# Development of a Bayesian multi-state degradation model for up-to-date reliability estimations of working industrial components

M. Compare[1,2], P. Baraldi[1], I. Bani[1], E. Zio[1,2,3], D. Mc Donnell[4]

[1]*Energy Department, Politecnico di Milano, Milano, Italy*

[2]*Aramis Srl, Milano, Italy*

[3]*Chair on Systems Science and Energetic Challenge Fondation EDF (Electricite' de France) CentraleSupelec, France*

[4]*School of Engineering, Trinity College Dublin, College Green, Dublin 2, Ireland*

**Abstract**

We consider a three-state continuous-time semi-Markov process with Weibull-distributed transition times to model the degradation mechanism of an industrial equipment. To build this model, an original combination of techniques is proposed for building a semi-Markov degradation model based on expert knowledge and few field data within the Bayesian statistical framework. The issues addressed are: i) the prior elicitation of the model parameters values from experts, avoiding possible information commitment; ii) the development of a Markov-Chain Monte Carlo algorithm for sampling from the posterior distribution; iii) the posterior inference of the model parameters values and, on this basis, the estimation of the time-dependent state probabilities and the prediction of the equipment remaining useful life. The developed Bayesian model offers the possibility of updating the system reliability estimation every time a new evidence is gathered. The application of the modeling framework is illustrated by way of a real industrial case study concerning the degradation of diaphragms installed in a production line of a biopharmaceutical industry.

*Key words*: Multi-state degradation modelling, Weibull distribution, Remaining useful life, Maintenance, Bayesian inference, MCMC algorithms

**Acronyms**

A-MCMC: Adaptive MCMC algorithm

CDF: Cumulative Distribution Function

EPDM: Ethylene Propylene Diene Monomer

i.i.d: independent and identically distributed

MCMC: Markov Chain Monte Carlo

MTS: Most Trustworthy Specification

N-RWMH: Normal Random Walk Metropolis-Hastings algorithm

PDF: Probability Density Function

RUL: Remaining Useful Life

# 1 Introduction

Multi-state degradation modelling is based on the discretization of the degradation process affecting an industrial equipment in three or more states, each one associated to a certain range of values of suitable degradation indicator variables (e.g., oxidized areas in gas turbine nozzle systems (Compare et al., 2015), electrical resistance values in electrical power switches (Baraldi et al., 2013b), linear extent of wear in bearing shells (Moghaddas & Zuo, 2013; Moghaddas et al., 2015)), performance levels (e.g., amount of power supplied by power generating systems (Liu et al., 2015)) or symptoms (e.g., vibrational signals (Baraldi et al., 2014a)). The main advantage of this modeling approach over the widely used binary model (i.e., considering only two operational states for the equipment, 'good' and 'failed') lies in its ability of more accurately describing the sequential phases of degradation, which can be even physically different (Levitin et al., 2003; Lisnianski & Levitin, 2003; Zuo et al., 2001).

The transition times from one state to another are often assumed to be Weibull-distributed (e.g., Giorgio et al., 2011; Moghaddas & Zuo, 2012; Baraldi et al., 2013a). This choice is due to the flexibility of the Weibull distribution and the possibility it gives of keeping memory of the time spent in a degradation state, which influences the next stochastic transition time. This property gives rise to a semi-Markov model (Limnios & Oprisan,2001).

The estimation of the model parameters and the characterization of the corresponding uncertainties are fundamental to properly set and use multi-state degradation models. To do this, different approaches have been proposed in the literature, which mainly depend on the available knowledge, information and data. Namely, when a substantial amount of collected data is available, techniques from statistical analysis can be adopted, which may be purely analytical (e.g., Giorgio et al., 2011) or numerical (e.g., Compare et al., 2015).

On the contrary, situations characterized by scarcity of data are common in industrial applications (e.g., very highly reliable components, new technology just introduced in the production system, etc.). In this case, expert opinion becomes a valuable source of information to be taken into account for developing semi-Markov degradation models. Different approaches have been proposed within probabilistic and non-probabilistic theoretical frameworks. Probabilistic approaches (e.g., Chirister et al., 1995; Christer et al., 1998; Jenkinson, 2005) are typically based on the elicitation of subjective expert judgements about the probabilities of occurrence of single events or about the probability distributions of uncertain quantities of interest. Often, the elicitation process is oriented to obtain a suitable prior distribution to be updated in light of the available data, according to the Bayesian paradigm (e.g., Coolen, 1996; Singpurwalla & Song, 1988; Garthwaite et al., 2005). On the other side, within the non-probabilistic approaches, Possibility Theory has been used to tackle the situation in which the knowledge on each Weibull parameter is available in terms of a set of nested intervals with corresponding confidence levels provided by an expert (Baraldi et al., 2013c; Baraldi et al., 2014b;). Similarly, Dempster-Shafer theory of Evidence has been applied to develop a semi-Markov degradation model in the situation in which an interval that is believed to contain the unknown parameter

value is asked to each member of a team of experts (Baraldi et al., 2013a). Fuzzy logic has also been used with the same aim in (Ge & Asgarpoor, 2010).

In this work, we want to build a semi-Markov degradation model in an intermediate situation between that of having a sizeable amount of field data (which would justify the application of traditional statistical techniques in the frequentist probability framework) and the opposite, of no data available (which has been treated with non-probabilistic techniques that avoid information commitment). The objective is to fully exploit all the available sources of information in a coherent and solid way.

To do this, we resort to the Bayesian statistics framework, which allows combining the prior knowledge of experts with the evidence coming from field data to build a degradation model useful for maintenance applications (Compare & Zio, 2014). The proposed methodology allows the elicitation of the prior distributions of the model parameters, avoiding possible commitment of the information provided by the expert. Then, an adaptive Markov Chain Monte Carlo algorithm is developed to estimate the posterior distributions of the multi-state model parameters, which encode both the prior knowledge and the evidence brought by the available dataset. Finally, the developed stochastic model is used to derive the expected probabilities of occupying the degradation states over time for new components along with the corresponding credibility intervals, and to estimate the remaining useful life of a new component, which will be detected in a given degradation state after a certain working time.

The proposed procedure is applied to a case study in the biopharmaceutical industry, concerning the Ethylene Propylene Diene Monomer (EDPM) diaphragm installed within a production line of a company leader in that field.

The original contribution of the proposed method mainly lies in 1) the original combination of techniques, taken from the scientific literature of different contexts, which have been adapted to propose a comprehensive development pathway for building a semi-Markov degradation model based on expert knowledge and few field data, 2) the use of a Bayesian semi-Markov degradation model to support maintenance planning.

Furthermore, within the proposed framework for the construction of the Bayesian semi-Markov degradation model, we have developed a novel procedure to sample multiple parameters from their joint posterior distribution. The procedure is based on the combined use of *i)* an adaptive Markov Chain Monte Carlo (MCMC) algorithm to set possibly acceptable values for the entries of the covariance matrix and *ii)* the Normal-Random Walk Metropolis Hastings (N-RWMH).

An important property of the proposed framework for supporting maintenance planning is that it allows updating the posterior parameter distribution taking into account the outcome of the last inspection performed on the industrial component under observation. This additional updating is useful in those situations characterized by scarcity of data, where adding a single observation can significantly improve the parameter estimation.

The remainder of the paper is organized as follows. Section 2 introduces the case study motivating the development of the proposed framework. Section 3 defines the assumptions at the basis of the model development. Section 4 illustrates the development of the whole methodology, from the elicitation of the prior distribution to the posterior inference. This is applied to the EDPM diaphragms in Section 5, where the results of the study are also discussed. Some final considerations about the applicability of the proposed framework to other case studies are drawn in Section 6. Section 7 concludes the paper.

# 2 Case study motivating the framework development

The case study motivating the development of the proposed methodological framework concerns the gradual degradation of Ethylene Propylene Diene Monomer (EPDM) diaphragms used in the production line of a pharmaceutical company, where they are subject to harsh environmental conditions. The EDPM diaphragm is the sealing element in flow control valves whose function is to maintain the integrity of the hermetically sealed environments within production bioreactors. Severe degradation states of EPDM diaphragms may result in undesired contaminations leading to a complete loss of recent production.

According to maintenance operators' practice, the degradation of EPDM diaphragms can be modeled using a three-state model, where each state corresponds to a different level of damage:

- state 1 – the component shows mild discolouration, melting and weir markings
- state 2 – the component shows signs of more severe melting, material flow, and surface creasing
- state 3 – the component shows melting, material flow, and material creasing is evident.

A dataset is available, containing the time at which 109 diaphragms have been replaced and the corresponding identified state of degradation (Table 1). It contains 73 components replaced when they were in state 1, 32 in state 2 and 4 in state 3. The detection of a component in the third (and last) degradation state is quite rare, since this corresponds to a failure. Notice that the identification of the degradation state of the EPDM diaphragms requires to perform a destructive test on the component.

In addition to the evidence available in the dataset, we have the possibility of interviewing an expert to gather additional information for properly setting the degradation model parameters.

**Table 1.** The available dataset

| #Diaphragm | Replacement time | Degradation state |
|---|---|---|
| 1 | 16.4 | 1 |
| 2 | 26.7 | 1 |
| … | … | … |
| N | 173 | 2 |

# 3 Modeling assumptions

In this work, we frame the case study illustrated above as a general parameter estimation problem in which we want to exploit all the available knowledge, information and data (Zio, 2016). For this, we first build a methodological framework and then we give some practical advices to readers interested in implementing the proposed procedure on other datasets.

We consider a three-state process (Figure 1), where:

- State 1 refers to a new or mildly degraded component;
- State 2 refers to a fairly degraded component;
- State 3 refers to a heavily degraded component.

It is assumed that direct transitions from state 1 to state 3 are not possible, which is typical of degradation processes caused by cumulative damage (Ruiz-Castro, 2014; Compare et al., 2015). The stochastic transition times from state 1 to 2 and from state 2 to 3 are indicated by $T_1$ and $T_2$, respectively. The first transition time

is assumed to obey a Weibull distribution conditional on the random parameters $\alpha_1$ (scale) and $\beta_1$ (shape), whereas the second transition time is Weibull-distributed, conditional on the random parameters $\alpha_2$ and $\beta_2$.

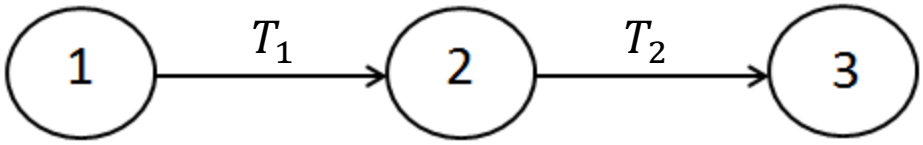



**Figure 1.** *Three-state model of the degradation process*

We assume that a dataset $(\boldsymbol{r}, \boldsymbol{d}) = \{(r_n, d_n)\}_{n=1,\dots,N}$ is available, containing $N$ observations. The two entries $(r_n, d_n)$ for the generic $n$-th record, $n = 1, \dots, N$, are the replacement time ($r_n$, i.e., the time elapsed from the component installation to its replacement) and the corresponding degradation state $d_n$. It is also assumed that each component has been replaced after inspection; that is, every record refers to a different component.

The stochastic transition times from state 1 to 2 and from state 2 to 3 of the $n^{th}$ component, $n = 1, \dots, N$ are indicated by $T_{1n}$ and $T_{2n}$, respectively. In particular, $T_{1n}$, $n = 1, \dots, N$ are considered as random variables independent and identically distributed (i.i.d.) as $T_1$, conditional on the parameters $\alpha_1$ and $\beta_1$. Similarly, $T_{2n}$, $n = 1, \dots, N$ are random variables independent and identically distributed as $T_2$, conditional on $\alpha_2$ and $\beta_2$. Obviously, if the $n^{th}$ component is found in degradation state $d_n = 1$, then it has not experienced the first nor the second transition. However, this does not mean that the stochastic transition times $T_{1n}$ and $T_{2n}$ do not exist; rather, it means that the replacement time $r_n$ has worked like a censoring time preventing the component from reaching states 2 and 3. In this respect, we assume that $r_n$ is a random variable independent from both $T_{1n}$ and $T_{2n}$. This assumption requires that the choice of the replacement time for each component in the dataset is not influenced by any information eventually collected during the component life (ongoing detections, measurements, etc.), which could reveal, directly or indirectly, the evolution of its degradation state. The same reasoning applies to a component found in degradation state $d_n = 2$.

These assumptions allow us considering the following Bayesian model:

$$\begin{cases} T_{11}, \dots, T_{1N} | \boldsymbol{\theta} \ \sim^{i.i.d.} \ Weibull(\alpha_1, \beta_1), T_{21}, \dots, T_{2N} | \boldsymbol{\theta} \ \sim^{i.i.d.} \ Weibull(\alpha_2, \beta_2) \\ \boldsymbol{\theta} := (\alpha_1, \beta_1, \alpha_2, \beta_2) \sim \pi(\boldsymbol{\theta}) \end{cases} \quad (1)$$

where $\alpha_1, \beta_1, \alpha_2, \beta_2$ are random variables, arranged in the parameter vector $\boldsymbol{\theta} \in \Theta := (\mathbb{R}_+)^w, w = 4$, with joint prior probability distribution $\pi(\boldsymbol{\theta})$. Additional assumptions are:

- $T_{1n} | \alpha_1, \beta_1$ is independent from $T_{2n} | \alpha_2, \beta_2 \ \forall \, n = 1, \dots, N$. That is, the second transition time is not influenced by the first one; rather, it is influenced by the sojourn time in state 2.
- The two random vectors $(\alpha_1, \beta_1)$ and $(\alpha_2, \beta_2)$ are a priori independent, i.e., we assume that the expert can give two different a priori distributions for these two pairs of parameters.

# 4 Method development

## 4.1 Prior elicitation

In the Bayesian analysis framework, the first important issue to address is the assignment of prior distributions to the model parameters. These distributions can be built starting either from data available from similar phenomena/experiments (Epifani et al., 2014) or from the opinions of experts (Bousquet, 2010). Otherwise, in complete absence of information, non-informative priors are used (e.g., Kass & Wasserman, 1996).

As mentioned in Section 2, we consider the case in which data for setting the priors are not available and, thus, we have to rely on the judgements of experts. In general, particular attention must be paid in eliciting the information from experts to avoid possible commitments; thus, the questions used to interview the experts should not be formulated in such a way that the answers are biased (Garthwaite et al., 2005).

In our case study, a set of preliminary interviews with the expert led us to consider the elicitation procedure proposed in (Bousquet, 2010), which is illustrated for the transition time $T_1$ only, the case of $T_2$ being analogous.

In the case of Weibull distributions, experts are expected to be more naturally led to convey information about the transition times, rather than about the scale and shape parameters, because these have not a straightforward physical interpretation (Bousquet, 2010). For this, the statistical information to be elicited directly concerns the prior predictive distribution, whose probability density function (PDF), in the case of $T_1$, reads:

$$f_{\pi(\alpha_1,\beta_1)}(t_1) = \int_{\mathbb{R}_+\times\mathbb{R}_+} f(t_1|\alpha_1,\beta_1)\pi(\alpha_1,\beta_1)d\alpha_1 d\beta_1 \qquad (2)$$

where $\pi(\alpha_1,\beta_1)$ is the marginal prior distribution to be elicited. According to (Bousquet, 2010), the expert is asked to provide estimates for a certain number $h_1$ of quantiles for the transition time $T_1$. Specifically, he/she is required to provide a set of pairs $\{(s_i,q_i)\}_{i=1,\dots,h_1}$ with $0 < s_i < s_{i+1}, 0 < q_i < q_{i+1} < 1 \; \forall\, i = 1,\dots,h_1 - 1$ answering to these $h_1$ explicit questions:

*"In your opinion, after how much working time $s_i$ do you expect to find that exactly $(q_i \times 100)\%$ of the components have undergone the transition from state 1 to state 2?" ,*

where the values $q_i$ can also be proposed directly by the analyst.
The marginal prior distribution $\pi(\alpha_1,\beta_1)$ to be elicited is that one which possibly solves all the $h_1$ equations:

$$P_{\pi(\alpha_1,\beta_1)}(T_1 \le s_i) = \int_0^{s_i} f_{\pi(\alpha_1,\beta_1)}(t_1)\, dt_1 = q_i\,, \qquad i = 1,\dots,h_1 \qquad (3)$$

In general, there may not be a distribution $\pi(\alpha_1,\beta_1)$ able to exactly fulfill all the $h_1$ Eqs. (3). Conversely, there is the possibility of satisfying exactly at least one of them. Thus, the expert can be finally asked to select among $\{(s_i,q_i)\}_{i=1,\dots,h_1}$ the pair $(s_{\bar{\imath}},q_{\bar{\imath}})$ which he/she feels more confident of. This is named *Most Trustworthy Specification* (MTS), and it is usually expected that $q_{\bar{\imath}}$ is too close neither to 0 nor to 1, as the

expert knowledge for central quantiles is usually better than that for those extremes (Garthwaite et al., 2005). The elicitation procedure, illustrated below, utilizes the MTS to ensure that $\pi(\alpha_1, \beta_1)$ exactly satisfies:

$$P_{\pi(\alpha_1,\beta_1)}(T_1 \leq s_{\bar{\imath}}) = \int_0^{s_{\bar{\imath}}} f_{\pi(\alpha_1,\beta_1)}(t_1)\, dt_1 = q_{\bar{\imath}} \tag{4}$$

According to (Bousquet, 2010), we can consider the marginal prior distribution $\pi(\alpha_1, \beta_1)$ as the posterior distribution that derives from the updating of the non-informative improper Jeffreys prior (Kass & Wasserman, 1996)

$$\pi^J(\alpha_1, \beta_1) \propto {\alpha_1}^{-1} \mathbb{I}_{\{\alpha_1 > 0\}} \mathbb{I}_{\{\beta_1 > \beta_1^0\}} \tag{5}$$

(where $\mathbb{I}$ denotes the indicator function) on the basis of a virtual sample $\tilde{\boldsymbol{s}}_{m_1} = (\tilde{s}_1, \ldots, \tilde{s}_{m_1})$ of $m_1$ i.i.d. uncensored observations of $T_1$, which are consistent with the expert opinions $\{(s_i, q_i)\}_{i=1,\ldots,h_1}$ (Figure 2). Notice that the value of $\beta_1^0 \geq 0$ has to be fixed according to the phenomenon under study, whereas the value of $m_1$ is derived from the optimization procedure described below.

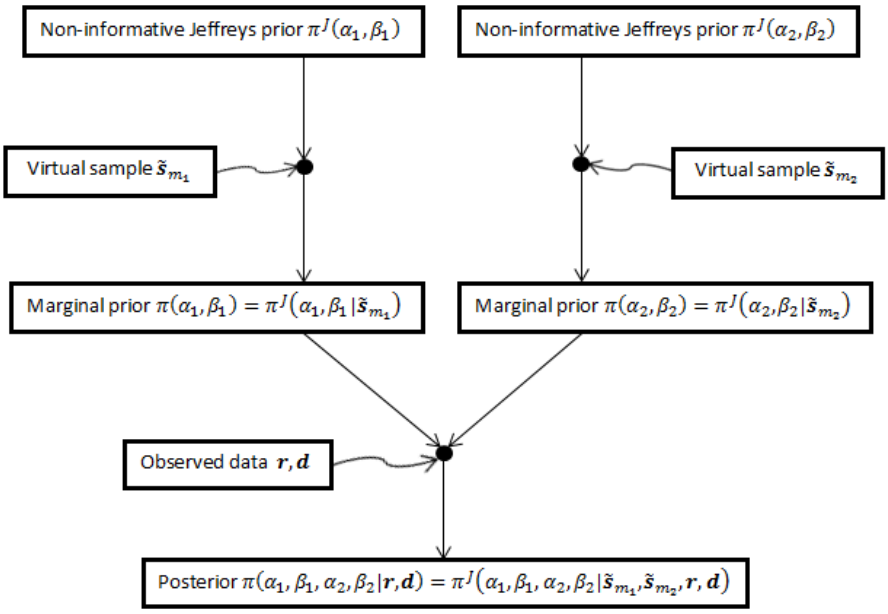



**Figure 2.** Scheme of the complete procedure

The application of the Bayes theorem yields:

$$\begin{aligned} \pi(\alpha_1, \beta_1) = \pi^J(\alpha_1, \beta_1|\tilde{\boldsymbol{s}}_{m_1}) \propto f(\tilde{\boldsymbol{s}}_{m_1}|\alpha_1, \beta_1)\pi^J(\alpha_1, \beta_1) = \\ = \prod_{i=1}^{m_1} \frac{\beta_1}{\alpha_1}\left(\frac{\tilde{s}_i}{\alpha_1}\right)^{\beta_1 - 1} \exp\left(-\left(\frac{\tilde{s}_i}{\alpha_1}\right)^{\beta_1}\right) {\alpha_1}^{-1} \mathbb{I}_{\{\alpha_1 > 0\}} \mathbb{I}_{\{\beta_1 > \beta_1^0\}} \end{aligned} \tag{6}$$

Eq. (6) can be handled as in (Bousquet, 2010) to get:

$$\pi(\alpha_1, \beta_1) = \pi(\alpha_1|\beta_1)\pi(\beta_1) \tag{7}$$

where:

- $\pi(\alpha_1|\beta_1)$ obeys to the Generalized Inverse Gamma (GIG) distribution, whose PDF is:

$$\pi(\alpha_1|\beta_1) \sim GIG(m_1, b, \beta_1) = \frac{b^{m_1}\beta_1}{\Gamma(m_1)} \frac{1}{{\alpha_1}^{m_1\beta_1+1}} \exp\left(-\frac{b}{{\alpha_1}^{\beta_1}}\right) \mathbb{I}_{\{\alpha_1>0\}} \tag{8}$$

- $\pi(\beta_1)$ is the left-truncated Gamma distribution, whose PDF is:

$$\begin{aligned} \pi(\beta_1) \sim G\left(m_1, \frac{m_1}{c_1}\right) \mathbb{I}_{\{\beta_1>\beta_1^0\}} = \\ = \left[1 - F_{\beta_1}(\beta_1^0)\right]^{-1} \cdot \frac{1}{\Gamma(m_1)} \left(\frac{m_1}{c_1}\right)^{m_1} {\beta_1}^{m_1-1} \exp\left(-\frac{m_1}{c_1}\beta_1\right) \mathbb{I}_{\{\beta_1>\beta_1^0\}} \end{aligned} \tag{9}$$

  where $\left[1 - F_{\beta_1}(\beta_1^0)\right]$ is the normalizing constant:

$$1 - F_{\beta_1}(\beta_1^0) = \int_{\beta_1^0}^{\infty} \frac{1}{\Gamma(m_1)} \left(\frac{m_1}{c_1}\right)^{m_1} {\beta_1}^{m_1-1} \exp\left(-\frac{m_1}{c_1}\beta_1\right) d\beta_1$$

- $b = b(m_1, \beta_1) = \left((1 - q_{\bar{\imath}})^{-\frac{1}{m_1}} - 1\right)^{-1} \cdot s_{\bar{\imath}}^{\beta_1}$ enters Eq. (8), and assures that the elicited prior exactly satisfies Eq. (4) (Bousquet, 2010);
- $m_1$ and $c_1$ are hyper-parameters to be properly fixed.

According to (Bousquet, 2010), the hyper-parameter $m_1$ could be set equal to the number of transition times $T_1$ observed by the expert, which are at the basis of his/her belief. This would be equivalent to the size of the virtual sample entering the procedure described above. Thus, if these earlier observations were all uncensored, then it would be sufficient to ask the expert the number of these observations and set $m_1$ to such value. Once $m_1$ is fixed, the calibration of $c_1(m_1)$ can be performed by minimizing the discrete Kullback-Leibler loss function between a generic PDF $f$ of $T_1$, satisfying the specifications $\{(s_i, q_i)\}_{i\neq\bar{\imath}}$, and the predictive PDF $f_\pi$ associated to the marginal prior distribution $\pi(\alpha_1, \beta_1)$ (Bousquet, 2010):

$$\begin{aligned} \mathcal{D}_{m_1}(f, f_\pi) = \sum_{i=0}^{\bar{\imath}-2} \left[(q_{i+1} - q_i) \log \frac{q_{i+1} - q_i}{q_{i+1}^{(e)} - q_i^{(e)}}\right] + \\ + (q_{\bar{\imath}+1} - q_{\bar{\imath}-1}) \log \frac{q_{\bar{\imath}+1} - q_{\bar{\imath}-1}}{q_{\bar{\imath}+1}^{(e)} - q_{\bar{\imath}-1}^{(e)}} + \sum_{i=\bar{\imath}+1}^{h_1} \left[(q_{i+1} - q_i) \log \frac{q_{i+1} - q_i}{q_{i+1}^{(e)} - q_i^{(e)}}\right] \end{aligned} \tag{10}$$

where:

– $$s_0 = 0,\ s_{h_1+1} = \infty,\ q_0 = q_0^{(e)} = 0,\ q_{h_1+1} = q_{h_1+1}^{(e)} = 1 \tag{11}$$

– $$q_i^{(e)} = \iint_{\mathbb{R}_+ \times \mathbb{R}_+} F_{T_1}(s_i|\alpha_1, \beta_1)\pi(\alpha_1|\beta_1)\pi(\beta_1)\, d\alpha_1 d\beta_1 \ \forall i = 1, \dots, h_1 \tag{12}$$

However, in the situation considered in this paper, the expert does not know exactly the time $T_1$ at which transitions have occurred. Rather, he/she knows the state of the component at the inspection times. The worth of this observation is smaller than that of an uncensored transition. To give due account to this, the size $m_1$ of the virtual sample must be smaller than the number of earlier observations seen by the expert and is very difficult to quantify. To overcome the issue, a joint calibration of the pair $(m_1, c_1)$ can be made via the following algorithm, proposed in (Bousquet, 2010).

| | |
|---|---|
| I. | Consider a set $M_1$ of possible values of $m_1$. |
| II. | $\forall m_1 \in M_1$ find $c_1^*(m_1) := \arg \min_{\pi(\cdot\,|c_1(m_1))} \mathcal{D}_{m_1}(f, f_\pi)$ |
| III. | find $m_1^* := \arg \min_{m_1 \geq 0} \mathcal{D}_{m_1}\left(f, f_{\pi(\cdot|c_1^*(m_1))}\right)$ |
| IV. | set $m_1 = m_1^*$ |
| V. | set $c_1 = c_1^*(m_1^*)$ |

Figure 3: Steps of the Joint calibration procedure

As suggested in (Bousquet, 2010), to efficiently solve the minimization problem at step II, the golden section search method combined with parabolic interpolation (Brent, 1973) is used, where the terms $q_i^{(e)}$ are estimated by substituting Eq. (8) and Eq. (9) in Eq. (12). After some manipulations, Eq. (12) becomes:

$$q_i^{(e)} = 1 - \int_{\beta_1^0}^{+\infty} \left(1 + \left[(1 - q_{\bar{\imath}})^{-\frac{1}{m_1}} - 1\right]\left(\frac{s_i}{s_{\bar{\imath}}}\right)^{\beta_1}\right)^{-m_1} \pi(\beta_1)\, d\beta_1. \tag{13}$$

Then, the terms $q_i^{(e)}$ can be estimated by applying the Monte-Carlo method (Zio, 2013):

$$q_i^{(e)} \approx 1 - \frac{1}{K}\sum_{k=1}^{K} \left(1 + \left[(1 - q_{\bar{\imath}})^{-\frac{1}{m_1}} - 1\right]\left(\frac{s_i}{s_{\bar{\imath}}}\right)^{\beta_1{}^{(k)}}\right)^{-m_1}, \tag{14}$$

where $\left(\beta_1{}^{(1)}, \dots, \beta_1{}^{(K)}\right)$ are $K$ i.i.d. samples from the truncated gamma distribution in Eq. (9), with $K \gg 1$. Notice that, once step II is taken, steps III – IV – V are straightforward. Notice also that the unicity of the solution $\left(m_1^*,\ c_1^*(m_1^*)\right)$ is guaranteed by the convexity of the loss function $\mathcal{D}$ with respect to $\pi$, along with the one-to-one continuous correspondence between $\pi\left(\beta_1|c_1(m)\right)$ and $c_1(m)$ (Bousquet, 2010). At the end of the calibration, one should verify whether the differences between the quantiles of the elicited prior and the correspondent expert specifications are small, possibly below the expert sensitivity degree. If the expert believes that the elicited quantiles are not well representative of his/her opinions, the use of a Weibull model should be reconsidered.

Large values of $m_1$ relate to an a priori distribution derived from updating the Jeffreys prior with a large number of data. For this, optimization procedures yielding large values of $m_1$ are indicative of informative priors.

Finally, all the considerations discussed in this Section apply also to $T_2$, including the calibration procedure. Additional details about the elicitation procedures can be found in (Garthwaite et al., 2005), where some methods, often based on psychological studies, for successfully asking and eliciting the information from the experts are reviewed.

## 4.2 Posterior Distributions

In this Section, a Markov Chain Monte Carlo (MCMC, Robert & Casella, 2004) algorithm is developed to estimate the posterior distributions of the model parameters. MCMC is a family of algorithms that allow drawing samples from a probability distribution $\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})$ (usually referred to as target distribution), which are produced by an ergodic Markov chain.

Under the assumptions mentioned in Section 2, the contribution $L(r_n, d_n|\boldsymbol{\theta})$ of the $n^{th}$ observation to the likelihood function

$$L(\boldsymbol{r},\boldsymbol{d}|\boldsymbol{\theta}) = \prod_{n=1}^{N} L(r_n, d_n|\boldsymbol{\theta}) \tag{15}$$

is such that:

- if $d_n = 1$, $L(r_n, d_n|\boldsymbol{\theta})$ is the probability of the event $\{T_{1n} > r_n|\boldsymbol{\theta}\}$:

$$L(r_n, d_n = 1|\boldsymbol{\theta}) = 1 - F_{T_1}(r_n|\alpha_1,\beta_1) \tag{16}$$

  where $F_{T_1}$ is the Cumulative Distribution Function (CDF) of the random variable $T_1|\ \alpha_1,\beta_1$. (Notice that this outcome does not provide any useful information about the realization of $T_2$).

- If $d_n = 2$, $L(r_n, d_n|\boldsymbol{\theta})$ is the probability of the event $\{T_{1n} \le r_n\ and\ T_{1n} + T_{2n} > r_n|\boldsymbol{\theta}\}$:

$$L(r_n, d_n = 2|\boldsymbol{\theta}) = \int_0^{r_n} \left[1 - F_{T_2}(r_n - \tau|\alpha_2,\beta_2)\right] f_{T_1}(\tau|\alpha_1,\beta_1)d\tau \tag{17}$$

  where $f_{T_1}$ is the PDF of $T_1|\ \alpha_1,\beta_1$, whereas $F_{T_2}$ is the CDF of $T_2|\ \alpha_2,\beta_2$.

- If $d_n = 3$, $L(r_n, d_n|\boldsymbol{\theta})$ is the probability of the event $\{T_{1n} + T_{2n} \le r_n|\boldsymbol{\theta}\}$:

$$L(r_n, d_n = 3|\boldsymbol{\theta}) = \int_0^{r_n} F_{T_2}(r_n - \tau|\alpha_2,\beta_2) f_{T_1}(\tau|\alpha_1,\beta_1)d\tau \tag{18}$$

The application of the Bayes theorem yields the following posterior distribution:

$$\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d}) \propto \prod_{n=1}^{N} L(r_n, d_n|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}) \tag{19}$$

Finally, notice that it is not possible to decompose the likelihood function in the form $L(\boldsymbol{r},\boldsymbol{d}|\boldsymbol{\theta}) = L(\boldsymbol{r},\boldsymbol{d}|\alpha_1,\beta_1)L(\boldsymbol{r},\boldsymbol{d}|\alpha_2,\beta_2)$, being the random vectors $(\alpha_1,\beta_1)$ and $(\alpha_2,\beta_2)$ not a-posteriori independent.

### 4.2.1 Random Walk Metropolis-Hastings algorithm

In this work, a Normal Random Walk Metropolis-Hastings (N-RWMH, Robert & Casella, 2004) algorithm is used to obtain samples from the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})$ of the model parameters. Formally, we indicate by $\Theta = (\mathbb{R}_+)^w$ the state space, by $\mathfrak{B}(\Theta)$ the Borel $\sigma$-algebra on $\Theta$ and by $q(\cdot\,;\boldsymbol{\mu},\Sigma) = \mathcal{N}(\cdot\,;\boldsymbol{\mu},\Sigma)$ a normal proposal density with mean $\boldsymbol{\mu}$ and covariance matrix $\Sigma$. The transition probability of the N-RWMH algorithm from the current state $\boldsymbol{\theta} \in \Theta$ to a set $E \in \mathfrak{B}(\Theta)$ is given by:

$$P_{q_\Sigma}(\boldsymbol{\theta},E) = \int_E \xi(\boldsymbol{\theta},\boldsymbol{\nu})q(\boldsymbol{\nu};\boldsymbol{\theta},\Sigma)d\boldsymbol{\nu} + \left(1 - \int_\Theta \xi(\boldsymbol{\theta},\boldsymbol{\nu})q(\boldsymbol{\nu};\boldsymbol{\theta},\Sigma)d\boldsymbol{\nu}\right)\mathbb{I}_{\{\boldsymbol{\theta}\in E\}} \qquad (20)$$

where $\boldsymbol{\nu}$ is the integration variable and $\xi(\boldsymbol{\theta},\boldsymbol{\nu}) = \min\left(1,\frac{\pi(\boldsymbol{\nu}|\boldsymbol{r},\boldsymbol{d})}{\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})}\right)$.

The Markov chain defined by Eq. $(20)$ has, under appropriate assumptions (e.g., Gilks at al., 1996), $\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})$ as its unique invariant distribution. A desired number $G$ of steps of this Markov chain can be simulated by the following well known procedure (e.g., Jackman, 2009):

- Choose a (symmetric positive definite) covariance matrix $\Sigma \in \mathbb{R}^{w\times w}$ and set an initial state $\boldsymbol{\theta}^{(0)} \in \Theta$
- for $g = 1,\dots,G$:
    1. sample a candidate new state $\boldsymbol{\theta}^*$ from $q\left(\cdot\,;\boldsymbol{\theta}^{(g-1)},\Sigma\right)$
    2. set $\varphi = \frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{r},\boldsymbol{d})}{\pi(\boldsymbol{\theta}^{(g-1)}|\boldsymbol{r},\boldsymbol{d})}$ by using Eq. $(19)$
    3. set $\xi = \min(1,\varphi)$
    4. sample $U \sim Unif(0,1)$
    5. if $U \le \xi$
        set $\boldsymbol{\theta}^{(g)} = \boldsymbol{\theta}^*$
       else
        set $\boldsymbol{\theta}^{(g)} = \boldsymbol{\theta}^{(g-1)}$
       end if

    end for

The proper setting of the entries of the covariance matrix $\Sigma$ is fundamental for the algorithm efficiency: small values will origin an inefficient algorithm due to a bad mixing of the chain, whereas large values will lead to an excessively low acceptance rate (step 5 of the algorithm above).

If there were only one parameter to be estimated, the definition of the variance $\Sigma$ could have been done via a trial-and-error procedure: the variance is increased if the acceptance ratio is too low and reduced otherwise, until a satisfactory value is found. In this respect, (Gelman et al., 1996) suggests the optimal acceptance rate is about 0.44 for $d = 1$ and 0.234 for $d \to +\infty$.

On the contrary, we consider the situation in which there are four parameters. In this case, the task of calibrating $\Sigma$ becomes harder, and one can resort to an adaptive MCMC algorithm.

In this work, the following version by (Andrieu & Thoms, 2008) of the adaptive algorithm (A-MCMC) described in (Haario et al., 2001) is employed:

- initialize $\boldsymbol{\theta}^{(0)} \in \Theta$, initialize $\boldsymbol{\eta}^{(0)} = \mathbf{0} \in \mathbb{R}^w$ and a (symmetric positive definite) covariance matrix $\Sigma^{(0)} \in \mathbb{R}^{w \times w}$
- for $s = 1, \dots, S$
    1. sample a candidate new state $\boldsymbol{\theta}^*$ from $q\left(\cdot\,; \boldsymbol{\theta}^{(s-1)}, \lambda\Sigma^{(s-1)}\right)$
    2. set $\varphi = \frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{r},\boldsymbol{d})}{\pi\left(\boldsymbol{\theta}^{(s-1)}|\boldsymbol{r},\boldsymbol{d}\right)}$
    3. set $\xi = \min(1, \varphi)$
    4. sample $U \sim Unif(0,1)$
    5. if $U \leq \xi$
        set $\boldsymbol{\theta}^{(s)} = \boldsymbol{\theta}^*$
       else
        set $\boldsymbol{\theta}^{(s)} = \boldsymbol{\theta}^{(s-1)}$
       end if
    6. update $\boldsymbol{\eta}^{(s)} = \boldsymbol{\eta}^{(s-1)} + \gamma^{(s)}\left(\boldsymbol{\theta}^{(s)} - \boldsymbol{\eta}^{(s-1)}\right)$
    7. update $\Sigma^{(s)} = \Sigma^{(s-1)} + \gamma^{(s)}\left(\left(\boldsymbol{\theta}^{(s)} - \boldsymbol{\eta}^{(s-1)}\right)\left(\boldsymbol{\theta}^{(s)} - \boldsymbol{\eta}^{(s-1)}\right)^T - \Sigma^{(s-1)}\right)$

  end for

where $\lambda = \frac{(2.38)^2}{w}$, $\left\{\gamma^{(s)} = \frac{1}{s^\delta}\right\}_{s=1,\dots,S}$, $\delta \in ((1+\lambda)^{-1}, 1]$.

The stochastic process defined by the A-MCMC algorithm is no longer Markovian, being each step dependent on all the past iterations. The loss of the Markov property is a common feature for the majority of adaptive algorithms and could undermine the convergence to the target distribution $\pi(\boldsymbol{\theta}|\boldsymbol{r}, \boldsymbol{d})$. For this reason, in this work we use the adaptive algorithm only to set possibly acceptable values for the entries of the covariance matrix of the proposal distribution to be used in the N-RWMH. In practice, the adaptive algorithm is stopped after a proper number $S$ of iterations, and the states visited by the chain are discarded. Then, the N-RWMH is run for $G$ iterations, which uses the values of the last covariance matrix $\Sigma^{(S)}$.

To set S, one can look at the stability of $\boldsymbol{\eta}^s$ and of the covariance matrix $\Sigma^{(s)}$ over $s$: S must be large enough that the changes in their values are acceptably small.

Notice that the integrals in Eqs. $(17)$ and $(18)$ are convolutions of a Weibull PDF with parameters $\alpha_1, \beta_1$ with the CDF and the complementary CDF of a Weibull distribution with parameters $\alpha_2, \beta_2$, respectively. The results of these convolutions are not available in closed form; then, we have to numerically compute these integrals within both the N-RWMH and A-MCMC algorithms, which is time-consuming. An alternative algorithm to estimate the posterior distribution, which avoids numerical estimations to sample from $\pi(\alpha_1, \beta_1, \alpha_2, \beta_2, T_{11}, \dots, T_{1N}, T_{21}, \dots, T_{2N}|\boldsymbol{r}, \boldsymbol{d})$, is the data-augmented Metropolis-within-Gibbs algorithm (e.g., Laird, 2013). This latter algorithm may become particularly advantageous if the number of degradation states considered in the model increases, as it avoids the numerical estimation of very complex integrals (e.g., Compare et al., 2015). However, the method has the potential drawback of dramatically increasing the Markov Chain state-space dimension in case of a large number $N$ of observations. The application of the Metropolis-within-Gibbs algorithm to multi-state degradation modeling and its comparison with N-RWMH will be investigated in further works.

### 4.2.2 MCMC diagnostics

Some diagnostic methods are executed on the $G$ draws, sampled by the procedure described above, in order to check:

1. The convergence of the chain to its invariant distribution, i.e., the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})$.
2. The algorithm efficiency.
3. The need of introducing/increasing the burn-in length, i.e., the number of initial iterations that have to be discarded as long as the chain is judged to have not yet converged. These samples must be removed from the chain, as they may bias the sampling from the posterior distribution.

The diagnostic methods used in this work are traceplots, autocorrelation plots, the Effective Sample Size and the Geweke test (e.g., Jackman,2009; Geweke, 1992). Notice that we are considering both qualitative and quantitative diagnostics, which are applied separately on the members of the vector $\boldsymbol{\theta}$**.** More details on the employed diagnostic methods can be found in the Appendix.

## 4.3 Posterior inference

Once *G* samples $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(G)}$ have been obtained from the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})$, we can use them to predict the evolution of the component degradation process. Depending on the maintenance approach, different information can be extracted from the posterior distribution of the parameters.

Scheduled maintenance is typically applied when the identification of the component degradation state requires to perform a destructive test. In this case, the component is preventively replaced when it reaches a predefined age, which can be set by considering the expected degradation path of a new generic component. To this purpose, the following information can be obtained by the obtained Bayesian semi-Markov model (Section 3.3.1):

- The probabilities of occupying the three degradation states over time.
- The distribution of the failure time, i.e., the time until the transition to state 3.

Differently, a predictive maintenance approach can be applied when there is the possibility of performing a non-destructive inspection of the component degradation state. In practice, the component is inspected, its degradation state identified and its failure time is predicted using the developed Bayesian semi-Markov model (Section 3.3.2). Notice that the stochastic quantity which we are interested in is the Remaining Useful Life (RUL), i.e., the remaining time until the transition toward state 3 of the specific component that we have observed being in a given state.

### 4.3.1 Posterior inference for scheduled maintenance

With reference to the probability of occupying the degradation states over time, we want to estimate the following values:

- The expected probability of occupying state 1 after working time $t \geq 0$:

$$\mathbb{E}_{\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})}[P(T_1 > t|\boldsymbol{\theta})] = \int_{\mathbb{R}_+\times\mathbb{R}_+} \left(1 - F_{T_1}(t|\alpha_1,\beta_1)\right)\pi(\alpha_1,\beta_1|\boldsymbol{r},\boldsymbol{d})d\alpha_1 d\beta_1 \qquad (21)$$

- The expected probability of occupying state 2 after working time $t \geq 0$:

$$\mathbb{E}_{\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})}[P(T_1 \leq t, T_1 + T_2 > t|\boldsymbol{\theta})] = \\ = \int_{\Theta} \left( \int_0^t \left[1 - F_{T_2}(t-\tau|\alpha_2,\beta_2)\right] f_{T_1}(\tau|\alpha_1,\beta_1) d\tau \right) \pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d}) d\boldsymbol{\theta} \quad (22)$$

- The expected probability of occupying state 3 after working time $t \geq 0$:

$$\mathbb{E}_{\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})}[P(T_1 + T_2 \leq t|\boldsymbol{\theta})] = \int_{\Theta} \left( \int_0^t F_{T_2}(t-\tau|\alpha_2,\beta_2) f_{T_1}(\tau|\alpha_1,\beta_1) d\tau \right) \pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d}) d\boldsymbol{\theta} \quad (23)$$

Moreover, we are also interested in estimating a $(0.05, 0.95)$ pointwise credibility band for each of the three probabilities, which describes the uncertainty on the estimated probability value. To do this, we first define a proper time grid $\{t_j\}_{j=1,\dots,J}$ at which the quantities in Eqs. $(21-23)$ will be calculated, $0 = t_1 < t_2 < \cdots < t_J$. Then, we exploit the MCMC samples to make inference about the probability of occupying the states (Christensen et al., 2010).

For example, with reference to degradation state 1, we create the matrix

$$\left[ \left\{ 1 - F_{T_1}\left(t_j|\alpha_1^{(g)},\beta_1^{(g)}\right) \right\}_{gj} \right], \qquad g = 1,\dots,G\,, \qquad j = 1,\dots,J$$

whose $g^{th}$ row contains the $J$ values of the Complementary CDF of the random variable $T_1|\boldsymbol{\theta} = \boldsymbol{\theta}^{(g)}$ evaluated on the time grid $\{t_j\}_{j=1,\dots,J}$. Unbiased estimates of the expected values (over time) in Eq. $(21)$ are given by the averages of the values in every matrix column, whereas the $(0.05, 0.95)$ pointwise credibility band for the probability of occupying degradation state 1 is defined by the quantiles of order $0.05$ and $0.95$ of the columns.

Similarly, the matrices

$$\left[ \left\{ \int_0^t \left[1 - F_{T_2}(t_j - \tau|\alpha_2^{(g)},\beta_2^{(g)})\right] f_{T_1}\left(\tau|\alpha_1^{(g)},\beta_1^{(g)}\right) d\tau \right\}_{gj} \right], \qquad g = 1,\dots,G\,, \qquad j = 1,\dots,J$$

and:

$$\left[ \left\{ \int_0^t F_{T_2}(t_j - \tau|\alpha_2^{(g)},\beta_2^{(g)}) f_{T_1}\left(\tau|\alpha_1^{(g)},\beta_1^{(g)}\right) d\tau \right\}_{gj} \right], \qquad g = 1,\dots,G\,, \qquad j = 1,\dots,J$$

are used to make inference about the probability of occupying states 2 and 3, respectively, where their entries have to be computed via numerical integration methods.

Notice that the available gathered evidence for estimating the probabilities in Eqs. $(21-23)$ is that of the original dataset $(\boldsymbol{r},\boldsymbol{d}) = \{(r_n, d_n)\}_{n=1,\dots,N}$ of $N$ records.

### 4.3.2 Posterior inference for predictive maintenance

Assume now that a $(N+1)^{th}$ component, currently operating in the production line, is detected in a state $d_{N+1} < 3$ after a working time $r_{N+1} > 0$. The proposed Bayesian framework allows predicting its RUL to support maintenance decisions about the component replacement. To do this, we compute the posterior density $\pi\big(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d},(r_{N+1},d_{N+1})\big)$, which updates $\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})$ with the new information contained in the observation $(r_{N+1},d_{N+1})$.

Then, in case $d_{N+1}=1$, the CDF of the arrival time in state 3, $T_1+T_2$, for the $(N+1)^{th}$ component is:

$$\begin{aligned} &F_{T_1+T_2}\big(t|\boldsymbol{r},\boldsymbol{d},(r_{N+1},d_{N+1}=1)\big)= \\ &\quad=\int_{\Theta} P(T_1+T_2\le t|\boldsymbol{r},\boldsymbol{d},(r_{N+1},d_{N+1}=1),\boldsymbol{\theta})\pi\big(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d},(r_{N+1},d_{N+1}=1)\big)d\boldsymbol{\theta} \end{aligned} \tag{24}$$

where

$$P(T_1+T_2\le t|\boldsymbol{r},\boldsymbol{d},(r_{N+1},d_{N+1}=1),\boldsymbol{\theta})=\int_{r_{N+1}}^{t} F_{T_2}(t-\tau|\boldsymbol{\theta})\frac{f_{T_1}(\tau|\boldsymbol{\theta})}{1-F_{T_1}(r_{N+1}|\boldsymbol{\theta})}d\tau \tag{25}$$

Eq. (25) derives from the definition of conditional probability, which applies to the first transition time distribution, only. From this, the estimation of the RUL can be done through the following procedure:

1. Draw $G$ samples $\boldsymbol{\theta}_{up}^{(1)},\dots,\boldsymbol{\theta}_{up}^{(G)}$ from the updated posterior distribution $\pi\big(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d},(r_{N+1},d_{N+1}=1)\big)$ using the MCMC algorithm.
2. Compute, via numerical integration methods, the vector of values:
$$\left\{\int_{r_{N+1}}^{t} F_{T_2}\left(t-\tau|\boldsymbol{\theta}_{up}^{(g)}\right)\frac{f_{T_1}\left(\tau|\boldsymbol{\theta}_{up}^{(g)}\right)}{1-F_{T_1}\left(r_{N+1}|\boldsymbol{\theta}_{up}^{(g)}\right)}d\tau\right\}_{g=1,\dots,G}$$
3. Subtract $r_{N+1}$ from the average of the above vector to get an unbiased RUL estimate.

Similarly, in the case $d_{N+1}=2$, the CDF of the arrival time in state 3 is:

$$\begin{aligned} &F_{T_1+T_2}\big(t|\boldsymbol{r},\boldsymbol{d},(r_{N+1},d_{N+1}=2)\big)= \\ &\quad=\int_{\Theta} P(T_1+T_2\le t|\boldsymbol{r},\boldsymbol{d},(r_{N+1},d_{N+1}=2),\boldsymbol{\theta})\pi\big(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d},(r_{N+1},d_{N+1}=2)\big)d\boldsymbol{\theta} \end{aligned} \tag{26}$$

where (see the Appendix for the proof):

$$\begin{aligned} &P(T_1+T_2\le t|\boldsymbol{r},\boldsymbol{d},(r_{N+1},d_{N+1}=2),\boldsymbol{\theta})= \\ &\qquad=\frac{\int_0^{r_{N+1}}\big[F_{T_2}(t-\tau|\boldsymbol{\theta})-F_{T_2}(r_{N+1}-\tau|\boldsymbol{\theta})\big]f_{T_1}(\tau|\boldsymbol{\theta})d\tau}{\int_0^{r_{N+1}}\big[1-F_{T_2}(r_{N+1}-\tau|\boldsymbol{\theta})\big]f_{T_1}(\tau|\boldsymbol{\theta})\,d\tau} \end{aligned} \tag{27}$$

In this case, the procedure to estimate the RUL becomes:

1. Draw $G$ samples $\boldsymbol{\theta}_{up}^{(1)}, \dots, \boldsymbol{\theta}_{up}^{(G)}$ from the updated posterior distribution $\pi\left(\boldsymbol{\theta}|\boldsymbol{r}, \boldsymbol{d}, (r_{N+1}, d_{N+1} = 2)\right)$ via a MCMC algorithm
2. Compute, with the help of numerical integration methods, the vector of values: $\left\{\frac{\int_0^{r_{N+1}}\left[F_{T_2}\left(t-\tau|\boldsymbol{\theta}_{up}^{(g)}\right)-F_{T_2}\left(r_{N+1}-\tau|\boldsymbol{\theta}_{up}^{(g)}\right)\right]f_{T_1}\left(\tau|\boldsymbol{\theta}_{up}^{(g)}\right)d\tau}{\int_0^{r_{N+1}}\left[1-F_{T_2}\left(r_{N+1}-\tau|\boldsymbol{\theta}_{up}^{(g)}\right)\right]f_{T_1}\left(\tau|\boldsymbol{\theta}_{up}^{(g)}\right)d\tau}\right\}_{g=1,\dots,G}$
3. An unbiased estimate for the $RUL$ can be obtained by subtracting $r_{N+1}$ to the mean of the above vector.

## 4.4 Sensitivity analysis

In the Bayesian framework, the choice of a proper prior distribution is one of the most delicate issues since it requires the elicitation of a distribution consistent with the prior knowledge (Gelman, 2002). Thus, sensitivity analysis methods are usually applied to quantify the effects of modifications of the parameter prior distribution and its hyper-parameters on the posterior distribution and on the predictions of interest (e.g., Roos et al., 2013; Oakley & O'Hagan, 2004). This study becomes particularly important in case of scarcity of field data, when the posterior predictions are expected to be strongly dependent from the prior distribution. In particular, within the framework proposed in this work, we are interested in assessing the influence of the values of the hyper-parameters $m_1$ and $m_2$, which can be interpreted as weights given to the expert judgements about transition times $T_1$ and $T_2$, respectively (Section 3.1).

To do this, the procedure for the computation of the posterior distribution and of the predictors of interest (Sections 3.2 and 3.3) can be repeated assigning different values to the hyper-parameters $m_1$ and $m_2$.

In practice, in the case in which only few field data are available and expert opinions on the transition times are vague, the final choice of the hyper-parameters $m_1$ and $m_2$ should take into account also considerations about risk aversion. Let us consider, for example, a critical component whose failure entails much larger economical losses than the cost of its replacement. In a situation in which the few field data available would suggest transition times longer than that suggested by the expert, a cautious decision-maker should overweigh the expert opinions and, thus, use large $m_1$ and $m_2$ values also on the basis of the findings of the sensitivity analysis.

# 5 Case study

In order to develop the three-state model of the degradation process (Section 2), the prior distributions of the model parameters have been elicited from an expert using a questionnaire. The expert was asked to provide the 0.10, 0.50, 0.90 quantiles of the transition time $T_1$ and, additionally, he had the possibility of estimating for some proposed $T_1$ values the correspondent quantile orders. Expert answers are reported in Table 2. Finally, the expert indicated as Most Trustworthy Specification (MTS) the pair $(s_{\bar{\imath}}, q_{\bar{\imath}}) = (227, 0.50)$.

| # | $s_i$ | $q_i$ |
|---|---|---|
| **1** | 50 | **0.01** |
| **2** | 100 | **0.05** |
| **3** | **143** | 0.1 |
| **4** | 160 | **0.15** |
| **5** | 195 | **0.30** |
| **6** | **227** | 0.5 |

| | | |
|---|---|---|
| **7** | 250 | **0.7** |
| **8** | **300** | 0.9 |

**Table 2.** Expert opinions for transition time $T_1$. The value $s_i$, expressed in terms of working hours, indicates the quantile of order $q_i$ of transition time $T_1$. Bold values correspond to the opinions provided by the expert, whereas the other values were proposed in the questionnaire

In this case study, it is reasonable to assume that the hazard function of the transition times between two consecutive states is a non-decreasing function. In fact, this degradation process is characterized by a continuous worsening of the conditions of the component, which leads to an increasing risk of having a transition into a successive degradation state. Hence, the choice of $\beta_1^0 \geq 1$ seems to be adequate. In particular, without further information, we assumed $\beta_1^0 = 1$.

Figure 4 shows the results of the calibration of the hyper-parameters $m_1$ and $c_1$ according to the procedure illustrated in Section 4.1. Notice that the value of $m_1$ which minimizes $\mathcal{D}_{m_1}\left(f, f_{\pi(\cdot|c_1^*(m_1))}\right)$ is $m_1^* \approx 3.5$ (Figure 4 right), which corresponds to an optimal value of $c_1$, $c_1^*(m_1^*) \approx 5.78$.

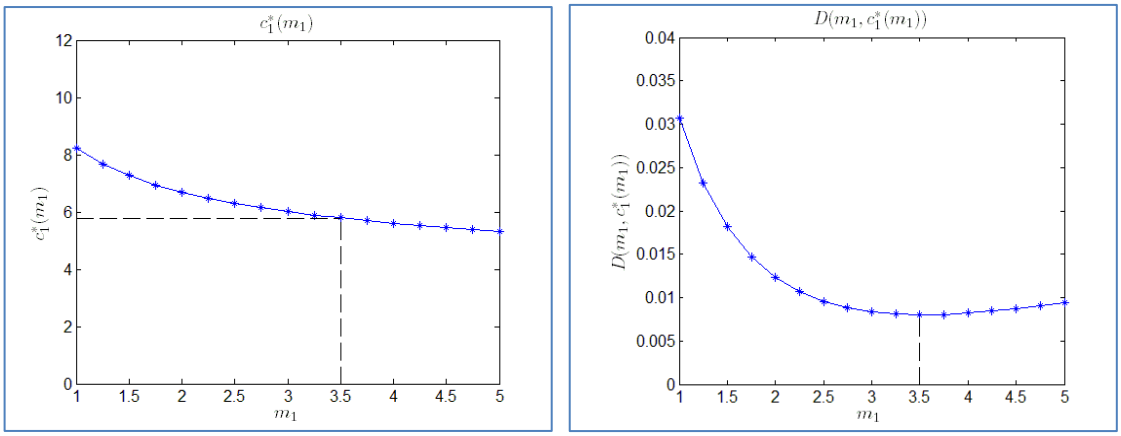


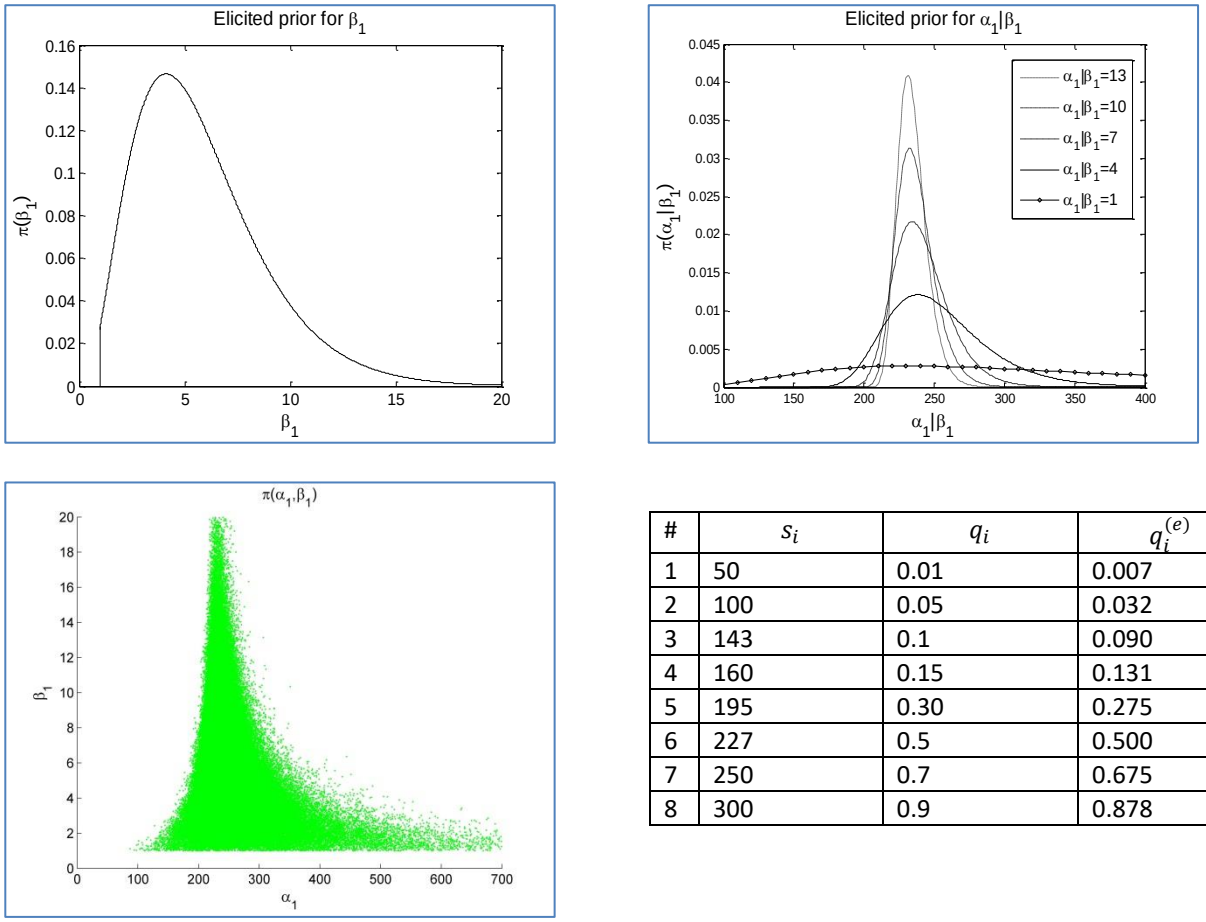

**Figure 4.** Results of step II (left) and III (right) of the joint calibration procedure for the pair $(m_1, c_1)$ (Figure 3)

The results of the elicitation of the prior distributions, with values $m_1 = 3.5$ and $c_1 = 5.78$ in Eq. (8) and Eq. (9), are shown in Figure 5. Notice that the distribution $\pi(\beta_1)$ is quite uninformative, since it embraces with non-negligible probability also values such as $\beta_1 > 5$: according to (Bousquet, 2010), these values can be judged as physically extreme in a context of reliability analysis of industrial installations. The distributions $\pi(\alpha_1|\beta_1 = y)$ tend to concentrate in a neighbourhood of $s_{\bar{i}}$ as $y$ increases. This can be observed also in Figure 5 lower left, where $2 \cdot 10^5$ i.i.d. samples from the marginal prior $\pi(\alpha_1, \beta_1)$ are shown. Finally, Figure 5 lower right, shows that the discrepancy between the elicited prior predictive distribution and the expert opinions is reasonably low.

With regards to the transition time $T_2$, the expert gave the opinions reported in Table 3, and he also specified that he felt very little confident about them. Furthermore, since very few components have been found in state 3 during the inspection, the dataset conveys scarce information about the parameters $\alpha_2$ and $\beta_2$. For this reason, in accordance with the expert feeling, instead of applying the proposed optimization procedure to find the hyper-parameter $m_2$, we prefer to assign to $m_2$ a small value (i.e., $m_2 = 0.5$), which corresponds to a weakly informative prior on the parameters $\alpha_2$ and $\beta_2$. Then, an analysis of the sensitivity of the prediction to the $m_2$ values has been performed by repeating the predictions with $m_2 = 3.7$ and $m_2 = 10$.

In correspondence of $m_2 = 0.5$ and with the pair (30, 0.50) chosen as $MTS$, the result of step II of the procedure in Figure 3 is $c_2^* \approx 10.05$. The obtained marginal prior distribution $\pi(\alpha_2, \beta_2)$ is shown in Figure 6.

Notice that, although 50 working hours is indicated by the expert as the 0.90 quantile, the choice of a small value for $m_2$ causes a prior distribution in which the probability of a larger scale parameter is not negligible, especially for those values of the shape parameter $\beta_2$ that are physically most admissible.




| # | $s_i$ | $q_i$ | $q_i^{(e)}$ |
|---|---|---|---|
| 1 | 50 | 0.01 | 0.007 |
| 2 | 100 | 0.05 | 0.032 |
| 3 | 143 | 0.1 | 0.090 |
| 4 | 160 | 0.15 | 0.131 |
| 5 | 195 | 0.30 | 0.275 |
| 6 | 227 | 0.5 | 0.500 |
| 7 | 250 | 0.7 | 0.675 |
| 8 | 300 | 0.9 | 0.878 |

**Figure 5.** Obtained marginal prior distribution $\pi(\beta_1)$ (upper left); the PDFs $\pi(\alpha_1|\beta_1 = y)$ for some possible values $y = 1,4,7,10,13$ that $\beta_1$ can assume (upper right); 200000 i.i.d. samples from the joint prior distribution $\pi(\alpha_1, \beta_1)$ (lower left); the Table comparing the quantiles of the predictive distribution arising from the elicited prior with the expert opinions (lower right)

| # | $s_i$ | $q_i$ |
|---|---|---|
| 1 | 50 | 0.10 |
| 2 | 100 | 0.50 |
| 3 | 143 | 0.90 |

**Table 3.** Expert opinions for transition time $T_2$. The value $s_i$, expressed in terms of working hours, indicates the quantile of order $q_i$ of transition time $T_2$. Bold values correspond to the opinions provided by the expert, whereas the other values were proposed in the questionnaire

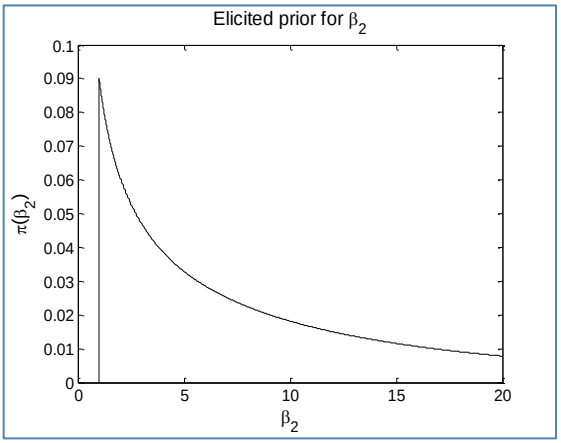
Once the prior distributions have been obtained, we have applied the following procedure to sample from the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{r}, \boldsymbol{d})$:

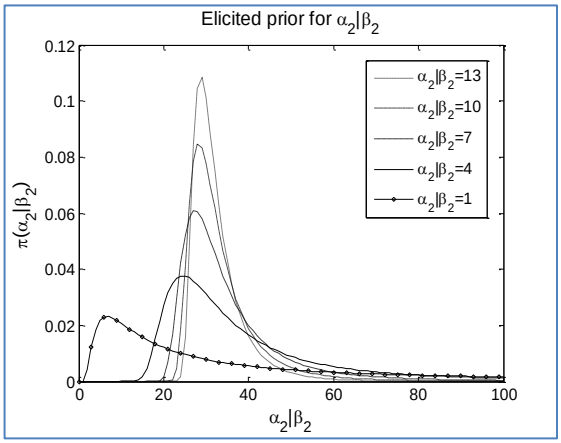
- $S = 100000$ iterations of the A-MCMC algorithm are run, for tuning the entries of the proposal distribution covariance matrix $\Sigma$.

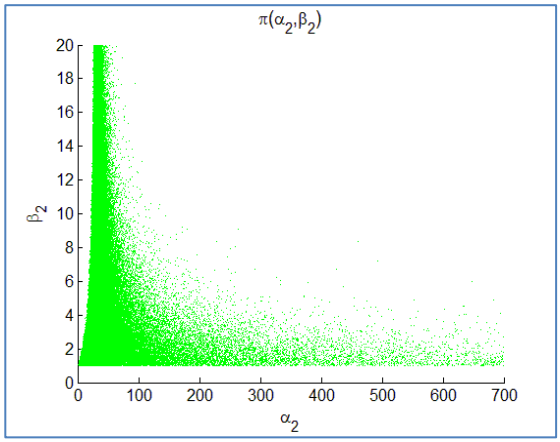
- $G = 400000$ iterations of the N-MHRW algorithm are run with the proposal distribution covariance matrix setted to $\Sigma^{(S)}$.

The obtained results of the MCMC simulations are shown in Figure 7.

To test the convergence of the chain, we have applied the convergence diagnostic tools described in Appendix. The results are summarized in Table 4. From the traceplots and the Geweke tests, there is no evidence of the need of a burn-in period (Figures 7a, 7b). The autocorrelation plots show that the algorithm is not extremely efficient, due to its rejection-based nature (with $w = 4$, the A-MCMC tunes the proposal density in order to obtain an optimal acceptance ratio of about $25\%$). However, the $400000$ iterations performed allow obtaining an effective sample size greater than or equal to $6475$ for all the parameters.




| # | $s_i$ | $q_i$ | $q_i^{(e)}$ |
|---|---|---|---|
| 1 | 50 | 0.10 | 0.074 |
| 2 | 100 | 0.50 | 0.50 |
| 3 | 143 | 0.90 | 0.871 |

**Figure 6.** Obtained marginal prior distribution $\pi(\beta_2)$ (upper left); the PDFs $\pi(\alpha_2|\beta_2 = y)$ for some possible values $y = 1,4,7,10,13$ that $\beta_2$ can assume (upper right); 200000 i.i.d. samples from the joint prior distribution $\pi(\alpha_2, \beta_2)$ (lower left); the Table comparing the quantiles of the predictive distribution arising from the elicited prior with the expert opinions (lower right)

| Chain | Effective sample size | Geweke test p-value |
|---|---|---|
| $\left\{\alpha_1^{(g)}\right\}_{g=1,\dots,G}$ | 19145 | 0.524 |
| $\left\{\beta_1^{(g)}\right\}_{g=1,\dots,G}$ | 19745 | 0.878 |

| | | |
|---|---|---|
| $\left\{\alpha_2^{(g)}\right\}_{g=1,\ldots,G}$ | 6475 | 0.113 |
| $\left\{\beta_2^{(g)}\right\}_{g=1,\ldots,G}$ | 17341 | 0.071 |

**Table 4.** The effective sample sizes and the p-values of the Geweke test from the simulated Markov chain

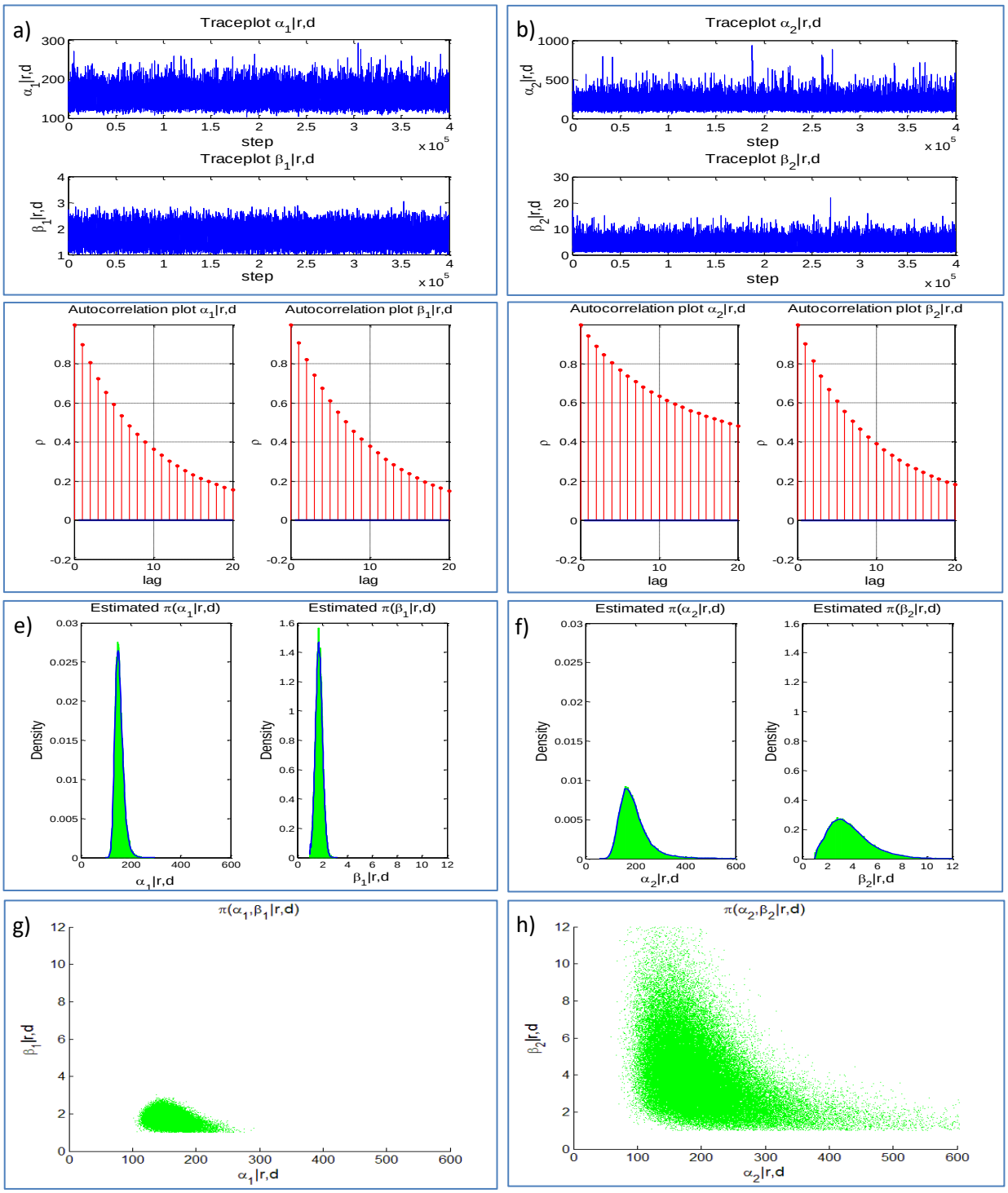



**Figure 7.** Results of the MCMC simulation from $\pi(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{d})$: traceplots (7a, 7b), autocorrelation plots (7c, 7d), histograms with the kernel density estimations of the marginal posterior densities for each parameter (7e, 7f), and 200000 MCMC samples out of 400000 from $\pi(\alpha_1,\beta_1|\boldsymbol{r},\boldsymbol{d})$ and $\pi(\alpha_2,\beta_2|\boldsymbol{r},\boldsymbol{d})$ (7g, 7h)

Figure 7 shows that the posterior marginal density $\pi(\alpha_1,\beta_1|\boldsymbol{r},\boldsymbol{d})$ has moved towards lower values of the scale parameter with respect to the corresponding prior marginal density $\pi(\alpha_1,\beta_1)$ (Figure 5). This is probably due to the fact that the expert opinions do not take into account the effects of some harsh operational conditions experienced by the diaphragms during their lives. Future work will be developed to embed the operational conditions into a Proportional Hazards model.

The comparison of the posterior marginal density $\pi(\alpha_2, \beta_2 | \boldsymbol{r}, \boldsymbol{d})$ of Figure 7 with the corresponding prior marginal density $\pi(\alpha_2, \beta_2)$ of Figure 6 shows that i) the uncertainty on the shape parameter remains quite large, ranging about from 1 to 10, ii) there is a consistent increment of the scale parameter. Thus, the collected data seem to suggest that the transition time $T_2$ is larger than what the expert believes.

Once an effective sample of parameters values taken from the posterior distribution becomes available, it can be used to estimate the a posteriori probabilities of being in the three states over time, with the corresponding $(0.05, 0.95)$ pointwise credibility bands. Figure 8 shows that the large uncertainty on parameters $\alpha_2$ and $\beta_2$ causes very large credibility bands for the probability of occupying states 2 and 3, especially after 150 working hours. This result is due to the fact that the majority (104 out of 109) of the components of the dataset were replaced within 173 working hours. The validity of the estimates on longer time windows relies on the assumption that the conditional Weibull model is adequate to describe the second transition time. This assumption can be fully verified only when a more consistent number of components observed in states 2 and 3 at longer working times will be available.

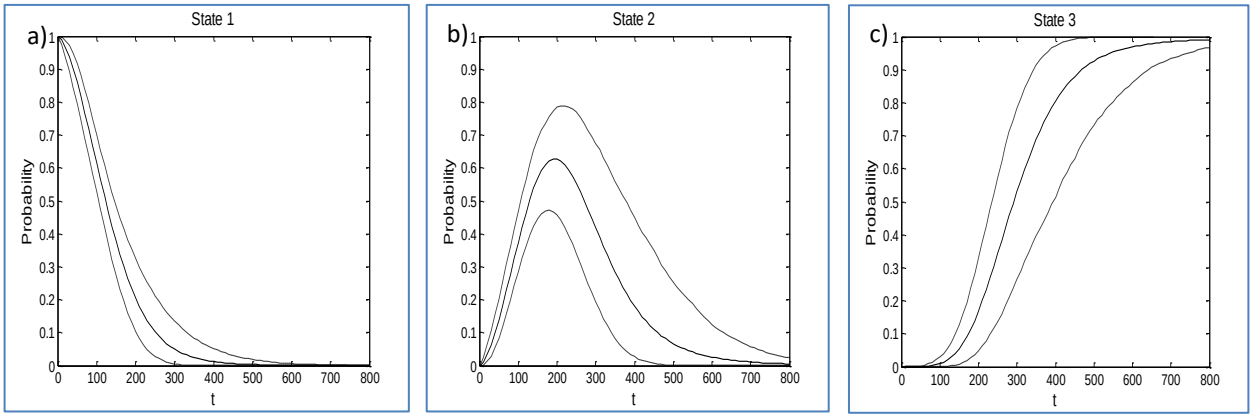



**Figure 8.** Probability that a component is in the three states over time (solid line), with the $(0.05, 0.95)$ pointwise credibility band (dashed lines)

Since the probability of being in state 3 over time coincides with the CDF of the EPDM diaphragms failure time, the predictions shown in Figure 8c are of particular interest for maintenance decision. This result can be helpful to determine the best replacement time value for the EPDM diaphragms, in order to achieve a satisfactory trade off between the minimization of the production costs and the reduction of the failure risk. With respect to the production cost, one should consider several factors such as the cost of the component replacement, which includes the cost of the diaphragm itself and of performing the technical actions, the economic loss due temporary unavailability of the system and the possibility of human errors during the repair. With respect to the failure risk, it has to be considered that a EPDM failure causes the loss of all the production and a long downtime of the entire plant for the purification and certification of the system. The identification of the optimal replacement time is typically performed by defining a cost function which properly takes into account the cost factors above listed. The defined cost function is then minimized with respect to the replacement time.

Finally, we have performed a sensitivity analysis on the hyper-parameter $m_2$, governing the prior distribution on the parameters $\alpha_2$ and $\beta_2$ of the transition time $T_2$. The objective is to evaluate the modifications of the predicted transition times as the parameter $m_2$ increases, i.e., more credit is given to prior opinions.

Figure 9 compares the expected probabilities of occupancy of the three states obtained with $m_2 = 0.5$ to those obtained with $m_2 = 3.7$ and $m_2 = 10$. In Figure 9a, we can see that the effect on the transition time

$T_1$ is very small (though not null, because $(\alpha_1, \beta_1)$ and $(\alpha_2, \beta_2)$ are a posteriori dependent), whereas a noticeable difference can be observed for transition time $T_2$ (Figure 9b and 9c). As expected, the second transition time tends to occur earlier as $m_2$ increases, since the expert predictions about $T_2$ seem to be shorter than those suggested by the data. This is particularly true for the low order quantiles of $T_2$, which are the most influenced (Figure 9c), since the expert judgement allocates probability mass on shorter transition times. The performed sensitivity analysis points out that the predictions are quite sensitive on the choice of the prior distribution for the parameters governing the second transition time. Eventually, the adoption of prudential policies might suggest to overweigh, in this case, the vague expert opinions.

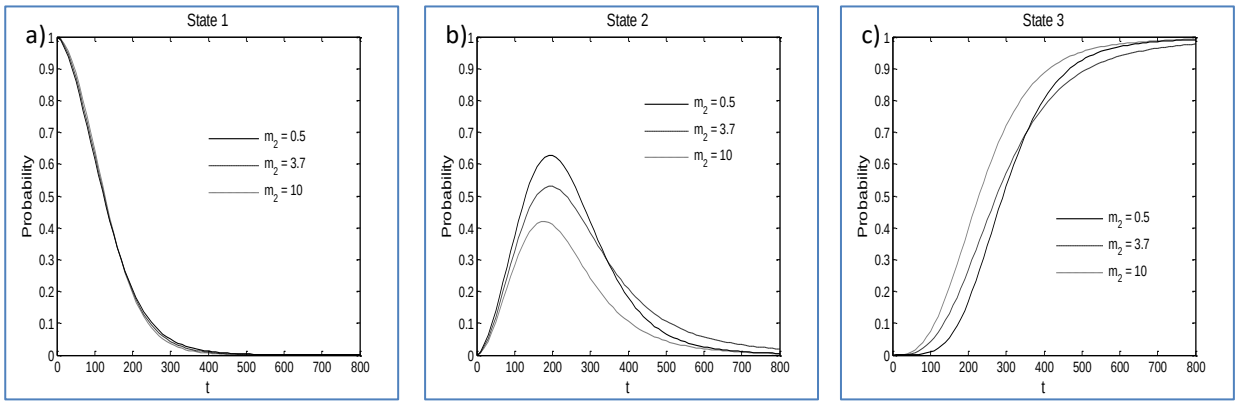



**Figure 9.** The expected probability of occupancy of the three states over time, in the three cases $m_2 = 0.5$, $m_2 = 3.7$ and $m_2 = 10$

In the remaining part of this section, we investigate the potential use of the proposed Bayesian framework in the case in which it were possile to periodically perform non destructive inspections of the component, according to Section 3.3. To this aim, we assume that an hypothetical $(N+1)^{th}$ diaphragm has been observed at time $r_{N+1} > 0$ and found in state $d_{N+1} < 3$. First of all, notice that the knowledge of the detection outcome allows updating the predictions for the failure time reported in Figure 8c, i.e., a new datum becomes available and it can be used to update the parameters' posterior distribution. Furthermore, in this case, the maintenance planner is interested in the prediction of the RUL of the specific $(N+1)^{th}$ component, which was inspected at time $r_{N+1} > 0$ and found in state $d_{N+1} < 3$. Observing the predicted RUL and the corresponding credibility band, it is possible to decide wheter and for how much more time the component can be exploited in the production line.

Figure 10 shows the obtained RUL predictions for three hypothetical components, which are found in state 2 at inspections performed at times 10, 250, 500, respectively.

It is interesting to observe that the RUL credibility intervals are not monotonically decreasing as the time $r_{N+1}$ at which the component is observed in state 2 increases. For example, the 95% quantile is larger when the inspection is performed at time 500 than when is performed at time 250. The reason is that the credibility intervals in these two cases are generated from two different posterior distributions, $\pi(\boldsymbol{\theta}|\boldsymbol{r}, \boldsymbol{d}, (500,2))$ and $\pi(\boldsymbol{\theta}|\boldsymbol{r}, \boldsymbol{d}, (250,2))$, respectively. The former is heavily influenced by the observation $(500,2)$, which corresponds to a component survived to state 3 after a very long working time: this rather 'optimistic' evidence generates predictions for the transition times that are larger than those obtained by updating with $(250,2)$. This is also confirmed by the comparison of Figures 10 and 11, which shows the predictions obtained without performing the updating, i.e., in the case the posterior distribution used for the inference is $\pi(\boldsymbol{\theta}|\boldsymbol{r}, \boldsymbol{d})$. The effect of the updating is very relevant for the case $(500,2)$, especially for high quantiles of the

credible interval. This remarks the importance of exploiting all the available information when dealing with small sample data.

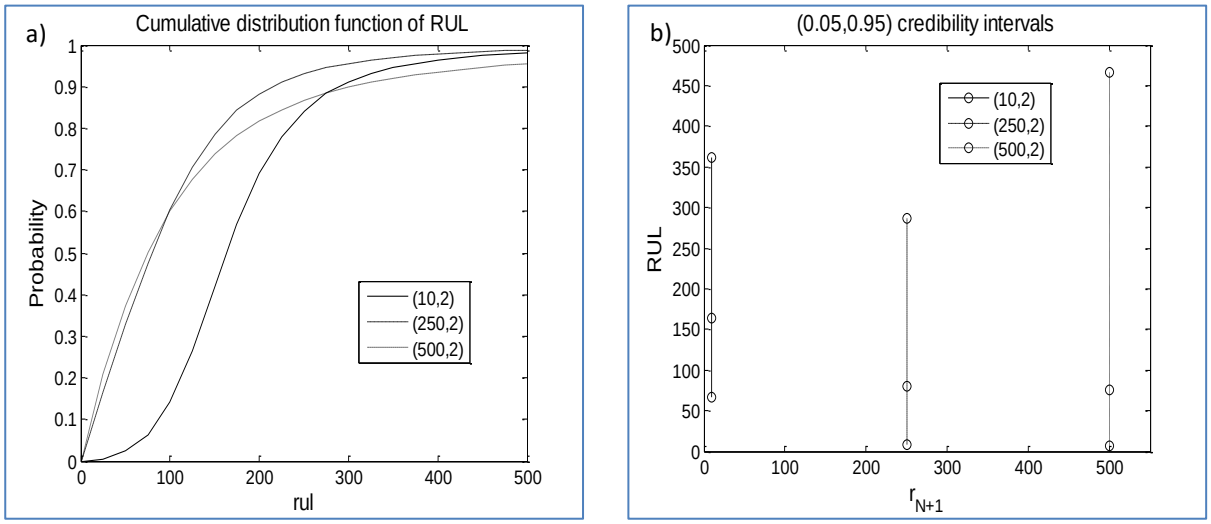



**Figure 10.** The expected CDFs of the RUL for three different components observed at different times in state 2, calculated by updating the posterior distribution (10a). The corresponding $(0.05, 0.95)$ credibility intervals are drawn in Figure 10b, with the quantiles of order 0.05, 0.50, 0.95 marked with an empty circle

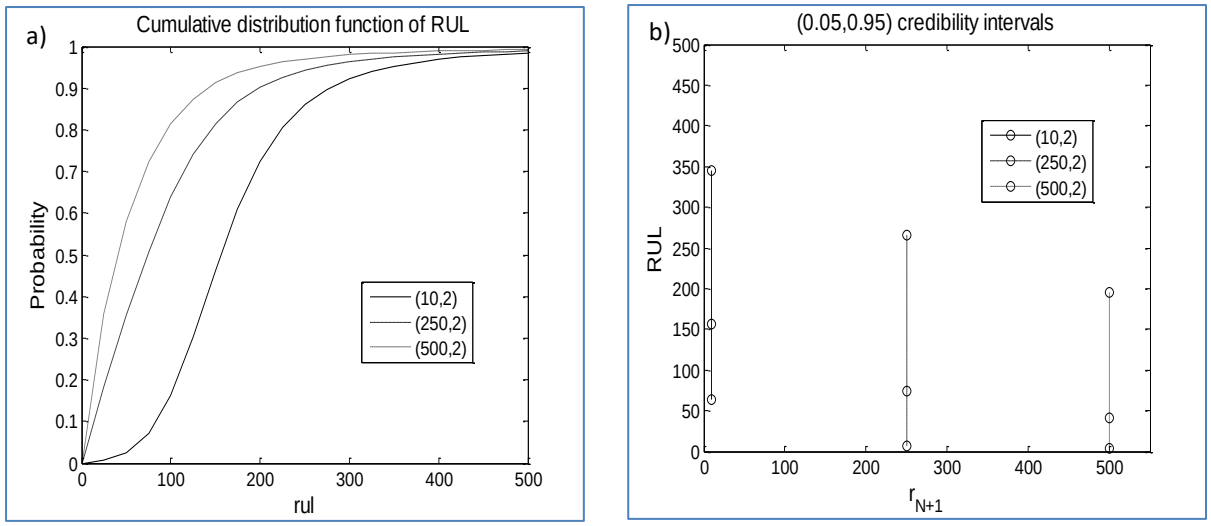



**Figure 11.** The expected CDFs of the RUL for three different components observed at different times in state 2, calculated without updating the posterior distribution (11a). The corresponding $(0.05, 0.95)$ credibility intervals are drawn in Figure 11b, with the quantiles of order 0.05, 0.50, 0.95 marked with an empty circle

To further support maintenance decision making, future research work will focus on the development of a sound maintenance cost model, which requires estimating the costs and benefits of inspecting and removing the membranes. These estimations are very difficult in the considered case study, as it pertains to a field (i.e., the biopharmaceutical industry) where there exists a variety of industry specific complications and strict regulations. For example, re-starting the production process after a shutdown due to failure takes approximately 30 days, due to the specific procedures required by the regulatory committees to guarantee the safety of the medicines. Thus, the revenue losses caused by the failure depend on several factors such as the seasonal period, the type of medicine not produced, the portion of the lot already produced, etc. which make difficult the estimation of the cost of the failure. Yet, maintenance inspections are performed in an opportunistic approach, the maintenance opportunities being to gain access to the diaphragms during a

programmed system shutdown. This entails that cost models based on periodic inspections are not applicable and more complex models need to be developed, which encode the variability of the time to produce a medicine lot.

Once the cost model is developed, we will also perform a sensitivity analysis to check whether some parameters can be set at default values without appreciable impact on optimality of the maintenance policy. This could simplify the parameter estimation problem, e.g., by showing that it would not be relevant to assume more complicated distributions than an exponential distribution (Coolen, F.P.A., Dekker,1995).

# 6 Discussion

The aim of this Section is to comment on the applicability of the proposed methodology to other case studies.

From a wide perspective, it seems fair to say that when the objective of a research investigation is to exploit all the knowledge, information and data available to solve a general issue such as the estimation of the parameters of a multi-state model, then the peculiarities of the considered case study always compromise the generality of the developed solution.

In this work, the peculiar characteristics of the case study concern i) the assumptions behind the degradation model; ii) the available dataset; iii) the information retrievable from experts through the elicitation process.

With respect to the degradation model, there are three main assumptions in this work:

1. The degradation process is a three-state process. This assumption comes from the practice of the maintenance operators working at the industrial partner, who use these three levels to qualitatively indicate the membrane health state. When the number of degradation states increases, one has to consider the additional transition times $T_i, i \geq 3$ and, correspondingly, the inspection outcomes of components found in degradation states $d$>3. To accommodate this evidence, we need to change the likelihood function in Equation (15), which depends on additional model parameters (i.e., the cardinality of $\boldsymbol{\theta}$ increases). For example, in (Compare et al., 2016), a four-state degradation process is considered and the likelihood function is derived, which depends on six parameters.
   On the other side, the additional model parameters must be encoded in the MCMC algorithm, thus leading to a rise in the computational burden, due to both the additional time needed to numerically calculate the likelihood function (see Compare et al., 2016) and the increase in the number of chains to be tracked and analyzed.
   In general, we can conclude that the larger the number of states, the more refined the description of the degradation process, the more complex the issue of estimating the model parameters.
2. The degradation process is homogeneous (Moghaddas & Zuo, 2015). That is, we assume that the sojourn time in any state does not impact on those in the other states. In this work, the assumption of homogeneous semi-Markov process is justified by the fact that the membrane degradation has been considered an additive process, where the magnitude of the degradation increments does not depend on time but, rather, on the number of cycles the membrane is exposed to. In general, however, the independence assumption of the sojourn times is questionable and, if not applicable, one has to change the likelihood function to encode transition rates that depend on both the sojourn time in the current state and the total time of the component (e.g., Moghaddas & Zuo, 2015). In this respect, finding a suitable model for the transition rates is not straightforward.
3. The degradation process is left-to-right. This assumption also derives from the specific characteristics of the considered case study. Namely, the membranes have always been

replaced upon inspection; then, transitions from right to left have never been experienced by the membranes. Moreover, the cumulative damage nature of the degradation process makes meaningless the direct jumps from one state to the states after the next degraded state: abrupt degradation is modeled as a rapid passing through the states, rather than by jumps. In general, one could consider multi-state systems with more complex structures (e.g., Fleming & Smit, 2008; Baraldi et al. 2013, Lin et al., 2015). In this case, the considerations above apply, regarding the consequence on the computational burden of increasing the complexity of the likelihood function and the number of parameters.

With respect to the inspection outcomes dataset, evidently the available evidence determines the structure of the likelihood function, which is just a representation of the probability of gathering that evidence. Then, different types of datasets lead to considering different types of likelihood functions. For example, in (Baraldi et al. 2014b), the likelihood function is derived for a four-state left-to-right degradation process, in which maintenance inspections of different refinement levels (i.e., requiring dis-assembling or not) can detect the component in different states.

In regards to the elicitation process, the procedure illustrated in Section 4 has been chosen based on a series of preliminary interviews with the experts, in which their ability in providing statements like those in Section 4 has been checked. The choice of this procedure is also motivated by the findings of (Teigen & Jorgensen, 2005), which show that the elicitation biases reduce when experts assign both the quantiles $s_i$ and the corresponding levels $q_i$ x 100%; on the contrary, when experts are asked to assign the $s_i$ values corresponding to a pre-fixed level $q_i$ x 100%, the correctness of the judgment worsens.

Obviously, if the elicited information changes, then the procedure adopted in this work can no longer be applied and a new combination of techniques is required to be developed to respond to the specific type of gathered information. This confirms that every case study can be considered as a unique issue, which deserves its own tailored solution.

Finally, there will always be some uncertainty in the choice of priors, especially in case the information from which these distributions are constructed is poor, or when there are several priors elicited, say, from different experts. It is of interest, then, to characterize the sensitivity of the posterior distribution to priors. To do this, several techniques have been developed (e.g., Hill & Spal, 1994; Lopes, H.F., Tobias, 2011;), which can be applied depending on the characteristics of the case study. In this work, we have developed a sensitivity analysis for the second transition time, only, where the scarcity of data and the lack of confidence on the provided information led us to choose uninformative priors and perform a sensitivity analysis on the hyper-parameter $m_2$ governing the prior distributions of parameters $\alpha_2$ and $\beta_2$. Again, if change the elicitation framework, this analysis on a single parameter can no longer be done.

# 7 Conclusions

In this paper, we have proposed a complete procedure for assessing the reliability of industrial equipment whose degradation process can be described through a three-state semi-Markov model. The method has been set within a Bayesian framework, that allows updating the prior distribution elicited from an expert with the available field data. In details, a Metropolis-Hastings algorithm, combined with numerical integration methods, has been proposed to sample from the complex posterior distribution, where the choice of the covariance matrix of the normal proposal density is automatically optimized in a preliminary

run of an adaptive MCMC algorithm. Then, the posterior samples allow the estimation of the probabilities of occupying the different degradation states over time, the component remaining useful time and the corresponding uncertainty.

One of the advantages in the use of multi-state models, compared to traditional binary models, lies in the possibility of making estimates about the RUL of a component on the basis of its degradation conditions. This makes the procedure usable in support of the planning of periodic maintenance actions, and for condition based and predictive Maintenance.

Besides, the application of the method to a real case study has shown the importance of iterating the posterior distribution updating each time a new observation is available, especially in situations of scarcity of data.

Future works will focus on the development of a Proportional Hazards model which takes into account the effects on the degradation process of the operating conditions in which the components are employed.

# 8 References


Andrieu, C., Thoms, J., "A tutorial on adaptive MCMC", Statistics and Computing, 18, pp. 343-373, 2008.

Baraldi, P., Compare, M., Zio, E., "Maintenance policy performance assessment in presence of imprecision based on Dempster–Shafer Theory of Evidence", Information Sciences. Vol. 245, pp. 112-131, 2013a.

Baraldi, P., Balestrero, A., Compare, M., Benetrix, L., Despujols, A., Zio, E. "A modeling framework for maintenance optimization of electrical components based on fuzzy logic and effective age", Quality and Reliability Engineering International, Vol. 29, No. 3, pp. 385-405, 2013b.

Baraldi, P., Compare, M., Despujols, A., Lair, W., Zio, E. "A practical analysis of the degradation of a nuclear component with field data", Safety, Reliability and Risk Analysis: Beyond the Horizon - Proceedings of the European Safety and Reliability Conference, ESREL 2013, pp. 1009-1014, 2014a.

Baraldi, P., Compare, M., Despujols, A., Zio, E., "Modelling the effects of maintenance on the degradation of a water-feeding turbo-pump of a nuclear power plant", Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, Vol. 225, No. 2, pp. 169-183, 2011.

Baraldi, P., Compare, M., Zio, E. "Uncertainty treatment in expert information systems for maintenance policy assessment", Applied Soft Computing Journal, Vol. 22, pp. 297-310, 2014b.

Baraldi, P., Compare, M., Zio, E., "Uncertainty analysis in degradation modeling for maintenance policy assessment", Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, Vol. 227, No. 3, pp. 267-278, 2013c.

Bousquet, N., "Elicitation of Weibull priors," http://arxiv.org/abs/1007.4740v2, 2010.

Brent, R., "Algorithms for minimization without derivatives", Prentice-Hall Englewood Cliffs N.J., 1973.

Christer A.H., Wang W. "A delay time based maintenance model of multi-component system," IIMA Journal of Management Mathematics, Vol. 6, pp. 205-222, 1995.

Christer, A.H., Wang W., Sharp J., and Baker R. "A Case Study of Modelling Preventive Maintenance of a Production Plant Using Subjective Data." The Journal of the Operational Research Society , Vol. 49, No. 3, pp. 210-19, 1998.

Christensen, R., Johnson, W.O., Branscum, A.J., Hanson, T.E., “Bayesian ideas and data analysis: An introduction for scientists and statisticians”, Boca Raton, FL: CRC Press., 2010.

Compare, M., Martini, F., Mattafirri, S. Carlevaro, F., Zio, E., “Semi-Markov model for the oxidation degradation mechanism in gas turbine nozzle”, IEEE Reliability Transactions, Vol. 99, pp. 1-8, 2016.

Coolen, F.P.A, “On Bayesian reliability analysis with informative priors and censoring”, Reliability Engineering and System Safety, Vol. 53, No. 1, pp. 91-98, 1996

Coolen, F.P.A., Dekker, R. “Analysis of a 2-Phase Model for Optimization of Condition-Monitoring Intervals”, IEEE Transactions On Reliability, Vol. 44, No. 3, pp. 505-511, 1995.

Cowles, M.K., Carlin, B.P., “Markov Chain Monte Carlo convergence diagnostics: a comparative review”, Journal of the American Statistical Association, Vol. 91, pp. 883-904, 1996.

Epifani, I., Ladelli, L., Pietavolo, A., “Bayesian estimation for a parametric Markov Renewal model applied to seismic data”, Electronic Journal of Statistics, Vol. 8, pp. 2264-2295, 2014.

Fleming, K.N., Smit, K. “Evaluation of design, leak monitoring, and nde strategies to assure pbmr helium pressure boundary reliability”, Proceedings of the 4th International Topical Meeting on High Temperature Reactor Technology, Vol. 2, pp. 135-142, 2008.

Garthwaite, P.H., Kadane, J.B., O’Hagan, A., “Statistical methods for eliciting prior distributions”, Journal of the Americal Statistical Association, Vol. 100, pp. 680-701, 2005.

Ge, H., Asgarpoor, S., “Reliability evaluation of equipment and substations with fuzzy Markov processes”, IEEE Transactions on Power Systems, Vol. 25, No. 3, pp. 1319-1328, 2010.

Gelman, A., “ Prior distribution”, Encyclopedia Environmetrics, vol. 3, pp.1634-1637. John Wiley & Sons, Chichester, 2002.

Gelman, A., Roberts, G.O., Gilks, W.R., “Efficient Metropolis jumping rules”, Bayesian Statistics, Oxford University Press, Vol. 5, pp. 599-607, 1996.

Geweke, J., “Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments”, in J. Bernardo, J. Berger, A. Dawid and F.Smith, eds., Bayesian Statistics, Oxford University Press, New York, pp. 169-193, 1992.

Gilks, W.R., Richardson, S., Spiegelharter, D.J., “Markov Chain Monte Carlo in practice”, Chapman and Hall, London, UK, 1996.

Giorgio, M., Guida, M., Pulcini, G., “An age- and state-dependent Markov model for degradation processes”, IIE Transactions, Vol. 43, No. 9, pp. 621-632, 2011.

Haario, H., Saksman, E., Tamminen, J., “An adaptive Metropolis algorithm”, Bernoulli 7(2), pp. 223-242, 2001.

Hill, S.D. Spal, J.C. “Sensitivity of a Bayesian Analysis to the Prior Distribution”, IEEE Transactions On Systems, Man, And Cybernetics, Vol. 24, No. 2, pp. 216-221, 1994.

Jackman, S., “Bayesian analysis for the social sciences”, Wiley, New York, 2009.

Jenkinson, D., “The Elicitation of Probabilities – A Review of the statistical literature”, BEEP working paper, University of Sheffield, 2005.

Kass, R.E., Wasserman, L., “The selection of prior distributions by formal rules”, Journal of the American Statistical Association, Vol. 91, pp. 1343-1370, 1996.

Laird, A., “Modeling a Progressive Disease Process Under Panel Observation”, Ph. D. thesis, University of Washington, 2013.

Levitin, G., Lisnianski, A., Ushakov I., “Reliability of Multi-State Systems: A Historical Overview”, Mathematical and statistical methods in reliability, Lindqvist, Doksum (eds.), World Scientific, pp. 123-137, 2003.

Limnios, N, Oprisan, G., “Semi-Markov Processes and Reliability”, Birkhauser, Boston, 2001.

Lin, Y.-H., Li, Y.-F., Zio, E. “Integrating random shocks into multi-state physics models of degradation processes for component reliability assessment”, IEEE Transactions on Reliability, Vol. 64, No. 1, pp. 154-166, 2015.

Link, W.A, Eaton, M.J., “On thinning of chains in MCMC”, Methods in Ecology and Evolution, Vol. 3, No.1, pp. 112-115, 2012.

Lisnianski, A., Levitin, G., “Multi-State System Reliability: Assessment, Optimization and Applications”, Series on Quality, Reliability and Engineering Statistics: Volume 6, World Scientific, 2003.

Liu, Y., Lin,P., Li, Y.F., Huang, H.Z., “Bayesian Reliability and Performance Assessment for Multi-State Systems”, in Reliability, IEEE Transactions on, Vol. 64, No. 1, pp. 394-409, March 2015.

Lopes, H.F., Tobias, J.L. “Confronting prior convictions: On issues of prior sensitivity and likelihood robustness in Bayesian analysis”, Annual Review of Economics, Vol. 3, pp. 107-131, 2011.

MacEachern, S.N., Berliner, L.M., “Subsampling the Gibbs sampler”, The American Statistician, Vol. 48, pp. 188–190, 1994.

Moghaddass, R., Zuo, M.J., “Condition-based Replacement Policy for a Device Using Interval-Censored inspection data”, Reliability and Maintainability Symposium (RAMS), 2013, Proceedings-Annual, pp. 1-6.

Moghaddass, R., Zuo, M.J., Liu, Y., Huang, H.Z., “Predictive analytics using a nonhomogeneous semi-Markov model and inspection data”, IIE Transactions, 47, no. 5, pp. 505-520, 2015.

Oakley, J., O’Hagan, A., “Probabilistic sensitivity analysis of complex models: a Bayesian approach”, Journal of the Royal Statistical Society, Series B, Vol 66, No. 3, 751-769, 2004.

Raftery, A. L. and Lewis, S., “How many iterations in the Gibbs sampler?” In Bayesian Statistics 4, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 763-74. Oxford University Press, 1992.

Robert, C.P., Casella, G., “Monte Carlo statistical methods”, New York: Springer, 2004.

Roos, M., Martins, T.G., Held, L., Rue, H. “Sensitivity analysis for Bayesian hierarchical models”, arXiv:1312.4797, 2013

Ruiz-Castro, J.E.,“Preventive maintenance of a multi-state device subject to internal failure and damage due to external shocks”, IEEE Transactions on Reliability, Vol. 63, No. 2, pp. 646-660, 2014.

Singpurwalla, N.D., Song, M.S., “Reliability Analysis using Weibull Lifetime Data and Expert Opinion”, IEEE Transactions on Reliability, Vol. 37, No. 3, pp. 340-347, 1988.

Teigen, K.H., Jorgensen, M. When 90% confidence intervals are only 50% certain: On credibility of credible intervals. Applied Cognitive Psychology. Vol. 19, pp. 455–475, 2005.

Thiebaux, H.J., Zwiers, F.W., “The Interpretation and Estimation of Effective Sample Size”, Journal of Applied Meteorology, Vol. 23, pp. 800-811, 1984.

Zio, E., “An introduction to the basics of reliability and risk analysis”, Singapore: World Scientific Publishing Company, 2007.

Zio, E., “The Monte Carlo Simulation Method for System Reliability and Risk Analysis”, Springer Series in Reliability Engineering, 2013.

Zio, E. “Challenges and opportunities in reliability engineering: the big KID (Knowledge, Information and Data)”, IEEE, Transactions on Reliability, In press (2016).

Zuo, M. J., Huang, J., Kuo, W., "Multi-state k-out-of-n Systems," Reliability Engineering Handbook, (H. Pham, ed.), Springer-Verlag, London, September 2001.

# 9 Appendix

## 9.1 MCMC diagnostics

Without loss of generality, let $\theta$ denote one of the parameters in the vector $\boldsymbol{\theta}$.

- Traceplot (e.g., Jackman, 2009)

  It is the plot of the time series $\left(i, \theta^{(i)}\right)_{i=1,\dots,G}$ of the states visited by the chain. It provides a pictorial view of the needed burn-in length (Figure 12). Moreover, a traceplot also gives a snapshot of the chain mixing (i.e., how quickly it explores the support of the target distribution), which is strictly related to the sampling efficiency. Besides, the thickness of a traceplot is a rough visual criterion to check if the number of executed iterations is large enough to obtain a significant sample from the posterior density: a little thick traceplot alerts about the possibility of obtaining a not very accurate sample and suggests to lengthen the drawn MCMC chain.

- Autocorrelation plot (e.g., Jackman, 2009)

An autocorrelation plot is the graph $(l, \hat{\rho}_l)_{l=0,\dots,L}$ of the empirical autocorrelations $\hat{\rho}_l = \text{cor}\left(\left(\theta^{(1)}, \dots, \theta^{(G-l)}\right), \left(\theta^{(1+l)}, \dots, \theta^{(G)}\right)\right)$ of the process at lag $l, l = 0, \dots, L$, where usually $L$ is set to about 20, and, obviously, $\rho_0 = 1$. The efficiency of the MCMC algorithms can be diagnosed by observing the values $\hat{\rho}_l \; \forall l = 1, \dots, L$, especially for small $l$: the more such values are close to 0, the larger the efficiency of the algorithm (in the limit case, if our sampling were i.i.d., we would obtain $\hat{\rho}_l \approx 0 \; \forall l = 1, \dots, L$). The practice of thinning the chain in order to reduce autocorrelation is not suggested, unless it is necessary due to computer memory constraints. Indeed, the discard of some MCMC samples has the effect of worsening the posterior estimates. (MacEachern & Berliner, 1994, Link & Eaton, 2012).

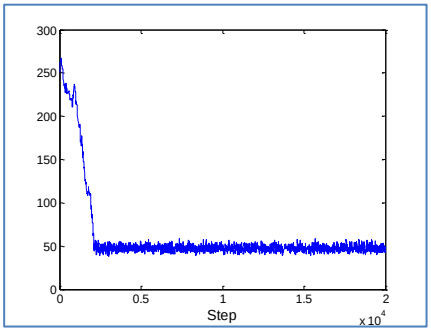



**Figure 12.** *Unlucky choices for the initial state $\theta^{(0)}$ can require many burn-in iterations. Here, a burn-in of at least 2500 iterations is needed*

- Effective Sample Size (e.g., Jackman, 2009)

The $G$ draws $\theta^{(1)}, \dots, \theta^{(G)}$ from $\pi(\theta|\boldsymbol{r}, \boldsymbol{d})$ through a MCMC sampler are not independent. Thus, the amount of information carried by such a sample is equivalent, actually, to that brought by a virtual i.i.d. sample from $\pi(\theta|\boldsymbol{r}, \boldsymbol{d})$ of a smaller size $G^* < G$. $G^*$ is called Effective Sample Size (ESS). In particular, the larger the exploration efficiency of the algorithm, the larger the ESS. The following considerations lead to a possible quantitative definition of the ESS (e.g., Thiebaux & Zwiers, 1984, Jackman, 2009). Let $\text{Var}(\bar{\theta}_G)$ be the variance of the sample mean $\bar{\theta}_G$ and, under the assumption that $\theta^{(1)}, \dots, \theta^{(G)}$ are all identically distributed as $\pi(\theta|\boldsymbol{r}, \boldsymbol{d})$ (i.e., sampled from a chain that has already reached convergence), let $\sigma^2 = \text{Var}\left(\theta^{(g)}\right)$ be the variance of each draw $\theta^{(g)}$. If $\theta^{(1)}, \dots, \theta^{(G)}$ were $G$ i.i.d. samples, we would obtain the well-known identity (Zio, 2013):

$$\text{Var}(\bar{\theta}_G) = \frac{\sigma^2}{G} \; . \tag{28}$$

Taking into account the autocorrelation introduced by MCMC methods, it can be shown (Jackman, 2009) that, for large $G$:

$$\text{Var}(\bar{\theta}_G) \approx \frac{\sigma^2}{G/(1 + 2\sum_{l=1}^{\infty} \rho_l)} \tag{29}$$

So, the Effective Sample Size can be defined as (Jackman, 2009):

$$G^* = \frac{G}{1 + 2\sum_{l=1}^{\infty} \rho_l} \tag{30}$$

and, substituting Eq. $(30)$ in Eq. $(29)$ it follows that:

$$\mathrm{Var}(\bar{\theta}_G) \approx \frac{\sigma^2}{G^*} \tag{31}$$

The direct application of equation $(30)$ is not possible; in fact, a chain of $G$ observations allows the estimation of the autocorrelation up to lag $G - 2$ only, instead of infinite. Moreover the empirical estimate $\hat{\rho}_l$ for $\rho_l$ becomes more uncertain with larger $l$.
According to (Thiebaux & Zwiers, 1984, Jackman, 2009), one can overcome this problem by setting:

$$\mathrm{Var}(\bar{\theta}_G) \approx \frac{S_0}{G} \tag{32}$$

where $S_0$ is the spectral density at frequency 0 of the employed chain. Eqs. $(31)$ and $(32)$ imply that one can estimate the ESS with:

$$G^* \approx G\frac{\widehat{\sigma^2}}{\widehat{S_0}} \tag{33}$$

being $\widehat{\sigma^2}$ the empirical variance of the sampled chain $\theta^{(1)}, \dots, \theta^{(G)}$ and $\widehat{S_0}$ its empirical spectral density at frequency 0. This latter may be obtained by fitting the sampled chain with an Auto-Regressive (AR) process (Jackman, 2009).

- Geweke test (Geweke, 1992)

In (Geweke, 1992), the null hypothesis that the mean $\bar{\theta}_A$ of the first $10\%$ of draws is equal to the mean $\bar{\theta}_B$ of the last $50\%$ (assuming that this last $50\%$ has already reached convergence) is tested to assess the convergence of the Markov chain. Thus, $G_A = \frac{G}{10}$ and $G_B = \frac{G}{2}$ are the cardinalities of the two considered tranches, whereas $S_0^A$ and $S_0^B$ are the estimators of the spectral density at frequency $0$ of the first $10\%$ and the last $50\%$, respectively, obtained by means of the fitting of an Auto-Regressive (AR) process. Under the null hypothesis, it holds:

$$\frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{G_A^{-1}S_0^A + G_B^{-1}S_0^B}} \sim AN(0,1) \tag{34}$$

where $AN(0,1)$ is the asymptotic standard normal distribution. So, if the test gives evidence against the null hypothesis $\bar{\theta}_A = \bar{\theta}_B$, then we can conclude that the burn-in length must be increased.

Other MCMC diagnostics methods (e.g., Gelman-Rubin, Heidelberger-Welch, Raftery-Lewis) can be found in (Cowles & Carlin, 1996).

## 9.2 Proof of Eq. (27)

Notice that $\forall\, t \geq r_{N+1} > 0$:

$$P(T_1 + T_2 \leq t|\boldsymbol{r}, \boldsymbol{d}, (r_{N+1}, d_{N+1} = 2), \boldsymbol{\theta}) = P(T_1 + T_2 \leq t|(r_{N+1}, d_{N+1} = 2), \boldsymbol{\theta}) =$$

$$= P(T_1 + T_2 \leq t|T_1 \leq r_{N+1}, T_1 + T_2 > r_{N+1}, \boldsymbol{\theta}) = \frac{P(r_{N+1} < T_1 + T_2 \leq t, T_1 \leq r_{N+1}|\boldsymbol{\theta})}{P(T_1 \leq r_{N+1}, T_1 + T_2 > r_{N+1}|\boldsymbol{\theta})}$$

where the first identity holds because we make the hypothesis, similarly to what we assumed in the Bayesian model presented in (1), that the two transition times of the new component are conditional independent of the transition times of the components in the dataset, given $\boldsymbol{\theta}$.

We now focus on computing the fraction. Its numerator can be calculated by integrating the joint PDF $f_{T_1,T_2|\boldsymbol{\theta}}(t_1, t_2|\boldsymbol{\theta})$ over the domain shown in Figure 13a:

$$P(r_{N+1} < T_1 + T_2 \leq t, T_1 \leq r_{N+1}|\boldsymbol{\theta}) = \int_0^{r_{N+1}} \int_{r_{N+1}-t_1}^{t-t_1} f_{T_1,T_2|\boldsymbol{\theta}}(t_1, t_2|\boldsymbol{\theta}) dt_2 dt_1 =$$

$$= \int_0^{r_{N+1}} \left( \int_{r_{N+1}-t_1}^{t-t_1} f_{T_2|\boldsymbol{\theta}}(t_2|\boldsymbol{\theta}) dt_2 \right) f_{T_1|\boldsymbol{\theta}}(t_1|\boldsymbol{\theta}) dt_1 = \int_0^{r_{N+1}} \left[ F_{T_2}(t - t_1|\boldsymbol{\theta}) - F_{T_2}(r_{N+1} - t_1|\boldsymbol{\theta}) \right] f_{T_1}(t_1|\boldsymbol{\theta}) dt_1,$$

where we exploited the fact that $f_{T_1,T_2|\boldsymbol{\theta}}(t_1, t_2|\boldsymbol{\theta}) = f_{T_1|\boldsymbol{\theta}}(t_1|\boldsymbol{\theta}) f_{T_2|\boldsymbol{\theta}}(t_2|\boldsymbol{\theta})$ due to the conditional independence of the two transition times of the new component, given $\boldsymbol{\theta}$.

While, to obtain the denominator, the integration must be performed on the domain illustrated in Figure 13b:

$$P(T_1 \leq r_{N+1}, T_1 + T_2 > r_{N+1}|\boldsymbol{\theta}) = \int_0^{r_{N+1}} \int_{r_{N+1}-t_1}^{+\infty} f_{T_1,T_2|\boldsymbol{\theta}}(t_1, t_2|\boldsymbol{\theta}) dt_2 dt_1 =$$

$$= \int_0^{r_{N+1}} \left( \int_{r_{N+1}-t_1}^{+\infty} f_{T_2|\boldsymbol{\theta}}(t_2|\boldsymbol{\theta}) dt_2 \right) f_{T_1|\boldsymbol{\theta}}(t_1|\boldsymbol{\theta}) dt_1 = \int_0^{r_{N+1}} \left[ 1 - F_{T_2}(r_{N+1} - t_1|\boldsymbol{\theta}) \right] f_{T_1}(t_1|\boldsymbol{\theta}) dt_1,$$

from which it follows Eq. (27).

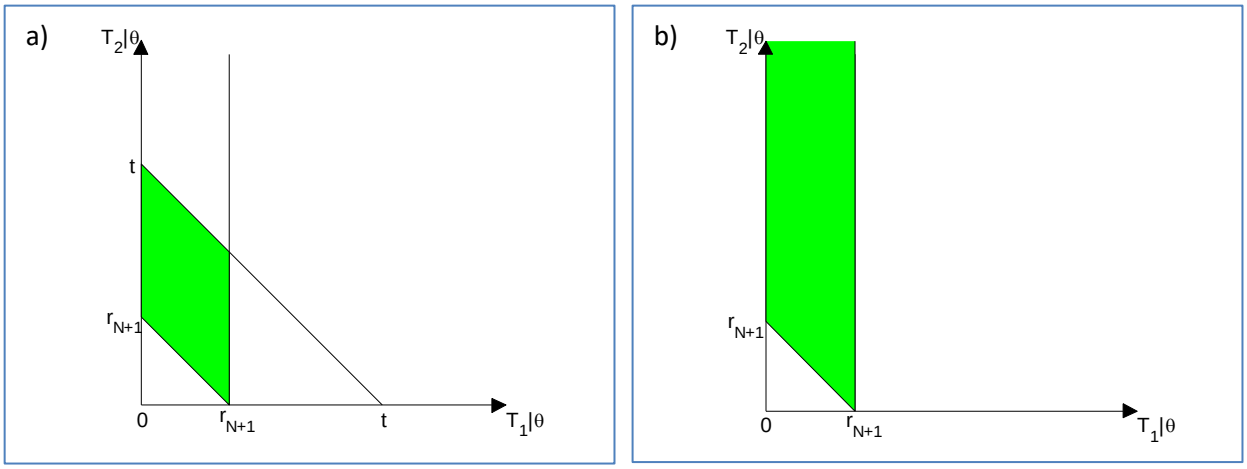



**Figure 13.** The integration domains used in the computations