Industrial Equipment Reliability Estimation: a Bayesian Weibull Regression Model with Covariate Selection

M. Compare[1,2], P. Baraldi[1], I. Bani[1], E. Zio[1,2,3,4], D. McDonnell[5]

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

[2]*Aramis Srl, Milano, Italy*

[3]*MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France*

[4]*Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea*

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

Abstract

A three-state continuous-time semi-Markov process is used to model the degradation of an industrial equipment. The transition times are assumed Weibull-distributed and influenced by a set of covariates. A Weibull Regression Model is developed within the Bayesian probability framework, to account for the influence of these covariates and estimate the model parameters with the related uncertainty, on the basis of few data and expert judgment. The number of covariates is reduced by a two-step selection procedure derived from the condition monitoring engineering practice. The developed model enables estimating reliability and time-dependent state probabilities for a component degrading in given operational and ambient conditions, represented by a vector of covariates. The model is illustrated by way of a real case study concerning the degradation process affecting diaphragm valves used in the biopharmaceutical industry.

Key words: Multi-state degradation modelling, Weibull Regressions model, Variable selection, Bayesian inference, MCMC algorithms

1 Introduction

Maintenance operators are familiar with the practice of classifying the component degradation into multiple discrete states, based on their mapping with suitable degradation indicator variables (e.g., [1], [2], [3], [4]). This justifies the interest in developing multi-state degradation models, whose parameters can be estimated from the field data gathered by the maintenance operators. These models allow tracking the evolution of the component degradation processes through progressive states of deterioration better than the traditional binary-state reliability models, where only two states ('good' vs. 'failed') are considered.

A difficulty to be considered is that the operational and ambient conditions (e.g., physical stresses, workload levels, environmental conditions, etc.) can influence the evolution of the degradation processes. Thus, the multi-state degradation modeling approaches need to encode the influencing factors. To this aim, different approaches have been proposed in the literature (e.g., [5], [6], [7], [8]), including the Proportional Hazard Models (PHMs, [9], [10]). These are semi-parametric models considering a baseline hazard function to describe the evolution of the degradation process in normal conditions and, then, introducing multiplicative

factors to consider the effects of covariates. We use the Weibull distribution as baseline distribution, which turns the PHM into a parametric model. This is referred to as Weibull Regression Model (WRM, [11]).

Although parametric models need less data than semi-parametric models for parameters estimation, the development of a WRM may, however, require a non-negligible amount of data: the larger the number of covariates, the larger the number of parameters appearing in the WRM model and, thus, the larger the amount of data required to tune the model ([11]). The aim of this work is to use the Bayesian probability framework to extend the applicability of WRM to situations where data are relatively scarce and, then, expert opinion becomes a valuable source of information for developing the degradation model. These situations are quite common in industrial practice (e.g., innovative and/or very high reliable components, with new technology, etc.). Nonetheless only a few sound methodological framework have been proposed for reliability modelling, in quite specific conditions (e.g.; [2], [12], [13]).

To reduce the number of parameters in the Bayesian WRM, several approaches can be applied (e.g., [14], [15]). According to the condition monitoring engineering practice (e.g., [16], [17], [18]), we develop a two-step covariate selection approach. First, a preliminary, unrefined selection is performed based on the ordinal logistic regression technique [19]. The set of selected covariates is, then, further reduced adopting the Log Pseudo-Marginal Likelihood (LPML) technique [20]. Finally, the WRM with the selected covariates and with the posterior distribution of its parameters estimated by an adaptive Markov Chain Monte Carlo (MCMC) algorithm, is used to compute the expected probabilities of occupying the degradation states over time for new components along with pointwise credibility intervals. This is a useful information to plan maintenance inspections.

The proposed methodological framework builds on that illustrated in [21], where a Bayesian three-state semi-Markovian model is developed to describe the component degradation process. Driven by the available knowledge base, the assumption of the model in [21] is that the components work under the same operational conditions.

In this work, we use the results of the procedure for the elicitation of the Weibull parameters proposed in [21]. Yet, the application example investigated in this work to show the methodology is the same of that proposed in [21], concerning the degradation of the Ethylene Propylene Diene Monomer (EPDM, i.e., a rubber diaphragm employed in the production line of a world leading company in that field), although it is here analyzed based on a different knowledge base, in which a set of degradation process covariates is available. This enhancement constitutes the contribution of the present work: a full-fledged development framework is proposed for building a semi-Markov degradation model based on expert knowledge and few field data, including covariates candidate to enter the model, from which the most influential have to be selected.

The remainder of the paper is organized as follows. Section 2 outlines the assumptions at the basis of the developed model, which are further formalized in Section 3. Section 4 describes the two-step covariates selection method. In Section 5, the developed theoretical framework is finally applied to the EDPM diaphragm case study to *i)* select the covariates that most influence the degradation process, and *ii)* to derive reliability estimates of these components in different operational conditions. Final remarks are drawn in Section 5.

2 Problem Setting

The problem setting outlined in this Section conceptualizes the characteristics of the case study considered in this paper, concerning EDPM diaphragms used in the biopharmaceutical industry. We consider a three-state degradation 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 a cumulative damage, and that the degradation process is influenced by a set of Φ covariates. Multiple identical components of the same type and taken from the same production lot are operated in the plant, in different locations and under different covariate settings. Destructive inspections are performed on the components when inspected, which are then replaced by new components.

The objective of the work is to develop a degradation model of the individual component. The information available is made by field data collected from all the plant and expert opinions. With respect to field data, a dataset $\left(\boldsymbol{r} \in \mathbb{R}^N, \boldsymbol{d} \in \mathbb{R}^N, \boldsymbol{X}' \in \mathbb{R}^{N\times\Phi}\right) = \{(r_n, d_n, \boldsymbol{x}'_n)\}_{n=1,\dots,N}$ is available, containing N observations. Given that components are always replaced upon the destructive inspection, every record of the dataset refers to a different component and, thus, the dataset refers to *N* components. The generic n-th record reports the degradation state d_n in which the n-th component was found at the replacement time r_n (i.e., the time interval between installation and replacement) and the corresponding vector $\boldsymbol{x}'_n = (x'_{n1}, \dots, x'_{n\Phi}) \in \mathbb{R}^{\Phi}$ of the values of the Φ covariates initially available.

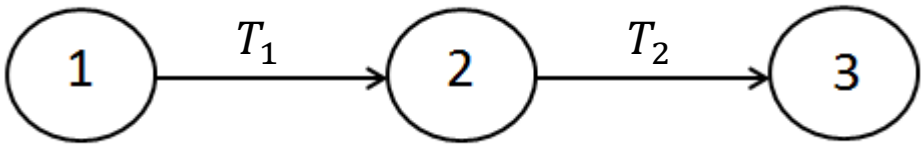


Figure 1. *Three-state model of the degradation process*

The expert opinion about the degradation process concern the transition times between states 1 and 2 and states 2 and 3, whereas the experts give no information about either the effect of the covariates on the degradation process, or their relative importance in influencing the transition times.

3 Model development

As baseline model (i.e., the model describing the behavior of a component associated to a null vector of covariates), we assume that the first transition time obeys a Weibull distribution (e.g., [3], [21], [22]) conditional on the random parameters α_1 (scale) and β_1 (shape), and the second transition time T_2 to be Weibull distributed conditional on the correspondent random parameters α_2 and β_2, viz.:

$$\begin{cases} T_1|(\alpha_1, \beta_1, \boldsymbol{x} = \boldsymbol{0}) \sim Weibull(\alpha_1, \beta_1) \\ T_2|(\alpha_2, \beta_2, \boldsymbol{x} = \boldsymbol{0}) \sim Weibull(\alpha_2, \beta_2) \end{cases} \quad (1)$$

where $\boldsymbol{x} \in \mathbb{R}^p$ is the value of the vector of the p covariates identified as the influencing factors out of the available Φ (i.e., the covariates selected from the two-step procedure described in Section 3) to be encoded

in the analysis. For clarity, we explicitly mention that we develop the Bayesian degradation model assuming that the selection of the most influential covariates has already been performed.

The stochastic transition times from states 1 to 2 and from states 2 to 3 of the n^{th} component, $n = 1, \dots, N$, are considered as independent random variables conditional on parameters $\alpha_1, \beta_1, \alpha_2, \beta_2$ and the vector of covariates $\boldsymbol{x}$. Obviously, if the n^{th} component is found in degradation state $d_n = 1$, then it has not experienced the first nor the second transitions. This 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 the random replacement time, r_n, is conditional independent from both the transition times, given $\boldsymbol{x}_n$, and that the parameters governing the distribution of r_n are different from those governing the distribution of T_1 and T_2. This ensures that the censoring mechanism is independent and non-informative with respect to the degradation process.

The WRM formulation in [9], which is here adopted for both the transition times T_1 and T_2 of the N considered components, reads:

$$\begin{cases} h(t_1|\alpha_1, \beta_1, \boldsymbol{\gamma}, \boldsymbol{x}) = e^{\boldsymbol{x}\boldsymbol{\gamma}} h(t_1|\alpha_1, \beta_1, \boldsymbol{x} = \boldsymbol{0}) \\ h(t_2|\alpha_2, \beta_2, \boldsymbol{\gamma}, \boldsymbol{x}) = e^{\boldsymbol{x}\boldsymbol{\gamma}} h(t_2|\alpha_2, \beta_2, \boldsymbol{x} = \boldsymbol{0}) \end{cases} \tag{2}$$

where $h\left(t_j|\alpha_j, \beta_j, \boldsymbol{x} = \boldsymbol{0}\right) = \frac{\beta_j}{\alpha_j}\left(\frac{t_j}{\alpha_j}\right)^{\beta_j - 1}$ is the baseline hazard function of the Weibull distribution of transition time T_j, $j = 1,2$, whereas $\boldsymbol{\gamma} \in \mathbb{R}^p$ is the vector of parameters that tune the effects of the selected covariates on both the transition times. In this respect, we assume that the effect of the covariates on the two transition times is the same (i.e., the same vector of coefficients $\boldsymbol{\gamma}$ is used for both transition times), as this allows reducing the number of parameters to be estimated. In fact, it is always possible to use two different sets of covariates for the two transition times. However, this would require estimating the corresponding vectors of tuning parameters. This can be done when the number n of observations is large enough and the number p of covariates is sufficiently small.

From Eqs. (2), it clearly appears that the hazard function of each transition time is proportional to the corresponding baseline hazard function, being the multiplicative factor, $e^{\boldsymbol{x}\boldsymbol{\gamma}}$, time-independent (i.e. parameters $\boldsymbol{\gamma}$ and covariates $\boldsymbol{x}$ do not depend on time). Under this hypothesis, usually referred to as 'PH assumption', it can be proved that the Weibull baseline hazard model yields a modified WRM (e.g., [24]):

$$\begin{cases} T_1|\alpha_1, \beta_1, \boldsymbol{\gamma}, \boldsymbol{x} \sim^{indep.} Weibull\left(\frac{\alpha_1}{e^{(\boldsymbol{x}\boldsymbol{\gamma}/\beta_1)}}, \beta_1\right) \\ T_2|\alpha_2, \beta_2, \boldsymbol{\gamma}, \boldsymbol{x} \sim^{indep.} Weibull\left(\frac{\alpha_2}{e^{(\boldsymbol{x}\boldsymbol{\gamma}/\beta_2)}}, \beta_2\right) \end{cases} \tag{3}$$

Notice also that the underlying assumption in model (3) is that $T_2|\alpha_2, \beta_2, \boldsymbol{\gamma}, \boldsymbol{x}$ is independent on $T_1|\alpha_1, \beta_1, \boldsymbol{\gamma}, \boldsymbol{x}$. That is, the second transition time is not influenced by the first one; rather, it is influenced by the sojourn time in state 2.

Finally, vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are arranged to form the rows of matrix $\boldsymbol{X}$.

3.1 Prior elicitation

To complete the Bayesian model, prior distributions have to be assigned to the 4+p parameters $(\alpha_1, \beta_1, \alpha_2, \beta_2, \boldsymbol{\gamma})$ appearing in Eq. (3). We assume that some prior information is available about neither the

effect of the covariates, nor their relative importance in influencing the transition times. This situation can be accommodated by assigning to the p coefficients $\boldsymbol{\gamma}$ a non-informative diffuse normal prior distribution, also referred to as ridge prior, with mean $\mathbf{0}$ and large variance matrix:

$$\pi(\boldsymbol{\gamma}) = N_p(\mathbf{0}, \sigma^2 I_p) \tag{4}$$

where N_p denotes a p-variate normal distribution, I_p is the identity matrix of size p (i.e., the number of selected covariates), and σ^2 is the variance value, which must be large enough to allow the coefficients of proportionality $\boldsymbol{\gamma}$ in the WRM to vary in a wide range.

We assume, however, that some prior information is available about parameters $(\alpha_1, \beta_1, \alpha_2, \beta_2)$. In this respect, this work builds on the results of [21], in which the procedure proposed in [25] was used to elicit information from an expert about the two baseline transition times $T_1|(\boldsymbol{x} = \mathbf{0})$ and $T_2|(\boldsymbol{x} = \mathbf{0})$, and derive the corresponding prior distributions. The procedure shown in [21] is not reported in this paper, which directly uses its outcomes. See also [26].

Finally notice also that the assumption that the prior distributions are elicited with reference to the baseline setting does not undermine the generality of the proposed methodology. In fact, if an expert feels comfortable to give opinions under a certain set of conditions $\boldsymbol{x} = \boldsymbol{z} = (z_1, z_2, \dots, z_p)$ different from $\boldsymbol{x} = \mathbf{0}$, then a simple transformation of matrix $\boldsymbol{X}$ can be done to avoid complexity in the elicitation procedure. For example, assume that the typical range of the first covariate is $[15, 25]$; then, the expert is expected to provide more reliable estimates under the condition $z_1 = 20$, rather than under $z_1 = 0$. In such cases, to avoid that also parameter γ_1 appears in the elicitation of the prior distribution for parameters $(\alpha_1, \beta_1, \alpha_2, \beta_2)$, we can apply the transformation $x_{n1} \leftarrow x_{n1} - z_1, \forall n = 1, \dots, N$.

4 Variable selection

In general, the selection of the most influencing covariates can have a considerable impact on the effectiveness of the resulting degradation model, since irrelevant and noisy covariates tend to unnecessarily increase the complexity of the model and, thus, degrade modeling performance. The larger the number of covariates, the larger is the number of parameters in the WRM model and, thus, the larger is the amount of data required to tune the model. Then, variable selection is strongly beneficial for analyses based on a small amount of data.

The objective of variable selection is to pick from the whole set of Φ available covariates, the subset of those p that allows obtaining the WRM model providing the best performance.

Generally speaking, methods to select relevant variables from a large set of alternatives, can be divided in filtering and wrapper methods ([27], [28]). The former select the alternatives independently from the model used for the prediction (WRM, in this case), whereas the latter look at the improvement in the predictive performance achieved on the available data when combinations of covariates are encoded in the prediction model.

In this Section, we apply an approach to variable selection, which is made up of two steps: a filter method followed by a wrapper method (Figure 2). Namely, the filter method, which is based on the ordinal logistic model technique [19], is used to preliminary discard from the Φ variables those which are not relevant for determining the transition behavior. Then, a wrapper method, based on the maximization of the LPML

function, is applied on the filtered set of $\tilde{p}$ covariates to check whether the prediction performance of the WRM containing a subset of p out of the $\tilde{p}$ filtered covariates can outperform that of the model containing all the $\tilde{p}$ filtered covariates.

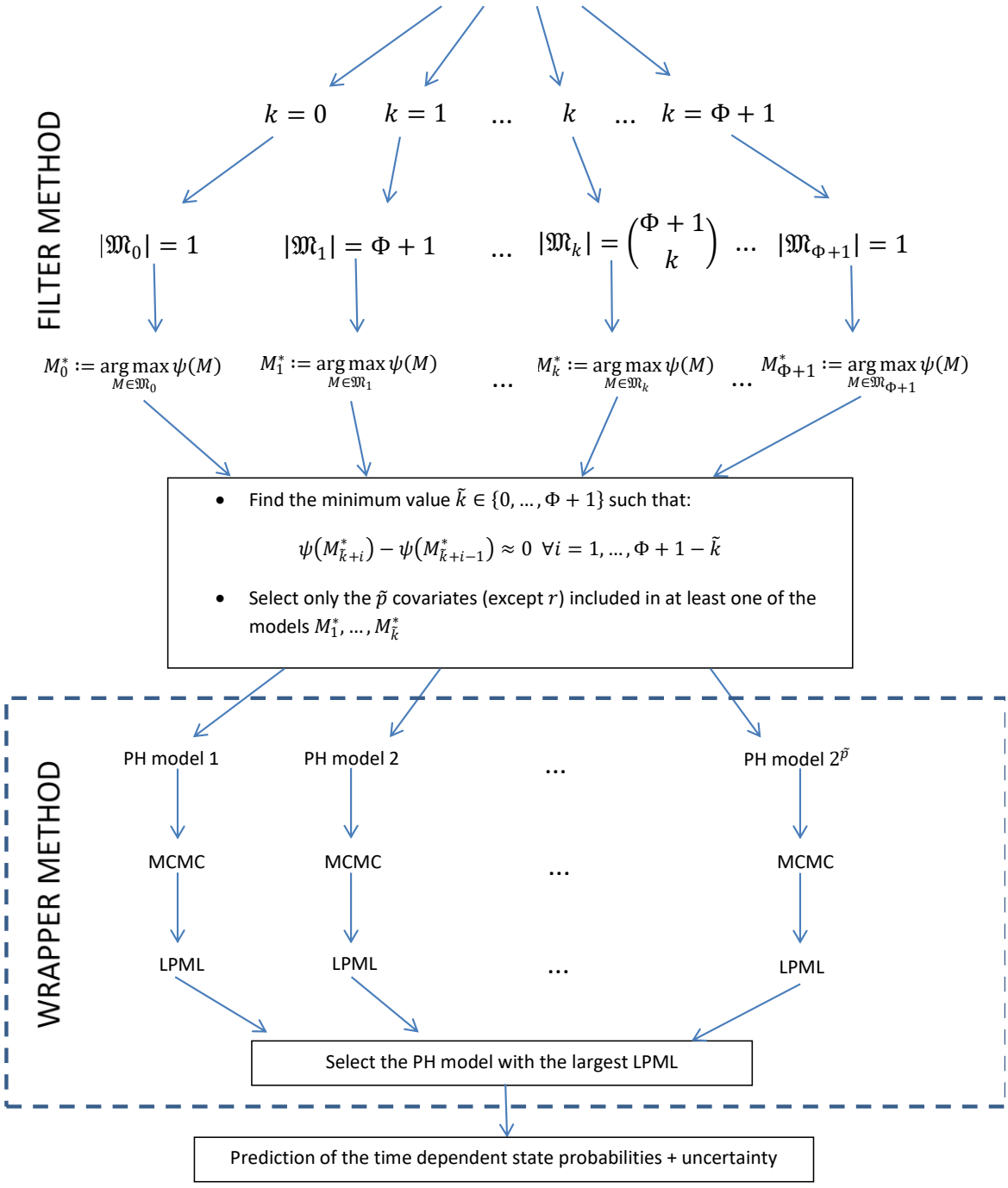


Figure 2. The variable selection procedure, divided in filter and wrapper methods.

Although the two-step approach is often adopted in condition monitoring applications for selecting the features of condition monitoring signals most relevant to classify the system health conditions, we recognize that the proposed method lacks a firm theoretical justification. Future research work will focus on strengthening the theoretical justifications of the approaches used in this study, whose main objective is, indeed, to report on the outcomes of a real industrial application, where the selection of the techniques has

also been motivated by diverse factors (e.g., computational time, familiarity of the experts with the techniques, etc.). Yet, the research work will also appraise the robustness of the proposed methodology.

It is worth noticing that the filter approach works within the frequentist probabilistic framework, whereas the wrapper approach and the final degradation model are built under the Bayesian one. In this respect, the use of the frequentist analysis to perform filtering selection does not entail a loss in prior information about the effect of the covariates, as we are assuming that no valuable expert information is available about them.

4.1 Filter method: ordinal logistic regression

Ordinal logistic regression is used to predict the realization of an ordinal variable, i.e. a categorical random variable with multiple ordered outcomes, on the basis of a set of covariates. In our case, the ordinal variable to be predicted is the degradation state $d = \{1,2,3\}$, whereas the Φ covariates $\boldsymbol{x}'$ are all used as explanatory variables, together with the replacement time r. This way, we are using a total of $\Phi + 1$ covariates in this regression model. We define the cumulative logits as [19]:

$$\begin{cases} \text{logit}\left[P\left(d_n \le l | r_n, \boldsymbol{x'}_n, \omega_l, \boldsymbol{\zeta}\right)\right] = \log\left(\dfrac{P\left(d_n \le l | r_n, \boldsymbol{x'}_n, \omega_l, \boldsymbol{\zeta}\right)}{1 - P\left(d_n \le l | r_n, \boldsymbol{x'}_n, \omega_l, \boldsymbol{\zeta}\right)}\right) = \omega_l + \zeta_0 r_n + \sum_{j=1}^{\Phi} \zeta_j x_{nj}\,, \quad l = 1,2 \\ P\left(d_n \le 3 | r_{n,} \boldsymbol{x'}_n\right) = 1 \end{cases} \tag{5}$$

where we have assumed that the vector of coefficients $\boldsymbol{\zeta} = (\zeta_0, \dots, \zeta_\Phi)$ is independent of l, i.e. that the impact of the covariates on the two (i.e., $l = 1,2$) logits is the same. This is coherent with the assumption of considering the same vector $\boldsymbol{\gamma}'$ to tune the effect of the covariates on the two transition times (Eq. 2).

According to [19], the contribution of the n^{th} observation to the likelihood function $\mathcal{L} = \prod_{n=1}^{N} \mathcal{L}\left(d_n | r_{n,} \boldsymbol{x'}_n, \omega_1, \omega_2, \boldsymbol{\zeta}\right)$ associated to the model defined by Eq. (5) is:

- if $d_n = 1$: $\mathcal{L}\left(d_n = 1 | r_{n,} \boldsymbol{x'}_n, \omega_1, \boldsymbol{\zeta}\right) = P\left(d_n = 1 | r_{n,} \boldsymbol{x'}_n, \omega_1, \boldsymbol{\zeta}\right)$
- if $d_n = 2$: $\mathcal{L}\left(d_n = 2 | r_{n,} \boldsymbol{x'}_n, \omega_1, \omega_2, \boldsymbol{\zeta}\right) = P\left(d_n \le 2 | r_{n,} \boldsymbol{x'}_n, \omega_2, \boldsymbol{\zeta}\right) - P\left(d_n = 1 | r_{n,} \boldsymbol{x'}_n, \omega_1, \boldsymbol{\zeta}\right)$
- if $d_n = 3$: $\mathcal{L}\left(d_n = 3 | r_{n,} \boldsymbol{x'}_n, \omega_2, \boldsymbol{\zeta}\right) = 1 - P\left(d_n \le 2 | r_{n,} \boldsymbol{x'}_n, \omega_2, \boldsymbol{\zeta}\right)$

  where: $$P\left(d_n = 1 | r_{n,} \boldsymbol{x'}_n, \omega_1, \boldsymbol{\zeta}\right) = \frac{\exp\left(\omega_1 + \zeta_0 r_n + \sum_{j=1}^{\Phi} \zeta_j x'_{nj}\right)}{1 + \exp\left(\omega_1 + \zeta_0 r_n + \sum_{j=1}^{\Phi} \zeta_j x'_{nj}\right)}$$

  $$P\left(d_n \le 2 | r_{n,} \boldsymbol{x'}_n, \omega_2, \boldsymbol{\zeta}\right) = \frac{\exp\left(\omega_2 + \zeta_0 r_n + \sum_{j=1}^{\Phi} \zeta_j x'_{nj}\right)}{1 + \exp\left(\omega_2 + \zeta_0 r_n + \sum_{j=1}^{\Phi} \zeta_j x'_{nj}\right)}$$

To check the goodness of fit of a logistic model, in this work we resort to the pseudo-R^2 measure proposed in [29]:

$$\psi := R^2_{CU} = \frac{1 - \left(\frac{\mathcal{L}_0}{\mathcal{L}}\right)^{2/N}}{1 - (\mathcal{L}_0)^{2/N}} \tag{6}$$

where $\mathcal{L}_0$ is the likelihood of the ordinal logistic model without covariates.

This measure can also be used to drive variable selection in the case where the number of covariates is relatively small (e.g., $\Phi \le 10$). In details, we first define the set $\mathfrak{M}$ of all the $2^{\Phi+1}$ possible ordinal logistic models M, each one characterized by a different combination of covariates. Then, $\mathfrak{M}$ is partitioned into the $\Phi + 2$ subsets $\mathfrak{M}_k = \{$ordinal logistic models M including exactly k covariates$\}$, $k = 0, \dots, \Phi + 1$.

Finally, for each $k = 0, \dots, \Phi + 1$, we perform the ordinal logistic regression for all the models $M \in \mathfrak{M}_k$ and identify the model $M_k^* := \underset{M \in \mathfrak{M}_k}{\arg\max}\, \psi(M)$ that best fits the data among those in $\mathfrak{M}_k$, and take the corresponding pseudo-R^2_{CU} value $\psi_k^* := \psi(M_k^*)$.

On this basis, to select the most influential variables, we look at the improvement in the fitting performance (summarized by ψ) brought by an increment in the number of variables encoded by the regression model. That is, we identify the minimum value $\tilde{k} \in \{0, \dots, \Phi + 1\}$ such that the differences $\psi^*_{\tilde{k}+i} - \psi^*_{\tilde{k}+i-1}$, $i = 1, \dots, \Phi + 1 - \tilde{k}$ give all negligible values. The number $\tilde{k}$ of most explanatory variables of the logistic model is expected to contain the replacement time r, which will not be used as a covariate in the WRM, but as the independent variable. For this, we consider that $\tilde{k} = \tilde{p} + 1$, where $\tilde{p}$ is the number of covariates selected from Φ, whereas 1 is added to take into account time. In this condition, we could select the covariates appearing in the model $M^*_{\tilde{k}}$, being confident on that there is no way of sensibly improving the fitting performance by including additional covariates in the regression model. However, to be conservative, we select the covariates appearing in at least one of the models with covariate number $M^*_k,\ k = 1, \dots, \tilde{k}$.

When the number of covariates is larger than 10, classical greedy procedures as stepwise regression can be used to perform variable selection [30]. These will be investigated in future works.

4.2 Wrapper method

The LPML criterion [20] is used to check whether the $\tilde{p}$ filtered covariates are all relevant in predicting the transition times. This is a leave-one-out cross-validation measure of the performance of the posterior predictions obtained from a Bayesian model. By adopting this approach, the search for the best model among the $2^{\tilde{p}}$ WRMs characterized by the different possible combinations of the filtered covariates, requires: *i)* to sample from the posterior distribution arising from each WRM, *ii)* to compute the corresponding LPML value and, finally, *iii)* to identify the model with the larger LPML value (Figure 2).

For generality, the methodology to derive and sample from the posterior distribution is shown in the following with respect to the WRM $\pi\left(\boldsymbol{\theta}, \widetilde{\boldsymbol{\gamma}} \,|\boldsymbol{r}, \boldsymbol{d}, \widetilde{\boldsymbol{X}}\right)$, including a number $p = \tilde{p}$ of filtered covariates and relative tuning parameters $\widetilde{\boldsymbol{\gamma}}$ (Section 3.1). On this basis, we define and calculate the corresponding LPML value, being $\widetilde{\boldsymbol{X}}$ the matrix with rows $\widetilde{\boldsymbol{x_n}} = (x_{n1}, \dots, x_{n\tilde{p}})$. This procedure can be applied similarly to compute the LPML of all the remaining $2^{\tilde{p}} - 1$ WR submodels.

4.2.1 Posterior distribution and sampling

In this Section, we derive the posterior distribution $\pi\left(\boldsymbol{\theta}, \widetilde{\boldsymbol{\gamma}} \,|\boldsymbol{r}, \boldsymbol{d}, \widetilde{\boldsymbol{X}}\right)$ of the WRM containing the subsets of the $\tilde{p}$ filtered covariates. Then, to sample from it, we resort to the Normal Random Walk Metropolis-Hastings (N-RWMH) algorithm, combined with its adaptive version proposed in [31].

For brevity, we group the four baseline parameters in a single vector $\boldsymbol{\theta} := (\alpha_1, \beta_1, \alpha_2, \beta_2) \in \Theta$, $\Theta := (\mathbb{R}_+)^w, w = 4$. Then, under the assumptions mentioned in Section 2, the contribution $L(r_n, d_n, \widetilde{\boldsymbol{x}}_n|\boldsymbol{\theta}, \widetilde{\boldsymbol{\gamma}})$ of the n^{th} observation to the likelihood function reads:

$$L\left(\boldsymbol{r}, \boldsymbol{d}, \widetilde{\boldsymbol{X}}|\boldsymbol{\theta}, \widetilde{\boldsymbol{\gamma}}\right) = \prod_{n=1}^{N} L(r_n, d_n, \widetilde{\boldsymbol{x}}_n|\boldsymbol{\theta}, \widetilde{\boldsymbol{\gamma}}) \tag{7}$$

which is such that:

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

$$L(r_n, d_n = 1, \tilde{\boldsymbol{x}}_n | \boldsymbol{\theta}, \tilde{\boldsymbol{\gamma}}) = 1 - F_{T_1}(r_n | \alpha_1, \beta_1, \tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{x}}_n) \tag{8}$$

where F_{T_1} is the Cumulative Distribution Function (CDF) of the random variable $T_1 |\, \alpha_1, \beta_1, \tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{x}}_n$. Notice also that for this type of outcome, no useful information can be deduced about the realization of T_2.

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

$$L(r_n, d_n = 2, \tilde{\boldsymbol{x}}_n | \boldsymbol{\theta}, \tilde{\boldsymbol{\gamma}}) = \int_0^{r_n} \left[1 - F_{T_2}(r_n - \tau | \alpha_2, \beta_2, \tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{x}}_n)\right] f_{T_1}(\tau | \alpha_1, \beta_1, \tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{x}}_n) d\tau \tag{9}$$

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

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

$$L(r_n, d_n = 3, \tilde{\boldsymbol{x}}_n | \boldsymbol{\theta}, \tilde{\boldsymbol{\gamma}}) = \int_0^{r_n} F_{T_2}(r_n - \tau | \alpha_2, \beta_2, \tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{x}}_n) f_{T_1}(\tau | \alpha_1, \beta_1, \tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{x}}_n) d\tau \tag{10}$$

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

$$\pi\left(\boldsymbol{\theta}, \tilde{\boldsymbol{\gamma}} | \boldsymbol{r}, \boldsymbol{d}, \tilde{\boldsymbol{X}}\right) \propto \prod_{n=1}^{N} L(r_n, d_n, \tilde{\boldsymbol{x}}_n | \boldsymbol{\theta}, \tilde{\boldsymbol{\gamma}}) \, \pi(\boldsymbol{\theta}) \pi(\tilde{\boldsymbol{\gamma}}) \tag{11}$$

Notice that the integral in Eq. (9) is the convolution of a Weibull PDF with parameters $\alpha_1, \beta_1, \tilde{\boldsymbol{\gamma}}$ with the complementary CDF of a Weibull distribution with parameters $\alpha_2, \beta_2, \tilde{\boldsymbol{\gamma}}$. Similarly, Eq. (10) is the convolution of the same PDF with the CDF of the second transition time. The results of these integrals are not available in closed form, whereby we have to numerically compute them at each iteration of the MCMC algorithm.

A Normal Random Walk Metropolis Hastings (N-RWMH) algorithm [32] is used to obtain samples from the posterior distribution $\pi\left(\boldsymbol{\theta}, \tilde{\boldsymbol{\gamma}} | \boldsymbol{r}, \boldsymbol{d}, \tilde{\boldsymbol{X}}\right)$. The algorithm efficiency heavily depends on the choice of the covariance matrix $\Sigma \in \mathbb{R}^{(w+\tilde{p})\times(w+\tilde{p})}$ of the normal proposal density q. For this reason, we dedicate a certain number S of initial iterations in the execution of an adaptive version of the algorithm ([31], [33]) that allows automatically tuning the covariance matrix entries. In particular, S must be large enough to allow the adaptive algorithm to tune the entries of the covariance matrix Σ of the normal proposal density on a range of values that possibly renders efficient the N-RWMH algorithm.

In details, the following algorithm, comprehensive of the preliminary adaptive run, is executed:

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

set $\left(\boldsymbol{\theta}^{(g)}, \widetilde{\boldsymbol{\gamma}}^{(g)}\right) = (\boldsymbol{\theta}^{*}, \widetilde{\boldsymbol{\gamma}}^{*})$

else

set $\left(\boldsymbol{\theta}^{(g)}, \widetilde{\boldsymbol{\gamma}}^{(g)}\right) = \left(\boldsymbol{\theta}^{(g-1)}, \widetilde{\boldsymbol{\gamma}}^{(g-1)}\right)$

end if

6. if $g \leq S$

update $\boldsymbol{\eta}^{(g)} = \boldsymbol{\eta}^{(g-1)} + \varepsilon^{(g)}\left(\left(\boldsymbol{\theta}^{(g)}, \widetilde{\boldsymbol{\gamma}}^{(g)}\right) - \boldsymbol{\eta}^{(g-1)}\right)$

update

$$\Sigma^{(g)} = \Sigma^{(g-1)} + \varepsilon^{(g)}\left[\left(\left(\boldsymbol{\theta}^{(g)}, \widetilde{\boldsymbol{\gamma}}^{(g)}\right) - \boldsymbol{\eta}^{(g-1)}\right)\left(\left(\boldsymbol{\theta}^{(g)}, \widetilde{\boldsymbol{\gamma}}^{(g)}\right) - \boldsymbol{\eta}^{(g-1)}\right)^{T} - \Sigma^{(g-1)}\right]$$

else

set $\Sigma^{(g)} = \Sigma^{(g-1)}$

end if

end for

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

At the end of the algorithm, we discard the first S iterations and relabel the remaining MCMC samples: $\left(\boldsymbol{\theta}^{(g)}, \widetilde{\boldsymbol{\gamma}}^{(g)}\right) \leftarrow \left(\boldsymbol{\theta}^{(g+S)}, \widetilde{\boldsymbol{\gamma}}^{(g+S)}\right)$ $g = 1, \dots, G$.

4.2.2 LPML

The LPML of the WRM including the $\tilde{p}$ filtered covariates is defined as:

$$LPML = \sum_{n=1}^{N} \log\left(CPO_n\right) \tag{12}$$

where CPO_n is the Conditional Predictive Ordinate for the n^{th} observation:

$$CPO_n := P\left(r_n, d_n | \left(\boldsymbol{r}, \boldsymbol{d}, \widetilde{\boldsymbol{X}}\right)_{-n}, \widetilde{\boldsymbol{x}}_n\right) = \int_{\Theta\times\Gamma} L(r_n, d_n, \widetilde{\boldsymbol{x}}_n | \boldsymbol{\theta}, \widetilde{\boldsymbol{\gamma}})\, \pi\left(\boldsymbol{\theta}, \widetilde{\boldsymbol{\gamma}} | \left(\boldsymbol{r}, \boldsymbol{d}, \widetilde{\boldsymbol{X}}\right)_{-n}\right) d\boldsymbol{\theta} d\widetilde{\boldsymbol{\gamma}} \tag{13}$$

where $\widetilde{\boldsymbol{X}}$ are the columns of $\boldsymbol{X}$ corresponding to the pre-selected $\tilde{p}$ covariates, $\widetilde{\boldsymbol{\gamma}}$ the corresponding tuning parameter and $\left(\boldsymbol{r}, \boldsymbol{d}, \widetilde{\boldsymbol{X}}\right)_{-n}$ the available dataset with the n^{th} observation removed. We can notice that CPO_n is the expected probability, based on all the dataset except the n^{th} record, that the n^{th} component is in state d_n after a working time equal to r_n. In this sense, LPML is a leave-one-out-cross validation measure of the prediction performance of the model. Certainly, the larger the CPO, the better the predictive performance of the WRM, which is encoded in the likelihood function.

Notice also from Eqs. $(12-13)$ that the calculation of LPML requires, in theory, to sample from the posterior distribution $\pi\left(\boldsymbol{\theta}, \widetilde{\boldsymbol{\gamma}} | \left(\boldsymbol{r}, \boldsymbol{d}, \widetilde{\boldsymbol{X}}\right)_{-n}\right)$ $\forall n = 1, \dots, N$. This is computationally expensive also for small sample sizes N. However, the following identity can be proved (e.g., [34]):

$$CPO_n = \left(\int_{\Theta\times\Gamma} \frac{\pi\left(\boldsymbol{\theta}, \widetilde{\boldsymbol{\gamma}} | \boldsymbol{r}, \boldsymbol{d}, \widetilde{\boldsymbol{X}}\right)}{L(r_n, d_n, \widetilde{\boldsymbol{x}}_n | \boldsymbol{\theta}, \widetilde{\boldsymbol{\gamma}})} d\boldsymbol{\theta} d\widetilde{\boldsymbol{\gamma}}\right)^{-1} \quad n = 1, \dots, N \tag{14}$$

where each CPO can be derived starting from a unique posterior distribution, i.e. $\pi\left(\boldsymbol{\theta}, \tilde{\boldsymbol{\gamma}} | \boldsymbol{r}, \boldsymbol{d}, \tilde{\boldsymbol{X}}\right)$. For this, a MCMC estimate of each CPO can be obtained by exploiting the samples $\left(\boldsymbol{\theta}^{(g)}, \tilde{\boldsymbol{\gamma}}^{(g)}\right)_{g=1,\dots,G}$ from the posterior distribution $\pi\left(\boldsymbol{\theta}, \tilde{\boldsymbol{\gamma}} | \boldsymbol{r}, \boldsymbol{d}, \tilde{\boldsymbol{X}}\right)$:

$$\widehat{CPO_n} = \left(\frac{1}{G} \sum_{g=1}^{G} \frac{1}{L(r_n, d_n, \tilde{\boldsymbol{x}}_n | \boldsymbol{\theta}^{(g)}, \tilde{\boldsymbol{\gamma}}^{(g)})} \right)^{-1} \quad n = 1, \dots, N \tag{15}$$

Consequently, the estimate for the LPML is:

$$\widehat{LPML} = \sum_{n=1}^{N} \log\left(\widehat{CPO_n}\right) \tag{16}$$

From Eqs. $(12 - 13)$, it clearly emerges that the larger is the LPML value, the better is the WRM from the prediction capability perspective. Therefore, we can select the WRM with $p \leq \tilde{p}$ covariates having the largest LPML as the best candidate to make posterior predictions for the time-dependent state probabilities of a new component with a generic vector of selected covariates $\boldsymbol{x}$.

Notice also that the proposed procedure actually requires $\tilde{p}$ to be quite small (i.e., $\tilde{p} \leq 4$), in order to avoid computational burden. Other possible approaches, which may be less computationally demanding than LPML, are the Stochastic Search Variable Selection ([35]), spike and slab priors ([36]), Bayesian LASSO ([37]). These will be investigated in future research work.

Once the best subset of covariates $\boldsymbol{X} \subseteq \tilde{\boldsymbol{X}} \subseteq \boldsymbol{X}'$ has been selected, the MCMC sample $\left\{\boldsymbol{\theta}^{(g)}, \boldsymbol{\gamma}^{(g)}\right\}_{g=1,\dots,G}$ from its corresponding posterior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{\gamma} | \boldsymbol{r}, \boldsymbol{d}, \boldsymbol{X})$ can be exploited to estimate the probabilities of occupying the three degradation states over time for a new component which has a generic vector $\boldsymbol{x}$ of covariates (see Appendix 8.2).

Notice that we used the harmonic mean estimator in Eq. (15) because of its simplicity, which is the main advantage of this method over other more specialized techniques [41].

5 Case study

As in [21], we consider the gradual degradation of EPDM diaphragms used in the production line of a pharmaceutical company as sealing elements of flow control valves that have 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. Differently from [21], we now consider the effect of the harsh environmental conditions in which these components operate on the degradation behaviour.

According to expert judgements, 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. This is a failure state, as it can result in undesired contaminations leading to a complete loss of recent production.

Notice that the expert assignments of a degraded condition to its state may suffer from imprecision and subjectiveness. Future works will investigate the development of other methods, which take into account that the observed levels of damage are clearly imprecise. For this, we will investigate the possibility of building on the methodology developed in [38] or in [39], within the fuzzy logic framework.

A dataset containing information about *N*=109 diaphragms employed in the past is available. Initially, we have Φ=11 covariates (Table 1), concerning the use of some physical and chemical agents that come into contact with EPDM diaphragms during the production (e.g. high temperature saturated steam, cleaning agents, chemical detergents, sparge gases and multiple product media). 73 components out of *N* have been retired from service when they were in state 1, 32 in state 2, and 4 in state 3. The detection of a component in the third degradation state is quite rare, since this corresponds to a failure.

The generic entry x_{nj} of the covariates matrix $\boldsymbol{X}$ is the rate of exposure of the n^{th} diaphragm to the j^{th} agent in the time interval between installation (time zero) and replacement (time r_n), $j = 1, \dots, p$. Each covariate is normalized to range in the interval $[0,1]$. The exposure to each agent is assumed constant during the whole lifetime of the diaphragms, so that we can assume that the covariates are time-independent.

Notice that the identification of the degradation state of the EPDM diaphragms requires to perform a destructive test on the component. This explains why it has not been possible to carry out more than one inspection per diaphragm.

# Diaphragm	Replacement Time ($\boldsymbol{r}$)	Degradation State ($\boldsymbol{d}$)	Covariates Matrix ($\boldsymbol{X}$)		
			X^1	…	X^{11}
1	16.4	1	0.89	…	0
2	26.7	1	0.05	…	1
…	…	…	…	…	…
109	173	2	0.09	…	0.78

Table 1. The available dataset

Notice that in the biopharmaceutical industry there exist a variety of industry specific complications explaining the independence between maintenance interventions and component degradation state, which are not rare in other industrial fields (e.g., [23]). Firstly, there is no known way to provide a quantitative assessment of the degradation state of the components in-situ: although some recommendations from the Original Equipment Manufacturer (OEM) are available, these are derived from accelerated lifetime testing under covariate settings whose matching with the real ones is doubtful, being the covariates not continuously monitored. Thus, the only information available to infer the degradation behaviour is that exploited in this work (degradation state and corresponding inspection time and covariate values). Given the lack of knowledge on the degradation process and the rigorous regulatory nature of the biopharmaceutical industry (e.g., all elements of the production process must be cleaned and hermetically sealed), companies adopt strict risk-averse maintenance policies.

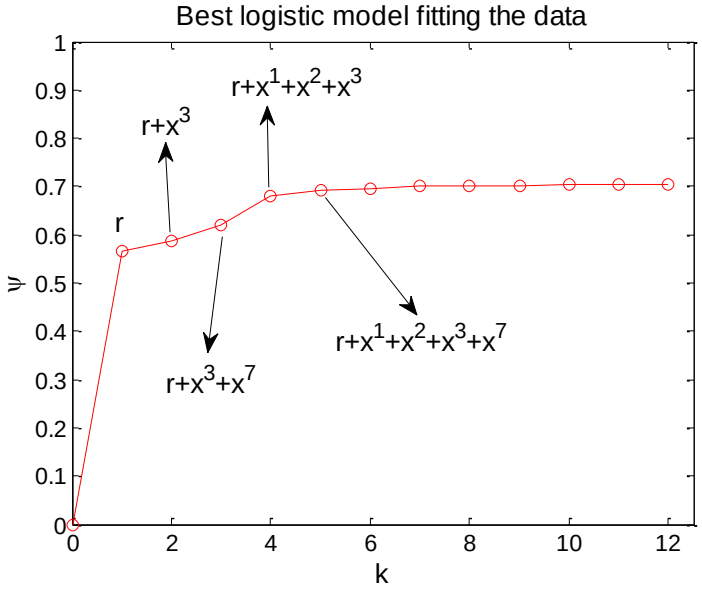
Secondly, the production process operates on a 24 hour 365 day basis; then, performing diaphragm inspections would require to shutdown the system. Due to the strict regulations of the biopharmaceutical industry, re-starting the production process after shutdown takes approximately 30 days. This tells us that shutdown could result in millions of Euro in lost revenue and, then, a maintenance strategy based on periodic inspections is not viable, as these cannot be performed when the production lines are producing a lot of medicines. On the contrary, maintenance must be performed in an opportunistic approach, the maintenance opportunities to gain access to the diaphragms being upon a programmed system shutdown. In turn, there is no financial or safety advantage from inspecting a diaphragm as opposed to replacing it, the completion of

these actions requiring the same amount of plant shutdown time. This issue is further highlighted considering the ratio comparing the cost of replacing a diaphragm to the cost of lost production due to a failed diaphragm is approximately 1:80000.

The procedure described in Section 4 is applied to select the most influencing covariates. This has taken almost seven minutes on a 4GB RAM, 2.27 GHz personal computer. Figure 3 shows that models with more than 3 covariates, in addition to the time (i.e., the baseline regression variable, see Eq. (5)), achieve negligible gains in comparison to the best model with the 4 variables (r, X^1, X^2, X^3). To be conservative, we select the first 4 covariates (i.e., X^1, X^2, X^3, X^7), as a preliminary set to be confirmed by the Bayesian WRM.

Notice that the heuristic rule of looking at the 'elbow' of the curve in Figure 3 is a common practice to address trade off problems such as in clustering analysis, where similarly to our problem, one should choose an optimal number so that adding another covariate does not give much better modeling of the data ([40][39]).

Figure 3. The best models with k covariates and their pseudo-R^2_{CU}



To develop the Bayesian three-state model of the degradation process, the prior distributions of parameters $(\alpha_1, \beta_1, \alpha_2, \beta_2)$ have been elicited from an expert using a questionnaire. The expert was asked to provide a set of quantiles of both transition times T_1 and T_2 under the condition $(x_1, x_2, x_3, x_7) = \mathbf{0}$.

The expert judgements and the outcomes of the prior elicitation procedure can be found in [21]. Here, we just highlight that these results were obtained forcing the two shape parameters to be not smaller than 1. Indeed, 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, because the degradation process of EPDM diaphragms is characterized by a continuous worsening of the conditions of the component, which leads to an increasing probability of having a transition into the successive degradation state.

Once the prior distributions have been obtained, we can proceed applying the second variable selection method, which is based on the LPML. The $\tilde{p}$=4 covariates pre-selected in the first step give rise to $2^4 = 16$ different models, from which we have to choose the best.

The computation of the LPML for each model is still feasible without an excessive computational effort if the number of diaphragms in the dataset found in states 2 or 3 is limited, being the numerical approximation of the likelihood functions of the components found in states 2 or 3 the most computational demanding step in

the proposed MCMC algorithm. Hence, we sample from the posterior distributions $\left(\boldsymbol{\theta}, \boldsymbol{\gamma} | \boldsymbol{r}, \boldsymbol{d}, M_j\right)$ $j = 0, \ldots, 15$, by applying the following procedure:

- $S = 100000$ iterations of the A-MCMC algorithm are run, with the only aim of tuning the entries of the proposal distribution covariance matrix Σ.
- $G = 400000$ iterations of the N-MHRW algorithm are run with the tuned proposal distribution covariance matrix set to $\Sigma^{(S)}$.

As described in Section 4, we can exploit the 16 sampled chains to estimate the LPML for each of the models (Figure 4). The model which includes all the four covariates X^1, X^2, X^3, X^7 has the highest LPML, i.e. it achieves better cross-validation predictions than the other models. In light of the obtained results, we choose the model including the four covariates X^1, X^2, X^3, X^7 to make predictions about the time-dependent state probabilities of new components. The wrapper approach has taken four hours for every model, for a total of 64 hours on the same personal computer (4GB RAM, 2.27GHz).

Notice also that the simpler model including the two covariates X^3, X^7 only shows good cross-validation predictions, and may be preferred to model M_{15} in the case in which the tracking of the covariate values become expensive.

To provide evidence that implementing the procedure several times yields the same set of selected covariates, we have performed the following robustness study.

From the $S = 100000$ iterations, we have selected 10 MC chains by applying the thinning procedure ([43]): the first chain contains samples 1, 11, 21, etc.; the second chain contains sample 2, 12, 22, etc; finally, the tenth chain contains samples 10, 20, 30, etc. If we re-implement the procedure illustrated above with these chains, then we get the results summarized in Figure 7: the set of covariates with the largest LMPL remains always the same. This proves the robustness of the proposed methodology in this case study.

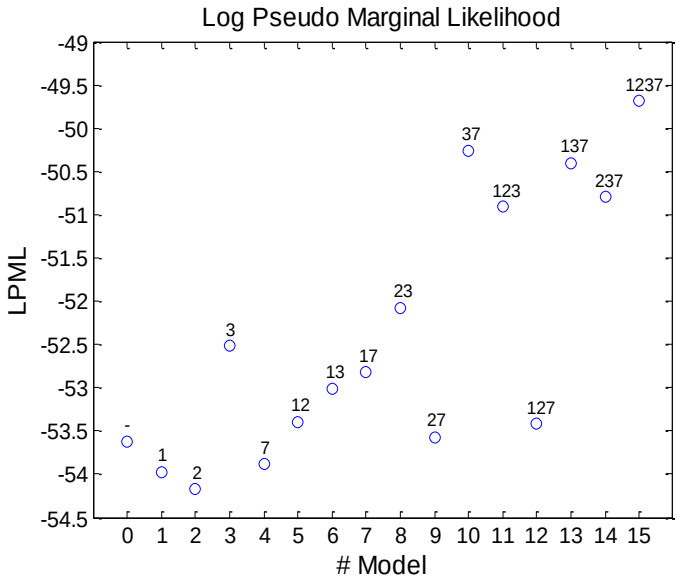


Figure 4. The estimated LPML for the 16 different models, labeled with the numbers of the included covariates

Finally, we can use the selected model with four covariates to make predictions about the time-dependent state probabilities of a new component with covariate vector $(X^1 = x_1, X^2 = x_2, X^3 = x_3, X^7 = x_7)$.

Although the exact exposure rate values x_1, x_2, x_3 and x_7 are available only upon replacement, such values are usually known with good precision at the moment of the component installation, as they depend on the type of production system where the component is installed into and on the production schedule. Therefore, the predictions about the time-dependent state probabilities of a new component can be made at installation. For example, in Figure 8, we can compare the estimates of the time dependent state probabilities for three different new components, with vectors of covariates $(0.1, 0, 0, 0)$, $(0.1, 0, 0, 0.5)$ and $(0.1, 0, 0, 1)$. In practice, we are changing the values of X^7 on its whole range $[0,1]$, while keeping constant the values of the other three covariates. Figure 6 shows a reduction of the two transition times as X^7 increases and, thus, an acceleration of the degradation process.

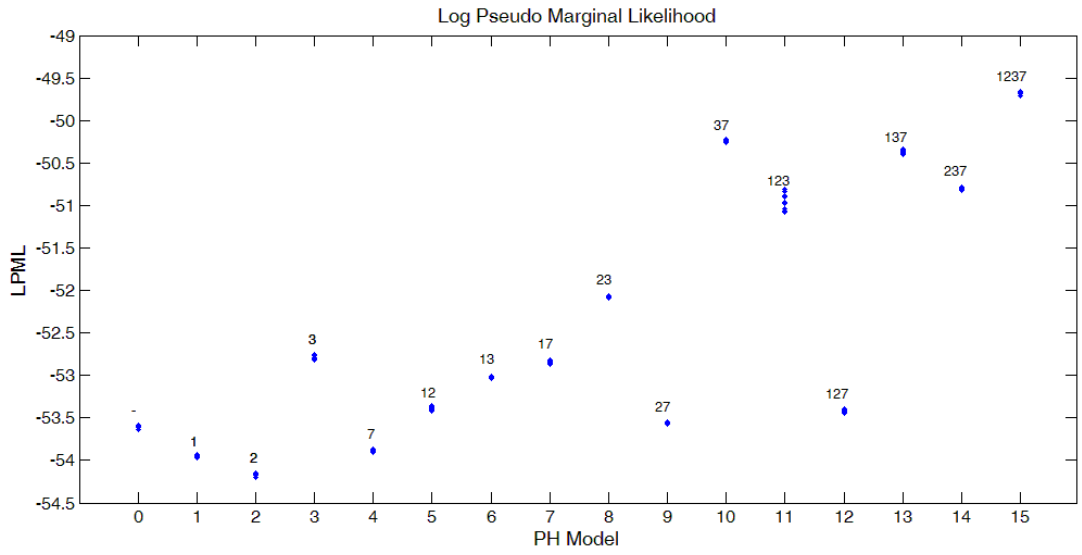


Figure 7. The LPML estimates for the 16 different models provided by 10 different chains.

The obtained results, and in particular the predicted probability of occupying state 3 (which is the CDF of the EPDM diaphragms failure time), can be used in support of maintenance decisions. Indeed, they can enter an optimization model to find the optimal replacement time for each installed EPDM diaphragm, depending on the values of the exposure rate to physical and chemical agents during the production. On one side, the optimization model has to give account to the costs of performing preventive maintenance actions, including the cost of the diaphragm itself, the technical actions, the economic loss due to temporary unavailability of the system, the possibility of human errors during the repair, etc. On the other side, the optimization model has to encode the expected cost of failure, as this causes a significant loss of production and a long downtime of the entire plant for the purification and certification of the system.

6 Conclusions

In this paper, we have proposed a complete procedure for assessing the reliability of industrial equipment whose degradation process can be described as a three-state semi-Markov model, with transition times influenced by operational or environmental conditions.

To select the most significant covariates, a two-step procedure has been proposed, which is made up of: *i)* a preliminary filter method to discard not significant covariates, which is based on the ordinal logistic regression; *ii)* a subsequent wrapper selection which identifies, among the pre-selected covariates, the subset that achieves the best cross-validated predictions when included in a Bayesian Proportional Hazard

model. In particular, to perform step *ii)*, a Metropolis-Hastings algorithm combined with numerical integration methods is developed, which is used to sample from the complex posterior distributions of the possible candidate models. Then, the model with the largest LPML (i.e. that with the best cross-validation predictive capability) is selected to make posterior predictions.

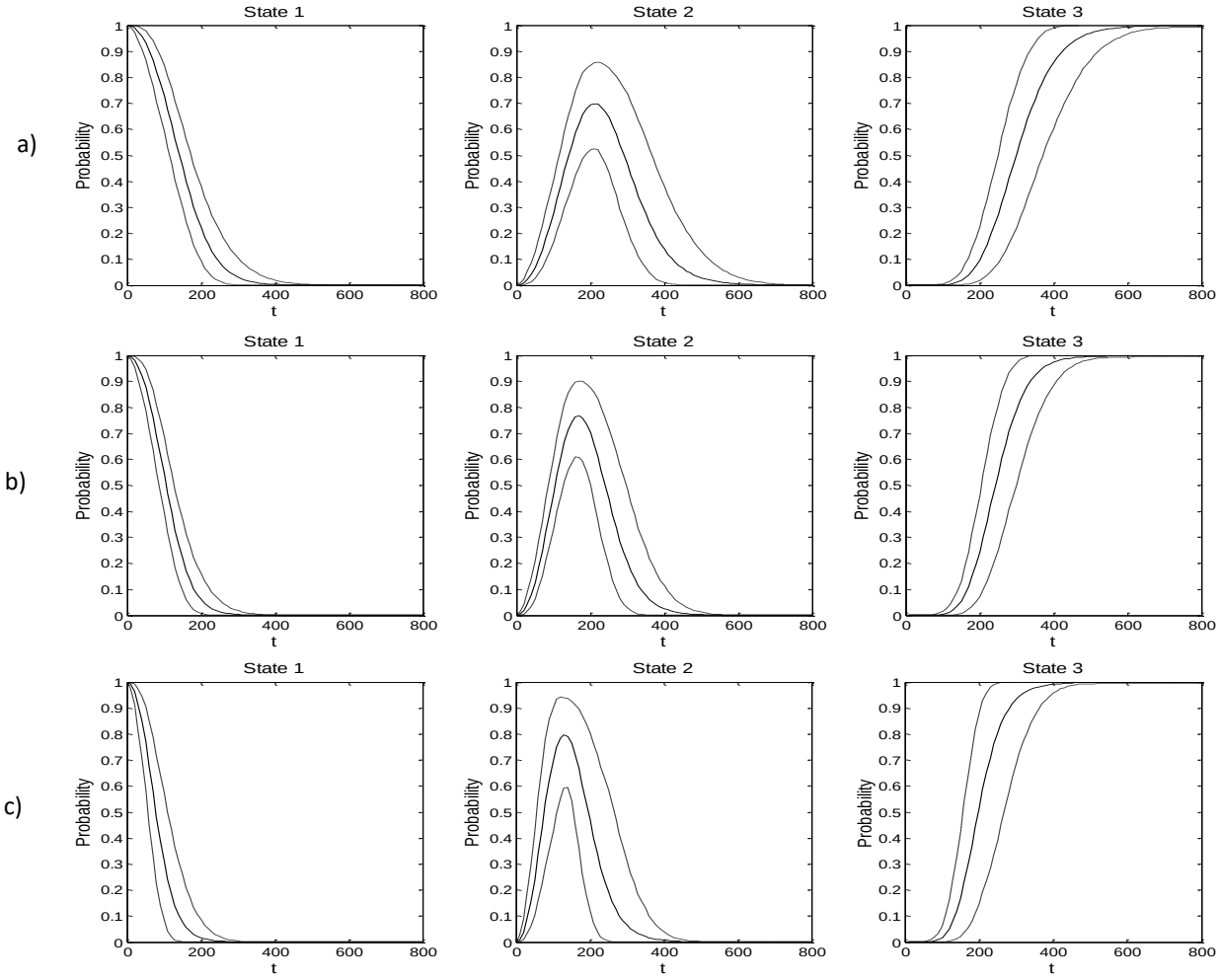


Figure 8. Time dependent state probabilities (solid lines) of the component with covariates $(0.1, 0, 0, 0)$ (6a), with covariates $(0.1, 0, 0, 0.5)$ (6b) and with covariates $(0.1, 0, 0, 1)$ (6c). The dashed lines are the corresponding $(0.05, 0.95)$ pointwise credibility bands.

The posterior samples of the selected model allow estimating the time-dependent state probabilities and the corresponding uncertainty. The obtained results can be used to establish the best replacement times for the components (depending on the operational and environmental conditions), which minimize the overall production costs, also in presence of uncertainty.

The proposed procedure is also helpful for a deeper understanding of the degradation mechanisms affecting the EPDM, as it allows identifying the most relevant influencing factors. In this sense, the method can be used as a validation tool of pre-existing physical degradation models or can be integrated with them.

Future works will focus on the development of a full Bayesian stochastic variable selection, which allows a computationally efficient selection of the most influent covariates also in the case a large number of them is available. Furthermore, the WRM will be extended to consider possible dependencies among the transition times.

7 References

[1] Compare, M., Martini, F., Zio, E., Carlevaro, F., Mattafirri, S., "Semi-Markov model for the oxidation degradation mechanism in gas turbine nozzles", 65(2), 7404290, pp. 574-581, 2016.

[2] Liu, Y. Lin P., Li, Y-F.; Huang, H-Z., "Bayesian Reliability and Performance Assessment for Multi-State Systems", IEEE Transactions on Reliability, 64 (1), 2015.

[3] Moghaddass, R., Zuo, M.J., "Multi-state degradation analysis for a condition monitored device with unobservable states", Proceedings of 2012 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, ICQR2MSE 2012, pp. 549-554, 2012.

[4] Jardine, A.K.S., Joseph, T., Banjevic, D., "Optimizing condition-based maintenance decisions for equipment subject to vibration monitoring", Journal of Quality and Maintenance Engineering, 5, pp. 1355-2511, 1999.

[5] Martón, I., Sánchez, A.I. and Martorell, S., "Ageing PSA incorporating effectiveness of maintenance and testing", Reliability Engineering & System Safety, 139(0), pp. 131-140, 2015.

[6] Doyen, L., Gaudoin, O., "Classes of imperfect repair models based on reduction of failure intensity or virtual age", Reliability Engineering & System Safety, 84(1), pp. 45-56, 2004.

[7] Ansell, J.I. and Philipps, M.J., "Practical aspects of modelling of repairable systems data using proportional hazards models", Reliability Engineering & System Safety, 58(2), pp. 165-171, 1997.

[8] Kumar, D. and Klefsjö, B., "Proportional hazards model: a review", Reliability Engineering & System Safety, 44(2), pp. 177-188, 1994.

[9] Cox, D.R., "Regression models and life tables", Journal of the Royal Statistical Society, Series B, 34, pp. 187-220, 1972.

[10] Cox, D.R., Oakes, D., "Analysis of survival data", Chapman and Hall, London, 1984.

[11] Smith, R.L., "Weibull regression models for reliability data", Reliability Engineering & System Safety, 34(1), pp. 55-76, 1991.

[12] Peng, W., Li, Y-F; Yang, Y-J., Mi, J., Huang, H-Z., "Leveraging Degradation Testing and Condition Monitoring for Field Reliability Analysis With Time-Varying Operating Missions", IEEE Transactions on Reliability, 64 (4), pp. 1367 – 1382, 2015.

[13] Elsayed, E.A., Zhang, H., "Design of PH-based accelerated life testing plans under multiple-stress-type", Reliability Engineering and System Safety, 92(3), pp. 286–292, 2007.

[14] Huda, S., Abdollahian, M., Mammadov, M. Yearwood, J., Ahmed, S., Sultan, I., "A hybrid wrapper–filter approach to detect the source(s) of out-of-control signals in multivariate manufacturing process", European Journal of Operational Research, 237 (3), pp. 857-870, 2014.

[15] Uncu, O., Türkşen, I.B., "A novel feature selection approach: Combining feature wrappers and filters", Information Sciences, 177 (2), pp. 449-466, 2007.

[16] Kamal, N.A.M., Bakar, A.A., Zainudin, S., "Filter-wrapper approach to feature selection of GPCR protein", Proceedings - 5th International Conference on Electrical Engineering and Informatics: Bridging the Knowledge between Academic, Industry, and Community, ICEEI 2015, pp. 693-698, 2015.

[17] Cadenas, J.M., Garrido, M.C., Martínez, R., "Feature subset selection Filter-Wrapper based on low quality data", Expert Systems with Applications, 40 (16), pp. 6241-6252, 2013.

[18] Hsu, H.-H., Hsieh, C.-W., Lu, M.-D., "Hybrid feature selection by combining filters and wrappers", Expert Systems with Applications, 38 (7), pp. 8144-8150, 2011.

[19] Agresti, A., "Categorical Data Analysis", 3rd Edition, Hoboken, NJ: John Wiley & Sons, 2013.

[20] Gelfand, A.E., Mallick, B.K., "Bayesian analysis of proportional hazards models built from monotone functions", Biometrics, 51, pp. 843-852, 1995.

[21] Compare, M., Baraldi, P., Bani, I., Zio, E., McDonnell, D., "A Bayesian Semi-Markovian Framework for Prognostics and Health Management of Industrial Components", Reliability Engineering and System Safety, 166, pp. 25-40, 2017.

[22] Boutros, T., Liang, M., "Detection and diagnosis of bearing and cutting tool faults using hidden Markov models", Mechanical Systems and Signal Processing, 25 (6), pp. 2102-2124, 2011.

[23] Compare, M., Di Maio, F., Zio, E., Carlevaro, F., Mattafirri, S., "Improving scheduled maintenance by missing data reconstruction: A double-loop Monte Carlo approach", Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 230 (5), pp. 502-511, 2016.

[24] Hosmer, D.W., Lemeshow, S., May, S., "Applied survival analysis", Wiley Blackwell, 2011.

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

[26] Bedrick, E.J., Christensen, R., and Johnson, W.O., "Bayesian accelerated failure time analysis with application to veterinary epidemiology", Statistics in Medicine, 19, 221-237, 2000.

[27] Kohavi, R., John, G., "Wrappers for feature subset selection", Artificial Intelligence 97, pp. 273-324, 1997.

[28] Zio, E., Baraldi, P., Pedroni, N., "Selecting Features for Nuclear Transitents Classification by Means of Genetic Algorithms", IEEE Trans. Nuclear Science, 53, 3, pp. 1479-1493, 2006.

[29] Cragg, J., Uhler, R., "The demand for automobiles", Canadian Journal of Economics, 3, pp. 386-406, 1970.

[30] Harrell, F. E., Lee, K. L., Mark, D. B., "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors", Statistics in Medicine, 15, pp. 361-387, 1996.

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

[32] Robert, C.P., Casella, G., "Monte Carlo statistical methods", New York: Springer, 2004.

[33] Haario, H., Saksman, E., Tamminen, J., "An adaptive Metropolis algorithm", Bernoulli 7(2), pp. 223-242, 2001.

[34] Gelfand, A.E., Dey, D.K., "Bayesian model choice: Asymptotics and exact calculations", Journal of the Royal Statistical Society, Series B, 56, pp. 501-514, 1994.

[35] George, E.I., McCullogh, R.E., "Variable selection via Gibbs sampling", Journal of the American Statistical Association, 88, pp.881-889, 1993.

[36] Mitchell, T.J., Beauchamp, J.J., "Bayesian variable selection in linear regression", Journal of the American Statistical Association, 83, pp. 1023-1036, 1988.

[37] Park, T., Casella, G., "The Bayesian lasso", Journal of the American Statistical Association, 103, pp. 681-686, 2008.

[38] Garca-Zattera, M.J., Jara, A., Lesaffre, E., Marshall, G., "Modeling of multivariate monotone disease processes in the presence of misclassification", Journal of the American Statistical Association, 107, 976–989, 2012.

[39] Denœux, T. "Maximum likelihood estimation from fuzzy data using the em algorithm", Fuzzy Sets and Systems, 183(1), pp. 72-91, 2011.

[40] Thorndike, R.L., "Who Belongs in the Family?", Psychometrika, 18(4): 267–27, 1953

[41] Raftery, A., Newton, M., Satagopan, J., Krivitsky, P., "Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity", Bayesian Statistics, 8, pp. 1–45, 2007.

[42] 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.

[43] Maceachern, S.N., Berliner, L.M., "Subsampling the Gibbs Sampler", The American Statistician, 48(3), pp. 188-190, 1994.

8 Appendix

Once the G samples $\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\gamma}^{(1)}\right), \dots, \left(\boldsymbol{\theta}^{(\mathrm{G})}, \boldsymbol{\gamma}^{(\mathrm{G})}\right)$ have been obtained from the posterior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{\gamma} | \boldsymbol{r}, \boldsymbol{d}, \boldsymbol{X})$, we can estimate the following values:

- The expected probability of occupying state 1 after working time $t \geq 0$:
  $\mathbb{E}_{\pi(\boldsymbol{\theta},\boldsymbol{\gamma}|\boldsymbol{r},\boldsymbol{d},\boldsymbol{X})}[P(T_1 > t | \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{x})] =$
  $$= \int_{\mathbb{R}_+ \times \mathbb{R}_+ \times \Gamma} \left(1 - F_{T_1}(t | \alpha_1, \beta_1, \boldsymbol{\gamma}, \boldsymbol{x})\right) \pi(\alpha_1, \beta_1, \boldsymbol{\gamma} | \boldsymbol{r}, \boldsymbol{d}, \boldsymbol{X}) d\alpha_1 d\beta_1 d\boldsymbol{\gamma} \tag{21}$$

- The expected probability of occupying state 2 after working time $t \geq 0$:
  $\mathbb{E}_{\pi(\boldsymbol{\theta},\boldsymbol{\gamma}|\boldsymbol{r},\boldsymbol{d},\boldsymbol{X})}[P(T_1 \leq t, T_1 + T_2 > t | \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{x})] =$
  $$= \int_{\Theta \times \Gamma} \left(\int_0^t \left[1 - F_{T_2}(t - \tau | \alpha_2, \beta_2, \boldsymbol{\gamma}, \boldsymbol{x})\right] f_{T_1}(\tau | \alpha_1, \beta_1, \boldsymbol{\gamma}, \boldsymbol{x}) d\tau \right) \pi(\boldsymbol{\theta}, \boldsymbol{\gamma} | \boldsymbol{r}, \boldsymbol{d}, \boldsymbol{X}) d\boldsymbol{\theta} d\boldsymbol{\gamma} \tag{22}$$

- The expected probability of occupying state 3 after working time $t \geq 0$:
  $\mathbb{E}_{\pi(\boldsymbol{\theta},\boldsymbol{\gamma}|\boldsymbol{r},\boldsymbol{d},\boldsymbol{X})}[P(T_1 + T_2 < t | \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{x})] =$
  $$= \int_{\Theta \times \Gamma} \left(\int_0^t F_{T_2}(t - \tau | \alpha_2, \beta_2, \boldsymbol{\gamma}, \boldsymbol{x}) f_{T_1}(\tau | \alpha_1, \beta_1, \boldsymbol{\gamma}, \boldsymbol{x}) d\tau \right) \pi(\boldsymbol{\theta}, \boldsymbol{\gamma} | \boldsymbol{r}, \boldsymbol{d}, \boldsymbol{X}) d\boldsymbol{\theta} d\boldsymbol{\gamma} \tag{23}$$

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. $(22 - 24)$ are 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 ([42]).

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

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

whose g^{th} row contains the J values of the Complementary CDF of the random variable $T_1 | \{\boldsymbol{\theta} = \boldsymbol{\theta}^{(g)}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{(g)}\}$ evaluated on the time grid $\{t_j\}_{j=1,\dots,J}$. Unbiased estimates of the expected values (over time) in Eq. (22) 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

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

and:

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

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