UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

THEORETICAL FOUNDATIONS OF UNCERTAIN EVENT PROGNOSIS AND PREDICTIVE BAYESIAN CRAMÉR-RAO BOUNDS

TESIS PARA OPTAR AL GRADO DE DOCTOR EN INGENIERÍA ELÉCTRICA
EN COTUTELA CON LA UNIVERSIDAD DE NOTTINGHAM

DAVID ESTEBAN ACUÑA URETA

PROFESOR GUÍA:
MARCOS EDUARDO ORCHARD CONCHA

PROFESOR CO-GUÍA:
PATRICK WILLIAM WHEELER

MIEMBROS DE LA COMISIÓN:
FERNANDO AUAT CHEEIN
JORGE SILVA SÁNCHEZ
PERICLE ZANCHETTA

SANTIAGO DE CHILE
ABRIL 2020

RESUMEN EN ESPAÑOL DE LA TESIS PARA OPTAR
AL GRADO DE DOCTOR EN INGENIERÍA ELÉCTRICA
POR: DAVID ESTEBAN ACUÑA URETA
FECHA: ABRIL 2020
PROF. GUÍA: MARCOS EDUARDO ORCHARD CONCHA
PROF. CO-GUÍA: PATRICK WILLIAM WHEELER

THEORETICAL FOUNDATIONS OF UNCERTAIN EVENT PROGNOSIS AND PREDICTIVE BAYESIAN CRAMÉR-RAO BOUNDS

La confiabilidad y la continuidad operativa de los sistemas se han vuelto cada vez más importantes en la medida en que el progreso tecnológico se traduce en productos que van paulatinamente siendo incorporados y estandarizados en la sociedad. Algunos ejemplos típicos que se pueden nombrar son el uso de satélites para comunicaciones y ubicación, dispositivos de ingeniería biomédica como resonancia magnética, computadoras, maquinaria para procesos de manufactura como tornos o fresadoras, vehículos, defensa, entre otros.

Por la presente, se revisan los fundamentos probabilísticos del problema del pronóstico de fallas, presentando críticas constructivas ante las inconsistencias encontradas en las estrategias que han seguido muchos investigadores. Además, se desarrolla una formalización teóricamente rigurosa, presentando así nuevas distribuciones de probabilidad semi-cerradas para el primer tiempo de ocurrencia de cualquier tipo de evento (no necesariamente una falla) en sistemas dinámicos de tiempo discreto y continuo. También se introduce un nuevo concepto de "*evento incierto*", que generaliza el enfoque típico de cruce de umbral para declarar la ocurrencia de eventos. Esta generalización permite incertidumbre en la definición de eventos, haciéndolos inciertos. Estos nuevos conceptos se ilustran a través de un caso de estudio de pronóstico de crecimiento de fracturas por fatiga.

Debido a la falta de formalismo matemático con respecto al problema del pronóstico de fallas, se desarrollaron métodos heurísticos para evaluar la calidad de los resultados. En este sentido, otras dos contribuciones presentadas por el presente apuntan a abordar este problema de una manera más formal, usando *Bayesian Cramér-Rao Lower Bounds* (BCRLBs) para la *Error Covariance Matrix* (ECM) de estimaciones predictivas. Ambas contribuciones se ilustran a la luz del problema del pronóstico de tiempo *End-of-Discharge* (EoD) de baterías de Ion-Litio.

Dentro del problema de pronóstico, la incertidumbre de los estados del sistema se propaga en el tiempo, produciendo distribuciones de probabilidad que terminan caracterizando el primer tiempo de ocurrencia de eventos futuros. Por lo tanto, hay nuevas BCRLBs asociadas con el ECM de estados futuros del sistema (en cada instante de tiempo futuro), así como con el *Mean Squared Error* (MSE) del primer tiempo de ocurrencia de eventos futuros (el concepto de ECM reduce MSE en una dimensión). Las primeras (relacionadas con futuros estados del sistema) se utilizan para proponer una metodología de diseño paso-a-paso para ajustar los hiper-parámetros de algoritmos de pronóstico, lo que permite garantizar que los resultados no violen límites fundamentales de precisión. Sin embargo, las últimas (relacionadas con la primera ocurrencia de eventos futuros) se presentan y analizan en términos de uso potencial en el análisis y diseño de algoritmos de pronósticos.

RESUMEN EN INGLÉS DE LA TESIS PARA OPTAR
AL GRADO DE DOCTOR EN INGENIERÍA ELÉCTRICA
POR: DAVID ESTEBAN ACUÑA URETA
FECHA: ABRIL 2020
PROF. GUÍA: MARCOS EDUARDO ORCHARD CONCHA
PROF. CO-GUÍA: PATRICK WILLIAM WHEELER

THEORETICAL FOUNDATIONS OF UNCERTAIN EVENT PROGNOSIS AND PREDICTIVE BAYESIAN CRAMÉR-RAO BOUNDS

The reliability and operational continuity of systems have become increasingly important as technological progress translates into products that are gradually being incorporated and standardised in society. Some typical examples that can be named are the use of satellites for communications and location, biomedical engineering devices such as magnetic resonance, computers, machinery for manufacturing processes such as lathes or milling machines, vehicles, defense, among others.

Hereby, the probabilistic fundamentals of the failure prognosis problem are revisited, presenting constructive grievance to inconsistencies found in strategies that have been followed by many researchers. Moreover, a theoretically rigorous formalisation is developed, thus presenting new semi-closed probability distributions for the first occurrence time of any sort of event (not necessarily a failure) in both discrete- and continuous-time dynamic systems. A new concept of "*uncertain event*" is introduced as well, generalising the typical threshold-crossing approach to declare the occurrence of events. This generalisation allows uncertainty in the definition of events, making them uncertain. These new concepts are illustrated through a case study of fatigue crack growth prognosis.

Due to a lack of mathematical formalism with respect to the failure prognosis problem, heuristic methods were developed to assess quality of results. In this regard, another two contributions presented hereby aim at tackling this issue in a more formal fashion, namely, by using *Bayesian Cramér-Rao Lower Bounds* (BCRLBs) for the *Error Covariance Matrix* (ECM) of predictive estimates. Both contributions are illustrated in the light of the problem of *End-of-Discharge* (EoD) time prognosis of Lithium-Ion batteries.

Within the prognosis problem, uncertainty of system states is propagated over time, yielding probability distributions that end up characterising the first occurrence time of future events. Therefore, there are novel BCRLBs associated with the ECM of future system states (at each future time instant) as well as to the *Mean Squared Error* (MSE) of the first occurrence time of future events (the concept of ECM reduces to MSE in one dimension). The former (related to future system states) are used to propose a step-by-step design methodology to adjust prognostic algorithms hyper-parameters, permitting to guarantee that results do not violate fundamental precision bounds. The latter (related to the first occurrence time of future events), however, are presented and analysed in terms of potential use in the analysis and design of prognostic algorithms.

Para la Gloria de Dios Padre, Dios Hijo y Dios Espíritu Santo.

Acknowledgements

First of all, I solemnly acknowledge that all the achievements and handicaps faced to complete this thesis conform part of God's Glory; God the Father, God the Son, and God the Holy Spirit. One God in three Divine persons. I hereby declare that I am merely a human, neither more nor less than others in terms of dignity, chosen by God not on my own merits to complete this task according to his Providence.

Either in moments of strength or weakness, my family has been the main tangible love of God in my life. I would like to thank my father Ángel, my mother María, and my brothers Ángel and Daniel, for all their support and love.

I would like to express my deepest appreciation to Dr. Marcos E. Orchard, whom has been a real friend and outstanding mentor all along my undergraduate and postgraduate studies. He believed in me even when I could not do it myself. I wish I had someday his respectful and positive attitude with others, as well as his teaching and academic skills. He will always be my mentor, no matter what God's Providence led us to in the future.

Finally, I would like to thank Dr. Pat Wheeler for giving me the chance to have international experience in Nottingham, England. I received great blessings over there.

This work was funded by CONICYT-PFCHA/Doctorado Nacional/2017-21171097.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The elaboration of this thesis encompasses a transdisciplinary problem which corresponds to the prediction of events. There might be many reasons to want to anticipate the occurrence of various events depending on the context, of course. In agriculture, for example, climate prediction is important to get the most out of the crops; in business, on the other hand, the prediction of demand for products provides valuable information on logistics for the management and supply of stock. There are many other examples that could be mentioned, and for this same reason is that all the mathematical criticism included in this thesis embraces a general perspective. However, all the work presented in the following chapters is motivated and contextualised in a very important problem in engineering systems: future time in which these systems can present catastrophic failures (time instant in which they cease to be operational). The importance of this problem is enormous, and some examples of this are detailed below:

- Excessive expenditure on maintenance of equipment or components: Normally a lot of anticipation is made to reduce risks.
- Loss of production in the industry due to the detention of machinery: Either due to the frequency with which maintenance is performed or due to the occurrence of a failure and the time it takes for the respective repair.
- Risk of loss of human life: Failures in transport systems (airplanes, ships, cars, etc.) or medical equipment (pacemakers, artificial respirators, etc.) are critical and can have fatal outcomes.

The prediction of events is embraced from different approaches depending on the system to be studied and the information available. Suppose you want to study the remaining useful life of a system (which is equivalent to studying the time it fails in the future). There are systems that are very simple, such as pencils or screws, and therefore the unit-to-unit differences are practically negligible; it can be assumed that all the elaborated units are statistically identical. Adding this to the low marginal cost of production, several units can be used until they fail and the cycle life of each of them can be recorded to later perform statistics. Could the same be done with satellites, airplanes, vehicles or submarines? Obviously this is not possible because, in contrast to pencils or screws, these are complex (it is impossible to neglect the unit-to-unit differences, which immediately makes the hypothesis of statistical

identity theoretically infeasible) and very expensive (it is also infeasible to make millions of them fail, and so do statistics). What can be done in these cases then? The answer to this last question is that the particular condition of the system under study must be monitored and the way to do it depends on the information available.

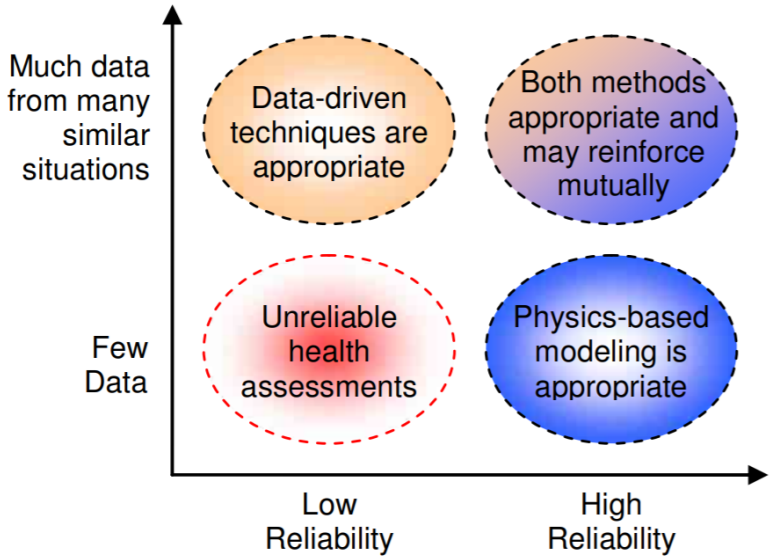


Figure 1.1: Illustration of how reliable would yield data-driven and physics-based approaches to health monitoring according to data availability [1]. Figure adapted from Inman *et al.* (2005) [2], p. 6.

When many historical data are available, as in the case of some industrial plants, it is recommended to use data-driven methods. Moreover, it is a great trend today due to the great interest that has arisen on *machine learning* and, more precisely, *deep learning* algorithms. However, in several other cases data is rather scarce, as in the case of satellites (usually no more than a single unit of the same type is produced). In this type of situation, it is essential to have an understanding of the physical phenomenology of the system which, if added to a good availability of measurement data (using sensors), can lead to reliable monitoring of the condition of the system under study. In Fig. 1.1 a scheme is shown that very clearly summarises the recommendation of the type of monitoring approach depending on the availability of data and physical models that account for the operability of the system in question.

Condition Monitoring

Condition monitoring tasks can be classified into two main fields: fault diagnosis and failure prognosis. Diagnostics, on the one hand, aims at assessing the current system health throughout a series of actions, including the detection of deviations from a desired operation pattern (anomaly detection), fault detection, fault isolation, and the assessment of the fault severity. Failure prognostics, on the other hand, consists on the extrapolation of system health indicators into the future, providing critical information about the risk of imminent catastrophic failures and thus helping to take preventive measures and maximise the system

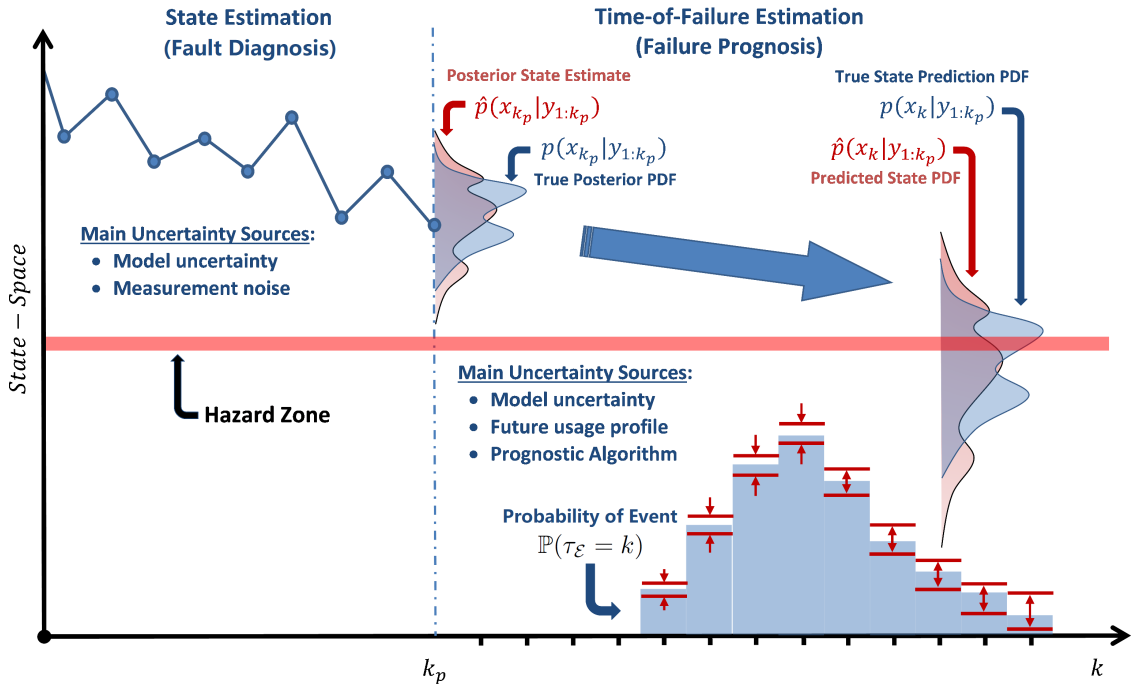


Figure 1.2: *The Failure Prognosis Problem.* Prognostics are executed in time k_p once an anomaly is detected during fault diagnosis. The current state posterior *Probability Density Function* (PDF) $p(x_{k_p}|y_{1:k_p})$, which yields from a set of collected measurements, is used as initial condition for a prognostic algorithm, which in turn propagates the uncertainty of the system state over time. Given the model dynamics, the sequence of PDFs for the system state in the future $\{p(x_k|y_{1:k_p})\}_{k>k_p}$ is fully determined by the system model. However, these PDFs are usually not tractable analytically, therefore prognostic algorithms aim at providing approximations whose performance depend on hyper-parameters that have to be tuned. A failure is typically declared when the system state gets into a particular region of the state-space, like under or above a threshold, for example. Considering the failure threshold depicted with an horizontal red line, $\mathbb{P}(\tau_{\mathcal{E}} = k)$ corresponds to the probability that an event $\mathcal{E}$ (failure condition in this case) was going to take place for the first time at the time instant $\tau_{\mathcal{E}} = k$. The aim of the failure prognosis problem is to compute the probability distribution of the random variable $\tau_{\mathcal{E}}$.

performance within a given prediction horizon. Please refer to Fig. 1.2 for a graphical illustration of condition monitoring. In aircraft applications, for example, prognostics algorithms need to provide information required to decide if an emergency landing is needed upon the detection of cracks in an airplane wing or in the blades of a turbine; in satellite applications, in contrast, the information provided by prognostic algorithms could help to decide whether to disturb a cubesat from its orbit (and make it fall to the Earth so that it does not remain in orbit as space junk) as a preventive measure or not; in industrial or military applications, however, prognostics could yield helpful for autonomous systems like a robot or a drone, where energetic constrains may lead to prioritising some tasks over others, so as to guarantee the fulfilment of the most important objectives and to allow its safe return to base [4].

Several failure prognostics techniques have been developed in the last decade, and the

selection of the most appropriate one depends on the particular application domain. Lately, machine learning and artificial intelligence approaches have become a trend in the scientific community. Unfortunately, because of their nature, the performance of data analytic tools relays strongly on the quality and availability of data. Additionally, failure data sets tend to be scarce, then it is usually necessary to focus on model-based approaches by using prior knowledge about the system behaviour, so as to design and implement proper predictive capabilities.

Prognostic algorithms may aim for either deterministic or probabilistic results in their analysis. The former approach (deterministic) aims at obtaining an estimate for the future specific time at which a particular event of interest (e.g., the *End-of-Life* (EoL) of a system) may take place [5]. Although these algorithms may seem to work well in some applications, it is important to note that the future is uncertain in nature. Deterministic approaches are thus solely useful in applications where incorrect (or unnecessary) preventive actions would only imply low-cost consequences (e.g., if a cellphone user needs to decide whether to charge it or not). Nonetheless, the scenario is completely different when catastrophic consequences can be expected due to inadequate corrective measures: the quantification of risk and low-probability events becomes tremendously valuable.

Prognostic results have been labeled as "validated" mostly using heuristic procedures due to the lack of standards and formalism for assessing algorithms performance. This lack of standards evidences the difficulties faced by the scientific community to agree on standardised performance metrics, since there has not been a consistent mathematical basis to sustain rigorous arguments. Specifically, the difficulties encountered in evaluating failure prognostic algorithms are due to the uncertainty associated with the failure time, and the number of experiments being too small to characterize its statistics. This is because it is expensive to experiment, so data is rather scarce, and also unit-to-unit differences weaken the hypothesis of the experiments being statistically identical. Compliance with these conditions is a requirement for metrics based on the Law of Large Numbers, for example.

1.1 Hypotheses

The findings presented in this thesis considered the following hypotheses:

1. The prognostic problem can be rigorously formalised using Probability Theory under a Bayesian perspective. Probability measures can be defined, given system dynamics and a proper characterisation of uncertainty sources.
2. Probability-based methods for event prognosis in nonlinear dynamic processes, and particularly Bayesian approaches, allow characterising the effect of the most significant sources of uncertainty affecting the accuracy and precision of risk estimates, conditional on acquired measurements.
3. Performance measures for prognostic algorithms can be built using the concept of *Bayesian Cramér-Rao Lower Bounds* (BCRLBs) for the *Error Covariance Matrix* (ECM) of predicted state trajectories and a probabilistic characterisation of many sources of uncertainty that affect their dynamics; including system initial conditions, model parameter uncertainty, measurement (sensor) uncertainty, and possible future operating

profiles.

4. It is possible to define guidelines and design procedures for prognostic algorithms by making use of performance measures that are inspired on the concept of BCRLB for the ECM of predicted state trajectories, ensuring a proper characterisation of the uncertainty related to the long-term prediction problem.
5. There is a BCRLB associated with the *Mean Squared Error* (MSE) incurred when estimating the first occurrence time of an event, which may be used to assess the performance of prognostic algorithms.

1.2 Objectives

This thesis focuses on theoretical aspects that help to significantly improve event prognosis frameworks currently available in literature. Among the objectives of this thesis, there are the following.

1.2.1 General Objectives

This research aims at providing a general and rigorous theoretical formalisation of the event prognosis problem by using Probability Theory, as well as at developing novel prognostic performance measures inspired on the concept of BCRLBs. The aforementioned metrics (or measures) must be capable of properly assessing the performance of implementations of prognostic approaches in terms of the following criteria, related to the random variable that characterises the first occurrence time of an event of interest: convergence to the true probability measure, precision and accuracy of the estimates.

1.2.2 Specific Objetives

1. To formalise the event prognosis problem in the light of Probability Theory by establishing a rigorous mathematical framework.
2. To prove the existence of fundamental limits for prognostic performance measures based on the Bayesian Cramér-Rao concept, conditional to a proper probabilistic characterisation of future operating profiles in specific applications.
3. To define guidelines and algorithm design procedures based on these measures, which should ensure proper characterisation of the uncertainty related to the long-term prediction problem.
4. To introduce rigour into the *Prognostics and Health Management* (PHM) community.

1.3 Contributions and Structure of the Thesis

The contributions of this thesis to the state-of-the-art are *enumerated* hereby accordingly with each of the next chapters; from Chapter 2 to Chapter 5. Brief descriptions of their contents are explained below:

- **Chapter 2:** Provided prognostics covers a wide range of scientific disciplines concerned with predicting, this chapter starts presenting the respective literature review in general

and then goes throughout the specific field of failure prognosis, which motivates the development of this thesis. Nonetheless, the literature review of this thesis is restricted to *probabilistic* approaches given the contributions presented in the next chapters. The first contribution of this thesis appears in this chapter and is the following:

1. *Criticism to the lack of rigour in failure prognosis, which has led to results that present philosophical contradictions with causality and Probability Theory. As another consequence, heuristic metrics and standards have been adopted. This motivates some of the next contributions that propose formal mathematical limits to address these issues.*

- **Chapter 3:** The aim of event prognosis is to characterise the probability distribution of the first time of occurrence of an event of interest (illustrated as $\tau_{\mathcal{E}}$ in Fig. 1.2). As it is shown in Chapter 2, most of the literature focuses on providing closed-form or numerically solvable expressions for specific types of stochastic processes where events are always triggered when crossing fixed thresholds. In the context of condition monitoring, closed-form expressions are mostly unsuitable as practical applications would hardly hold all the specific requirements, whereas numerically solvable expressions provide generality but are computationally expensive and infeasible in real-time applications. The contributions in this chapter are the following:

  2. *Analytical semi-closed expressions for calculating the first time of occurrence probability distribution of any type of event are derived rigorously using Probability Theory in both discrete- and continuous-time systems. These expressions are completely general and suitable for being used in real-time prognostic routines.*
  3. *The classical notion of threshold crossing that depicts the occurrence of events is generalised and, hence, a new concept of "uncertain event" is introduced. This concept allows the incorporation of uncertainty over the notion of event occurrence. Besides, it corresponds to the idea of "uncertain hazard zone" broadly known but never formalised in the PHM community.*

  All the concepts are illustrated throughout an example of fatigue crack growth prognosis. Additionally, this chapter sets the theoretical basis over which Chapters 4 and 5 are developed.

- **Chapter 4:** The general consensus is that better algorithms will be more accurate (related to biases in estimates) and more precise (related to variances in estimates). This idea sounds intuitive and natural. However, it is simple to arbitrarily "improve" the precision of an algorithm by modifying the model's hyper-parameters that define the state's evolution over time (state transition model). It is logical to wonder if there exists a fundamental limit for these "improvements". Could precise, although biased, time-of-failure estimates be generated? In terms of decision-making processes linked to maintenance scheduling, the latter could be catastrophic. The contributions in this chapter are the following:

  4. *In Fig. 1.2 it is shown that the PDF of future system states, $p(x_k|y_{1:k_p})$, is approximated by $\hat{p}(x_k|y_{1:k_p})$, which could in turn be obtained by employing a prognostic algorithm. The similarity between both PDFs depends on the hyper-parameters of the algorithm, and could be measured in terms of ECM. This chapter introduces predictive BCRLBs for the predicted states' ECMs, and uses them as fundamental limits restricting the space of feasible hyper-parameters for prognostic algorithms design. It is a lower bound for the ECM that yields for each $k > k_p$, thus con-*

forming a sequence of bounds.

5. *A new step-by-step design methodology is proposed to fine-tune discrete-time prognostic algorithms' hyper-parameters to ensure that the results obtained do not infringe fundamental precision limits.*
6. *Particle-filtering-based prognostic algorithms have been adopted as the "de facto" alternative to Monte Carlo (MC) simulations. It is the very first time one of these algorithms is fine-tuned by following a formal design methodology.*

For illustrative purposes, the contributions are employed in an application example of Lithium-Ion (Li-Ion) batteries, where the aim is to design prognostic algorithms to predict their *End-of-Discharge* (EoD) time.

- **Chapter 5:** This chapter is motivated by the same ideas of Chapter 4. However, in contrast to Chapter 4, the contribution explores a precision limits for the first time of occurrence of future events, and is the following:

  7. *New BCRLBs in both discrete- and the continuous-time systems for the MSE associated with the first occurrence time of future events, which is denoted as the random variable $\tau_{\mathcal{E}}$ in Fig. 1.2. It is a lower bound for the MSE obtained when estimating $\tau_{\mathcal{E}}$ and is related to the maximum precision achievable by prognostic algorithms.*

  Same as in Chapter 4, EoD time prediction of Li-Ion batteries is chosen as case study for illustrative purposes.

Finally, conclusions are presented in Chapter 6.

Chapter 2

Background and Literature Review

As it was stated in Chapter 1, prognostics are performed in a vast amount of scientific disciplines. This chapter is oriented to do a literature review regarding probabilistic approaches to event prognosis. It starts reviewing the problem from a transdisciplinary standpoint, and then carries on specifically with the field of failure prognosis, where some criticism and discussion is presented. Finally, a brief explanation and the state-of-the-art of Cramér-Rao bounds are provided at the end of the chapter, settling theoretical bases for Chapters 4 and 5.

2.1 Event Prognosis: A Transdisciplinary Problem

One of the reasons behind the use of mathematical models to describe the evolution of dynamic systems over time is to provide the means to foresee and anticipate possible future critical events. For this reason it is possible to use many different mathematical structures, and the "right" option for a model structure will depend largely on the specific application domain. Dynamic models including an explicit characterisation of sources of uncertainty (e.g., those involving stochastic equations) are particularly suitable for quantifying the risk associated with the occurrence of events as they provide a rigorous mathematical framework for the calculation of probability measures. In this regard, the conditions determining the occurrence of events have usually been specified as a "threshold", so that the event of interest is always triggered when the system states first exceed this threshold. Certainly, this implies the assumption that a deterministic function of the system states can always reflect the requirements needed to cause the occurrence of events. As the system states evolve randomly over time (i.e. the condition indicator is a stochastic process), a probability distribution for the first hitting time is thus induced: the *First-Passage Time* (FPT) [6–9] or *First-Hitting Time* (FHT) [10, 11] probability distribution. On the one hand, this definition is similar to *duration models* [12, 13] and is also analogous to *survival probability* in statistics [14–16] but interpreted in a different context. On the other hand, these principles are related to the *Remaining Useful Life* (RUL), *End-of-Life* (EoL), *Time-of-Failure* (ToF) and *Time-to-Failure* (TtF) probability distributions [17–19] in the engineering discipline of *Prognostics and Health Management* (PHM).

Efforts have been made to find analytical expressions for FPT probability distributions in many disciplines and fields of application such as chemistry [20, 21], physics [22, 23], biology [24, 25], neurobiology [26, 27], epidemiology [28], psychology [29], finance [30, 31], economy [32, 33], reliability theory [34, 35], among others [6, 7]. However, it is important to emphasise the fact that most of these research efforts focused on continuous-time [36–42] rather than discrete-time systems [32, 43–45] (with the exception of autoregressive models [44, 46–53]). In continuous-time systems, the FPT probability distribution is the solution to a particular *Stochastic Differential Equation* (SDE) with boundary conditions, which is generally solved with transformations [54–56] or eigenfunction expansions [37, 55] (mostly approximated numerically). Direct expression derivations are restricted to a few standard cases related to Brownian motion, such as in [57], and some other direct approximations [58–70]. While it may be natural to think that events occur when a variable (or condition indicator) that evolves in time reaches some threshold or region, in some cases it is not straightforward to determine a suitable value for this threshold. In this regard, and to the best of the author's knowledge, only a handful of contributions were intended to incorporate the concept of random thresholds [71–76] (and only for very specific types of stochastic processes).

One of the fundamental issues of interest in the modern engineering discipline PHM is the prediction of system failures (i.e., failure prognosis), where there is a clear distinction between "faults" (anomalous conditions under which the systems are still operational) and catastrophic failures (which implies total system inoperability). As a consequence, the *hazard zone* [77] concept emerged as an extension of the typical threshold point of view by defining a likelihood over the state-space in regions that suggest faulty conditions. There was an attempt to provide a more general perspective in [78, 79], but it was limited to Markov processes. The proposed probability measures also contained an underlying hypothesis of statistical independence; although these measures have still proved useful in defining a functional cost criterion for the design of the prognostic algorithms [80]. Furthermore, despite the fact that *hazard zones* are well-known and accepted in the PHM community [18], the current state-of-the-art formalisation of the failure prognosis problem [81, 82] still defines failure events with the classical deterministic threshold approach, is limited to events over Markov processes, lacks of mathematical demonstrations, and has resulted in inconsistencies in computing FPT probability distributions with methods other than those simulating complete system state trajectories [78].

2.2 Failure Prognosis

Critical failures of systems has been a subject of great interest which caught the attention of researchers for decades, especially from the military and industrial sectors. In consideration of the increase in the sophistication of infrastructure and facilities, as well as their operating costs due to faults, a scientific discipline was required that could address these issues in a systematic manner, including factors such as reliability, repair and safety. Meeting such requirements was the underlying motivation of the emergence of reliability engineering, with a clear aim at evaluating and demonstrating safety of components, equipment, and systems, and also reliability, maintainability, availability; all those throughout the research on novel methods and tools [83].

Reliability engineering adopts a frequentist standpoint [84–91], which is explained below. Given many statistically identical items (of specific kind of equipment, component, or system) operating from the same initial time, the probability of successfully fulfilling their design final goal along a specified time span, which corresponds to their reliability over time, can be approximated as the ratio of the number of successful items over the total amount items, due to the *Law of Large Numbers*.

Although in some specific cases the hypothesis of statistically identical items (or independent identically distributed items, i.i.d.) followed by reliability engineering may be recognised as a good approximation, it does not usually apply because it is not possible to make exact copies of an object at all; there are always unit-to-unit variations even in serial manufacturing. In addition, when considering series production, the marginal cost associated with each unit is usually small relative to, for example, satellite or aircraft manufacturing; in this respect, the volume must be limited to very small amounts. Indeed, because of the sophistication of such systems it seems not possible to find only two of them equal.

Prognostics and Health Management (PHM) is a modern engineering discipline that seeks to preserve the operational behaviour and functionality of systems within their specific characteristics in order to ensure their mission success, safety and efficiency. Distinguishing between two general divisions adopted by PHM is essential, which are fault diagnosis and failure prognosis (see Fig. 1.2). On the one hand, fault diagnosis is linked to system monitoring by detecting and isolating faults in a particular system. Once a failure has been identified, on the other hand, failure prognosis involves predicting the system's *Remaining Useful Life* (RUL) or *Time-of-Failure* (ToF) throughout the analysis of future system states along with characterising sources of uncertainty that influence their evolution over time.

Differences between reliability engineering and PHM rely on the paradigm from which systems are thought. In the first case (reliability engineering), prognostics provide knowledge on the behaviour of many brand new or newly defined but statistically identical systems that are viewed as samples, so it is based on historical data and classical objective probability; whereas in the second case (PHM) it is a single and specific system whose condition is especially tracked under Bayesian subjectivity, which differs from the former approach. In the latter case, monitoring by aiming at optimal maintenance (only when needed) leads to the well-known *Condition-Based Maintenance* (CBM) [92].

2.2.1 A Probabilistic Perspective

Over the past decades, multiple failure prognostic strategies have been developed and the choice of the most suitable one depends on the specific application domain. Under the tag of *data mining* and *predictive analytics*, machine learning and artificial intelligence methods have recently become a trend in industrial and academic communities. Unfortunately, the success of data analytics software is strongly related to data quality and availability due to their nature. When data quality is not sufficient or failure data sets are limited, relying on model-based solutions and using prior knowledge of system behaviour instead is essential to design and implement acceptable predictive capabilities.

In their analysis, prognostic algorithms may be targeted as either deterministic or probabilistic depending on their outcomes. The former (deterministic) approach attempts to

achieve an approximation for a specific future time (a scalar) at which a particular event of interest (e.g., a system's *End-of-Life* (EoL)) may occur. While in some implementations these algorithms may seem to work perfectly, it is important to note that the future is inherently uncertain. Therefore, deterministic methods are only effective in situations where inappropriate (or unnecessary) preventive measures would only conduce to low-cost implications (e.g. when a cell phone user has to consider whether or not to charge his/her device). Nevertheless, the situation becomes entirely different if catastrophic consequences can be foreseen as a result of deficient corrective measures: risk quantification of low-probability incidents becomes extremely valuable.

2.2.2 The Bayesian Approach

The manner in which uncertainty and risk are quantified is one of the most critical aspects to address in the design of decision-making processes. The final costs associated with all available choices can thus end up effected by these sources of uncertainty; costs that sometimes mean sacrificing operational continuity of industrial systems. In this regard, and from different standpoints, many methods intend to address the problem of characterisation of uncertainty. There are possibility-based or epistemic approaches, such as fuzzy logic or evidence theory [93–96]. Nonetheless, most researchers have favoured probability-based strategies when implementing real-time failure prognostic algorithms [97], since these techniques allow the notion of uncertainty to be included under a broadly known and accepted mathematical formulation.

Fault Diagnosis

Bayesian methods [98] appear as a suitable alternative for most real-time learning models in the context of probability theory and are widely used in fault diagnosis. Bayesian schemes allow filtering algorithms (also known as Bayesian processors) to be applied in nonlinear dynamic processes for uncertainty characterisation [99–101]. This task is achieved by determining the state vector's posterior estimates, where both prior knowledge (provided by the model of a system) and measurements (acquired in real-time) are used efficiently. Usually, in fault diagnosis these states refer to critical variables whose potential progression over time may substantially affect the health of the system, thereby experiencing a condition of failure. Nonetheless, it is not possible to obtain a closed-form solution for the filtering problem except for a small number of cases. When the involved system is linear and Gaussian, a well-known solution is the *Kalman Filter* (KF) [102, 103], which gives an optimal posterior estimate in the sense of minimising the *Mean Squared Error* (MSE). Despite the feasibility of approximating nonlinear models by linearised variants within bounded regions of the state-space, KF implementations usually provide suboptimal solutions. In addition, underlying uncertainties may vary from a Gaussian *Probability Density Function* (PDF). In this respect, *Sequential Monte Carlo* (SMC) methods, a.k.a. *Particle Filters* (PFs) [98], provide effective solutions to the problem of filtering in frameworks involving nonlinear non-Gaussian systems. PFs represent the underlying PDF through a sequencially updated set of weighted samples, generating empirical probability distributions from which statistical inference can be done.

Failure Prognosis

Though state estimation is commonly applied in fault diagnosis, it ends up playing an essential role in failure prognosis as well, because the result of the filtering phase defines the initial conditions for the problem of long-term predictions (refer to Fig. 1.2 for a graphical illustration). Although state estimation algorithms are needed to describe the system's current health status, other important aspects also have to be addressed in order to effectively implement prognostic algorithms. On the one hand, whenever a fault condition is detected, isolated, and identified at early stages preferably through the implementation of fault diagnosis methods, uncertainty must be effectively (and efficiently) propagated over time with the implementation of appropriate degradation models. On the other hand, future loading (or stress) profiles must be probabilistically characterised as well as the impact of those profiles on degradation processes that may occur.

The concern in prognosis is centered on understanding the future evolution of uncertainty [104], all to take preventive actions to avoid catastrophic events [77]. Even though this problem can be solved theoretically through Monte Carlo simulations [105], the disadvantage is the related computational cost that cannot be afforded in real-time applications. Particle-filtering-based algorithms [106] have been chosen by the PHM community as the *de facto* state-of-the-art technique in this matter, as they constitute efficient alternatives to Monte Carlo simulation. Prognostic algorithms based on PFs [77] have frequently been used as a standard choice in different applications [107] such as prediction of damage growth [108], analysis of failures in analog electronic circuits [109], design of fault-tolerant components [110], electrical machines' failure analysis [111], State-of-Charge and State-of-Health prognosis of batteries [3, 112–114], among others [115]. Nonetheless, there are some disadvantages to this class of algorithms. First, it does not consider exogenous factors that can impact long-term predictions. Second, PF-based prognostic methods usually use a fixed number of -occasionally equally weighted- samples randomly positioned in the space-state, and it is hard to ensure that these samples are suitable to adequately characterise the tails of future state PDFs as the horizon for prediction increases. Such considerations inspire the development of new prognostic algorithms and performance assessment measures that could overcome such limitations in design and implementation.

2.2.3 Lack of Formalism and Standardisation

Up to now, the evaluation of failure prognostic algorithms had generally followed standard procedures involving single run-to-failure experiments in which it has been assessed the ability to "guess" the actual failure time [116–119]. In this respect, it would be interesting to ask if the failure time reported in those experiments can be used as the "*ground truth*". The answer can differ significantly depending on whether it is referred to probabilistic or deterministic methods to predict failures. The result of deterministic methods is a number, a time of failure, so in those situations it can be concluded that the *ground truth* is the time of failure in experiments. Nonetheless, the output of probabilistic algorithms is a collection of probability distributions for the system's states; the uncertainty of the health indicator is characterised at specific future times (inducing a probability distribution of failure on the time axis) corresponding to each element of the collection. But how could the output of probabilistic algorithms be analysed when an experiment offers only one realisation of the

corresponding stochastic process? Would it be correct to compare a single realisation to a probability distribution (i.e., a number vs a PDF)?

Many prognostic measures [119] (not metrics, given that it is infeasible to prove triangular inequality in many cases) were conveniently described in terms of end-user demands: whereas a maintainer would concentrate on cost-benefit indicators that could help to reduce interruption, operators would be inclined towards precision and accuracy information that could help to advise actions to be taken while re-planning process operation (or mission) during an emergency. Alternatively, if a functional grouping is found, it should be understood that a set of measures focuses on computational performance and algorithm efficiency (precision, accuracy, trajectory, robustness), whereas others are oriented to conduct cost-benefit analysis (mean-time between replacement and failure, life-cycle cost, total value, technical value, return on investment).

2.2.4 Inconsistencies in Classic Formulations of The Failure Prognosis Problem

The problem of failure prognosis can be defined in several ways. However, the mainstay over which they are built is the *Time-of-Failure* (ToF) concept, typically defined as follows [81, 82]:

Definition 2.1 [Time-of-Failure (widely accepted definition)] The time of the first system failure τ_F is currently defined as

$$\tau_F(k_p) := \inf\{k \in \mathbb{N} : \{k > k_p\} \wedge \{\text{System Failure at } k\}\}. \tag{2.1}$$

Note that some other notions are available in the literature, such as *Remaining Useful Life* (RUL), *End-of-Life* (EoL), or *Time-to-Failure* (TtF), which in terms of mathematical implications are equivalent to ToF. Hence, below it is only addressed and analysed the consistency of the already introduced ToF definition.

Bayesian approaches [98] provide an appropriate solution for real-time monitoring of degradation processes while characterising uncertainty sources affecting them through a state-space representation. Indeed, let us consider a Markov process $\{X_k, Y_k\}_{k\in\mathbb{N}}$ and a time k_p, which denotes an initial time instant for prognostics. Probability distributions associated with ToF prediction have been computed for more than two decades as shown below:

$$\mathbb{P}(\tau_F \leq k) = \int_{\mathbb{X}} \mathbb{1}_{\mathbb{X}_F}(x_k) p(x_k|y_{1:k_p}) \mathrm{d}x_k, \tag{2.2}$$

with

$$\mathbb{P}(\tau_F = k) = \frac{1}{\gamma(X_{k_p+1:+\infty}|y_{1:k_p})} \big(\mathbb{P}(\tau_F \leq k) - \mathbb{P}(\tau_F \leq k-1)\big), \tag{2.3}$$

where $\mathbb{X}_F \subset \mathbb{X}$ is such that $\mathbb{X}_F = \{x \in \mathbb{X} : x \leq \underline{x}\}$, or, alternatively, $\mathbb{X}_F = \{x \in \mathbb{X} : x \geq \bar{x}\}$. Thus, the integral of Eq. (2.2) accounts for the probability mass of $p(x_k|y_{1:k_p})$ that crosses a failure threshold ($\underline{x}$ or $\bar{x}$) at time k. As it is known from basic Probability Theory, there is a normalising constant $\gamma(X_{k_p+1:+\infty}|y_{1:k_p})$ in Eq. (2.2) that should not be included if $\mathbb{P}(\tau_F \leq k)$

was well-defined. It had been included so far because otherwise probability axioms do not hold.

Let us concentrate on the following among the potential inconsistencies that may result from the above concept of ToF probability measures and the underlying philosophy embraced by the PHM community:

a) **Determinism:** Determinism is the first major incoherence. This notion has already been mentioned in Section 2.2.3: philosophically speaking, it has been related things that are entirely different in nature all along these years. The state-of-the-art has been developed in practice by utilising degradation and, subsequently, ToF data from only a few or single run-to-failure experiments. Such data sets are employed to validate probabilistic algorithms and mathematical tools. However, though these algorithms and tools provide probabilistic outcomes, data is deterministic in nature. Would it be appropriate to compare a single sample with a probability distribution? Naturally not. Scientists have tried to use statistics instead of probability distributions to prevent this fundamental problem.

b) **Non-causality:** It is stated in Eq. (2.3) that the ToF probability demands computing a normalisation constant. This requirement means that, in order to calculate $\mathbb{P}(\tau_F \leq k)$, it would be needed to describe the way the systems evolve till the "end of time", just to find the correct constant of normalisation. When this claim is closely examined, it means that this constant of normalisation is a function of time instants beyond k, which is inconsistent. Furthermore, assume it is of concern to quantify the risk of failure up to a certain fixed time $\overline{k}$, $\overline{k} > k_p$. Would it be right to think that events after time $\overline{k}$ have any kind of effect on what would have happened from k_p to $\overline{k}$? The response is "no" due to the causality of the system.

c) **Need of normalisation constants:** When the ToF probability distribution is obtained throughout the "usual" procedure, the need for a normalising constant implies that the result is not *per se* a probability distribution, but something else that is forced to fulfil axioms of probability. Normalisation is generally used to turn likelihood functions into probability distributions.

d) **Non-increasing Cumulative Distribution Functions:** It is worth noting that Eq. (2.3) was used to determine the ToF probability distribution on systems undergoing monotonic processes of degradation. Because the expression in Eq. (2.3) was interpreted as cumulative probability, it should be an increasing function of time. This property will not hold if the process of degradation is not monotonic. In addition, in the general case, it is possible to demonstrate empirically that this idea is ill-defined. As a side note, degradation processes with regeneration phenomena can be found relatively easily: consider, for instance, capacity regeneration phenomena in Lithium-Ion (Li-Ion) batteries.

Such statements indicate an ill-conditioned definition of ToF probability distribution, and the PHM community should acknowledge and examine them.

2.3 Bayesian Cramér-Rao Lower Bounds

The Cramér-Rao Lower Bound (CRLB) [120, 121] is a fundamental mathematical limit that lower bounds the *Error Covariance Matrix* (ECM) of any parameters vector estimator (ECM is equivalent to *Mean Squared Error* (MSE) in one dimension). Its conventional version was originally developed for performance evaluation of unbiased estimators in settings of unknown deterministic parameters. A Bayesian version of it was later developed by Van Trees, being now applicable to settings of random parameters estimation: the so called *Bayesian Cramér-Rao Lower Bound* (BCRLB) [122]. This later version does not require estimators to be unbiased.

2.3.1 The Bayesian Cramér-Rao bound

Let $x \in \mathbb{R}^{n_x}$ and $y \in \mathbb{R}^{n_y}$ be random vectors of parameters (to be estimated) and observations, respectively. Let also $\hat{x}(y)$ be an estimator of x conditional to the observations y. The *Bayesian Cramér-Rao inequality* [122] states that

$$\mathbb{E}_{p(x,y)}\{[\hat{x}(y) - x][\hat{x}(y) - x]^T\} \geq J^{-1} \tag{2.4}$$

where $p(x, y)$ is a joint PDF and J is known as the *Bayesian Information Matrix* (BIM) (called *Fisher Information Matrix* (FIM) in the conventional deterministic parameter estimation setting), which is defined as

$$J = \mathbb{E}_{p(x,y)}\{-\Delta_x^x \log p(x, y)\}, \tag{2.5}$$

where Δ denotes the second-order derivative operator

$$\Delta_x^y = \nabla_x \nabla_y{}^T, \tag{2.6}$$

and ∇ denotes the gradient operator.

Although the BCRLB was presented for parameter estimation, the same concepts can be conceived in state estimation, as shown below.

2.3.2 BCRLBs in Dynamic Systems

The state-space representation of systems that is used by Bayesian processors (a.k.a. Bayesian *filters*) [98] is specially suitable to characterise and quantify uncertainty of system states in a sequential fashion. Therefore, consider a Markov process $\{X_k, Y_k\}_{k\in\mathbb{N}}$ such that

$$x_k = f(x_{k-1}, \omega_{k-1}) \tag{2.7}$$
$$y_k = g(x_k, \nu_k), \tag{2.8}$$

where ω_k and v_k denote random vectors that are independent though not necessarily Gaussian.

There are four different versions of BCRLBs that can be used in discrete-time dynamic systems [123] that are presented below. Let us denote a collection of state vectors as $x_{0:k} = \begin{bmatrix} x_0{}^T & x_1{}^T & \dots & x_k{}^T \end{bmatrix}^T$, and let also $y_{1:k} = \begin{bmatrix} y_1{}^T & y_2{}^T & \dots & y_k{}^T \end{bmatrix}^T$ denote a collection

of measurement vectors, both up to time k. In addition, let $\hat{x}_k(y_{1:k})$ denote an estimator of x_k conditional to the measurements collection $y_{1:k}$ and, on the other hand, let $\hat{x}_{0:k}(y_{1:k})$ denote an estimator of the whole states collection $x_{0:k}$. There are four different Bayesian Cramér-Rao inequalities since the estimation could be *joint* or *marginal*, depending on whether it is oriented to estimate $x_{0:k}$ or x_k, respectively. At the same time, if there is a collection of measurements $y_{1:k-1}$ available, the estimation is said to be *conditional* to it, whereas it is said to be *unconditional* otherwise as measurements are accounted as random variables. The inequalities are shown below:

A. *Joint unconditional BCRLB*

$$\mathbb{E}_{p(x_{0:k},y_{1:k})}\{[\hat{x}_{0:k}(y_{1:k}) - x_{0:k}][\hat{x}_{0:k}(y_{1:k}) - x_{0:k}]^T\} \geq J_{0:k}^{-1} \tag{2.9}$$

$$J_{0:k}^{-1} = \mathbb{E}_{p(x_{0:k},y_{1:k})}\{-\Delta_{x_{0:k}}^{x_{0:k}} \log p(x_{0:k}, y_{1:k})\} \tag{2.10}$$

B. *Marginal unconditional BCRLB*

$$\mathbb{E}_{p(x_k,y_{1:k})}\{[\hat{x}_k(y_{1:k}) - x_k][\hat{x}_k(y_{1:k}) - x_k]^T\} \geq J_k^{-1} \tag{2.11}$$

$$J_k^{-1} = \mathbb{E}_{p(x_k,y_{1:k})}\{-\Delta_{x_k}^{x_k} \log p(x_k, y_{1:k})\} \tag{2.12}$$

C. *Joint conditional BCRLB*

$$\mathbb{E}_{p(x_{0:k},y_k|y_{1:k-1})}\{[\hat{x}_{0:k}(y_{1:k}) - x_{0:k}][\hat{x}_{0:k}(y_{1:k}) - x_{0:k}]^T\} \geq J_{0:k}(y_{1:k-1})^{-1} \tag{2.13}$$

$$J_{0:k}(y_{1:k-1})^{-1} = \mathbb{E}_{p(x_{0:k},y_k|y_{1:k-1})}\{-\Delta_{x_{0:k}}^{x_{0:k}} \log p(x_{0:k}, y_k|y_{1:k-1})\} \tag{2.14}$$

D. *Marginal conditional BCRLB*

$$\mathbb{E}_{p(x_k,y_k|y_{1:k-1})}\{[\hat{x}_k(y_{1:k}) - x_k][\hat{x}_k(y_{1:k}) - x_k]^T\} \geq J_k(y_{1:k-1})^{-1} \tag{2.15}$$

$$J_k(y_{1:k-1})^{-1} = \mathbb{E}_{p(x_k,y_k|y_{1:k-1})}\{-\Delta_{x_k}^{x_k} \log p(x_k, y_k|y_{1:k-1})\} \tag{2.16}$$

In order to avoid manipulation and recomputation of large matrices at each time instant k, a recursive way to compute the marginal unconditional BCRLB J_k^{-1} (see Eq. (2.12)) was introduced in [124], thus providing an elegant and efficient solution in the following manner:

$$J_{k+1} = D_k^{22} - D_k^{21}(J_k + D_k^{11})^{-1}D_k^{12}, \tag{2.17}$$

where

$$\begin{aligned}
D_k^{11} &= \mathbb{E}\{-\Delta_{x_k}^{x_k} \log p(x_{k+1}|x_k)\} && (2.18)\\
D_k^{12} &= \mathbb{E}\{-\Delta_{x_k}^{x_{k+1}} \log p(x_{k+1}|x_k)\} = (D_k^{21})^T && (2.19)\\
D_k^{22} &= \mathbb{E}\{-\Delta_{x_{k+1}}^{x_{k+1}} [\log p(x_{k+1}|x_k) + \log p(y_{k+1}|x_{k+1})]\} && (2.20)\\
&= D_k^{22,a} + D_k^{22,b}, && (2.21)
\end{aligned}$$

with expectations taken with respect to $p(x_{0:k+1}, y_{1:k+1})$. However, it is known that marginal unconditional BCRLBs consider measurements only as random variables. An elegant recursive way to compute the marginal conditional BCRLB $J_k(y_{1:k-1})^{-1}$ (see Eq. (2.16)) was

developed and introduced in [125] so as to include measurements acquired sequentially in time (typical optimal filtering settings):

$$J_{k+1}(y_{1:k}) = B_k^{22} - B_k^{21}(J_k^A(Y_k) + B_k^{11})^{-1}B_k^{12}, \tag{2.22}$$

where

$$\begin{aligned} B_k^{11} &= \mathbb{E}\{-\Delta_{x_k}^{x_k} \log p(x_{k+1}|x_k)\} && (2.23)\\ B_k^{12} &= \mathbb{E}\{-\Delta_{x_k}^{x_{k+1}} \log p(x_{k+1}|x_k)\} = (B_k^{21})^T && (2.24)\\ B_k^{22} &= \mathbb{E}\{-\Delta_{x_{k+1}}^{x_{k+1}} [\log p(x_{k+1}|x_k) + \log p(y_{k+1}|x_{k+1})]\} && (2.25)\\ &= B_k^{22,a} + B_k^{22,b}, && (2.26) \end{aligned}$$

with expectations taken with respect to $p(x_{0:k+1}, y_{k+1}|y_{1:k})$. The $n_x \times n_x$ matrix $J_k^A(y_{1:k})$ denotes an auxiliary BIM matrix for x_k whose inverse corresponds exactly to the lower-right block of the inverse of the auxiliary BIM matrix $I_k^A(y_{1:k})$, where

$$I_k^A(y_{1:k}) = \mathbb{E}_{p(x_{0:k}|y_{1:k})}\{-\Delta_{x_{0:k}}^{x_{0:k}} \log p(x_{0:k}|y_{1:k})\}. \tag{2.27}$$

Chapter 3

Uncertain Event Prognosis

In this chapter, the classical deterministic threshold crossing approach with which events are typically triggered is generalised to a probabilistic notion of *uncertain event*, analogous to that of *uncertain hazard zone* in PHM [77]. A mathematically rigorous formalisation is provided for the first time of occurrence of uncertain future events characterised by discrete- and continuous-time stochastic processes. For the corresponding probability measures, explicit semi-closed expressions are derived and demonstrated throughout Section 3.1. In addition to the aforementioned theoretical contribution, Section 3.2 concentrates on discrete-time applications by showing how to compute such probability measures and validate the mentioned procedures with results compared to those obtained by using Monte Carlo simulations as a benchmark; all in the scope of an illustrative fatigue crack growth prognosis problem. Finally, Section 3.3 summarises the conclusions of this chapter.

3.1 Probability of Uncertain Future Events

Let us consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable space $(\mathbb{X}, \Sigma)$. Also, let $X : \mathbb{T} \cup \{0\} \times \Omega \to \mathbb{X}$, $\mathbb{T} \in \{\mathbb{N}, \mathbb{R}_+\}$, be a stochastic process; and F_{X_τ} denote the respective probability measure in $(\mathbb{X}, \Sigma)$ induced by X_τ, $\tau \in \mathbb{T}$.

Definition 3.1 [Uncertain Event Process & Likelihood] Let $\mathcal{E}$ denote an event of interest. An **uncertain event process** is defined as a function $E : \mathbb{T} \times \mathbb{X} \to \{\mathcal{E}, \mathcal{E}^c\}$ such that

$$\mathbb{P}(E_\tau = \mathcal{E}) = \int_\Omega \mathbb{P}(E_\tau = \mathcal{E}, X_\tau(\omega))\mathrm{d}\mathbb{P}(\omega) \tag{3.1}$$

$$= \int_{\mathbb{X}_\tau} \mathbb{P}(E_\tau = \mathcal{E}|x_\tau)\mathrm{d}F_{X_\tau}(x_\tau), \ \forall \tau \in \mathbb{T}. \tag{3.2}$$

The likelihood $\mathbb{P}(E_\tau = \mathcal{E}|x) = 1 - \mathbb{P}(E_\tau = \mathcal{E}^c|x)$, as a function of $x \in \mathbb{X}$ and $\tau \in \mathbb{T}$, is called **uncertain event likelihood function**.

Therefore, an uncertain event process $\{E_\tau\}_{\tau \in \mathbb{T}}$ represents a binary random variable evolv-

ing in time associated with occurrence of an uncertain event whose statistics depend on those of a stochastic process $\{X_\tau\}_{\tau\in\mathbb{T}\cup\{0\}}$ evaluated at the same time instant. Indeed, provided $\mathrm{i}, j \in \mathbb{T}$, $\mathrm{i} \neq j$, then

$$\mathbb{P}(E_\mathrm{i}, E_j|\{X_\tau\}_{\tau\in\mathbb{T}}) = \mathbb{P}(E_\mathrm{i}|X_\mathrm{i})\mathbb{P}(E_j|X_j). \tag{3.3}$$

In general, E_τ is independent from any other parameter as long as it is conditioned on X_τ, $\forall\tau \in \mathbb{T}$. This is a very important property that is used later.

Remark 3.1 [Particular Case: Threshold] According to Definition 3.1, expressing $\mathbb{P}(E_\tau = \mathcal{E}\big|x)$ as a function of $x \in \mathbb{X}$ corresponds exactly to a probabilistic description of the occurrence of an uncertain event $\mathcal{E}$. By assuming $\mathbb{X} = \mathbb{R}$, for example, and defining $\mathbb{P}(E_\tau = \mathcal{E}|x) = \mathbb{1}_{\mathbb{X}_\mathcal{E}}(x)$, with $\mathbb{X}_\mathcal{E} = \{x \in \mathbb{R} : x > c, c \in \mathbb{R}\}$, it is obtained the classical threshold crossing event setting studied so far in the literature, where $E_\tau(x)$ conditional to a fixed $x \in \mathbb{R}$ is no longer a random variable:

$$E_\tau(x) = \begin{cases} \mathcal{E}, & x \in \mathbb{X}_\mathcal{E} \\ \mathcal{E}^c, & \sim . \end{cases} \tag{3.4}$$

Remark 3.2 [Uncertain Hazard Zone] The *uncertain event likelihood function* $\mathbb{P}(E_\tau = \mathcal{E}|x)$ only as a function of $x \in \mathbb{X}$ is exactly what is conceived as *uncertain hazard zone* [77] in the PHM discipline.

Now, let us introduce a formal definition for the first occurrence time of an *uncertain event*.

Definition 3.2 [First Occurrence Time of an Event] Let $\{X_\tau\}_{\tau\in\mathbb{T}\cup\{0\}}$ be a stochastic process and $\{E_\tau\}_{\tau\in\mathbb{T}}$ be an uncertain event process, respectively. The first time of occurrence of an event $\mathcal{E}$ after a time instant $\tau_p \in \mathbb{T}\cup\{0\}$ is defined as

$$\tau_\mathcal{E}(\tau_p) := \inf\{\tau \in \mathbb{T} : \{\tau > \tau_p\} \wedge \{E_\tau = \mathcal{E}\}\}. \tag{3.5}$$

With these few definitions, semi-closed expressions of the probability distribution associated with $\tau_\mathcal{E}$ can be formally derived for both discrete- and continuous-time stochastic processes in a general manner as follows.

3.1.1 Discrete-Time Stochastic Processes

Let $\{X_k\}_{k\in\mathbb{N}\cup\{0\}}$ be a stochastic process and $\{E_k\}_{k\in\mathbb{N}}$ be an uncertain event process. The probability mass function associated with $\tau_\mathcal{E} = \tau_\mathcal{E}(k_p)$, $k_p \in \mathbb{N}\cup\{0\}$, is obtained below, which in turn coincides in structure with an expression found in *survival probability*:

$$\mathbb{P}(\tau_\mathcal{E} = k) := \mathbb{P}\left(\{E_k = \mathcal{E}\}, \{E_j = \mathcal{E}^c\}_{j=k_p}^{k-1}\right) \tag{3.6}$$

$$= \mathbb{P}\left(E_k = \mathcal{E}\big|\{E_j = \mathcal{E}^c\}_{j=k_p}^{k-1}\right)\mathbb{P}\left(\{E_j = \mathcal{E}^c\}_{j=k_p}^{k-1}\right) \tag{3.7}$$

$$\vdots$$

$$= \mathbb{P}\left(E_k = \mathcal{E} \middle| \{E_j = \mathcal{E}^c\}_{j=k_p}^{k-1}\right) \prod_{j=k_p+1}^{k-1} \mathbb{P}\left(E_j = \mathcal{E}^c \middle| \{E_{\mathrm{i}} = \mathcal{E}^c\}_{\mathrm{i}=k_p}^{j-1}\right) \cancelto{1}{\mathbb{P}\left(E_{k_p} = \mathcal{E}^c\right)} \tag{3.8}$$

$$= \mathbb{P}\left(E_k = \mathcal{E} | \tau_{\mathcal{E}} \geq k\right) \prod_{j=k_p+1}^{k-1} \mathbb{P}\left(E_j = \mathcal{E}^c | \tau_{\mathcal{E}} \geq j\right) \tag{3.9}$$

Conversely, $\mathbb{P}(\tau_{\mathcal{E}} = k)$ according to [32], defined below but under the generalised notion of *uncertain event* given in Definition 3.1, can be expressed in recursive form. If an event occurs at time k, it implies that $k_p < \tau_{\mathcal{E}} \leq k$, and by the *Law of Total Probability* it can be obtained

$$\mathbb{P}(E_k = \mathcal{E}) = \sum_{j=k_p+1}^{k} \mathbb{P}(E_k = \mathcal{E} | \tau_{\mathcal{E}} = j) \mathbb{P}(\tau_{\mathcal{E}} = j) \tag{3.10}$$

$$= \mathbb{P}(E_k = \mathcal{E} | \tau_{\mathcal{E}} = k) \mathbb{P}(\tau_{\mathcal{E}} = k) + \sum_{j=k_p+1}^{k-1} \mathbb{P}(E_k = \mathcal{E} | \tau_{\mathcal{E}} = j) \mathbb{P}(\tau_{\mathcal{E}} = j) \tag{3.11}$$

Hence, given that $\mathbb{P}(E_k = \mathcal{E} | \tau_{\mathcal{E}} = k) = 1$, it yields

$$\mathbb{P}(\tau_{\mathcal{E}} = k) = \mathbb{P}(E_k = \mathcal{E}) - \sum_{j=k_p+1}^{k-1} \mathbb{P}(E_k = \mathcal{E} | \tau_{\mathcal{E}} = j) \mathbb{P}(\tau_{\mathcal{E}} = j) \tag{3.12}$$

Let us now prove the equivalence with the following lemma of the two probability distributions presented above in Eqs. (3.9) and (3.12).

Lemma 3.1 Let $\{X_k\}_{k \in \mathbb{N} \cup \{0\}}$ be a stochastic process and $\{E_k\}_{k \in \mathbb{N}}$ be an uncertain event process, respectively. Let also $\tau_{\mathcal{E}} = \tau_{\mathcal{E}}(k_p)$, $k_p \in \mathbb{N} \cup \{0\}$. The mapping $\mathbb{P}(\tau_{\mathcal{E}} = \cdot) : \mathbb{N} \to [0, 1]$ can be either defined as

$$\mathbb{P}(\tau_{\mathcal{E}} = k) := \mathbb{P}\left(E_k = \mathcal{E} | \tau_{\mathcal{E}} \geq k\right) \prod_{j=k_p+1}^{k-1} \mathbb{P}\left(E_j = \mathcal{E}^c | \tau_{\mathcal{E}} \geq j\right) \tag{3.13}$$

as well as

$$\mathbb{P}(\tau_{\mathcal{E}} = k) := \mathbb{P}(E_k = \mathcal{E}) - \sum_{j=k_p+1}^{k-1} \mathbb{P}(E_k = \mathcal{E} | \tau_{\mathcal{E}} = j) \mathbb{P}(\tau_{\mathcal{E}} = j). \tag{3.14}$$

Proof.

Using Eq. (3.14), Eq. (3.13) can be obtained as shown below

$$\mathbb{P}(\tau_{\mathcal{E}} = k) = \mathbb{P}(E_k = \mathcal{E}) - \sum_{j=k_p+1}^{k-1} \mathbb{P}(E_k = \mathcal{E} | \tau_{\mathcal{E}} = j) \mathbb{P}(\tau_{\mathcal{E}} = j) \tag{3.15}$$

$$= \mathbb{P}(E_k = \mathcal{E}) - \sum_{j=k_p+1}^{k-1} \mathbb{P}(\tau_\mathcal{E} = j | E_k = \mathcal{E}) \mathbb{P}(E_k = \mathcal{E}) \tag{3.16}$$

$$= \mathbb{P}(E_k = \mathcal{E}) \left(1 - \sum_{j=k_p+1}^{k-1} \mathbb{P}(\tau_\mathcal{E} = j | E_k = \mathcal{E}) \right) \tag{3.17}$$

$$= \mathbb{P}(E_k = \mathcal{E}) \left(1 - \mathbb{P}(\tau_\mathcal{E} < k | E_k = \mathcal{E}) \right) \tag{3.18}$$

$$= \mathbb{P}(E_k = \mathcal{E}) \mathbb{P}(\tau_\mathcal{E} \geq k | E_k = \mathcal{E}) \tag{3.19}$$

$$= \mathbb{P}(E_k = \mathcal{E} | \tau_\mathcal{E} \geq k) \mathbb{P}(\tau_\mathcal{E} \geq k) \tag{3.20}$$

Nevertheless, note that

$$\mathbb{P}(\tau_\mathcal{E} \geq k) = 1 - \mathbb{P}(\tau_\mathcal{E} < k) \tag{3.21}$$

$$= 1 - \mathbb{P}(\tau_\mathcal{E} < k-1) - \mathbb{P}(\tau_\mathcal{E} = k-1) \tag{3.22}$$

$$= \mathbb{P}(\tau_\mathcal{E} \geq k-1) - \mathbb{P}(\tau_\mathcal{E} = k-1) \tag{3.23}$$

$$= \mathbb{P}\left(\{E_j = \mathcal{E}^c\}_{j=k_p}^{k-2} \right) - \mathbb{P}\left(\{E_{k-1} = \mathcal{E}\}, \{E_j = \mathcal{E}^c\}_{j=k_p}^{k-2} \right) \tag{3.24}$$

$$= \mathbb{P}\left(\{E_j = \mathcal{E}^c\}_{j=k_p}^{k-2} \right) - \mathbb{P}\left(E_{k-1} = \mathcal{E} \middle| \{E_j = \mathcal{E}^c\}_{j=k_p}^{k-2} \right) \mathbb{P}\left(\{E_j = \mathcal{E}^c\}_{j=k_p}^{k-2} \right) \tag{3.25}$$

$$= \mathbb{P}\left(\{E_j = \mathcal{E}^c\}_{j=k_p}^{k-2} \right) \left(1 - \mathbb{P}\left(E_{k-1} = \mathcal{E} \middle| \{E_j = \mathcal{E}^c\}_{j=k_p}^{k-2} \right) \right) \tag{3.26}$$

$$= \mathbb{P}(\tau_\mathcal{E} \geq k-1) \mathbb{P}\left(E_{k-1} = \mathcal{E}^c \middle| \tau_\mathcal{E} \geq k-1 \right) \tag{3.27}$$

Since Eq. (3.27) shows that $\mathbb{P}(\tau_\mathcal{E} \geq k)$ can be obtained using $\mathbb{P}(\tau_\mathcal{E} \geq k-1)$, by iterating this result it yields

$$\mathbb{P}(\tau_\mathcal{E} \geq k) = \prod_{j=k_p+1}^{k-1} \mathbb{P}\left(E_j = \mathcal{E}^c | \tau_\mathcal{E} \geq j \right) \tag{3.28}$$

□

Please note the following before presenting the preceding results in a formal theorem. Considering that each E_k depends on X_k, the concept of *uncertain event likelihood function* introduced in Definition 3.1 arises below by using the property illustrated with Eq. (3.3):

$$\mathbb{P}(\tau_\mathcal{E} = k) = \mathbb{P}\left(E_k = \mathcal{E} | \tau_\mathcal{E} \geq k \right) \prod_{j=k_p+1}^{k-1} \mathbb{P}\left(E_j = \mathcal{E}^c | \tau_\mathcal{E} \geq j \right) \tag{3.29}$$

$$= \int_{\mathbb{X}_{k_p+1:k}} \mathbb{P}\left(E_k = \mathcal{E} | \{\tau_\mathcal{E} \geq k\}, x_{k_p+1:k} \right) \prod_{j=k_p+1}^{k-1} \mathbb{P}\left(E_j = \mathcal{E}^c | \{\tau_\mathcal{E} \geq j\}, x_{k_p+1:k} \right) \mathrm{d}F_{X_{k_p+1:k}}(x_{k_p+1:k}) \tag{3.30}$$

$$= \int_{\mathbb{X}_{k_p+1:k}} \mathbb{P}\left(E_k = \mathcal{E} | x_k \right) \prod_{j=k_p+1}^{k-1} \mathbb{P}\left(E_j = \mathcal{E}^c | x_j \right) \mathrm{d}F_{X_{k_p+1:k}}(x_{k_p+1:k}). \tag{3.31}$$

This expression is used in Section 3.2 to introduce a procedure for estimating such probabilities based on numerical methods.

Theorem 3.1 [First Event Time in Stochastic Processes] Let $\{X_k\}_{k\in\mathbb{N}\cup\{0\}}$ be a stochastic process and $\{E_k\}_{k\in\mathbb{N}}$ be an uncertain event process, respectively. If the first time of occurrence of the event $\mathcal{E}$, $\tau_{\mathcal{E}} = \tau_{\mathcal{E}}(k_p)$, with $k_p \in \mathbb{N}\cup\{0\}$, is such that $\tau_{\mathcal{E}} < +\infty$, $\mathbb{P}-a.s.$, then the mapping $\mathbb{P}(\tau_{\mathcal{E}} = \cdot) : \mathbb{N} \to [0,1]$ exists and is well-defined in terms of its *uncertain event likelihood function* as

$$\mathbb{P}(\tau_{\mathcal{E}} = k) := \int_{\mathbb{X}_{k_p+1:k}} \mathbb{P}\left(E_k = \mathcal{E}|x_k\right) \prod_{j=k_p+1}^{k-1} \mathbb{P}\left(E_j = \mathcal{E}^c|x_j\right) \mathrm{d}F_{X_{k_p+1:k}}(x_{k_p+1:k}). \tag{3.32}$$

Therefore,

$$\mathbb{P}_{\mathcal{E}}(A) = \sum_{k\in A} \mathbb{P}(\tau_{\mathcal{E}} = k),\ \forall A \in 2^{\mathbb{N}}, \tag{3.33}$$

is a probability measure that defines the probability space $(\mathbb{N}, 2^{\mathbb{N}}, \mathbb{P}_{\mathcal{E}})$. Indeed, the following conditions hold:

1) $\mathbb{P}_{\mathcal{E}}(\mathbb{N}) = 1$.
2) $0 \le \mathbb{P}_{\mathcal{E}}(A) \le 1$, $\forall A \in 2^{\mathbb{N}}$.
3) $\mathbb{P}_{\mathcal{E}}(\cup_{k\in\mathbb{N}} A_k) = \sum_{k\in\mathbb{N}} \mathbb{P}_{\mathcal{E}}(A_k)$, $\forall \{A_k \in 2^{\mathbb{N}}\}_{k\in\mathbb{N}}$, with $A_{\mathrm{i}} \cap A_j = \phi$, $\forall \mathrm{i} \neq j$.

Proof.

1) Let us define $\{A_k\}_{k\in\mathbb{N}}$, $A_k = \{1,\ldots,k\}$, such that $A_k \nearrow \mathbb{N}$

$$\mathbb{P}_{\mathcal{E}}(A_k) = \sum_{j=1}^{k} \mathbb{P}(\tau_{\mathcal{E}} = j) = \mathbb{P}(\tau_{\mathcal{E}} < k+1) \tag{3.34}$$

$$\Rightarrow \lim_{k\to+\infty} \mathbb{P}_{\mathcal{E}}(A_k) = \lim_{k\to+\infty} \mathbb{P}(\tau_{\mathcal{E}} < k+1) \tag{3.35}$$

$$\Rightarrow \mathbb{P}_{\mathcal{E}}(\mathbb{N}) = \mathbb{P}(\tau_{\mathcal{E}} < +\infty) = 1, \tag{3.36}$$

due to the continuity property of probability measures and because $\tau_{\mathcal{E}} < +\infty$, $\mathbb{P}-a.s.$

2) By definition, because

$$0 \le \mathbb{P}(\tau_{\mathcal{E}} = k),\ \forall k \in \mathbb{N} \ \Rightarrow\ 0 \le \mathbb{P}_{\mathcal{E}}(A),\ \forall A \in 2^{\mathbb{N}}, \tag{3.37}$$

and, on the other hand,

$$A \subseteq \mathbb{N} \ \Rightarrow\ \sum_{k\in A} \mathbb{P}(\tau_{\mathcal{E}} = k) \le \sum_{k\in\mathbb{N}} \mathbb{P}(\tau_{\mathcal{E}} = k) \tag{3.38}$$

$$\Leftrightarrow\ \mathbb{P}_{\mathcal{E}}(A) \le \mathbb{P}_{\mathcal{E}}(\mathbb{N}) = 1,\ \forall A \in 2^{\mathbb{N}}. \tag{3.39}$$

3) Let $\{A_k \in 2^{\mathbb{N}}\}_{k\in\mathbb{N}}$ such that $A_{\mathrm{i}} \cap A_j = \phi$, $\forall \mathrm{i} \neq j$. By definition,

$$\mathbb{P}_{\mathcal{E}}\left(\cup_{k\in\mathbb{N}} A_k\right) = \sum_{j\in\cup_{k\in\mathbb{N}}A_k} \mathbb{P}(\tau_{\mathcal{E}} = j) \tag{3.40}$$

$$= \sum_{k\in\mathbb{N}} \sum_{j\in A_k} \mathbb{P}(\tau_{\mathcal{E}} = j) \tag{3.41}$$

$$= \sum_{k\in\mathbb{N}} \mathbb{P}_{\mathcal{E}}(A_k) \tag{3.42}$$

□

3.1.2 Continuous-Time Stochastic Processes

Let $\{X_t\}_{t\in\mathbb{R}_+\cup\{0\}}$ be a stochastic process and $\{E_t\}_{t\in\mathbb{R}_+}$ be an uncertain event process. By definition, if there was a probability density function associated with $\tau_{\mathcal{E}} = \tau_{\mathcal{E}}(t_p)$, $t_p \in \mathbb{R}_+ \cup \{0\}$, then it could be obtained as shown below, which in turn coincides in structure with an expression found in *survival probability*:

$$p(\tau_{\mathcal{E}} = t) := \mathbb{P}\left(\{E_t = \mathcal{E}\}, \{E_\tau = \mathcal{E}^c\}_{\tau\in(t_p,t)}\right) \tag{3.43}$$

$$= \mathbb{P}\left(\{E_t = \mathcal{E}\}, \{\tau_{\mathcal{E}} \geq t\}\right) \tag{3.44}$$

$$= \mathbb{P}\left(E_t = \mathcal{E} | \tau_{\mathcal{E}} \geq t\right) \mathbb{P}\left(\tau_{\mathcal{E}} \geq t\right) \tag{3.45}$$

with $\mathbb{P}(\tau_{\mathcal{E}} \geq t) = 1 - \mathbb{P}(\tau_{\mathcal{E}} < t)$. Let $\mathcal{B}(\mathbb{R}_+)$ and $\lambda = \lambda(\mathbb{R}_+)$ denote the Borel σ-algebra and Lebesgue measure in $\mathbb{R}_+$, respectively. Let also $F_{\tau_{\mathcal{E}}}(t) = \mathbb{P}(\tau_{\mathcal{E}} \leq t)$ be a probability measure in the measurable space $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ such that $F_{\tau_{\mathcal{E}}} << \lambda$. According to the Theorem of Radon-Nikodym, there is a probability density function $p(\tau_{\mathcal{E}} = t) := \frac{\mathrm{d}F_{\tau_{\mathcal{E}}}}{\mathrm{d}\lambda}(t)$, $t \in \mathbb{R}_+$, such that

$$\mathbb{P}(\tau_{\mathcal{E}} < t) = \int_{(t_p,t)} \mathrm{d}F_{\tau_{\mathcal{E}}}(\tau) \tag{3.46}$$

$$= \int_{(t_p,t)} p(\tau_{\mathcal{E}} = \tau)\mathrm{d}\tau \tag{3.47}$$

$$= \int_{(t_p,t)} \mathbb{P}\left(E_\tau = \mathcal{E} | \tau_{\mathcal{E}} \geq \tau\right) \mathbb{P}\left(\tau_{\mathcal{E}} \geq \tau\right) \mathrm{d}\tau \tag{3.48}$$

Given that the aforementioned probability density function exists, $\mathbb{P}(\tau_{\mathcal{E}} \geq t)$ has to be differentiable:

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(\tau_{\mathcal{E}} \geq t) = -\mathbb{P}\left(E_t = \mathcal{E} | \tau_{\mathcal{E}} \geq t\right) \mathbb{P}\left(\tau_{\mathcal{E}} \geq t\right), \tag{3.49}$$

because $\mathbb{P}\left(E_{t_p} = \mathcal{E} | \tau_{\mathcal{E}} \geq t_p\right) = \mathbb{P}\left(E_{t_p} = \mathcal{E}\right) = 0$ ($\tau_{\mathcal{E}} > t_p$, by definition of $\tau_{\mathcal{E}}$). By integrating over time:

$$-\int_{(t_p,t)} \mathbb{P}\left(E_\tau = \mathcal{E} | \tau_{\mathcal{E}} \geq \tau\right) \mathrm{d}\tau = \int_{(t_p,t)} \frac{1}{\mathbb{P}\left(\tau_{\mathcal{E}} \geq \tau\right)} \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbb{P}\left(\tau_{\mathcal{E}} \geq \tau\right) \mathrm{d}\tau \tag{3.50}$$

$$= \int_{(t_p,t)} \frac{\mathrm{d}}{\mathrm{d}\tau} \log \mathbb{P}\left(\tau_{\mathcal{E}} \geq \tau\right) \mathrm{d}\tau \tag{3.51}$$

$$= \log \mathbb{P}\left(\tau_{\mathcal{E}} \geq t\right) - \cancelto{0}{\log \mathbb{P}\left(\tau_{\mathcal{E}} \geq t_p\right)}. \tag{3.52}$$

Therefore,

$$\Rightarrow \mathbb{P}\left(\tau_{\mathcal{E}} \geq t\right) = \mathrm{e}^{-\int_{t_p}^{t} \mathbb{P}(E_\tau = \mathcal{E} | \tau_{\mathcal{E}} \geq \tau) \mathrm{d}\tau}. \tag{3.53}$$

Please note the following before presenting the preceding results in a formal theorem. Given the dependence of E_t on X_t, by using the concept of *uncertain event likelihood function* introduced in Definition 3.1 and the property illustrated with Eq. (3.3), it can be obtained:

$$\begin{aligned} p(\tau_{\mathcal{E}} = t) &= \mathbb{P}\left(E_t = \mathcal{E} | \tau_{\mathcal{E}} \geq t\right) \mathrm{e}^{-\int_{t_p}^{t} \mathbb{P}(E_\tau = \mathcal{E} | \tau_{\mathcal{E}} \geq \tau) \mathrm{d}\tau} \\ &= \int_{\mathbb{X}_{(t_p,t]}} \mathbb{P}\left(E_t = \mathcal{E} | \{\tau_{\mathcal{E}} \geq t\}, x_{(t_p,t]}\right) \mathrm{e}^{-\int_{t_p}^{t} \mathbb{P}\left(E_\tau = \mathcal{E} | \{\tau_{\mathcal{E}} \geq \tau\}, x_{(t_p,t]}\right) \mathrm{d}\tau} \mathrm{d}F_{\mathbb{X}_{(t_p,t]}}(x_{(t_p,t]}) \qquad (3.54) \\ &= \int_{\mathbb{X}_{(t_p,t]}} \mathbb{P}\left(E_t = \mathcal{E} | x_t\right) \mathrm{e}^{-\int_{t_p}^{t} \mathbb{P}(E_\tau = \mathcal{E} | x_\tau) \mathrm{d}\tau} \mathrm{d}F_{\mathbb{X}_{(t_p,t]}}(x_{(t_p,t]}). \qquad (3.55) \end{aligned}$$

Theorem 3.2 [First Event Time in Stochastic Processes] Let $\mathcal{B}(\mathbb{R}_+)$ and $\lambda = \lambda(\mathbb{R}_+)$ denote the Borel σ-algebra and Lebesgue measure in $\mathbb{R}_+$, respectively. Let also $\{X_t\}_{t \in \mathbb{R}_+ \cup \{0\}}$ be a stochastic process and $\{E_t\}_{t \in \mathbb{R}_+}$ be an uncertain event process. If the first time of occurrence of the event $\mathcal{E}$, $\tau_{\mathcal{E}} = \tau_{\mathcal{E}}(t_p)$, with $t_p \in \mathbb{R}_+ \cup \{0\}$, is such that $\tau_{\mathcal{E}} < +\infty$, $\mathbb{P} - a.s.$, and $F_{\tau_{\mathcal{E}}} << \lambda$, then the mapping $p(\tau_{\mathcal{E}} = \cdot) : \mathbb{R}_+ \to [0, 1]$ exists and is well-defined in terms of its *uncertain event likelihood function* as

$$\mathbb{P}(\tau_{\mathcal{E}} = t) := \int_{\mathbb{X}_{(t_p,t]}} \mathbb{P}\left(E_t = \mathcal{E} | x_t\right) \mathrm{e}^{-\int_{t_p}^{t} \mathbb{P}(E_\tau = \mathcal{E} | x_\tau) \mathrm{d}\tau} \mathrm{d}F_{\mathbb{X}_{(t_p,t]}}(x_{(t_p,t]}). \tag{3.56}$$

Hence,

$$\mathbb{P}_{\mathcal{E}}(B) = \int_B \mathrm{d}F_{\tau_{\mathcal{E}}}(\tau) = \int_B p(\tau_{\mathcal{E}} = \tau) \mathrm{d}\tau, \ \forall B \in \mathcal{B}(\mathbb{R}_+), \tag{3.57}$$

is a probability measure defining a probability space $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \mathbb{P}_{\mathcal{E}})$. Indeed, the following conditions hold:

1) $\mathbb{P}_{\mathcal{E}}(\mathbb{R}_+) = 1$.
2) $0 \leq \mathbb{P}_{\mathcal{E}}(B) \leq 1, \forall B \in \mathcal{B}(\mathbb{R}_+)$.
3) $\mathbb{P}_{\mathcal{E}}(\cup_{k \in \mathbb{N}} B_k) = \sum_{k \in \mathbb{N}} \mathbb{P}_{\mathcal{E}}(B_k)$, $\forall \{B_k \in \mathcal{B}(\mathbb{R}_+)\}_{k \in \mathbb{N}}$, with $B_{\mathrm{i}} \cap B_j = \phi$, $\forall \mathrm{i} \neq j$.

Proof.

1) Let us define $B_t = (0, t)$, $t \in \mathbb{R}_+$, such that $B_t \nearrow \mathbb{R}_+$

$$\mathbb{P}_{\mathcal{E}}(B_t) = \int_{(0,t)} p(\tau_{\mathcal{E}} = \tau) \mathrm{d}\tau = \mathbb{P}(\tau_{\mathcal{E}} < t) \tag{3.58}$$

$$\Rightarrow \lim_{t\to+\infty} \mathbb{P}_{\mathcal{E}}(B_t) = \lim_{t\to+\infty} \mathbb{P}(\tau_{\mathcal{E}} < t) \tag{3.59}$$

$$\Rightarrow \mathbb{P}_{\mathcal{E}}(\mathbb{R}_+) = \mathbb{P}(\tau_{\mathcal{E}} < +\infty) = 1, \tag{3.60}$$

due to the continuity property of probability measures and because $\tau_{\mathcal{E}} < +\infty$, $\mathbb{P} - a.s.$

2) By definition, because

$$0 \leq \mathbb{P}(\tau_{\mathcal{E}} = t),\ \forall t \in \mathbb{R}_+ \ \Rightarrow\ 0 \leq \mathbb{P}_{\mathcal{E}}(B),\ \forall B \in \mathcal{B}(\mathbb{R}_+), \tag{3.61}$$

and, on the other hand,

$$B \subseteq \mathbb{R}_+ \ \Rightarrow\ \int_B p(\tau_{\mathcal{E}} = \tau)\mathrm{d}\tau \leq \int_{\mathbb{R}_+} p(\tau_{\mathcal{E}} = \tau)\mathrm{d}\tau \tag{3.62}$$

$$\Leftrightarrow\ \mathbb{P}_{\mathcal{E}}(B) \leq \mathbb{P}_{\mathcal{E}}(\mathbb{R}_+) = 1,\ \forall B \in \mathcal{B}(\mathbb{R}_+). \tag{3.63}$$

3) Let $\{B_k \in \mathcal{B}(\mathbb{R}_+)\}_{k\in\mathbb{N}}$ such that $B_{\mathrm{i}} \cap B_j = \phi$, $\forall \mathrm{i} \neq j$. By definition,

$$\mathbb{P}_{\mathcal{E}}(\cup_{k\in\mathbb{N}} B_k) = \int_{\cup_{k\in\mathbb{N}} B_k} p(\tau_{\mathcal{E}} = \tau)\mathrm{d}\tau \tag{3.64}$$

$$= \sum_{k\in\mathbb{N}} \int_{B_k} p(\tau_{\mathcal{E}} = \tau)\mathrm{d}\tau \tag{3.65}$$

$$= \sum_{k\in\mathbb{N}} \mathbb{P}_{\mathcal{E}}(B_k) \tag{3.66}$$

□

3.2 Application to Fatigue Crack Prognosis

The theoretical contributions provided in Section 3.1 include formal mathematical demonstrations and therefore do not require further verification. In this section, however, the aim is to explain how to use these abstract mathematical statements in practical applications of engineering: characterisation of *Time-of-Failure* (ToF) probability distributions; and more precisely, the problem of fatigue crack growth prognosis. As a function of loading cycles, a simplified stochastic degradation model is used for this purpose to describe the growth of a fatigue crack in a test coupon. The event $\mathcal{E}$ refers to critical failures that may occur in mechanical systems with components undergoing fatigue crack processes, although it may not be obvious what particular crack lengths may cause such events. The problem is approached using the notion of *uncertain event* and is contrasted to the classical threshold-crossing-based events (i.e., when critical failure is declared once the length of the crack reaches a *known* particular value). Probability distributions are shown in both cases, so as to develop a further discussion.

3.2.1 Crack Growth Model

It has been opted to use the condensed discrete-time crack growth model described in [126] to explain both the problem and the application of the theory of *uncertain events*. It is of paramount importance to highlight the fact that the purpose of this application example is

to show how the conceptual contributions presented can be applied, rather than contributing to the state-of-the-art in terms of the topic of crack length prognostics. More information on the specifics of fatigue crack growth in alloy test coupons can be found in [126].

The crack length can be depicted by a stochastic process $\{X_k\}_{k\in\mathbb{N}\cup\{0\}}$, according to the mathematical notation introduced in Section 3.1. Note that the k indexing variable usually denotes time, though in this case more specifically, it denotes a *cycle number*. The material undergoes several instances of compression and decompression. Also, $\mathbb{X} = \mathbb{R}_+$ and $\Sigma = \mathcal{B}(\mathbb{R}_+)$ (Borel sets in $\mathbb{R}_+$) due to the positiveness of length measures. In arbitrary units, the following discrete-time model defines the crack length:

$$x_{k+1} = x_k + \mathrm{e}^{\omega_k} C(\beta\sqrt{x_k})^n, \tag{3.67}$$

where $\omega_k \sim \mathcal{N}(0, \sigma_w^2)$ is a random variable describing white Gaussian noise, and C, β and n are fixed constants. All the model parameters values are summarised in Table 3.1.

	C	β	n	σ_w^2
Values	0.005	1	1.3	2.98

Table 3.1: Model parameters and their values.

3.2.2 Uncertain Event Definition

As specified in Definition 3.1, the statistics of an uncertain event $\mathcal{E}$ (in this case, critical failures in mechanical systems with components that undergo fatigue crack processes) are defined by an uncertain event likelihood function $\mathbb{P}(E_k = \mathcal{E}|x)$, which expresses how likely is the event $\mathcal{E}$ conditional to a particular fixed crack length x at a given time k. Without loss of generality, let us say that critical failure occurrences are correlated with crack lengths of about $\bar{x} = 100$. So the uncertain event likelihood function can be defined as

$$\mathbb{P}(E_k = \mathcal{E}|x) = \frac{1}{1 + \mathrm{e}^{-\alpha(x-\bar{x})}},\ \alpha > 0, \tag{3.68}$$

and therefore characterises the uncertainty associated with the occurrence of critical failure events as a function of the condition of the test coupon. In fact, it is still possible to return to a threshold-based failure definition when using this critical failure likelihood (see Remark 3.1) by merely observing the limit

$$\lim_{\alpha\to+\infty} \frac{1}{1 + \mathrm{e}^{-\alpha(x-\bar{x})}} = \mathbb{1}_{\{x\in\mathbb{R}:x>\bar{x}\}}(x). \tag{3.69}$$

3.2.3 Method of Monte Carlo Simulations

The *uncertain event* definition introduces a new degree of freedom regarding uncertainty. To see how this new source of uncertainty might influence event prognosis, its effects on the probability distribution associated with $\tau_{\mathcal{E}}$ have to be analysed. Monte Carlo simulations are used below to perform the necessary approximations due to their ability to estimate expectations with arbitrary precision by simply increasing the number of simulations, defined as $N \in \mathbb{N}$. Additionally, let $x^{(\mathrm{i})}_{k_p+1:k} = \{x^{(\mathrm{i})}_j\}_{j=k_p+1}^{k}$ denote the i-th random state trajectory

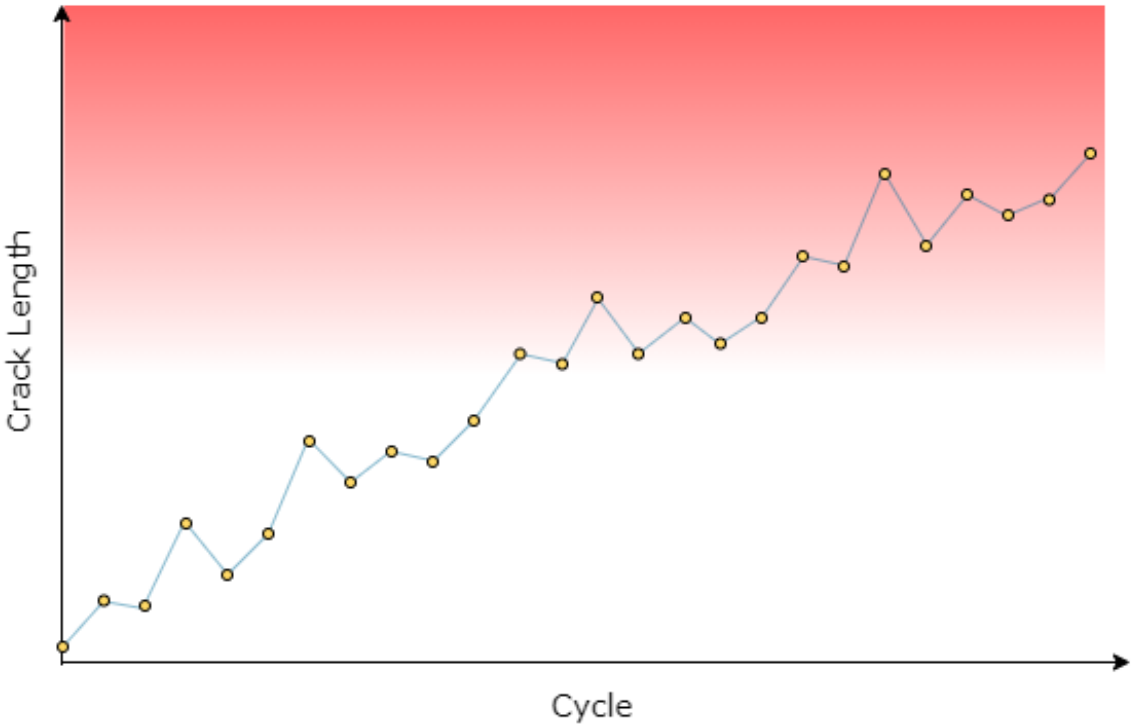


Figure 3.1: Single realisation of the stochastic process associated with crack length growth. The red color indicates the likelihood of an uncertain event; the greater the opacity the greater the likelihood (see Definition 3.1). The classical approach to the first occurrence time prediction on the other hand, would not have shown a color gradient, but a sudden and discontinuous shift from white to red.

or realisation of the stochastic process simulated from cycle k_p up to cycle k (see Fig. 3.1), described by the crack growth model (Markov process) of Eq. (3.67), with $\mathrm{i} \in \{1, \ldots, N\}$, $N >> 1$. The probability $\mathbb{P}(\tau_{\mathcal{E}} = k)$, with $\tau_{\mathcal{E}} = \tau_{\mathcal{E}}(k_p)$, can be approximated according to Theorem 3.1 as:

$$\mathbb{P}(\tau_{\mathcal{E}} = k) = \int_{\mathbb{X}_{k_p+1:k}} \mathbb{P}\left(E_k = \mathcal{E}|x_k\right) \prod_{j=k_p+1}^{k-1} \mathbb{P}\left(E_j = \mathcal{E}^c|x_j\right) \mathrm{d}F_{X_{k_p+1:k}}(x_{k_p+1:k}) \tag{3.70}$$

$$= \int_{\mathbb{X}_{k_p+1:k}} \mathbb{P}\left(E_k = \mathcal{E}|x_k\right) \prod_{j=k_p+1}^{k-1} \left(1 - \mathbb{P}\left(E_j = \mathcal{E}|x_j\right)\right) \mathrm{d}F_{X_{k_p+1:k}}(x_{k_p+1:k}) \tag{3.71}$$

$$\approx \frac{1}{N} \sum_{\mathrm{i}=1}^{N} \mathbb{P}\left(E_k = \mathcal{E}|x_k^{(\mathrm{i})}\right) \prod_{j=k_p+1}^{k-1} \left(1 - \mathbb{P}\left(E_k = \mathcal{E}|x_j^{(\mathrm{i})}\right)\right) \tag{3.72}$$

As shown in Fig. 3.1, each realisation of the stochastic process describing the development of crack growth throughout usage cycles (for example, the i-th), determines the likelihood of an uncertain event (in this case material futility). The figure shows an example of magnitudes taken by an uncertain event likelihood event function (similar to the one defined in Eq. (3.68)) expressed in terms of an hue over the crack length axis with colors ranging from clear white to red slowly. In this case, the smoothness of these changes in color would depend on the

α parameter with which the uncertain event likelihood function was defined in Eq. (3.68). To test the effect of this parameter on the probability mass distribution $\mathbb{P}(\tau_{\mathcal{E}} = \cdot)$, some variations are explored with the values shown in Table 3.2.

	α_1	α_2	α_3	α_4	$\alpha_{+\infty}$
Values	0.1	0.3	1.0	3.3	$\alpha \to +\infty$

Table 3.2: Values considered for the parameter α in the definition of the uncertain event likelihood $\mathbb{P}(E_k = \mathcal{E}|x)$ shown in Eq. (3.68).

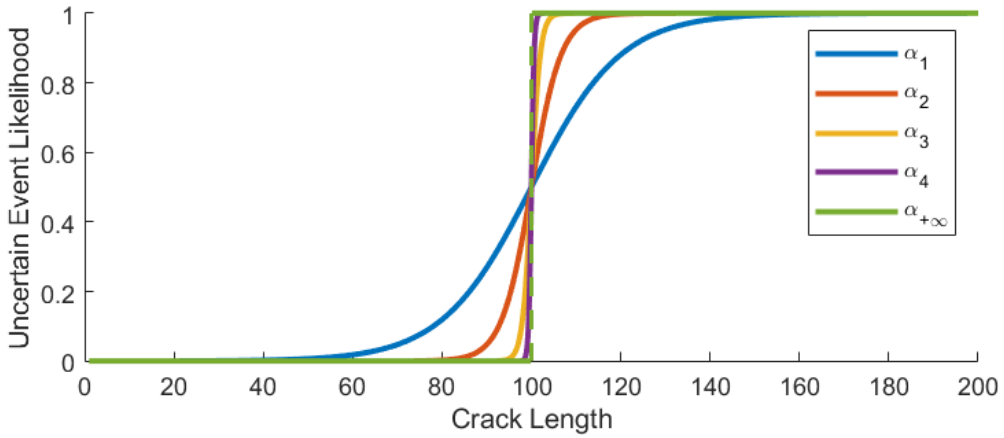


Figure 3.2: Uncertain event likelihood $\mathbb{P}(E_k = \mathcal{E}|x)$ as function of crack lengths $x \in \mathbb{R}_+$ considering different possible α parameters in Eq. (3.68), which are summarised in Table 3.2. The $\alpha_{+\infty}$ parameter represents a limit behaviour of the function when $\alpha \to +\infty$ (see Eq. (3.69)). The less the value of the parameter α, the greater the uncertainty about an specific crack length describing material futility.

For purposes of clarity, Fig. 3.2 reveals how the α parameter changes the way the uncertain event likelihood function looks like. There is a special case when $\alpha \to +\infty$ that is denoted by $\alpha_{+\infty}$, which corresponds to the generic threshold crossing notion that is widely assumed in the literature to trigger events (see Eq. (3.69)).

3.2.4 Simulation Results

Let us assume the predictions begin with the cycle number $k_p = 100$, where an initial crack almost negligible is observed (considered as $x_{k_p} = \mathrm{e}^{-10}$ for simulations) and the cycle number $k_h = 1000$ at which simulations are halted. Fig. 3.3 shows graphically how one hundred random crack growth trajectories would look like. Nevertheless, the Monte Carlo method mentioned in Section 3.2.3 to approximate $\mathbb{P}(\tau_{\mathcal{E}} = k)$, with $k \in \mathbb{N}$, demands the amount of simulations to be such that $N \to +\infty$, which is not feasible in real applications, but effective approximations can be obtained if N is "sufficiently large" (where "sufficiently large" depends on the state vector dimension, sources of uncertainty, model complexity, etc.). In this regard, the results for each of the α parameters obtained by performing an amount of $N = 10^7$ Monte Carlo simulations are included in Table 3.3 and Fig. 3.4. Greater values for N have not been considered because they have negligible effects on the results.

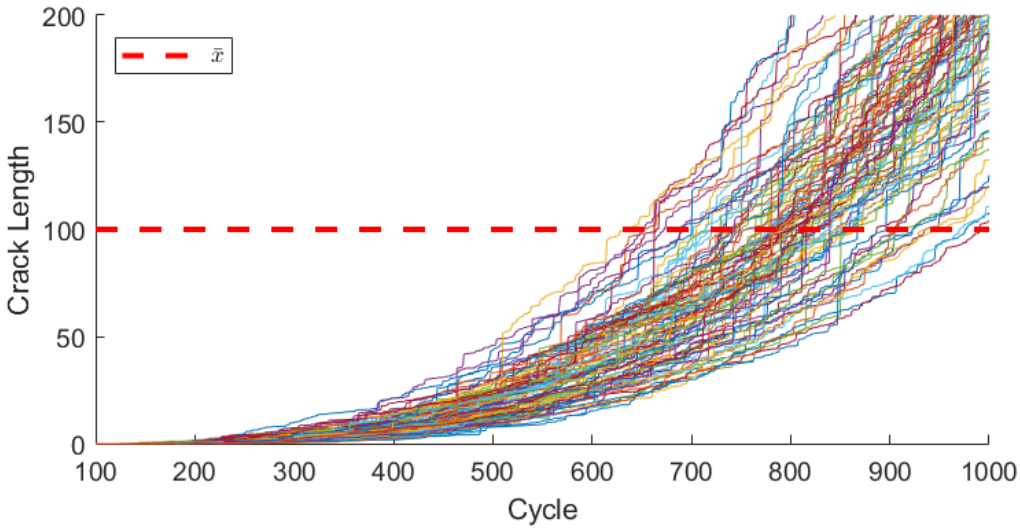


Figure 3.3: Graphical illustration of 100 realisations of the stochastic process described in Eq. (3.67), which corresponds to a fatigue crack growth model. The dashed horizontal line indicates a crack length of $\bar{x} = 100$ at which the material would incur critical failure, though there is uncertainty about it (see Section 3.2.2).

	α_1	α_2	α_3	α_4	$\alpha_{+\infty}$
$\mathbb{E}\{\tau_{\mathcal{E}}\}$	660.8835	766.3128	783.6094	786.7342	787.4333
$\mathbb{S}td\{\tau_{\mathcal{E}}\}$	102.6699	82.0342	82.7552	82.9145	82.9521
$\sum_{k=k_p}^{k_h} \mathbb{P}(\tau_{\mathcal{E}} = k)$	1.0000	0.9988	0.9970	0.9964	0.9962

Table 3.3: Results in terms of expected values, standard deviations and probability mass within a cycle span between k_p and k_h. The information is provided for each of the values considered for the parameter α in the definition of the uncertain event likelihood $\mathbb{P}\left(E_k = \mathcal{E}|x\right)$ of Eq. (3.68), which are shown in Table 3.2.

The probability distributions of $\tau_{\mathcal{E}}$ shown in Fig. 3.4 are pretty illustrative regarding consequences of uncertainty on the relationship between crack lengths and the eventuality of critical failures. Since the outline of these probability distributions is comparable to Gaussian bells, the expected values and standard deviations presented in Table 3.3 condense essentially all of the information needed to analyse the results properly.

The current standard threshold crossing methodology used in the literature is described precisely by $\alpha_{+\infty}$. It is straightforward to note from the expected values that, as α is reduced, the probability distributions of $\tau_{\mathcal{E}}$ are moved to the left. The standard deviations are increased in parallel, extending probabilities over a larger range of cycle numbers. Clearly, this behaviour is caused by any source of uncertainty indicating the possibility of earlier events. In fact, the *uncertain event likelihood function* definition (see section 3.2.2) implies that crack lengths below $\bar{x}$ may also trigger critical failures in mechanical components. It means that in a smaller amount of loading cycles it is likely to experience critical failures, which reflects the behaviour of the expected values in Table 3.3. On the other hand, the standard deviations are calculated as the result of introducing a new source of uncertainty

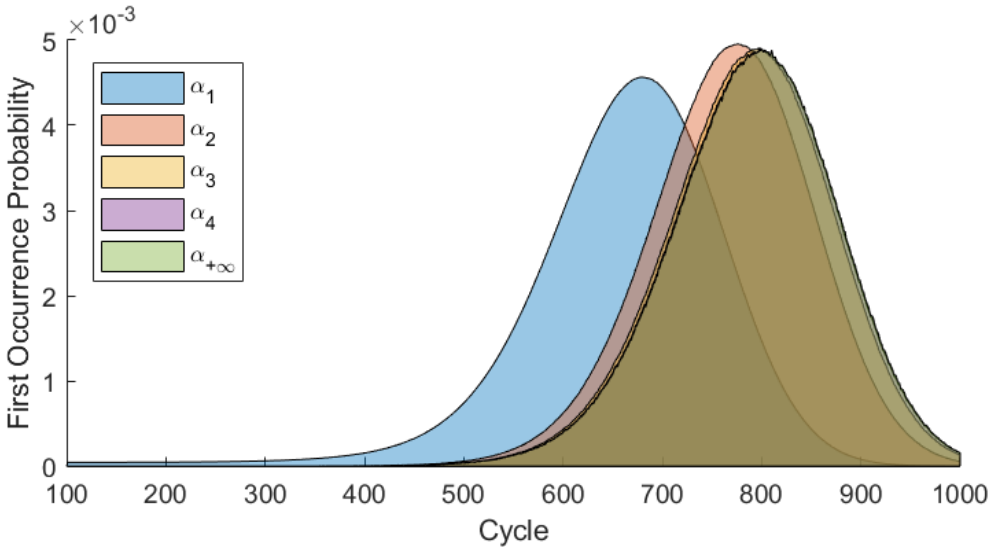


Figure 3.4: Probability distributions of the first occurrence time of uncertain future events under different definitions of uncertain event likelihood $\mathbb{P}(E_k = \mathcal{E}|x)$, which changes according to the different values of the α parameter (see Eq. (3.68)) presented in Table 3.2. The $\alpha_{+\infty}$ parameter describes the limit behaviour of the function when $\alpha \to +\infty$ (see Eq. (3.69)). The less the value of the α parameter, the greater the uncertainty about the cycle number at which cracks could lead to critical failures.

into the analysis. Finally, the results obtained by using α_3, α_4 and $\alpha_{+\infty}$ show consistency in similarity considering that their uncertain event likelihood functions are quite similar as well, as depicted in Fig. 3.2.

3.3 Summary

Scientists from several backgrounds have been addressing for more than fifty years the question of forecasting the time of occurrence of future events. They have therefore discussed this concept assuming a broad spectrum of stochastic processes. The common approach, though, has always been to trigger an event occurrence once a certain threshold or area in a higher dimensional space had been crossed. The fundamental explanation is based primarily on a goal of obtaining closed-form mathematical expressions. Uncertainty on this limit or higher dimensional area has been discussed so far only for a reduced amount of stochastic processes.

The concept of "*uncertain event*" is introduced in this chapter, which generalises the classic event definition by incorporating uncertainty on it. Even though the underlying idea was preconceived, its formalisation using Probability Theory is one of the main contributions presented in this chapter. In addition, the "*uncertain hazard zone*" definition established in the *Prognostics and Health Management* discipline has also finally been formalised. On the other hand, the second -and no less significant- contribution is the illustration of its applicability with a practical example of fatigue crack growth where concrete guidance and consequences of the proposed new concepts were presented and discussed.

Chapter 4

Predictive Bayesian Cramér-Rao Bounds of Future System States

As it was stated in Chapter 1, the event prognosis problem requires to predict the future states of a system, which can be described in terms of probability distributions (see Fig. 1.2). Such probability distributions are approximated by prognostic algorithms and are mainly determined by initial conditions (information about the system condition at the beginning of prognostics), dynamic model of the system, future inputs, characterisation and quantification of uncertainty sources, and hyper-parameters of the prognostic algorithm itself. An important question that arises in this respect is: *are there fundamental limits that may bound the hyper-parameters space when tuning a prognostic algorithm?* This question is embraced in this chapter by developing a conditional predictive version of the *Bayesian Cramér-Rao Lower Bound* (CP-BCRLB) (see Section 2.3 for a brief introduction). It establishes fundamental limits to the *Error Covariance Matrix* (ECM) incurred by estimators of future system states at each time in the future. This sequence of bounds (one per each future state vector predicted) can yield useful for hyper-parameters tuning of prognostic algorithms, as it is shown below.

This chapter begins by introducing theoretical contributions in Section 4.1 that are later used to propose a design methodology for prognostic algorithms in Section 4.2. Numerical methods may be needed to compute CP-BCRLBs, so Section 4.3 presents simplifications that can be made in particular prediction settings to get analytical expressions with which minimise computational efforts. All the contributions included in this chapter are then illustrated in Section 4.4 with a case study of *End-of-Discharge* (EoD) time prognosis of Lithium-Ion (Li-Ion) batteries. A summary is finally presented in Section 4.5.

Most of the content of this chapter was originally published in [80] and later extended as a book chapter in [127].

4.1 Conditional Predictive Bayesian Cramér-Rao Lower Bounds

Before presenting all the mathematical results based on the concept of CP-BCRLB (inspired by previous results presented in [125, 128]), it is important to clarify the idea behind these developments. The main goal is to lower bound the ECM incurred by any estimator $\hat{x}_k(y_{1:k_p})$ when estimating the system state vector x_k (marginal estimation), for all future time instants k greater than k_p (beginning of prognostics; see Fig. 1.2). This lower bound is formalised in Theorem 4.2 (mathematical demonstrations available in Appendix A.2) and later used in Section 4.2 for prognostic algorithm design. It can be computed in a recursive fashion, so it turns to be quite cheap in terms of computational resources. However, such recursion comes from a previous step where a different CP-BCRLB is derived for the whole future state trajectory (joint estimation), and whose results are summarised in Theorem 4.1 (mathematical demonstrations available in Appendix A.1).

Mathematical Results

Let $x_{k_p:k} = \begin{bmatrix} {x_{k_p}}^T & {x_{k_p+1}}^T & \dots & {x_k}^T \end{bmatrix}^T$ and also x^{i}, $\mathrm{i} = 1, 2, \dots, (k - k_p + 1)n_x$, be the i-th element of the vector $x_{k_p:k}$. The first step towards finding the marginal CP-BCRLB associated with x_k, is to explore the joint case of $x_{k_p:k}$ estimation.

Given a set of measurements $y_{1:k_p}$, let $\hat{x}_{k_p:k}(y_{1:k_p})$ denote an estimator of $x_{k_p:k}$. Also, let

$$\tilde{x}_{k_p:k} \triangleq \hat{x}_{k_p:k}(y_{1:k_p}) - x_{k_p:k} \tag{4.1}$$

denote the respective estimation error and

$$p_k^{cp} \triangleq p(x_{k_p:k}|y_{1:k_p}). \tag{4.2}$$

The second order derivative can be denoted as

$$\Delta_x^y = \nabla_x {\nabla_y}^T, \tag{4.3}$$

where $\nabla_x = \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \dots, \frac{\partial}{\partial x_{n_x}}\right]$ represents a gradient operator with dimensionality $1 \times n_x$.

Definition 4.1 [Conditional Predictive Bayesian Information Matrix] The *Conditional Predictive Bayesian Information Matrix* (CP-BIM) is defined as

$$I_{cp}(x_{k_p:k}|y_{1:k_p}) \triangleq \mathbb{E}_{p_k^{cp}}\{\left[\nabla_{x_{k_p:k}}{}^T \log p_k^{cp}\right]\left[\nabla_{x_{k_p:k}} \log p_k^{cp}\right]\}. \tag{4.4}$$

Two theorems lead to the mathematical notions of both joint and marginal CP-BCRLB variants. The joint variant lower bounds the ECM of complete state trajectories $x_{k_p:k}$ and demands expensive matrix computations. The marginal variant, in contrast, lower bounds the ECM of estimators of x_k (at a single future time instant) and is computed recursively with ease.

Theorem 4.1 [Joint Conditional Predictive BCRLB] Let us assume the following conditions about the density p_k^{cp} (see Eq. (4.2)):

1. p_k^{cp} is absolutely continuous and $\frac{\partial p_k^{cp}}{\partial x^{\mathrm{i}}}$ is absolutely integrable with respect to $x_{k_p:k}$, this is

$$\int \Big|\frac{\partial p_k^{cp}}{\partial x^{\mathrm{i}}}\Big| \mathrm{d}x_{k_p:k} < +\infty. \tag{4.5}$$

2. For each x^{i}, with $\mathrm{i} = 1, 2, \ldots, (k - k_p + 1)n_x$,

$$\lim_{x^{\mathrm{i}} \to +\infty} x^{\mathrm{i}} p(x_{k_p:k}) \;=\; \lim_{x^{\mathrm{i}} \to -\infty} x^{\mathrm{i}} p(x_{k_p:k}) = 0. \tag{4.6}$$

The ECM associated with any estimator $\hat{x}_{k_p:k}(y_{1:k_p})$ of the state trajectory $x_{k_p:k}$ is lower bounded as

$$\mathbb{E}_{p_k^{cp}}\{\tilde{x}_{k_p:k}\tilde{x}_{k_p:k}^T | y_{1:k_p}\} \geq I_{cp}^{-1}(x_{k_p:k}|y_{1:k_p}), \tag{4.7}$$

where $I_{cp}^{-1}(x_{k_p:k}|y_{1:k_p})$ is referred to as the *Joint Conditional Predictive BCRLB* (JCP-BCRLB).

Theorem 4.2 [Marginal Conditional Predictive BCRLB] Let us define

$$S_{\mathrm{i}+1}^{\mathrm{i}} = \mathbb{E}\{-\Delta_{x_{\mathrm{i}}}^{x_{\mathrm{i}}} \log p(x_{\mathrm{i}+1}|x_{\mathrm{i}})\}, \tag{4.8}$$

$$S_{\mathrm{i}+1}^{\mathrm{i},\mathrm{i}+1} = \mathbb{E}\{-\Delta_{x_{\mathrm{i}}}^{x_{\mathrm{i}+1}} \log p(x_{\mathrm{i}+1}|x_{\mathrm{i}})\}, \tag{4.9}$$

$$S_{\mathrm{i}+1}^{\mathrm{i}+1} = \mathbb{E}\{-\Delta_{x_{\mathrm{i}+1}}^{x_{\mathrm{i}+1}} \log p(x_{\mathrm{i}+1}|x_{\mathrm{i}})\}, \tag{4.10}$$

with $S_{\mathrm{i}+1}^{\mathrm{i}+1,\mathrm{i}} = S_{\mathrm{i}+1}^{\mathrm{i},\mathrm{i}+1^T}$, $\mathrm{i} = k_p, k_p + 1, \ldots, k$. The ECM associated with x_k, is lower bounded as

$$\mathbb{E}_{p_k^{cp}}\{\tilde{x}_k\tilde{x}_k^T | y_{1:k_p}\} \geq C_k^{22}, \tag{4.11}$$

where C_k^{22} is named as *Marginal Conditional Predictive BCRLB* (MCP-BCRLB), and can be recursively computed as

$$[C_k^{22}]^{-1} = S_k^k - S_k^{k,k-1}[[C_{k-1}^{22}]^{-1} + S_k^{k-1}]^{-1} S_k^{k-1,k}, \tag{4.12}$$

with the initial condition $[C_{k_p}^{22}]^{-1} = S_{k_p}^{k_p} = \mathbb{E}\{-\Delta_{x_{k_p}}^{x_{k_p}} \log p(x_{k_p}|y_{1:k_p})\}$.

Bayesian approaches to system monitoring (state estimation) assume that exogenous inputs are known. As a consequence, the latter are omitted in mathematical notation as it is concerned with describing state vector probability distributions. However, in event prognosis there is no certainty about future system operation; i.e., a random process characterises future exogenous inputs. This uncertainty becomes important in long-term predictions as its characterisation affects the whole prognosis problem and, thus, the computation of CP-BCRLBs. This fact can be easily sorted out by embedding the exogenous inputs vector into an augmented state vector before performing prognostics. This procedure is explained in Section 4.4.5.

4.2 Methodology for Prognostic Algorithm Design

Although it was mentioned that the concept of MCP-BCRLB (formalised in Theorem 4.2) would be used in this section to describe a design methodology, it is not applied straightforward. If there was an unbiased estimator $\hat{x}_k^*$ with zero covariance satisfying the conditions presented in Theorem 4.1, then the ECM would be minimum and the Bayesian Cramér-Rao inequality of Theorem 4.2 would become

$$\mathbb{E}_{p_k^{cp}}\{\tilde{x}_k\tilde{x}_k^T|y_{1:k_p}\} = \mathbb{C}ov_{p(x_k|y_{1:k_p})}\{x_k\} \geq C_k^{22}. \tag{4.13}$$

In the following, a design methodology for prognostic algorithms is described where the last inequality is used as criteria (in terms of predicted covariance rather than predicted ECM) to discard possible hyper-parameters of an algorithm whose covariance in time might violate this fundamental limit. Theorem 4.2 defines the lower bound C_k^{22}.

Design Methodology

Let us say that a probability-based prognostic algorithm is needed to assess the risk of future use for equipment failure. Let $\theta \in \Theta \subseteq \mathbb{R}^{n_\theta}$ denote an hyper-parameters vector determining any possible implementation of the aforementioned probability-based prognostic algorithm. A step-by-step methodology is defined for tuning these hyper-parameters, seeking to maximise the efficacy of the algorithm while taking into account specific efficiency restrictions (commonly imposed by computational cost and/or maximum processing time). Nevertheless, it is worth noting that some hyper-parameters may have a positive influence on the algorithm's efficacy, while others on its efficiency. Because of this, the hyper-parameters vector θ could be split into two by grouping those hyper-parameters primarily affecting efficiency (conveniently arranged as $\theta_A \in \Theta_A \subseteq \mathbb{R}^{n_{\theta_A}}, n_{\theta_A} < n_\theta$), and those primarily impacting on quality of results (conveniently arranged as $\theta_B \in \Theta_B \subseteq \mathbb{R}^{n_{\theta_B}}$, where $\theta^T = [\theta_A^T \quad \theta_B^T]$ and $n_{\theta_A} + n_{\theta_b} = n_\theta$). The θ_A parameter vector is commonly adjusted to satisfy efficiency limitations (e.g. maximum processing time); nevertheless, the actual question should be how to select suitable values for the θ_B parameter vector.

Below it is presented a methodology for the design of prognostic algorithms where CP-BCRLBs are implemented to lower bound covariances at each future time instant when predicting system states (see Eq. (4.13)). A feasibility region $\overline{\Theta_B} \subset \Theta_B$ is thus determined for the values of the hyper-parameters vector θ_B, where θ_A is assumed to meet efficiency restrictions. Therefore, this feasibility region is defined as all θ_B such that predictive state covariances do not violate MCP-BCRLBs (see Theorem 4.2), which constitute fundamental mathematical limits. The aforementioned design methodology is summarised as follows:

1) Meet efficiency requirements by choosing appropriate values for θ_A. Recursively compute MCP-BCRLBs for each predicted state in the future, beginning at time k_p and finishing at k_h, $k_h > k_p$, which defines a prediction horizon.
2) Choose different possible values for the hyper-parameters vector $\theta_B \in \Theta_B$. These can be obtained by sampling from a prior probability distribution, for example.
3) Conditional to each of the possible values for θ_B, execute prognostics. Each configuration of θ_B violating MCP-BCRLBs must be discarded, at any time $k_p < k < k_h$.

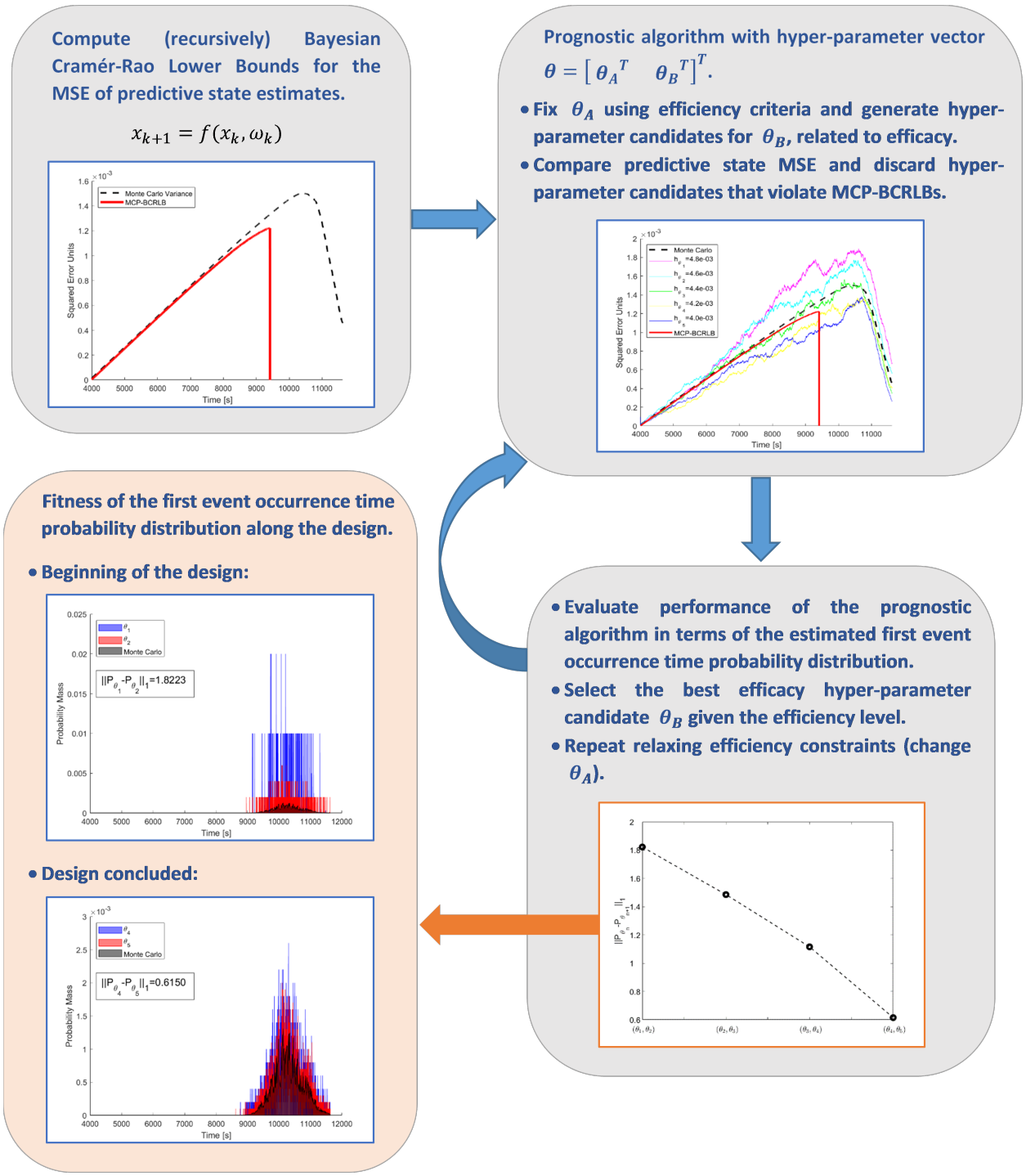


Figure 4.1: Methodology for the design of prognostic algorithms using MCP-BCRLBs.

4) For each of those non-discarded configurations of θ used in Step 3), compute the ℓ^1-distance, element by element, between the sequence of covariances $\{\mathbb{C}ov(x_k)\}_{k=k_p}^{k_h}$ (covariances as a function of time) and its corresponding MCP-BCRLB sequence. Calculate a weighted average with those distances. Select $[\theta_A^T \quad \hat{\theta}_B^T]$ such that this weighted average was minimised. Obtain the respective first event occurrence time probability

distribution associated with the hyper-parameters selection $[\theta_A^T \quad \hat{\theta}_B^T]$.

5) Explore how results are affected by relaxing soft efficiency constraints: Change θ_A to allow less efficient implementations and repeat Steps 1)-4). Evaluate the impact on the first event occurrence time probability distribution approximation by using any metric of choice. Iterate this procedure until changes in hyper-parameters yield negligible effects on this probability distribution.

4.3 Analytic Computation of MCP-BCRLBs

MCP-BCRLBs computation involves calculation of expectations over the probability distributions of predictive system states. This means that the designer might have to run Monte Carlo simulations (or alternative numerical methods) in order to tune the hyper-parameters of a particular prognostic algorithm. It would be better to avoid such situations whenever possible, as it is computationally expensive and time consuming. Fortunately, MCP-BCRLBs can be determined analytically in systems that are linear with respect to their state vectors and whose process noise is additive; i.e.,

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + \omega_k && (4.14)\\ &= A_k(u_k) \cdot x_k + B_k(u_k) + \omega_k, && (4.15) \end{aligned}$$

where u_k denotes system input, $A_k(u_k)$ and $B_k(u_k)$ correspond to an n_x-dimensional square matrix and an $n_x \times 1$ matrix, and ω_k represents an n_x-dimensional zero mean Gaussian noise vector. Indeed, if ω_k has Σ_k as covariance matrix:

$$-\log p(x_{\mathrm{i}+1}|x_{\mathrm{i}}) = c + \frac{1}{2}[x_{\mathrm{i}+1} - f(x_{\mathrm{i}}, u_{\mathrm{i}})]^T \Sigma_{\mathrm{i}}^{-1} [x_{\mathrm{i}+1} - f(x_{\mathrm{i}}, u_{\mathrm{i}})] \quad (4.16)$$

$$\begin{aligned} -\nabla_{x_{\mathrm{i}}} \nabla_{x_{\mathrm{i}}}{}^T \log p(x_{\mathrm{i}+1}|x_{\mathrm{i}}) &= A_{\mathrm{i}}(u_{\mathrm{i}})^T \Sigma_{\mathrm{i}}^{-1^T} \nabla_{x_{\mathrm{i}}} f(x_{\mathrm{i}}, u_{\mathrm{i}}) && (4.17)\\ &= A_{\mathrm{i}}(u_{\mathrm{i}})^T \Sigma_{\mathrm{i}}^{-1} A_{\mathrm{i}}(u_{\mathrm{i}}) && (4.18) \end{aligned}$$

$$\begin{aligned} -\nabla_{x_{\mathrm{i}+1}} \nabla_{x_{\mathrm{i}}}{}^T \log p(x_{\mathrm{i}+1}|x_{\mathrm{i}}) &= -A_{\mathrm{i}}(u_{\mathrm{i}})^T \Sigma_{\mathrm{i}}^{-1^T} \nabla_{x_{\mathrm{i}+1}} x_{\mathrm{i}+1} && (4.19)\\ &= -A_{\mathrm{i}}(u_{\mathrm{i}})^T \Sigma_{\mathrm{i}}^{-1} && (4.20) \end{aligned}$$

$$\begin{aligned} -\nabla_{x_{\mathrm{i}+1}} \nabla_{x_{\mathrm{i}+1}}{}^T \log p(x_{\mathrm{i}+1}|x_{\mathrm{i}}) &= \Sigma_{\mathrm{i}}^{-1^T} \nabla_{x_{\mathrm{i}+1}} x_{\mathrm{i}+1} && (4.21)\\ &= \Sigma_{\mathrm{i}}^{-1} && (4.22) \end{aligned}$$

Note that $\Sigma_{\mathrm{i}}^{-1^T} = \Sigma_{\mathrm{i}}^{-1}$ due to the symmetry assumed in the covariance matrix and $\nabla_y \nabla_x{}^T = \Delta_x^y$. In addition, it is worth underlying that the omission of the dependency on inputs u_k in mathematical expressions such as $p(x_{\mathrm{i}+1}|x_{\mathrm{i}})$ is a typical practice in engineering systems due to the assumption that inputs are accurately and precisely measured (therefore they should not afect, in this case, state transitions). Although not always true in prognosis, this chapter incorporates this omission practice unless it was necessary to make explicit the aforementioned dependency. Hence, analytical expressions for computing MCP-BCRLBs can be obtained from Eqs. (4.8)-(4.10):

$$\begin{aligned} S_{\mathrm{i}+1}^{\mathrm{i}} &= \mathbb{E}\{-\Delta_{x_{\mathrm{i}}}^{x_{\mathrm{i}}} \log p(x_{\mathrm{i}+1}|x_{\mathrm{i}})\} = A_{\mathrm{i}}(u_{\mathrm{i}})^T \Sigma_{\mathrm{i}}^{-1} A_{\mathrm{i}}(u_{\mathrm{i}}) && (4.23)\\ S_{\mathrm{i}+1}^{\mathrm{i},\mathrm{i}+1} &= \mathbb{E}\{-\Delta_{x_{\mathrm{i}}}^{x_{\mathrm{i}+1}} \log p(x_{\mathrm{i}+1}|x_{\mathrm{i}})\} = -A_{\mathrm{i}}(u_{\mathrm{i}})^T \Sigma_{\mathrm{i}}^{-1} && (4.24)\\ S_{\mathrm{i}+1}^{\mathrm{i}+1} &= \mathbb{E}\{-\Delta_{x_{\mathrm{i}+1}}^{x_{\mathrm{i}+1}} \log p(x_{\mathrm{i}+1}|x_{\mathrm{i}})\} = \Sigma_{\mathrm{i}}^{-1} && (4.25) \end{aligned}$$

4.4 Case Study: End-of-Discharge Time Prognosis of Lithium-Ion Batteries

The proposed design methodology for hyper-parameters tuning of prognostic algorithms is now being tested on an illustrative case study, which refers to *End-of-Discharge* (EoD) time prognosis of Lithium-Ion (Li-Ion) batteries. It is assumed in this case study that the first stage of fault diagnosis is carried out throughout state estimation via a particle filtering algorithm as suggested in [3] (regarding the amount of particles to be utilised, etc.). On the other hand, posterior estimates of the states of the system, namely the State-of-Charge (SoC, defined as the actual percentage of energy available with respect to the maximum storage capacity, denoted as E_{crit}) and internal polarisation resistance, are assumed to be always available whenever the prognostic algorithm was executed. The system is said to have undergone failure whenever SoC levels fall underneath 10%.

Section 4.4.4 illustrates how to apply the design methodology described in Section 4.2 in a simplified framework where future exogenous inputs are assumed to be known. In contrast, Section 4.4.5 incorporates uncertainty about them and explores their impact on the computation of predictive BCRLBs. Notwithstanding, let us firstly explain the system model and the prognostic algorithm to be designed.

4.4.1 State-Space Model

State-space models are useful in both estimation and prognosis stages as they provide representation of system dynamics. Voltage, in this application example of Li-Ion battery monitoring, is modelled as a function of i) SoC, ii) internal impedance of the battery, and iii) discharge current (which corresponds to the exogenous system input). The aim in this case study is oriented to prognosticate the time of depletion of the battery, defined to take place whenever its SoC reaches values below 10%.

Since actual failure prognostic algorithms demand the characterisation of future exogenous inputs, referring to future operating profiles (discharge currents) of a Li-Ion battery in this case, a probabilistic characterisation is proposed in [3] specifically for EoD time prognosis via Markov Chains. It is worth noting that, without loss of generality, the aforementioned characterisation of future system inputs and the performance assessment of prognostic algorithms can be addressed separately [113]. Moreover, conditional to a single realisation of the stochastic process characterising future system inputs, it is always feasible to assess prognostic algorithm performance. The impact of the uncertainty on future exogenous input characterisation can be later incorporated and evaluated by conditioning on different realisations and applying the *Law of Total Probability*. A known future battery usage profile is assumed then in a first instance, isolating the problem of EoD time probability distribution computation.

The relationship between *Open Circuit Voltage* (OCV) and SoC could be characterised through an affine function for most of the operating range of a Li-Ion battery, as depicted by "*zone 2*" in Fig. 4.2. Nevertheless, a state-space model that incorporates nonlinear operational behaviour was proposed in [3], which can be noted in "*zone 1*" and "*zone 3*" in the same figure. Besides, the polarisation resistance of the battery can be modelled as

a function of the discharge battery according to [129]. A state-space representation is thus expressed as:

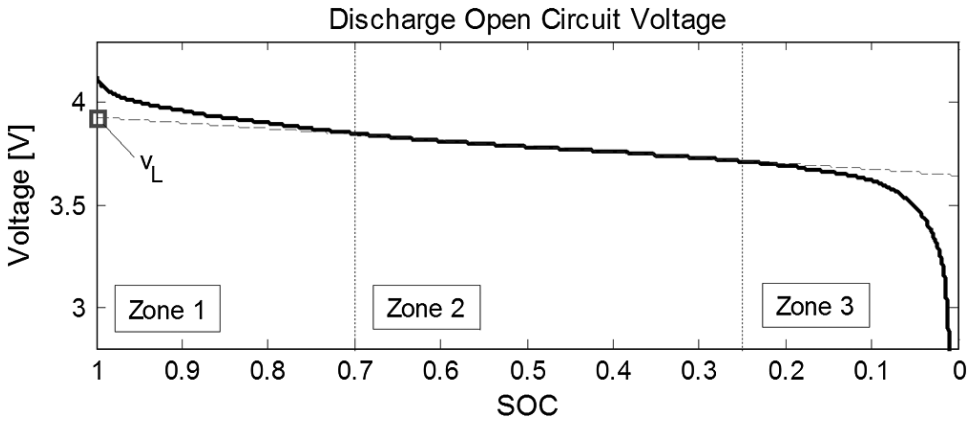


Figure 4.2: OCV curve of a Li-Ion cell (black line) and the projection of its linear operational range (dashed gray line) as a function of SoC [3].

State Transition Model

$$x_{k+1} \quad = \quad x_k - v_{oc}(x_k) \cdot u_k \cdot \frac{T_s}{E_{crit}} + \omega_k \tag{4.26}$$

Measurement Model

$$y_k = v_{oc}(x_k) - u(k) \cdot R_{int}(x_k, u_k) + \eta_k, \tag{4.27}$$

with

$$\begin{aligned} v_{oc}(x_k) =& v_L + (v_0 - v_L) \cdot \mathrm{e}^{\gamma \cdot (x_2(k)-1)} + \alpha \cdot v_L \cdot (x_2(k) - 1) \ldots \\ & \ldots + (1 - \alpha) \cdot v_L \cdot (\mathrm{e}^{-\beta} - \mathrm{e}^{-\beta \cdot \sqrt{x_2(k)}}) \end{aligned} \tag{4.28}$$

and

$$R_{int}(x_k, u_k) = r_0(u_k) + r_1(u_k) \cdot x_k + r_2(u_k) \cdot {x_k}^2. \tag{4.29}$$

The input $u_k = \mathrm{i}_k$ and output $y_k = v_k$ of the system are discharge current (measured in Amperes) and voltage (measured in Volts) between battery terminals, respectively. There is a single system state describing battery SoC, denoted as x_k, whose calculation is made as a fraction between actual energy available in the battery and its maximum storage capacity E_{crit}. The function $R_{int}(x_k, u_k)$, on the other hand, represents the absolute value of the battery internal impedance. Regarding ω_k and η_k, they denote process and measurement noise, respectively, and follow zero mean Gaussian distributions. Finally, sample time is measured in seconds and expressed as T_s, and v_0, v_L, α, β and γ refer to model parameters that must be obtained via batch estimation (see [3] for more details).

According to Definition 3.1 in Chapter 3, triggering a failure event whenever the battery SoC reaches values that fall underneath a threshold of 10% of available energy can be mathematically expressed in terms of an *uncertain event likelihood function* (see Remark 3.1) as

$$\mathbb{P}(E_\tau = \mathcal{E}|x) = \mathbb{1}_{\{x \in \mathbb{R}: x < 0.1\}}(x). \tag{4.30}$$

4.4.2 Prognostic Algorithm

To illustrate how to apply the design methodology presented in Section 4.2, a particle-filtering-based algorithm [77] is chosen for prognostic purposes. Predictions start at time k_p with an initial condition given by an empirical state posterior distribution yielded from a particle filter used for state estimation (fault diagnosis). This state posterior distribution is denoted as

$$p(x_{k_p}|y_{1:k_p}) = \sum_{\mathrm{i}=1}^{N_p} w_{k_p}^{(\mathrm{i})} \delta_{x_{k_p}^{(\mathrm{i})}}(x_{k_p}), \tag{4.31}$$

where the number of samples is denoted as N_p. The particle-filtering-based prognostic algorithm is summarised below:

0) Resample $p(x_{k_p}|y_{1:k_p})$ to obtain a set of N_θ particles equally weighted.

Perform the next steps for each future time k, $k > k_p$:

1) Compute expected state transitions $x_k^{*(\mathrm{i})} = \mathbb{E}\{f(x_{k-1}{}^{(\mathrm{i})}, u_{k-1}, \omega_{k-1})\}$, $\forall \mathrm{i} \in \{1, \ldots, N_\theta\}$, and the empirical covariance matrix

$$\hat{S}_k = \frac{1}{N_\theta - 1} \sum_{\mathrm{i}=1}^{N_\theta} [x_k^{*(\mathrm{i})} - \overline{x}_k^*][x_k^{*(\mathrm{i})} - \overline{x}_k^*]^T, \tag{4.32}$$

with $\overline{x}_k^* = \frac{1}{N_\theta} \sum_{\mathrm{i}=1}^{N_\theta} x_k^{*(\mathrm{i})}$.

2) Compute $\hat{D}_k$ such that $\hat{D}_k \cdot \hat{D}_k^T = \hat{S}_k$.

3) Update the samples

$$x_k^{(\mathrm{i})} = x_k^{*(\mathrm{i})} + h_\theta \cdot \hat{D}_k \cdot \varepsilon_k^{(\mathrm{i})},\ \varepsilon_k^{(\mathrm{i})} \sim \mathcal{E}, \tag{4.33}$$

where $\mathcal{E}$ denotes an Epanechnikov kernel with bandwidth h_θ.

Therefore, the hyper-parameters vector to be tuned in the prognostic algorithm ends up being $\theta^T = \begin{bmatrix} N_\theta & h_\theta \end{bmatrix}$ (number of particles and Epanechnikov kernel bandwidth).

4.4.3 Avoiding Monte Carlo Simulations to Estimate Fundamental Limits

According to Section 4.3, analytical expressions of MCP-BCRLBs can save unnecessary computations by avoiding Monte Carlo simulations (or similar numerical methods), like linear systems with additive process noise, for example. Hence, in the following it is intended to provide an approximation of the battery discharge model presented in Section 4.4.1, thus meeting the aforementioned requisites.

The state transition equation presented below can be used as a battery discharge model to later obtain mathematical expressions for MCP-BCRLBs computations:

$$x_{k+1} \;=\; x_k - v_{oc}(x_k) \cdot u_k \cdot \frac{T_s}{E_{crit}} + \omega_k. \tag{4.34}$$

It shows nonlinear behaviour with respect to the SoC state x_k due to the term $v_{oc}(x_k)$. However, from Fig. 4.2 it can be acknowledged that $v_{oc}(x_k)$ depicts a linear behaviour in most of its operational range. Furthermore, if $v_{oc}(x_k)$ is linearised around $x_o = 0.5$, then

$$v_{oc}(x_k) \approx v_{oc}(x_o) + \frac{\partial v_{oc}(x_k)}{\partial x_k}\Big|_{x_k=x_o} \cdot (x_k - x_o). \tag{4.35}$$

As a consequence, an approximation of the state transition equation can be obtained:

$$\begin{aligned} x_{k+1} = &\underbrace{\left(1 - \frac{\partial v_{oc}(x_k)}{\partial x_k}\Big|_{x_k=x_o} \cdot u_k \cdot \frac{T_s}{E_{crit}}\right)}_{A_k(u_k)} \cdot x_k \ldots \\ &\ldots + \underbrace{\left(\frac{\partial v_{oc}(x_k)}{\partial x_k}\Big|_{x_k=x_o} \cdot x_o - v_{oc}(x_o)\right) \cdot u_k \cdot \frac{T_s}{E_{crit}}}_{B_k(u_k)} + \omega_k, \end{aligned} \tag{4.36}$$

which in turn expresses the required form $x_{k+1} = A_k(u_k) \cdot x_k + B_k(u_k) + \omega_k$.

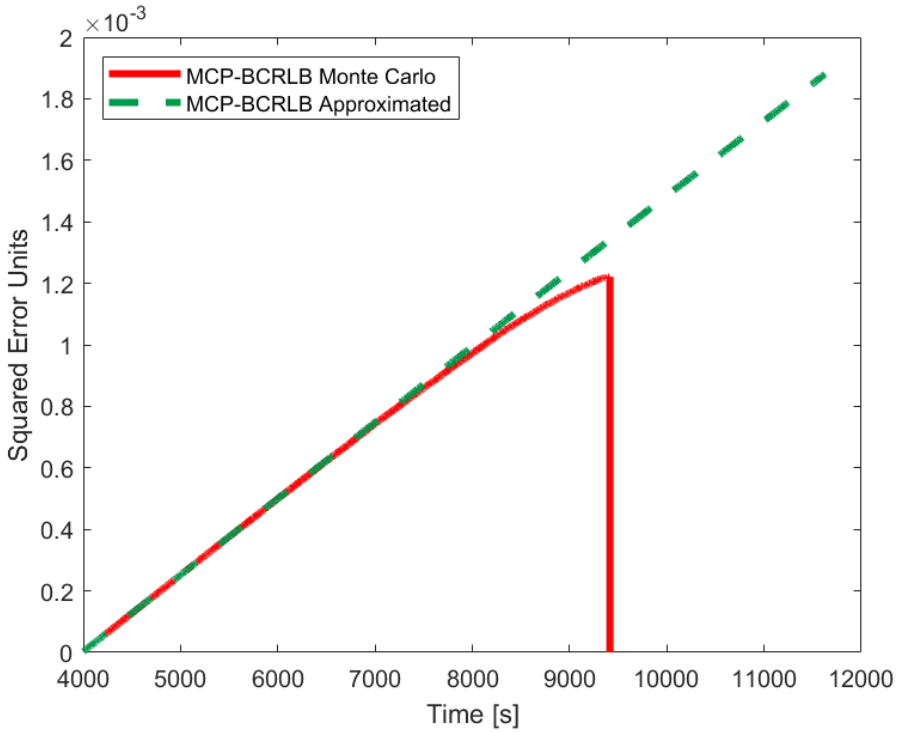


Figure 4.3: Differences between actual MCP-BCRLBs and an analytical, though simplified, way of computing MCP-BCRLBs by taking advantage of the linear behaviour experienced by Li-Ion batteries in a wide operating range.

Fig. 4.3 shows how the previous linearisation affects the computation of MCP-BCRLBs. This simplification saves a lot of computations, though its drawback relies on the lack of nonlinear characterisation and may skew the design methodology. It shows itself as an attractive alternative to proceed with design and avoid hard computations as long as its restricted validity region was considered.

4.4.4 Prognostic Algorithm Design: Known Future Operating Profiles

The prognostic algorithm presented in Section 4.4.2 is applied in EoD time prognosis of Li-Ion batteries (see Section 4.4.1). Hence, its hyper-parameters N_θ and h_θ can now be tuned by following the methodology described in Section 4.2. An important remark is to note that N_θ directly impacts on the computational effort of the algorithm (i.e., $\theta_A = N_\theta$), whereas h_θ has significant impact on the quality of probability distributions approximated by the prognostic algorithm (i.e., $\theta_B = h_\theta$).

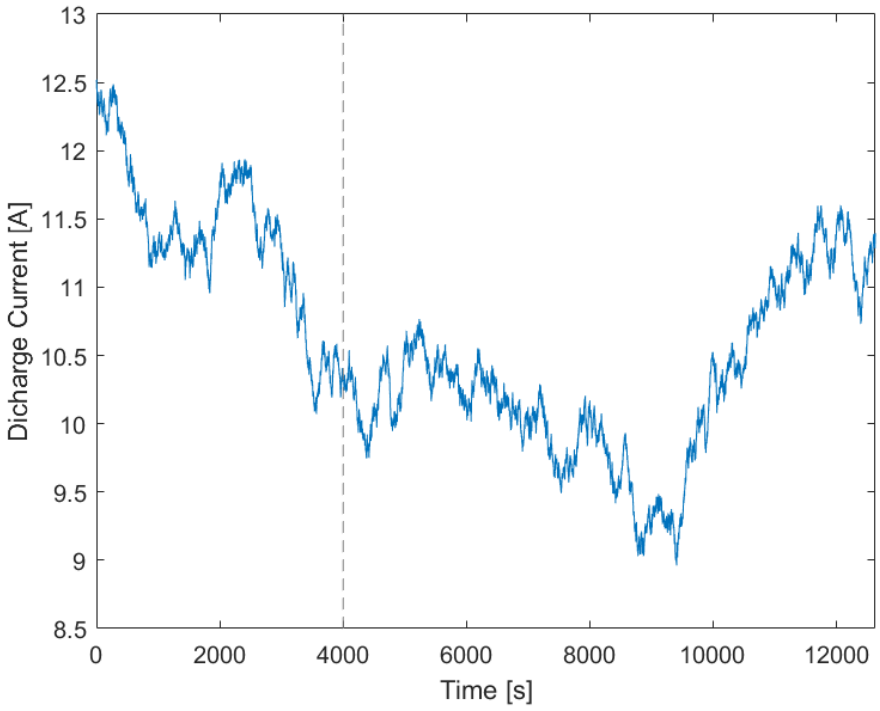
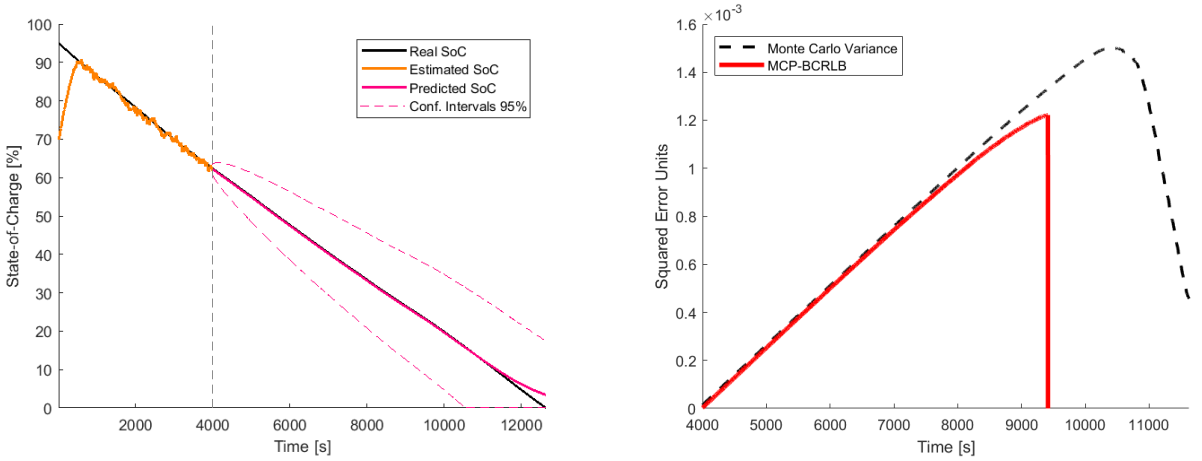


Figure 4.4: Discharge currents assumed as known. Prognostics are executed at time $k_p = 4000[s]$.

Consider that prognostics are executed at $k_p = 4000[s]$ and that predictions of future discharge currents (exogenous inputs) are assumed to be known (see Figure 4.4). In addition, no model simplifications are considered (in contrast to those mentioned in Section 4.4.3), embracing therefore the nonlinear behaviour shown in Fig. 4.2.

Let us now proceed with the design methodology proposed, step-by-step.

Step 1: Generation of MCP-BCRLBs Compute the MCP-BCRLBs from k_p (initial condition whose calculation was reported in [123]) onwards up to the prediction horizon k_h. Provided the computation of MCP-BCRLBs demands the calculation of analytically intractable integrals, ten million random trajectories are simulated using Eq. (4.26). The results of such predictions are shown in Fig. 4.5. Set N_θ to 100 particles, as recommended in [3].



(a) SoC estimation and prognosis.

(b) ECM and MCP-BCRLB curves.

Figure 4.5: Example results for $k_p = 4000[s]$. Figure 4.5(a) shows the results for battery SoC filtering and prediction stages. The estimation stage assumes an incorrect initial condition of 70% for the SoC, and is executed using a particle filter with 100 particles [3]. Long-term predictions are built simulating ten million random state trajectories. These predictions are used to compute covariance and MCP-BCRLB curves in Figure 4.5(b), thus empirically verifying Eq. (4.13).

Step 2: Choose hyper-parameter candidates for the algorithm There is an optimal bandwidth value for Epanechnikov kernels, denoted as h_{opt}, though it only applies when particles are obtained by independently sampling a Gaussian distribution having unity covariance matrix (see Eq. (4.37)) [77]. Nonetheless, despite the fact that such requirements rarely apply in nonlinear systems, h_{opt} has been used as educated guess in implementations of particle-filtering-based prognostic algorithms. In this case study, for instance, this educated guess would be $h_{opt} = 0.8529$ provided $n_x = 1$ and

$$h_{opt} = A \cdot N_\theta^{-\frac{1}{n_x+4}}, \; A = \left(8 \cdot c_{n_x}^{-1} \cdot (n_x + 4) \cdot (2 \cdot \sqrt{\pi})^{n_x}\right)^{\frac{1}{n_x+4}} \tag{4.37}$$

However, setting $h_\theta = h_{opt}$ leads to poor performance results regarding predicted state covariances (values of MCP-BCRLB are far lower than those of predicted state covariances). This fact suggests the search for new criteria to propose hyper-parameter candidates. The following Epanechnikov kernel bandwidths are considered as candidates henceforth in this chapter: $h_{\theta,1} = 0.0048$, $h_{\theta,2} = 0.0046$, $h_{\theta,3} = 0.0044$, $h_{\theta,4} = 0.0042$ and $h_{\theta,5} = 0.0040$.

Step 3: Discard hyper-parameter candidates that violate MCP-BCRLBs Assuming $N_\theta = 100$, Fig. 4.6 shows curves of predicted state variances related to each hyper-parameter candidate associated with h_θ. It can be noted that the MCP-BCRLB curve is violated by implementations that assume $h_{\theta,3}$, $h_{\theta,4}$ and $h_{\theta,5}$. In the case of $h_{\theta,3}$, however, it happens during short time spans, so it might not be discarded as increasing the hyper-parameter N_θ (related to efficiency) might eventually change this fact.

Step 4: Choose the most appropriate hyper-parameter candidate by using the ℓ^1-norm Only two candidates remain after Step 3): $h_{\theta,1}$ and $h_{\theta,2}$. A weighted

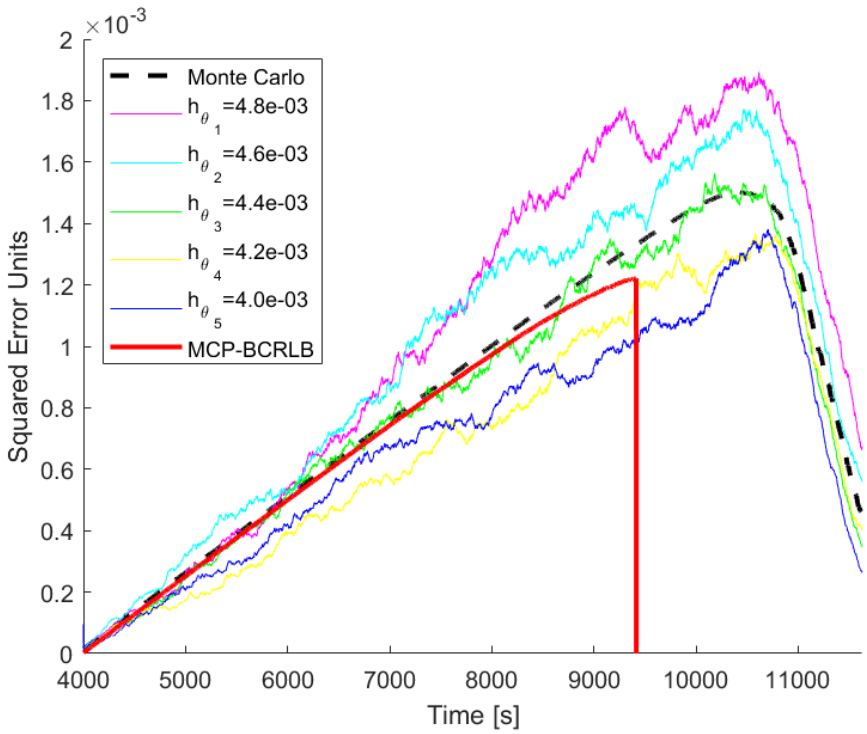


Figure 4.6: Predicted state variances obtained considering each of the possible candidates for the hyper-parameter h_θ whereas assuming $N_\theta = 100$. These curves can be compared to the actual variances (approximated via Monte Carlo simulations) and to the fundamental limits imposed by MCP-BCRLBs. Prognostics are executed at time $k_p = 4000[s]$.

average of the ℓ^1-norm is proposed as a notion of distance, applied element by element, between the sequences of matrices related to MCP-BCRLB and predicted state covariances, so as to be able to discriminate between candidates. However, since the system state is a scalar, then the criteria is reduced to choose the hyper-parameter whose sequence of variances in future times reaches minimum ℓ^1-distance with the sequence of MCP-BCRLBs:

$$\left\| \mathbb{V}ar_{h_{\theta,1}} - \text{MCP-BCRLB} \right\|_1 = 4.4223 \tag{4.38}$$

$$\left\| \mathbb{V}ar_{h_{\theta,2}} - \text{MCP-BCRLB} \right\|_1 = 3.8738. \tag{4.39}$$

A suitable candidate should be $h_{\theta,2} = 0.0046$. But what happens when relaxing efficiency constraints (i.e., allowing N_θ to adopt greater values and, hence, a larger number of particles)? The answer is not clear and may lead to a different candidate selection.

Step 5: Efficiency soft-constraints relaxation In order to understand the relationship between algorithmic efficacy (quality of results) and soft-constraints related to algorithmic efficiency (number of particles N_θ in this case), those soft-constraints may be relaxed and Steps 1)-4) may be repeated once more. Indeed, if $N_\theta = 500$ now, then the selected hyper-parameters vector would have been (see Table 4.1):

$$\theta^T = \begin{bmatrix} N_\theta & h_{\theta,2} \end{bmatrix} = \begin{bmatrix} 500 & 0.0046 \end{bmatrix}. \tag{4.40}$$

Although a criteria for choosing among hyper-parameter candidates could be only based on predicted state variance curves, to be fair, it should also include the true outcome

	$h_{\theta,1}$	$h_{\theta,2}$	$h_{\theta,3}$	$h_{\theta,4}$	$h_{\theta,5}$
$N_\theta = 100$	4.4223	3.8738✓	2.8580✗	3.1787✗	2.8415✗
$N_\theta = 500$	4.7344	3.9532	3.4707✓	2.9609✗	2.8519✗
$N_\theta = 1000$	4.2415	3.6688	3.0113✓	3.0336✗	2.8367✗
$N_\theta = 5000$	4.3668	3.9204	3.1596✓	2.7492✗	2.8441✗
$N_\theta = 10000$	4.3299	3.6690	3.1424✓	2.8502✗	2.8386✗

Table 4.1: Dissimilarity between predicted state variance and MCP-BCRLB curves (ℓ^1-distance). Candidates that were discarded in Step 3) are marked with a ✗-symbol. Selected candidates associated with minimum distances are marked with a ✓-symbol.

yielded from probabilistic prognostic algorithms, which is a probability distribution for the first occurrence time of an event (EoD, in this case). Therefore, another metric is required to discriminate between such probability distributions. Again, an ℓ^1-distance may yield useful for this purpose.

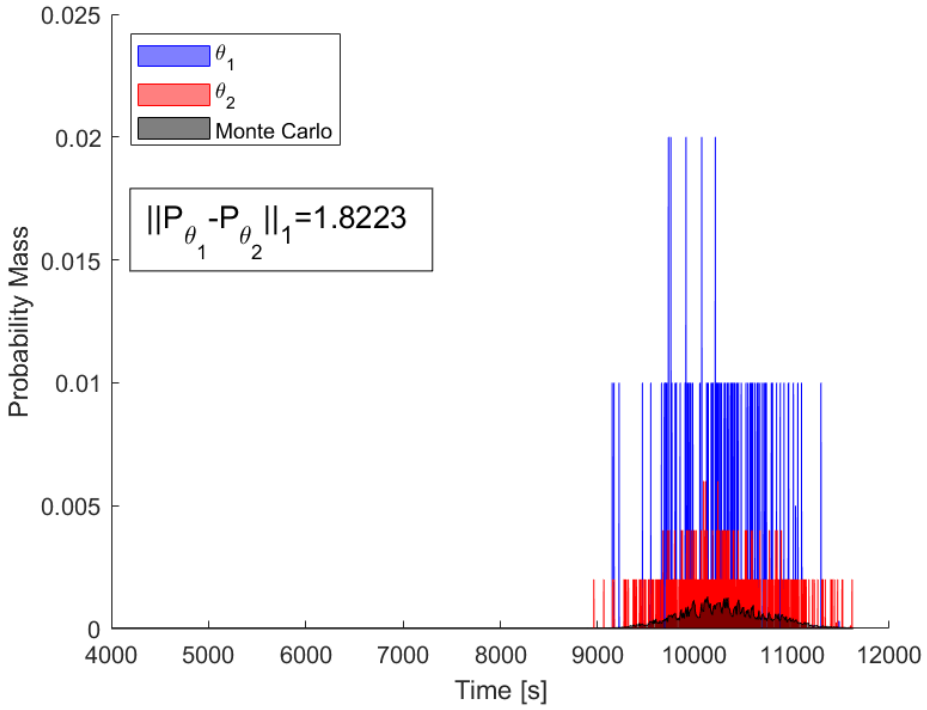


Figure 4.7: EoD time probability distributions for two choices of algorithm hyper-parameters $\theta_1^T = [100 \;\; 0.0046]$ and $\theta_2^T = [500 \;\; 0.0044]$. Prognostics are executed at time $k_p = 4000[s]$. The gray area shows the "ground truth" EoD time probability distribution, which is approximated by ten million simulations of state trajectories (convergence is reached with less than one million simulations).

As an illustrative example, Fig. 4.7 shows the output of the particle-filtering-based prognostic algorithm after going through Steps 1)-4) two times and selecting $\theta_1^T = [100 \;\; 0.0046]$ and $\theta_2^T = [500 \;\; 0.0044]$ as hyper-parameters at different efficiency levels. Differences in terms of N_θ suggest different efficiency levels that must be considered. On the other hand, it can be noted, by taking the ℓ^1-distance between both outputs,

that differences in efficiency constraints lead to significant impact regarding quality of results when comparing to the “ground truth" EoD time probability distribution obtained via Monte Carlo simulations. Results improve as the number of particles N_θ gets larger.

After taking into account all of the above, the impact on algorithmic efficacy can be evaluated with respect to efficiency (larger numbers of particles N_θ). A set of hyper-parameters candidates (obtained by following the previous design methodology) maximising efficacy conditional to given efficiency constraints is shown below:

$$\{\theta_n\}_{n=1}^{5} = \left\{ \begin{bmatrix} N_{\theta_n} \\ h_{\theta_n} \end{bmatrix} \right\}_{n=1}^{5} \tag{4.41}$$

$$= \left\{ \begin{bmatrix} 100 \\ 0.0046 \end{bmatrix}, \quad \begin{bmatrix} 500 \\ 0.0044 \end{bmatrix}, \quad \begin{bmatrix} 1000 \\ 0.0044 \end{bmatrix}, \quad \begin{bmatrix} 5000 \\ 0.0044 \end{bmatrix}, \quad \begin{bmatrix} 10000 \\ 0.0044 \end{bmatrix} \right\}. \tag{4.42}$$

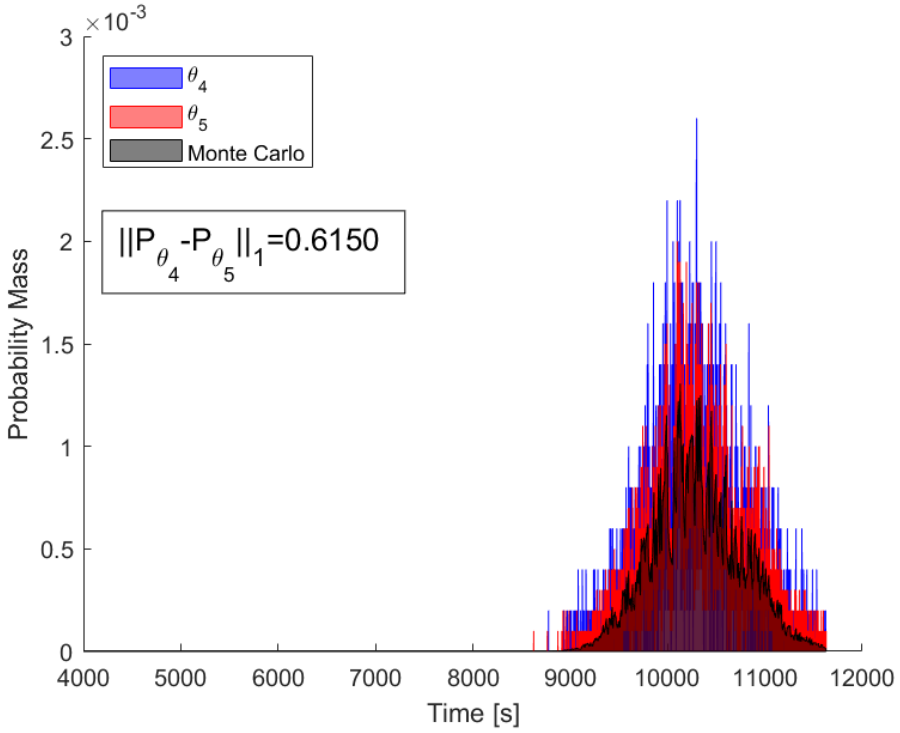


Figure 4.8: EoD time probability distributions for two choices of algorithm hyper-parameters $\theta_4^T = [5000 \quad 0.0044]$ and $\theta_5^T = [10000 \quad 0.0044]$. Prognostics are executed at time $k_p = 4000[s]$. The gray area shows the “ground truth" EoD time probability distribution, which is approximated by ten million simulations of state trajectories (convergence is reached with less than one million simulations).

Improvements on prognostic results are clearly reached by increasing the number of particles, as shown in Figs. 4.7 and 4.8. It is of special interest to characterise the left tail of probability distributions for the first occurrence time of the event (EoD in this case), because it constitutes valuable information to quantify future operational risk. It can be noted that efficacy does not increase significantly for number of particles larger

than $N_\theta = 5000$. Thus, the recommendation would be to select $\theta_4^T = [5000 \quad 0.0044]$ as a final hyper-parameter design in this specific case study. It offers the best efficacy performance without incurring in unnecessary computations.

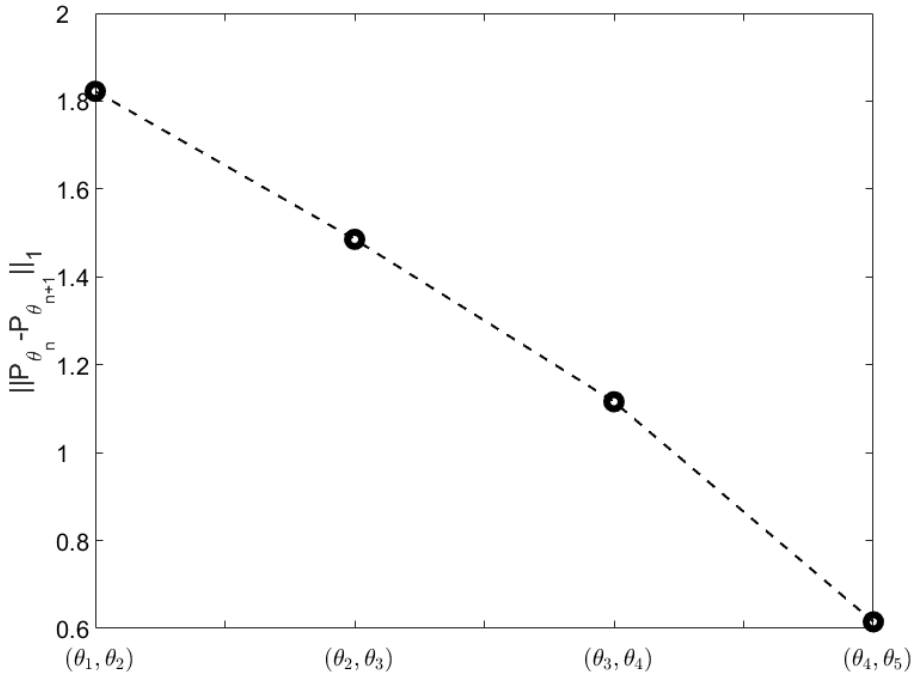


Figure 4.9: Summary of ℓ^1-distances between two consecutive EoD time probability distributions obtained throughout the design methodology. The hyper-parameters vector θ_n corresponds to the selection taken at the n-th iteration of the design procedure.

Fig. 4.9 shows how the ℓ^1-distances between consecutive prognostic algorithm's outputs evolve when iterating on the design methodology. It quantifies the analysis previously described from observing Figs. 4.7 and 4.8.

4.4.5 Prognostic Algorithm Design: Statistical Characterisations of Future Operating Profiles

Although the design methodology proposed in Section 4.2 was already illustrated in Section 4.4.4, it assumed that the future operating profile was known, which is seldom the case. What happens if future system inputs are unknown?

Uncertainty about future exogenous inputs can be incorporated during the prognosis stage by embedding system inputs into an augmented state vector. Thus, the latter would be defined as

$$\breve{x}_k = \begin{bmatrix} x_k \\ u_k \end{bmatrix}. \tag{4.43}$$

The augmented state vector now requires the incorporation of the future inputs dynamics into an augmented state transition system of equations. In this particular example of EoD

time prognosis, u_k (discharge current and system input) is characterised by a random walk model. As a consequence, the augmented state transition system of equations yields:

$$\underbrace{\begin{bmatrix} x_{k+1} \\ u_{k+1} \end{bmatrix}}_{\breve{x}_{k+1}} = \underbrace{\begin{bmatrix} x_k - v_{oc}(x_k) \cdot u_k \cdot \frac{T_s}{E_{crit}} \\ u_k \end{bmatrix}}_{\breve{f}(\breve{x}_k)} + \underbrace{\begin{bmatrix} \omega_k \\ \nu_k \end{bmatrix}}_{\breve{\omega}_k}, \tag{4.44}$$

where ν_k represents zero mean Gaussian distribution. Measurement equations are not required in this case since the state augmentation here would only been adopted in prognostics (which only propagates prior information in time), and predictions are thus not corrected by actual measurements.

According to Theorem 4.2, the elements of the recursive formula of MCP-BCRLB computations would become:

$$\breve{S}_{\mathrm{i}+1}^{\mathrm{i}} = \mathbb{E}\{-\Delta_{\breve{x}_{\mathrm{i}}}^{\breve{x}_{\mathrm{i}}} \log p(\breve{x}_{\mathrm{i}+1}|\breve{x}_{\mathrm{i}})\} = \begin{bmatrix} -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial {x_k}^2} & -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial x_k \partial u_k} \\ -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial u_k \partial x_k} & -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial {u_k}^2} \end{bmatrix}, \tag{4.45}$$

$$\breve{S}_{\mathrm{i}+1}^{\mathrm{i},\mathrm{i}+1} = \mathbb{E}\{-\Delta_{\breve{x}_{\mathrm{i}}}^{\breve{x}_{\mathrm{i}+1}} \log p(\breve{x}_{\mathrm{i}+1}|\breve{x}_{\mathrm{i}})\} = \begin{bmatrix} -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial x_k \mathrm{d} x_{k+1}} & -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial x_k \partial u_{k+1}} \\ -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial u_k \partial x_{k+1}} & -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial u_k \partial u_{k+1}} \end{bmatrix}, \tag{4.46}$$

$$\breve{S}_{\mathrm{i}+1}^{\mathrm{i}+1} = \mathbb{E}\{-\Delta_{\breve{x}_{\mathrm{i}+1}}^{\breve{x}_{\mathrm{i}+1}} \log p(\breve{x}_{\mathrm{i}+1}|\breve{x}_{\mathrm{i}})\} = \begin{bmatrix} -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial {x_{k+1}}^2} & -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial x_{k+1} \partial u_{k+1}} \\ -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial u_{k+1} \partial x_{k+1}} & -\frac{\partial^2 \log p(\breve{x}_{k+1}|\breve{x}_k)}{\partial {u_{k+1}}^2} \end{bmatrix}. \tag{4.47}$$

The final element to consider in MPC-BCRLB computations is a proper definition for the initial condition of the recursion. It requires information about error precision with which states and inputs are characterised at k_p (time at which prognostics are executed). The first one related to posterior estimates of the state vector x_{k_p} was already review in Section 4.4.4. The second one, however, refers to u_{k_p} and its error precision is given the capacity of the corresponding sensors, which are imperfect and associated with measurement uncertainty. Therefore, the initial condition for the recursive computation of MCP-BCRLBs is given by:

$$[\breve{C}_{k_p}^{22}]^{-1} = \begin{bmatrix} [C_{k_p}^{22}]^{-1} & 0 \\ 0 & \varepsilon_I^{-1} \end{bmatrix}, \tag{4.48}$$

with $\varepsilon_I > 0$. Note that the definition of the matrix $[\breve{C}_{k_p}^{22}]^{-1}$ is expressed as an inverse and, therefore, has to be non-singular. Proper selection of the constant ε_I may allow to force this invertibility property. This positive and arbitrary small constant ε_I sets a lower bound for measurement precision at time k_p; i.e., I_{k_p}.

4.5 Summary

Inspired by *Bayesian Cramér-Rao Lower Bounds*, this chapter has introduced new formal performance metrics to assess and design prognostic algorithms. These establish fundamental limits to predicted state covariances, thus allowing the development of a step-by-step methodology to tune hyper-parameters of prognostic algorithms, discarding those violating

the aforementioned limits. Both the design methodology as well as the metric are validated with mathematical demonstrations and empirical verification in the context of a problem of EoD time prognosis of Li-Ion batteries.

The design methodology presented in this chapter splits the hyper-parameters vector of a prognostic algorithm, distinguishing between those hyper-parameters with greater impact on efficiency, and those affecting efficacy. Efficiency hyper-parameters, on the one hand, are related to computational resources required in practical implementations, whereas the adjustment of efficacy hyper-parameters influence quality of results. Generally, efficiency resources are limited, so algorithm design must provide the best performance conditional to such limitation. The aforementioned design methodology is thus a structured procedure to explore and achieve the best efficacy given certain level of efficiency restriction.

From the point of view of prognostic algorithm designers, it is always feasible to assume future operating profiles as known: the algorithm will serve well regarding fulfilling its design purpose, though conditional to a proper characterisation of future system's inputs. However, this characterisation of future inputs can be conceived as a separate problem.

Chapter 5

Predictive Bayesian Cramér-Rao Bounds of Event Occurrence Time

According to the description of the event prognosis problem presented in Chapter 1, prognostic algorithms propagate uncertainty about the current state of the system into the future, which generates a sequence of probability distributions for each of the future states of the system (see Fig. 1.2). Chapter 4 shows that there are fundamental limits for these that allow us to lower bound the *Error Covariance Matrix* (ECM) or, in some cases, the covariance with which they are predicted. These fundamental limits are based on the concept of *Bayesian Cramér-Rao Lower Bound* (BCRLB) and, as seen in Chapter 4, they can be used as criteria to design prognostic algorithms. However, the true output of prognostic algorithms consists of a probability distribution for the first occurrence time of an event of interest, as illustrated in Fig. 1.2. It is reasonable, therefore, to think that such bounds could also exist for the time of occurrence of events, and this is what this chapter is about. Intuitively, these would constitute fundamental limits for the accuracy and precision with which the occurrence time of events can be predicted.

In the context of failure prognosis (where the event of interest is the occurrence of a failure), the scientific community has repeatedly promoted the validation of prognostic algorithms using the concept of "accuracy" (i.e., hoping to provide *Time-of-Failure* (ToF) estimates as close as possible to the actual value that was obtained in a single run-to-failure experiment) or "precision" (the smaller variance of obtained ToF *Probability Density Functions* (PDFs), the better). But, is this assessment theoretically correct? On the one hand, the actual ToF in a failure data set corresponds to a single realisation of a stochastic process, so studying the accuracy of the algorithm should be about how it approaches the expected ToF with respect to the actual statistics of the "ground truth" theoretical ToF probability distribution. The experimental ToF could even adopt an unlikely value distant from the expectation. Therefore, the notion of accuracy as a metric so far does not have a valid theoretical basis. On the other hand, is it actually possible to achieve an arbitrary high precision? Is there a minimum achievable *Mean Squared Error* (MSE)?

This chapter intends to approach the previous questions by associating concepts of accuracy and precision of prognostic algorithms with the first two moments of an underlying

"ground truth" first occurrence time probability distribution. BCRLBs are derived to lower bound the MSE (equivalent to ECM in one dimension, which is the case in $\tau_{\mathcal{E}}$ estimation) of any first occurrence time estimator. As this lower bound yields independent of the utilised prognostic algorithm, it may help to build objective performance metrics to discriminate poorly designed prognostic algorithms: particularly those promising higher precision than it is possible to achieve. The main contributions of this chapter are (i) to show the existence of such fundamental mathematical limits, and (ii) to provide insights on their potential use.

The aforementioned BCRLB is introduced below for both discrete- and continuous-time systems in Sections 5.1.1 and 5.1.2, respectively, while the corresponding mathematical proofs can be found in Appendix B. The same case study of *End-of-Discharge* (EoD) time prognosis of Lithium-Ion (Li-Ion) batteries that is previously chosen for illustrative purposes in Chapter 4 is used here as well. A brief summary of the main results of this chapter can be found in Section 5.3.

5.1 Mathematical Formalisations

Depending on the nature of time (discrete or continuous), the theorems that formalise BCRLBs that lower bound the MSE associated with the estimation of the first occurrence time of an event $\mathcal{E}$ are shown below.

5.1.1 Discrete-Time Stochastic Systems

Before attempting to introduce the BCRLB associated with the first occurrence time of future events in the discrete-time case, it is convenient for pedagogical purposes to describe some of the most important constituent elements for the theorem presented below.

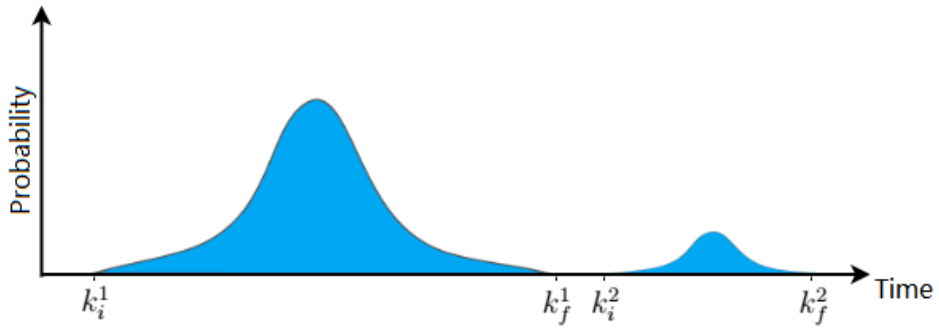


Figure 5.1: Example of probability mass distribution of $\tau_{\mathcal{E}}$ with disconnected support. The support of $\tau_{\mathcal{E}}$, denoted as $supp(\tau_{\mathcal{E}}) \subseteq \mathbb{N}$, is defined as the union of time intervals where the probability distribution of $\tau_{\mathcal{E}}$ is greater than zero. In this example, $supp(\tau_{\mathcal{E}}) = \{k_{\mathrm{i}}^1, k_{\mathrm{i}}^1 + 1, ..., k_f^1\} \cup \{k_{\mathrm{i}}^2, k_{\mathrm{i}}^2 + 1, ..., k_f^2\}$.

The stochastic process denoted as $\{X_k\}_{k \in \mathbb{N} \cup \{0\}}$ represents a dynamic system evolving in time. If prognostics are performed time k_p, then the probability distribution associated with the first occurrence time of a future event $\mathcal{E}$, $\tau_{\mathcal{E}} = \tau_{\mathcal{E}}(k_p)$, is denoted as $\mathbb{P}(\tau_{\mathcal{E}} = \cdot)$. In this

section, $\mathbb{P}(\tau_{\mathcal{E}} = \cdot)$ is referred to as the "*ground truth*". Although the statistics of $\tau_{\mathcal{E}}$ are completely determined given a system dynamic model, in practice some approximations must be done due to the intractable analytic expressions that result from the nonlinear nature of the uncertainty propagation process. Consequently, an approximated random variable $\hat{\tau}_{\mathcal{E}}$ (estimator of $\tau_{\mathcal{E}}$) is computed by the prognostic algorithm, providing approximate statistical information via $\mathbb{P}(\hat{\tau}_{\mathcal{E}} = \cdot)$. The mathematical expression $\hat{\tau}_{\mathcal{E}} \perp \tau_{\mathcal{E}}$ stands for the independence between $\hat{\tau}_{\mathcal{E}}$ and $\tau_{\mathcal{E}}$: the prognostic algorithm is conditioned on the same system dynamics, initial conditions and uncertainty sources that determine $\mathbb{P}(\tau_{\mathcal{E}} = \cdot)$, although the prognostic algorithm propagates uncertainty in an independent manner (yielding a probability distribution $\mathbb{P}(\hat{\tau}_{\mathcal{E}} = \cdot)$). The support of $\tau_{\mathcal{E}}$, $supp(\tau_{\mathcal{E}})$, is defined as an union of disjoint sets, as illustrated in Fig. 5.1.

Theorem 5.1 [Discrete-Time Bayesian Cramér-Rao Inequality] Let $\{X_k\}_{k\in\mathbb{N}\cup\{0\}}$ be a stochastic process and $\{E_k\}_{k\in\mathbb{N}}$ be an uncertain event process, respectively. According to Theorem 3.1, the first time of occurrence of an event $\mathcal{E}$ is characterised by the random variable $\tau_{\mathcal{E}} = \tau_{\mathcal{E}}(k_p)$, with $k_p \in \mathbb{N} \cup \{0\}$. On the other hand, if $\hat{\tau}_{\mathcal{E}}$ is an estimator of $\tau_{\mathcal{E}}$, $\hat{\tau}_{\mathcal{E}} \perp \tau_{\mathcal{E}}$, then it can be defined

$$\mathbb{P}^*(\hat{k}, k) = \mathbb{P}(\hat{\tau}_{\mathcal{E}} = \hat{k}, \tau_{\mathcal{E}} = k) = \mathbb{P}(\hat{\tau}_{\mathcal{E}} = \hat{k})\mathbb{P}(\tau_{\mathcal{E}} = k). \tag{5.1}$$

Defining recursively

$$\begin{aligned}
&k_{\mathrm{i}}^1 := \min\{k \geq 0 : \mathbb{P}(\tau_{\mathcal{E}} = k) > 0\}, \\
&k_{\mathrm{i}}^c := \min\{k > k_{\mathrm{i}}^{c-1} : \mathbb{P}(\tau_{\mathcal{E}} = k) > 0, \mathbb{P}(\tau_{\mathcal{E}} = k-1) = 0\}, && c > 1 \\
&k_f^c := \min\{k > k_{\mathrm{i}}^c : \mathbb{P}(\tau_{\mathcal{E}} = k) > 0, \mathbb{P}(\tau_{\mathcal{E}} = k+1) = 0\}, && c \geq 1
\end{aligned}$$

then the support of $\tau_{\mathcal{E}}$, $supp(\tau_{\mathcal{E}})$, can be expressed as

$$supp(\tau_{\mathcal{E}}) = \bigcup_{c\geq 1} \{k_{\mathrm{i}}^c, k_{\mathrm{i}}^c + 1, ..., k_f^c\} \text{ (Union of disjoint sets in } \mathbb{N}).$$

Thus, the MSE of any estimator $\hat{\tau}_{\mathcal{E}}$ is lower bounded as

$$\mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})^2\} \geq \Theta_{\mathcal{E}} \cdot \left(1 - \sum_{c\geq 1} \left(\mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{\hat{\tau}_{\mathcal{E}}\} - (k_{\mathrm{i}}^c - 1)\right) \mathbb{P}(\tau_{\mathcal{E}} = k_{\mathrm{i}}^c)\right)^2 \tag{5.2}$$

where $\Theta_{\mathcal{E}}$ is well-defined and equivalent to the Bayesian Cramér-Rao lower bound in event prognosis:

$$\Theta_{\mathcal{E}} = \left(\sum_{k\in supp(\tau_{\mathcal{E}})} \left(\frac{\mathbb{P}(\tau_{\mathcal{E}} = k+1)}{\mathbb{P}(\tau_{\mathcal{E}} = k)} - 1\right)^2 \mathbb{P}(\tau_{\mathcal{E}} = k)\right)^{-1}. \tag{5.3}$$

As it can be seen in Eq. (5.2), a multiplying factor affects $\Theta_{\mathcal{E}}$. Nonetheless, this multiplying factor is generally almost equal to 1 provided the probability distribution of $\tau_{\mathcal{E}}$ tends to be "continuous" in $\partial supp(\tau_{\mathcal{E}})$. In other words, $\mathbb{P}(\tau_{\mathcal{E}} = k_{\mathrm{i}}^c) \approx 0$, $\forall c$.

5.1.2 Continuous-Time Stochastic Systems

Analogous to the case of discrete-time stochastic processes presented in Section 5.1.1, the continuous-time case refers to a probability density $p(\tau_{\mathcal{E}} = t)$, which statistically characterises $\tau_{\mathcal{E}}$, the random variable denoting the first occurrence time of an event $\mathcal{E}$.

Theorem 5.2 [Continuous-Time Bayesian Cramér-Rao Inequality] Let $\{X_t\}_{t\in\mathbb{R}_+\cup\{0\}}$ be a stochastic process and $\{E_t\}_{t\in\mathbb{R}_+}$ be an uncertain event process, respectively. According to Theorem 3.2, the first time of occurrence of an event $\mathcal{E}$ is characterised by the random variable $\tau_{\mathcal{E}} = \tau_{\mathcal{E}}(t_p)$, with $t_p \in \mathbb{R}_+ \cup \{0\}$. On the other hand, if $\hat{\tau}_{\mathcal{E}}$ is an estimator of $\tau_{\mathcal{E}}$, $\hat{\tau}_{\mathcal{E}} \perp \tau_{\mathcal{E}}$, then it can be defined

$$p^*(\hat{t}, t) = p(\hat{\tau}_{\mathcal{E}} = \hat{t}, \tau_{\mathcal{E}} = t) = p(\hat{\tau}_{\mathcal{E}} = \hat{t})p(\tau_{\mathcal{E}} = t). \tag{5.4}$$

If the following conditions hold:

1. $p(\tau_{\mathcal{E}} = t)$ is absolutely continuous and $\frac{\mathrm{d}}{\partial t}p(\tau_{\mathcal{E}} = t)$ is absolutely integrable with respect to $t \in \mathbb{R}_+$, this is
$$\int \left|\frac{\partial}{\mathrm{d}t}p(\tau_{\mathcal{E}} = t)\right|\mathrm{d}t < +\infty, \tag{5.5}$$
2. and, on the other hand,
$$\lim_{t\to+\infty} tp(\tau_{\mathcal{E}} = t) = 0, \tag{5.6}$$

then the MSE of any estimator $\hat{\tau}_{\mathcal{E}}$ is lower bounded as

$$\mathbb{E}_{p^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})^2\} \geq \Theta_{\mathcal{E}}, \tag{5.7}$$

where $\Theta_{\mathcal{E}}$ is well-defined and equivalent to the Bayesian Cramér-Rao lower bound in event prognosis:

$$\Theta_{\mathcal{E}} = \left(\mathbb{E}_{p(\tau_{\mathcal{E}}=\cdot)}\{\left(\frac{\mathrm{d}}{\mathrm{d}t}\log p(\tau_{\mathcal{E}} = t)\right)^2\}\right)^{-1}. \tag{5.8}$$

5.2 Illustrative Example: Battery End-of-Discharge Time Prognosis

The theoretical results shown in this chapter are examined in the light of the same case study and the same probabilistic prognostic algorithm considered in Chapter 4, which are presented in Section 4.4.1 and 4.4.2, respectively.

Results and analysis

Both Chapter 4 and this chapter develop BCRLB in the context of event prognosis; the former refers to the prediction of system states, whereas the latter refers to the first occurrence time of the event of interest. They are different, but still they have in common the use of BCRLBs and, therefore, they can be analysed in a similar way.

According to Theorem 5.1, and following the same analytical ideas exposed in Section 4.2, it can be noted that the multiplying factor of $\Theta_{\mathcal{E}}$ typically adopts a value near 1. Hence, the *Bayesian Cramér-Rao Inequality* can be approximated by the following expression:

$$\mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})^2\} \geq \Theta_{\mathcal{E}}, \tag{5.9}$$

where the MSE can be rewritten as:

$$\mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})^2\} = \mathbb{V}ar_{\mathbb{P}(\tau_{\mathcal{E}}=\cdot)}\{\tau_{\mathcal{E}}\} + \mathbb{V}ar_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{\hat{\tau}_{\mathcal{E}}\} + Bias(\hat{\tau}_{\mathcal{E}}, \tau_{\mathcal{E}})^2. \tag{5.10}$$

The value of $\Theta_{\mathcal{E}}$ is a lower bound for the MSE. Notwithstanding, it can be seen from Eq. (5.10) that if $\hat{\tau}_{\mathcal{E}}$ is an unbiased estimator of $\tau_{\mathcal{E}}$ with negligible variance, then the minimum achievable MSE corresponds exactly to the variance of $\tau_{\mathcal{E}}$. Indeed, it would be obtained

$$\mathbb{V}ar_{\mathbb{P}(\tau_{\mathcal{E}}=\cdot)}\{\tau_{\mathcal{E}}\} \geq \Theta_{\mathcal{E}}. \tag{5.11}$$

The random variable $\hat{\tau}_{\mathcal{E}}$ can be any sort of estimator given its independence from $\tau_{\mathcal{E}}$, but it should be defined wisely so that Eq. (5.9) becomes useful, where the best scenario is given by Eq. (5.11). An important question that arises now from Eq. (5.11) is: *how far is* $\mathbb{V}ar_{\mathbb{P}(\tau_{\mathcal{E}}=\cdot)}\{\tau_{\mathcal{E}}\}$ *from the lower bound* $\Theta_{\mathcal{E}}$*?*

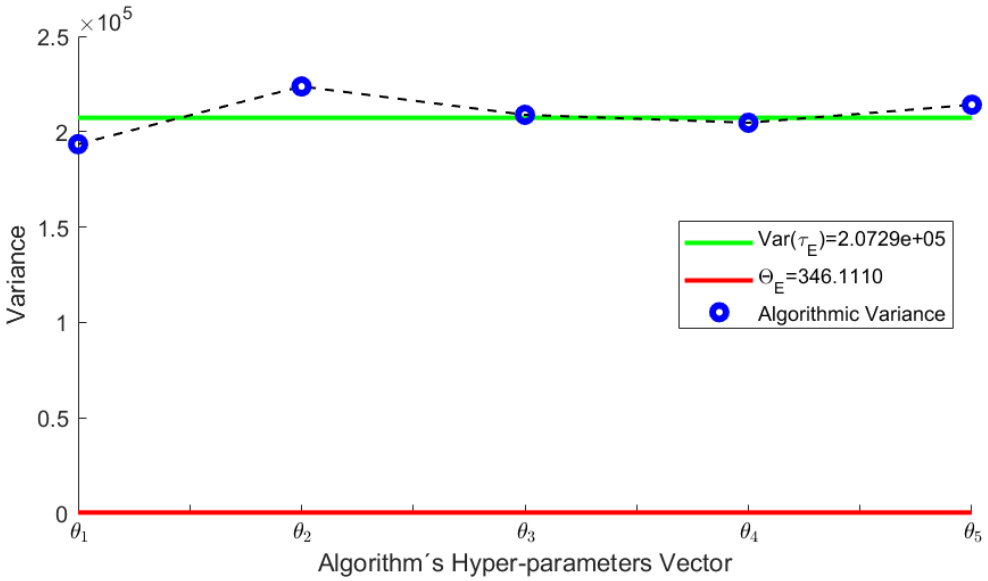


Figure 5.2: EoD time prognosis: variance estimations of $\tau_{\mathcal{E}}$ along the prognostic algorithm design described in Section 4.4.4. Each estimation assumes one of the hyper-parameters vector shown in Eq. (4.42). These results are compared to the actual variance of $\tau_{\mathcal{E}}$ and the BCRLB $\Theta_{\mathcal{E}}$, which are obtained via ten million Monte Carlo simulations.

Simulations results are shown in Fig. 5.2. Conditional to each of the hyper-parameters selected along the design methodology for prognostic algorithms in Section 4.4.4, estimators of the variance $\mathbb{V}ar_{\mathbb{P}(\tau_{\mathcal{E}}=\cdot)}\{\tau_{\mathcal{E}}\}$ are computed and yield quite close to the actual variance of $\tau_{\mathcal{E}}$ as the design goes on. The most important observation is how distant is the BCRLB $\Theta_{\mathcal{E}}$

from $\mathbb{V}ar_{\mathbb{P}(\tau_{\mathcal{E}}=\cdot)}\{\tau_{\mathcal{E}}\}$. This fact indicates at a first glance that the BCRLB presented in this chapter cannot be used as criteria for prognostic algorithm design.

Despite interpretation issues, perhaps the main constraint associated with the potential use of $\Theta_{\mathcal{E}}$, is that it is not straightforward to compute [130]. Furthermore, there is ongoing research on how to compute BCRLBs [131] in the state estimation setting [123–125, 128]. However, this concept is completely novel in the discipline of prognostics, provided that there was no proper theoretical basis on which this concept could have been developed before (in terms of the semi-closed form formalisation of uncertain event prognosis presented in Chapter 3), until the results presented in [78] that were further improved in this thesis.

As it has been shown recently, $\Theta_{\mathcal{E}}$ for the system studied hereby represents an interesting mathematical result, although in practice this bound and the actual variance of $\tau_{\mathcal{E}}$ are not necessarily tight, but the opposite. Nevertheless, it might still have potential use in objective functions for optimisation purposes, for example, or enable the access to an information theoretic discussion of prognostics that is part of future research efforts.

5.3 Summary

In this chapter it has been presented new theoretical results that expand the theory of uncertain event prognosis recently formalised in a rigorous fashion in Chapter 3, which are *Bayesian Cramér-Rao lower bounds* for the MSE of estimators of the first occurrence time of uncertain events in both, discrete- and continuous-time dynamic systems. The results are proven throughout mathematical demonstrations and a practical illustration is analysed in a case study of EoD time prognosis.

The illustrative example is constituted from a very simplified model for the State-of-Charge of batteries. Simulations show that the theoretical bounds presented hereby seem to be far distant from the MSE levels reached by a specific prognostic algorithm under different hyper-parameter configurations. Therefore, the use of these bounds as performance metrics in uncertain event prognosis does not seem evident at a first glance. These theoretical results expand and enrich the current theory of uncertain event prognosis and may provide useful information under certain conditions, thus enabling a wider information theoretic analysis and discussion.

Chapter 6

Conclusion

Event prognosis is a problem that has caught the attention of researchers from several scientific disciplines for more than fifty years. Regarding probabilistic approaches, most of the research efforts have focused on finding closed-form mathematical expressions to the probability distribution of the first occurrence time of events. In this regard, related literature has been enriched with problem-specific solutions that have been mostly reported for continuous-time systems, where a common factor has been to declare the occurrence of an event when some variables first cross particular thresholds.

The motivation of this thesis relies on the interest in predicting catastrophic failures. These are part of one the most important types of event to be addressed, for several reasons ranging from inefficient maintenance of infrastructure, to the protection of people's physical integrity. This research field is motivated by a practical need focused on applications, which has led to adopting non-rigorous heuristic approaches that have neglected the formal use of mathematics. Moreover, most of the efforts have been put rather to diagnose failures than to prognose them; the scientific community has been in debt in this regard.

The main contribution presented in this thesis is precisely a rigorous formalisation of the event prognosis problem, seeking to establish a consistent and general theoretical basis that has been lacking over some decades. The detail of the contributions to the state-of-the-art presented in this thesis is as follows:

- Before proposing any changes, it is appropriate to analyse the correctness of current approaches. Regarding failure prognosis, some arguments are presented in this thesis to support the idea that state-of-the-art procedures present serious theoretical inconsistencies, which motivates a rethinking of theoretical bases.
- Although the literature regarding event prognosis is quite rich, it is mostly limited to continuous-time systems (being mainly motivated by physical phenomena), but also provides solutions to very specific problems. The second contribution of this thesis is the derivation and formal demonstration of new semi-closed probability measures for the time when an event first occurs; for both discrete- and continuous-time systems. The quality of being semi-closed expressions gives greater generality, since there is no restriction regarding the characteristics of the stochastic process that models system

dynamics.

- At least as regarding failure prognosis, the idea of "*uncertain hazard zone*" was conceived several years ago in the scientific community, but it was never formalised and still gives glimpses of the inconsistency of the probability measures used so far. The third contribution of this thesis is the introduction of a new notion of "*uncertain event*", which is rigorously proposed and used successfully in a case study of fatigue crack growth. This new notion formalises the idea of "*uncertain hazard zone*" perfectly and in a consistent manner.
- Having formalised the event prognosis problem, it is interesting to study the existence of fundamental limits since they can account for the best possible performance that prognostic algorithms might achieve. In this thesis, fundamental limits are derived based on the concept of *Bayesian Cramér-Rao lower bounds* to limit the maximum accuracy and precision with which future states of a system can be predicted. These limits are rigorously demonstrated and used in a methodology that is proposed for the design of prognostic algorithms, which seeks to obtain the best quality of results conditional to limited computational resources. All of this theory is later successfully applied to the problem of *End-of-Discharge* time prognosis of Lithium-Ion batteries.
- Another interesting idea that emerges from the previous point is the following: What about the maximum accuracy and precision with which the future time of occurrence of an event can be estimated? Using once again the concept of *Bayesian Cramér-Rao lower bounds*, fundamental limits are formally derived that attempt to answer the recently asked question, but unfortunately the limits lie well below the maximum attainable precision and do not provide the expected utility. Nonetheless, the existence of such fundamental limits and their potential use in future research are still valuable contributions. These could allow a greater deepening in the understanding of theoretical aspects of event prognosis, like their potential use in objective functions for optimisation purposes, to give an example.

The contributions presented in this doctoral thesis are mainly of a theoretical nature and propose substantial changes to the theoretical bases with which scientists have been educated over several years around the world. It is a stimulating paradigm shift that requires a critical vision and that also evokes the incorporation of greater rigour in research. The paradigm changes are always difficult, but it is still expected to happen over the next few years. Ideas and intuition can be discussed, but one of the advantages of mathematics is that, given certain assumptions, there are implications that constitute absolute truths, philosophically speaking.

Bibliography

[1] R. Patrick-Aldaco. A Model Based Framework for Fault Diagnosis and Prognosis of Dynamical Systems with an Application to Helicopter Transmissions. PhD thesis, Georgia Institute of Technology, 2007.

[2] D. J. Inman, C. R. Farrar, V. Lopes Junior, and V. Steffen Junior. Damage Prognosis For Aerospace, Civil and Mechanical Systems. John Wiley & Sons, 2005.

[3] D. A. Pola, H. F. Navarrete, M. E. Orchard, R. S. Rabié, M. A. Cerda, B. E. Olivares, J. F. Silva, P. A. Espinoza, and A. Pérez. 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, 2015.

[4] B. Zhang, L. Tang, J. DeCastro, M. Roemer, and K. Goebel. Autonomous vehicle battery state-of-charge prognostics enhanced mission planning. International Journal of Prognostics and Health Management, 8, 2014.

[5] J. Lee, C. Jin, Z. Liu, and H. Davari Ardakani. Introduction to data-driven methodologies for prognostics and health management. In S. Ekwaro-Osire, A. Gonçalves, and F. Alemayehu, editors, Probabilistic Prognostics and Health Management of Energy Systems, pages 9–32. Springer, Cham, 2017.

[6] S. Redner. A guide to first-passage processes. Cambridge University Press, Cambridge, 2001.

[7] R. Metzler, G. Oshanin, , and S. Redner. First-passage phenomena and their applications. World Scientific, Singapore, 2014.

[8] J. Drugowitsch. Fast and accurate Monte Carlo sampling of first-passage times from Wiener diffusion models. Scientific Reports, 6, 2016.

[9] S. Herrmann and C. Zucca. Exact simulation of the first-passage time of diffusions. arXiv e-prints, page arXiv:1705.06881, 2017.

[10] M.-L. T. Lee and G. A. Whitmore. Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary. Statistical Science, 21(4):501–513, 2006.

[11] A. Lipton and V. Kaushansky. On the first hitting time density of an Ornstein-

Uhlenbeck process. arXiv e-prints, page arXiv:1810.02390, 2018.

[12] T. Lancaster. A stochastic model for the duration of a strike. J. Roy. Statist. Soc. Ser. A, 135(2):257–271, 1972.

[13] G. A. Whitmore. First-passage-time models for duration data: Regression structures and competing risks. Journal of the Royal Statistical Society. Series D (The Statistician), 35(2):207–219, 1986.

[14] M. A. Muñoz, G. Grinstein, and Y. Tu. Survival probability and field theory in systems with absorbing states. Phys. Rev. E, 56(5):5101–5105, 1997.

[15] L. V. Ballestra and G. Pacelli. Computing the survival probability density function in jump-diffusion models: A new approach based on radial basis functions. Engineering Analysis with Boundary Elements, 35(9):1075–1084, 2011.

[16] N. Levernier, M. Dolgushev, O. Bénichou, R. Voituriez, and T. Guérin. Survival probability of stochastic processes beyond persistence exponents. Nature Communications, 10(2990), 2019.

[17] M. G. Pecht and M. Kang. Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things. Wiley-IEEE Press, 2018.

[18] K. Goebel, Ma. Daigle, A. Saxena, S. Sankararaman, I. Roychoudhury, and J. Celaya. Prognostics: The Science of Making Predictions. 2017.

[19] G. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess, and B. Wu. Intelligent Fault Diagnosis and Prognosis for Engineering Systems. Wiley, 2006.

[20] A. Szabo, K. Schulten, and Z. Schulten. First passage time approach to diffusion controlled reactions. The Journal of Chemical Physics, 72(8):4350–4357, 1980.

[21] M. Razbin, P. Benetatos, and A. A. Moosavi-Movahedi. A first-passage approach to the thermal breakage of a discrete one-dimensional chain. Soft Matter, 15:2469–2478, 2019.

[22] M. Nyberg, T Ambjörnsson, and L. Lizana. A simple method to calculate first-passage time densities with arbitrary initial conditions. New Journal of Physics, 18, 2016.

[23] S. Ahmad, I. Nayak, A. Bansal, A. Nandi, and D. Das. First passage of a particle in a potential under stochastic resetting: A vanishing transition of optimal resetting rate. Phys. Rev. E, 99(2):022130, 2019.

[24] Z. Hu, L. Cheng, and B. J. Berne1. First passage time distribution in stochastic processes with moving and static absorbing boundaries with application to biological rupture experiments. The Journal of Chemical Physics, 133(3), 2010.

[25] T. Chou and M. R. D'Orsogna. First passage problems in biology. arXiv e-prints, page

arXiv:1408.4518, 2014.

[26] H. C. Tuckwell. Introduction to theoretical neurobiology. Volume 2: Nonlinear and Stochastic Theories. Cambridge University Press, Cambridge, 1988.

[27] D. Holcman, A. Marchewka, and Z. Schuss. Survival probability of diffusion with trapping in cellular neurobiology. Phys. Rev. E, 72(3), 2005.

[28] A. L. Lloyd and R. M. May. How viruses spread among computers and people. Science, 292(5520):1316–1317, 2011.

[29] D. J. Navarro and I. G. Fuss. Fast and accurate calculations for first-passage times in Wiener diffusion models. Journal of Mathematical Psychology, 53(4):967–980, 2009.

[30] G. Bakshi and G. Panayotov. First-passage probability, jump models, and intra-horizon risk. Journal of Financial Economics, 95:20–40, 2010.

[31] J. Janssen, O. Manca, and R. Manca. Applied Diffusion Processes from Engineering to Finance. Wiley-ISTE, Hoboken, 2013.

[32] M. Jaskowski and D. van Dijk. First-passage-time in discrete time. Research Seminar, Erasmus School of Economics (ESE), Erasmus University Rotterdam, 2015.

[33] J. H. Abbring and T. Salimans. The likelihood of mixed hitting times. arXiv e-prints, page arXiv:1905.03463, 2019.

[34] Y. S. Sherif and M. L. Smith. First-passage time distribution of Brownian motion as a reliability model. IEEE Transactions on Reliability, R-29(5):425–426, 1980.

[35] V. Pieper, M. Dominé, and P. Kurth. Level crossing problems and drift reliability. Mathematical Methods of Operations Research, 45(3):347–354, 1997.

[36] V. Srivastava, S. F. Feng, J. D. Cohen, N. Ehrich Leonard, and A. Shenhav. A martingale analysis of first passage times of time-dependent Wiener diffusion models. Journal of Mathematical Psychology, 77:94–110, 2017.

[37] O. Artime, N. Khalil, R. Toral, and M. San Miguel. First-passage distributions for the one-dimensional Fokker-Planck equation. Phys. Rev. E, 98(4), 2018.

[38] C. Buchera and M. Di Paola. Efficient solution of the first passage problem by path integration for normal and Poissonian white noise. Probabilistic Engineering Mechanics, 41:121–128, 2015.

[39] W. Li, W. Xu, J. Zhao, and Y. Jin. First-passage problem for strong nonlinear stochastic dynamical systems. Chaos, Solitons & Fractals, 28(2):414–421, 2006.

[40] M. J. Kearney and S. N. Majumdar. On the area under a continuous time Brownian motion till its first-passage time. Journal of Physics A: Mathematical and General, 38 (19), 2005.

[41] H. Li and M. Shaked. On the first passage times for Markov processes with monotone convex transition kernels. Stochastic Processes and their Applications, 58(2):205–216, 1995.

[42] J. B. Roberts. First-passage probabilities for randomly excited systems: Diffusion methods. Probabilistic Engineering Mechanics, 1(2):66–81, 1986.

[43] M. Burnashev and A. Tchamkerten. Estimating a random walk first-passage time from noisy or delayed observations. IEEE Transactions on Information Theory, 58(7):4230–4243, 2012.

[44] S. N. Majumdar. Universal first-passage properties of discrete-time random walks and Lévy flights on a line: Statistics of the global maximum and records. Physica A: Statistical Mechanics and its Applications, 389(20):4299–4316, 2010.

[45] H. Li and M. Shaked. Ageing first-passage times of Markov processes: A matrix approach. Journal of Applied Probability, 34(1):1–13, 1997.

[46] T. Koski, B. Jung, and G. Högnäs. Exit times for ARMA processes. Advances in Applied Probability, 50(A):191–195, 2018.

[47] C. Baumgarten. Survival probabilities of autoregressive processes. ESAIM: Probability and Statistics, 18:145–170, 2014.

[48] B. Jung. Exit times for multivariate autoregressive processes. Stochastic Processes and their Applications, 123(8):3052–3063, 2013.

[49] X. Guo, A. Hernández del Valle, and O. Hernández-Lerma. First passage problems for nonstationary discrete-time stochastic control systems. European Journal of Control, 18(6):528–538, 2012.

[50] E. Di Nardo. On the first passage time for autoregressive processes. Scientiae Mathematicae Japonicae, 67:137–152, 2010.

[51] A. Novikov and N. Kordzakhia. Martingales and first passage times of AR(1) sequences. Stochastics: An International Journal of Probability and Stochastic Processes, 80(2-3): 197–210, 2008.

[52] H. Larralde. A first passage time distribution for a discrete version of the Ornstein–Uhlenbeck process. Journal of Physics A: Mathematical and General, 37(12), 2004.

[53] G. K. Basak and K.-W. R. Ho. Level-crossing probabilities and first-passage times for linear processes. Advances in Applied Probability, 36(2):643–666, 2004.

[54] L. M. Ricciardi and S. Sato. First-passage-time density and moments of the Ornstein-Uhlenbeck process. Journal of Applied Probability, 25(1):43–57, 1988.

[55] V. Linetsky. Computing hitting time densities for CIR and OU diffusions: applications

to mean-reverting models. Journal of Computational Finance, 7(4):1–22, 2004.

[56] L. Alili, P. Patie, and J. L. Pedersen. Representations of the first hitting time density of an Ornstein-Uhlenbeck process. Stochastic Models, 21(4):967–980, 2005.

[57] H. R. Lerche. Boundary crossing of Brownian motion. Lecture Notes in Statistics, 40, 1986.

[58] S. Ghazizadeh, M. Barbato, and E. Tubaldi. New analytical solution of the first-passage reliability problem for linear oscillators. Journal of Engineering Mechanics, 138(6):695–706, 2012.

[59] J. Song and A. D. Kiureghian. Joint first-passage probability and reliability of systems under stochastic excitation. Journal of Engineering Mechanics, 132(1):65–77, 2006.

[60] S. K. Au and J. L. Beck. First excursion probabilities for linear systems by very efficient importance sampling. Probabilistic Engineering Mechanics, 16(3):193–207, 2001.

[61] A. Naess. Crossing rate statistics quadratic transformations of Gaussian processes. Probabilistic Engineering Mechanics, 16(3):209–217, 2001.

[62] E. Di Nardo, A. G. Nobile, E. Pirozzi, and L. M. Ricciardi. A computational approach to first-passage-time problems for Gauss-Markov processes. Advances in Applied Probability, 33(2):453–482, 2001.

[63] V. Bayer and C. Bucher. Importance sampling for the first passage problems of nonlinear structures. Probabilistic Engineering Mechanics, 14(1-2):27–32, 1999.

[64] S. Engelund, R. Rackwitz, and C. Lange. Approximations of first-passage times for differentiable processes based on higher-order threshold crossings. Probabilistic Engineering Mechanics, 10(1):53–60, 1995.

[65] R. S. Langley. A first passage approximation for normal stationary random processes. Journal of Sound and Vibration, 122(2):261–275, 1988.

[66] P. H. Madsen and S. Krenk. An integral equation method for the first-passage problem in random vibration. J. Appl. Mech., 51(3):674–679, 1984.

[67] P. Hänggi and P. Talkner. Non-Markov processes: The problem of the mean first passage time. Zeitschrift für Physik B Condensed Matter, 45(1):79–83, 1981.

[68] S. H. Crandall. First-cross probabilities of the linear oscillator. Journal of Sound and Vibration, 12(3):285–299, 1970.

[69] J. J. Coleman. Reliability of aircraft structures in resisting chance failure. Probabilistic Engineering Mechanics, 7(5):639–645, 1959.

[70] D. A. Darling and A. J. F. Siegert. The first passage problem for a continuous Markov process. The Annals of Mathematical Statistics, 24(4):624–639, 1953.

[71] B. Sildnes and B. H. Lindqvist. Modeling of semi-competing risks by means of first passage times of a stochastic process. Lifetime Data Analysis, 24(1):153–175, 2018.

[72] Y. Shu, Q. Feng, and D. W. Coit. Life distribution analysis based on Lévy subordinators for degradation with random jumps. Naval Research Logistics (NRL), 62(6):483–492, 2015.

[73] C. Paroissin and A. Salami. Failure time of non homogeneous Gamma process. Communications in Statistics - Theory and Methods, 43(15):3148–3161, 2014.

[74] A. Lehmann. Joint modeling of degradation and failure time data. Journal of Statistical Planning and Inference, 139(5):1693–1706, 2009.

[75] Y. Yang and G.-A. Klutke. Lifetime-characteristics and inspection-schemes for Lévy degradation processes. IEEE Transactions on Reliability, 49(4):377–382, 2000.

[76] M. Abdel-Hameed. Life distribution properties of devices subject to a Lévy wear process. Mathematics of Operations Research, 9(4):479–634, 1984.

[77] M. Orchard and G. Vachtsevanos. A particle-filtering approach for on-line fault diagnosis and failure prognosis. Trans. Inst. Meas. Control 2009, 31:221–246, 2009.

[78] D. E. Acuña and M. E. Orchard. A theoretically rigorous approach to failure prognosis. Annual Conference of the Prognostics and Health Management Society 2018, 10(1), 2018.

[79] D. E. Acuña and M. E. Orchard. Particle-filtering-based failure prognosis via sigma-points: Application to lithium-ion battery state-of-charge monitoring. Mech. Syst. Signal Process., 85:827–848, 2017.

[80] D. E. Acuña, R. J. Saona, and M. E. Orchard. Conditional predictive Bayesian Cramér-Rao lower bounds for prognostic algorithms design. Applied Soft Computing, 72:647–665, 2018.

[81] M. Daigle and K. Goebel. Model-based prognostics with concurrent damage progression processes. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(4):535–546, 2013.

[82] M. Daigle and S. Sankararaman. Advanced methods for determining prediction uncertainty in model-based prognostics with application to planetary rovers. Annual Conference of the Prognostics and Health Management Society 2013, 2013.

[83] A. Birolini. Reliability Engineering: Theory and Practice. New York: Springer-Verlag, 2007.

[84] K. Yang and J. Xue. Continuous state reliability analysis. 1996 Annual Reliability and Maintainability Symposium (RAMS '96), International Symposium on Product Quality and Integrity, pages 251–257, 1996.

[85] P. Lall, M. Pecht, and E. Harkim. Influence of Temperature on Microelectronics and System Reliability. CRC Press, New York, 1997.

[86] M. Pecht, D. Das, and A. Ramakrishnan. The IEEE standards on reliability program and reliability prediction methods for electronic equipment. Microelectronic Reliability, 42(9):1259–1266, 2002.

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

[88] N. Vichare, P. Rodgers, V. Eveloy, and M. Pecht. 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, 2004.

[89] Z. Chen and S. Zheng. Lifetime distribution based degradation analysis. IEEE Transactions on Reliability, 54(1):3–10, 2005.

[90] J. P. Kharoufeh and S. M. Cox. Stochastic models for degradaton-based reliability. IIE Transactions, 37:533–542, 2005.

[91] D. Xu and W. Zhao. Reliability prediction using multivariate degradation data. Annual Reliability and Maintainability Symposium (RAMS 2005), pages 337–341, 2005.

[92] R. Ahmad and S. Kamaruddin. An overview of time-based and condition-based maintenance in industrial application. Computers & Industrial Engineering, 63(1):135–149, 2012.

[93] D. Dubois. Possibility theory and statistical reasoning. Computational Statistics & Data Analysis, 51:47–96, 2006.

[94] G. Shafer. A Mathematical Theory of Evidence. N.J: Princeton University Press, Princeton, 1976.

[95] L. A. Zadeh. Probability measures of fuzzy events. Journal of Mathematical Analysis and Applications, 23(2):421–427, 1968.

[96] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, pages 9–34, 1977.

[97] D. Liu, V. Luo, and Y. Peng. Uncertainty processing in prognostics and health management: An overview. 2012 IEEE Conference on Prognostics and System Health Management (PHM), 2012.

[98] A. Doucet, N. de Freitas, and N. Gordon, editors. Sequential Monte Carlo Methods in Practice. New York: Springer-Verlag, 2001.

[99] U. D. Hanebeck, K. Briechle, and A. Rauh. Progressive Bayes: A new framework for nonlinear state estimation. Proc. of the SPIE AeroSense Symposium 2003, 5099,

Orlando, USA, page , 2003.

[100] A. Rauh, K. Briechle, and U. D. Hanebeck. Nonlinear measurement update and prediction: Prior density splitting mixture estimator. 2009 IEEE Control Applications, (CCA) & Intelligent Control, (ISIC), St. Petersburg, pages 1421–1426, 2009.

[101] U. D. Hanebeck, M. F. Huber, and V. Klumpp. Dirac mixture approximation of multivariate Gaussian densities. Proceedings of the 48th IEEE Conference on Decision and Control, 2009, held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009., pages 3851–3858, 2009.

[102] R. E. Kalman. A new approach to linear filtering and prediction problems. J. Basic Eng., 82(1):35–45, 1960.

[103] R. Faragher. Understanding the basis of the Kalman filter via a simple and intuitive derivation [lecture notes]. IEEE Signal Processing Magazine, 29(5):128–132, 2012.

[104] S. Engel, B. Gilmartin, K. Bongort, and A. Hess. Prognostics, the real issues involved with predicting life remaining. Proc. 2000 IEEE Aerospace Conf., 6:457–469, 2000.

[105] D. J. C. MacKay. Introduction to Monte Carlo methods. Learning in Graphical Models, 1998.

[106] M. Orchard, G. Kacprzynski, K. Goebel, B. Saha, and G. Vachtsevanos. Advances in uncertainty representation and management for particle filtering applied to prognostics. Proc. Int. Conf. Prognostics and Health Manag., pages 1–6, 2008.

[107] A. Dawn, J.-H. Choi, and N. H. Kim. Prognostics 101: a tutorial for particle filter-based prognostics algorithm using matlab. Reliability Engineering and System Safety, pages 161–169, 2013.

[108] E. Zio and G. Peloni. Particle filtering prognostic estimation of the remaining useful life of nonlinear components. Reliability Engineering and System Safety, pages 403–409, 2010.

[109] A. S. Sarathi, B. Long, and M. Pecht. Prognostics method for analog electronic circuits. Annual Conference of the Prognostics and Health Management Society 2012, pages 1–7, 2012.

[110] Y. Hu, P. Baraldi, F. Di Maio, and E. Zio. A particle filtering and kernel smoothing-based approach for new design component prognostics. Reliability Engineering and System Safety, pages 19–31, 2014.

[111] M. Yu, D. Wang, A. Ukil, V. Vaiyapuri, and N. Sivakumar. Model-based failure prediction for electric machines using particle filter. 13th International Conference on Control, Automation, Robotics & Vision, pages 1811–1816, 2014.

[112] M. Orchard, M. Cerda, B. Olivares, and J. Silva. Sequential Monte Carlo methods for discharge time prognosis in lithium-ion batteries. International Journal of Prognostics

and Health Management, 3:1–12, 2012.

[113] B. E. Olivares, M. A. Cerda, M. E. Orchard, and J. F. Silva. Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena. IEEE Transactions on Instrumentation and Measurement, 62(2):364–376, 2013.

[114] D. Acuña, M. Orchard, J. Silva, and A. Pérez. Multiple-imputation-particle-filtering for uncertainty characterization in battery state-of-charge estimation problems with missing measurement data: Performance analysis and impact on prognostic algorithms. International Journal of Prognostics and Health Management, 6(008):1–12, 2015.

[115] C. Chen, G. Vachtsevanos, and M. Orchard. Machine remaining useful life prediction: an integrated adaptive neuro-fuzzy and high-order particle filtering approach. Mechanical Systems and Signal Processing, 28:597–607, 2012.

[116] K. Goebel and P. Bonissone. Prognostic information fusion for constant load systems. 7th Annual Conference on Information Fusion, 2005.

[117] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel. Evaluating algorithmic performance metrics tailored for prognostics. IEEE Aerospace Conference, 2009.

[118] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel. On applying the prognostics performance metrics. Annual Conference of the Prognostics and Health Management Society, 2009.

[119] K. Goebel, A. Saxena, S. Saha, B. Saha, and J. Celaya. Data Mining in Systems Health Management: Detection, Diagnostics, and Prognostics, chapter Prognostic Performance Metrics, pages 149–171. Chapman & Hall/CRC Press, 2011.

[120] C. R. Rao. Information and the accuracy attainable in the estimation of statistical parameters. Bulletin of the Calcutta Mathematical Society, 37:81–89, 1945.

[121] H. Cramér. Mathematical Methods of Statistics. Princeton, NJ: Princeton Univ. Press, 1st edition, 1946.

[122] H. L. Van Trees. Detection, Estimation and Modulation Theory, volume 1. New York: Wiley, 1968.

[123] C. Fritsche, E. Ozkan, L. Svensson, and F. Gustafsson. A fresh look at Bayesian Cramér-Rao bounds for discrete-time nonlinear filtering. 17th International Conference on Information Fusion - FUSION 2014, 2014.

[124] P. Tichavský, C. H. Muravchik, and A. Nehorai. Posterior Cramér-Rao bounds for discrete-time nonlinear filtering. IEEE Trans. Signal Process., 46(5):1386–1396, 1998.

[125] L. Zuo, R. Niu, and P. Varshney. Conditional posterior Cramér-Rao lower bounds for nonlinear sequential Bayesian estimation. IEEE Transactions on Signal Processing, 59, 2011.

[126] F. Cadini, E. Zio, and D. Avram. Monte Carlo-based filtering for fatigue crack growth estimation. Probabilistic Engineering Mechanics, 24(3):367–373, 2009.

[127] M. E. Orchard and D. E. Acuña. On prognostic algorithm design and fundamental precision limits in long-term prediction. In E. Lughofer and M. Sayed-Mouchaweh, editors, Predictive Maintenance in Dynamic Systems: Advanced Methods, Decision Support Tools and Real-World Applications, pages 355–379. Springer, Cham, 2019.

[128] M. Šimandl, J. Královec, and P. Tichavský. Filtering, predictive, and smoothing Cramér-Rao bounds for discrete-time nonlinear dynamic systems. Automatica, 37: 1703–1716, 2001.

[129] C. Burgos, M. E. Orchard, M. Kazerani, R. Cárdenas, and D. Sáez. Particle-filtering-based estimation of maximum available power state in lithium-ion batteries. Applied Energy, 161:349–363, 2016.

[130] J. Dauwels. Computing Bayesian Cramér-Rao bounds. Proceedings of the International Symposium on Information Theory, 2005.

[131] L. Bacharach, C. Fritsche, U. Orguner, and E. Chaumette. A tighter Bayesian Cramér-Rao bound. 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.

[132] T. K. Moon and W. C. Stirling. Mathematical methods and algorithms for signal processing. Upper Saddle River, NJ: Prentice-Hall; London: Prentice-Hall International (UK), 2000.

Appendix A

Demonstrations: Chapter 4

A.1 Proof of Theorem 4.1

PROOF. (Theorem 4.1)

Let $\hat{x}^{\mathrm{i}} = \hat{x}^{\mathrm{i}}(y_{1:k_p})$ be an estimator of x^{i}. Since p_k^{cp} is absolutely continuous, $\frac{\partial p_k^{cp}}{\partial x^{\mathrm{i}}}$ is well-defined, moreover, by hypothesis, it is absolutely integrable. Therefore,

$$\frac{\partial}{\partial x^{\mathrm{i}}} \int\limits_{\mathbb{R}^{(k-k_p+1)n_x}} p_k^{cp}(x_{k_p:k})\mathrm{d}x_{k_p:k} = \int\limits_{\mathbb{R}^{(k-k_p+1)n_x}} \frac{\partial p_k^{cp}}{\partial x^{\mathrm{i}}}(x_{k_p:k})\mathrm{d}x_{k_p:k} \tag{A.1}$$

$$= \int\limits_{\mathbb{R}^{(k-k_p+1)n_x-1}} (p_k^{cp}(x_{k_p:k\setminus\mathrm{i}}, x^{\mathrm{i}}))\Big|_{-\infty}^{+\infty}\mathrm{d}x_{k_p:k\setminus\mathrm{i}} \tag{A.2}$$

$$= 0\,. \tag{A.3}$$

where the last identity is implied by the hypothesis of

$$\lim_{x^{\mathrm{i}}\to+\infty} x^{\mathrm{i}}p(x_{k_p:k}) \;=\; \lim_{x^{\mathrm{i}}\to-\infty} x^{\mathrm{i}}p(x_{k_p:k}) = 0$$

Then, multiplying by $\hat{x}^{\mathrm{i}}(y_{1:k_p})$ it is obtained

$$\int \hat{x}^{\mathrm{i}}(y_{1:k_p})\frac{\partial p_k^{cp}}{\partial x^{\mathrm{i}}}(x_{k_p:k})\mathrm{d}x_{k_p:k} = 0. \tag{A.4}$$

On the other hand, integrating by parts one gets that

$$\int_{-\infty}^{+\infty} x^{\mathrm{i}}\frac{\partial p_k^{cp}}{\partial x^{\mathrm{i}}}\mathrm{d}x^{\mathrm{i}} = (x^{\mathrm{i}}p_k^{cp})\Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} p_k^{cp}\mathrm{d}x^{\mathrm{i}}\,. \tag{A.5}$$

Due to the second condition (see Eq.(4.6)), $(x^{\mathrm{i}}p_k^{cp})\Big|_{-\infty}^{+\infty} = 0$. Integrating with respect to $x_{k_p:k\setminus\mathrm{i}}$ ($x_{k_p:k}$ omitting the i-th element), then

$$\int x^{\mathrm{i}}\frac{\partial p_k^{cp}}{\partial x^{\mathrm{i}}}\mathrm{d}x_{k_p:k} = -1\,. \tag{A.6}$$

Subtracting Eq. (A.6) to Eq. (A.4) it yields

$$\int (\hat{x}^{\mathrm{i}}(y_{1:k_p}) - x^{\mathrm{i}}) \frac{\partial p_k^{cp}}{\partial x^{\mathrm{i}}} \mathrm{d}x_{k_p:k} = \int (\hat{x}^{\mathrm{i}}(y_{1:k_p}) - x^{\mathrm{i}}) \frac{\partial \log p_k^{cp}}{\partial x^{\mathrm{i}}} p_k^{cp} \mathrm{d}x_{k_p:k} = 1\,. \tag{A.7}$$

Similarly, for $j \in \{1, 2, ..., (k - k_p + 1)n_x\} \setminus \{\mathrm{i}\}$ it can be obtained

$$\int (\hat{x}^{\mathrm{i}}(y_{1:k_p}) - x^{\mathrm{i}}) \frac{\partial \log p_k^{cp}}{\partial x^{j}} p_k^{cp} \mathrm{d}x_{k_p:k} = 0\,. \tag{A.8}$$

Thus, combining Eqs. (A.7)-(A.8) in a matrix form

$$\int (\hat{x}_{k_p:k}(y_{1:k_p}) - x_{k_p:k}) \Big[\nabla_{x_{k_p:k}} \log p_k^{cp}\Big] p_k^{cp} \mathrm{d}x_{k_p:k} = I_{(k-k_p+1)n_x}, \tag{A.9}$$

where $I_{(k-k_p+1)n_x}$ is the $(k - k_p + 1)n_x$-dimensional identity matrix. Pre-multiplying and post-multiplying the last equation by a^T and b, $a, b \in \mathbb{R}^{(k-k_p+1)n_x}$, respectively

$$\begin{aligned} a^T b &= \int a^T (\hat{x}_{k_p:k}(y_{1:k_p}) - x_{k_p:k}) \Big[\nabla_{x_{k_p:k}} \log p_k^{cp}\Big] b p_k^{cp} \mathrm{d}x_{k_p:k} && \text{(A.10)} \\ &= \int a^T (\hat{x}_{k_p:k}(y_{1:k_p}) - x_{k_p:k}) \sqrt{p_k^{cp}} \Big[\nabla_{x_{k_p:k}} \log p_k^{cp}\Big] b \sqrt{p_k^{cp}} \mathrm{d}x_{k_p:k}\,. && \text{(A.11)} \end{aligned}$$

Applying Cauchy-Schwarz inequality

$$\begin{aligned} (a^T b)^2 \;\leq\; & \Big(\int a^T (\hat{x}_{k_p:k}(y_{1:k_p}) - x_{k_p:k})(\hat{x}_{k_p:k}(y_{1:k_p}) - x_{k_p:k})^T a p_k^{cp} \mathrm{d}x_{k_p:k} \Big) \\ & \ldots \cdot \Big(\int \Big[\nabla_{x_{k_p:k}} \log p_k^{cp}\Big] b b^T \Big[\nabla_{x_{k_p:k}} \log p_k^{cp}\Big]^T p_k^{cp} \mathrm{d}x_{k_p:k} \Big)\,. \end{aligned} \tag{A.12}$$

Expressing this in terms of conditional expectation, it is obtained

$$\begin{aligned} (a^T b)^2 \;\leq\; & \Big(a^T \mathbb{E}_{p_k^{cp}}\{\tilde{x}_{k_p:k} \tilde{x}_{k_p:k}^T | y_{1:k_p}\} a \Big) \\ & \ldots \cdot \Big(\int \Big[\nabla_{x_{k_p:k}} \log p_k^{cp}\Big] b b^T \Big[\nabla_{x_{k_p:k}} \log p_k^{cp}\Big]^T p_k^{cp} \mathrm{d}x_{k_p:k} \Big)^T && \text{(A.13)} \\ = \; & \Big(a^T \mathbb{E}_{p_k^{cp}}\{\tilde{x}_{k_p:k} \tilde{x}_{k_p:k}^T | y_{1:k_p}\} a \Big) \\ & \ldots \cdot \Big(b^T \mathbb{E}_{p_k^{cp}}\{ \Big[\nabla_{x_{k_p:k}}{}^T \log p_k^{cp}\Big] \Big[\nabla_{x_{k_p:k}} \log p_k^{cp}\Big] \} b \Big)\,. && \text{(A.14)} \end{aligned}$$

Defining

$$I_{cp}(x_{k_p:k}|y_{1:k_p}) \triangleq \mathbb{E}_{p_k^{cp}}\{ \Big[\nabla_{x_{k_p:k}}{}^T \log p_k^{cp}\Big] \Big[\nabla_{x_{k_p:k}} \log p_k^{cp}\Big] \} \tag{A.15}$$

and choosing $b = I_{cp}^{-1}(x_{k_p:k}|y_{1:k_p}) a$, then

$$a^T \Big(\mathbb{E}_{p_k^{cp}}\{\tilde{x}_{k_p:k} \tilde{x}_{k_p:k}^T | y_{1:k_p}\} - I_{cp}^{-1}(x_{k_p:k}|y_{1:k_p}) \Big) a \geq 0\,. \tag{A.16}$$

Given that $a \in \mathbb{R}^{(k-k_p+1)n_x}$ is arbitrary, $\mathbb{E}_{p_k^{cp}}\{\tilde{x}_{k_p:k} \tilde{x}_{k_p:k}^T | y_{1:k_p}\} - I_{cp}^{-1}(x_{k_p:k}|y_{1:k_p})$ must necessarily be a semi-definite positive matrix. □

A.2 Proof of Theorem 4.2

Some previous results are needed before demonstrating Theorem 4.2, which were extracted from [132].

Theorem A.1 Defining the score function $s(\theta, z) = \nabla_\theta \log p(\theta|z)$ and taking $t(\cdot, \cdot)$ a vectorial function of z and θ with values in $\mathbb{R}^{n_\theta}$, and assuming some regularity, the following identity holds

$$\mathbb{E}_{p(\theta|z)}\{t(\theta, z)s(\theta, z)\} = \nabla_\theta \mathbb{E}_{p(\theta|z)}\{t(\theta, z)\} - \mathbb{E}_{p(\theta|z)}\{\nabla_\theta t(\theta, z)\}. \tag{A.17}$$

Proof.
Since

$$\mathbb{E}_{p(\theta|z)}\{t(\theta, z)\} = \int t(\theta, z)p(\theta|z)\mathrm{d}z, \tag{A.18}$$

the gradient operator can be applied at both sides and assuming that differentiation conditions are fulfilled so that the gradient can get into the integral:

$$\begin{aligned}
\nabla_\theta \mathbb{E}_{p(\theta|z)}\{t(\theta, z)\} &= \nabla_\theta \int t(\theta, z)p(\theta|z)\mathrm{d}z && \text{(A.19)}\\
&= \int [\nabla_\theta t(\theta, z)p(\theta|z)]\mathrm{d}z && \text{(A.20)}\\
&= \int t(\theta, z)[\nabla_\theta p(\theta|z)]\mathrm{d}z + \int [\nabla_\theta t(\theta, z)]p(\theta|z)\mathrm{d}z && \text{(A.21)}\\
&= \int t(\theta, z)[\nabla_\theta \log p(\theta|z)]p(\theta|z)\mathrm{d}z + \int [\nabla_\theta t(\theta, z)]p(\theta|z)\mathrm{d}z && \text{(A.22)}\\
&= \mathbb{E}_{p(\theta|z)}\{t(\theta, z)s(\theta, z)\} + \mathbb{E}_{p(\theta|z)}\{\nabla_\theta t(\theta, z)\} && \text{(A.23)}
\end{aligned}$$

□

Corollary A.1 If $s(\theta, z)$ is the score function of a **differentiable** likelihood $p(\theta|z)$, then

$$\mathbb{E}_{p(\theta|z)}\{s(\theta, z)\} = 0. \tag{A.24}$$

Proof.
By Theorem A.1, and because of Eq. (A.17), for each constant vector t:

$$\begin{aligned}
t\mathbb{E}_{p(\theta|z)}\{s(\theta, z)\} &= \mathbb{E}_{p(\theta|z)}\{ts(\theta, z)\} && \text{(A.25)}\\
&= \nabla_\theta \mathbb{E}_{p(\theta|z)}\{t\} - \mathbb{E}_{p(\theta|z)}\{\nabla_\theta t\} && \text{(A.26)}\\
&= 0. && \text{(A.27)}
\end{aligned}$$

As the latter expression is valid for all t, it follows that $\mathbb{E}_{p(z|\theta)}\{s(\theta, z)\} = 0$. □

Lemma A.1 If the **score** function $s(y_{1:k_p}, x_{k_p:k}) = \nabla_{x_{k_p:k}} \log p(x_k|y_{1:k_p})$ is differentiable, then $I_{cp}(x_{k_p:k}|y_{1:k_p})$ can be expressed as

$$I_{cp}(x_{k_p:k}|y_{1:k_p}) = \mathbb{E}_{p_k^{cp}}\{-\Delta_{x_{k_p:k}}^{x_{k_p:k}} \log p_k^{cp}\}. \tag{A.28}$$

PROOF.
By Theorem A.1 and Corollary A.1, substituting $\theta = x_{k_p:k}$ and $z = y_{1:k_p}$, then

$$\begin{aligned}&\mathbb{E}_{p_k^{cp}}\{t(y_{1:k_p}, x_{k_p:k})s(y_{1:k_p}, x_{k_p:k})\}\\ &\quad = \nabla_{x_{k_p:k}}\mathbb{E}_{p_k^{cp}}\{t(y_{1:k_p}, x_{k_p:k})\} - \mathbb{E}_{p_k^{cp}}\{\nabla_{x_{k_p:k}}t(y_{1:k_p}, x_{k_p:k})\}.\end{aligned} \tag{A.29}$$

Taking the function $t(y_{1:k_p}, x_{k_p:k}) = s(y_{1:k_p}, x_{k_p:k})^T = (\nabla_{x_{k_p:k}} \log p(x_k|y_{1:k_p}))^T$ and using the previous result, it follows that

$$\mathbb{E}_{p_k^{cp}}\{t(y_{1:k_p}, x_{k_p:k})s(y_{1:k_p}, x_{k_p:k})\} = -\mathbb{E}_{p_k^{cp}}\{\nabla_{x_{k_p:k}}t(y_{1:k_p}, x_{k_p:k})\}, \tag{A.30}$$

and the result is straightforward. □

PROOF. (Theorem 4.2)

Taking into account these previous results, it is now possible to prove Theorem 4.2:

Provided

$$\log p(x_{k_p:k}|y_{1:k_p}) = \log p(x_{k_p}|y_{1:k_p}) + \sum_{\mathrm{i}=k_p+1}^{k} \log p(x_{\mathrm{i}}|x_{\mathrm{i}-1}), \tag{A.31}$$

and considering that $D_{\mathrm{i}} = S_{\mathrm{i}}^{\mathrm{i}} + S_{\mathrm{i}+1}^{\mathrm{i}}$, $I_{cp}(x_{k_p:k}|y_{1:k_p})$ can be decomposed into a matrix of four blocks in the following manner

$$I_{cp}(x_{k_p:k}|y_{1:k_p}) = \left[\begin{array}{ccc|c} D_{k_p} & S_{k_p+1}^{k_p,k_p+1} & & \\ S_{k_p+1}^{k_p+1,k_p} & \ddots & \ddots & \\ & \ddots & D_{k-1} & S_k^{k-1,k} \\ \hline & & S_k^{k,k-1} & S_k^k \end{array}\right] = \left[\begin{array}{c|c} J_k^{11} & J_k^{12} \\ \hline J_k^{21} & J_k^{22} \end{array}\right], \tag{A.32}$$

where empty spaces represent zeros. Considering the previous definitions, it can be verified that

$$I_{cp}(x_{k_p:k}|y_{1:k_p}) = \left[\begin{array}{cc|c} J_{k-1}^{11} & J_{k-1}^{12} & \\ J_{k-1}^{21} & J_{k-1}^{22} + S_k^{k-1} & S_k^{k-1,k} \\ \hline & S_k^{k,k-1} & S_k^k \end{array}\right]. \tag{A.33}$$

On the other hand,

$$\mathbb{E}_{p_k^{cp}}\{\tilde{x}_{k_p:k}\tilde{x}_{k_p:k}^T|y_{1:k_p}\} \geq \left[\begin{array}{c|c} J_k^{11} & J_k^{12} \\ \hline J_k^{21} & J_k^{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} C_k^{11} & C_k^{12} \\ \hline C_k^{21} & C_k^{22} \end{array}\right]. \tag{A.34}$$

Then, the ECM associated to any estimator of x_k is lower bounded by

$$\mathbb{E}_{p_k^{cp}}\{\tilde{x}_k\tilde{x}_k^T|y_{1:k_p}\} \geq C_k^{22}. \tag{A.35}$$

Taking into account Eqs. (A.33)-(A.34), then the matrix inversion rule can be employed to conclude the demonstration:

$$[C_k^{22}]^{-1} \;=\; J_k^{22} - J_k^{21}[J_k^{11}]^{-1}J_k^{12} \tag{A.36}$$

$$= S_k^k - S_k^{k,k-1}[J_{k-1}^{22} + S_k^{k-1} - J_{k-1}^{21}[J_{k-1}^{11}]^{-1}J_{k-1}^{12}]^{-1}S_k^{k-1,k} \quad (A.37)$$

$$= S_k^k - S_k^{k,k-1}[[C_{k-1}^{22}]^{-1} + S_k^{k-1}]^{-1}S_k^{k-1,k}, \quad (A.38)$$

with the initial condition $[C_{k_p}^{22}]^{-1} = S_{k_p}^{k_p} = \mathbb{E}\{-\Delta_{x_{k_p}}^{x_{k_p}} \log p(x_{k_p}|y_{1:k_p})\}$. □

A.3 MCP-BRCLB Initial Condition and Recursion Elements

A.3.1 Initial Condition

Below it is developed a series of equations for obtaining the initial condition for the computation of the MCP-BCRLB recursive sequence:

$$[C_{k_p}^{22}] = \mathbb{E}\{-\Delta_{x_{k_p}}^{x_{k_p}} \log p(x_{k_p}|y_{1:k_p})\}^{-1}. \quad (A.39)$$

Following the indications reported in [123], it can be noted that

$$p(x_k|y_{1:k}) = \frac{p(x_k, y_k, y_{1:k-1})}{p(y_k, y_{1:k-1})} \quad (A.40)$$

$$= \frac{p(y_k|x_k, y_{1:k-1})p(x_k|y_{1:k-1})p(y_{1:k-1})}{p(y_k|y_{1:k-1})p(y_{1:k-1})} \quad (A.41)$$

$$= \frac{\overbrace{p(y_k|x_k)}^{likelihood}\overbrace{p(x_k|y_{1:k-1})}^{prior}}{\underbrace{p(y_k|y_{1:k-1})}_{evidence}} \quad (A.42)$$

Thus,

$$-\log p(x_k|y_{1:k}) = -\log p(y_k|x_k) - \log p(x_k|y_{1:k-1}) + \log p(y_k|y_{1:k-1}) \quad (A.43)$$

It is required to get second derivatives with respect to x_k before applying expectation. Let us proceed term by term:

- **Likelihood**:

$$-\log p(y_k|x_k) = -\log \frac{1}{\sqrt{2\pi}\sigma_\nu} \mathrm{e}^{-\frac{1}{2}\frac{\left(y_k - (v_{oc}(x_k) - I_k R_{int}(x_k, I_k))\right)^2}{\sigma_\nu^2}} \quad (A.44)$$

$$= c_0 + \frac{1}{2}\frac{\left(y_k - (v_{oc}(x_k) - I_k R_{int}(x_k, I_k))\right)^2}{\sigma_\nu^2} \quad (A.45)$$

$$\Rightarrow -\frac{\partial \log p(y_k|x_k)}{\partial x_k} = -\Big(\frac{y_k - (v_{oc}(x_k) - I_k R_{int}(x_k, I_k))}{\sigma_\nu^2}\Big) \cdot \Big(\frac{\partial v_{oc}(x_k)}{\partial x_k} - I_k \frac{\partial R_{int}(x_k, I_k)}{\partial x_k}\Big) \quad (A.46)$$

$$\Rightarrow -\frac{\partial^2 \log p(y_k|x_k)}{\partial {x_k}^2} = \frac{1}{\sigma_\nu^2}\Big(\frac{\partial v_{oc}(x_k)}{\partial x_k} - I_k\frac{\partial R_{int}(x_k, I_k)}{\partial x_k}\Big)^2 \ldots - \Big(\frac{y_k - (v_{oc}(x_k) - I_k R_{int}(x_k, I_k))}{\sigma_\nu^2}\Big) \ldots \cdot \Big(\frac{\partial^2 v_{oc}(x_k)}{\partial {x_k}^2} - I_k\frac{\partial^2 R_{int}(x_k, I_k)}{\partial {x_k}^2}\Big) \quad (A.47)$$

- **Prior**: Firstly, note that

$$p(x_k|y_{1:k-1}) = \int_{\mathcal{X}_{k-1}} p(x_k, x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1} \quad (A.48)$$
$$= \int_{\mathcal{X}_{k-1}} p(x_k|x_{k-1}, y_{1:k-1})p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1} \quad (A.49)$$
$$= \int_{\mathcal{X}_{k-1}} p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1} \quad (A.50)$$

Then, since $p(\cdot|x_{k-1})$ is sufficiently regular,

$$-\log p(x_k|y_{1:k-1}) = -\log\Big(\int_{\mathcal{X}_{k-1}} p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1}\Big) \quad (A.51)$$

$$\Rightarrow -\frac{\partial \log p(x_k|y_{1:k-1})}{\partial x_k} = -\frac{1}{p(x_k|y_{1:k-1})}\frac{\partial p(x_k|y_{1:k-1})}{\partial x_k} \quad (A.52)$$
$$= -\frac{\int_{\mathcal{X}_{k-1}} \frac{\partial p(x_k|x_{k-1})}{\partial x_k} p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1}}{\int_{\mathcal{X}_{k-1}} p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1}} \quad (A.53)$$

$$\Rightarrow -\frac{\partial^2 \log p(x_k|y_{1:k-1})}{\partial {x_k}^2} = -\frac{\int_{\mathcal{X}_{k-1}} \frac{\partial^2 p(x_k|x_{k-1})}{\partial {x_k}^2} p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1}}{\int_{\mathcal{X}_{k-1}} p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1}} \ldots + \Big(\frac{\int_{\mathcal{X}_{k-1}} \frac{\partial p(x_k|x_{k-1})}{\partial x_k} p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1}}{\int_{\mathcal{X}_{k-1}} p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1}}\Big)^2 \quad (A.54)$$

On the other hand

$$p(x_k|x_{k-1}) = \frac{1}{\sqrt{2\pi}\sigma_\omega}\mathrm{e}^{-\frac{1}{2}\frac{\left(x_k - (x_{k-1} - I_{k-1}v_{oc}(x_{k-1})\frac{T_s}{E_{crit}})\right)^2}{\sigma_\omega^2}} \quad (A.55)$$

$$\Rightarrow \frac{\partial p(x_k|x_{k-1})}{\partial x_k} = -\frac{1}{\sqrt{2\pi}\sigma_\omega^3}\mathrm{e}^{-\frac{1}{2}\frac{\left(x_k - (x_{k-1} - I_{k-1}v_{oc}(x_{k-1})\frac{T_s}{E_{crit}})\right)^2}{\sigma_\omega^2}} \ldots \cdot \big(x_k - (x_{k-1} - I_{k-1}v_{oc}(x_{k-1})\frac{T_s}{E_{crit}})\big) \quad (A.56)$$

$$\Rightarrow \frac{\partial^2 p(x_k|x_{k-1})}{\partial {x_k}^2} = \frac{1}{\sqrt{2\pi}\sigma_\omega^5} \mathrm{e}^{-\frac{1}{2}\frac{\left(x_k-(x_{k-1}-I_{k-1}v_{oc}(x_{k-1})\frac{T_s}{E_{crit}})\right)^2}{\sigma_\omega^2}} \ldots \cdot \left(x_k-(x_{k-1}-I_{k-1}v_{oc}(x_{k-1})\frac{T_s}{E_{crit}})\right)^2 \ldots - \frac{1}{\sqrt{2\pi}\sigma_\omega^3} \mathrm{e}^{-\frac{1}{2}\frac{\left(x_k-(x_{k-1}-I_{k-1}v_{oc}(x_{k-1})\frac{T_s}{E_{crit}})\right)^2}{\sigma_\omega^2}} \tag{A.57}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_\omega^3} \mathrm{e}^{-\frac{1}{2}\frac{\left(x_k-(x_{k-1}-I_{k-1}v_{oc}(x_{k-1})\frac{T_s}{E_{crit}})\right)^2}{\sigma_\omega^2}} \ldots \cdot \left(\frac{\left(x_k-(x_{k-1}-I_{k-1}v_{oc}(x_{k-1})\frac{T_s}{E_{crit}})\right)^2}{\sigma_\omega^2}-1\right) \tag{A.58}$$

- **Evidence:** Since $p(y_k|y_{1:k-1})$ does not depend on x_k, it follows that

$$\Rightarrow \frac{\partial^2 \log p(y_k|y_{1:k-1})}{\partial {x_k}^2} = 0 \tag{A.59}$$

Therefore, the MCP-BCRLB at the prognostic time instant k_p can be approximated considering the state posterior distributions at times k_p and $k_p - 1$, which are of the form:

$$p(x_{k_p}|y_{1:k_p}) \approx \sum_{\mathrm{i}=1}^{N_p} w_{k_p}^{(\mathrm{i})} \delta_{x_{k_p}^{(\mathrm{i})}}(x_{k_p}) \tag{A.60}$$

$$p(x_{k_p-1}|y_{1:k_p-1}) \approx \sum_{\mathrm{i}=1}^{N_p} w_{k_p-1}^{(\mathrm{i})} \delta_{x_{k_p-1}^{(\mathrm{i})}}(x_{k_p-1}) \tag{A.61}$$

Nonetheless, all these calculations require the following equations:

$$x_{k+1} = x_k - I_k v_{oc}(x_k)\frac{T_s}{E_{crit}} + \omega_k \tag{A.62}$$

$$y_k = v_{oc}(x_k) - I_k R_{int}(x_k, I_k) + \nu_k \tag{A.63}$$

$$v_{oc}(x_k) = v_L + (v_0 - v_L)\mathrm{e}^{\gamma(x_k-1)} + \alpha v_L(x_k - 1) + (1-\alpha)v_L(\mathrm{e}^{-\beta} - \mathrm{e}^{-\beta\sqrt{x_k}}) \tag{A.64}$$

$$\frac{\partial v_{oc}(x_k)}{\partial x_k} = (v_0 - v_L)\gamma\mathrm{e}^{\gamma(x_k-1)} + \alpha v_L + (1-\alpha)v_L\mathrm{e}^{-\beta\sqrt{x_k}}\frac{\beta}{2\sqrt{x_k}} \tag{A.65}$$

$$\frac{\partial^2 v_{oc}(x_k)}{\partial {x_k}^2} = (v_0 - v_L)\gamma^2\mathrm{e}^{\gamma(x_k-1)} - (1-\alpha)v_L\mathrm{e}^{-\beta\sqrt{x_k}}\frac{\beta}{4}\left(\frac{\beta}{x_k} + \frac{1}{{x_k}^{3/2}}\right) \tag{A.66}$$

$$R_{int}(x_k, I_k) = p_0(I_k) + p_1(I_k)x_k + p_2(I_k)x_k^2 \tag{A.67}$$

$$\frac{\partial R_{int}(x_k, I_k)}{\partial x_k} = p_1(I_k) + 2p_2(I_k)x_k \tag{A.68}$$

$$\frac{\partial^2 R_{int}(x_k, I_k)}{\partial {x_k}^2} = 2p_2(I_k) \tag{A.69}$$

A.3.2 Elements of the Recursion

$$-\log p(x_{k+1}|x_k) = -\log \frac{1}{\sqrt{2\pi}\sigma_\omega} \mathrm{e}^{-\frac{1}{2}\frac{\left(x_{k+1}-(x_k-I_k v_{oc}(x_k)\frac{T_s}{E_{crit}})\right)^2}{\sigma_\omega^2}} \tag{A.70}$$

$$= c_1 + \frac{1}{2}\frac{\left(x_{k+1}-(x_k-I_k v_{oc}(x_k)\frac{T_s}{E_{crit}})\right)^2}{\sigma_\omega^2} \tag{A.71}$$

$$-\frac{\partial \log p(x_{k+1}|x_k)}{\partial x_k} = \frac{1}{\sigma_\omega^2}\big(x_{k+1}-x_k+I_k v_{oc}(x_k)\frac{T_s}{E_{crit}}\big)\big(-1+I_k\frac{\mathrm{d}v_{oc}(x_k)}{\mathrm{d}x_k}\frac{T_s}{E_{crit}}\big) \tag{A.72}$$

$$-\frac{\partial \log p(x_{k+1}|x_k)}{\partial x_{k+1}} = \frac{1}{\sigma_\omega^2}\big(x_{k+1}-x_k+I_k v_{oc}(x_k)\frac{T_s}{E_{crit}}\big) \tag{A.73}$$

$$\Rightarrow -\frac{\partial^2 \log p(x_{k+1}|x_k)}{\partial {x_k}^2} = \frac{1}{\sigma_\omega^2}\Big(\big(-1+I_k\frac{\mathrm{d}v_{oc}(x_k)}{\mathrm{d}x_k}\frac{T_s}{E_{crit}}\big)^2 + \ldots + \big(x_{k+1}-x_k+I_k v_{oc}(x_k)\frac{T_s}{E_{crit}}\big)I_k\frac{\mathrm{d}^2 v_{oc}(x_k)}{\mathrm{d}{x_k}^2}\frac{T_s}{E_{crit}}\Big) \tag{A.74}$$

$$\Rightarrow -\frac{\partial^2 \log p(x_{k+1}|x_k)}{\partial x_k \partial x_{k+1}} = \frac{1}{\sigma_\omega^2}\big(-1+I_k\frac{\mathrm{d}v_{oc}(x_k)}{\mathrm{d}x_k}\frac{T_s}{E_{crit}}\big) \tag{A.75}$$

$$\Rightarrow -\frac{\partial^2 \log p(x_{k+1}|x_k)}{\partial {x_{k+1}}^2} = \frac{1}{\sigma_\omega^2} \tag{A.76}$$

Appendix B

Demonstrations: Chapter 5

B.1 Demonstration of Theorem 5.1

PROOF. (Theorem 5.1)

For $k \in \mathbb{N}$, let us define $\Delta_k : \mathcal{F}(\mathbb{N}, \mathbb{N}) \to \mathbb{R}$ as $\Delta_k(f(\cdot, k)) = f(\cdot, k+1) - f(\cdot, k)$, for any $f \in \mathcal{F}$. Then,

$$\Delta_k \mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})\mathbb{1}_{\tau_{\mathcal{E}}=k}\} := \mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})\mathbb{1}_{\tau_{\mathcal{E}}=k+1}\} - \mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})\mathbb{1}_{\tau_{\mathcal{E}}=k}\} \tag{B.1}$$

$$= \sum_{\hat{k}\in\mathbb{N}} (\hat{k} - (k+1))\mathbb{P}^*(\hat{k}, k+1) - (\hat{k} - k)\mathbb{P}^*(\hat{k}, k) \tag{B.2}$$

$$= -\sum_{\hat{k}\in\mathbb{N}} \mathbb{P}^*(\hat{k}, k+1) + \sum_{\hat{k}\in\mathbb{N}} (\hat{k} - k)\left(\mathbb{P}^*(\hat{k}, k+1) - \mathbb{P}^*(\hat{k}, k)\right) \tag{B.3}$$

$$= -\mathbb{P}(\tau_{\mathcal{E}} = k+1) + \sum_{\hat{k}\in\mathbb{N}} (\hat{k} - k)\Delta_k \mathbb{P}^*(\hat{k}, k). \tag{B.4}$$

Summing with respect to $k \in supp(\tau_{\mathcal{E}} - 1)$, it is obtained

$$\underbrace{\sum_{k\in supp(\tau_{\mathcal{E}}-1)} \Delta_k \mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})\mathbb{1}_{\tau_{\mathcal{E}}=k}\}}_{(\star)} = -1 + \underbrace{\sum_{k\in supp(\tau_{\mathcal{E}}-1)} \sum_{\hat{k}\in\mathbb{N}} (\hat{k} - k)\Delta_k \mathbb{P}^*(\hat{k}, k)}_{(\star\star)}. \tag{B.5}$$

Therefore, on the one hand,

$$(\star) = \sum_{c\geq 1} \sum_{k=k_{\mathrm{i}}^c - 1}^{k_f^c - 1} \Delta_k \mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})\mathbb{1}_{\tau_{\mathcal{E}}=k}\} \tag{B.6}$$

$$= \sum_{c\geq 1} \mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})\mathbb{1}_{\tau_{\mathcal{E}}=k}\}\Big|_{k=k_{\mathrm{i}}^c - 1}^{k=k_f^c} \tag{B.7}$$

$$= \sum_{c\geq 1} \mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{\hat{\tau}_{\mathcal{E}} - k\}\mathbb{P}(\tau_{\mathcal{E}} = k)\Big|_{k=k_{\mathrm{i}}^c - 1}^{k=k_f^c} \tag{B.8}$$

$$= \sum_{c\geq 1} \mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{\hat{\tau}_{\mathcal{E}} - k_f^c\}\mathbb{P}(\tau_{\mathcal{E}} = k_f^c) \tag{B.9}$$

Because $\mathbb{P}(\tau_{\mathcal{E}} = k_{\mathrm{i}}^c - 1) = 0$ by definition of k_{i}^c, $\forall c \geq 1$.

On the other hand, by independence

$$\Delta_k \mathbb{P}^*(\hat{k}, k) = \mathbb{P}(\hat{\tau}_{\mathcal{E}} = \hat{k})\big(\mathbb{P}(\tau_{\mathcal{E}} = k+1) - \mathbb{P}(\tau_{\mathcal{E}} = k)\big), \tag{B.10}$$

and defining

$$\alpha_k := \frac{\mathbb{P}(\tau_{\mathcal{E}} = k+1)}{\mathbb{P}(\tau_{\mathcal{E}} = k)},\ \forall k \in supp(\tau_{\mathcal{E}}), \tag{B.11}$$

then,

$$(\star\star) = \sum_{k\in supp(\tau_{\mathcal{E}}-1)} \sum_{\hat{k}\in\mathbb{N}} (\hat{k}-k)\mathbb{P}(\hat{\tau}_{\mathcal{E}} = \hat{k})\big(\mathbb{P}(\tau_{\mathcal{E}} = k+1) - \mathbb{P}(\tau_{\mathcal{E}} = k)\big) \tag{B.12}$$

$$= \sum_{k\in supp(\tau_{\mathcal{E}})} \sum_{\hat{k}\in\mathbb{N}} (\hat{k}-k)\mathbb{P}(\hat{\tau}_{\mathcal{E}} = \hat{k})\big(\mathbb{P}(\tau_{\mathcal{E}} = k+1) - \mathbb{P}(\tau_{\mathcal{E}} = k)\big)$$

$$\ldots - \sum_{c\geq 1}\sum_{\hat{k}\in\mathbb{N}} (\hat{k}-k_f^c)\mathbb{P}(\hat{\tau}_{\mathcal{E}} = \hat{k})\big(\mathbb{P}(\tau_{\mathcal{E}} = k_f^c+1) - \mathbb{P}(\tau_{\mathcal{E}} = k_f^c)\big)$$

$$\ldots + \sum_{c\geq 1}\sum_{\hat{k}\in\mathbb{N}} (\hat{k}-(k_{\mathrm{i}}^c-1))\mathbb{P}(\hat{\tau}_{\mathcal{E}} = \hat{k})\big(\mathbb{P}(\tau_{\mathcal{E}} = k_{\mathrm{i}}^c) - \mathbb{P}(\tau_{\mathcal{E}} = k_{\mathrm{i}}^c-1)\big) \tag{B.13}$$

$$= \sum_{k\in supp(\tau_{\mathcal{E}})} \sum_{\hat{k}\in\mathbb{N}} (\hat{k}-k)\mathbb{P}(\hat{\tau}_{\mathcal{E}} = \hat{k})\mathbb{P}(\tau_{\mathcal{E}} = k)(\alpha_k - 1)$$

$$\ldots + \sum_{c\geq 1}\sum_{\hat{k}\in\mathbb{N}} (\hat{k}-k_f^c)\mathbb{P}(\hat{\tau}_{\mathcal{E}} = \hat{k})\mathbb{P}(\tau_{\mathcal{E}} = k_f^c) \tag{B.14}$$

$$\ldots + \sum_{c\geq 1}\sum_{\hat{k}\in\mathbb{N}} (\hat{k}-(k_{\mathrm{i}}^c-1))\mathbb{P}(\hat{\tau}_{\mathcal{E}} = \hat{k})\mathbb{P}(\tau_{\mathcal{E}} = k_{\mathrm{i}}^c) \tag{B.15}$$

$$= \sum_{k\in supp(\tau_{\mathcal{E}})} \mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{(\hat{\tau}_{\mathcal{E}} - k)(\alpha_k - 1)\}\mathbb{P}(\tau_{\mathcal{E}} = k)$$

$$\ldots + \sum_{c\geq 1} \mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{\hat{\tau}_{\mathcal{E}} - k_f^c\}\mathbb{P}(\tau_{\mathcal{E}} = k_f^c)$$

$$\ldots + \sum_{c\geq 1} \mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{\hat{\tau}_{\mathcal{E}} - (k_{\mathrm{i}}^c-1)\}\mathbb{P}(\tau_{\mathcal{E}} = k_{\mathrm{i}}^c). \tag{B.16}$$

Considering the expressions for $(\star)$ and $(\star\star)$, (B.5) becomes

$$1 = \sum_{k\in supp(\tau_{\mathcal{E}})} \mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{(\hat{\tau}_{\mathcal{E}} - k)(\alpha_k - 1)\}\mathbb{P}(\tau_{\mathcal{E}} = k)$$

$$\ldots + \sum_{c\geq 1} \mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{\hat{\tau}_{\mathcal{E}} - (k_{\mathrm{i}}^c-1)\}\mathbb{P}(\tau_{\mathcal{E}} = k_{\mathrm{i}}^c). \tag{B.17}$$

Applying the Cauchy-Schwarz inequality,

$$\left(\sum_{k\in supp(\tau_{\mathcal{E}})}\mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{(\hat{\tau}_{\mathcal{E}}-k)(\alpha_k-1)\}\mathbb{P}(\tau_{\mathcal{E}}=k)\right)^2$$

$$\leq\left(\sum_{k\in supp(\tau_{\mathcal{E}})}\mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{(\hat{\tau}_{\mathcal{E}}-k)^2\}\mathbb{P}(\tau_{\mathcal{E}}=k)\right)$$

$$\ldots\cdot\left(\sum_{k\in supp(\tau_{\mathcal{E}})}\mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{(\alpha_k-1)^2\}\mathbb{P}(\tau_{\mathcal{E}}=k)\right) \tag{B.18}$$

$$=\mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}}-\tau_{\mathcal{E}})^2\}\cdot\left(\sum_{k\in supp(\tau_{\mathcal{E}})}(\alpha_k-1)^2\mathbb{P}(\tau_{\mathcal{E}}=k)\right). \tag{B.19}$$

The *Bayesian Cramér-Rao lower bound* defined as

$$\Theta_{\mathcal{E}}=\left(\sum_{k\in supp(\tau_{\mathcal{E}})}(\alpha_k-1)^2\mathbb{P}(\tau_{\mathcal{E}}=k)\right)^{-1} \tag{B.20}$$

$$=\left(\sum_{k\in supp(\tau_{\mathcal{E}})}\left(\frac{\mathbb{P}(\tau_{\mathcal{E}}=k+1)}{\mathbb{P}(\tau_{\mathcal{E}}=k)}-1\right)^2\mathbb{P}(\tau_{\mathcal{E}}=k)\right)^{-1} \tag{B.21}$$

is well-defined. Thus, Eqs. (B.18)-(B.19) lead to

$$\mathbb{E}_{\mathbb{P}^*}\{(\hat{\tau}_{\mathcal{E}}-\tau_{\mathcal{E}})^2\}\geq\Theta_{\mathcal{E}}\cdot\left(1-\sum_{c\geq1}\mathbb{E}_{\mathbb{P}(\hat{\tau}_{\mathcal{E}}=\cdot)}\{\hat{\tau}_{\mathcal{E}}-(k_{\mathrm{i}}^c-1)\}\mathbb{P}(\tau_{\mathcal{E}}=k_{\mathrm{i}}^c)\right)^2. \tag{B.22}$$

□

B.2 Demonstration of Theorem 5.2

Proof. (Theorem 5.2)

Let $\hat{\tau}_{\mathcal{E}}$ be an estimator of $\tau_{\mathcal{E}}$. Since $p(\tau_{\mathcal{E}}=t)$ is absolutely continuous, $\frac{\mathrm{d}}{\partial t}p(\tau_{\mathcal{E}}=t)$ is well-defined, moreover, by hypothesis, it is absolutely integrable. Therefore, we have the following identity

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbb{E}_{p^*}\{(\hat{\tau}_{\mathcal{E}}-\tau_{\mathcal{E}})\mathbb{1}_{\tau_{\mathcal{E}}=t}\}\right):=\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathbb{R}_+}(\hat{t}-t)p^*(\hat{t},t)\mathrm{d}\hat{t}\right) \tag{B.23}$$

$$=\int_{\mathbb{R}_+}\left(-p^*(\hat{t},t)+(\hat{t}-t)\frac{\mathrm{d}}{\mathrm{d}t}p^*(\hat{t},t)\right)\mathrm{d}\hat{t} \tag{B.24}$$

$$=-\int_{\mathbb{R}_+}p^*(\hat{t},t)\mathrm{d}\hat{t}+\int_{\mathbb{R}_+}(\hat{t}-t)\frac{\mathrm{d}}{\mathrm{d}t}p^*(\hat{t},t)\mathrm{d}\hat{t} \tag{B.25}$$

$$= -p(\tau_{\mathcal{E}} = t) + \int_{\mathbb{R}_+} (\hat{t} - t) \frac{\mathrm{d}}{\mathrm{d}t} p^*(\hat{t}, t) \mathrm{d}\hat{t} \tag{B.26}$$

Integrating with respect to $t \in \mathbb{R}_+$, it is obtained

$$\underbrace{\int_{\mathbb{R}_+} \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}_{p^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}}) \mathbb{1}_{\tau_{\mathcal{E}} = t}\} \mathrm{d}t}_{(\star)} = -1 + \underbrace{\int_{\mathbb{R}_+} \int_{\mathbb{R}_+} (\hat{t} - t) \frac{\mathrm{d}}{\mathrm{d}t} p^*(\hat{t}, t) \mathrm{d}\hat{t}\mathrm{d}t}_{(\star\star)}. \tag{B.27}$$

Therefore, on the one hand,

$$(\star) = \int_{\mathbb{R}_+} \frac{\mathrm{d}}{\mathrm{d}t} \left(\mathbb{E}_{p(\hat{\tau}_{\mathcal{E}} = \cdot)}\{\hat{\tau}_{\mathcal{E}} - t\} p(\tau_{\mathcal{E}} = t) \right) \mathrm{d}t \tag{B.28}$$

$$= \lim_{t \to +\infty} \mathbb{E}_{p(\hat{\tau}_{\mathcal{E}} = \cdot)}\{\hat{\tau}_{\mathcal{E}} - t\} p(\tau_{\mathcal{E}} = t) - \mathbb{E}_{p(\hat{\tau}_{\mathcal{E}} = \cdot)}\{\hat{\tau}_{\mathcal{E}}\} \cancelto{0}{p(\tau_{\mathcal{E}} = 0)} \tag{B.29}$$

$$= 0, \tag{B.30}$$

because $\lim_{t \to +\infty} t p(\tau_{\mathcal{E}} = t) = 0$. On the other hand, by independence

$$\frac{\mathrm{d}}{\mathrm{d}t} p^*(\hat{t}, t) = p(\hat{\tau}_{\mathcal{E}} = \hat{t}) \frac{\mathrm{d}}{\mathrm{d}t} p(\tau_{\mathcal{E}} = t). \tag{B.31}$$

Thus,

$$(\star\star) = \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} (\hat{t} - t) p(\hat{\tau}_{\mathcal{E}} = \hat{t}) \frac{\mathrm{d}}{\mathrm{d}t} p(\tau_{\mathcal{E}} = t) \mathrm{d}\hat{t}\mathrm{d}t \tag{B.32}$$

$$= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} (\hat{t} - t) p(\hat{\tau}_{\mathcal{E}} = \hat{t}) \frac{\mathrm{d}}{\mathrm{d}t} \log p(\tau_{\mathcal{E}} = t) p(\tau_{\mathcal{E}} = t) \mathrm{d}\hat{t}\mathrm{d}t \tag{B.33}$$

$$= \mathbb{E}_{p^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}}) \frac{\mathrm{d}}{\mathrm{d}t} \log p(\tau_{\mathcal{E}} = t)\}. \tag{B.34}$$

Considering the expressions for $(\star)$ and $(\star\star)$, (B.27) becomes

$$\mathbb{E}_{p^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}}) \frac{\mathrm{d}}{\mathrm{d}t} \log p(\tau_{\mathcal{E}} = t)\} = 1. \tag{B.35}$$

Applying the Cauchy-Schwarz inequality,

$$\left(\mathbb{E}_{p^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}}) \frac{\mathrm{d}}{\mathrm{d}t} \log p(\tau_{\mathcal{E}} = t)\} \right)^2 \leq \mathbb{E}_{p^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})^2\} \cdot \mathbb{E}_{p^*}\{\left(\frac{\mathrm{d}}{\mathrm{d}t} \log p(\tau_{\mathcal{E}} = t) \right)^2\}. \tag{B.36}$$

The *Bayesian Cramér-Rao lower bound* defined as

$$\Theta_{\mathcal{E}} = \left(\mathbb{E}_{p(\tau_{\mathcal{E}} = \cdot)}\{\left(\frac{\mathrm{d}}{\mathrm{d}t} \log p(\tau_{\mathcal{E}} = t) \right)^2\} \right)^{-1} \tag{B.37}$$

is well-defined. Thus, Eq. (B.36) leads to

$$\mathbb{E}_{p^*}\{(\hat{\tau}_{\mathcal{E}} - \tau_{\mathcal{E}})^2\} \geq \Theta_{\mathcal{E}}. \tag{B.38}$$

□