Advances in Reliability, Safety and Security, Part 3 – Kołowrocki, Dąbrowska (eds.)
© 2024 Polish Safety and Reliability Association, Gdynia, ISBN 978-83-68136-15-9 (printed), ISBN 978-83-68136-02-9 (electronic)

Advances in Reliability, Safety and Security

ESREL 2024
Monograph Book Series

Model Selection For Bounded Transformed Gamma Processes: Bayesian Approach

Massimiliano Giorgio[a], Fabio Postiglione[b], Gianpaolo Pulcini[c]

[a] *Dipartimento di Ingegneria Industriale, Università di Napoli Federico II, Napoli, Italy*
[b] *Dipartimento di Ingegneria dell'Informazione ed Elettrica e Matematica Applicata, Università degli Studi di Salerno, Fisciano (SA), Italy*
[c] *Istituto di Scienze e Tecnologie per Energia e Mobilità Sostenibili (STEMS), Consiglio Nazionale delle Ricerche (CNR), Napoli, Italy*

Abstract

The bounded transformed gamma process (BTGP) was recently proposed to describe degradation phenomena where the degradation growth is monotonic increasing and intrinsically bounded above. The BGTP is ruled by two monotone increasing functions, namely age and bounded state functions, whose functional form influences the behavior of the degradation model. In this paper, a Bayesian model selection procedure, based on the Bayes factor, is introduced to select the functional form of the bounded state function maximizing the marginal likelihood and thus providing the best fit to the available degradation data among suitable alternatives. The Bayesian model selection procedure involves prior information on the upper bound of the degradation phenomenon and on the behavior of the mean degradation function, and is performed by adopting some Markov Chain Monte Carlo procedures. The proposed approach is applied to a set of real data consisting of the wear measurements of the liners of an 8-cylinder Diesel engine for marine propulsion.

Keywords: model selection, Bayes factor, transformed gamma process, bounded degradation processes, Markov chain monte carlo method

1. Introduction

Many degradation phenomena affecting real-world technological units are intrinsically bounded, at least due to the finite size of deteriorating materials composing them. Nonetheless, only recently this aspect has been considered in the literature (Giorgio et al., 2015; Ling et al., 2015; Deng and Pandey, 2017), and detailed in (Fouladirad et al., 2023), where a new model called the bounded transformed gamma process (BTGP) has been proposed. The BTGP aims at modelling a monotonic increasing degradation phenomenon, where the degradation level can not exceed an upper limit U. The distinguishing feature of the BTGP, with respect to other existing bounded degradation models (Giorgio et al., 2015; Ling et al., 2015; Deng and Pandey, 2017), is that in the BTGP the bound U is treated as an unknown parameter which must be estimated from the available data.

The behavior of the BTGP is ruled by two specific functions, named bounded state function and age function. A maximum likelihood estimation (MLE) approach for the BTGP parameters has been presented in (Fouladirad et al., 2023), where different suitable functional models for the state and age functions have been evaluated and compared. In (Giorgio et al., 2023), a Bayesian inferential approach, based on Markov Chain Monte Carlo (MCMC) methods, has been introduced for the estimation of the parameters and functions thereof of the BTGP.

In this paper, we consider the issue of selecting the BTGP model providing the best fit to the degradation data under study in a Bayesian framework. More specifically, given some prior information on the upper bound and on the behavior of the mean degradation function, we compare BTGP models with different functional forms of the bounded state function and, by computing the Bayes factor, we select the one providing the best fit to the observed data. In particular, the Bayes factor quantifies the statistical evidence to prefer one model over the other (Morey et al., 2016). It can be thought of as a Bayesian analog to the likelihood-ratio test but, since it uses the (integrated) marginal likelihood rather than the maximized likelihood, decisions made on the base of these tests generally do not coincide (Lasaffre and Lawson, 2012). Moreover, differently from the likelihood ratio test, in addition to using prior information that the analyst possesses, the Bayes factor provides the evaluation of the evidence in favor of a competing model, rather than only allowing the analyst to reject the null hypothesis model or not (Ly et al., 2020). MCMC procedures are used for this purpose because the multivariate integrations involved in the computation of the marginal likelihoods can not be solved analytically. The proposed selection approach is finally applied to a set of real data consisting of wear measurements on the liners of an 8-cylinder engine of a cargo ship, already analyzed in (Fouladirad et al., 2023) and (Giorgio et al., 2023).

2. The bounded transformed gamma process

The bounded transformed gamma process (BTGP) $\{W(t);\ t \geq 0\}$ is a Markovian asymptotically bounded above monotonic increasing process with dependent increments (Fouladirad et al., 2023). The Markovian property implies that the BTGP is completely defined an initial condition, here $W(t) = W(0) = 0$, and by the conditional probability density function (pdf) of its increment $\Delta W(t, t+\tau)$ over the time interval $(t, t+\tau)$, given the current state $W(t) = w_t$, that is defined as:

$$f_{\Delta W(t,t+\Delta t)|W(t)}(\delta|w_t) = g'(w_t + \delta)\frac{[\Delta g(w_t, w_t + \delta)]^{\Delta\eta(t,t+\tau)-1}}{\Gamma[\Delta\eta(t,t+\tau)]}e^{-\Delta g(w_t,w_t+\delta)}, \qquad 0 < \delta < U - w_t\ , \tag{1}$$

where U is the upper bound, $g(w)$ is a non-negative, monotone increasing and differentiable function of the degradation level w, defined over the domain $0 \leq w < U$, with $g(0) = 0$ and $\lim_{w\to U} g(w) = \infty$, $g'(w_t + \delta)$ is the first derivative of $g(\cdot)$ evaluated at $w_t + \delta$, $\Delta g(w_t, w_t + \delta) = g(w_t + \delta) - g(w_t)$, $\eta(t)$ is a non-negative, monotone increasing function, defined on the domain $0 \leq t < \infty$, with $\eta(0) = 0$ and $\lim_{t\to\infty}\eta(t) = \infty$, $\Delta\eta(t, t+\tau) = \eta(t+\tau) - \eta(t)$, and $\Gamma[\Delta\eta(t, t+\tau)]$ is the complete gamma function.

The functions $\eta(\cdot)$ and $g(\cdot)$ are called age and bounded state function, respectively. Note that the function $g(\cdot)$ is not bounded above, but it is called "bounded" because it is the state function of a bounded process. From (1), the conditional cumulative distribution function (Cdf) of $\Delta W(t, t+\tau)$, given $W(t) = w_t$, is:

$$F_{\Delta W(t,t+\tau)|W(t)}(\delta|w_t) = \frac{\gamma\big(\Delta g(w_t,w_t+\delta); \Delta\eta(t,t+\tau)\big)}{\Gamma\big(\Delta\eta(t,t+\tau)\big)}, 0 < \delta < U, \tag{2}$$

where $\gamma(x; a) = \int_0^x z^{a-1}e^{-z}\mathrm{d}z$ is the (lower) incomplete gamma function. From (1) and (2), by substituting $\Delta g(w_t, w_t + \delta)$, $g'(w_t + \delta)$, and $\Delta\eta(t, t+\tau)$ by $g(w)$, $g'(w)$, and $\eta(t)$, respectively, the pdf $f_{W(t)}(w)$ of the degradation level $W(t)$ of a new unit, being $W(t) = \Delta W(0, t)$, can be immediately retrieved:

$$f_{W(t)}(w) = g'(w)\frac{[g(w)]^{\eta(t)-1}}{\Gamma(\eta(t))}e^{-g(w)},\ \ 0 < w < U\ \ , \tag{3}$$

while the Cdf $F_{W(t)}(w)$ is given by:

$$F_{W(t)}(w) = \begin{cases} \gamma\big(g(w); \eta(t)\big)/\Gamma\big(\eta(t)\big)\,, & 0 < w < U \\ 1 & w \geq U \end{cases}\ . \tag{4}$$

To completely define the BTGP, we need to assign functional forms to $\eta(\cdot)$ and $g(\cdot)$. Different suitable forms for the age function are available for the transformed gamma processes and can be used also for the BTGP, see (Fouladirad et al., 2023). In this paper, we focus on the well-known and ductile power-law function:

$$\eta(t) = (t/a)^b\ , \tag{5}$$

where the parameter a is the (time) scale parameter, while the exponent b governs the shape of the age function and is a key parameter ruling the behavior of the mean degradation over time t, as discussed below.

Regarding the state function, the following three functional forms are proposed for the bounded state function among other possible choices, see also (Fouladirad et al., 2023):

$$g_1(w) = -\beta \ln\left(1 - \frac{w}{U}\right),\ \ 0 \leq w < U, \tag{6}$$

$$g_2(w) = \beta\ \frac{w/U}{1-w/U}\ ,\ \ \ 0 \leq w < U\ , \tag{7}$$

$$g_3(w) = \beta \tan\left(\frac{\pi}{2}\frac{w}{U}\right),\ \ 0 \leq w < U\ , \tag{8}$$

providing a reasonable trade-off between flexibility and model simplicity. Their first derivatives are:

$$g_1'(w) = \frac{\beta}{U-w}\,,\ \ 0 \leq w < U\ , \tag{9}$$

$$g_2'(w) = \frac{\beta\, U}{(U-w)^2}\,, \quad 0 \le w < U\,, \tag{10}$$

$$g_3'(w) = \frac{\beta}{U}\frac{\pi/2}{\cos^2\left(\frac{\pi w}{2U}\right)}\,, \quad 0 \le w < U, \tag{11}$$

respectively. Henceforth, we refer to as $BTGP_1$, $BTGP_2$, and $BTGP_3$ models with the age function in (5) and bounded state function $g_1(w)$, $g_2(w)$, and $g_3(w)$ reported in (6), (7), (8), respectively. It is worth noting that the parameter U acts as scale parameter on the wear axis, while β is a multiplicative constant.

With this state function, mean and variance of the degradation level $W(t)$ are not available in a closed form. However, they can be easily computed by univariate numerical integrations as follows:

$$E\{W(t)\} = \int_0^U w\, f_{W(t)}(w)\mathrm{d}w\,, \tag{12}$$

$$V\{W(t)\} = \int_0^U w^2 f_{W(t)}(w)\mathrm{d}w - E^2\{W(t)\}\,. \tag{13}$$

In Figure 1, the behavior of the mean degradation function $E\{W(t)\}$ of the model $BTGP_1$ is depicted for $U = 10$, $a = 1$, and some selected values of β and b.

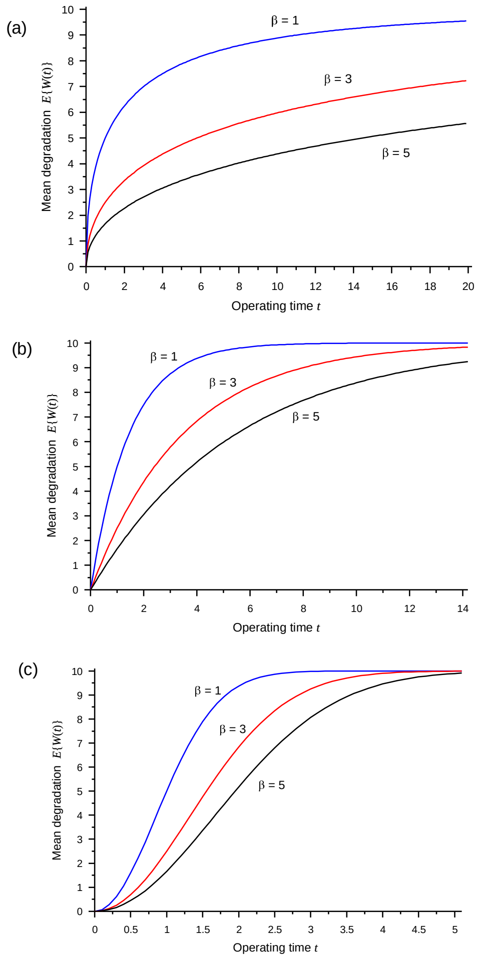

Fig 1. The behavior of the mean function of the BTGP model M_1,where $U = 10$, $a = 1$, for some selected values of β, and for (a) $b = 0.5$; (b) $b = 1.0$; (c) $b = 2.0$.

It is possible to notice that the mean function $E\{W(t)\}$ is concave from the beginning when $b < 1$, and its first derivative with respect to t decreases monotonically with t (see Figure 1a), while it initially grows almost linearly, and then becomes concave, when $b = 1$ (see Figure 1b). Remarkably, the mean function shows an inflection point when $b > 1$ (see Figure 1c), i.e. when $\eta(t)$ is convex, while the parameter β has no effect on the behavior of the mean function, because it acts almost like a time scale parameter. A similar behavior is exhibited by the models BTGP$_2$ and BTGP$_3$ highlighting that the parameter b is a key parameter governing the shape of $E\{W(t)\}$ over t.

3. The Bayesian procedure

To incorporate the prior knowledge of the analyst derived from her/his experience with similar degradation processes into the estimation procedure, we adopt a Bayesian model selection approach where different kind of prior information on the observed phenomenon can be included. After presenting the likelihood function of wear data, both vague and informative priors are proposed, depending on the available prior information. All the proposed priors are proper distributions because the Bayes factor can be computed only when all priors are proper.

3.1. The likelihood function

Let us suppose that m identical degrading units are operating under the same working conditions and that the unit i $(i = 1, \dots, m)$ is inspected n_i times at the ages $t_{i,j}$ $(j = 1, \dots, n_i)$. Let $w_{i,j}$ be observed value of the degradation level $W(t_{i,j})$ of the unit i measured at the inspection time $t_{i,j}$. The conditional pdf of $\Delta W(t_{i,j-1}, t_{i,j})$, given $W(t_{i,j-1}) = w_{i,j-1}$, can be derived from (1) as follows:

$$f_{\Delta W(t_{i,j-1},t_{i,j})|W(t_{i,j-1})}\left(\delta_{i,j}\middle|w_{i,j-1}\right) = g'\left(w_{i,j-1} + \delta_{i,j}\right)\frac{\left(\Delta g_{i,j}\right)^{\Delta\eta_{i,j}-1}}{\Gamma(\Delta\eta_{i,j})}e^{-\Delta g_{i,j}}, \quad 0 < \delta_{i,j} < U - w_{i,j-1}\ , \tag{14}$$

where $g'(w)$ is one of the derivatives in (9)-(11), $\delta_{i,j} = w_{i,j} - w_{i,j-1}$, $\Delta g_{i,j} = \Delta g\left(w_{i,j-1}, w_{i,j-1} + \delta_{i,j}\right)$ where $g(w)$ has one of the functional forms in (6)-(8) and $\Delta\eta_{i,j} = (t_{i,j}/a)^b - (t_{i,j-1}/a)^b$, with $w_{i,0} = t_{i,0} = 0$ for all i.

Given the data vector $\mathbf{w} = (w_{1,1}, \dots, w_{i,n_i}, \dots, w_{m,1},\ \dots\ ,\ w_{m,n_i})$, the likelihood function is then:

$$L(\mathbf{w};\boldsymbol{\theta}) = \prod_{i=1}^{m}\prod_{j=1}^{n_i} f_{\Delta W(t_{i,j-1},t_{i,j})|W(t_{i,j-1})}\left(\delta_{i,j}\middle|w_{i,j-1}\right), \tag{15}$$

where $\boldsymbol{\theta} = (a, b, \beta, U)$ is the vector of model parameters.

3.2. The prior information

The prior information that we assume the analyst may possess refers to some physical characteristics that are relevant for a bounded degradation process and can be retrieved from analyst experience.

In particular, for bounded degradation processes, the upper bound U is a key parameter. One basic information available for U is that it must be greater than the maximum degradation level w_{M} observed in the current study or in past studies involving similar degrading units. Other prior information might be eventually available depending on the application.

Several proper prior distributions on U can be reasonably proposed for technological applications, depending on the degrees of knowledge of the analyst on the application itself. Some of them are listed below:

- no information is available on U, except that $U > w_{\mathrm{M}}$, and hence the (vague) 2-parameter exponential prior, with location parameter equal to w_{M}, is used:

  $$\pi(U) = \lambda \exp[-\lambda\,(U - w_M)], \qquad U > w_{\mathrm{M}}\ , \tag{16}$$

  where the parameter λ is sufficiently small with respect to w_{M} (say, $\lambda = w_{\mathrm{M}}/50$) in order to ensure that the prior (18) is truly vague;
- an interval $(w_{\mathrm{L}}, w_{\mathrm{U}})$ of equally probable values for U, with $w_{\mathrm{L}} \geq w_{\mathrm{M}}$, can be formulated on U, and hence the uniform prior over the interval $(w_{\mathrm{L}}, w_{\mathrm{U}})$ is used:

  $$\pi(U) = \frac{1}{w_{\mathrm{U}} - w_{\mathrm{L}}}, \quad w_{\mathrm{L}} \leq U \leq w_{\mathrm{U}}\ ; \tag{17}$$
- the analyst can provide a prior value of the mean $E\{U\}$ and the variance $V\{U\}$ of U, under the constraint $U > w_{\mathrm{M}}$. Thus, the following 3-parameter gamma distribution, with location parameter equal to w_{M}, is used:

$$\pi(U) = \frac{q^p (U - w_M)^{p-1}}{\Gamma(p)} e^{-q(U - w_M)}, \quad U > w_M, \tag{18}$$

where the prior parameters q and p can be computed as $q = (E\{U\} - w_M)/V\{U\}$ and $p = q\,(E\{U\} - w_M)$.

Another kind of information the analyst can derive from her/his experience is pertinent to the shape of the mean degradation function $E\{W(t)\}$, more specifically she/he can provide some hints on the presence of an inflection point of $E\{W(t)\}$. This information can be converted into a prior information on the shape parameter b of the age function $\eta(t) = (t/a)^b$ in (5) because, as discussed in Section 2, the mean function $E\{W(t)\}$ shows an inflection point when b is larger than 1, it is concave from the beginning when b is lower than 1, while it initially increases linearly when b is equal to 1. Several proper prior distributions on b can be thus proposed based on the degree of knowledge on the shape of the mean function $E\{\Delta W(t)\}$, as follows:

- the only prior information on $E\{\Delta W(t)\}$ is that it has no inflection point, and thus the uniform prior is used:

$$\pi(b) = 1, \qquad 0 \le b \le 1; \tag{19}$$

- it is known that the mean function $E\{\Delta W(t)\}$ has no inflection point and it is also possible to formulate a prior mean $E\{b\}$ and a prior variance $V\{b\}$ of b. Thus, assumed $b \le 1$, the following Beta prior is adopted:

$$\pi(b) = \frac{b^{r-1}(1-b)^{s-1}}{\mathrm{B}(r,s)}, \qquad 0 \le b \le 1, \tag{20}$$

  where the parameters r and s can be computed by $r = E^2\{b\}(1 - E\{b\})/V\{b\} - E\{b\}$ and $s = r/E\{b\} - r$, respectively;

- the analyst knows that $E\{\Delta W(t)\}$ has an inflection point and she/he can also provide a value for the prior mean $E\{b\}$ and the variance $V\{b\}$ of b. Then, assumed $b > 1$, the following 3-parameter gamma distribution, with location parameter equal to 1, is adopted:

$$\pi(b) = \frac{s^r (b-1)^{r-1}}{\Gamma(r)} e^{-s(b-1)}, \quad b > 1, \tag{21}$$

  where the parameters s and r can be computed as $s = (E\{b\} - 1)/V\{b\}$ and $r = s\,(E\{b\} - 1)$.

- the analyst knows that $E\{\Delta W(t)\}$ initially increases almost linearly and she/he can also provide a value for the variance $V\{b\}$ of b. Thus, the following 1-parameter gamma distribution, with unit mean and variance $V\{b\}$, is adopted:

$$\pi(b) = \frac{r^r b^{r-1}}{\Gamma(r)} e^{-rb}, \quad b > 0, \tag{22}$$

  where the parameter r is given by $r = 1/V\{b\}$.

Regarding the other two process parameters β and a, we assume that no information is available corresponding to uniform vague prior pdfs over the intervals $(0, \beta_U)$ and $(0, a_U)$ for β_U and a_U large enough values, respectively:

$$\pi(\beta) = 1/\beta_U \quad \text{and} \quad \pi(a) = 1/a_U\,. \tag{23}$$

4. Model selection and Bayes factor

The Bayesian inferential approach discussed in Section 3 can be applied to any bounded transformed bounded process. For a given data set, we have three competing degradation models, say BTGP_1, BTGP_2 and BTGP_3, to compare. More specifically, we want to select the model which provides the best fit for the available data and hence good estimates and predictions. Obviously, the proposed approach can be easily extended to compare other BTGP models with state or age functions different from those analyzed in this paper.

For model selection in a Bayesian framework, different strategies are available. Some of them focus on the predictive accuracy of the models, measured by cross-validation approaches, or by computing some popular information-based metrics among which we mention the Deviance Information Criterion (DIC) presented in (Spiegehalter et al., 2002), and the Watanabe-Akaike or Wide Applicable Information Criterion (WAIC) introduced in (Watanabe, 2010). For further readings, see (Gelman et al., 2014).

Other ways to compare models is through the quantification of the evidence in favor of one model with respect to the others, typically based on the Bayes factor, introduced by (Jeffreys, 1961). Given two competing models, say M_1 and M_2, that may have generated the observed data $\mathbf{w}$, a Bayesian approach to model selection between the two models relies on the posterior probabilities:

$\Pr\{M_1|\mathbf{w}\}$ and $\Pr\{M_2|\mathbf{w}\} = 1 - \Pr\{M_1|\mathbf{w}\}$, (24)

that represent a measure of the evidence in favor of models M_1 and M_2, respectively, given the data $\mathbf{w}$. After assigning the prior probabilities $\Pr\{M_1\}$ and $\Pr\{M_2\} = 1 - \Pr\{M_1\}$ of the two models, it is possible to compute the posterior probability of model M_k ($k = 1, 2$) as follows:

$$\Pr\{M_k|\mathbf{w}\} = \frac{\pi\{\mathbf{w}|M_k\} \Pr\{M_k\}}{\pi\{\mathbf{w}|M_1\} \Pr\{M_1\} + \pi\{\mathbf{w}|M_2\} \Pr\{M_2\}} \ , \ k = 1, 2 \ , \tag{25}$$

where $\pi\{\mathbf{w}|M_k\}$, frequently called the marginal likelihood for the model M_k, is the marginal probability (distribution) of the observed data $\mathbf{w}$ under the model M_k ($k = 1, 2$). The marginal likelihood $\pi\{\mathbf{w}|M_k\}$ can be computed by a (multiple) integration over the parameter space $\Theta^{(k)}$ of the product of the likelihood function $L(\mathbf{w}|\boldsymbol{\theta}^{(k)}; M_k)$ under M_k and the (joint) prior distribution $\pi(\boldsymbol{\theta}^{(k)}|M_k)$ of the vector $\boldsymbol{\theta}^{(k)}$ of the parameters of the model M_k, viz.

$$\pi\{\mathbf{w}|M_k\} = \int_{\Theta^{(k)}} L(\mathbf{w}|\boldsymbol{\theta}^{(k)}; M_k)\, \pi(\boldsymbol{\theta}^{(k)}|M_k)\, \mathrm{d}\boldsymbol{\theta}^{(k)} \ , \quad k = 1, 2. \tag{26}$$

Using the posterior probabilities (25), a measure of the evidence provided by the available data $\mathbf{w}$ in favor of model M_1 over model M_2 is the Bayes factor B_{12} that is defined as the ratio of posterior odds of M_1 (with respect to M_2) and its prior odds, see for instance (O'Hagan and Forster, 2004):

$$B_{12} = \frac{\pi\{\mathbf{w}|M_1\}}{\pi\{\mathbf{w}|M_2\}} = \frac{\Pr\{M_1|\mathbf{w}\} / \Pr\{M_2|\mathbf{w}\}}{\Pr\{M_1\} / \Pr\{M_2\}} \ , \tag{27}$$

where the second equality holds after (25) and suggests that the Bayes factor can be also expressed in terms of ratio between posterior $\Pr\{M_1|\mathbf{w}\} / \Pr\{M_2|\mathbf{w}\}$ and prior $\Pr\{M_1\} / \Pr\{M_2\}$ odds of model M_1 and M_2, see also (Gill, 2002; Campbell and Gustafson, 2022). It is worth noting that the marginal probabilities (26) depend on the prior distribution $\pi(\boldsymbol{\theta}^{(k)}|M_k)$ of the model parameters.

Large values of the Bayes factor B_{12} provide evidence in favor of model M_1, while small values (or, equivalently, large values of $B_{21} = 1/B_{12}$) provide evidence against M_1. Different criteria to evaluate the evidence in favor of model M_1 over M_2 are available. The following guidelines, based on the metric $2\ln(B_{12})$, are provided in (Kass and Raftery, 1995):

- if the value of $2\ln(B_{12})$ is in the range [0, 2], the evidence in favor of M_1 is not worth than a bare mention;
- for a value of $2\ln(B_{12})$ in the range [2, 6] there is positive evidence in favor of M_1;
- there is strong evidence for a value of $2\ln(B_{12})$ in the range [6, 10];
- there is very strong evidence in favor of M_1 for $2\ln(B_{12}) > 10$.

Obviously, working with B_{21} is possible to evaluate the evidence in favor of M_2 (i.e., against M_1), by using the same rules. To correctly adopt the Bayes factor, some recommendations are necessary. The prior distributions for the parameters of each model must be specified and, as anticipated in Section 3, all the prior distributions must be proper to compute the marginal likelihoods $\pi\{\mathbf{w}|M_k\}$ of the data $\mathbf{w}$ under the model M_k ($k = 1, 2$). Otherwise, all the $\pi\{\mathbf{w}|M_k\}$, and thus the Bayes factor, would be computed up to an undefined multiplicative constant. Consequently, if no information for a model parameter is available, a proper non-informative or weakly informative prior distribution must be used. Moreover, the Bayes factor is quite sensitive to the choices of prior distributions. For instance, if the prior distribution of a parameter of the model M_1 is strong but wrong, the Bayes factor B_{12} can provide evidence against the model M_1 even if this model is the most appropriate to describe the observed data. Thus, strong prior distributions are suggested to be used only if the prior distribution is expressed in terms of a parameter that indexes both the models, while the prior distribution on parameters indexing only one of the competing models must be non-informative or weakly informative.

5. A Markov Chain Monte Carlo procedure for Bayes factor estimation

In most practical scenarios, Bayesian methods require challenging and highly time-consuming numerical integrations, like those involved in the calculation of the marginal probabilities (26). To circumvent this problem, different procedures are available to compute the Bayes factors by using MCMC sampling methods, see for instance (Christensen et al., 2011; Lodewyckx et al., 2011). Some of them, named transdimensional MCMC, rely on combining the K models to be compared within a hierarchical supermodel forming a mixture model where each model is indexed by M_k (Carlin and Chib, 1995; Green, 1995), whose prior distribution is $\pi(k) = \Pr\{M_k\}$, $k \in \{1, \dots, K\}$. In general transdimensional MCMC settings, models can have parameters $\boldsymbol{\theta}^{(k)}$ defined on parameter spaces $\Theta^{(k)}$ ($k = 1, \dots, K$) that may differ from model to model. In this case, all the parameters can be arranged in a parameter vector $\boldsymbol{\theta} = \left(\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(K)}\right)$, that is an element of the parameter space $\Theta = \Theta^{(1)} \times \ldots \times \Theta^{(K)}$.

However, in this paper we consider BTGP models with the same parameters, and the overall parameter vector is simply that of single models, i.e. $\boldsymbol{\theta} = \boldsymbol{\theta}^{(1)} = \cdots = \boldsymbol{\theta}^{(K)}$. The full mixture supermodel of data and model parameters can be written as:

$$\pi\{\mathbf{w}, \boldsymbol{\theta}\} = \sum_{k=1}^{K} \pi\{\mathbf{w}, \boldsymbol{\theta} | M_k\}\, \pi(k), \tag{28}$$

where $\pi\{\mathbf{w}, \boldsymbol{\theta} |\, M_k\} = L(\mathbf{w}|\boldsymbol{\theta}; M_k)\pi(\boldsymbol{\theta}|M_k)$ is the joint probability distribution due to the Bayesian model M_k. Similarly to (26), the marginal likelihood $\pi\{\mathbf{w}|M_k\}$ of the BTGP model M_k is given by:

$$\pi\{\mathbf{w}|M_k\} = \int_{\Theta} L(\mathbf{w}|\boldsymbol{\theta}; M_k)\, \pi(\boldsymbol{\theta}|M_k) \mathrm{d}\boldsymbol{\theta}. \tag{29}$$

From (27), for mixture model in (28), the Bayes factor B_{kh} of model M_k with respect to model M_h ($k, h = 1, \ldots, K$ and $k \neq h$) can be calculated by taking the ratio between $\pi\{\mathbf{w}|M_k\}$ and $\pi\{\mathbf{w}|M_h\}$.

However, to avoid the high-dimensional numerical integrations in (29) by exploiting the second equality in (27), we compute the Bayes factor as:

$$B_{kh} = \frac{\pi\{M_k|\mathbf{w}\}/\, \pi\{M_h|\mathbf{w}\}}{\pi\{k\}/\, \pi\{h\}}, \tag{30}$$

where posterior model probabilities are easily estimated by MCMC posterior sampling methods.

Indeed, a simple estimator of the posterior probability of each model M_k ($k = 1, \ldots, K$) is:

$$\widehat{\Pr}\{M_k|\mathbf{w}\} = \frac{\text{number of occurences of model } M_k}{\text{total number of iterations}}, \tag{31}$$

as suggested in (Lodewyckx et al., 2011).

A convenient common choice for prior probability of model M_k is to assume that $\pi(k) = 1/K$, $\forall k \in \{1, \ldots, K\}$, accounting for equally likely models.

More specifically, for each model M_k ($k = 1, \ldots, K$), the MCMC algorithm generates a five-dimensional pseudo-random vector sample of size N, that is $\left(\boldsymbol{\theta}_j, M_{k,j}\right) = \left(a_j, b_j, \beta_j, U_j, M_{k,j}\right)$ ($j = 1, \ldots, N$), from the joint posterior pdf $\pi(a, b, \beta, U, M_k|\boldsymbol{w}) = \pi(a, b, \beta, U|M_k, \boldsymbol{w})\, \pi(M_k|\boldsymbol{w})$, where the joint posterior pdf of the model parameters is:

$$\pi(a, b, \beta, U|M_k, \boldsymbol{w}) \propto L(\boldsymbol{w}|a, b, \beta, U, M_k)\, \pi(U|M_k)\, \pi(b|M_k), \tag{32}$$

where priors $\pi(U|M_k)$ and $\pi(b|M_k)$, in the considered case, do not depend on k.

The sample is collected after a sufficiently large burn-in period to make negligible the influence of the starting point of the MCMC and of the choice of $\pi(k)$, which the Bayes factor does not depend on.

After collecting the vector sample $(\boldsymbol{\theta}_j, M_{k,j})$, we can estimate the posterior probability of each model M_k by (31), and then we can estimate the Bayes factor by using (30).

6. Numerical application

We apply the proposed approach to select the model providing the best fit, between the three competing statistical models BTGP_1, BTGP_2 and BTGP_3, for the wear measurements of the liners of an 8-cylinder marine engine in Table 1. Observed data are depicted in Figure 2, where data pertaining to the same wear path are connected by lines.

We then assume that the analyst, based on previously observed similar degradation phenomena, knows that the upper bound U is surely larger that $w_{\mathrm{M}} = 4.3$ mm, and the mean degradation function has an inflection point, guaranteeing that $b > 1$. Moreover, the analyst can also provide the prior values of the mean and variance of U and of b: $E\{U\} = 4.6$ mm and $V\{U\} = 0.09$ mm^2, $E\{b\} = 1.5$ and $V\{b\} = 0.09$. Thus, the prior pdfs for the BTGP parameters are (Giorgio et al., 2023):

- the 3-parameter gamma prior (18) on U, with parameters $q = (E\{U\} - w_{\mathrm{M}})/V\{U\} = 3.333$ and $p = q(E\{U\} - w_{\mathrm{M}}) = 1.0$, and
- the 3-parameter gamma prior (21) on b, with parameters $s = (E\{b\} - 1)/V\{b\} = 5.556$ and $r = s(E\{b\} - 1) = 2.778$.

The prior parameters β_U and α_U are set equal to 200 and 50,000 h, respectively, to guarantee that the prior distributions on α and β include a very large portion of the range supported by the likelihood. Finally, the $K = 3$ BTGP models are considered equally likely, and thus the prior probabilities for M_k are $\pi(k) = 1/3$, $k = 1, 2, 3$.

Thus, the assumed joint prior pdf of a, b, β, and U, given M_k, is:

$$\pi(a,b,\beta,U|M_k) \propto \frac{q^p (U-w_{\mathrm{M}})^{p-1}}{\Gamma(p)} \frac{s^r (b-1)^{r-1}}{\Gamma(r)} e^{-q(U-w_{\mathrm{M}})-s(b-1)}, \quad a < 50.000,\ b > 1, \beta < 200,\ U > w_{\mathrm{M}}, \tag{33}$$

with $p = 1.0$, $q = 3.333$, $w_{\mathrm{M}} = 4.3$ mm, $r = 2.778$, and $s = 5.556$.

Table 1. Wear $w_{i,k}$ [mm] accumulated by liner i up to the inspection time $t_{i,k}$ [h].

i	$t_{i,1}$	$w_{i,1}$	$t_{i,2}$	$w_{i,2}$	$t_{i,3}$	$w_{i,3}$	$t_{i,4}$	$t_{i,4}$
1	11,300	0.90	14,680	1.30	31,270	2.85		
2	11,300	1.50	21,970	2.00				
3	12,300	1.00	16,300	1.35				
4	14,810	1.90	18,700	2.25	28,000	2.75		
5	10,000	1.20	30,450	2.75	37,310	3.05		
6	6,860	0.50	17,200	1.45	24,710	2.15		
7	2,040	0.40	12,580	2.00	16,620	2.35		
8	7,540	0.50	8,840	1.10	9,770	1.15	16,300	2.10

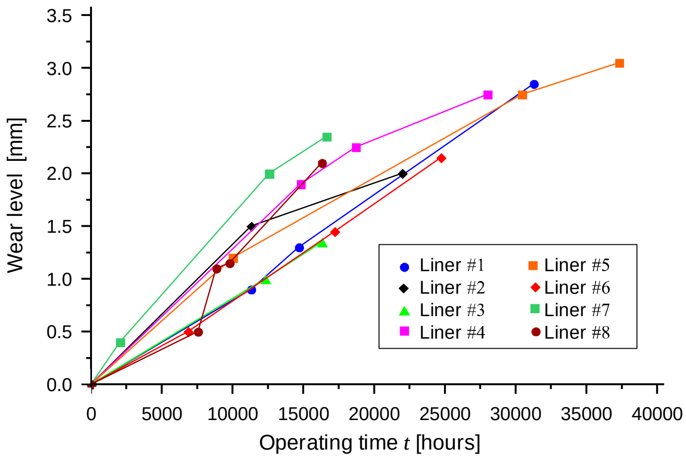

Fig. 2. Observed wear paths of the liners.

Table 2. Bayes factors estimates for different couples of BTGP models.

Comparing models	$\hat{B}_{kh}$	$2\ln(\hat{B}_{kh})$	Interpretation
2 vs 1	9.755	4.556	positive evidence in favor of BTGP_2 against BTGP_1
3 vs 1	3.440	2.471	positive evidence in favor of BTGP_3 against BTGP_1
2 vs 3	2.836	2.085	positive evidence in favor of BTGP_2 against BTGP_3

To collect posterior samples of (a,b,β,U,M_k) composed by $N = 3 \cdot 10^5$ five-dimensional vector elements, we used a burn-in period of 10^5 iterations and a thinning interval equal to 200, guaranteeing convergence of the MCMC algorithm to the target distribution, and very good mixing.

From (31), we compute the posterior probabilities for the three BTGP models and compute all the Bayes factors for each couple of models by (30) that can be rewritten as

$$\hat{B}_{kh} = \frac{\widehat{\Pr}\{M_k|\mathbf{w}\}}{\widehat{\Pr}\{M_h|\mathbf{w}\}}, \ \forall k,h \in \{1,2,3\}, \text{ and } k \neq h\,, \tag{34}$$

since all the three models have the same prior probability. The estimated Bayes factors in Table 2 show that, according to the rules recalled in Section 4, positive evidence exists in favor of the BTGP_2 model (that is, the BGTP model with the bounded state function (7)), against the other two competing models having state functions (6) and (8), respectively, and hence we can conclude that, given the selected prior distributions, the BTGP_2 model provides the best fit to the wear data in Table 1.

Computations to obtain the results reported in Table II have been performed by routines implemented in OpenBUGS (Lunn et al., 2009).

7. Conclusions

In this work, a Bayesian procedure based on the Bayes factor has been developed for selecting, among several bounded transformed gamma processes (BTGP) with different state functions, the model that provides the best fit for an observed data set. Some prior information on the upper bound U for the degradation level and on the shape of the mean degradation function are assumed to be available. Computations have been performed by adopting a transdimensional Monte Carlo Markov Chain technique. The proposed approach has been applied to a set of real wear data of liners of an 8-cylinder marine engine, showing the feasibility of the suggested Bayesian model selection procedure.

Acknowledgements

The Authors would like to thank the Editor and the anonymous referees for their suggestions, which help to improve the quality of the manuscript.

References

Campbell, H., Gustafson, P. 2022. Bayes factors and posterior estimations: two sides of the very same coin. The American Statistician 77(3), 248-258.

Carlin, B.P., Chib, S. 1995. Bayesian model choice via Markov chain Monte Carlo methods. Journal of the Royal Statistical Society, Series B, 57, 473–484.

Christensen, R., Johnson, W., Branscum, W., Hanson, T.E. 2011. Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians. CRC Press, Boca Raton.

Deng, Y., Pandey, M.G. 2017. Modelling of a bounded stochastic degradation process based on a transformed gamma process. Walls, L., Revei, M., Bedford, T. (Eds.), Risk, Reliability and Safety: Innovating Theory and Practice, Proceedings of ESREL 2016. CRC Press, Taylor & Francis Group, London, 1417–1424.

Fouladirad, M., Giorgio, M., Pulcini, G. 2023. A transformed gamma process for bounded degradation phenomena. Quality and Reliability Engineering International 39(2), 546–564.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B. 2014. Bayesian Data Analysis. Third Edition. CRC Press, Boca Raton.

Gill, J. 2015. Bayesian Hypothesis Testing and the Bayes Factor. Bayesian Methods: A Social and Behavioral Sciences Approach, Third Edition. CRC Press, Boca Raton, 207–246.

Giorgio, M., Guida, M., Pulcini, G. 2015. A new class of Markovian processes for deteriorating units with state dependent increments and covariates. IEEE Transactions on Reliability 64(2), 562–578.

Giorgio, M., Postiglione, F., Pulcini, G. 2023. Bayesian inference for the bounded transformed gamma process. Brito, M.P., Terje, A., Baraldi, P., Cepin, M., Zio E. (Eds.), Proceedings of ESREL 2023. Research Publishing, Singapore, 1192–1199.

Green, P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Jeffreys, H. 1961. Theory of Probability. Oxford University Press, Oxford.

Kass, R.E., Raftery, A.E. 1995. Bayes factors. Journal of the American Statistical Association 90(430), 773-795.

Lesaffre, E., Lawson, A.B. 2012. Bayesian hypothesis testing. Bayesian Biostatistics. John Wiley and Sons, Somerset, 72–78.

Ling, M., Tsui, H.K.L., Balakrishnan N. 2015. Accelerated degradation analysis for the quality of a system based on the gamma process. IEEE Transactions on Reliability 64(1), 463–472.

Lodewyckx, T., Kim, W., Lee, M.D., Tuerlinckx, F., Kuppens, P., Wagenmakers, E.-J. 2011. A tutorial on Bayes factor estimation with the product space method. Journal of Mathematical Psychology 55(11), 331-347.

Lunn, D., Spiegelhalter, D.J., Thomas, A., Best, N.G. 2009. The BUGS project: Evolution, critique and future directions (with discussion). Statistics in Medicine 28(25), 3049–3082.

Ly, A., Stefan, A., van Doorn, J., Dablander, F., van den Bergh, D., Sarafoglou, A., Kucharsky, S., Derks, K., Gronau, Q.F., Raj, A., Boehm, U., van Kesteren, E.-J., Hinne, M., Matzke, D., Marsman, M., Wagenmakers, E.-J. 2020. The Bayesian Methodology of Sir Harold Jeffreys as a Practical Alternative to the P Value Hypothesis Test. Computational Brain and Behavior 3(2), 153–161.

Morey, R.D., Romeijn, J.-W., Rouder, J.N. 2016. The philosophy of Bayes factors and the quantification of statistical evidence. Journal of Mathematical Psychology 72, 6–18.

O'Hagan, A., Forster, J. 2004. Kendall's advanced theory of statistics, Volume 2B: Bayesian inference. Second Edition. Arnold Press, London.

Spiegehalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. 2002. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society B 44, 583-639.

Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research 11, 3571-3594.