Phase-type distributions in structural degradation: fitting to real data and reliability estimation

Javier Riascos-Ochoa, Mauricio Sánchez-Silva.
Department of Civil and Environmental Engineering, Universidad de Los Andes.
Cr 1E 19A-40, ML-630
Bogotá, Colombia

Raha Akhavan-Tabatabaei.
Department of Civil and Environmental Engineering, Universidad de Los Andes.
Cr 1E 19A-40, ML-317
Bogotá, Colombia

Abstract

One of the main problems in the reliability assessment of structures and infrastructure systems is the adequate modeling of the degradation process over the time. There are two challenges: fitting the parameters and variables of specific stochastic models to real data is not easy and the reliability estimation is numerically intractable for most of the cases. This paper presents a deterioration model based on the formalism of Phase-type (PH) distributions that solves these two problems. This is possible because PH distributions can fit any (positive) dataset or distribution (by means of PH-fitting algorithms) and their matrix-geometric properties allow to estimate easily the moments and the distribution of the lifetime. The model supposes degradation by cumulated shocks with inter-arrival times and sizes of shocks PH-distributed. Excellent accuracy and computation time of the proposed model were obtained by comparing the results with simulated (Monte Carlo) cases. These results suggest that the use of PH distributions is a convenient and practical approach to address the problems of uncertainty, fitting and the reliability estimation of infrastructures under degradation.

Keywords: Time-dependent reliability, lifetime, shock-based degradation, fitting algorithms, matrix-geometric approach.

1. Introduction

The life-cycle analysis and time-dependent reliability assessment in structures and infrastructure systems require the evaluation of uncertain future performance (e.g. structural capacity, stiffness). For this purpose it is of great importance the description and modelling of the processes leading to failure; described commonly as structural deterioration (Frangopol et al. 2004).

A comprehensive deterioration model should take into consideration the different degradation mechanisms, and their random evolution over time. One of these mechanisms is the *shock-based model* which accounts for the effect of sudden events or *shocks* (e.g. earthquakes, blasts, hurricanes) on the system (Aven & Jansen (1999), Nakagawa (2007) and Sanchez-Silva et al. (2011)). This model is characterized by two *stochastic processes*: the *inter-arrival times* T_i and the *shock sizes* Y_i.

Two main challenges still exist for shock-based models. First, there is no enough and dependable data in order to appropriately fit the random variables T_i and Y_i. And second, even if an accurate fitting is performed, in most of the cases the reliability estimation is numerically intractable. The main difficulty arises in the computation of the *convolutions* that appear in the probabilistic description of quantities such as the total time and the cumulated damage.

In order to solve these challenges this paper presents a model for the reliability estimation of systems under shock-degradation based on the formalism of Phase-type (PH) distributions. Thanks to the matrix-geometric properties of PH distributions (Neuts 1981), the proposed *PH-shock model* gives closed-form solutions for the reliability quantities (i.e. the reliability function, the probability density and the n-moments of the lifetime). Moreover, as PH distributions can fit any probability distribution with positive support, they can approximate general datasets of T_i or Y_i by means of different fitting algorithms (e.g. moment matching (MM) or expectation maximization (EM)). Excellent accuracy and efficiency (execution times) of the fitting algorithms and the PH-shock model are obtained compared with Monte Carlo simulations. In addition, the proposed model generalizes previous shock-models based on PH distributions (Montoro-Cazorla et al. (2006, 2009, 2012), Segovia 2013), in the sense that the failure mechanisms considered in these studies are not cumulative as proposed in this paper.

This paper is structured as follows: Section 2 reviews the problem of time-dependent reliability. The shock-based degradation model is described in Section 3, while in Section 4 we present the theory of PH distributions and the fitting algorithms. The proposed PH-shock model and the details of the reliability computation are presented in Section 5 with an illustrative example in Section 6. Finally the conclusions.

2. The problem of reliability and degradation of systems

This section formulates the problem of time-dependent reliability estimation and explains the different degradation mechanisms.

2.1 Mathematical description of the system's failure and reliability

The *system's performance indicator* $V(t)$ represents any physical quantity (e.g. stiffness, structural capacity) that describes the state of the system at any time $t \geq 0$. *Deterioration*, $D(t)$, is the decay (in performance units) of $V(t)$ over time t. In order to model the random nature of the phenomena, the approach followed in this paper is to assume that $\{V(t)\}_{t\geq 0}$ and $\{D(t)\}_{t\geq 0}$ constitute *stochastic processes* of the performance and the deterioration, respectively. The following relation holds:

$$V(t) = \max(z - D(t), a^*), \tag{1}$$

where $z = V(0)$ is the performance at construction (i.e. at $t = 0$), and a^* is a threshold or a limit state (e.g., serviceability, ultimate). If the system's performance falls bellow this threshold the system fails, and the length of time required for a failure to happen is known as the *system's lifetime L*, i.e.:

$$L = \inf\{t \geq 0 : V(t) \leq a^*\}. \tag{2}$$

Consequently, the (time-dependent) reliability of the system can be expressed as:

$$R(t) = P(L > t) = P(V(t) > a^*) = P(D(t) < z - a^*), \tag{3}$$

where *t* is the time at which the system is evaluated. From this expression the probability density function (PDF) of the lifetime L and their *n*-moments can be obtained ($\mathrm{E}(\cdot)$ is the expectation operator):

$$f(t) = -\frac{d}{dt}R(t); \qquad \mathrm{E}(L^n) = \int_0^\infty t^n f(t)dt. \tag{4}$$

It can be noticed from Eq. (3) that the central element of the analysis is the appropriate evaluation of deterioration, i.e., $D(t)$, which depends on the degradation mechanisms involved.

2.2 Degradation mechanisms

Two main degradation mechanisms can be identified (Klutke et al. 2002, Sanchez-Silva et al. 2011, Iervolino et al. 2013): *progressive degradation*, caused by processes like aging or corrosion, which continuously deteriorates the system's property over the time; and *shock-based degradation* that accounts for the effect of sudden events (e.g. earthquakes, blasts, over-currents) on the system.

Progressive deterioration can be modelled with a deterministic deterioration rate $\lambda(\Phi, t)$, i.e. $D_p(t) = \int_0^t \lambda(\Phi, \tau)d\tau$; as a random-variable (RV) model, for which the vector parameter Φ in $\lambda(\Phi, t)$ is random (Pandey et al. 2009); or as a stochastic process (e.g. the gamma process in van Noortwijk 2009).

Shock-based models are characterized by two stochastic processes: the arrival times and the magnitudes of shocks, which define the random variables of *inter-arrival time* T_i, and *size* (*damage*) Y_i of each i^{th} shock ($i = 1,2, \ldots$). Two classes of models can be identified: the *catastrophic-shock model*, for which failure occurs when a shock with enough damage ($Y_i > z - a^*$) arrives; and the *cumulative-shock model* for cases when each shock cumulates damage on the system, i.e. the deterioration at time t is given by:

$$D_s(t) = \sum_{i=1}^{N_t} Y_i, \tag{5}$$

where N_t is the number of shocks by time t. The performance $V(t)$ evolution over the time t for this case is sketched in Figure 1 for a specific sample path.

In addition, cumulative-shock models can model both, the catastrophic-shock model and the progressive degradation (as an approximation). The former is an especial case of the cumulative-shock model that can be modeled as: $D_s(t) = \sum_{i=1}^{N_t} Y_i \mathbf{1}_{\{Y_i > V_0 - a^*\}}$, where $\mathbf{1}_{\{\cdot\}}$ is the indicator function. Finally, progressive degradation can be approximated by making inter-arrival times T_i and shock-damage Y_i enough small. Therefore, in this paper we will concentrate only on the cumulative-shock degradation which will be referred in general as a shock-based model. Next section is devoted to the reliability analysis and the main challenges of this kind of models.

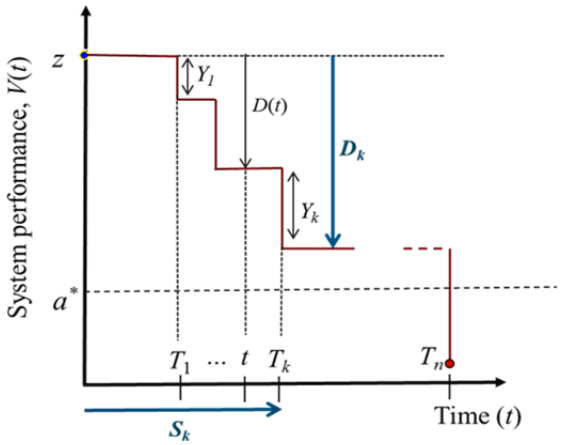

Figure 1. Degradation caused by cumulated-shocks

3. Reliability in shock-based degradation models

3.1 General expressions for the reliability quantities

For the reliability analysis of shock-based degradation two other processes are also required (Figure 1): the cumulated deterioration D_k and the total time S_k until the k^{th} shock, defined by:

$$D_k = \sum_{i=1}^{k} Y_i \; ; \qquad S_k = \sum_{i=1}^{k} T_i. \tag{6}$$

In a general approach, the shock-based degradation problem can be formulated as follows. Suppose that the random variables Y_i, T_i, D_k and S_k follow cumulative distributions functions (CDF): F_{Y_i}, F_{T_i}, F_{D_i}, and F_{S_i} and density functions (PDF): f_{Y_i}, f_{T_i}, f_{D_i}, and f_{S_i}. Furthermore, assume that the processes $\{Y_i\}_{i\geq 1}$ and $\{T_i\}_{i\geq 1}$ are independent, and that T_i and T_j are independent between them for $i \neq j$, as well as Y_i and Y_j. Thus, according to Eq. (4) and (5) the reliability can be expressed as:

$$R(t) = P\left(\sum_{i=1}^{N_t} Y_i < z - a^*\right) = P(N_t = 0) + \sum_{k=1}^{\infty} \bar{P}_k(z - a^*) P(N_t = k), \tag{7}$$

where $\bar{P}_k(z-a^*)$ is the probability of survival until the k^{th} shock given an initial capacity z, and $P(N_t=k)$ is the probability of having exactly k shocks by time t. These quantities are given by:

$$\bar{P}_k(z-a^*) = P\left(\sum_{i=1}^{k} Y_i < z-a^*\right) = P(D_k < z-a^*) = F_{D_k}(z-a^*), \tag{8}$$

$$P(N_t=k) = P\left(\sum_{i=1}^{k} T_i < t\right) - P\left(\sum_{i=1}^{k+1} T_i < t\right) = F_{S_k}(t) - F_{S_{k+1}}(t). \tag{9}$$

As D_k and S_k are the sum of independent random variables (Eq. (6)), their distributions $F_{D_k}(z)$ and $F_{S_k}(t)$ are given by the following iterative expressions in terms of the *convolution* of the distributions of D_{k-1} with Y_k and S_{k-1} with T_k, respectively:

$$F_{D_k}(y) = \int_0^\infty F_{D_{k-1}}(y-w)dF_{Y_k}(w); \qquad F_{S_k}(t) = \int_0^\infty F_{S_{k-1}}(t-\tau)dF_{T_k}(\tau). \tag{10}$$

For simplicity let $a^*=0$. Then, the PDF of L and their n-moments (Eq. (4)) are given by:

$$\begin{aligned} f(t) &= -\frac{d}{dt}\sum_{k=1}^{\infty}\bar{P}_k(z)P(N_t=k) = -\sum_{k=1}^{\infty}\bar{P}_k(z)\frac{d}{dt}P(N_t=k) \\ &= -\sum_{k=1}^{\infty}\bar{P}_k(z)\left[f_{S_k}(t) - f_{S_{k+1}}(t)\right] \end{aligned} \tag{11}$$

$$\begin{aligned} E(L^n) &= \int_0^\infty t^n f(t)dt = -\sum_{k=1}^{\infty}\bar{P}_k(z)\int_0^\infty t^n\left[f_{S_k}(t) - f_{S_{k+1}}(t)\right]dt \\ &= -\sum_{k=1}^{\infty}\bar{P}_k(z)[\mathrm{E}(S_k^n) - \mathrm{E}(S_{k+1}^n)] \end{aligned} \tag{12}$$

3.2 Challenges in the reliability estimation of shock-based models

Two main problems arise in the reliability analysis of the previous shock-based model:

1. To find the appropriate probabilistic model for the random variables: inter-arrival times T_i and shock sizes Y_i .
2. To compute numerically the reliability quantities by solving the convolution integrals, Eq. (10).

The first problem is closely related with the (physical) modeling of the random variables T_i or Y_i. The other option is fitting these variables to known probabilistic models from real (usually sparse) data. In both cases appropriate *fitting algorithms* are needed for general probabilistic models.

However, even if these distributions are determined, the numerical computation of the reliability quantities is a difficult task due to the convolution integrals in Eq. (12). This problem is usually overcome by assuming easy to handle probabilistic models (shocks sizes and inter-arrival times distributed exponential), leaving aside the actual description of the phenomena. Another alternative is to use simulations (Monte Carlo) which can be time-consuming and no reliable for general complex cases. It is clear

that a need still exists for proposing and solving general stochastic deterioration models by appropriate numerical and statistical methods.

In the next section we will describe the use of Phase-type (PH) distributions to handle the previous problems when modelling damage accumulation.

4. Phase-type distributions

4.1 Basic concepts

Phase-type (PH) distributions are based on the method of stages, first introduced by A. K. Erlang and later formalized by Neuts in 1981 (Neuts 1981). This family of distributions model random time intervals (or other discrete or continuous quantities) as being made up of a number of exponentially distributed segments and exploit the resulting Markovian structure to simplify the analysis. Then, the exponential distribution is used as the building block to construct more complex distributions by considering simple operations like convolutions and finite mixtures. For instance, the Erlang distribution $E(n,\lambda)$ is the distribution of the sum of n independent exponential random variables, each with parameter λ. When a time interval is distributed $E(n,\lambda)$, it can be thought of as consisting of n successive phases (or stages), each taking an independent but identically distributed (exponential) amount of time.

Consider a Markov process $\{X(t)\}_{t\geq 0}$ defined on the finite state-space $S = \{0,1,2,...,N\}$ where the last N states are transients and state 0 is an absorbing state. Then, a random variable is PH distributed if it spends exponential amounts of time in eacg transient states before finally being absorbed in state 0 (Latouche & Ramaswami 1999). The Continuous Markov process is then defined by an initial probability vector of the Markov chain (τ^0, $\boldsymbol{\tau}$) and its infinitesimal generator matrix is:

$$Q = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{t} & \mathbf{T} \end{bmatrix}, \tag{13}$$

where $\boldsymbol{\tau}$ is a row vector of size N, $\mathbf{T}$ is an $N\times N$ matrix, and $\mathbf{t}$ is a column vector of size N. Note that since Q is the generator of a Markov chain, $T^{ii} < 0$, $t^i \geq 0$, $T^{ij} \geq 0$ for $1 \leq i \neq j \leq N$, and $\mathbf{T}\cdot\mathbf{1}+\mathbf{t}=\mathbf{0}$; where $\mathbf{1}$ is a N-column vector of 1's. We also have $\tau^0 + \boldsymbol{\tau} \cdot \mathbf{1} = 1$. Then, the distribution of time until absorption is distributed $PH(\boldsymbol{\tau}, \mathbf{T})$ of order N (i.e., with N phases).

There are three properties of PH distributions that are of particular use for modelling degradation and which solve the *challenges* exposed in the previous section:

1. PH distributions are dense in the set of continuous density functions with support on $[0, \infty)$, making it possible to find a PH distribution arbitrarily close to any positive distribution (Challenge 1).
2. PH distributions are closed under convolution; this means that the sum of PH distributed random variables is also PH distributed (Challenge 2).
3. There is a closed expressions for the CDF, PDF and moments of a Phase-Type (Challenge 2); these are:

- CDF:

$$F(x) = P(X \leq x) = 1 - \boldsymbol{\tau} \cdot \exp(\mathbf{T}x) \cdot \mathbf{1} \tag{14}$$

- PDF:

$$f(x) = P(X \leq x) = \boldsymbol{\tau} \cdot \exp(\mathbf{T}x) \cdot \mathbf{t} \tag{15}$$

- Moments:

$$\mathrm{E}(x^n) = (-1)^n n! \cdot \boldsymbol{\tau} \cdot \mathbf{T}^{-n} \cdot \mathbf{1} \tag{16}$$

4.2 Fitting algorithms

A number of efficient *algorithms* have been proposed in the literature to *fit* a PH distribution to arbitrary datasets (numerically generated from any continuous distribution or from field measurements); see for example, Bobbio et al. (2005), Osogami and Harchol-Balter (2003). In this paper two fitting algorithms were applied: the Moment-matching (MM) algorithm proposed in Kharoufeh et al. (2010) and the Expectation-maximization (EM) algorithm by Thummler et al. (2006).

The MM algorithm uses the *coefficient of variation* (COV) of the dataset to choose between different PH distributions (Figure 2) and finds the parameters that better fit the first two (or three) moments of the dataset. Such parameters are: number of states *N,* rate λ and initial probability a, for the case $0 < \mathrm{COV}^2 \leq 0.5$; and initial probability a and rates λ_1, λ_2, for the case $0 < \mathrm{COV}^2 \leq 0.5$ (Figure 2).

On the other hand, the EM algorithm uses a hyper-Erlang distribution for the fitting. These are special PH distributions that are also dense in the distributions of positive support. By maximizing the Log-Likelihood of the whole data, the EM algorithm finds the best *N*-state hyper-Erlang structure (i.e. the number of Erlang branches *M* and the vector $\boldsymbol{r} = [r_1, r_2, \dots, r_M]$ of the number of states r_i in each *i*-Erlang branch), the initial probability vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_M]$, and the rates $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_M]$ (Figure 2).

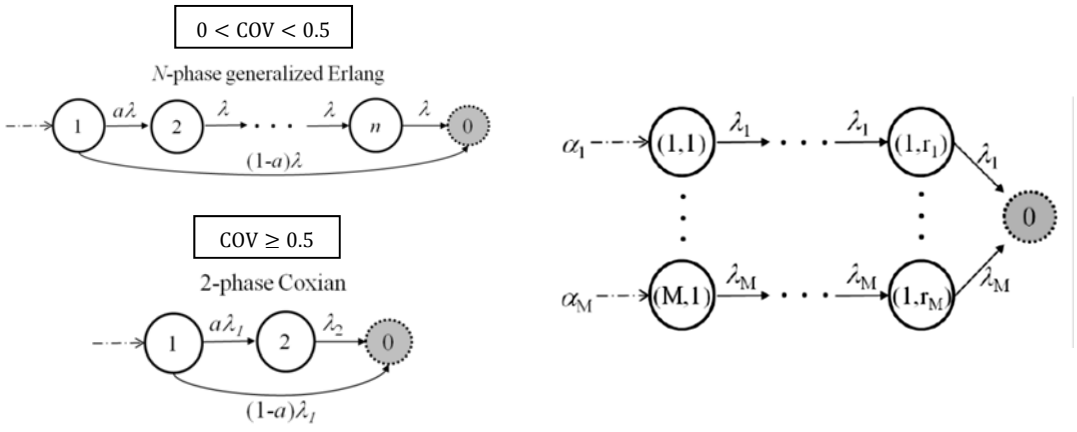

Figure 2. Structure of the PH distributions used in the MM fitting (left) and the hyper-Erlang distribution used in the EM fitting (right).

5. Modelling shock-based degradation using PH distributions: The PH-shock model

5.1 Model assumptions

In order to use PH distributions to handle degradation the following assumptions were made (Riascos-Ochoa et al., 2014):

Assumption 1*:* The inter-arrival times between shocks Ti, $i \geq 1$, are independent but not necessarily identically distributed and follow a PH distribution with representation $T_i \sim PH(\boldsymbol{\tau}_i, \mathbf{T})$ of order n_i and generator matrix:

$$Q_{T_i} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{t}_i & \mathbf{T}_i \end{bmatrix}$$

Note: The probability that the Markov chain associated with the PH distribution of T_i starts in its absorbent state is zero; i.e., $\tau_i^0 = 0, \forall i \geq 1$. Then, the time until absorption (inter-arrival time) is always strictly greater than zero.

Assumption 2: The size of shock *i*, denoted by Y_i with $i \geq 1$, are assumed independent but not necessarily identically distributed and follow a PH distribution with representation $Y_i \sim PH(\boldsymbol{\gamma}_i, \mathbf{Y}_i)$ of order m_i and generator matrix:

$$Q_{Y_i} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{y}_i & \mathbf{Y}_i \end{bmatrix}$$

Note: The probability that the Markov chain associated with the PH distribution of Y_i starts in its absorbent state is zero; i.e., $\gamma_i^0 = 0, \ \forall i \geq 1$ and, therefore, shock sizes are always strictly greater than zero.

Assumption 3: the time occurrence of shocks, T_i, and shock sizes, Y_i, are independent.

5.2 Closed-expressions for the reliability quantities

As it was mentioned in section 3, the main interest is to find easy-to-evaluate expressions for the reliability quantities $R(t)$, $f(t)$, and $\mathrm{E}(L^n)$, Eqs. (7), (11) and (12). In order to do so, expressions for $\bar{P}_k(z)$, $P(N_t = k)$ and $\frac{d}{dt}P(N_t = k)$ (Eqs. (8) and (9)) are needed, which requires the computation of the distributions F_{D_k} of the cumulated damage D_k and the distributions F_{S_k} (and f_{S_k}) of the time S_k until the k^{th} shock, which involves the convolutions (10). As D_k and S_k are the sum of independent PH distributed variables $\{Y_i\}_{i \leq k}$ and $\{T_i\}_{i \leq k}$, respectively, they are also PH distributed according to Property 2 in subsection 4.1. They have representations $D_k \sim PH(\mathbf{d}_k, \mathbf{D}_k)$ and $S_k \sim PH(\mathbf{s}_k, \mathbf{S}_k)$ with (Riascos-Ochoa et al., 2014):

$$\mathbf{D}_k = \begin{bmatrix} \mathbf{Y}_1 & \boldsymbol{y}_1 \cdot \boldsymbol{\gamma}_2 & \mathbf{0} & & \\ \mathbf{0} & \mathbf{Y}_1 & \boldsymbol{y}_2 \cdot \boldsymbol{\gamma}_3 & \mathbf{0} & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \mathbf{0} & \mathbf{Y}_{k-1} & \boldsymbol{y}_{k-1} \cdot \boldsymbol{\gamma}_k \\ & & & \mathbf{0} & \mathbf{Y}_k \end{bmatrix} \quad \mathbf{S}_k = \begin{bmatrix} \mathbf{T}_1 & \mathbf{t}_1 \cdot \boldsymbol{\tau}_2 & \mathbf{0} & & \\ \mathbf{0} & \mathbf{T}_1 & \mathbf{t}_2 \cdot \boldsymbol{\tau}_3 & \mathbf{0} & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \mathbf{0} & \mathbf{T}_{k-1} & \mathbf{t}_{k-1} \cdot \boldsymbol{\tau}_k \\ & & & \mathbf{0} & \mathbf{T}_k \end{bmatrix}$$

$$\mathbf{d}_k = [\boldsymbol{\gamma}_1, \mathbf{0}, \dots \mathbf{0}] \qquad \mathbf{s}_k = [\boldsymbol{\tau}_1, \mathbf{0}, \dots \mathbf{0}]$$

Figure 3 shows the PH-structure for D_k by connecting the Y_i's PH-structures. A similar structure is followed by S_k as the joining of the corresponding T_i's structures.

Then, by using Property 3, Eqs. (14),(15), and Eqs. (10) and (11), expressions for $\bar{P}_k(z)$ and $P(N_t = k)$ are obtained. Also, the *n*-moments of S_k are obtained with Eq. (16), which are needed for the moments of the lifetime, Eq. (12):

$$\bar{P}_k(z) = 1 - \mathbf{d}_k \cdot \exp(\mathbf{D}_k z) \cdot \mathbf{1}$$
$$P(N_t = k) = \mathbf{s}_{k+1} \cdot \exp(\mathbf{S}_{k+1} t) \cdot \mathbf{1} - \mathbf{s}_k \cdot \exp(\mathbf{S}_k t) \cdot \mathbf{1}$$
$$\frac{d}{dt} P(N_t = k) = \mathbf{s}_{k+1} \cdot \exp(\mathbf{S}_{k+1} t) \cdot (\mathbf{S}_{k+1} \cdot \mathbf{1}) - \mathbf{s}_k \cdot \exp(\mathbf{S}_k t) \cdot (\mathbf{S}_k \cdot \mathbf{1})$$
$$\mathrm{E}(S_k^n) = (-1)^n n! \cdot \mathbf{s}_k \cdot \mathbf{S}_k^{-n} \cdot \mathbf{1}$$

Clearly these expressions are much less complex than the original ones involving the convolutions. By replacing these into Eqs. (7), (11) and (12) easy-to-evaluate expressions for $R(t)$, $f(t)$, and $\mathrm{E}(L^n)$ are obtained.

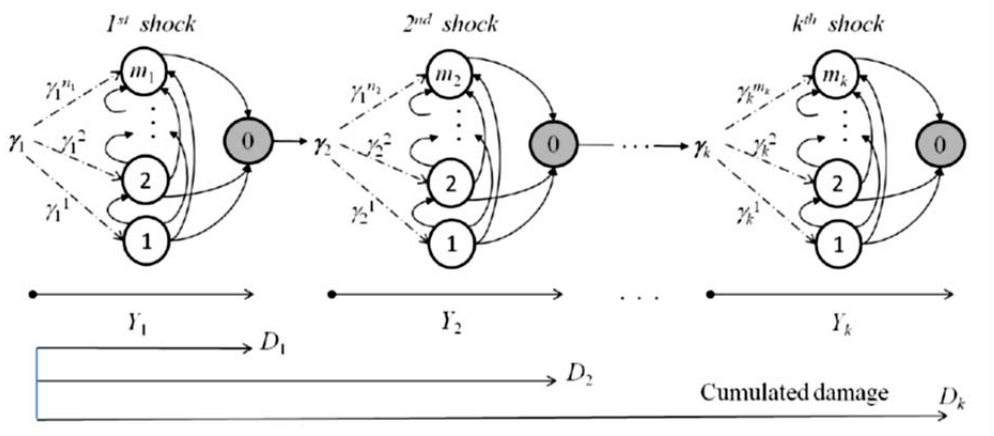

Figure 3. Structure of the PH distribution for the cumulated damage D_k

5.3 Numerical evaluation

Note that the expressions (7), (11) and (12) for $R(t)$, $f(t)$, and $\mathrm{E}(L^n)$, respectively, involve the evaluation of infinite sums. In order to overcome this issue, these sums are truncated in the K^{th} term. In other words the evaluation of the reliability quantities is done until the K^{th} shock and approximations $R_K(t)$, $f_K(t)$, and $\mathrm{E}_K(L^n)$ of the real quantities are employed:

$$R_K(t) = P(N_t = 0) + \sum_{k=1}^{K} \bar{P}_k(z - a^*) P(N_t = k)$$
$$f_K(t) = -\sum_{k=1}^{K} \bar{P}_k(z) \left[f_{S_k}(t) - f_{S_{k+1}}(t) \right]$$
$$\mathrm{E}_K(L^n) = -\sum_{k=1}^{K} \bar{P}_k(z) [\mathrm{E}(S_k^n) - \mathrm{E}(S_{k+1}^n)]$$

It can be shown (Riascos-Ochoa et al., 2014) that the correct reliability $R(t)$, for all $t > 0$, satisfies $R_K(t) \leq R(t) \leq R_{K-1}(t) + \varepsilon$ for a given $\varepsilon > 0$ and with the condition $K = \min(k, \bar{P}_k(z) < \varepsilon)$. In other words, the K^{th} shock for truncation is determined by the minimum shock for which the probability of survival until that shock is less than ε.
Such K always exists, and the error in the reliability is less than ε. In the numerical computation we set $\varepsilon = 10^{-6}$ and make the calculations until the K^{th} shock determined by the previous condition.

The numerical evaluation of the PH-shock model and the reliability quantities consists in four stages that are described in the following pseudo-code:

STAGE 1: DEFINITION OF THE SYSTEM (FITTING)
(1) **PH-representation**: Define the PH-representation of the inter-arrival times and shock sizes: $\{\tau_k, T_k\}$, $\{\gamma_k, Y_k\}$
(2) Define ε, z and $\{t_i\}$ with $t_i > 0$
STAGE 2: COMPUTATION OF $\overline{P}_k(z)$
(3) Set $k = 0$, $\overline{P}_0(z) = 1$
(4) WHILE $\overline{P}_k(z) > \varepsilon$
(5) Increase k in 1
(6) Compute the PH-representation of the cumulative damage until k^{th} shock: $\{\mathbf{d}_k, D_k\}$
(7) Compute $\overline{P}_k(z) = 1 - \mathbf{d_k} \cdot \exp(D_k z) \cdot \mathbf{1}$
(8) Set the number of computed shocks $K = k$
STAGE 3: COMPUTATION OF $P\{N_t = k\}$ **AND** $\frac{d}{dt}P\{N_t = k\}$
(10) FOR $k = 1$ to $k = K + 1$
(11) Compute the PH-representation of the time until k^{th} shock: $\{\mathbf{s}_k, S_k\}$
(12) FOR all $t \in \{t_i\}$
(13) Set $F_{S_0}(t) = 1$, $f_{S_0}(t) = 0$
(14 Compute $F_{S_k}(t) = 1 - \mathbf{s_k} \cdot \exp(S_k t) \cdot \mathbf{1}$ and $f_{S_k}(t) = \mathbf{s_k} \cdot \exp(S_k t) \cdot (-S_k \cdot \mathbf{1})$
(15) Compute $P\{N_t = k-1\} = F_{S_{k-1}}(t) - F_{S_k}(t)$ and $\frac{d}{dt}P\{N_t = k-1\} = f_{S_{k-1}}(t) - f_{S_k}(t)$
(16) END
(17) END
STAGE 4: COMPUTATION OF THE RELIABILITY QUANTITIES
(18) FOR all $t \in \{t_i\}$
(19) Compute $R_K(t) = \sum_{k=0}^{K} \overline{P}_k(z) P\{N_t = k\}$
(20) $f_K(t) = -\sum_{k=0}^{K} \overline{P}_k(z) \frac{d}{dt}P\{N_t = k\}$
(21) END
(22) Compute $E[L^n]_K = \sum_{k=0}^{K} \overline{P}_k(z) \left[E[S_{k+1}^n] - E[S_k^n]\right]$

Figure 4. Pseudo-code for the numerical evaluation of the reliability quantities with the proposed PH-shock model

6. Illustrative example

In order to show the efficiency of the PH-shock model, a study comparing PH distributions and Monte Carlo Simulation (10^5 realizations) was carried out. Two-PH distributions were developed using the MM and EM fitting algorithms (Table I). The times between shocks T_i are independent and identically distributed (i.i.d) following an exponential distribution (mean 10 years). Shock sizes Y_i are also i.i.d and follow a lognormal distribution with mean 20 (performance units); the results were obtained for various coefficients of variation COV_Y of shock sizes (Table I). All the evaluations were run under the same hardware specifications with MATLAB® 2012.

The results in Table I show that, in all cases, the mean time to failure $\mathrm{MTTF} = \mathrm{E}(L)$ and the $\mathrm{COV} = \sqrt{\mathrm{E}(L^2) - \mathrm{E}(L)^2}/\mathrm{E}(L)$ of the lifetime L are in good agreement. However, the computational cost using the PH-shock model is between 20 and 30 times faster than Monte Carlo.

Figure 4 shows the results for the PDF $f(t)$ of the lifetime L with Monte Carlo simulation and the PH-shock model with both fittings algorithms. It is clear the good agreement between the curves, being the EM algorithm with the best performance. This is because the EM algorithm uses a greater number of states *N* and finds the best configuration *r* (Table I and Figure 4), giving a better fitting of the original shock-size Y_i distribution.

Table I: Comparison of Monte Carlo Simulation and the PH-shock model with the EM and MM fitting algorithms applied to Y_i~LN(mean=20, COV_Y).

Monte Carlo Simulation	$\mathrm{COV}_Y = 0.2$	$\mathrm{COV}_Y = 0.5$	$\mathrm{COV}_Y = 0.8$	$\mathrm{COV}_Y = 1.0$	$\mathrm{COV}_Y = 2.0$
ET: Execution time (s)	3.2	3.0	2.6	2.8	3.3
$MTTF$	54.8	56.2	58.0	60.1	69.4
COV_L	0.44	0.47	0.50	0.52	0.59
PH shock model with MM algorithm	$\mathrm{COV}_Y = 0.2$	$\mathrm{COV}_Y = 0.5$	$\mathrm{COV}_Y = 0.8$	$\mathrm{COV}_Y = 1.0$	$\mathrm{COV}_Y = 2.0$
N: Number of states	25	4	2	2	2
K: Number of shocks	8	12	17	20	22
ET: Execution time (s)	0.1	0.11	0.11	0.14	0.15
$MTTF$ (%Error)	55.2 (**0.7%**)	56.2 (**0.1%**)	58.1 (**0.2%**)	60.0 (**0.2%**)	67.4 (**2.9%**)
COV_L (%Error)	0.44 (**0.1%**)	0.47 (**0.2%**)	0.52 (**3.4%**)	0.55 (**5.9%**)	0.54 (**7.9%**)
PH shock model with EM algorithm	$\mathrm{COV}_Y = 0.2$	$\mathrm{COV}_Y = 0.5$	$\mathrm{COV}_Y = 0.8$	$\mathrm{COV}_Y = 1.0$	$\mathrm{COV}_Y = 2.0$
N: Number of states	25	10	10	10	10
r: Best configuration	[25]	[4, 6]	[2, 3, 5]	[2, 3, 5]	[1, 2, 3, 4]
K: Number of shocks	8	11	14	18	22
ET: Execution time (s)	0.1	0.12	0.21	0.22	0.36
$MTTF$ (%Error)	55.2 (**0.7%**)	56.2 (**0.1%**)	58.8 (**1.3%**)	59.9 (**0.2%**)	69.6 (**0.3%**)
COV_L (%Error)	0.44 (**0.1%**)	0.47 (**0.3%**)	0.50 (**0.2%**)	0.53 (**1.2%**)	0.60 (**0.9%**)

6. Conclusions

Optimum solutions for the design, operation and maintenance of engineered systems require the selection of appropriate degradation models. In particular, this paper focused in the problem of shock-based degradation, which also can be applied to progressive degradation. Three important problems were addressed: the definition of adequate deterioration models, the fitting of their parameters and (random) variables to real (or simulated) data of the damaging events (shocks), and the reliability estimation of systems under this kind of degradation. This last task is especially complex in general shock-models due to the convolutions involved, relying too much in Monte Carlo simulation.

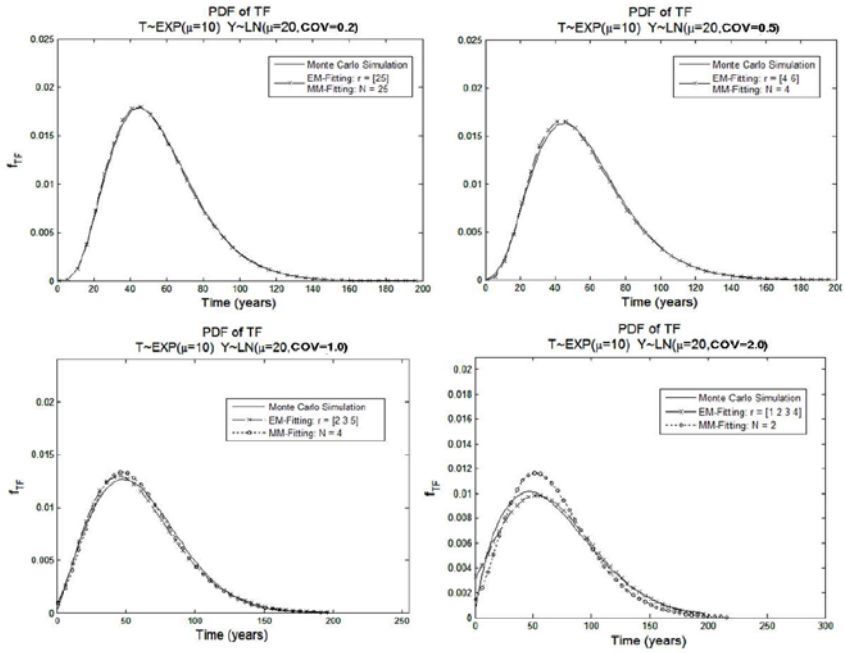

Figure 4. PDF's *f*(*t*) of the lifetime *L* via Monte Carlo simulation and the PH shock model (with the MM and EM algorithms for the fitting).

An approach to solve these problems was presented by using a special class of probability distributions known as Phase-type (PH). The proposed PH-shock model uses a PH representation of the random variables for inter-arrival times $\{T_i\}_{i\geq 1}$ and shock sizes $\{Y_i\}_{i\geq 1}$ to easily evaluate the main reliability quantities: the reliability function $R(t)$, the PDF $f(t)$ of the system's lifetime L and the n-moments $E(L^n)$ of L. The matrix-geometric properties of the PH distributions allow this easy and efficient computation compared with the direct evaluation of the convolutions or the Monte Carlo simulations. In addition, the proposed model is versatile and general. First, it only assumes independence of the processes $\{T_i\}_{i\geq 1}$ and $\{Y_i\}_{i\geq 1}$, between the random variables T_j and T_i, and between Y_j and Y_i (for $i \neq j$), although these variables can be not-identically distributed. And finally, PH distributions are dense in the set of positive distributions. By means of fitting algorithms (in particular the MM or the EM algorithm) a wide range of probability models can be approximated by adequate PH distributions. These results suggest that the use of PH distributions is a convenient and practical approach to address the problems of uncertainty, fitting and the reliability estimation of systems under degradation processes.

References

Aven, T. J. and Jansen, U. (1999) *Stochastic Models in Reliability*. Applications of Mathematics: stochastic modeling and applied probability series 41, Springer, New York.

Bobbio, A., Horváth, A., and Telek, M. (2005) Matching three moments with minimal acyclic phase type distributions. *Elsevier Science B*, vol 21(1), pp. 303–326.

Frangopol, D., Kallen, M.-J. and van Noortwijk, J. (2004) Probabilistic models for life-cycle performance of deteriorating structures: review and future directions.

Progress in Structural Engineering and Materials, vol. 6(4), pp. 197–212.

Iervolino, I., Giorgio, M., and Chioccarelli, E. (2013) Gamma degradation models for earthquake resistant structures. *Structural Safety*, vol 45, pp. 48–58.

Kharoufeh, J. P., Solo, C. J., and Ulukus, M. Y. (2010). Semi-markov models for degradation based reliability. *IIE Transactions,* vol 42(8), pp. 599–612.

Klutke, G.-A. and Yang, Y. (2002) The availability of inspected systems subject to shocks and graceful degradation. *IEEE Transactions on Reliability*, vol 51(3), pp. 371–374.

Latouche, G. and Ramaswami, V. (1999) *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Philadelphia: Society for Industrial and Applied Mathematics.

Montoro-Cazorla, D. and R. Pérez-Ocón (2006). Replacement times and costs in a degrading system with several types of failure: The case of phase-type holding times. *European Journal of Operational Research*, vol 175, pp. 1193–1209.

Montoro-Cazorla, D. and R. Pérez-Ocón (2012). A shock and wear system under environmental conditions subject to internal failures, repair, and replacement. *Reliability Engineering and System Safety*, vol 99, pp. 55–61.

Montoro-Cazorla, D., Pérez-Ocón, R., and Segovia, M. C. (2009). Shock and wear models under policy n using phase-type distributions. *Applied Mathematical Modelling*, vol 33, pp. 543–554.

Nakagawa, T. (2007). *Shock and Damage Models in Reliability*. London: Springer.

Neuts, M. F. (1981). *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. New York: Courier Dover Publications.

Osogami, T. and Harchol-Balter, M. (2003) A closed-form solution for mapping general distributions to minimal PH distributions. *Computer Performance Evaluation. Modelling Techniques and Tools*, vol 63(6), pp. 200–217.

Pandey, M. D., Yuan, X. X. and van Noortwijk, J. M. (2009) The influence of temporal uncertainty of deterioration on life-cycle management of structures. *Structure and Infrastructure Engineering*, vol. 5(2), pp. 145–156.

Riascos-Ochoa, J., Sanchez-Silva, M. and Akhavan-Tabatabaei, R. (2014). Stochastic shock models for degrading infrastructure systems via phase-type distributions. [Under review].

Sanchez-Silva, M., Klutke, G.-A. and Rosowsky, D. (2011). Life-cycle performance of structures subject to multiple deterioration mechanisms. *Structural Safety*, vol 33(3), pp. 206-217.

Segovia, M. and Labeau, P. (2013). Reliability of a multi-state system subject to shocks using phase-type distributions. *Applied Mathematical Modelling,* vol 37, pp. 4883–4904.

Thummler, A., Buchholz, P., and Telek, M. (2006). A novel approach for phase-type fitting with the EM algorithm. *IEEE Transactions on Dependable and Secure Computing,* vol 3(3), pp. 245–258.

van Noortwijk, J. (2009) A survey of the application of gamma processes in maintenance. *Reliability Engineering and Systems and Safet*y. vol 94, pp. 2–21.