On alternative Monte Carlo methods for parameter estimation in gamma process models with intractable likelihood

Daniel Z. Herr, Radislav Vaisman, Mitchell Scovell, Nikolai Kinaev

***Abstract*—Because of stochastic gamma processes adaptability, they are now widely used to mimic a variety of degradation events. However, in certain situations, measurement errors are present in the degradation data, and an intractable probability setting is emerging. When completing inference tasks, its intractableness causes a significant practical difficulty. In order to overcome the difficulty of producing maximum likelihood estimators and the related confidence intervals for the model parameters, we propose a new technique. The rare-event problem, which has a significant influence on the estimator efficiency and, consequently, on the whole inference process, plagues previously employed Monte Carlo approaches for intractable likelihood estimation. We suggest using an alternative Monte Carlo method to address this, while avoiding the establishment of a rare-event issue. The Cross Entropy optimization approach, which can handle objective functions that are tainted by noise, is then added to this technique. We demonstrate that the suggested mix can be implemented within an acceptable computation time and lays the foundation for efficient, generic, and scalable inference processes under the intractable likelihood scenario. Our results show that, given the stochastic gamma process degradation model assumption, the proposed technique may yield high-quality inference results.**

***Index Terms*—Degradation modeling, Stochastic Gamma process, Maximum likelihood parameter estimation, Bootstrap, Monte Carlo, Cross Entropy Method**

I. Introduction

MOST intricate products and systems that fuel our daily activities are prone to decay. This has an important impact on the system's lifespan, the service's quality, and the usage's associated safety. Consequently, the consideration of reliability management and prognostic programs is of utmost significance. In order to understand different failure mechanisms, a reliability engineer requires an effective solution to the statistical inference problems that arise in complex degradation models. Such information eventually improves the reliability of failure predictions and the capacity to take extreme situations into consideration. Furthermore, to increase the *system's safety*, implement the best maintenance *planning regimes*, and reduce *downtimes*, a knowledge of failure mechanisms is of critical importance. Given its adaptability for modeling deterioration, stochastic *gamma processes* (GPs) are at the forefront of the most recent state-of-the-art research [1]–[3].

However, in many real-life problems, practitioners encounter *measurement errors* and the latter is usually accompanied by *intractable likelihoods* [4], [5]. Intractable likelihoods impose a serious challenge [6], [7]. Our study addresses this problem by combining a new *Monte Carlo* (MC) approach and the *Cross Entropy* (CE) optimization method, which is capable of working with objective functions corrupted with noise [8]. This blend allows one to complete the inference task, namely, to deliver both the maximum likelihood *point estimates* and the set of the corresponding *confidence intervals* for the model parameters. The proposed framework provides a novel approach for handling inference tasks subject to the intractable likelihood setting.

There are countless examples of the degradation phenomena in many different fields, and degradation processes have a significant influence on our daily lives. Degradation processes, such as the deterioration of forests, the decline of fish populations, and the deterioration of water quality brought on by pollution, are frequent in ecology [9]–[12]. In manufacturing, it is of interest to consider corrosion, the loss of solar panel production, and aircraft engines [13]–[16]. Additional examples include a micro-mechanical model for initiation of hydrogen embrittlement in high strength steels [17], [18], medicine (shelf life of pharmaceuticals and vaccines) [19], [20], progression of chemical reactions [21], social science (age-related cognitive decline) [22], [23], and many more [24]–[26].

Coastal erosion and *agricultural land deterioration* [27], [28] are two important environmental and economic problems. While coastal erosion is mainly created by natural factors such as ocean waves, agricultural land degradation is caused by various human activities, including the production of food and fibres by the farming sector. Therefore, the development of sustainable risk management plans is of great interest for the protection of the local environment. Additionally, degradation mechanisms are commonly seen in medicine. The Food and Drug Administration in United States specifically mandates that the immediate container label for new drugs include the matching *shelf life* information. Because the active medication components are prone to degradation, this condition is necessary. The mRNA-based immunization is a more recent illustration [20]. Although mRNA vaccines are in a strong position to emerge as the most effective COVID-19 vaccine candidates, mRNA molecules have been observed to spontaneously decay even when stored in the refrigerator. Therefore, it is critical to study the mRNA breakdown process accurately.

1975 saw the introduction of stochastic GPs to the field of system reliability [29], [30]. GPs were recognized as being useful in identifying good inspection and maintenance options.

As a consequence, these processes were discovered to be well suited for characterizing a variety of degradation scenarios [31], [32]. For example, GPs were used to find optimal sand nourishment sizes [33], optimal dike heightening [34], and more [35]–[38]. Stochastic GP modeling is an active field of research. For some recent examples, we refer to wheel wear modeling [39], remaining useful life prediction [40], and space-time-dependent reliability analysis [41]. Due to the importance of GPs in the reliability domain, a natural question of parameter inference comes to light.

For a *perturbed gamma process model*, which combines two stochastic processes: a Brownian motion and a homogeneous gamma process, Bordes et al. [42] provide both the parameters estimation procedure and the corresponding asymptotic properties of the estimators, including consistency and asymptotic normality. In addition, the Expectation–Maximization (EM) algorithm, was applied for less general problems, such as, for example, for GPs with random initial degradation but without measurement errors [3]. In this manuscript, we propose to apply a simpler and more general method, that can handle multi-extremal optimization [43], is easier to implement, and spares the potentially difficult development of the EM algorithm [40], [44]. For additional recent examples, we refer to a Markov Chain Monte Carlo (MCMC) method for parameter estimation of the non-homogeneous GP, which was examined by Guida et al. [45]. In addition, a non-homogeneous GP with measurement errors was considered by Hazra et al. [1]. Moreover, Hazra et al. provided both a Likelihood MCMC (L-MCMC) version of the algorithm, and introduced a novel approach based on *approximate* Bayesian computation (ABC-MCMC). While ABC-MCMC avoids the noisy likelihood estimation problem, which is an essential ingredient of the L-MCMC method, a certain compromise on accuracy should be taken into account.

The major problem with the L-MCMC approach is the usage of the Crude Monte Carlo (CMC) algorithm for intractable likelihood estimation [46]. Specifically, the CMC method introduces large variance. That is, the rare-event setting is introduced by a naive CMC estimator; this will be considered in details in Section II. In this paper, we propose a simple idea that helps to overcome the rare-event problem. In particular, we provide an alternative (CMC_2) estimator, which can be used as an objective function for an optimization procedure with the view of obtaining maximum likelihood estimators (MLEs). In this case, we propose to apply the CE optimization method [47], [48], which has already been shown to perform well under the noisy optimization setting [43]. We give a brief overview of the CE method in Section III.

The major contribution introduced in this study is as follows.

Our first contribution is that we rigorously demonstrate that it is preferable not to use the naive CMC method for the noisy likelihood estimation under the GP model setting, since this approach suffers from the rare-event problem phenomena. Instead, the proposed alternative (CMC_2) estimator avoids the rare-event setting. Our second contribution is that we show that one can apply a well-established noisy optimization CE method (combined with the CMC_2 algorithm) to produce maximum likelihood point estimates of the GP process parameters. Moreover, by blending CE with a bootstrap method, one can easily complete the inference task by providing the corresponding confidence intervals. The alternative Monte Carlo framework of the CMC_2 algorithms opens the way to future research. For example, the proposed CMC_2 estimator can be combined with likelihood-based MCMC methods (such as the L-MCMC) [1]. Finally, a research software package capable of handling real-life degradation inference problems is provided. We show that the software is capable of achieving good results while using a reasonable computation effort.

To summarize, this study proposes an efficient method for handling inference under GP setting with measurement errors. In particular, we present a general technique for obtaining both point estimators and confidence intervals for the GP model parameters. The suggested combination of a simple Monte Carlo method (CMC_2) and the CE algorithm, provides a viable alternative to the recently proposed ABC-MCMC method of Hazra et al. [1]. For synthetically generated datasets, the MLE and the corresponding confidence intervals provide an accurate estimates of the actual model parameters. In addition, we show that our method can work with real-life datasets. One limitation of our study is that the provided research package is single threaded. Therefore, in order to handle larger data sizes, one would need to consider a parallel implementation of the CE method.

This manuscript has the following organization. In Section II, we formally define the inference problem and show that the naive CMC method fails to provide an adequate instrument for the corresponding intractable likelihood estimation. In Section III, we develop the CMC_2 estimator which does not suffer from the rare-event problem, and provide a rigorous description of the CE method. These will be used to obtain MLEs and confidence intervals. Section IV presents an experimental study which demonstrates the performance of the proposed method when applied to both synthetic and real-life degradation data instances. We conclude with Section V, where our findings are summarized and possible directions for future research are outlined.

II. Problem definition

A. The degradation model

In this section, we formally define the degradation problem under consideration. Following Hazra et al. [1], let us consider a system which consists of $m \in \mathbb{N}$ identical components. At time $t = 0$, each component is assumed to have some (random) initial degradation A_i for $1 \leq i \leq m$. With time, the system's health starts to degrade. For the ith component, the true degradation at time t is given by $A_i + X_i(t)$. The measurement error, provided that a measurement for the ith component was performed at time t, is denoted by $Z_i(t)$; we assume that errors are time invariant, that is, that the variance of $Z_i(t)$ does not depend on t. Then, the measured degradation of component i at time t, $Y_i(t)$, is given by:

$$Y_i(t) = A_i + X_i(t) + Z_i(t).$$

We next make the following distribution assumptions. The initial degradation of component i, A_i, is a Gaussian random variable such that $A_i \sim \mathsf{N}(\mu_A, \sigma_A)$, where μ_A and σ_A are the corresponding mean and the standard deviation, respectively. The measurement error $Z_i(t)$ is assumed to be a zero-mean Gaussian random variable such that $Z_i(t) \sim \mathsf{N}(0, \sigma_Z)$, where σ_Z is the standard deviation of the measurement error, and $Z_i(0) \stackrel{\text{def}}{=} 0$.

Following Hazra et al. and Van Noortwijk [1], [30], the true degradation of component i at time t, excluding the initial degradation A_i, namely $X_i(t)$ for $t \geq 0$, is modeled as a general monotonic stochastic process. Here, we utilize a nonstationary GP with a non-linear time-dependent shape function $\alpha(t)$ and a scale parameter β. The function $\alpha(t)$ is assumed to be non-decreasing, right-continuous, and real-valued. Specifically, this paper examines a commonly used power law function $\alpha(t) = \alpha t^{\eta}$, $\alpha > 0$, $\eta > 0$, and $\alpha(0) = 0$ [1]. Suppose that for a component i, $1 \leq i \leq m$, $n_i \in \mathbb{N}$ measurements at times $0 \leq t_{i,1} \leq t_{i,2} \leq \cdots \leq t_{i,n_i}$ are executed. Then, the degradation increments for the ith component (between time $t_{i,j-1}$ and $t_{i,j}$), are defined via $\Delta X_{i,j} = X_i(t_{i,j}) - X_i(t_{i,j-1})$, where $1 \leq j \leq n_i$; it is assumed that $X_i(t_{i,0}) = X_i(0) = 0$. The increment $\Delta X_{i,j}$ is an independent random variable which follows Gamma distribution such that $\Delta X_{i,j} \sim \mathsf{Gamma}(\alpha(t_{i,j}) - \alpha(t_{i,j-1}), \beta)$ for $1 \leq i \leq m$ and $1 \leq j \leq n_i$. The complete problem setting, including the rigorous formulation of probability density functions (PDFs) of random variables, is compactly described in Table I.

B. The inference task and the likelihood function

The inference task is defined as follows. Given the measurement data $\mathbf{Y} = \{Y_i(t) : t \in \{t_{i,1}, \ldots, t_{i,n_i}\}\}_{i=1}^{m}$, find the most plausible estimator for the set of the process parameters $\boldsymbol{\theta} = (\alpha, \eta, \beta, \mu_A, \sigma_A, \sigma_Z)$. Following Hazra et al. [1], we will be concerned with the maximization of the likelihood function $\mathcal{L}(\boldsymbol{\theta}; \Delta\mathbf{y})$, where $\Delta\mathbf{y}$ is the realization of observation increments that is defined via $\Delta\mathbf{y} = (\Delta\mathbf{y}_1, \ldots, \Delta\mathbf{y}_m)^{\top}$. Here, we have that $\Delta\mathbf{y}_i = (\Delta y_{i,1}, \ldots, \Delta y_{i,n_i})^{\top}$, $\Delta y_{i,j} = y_i(t_{i,j+1}) - y_i(t_{i,j})$, $y_i(t_{i,0}) = a_i$ (where a_i is a realization of A_i), and $\Delta Z_{i,j} = Z_i(t_{i,j}) - Z_i(t_{i,j-1})$ for all $1 \leq i \leq m$ and $1 \leq j \leq n_i$. The system components are assumed to be independent and therefore it holds that:

$$\mathcal{L}(\boldsymbol{\theta}; \Delta\mathbf{y}) = \prod_{i=1}^{m} \mathcal{L}(\boldsymbol{\theta}; \Delta\mathbf{y}_i).$$

Note that for all $1 \leq i \leq m$ and $1 \leq j \leq n_i$,

$$\begin{aligned}\Delta Y_{i,j} &= Y_i(t_{i,j}) - Y_i(t_{i,j-1}) = A_i + X_i(t_{i,j}) + Z_i(t_{i,j}) \\ &\quad - (A_i + X_i(t_{i,j-1}) + Z_i(t_{i,j-1})) = \Delta X_{i,j} + \Delta Z_{i,j},\end{aligned}$$

holds. That is, using a vector notation, for any component $1 \leq i \leq m$, one can write: $\Delta\mathbf{Y}_i = \Delta\mathbf{X}_i + \Delta\mathbf{Z}_i$, where $\Delta\mathbf{X}_i = (\Delta X_{i,1}, \ldots, \Delta X_{i,n_i})^{\top}$ and $\Delta\mathbf{Z}_i = (\Delta Z_{i,1}, \ldots, \Delta Z_{i,n_i})^{\top}$. Since $\Delta\mathbf{Y}_i = \Delta\mathbf{X}_i + \Delta\mathbf{Z}_i$, it is of interest to consider the distribution of $\Delta\mathbf{X}_i$ and $\Delta\mathbf{Z}_i$; these are discussed next. The increments of the GP are independent, and thus it holds that [1]:

$$\begin{aligned}f_{\Delta\mathbf{X}_i}(\Delta\mathbf{x}_i) &= \prod_{j=1}^{n_i} f_{\mathsf{G}}(\Delta x_{i,j}; \alpha(t_{i,j}) - \alpha(t_{i,j-1}), \beta) \qquad (1)\\ &= \prod_{j=1}^{n_i} \frac{\beta^{-(\alpha(t_{i,j}) - \alpha(t_{i,j-1}))}}{\Gamma(\alpha(t_{i,j}) - \alpha(t_{i,j-1}))} \Delta x_{i,j}^{\alpha(t_{i,j}) - \alpha(t_{i,j-1}) - 1} \mathrm{e}^{-\frac{\Delta x_{i,j}}{\beta}} \\ &= \prod_{j=1}^{n_i} \frac{\beta^{-\left(\alpha t_{i,j}^{\eta} - \alpha t_{i,j-1}^{\eta}\right)}}{\Gamma\left(\alpha t_{i,j}^{\eta} - \alpha t_{i,j-1}^{\eta}\right)} \Delta x_{i,j}^{\alpha t_{i,j}^{\eta} - \alpha t_{i,j-1}^{\eta} - 1} \mathrm{e}^{-\frac{\Delta x_{i,j}}{\beta}} \mathbb{1}_{\{\Delta x_{i,j} \geq 0\}}.\end{aligned}$$

To derive the PDF $f_{\Delta\mathbf{Z}_i}(\Delta\mathbf{z}_i)$, we follow [46] and note that $\mathbf{Z}_i = (Z_i(t_{i,1}), \ldots, Z_i(t_{i,n_i}))^{\top}$ has the multivariate Gaussian distribution with mean zero and covariance matrix $\Sigma_i = \sigma_Z^2 \mathbf{I}_{n_i \times n_i}$, where $\mathbf{I}_{n_i \times n_i}$ is the identity matrix of size $n_i \times n_i$. In addition, $\Delta\mathbf{Z}_i$ is a linear transformation of $\mathbf{Z}_i$, such that $\Delta\mathbf{Z}_i = \boldsymbol{J}_i \mathbf{Z}_i$, where

$$\boldsymbol{J}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & \cdots & \cdots & -1 & 1 \end{pmatrix}_{n_i \times n_i},$$

is an $n_i \times n_i$ matrix. Therefore, $\Delta\mathbf{Z}_i$ follows a multivariate Gaussian distribution with mean zero and covariance matrix $\Sigma_{\Delta\mathbf{Z}_i} = \boldsymbol{J}_i \Sigma_i \boldsymbol{J}_i^{\top}$, where $\Sigma_{\Delta\mathbf{Z}_i}$ is an $n_i \times n_i$ dimensional matrix [8][Theorem 3.6]. Specifically, we have that

$$\Sigma_{\Delta\mathbf{z}_i} = 2\sigma_Z^2 \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & 0 & \cdots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \cdots & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & \cdots & 0 & -\frac{1}{2} & 1 \end{pmatrix}_{n_i \times n_i}, \qquad (2)$$

and, $\Delta\mathbf{Z}_i \sim \mathsf{MVN}(\mathbf{0}, \Sigma_{\Delta\mathbf{Z}_i})$ holds. Finally, since $\Delta\mathbf{X}_i$ and $\Delta\mathbf{Z}_i$ are independent, the likelihood function for the ith component data is a convolution integral which takes the form

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}; \Delta\mathbf{y}_i) = \int_{a_i} \int_{\Delta\mathbf{z}_i} & f_{\Delta\mathbf{X}_i}(\Delta\mathbf{y}_i - \Delta\mathbf{z}_i | a_i) f_{\mathsf{N}}(a_i) \\ & f_{\Delta\mathbf{Z}_i}(\Delta\mathbf{z}_i)\, da_i\, d\Delta\mathbf{z}_i,\end{aligned}$$

and we conclude that

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}; \Delta\mathbf{y}) = \prod_{i=1}^{m} \int_{a_i} \int_{\Delta\mathbf{z}_i} & f_{\Delta\mathbf{X}_i}(\Delta\mathbf{y}_i - \Delta\mathbf{z}_i | a_i) f_{\mathsf{N}}(a_i) \qquad (3) \\ & f_{\Delta\mathbf{Z}_i}(\Delta\mathbf{z}_i)\, da_i\, d\Delta\mathbf{z}_i.\end{aligned}$$

This likelihood involves the calculation of $m + \sum_{i=1}^{m} n_i$ dimensional integral, and consequently, the computational intractability suggests a Monte Carlo approach, which is discussed in Section II-C.

C. The Crude Monte Carlo estimator for the intractable gamma process likelihood

A straight-forward CMC approach is almost an immediate consequence of (3). Specifically, let C_i be a random variable,

TABLE I The nomenclature.

Notation	Meaning
$m \in \mathbb{N}$	number of identical components
$n_i \in \mathbb{N}$	number of measurements performed for component i
$A_i \in \mathbb{R}$	initial degradation of the ith component
$Y_i(t) \in \mathbb{R}$	observed degradation measurement of the ith component at time t
$X_i(t) \in \mathbb{R}^+$	true degradation of the ith component at time t (excluding A_i)
$Z_i(t) \in \mathbb{R}$	time invariant measurement error for the ith component at time t, where $Z_i(0) \stackrel{\text{def}}{=} 0$
$\mathbf{Y} = \{Y_i(t) : t \in \{t_{i,j}\}_{j=1}^{n_i}\}_{i=1}^{m}$	observed degradation data with measurement errors
$\Delta Z_{i,j} = Z_i(t_{i,j}) - Z_i(t_{i,j-1})$	measurement error increment of the ith component for $1 \le j \le n_i$
$\Delta X_{i,j} = X_i(t_{i,j}) - X_i(t_{i,j-1})$	true degradation increment of the ith component for $1 \le j \le n_i$
$\Delta Y_{i,j} = Y_i(t_{i,j}) - Y_i(t_{i,j-1})$	observation increment of the ith component for $1 \le j \le n_i$
$\Delta\mathbf{Y} = (\Delta\mathbf{Y}_1, \ldots, \Delta\mathbf{Y}_m)^\top$	observed increments
$\Delta\mathbf{Y}_i = (\Delta Y_{i,1}, \ldots, \Delta Y_{i,n_i})^\top$	observed increments of the ith component
$\alpha(t) = \alpha t^\eta$	non-decreasing, right-continuous, time-dependent shape function
	(corresponds to Gamma PDF shape parameter)
$\beta > 0$	scale parameter (corresponds to Gamma PDF)
$W \sim \mathsf{N}(\mu, \sigma)$	univariate Gaussian (normal) distribution
	$f_{\mathsf{N}}(w; \mu, \sigma) = (2\pi\sigma^2)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\left(\frac{w-\mu}{\sigma}\right)^2}, \quad w \in \mathbb{R}$
$\mathbf{W} \sim \mathsf{MVN}(\boldsymbol{\mu}, \Sigma_{d\times d})$	multivariate Gaussian (normal) distribution
	$f_{\mathsf{MVN}}(\mathbf{w}; \boldsymbol{\mu}, \Sigma_{d\times d}) = (2\pi)^{-\frac{k}{2}} \lvert\Sigma_{d\times d}\rvert^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})^\top \Sigma_{d\times d}^{-1}(\mathbf{w}-\boldsymbol{\mu})}, \ \mathbf{w} \in \mathbb{R}^d$
$W \sim \mathsf{Gamma}(\alpha, \beta)$	Gamma distribution
	$f_{\mathsf{G}}(w; \alpha, \beta) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)} w^{\alpha-1} \mathrm{e}^{-\frac{w}{\beta}} \mathbb{1}_{\{w\ge 0\}}$, where $\mathbb{1}$ is the indicator function
$\boldsymbol{\theta} = (\alpha, \eta, \beta, \mu_A, \sigma_A, \sigma_Z)$	the set of process parameters
$\mathcal{L}(\boldsymbol{\theta}; \Delta\mathbf{y})$	the likelihood function

which is defined via

$$C_i = \prod_{j=1}^{n_i} \frac{\beta^{-\left(\alpha t_{i,j}^\eta - \alpha t_{i,j-1}^\eta\right)}}{\Gamma\left(\alpha t_{i,j}^\eta - \alpha t_{i,j-1}^\eta\right)} \Delta X_{i,j}^{\alpha t_{i,j}^\eta - \alpha t_{i,j-1}^\eta - 1} \mathrm{e}^{-\frac{\Delta X_{i,j}}{\beta}} \mathbb{1}_{\{\Delta X_{i,j} \ge 0\}}, \tag{4}$$

where $\Delta X_{i,j} = \Delta y_{i,j} - \Delta Z_{i,j}$. Note that $\Delta y_{i,j}$ is observed, and that $\Delta Z_{i,j}$ for $1 \le j \le n_i$, can be drawn from $\mathsf{MVN}(\mathbf{0}, \Sigma_{\Delta\mathbf{Z}_i})$. Then, it holds that

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}; \Delta\mathbf{y}) &\underbrace{=}_{(3)} \prod_{i=1}^{m} \int_{a_i} \int_{\Delta\mathbf{z}_i} f_{\Delta\mathbf{X}_i}(\Delta\mathbf{y}_i - \Delta\mathbf{z}_i | a_i) \\ &\qquad f_{\mathsf{N}}(a_i) f_{\Delta\mathbf{Z}_i}(\Delta\mathbf{z}_i)\, da_i\, d\Delta\mathbf{z}_i \\ &= \prod_{i=1}^{m} \mathbb{E}_{A_i, \Delta\mathbf{Z}_i}[C_i] = \mathbb{E}_{\{A_i, \Delta\mathbf{Z}_i\}_{i=1}^m}\left[\prod_{i=1}^{m} C_i\right],\end{aligned} \tag{5}$$

where the last equality follows from the independence of C_is for $1 \le i \le m$. Consequently, (5) suggests a CMC method for the estimation of $\mathcal{L}(\boldsymbol{\theta}; \Delta\mathbf{y})$. Specifically, let $N \in \mathbb{N}$ be a predefined sample size, and let $A_i \sim \mathsf{N}(\mu_A, \sigma_A)$ and $\Delta\mathbf{Z}_i = (\Delta Z_{i,1}, \ldots, \Delta Z_{i,n_i}) \sim \mathsf{MVN}(\mathbf{0}, \Sigma_{\Delta\mathbf{Z}_i})$, where $\Sigma_{\Delta\mathbf{Z}_i}$ is given in (2), be N independent realizations of A and $\Delta\mathbf{Z}$. Finally, by defining

$$C_i^{(k)} \leftarrow \prod_{j=1}^{n_i} \frac{\beta^{-\left(\alpha t_{i,j}^\eta - \alpha t_{i,j-1}^\eta\right)}}{\Gamma\left(\alpha t_{i,j}^\eta - \alpha t_{i,j-1}^\eta\right)} \Delta X_{i,j}^{\alpha t_{i,j}^\eta - \alpha t_{i,j-1}^\eta - 1} \mathrm{e}^{-\frac{\Delta X_{i,j}}{\beta}} \mathbb{1}_{\{\Delta X_{i,j} \ge 0\}},$$

where $\Delta Y_{i,j} = y_i(t_{i,j}) - y_i(t_{i,j-1})$ for $1 \le j \le n_i$ ($\Delta y_{i,1} = y_{i,1} - A_i$) and $\Delta\mathbf{X}_i = (\Delta X_{i,1}, \ldots, \Delta X_{i,n_i})^\top$; here: $(\Delta Y_{i,1}, \ldots, \Delta Y_{i,n_i})^\top - (\Delta Z_{i,1}, \ldots, \Delta Z_{i,n_i})^\top$, the $\widehat{\mathcal{L}}_{\mathrm{CMC}}(\boldsymbol{\theta}; \Delta\mathbf{y})$ estimator is given by:

$$\widehat{\mathcal{L}}_{\mathrm{CMC}}(\boldsymbol{\theta}; \Delta\mathbf{y}) = \frac{1}{N} \sum_{k=1}^{N} \prod_{i=1}^{m} C_i^{(k)}. \tag{6}$$

The major issue with $\widehat{\mathcal{L}}_{\mathrm{CMC}}(\boldsymbol{\theta};\Delta\mathbf{y})$, is the so-called rare-event problem, which was also noted by Lu et al. [46]. In particular, provided that at least one of $\Delta X_{i,j}$s for $1 \leq j \leq n_i$ is negative (which happens if $\Delta Y_{i,j} < \Delta Z_{i,j}$ holds), we have that $C_i^{(k)} = 0$ and that $C^{(k)} = \prod_{i=1}^m C_i^{(k)} = 0$. When the event $\{C^{(k)} > 0\}$ is relatively non-frequent, the efficiency of (6) starts to deteriorate due to the increase in the variance of $\widehat{\mathcal{L}}_{\mathrm{CMC}}(\boldsymbol{\theta};\Delta\mathbf{y})$; please refer to Section II-D for a rigorous discussion. This setting, which is common to many hard estimation tasks, corresponds to the rare event problem in Monte Carlo [49].

D. The rare-event problem

Consider the below context to have a better grasp of the rare-event phenomena [7]. Suppose that $\mathcal{X}$ and $\mathcal{X}^* \subseteq \mathcal{X}$ are sets and say that we are interested in the estimation of the probability $\ell = |\mathcal{X}^*|/|\mathcal{X}|$; here, we assume that $|\mathcal{X}^*| \ll |\mathcal{X}|$. Let X be a random variable and suppose that X is uniformly distributed over the $\mathcal{X}$ set. In this case, $\ell = \mathbb{E}[\mathbb{1}_{\{X\in\mathcal{X}^*\}}]$, and the corresponding estimator is: $\widehat{\ell}_{\mathrm{CMC}} = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{X_i\in\mathcal{X}^*\}}$. The collection $\{\mathbb{1}_{\{X_i\in\mathcal{X}^*\}}\}_{i=1}^N$s is the set of independent and identically distributed (i.i.d.) random variables, where each such variable is Bernoulli distributed. Thus, $\widehat{\ell}_{\mathrm{CMC}}$ is an unbiased estimator of ℓ. Moreover, the variance $\mathrm{Var}\left(\widehat{\ell}_{\mathrm{CMC}}\right)$ is: $\mathrm{Var}\left(\widehat{\ell}_{\mathrm{CMC}}\right) = \frac{\ell(1-\ell)}{N}$.

The efficiency, of the CMC estimator, is examined via the relative error (RE) [50] and $\widehat{\ell}_{\mathrm{CMC}}$'s RE is:

$$\mathrm{RE} = \mathrm{RE}\left(\widehat{\ell}_{\mathrm{CMC}}\right) = \frac{\sqrt{\mathrm{Var}\left(\widehat{\ell}_{\mathrm{CMC}}\right)}}{\mathbb{E}\left(\widehat{\ell}_{\mathrm{CMC}}\right)} = \sqrt{\ell(1-\ell)/N}/\ell.$$

Since rare-event setting implies $\ell \ll 1$, it holds that:

$$\mathrm{RE} \approx (N\ell)^{-0.5}. \tag{7}$$

Equation (7) sets a serious computational problem. For example, consider the (rare-event) probability $\ell \approx 10^{-12}$, and suppose that we are concerned in a humble RE of 10%. Then, (7) implies that the necessary number of samples N is about 10^{14} [7]. We next show that (6) suffers from the rare-event setting. In particular, we show that this CMC estimator is not suitable to handle certain parameter ranges.

The random variable C_i in (4) is greater than zero if $\Delta X_{i,j}$s are non-negative for all $1 \leq j \leq n_i$. Since $\Delta Y_{i,j} = \Delta X_{i,j} + \Delta Z_{i,j}$, it holds that $\prod_{i=1}^m C_i > 0$ if $\Delta Z_{i,j} \leq \Delta Y_{i,j}$ for all $1 \leq i \leq m$ and $1 \leq j \leq n_i$. Next, we provide an upper bound for the probability of the event $\{\prod_{i=1}^m C_i > 0\}$. For simplicity, suppose that $A_i = 0$ and that n_i is even for all $1 \leq i \leq m$. Then,

$$\begin{aligned}\mathbb{P}\left(\prod_{i=1}^m C_i > 0\right) &= \mathbb{P}(\cap_{1\leq i\leq m,\, 1\leq j\leq n_i}\{\Delta Z_{i,j} \leq \Delta Y_{i,j}\}) \qquad (8)\\ &\leq \prod_{i=1}^m \prod_{j=1}^{n_i/2} \mathbb{P}(\Delta Z_{i,2j} \leq \Delta Y_{i,2j}),\end{aligned}$$

where the last inequality follows from the fact that $\Delta Z_{i,2j}$ for $1 \leq j \leq n_i/2$, namely $(\Delta Z_{i,2} = Z_i(t_{i,2}) - Z_i(t_{i,1}), \Delta Z_{i,4} = Z_i(t_{i,4}) - Z_i(t_{i,3}), \ldots, \Delta Z_{i,n_i} = Z_i(t_{i,n_i}) - Z_i(t_{i,n_i-1}))$, are independent random variables. Since $\Delta Z_{i,2j}$ is a difference between two univariate Gaussian random variables each having mean zero and standard deviation σ_Z, it holds that $\Delta Z_{i,2j} \sim \mathsf{N}(0, \sqrt{2}\sigma_Z)$. By defining $\overline{\Delta Y_i} = \max\{\Delta Y_{i,2j}\}_{j=1}^{n_i/2}$, $\overline{\Delta Y} = \max\{\overline{\Delta Y_i}\}_{i=1}^m$, and combining this with (8), we arrive at

$$\begin{aligned}\mathbb{P}\left(\prod_{i=1}^m C_i > 0\right) &\leq \prod_{i=1}^m \prod_{j=1}^{n_i/2} \mathbb{P}(\Delta Z_{i,2j} \leq \Delta Y_{i,2j}) \qquad (9)\\ &\leq \prod_{i=1}^m \prod_{j=1}^{n_i/2} \mathbb{P}(\Delta Z_{i,2j} \leq \overline{\Delta Y})\\ &= \left[\mathbb{P}\left(W \leq \frac{\overline{\Delta Y}}{\sqrt{2}\sigma_Z}\right)\right]^{\sum_{i=1}^m n_i/2}\\ &\leq \left[\mathbb{P}\left(W \leq \frac{|\overline{\Delta Y}|}{\sqrt{2}\sigma_Z}\right)\right]^{\sum_{i=1}^m n_i/2},\end{aligned}$$

where $W \sim \mathsf{N}(0,1)$. Thus, we can conclude that the probability that $C^{(k)} = 0$ satisfies:

$$\begin{aligned}\mathbb{P}\left(\prod_{i=1}^m C_i = 0\right) &= 1 - \mathbb{P}\left(\prod_{i=1}^m C_i > 0\right)\\ &\geq 1 - \left[\mathbb{P}\left(W \leq \frac{|\overline{\Delta Y}|}{\sqrt{2}\sigma_Z}\right)\right]^{\sum_{i=1}^m n_i/2}.\end{aligned}$$

When $\sigma_Z \to \infty$, $\frac{|\overline{\Delta Y}|}{\sqrt{2}\sigma_Z} \to 0$, and, if it also holds that $n_i \to \infty$, we arrive at $\mathbb{P}\left(\prod_{i=1}^m C_i = 0\right) \to 1 - 0.5^{\sum_{i=1}^m n_i/2} \to 1$. While this result is asymptotic, the situation can quickly become quite dreadful in practice. Consider a data degradation example with $m = 20$ components and $n_i = 10$ examination times. Then, for $\sigma_Z \geq \left|\overline{\Delta Y}\right|$, we have that

$$\begin{aligned}\mathbb{P}\left(\prod_{i=1}^m C_i = 0\right) &\geq 1 - \Phi\left(\frac{|\overline{\Delta Y}|}{\sqrt{2}\sigma_Z}\right)^{20\times 5}\\ &\geq 1 - \Phi\left(1/\sqrt{2}\right)^{20\times 5} \approx 1 - 1.246\times 10^{-12},\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard univariate Gaussian random variable with mean 0 and standard deviation 1. That is, the probability to get a non zero product of C_is is a rare event which happens with probability 1.246×10^{-12}, thus, the corresponding CMC estimator is prohibitive from the computation point of view. In order to overcome the rare-event problem, Buist [7] proposed to apply the Sequential Monte Carlo approach. While this helps to achieve the desired variance minimization, the overall procedure is computationally demanding. In this work, we propose a simpler Monte Carlo estimator which avoids the above rare-event setting. This estimator, combined with the CE optimization method which is capable of working with noisy objective functions, opens the way for efficient inference procedure in GPs with intractable likelihood. The details are provided in Section III.

III. Methods

A. An alternative Crude Monte Carlo estimator

We start with the development of a simple alternative to the CMC estimator from (6). To do so, consider the joint distribution of $\Delta\mathbf{X}_i$, $\Delta\mathbf{Y}_i$ and A_i. The degradation increments

are independent of the initial degradation, and thus it holds that:

$$f_{\Delta\mathbf{X}_i,\Delta\mathbf{Y}_i,A_i}(\Delta\mathbf{x}_i,\Delta\mathbf{y}_i,a_i) = f_{\Delta\mathbf{Y}_i|\Delta\mathbf{X}_i,A_i}(\Delta\mathbf{y}_i|\Delta\mathbf{x}_i,a_i) \quad (10)$$
$$f_{\Delta\mathbf{X}_i|A_i}(\Delta\mathbf{x}_i|a_i)f_{A_i}(a_i)$$
$$= f_{\Delta\mathbf{Y}_i|\Delta\mathbf{X}_i,A_i}(\Delta\mathbf{y}_i|\Delta\mathbf{x}_i,a_i)f_{\Delta\mathbf{X}_i}(\mathbf{x}_i)f_{A_i}(a_i).$$

Since $\Delta\mathbf{Y}_i = \Delta\mathbf{X}_i + \Delta\mathbf{Z}_i$ for all $1 \leq i \leq m$, where $\Delta\mathbf{Z}_i \sim \mathsf{MVN}(\mathbf{0}, \Sigma_{\Delta\mathbf{Z}_i})$, it holds (again, via [8][Theorem 3.6]), that $\Delta\mathbf{Y}|A_i, \Delta\mathbf{X}_i \sim \mathsf{MVN}(\Delta\mathbf{X}_i, \Sigma_{\Delta\mathbf{Z}_i})$. That is, the marginal distribution of $\Delta\mathbf{Y}_i$ satisfies:

$$f_{\Delta\mathbf{Y}_i}(\Delta\mathbf{y}_i) = \int_{a_i}\int_{\Delta\mathbf{x}_i} f_{\Delta\mathbf{X}_i,\Delta\mathbf{Y}_i,A_i}(\Delta\mathbf{x}_i,\Delta\mathbf{y}_i,a_i)da_i d\Delta\mathbf{x}_i$$
$$\underbrace{=}_{(10)} \int_{a_i}\int_{\Delta\mathbf{x}_i} f_{\Delta\mathbf{Y}_i|\Delta\mathbf{X}_i,A_i}(\Delta\mathbf{y}_i|\Delta\mathbf{x}_i,a_i)f_{A_i}(a_i)f_{\Delta\mathbf{X}_i}(\mathbf{x}_i)da_i d\Delta\mathbf{x}_i$$
$$= \mathbb{E}_{A_i,\Delta\mathbf{X}_i}\left[f_{\Delta\mathbf{Y}_i|\Delta\mathbf{X}_i,A_i}(\Delta\mathbf{y}_i|\Delta\mathbf{X}_i,A_i)\right], \quad (11)$$

where $f_{\Delta\mathbf{Y}_i|\Delta\mathbf{X}_i,A_i}(\Delta\mathbf{y}_i|\Delta\mathbf{X}_i,A_i)$ is the PDF of the multivariate Gaussian distribution with mean $\Delta\mathbf{X}_i$ and the covariance matrix $\Sigma_{\Delta\mathbf{Z}_i}$. That is, one can conclude that

$$\mathcal{L}(\boldsymbol{\theta};\Delta\mathbf{y}) = \prod_{i=1}^{m}\mathcal{L}(\boldsymbol{\theta};\Delta\mathbf{y}_i) = \prod_{i=1}^{m} f_{\Delta\mathbf{Y}_i}(\Delta\mathbf{y}_i) \quad (12)$$
$$= \prod_{i=1}^{m}\mathbb{E}_{A_i,\Delta\mathbf{X}_i}\left[f_{\Delta\mathbf{Y}_i|\Delta\mathbf{X}_i,A_i}(\Delta\mathbf{y}_i|\Delta\mathbf{X}_i,A_i)\right]$$
$$= \mathbb{E}_{\{A_i,\Delta\mathbf{X}_i\}_{i=1}^m}\left[\prod_{i=1}^{m} f_{\Delta\mathbf{Y}_i|\Delta\mathbf{X}_i,A_i}(\Delta\mathbf{y}_i|\Delta\mathbf{X}_i,A_i)\right]$$
$$= \mathbb{E}_{\{A_i,\Delta\mathbf{X}_i\}_{i=1}^m}\left[\prod_{i=1}^{m} W_i\right],$$

where W_i is the value of the multivariate Gaussian PDF with mean $\Delta\mathbf{X}_i$ and the covariance matrix $\Sigma_{\Delta\mathbf{Z}_i}$ evaluated at $\Delta\mathbf{Y}_i$ (please see line 6 of Algorithm 1). The last equality in (12) follows from the independence of W_is for $1 \leq i \leq m$. The corresponding alternative Monte Carlo method which follows immediately from (12), and from the fact that we can easily sample from $f_{A_i}(a_i)$ and from $f_{\Delta\mathbf{X}_i}(\mathbf{x}_i)$, is summarized in Algorithm 1.

Algorithm 1 avoids the rare-event problem present in (6), since $W_i^{(k)}$ in line 6 of Algorithm 1 is never equal to zero for all $1 \leq i \leq m$. This desirable property makes the $\widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\boldsymbol{\theta};\Delta\mathbf{y})$ estimator to be a suitable candidate for the noisy MLE optimization task. The latter will be performed via the CE optimization procedure which is discussed in Section III-B.

Algorithm 1: The alternative CMC algorithm (CMC_2) for the estimation of $\mathcal{L}(\boldsymbol{\theta};\Delta\mathbf{y})$

Input: m — the number of components, $\{n_i\}_{i=1}^m$ — the number of measurements for each component, $\{\{t_{i,j}\}_{j=1}^{n_i}\}_{i=1}^m$ — measurement times, a vector of parameters $\boldsymbol{\theta} = (\alpha, \eta, \beta, \mu_A, \sigma_A, \sigma_Z)$, the observed data $\mathbf{y} = \{y_i(t) : t \in \{t_{i,1}, \ldots, t_{i,n_i}\}\}_{i=1}^m$, and a sample size $N \in \mathbb{N}$

Output: A CMC estimator $\widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\boldsymbol{\theta};\Delta\mathbf{y})$ of $\mathcal{L}(\boldsymbol{\theta};\Delta\mathbf{y})$

1 **for** $k \leftarrow 1$ **to** N **do**
2 **for** $i \leftarrow 1$ **to** m **do**
3 Draw $A_i \sim \mathsf{N}(\mu_A, \sigma_A)$
4 Calculate $\Delta\mathbf{Y}_i = (\Delta Y_{i,1}, \ldots, \Delta Y_{i,n_i})^\top$; here $\Delta Y_{i,j} = y_i(t_{i,j}) - y_i(t_{i,j-1})$ for $1 \leq k \leq n_i$, where $\Delta Y_{i,1} = y_{i,1} - A_i$
5 Draw $\Delta\mathbf{X}_i = (\Delta X_{i,1}, \ldots, \Delta X_{i,n_i})$, where $\Delta X_{i,j} \sim \mathsf{Gamma}(\alpha(t_{i,j}) - \alpha(t_{i,j} - t_{i,j-1}), \beta))$
6 $W_i^{(k)} \leftarrow f_{\mathsf{MVN}}(\Delta\mathbf{Y}_i; \Delta\mathbf{X}_i, \Sigma_{\Delta\mathbf{Z}_i})$, where $\Sigma_{\Delta\mathbf{Z}_i}$ is given in (2)
7 **end**
8 $W^{(k)} \leftarrow \prod_{i=1}^m W_i^{(k)}$
9 **end**
10 $\widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\boldsymbol{\theta};\Delta\mathbf{y}) = \frac{1}{N}\sum_{k=1}^N W^{(k)}$
11 **return** $\widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\boldsymbol{\theta};\Delta\mathbf{y})$

B. The Cross Entropy method

The CE technique is a sequential procedure used to progressively alter a random search's sampling distribution such that the best result is more likely to appear throughout subsequent phases of the algorithm's execution. However, it is noteworthy that the CE approach makes use of strict information theory and stochastic simulation principles [51]–[53], setting it apart from similar evolutionary computation methods. The CE method may be used for a wide range of estimation and optimization tasks, making it extremely flexible. In particular, CE is useful for discrete, continuous, and noisy optimization, and rare-event estimation [48].

Recall that we are interested in the maximum likelihood estimator and thus, under our setting, we deal with noisy continuous optimization, since we only have access to an estimator of $\mathcal{L}(\boldsymbol{\theta};\Delta\mathbf{y})$. To set the stage, consider a general unconstrained optimization problem: $\max_{\mathbf{x}\in\mathcal{X}} S(\mathbf{x})$, where $S : \mathcal{X} \to \mathbb{R}$ is a fitness (or an objective) function. Let

$$\mathcal{X}^* = \{\mathbf{x} \in \mathcal{X} : S(\mathbf{x}) = \max_{\mathbf{x}\in\mathcal{X}} S(\mathbf{y}),\ \mathbf{y} \in \mathcal{X}\},$$

be the set of optimal solutions. A CE framework for *continuous optimization* is depicted in Fig. 1. For the *discrete optimization* setting, we refer to [54].

We closely follow [54] and consider the CE algorithm that will be derived from Fig. 1. The CE method begins with the *initialization* step in which a PDF $g_1(\mathbf{x})$ for $\mathbf{X} \in \mathcal{X}$ is defined. Since $\mathcal{X}$ is a continuous space, a logical choice for $g_1(\mathbf{x})$ can be, for example, a (truncated) Gaussian, or a continuous uniform distribution over the set $\mathcal{X}$. During the initialization step, we also set the $(1-\rho)$-th quantile of the fitness function at time zero — γ_0, to minus infinity and initialize the iteration counter t. In the next step, we calculate the $(1-\rho)$-th quantile of the fitness function $Y = S(\mathbf{X})$ at time t; note that $\mathbf{X} \sim g_t(\mathbf{x})$ and assume that the CDF of Y is $F_Y^{(t)}(y)$; here, t is the current iteration counter. Using the obtained $(1-\rho)$-th quantile, we proceed to the *update the sampling distribution* step. In order to increase the frequency of sampling of better solutions, we skip the specifics of the update phase at this point (they will be covered below). However, the modified distribution will be employed in the

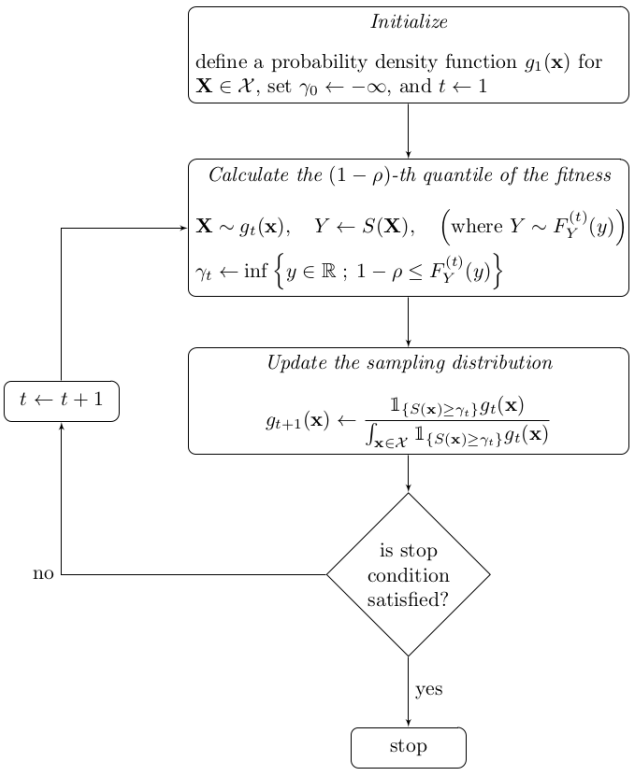


Fig. 1: The Cross Entropy continuous optimization framework.

following iteration. When a predetermined stopping condition is satisfied, the CE procedure ends. For example, since we deal with noisy optimization, one might want to stop when the γ_t quantile does not change considerably for several iterations.

Under some (not very realistic) assumptions, namely, provided that $S(\mathbf{X})$ can be calculated without noise, and provided that an infinite number of samples can be obtained from $\mathbf{X} \sim g_t(\mathbf{x})$, one can observe the following. In every iteration (t), the fitness function value of $\mathbf{X} \sim g_t(\mathbf{x})$ satisfies $S(\mathbf{X}) \ge \gamma_{t-1}$. Moreover, the sequence $(\gamma_0, \gamma_1, \ldots)$ is strictly increasing. That is, the procedure will eventually converge to some $\mathbf{x} \in \mathcal{X}^*$. However, in practice, we neither have access to the infinite sample size, nor exact $S(\mathbf{X})$ value is available. That is, from the practical point of view, we will need to resolve several issues with the framework from Fig. 1; these are detailed next. First of all, the $\mathcal{X}$ space is infinite, and thus, *exact* computation of the $(1-\rho)$-th quantile is infeasible. Moreover, as mentioned above, $S(\mathbf{x})$ is not available analytically, namely, we can only hope to achieve an estimate of $S(\mathbf{x})$. Finally, the third issue is the sampling from the PDF $g_t(\mathbf{x})$ for all $t > 1$.

The quantile problem is not very hard to resolve. Specifically, provided that a finite sample from $g_t(\mathbf{x})$ is available, an approximate fitness quantile (or, in other words, an $(1-\rho)$-th *sample fitness quantile*) $\widehat{\gamma}_t$, can be readily obtained from $\{S(\mathbf{X}_i)\}_{i=1}^N$, where $\mathbf{X}_l \sim g_t(\mathbf{x})$ for $1 \le l \le N$. While $S(\mathbf{x})$ is not available analytically, we further assume that an estimate of $S(\mathbf{x})$ can be obtained using a Monte Carlo algorithm. The sampling from $g_t(\mathbf{x})$ can be a very hard problem. It will be rigorously addressed in Section III-C. We proceed with the definition of a rigorous CE procedure, while providing solutions to the above-mentioned problems.

C. Cross Entropy method for noisy maximum likelihood optimization

Under our setting, define $\mathcal{X} = \mathbb{R}^k$, where $k = |\boldsymbol{\theta}|$. That is, $\mathbf{x} = (x_1, \ldots, x_k)^\top \in \mathcal{X}$, and let the objective function be $S(\mathbf{x}) = \mathcal{L}_{\mathrm{CMC}_2}(\boldsymbol{\theta}; \Delta\mathbf{y})$. Please note that we do not have access to the exact value of $\mathcal{L}(\boldsymbol{\theta}; \Delta\mathbf{y})$, and thus, an estimator $S(\mathbf{x})$ is used instead. Consider now the definition of the probability distribution of the random variable $\mathbf{X} = (X_1, \ldots, X_k)^\top$, where $X_i \in \mathbb{R}$ for $i = 1, \ldots, k$. The sampling from the PDF defined in the *Update the sampling distribution* phase in Fig. 1, namely, from

$$g_{t+1}(\mathbf{x}) = \frac{\mathbb{1}_{\{S(\mathbf{x})\ge\gamma_t\}} g_t(\mathbf{x})}{\sum_{\mathbf{x}\in\mathcal{X}} \mathbb{1}_{\{S(\mathbf{x})\ge\gamma_t\}} g_t(\mathbf{x})}, \tag{13}$$

is generally hard. In order to resolve this problem, we approximate (13) using a parametric family

$$f(\mathbf{x}; \boldsymbol{\mu}_{t+1}, \boldsymbol{\sigma}_{t+1}) = \prod_{d=1}^{k} \left(2\pi\sigma_{t+1,i}^2\right)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\sigma_{t+1,d}^2}\left(x_i - \mu_{t+1,d}\right)^2}, \tag{14}$$

where $\boldsymbol{\mu}_{t+1} = (\mu_{t+1,1}, \ldots, \mu_{t+1,k})$, and $\boldsymbol{\sigma}_{t+1} = (\sigma_{t+1,1}, \ldots, \sigma_{t+1,k})$. The PDF in (14) is a joint PDF of k independent univariate Gaussian random variables. That is, in order to approximate the sequence of PDFs $\{g_t\}_{t\in\mathbb{N}\setminus\{0\}}$ in (13), we need to obtain the corresponding parameter vectors $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$ for all $t \in \mathbb{N} \setminus \{0\}$. As soon as these parameters are available, the sampling from (14) is straight-forward, and we can proceed to the resolution of CE issues from the end of Section III-B. Specifically, we show how to obtain the sample fitness quantile, and the sequence of parameters $\{\boldsymbol{\mu}_t, \boldsymbol{\sigma}_t\}_{t\in\mathbb{N}\setminus\{0\}}$.

The calculation of sample quantiles. Provided that $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$ are readily available, the calculation of $\widehat{\gamma}_t$ is not very hard. In particular, given a sample of independent random variables $\{\mathbf{X}_l\}_{l=1}^N$, such that $\mathbf{X}_l \sim f(\mathbf{x}; \boldsymbol{\mu}_t, \boldsymbol{\sigma}_t)$ for $1 \le l \le N$, and the corresponding set $\{S(\mathbf{X}_l)\}_{l=1}^N$, let $S_{(1)} \le \cdots \le S_{(N)}$ be the ascending ordering of the $\{S(\mathbf{X}_l)\}_{l=1}^N$ set. Then, the sample fitness $(1-\rho)$-th quantile $\widehat{\gamma}_t$ is equal to $S_{(\lceil N\times(1-\rho)\rceil)}$.

The approximation of the sampling PDF. Here, our objective is to approximate the sampling PDF (13) via $f(\mathbf{x}; \boldsymbol{\mu}_{t+1}, \boldsymbol{\sigma}_{t+1})$, where $f(\mathbf{x}; \boldsymbol{\mu}_{t+1}, \boldsymbol{\sigma}_{t+1})$ belongs to the parametric family defined in (14). In other words, we aim to recover $\boldsymbol{\mu}_{t+1}$ and $\boldsymbol{\sigma}_{t+1}$, using samples from iteration t. This task is completed by minimizing the relative entropy (divergence of Kullback-Leibler) of $f(\mathbf{x}; \boldsymbol{\mu}_{t+1}, \boldsymbol{\sigma}_{t+1})$ with respect to $g_{t+1}(\mathbf{x})$ [51]. Definition 1 provides the formal characterization of the relative entropy.

Definition 1 (Relative entropy)**.** *The relative entropy of a PDF $f(\cdot)$ with respect to a PDF $g(\cdot)$ is given by:*

$$\begin{aligned}\mathcal{D}(g, f) &= \mathbb{E}_g \ln\left(\frac{g(\mathbf{X})}{f(\mathbf{X})}\right) = \int \ln\left(\frac{g(\mathbf{x})}{f(\mathbf{x})}\right) g(\mathbf{x})\, d\mathbf{x} \\ &= \int g(\mathbf{x}) \ln g(\mathbf{x})\, d\mathbf{x} - \int g(\mathbf{x}) \ln f(\mathbf{x})\, d\mathbf{x}.\end{aligned}$$

Definition 1 implies that:

$$\min_{\boldsymbol{\mu}_{t+1}, \boldsymbol{\sigma}_{t+1}} \mathcal{D}(g_{t+1}(\mathbf{x}), f(\mathbf{x}; \boldsymbol{\mu}_{t+1}, \boldsymbol{\sigma}_{t+1})) \tag{15}$$

$$= \min_{\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}} \Bigg(\int g_{t+1}(\mathbf{x}) \ln g_{t+1}(\mathbf{x})\, d\mathbf{x} - \underbrace{\int g_{t+1}(\mathbf{x}) \ln f(\mathbf{x};\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1})\, d\mathbf{x}}_{(*)} \Bigg).$$

By noting that the optimization problem (15) is with respect to the $\boldsymbol{\mu}_{t+1}$ and the $\boldsymbol{\sigma}_{t+1}$ parameters, we arrive to the conclusion that (15) is equivalent to the *maximization* problem of the $(*)$ term with respect to $\boldsymbol{\mu}_{t+1}$ and $\boldsymbol{\sigma}_{t+1}$. Namely, one can consider:

$$\min_{\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}} \mathcal{D}(g_{t+1}(\mathbf{x}), f(\mathbf{x};\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1})) = \max_{\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}} \int g_{t+1}(\mathbf{x}) \ln f(\mathbf{x};\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1})\, d\mathbf{x}. \tag{16}$$

By adopting $f(\mathbf{x};\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t)$ as an approximation of $g_t(\mathbf{x})$, (16) can be written as:

$$\begin{aligned}
&\max_{\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}} \int g_{t+1}(\mathbf{x}) \ln f(\mathbf{x};\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}) d\mathbf{x} \qquad (17)\\
&= \max_{\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}} \int \frac{\mathbb{1}_{\{S(\mathbf{x})\geq\gamma_t\}} g_t(\mathbf{x})}{\sum_{\mathbf{x}\in\mathcal{X}} \mathbb{1}_{\{S(\mathbf{x})\geq\gamma_t\}} g_t(\mathbf{x})} \ln f(\mathbf{x};\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}) d\mathbf{x}\\
&= \max_{\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}} \int \frac{\mathbb{1}_{\{S(\mathbf{x})\geq\gamma_t\}} f(\mathbf{x};\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t)}{\int \mathbb{1}_{\{S(\mathbf{x})\geq\gamma_t\}} f(\mathbf{x};\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t) d\mathbf{x}}\\
&\qquad \times \ln\left(\prod_{d=1}^{k} \left(2\pi\sigma_{t+1,d}^2\right)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\sigma_d^2}\left(x_d-\mu_{t+1,d}\right)^2} \right) d\mathbf{x}\\
&= \max_{\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}} \int \mathbb{1}_{\{S(\mathbf{x})\geq\gamma_t\}} f(\mathbf{x};\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t)\\
&\qquad \times \ln\left(\prod_{d=1}^{k} \left(2\pi\sigma_{t+1,d}^2\right)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\sigma_d^2}\left(x_d-\mu_{t+1,d}\right)^2} \right) d\mathbf{x}\\
&= \max_{\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}} \mathbb{E}_{f(\mathbf{x};\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t)} \mathbb{1}_{\{S(\mathbf{X})\geq\gamma_t\}}\\
&\qquad \times \ln\left(\prod_{d=1}^{k} \left(2\pi\sigma_{t+1,d}^2\right)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\sigma_d^2}\left(x_d-\mu_{t+1,d}\right)^2} \right).
\end{aligned}$$

The correctness of (17) is due to the fact that the denominator $\int \mathbb{1}_{\{S(\mathbf{x})\leq\gamma_t\}} f(\mathbf{x};\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t) d\mathbf{x}$ is constant which does not depend on the parameters $\boldsymbol{\mu}_{t+1}$ and $\boldsymbol{\sigma}_{t+1}$.

While the exact solution of (17) is generally not available analytically, it can be approximated via samples from $f(\mathbf{x};\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t)$. In fact, this is common to consider the so-called *stochastic counterpart* [48], and to examine the solution of (17), which can be approximated via:

$$\max_{\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}} \frac{1}{N} \sum_{l=1}^{N} \mathbb{1}_{\{S(\mathbf{X}_l)\geq\widehat{\gamma}_t\}} \times \ln\left(\prod_{d=1}^{k} \left(2\pi\sigma_{t+1,d}^2\right)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\sigma_{t+1,d}^2}\left(X_{l,d}-\mu_{t+1,d}\right)^2} \right), \tag{18}$$

where $\mathbf{X}_l = (X_{l,1},\ldots,X_{l,k})^\top \sim f(\mathbf{x};\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t)$ for $l = 1,\ldots,N$. Given the sample $\{\mathbf{X}_l\}_{l=1}^N$, consider a subset $E = \{\mathbf{X}\in\{\mathbf{X}_l\}_{l=1}^N : S(\mathbf{X})\geq\widehat{\gamma}_t\}$. Then, (18) is equivalent to

$$\max_{\boldsymbol{\mu}_{t+1},\boldsymbol{\sigma}_{t+1}} \sum_{\mathbf{X}\in E} \ln\left(\prod_{d=1}^{k} \left(2\pi\sigma_{t+1,d}^2\right)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\sigma_{t+1,d}^2}\left(X_d-\mu_{t+1,d}\right)^2} \right),$$

which, for any coordinate $1\leq d\leq k$, can be recognized as the log-likelihood maximization problem under the univariate Gaussian model setting with unknown mean and variance. That is, (see [55, Example 7.2.12]), the optimal parameters $\boldsymbol{\mu}_{t+1}^* = (\mu_{t+1,1}^*,\ldots,\mu_{t+1,k}^*)$ and $\boldsymbol{\sigma}_{t+1}^* = (\sigma_{t+1,1}^*,\ldots,\sigma_{t+1,k}^*)$ which maximize (18), are given by:

$$\mu_{t+1,d}^* = \frac{\sum_{l=1}^N \mathbb{1}_{\{S(\mathbf{X}_l)\geq\gamma_t\}} X_{l,d}}{\sum_{l=1}^N \mathbb{1}_{\left\{S(\mathbf{X}_j)\geq\gamma_t\right\}}}$$

and

$$\sigma_{t+1,d}^* = \left(\frac{\sum_{l=1}^N \mathbb{1}_{\{S(\mathbf{X}_l)\geq\gamma_t\}} (X_{j,d}-\mu_{t+1,d}^*)^2}{\sum_{l=1}^N \mathbb{1}_{\{S(\mathbf{X}_l)\geq\gamma_t\}}} \right)^{\frac{1}{2}} \quad \forall\, 1\leq d\leq k.$$

In other words, for any coordinate $1\leq d\leq k$, the optimal $\mu_{t+1,d}^*$ and $\sigma_{t+1,d}^*$ are the sample mean and the sample standard deviation of the samples from the $E = \{\mathbf{X}\in\{\mathbf{X}_l\sim f(\mathbf{x};\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t)\}_{l=1}^N : S(\mathbf{X})\geq\widehat{\gamma}_t\}$ set, respectively. That is, the $t+1$ CE iteration parameters ($\boldsymbol{\mu}_{t+1}$ and $\boldsymbol{\sigma}_{t+1}$), can be calculated from the set of best-performing samples from the current iteration t. This discussion combined with the framework from Fig. 1, leads to the CE method for the maximum likelihood problem. The corresponding CE algorithm is summarized in Algorithm 2.

We now explore the smoothing step in lines 11-13 and the termination condition in line 2 to wrap up the discussion of the CE method. Given that we are working with noisy likelihood estimates, we can declare the algorithm terminated if the sample quantile sequence stops moving significantly for some $t > t'$. Alternatively, we can limit the number of iterations performed by Algorithm 2, namely, define $\mathbb{1}_{\{t>T\}}$ to be the stopping criteria for some predefined threshold parameter T, say $T = 300$.

The *smoothing* step in lines 11-13 of Algorithm 2 was shown to help the CE algorithm to converge to optimal solutions [43]. In particular, note that smoothing is just a combination of the previous iteration Gaussian distribution parameters ($\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$) and the solution of the stochastic counterpart optimization problem ($\tilde{\boldsymbol{\mu}}_{t+1}$ and $\tilde{\boldsymbol{\sigma}}_{t+1}$). We do not use the newly obtained solution $(\tilde{\boldsymbol{\mu}}_{t+1},\tilde{\boldsymbol{\sigma}}_{t+1})$ directly in the $t+1$-th iteration, as the stochastic counterpart solution may be obtained from the possibly some that *unfavorable* sample set at a specific iteration. This (if no smoothing is performed), may cause the CE algorithm to focus on a sub-optimal part of the search space, and thus the CE method can terminate its execution in such an inadequate region. This unfavorable scenario can happen especially if the sample size N_{CE} is too small. In other words, by slowing down the update of the probability distribution parameters, the smoothing step enables control of the pace of convergence of the CE algorithm. For additional details about the smoothing step, we refer to [43], [47], [48].

Recall that $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ can be immediately recovered from the mean $\boldsymbol{\mu}_t$ of the sampling distribution $f(\mathbf{x};\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t)$ in the last iteration of the CE Algorithm 2, since

$$\boldsymbol{\mu}_t = (\widehat{\alpha}_{(\mathrm{MLE})},\widehat{\eta}_{(\mathrm{MLE})},\widehat{\beta}_{(\mathrm{MLE})},\widehat{\mu}_{A(\mathrm{MLE})},\widehat{\sigma}_{A(\mathrm{MLE})},\widehat{\sigma}_{Z(\mathrm{MLE})}).$$

That is, Algorithm 2 can be readily used for obtaining maximum likelihood estimators for the GP problem under consideration. However, CE can be also utilized for acquiring confidence intervals (CIs). Specifically, we propose to use the non-parametric bootstrap method for the construction of the corresponding CIs [56], [57]. To do so, one is required to

Algorithm 2: The Cross Entropy algorithm

Input: m — the number of components, $\{n_i\}_{i=1}^m$ — the number of measurements for each component, $\{\{t_{i,j}\}_{j=1}^{n_i}\}_{i=1}^m$ — measurement times, a vector parameters $\boldsymbol{\theta} = (\alpha, \eta, \beta, \mu_A, \sigma_A, \sigma_Z)$, $k = |\boldsymbol{\theta}|$, the observed data $\mathbf{y} = \{y_i(t) : t \in \{t_{i,1}, \ldots, t_{i,n_i}\}\}_{i=1}^m$, a sample size $N_{\mathrm{CMC}_2} \in \mathbb{N}$, a CE rarity parameter $\rho \in (0,1)$, a CE smoothing parameters $\upsilon_1, \upsilon_2 \in (0,1)$, and a CE sample size $N_{\mathrm{CE}} \in \mathbb{N}$.

Output: Approximate maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$.

Set $t \leftarrow 1$ and some initial $\boldsymbol{\mu}_t \leftarrow (\mu_{t,1}, \ldots, \mu_{t,k})$, $\boldsymbol{\sigma}_t \leftarrow (\sigma_{t,1}, \ldots, \sigma_{t,k})$.

while *termination criterion is not satisfied* **do**

  Sample $\mathbf{X}_l \sim f(\mathbf{x}; \boldsymbol{\mu}_t, \boldsymbol{\sigma}_t)$, and calculate $S(\mathbf{X}_l) = \widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\mathbf{X}_l; \Delta\mathbf{y})$ via Algorithm 1 with N_{CMC_2} sample size for $1 \le l \le N_{\mathrm{CE}}$.

  Let $S_{(1)} \le \cdots \le S_{(N_{\mathrm{CE}})}$ be the elements of the $\{S(\mathbf{X}_l)\}_{l=1}^{N_{\mathrm{CE}}}$ set sorted in an ascending order.

  $\widehat{\gamma}_t \leftarrow S_{(\lceil N_{\mathrm{CE}} \times \rho \rceil)}$ /* Sample fitness' ρ-quantile */

  /* Calculate $\boldsymbol{\mu}_{t+1}$ and $\boldsymbol{\sigma}_{t+1}$ */

  for $d = 1, \ldots, k$ **do**

$$\tilde{\mu}_{t+1,d} = \frac{\sum_{l=1}^{N_{\mathrm{CE}}} \mathbb{1}_{\{S(\mathbf{X}_l) \ge \gamma_t\}} X_{l,d}}{\sum_{l=1}^{N_{\mathrm{CE}}} \mathbb{1}_{\{S(\mathbf{X}_l) \ge \gamma_t\}}}$$

and

$$\tilde{\sigma}_{t+1,d} = \left(\frac{\sum_{l=1}^{N_{\mathrm{CE}}} \mathbb{1}_{\{S(\mathbf{X}_l) \ge \gamma_t\}} (X_{l,d} - \tilde{\mu}_{t+1,d})^2}{\sum_{l=1}^{N_{\mathrm{CE}}} \mathbb{1}_{\{S(\mathbf{X}_l) \ge \gamma_t\}}} \right)^{\frac{1}{2}}.$$

    /* note that $X_{l,d}$ is the d-th component of $\mathbf{X}_l$ */

  end

$$\tilde{\boldsymbol{\mu}}_{t+1} \leftarrow (\tilde{\mu}_{t+1,i}, \ldots, \tilde{\mu}_{t+1,k}), \text{ and}$$
$$\tilde{\boldsymbol{\sigma}}_{t+1} \leftarrow (\tilde{\sigma}_{t+1,i}, \ldots, \tilde{\sigma}_{t+1,k}).$$

  /* The smoothing step */

  $\boldsymbol{\mu}_{t+1} \leftarrow \upsilon_1 \tilde{\boldsymbol{\mu}}_{t+1} + (1 - \upsilon_1) \boldsymbol{\mu}_t$

  $B_{\mathrm{mod}} \leftarrow \upsilon_2 - \upsilon_2 \left(1 - \frac{1}{t}\right)^5$

  $\boldsymbol{\sigma}_{t+1} \leftarrow B_{\mathrm{mod}} \tilde{\boldsymbol{\sigma}}_{t+1} + (1 - B_{\mathrm{mod}}) \boldsymbol{\sigma}_t$

  $t \leftarrow t + 1$

end

/* Note that $\boldsymbol{\mu}_t$ is the desired vector $(\widehat{\alpha}_{(\mathrm{MLE})}, \widehat{\eta}_{(\mathrm{MLE})}, \widehat{\beta}_{(\mathrm{MLE})}, \widehat{\mu}_{A(\mathrm{MLE})}, \widehat{\sigma}_{A(\mathrm{MLE})}, \widehat{\sigma}_{Z(\mathrm{MLE})})$ */

Set $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} \leftarrow \boldsymbol{\mu}_t$

return $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$

sample m degradation trajectories with replacement from the original dataset $\mathbf{Y} = \{Y_i(t) : t \in \{t_{i,1}, \ldots, t_{i,n_i}\}\}_{i=1}^m$. Then, using some number (say $B = 100$) of such bootstrap datasets and Algorithm 2, we can obtain the maximum likelihood estimates $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(B)}$. The $\{\boldsymbol{\theta}^{(b)}\}_{b=1}^B$ set can be readily used for the CIs construction. In this paper, we utilize the 95% Normal (CI_{N}), the 95% pivotal ($\mathrm{CI}_{\mathrm{piv}}$), and the 95% percentile ($\mathrm{CI}_{\mathrm{per}}$) confidence intervals [56].

IV. Experimental study

In this section, we concentrate on the performance evaluation of the proposed combination of the CMC_2 algorithm and the CE method. For each model under consideration, we apply our algorithm for parameter estimation. The experimental study shows that the proposed paradigm is very effective and that it is scalable in practice in the sense that it is able to provide reliable estimates for real-sized data instances. Specifically, we consider two synthetic datasets along with one real-life application. The following case studies are examined.

1) A synthetic *stationary* model for data similar to the one presented by Hazra et al. [1], is considered in the first case study. The degradation data was generated for $m = 5$ components, and the corresponding measurements were taken at times $t \in \{5, 10, 15\}$ for each component. The set of parameters used to generate the data is $\boldsymbol{\theta}^* = (\alpha^*, \eta^*, \beta^*, \mu_A^*, \sigma_A^*, \sigma_Z^*) = (4, 1, 0.015, 0.0, 0.0, 0.1)$. Here the focus is on a synthetic dataset, since for such data, the availability of the true parameter set $\boldsymbol{\theta}^*$, allows us to benchmark the performance of the method proposed in this manuscript.

2) Similarly to the first case study, we follow Hazra et al. [1] and consider a synthetic dataset generated from a *non-stationary* model. The degradation data was simulated for $m = 10$ components. For every component, the measurements were taken at times $t \in \{2, 4, 6\}$. In this case study, we used the set of parameters: $\boldsymbol{\theta}^* = (\alpha^*, \eta^*, \beta^*, \mu_A^*, \sigma_A^*, \sigma_Z^*) = (2, 2.5, 0.01, 0.5, 0.1, 0.1)$.

3) The use of the GP degradation model with actual data is the subject of our third case study. In particular, we consider the drug potency degradation dataset from Hamada et al. and Chow et al. [58], [59]. This data reflects the so-called drug's shelf life (SL). Specifically, since a drug potency generally degrades over time, its SL is determined by the time at which the drug's potency reaches some predefined threshold value (in our case, we set the threshold to be 90%). There is a great practical importance when considering drug SLs. For example, if the actual SL (say with 95% confidence), is smaller than the declared one, the product has to be recalled [59]. A shorter SL can also indicate that a manufacturing problem is present. In addition, if the actual SL is longer than stated, an adjusted labeled SL for the product may be applied, thus giving a clear financial benefit to the manufacturer.

The experimental setup. Both the CMC_2 algorithm and the CE method were implemented in Python (version 3.9.16). All the software and the research data are freely available on request from the authors. No optimization was undertaken and the implementation is single-threaded. While the parallelization of the CE algorithm is (relatively) easy to implement, our objective is to demonstrate that even a relatively slow-performing scripting language can already provide satisfactory

results. The tests were executed on Intel Core i7-6920HQ CPU 2.90GHz processor with 32GB of RAM running 64 bit Debian 11 "bullseye". A crucial aspect of the proposed method is the CE method (input) parameter choice, which is important from both the computational complexity and the inference accuracy points of view. Using several preliminary experiments, we obtained a set of parameters that work well for the case studies under consideration; please see Remark 1 for additional information regarding these preliminary experiments. Since the CE algorithm is capable of working with noisy objective functions, we set $N_{\mathrm{CMC}_2} = 1$. The additional CE parameters are as follows: the rarity parameter, the smoothing parameters, and the CE sample size, are set to be $\rho = 0.1$, $\upsilon_1 = \upsilon_2 = 0.7$, and $50 \times |\boldsymbol{\theta}|$, respectively. That is, for case study 1, where we consider 3 unknown (inference) parameters, namely $(\alpha, \beta, \sigma_Z)$, $N_{\mathrm{CE}} = 50 \times 3 = 150$. For the full inference procedure, and in particular for case studies 2 and 3, we set $N_{\mathrm{CE}} = 50 \times 6 = 300$. For all case studies, we execute the CE algorithm for 300 iterations; namely, the termination criteria in line 2 of Algorithm 2 is $\mathbb{1}_{\{t>300\}}$. In order to obtain confidence intervals, we use the non-parametric bootstrap method with $B = 100$ bootstrap rounds [56], [57].

A. Case study 1: the synthetic stationary data

For the first case study, we follow Hazra et al. [1], and consider a synthetic data with $m = 5$ fictitious components. The data was generated using the parameter set

$$\boldsymbol{\theta}^* = (\alpha^*, \eta^*, \beta^*, \mu_A^*, \sigma_A^*, \sigma_Z^*) = (4, 1, 0.015, 0.0, 0.0, 0.1);$$

the inspection times for every component are $t \in \{5, 10, 15\}$. For this case study, we assume that η, μ_A and σ_A are fixed and known in advance. That is, we are interested in the MLE for three remaining parameters α, β, and σ_Z.

To benchmark the CE method, we first execute the CMC Algorithm 1 to obtain the $\log \widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\boldsymbol{\theta}^*; \Delta\mathbf{y})$ estimator for $\boldsymbol{\theta}^* = (4, 1, 0.015, 0.0, 0.0, 0.1)$. In this case, it was found that a sample size of $N = 100,000$ is sufficient to obtain a relative error (RE) [50], which is smaller than 2%. Specifically, we found that $\log \widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\boldsymbol{\theta}^*; \Delta\mathbf{y}) = 8.2010$ and that $\mathrm{RE} = 0.0122$. Fig. 2 shows the typical dynamics of the CE algorithm.

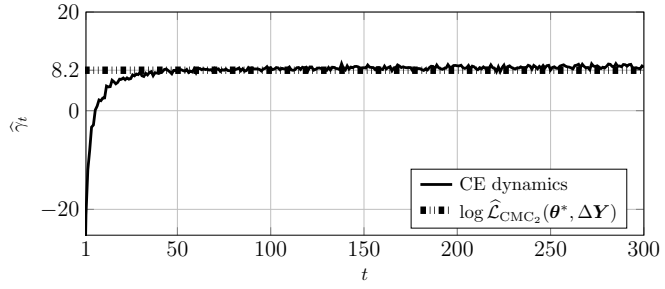


Fig. 2: The typical dynamics of the CE algorithm for the first case study. The figure depicts the performance of the CE method $(\widehat{\gamma}_t)$ as a function of the CE iteration t. The runtime of the CE algorithm is about 7.9 seconds.

One can observe that the CE method reaches a log-likelihood value which is close to 8.2 in about 50 iteration. The CE method delivers an MLE estimator

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} &= (\widehat{\alpha}_{(\mathrm{MLE})}, \widehat{\eta}_{(\mathrm{MLE})}, \widehat{\beta}_{(\mathrm{MLE})}, \widehat{\mu}_{A(\mathrm{MLE})}, \widehat{\sigma}_{A(\mathrm{MLE})}, \widehat{\sigma}_{Z(\mathrm{MLE})}) \\ &= (3.5086, 1, 0.0184, 0, 0, 0.1098),\end{aligned}$$

and, when executing Algorithm 1 with $N = 100,000$ sample size, we obtain $\log \widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}; \Delta\mathbf{y}) = 8.2966$ with $\mathrm{RE} = 0.0091$. This result is comparable to the one obtained with the true parameter set, namely, to $\log \widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\boldsymbol{\theta}^*; \Delta\mathbf{y})$.

Remark 1 (CE parameters)**.** Fig. 2 is instructive in the sense that similar CE dynamics plots can be used in preliminary experiments to determine a suitable set of the CE algorithm parameters. Specifically, by running the CE method several times and observing the corresponding dynamics, one can determine both the required sample size and the number of CE iterations needed to achieve a good solution. To do so, it is advisable to monitor the performance $\widehat{\gamma}_t$ to determine when it ceases to improve. While the proposed rarity parameter ρ and the set of smoothing parameters v_1 and v_2 can be found in a similar fashion, namely, by monitoring the convergence of $\widehat{\gamma}_t$, we used a common set of values, specifically, $\rho = 0.1$ and $v_1 = v_2 = 0.7$. These, according to our experience, usually work well in practice [48]. Finally, it is advisable to run the algorithm with different random generator seeds to ensure that the algorithm converges to solution of similar quality.

By utilizing the non-parametric bootstrap method with $B = 100$ replications, we obtain confidence intervals for the set of parameter of interest. Table II provides a summary of confidence intervals. For this case study, all types of CIs, namely, CI_{N}, $\mathrm{CI}_{\mathrm{piv}}$, and $\mathrm{CI}_{\mathrm{per}}$, perform similarly and contain all the true model parameters α^*, β^*, and σ_Z^*.

Remark 2 (The number of bootstrap rounds B)**.** Our numerical experiments indicate that $B = 100$ bootstrap rounds are sufficient for obtaining adequate CIs. To see this, consider Figure 3 which shows the changes to the CIs widths as a function of the number of bootstrap rounds B. Similar sensitivity results (not presented here), were obtained for case study 2 and case study 3.

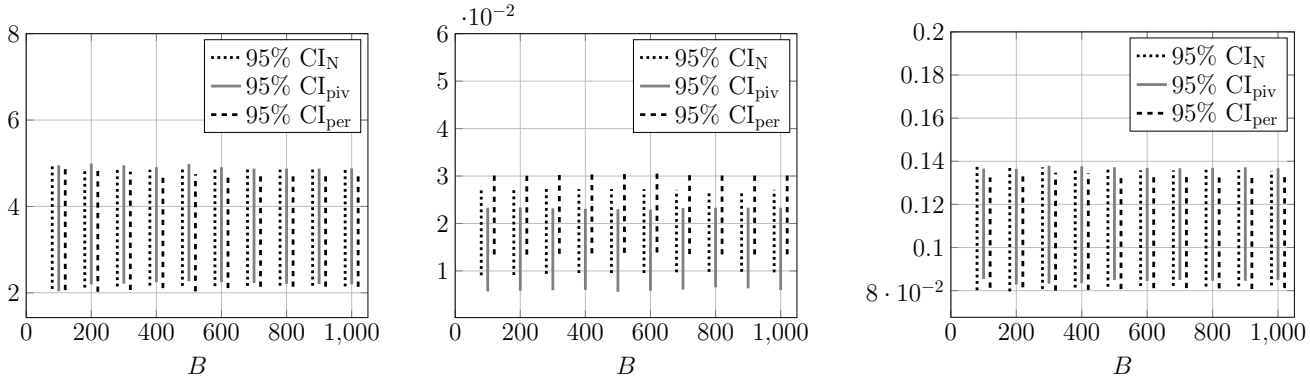
B. Case study 2: the synthetic non-stationary data

In this case study, we again consider the experiment from Hazra et al. [1], by examining a synthetic dataset of $m = 10$ fictitious components. The data was generated using the parameter set $\boldsymbol{\theta}^* = (\alpha^*, \eta^*, \beta^*, \mu_A^*, \sigma_A^*, \sigma_Z^*) = (2, 2.5, 0.01, 0.5, 0.1, 0.1)$, where we assume that the inspection times for each component are $t \in \{2, 4, 6\}$. Similar to the first case study, Algorithm 1 is executed and we obtain $\log \widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\boldsymbol{\theta}^*; \Delta\mathbf{y}) = 12.6967$ with $\mathrm{RE} = 0.0103$. To obtain an RE that is smaller than 2%, Algorithm 1 requires $N = 10,000,000$ sample size.

Fig. 4 shows the typical dynamics of the CE algorithm. Here, the CE method reaches a log-likelihood value which is

TABLE II Case study 1 CIs based on $B = 100$ bootstrap rounds. The total execution time is 867.9 seconds.

parameter name	$\boldsymbol{\theta}^*$	$\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$	95% $\mathrm{CI_N}$	95% $\mathrm{CI_{piv}}$	95% $\mathrm{CI_{per}}$
α	4.0	3.5086	(2.0896, 4.9276)	(2.0369, 4.9546)	(2.0627, 4.9804)
η	1.0	1.0	—	—	—
β	0.015	0.0184	(0.0091, 0.0277)	(0.0057, 0.0233)	(0.0135, 0.0311)
μ_A	0	0	—	—	—
σ_A	0	0	—	—	—
σ_Z	0.1	0.1098	(0.0802, 0.1377)	(0.0854, 0.1366)	(0.0814, 0.1325)



(a) α (b) β (c) σ_Z

Fig. 3: Case study 1 CI width as a function of the number of bootstrap rounds for parameters α, β, and σ_Z.

close to 12.7. The MLE estimator obtained by the CE method is:

$$\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} = (\widehat{\alpha}_{(\mathrm{MLE})}, \widehat{\eta}_{(\mathrm{MLE})}, \widehat{\beta}_{(\mathrm{MLE})}, \widehat{\mu}_{A(\mathrm{MLE})}, \widehat{\sigma}_{A(\mathrm{MLE})}, \widehat{\sigma}_{Z(\mathrm{MLE})}) = (3.3128, 2.5586, 0.0055, 0.5377, 0.03499, 0.1270).$$

It is interesting to note that Algorithm 1 with $N = 100,000$ sample size, delivers $\log \widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}; \Delta\mathbf{y}) = 13.5679$ with $\mathrm{RE} = 0.0155$. That is, we have that $\log \widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}; \Delta\mathbf{y}) > \log \widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\boldsymbol{\theta}^*; \Delta\mathbf{y})$.

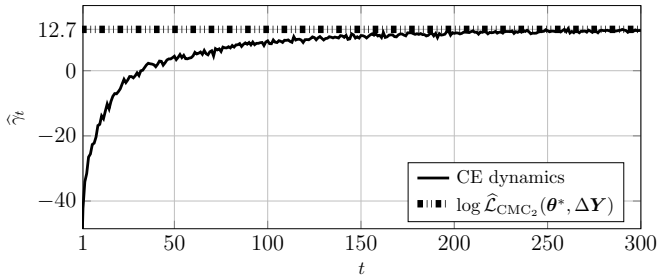


Fig. 4: The typical dynamics of the CE algorithm for the second case study. The figure depicts the performance of the CE method $(\widehat{\gamma}_t)$ as a function of the CE iteration t. The runtime of the CE algorithm is about 23.8 seconds.

Similar to the first case study, we use a non-parametric bootstrap method with $B = 100$ bootstrap rounds to obtain confidence intervals for the set of parameters of interest. Table III provides a summary of confidence intervals. Our experiments (not reported here), indicate that $B = 100$ bootstrap rounds are sufficient for obtaining adequate CIs.

C. Case study 3: the drug potency data

Here we examine the drug potency data from Chow et al. [59]. In this case, we do not know the set of true parameters $\boldsymbol{\theta}^*$. The typical dynamics of three independent runs of the CE algorithm is shown in Fig. 5. Note that CE quickly reaches a point after which, no large improvement in performance is observed. All three independent runs of the CE method reach the log-likelihood value of about $-10^{2.135} \approx -136$, and a typical obtained MLE estimator is:

$$\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} = (\widehat{\alpha}_{(\mathrm{MLE})}, \widehat{\eta}_{(\mathrm{MLE})}, \widehat{\beta}_{(\mathrm{MLE})}, \widehat{\mu}_{A(\mathrm{MLE})}, \widehat{\sigma}_{A(\mathrm{MLE})}, \widehat{\sigma}_{Z(\mathrm{MLE})}) = (4.3689, 1.0753, 0.0414, -4.5703, 0.1918, 1.3191).$$

By executing Algorithm 1 with $N = 1,000,000$ sample size, we obtain $\log \widehat{\mathcal{L}}_{\mathrm{CMC}_2}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}; \Delta\mathbf{y}) = -136.5$ with $\mathrm{RE} = 0.0096$.

As usual, we apply the non-parametric bootstrap method with $B = 100$ bootstrap rounds, to obtain confidence intervals for the parameter set of interest. Table IV provides a summary of confidence intervals.

As for the first and the second case studies, our experiments indicate that $B = 100$ bootstrap rounds are sufficient for obtaining adequate CIs.

We proceed with the calculation of the true degradation and the corresponding confidence intervals. Recall that the true degradation at time t is given by $A + X(t)$. The

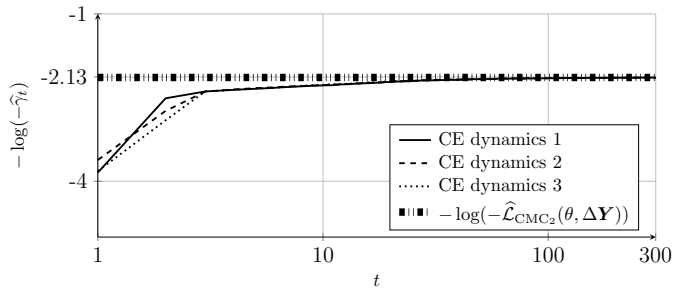


Fig. 5: The typical dynamics of the CE algorithm for the third case study. The figure depicts the logarithm of performance of the CE method ($-\log(-\widehat{\gamma}_t)$) as a function of the CE iteration t for three independent runs of the CE method. The runtime of the CE algorithm is about 42.1 seconds.

mean degradation is given by $\mu_A + \beta\alpha t^{\eta}$, since $\mathbb{E}A = \mu_A$ and $\mathbb{E}X(t) = \beta\alpha t^{\eta}$ [1]. Thus, given $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$, the average degradation can be readily estimated using the corresponding parameters from $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$. Specifically, the average degradation at time t is given by $\widehat{\mu}_{A(\mathrm{MLE})} + \widehat{\beta}_{(\mathrm{MLE})}\widehat{\alpha}_{(\mathrm{MLE})}t^{\widehat{\eta}_{(\mathrm{MLE})}}$. To obtain the corresponding CIs, we can again apply the non-parametric bootstrap approach. Specifically, given a bootstrapped sample $\boldsymbol{\theta}^{(b)} = (\alpha^{(b)}, \eta^{(b)}, \beta^{(b)}, \mu_A^{(b)}, \sigma_A^{(b)}, \sigma_Z^{(b)})$ for $1 \leq b \leq B$, the average bootstrapped degradation is $\mu_A^{(b)} + \beta^{(b)}\alpha^{(b)}t^{\eta^{(b)}}$. These bootstrapped degradation values can be used in the usual way to obtain the CIs. For the drug potency data, the mean degradation and the corresponding CIs are depicted in Fig. 6. Again, it appears that $B = 100$ bootstrap rounds are sufficient and similar to Chow et al. [59], we arrive at the conclusion that the shelf-life is longer when 36 months.

V. Conclusion

In this work, we introduced a computationally efficient method which is suitable for the GP inference task. The proposed approach avoids the rare-event problem that emerges from the naive Monte Carlo estimation of the intractable likelihood function. Our experimental evaluation is comparable to, and sometimes improves upon, the previously published results. The work we propose is significant for several reasons. Firstly, it advances the domain of inference under the intractable likelihood function setting, and therefore, its benefits may extend beyond the study of degradation processes. Additionally, the proposed alternative Monte Carlo estimator can be applied to other likelihood-based Markov Chain Monte Carlo algorithms.

However, our study is subject to a few limitations. The primary limitation is that practitioners will need to identify reasonable input parameters for the Cross Entropy algorithm, such as sample sizes, smoothing, and rarity parameters, depending on the problem at hand. These parameters will ultimately determine the quality of inference and computational time required. Additionally, our study presents a single-threaded implementation of the proposed methods. A parallel implementation will allow to handle larger real-world datasets. Lastly, while our results with synthetic data demonstrated high-quality inference outcomes, the algorithm was unable to recover the standard deviation of the initial degradation in the second case study. Nevertheless, the important set of degradation parameters needed for the unobserved degradation approximation appeared to be estimated in an accurate fashion.

As for the future research, we believe that the following directions are of interest.

While in this paper we concentrated solely on the alternative Monte Carlo estimator, it is of interest to cast the problem into the state-space model setting (Feynman-Kac formalism). This can potentially allow to minimize the Monte Carlo algorithm variance and introduce a considerable computational improvement to any inference method that will eventually apply such an estimator. In addition, it will be of interest to design an Expectation-Maximization algorithm and compare its performance with the proposed Cross-Entropy method.

The proposed Monte Carlo method that was used as an estimator for the intractable likelihood function avoids the rare-event trap in contrast to its counterpart, and in this paper, we used the Cross Entropy method for the corresponding intractable likelihood optimization. However, our Monte Carlo approach opens the way to reliable estimation with likelihood-based Markov Chain Monte Carlo methods. It is thus of interest to explore the performance of likelihood based Markov Chain Monte Carlo and other global optimization methods and compare these to the proposed Cross Entropy approach.

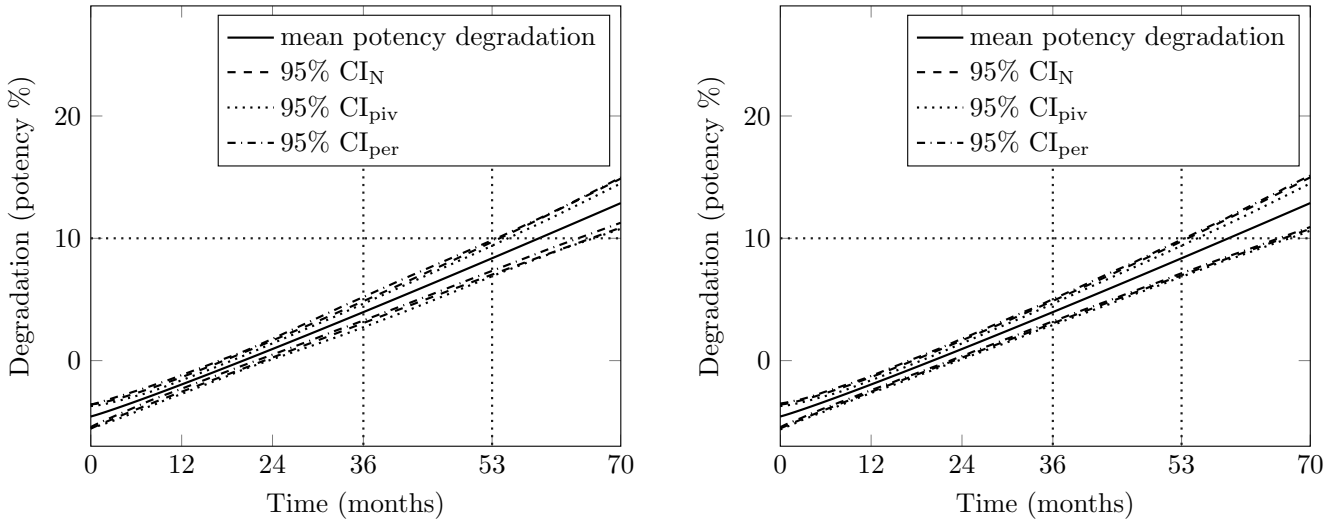
We showed several computational experiments that involve specific gamma distribution shape parameters. Thus, the robustness of the proposed algorithmic method under additional important gamma distribution shape parameters setting must be considered.

TABLE III Case study 2 CIs based on $B = 100$ bootstrap rounds. The total execution time is 2201.7 seconds.

parameter name	$\boldsymbol{\theta}^*$	$\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$	95% CI_{N}	95% $\mathrm{CI}_{\mathrm{piv}}$	95% $\mathrm{CI}_{\mathrm{per}}$
α	2.0	3.3128	(2.0161, 4.6095)	(2.3500, 4.9758)	(1.6499, 4.2756)
η	2.5	2.5586	(2.0748, 3.0425)	(1.8960, 2.8855)	(2.2318, 3.2212)
β	0.01	0.0055	(0.0022, 0.0087)	(0.001, 0.0079)	(0.0031, 0.0099)
μ_A	0.5	0.5377	(0.4510, 0.6244)	(0.4407, 0.6010)	(0.4744, 0.6347)
σ_A	0.1	0.03499	(0.0267, 0.0433)	(0.0337, 0.0488)	(0.0212, 0.0363)
σ_Z	0.1	0.1270	(0.1060, 0.1480)	(0.1150, 0.1557)	(0.0984, 0.1390)

TABLE IV Case study 3 CIs based on $B = 100$ bootstrap rounds. The execution time is 3614.9 seconds.

parameter name	$\widehat{\boldsymbol{\theta}}_{\text{MLE}}$	95% CI_{N}	95% CI_{piv}	95% CI_{per}
α	4.3689	(2.3074, 6.4305)	(2.4413, 6.1994)	(2.5385, 6.2966)
η	1.0753	(0.9097, 1.2408)	(0.9081, 1.2141)	(0.9364, 1.2424)
β	0.0414	(0.0150, 0.0679)	(0.0066, 0.0572)	(0.0256, 0.0763)
μ_A	-4.5703	(-5.5549, -3.5857)	(-5.539, -3.7124)	(-5.4282, -3.6017)
σ_A	0.1918	(0.1093, 0.2742)	(0.0736, 0.2414)	(0.1422, 0.3099)
σ_Z	1.3191	(0.9933, 1.6450)	(0.9990, 1.6148)	(1.0236, 1.6394)



(a) $B = 100$ bootstrap rounds (b) $B = 1000$ bootstrap rounds

Fig. 6: The left (a) and the right (b) panel show the average degradation and CIs for the drug potency data obtained using $B = 100$ and $B = 1000$ bootstrap rounds, respectively. The degradation threshold is set to be 10%.

In this paper, we worked under the assumption of the measurement error normality. It is therefore important to consider the development of additional Monte Carlo estimators that both avoids the rare event trap and can handle non Gaussian measurement errors.

The proposed approach also has implications for many areas of the human and social sciences. Social scientists are interested in explaining and predicting many phenomena that can degrade over time (e.g., social cohesion, mental health, economic contraction/decline). In social science, the observable data that track degradation processes are often particularly noisy as it relies upon a proxy, or several proxies, that track a fundamentally unobservable variable (i.e., a latent variable). As mentioned in the introductory section, one important example of this is age-related cognitive decline. Age-related cognitive decline is explained by a complex interaction of many factors, but due to this complexity it may be impossible to ever specify a theoretically informed model. There are, however, several psychological tests that can track this phenomenon indirectly that could be used to build a model to help explain and predict the rate of cognitive decline. In summary, this approach has utility in many areas of the human and social sciences when there is an extremely complex and opaque data-generating mechanism.

Although we currently offer a single-threaded implementation of the inference algorithms, it would be highly beneficial to create a parallel software implementation that can operate on multiple CPUs or a GPU. Such software will allow practitioners to handle large real-life degradation data instances. Specifically, both Algorithms 2 and 3 may benefit from modifications that leverage tensor data structures. In particular, tensors may enhance Algorithm 2 with improved data storage for the observed data and the intermediate variables and results generated during the estimation process. In addition, utilizing tensor-based operations allows for parallel sampling which can generate multiple samples simultaneously. This results in a more efficient process for generating and processing large volumes of samples, potentially improving estimation accuracy in certain situations. As for Algorithm 3, rather than using individual variables to store parameters, they can be represented as tensor objects which are efficient for storing and manipulating multidimensional parameter arrays. Again, tensors capacity to facilitate vectorized operations (i.e. their ability to perform operation on entire arrays or vectors at once) is helpful to parallelize the sampling step of the algorithm where random samples are generated based on the

current parameter estimates. Therefore, this adjustment would help avoid the need for explicit loops and, again, improve computational efficiency.

REFERENCES

[1] I. Hazra, M. D. Pandey, and N. Manzana, "Approximate Bayesian computation (ABC) method for estimating parameters of the gamma process using noisy data," *Reliability engineering & system safety*, vol. 198, p. 106780, 2020.

[2] S.-T. Tseng, N. Balakrishnan, and C.-C. Tsai, "Optimal step-stress accelerated degradation test plan for gamma degradation processes," *IEEE transactions on reliability*, vol. 58, no. 4, p. 611–618, 2009.

[3] L. Shen, Y. Wang, Q. Zhai, and Y. Tang, "Degradation modeling using stochastic processes with random initial degradation," *IEEE transactions on reliability*, vol. 68, no. 4, pp. 1320–1329, 2019.

[4] X. Yuan, "Stochastic modeling of deterioration in nuclear power plant components," PhD thesis, Civil Engineering, University of Waterloo, Ontario, Canada, April 2007.

[5] D. Lu, "Estimation of stochastic degradation models using uncertain inspection data," PhD thesis, Civil Engineering, University of Waterloo, Ontario, Canada, May 2012.

[6] T. Matsubara, J. Knoblauch, F. Briol, and C. J. Oates, "Robust generalised bayesian inference for intractable likelihoods," *Journal of the Royal Statistical Society. Series B, Statistical methodology*, vol. 84, no. 3, pp. 997–1022, 2022.

[7] M. Buist, "A Comparison of the Likelihood Estimation of Crude Monte Carlo Against Sequential Monte Carlo for Gamma Process-Based Degradation Models With Noisy Measurement Data," Eindhoven, Netherlands, Master Thesis, May 2023.

[8] D. P. Kroese and J. C.C. Chan, *Statistical Modeling and Computation*, 1st ed. New York, NY: Springer New York, 2014.

[9] A. M. Negm, A. M. F. Al-Quraishi, and Y. T. Mustafa, *Environmental degradation in Asia: land degradation, environmental contamination, and human activities*, ser. Earth and environmental sciences library. Cham, Switzerland: Springer, 2022.

[10] R. Yu and C. Zhang, "Early warning of water quality degradation: A copula-based bayesian network model for highly efficient water quality risk assessment," *Journal of environmental management*, vol. 292, pp. 112 749–112 749, 2021.

[11] A. Jahani, J. Feghhi, M. F. Makhdoum, and M. Omid, "Optimized forest degradation model (ofdm): an environmental decision support system for environmental impact assessment using an artificial neural network," *Journal of environmental planning and management*, vol. 59, no. 2, pp. 222–244, 2016.

[12] A. J. A. J. Lawrence and K. Hemingway, *Effects of pollution on fish: molecular effects and population responses*. Oxford, U.K. ; Carlton, Vic.: Blackwell Science, 2003.

[13] M. Kharaji Manouchehrabadi, S. Yaghoubi, and J. Tajik, "Optimal scenarios for solar cell supply chain considering degradation in powerhouses," *Renewable energy*, vol. 145, pp. 1104–1125, 2020.

[14] Z. Zhao, B. Liang, X. Wang, and W. Lu, "Remaining useful life prediction of aircraft engine based on degradation pattern learning," *Reliability engineering & system safety*, vol. 164, pp. 74–83, 2017.

[15] N. Caballé, I. Castro, C. Pérez, and J. Lanza-Gutiérrez, "A condition-based maintenance of a dependent degradation-threshold-shock model in a system with multiple degradation processes," *Reliability engineering & system safety*, vol. 134, pp. 98–109, 2015.

[16] G. Ni, J. Chen, and H. Wang, "Degradation assessment of rolling bearing towards safety based on random matrix single ring machine learning," *Safety science*, vol. 118, pp. 403–408, 2019.

[17] M. Dadfarnia, A. Nagao, S. Wang, M. L. Martin, B. P. Somerday, and P. Sofronis, "Recent advances on hydrogen embrittlement of structural materials," *International journal of fracture*, vol. 196, no. 1-2, pp. 223–243, 2015.

[18] P. Novak, R. Yuan, B. Somerday, P. Sofronis, and R. Ritchie, "A statistical, physical-based, micro-mechanical model of hydrogen-induced intergranular fracture in steel," *Journal of the mechanics and physics of solids*, vol. 58, no. 2, pp. 206–226, 2010.

[19] R. C. Portnoy, *Medical plastics degradation resistance & failure analysis*, ser. Plastics Design Library. Norwich, NY: Plastic Design Library, 1998.

[20] E. Oude Blenke, E. Ornskov, C. Schöneich, G. A. Nilsson, D. B. Volkin, E. Mastrobattista, O. Almarsson, and D. J. A. Crommelin, "The storage and in-use stability of mrna vaccines and therapeutics: Not a cold case," *Journal of pharmaceutical sciences*, 2023.

[21] F. P. Glasser, J. Marchand, and E. Samson, "Durability of concrete — degradation phenomena involving detrimental chemical reactions," *Cement and concrete research*, vol. 38, no. 2, pp. 226–246, 2008.

[22] I. J. Deary, J. Corley, A. J. Gow, S. E. Harris, L. M. Houlihan, R. E. Marioni, L. Penke, S. B. Rafnsson, and J. M. Starr, "Age-associated cognitive decline," *British medical bulletin*, vol. 92, no. 1, pp. 135–152, 2009.

[23] T. A. Salthouse, "When does age-related cognitive decline begin?" *Neurobiology of aging*, vol. 30, no. 4, pp. 507–514, 2009.

[24] C.-C. Tsai, S.-T. Tseng, and N. Balakrishnan, "Optimal design for degradation tests based on gamma processes with random effects," *IEEE transactions on reliability*, vol. 61, no. 2, pp. 604–613, 2012.

[25] M. H. Ling, K. L. Tsui, and N. Balakrishnan, "Accelerated degradation analysis for the quality of a system based on the gamma process," *IEEE transactions on reliability*, vol. 64, no. 1, pp. 463–472, 2015.

[26] C. Paroissin, "Online estimation methods for the gamma degradation process," *IEEE transactions on reliability*, vol. 66, no. 4, pp. 1361–1367, 2017.

[27] M. Sanò, J. Jiménez, R. Medina, A. Stanica, A. Sanchez-Arcilla, and I. Trumbic, "The role of coastal setbacks in the context of coastal erosion and climate change," *Ocean & coastal management*, vol. 54, no. 12, pp. 943–950, 2011.

[28] C. Barrow, "Biochar: Potential for countering land degradation and for improving agriculture," *Applied geography (Sevenoaks)*, vol. 34, pp. 21–28, 2012.

[29] M. Abdel-Hameed, "A gamma wear process," *IEEE transactions on reliability*, vol. R-24, no. 2, pp. 152–153, 1975.

[30] J. van Noortwijk, "A survey of the application of gamma processes in maintenance," *Reliability engineering & system safety*, vol. 94, no. 1, pp. 2–21, 2009.

[31] R. Dekker, "Applications of maintenance optimization models: a review and analysis," *Reliability engineering & system safety*, vol. 51, no. 3, pp. 229–240, 1996.

[32] R. Dekker and P. A. Scarf, "On the impact of optimisation models in maintenance decision making: the state of the art," *Reliability engineering & system safety*, vol. 60, no. 2, pp. 111–119, 1998.

[33] J. M. v. Noortwijk and E. B. Peerbolte, "Optimal sand nourishment decisions," *Journal of waterway, port, coastal, and ocean engineering*, vol. 126, no. 1, pp. 30–38, 2000.

[34] L. J. P. Speijker, J. M. van Noortwijk, M. Kok, and R. M. Cooke, "Optimal maintenance decisions for dikes," *Probability in the engineering and informational sciences*, vol. 14, no. 1, pp. 101–121, 2000.

[35] M. Kallen and J. van Noortwijk, "Optimal maintenance decisions under imperfect inspection," *Reliability engineering & system safety*, vol. 90, no. 2, pp. 177–185, 2005.

[36] J. M. van Noortwijk and P. H. van Gelder, "Optimal maintenance decisions for berm breakwaters," *Structural safety*, vol. 18, no. 4, pp. 293–309, 1996.

[37] M. Crowder and J. Lawless, "On a scheme for predictive maintenance," *European journal of operational research*, vol. 176, no. 3, pp. 1713–1722, 2007.

[38] J. van Noortwijk and H. Klatter, "Optimal inspection decisions for the block mats of the eastern-scheldt barrier," *Reliability engineering & system safety*, vol. 65, no. 3, pp. 203–211, 1999.

[39] X. Dai, S. Qu, H. Sui, and P. Wu, "Reliability modelling of wheel wear deterioration using conditional bivariate gamma processes and bayesian hierarchical models," *Reliability engineering & system safety*, vol. 226, p. 108710, 2022.

[40] K. Song and L. Cui, "A common random effect induced bivariate gamma degradation process with application to remaining useful life prediction," *Reliability engineering & system safety*, vol. 219, p. 108200, 2022.

[41] Z.-H. Lu, H.-P. Qiao, X.-Y. Zhang, and Y.-G. Zhao, "An innovative method for space-time-dependent reliability analysis," *Structural Safety*, vol. 102, p. 102326, 2023.

[42] L. Bordes, C. Paroissin, and A. Salami, "Parametric inference in a perturbed gamma degradation process," *Communications in statistics. Theory and methods*, vol. 45, no. 9, pp. 2730–2747, 2016.

[43] D. P. Kroese, S. Porotsky, and R. Y. Rubinstein, "The cross-entropy method for continuous multi-extremal optimization," *Methodology and computing in applied probability*, vol. 8, no. 3, pp. 383–407, 2006.

[44] F. Duan and G. Wang, "Planning of step-stress accelerated degradation test based on non-stationary gamma process with random effects," *Computers & industrial engineering*, vol. 125, pp. 467–479, 2018.

[45] M. Guida, F. Postiglione, and G. Pulcini, "A bayesian approach for non-homogeneous gamma degradation processes," *Communications in statistics. Theory and methods*, vol. 48, no. 7, pp. 1748–1765, 2019.

[46] D. Lu, M. D. Pandey, and W.-C. Xie, "An efficient method for the estimation of parameters of stochastic gamma process from noisy degradation measurements," *Proceedings of the Institution of Mechanical Engineers. Part O, Journal of risk and reliability*, vol. 227, no. 4, pp. 425–433, 2013.

[47] R. Y. Rubinstein and D. P. Kroese, *The Cross-Entropy Method : A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*, 1st ed., ser. Information Science and Statistics. New York, NY: Springer New York, 2004.

[48] R. Y. Rubinstein, A. Ridder, and R. Vaisman, *Fast Sequential Monte Carlo Methods for Counting and Optimization*, 1st ed. Hoboken, New Jersey: Wiley Publishing, 2013.

[49] G. Rubino and B. Tuffin, *Rare Event Simulation using Monte Carlo Methods*, 1st ed. Hoboken: Wiley, 2009.

[50] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method*, 3rd ed. New York: John Wiley & Sons, 2017.

[51] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*. New York, NY, USA: Cambridge University Press, 2002.

[52] B. D. Ripley, *Stochastic Simulation*. New York, NY, USA: John Wiley & Sons, Inc., 1987.

[53] S. Asmussen and P. Glynn, *Stochastic Simulation: Algorithms and Analysis*, ser. Stochastic Modelling and Applied Probability. Springer New York, 2007.

[54] R. Vaisman, "Finding minimum label spanning trees using cross-entropy method," *Networks*, vol. 79, no. 2, pp. 220–235, 2022.

[55] R. L. Berger, G. Casella, and R. L. Berger, *Statistical inference*, 2nd ed., ser. Duxbury advanced series. Pacific Grove, CA: Duxbury/Thomson Learning, 2002.

[56] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*, ser. Springer Texts in Statistics. New York, NY: Springer, 2013.

[57] B. Efron and R. Tibshirani, *An introduction to the bootstrap*, ser. Monographs on statistics and applied probability (Series) 57. New York: Chapman & Hall, 1993.

[58] M. Hamada, *Bayesian reliability*, 1st ed., ser. Springer series in statistics. New York, NY: Springer, 2008.

[59] S. Chow and J. Shao, "Estimating drug shelf-life with random batches," *Biometrics*, vol. 47, no. 3, pp. 1071–1079, 1991.