016) 2–31. doi:10.1016/j.ymssp.2015.11.008.
[28] R. Gouriveau, M. Hilairet, D. Hissel, S. Jemeï, M. Jouin, E. Lechartier, S. Morando, E. Pahon, M.-C. Péra, N. Zerhouni, IEEE PHM 2014 Data Challenge - Outline, Experiments, Scoring of results, Winners (2014). URL http://eng.fclab.fr/ieee-phm-2014-data-challenge/
[29] J. Larminie, A. Dicks, Fuel cell systems explained, in: Fuel Cell Syst. Explain., 2nd Edition, John Wiley & Sons Ltd, Chichester, 2003, pp. 45–66. doi:10.1002/9781118878330.
[30] D. Zhang, C. Cadet, N. Yousfi-Steiner, F. Druart, C. Bérenguer, PHM-oriented Degradation Indicators for Batteries and Fuel Cells, Fuel Cells 17 (2) (2017) 268–276. doi:10.1002/fuce.201600075.
[31] M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process. 50 (2) (2002) 174–188. doi:10.1109/78.978374.
[32] S. C. Patwardhan, S. Narasimhan, P. Jagadeesan, B. Gopaluni, S. L. Shah, Nonlinear Bayesian state estimation: A review of recent developments, Control Eng. Pract. 20 (10) (2012) 933–953. doi:10.1016/j.conengprac.2012.04.003.
[33] T. Li, M. Bolic, P. M. Djuric, Resampling Methods for Particle Filtering (2015). doi:10.1109/MSP.2014.2330626.
[34] P. P. Bonissone, F. Xue, R. Subbu, Fast meta-models for local fusion of multiple predictive models, Appl. Soft Comput. J. 11 (2011) 1529–1539. doi:10.1016/j.asoc.2008.03.006.
[35] Robert T. Clemen, R. L. Winkler, Aggregating Probability Distributions, in: W. Edwards, J. Ralph F. Miles, D. von Winterfeldt (Eds.), Adv. Decis. Anal., 1st Edition, Cambridge University Press, New York, 2007, Ch. 9, pp. 172–194.
[36] L. Berliner, J. Brynjarsdottir, A framework for multi-model ensembling, SIAM-ASA Journal on Uncertainty Quantification 4 (1) (2016) 902–923. doi:10.1137/130928339.
[37] D. Bolger, B. Houlding, Reliability updating in linear opinion pooling for multiple decision makers, Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 230 (3) (2016) 309–322. doi:10.1177/1748006X16631866.
[38] D. Bolger, B. Houlding, Deriving the probability of a linear opinion pooling method being superior to a set of alternatives, Reliability Engineering and System Safety 158 (October 2016) (2017) 41–49. doi:10.1016/j.ress.2016.10.008.
[39] J. M. van Noortwijk, A survey of the application of gamma processes in maintenance, Reliab. Eng. Syst. Saf.doi:10.1016/j.ress.2007.03.019.
[40] M. D. Pandey, X. X. Yuan, J. M. van Noortwijk, The influence of temporal uncertainty of deterioration on life-cycle management of structures, Struct. Infrastruct. Eng. 5 (September 2014) (2009) 145–156. doi:10.1080/15732470601012154.
[41] D. Gamerman, H. F. Lopes, Markov Chains, in: Markov Chain Monte Carlo-Stochastic Simul. Bayesian Inference, 2nd Edition, Chapman & Hall/CRC, Boca Raton, 2006, Ch. 4, pp. 113–136.
[42] L. Devroye, Complexity questions in non-uniform random variate generation, in: Non- Uni form Random Variate Gener., Springer Verlag, New York, 1986, pp. 586–587.
[43] A. Saxena, J. Celaya, B. Saha, S. Saha, K. Goebel, Evaluating prognostics performance for algorithms incorporating uncertainty estimates, in: IEEE Aerosp. Conf. Proc., 2010, pp. 1–11. doi:10.1109/AERO.2010.5446828.
[44] A. Saxena, J. Celaya, B. Saha, S. Saha, K. Goebel, Metrics for Offline Evaluation of Prognostic Performance, Int. J. Progn. Heal. Manag. 1 (1) (2010) 1–20.
[45] A. Saxena, J. Celaya, I. Roychoudhury, S. Saha, B. Saha, K. Goebel, Designing data-driven battery prognostic approaches for variable loading profiles: Some lessons learned, Eur. Conf. Progn. Heal. Manag. Soc (2012) 1–11.
[46] Y. Hu, P. Baraldi, F. Di Maio, E. Zio, Online Performance Assessment Method for a Model-Based Prognostic Approach, IEEE Trans. Reliab. 65 (2) (2016) 718–735. doi:10.1109/TR.2015.2500681.
[47] M. Rigamonti, P. Baraldi, E. Zio, D. Astigarraga, A. Galarza, Particle Filter-Based Prognostics for an Electrolytic Capacitor Working in Variable Operating Conditions, IEEE Trans. Power Electron. 31 (2) (2016) 1567–1575. doi:10.1109/TPEL.2015.2418198.
[48] W. Kahle, S. Mercier, C. Paroissin, Gamma Processes, in: Degrad. Process. Reliab., 1st Edition, Wiley-ISTE,

2016, pp. 238–243. doi:10.1002/9781119307488.ch2.
[49] S. Mercier, H. H. Pham, A preventive maintenance policy for a continuously monitored system with correlated wear indicators, Eur. J. Oper. Res. 222 (2) (2012) 263–272. doi:10.1016/j.ejor.2012.05.011.
[50] H. H. Pham, S. Mercier, An imperfect replacement policy for a periodically tested system with two dependent wear indicators, in: Safety, Reliab. Risk Anal. Beyond Horiz., CRC Press, Amsterdam, 2013, pp. 1033–1041.

Appendix 1. Prognostic Metrics

In order to evaluate the average performance of RUL predictions, the common way is to apply several RUL predictions at different time steps to obtain a sequence of predicted RULs [43, 44]. To evaluate the quality of prognostic outcomes, a synthesis of the prognostic metrics is used [21, 45, 46, 47, 24].

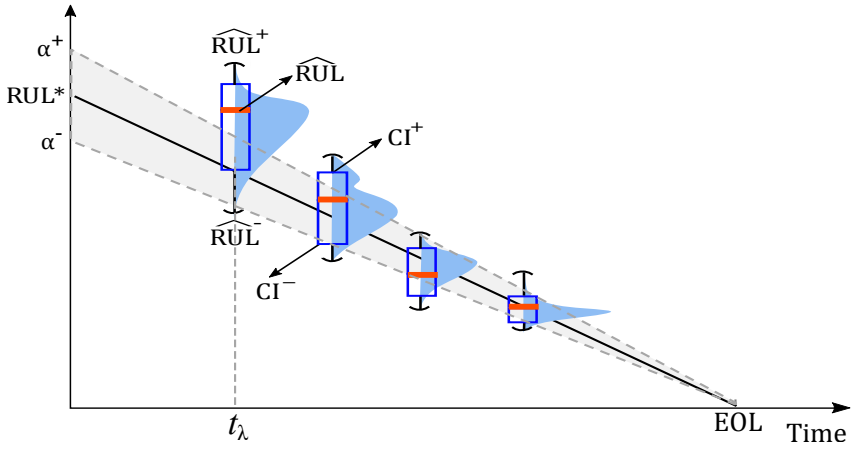

Figure 20: RUL predictions at different prediction times t_λ.

Figure 20 illustrates the RUL predictions with uncertainties at different prediction time steps t_λ. The uncertainties are represented by the Probability Density Function (PDF). The accuracy bounds of a width of 2α shrinks with the prediction time index t_λ, which creates the α-λ accuracy zone covering the true residual life RUL^*. The upper bounds and the lower bounds of the α-λ accuracy zone:

$$\begin{aligned} \alpha^+ &= RUL_t^* \cdot (1+\alpha) \\ \alpha^- &= RUL_t^* \cdot (1-\alpha) \end{aligned} \tag{21}$$

$\widehat{RUL}^+$ and $\widehat{RUL}^-$ are the upper and lower bounds of the predicted RUL uncertainties, whereas CI^+ and CI^- are the bounds of the confidence interval.

Based on those characteristics, different metrics are described as the follows:

- The accuracy index Acc_t directly reflects the prediction errors relative to the true RUL:

$$Acc_t = 1 - \frac{|RUL_t^* - \widehat{RUL}_t|}{RUL_t^*} \tag{22}$$

  where RUL_t^* the true RUL and $\widehat{RUL}_t$ the median value of predicted RULs at prediction time t_t. A larger value of Acc_t indicates a better accuracy.

- The α-λ metric considers whether the predicted $\widehat{RUL}$ lies within the $\pm\alpha$ interval stating whether the required accuracy is met at a given time t_λ. As being illustrated in Figure 20, the probability of lying within the α-λ accuracy zone is described by Equation (23):

$$\alpha Ac_t = p\left(\alpha_t^- \leq \widehat{RUL}_t \leq \alpha_t^+\right) \tag{23}$$

where α_t^+ and α_t^- are the upper and lower bounds of the accuracy zone. A higher value represents a better performance.

- The coverage index Cvg_t considers whether the true RUL lies within the RUL prediction interval at time index λ for each trajectory:

$$Cvg_t = p\left(\widehat{RUL}_t^{CI-} \leq RUL_t^* \leq \widehat{RUL}_t^{CI+}\right) \tag{24}$$

The value of Cvg close to 80% indicates a good representation of the uncertainty [21].

- The precision index Prc_t computes the relative width of the prediction interval, which is defined by:

$$Prc_t = \frac{\widehat{RUL}_t^{CI+} - \widehat{RUL}_t^{CI-}}{RUL_t^*} \tag{25}$$

where $\widehat{RUL}_t^{CI+}$ and $\widehat{RUL}_t^{CI-}$ are the upper and lower bounds of the Confidence Interval (CI) of the predicted RULs distribution (*e.g.* CI = 50%) while RUL_t^* is the corresponding true RUL. Smaller values of Prc_t indicate more precise predictions.

- The steadiness index Std_t measures the variance of the estimated value of the End of Life (EOL) when new measurements become available. It is defined as:

$$Std_t = \frac{\sqrt{var(\widehat{EOL}_{(t-L):t})}}{EOL^*} \tag{26}$$

where L is the length of a sliding time window filtering the variances of the predicted EOL. Smaller values of Std_t indicate better performance.

- The risk index Rsk_t is the probability of obtaining an estimated RUL larger than the true RUL:

$$Rsk_t = p(\widehat{RUL}_t > RUL_t^*) \tag{27}$$

This index indicates the probability of receiving a later notification of a failure such that scheduling a maintenance after the failure is risky. Lower values correspond to a lower risk, which means a better performance.

Appendix 2. Introducing and controlling dependence in simulated deterioration trajectories

This appendix gives technical details on the simulation of dependent trajectories of the two deterioration indicators γ and V_{st}. Two approaches have been considered in this work.

Approach 1

The two indicators V_{st} and γ are generated from the same realization of a Gamma degradation process, with different additive noises, which introduces some dependence between both indicators. Different parameters α^i and β_i are used for the i^{th} stack. The objective of the simulation procedure is also to represent stack-to-stack variability around the average behavior given by:

$$\overline{\gamma}(t) = \overline{\gamma}(t-1) + \Delta\overline{\gamma}(t) \tag{28}$$

where $\Delta\overline{\gamma}(t)$ follows a Gamma law $\mathcal{G}(\overline{\alpha}\Delta t, \overline{\beta})$. The average values of the Gamma process parameters $\overline{\alpha}$ and $\overline{\beta}$ are preset according to our knowledge of PEMFC stack degradation in the following way. The failure threshold FT_γ (degradation rate) is set to 0.15, the average End of Life $\overline{EOL}$ is set to 1000 hours, and the slope of the degradation path is fixed to the computed value of $\overline{\alpha} \cdot \overline{\beta} = \frac{FT_\gamma}{\overline{EOL}}$, whereas the degradation variance is the value of $\overline{\alpha} \cdot \overline{\beta}^2$. Thus, for the i^{th} PEMFC stack:

$$\gamma^i(t) = \gamma^i(t-1) + \Delta\gamma^i(t) \tag{29}$$

where $\Delta\gamma^i(t)$ follows a Gamma law $\mathcal{G}(\alpha^i\Delta t, \beta^i)$, and α^i (resp. β^i) is drawn from a normal distribution around $\overline{\alpha}$ (resp. $\overline{\beta}$) with 5% variation. The measurements data are simulated according to Algorithm 3.

Algorithm 3 Data simulation Approach 1

1: Choose $FT_\gamma, \overline{EOL}, \overline{\alpha}, \overline{\beta}$
2: **for** i = 1 : number of simulated stacks
3: Draw α^i, β^i, from normal distributions with average values $\overline{\alpha}$ and $\overline{\beta}$
4: Generate realization $\gamma^i(t)$ of a Gamma process with parameters (α^i, β^i)
5: Add noises to $\gamma^i(t)$ to obtain the SOH degradation indexes for building the signals for Model 1 and Model 2:
$\gamma_1^i(t) = \gamma^i(t) + \varepsilon_1 \quad$ where $\quad \varepsilon_1 \sim \mathcal{N}(0, \sigma_1^2(t))$
$\gamma_2^i(t) = \gamma^i(t) + \varepsilon_2 \quad$ where $\quad \varepsilon_2 \sim \mathcal{N}(0, \sigma_2^2(t))$
6: Generate V_{st}^i index via Equations (2) and (3) using $\gamma_2^i(t)$
7: Add measurement noises:
$\gamma_{meas}^i(t) = \gamma_1^i(t) + \varepsilon_{meas,1} \quad$ where $\quad \varepsilon_{meas,1} \sim \mathcal{N}(0, \sigma_{meas,1}^2(t))$
$V_{st}{}_{meas}^i(t) = V_{st}^i(t) + \varepsilon_{meas,2} \quad$ where $\quad \varepsilon_{meas,2} \sim \mathcal{N}(0, \sigma_{meas,2}^2(t))$
8: **end for**

Approach 2

The two indicators V_{st} and γ are simulated from two different degradation processes, dependent by construction. To this aim, a bivariate dependent Gamma process is constructed by trivariate reduction in the case of bivariate Gamma random vectors [42].

Let us first recall that an univariate Gamma process [48] with parameters (α, β) (where α, $\beta > 0$) is a subordinator such that for every $t \geq 0$, the random variable $G(t)$ is Gamma-distributed $(\alpha t, \beta)$ with probability density function :

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad for \quad x, \alpha, \beta > 0 \tag{30}$$

The principle of this second approach is to generate a bivariate Gamma process $\boldsymbol{\gamma}_t$ consisting of two dependent deterioration process $\gamma_{1,t}$ and $\gamma_{2,t}$. $\gamma_{1,t}$ is then used directly as the SOH deterioration index measurements, and $\gamma_{2,t}$ to generate the stack voltage measurement V_{st}, see Algorithm 4.

Starting from three independent univariate Gamma processes g_t^j with (a_j, b_j) for $j = 1, 2, 3$, one can build two dependent Gamma processes (or a bivariate dependent Gamma process) by trivariate reduction:

$$\begin{aligned} \gamma_{1,t} &= g_{1,t} + g_{3,t} \\ \gamma_{2,t} &= g_{2,t} + g_{3,t} \end{aligned} \tag{31}$$

The process $\boldsymbol{\gamma}_t = (G_{1,t}, G_{2,t})$ is, then, a bivariate subordinator [49] with Gamma marginal processes and marginal parameters (α_j, β_j) where $\alpha_j = a_j + a_3$ for $j = 1, 2$. The linear correlation between the two random variables $G_{1,t}$ and $G_{2,t}$ is independent of time t and described by the Pearson's correlation coefficient [49, 50]:

$$\rho = \frac{a_3}{\sqrt{\alpha_1 \alpha_2}} \tag{32}$$

where ρ is the Pearson's correlation coefficient, α_1 and α_2 are the marginal gamma parameters. Consequently, we have the following link between the two parametrizations (a_1, a_2, a_3) and $(\alpha_1, \alpha_2, \rho)$:

$$\begin{aligned} a_1 &= \alpha_1 - \rho\sqrt{\alpha_1 \alpha_2} \\ a_2 &= \alpha_2 - \rho\sqrt{\alpha_1 \alpha_2} \\ a_3 &= \rho\sqrt{\alpha_1 \alpha_2} \end{aligned} \tag{33}$$

where $0 \leq \rho \leq \frac{min(\alpha_1, \alpha_2)}{\sqrt{\alpha_1 \alpha_2}}$.

This link allows to choose a_1, a_2 and a_3 so as to generate a bivariate Gamma process with desired α_1, α_2 and ρ. Within the range $0 \leq \rho \leq \frac{min(\alpha_1, \alpha_2)}{\sqrt{\alpha_1 \alpha_2}}$, trivariate reduction leads to one of the fastest algorithms known to date for bivariate Gamma distributions [42]. As in Approach 1, the desired parameters α_1, α_2, β are determined to be consistent with the available knowledge of PEMFC stack degradation : the failure threshold FT_γ (on the SOH level) and the average End of Life $\overline{EOL}$.

Appendix 3. Data Size Analysis

This appendix proposed the results of a sensitivity analysis applied to decide for the data size for the experiment presented in this work. Two parameters need to be taken into account: 1) the number of training trajectories K, and 2) the number of selected nearest neighbor trajectories N. The following analysis are carried out under data generation with medium variance ($\alpha = 0.1, \beta = 1.5e{-}3$) and dependency $\rho = 0.5\rho_{max}$.

Figure 21 shows the accuracy surface of the ensemble with different training size K from 2 to 50, and neighbor size N from 1 to 15. The optimal values are not strict. We can obtain best accuracy outcomes with a larger number of training size and a larger number of nearest neighbors.

During the simulation, it is noticed that the number of the nearest neighbors N strongly affects the simulation time (changes exponentially). Thus, we would prefer smaller N while maintaining a good accuracy.

Algorithm 4 Data simulation Approach 2

Choose $FT_\gamma, \overline{EOL}$
Determine α_1, α_2 and β for marginal Gamma distributions
Given ρ, $0 \leq \rho \leq \frac{min(\alpha_1,\alpha_2)}{\sqrt{\alpha_1\alpha_2}}$
for i=1 : number of simulated stacks
    Generate the realizations of Gamma processes with parameters
    $a_1 = \alpha_1 - \rho\sqrt{\alpha_1\alpha_2}$, $a_2 = \alpha_2 - \rho\sqrt{\alpha_1\alpha_2}$, $a_3 = \rho\sqrt{\alpha_1\alpha_2}$, $b = \beta$:
    $g_1^i(t)$: Gamma process realization with parameters (a_1, b)
    $g_2^i(t)$: Gamma process realization with parameters (a_2, b)
    $g_3^i(t)$: Gamma process realization with parameters (a_3, b)
    Generate dependent SOH indexes by trivariate reduction:
    $\gamma_1^i(t) = g_1^i(t) + g_3^i(t)$
    $\gamma_2^i(t) = g_2^i(t) + g_3^i(t)$
    Generate V_{st}^i index via Equations (2) and (3) using $\gamma_2^i(t)$
    Add measurement noises:
    $\gamma_{meas}^i(t) = \gamma_1^i(t) + \varepsilon_{meas,1}$ where $\varepsilon_{meas,1} \sim \mathcal{N}(0, \sigma_{meas,1}^2(t))$
    ${V_{st}}_{meas}^i(t) = V_{st}^i(t) + \varepsilon_{meas,2}$ where $\varepsilon_{meas,2} \sim \mathcal{N}(0, \sigma_{meas,2}^2(t))$
end for

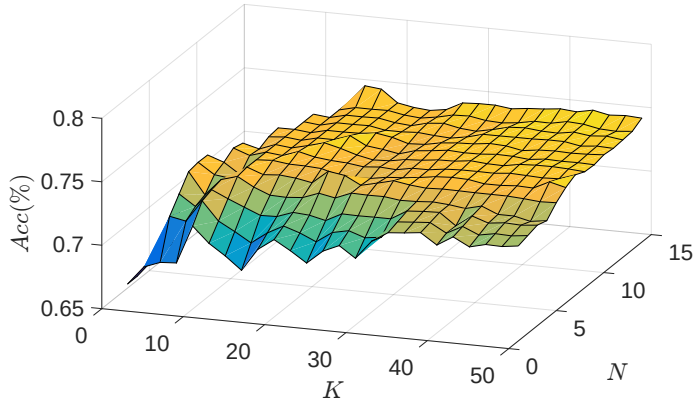

Figure 21: Accuracy surface with different training size K and neighbor size N.

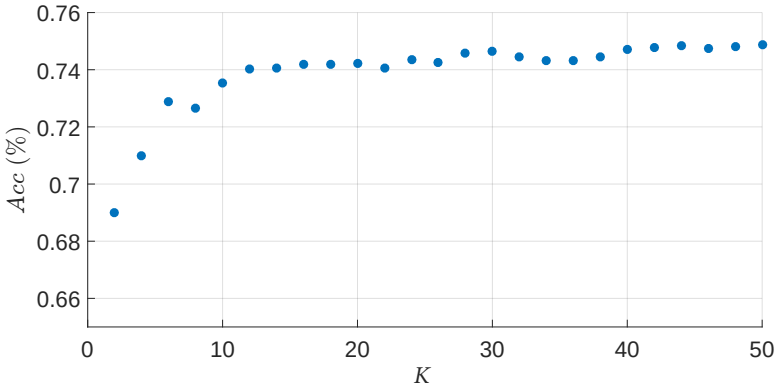

Figure 22: Accuracy Acc vs number of training trajectories K (with $N = 5$).

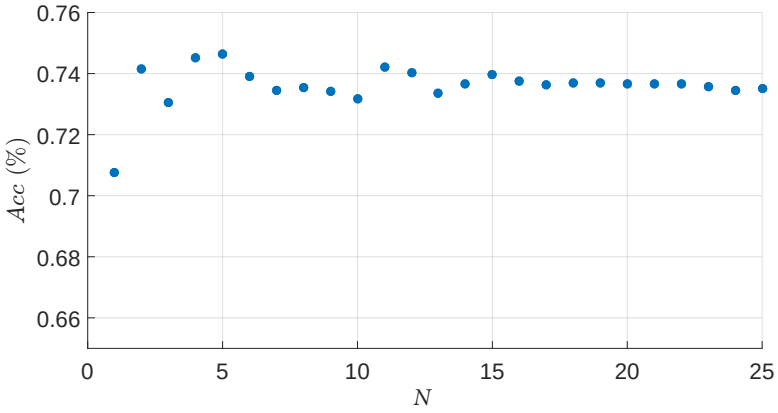

Figure 23: Accuracy Acc vs number of nearest neighbors N (with $K = 50$).

Figure 22 shows the impact of training data size K on the prognostic accuracy. It implies that larger size gives better accuracy.

Figure 23 shows the number of nearest neighbors K with the accuracy. It can be noticed that, after more than around 15 nearest trajectories, the accuracy does not vary significantly. With $K = 50$, a better accuracy can be obtained with a smaller N around 5.

As been mentioned before, a smaller N is preferred and compared with it, the training size K does not affect the simulation time too much. Therefore, we set the number of training trajectories $K = 50$ and the number of nearest neighbor trajectories $N = 5$ for the experiment presented in this work.