



HAL
open science

Model selection for degradation modeling and prognosis with health monitoring data

Thi Phuong Khanh Nguyen, Mitra Fouladirad, Antoine Grall

► **To cite this version:**

Thi Phuong Khanh Nguyen, Mitra Fouladirad, Antoine Grall. Model selection for degradation modeling and prognosis with health monitoring data. Reliability Engineering and System Safety, 2018, 169, pp.105-116. 10.1016/j.ress.2017.08.004 . hal-01886819

HAL Id: hal-01886819

<https://hal.science/hal-01886819>

Submitted on 3 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in: <http://oatao.univ-toulouse.fr/19980>

Official URL: <http://dx.doi.org/10.1016/j.res.2017.08.004>

To cite this version:

Nguyen, Thi Phuong Khanh  and Fouladirad, Mitra and Grall, Antoine *Model selection for degradation modeling and prognosis with health monitoring data.* (2017) Reliability Engineering and System Safety, 169. 105-116. ISSN 0951-8320

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Model selection for degradation modeling and prognosis with health monitoring data

Khanh T.P. Nguyen*, Mitra Fouladirad, Antoine Grall

ROSAS Department, University of Technology of Troyes, France

A B S T R A C T

Keywords:

Model selection
Prognostic prediction
Residual life prediction
Degradation process
Lévy process

Health monitoring data are increasingly collected and widely used for reliability assessment and lifetime prediction. They not only provide information about degradation state but also could trace failure mechanisms of assets. The selection of a deterioration model that optimally fits in with health monitoring data is an important issue. It can enable a more precise asset health prognostic and help reducing operation and maintenance costs. Therefore, this paper aims to address the problem of degradation model selection including goals, procedure and evaluation criteria. Focusing on continuous degradation modeling including some currently used Lévy processes, the performance of classical and prognostic criteria are discussed through numerous numerical examples. We also investigate in what circumstances which methods perform better than others. The efficiency of a new hybrid criterion is highlighted that allows to take into account the information of goodness-of-fit of observation data when evaluating prognostic measure.

1. Introduction

Degradation modeling in the presence of health monitoring data is extremely important for lifetime prognosis and maintenance planning. Complex models permit to take advantage of all available information and describe precisely the dynamics of degradation. However, these models are not easily tractable, and their calibration in the presence of data is a burdensome task. On the other hand, a very simple model which can be easily fitted to data but can underestimate or overstate the uncertainty around the lifetime prediction. This latter can induce risks and additional costs in prognosis based decision making and maintenance. A useful and suitable degradation model leads to a balance between accuracy and tractability, [1,2].

The degradation considered as a random phenomenon often has a gradual time-continuous trajectory. Regarding the system under consideration, the degradation can take values in discrete or continuous space. For instance, in a crack growth phenomenon, the crack length can take infinite possible values as soon as it begins to grow. Similarly, a deteriorating production process can have several quality states which will impact the production and result of gain or losses. In these two cases, the modeling procedure should take into account the phenomenon under consideration, see for instance [3–5].

This paper focuses on the gradual degradation modeling and prognosis with health monitoring data. When data is available, the important issue is to select the model which describes the underlying degradation

phenomenon in the best possible way. The data are collected under given environmental conditions and may not represent the average behavior of the deteriorating system. A suitable model is one who can take into account the possibility of extreme behaviors during data collection without losing in perspective the real average degradation behavior. The final use of the degradation model can largely impact the way data is handled, and a model is favored. If the result of modeling has a large impact on safety issues, the choice and the procedure are not carried out in the same way as if only economic profits or losses are taken into account. Similarly, the cost induced by decision policies based on different models may influence the model selection. Regarding statistical properties, the best candidate is well characterized from data trajectories. It can permit fast calibration and straightforward lifetime estimation. For more details and examples, refer to [6,7].

To be able to discard irrelevant models it is necessary to have some prior knowledge about the degradation phenomenon under consideration. In the presence of such information and degradation data, the goal of degradation modeling is to select one model from a set of competing models that best captures the underlying degradation process. As it is mentioned before the selection criteria depend mainly on the specific purpose for which the model can be used, see for instance [5,8].

This paper deals partly with Lévy processes [9] and focuses on the most commonly used which are Wiener and Gamma processes. Model calibration and data fitting of these models have been widely addressed

* Corresponding author.

E-mail addresses: thi_phuong_khanh.nguyen@utt.fr, nguyenthpk85@gmail.com (K.T.P. Nguyen), mitra.fouladirad@utt.fr (M. Fouladirad), antoine.grall@utt.fr (A. Grall).

in the fields of finance, biology and engineering [10,11]. However, the model selection criteria in the finance and biology field have some differences with engineering domain where maintenance and safety constraints are significant concerns [5,12,13]. The model selection for engineering prediction problems is an important issue but has not been generalized and extensively addressed, [14].

This paper proposes different criteria for model selection, to avoid system failure and with the most reasonable calculation time. First, an overview of the considered stochastic processes is given, and their use in degradation modeling is underlined. Afterwards, to provide a first selection criterion and to continue to outline a methodological guide, some widely addressed and well known statistical data fitting criteria are pointed out. Their limits and performances are highlighted, and new prognostic basis criteria are introduced. To complete the methodological task, the proposed procedure is applied to different simulated data sets. The behavior, weakness, and performances of the model selection are analyzed and discussed.

The remainder of the paper is as follows. Section 2 describes the set of models under consideration. In Section 3, the criteria for the data-based model selection are exposed. Section 4 gives some prognosis criteria. Eventually, in Section 5 the proposed models and criteria are tested on simulated data.

2. Stochastic degradation modeling and parameter estimation

This section focuses on the considered Levy processes in the framework of degradation modeling, [12,15,16]. A Lévy process is a stochastically continuous process with stationary and independent increments. It can be decomposed into the sum of a drifted Brownian motion and a jump process such as Gamma or Poisson process. These properties make this class of processes a good candidate for degradation modeling. In this paper, the widely used Levy processes, particularly Brownian and Gamma family, are the models under consideration to describe the degradation behavior. They are presented in the following.

Let us introduce some notations. Let X_t be the degradation level of the system at time t . Let L be the failure threshold in the sense that the system is supposed to be failed if the degradation level exceeds the level L . Let $t_{ps}(L)$ be the first passage time of the degradation process to level L :

$$t_{ps}(L) = \inf\{t \in \mathbb{R}^+, X_t \geq L\}. \quad (1)$$

The residual useful life time (RUL) at time t given $X_t = x$ denoted by $RUL_{(x,t)}$ is the time duration before the first passage time $t_{ps}(L)$ starting from the degradation level x at t . In the following $t_{ps}(L)$ will be denoted t_{ps} except in case of ambiguity.

2.1. Lévy type diffusion process

To consider a general degradation modeling framework and take into account the possible existing physical models, it is natural to introduce stochastic differential equations (SDE) based on a standard Brownian motions B_t :

$$dX_t = m(X_t, t)dt + \sigma(X_t, t)dB_t,$$

$(m, \sigma) : \mathbb{R} \times \mathbb{R}^+ \mapsto \mathbb{R}$ are respectively the drift and the diffusion coefficient. These equations appear at the beginning of 20th century in statistical mechanics and have been thoroughly formulated by Itô [17,18]. Such equations can be derived directly from existing physical models by adding Gaussian “white noises” on measurements. They permit a wide range of degradation modeling due to the flexibility of the structure and functional parameters.

In this section, some specific Levy diffusions processes are presented, and their interest in degradation modeling is underlined. For each case, the differential equation, the related distribution functions, and some statistical properties are exposed.

2.1.1. Wiener process

The Wiener process is very popular deterioration modeling when observations increments vary non-monotonically. The statistical properties of the failure time in the case of a Wiener process are studied in [19,20]. It has been considered in reliability and lifetime analysis widely since the 1970s. Authors in [21] used the Wiener process with drift to model accelerated life testing data. In [22,23] the impact of measurement errors on the Wiener degradation model of self-regulating heating cables is analyzed. Authors in [24–26] also focused on the stopping time (failure time) of Wiener deterioration models and expanded the existing theoretical results in this domain. The Brownian motion with non-linear drift has attracted more attention in engineering problems and residual lifetime estimation, see for example [27–29].

More precisely, a diffusion process in Brownian motion family has the following properties. The increments are independent, X_t is solution of the SDE $dX_t = \mu(t)dt + \sigma dB_t$, where $\mu(t)$ is a function of t and B_t is a standard Brownian motion. The transition probability to $X_t = x$ knowing that $X_s = y$ is given by:

$$p(x, t|y, s) = \frac{1}{\sqrt{4\pi\sigma(t-s)}} \exp\left(-\frac{(x + M(t, s) - y)^2}{4\sigma^2(t-s)}\right), \quad (2)$$

where $M(t, s)$, $\gamma(t, s)$ are given by:

$$M(t, s) = -\int_s^t \mu(u)du, \quad (3)$$

The mean and variance values of X_t are given by:

$$\mathbb{E}[X_t] = -M(t, 0), \text{Var}[X_t] = \sigma^2 t \quad (4)$$

M_1 : *Wiener process with linear drift*. This process is the special case of a Wiener process when the drift and the variance are not time dependent ($\mu(t) = \mu$ is constant). This diffusion process which is also a Lévy process is suitable for fluctuating degradation records linearly increasing in time. It will be referred as M_1 in Section 5.

The RUL cumulative distribution function (cdf) for a drifted Brownian motion given the observation value $X_t = x$ at the observation time t are given as follows [30]:

$$F_{RUL_{(x,t)}}(u) = \Phi\left(\frac{-L + \mu u + x}{\sigma\sqrt{u}}\right) + e^{\frac{2\mu}{\sigma^2}(L-x)} \Phi\left(\frac{-L - \mu u + x}{\sigma\sqrt{u}}\right) \quad (5)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

M_2 : *Wiener process with time-dependent drift*. This process is the particular case of a diffusion process when the degradation process is exponentially increasing in time ($\mu(t) = at^b$). It will be referred as M_2 in Section 5. In this case, the ratio between its drift and diffusion is not a constant and also depends on the time. Therefore, it is difficult to derive the explicit expression of the RUL distribution. Its evaluation requires solving a non-singular Volterra Integral Equation. It can be done numerically, see e.g. [31].

2.1.2. M_3 : diffusion process with purely time-dependent drift and diffusion

This process is the particular case of a diffusion process when $\mu(t)$ and $\sigma(t)$ are time dependent functions independent of X_t . This is suitable for a degradation process including random walks with time-dependent drift and diffusion terms. In this paper, we consider a special case of the purely time dependent drift and diffusion Brownian motion: $\mu(t) = cat^b$ and $\sigma(t) = \sqrt{2at^b}$. It will be denoted as M_3 in Section 5. As the power-law drift is proportional with drift, according to [32], the RUL CDF of the process at time t given a degradation level at time t , $X_t = x$, is derived:

$$F_{RUL_{(x,t)}}(u) = \Phi\left(\frac{-L - c\gamma(t_{ps}, t) + x}{\sqrt{2\gamma(t_{ps}, t)}}\right) + e^{c(L-x)} \Phi\left(\frac{-L - c\gamma(t_{ps}, t) + x}{\sqrt{2\gamma(t_{ps}, t)}}\right) \quad (6)$$

where $\gamma(t_{ps}, t)$ is given by Eq. (3).

2.1.3. M_4 : Ornstein–Uhlenbeck process

A time-dependent Ornstein–Uhlenbeck (OU) Process is widely used to describe physical dynamics of systems which stabilize at its equilibrium point. In the field of reliability modeling, OU process could be a good candidate for modeling the degradation process when the drift is time-dependent and also depends on the degradation state [31,33]. Indeed for commonly used processes such as drifted Brownian motion [30], the mean or the drift trend can be chosen quite freely, but the variance strongly depends on the stochastic process properties. To control the variance function independently of the mean tendency, a time-dependent OU process is a good candidate. The increments of Ornstein–Uhlenbeck process are independent, X_t is solution of the SDE: $dX_t = (\lambda(t)X_t + \mu(t))dt + \sigma(t)dB_t$, where $\lambda(t)$, $\mu(t)$ and $\sigma(t)$ are functions in t . The transition probability density function from $X_s = y$ to $X_t = x$ is given by:

$$p(x, t|y, s) = \frac{e^{\alpha(t,s)}}{\sqrt{4\pi\gamma(t,s)}} \exp\left(-\frac{(xe^{\alpha(t,s)} + M(t,s) - y)^2}{4\gamma(t,s)}\right) \quad (7)$$

where $\alpha(t, s)$, $M(t, s)$, $\gamma(t, s)$ are given by:

$$\alpha(t, s) = -\int_s^t \lambda(u)du, \quad M(t, s) = -\int_s^t \mu(u)e^{\alpha(u,s)}du, \quad \gamma(t, s) = \int_s^t \frac{\sigma^2(u)e^{2\alpha(u,s)}}{2} du$$

In this paper, we consider a special case of OU process, denoted as M_4 and defined by:

$$\lambda(t) = a, \quad \mu(t) = m'(t) - a \cdot m(t), \quad m(t) = bt^c, \quad c \geq 1, \quad \sigma(t) = d \quad (8)$$

The remaining useful lifetime pdf of this OU process can be numerically evaluated as the solution of a non-singular Volterra integral equation, see [33].

2.2. Gamma process

The gamma process is a positive stochastic process with independent increments. Hence it is useful to describe the deterioration caused by the accumulation of wear [34]. This process is a jump process which can be roughly considered as a succession of the frequent arrival of tiny increments. This rough description makes it relevant to model gradual deterioration such as corrosion, erosion, wear of structural components, concrete creep, crack growth [35]. Moreover, the gamma process has an exact probability distribution function which permits feasible mathematical developments. This process has been widely used in deterioration modeling for condition-based maintenance (see [12]).

The increments of a gamma process (X_t), $X_t - X_s$, are independent, $X_t - X_s \sim Ga(\alpha(t) - \alpha(s), \beta)$ with the transition probability density function from $X_s = y$ to $X_t = x$ as follows:

$$p(x, t|y, s) = \frac{\beta^{\alpha(t)-\alpha(s)}}{\Gamma(\alpha(t) - \alpha(s))} (x - y)^{\alpha(t)-\alpha(s)-1} e^{-(x-y)\beta} \quad (9)$$

where the shape function $\alpha(t)$ is an increasing function defined on \mathbb{R}^+ . Γ is the Euler's Gamma function. and β the scale parameter. The mean and variance values of X_t are given by:

$$\mathbb{E}[X_t] = \frac{\alpha(t)}{\beta}, \quad \text{Var}[X_t] = \frac{\alpha(t)}{\beta^2} \quad (10)$$

The choice of $\alpha(\cdot)$ and β allows to model various deterioration behaviors from almost deterministic to very chaotic. Note that, based on the form of $\alpha(t)$, the Gamma process can be:

- Homogeneous Gamma Process if $\alpha(t)$ is a linear function in t : $\alpha(t) = at$. This process is denoted M_5 .
- Non-homogeneous Gamma Process if $\alpha(t)$ is a non-linear function: for instance $\alpha(t) = at^b$, $a > 0$, $b > 1$ for the process referred as M_6 in the following.

Given a degradation level $X_t = y$ at time t , the cumulative distribution of the remaining useful life of a gamma process is given by [36]:

$$F_{RU_{L(y,t)}}(u) = \frac{\Gamma(\alpha(u+t) - \alpha(t), (L - y_t)/\beta)}{\Gamma(\alpha(u+t) - \alpha(t))}; \quad (11)$$

$\Gamma(\cdot, \cdot)$ is the upper incomplete Gamma function.

2.3. Parameter estimation

Parameter estimation refers to the process of using sample data (degradation records at observation times) to determine the values of the model parameters that best fits the data. The parameter estimation of Lévy processes is discussed in [37]. Among numerous parameter estimation methods, the Maximum Likelihood Estimation (MLE) is one of the most widely used [38]. It proposes a unified approach to select the set of values of the model parameters through the maximization of the likelihood function. Intuitively, this maximizes the matching between the chosen model and the observed data.

Suppose that the degradation records have been obtained from n identical independent components. Let m_l be the number of records collected on the l th component ($1 \leq l \leq n$) and assume $m_l \geq 2$. Then, x_{lj} is the degradation records of the l th component at the j th observation time ($1 \leq j \leq m_l$). Recall that $p_\theta(x_{lj}, t_{lj} | x_{l(j-1)}, t_{l(j-1)}; \theta)$ is the transition probability density function of the degradation process knowing that θ is the vector of model parameters. The log-likelihood function for the whole uncensored sample data set is given by:

$$L(\theta) = \sum_{l=1}^n \log(p_\theta(x_{l1}, t_{l1} | X_0, 0); \theta) + \sum_{l=1}^n \sum_{j=2}^{m_l} \log(p_\theta(x_{lj}, t_{lj} | x_{l(j-1)}, t_{l(j-1)}; \theta))$$

The objective of MLE method is to find the parameter vector which maximizes the log-likelihood function: $\theta^* = \arg \max_\theta \log L(\theta)$. Deriving analytical treatment is difficult. The optimization problem is usually solved by numerical algorithms. In this paper, we propose to use the Nelder–Mead (NM) method to estimate the model parameters.

3. Degradation model selection criteria

The degradation process is considered as a random phenomenon, presented by time-continuous trajectories. The criteria considered for model selection have to take into account the stochastic characteristics of the degradation process variability. The problem is posed on the choice of the best model among a class of competing models given a data set according to different goals. The primary objective of model selection is to choose the model that best fits a particular observation data, i.e., with maximal goodness-of-fit. However, for some reasons such as constraints about the time of parameter estimation or the model manageability a sub-optimal simple model could be favored. In these cases, the aim of the model selection is to balance goodness-of-fit and model complexity. Thus, a simple slight model can be preferred to an unnecessarily complicated one. On the other hand, the purpose of model selection should also include the ability to predict the future degradation behavior and assess its associated uncertainty from a particular degradation path. For this purpose, errors of predictions have to be taken into account in the selection criteria. In summary, selection criteria for degradation modeling and prognosis can take into account three different objectives which are the goodness-of-fit, the complexity level and the relevance for prognosis.

3.1. Goodness-of-fit

The goodness of fit measures characterizes the model's ability to match the observation data. They describe the discrepancy between observed values and the values expected under the considered model. For regression analysis of deterministic models, goodness-of-fit measures are typically based on the evaluation of residual values, such as the mean squared errors (MSE), the mean absolute errors (MAE), between theoretical values generated by the model and observation values.

For a degradation process characterized by stochastic models (for example, the models presented in Section 2), goodness-of-fit measures are typically dedicated to the distribution of degradation data. In detail, the goodness-of-fit tests are introduced to assess whether a specific distribution is matched to a frequency data set. Among them, Pearson's Chi-squared test and Kolmogorov–Smirnov test are widely

used in the statistics. For more details about the goodness-of-fit test for Lévy processes, refer to [39]. However, these tests are not relevant to select a model among a list of competing models. Then the distribution error measure (Pearson's Chi-square or Kolmogorov–Smirnov error) is introduced as one of the different model selection criteria.

3.1.1. Pearson's Chi-square error

Pearson's Chi-squared error (PCSE) is used to examine the discrepancy between the number of observations in each category with the theoretical expectation number. It strictly depends on the sample data size. In general, the PCSE is evaluated through the following steps considering that the degradation process is time homogeneous and that the inspections are periodic:

1. Divide observation data (increments of degradation) into k bins (e.g. k regular intervals)
2. Let O_i and E_i be respectively the observed data frequency in bin i and the probability that an increment is in bin i for the considered model. The PCSE is given by:

$$\chi^2 = \frac{1}{k} \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (12)$$

which follows asymptotically when the sample data size tends to infinity a chi-square distribution with $k - 1$ degrees of freedom.

3.1.2. Kolmogorov–Smirnov error

The Kolmogorov–Smirnov error (KSE) characterizes the discrepancy between the empirical distribution function (ECDF) and the theoretical distribution estimated from a given model. The Kolmogorov–Smirnov statistic is given by:

$$KSE = \sup_x \left| \frac{\sum_{i=1}^n \mathbf{1}_{\{x_i \leq x\}}}{n} - F(x) \right|$$

where $\mathbf{1}_{\{x_i \leq x\}} = 1$ if $x_i \leq x$ and 0 otherwise and $F(x) = \mathbb{P}(\Delta X \leq x)$ is the cumulative distribution function obtained from the model for the increments distribution. Under the hypothesis that the observed sample comes from the distribution F , $\sqrt{n}KSE$ converges to the Kolmogorov distribution.

Both of PCSE and KSE could be implemented for the selection degradation model as goodness-of-fit measures. In fact, consider the homogeneous degradation models, such as the drifted Brownian or the homogeneous gamma model, the degradation increments are independent of the time and the degradation state. Under periodic observation policy, the degradation increases between two consecutive observation times are identical and follow the same distribution having the same parameters. Therefore, they could be divided into k bins to evaluate the distribution error measures. However, it is difficult to assess these distribution errors in the following cases when the degradation increments between two consecutive observation are not related to iid random variables i.e. when:

- degradation states are non-periodically inspected,
- degradation processes are assumed to follow non-stationary models, such as non-homogeneous gamma process, time-dependent drift or diffusion Brownian model, or OU process whose future increments are also dependent on the current degradation state.

Therefore, it is necessary to introduce more general goodness-of-fit measures that could be easily implemented for all type of parametric models with periodic and non-periodic observation data.

3.1.3. Empirical average log likelihood function

The likelihood function is a most popular goodness-of-fit measure, widely used in model selection because it can be evaluated from both of periodic or non-periodic observation data for all type of models. The likelihood function of the sample data is the probability measure that a particular data set is obtained given the chosen probability distribution

model [40]. To compare the likelihood function values when considering different sizes of data, it prefers to evaluate its empirical average values. Besides, as the log is a monotonic transformation and allows to easily evaluate the sum instead of the product of the likelihood, the Empirical Average Log-Likelihood (EAL) is assessed in this paper. However, its values are often negative, that's why in Eq. (13) we add the minus sign to obtain the positive ones. Therefore, we use the minimal log-likelihood instead of the maximal likelihood concept in our code for the examples given later in the paper.

Let n be the total number of data observation points, x_i be the observation value at observation time t_i , then the EAL measure of the model defined by parameter set θ is given by:

$$EAL = \frac{-\log p_\theta(x_1, t_1 | X_0, 0; \theta) - \sum_{i=2}^n \log p_\theta(x_i, t_i | x_{i-1}, t_{i-1}; \theta)}{n} \quad (13)$$

where p_θ is the process transition probability from given state to another state. As EAL could be easily applied for all type of parametric models with periodic or non-periodic observation data, hereafter, we chose this measure to quantify the goodness-of-fit.

3.2. Parametric complexity

The criteria presented in the Section 3.1 are only based on the discrepancy between the given data set and the considered model. It can lead to the choice of a complex model that overfits the data set. In order to overcome this problem, criteria that take into account complexity for model selection are introduced. The importance of complexity in model selection is highlighted in [41].

3.2.1. Minimum description length

Minimum Description Length (MDL) originates from the field of information theory in computer science. It was firstly proposed for model selection by Rissanen in 1978. It is based on the shortest encoding length principle of the data and the model itself.

Let D , M represent respectively the observed data set and a model from the list of models candidates to be selected. The "best" model M^* can be chosen as that which minimizes the MDL measure defined by $MDL = L(M) + L(D|M)$, [42] where:

- $L(M)$ is given by (see [43,44]):

$$L(M) = \frac{k}{2} \ln \left(\frac{n}{2\pi} \right) + \ln \int \sqrt{|I(\theta)|} d\theta \quad (14)$$

It depends on the number of parameters in the model k , on the number of observations in the data set n and also on the Fisher information $I(\theta)$. In the cases of universal priors where there is no information on the distribution of data and model, the Eq. (14) is simplified by Ramos [42] as follows:

$$L(M) = (k + 1) \log_2 k + \frac{k}{2} \log_2 n \quad (15)$$

- $L(D|M)$ is the log likelihood function of the data given the model M with parameter θ :

$$L(D|M) = -\log p_\theta(x_1, t_1 | X_0, 0; \theta) - \sum_{i=2}^n \log p_\theta(x_i, t_i | x_{i-1}, t_{i-1}; \theta) \quad (16)$$

3.2.2. Akaike information criterion

Akaike Information Criterion (AIC) is based on information theory, firstly introduced in [45] and widely used for model selection [46]. It is dedicated to select the model with the closest distribution to the true distribution in the Kullback–Leibler sense and having the smallest number of parameter. In fact, considering the data set D , AIC of model M having k parameters is given by:

$$AIC = 2L(D|M) + 2k \quad (17)$$

where $L(D|M)$ is the logarithm of the model's maximum likelihood. Given this criteria, among candidate models, the one with minimum AIC will be chosen.

3.2.3. Bayesian information criterion

Bayesian Information Criterion (BIC) firstly derived by Schwarz in 1978 is an approximation for the posterior distribution of Bayesian methods using the assumption that the priors are equal. It provides a measure for model selection that penalizes the model having more parameters, refer to [47,48]. In fact, BIC is given by:

$$BIC = 2L(D|M) + k \ln(n) \quad (18)$$

where $L(D|M)$ is the logarithm of the model's maximum likelihood. According to the BIC criterion, the model having the lowest value of BIC is the best. Besides the number of parameter k , the BIC also takes into account the size of the data set n . When n increases, the model having a smaller number of parameters k is favored by the BIC which is not necessarily the case for the AIC.

The three criteria AIC, BIC and MDL deal with the trade-off between the model goodness of fit and the model complexity, but only MDL considers the model's functional form given in (14). Therefore, it is expected to give the more accurate decision for model selection. However, it is not always easy to evaluate the integral term of the Fisher information.

3.3. Robustness - generalizability

Three criteria AIC, BIC and MDL penalize the complexity of the model and favor the simpler one. In fact, the effects of model complexity on model fit strictly depend on the mutual influence of complexity and generalizability. Generalizability is described in [49] as the "model's ability to accurately predict future, as yet unseen, data samples from the same process that generated the currently observed sample". The concept of generalizability is close to the robustness which is the insensitivity to outliers in the observed dataset. Considering the generalizability, a model which does not only fit to a given data set but also has the high ability to match other data generated by the same underlying distribution will be prioritized. Therefore, in this section, selection measures that characterize the robustness or the generalizability of the model are considered.

3.3.1. Bootstrap method

The main idea of Bootstrap methods, introduced by Efron (1979), is to draw repeated samples from the initial data set (that usually has small size), a large number of times to create a great size of "phantom samples". Then, based on these samples, the underlying distribution of population of interest is studied, for more details see [50]. Following the idea of using bootstrap method for model selection [51], we present the procedure to evaluate the bootstrap metric (BS) that characterizes the generalizability and robustness of a proposed degradation model (M) given a data set D having n samples as follows:

1. For $j \in K : \{1, \dots, k\}$ where k is a large number.
 - Randomly draw m samples from the data set D , ($m < n$) to create a bootstrap data set D_j ,
 - Estimate the model parameters from this data set D_j , $\hat{\theta}(D_j)$, by maximizing the log likelihood $L(D_j)$,
 - Evaluate the natural logarithm of the predictive model's likelihood for the initial data set with $\hat{\theta}(D_j)$: $L_{\hat{\theta}(D_j)}(D)$.
2. The average BS metric of model M_j is given by : $BS = \frac{1}{k} \sum_{j=1}^k L_{\hat{\theta}(D_j)}(D)$

The BS method is suitable for the small size data set because its implementation time is an issue. When the number of observation points is large enough, the Cross Validation method is preferred than the BS method.

3.3.2. Cross validation method

Cross validation (CV) method is a computer science technique for assessing how accurately a predictive model will match to data in practice [52]. Its principle is to split the data into subsamples including a

training sample set for estimation of model parameters and a validation sample set for testing the performance of predictive model in order to limit overfitting problem. Among CV techniques, k -fold CV presented by Boyce et al. [53] is widely used for model selection [54]. In the application of degradation model selection, we propose to evaluate the k -fold CV measure of model M as follows:

1. Divide the data set D in to k folds
2. For $j \in K : \{1, \dots, k\}$:
 - Estimate model parameters ($\hat{\theta}|D_{\{K \setminus j\}}$) using all data that do not belong to fold j ,
 - Evaluate the natural logarithm of the predictive model's likelihood for data that belong to fold j : $L(D_j|(\hat{\theta}|D_{\{K \setminus j\}}))$
3. The CV value of model M is given by:

$$CV = \sum_{j=1}^k L(D_j|(\hat{\theta}|D_{\{K \setminus j\}})) \quad (19)$$

where n is the size of data set D .

The CV measure gives an idea of the robustness of the considered model by examining how the model fits to an independent unknown dataset. The model having the minimum CV will be chosen.

4. Prognostic metrics

In the above section, the model selection criteria have introduced that measure how candidate models are close enough to available degradation data. In reliability engineering, one of the primary purposes of degradation modeling is to predict the system reliability. The Residual Useful Lifetime (RUL) is considered extensively for maintenance decision and planning. Therefore, the degradation models can be selected based on their lifetime prognostic performance. Authors in [55] introduced a set of metrics to assess the prognostic ability of a given model when degradation data and true failure time (t_F) are known. Inspired by this work we consider, in this section, three measures to evaluate the prognostic performances for degradation model selection.

4.1. Prognostic horizon

The Prognostic Horizon (PH) is introduced in [56]. Let t be the last observation time and RUL_t be the predicted RUL of component at t . Let t_F be the actual failure time. For two given values $\epsilon > 0$ and $\zeta > 0$, let define $t_{\epsilon, \zeta}$ by:

$$t_{\epsilon, \zeta} = \inf \{t, \mathbb{P}\{(1 - \epsilon)t_F - t \leq RUL_t \leq (1 + \epsilon)t_F - t\} \geq \zeta\}$$

The Prognostic Horizon is defined as $PH = t_F - t_{\epsilon, \zeta}$. For model selection, the model having the greatest PH will be chosen. In other words, the prognostic horizon criterion (PHC) is also characterized by the minimal value of the observation time, $PHC = t_{\epsilon, \zeta}$ such that the estimated RUL distribution satisfies the above requirement. The model having the minimal value of $t_{\epsilon, \zeta}$, can be chosen as it estimates faster and more precisely the failure time.

4.2. Prognostic accuracy

Let L be the failure threshold and t_τ be the first time the degradation level exceeds the threshold $L_\tau = \tau \cdot L$ with $\tau \in [0, 1]$. The prognostic accuracy criterion (PAC) is defined in this paper as a prognostic measure that evaluates the probability mass of predicted RUL of component at time t_τ within the ϵ -bounds $\epsilon^+ = (1 + \epsilon)t_F - t_\tau$ and $\epsilon^- = (1 - \epsilon)t_F - t_\tau$. The model having the maximal value of PAC is the best one. Then, the PAC is given by:

$$PAC = \mathbb{P}\{\epsilon^- \leq RUL_{t_\tau} \leq \epsilon^+\} \quad (20)$$

For PHC and PAC metrics, the choice of ϵ , ζ , and τ depend on the operation requirements of the considered system. Note that, the

Table 1
Summary of the model selection criteria investigated in numerical examples.

Purposes	Selection criteria	Features
To assess how candidate models are close enough to available degradation data.	Empirical Average Log-Likelihood (<i>EAL</i>)	Classical selection criteria to quantify the goodness-of fit.
To assess how candidate models are close enough to available degradation data.	Akaike Information Criterion (<i>AIC</i>)	Classical selection criteria to quantify the goodness-of fit and the model complexity.
To assess how candidate models are close enough to available degradation data.	Bayesian Information Criterion (<i>BIC</i>)	Classical selection criteria to quantify the goodness-of fit and the model complexity, taking into account data size.
To assess how candidate models are close enough to available degradation data.	Minimum Description Length (<i>MDL</i>)	Classical selection criteria to quantify the goodness-of fit and the model complexity, taking into account data size and model formality.
To assess how candidate models are close enough to available degradation data.	Cross Validation (<i>CV</i>)	Classical selection criteria to quantify the goodness-of fit and the generalizability.
To assess the prognostic ability of candidate models when degradation data and true failure time are known.	Prognostic Horizon Criterion (<i>PHC</i>)	Prognostic criterion to investigate which model could predict an acceptable RUL distribution as soon as possible.
To assess the prognostic ability of candidate models when degradation data and true failure time are known.	Prognostic Accuracy Criterion (<i>PAC</i>)	Prognostic criterion to investigate the precision of the RUL prediction of candidate models.
To assess the prognostic ability of candidate models when degradation data and true failure time are known.	Hybrid Criterion (<i>HyC</i>)	New criterion to taking into account the goodness-of-fit information of the observation data when investigating the precision of the RUL prediction of candidate models.

$(PAC)_{\tau=30\%}$ characterizes the value of *PAC* criterion when the degradation data is observed until $\tau = 30\%$ of the failure threshold ($\tau \cdot L$). The predicted RUL distribution is evaluated by the mathematical formulas as given e.g. in Eqs. (5), (6), or (11).

4.3. Hybrid criterion

In this subsection, we propose a new criterion called Hybrid Criterion (*HyC*) to take into account the goodness-of-fit information of the observation data when evaluating the prognostic measure. In fact, *HyC* is calculated by the weighted sum between the weighted probability mass of predicted RUL at time t_r (as defined in the previous section) and the log-likelihood of the observation data that does not exceed the threshold L_r (defined in the previous section).

$$HyC = -w \cdot \log(\mathbb{P}(\epsilon^- \leq RUL_{t_r} \leq \epsilon^+)) + (w-1) \left\{ \log p(x_1, t_1 | X_0, 0; \theta) + \sum_{i=2}^{n_r} \log p(x_i, t_i | x_{i-1}, t_{i-1}; \theta) \right\} \quad (21)$$

where $n_r = \max\{n, x_n \leq \tau \cdot L\}$ and $w \in [0, 1]$ is the weight parameter for the prognostic measure. In this paper, w is proposed to be the ratio between the number of observation periods from the last observation time to the failure time and the total number of observation periods from initial moment to the failure time. As a consequence, it increases with the prognostic horizon. For model selection, the model having the minimal value of *HyC* will be chosen. Similar to *PAC*, $(HyC)_{\tau=30\%}$ characterizes the value of *HyC* criterion when the degradation data is observed until $\tau = 30\%$ of the failure threshold ($\tau \cdot L$).

Table 1 summarizes the features of the selection criteria that are evaluated and discussed in numerical examples. These criteria are applied to choose the best model for degradation data generated by one among the Brownian motion (M_1, M_2, M_3, M_4) and the Gamma (M_5, M_6) family.

5. Numerical experiments

5.1. Design of numerical experiments

Consider six models presented in Section 2 which are respectively M_1 a drifted Brownian motion, M_2 a time-dependent drifted Brownian motion, M_3 a time diffusion-dependent drifted Brownian motion, M_4 a particular OU process as presented by Eq. (8), M_5 a homogeneous gamma process and M_6 a non-homogeneous gamma process. These models are classified into two principal groups: diffusion processes (Brownian and O.U.) and Gamma processes. In detail, M_1 is a particular

case of M_2 and M_4 while M_5 can be nested into M_6 (when choosing appropriate parameters). Corresponding to the classic criteria introduced in Section 3 and to the prognostic criteria introduced in Section 4, the best model is chosen for given simulated data set.

The selection procedure follows the following steps:

Data generation One of above models is chosen to generate data. Its parameters are chosen in order to have the mean degradation level 100 at time $t = 100$: $E[X_{100}] = 100$ with different values of coefficient of variation at time 100, $vc_{100} \in \{10\%, 30\%, 50\%, 70\%\}$. The maximum simulation time t_m is 500. For each simulation, the evolution of a degradation process from 0 to t_m is generated with the time step $\Delta t = 0.01$.

Data collection and parameter estimation Consider a failure threshold $L = 80$, the first time when degradation process exceeds this threshold is the true failure time t_f . The degradation data are recorded until this time through periodic inspection policy (at every period $T = k \cdot \Delta t$). Then, following the methods presented in Section 2.3, the recorded observation data are used to estimate parameters for all six models.

Evaluation of comparison criteria and selection of the best model corresponding to every criterion The comparison criteria are evaluated for every model with their estimated parameters. Among criteria presented in Sections 3 and 4, we only evaluate the characterized criteria as follows:

- Empirical Average Log-Likelihood (*EAL*) that characterizes the goodness-of-fit measures and could be used for all above 6 types degradation models with periodic and non-periodic observation data.
- Akaike Information Criteria (*AIC*) that takes into account the goodness-of-fit measures and the model complexity presented by the number of parameters.
- Bayesian Information Criteria (*BIC*) that does not only take into account the goodness-of-fit measures, the number of parameters but also the number of observation data.
- Minimum Description Length (*MDL*) that take into account the goodness-of-fit measures, the number of observation data, the model complexity presented through parameter number and model formulation.
- Cross Validation (*CV*) that verifies the model robustness. In this paper, the 5-fold-CV is considered. In detail, observation data are divided into five subsets (four subsets for parameter estimation and one subset for testing model).
- Prognostic Horizon (*PHC*) is a measure that considers which model gives us an acceptable estimation of RUL in the best time (as soon

Table 2

Model selection for periodic observation data ($T = 0.2$) of a degradation process generated by a Wiener process with drift (M_1) with different coefficient of variation vc_r . The numbers in the table represents the percentage that M_i is chosen according to every selection criterion.

	vct = 10%							vct = 30%							vct = 50%									
	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC
M1	22	23	76	98	100	42	45	38	1	5	43	92	100	10	26	25	3	1	43	92	100	6	21	19
M2	48	23	5	1	0	13	17	17	4	5	3	3	0	4	31	33	12	20	3	2	0	3	44	51
M3	7	16	4	0	0	37	36	42	0	12	9	1	0	6	36	34	5	8	9	0	0	4	27	23
M4	23	38	15	1	0	6	2	3	95	78	45	4	0	56	7	8	80	71	45	6	0	21	8	7
M5	0	0	0	0	0	2	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	54	0	0
M6	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	12	0	0

Table 3

Model selection for periodic observation data ($T = 0.2$) of a degradation process generated by a Wiener process with time-dependent drift (M_2) with different coefficient of variation vc_r . The numbers in the table represents the percentage that M_i is chosen according to every selection criterion.

	vct = 10%							vct = 30%							vct = 50%									
	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC
M1	0	0	0	1	19	0	2	1	0	0	9	69	97	1	10	8	0	0	20	84	99	0	9	11
M2	88	83	90	97	81	71	46	49	7	12	23	24	3	43	36	42	30	26	37	12	1	34	43	44
M3	0	1	1	1	0	1	33	32	1	2	1	1	0	1	33	31	0	4	4	1	0	0	20	21
M4	12	17	9	1	0	4	19	18	92	86	67	6	0	38	21	19	70	70	39	3	0	28	28	24
M5	0	0	0	0	0	18	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	28	0	0
M6	0	0	0	0	0	6	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	10	0	0

Table 4

Model selection for periodic observation data ($T = 0.2$) of a degradation process generated by a Brownian motion with time dependent drift and diffusion coefficient (M_3) with different coefficient of variation vc_r . The numbers in the table represents the percentage that M_i is chosen according to every selection criterion.

	vct = 10%							vct = 30%							vct = 50%									
	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC
M1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	4	6	0	0	0	0	0	1	7	5
M2	0	0	0	0	0	12	14	15	0	0	0	0	0	14	21	19	0	0	0	0	0	0	19	16
M3	100	100	100	100	100	67	83	80	100	100	100	100	100	20	64	66	100	100	100	100	100	12	63	72
M4	0	0	0	0	0	2	5	0	0	0	0	0	0	24	10	9	0	0	0	0	0	42	8	7
M5	0	0	0	0	0	20	1	0	0	0	0	0	0	36	1	0	0	0	0	0	0	45	3	0
M6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

as possible). In this section, we chose arbitrarily $\epsilon = 0.1$, $\zeta = 0.7$ (results with different value of ζ are discussed later).

- Prognostic accuracy criterion (PAC) is a measure that evaluates the precision of RUL estimation corresponding to an amount of accumulated observation data with a fixed observation period. In this section, we chose arbitrarily $\epsilon = 0.01$, $\tau = 0.3$ (results with different value of τ are discussed later).
- Hybrid Criterion (HyC) is a measure that allows us to take into account the goodness-of-fit information of observation data in the evaluation of the RUL estimation accuracy. The value of ϵ and τ are chosen similar to the PAC (results with different values of τ are discussed later).

Repeat the procedure for N times, where N is an enough large number, and analyze the statistical results

5.2. Analysis of numerical experiments

In this section, we consider a degradation process that is inspected at every period $T = 0.2$ time unit (i.e $k = 20$). The model selection is performed following the steps presented in the above section. For every simulation case, the best model corresponding to every comparison criteria is recorded to evaluate the percentage of simulation cases when model M_i is chosen. The results are then presented in Tables 1–6.

5.2.1. Dissociation of the two families

When data are generated from one family (Gamma processes or diffusion processes), the classical criteria chose very often a model in the same family. In this sense, the classical criteria are better than the prognostic criteria. For example, consider Table 2, when data are generated by M_1 with the coefficient of variation varying from 10% to

50%, all the classical criteria can identify that the data are generated by a diffusion process while according to the prognostic criteria a Gamma process could be chosen.

5.2.2. Identification within a family

Thanks to a good estimation of the model parameters, within the same group the obtained models are similar to the model used for data generation. Furthermore:

- In the first group, M_1 is a particular case of M_2 and M_4 (when choosing appropriate parameters). Therefore, consider Table 2, when M_1 is the true model, especially with high coefficients of variation, M_2 and M_4 are also frequently chosen to be the best model. In these cases, according to the estimated parameters, M_2 and M_4 are almost the same as M_1 .
- In the first group, M_2 could be nested into M_4 , and M_1 could be considered as a simple particular case of M_2 . Table 3 shows that when M_2 is the true model, M_2 and M_4 are chosen to be the best model in almost all cases. However, with the high coefficient of variation, according to BIC and MDL criteria, a simpler model as M_1 is more favorable. Each time another model (M_1 or M_4) is chosen, its estimated parameters make these models very close to M_2 .
- In the first group, M_3 is a particular case of the Brownian family where the drift and the diffusion also depend on time. Therefore, consider Table 4, according to all of the selection criteria in the most of cases (except the PHC in the case of high coefficient of variation), M_3 is the best one that fits the data. This model is very flexible for the parameter estimation and data fitting, even if this model is chosen for data generated from M_i , $i = 1, 2, 3$, the estimated parameters of M_3 lead to a model which is very close to the original one.

Table 5

Model selection for periodic observation data ($T = 0.2$) of a degradation process generated by a OU process (M_4) with different coefficient of variation vc_r . The numbers in the table represents the percentage that M_i is chosen according to every selection criterion.

	vct = 10%							vct = 30%							vct = 50%									
	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC
M1	0	0	0	3	37	0	2	2	0	0	14	74	98	0	7	7	0	0	15	86	99	1	6	10
M2	1	1	13	63	63	14	31	33	5	6	18	19	2	0	47	44	5	4	20	10	1	0	64	57
M3	0	0	0	1	0	27	44	45	0	2	1	1	0	3	27	32	0	1	4	1	0	0	20	22
M4	99	99	87	33	0	46	23	20	95	92	67	6	0	69	19	17	95	95	61	3	0	52	10	11
M5	0	0	0	0	0	10	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	38	0	0
M6	0	0	0	0	0	3	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	9	0	0

Table 6

Model selection for periodic observation data ($T = 0.2$) of a degradation process generated by a homogeneous gamma process (M_5) with different coefficient of variation vc_r . The numbers in the table represents the percentage that M_i is chosen according to every selection criterion.

	vct = 10%							vct = 30%							vct = 50%									
	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC
M1	0	0	0	0	0	7	16	0	0	0	0	0	0	1	33	12	0	0	0	0	0	10	26	19
M2	0	0	0	0	0	21	13	3	0	0	0	0	0	6	19	2	0	0	0	0	0	10	14	4
M3	0	0	0	0	0	21	21	8	0	0	0	0	0	6	12	5	0	0	0	0	0	9	22	4
M4	0	0	0	0	0	6	6	5	0	0	0	0	0	17	6	3	0	0	0	0	0	9	10	8
M5	46	0	89	99	0	17	16	47	66	5	63	85	98	55	19	44	91	35	65	80	90	52	13	33
M6	54	100	11	1	100	28	28	37	34	95	37	15	2	15	11	34	9	65	35	20	10	10	15	32

Table 7

Model selection for periodic observation data ($T = 0.2$) of a degradation process generated by a non-homogeneous gamma process (M_6) with different coefficient of variation vc_r . The numbers in the table represents the percentage that M_i is chosen according to every selection criterion.

	vct = 10%							vct = 30%							vct = 50%									
	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC	CV	EAL	AIC	BIC	MDL	PHC	PAC	HyC
M1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	4	5	0	0	1	2	2	4	2	5
M2	0	0	0	0	0	52	29	6	0	2	0	0	0	29	10	1	0	2	1	0	0	21	15	4
M3	0	0	0	0	0	21	13	6	0	0	0	0	0	3	6	1	0	0	0	0	0	0	9	1
M4	0	0	0	0	0	1	14	6	0	0	0	0	0	20	15	3	0	0	0	0	0	30	9	4
M5	40	5	5	5	5	0	3	3	5	1	4	6	15	0	17	25	21	1	2	7	14	2	32	41
M6	60	95	95	95	95	26	41	79	95	97	96	94	85	46	48	65	79	97	96	91	84	43	33	45

- In the first group, when data are generated by M_4 , see Table 5, according to CV and EAL criteria, M_4 is the best model in almost case, whereas AIC, BIC, MDL prefer a simpler model M_1 or M_2 . Considering prognostic criteria, in the case of high coefficient of variation, the PAC and HyC (with $\tau = 30\%$) also prefer M_2 while according to the PHC, the gamma family (M_5, M_6) could be chosen as the best one. With high variance the identifiability is difficult, and the model simplicity and reactivity criteria take over.
- Among the Gamma family, M_5 is considered as a particular case of M_6 when $b = 1$. Therefore, for data generated by M_5 and according to the classic criteria, either M_5 or M_6 is the best model that fits the data, see Table 6. When data are generated by M_6 , the classical criteria can select the good model in almost all cases, see Table 7. However, according to the prognostic criteria that are based on the predicted RUL, a model based on diffusion process could also be chosen when data are generated by M_5 or M_6 . Because of a good parameter estimation, the degradation process made by Brownian/OU models are also close to the one generated by Gamma models. Thus, the predicted RUL of Brownian/OU models is also close to the real value of RUL.

5.2.3. Considering classical criteria

- **Better data fitting for models with a high number of parameters:** A model with many parameters which is consequently a more complex model can fit an observation data set better than a simpler model with few parameters, even if the latter has generated the data. It is related to the well-known overfitting phenomenon. Therefore, according to goodness-of-fit criterion (EAL), an excessively complex model could be chosen more often than the true model. For example consider Table 2 with the coefficient of variation 30%. According to

the EAL criterion, M_4 is chosen as the best model for 78 scenarios while the true model M_1 is only selected for five scenarios. Similarly, consider Table 6 with the coefficient of variation 30%. According to the EAL criterion, M_6 is chosen to be the best model for 95 scenarios while the true model M_5 is only chosen for five scenarios.

- **Number of models parameters versus criteria:** In most of the cases, compared to EAL criterion the CV criterion promotes a general model while AIC and BIC criteria favor a simpler model. For example, see Table 2 with the coefficient of variation 50%. According to this CV and for 80% of the scenarios, M_4 is the best model while according EAL, AIC and BIC, M_4 is the best model for respectively 71%, 45% and 6% of the scenarios. The model complexity is more penalized by BIC measure than AIC measure. Therefore, according to BIC, the simpler model is more favored (M_1 is chosen for 92% of scenarios with BIC instead of 43% of scenarios with AIC).
- **Prior knowledge in a criteria:** Using hypothesis of universal priors when there are not any knowledge about the distribution of the data and the model, the MDL criterion given by Eq. (15) promotes a simple model. For example, consider Table 3. According to MDL, M_1 is the best model that fits the observation data generated by M_2 in almost all scenarios, especially when the coefficient of variation is high. Hence it is necessary to improve the formulation of MDL by taking into account the prior information about the data distribution and the model.

5.2.4. Considering prognostic criteria

The considered prognosis criteria are based on the RUL prediction of the component.

- **PHC and identification problem:** It is difficult to identify the underling model of the observed data based on the PHC, especially

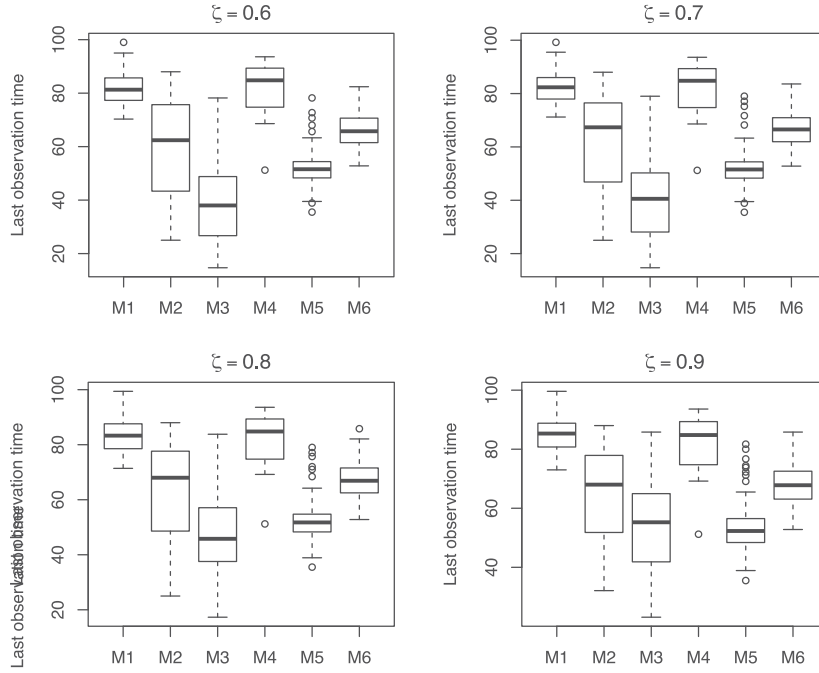


Fig. 1. Box plot of the estimated failure time $t_{\epsilon, \zeta}$ for every model with data generated with M_3 with coefficient of variation 10%.

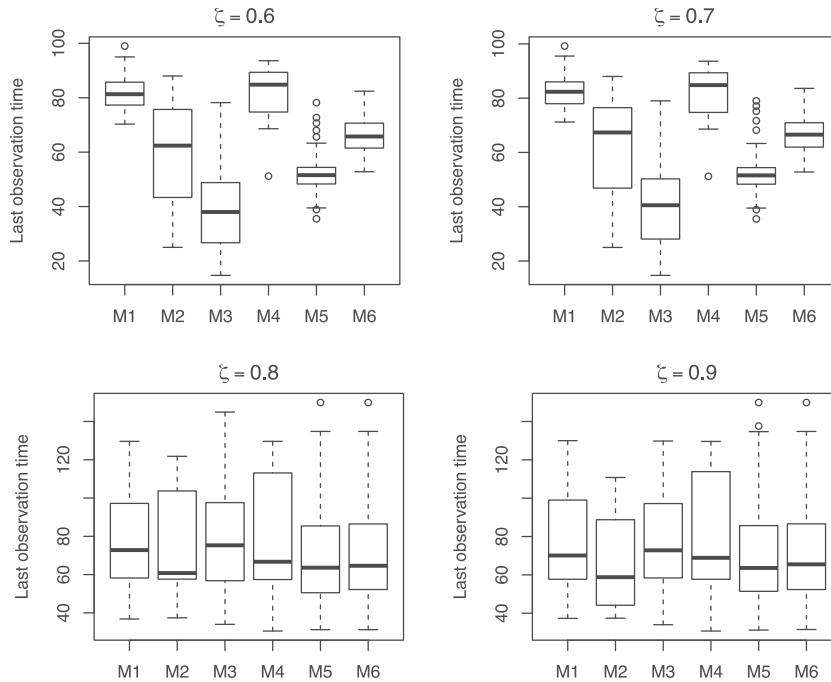


Fig. 2. Box plot of the estimated failure time $t_{\epsilon, \zeta}$ for every model with data generated from M_3 with coefficient of variation 50%.

in the case of high coefficients of variation. In fact, an efficient estimation parameter methods lead to obtaining very similar values of *PHC* criterion for all the models.

For example, consider the case where data are generated by M_3 with a coefficient of variation equals to 10% (see Fig. 1). The box plot of $t_{\epsilon, \zeta}$ with model M_3 is quite lower than the ones with other models. According to *PHC* when $\nu_{c_t} = 10\%$, M_3 is chosen to be the best model for 67% of scenarios (see Table 4). In the case of high coefficient ($\nu_{c_t} = 50\%$), the box plot of $t_{\epsilon, \zeta}$ for M_3 is approximately equal to the ones with other models (for $\zeta = 0.6$ and 0.7) and is quite higher than the ones of M_5 and M_6 (for $\zeta = 0.8$ and 0.9), see

Fig. 2. Therefore, according to *PHC* when $\nu_{c_t} = 50\%$ and $\zeta = 0.7$ M_5 is chosen to be the best model for 77% of scenarios (see Table 4).

- **Comparison of *PHC* and *PAC*:** Compared to the *PHC*, the *PAC* gives better results. Consider for example Table 4. For a high coefficient of variation (50%), the true model M_3 is detected for only 12% of scenarios with *PHC* while according to *PAC*, this ratio is increasing to 63%.
- **Influence of the observation duration:** A paradox result is recognized. When the threshold of observation data τ is increasing (i.e., for more gathered data), a model that belongs to another group could be selected to be the best model according to *PAC*.

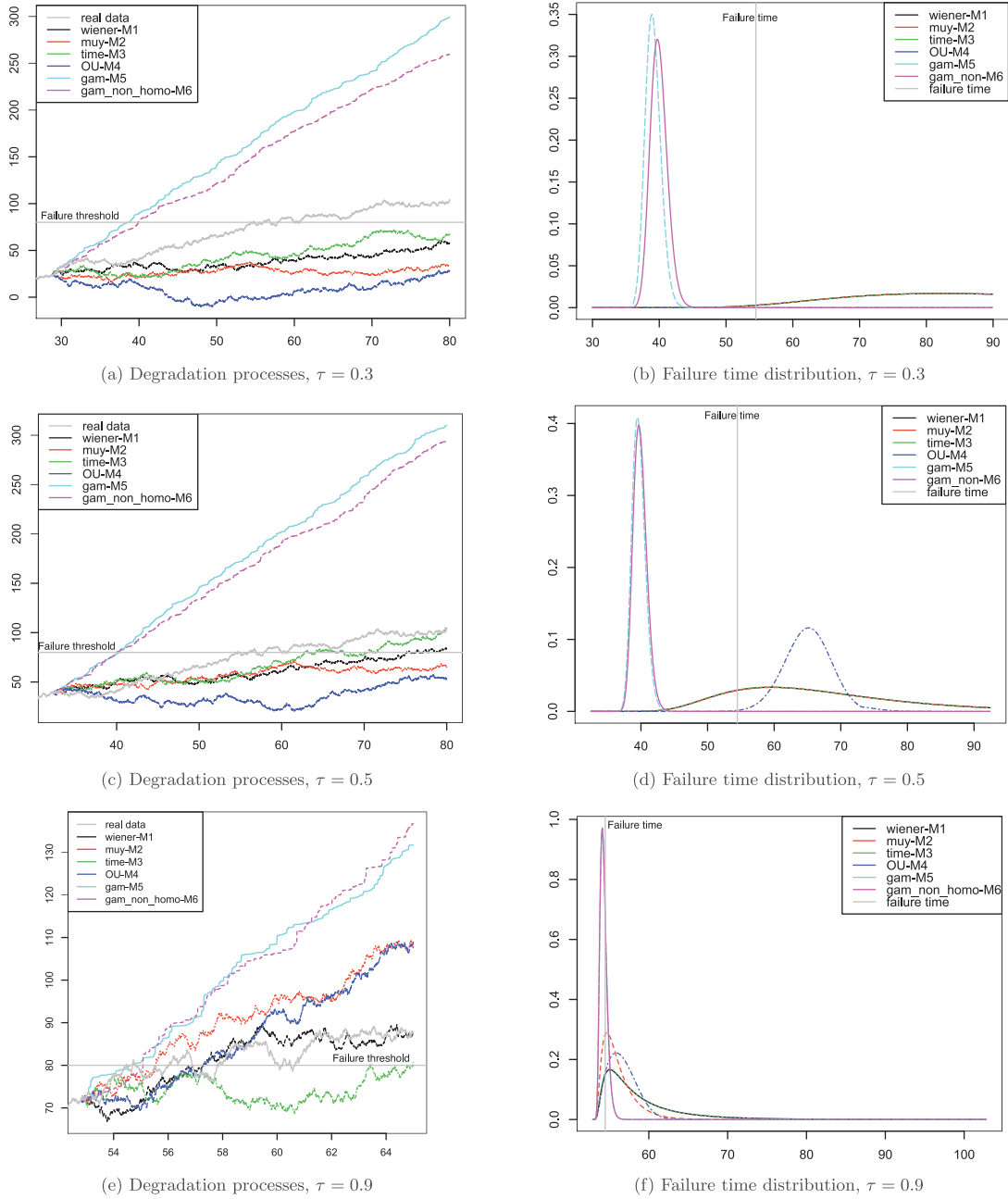


Fig. 3. Degradation processes generated (after parameter estimations from observation data generated by M_1 with $v_{c_t} = 30\%$) and failure time distributions estimated by models (M_1, \dots, M_6) with different values of τ .

For example, consider Table 8 that presents the results of model selection according to the PAC and PHC with different value of τ for data generated by M_1 . When τ is increasing from 30% to 90%, the percentage of the choice of Gamma family is increasing, especially with high values of coefficient of variation.

In order to explain this paradox result, we consider an example in which data are generated by M_1 with the coefficient of variation 30% (see Fig. 3). Based on the observation data with different values of the threshold τ , the parameters of every model are estimated (see Fig. 3(a), (c), (e)). Their corresponding RUL distributions are estimated (see Fig. 3(b), (d), (f) for estimated failure time distribution). When τ is increasing (i.e., for more gathered observation data), the RUL distributions are more precise. Whatever their shapes, they are more peaky, with a low variance and centered on the actual

failure time. Remark that this is mainly because the prognostic horizon is getting shorter (i.e., the last observation time is closer to the failure time) as the value of τ increases. Even the degradation process generated by Gamma models (M_5 or M_6) give results which are close to real data on a short time prediction horizon. It can be observed e.g. on Fig. 3(f). The first passage times of degradation processes generated by M_5 and M_6 are close to the failure time. Then the failure time distribution estimated by Gamma models are the most accurate compared to the real failure time. For a short term horizon, these processes give good lifetime estimation with low variance. That is the reason why a Gamma model could be chosen to be the best model according to PAC when more measurements are obtained, and the prognostic horizon is short.

- **Efficiency of the hybrid criterion:** The HyC that takes into account the goodness-of-fit of observation data is the best criterion among

Table 8
Model selection according to *PAC* and *HyC* with different thresholds of observed data generated by M_1 .

vc_τ	Criteria	M_1 (true model)	M_2	M_3	M_4	M_5	M_6
10%	(<i>PAC</i>) $_{\tau=30\%}$	45%	17%	36%	2%	0	0
	(<i>HyC</i>) $_{\tau=30\%}$	39%	17%	41%	3%	0	0
	(<i>PAC</i>) $_{\tau=50\%}$	37%	14%	30%	19%	0	0
	(<i>HyC</i>) $_{\tau=50\%}$	33%	16%	34%	17%	0	0
	(<i>PAC</i>) $_{\tau=70\%}$	35%	21%	28%	16%	0	0
	(<i>HyC</i>) $_{\tau=70\%}$	34%	23%	27%	16%	0	0
	(<i>PAC</i>) $_{\tau=90\%}$	21%	8%	25%	35%	5%	6%
	(<i>HyC</i>) $_{\tau=90\%}$	25%	12%	29%	34%	0	0
30%	(<i>PAC</i>) $_{\tau=30\%}$	26%	31%	36%	7%	0	0
	(<i>HyC</i>) $_{\tau=30\%}$	25%	33%	34%	8%	0	0
	(<i>PAC</i>) $_{\tau=50\%}$	25%	22%	23%	30%	0	0
	(<i>HyC</i>) $_{\tau=50\%}$	20%	30%	21%	29%	0	0
	(<i>PAC</i>) $_{\tau=70\%}$	11%	18%	16%	54%	1%	0
	(<i>HyC</i>) $_{\tau=70\%}$	13%	21%	16%	50%	0	0
	(<i>PAC</i>) $_{\tau=90\%}$	2%	9%	3%	35%	26%	25%
	(<i>HyC</i>) $_{\tau=90\%}$	3%	16%	13%	68%	0	0
50%	(<i>PAC</i>) $_{\tau=30\%}$	21%	44%	27%	8%	0	0
	(<i>HyC</i>) $_{\tau=30\%}$	19%	51%	23%	7%	0	0
	(<i>PAC</i>) $_{\tau=50\%}$	7%	33%	36%	23%	0	1%
	(<i>HyC</i>) $_{\tau=50\%}$	9%	37%	32%	22%	0	0
	(<i>PAC</i>) $_{\tau=70\%}$	4%	33%	20%	37%	4%	2%
	(<i>HyC</i>) $_{\tau=70\%}$	3%	26%	23%	38%	0	0
	(<i>PAC</i>) $_{\tau=90\%}$	1%	11%	3%	7%	44%	34%
	(<i>HyC</i>) $_{\tau=90\%}$	1%	23%	18%	58%	0	0
70%	(<i>PAC</i>) $_{\tau=30\%}$	14%	59%	22%	4%	0	1%
	(<i>HyC</i>) $_{\tau=30\%}$	13%	64%	19%	4%	0	0
	(<i>PAC</i>) $_{\tau=50\%}$	9%	60%	18%	8%	3%	2%
	(<i>HyC</i>) $_{\tau=50\%}$	11%	62%	19%	8%	0	0
	(<i>PAC</i>) $_{\tau=70\%}$	5%	37%	12%	8%	22%	16%
	(<i>HyC</i>) $_{\tau=70\%}$	8%	56%	23%	13%	0	0
	(<i>PAC</i>) $_{\tau=90\%}$	0	3%	0	2%	44%	51%
	(<i>HyC</i>) $_{\tau=90\%}$	1%	46%	22%	31%	0	0

prognostic criteria. In fact, consider all the results on Tables 2–7. Based on *HyC*, the underlying model family (whether diffusion-based or Gamma-based processes) can be correctly identified in most of the cases, even if when the coefficient of variation is high. Moreover, the paradox result discussed for *PAC* is also bypassed. Looking for example at Table 8, it comes that the underlying model family is well identified with the different thresholds of τ even if the coefficient of variation is high.

6. Conclusion

In this paper, we have discussed characteristics of selection criteria for degradation models. The selection criteria are classified into two groups: (1) classical statistical criteria that are based on the discrepancy between observation degradation data and the values expected under the considered model and (2) prognostic criteria that are based on the relevance between failure time and its expected distribution under the considered model. The advantages and disadvantages of these criteria are considered through numerous numerical examples for model selection between solutions of stochastic differential equations and Gamma processes.

In general, the classic criteria is better than the prognostic criteria for the purpose of dissociation of the two families (Gamma vs. Brownian/OU models). In detail among classical criteria, compared to *EAL* criterion, *CV* criterion favors general models with high parameters number while *AIC* and *BIC* prefers a simple model having fewer parameters. As the model complexity is more strongly penalized by *MDL* criterion with the hypothesis of universal priors, then *MDL* criterion frequently favors a simple model. Hence, it is necessary to improve the formulation of *MDL* by taking into account the prior information about the data distribution and the model formality. On the other hand, among prognostic criteria,

PHC is the worst criteria to identify the underlying model family of observed data. When model parameters are well estimated, the prognostic measure assessment of each model is approximately equal. Considering *PAC*, an interesting paradox result is recognized: the longer the observation duration is, the more frequently a wrong underlying model is selected. In fact, when the prognostic horizon is small it means that the last observation time is close to the failure time. The degradation process generated by another model can be close to real data during a short period of time. Therefore, it may be counterproductive to increase the observation time for a model selection based on the *PAC* criterion. Thus, we proposed new criteria, *HyC* that allows taking into account the goodness-of-fit of observation data when evaluating the prognostic measures. Numerical experiments highlight the performance of *HyC*. Particularly the underlying model family of an observation data set, whether it is diffusion (Brownian/OU) family or Gamma family, can be correctly identified in most of the cases, even if the coefficient of variation is high.

In summary, the classical statistical methods for model calibration are very efficient. These methods are usually powerful with a large set of data and they do not take into account the possible posterior information on the system operation conditions. The prognostic criteria are more strict and very sensitive to the coefficient for variation of data and the decision parameters. Moreover, these criteria are more useful when the operational conditions are different from the available training degradation data. In this paper, the same degradation data are considered for classical model selection and prognostic criteria. Therefore, the results of model selection bring out the interest of classical methods. Regarding the sensitivity of the prognostic criteria and the precision sought by the latter, they are more recommended when the operational conditions after the last observation are very fluctuating.

The selection criteria presented in this paper primarily take account of the sampling errors in parameter estimation. Further works will consider the influence of confidence intervals of estimated parameters on a model comparison. On the other hand, the impact of some characteristic features of the observed data set on model selection will be investigated, particularly such as the necessary amount of data, the frequency of the observations, periodic or not. Moreover, new appropriate criteria that allow comparing parametric and non-parametric models efficiently could be developed. The set of models can also be extended, and some other models such as Variance-Gamma or jump diffusion models can be included.

References

- [1] Liao CM, Tseng ST. Optimal design for step-stress accelerated degradation tests. *IEEE Trans Reliab* 2006;55(1):59–66.
- [2] Xu Z, Ji Y, Zhou D. Real-time reliability prediction for a dynamic system based on the hidden degradation process identification. *IEEE Trans Reliab* 2008;57(2):230–42.
- [3] Nikulin M, Limnios N, Balakrishnan N, Kahle W, Huber-Carol C. *Advances in degradation modeling: applications to reliability, survival analysis, and finance*. Birkhauser Boston; 2010.
- [4] Ye Z-S, Xie M. Stochastic modelling and analysis of degradation for highly reliable products. *Appl Stochastic Models Bus Ind* 2015;31:16–32.
- [5] Zhou RR, Serban N, Gebraeel N. Degradation modeling applied to residual lifetime prediction using functional data analysis. *Ann Appl Stat* 2011;5(2B):1586–610.
- [6] Kaipo J, Somersalo E. *Statistical and computational inverse problems, theory and methods for parameter estimation*. New York: Springer; 2005.
- [7] Muller H, Zhang Y. Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics* 2005;61:1064–75.
- [8] Bousquet N, Fouladirad M, Grall A, Paroissin C. Bayesian gamma processes for optimizing condition-based maintenance under uncertainty. *Appl Stochastic Models Bus Ind* 2015;31(3):360–79.
- [9] Bertoin J. *Lévy processes*, vol. 121. Cambridge University Press; 1998. Part of Cambridge Tracts in Mathematics.
- [10] Peng R, Li Y, Zhang W, Hu Q. Testing effort dependent software reliability model for imperfect debugging process considering both detection and correction. *Reliab Eng Syst Saf* 2014;126:37–43.
- [11] Peng R, Zhai Q. Modeling of software fault detection and correction processes with fault dependency. *Eksplotacja i Niezawodność Maint Reliab* 2017;19(3):467–75.
- [12] van Noortwijk JM. A survey of the application of gamma processes in maintenance. *Reliab Eng Syst Saf* 2009(1):2–21.
- [13] Lorton A, Fouladirad M, Grall A. Computation of remaining useful life on a physics-based model and impact of a prognosis on the maintenance process. *J Risk Reliab* 2013;227:434–49.

- [14] Cross RJ, Makeev A, Armanios E. A comparison of predictions from probabilistic crack growth models inferred from Virkler's data. *J ASTM Int* 2006;3(10):1–11.
- [15] Li H, Pan D, Chen C. Reliability modeling and life estimation using an expectation maximization based wiener degradation model for momentum wheels. *IEEE Trans Cybern* 2015;45(5):955–63.
- [16] Tamura Y, Yamada S. Reliability analysis based on a jump diffusion model with two wiener processes for cloud computing with big data. *Entropy* 2015;17(7):33–45.
- [17] Itô. *Stochastic processes*. Berlin Heidelberg New York: Springer; 2004.
- [18] Tseng ST, Peng CY. Stochastic diffusion modeling of degradation data. *J Data Sci* 2007;5:315–33.
- [19] Cox D, Oakes D. *Analysis of survival data*. Chapman & Hall; 1984.
- [20] Chhikara R, Folks J. The inverse Gaussian distribution as a lifetime model. *Technometrics* 1977;19(4):461–8.
- [21] Doksum K, Hoyland A. Models for variable-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution. *Technometrics* 1992;34(1):74–82.
- [22] Whitmore G. Estimation degradation by a Wiener diffusion process subject to measurement error. *Lifetime Data Anal* 1995;1:307–19.
- [23] Whitmore G, Scenkelberg F. Modelling accelerated degradation data using Wiener diffusion with a time scale transformation. *Lifetime Data Analysis* 1997;3:27–45.
- [24] Padgett WJ, Tomlinson M. Inference from accelerated degradation and failure data based on Gaussian process models. *Lifetime Data Anal* 2004;10:191–206.
- [25] Park C, Padgett WJ. Stochastic degradation models with several accelerating variables. *IEEE Trans Reliab* 2006;55:379–90.
- [26] Whitmore GA, Ramsay T, Aaron S. Recurrent first hitting times in Wiener diffusion under several observation schemes. *Lifetime Data Anal* 2012;18:157–76.
- [27] Wang X. Wiener processes with random effects for degradation data. *J Multivariate Anal* 2010;101(2):340–51.
- [28] Si X, Wang W, Hu C, Zhou D. Remaining useful life estimation—a review on the statistical data driven approaches. *Eur J Oper Res* 2011;213(1):1–14.
- [29] Wang W, Carr M, Xu W, Kobbacy K. A model for residual life prediction based on Brownian motion with an adaptive drift. *Microelectron Reliab* 2011;51(2):285–93.
- [30] Baussaron J, Mihaela B, Léo G-R, Fabrice G, Paul S. Reliability assessment based on degradation measurements: how to compare some models? *Reliab Eng Syst Saf* 2014;131:236–41.
- [31] Deng Y, Barros A, Grall A. Residual useful life estimation based on a time-dependent Ornstein-Uhlenbeck process. *Chem Eng Trans* 2013;33:325–30.
- [32] Molini A, Talkner P, Katul GG, Porporato A. First passage time statistics of Brownian motion with purely time dependent drift and diffusion. *Physica A* 2011;390(11):1841–52. doi:10.1016/j.physa.2011.01.024.
- [33] Deng Y, Barros A, Grall A. Degradation modeling based on a time-dependent Ornstein-Uhlenbeck process and residual useful lifetime estimation. *IEEE Trans Reliab* 2016;65(1):126–40.
- [34] Avramidis AN, L'ecuyer P, Tremblay PA. Efficient simulation of gamma and variance-gamma processes. In: *Proceedings of the 2013 winter simulation conference*, vol. 1; 2003. p. 319–26.
- [35] Çınlar E, Bažant Z, Osman E. Stochastic process for extrapolating concrete creep. *J Eng Mech Div* 1977;103(EM6):1069–88.
- [36] Parioissin C, Salami A. Failure time of non homogeneous gamma process. *Commun Stat - Theory Methods* 2014;43(15):3148–61.
- [37] Liptser RS, Shiryaev AN. *Statistics of random processes: II applications. Applications of mathematics*. 2nd. Berlin, Heidelberg, New York: Springer-Verlag; 2001.
- [38] Myung J. Tutorial on maximum likelihood estimation. *J Math Psychol* 2003;47(1):90–100.
- [39] Barndorff-Nielsen O, Mikosch T, Resnick S. *Lévy processes: theory and applications*. Basel: Birkhäuser; 2001.
- [40] Nishii R. Maximum likelihood principle and model selection when the true model is unspecified. *J Multivariate Anal* 1988;27(2):392–403.
- [41] Myung IJ. The importance of complexity in model selection. *J Math Psychol* 2000;44(1):190–204.
- [42] Ramos AA. The minimum description length principle and model selection in spectropolarimetry. *Astrophys J* 2006;646(2):14–45.
- [43] Grünwald P. Model selection based on minimum description length. *J Math Psychol* 2000;44(1):133–52.
- [44] Myung JI, Pitt MA. Model comparison methods. In: *Methods in enzymology, BT*, vol. 383 of numerical computer methods, part D. Academic Press; 2004. p. 351–66.
- [45] Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control* 1974;19(6):716–23.
- [46] Bozdogan H. Akaike's information criterion and recent developments in information complexity. *J Math Psychol* 2000;44(1):62–91. doi:10.1006/jmps.1999.1277.
- [47] Wasserman L. Bayesian model selection and model averaging. *J Math Psychol* 2000;44(1):92–107.
- [48] Busemeyer JR, Wang Y-M. Model comparisons and model selections based on generalization criterion methodology. *J Math Psychol* 2000;44(1):171–89.
- [49] Liu CC, Aitkin M. Bayes factors: prior sensitivity and model generalizability. *J Math Psychol* 2008;52(6):362–75.
- [50] Davison AC, Hinkley DV. *Bootstrap methods and their application*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press; 1997.
- [51] Simon G, Lendasse A, Verleysen M. Bootstrap for model selection: linear approximation of the optimism. In: Mira J, Álvarez JR, editors. *Computational methods in neural modeling, lecture notes in computer science*. Berlin Heidelberg: Springer; 2003. p. 182–9. doi:10.1007/3-540-44868-3_24.
- [52] Browne MW. Cross-validation methods. *J Math Psychol* 2000;44(1):108–32.
- [53] Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FK. Evaluating resource selection functions. *Ecol Modell* 2002;157(2–3):281–300.
- [54] Wiens TS, Dale BC, Boyce MS, Kershaw GP. Three way k-fold cross-validation of resource selection functions. *Ecol Modell* 2008;212(3–4):244–55.
- [55] Saxena A, Celaya J, Saha B, Saha S, Goebel K. Metrics for offline evaluation of prognostic performance. *Int J PHM* 2010;1(001):1–20.
- [56] Saxena A, Celaya J, Balaban E, Goebel K, Saha B, Saha S, et al. Metrics for evaluating performance of prognostic techniques. In: *International conference on prognostics and health management*; 2008. p. 1–17. doi:10.1109/PHM.2008.4711436.