A Theoretically Rigorous Approach to Failure Prognosis

David E. Acuña[1] and Marcos E. Orchard[2]

[1,2] *Department of Electrical Engineering, University of Chile, Av. Tupper 2007, Santiago, Chile*
davacuna@ing.uchile.cl
morchard@ing.uchile.cl

[1] *Power Electronics, Machines and Control Group, University of Nottingham, Nottingham, NG7 2RD, UK*

[1] *Resilience Engineering Research Group, University of Nottingham, Nottingham, NG7 2RD, UK*

ABSTRACT

For more than twenty years, we have witnessed a continuous and significant growth in the scope and quality of research in Prognostics and Health Management (PHM). Prognostic algorithms and risk assessment metrics naturally play a critical role in this regard, since they provide the necessary information to take preventive measures and avoid catastrophic system failures. Unfortunately, the problem of failure prognostics has been treated many times from a heuristic, and mostly intuitive, standpoint. Indeed, the PHM community has often validated contributions to the state-of-the-art solely based on the performance experienced under specific run-to-failure experiments, and accepted lack of mathematical rigor in the formulation of the prediction problem itself. In this paper, we revisit the fundamentals of the prognostic problem, providing constructive criticism to inconsistencies found in approaches that have been adopted by many researchers within the PHM community. In addition, we propose a rigorous mathematical framework for failure prognostics, introducing failure probability measures for both discrete- and continuous-time dynamical systems that truly formalize the prognostic problem. We further discuss the philosophical implications of these novel notions in the context of a paradigm change, using as an illustrative example the problem of Lithium-Ion battery condition monitoring.

NOMENCLATURE

Ω	Sample space
$\mathcal{B}$	σ-algebra
$\mathcal{P}(\cdot)$	Probability mass function
$p(\cdot)$	Probability density function
τ_F	Time-of-Failure
$\mathcal{H}_\tau$	Operative system condition at time τ
$\mathcal{F}_\tau$	Catastrophic system failure at time τ
$h_F(\cdot)$	Hazard zone function
$F(\cdot)$	System failure function
$\mathcal{P}_A(\cdot)$	Acuña's probability mass function
$p_A(\cdot)$	Acuña's probability density function
$\mathcal{R}_A(\cdot)$	Acuña's risk function

1. INTRODUCTION

Systems failure has been a topic of great relevance that has been capturing the attention of researchers, and specially from military and industrial sectors, from the early 60's. Given the complexity of equipment and industrial processes, and the operational cost associated with failures and malfunction, it was required to systematically address these issues to include aspects such as reliability, maintainability, and safety. Traditional reliability engineering emerged to fulfill these requirements throughout the development of different methods and tools (Birolini, 2007).

Traditional reliability engineering considered a probabilistic framework from a "frequentist" standpoint (Yang & Xue, 1996; Lall, Pecht, & Harkim, 1997; Pecht, Das, & Ramakrishnan, 2002; Girish, Lam, & Jayaram, 2003; Vichare, Rodgers, Eveloy, & Pecht, 2004; Chen & Zheng, 2005; Kharoufeh & Cox, 2005; Xu & Zhao, 2005): Provided that several statistically identical items (equipment, components, or systems) initiate operation at the same moment, then the probability of successfully accomplishing their purpose along a stated time interval (i.e., the reliability function) can be estimated as the ratio among the number of items that succeeded the purpose over the total amount items (because of the Law of Large Numbers).

Even though the hypothesis of statistically identical items (or independent identically distributed, i.i.d.) adopted by traditional reliability may be accepted as a good approximation in some particular cases, it does not apply in general because

constructing exact copies of an item is not possible at all: there are always unit-to-unit differences even in series production. Moreover, the marginal cost associated to each unit in series (or mass) production is typically small compared to, for example, cases in which the product involves certain customization (e.g., satellites or aircraft). In cases where the production is limited to a few dozens, and mainly due to the complexity of these systems, it seems impossible to hold on to the Law of Large Numbers (or to find two statistically identical systems). In these cases, traditional reliability approaches are not suitable and, thus, a different approach that accounted for the reliability or system health is required.

Prognostics and Health Management (PHM) is a modern engineering discipline that aims at supervising and maintaining the operative continuity of industrial processes, considering its subtleties and particularities, and assuring system safety and quality of end products. It is important to distinguish between two general branches embraced by PHM, which are fault diagnosis and failure prognosis. Fault diagnosis, on the one hand, is related to system monitoring through the detection and isolation of faults in a given system. Fault prognosis, on the other hand, is concerned with the estimation of the Remaining Useful Life (RUL) or Time-of-Failure (ToF) of a faulty system. The latter is achieved by analyzing the evolution in time of system states alongside a characterization of all associated uncertainty sources. One of the main differences among traditional reliability centered analysis and PHM results from the manner in which systems are conceived. In the former case (traditional reliability), prognostics provide information about the operation of several, brand new or recently fixed, statistically identical pieces of equipment that are treated as samples from a given probability distribution (simplifications allow to use historical data for this purpose). In the latter case (PHM), however, we have an unique piece of equipment that is being specifically supervised. Monitoring, in PHM, requires the implementation of online (or, at least, real-time) procedures, so that an optimal maintenance policy could be implemented; i.e., Condition-Based Maintenance (CBM).

2. FAILURE PROGNOSTICS FROM THE STANDPOINT OF THE PHM COMMUNITY

PHM researchers had to face the problem of failure prognostics subject to strong constraints on the amount and quality of data. Most complex systems where it is worth to implement PHM methods have high cost associated with the occurrence of a catastrophic failure. Therefore, failure data is scarce and there are financial and safety constraints on the design run-to-failure experiments. This fact has led to a biased standpoint and a incorrect adoption of the prognosis notion that has been -unintentionally- accepted for more than two decades. This section intends to expose the most critical weaknesses associated with some notions about failure prognosis. Notions that may have seemed right at a first glance (and thus, widely accepted), although theoretically incorrect.

2.1. The assessment of failure prognostic algorithms

The assessment of failure prognostic algorithms typically follows standard procedures where we measure the capability of "guessing" the actual failure time in run-to-failure experiments. In this regard, it would be interesting to wonder if it is possible to use the failure time recorded on those experiments as the "*ground truth*". Depending on whether we use deterministic or a probabilistic approaches to failure prognostics, the answer may differ significantly. The output of deterministic approaches is a number, a failure time, so in those cases we may assume that the failure time in experiments is the *ground truth*. Notwithstanding, the output of probabilistic prognostic algorithms is a sequence of probability distributions for the system state; each element of the sequence characterizing the uncertainty associated with the health indicator at specific future time instants (inducing a failure probability distribution on the time axis). Therefore, how could we even assess the performance of probabilistic prognostic algorithms if an experiment only provides a single realization of the associated stochastic process? Would it be right to compare probability distributions to a single realization (i.e., a Probability Density Function (PDF) vs. a number)? Many performance metrics for prognostic algorithms have, in vain, tried to provide a valid answer to the latter inquiry. We will provide a formal and rigorous answer to this question.

2.2. Inconsistencies in the classic formulation of failure prognosis problem

The failure prognosis problem can be defined in a number of ways. Nonetheless, the mainstay in all of them is the concept of *Time-of-Failure (ToF)*.

Definition 1. ***[Time-of-Failure (widely accepted definition)]*** *The time of the first system failure τ_F is currently defined as*

$$\tau_F := \inf\{k \in \mathsf{N} : \textit{System Failure at } k\}.$$

Note that there are some other notions available in the literature such as Time-to-Failure (TtF), End-of-Life (EoL), or Remaining Useful Life (RUL), which may be equivalent to the ToF in terms of logical/mathematical implications. Therefore, below we only discuss and analyze the consistency of the ToF definition already introduced.

Bayesian approaches (Doucet, de Freitas, & Gordon, 2001) constitute a suitable option for online characterization of uncertainty sources and degradation processes that affect the condition of nonlinear dynamic processes, via a state-space representation. In this regard, let us consider $\{X_k, k \in \mathsf{N}\}$ a first order Markov process denoting a n_x-dimensional system state vector with initial distribution $p(x_0)$ and transition

probability $p(x_k|x_{k-1})$. Also, let $\{Y_k, k \in \mathbb{N} \setminus \{0\}\}$ denote n_y-dimensional conditionally independent noisy observations. Then,

$$x_k = f(x_{k-1}, \omega_{k-1}) \quad (1)$$
$$y_k = g(x_k, \nu_k), \quad (2)$$

where ω_k and ν_k denote independent, not necessarily Gaussian, random vectors.

Considering k_p as the time at which prognostics are executed, this stochastic state-space representation of the system is used later to compute the cumulative probability of the ToF (as understood in Definition 1):

$$\mathcal{P}(\tau_F \leq k) \propto \int_{\mathbb{R}^{n_x}} \mathcal{P}(failure|x_k) p(x_k|y_{1:k_p}) dx_k, \quad (3)$$

where $\mathcal{P}(failure|X)$ corresponds to the probability of system failure, conditional to a value of the state vector $x \in \mathbb{R}^{n_x}$ ("fault dimension" if the concept of *hazard zone* (Saxena, Celaya, Saha, Saha, & Goebel, 2010; Tang, Orchard, Goebel, & Vachtsevanos, 2011) is used instead). In discrete-time system, differentiation is replaced by time differences. Roughly, this is the manner in which the probability distribution for the ToF has been computed (or approximated) for more than two decades.

Among the possible contradictions that may arise from the aforementioned definition of ToF probability measures, and the underlying philosophy adopted by the PHM community, we would like to focus on the following:

a) Determinism: The first big inconsistency is determinism. This idea was already introduced in Section 2.1: all these years, we have been comparing elements that are absolutely different in nature, philosophically speaking. In practice, the state-of-the-art has been built using degradation and, consequently, ToF data from single (or just a few) run-to-failure experiments. These data sets are used to validate the performance of mathematical tools and algorithms whose output are probability distributions. Data is deterministic in nature, whereas the mathematical tools and algorithms provide a probabilistic outcome. Would it be right to compare a probability distribution to a single sample? Of course not. Researchers have tried to use result statistics to avoid this fundamental problem.

b) Non-causality: The expression in Eq. (3) states that the cumulative probability of the ToF can be computed, except by a normalization constant. This assumption implies that, to compute $\mathcal{P}(\tau_F \leq k)$, you need to characterize the manner in which the systems evolves until the "the end of time", just to find the right normalization constant. If you analyze this statement carefully, it is implied that this normalization constant is a function of the time instants beyond k, which is contradictory. Indeed, suppose it is of interest to compute the risk of failure up to a particular fixed time $\overline{k}$, $\overline{k} > k_p$. Would it be correct to think that events happening after $\overline{k}$ have any type of influence on what would have happened from k_p to $\overline{k}$? The answer is "no", because of system causality.

c) Need of normalization constants: The need for a normalizing constant when obtaining the ToF probability distribution throughout the "usual" procedure implies that the outcome is not a probability distribution *per se*, but something else being forced to fulfill probability axioms. In fact, in Probability Theory, probability distributions are not normalized. Normalization is usually employed to compute probability distributions from likelihood functions.

d) Non-increasing Cumulative Probability Functions: It is interesting to note that Eq. (3) has been used to compute the ToF probability distribution on systems that undergo monotonic degradation processes. The latter, because the expression in Eq. (3) has been interpreted as cumulative probability function and thus, it should be an increasing function of time. This property does not hold if the degradation process is not monotonic. Moreover, in the general case (for any type of dynamic system and *hazard zone* definition), it is possible to empirically demonstrate that this concept of cumulative probability is ill-defined. As a side note, it is fairly easy to find degradation processes with regeneration phenomena: consider, for example, capacity regeneration phenomena in Lithium-Ion (Li-Ion) batteries.

These arguments are indicators of a ill-conditioned definition of ToF probability distribution and should be considered and analyzed by the PHM community. This article aims at solving all these inconsistencies by building new foundations to properly define the failure prognosis problem through the application a rigorous mathematical properties and probability theory.

3. THE FAILURE PROGNOSTIC PROBLEM: A RIGOROUS MATHEMATICAL FORMULATION

The main task assigned to failure prognostic algorithms, in the context of online condition monitoring systems, is to characterize damage progression once the fault condition is detected, isolated and identified. On the one hand, data-driven prognostic methods typically perform this task by training empirical models able to characterize damage progression at early stages of the fault, extrapolating the future behavior through the implementation of deterministic n-step prediction routines with the promise of generating an *educated guess* for the failure time of the system. On the other hand, model-based approaches may utilize a probabilistic perspective where empirical information is solely used to provide estimates and uncertainty bounds for parameters in a given

degradation model structure (inspired on prior knowledge about the system dynamics); allowing to propagate uncertainty sources via the implementation of discrete-time diffusion equations.

A serious formalization of the failure prognostic problem must embrace not only the discrete-time case, which is the most common context due to the usage of computers, but also the continuous-time case. For pedagogic purposes, we start formulating it in Section 3.1 assuming discrete-time systems, because in that way it is easier to develop intuition in our readers. Then, results for the continuous-time case are presented in Section 3.3.

3.1. A probabilistic perspective to the failure prognostic problem

For pedagogic purposes, we formulate the failure prognostic problem using a mathematical notation that eases understanding and intuition in our readers. This notation was first introduced in (Acuña & Orchard, 2017) for the discrete-time systems (Allison, 1982).

Let us imagine a system that can incur into a catastrophic failure condition only once. Similarly to the experiment of tossing coin, at each time instant k the system may continue operating or not. We denote $\mathcal{H}_k$ as the event of *being in a faulty, although operative*, condition at time k, whereas $\mathcal{F}_k$ denotes the event of *undergoing a catastrophic failure* at time k.

Thus, we can define a probability space $(\Omega, \mathcal{B}, \mathcal{P})$, where

- $\Omega = \{ \bigcap_{j=k_p}^{k-1} \mathcal{H}_j \bigcap \mathcal{F}_k | \; k \in \mathbb{N},\; 0 < k_p < k\}$ is the sample space that determines all possible sequences where the system remains operative until the time instant "k", when it actually undergoes a catastrophic failure,
- $\mathcal{B} = \sigma(\Omega)$ is the σ-algebra generated by Ω,
- $\mathcal{P}$ is a function that assigns a probability measure to every event in the σ-algebra $\mathcal{B}$.

Given the above, it is possible to characterize the true probability of failure at the k-th time instant, $\mathcal{P}(\mathcal{F}_k)$. Indeed, denoting $\mathcal{H}_{k_p:k} = \bigcap_{j=k_p}^{k} \mathcal{H}_j$, and according to the definition of conditional probability, it follows that:

$$\mathcal{P}(\mathcal{F}_k) = \frac{\mathcal{P}(\mathcal{F}_k, \mathcal{H}_{k_p:k-1})}{\mathcal{P}(\mathcal{H}_{k_p:k-1}|\mathcal{F}_k)}, \quad \forall k > k_p, \tag{4}$$

since $\mathcal{P}(\mathcal{H}_{k_p:k-1}|\mathcal{F}_k)$ corresponds to the probability of staying operative until time $k-1$, given that the failure occurred at time k, it is important to note that $\mathcal{P}(\mathcal{H}_{k_p:k-1}|\mathcal{F}_k) = 1$ (it is assumed that the system can only experience one catastrophic failure). Hence, Equation (4) is reduced to

$$\mathcal{P}(\mathcal{F}_k) = \mathcal{P}(\mathcal{F}_k|\mathcal{H}_{k_p:k-1})\mathcal{P}(\mathcal{H}_{k_p:k-1}). \tag{5}$$

Since $\mathcal{P}(\mathcal{H}_{k_p:k-1})$ is the probability of having the system still operative at the $(k-1)$-th time instant -which means a finite union of events-, and using the basic properties of conditional probabilities, we can write:

$$\begin{aligned}\mathcal{P}(\mathcal{H}_{k_p:k-1}) &= \mathcal{P}(\mathcal{H}_{k-1}|\mathcal{H}_{k_p:k-2})\mathcal{P}(\mathcal{H}_{k_p:k-2}) \\ &= \mathcal{P}(\mathcal{H}_{k-1}|\mathcal{H}_{k_p:k-2})\mathcal{P}(\mathcal{H}_{k-2}|\mathcal{H}_{k_p:k-3})\mathcal{P}(\mathcal{H}_{k_p:k-3}) \\ &\vdots \\ &= \prod_{j=k_p+1}^{k-1} \mathcal{P}(\mathcal{H}_j|\mathcal{H}_{k_p:j-1}).\end{aligned}$$

Additionally, provided

$$\mathcal{P}(\mathcal{H}_j|\mathcal{H}_{k_p:j-1}) = 1 - \mathcal{P}(\mathcal{F}_j|\mathcal{H}_{k_p:j-1}),\ \forall j > k_p,$$

because of the mutual exclusion among operative and failed conditions, it follows that:

$$\mathcal{P}(\mathcal{F}_k) = \mathcal{P}(\mathcal{F}_k|\mathcal{H}_{k_p:k-1}) \prod_{j=k_p+1}^{k-1} 1 - \mathcal{P}(\mathcal{F}_j|\mathcal{H}_{k_p:j-1}) \ . \tag{6}$$

As it can be observed in Eq. (6), any failure probability measure is fully determined by understanding the meaning of $\mathcal{P}(\mathcal{F}_k|\mathcal{H}_{k_p:k-1})$, for all k. It is important to note that in this mathematical notation we assume and omit, on purpose, the conditional on the set of measurements $y_{1:k_p}$ in all expressions.

State-Space Uncertainty Characterization and Failure Time Probability

Let us consider, for a moment, that the true state vector trajectory is known. From a Bayesian standpoint, we should recognize that *the likelihood of a catastrophic failure increases as the system trajectory approaches specific regions of the state-space*. These regions, also referred to as *hazard zones*, can be described by a mapping $h_F(x) : \mathbb{R}^{n_x} \to [0, 1]$:

$$h_F(x) \coloneqq \mathcal{P}(failure|x). \tag{7}$$

The simplest hazard zone corresponds to the case where we claim that the system undergoes a catastrophic failure if the state trajectory reaches a particular manifold on the state-space, denoted by $R \subset \mathbb{R}^{n_x}$. In this case, the hazard zone can be characterized via the indicator function:

$$h_F(x) = \mathbb{1}_R(x), \quad x \in \mathbb{R}^{n_x}, \tag{8}$$

which denotes a discriminant between two classes. Notwithstanding, in most cases the manifold that represents the hazard zone itself is uncertain, and thus it must be represented by an appropriate *likelihood function* $h_F(x) : \mathbb{R}^{n_x} \to [0, 1]$.

Unfortunately, actual implementations of failure prognostic algorithms also require to estimate the state trajectory over time (state trajectory is uncertain). Bayesian Processors (BPs) play a critical role in this task, merging prior information on the system model with noisy measurements (acquired in real-time). BPs provide a probabilistic characterization of the system states that consists of an approximation of the state posterior density; i.e., $p(x_k|y_{1:k}) \approx \hat{p}(x_k|y_{1:k}), \forall k \in \mathsf{N}$, considering for this purpose the state-space representation in Eqs. (1)-(2). The system state vector is no longer deterministic and must be characterized by a probability density on the state-space. Additionally, $h_F(x)$ becomes a function of a random variable. As a consequence, we now need to incorporate the concept of expectation to understand the risk associated with catastrophic failures that may occur in a moment infinitesimally larger than the *current* time instant k:

$$\mathsf{E}_{p(x_k|y_{1:k})}\{h_F(x_k)\} = \int_{\mathsf{R}^{n_x}} h_F(x_k)p(x_k|y_{1:k})dx_k, \quad k \geq 0. \tag{9}$$

Eq. (9) is, by definition, the expectation of the failure likelihood at time k. In Section 3.2, we will provide an *ad-hoc* mathematical notation that will clarify why the notion of expectation, when properly used, becomes important to quantify the risk of failure.

Let us now incorporate the concept of "long-term prediction" in this analysis. Assume that a fault condition is detected and diagnosed at time k_p, moment at which we will desire to execute prognostic algorithms to provide a sequence of prior state probability densities $\hat{p}(x_k|y_{1:k_p})$ with $k \geq k_p$. Then, for each future time instant $k \geq k_p$, we may compute the expected failure likelihood:

$$\mathsf{E}_{p(x_k|y_{1:k_p})}\{h_F(x_k)\} = \int_{\mathsf{R}^{n_x}} h_F(x_k)p(x_k|y_{1:k_p})dx_k. \tag{10}$$

Note that in failure prognosis, all elements in the sequence of expected failure likelihoods are conditioned on the same evidence set (measurements acquired until time k_p). A natural question that arises from this observation is: How do two consecutive elements of the sequence relate to each other?

The answer for this question is, obviously, "two consecutive elements in the sequence are linked through the state transition equation". But this answer has an underlying assumption: *the system has not failed yet, and thus the state transition is still possible*. A catastrophic failure is ruled out since, in that condition, the state transition model become useless. With these ideas in our minds, we can finally provide a proper expression for the term $\mathcal{P}(\mathcal{F}_k|\mathcal{H}_{k_p:k\ \ 1})$ in Eq. (6):

$$\mathcal{P}(\mathcal{F}_k|\mathcal{H}_{k_p:k\ \ 1}) := \mathsf{E}_{p(x_k|y_{1:k_p})}\{h_F(x_k)\}. \tag{11}$$

Eqs. (11) and (6) are the foundation stone for the failure probability measure proposed by (Acuña & Orchard, 2017). For completeness purposes, we now present a theorem that formalizes this result and establishes some necessary conditions for this probability measure.

Towards a Rigorous Formulation of a Failure Probability Measure

We now proceed to introduce a notation that eases the definition of a mathematically rigorous failure probability measure. Let us assume, for these purposes, a first order Markovian system model (see Section 2.2, Eqs. (1)-(2)).

Definition 2. ***[System Failure Function]*** *A system failure is characterized by the function*

$$\begin{aligned} F :& \mathsf{R}^{n_x} \times \Omega \to \{0,1\} \\ & (x,\omega) \mapsto F(x,\omega) = \mathsf{1}_{\textit{System Failure in } x}(\omega), \end{aligned}$$

where $F(X_k)(\cdot) := F(X_k,\cdot)$ corresponds to a binary random variable indicating whether the system is in a failure condition or not, at the k-th time instant.

The function F is introduced as a random variable, although typically it solely corresponds to an indicator function (see the concept of hazard zone in Eq. (7)).

We have already discussed the random nature of system failures. We still need, however, to incorporate the concept of failure *observability*. A failure is said to be *observable* if it is possible to recognize that a failure event takes place based on the analysis of the information provided by sensor measurements.

Definition 3. ***[Failure Observability]*** *A failure is observable if and only if $\exists G : \mathsf{R}^{n_y} \times \Omega \to \{0,1\}$ such that*

$$F(X_k) = G(Y_k) \quad \mathcal{P}-a.s., \quad \forall k \geq 1.$$

Catastrophic failures are *almost surely* observable from system measurements, since they are related to an interruption of the operation. The concept of failure observability is helpful to justify the implementation of BPs for condition monitoring.

In prognostics, though, measurements are not available for $k \geq k_p$ (it is assumed that the prognostic algorithm is executed at time k_p). Given the random nature of failure events, the characterization of the risk associated with a failure event requires the definition of a probability measure for the time of occurrence of the first system failure, as follows.

Definition 4. ***[Time-of-Failure, ToF]*** *The time of the first system failure τ_F is defined as*

$$\tau_F := \inf\{k \in \mathsf{N} : F(X_k) = 1\}.$$

Note that the system could undergo regenerative phenomena, so neither $F(X_k)$ is necessarily increasing (as a function of the time index k) nor τ_F is *almost surely* finite. Therefore, a definite failure probability measure needs to be defined under very strict assumptions. Before we proceed with this formal statement, let us summarize the notation associated with the main constitutive elements of the failure prognosis problem:

- The hazard zone:

$$h_F(x_k) = \mathcal{P}(F(x_k) = 1). \tag{12}$$

- The probability of failure given an operative previous condition:

$$\begin{aligned}\mathcal{P}(F_k|H_{0:k-1}) &= \int_{\mathbb{R}^{n_x}} h_F(x_k)p(x_k|y_{1:k_p})dx_k \\ &= \int_{\mathbb{R}^{n_x}} \mathcal{P}(F(x_k) = 1)p(x_k|y_{1:k_p})dx_k \\ &= \mathcal{P}(F(X_k) = 1|y_{1:k_p}) \\ &= \mathbb{E}_{p(x_k|y_{1:k_p})}\{\mathcal{P}(F(X_k))\}.\end{aligned}$$

  Indeed, provided $F(X_k)$ is a random binary variable, we have:

$$\begin{aligned}\mathbb{E}_{p(x_k|y_{1:k_p})}\{\mathcal{P}(F(X_k))\} &= 1 \cdot \mathcal{P}(F(X_k) = 1|y_{1:k_p}) \\ &\quad \ldots + 0 \cdot \mathcal{P}(F(X_k) = 0|y_{1:k_p}) \\ &= \mathcal{P}(F(X_k) = 1|y_{1:k_p}).\end{aligned}$$

- The event of failure at time k:

$$\mathcal{F}_k = \{\tau_F = k\}. \tag{13}$$

3.2. The failure prognostic problem in discrete-time dynamic systems

Before we proceed to derive *ad-hoc* probability measures, we need to take some considerations. Firstly, given that $\mathcal{F}_k = \{\tau_F = k\} \in \Omega$, then $\cup_{i=0}^{k-1}\mathcal{F}_i = \cup_{i=0}^{k-1}\{\tau_F = i\} \in \sigma(\Omega)$ and its complement, $\cap_{i=0}^{k-1}\{\tau_F \neq i\} = \cap_{i=0}^{k-1}\mathcal{H}_i = \mathcal{H}_{0:k-1}$, are measurable sets. Applying the Theorem of Radon-Nikodym, $\mathcal{P}(\cup_{i=0}^{k-1}\mathcal{F}_i)$ as a function of time $k \in \mathbb{N}$ is absolutely continuous with respect to the *counting measure* in $\mathbb{N}$ (which is always true since every measure is absolutely continuous with respect to the counting measure on a discrete space), then *there exists a unique probability mass function* $\mathcal{P}(\mathcal{F}_k)$ such that

$$\begin{aligned}\mathcal{P}(\mathcal{H}_{0:k-1}) &= 1 - \mathcal{P}(\cup_{i=0}^{k-1}\mathcal{F}_i) \\ &= 1 - \sum_{i=0}^{k-1}\mathcal{P}(\mathcal{F}_i).\end{aligned}$$

Note that $\mathcal{F}_k = \{\tau_F = k\} = \{\tau_F = k\} \cap \cap_{i=0}^{k-1}\{\tau_F \neq i\} = \mathcal{F}_k \cap \mathcal{H}_{0:k-1}$, because failing for a first time at k implies the system was necessarily healthy up to just one time instant before, then

$$\begin{aligned}\mathcal{P}(\mathcal{F}_k) &= \mathcal{P}(\mathcal{F}_k, \mathcal{H}_{0:k-1}) \\ &= \mathcal{P}(\mathcal{F}_k|\mathcal{H}_{0:k-1})\mathcal{P}(\mathcal{H}_{0:k-1}).\end{aligned}$$

Thus,

$$\mathcal{P}(\mathcal{H}_{0:k-1}) = 1 - \sum_{i=0}^{k-1}\mathcal{P}(\mathcal{F}_i|\mathcal{H}_{0:i-1})\mathcal{P}(\mathcal{H}_{0:i-1}).$$

Applying the difference operator Δ to $\mathcal{P}(\mathcal{H}_{0:k-1})$, then it follows that

$$\begin{aligned}\Delta\mathcal{P}(\mathcal{H}_{0:k-1}) &= -\mathcal{P}(\mathcal{F}_k|H_{0:k-1})\mathcal{P}(\mathcal{H}_{0:k-1}) \\ \mathcal{P}(\mathcal{H}_{0:k}) - \mathcal{P}(\mathcal{H}_{0:k-1}) &= -\mathcal{P}(\mathcal{F}_k|\mathcal{H}_{0:k-1})\mathcal{P}(\mathcal{H}_{0:k-1}) \\ \mathcal{P}(\mathcal{H}_{0:k}) &= (1 - \mathcal{P}(\mathcal{F}_k|\mathcal{H}_{0:k-1}))\mathcal{P}(\mathcal{H}_{0:k-1}).\end{aligned}$$

Applying natural logarithm we get

$$\begin{aligned}\log\mathcal{P}(\mathcal{H}_{0:k}) - \log\mathcal{P}(\mathcal{H}_{0:k-1}) &= \log(1 - \mathcal{P}(\mathcal{F}_k|\mathcal{H}_{0:k-1})) \\ \Delta\log\mathcal{P}(\mathcal{H}_{0:k-1}) &= \log(1 - \mathcal{P}(\mathcal{F}_k|\mathcal{H}_{0:k-1})).\end{aligned}$$

Now summing from $k_p + 1$ up to k,

$$\sum_{i=k_p+1}^{k}\Delta\log\mathcal{P}(\mathcal{H}_{0:i-1}) = \sum_{i=k_p+1}^{k}\log(1 - \mathcal{P}(\mathcal{F}_i|\mathcal{H}_{0:i-1}))$$

$$\log\mathcal{P}(\mathcal{H}_{0:k}) - \log\mathcal{P}(\mathcal{H}_{0:k_p}) = \log\prod_{i=k_p+1}^{k}(1 - \mathcal{P}(\mathcal{F}_i|\mathcal{H}_{0:i-1})).$$

Note also that $\mathcal{P}(\mathcal{H}_{0:k_p}) = 1$, because the system is assumed to be operative at the beginning of prognostics.

$$\Rightarrow \mathcal{P}(\mathcal{H}_{0:k}) = \prod_{i=k_p+1}^{k}(1 - \mathcal{P}(\mathcal{F}_i|\mathcal{H}_{0:i-1})).$$

Finally, and considering all these elements, the failure probability mass function for discrete-time systems can be defined as follows.

A well-defined risk-of-failure function (dicrete-time systems)

Provided the previous results defined the probability mass $\mathcal{P}(\{\tau_F = k\}) = \mathcal{P}(\mathcal{F}_k)$, let us show that it is well-defined.

Theorem 1. ***[Acuña's Failure Probability Mass Function]*** *Considering the probability space $(\mathbb{N}, \sigma(\mathbb{N}), \mathcal{P})$ and given that $\mathcal{P}(\cup_{i=0}^{k-1}\{\tau_F = i\}|y_{1:k_p})$ as a function of time $k \in \mathbb{N}$ is always absolutely continuous with respect to the counting measure in $\mathbb{N}$, if the following conditions hold:*

- *$\tau_F < +\infty$, $\mathcal{P}(\tau_F = \cdot|y_{1:k_p})$-a.s.*

- $\tau_F > k_p$.

Then the mapping $\mathcal{P}(\tau_F = \cdot|y_{1:k_p}) : \mathbb{N} \to [0,1]$ *defines the Acuña's Time-of-Failure Probability Mass Function as*

$$\mathcal{P}_A(k|y_{1:k_p}) \coloneqq \mathcal{P}(\tau_F = k|y_{1:k_p}) \tag{14}$$

$$= \mathcal{P}(F(X_k) = 1|y_{1:k_p}) \prod_{j=k_p+1}^{k-1} 1 - \mathcal{P}(F(X_j) = 1|y_{1:k_p}) \ , \tag{15}$$

with

$$\mathcal{P}(F(X_k) = 1|y_{1:k_p}) = \int_{\mathbb{R}^{n_x}} \mathcal{P}(F(x_k) = 1) p(x_k|y_{1:k_p}) dx_k. \tag{16}$$

This probability measure is well-defined (satisfies the conditions of probability mass function) and corresponds to the ***unique*** *Bayesian probability function that can characterize the risk of future failures in discrete-time systems. Therefore, it holds that*

1) $\mathcal{P}_A(k|y_{1:k_p}) = 0, \ \forall k \in \mathbb{N}, \ k \leq k_p$.

2) $0 \leq \mathcal{P}_A(k|y_{1:k_p}) \leq 1, \ \forall k \in \mathbb{N}$.

3) $\sum_{i=0}^{+1} \mathcal{P}_A(k|y_{1:k_p}) = 1$.

Proof.

1) By definition. The system is guaranteed to be healthy at least till time k_p.

2) $0 \leq \mathcal{P}(F(X_k) = 1|y_{1:k_p}) \leq 1 \Rightarrow 0 \leq \prod_{i=k_p+1}^{k-1}(1 - \mathcal{P}(F(X_i) = 1|y_{1:k_p})) \leq 1, \ \forall k \in \mathbb{N}, \ k > k_p \geq 0$. Then it imples that $0 \leq \mathcal{P}_A(k|y_{1:k_p}) \leq 1, \ \forall k \in \mathbb{N}$.

3) Since $\mathcal{P}(\cup_{i=0}^{k-1}\{\tau_F = i\}|y_{1:k_p})$ as a function of time $k \in \mathbb{N}$ is always absolutely continuous with respect to the counting measure in $\mathbb{N}$,

$$\begin{aligned} \sum_{i=0}^{k-1} \mathcal{P}_A(i|y_{1:k_p}) &= \sum_{i=0}^{k-1} \mathcal{P}(\tau_F = i|y_{1:k_p}) \\ &= \mathcal{P}(\cup_{i=0}^{k-1}\{\tau_F = i\}|y_{1:k_p}) \\ &= \mathcal{P}(\tau_F < k|y_{1:k_p}). \end{aligned}$$

Taking the limit when $k \to +\infty$ we have that

$$\lim_{k \to +1} \mathcal{P}(\tau_F < k|y_{1:k_p}) = \mathcal{P}\left(\lim_{k \to +1} \{\tau_F < k\}|y_{1:k_p}\right) \ ,$$

because $\{\tau_F < k\}_{k \in \mathbb{N}}$ is an increasing sequence of measurable sets so the continuity property of probability measures hold. Additionally, since $\tau_F < +\infty$, $\mathcal{P}(\tau_F = \cdot|y_{1:k_p})$-a.s.

$$\mathcal{P}\left(\lim_{k \to +1} \{\tau_F < k\}|y_{1:k_p}\right) = 1.$$

Thus, it yields

$$\Rightarrow \sum_{i=0}^{+1} \mathcal{P}_A(k|y_{1:k_p}) = 1.$$

□

Acuña's Time-of-Failure Probability Mass Function allows to build a metric that quantifies future risk in the operation of a dynamic non-linear system that undergoes a fault condition. This risk measure is to be understood as the final outcome of any prognostic algorithm, and the main objective behind the implementation of routines for the quantification of system uncertainty:

Definition 5. ***[Acuña's Discrete-Time Risk-of-Failure]*** *If the following conditions are fulfilled:*

- $\tau_F < +\infty$, $\mathcal{P}(\tau_F = \cdot|y_{1:k_p})$*-a.s.*
- $\tau_F > k_p$.

Then the risk of incurring into a future catastrophic failure in discrete-time systems is defined as:

$$\mathcal{R}_A(k|y_{1:k_p}) \coloneqq \mathcal{P}(\tau_F \leq k|y_{1:k_p}) \tag{17}$$

$$= \sum_{i=k_p+1}^{k} \mathcal{P}_A(i|y_{1:k_p}). \tag{18}$$

3.3. The failure prognostic problem in continuous-time dynamic systems

In this section we show the formalization of the failure prognostic problem in continuous-time dynamic systems, following a similar line of though than in the case of discrete-time systems (see Section 3.2).

Analogously to the discrete-time case, we denote $\mathcal{H}_t$ as the event of *being in a faulty, although operative*, condition at time t, whereas $\mathcal{F}_t$ denotes the event of *undergoing a catastrophic failure* at time t.

Thus, we can define a probability space $(\Omega, \mathcal{B}, \mathcal{P})$, where

- $\Omega = \{ \bigcap_{\tau \in [t_p,t)} \mathcal{H}_\tau \bigcap \mathcal{F}_t | \ t \in \mathbb{R}, \ 0 < t_p < t\}$ is the sample space that determines all possible sequences where the system remains operative until the time instant "t", when it actually undergoes a catastrophic failure,
- $\mathcal{B} = \sigma(\Omega)$ is the σ-algebra generated by Ω,
- $\mathcal{P}$ is a function that assigns a probability measure to every event in the σ-algebra $\mathcal{B}$.

Given the above, it is possible to characterize the true failure probability density at time t, $p(\mathcal{F}_t)$. Indeed, denoting $\mathcal{H}_{t_p:t} = \bigcap_{\tau \in [t_p,t)} \mathcal{H}_\tau$, it follows that:

$$p(\mathcal{F}_t) = \mathcal{P}(\mathcal{F}_t|\mathcal{H}_{t_p:t}) e^{\int_{t_p}^{t} P(F|H_{t_p:})d\tau}, \tag{19}$$

with

$$\mathcal{P}(\mathcal{F}_t|\mathcal{H}_{t_p:t}) \coloneqq \mathsf{E}_{p(x_t|y_{1:t_p})}\{h_F(x_t)\}. \tag{20}$$

Considering Definition 2 for the system failure function, given that $\mathcal{F}_t = \{\tau_F = t\} \in \Omega$, then $\cup_{\tau \geq [0,t)}\mathcal{F}_\tau = \cup_{\tau \geq [0,t)}\{\tau_F = \tau\} \in \sigma(\Omega)$ and its complement, $\cap_{\tau \geq [0,t)}\{\tau_F \neq \tau\} = \cap_{\tau \geq [0,t)}\mathcal{H}_\tau = \mathcal{H}_{[0,t)}$, are measurable sets.

Applying the Theorem of Radon-Nikodym, and assuming that $\mathcal{P}(\cup_{\tau \geq [0,t)}\mathcal{F}_\tau)$ as a function of time $t \in \mathsf{R}_+$ is absolutely continuous with respect to the *Lebesgue measure* in R_+, then *there exists a unique probability density function* $p(\mathcal{F}_t)$ such that

$$\begin{aligned}\mathcal{P}(\mathcal{H}_{[0,t)}) &= 1 - \mathcal{P}(\cup_{\tau \geq [0,t)}\mathcal{F}_\tau) \\ &= 1 - \int_0^t p(\mathcal{F}_\tau)d\tau.\end{aligned}$$

Note that $\mathcal{F}_t = \{\tau_F = t\} = \{\tau_F = t\} \cap \ \cap_{\tau \geq [0,t)} \{\tau_F \neq \tau\} \ = \mathcal{F}_t \cap \mathcal{H}_{[0,t)}$, because failing for a first time at t implies the system was necessarily healthy up to just one time instant before, then

$$\begin{aligned}p(\mathcal{F}_t) &= \mathcal{P}(\mathcal{F}_t, \mathcal{H}_{[0,t)}) \\ &= \mathcal{P}(\mathcal{F}_t|\mathcal{H}_{[0,t)})\mathcal{P}(\mathcal{H}_{[0,t)}).\end{aligned}$$

Thus,

$$\mathcal{P}(\mathcal{H}_{[0,t)}) = 1 - \int_0^t \mathcal{P}(\mathcal{F}_\tau|\mathcal{H}_{[0,\tau)})\mathcal{P}(\mathcal{H}_{[0,\tau)})d\tau.$$

Assuming $\mathcal{P}(\mathcal{H}_{[0,t)})$ to be differentiable with respect to t, then by the Fundamental Theorem of Calculus it follows that

$$\frac{d}{dt}\mathcal{P}(\mathcal{H}_{[0,t)}) = -\mathcal{P}(\mathcal{F}_t|\mathcal{H}_{[0,t)})\mathcal{P}(\mathcal{H}_{[0,t)}).$$

Then,

$$\frac{1}{\mathcal{P}(\mathcal{H}_{[0,t)})}d\mathcal{P}(\mathcal{H}_{[0,t)}) = -\mathcal{P}(\mathcal{F}_t|\mathcal{H}_{[0,t)})dt.$$

Integrating between t_p and t, we get

$$\begin{aligned}\int_{P(H_{[0;t_p)})}^{P(H_{[0;t)})} \frac{1}{\mathcal{P}(\mathcal{H}_{[0,\tau)})}d\mathcal{P}(\mathcal{H}_{[0,\tau)}) &= -\int_{t_p}^t \mathcal{P}(\mathcal{F}_\tau|\mathcal{H}_{[0,\tau)})d\tau \\ \log\mathcal{P}(\mathcal{H}_{[0,t)}) - \log\mathcal{P}(\mathcal{H}_{[0,t_p)}) &= -\int_{t_p}^t \mathcal{P}(\mathcal{F}_\tau|\mathcal{H}_{[0,\tau)})d\tau.\end{aligned}$$

Note also that $\mathcal{P}(\mathcal{H}_{[0,t_p)}) = 1$, by the same argument of system health at the beginning of prognostics.

$$\Rightarrow \mathcal{P}(\mathcal{H}_{[0,t)}) = e^{\int_{t_p}^t P(F\ jH_{[0;\ )})d\tau}.$$

Finally, the failure probability density function for continuous-time systems follows next.

A well-defined risk-of-failure function (continuous-time systems)

Let us demonstrate that $\mathcal{P}(\{\tau_F = t\}) = \mathcal{P}(\mathcal{F}_t)$ is an appropriate definition for the probability of failure.

Theorem 2. ***[Acuña's Failure Probability Density Function]*** *Considering the probability space* $(\mathsf{R}_+, \sigma(\mathsf{R}_+), \mathcal{P})$ *and assuming that* $\mathcal{P}(\cup_{\tau \geq [0,t)}\{\tau_F = \tau\}|y_{1:t_p})$ *as a function of time* $t \in \mathsf{R}_+$ *is absolutely continuous with respect to the Lebesgue measure in* R_+*, if the following conditions hold:*

- $\tau_F < +\infty$, $p(\tau_F = \cdot|y_{1:t_p})$*-a.s.*
- $\tau_F > t_p$.

Then the mapping $p(\tau_F = \cdot|y_{1:t_p}) : \mathsf{R}_+ \to [0,1]$ *defines the Acuña's Time-of-Failure Probability Density Function as*

$$p_A(t|y_{1:t_p}) \coloneqq p(\tau_F = t|y_{1:t_p}) \tag{21}$$

$$= \mathcal{P}(F(X_t) = 1|y_{1:t_p})e^{\int_{t_p}^t P(F(X\ ) = 1jy_{1:t_p})d\tau}, \tag{22}$$

with

$$\mathcal{P}(F(X_t) = 1|y_{1:t_p}) = \int_{\mathsf{R}^{n_x}} \mathcal{P}(F(x_t) = 1)p(x_t|y_{1:t_p})dx_t. \tag{23}$$

This probability density is well-defined (satisfies the conditions of probability density function) and corresponds to the ***unique*** *Bayesian probability function that can characterize the risk of future failures in continuous-time systems. Therefore, it holds that*

1) $p_A(t|y_{1:t_p}) = 0,\ \forall t \in \mathsf{R}_+,\ t \leq t_p$.

2) $0 \leq p_A(t|y_{1:t_p}) \leq 1,\ \forall t \in \mathsf{R}_+$.

3) $\int_0^{+1} p_A(t|y_{1:t_p})d\tau = 1$.

Proof.

1) By definition. The system is guaranteed to be healthy at least till time t_p.

2) $0 \leq \mathcal{P}(F(X_t) = 1|y_{1:t_p}) \leq 1 \Rightarrow 0 \leq e^{\int_{t_p}^t P(F(X\ ) = 1jy_{1:t_p})d\tau} \leq 1,\ \forall t \in \mathsf{R}_+,\ t > t_p \geq 0$. Then it implies that $0 \leq p_A(t|y_{1:t_p}) \leq 1,\ \forall t \in \mathsf{R}_+$.

3) Since $\mathcal{P}(\cup_{\tau \geq [0,t)}\{\tau_F = \tau\}|y_{1:t_p})$ as a function of time $t \in \mathsf{R}_+$ is assumed to be absolutely continuous with respect to the Lebesgue measure in R_+,

$$\begin{aligned}\int_0^t p_A(\tau|y_{1:t_p})d\tau &= \int_0^t p(\tau_F = \tau|y_{1:t_p})d\tau \\ &= \mathcal{P}(\cup_{\tau \geq [0,t)}\{\tau_F = \tau\}|y_{1:t_p}) \\ &= \mathcal{P}(\tau_F < t|y_{1:t_p}).\end{aligned}$$

Taking the limit when $t \to +\infty$ we have that

$$\lim_{t \to +\infty} \mathcal{P}(\tau_F < t | y_{1:t_p}) = \mathcal{P}\left(\lim_{t \to +\infty} \{\tau_F < t\} | y_{1:t_p}\right),$$

because $\{\tau_F < t\}_{t \in \mathbb{N}}$ is an increasing sequence of measurable sets so the continuity property of probability measures hold. Additionally, since $\tau_F < +\infty$, $p(\tau_F = \cdot | y_{1:t_p})$-a.s.

$$\mathcal{P}\left(\lim_{t \to +\infty} \{\tau_F < t\} | y_{1:t_p}\right) = 1.$$

Thus, it yields

$$\Rightarrow \int_0^{+\infty} p_A(t | y_{1:t_p}) d\tau = 1.$$

□

The *Acuña's Time-of-Failure Probability Density Function* allows to build a metric that quantifies future risk in the operation of a dynamic non-linear system that undergoes a fault condition. This risk measure is to be understood as the final outcome of any prognostic algorithm, and the main objective behind the implementation of routines for the quantification of system uncertainty:

Definition 6. ***[Acuña's Continuous-Time Risk-of-Failure]*** *If the following conditions are fulfilled:*

- $\tau_F < +\infty$, $p(\tau_F = \cdot | y_{1:t_p})$*-a.s.*
- $\tau_F > t_p$.

Then, the risk of incurring into a future catastrophic failure in continuous-time systems is defined as:

$$\mathcal{R}_A(t | y_{1:t_p}) := \mathcal{P}(\tau_F \leq t | y_{1:t_p}) \tag{24}$$

$$= \int_{t_p}^{t} p_A(t | y_{1:t_p}) d\tau. \tag{25}$$

4. DEFINITION OF FAILURE PROBABILITY IN PHM AND RELIABILITY ENGINEERING

Theorems 1 and 2 provide a link between definitions in the disciplines of Prognostics and Health Management and traditional reliability engineering. In fact, let us consider the following:

- **Reliability, R:** Probability that the component or system experiences no failures up to some time instant given that the component or system was repaired to a like new condition or was functioning at time zero.
- **Unreliability, F:** Probability that the component or system experiences the first failure or has failed one or more times up to some time instant given that it was operating or repaired to a like new condition at time zero.

These concepts are related as

$$\mathsf{R} = 1 - \mathsf{F}.$$

In most of the reliability engineering literature, risk is quantified by estimating probability densities in continuous-time. In this regard, two more concepts arise:

- **Failure Density, $\mathrm{f}(t)$:** The failure density of a component or system, $\mathrm{f}(t)$, is defined as the probability per unit time that the component or system experiences its first failure at time t, given that the component or system was operating at time zero.
- **Failure Rate, $\mathrm{r}(t)$:** The failure rate of a component or system, $\mathrm{r}(t)$, is defined as the probability per unit time that the component or system experiences a failure at time t, given that the component or system was operating at time zero and has survived to time t.

which are related to $\mathsf{R}(t)$ and $\mathsf{F}(t)$ as:

$$\mathsf{F}(t) = \int_0^t \mathrm{f}(\tau) d\tau, \quad \mathsf{R}(t) = e^{\int_0^t \mathrm{r}(\tau) d\tau},$$

whereas

$$\mathrm{f}(t) = \mathrm{r}(t) e^{\int_0^t \mathrm{r}(\tau) d\tau}.$$

As it can be seen in the previous equations, the relationship between PHM and reliability engineering in continuous-time is given by the analogy:

$$\mathrm{r}(t) \leftrightarrow \mathcal{P}(F(X_t) = 1 | y_{1:t_p}).$$

Remark 1. *Note that integrals in PHM and reliability engineering definitions start from t_p and 0, respectively. This is because in reliability engineering it is assumed $t_p = 0$; but that condition doesn't hold in general, as in the case of PHM, where prognostics is performed just once an anomaly is detected.*

Remark 2. *In this case, Acuña's Failure Probability Density Function, $p_A(t | y_{1:t_p})$, coincides with the well-known continuous-time Weibull distribution*

$$p_{Wb}(t) = \frac{\gamma}{\lambda} \left(\frac{t}{\lambda}\right)^{\gamma - 1} e^{-\left(\frac{t}{\lambda}\right)^{\gamma}},$$

where $\gamma > 0$ is a ***shape*** *parameter and $\lambda > 0$ is a* ***scale*** *parameter, if $\mathcal{P}(F(X_t) = 1 | y_{1:t_p})$ is such that*

$$\mathcal{P}(F(X_t) = 1 | y_{1:t_p}) = \frac{\gamma}{\lambda} \left(\frac{t}{\lambda}\right)^{\gamma - 1}.$$

Similarly to the continuous-time case, in discrete-time we have:

$$\mathsf{F}(k) = \sum_{i=0}^{k} \mathrm{f}(i), \quad \mathsf{R}(k) = \prod_{i=1}^{k} (1 - \mathrm{r}(i)),$$

whereas

$$\mathrm{f}(k) = \mathrm{r}(k) \prod_{i=1}^{k-1} (1 - \mathrm{r}(i)).$$

Finally, the relationship between PHM and reliability engineering in discrete-time is given by the analogy:

$$\mathrm{r}(k) \leftrightarrow \mathcal{P}(F(X_k) = 1 | y_{1:k_p}).$$

Remark 3. *The sums and products in the recent expressions start from* 1 *instead of* $k_p + 1$ *because traditional reliability assumes* $k_p = 0$*. It is the same argument explained in Remark 1 but for discrete-time.*

Remark 4. *The Acuña's Failure Probability Mass Function,* $\mathcal{P}_A(k|y_{1:k_p})$*, coincides with the well-known discrete-time Weibull distribution*

$$P_{Wb}(k) = q^{k^{\beta}} - q^{(k+1)^{\beta}},$$

where q *is a parameter such that* $0 < q < 1$ *and* $\beta > 0$ *is a* ***shape*** *parameter, if* $\mathcal{P}(F(X_k) = 1|y_{1:k_p})$ *is such that*

$$\mathcal{P}(F(X_k) = 1|y_{1:k_p}) = \log q^{k^{(k+1)^{\beta}}}.$$

5. CASE STUDY: BATTERY END-OF-DISCHARGE TIME PROGNOSIS

Although it seems natural to think of a failure as a physical event (and that is why concepts like RUL or EoL are commonly found in the literature), failures are conceived abstractly from the standpoint if this research work and, thus, it is up to the user how to define a failure in a particular system. For illustrative purposes, we hereby present the problem of End-of-Discharge (EoD) time prognostics in lithium-ion (Li-Ion) batteries, where the interest is on monitoring the remaining energy available in a battery (rather than predicting its RUL). Contexts in which this problem arises may be related to problem of energy autonomy, such as in the operation of unmanned aerial vehicles (UAVs) or electric vehicles (EVs), where under certain decisions may compromise the feasibility of successfully achieving specific tasks.

For most of the battery operating range, the relationship between State-of-Charge (SoC, defined as the ratio between the actual available energy and the maximum battery storage capacity E_{crit}) and the *Open Circuit Voltage* (OCV) curve can be well characterized by an affine function. Also, we have adopted a structure proposed in (Burgos, Orchard, Kazerani, Cárdenas, & Sáez, 2016) to model the dependency between the polarization resistance and the battery discharge current. Thus, the resulting state-space model of the system is:

State Transition Model

$$x_{k+1} = x_k - v_{oc}(x_k) \cdot u_k \cdot \frac{T_s}{E_{crit}} + \omega_k \qquad (26)$$

Measurement Model

$$y_k = v_{oc}(x_k) - u(k) \cdot R_{int}(x_k, u_k) + \eta_k, \qquad (27)$$

with

$$\begin{aligned} v_{oc}(x_k) =& v_L + (v_0 - v_L) \cdot e^{\gamma\,(x_2(k)\ \ 1)} \\ & \ldots + \alpha \cdot v_L \cdot (x_2(k) - 1) \\ & \ldots + (1 - \alpha) \cdot v_L \cdot (e^{\ \beta} - e^{\ \beta\sqrt{x_2(k)}}) \end{aligned} \qquad (28)$$

and

$$R_{int}(x_k, u_k) = r_0(u_k) + r_1(u_k) \cdot x_k + r_2(u_k) \cdot {x_k}^2. \qquad (29)$$

In this representation, the input to the system $u_k = i_k[A]$ is defined as the discharge current, while $y_k = v_k[V]$ is the voltage at the battery terminals. The state x_k is the battery SoC measured with respect to E_{crit}, the expected total energy delivered by the battery; whereas the absolute value of the internal impedance is represented by the function $R_{int}(x_k, u_k)$. The process noise ω_k and the measurement noise η_k assume a zero mean Gaussian distribution. Finally, $T_s[s]$ is the sample time and $v_0,\ v_L,\ \alpha,\ \beta$ and γ are model parameters to be estimated offline (see (Pola et al., 2015) for more details).

In this case, "failure" is defined as a condition where the SoC reaches a lower threshold of 10%. Thus, Eq. (12) becomes:

$$\mathcal{P}(F(x_k) = 1) = \mathbf{1}_{fx2\mathrm{R}:x<0.1g}(x_k). \qquad (30)$$

Fig. 1 illustrates the problem of battery EoD time prognosis, allowing to compare the manner in which probability measures for the ToF have been computed until now by most PHM researchers with respect to the way they should be computed. A probabilistic prognostic algorithm generates a sequence of PDFs for the future system states (the SoC in this particular example) as a function of time. However, since prognostic algorithms themselves are out of the scope of this article, in Fig. 1 we show the results when performing Monte Carlo simulations. Having available the sequence of future system states PDFs, the next step is to, respectively, compute:

1. $\mathcal{P}(F(X_k) = 1|y_{1:k_p})$.
2. $\mathcal{P}_A(k|y_{1:k_p})$.
3. $\mathcal{R}_A(k|y_{1:k_p})$.

Regarding the interpretation of the aforementioned functions, it is important to note that $\mathcal{P}(F(X_k) = 1|y_{1:k_p})$ (probability mass obtained when performing integration of $p(x_k|y_{1:k_p})$ for values of x_k below the threshold of 10% SoC in this example) has been typically misunderstood as $\mathcal{R}_A(k|y_{1:k_p})$ within the PHM community. Although both expressions represent probability of failure in the future, $\mathcal{P}(F(X_k) = 1|y_{1:k_p})$ accounts for the probability of failure **conditional** to the fact that the system did not undergo a "catastrophic" failure at time instants prior to $k > k_p$, whereas and $\mathcal{R}_A(k|y_{1:k_p})$ character-

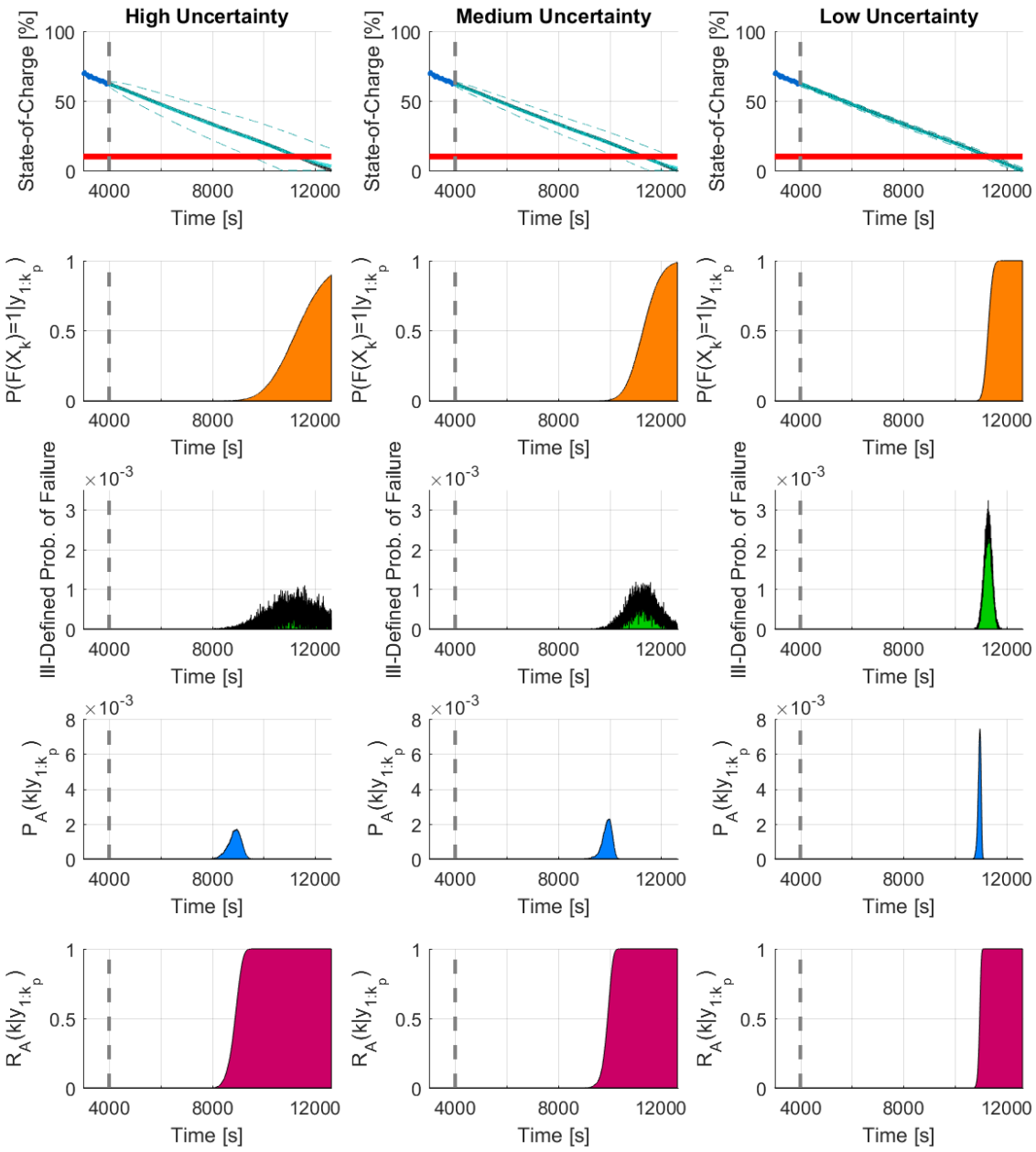

Figure 1. (Battery condition monitoring) A particle-filter algorithm is chosen as the Bayesian processor for estimating the SoC probability distribution during the *filtering* stage. At time $k_p = 4000[s]$ (marked with a vertical dashed line), Monte Carlo simulations are performed. It is considered that a SoC below 10% corresponds to a catastrophic failure, which is depicted as a threshold with a red horizontal line in the first row of graphs. Monte Carlo simulations generate a sequence of future prior state PDFs for the state of the system. This sequence is afterwards used to compute the future failure likelihood $\mathcal{P}(F(X_k) = 1|y_{1:kp})$, which has been understood as cumulative probability so far, leading to an ill-defined probability mass function that is typically computed by taking differences in time. The probability mass function $\mathcal{P}_A(k|y_{1:kp})$ for the random variable τ_F, the EoD time, and its respective risk function $\mathcal{R}_A(k|y_{1:kp})$ are depicted.

izes the probability of experiencing a single "catastrophic" failure event either at or previous to the time instant "k". Another important difference is that $\mathcal{P}(F(X_k) = 1|y_{1:k_p})$ is not a cumulative mass function since it is not necessarily an increasing function of time (consider, for example, regenerative systems), among some other cases. This issue has not been detected, possibly because simulations typically show that $\mathcal{P}(F(X_k) = 1|y_{1:k_p})$ increases (apparently) monotonically in time. In contrast, the Acuña's Risk-of-Failure function $\mathcal{R}_A(k|y_{1:k_p})$ is guaranteed to be a cumulative mass function according to the Theorem 1, and it corresponds to the information that is truly required to perform decision-making. Moreover, when comparing $\mathcal{P}(F(X_k) = 1|y_{1:k_p})$ to $\mathcal{R}_A(k|y_{1:k_p})$ in Fig. 1, it can be noticed that misunderstanding $\mathcal{P}(F(X_k) = 1|y_{1:k_p})$ as a measure of risk may lead to actually riskier actions, especially when the characterization of system dynamics is highly uncertain (which is the most usual scenario when studying degradation processes).

Perhaps the most important contribution of this work is to demonstrate that we require to change the manner in which we understand and conceive the concept of *future failure events*; in other words, a change of paradigm of the failure prognostic problem is needed. As Fig. 1 shows, the widely-used, and ill-defined, ToF probability distribution (in green) provides an estimate for the ToF expectation that is relatively insensitive to the uncertainty associated to the system characterization. Moreover, when using this ill-defined probability measure, an increment on the uncertainty in the system (for example a greater variance in the process noise) turns into an increment in the uncertainty of the ToF, a fact that could be considered as "intuitive" (and thus, accepted) by researchers. However, when analyzing the proposed definition, $\mathcal{P}_A(k|y_{1:k_p})$ (in blue), it can be noticed that increasing the uncertainty of the system dynamics will change the variance of the ToF probability distribution and the expectation of the ToF: the risk has increased as it should! **The failure prognostic problem has been conceived, for many years, as a problem where you need to "guess" the actual system ToF**. However, a Bayesian approach to the failure prognostic problem asserts causality and, therefore, **from a Bayesian standpoint, the future is uncertain in nature; future cannot be characterized in deterministic way. The only thing we can do is not to guess when an event may occur ("ground truth" for that question does not even exist), but to characterize and measure future uncertainty as accurately as possible and make decisions based on risk**. This philosophy is actually induced by the rigorous mathematical framework presented in this manuscript that redefines the failure prognostic problem. The probability distribution $\mathcal{P}_A(k|y_{1:k_p})$ changes its variance as well as its expected value as a function of the uncertainty levels because the support of the ToF distribution does not have anything to do with the actual failure time. The most valuable information for decision-making purposes is included in $\mathcal{R}_A(k|y_{1:k_p})$, since it accounts for the risk of failure. The more uncertainty in the system, the more conservative this measure is.

Another important remark to consider is that as uncertainty in the system dynamics diminishes, these tend to evolve deterministically in time. As a consequence, $\mathcal{P}_A(k|y_{1:k_p})$ tends in turn to concentrate the probability mass around the time where the expected future system state crosses the failure threshold. This probability mass always approaches from the left (from previous time instants) since more uncertainty implies less knowledge about the system dynamics and these could yield to a system failure before it is expected. This fact confirms that under this new formalization for the failure prognostic problem there is consistency regarding the transition from a probabilistic to a deterministic framework and viceversa.

Why have prognostic algorithms "worked out" all this time then?

Under the conventional paradigm, and as it has been mentioned previously, the main goal in prognostic algorithm design was to "guess" the actual ToF. Prognostic algorithms have been validated so far in terms of their capability of accurately guessing the actual ToF for specific realizations of a stochastic process. This has led to think that it is appropriate to measure accuracy of an algorithm in terms of how similar the outcome of an experiment is with respect to the ToF conditional expectation. As a result, quality of ToF estimates depends heavily on the accuracy of the state transition model (which is the one who determines the expected future states trajectory), regardless of the manner in which the algorithm characterizes future uncertainty. Moreover, any prognostic algorithm would be validated as long as it interfered on the system dynamics in a negligible way.

6. CONCLUSIONS

In this paper we review the failure prognostic problem in the context of real-time systems monitoring from a probabilistic-based point of view. We have demonstrated that some expressions widely used to characterize the probability of failure were inconsistent, and that they are actually useless to perform validation of failure prognostic algorithms. In contrast, we present a rigorous mathematical development that was utilized to derive and demonstrate two theorems that embrace the definitions and hypotheses in a rightful manner, with the hope that the PHM community could use them, henceforth, to address the problem of failure prognostics in both discrete- and continuous-time dynamic systems. In addition, we compare both paradigms using the battery EoD time prognostics problem as an illustrative example. Philosophic implications arising from both approaches are explained and discussed,

leading to a proposal that aims at establishing a new paradigm regarding how the failure prognosis problem is understood.

The contributions of this manuscript are expected to become a theoretical basis and reference for future research in related topics as well as to promote the development of more rigorous procedures and standards for research work assessment.

An important remark that is not stated through the manuscript is that, although the mathematical developments presented hereby are explained in terms of future failure occurrence, these results can be also be understood from a more general perspective that is not limited to failures, but any future event that forces us to provide a probabilistic characterization of the system operational risk.

ACKNOWLEDGMENT

This work has been supported by FONDECYT Chile Grant Nr. 1170044, and the Advanced Center for Electrical and Electronic Engineering, AC3E, Basal Project FB0008, CONICYT. Also, the authors want to thank project CONICYT PIA ACT1405 and CONICYT-PFCHA/Doctorado Nacional/2017-21171097 for supporting David E. Acuña doctoral studies.

REFERENCES

Acuña, D. E., & Orchard, M. E. (2017). Particle-filtering-based failure prognosis via sigma-points: Application to lithium-ion battery state-of-charge monitoring. *Mech. Syst. Signal Process.*, *85*, 827-848.

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, *13*, 61-98.

Birolini, A. (2007). *Reliability engineering: Theory and practice*. New York: Springer-Verlag.

Burgos, C., Orchard, M. E., Kazerani, M., Cárdenas, R., & Sáez, D. (2016). Particle-filtering-based estimation of maximum available power state in lithium-ion batteries. *Applied Energy*, *161*, 349-363.

Chen, Z., & Zheng, S. (2005). Lifetime distribution based degradation analysis. *IEEE Transactions on Reliability*, *54*(1), 3-10.

Doucet, A., de Freitas, N., & Gordon, N. (Eds.). (2001). *Sequential monte carlo methods in practice*. New York: Springer-Verlag.

Girish, T., Lam, S. W., & Jayaram, S. R. (2003). Reliability prediction using degradation data - a preliminary study using neural network-based approach. *European Safety & Reliability International Conference (ESREL 2003)*, 681-688.

Kharoufeh, J. P., & Cox, S. M. (2005). Stochastic models for degradaton-based reliability. *IIE Transactions*, *37*, 533-542.

Lall, P., Pecht, M., & Harkim, E. (1997). *Influence of temperature on microelectronics and system reliability*. CRC Press, New York.

Pecht, M., Das, D., & Ramakrishnan, A. (2002). The ieee standards on reliability program and reliability prediction methods for electronic equipment. *Microelectronic Reliability*, *42*(9), 1259-1266.

Pola, D. A., Navarrete, H. F., Orchard, M. E., Rabié, R. S., Cerda, M. A., Olivares, B. E., ... Pérez, A. (2015). Particle-filtering-based discharge time prognosis for lithium-ion batteries with a statistical characterization of use profiles. *IEEE Transactions on Reliability*, *64*(2), 710-721.

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Evaluating prognostics performance for algorithms incorporating uncertainty estimates. *Proc. 2010 IEEE Aerospace Conf.*, 1-11.

Tang, L., Orchard, M. E., Goebel, K., & Vachtsevanos, G. (2011). Novel metrics and methodologies for the verification and validation of prognostic algorithms. *Proc. IEEE Aerosp. Conf.*, 5-12.

Vichare, N., Rodgers, P., Eveloy, V., & Pecht, M. (2004). In situ temperature measurement of a notebook computer - a case study of health and usage monitoring of electronics. *IEEE Transactions on Device and Materials Reliability*, *4*(4), 658-663.

Xu, D., & Zhao, W. (2005). Reliability prediction using multivariate degradation data. *Annual Reliability and Maintainability Symposium (RAMS 2005)*, 337-341.

Yang, K., & Xue, J. (1996). Continuous state reliability analysis. *1996 Annual Reliability and Maintainability Symposium (RAMS '96), International Symposium on Product Quality and Integrity*, 251-257.