A Bayesian Population Variability Analysis for Estimation of the Work Time Loss Distributions due to Occupational Accidents

M.C. Moura, R. Azevedo, E.L. Droguett, Isis Lins
CEERMA/UFPE, Brazil

R. Vilela
CHESF, Brazil

1. INTRODUCTION

Occupational accidents pose several negative consequences to employees, employers, environment and people surrounding the local where the accident takes place [1]. Therefore, it is imperative to develop adequate safety policies and operational procedures that minimize the probability of occurrence of occupational accidents. To do that, it is required to predict measures such as the expected work time loss.

As in failures analysis this process consists of two main steps: (i) inference of reliability quantities that explain the stochastic mechanism that originate the accident/recovery process such as accidents and recovery rates and (ii) characterization of accidental behavior by applying stochastic process analysis methods (e.g. Markov process).

A lot of research has been done on the quantitative analysis of occupational accident data. This includes modeling of occupational accident frequency [1 – 3] and work time losses [1, 4]. Several authors have used descriptive statistics and factor analysis, as Babu & Nayak [5]. From the 90's, Bayesian inference methods have long been used for the analysis of accident databases [6] to model uncertainties and enable future predictions but very few studies deal with prediction of work time loss [1].

For Meel et al. [3], the use of advanced Bayesian prediction methods is more suitable due to the scarcity and censorship of accident data. Meel & Seider [7] and Meel et al. [3] developed a Bayesian approach for the assessment of accident frequency in petrochemical industries in Europe. Marcoulaki et al. [1] extended the Meel's model by including analysis on work time loss and unavailable due occupational accidents.

Marcoulaki et al. [1] assumed that all workers have the same accident and recovery rates. Thus, the authors used data from all workers together and treated them as homogenous in order to formulate the likelihood function of the Bayesian model. However, due to the existence of subjective and individual characteristics it is expected that a same class of employees have a unique accident and recovery rates even though they have similar functions in the workplace or are allocated in the same occupational environment as other employees. Therefore, a population variability assessment over the rates is more appropriate for accidents analysis.

2. OBJETIVOS DO TRABALHO

Thus, this paper proposes a Bayesian population variability analysis-based method for this problem. Bayesian population variability analysis is an estimation procedure for the assessment of the variability of reliability measures among a group of similar systems/employees. Moreover, a Markov-based model is proposed for investigating future trends regarding occupational accidents in the workplace as well as enabling a better management of the labor force from the results of the population variability analysis. In fact, the key performance indicators here estimated by the Bayesian models will be the expected unavailability of the labor force and, consequently, the expected recovery time from an accident which will be computed by a Markov-based model. Thus, the Markov based model will work as a simulation algorithm to generate the accident and recovery times for the workers. To do that the numerical method proposed by Moura & Droguett [8] will be used because it is efficient to estimate the expected unavailability curve and associated uncertainty.

The remainder of this paper is organized as follows. Section 3 presents the theoretical background about the Bayesian Population Variability Analysis. Section 4 presents the proposed model, while Section 5 discusses the numerical results using evidence from a real accident database of a hydroelectric power company in Brazil. Finally, section 6 presents concluding remarks.

3. THEORETICAL BACKGROUND

3.1 *Bayesian population variability assessment for accident analysis.*

The failure and repair rates of a system are often used to characterize our expectations regarding the system's ability to perform its intended function. Likewise, we can use these measures to represent our prospects about the workers safety making the accident an expression of our personal uncertainty about the workers accidental behavior [9]. Thus, let ρ be a random variable that defines the accident or recovery rate of an employee and $\phi(\rho) = (\rho|\theta_1, \dots, \theta_r)$ denote the parametric variability distribution model. Then, a probability distribution $\pi(\theta_1, \dots, \theta_r)$ over the parameters $\theta_1, \dots, \theta_r$ of the model can be used to describe the uncertainty about the population variability distribution. The estimated population variability density is taken as

$$\hat{p}(\rho) = \int \dots \int_{\theta_1,\dots,\theta_r} \varphi(\rho|\theta_1, \dots, \theta_r).\pi(\theta_1, \dots, \theta_r).d\theta_1 \dots d\theta_r \quad (1)$$

Therefore, the estimated density function consists of a weighted mix of distributions of the chosen model. This is opposed to estimation methods as the Maximum Likelihood Estimators which is based in a single 'best' distribution chosen from the set of distributions.

The Bayesian population variability analysis of ρ assumes that the distribution $\pi(\theta_1,\dots,\theta_r)$ is uncertain and a Bayesian method is applied to infer it. For this, in addition to the collected accident data is necessary to have prior information about the population variability. Let E_0 be prior evidence that provides information about the π distribution, i.e. E_0 is the prior knowledge over π without considering data recorded (e.g. the expert's estimates about the probability distributions of $(\theta_1,\dots,\theta_r)$. Then, $\pi_0(\theta_1,\dots,\theta_r) = \pi(\theta_1,\dots,\theta_r \mid E_0)$ is the prior probability distribution of the population variability distribution parameters $(\theta_1,\dots,\theta_r)$ and $\hat{p}_0(\rho \mid E_0) = \int\dots\int_{\theta_1,\dots,\theta_r} \varphi(\rho \mid \theta_1,\dots,\theta_r) \times \pi_0(\theta_1,\dots,\theta_r).d\theta_1\dots d\theta_r$ is the prior population variability distribution of ρ.

Let the evidence E_1 include the information about the accidents database i.e. the information obtained from accidents records. Then, the employees' population distribution over ρ, conditional on E_0 and E_1 is given by

$$\hat{p}_1(\rho|E_0, E_1) = \frac{P(E_1|\rho, E_0).\int_{\underline{\theta}} \varphi(\rho|\underline{\theta}).\pi_0(\underline{\theta}).d\underline{\theta}}{\int_\rho P(E_1|\rho, E_0).d\rho.\int_{\underline{\theta}} \varphi(\rho|\underline{\theta}).\pi_0(\underline{\theta}).d\underline{\theta}} \quad (2)$$

where $\underline{\theta}=\{\theta_1,\dots,\theta_r\}$ and $P(E_1 \mid \rho, E_0)$ is the likelihood of the evidence E_1. The posterior distribution of the population variability parameters based on information types E_0 and E_1 is developed by applying the Bayes' theorem:

$$\pi_1(\underline{\theta}|E_0, E_1) = \frac{P(E_1|\underline{\theta}, E_0).\pi_0(\underline{\theta})}{\int_{\underline{\theta}} P(E_1|\underline{\theta}, E_0).\pi_0(\underline{\theta}).d\underline{\theta}} \quad (3)$$

Therefore, it's possible to evaluate the posterior population variability distribution of the accident or recovery rate by making $\hat{p}_1(\rho \mid E_0, E_1) = \int \underline{\theta}\ \varphi(\rho \mid \underline{\theta}) \times \pi_1(\underline{\theta} \mid E_0,E_1).d\underline{\theta}$ [10, 11].

Assuming that the accidents occurring in each employee are independent, the likelihood function of the exposure data becomes

$$P(E_1|\underline{\theta}, E_0) = \prod_{i=1}^{n} P(E_{1i}|\underline{\theta}, E_0) \quad (4)$$

where $P(E_{1i} \mid \underline{\theta},E_0)$ is the probability of observing evidence E_{1i} for the i-th out of n employees. Note that the accident measure for the i-th worker, ρ_i, is not known exactly and therefore $P(E_{1i} \mid \underline{\theta},E_0)$ is given as a function of ρ_i ($P(E_{1i} \mid \underline{\theta},E_0, \rho_i)$). We know that ρ_i is one of possibly many values of variable ρ. Moreover, according to our model, ρ is distributed according to $\varphi(\rho \mid \underline{\theta})$, and $\underline{\theta}$ is also unknown. Therefore, we calculate the probability of observing the information E_{1i} by allowing the accident measure to assume all possible values i.e. by averaging $P(E_{1i} \mid \underline{\theta},E_0, \rho)$ over the distribution of ρ:

$$P(E_{1i}|\underline{\theta}, E_0) = \int_{\rho} P(E_{1i}|\underline{\theta}, E_0, \rho).\varphi(\rho|\underline{\theta}).d\rho \quad (5)$$

4. A POPULATION VARIABILITY AND MARKOV BASED INTEGRATED MODEL

4.1 The timeline of the worker.

The assessment of the likelihood function (Eq. 4) is mainly based on the exposure data i.e. the number of accidents or recoveries by the operational time of a worker (runtime data). In general, the employment time is not necessarily the same for all workers. Therefore, only knowing the number of accidents suffered in an observation period is not sufficient. The observation of the timeline for each worker allows obtaining exposure data [1]. Then, consider that any occupational accident is reported and recorded. Typically, the registration and reporting forms of occupational accidents contains the accident date and the duration of the recovery period following the accident. This information, in addition the admission and dismissal dates, sets the timeline for each work as illustrated in Figure 1.

Figure 1 shows the timeline for a workplace with 12 workers during a period of observation starting at time T_S and terminating at T_F, where $T_{Si} \geq T_S$ and $T_{Fi} \leq T_F$ are, respectively, the starting and finishing times of i-th worker. If $T_{Si} > T_S$ then the worker was admitted after the initial observation period and, similarly, if $T_{Fi} < T_F$ then the worker was fired before the final observation period. In the interval [T_{Si}; T_{Fi}] the worker is involved in K_i accidents. The working time of worker n between k_{n-1} and k_n (k_n = 1, 2, …,K_n) is denoted by t_{n,K_n}.The respective recovery time is denoted by r_{n,K_n}. The time range between the recovery from the last accident, K_n, and T_{Fn} is denoted by s_n.

The staring times T_{S1}, T_{S2}, T_{S3}, T_{S6}, T_{S7}, T_{S9}, T_{S10} and T_{S11} coincide with T_S; this means these workers were hired before or at T_S. The finishing times T_{F1}, T_{F2}, T_{F3}, T_{F4}, T_{F5}, T_{F7}, T_{F10} and T_{F11} coincide with T_F i.e. the workers 1, 2, 3, 4, 5, 7, 10 and 11 remained working after T_F. Only workers 1, 2, 3, 7, 10 and 11 are employed over the whole exposure time, while the others work only for a fraction of this time. Workers 3, 5, 8 and 11 have no accidents, so $s_3 = T_{F3} - T_{S3}$, $s_5 = T_{F5} - T_{S5}$, $s_8 = T_{F8} - T_{S8}$ and $s_{11} = T_{F11} - T_{S11}$. The contract of worker 9 expires when he recovers from his last accident, thus $s_9 = 0$ (it is noteworthy that some regulations protect workers from this type of situations, giving him a labor warranty for a specified period after the recovery from an accident. However, this example is useful for illustration). Ultimately, the finish time (T_F) occurs before the worker 10 back to work, therefore the recovery time $r_{10,1}$ is said to be censored and there is no s_{10}.

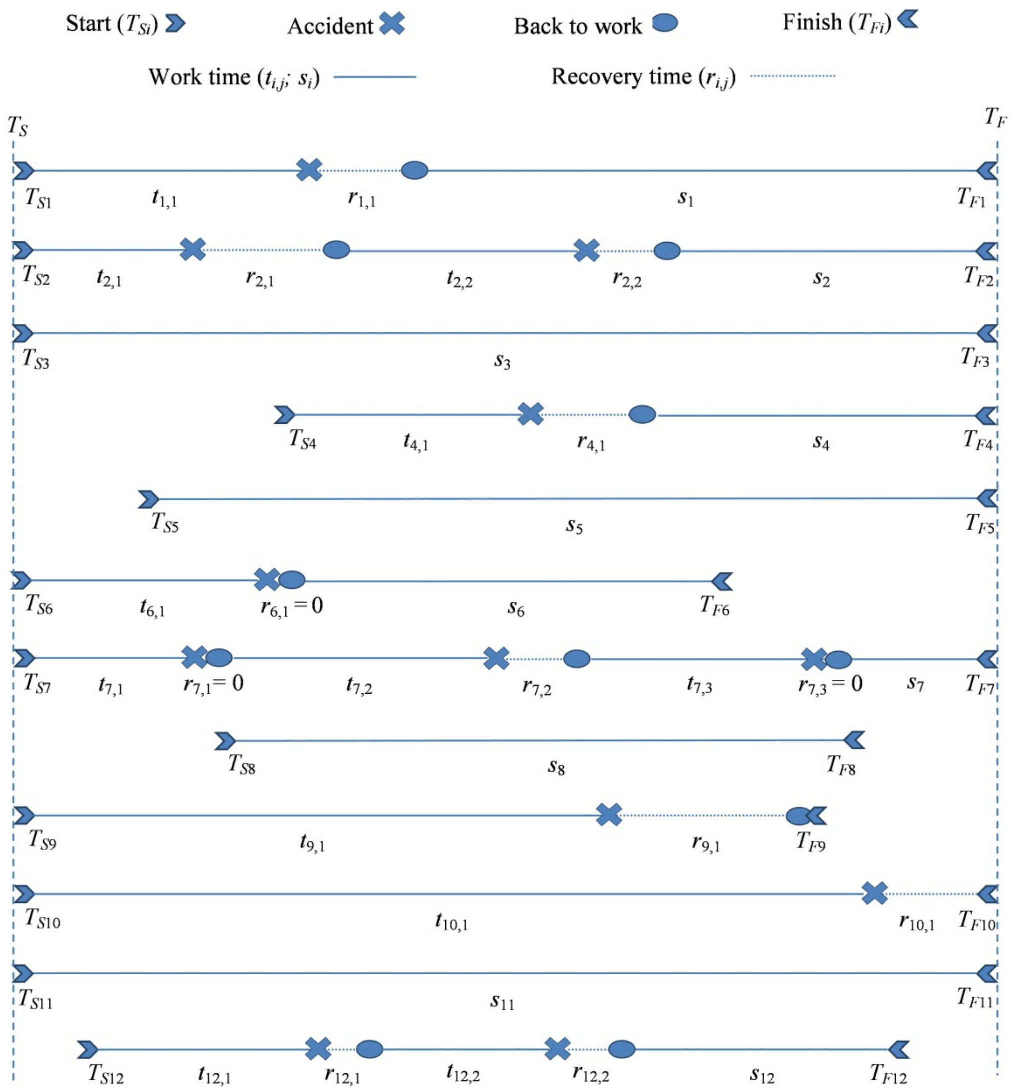

Figure 1 – Timelines for a workplace with 12 workers.
Source: Adapted from Marcoulaki et al (2012).

Note that $T_n = (\sum_{i=1}^{K_n} t_{n,k_i}) + s_n = T_{Fn} - T_{Sn} - R_n$ (where $R_n = \sum_{i=1}^{K_n} r_{n,k_i}$) represents the total time for which the worker n was submitted to the risks of occupational accidents. Thus, we can determine the exposure data of each work by the pair (K_n, T_n). Similarly we can determine the exposure data for the recovery data by the pair (K_{Rn}, R_n), where K_{Rn} is the number of recoveries of the worker n. Thus, it is possible to use the censored data found in the database available. Table 1 illustrates the exposure data extracted from in Figure 1. The information in Table 1 provides a clear view of the E_1 evidence needed for the likelihood function construction in the population variability model ($E_{1\lambda}$ corresponds to E_1 type evidence for the accidents rate analysis and $E_{1\mu}$ corresponds to E_1 type evidence for the recovery rate analysis).

Table 1 – Exposure data (E1 evidence) for Figure 1 example.

Worker	E_1^{λ}	E_1^{μ}

(i)	K_i	T_i	K_{Ri}	R_i
1	1	$t_{1,1}+s_1$	1	$r_{1,1}$
2	2	$t_{2,1}+t_{2,2}+s_2$	2	$r_{2,1}+r_{2,2}$
3	0	s_3	-	-
4	1	$t_{4,1}+s_4$	1	$r_{4,1}$
5	0	s_5	-	-
6	1	$t_{6,1}+s_6$	1	$r_{6,1}$
7	3	$t_{7,1}+t_{7,2}+t_{7,3}+s_7$	3	$r_{7,1}+r_{7,2}+r_{7,3}$
8	0	s_8	-	-
9	1	$t_{9,1}$	1	$r_{9,1}$
10	1	$t_{10,1}$	0	$r_{10,1}$
11	0	s_{11}	-	-
12	2	$t_{12,1}+t_{12,2}+s_{12}$	2	$r_{12,1}+r_{12,2}$

4.2 The population variability analysis construction

4.2.1 Mixed likelihood model

In order to perform a population variability analysis of accident or recovery rate we need to specify an appropriate probability distribution to describe the underlying variability of the measure of interest, $\varphi(\rho \mid \underline{\theta})$, as well as construct the likelihood function $P(E_1 \mid \underline{\theta}, E_0)$. The specification of the probability distribution describing the variability may be guided by the nature of the accident measure. According to Droguett et al [12], the use of the Gamma or Lognormal distribution is suitable for representing the variability in the number of failure of the various systems with different failure rates. Thus, let us consider the population variability of the accident measure is given by a lognormal distribution i.e.

$$\varphi(\rho|v,\tau) = \frac{1}{\sqrt{2\pi}\rho\tau} e^{-\frac{1}{2}\left(\frac{\ln\rho - v}{\tau}\right)^2} \tag{6}$$

where v and τ are the parameters (mean and standard deviation). The Lognormal distribution can be written in terms of median (ρ_{50}) and error factor ($EF\rho$) of the distribution by making the following changes: $\rho_{50} = e^{v}$ and $EF\rho = e^{1.645\tau}$, as illustrated in Equation (7). The posterior of the variability distribution parameters in Equation (3) can be written as Equation (8).

$$\varphi(\rho|\rho_{50}, EF_\rho) = \frac{1}{\sqrt{2\pi}.\rho.\frac{\ln EF_\rho}{1.645}} e^{-\frac{1}{2}\left(\frac{\ln\rho - \ln\rho_{50}}{\ln EF_\rho/1.645}\right)^2} \tag{7}$$

$$\pi_1(\rho_{50}, EF_\rho|E_0, E_1) = \frac{P(E_1|\rho_{50}, EF_\rho, E_0).\pi_0(\rho_{50}, EF_\rho)}{\int_{\rho_{50}}\int_{EF_\rho} P(E_1|\rho_{50}, EF_\rho, E_0).\pi_0(\rho_{50}, EF_\rho).dEF_\rho.d\rho_{50}} \tag{8}$$

The likelihood construction is an evidence-driven process i.e. it is dependent on the type of the available evidence E_1. In the analysis of accidents is difficult to obtain expert estimates on the accidents rate of each worker individually. In general, the exposure data (as in Table 1) are the only available evidences. Moreover, due to non-homogeneity among workers (as discussed in Section 3), it is not representative to obtain such estimates directly from the accidents databases.

In this way, the likelihood can be constructed. If we know the accident rate $\lambda_i = \lambda$ or the recovery rate $\mu_i = \mu$ of each employee (ρ represents both rates), we can use the Poisson distribution to estimate the likelihood of observing q_i events (K_i accidents or K_{Ri} recoveries) in the time w_i (T_i or R_i):

$$P(q_i|w_i,\rho,E_0) = \frac{(\rho.w_i)^{q_i}.e^{-\rho.w_i}}{\Gamma(q_i+1)} \quad (9)$$

As we only know that ρ is one of the possible values of the rate represented by its population variability distribution $\varphi(\rho\,|\,\rho_{50},EF\rho)$, we average the likelihood given by Equation (5) over all possible values of ρ in order to calculate the probability of the data unconditional on the unknown value of ρ:

$$P\left(q_i,w_i|\rho_{50},EF_\rho,E_0\right) = P_i = \int_\rho P(q_i|w_i,\rho,E_0).\varphi\left(\rho|\rho_{50},EF_\rho\right).d\rho \quad (10)$$

Replacing Equations (7) and (9) into Equation (10):

$$P_i = \int_\rho \frac{(\rho.w_i)^{q_i}.e^{-\rho.w_i}}{\Gamma(q_i+1)}.\frac{1}{\sqrt{2\pi}.\rho.\frac{\ln EF_\rho}{1.645}}e^{-\frac{1}{2}\left(\frac{\ln\rho-\ln\rho_{50}}{\ln EF_\rho/1.645}\right)^2}d\rho \quad (11)$$

The total likelihood is then obtained by replacing Equation (11) into Equation (4).

$$P\left(E_1|\underline{\theta},E_0\right) = \prod_{i=1}^{n}\int_\rho \frac{(\rho.w_i)^{q_i}.e^{-\rho.w_i}}{\Gamma(q_i+1)}.\frac{1}{\sqrt{2\pi}.\rho.\frac{\ln EF_\rho}{1.645}}e^{-\frac{1}{2}\left(\frac{\ln\rho-\ln\rho_{50}}{\ln EF_\rho/1.645}\right)^2}d\rho \quad (12)$$

4.2.2 A prior distribution specification

Application of population variability analysis re-quires specifying the prior distribution defined over the parameter space of the chosen variability model, representing the prior belief about the variability distribution before the evidence (accident database) be-comes available. Some authors propose that the prior be specified in the form of a discrete prior distribution [10, 13].

The algorithm here adopted involves the specification of an informed continuous prior over the parameter space of the variability model. The analyst is required to provide initial estimates in terms of a central value (median) and the extent of variability (error factor) in the population variability distribution. As discussed in Droguett et al [12], these estimates take the form of Lognormal distributions. If the parameters v and τ of the population variability distribution of ρ are distributed as a Lognormal distribution, then the distributions of $\rho_{50} = e^v$ and $EF\rho = e^{1.645\tau}$ take the forms:

$$f(\rho_{50}) = \frac{1}{\sqrt{2\pi}.\rho_{50}.\frac{\ln\varepsilon_{\rho_{50}}}{1.645}}e^{-\frac{1}{2}\left(\frac{\ln\rho_{50}-\ln\gamma_{\rho_{50}}}{\ln\varepsilon_{\rho_{50}}/1.645}\right)^2} \quad (13)$$

$$f\left(EF_\rho\right) = \frac{1}{\sqrt{2\pi}.EF_\rho.\frac{\ln\varepsilon_{EF_\rho}}{1.645}}e^{-\frac{1}{2}\left(\frac{\ln EF_\rho-\ln\gamma_{EF_\rho}}{\ln\varepsilon_{EF_\rho}/1.645}\right)^2} \quad (14)$$

where γ_x and ε_x are the median and error factor of the distribution of x (where x can represents both the median and error factor of ρ). Considering the independence between v and τ then the prior distribution of variability model is $\pi_0(\rho_{50}, EF\rho) = f(\rho_{50}) \times f(EF\rho)$. The definition of prior distributions consists in to determine the values $\gamma_{\rho 50}$, $\varepsilon_{\rho 50}$, $\gamma_{EF\rho}$ and $\varepsilon_{EF\rho}$ (e.g. by specialist's opinions):

$$\pi_0\left(\rho_{50},EF_\rho\right) = f\left(\rho_{50}|\gamma_{\rho_{50}},\varepsilon_{\rho_{50}}\right).f\left(EF_\rho\middle|\gamma_{EF_\rho},\varepsilon_{EF_\rho}\right) \quad (15)$$

So the set $\{\gamma_{\lambda 50}, \varepsilon_{\lambda 50}, \gamma_{EF\lambda}, \varepsilon_{EF\lambda}\}$ corresponds to the E_0 evidence for accident rate analysis and the set $\{\gamma_{\mu 50}, \varepsilon_{\mu 50}, \gamma_{EF\mu}, \varepsilon_{EF\mu}\}$ is the E_0 evidence for recovery rate analysis.

4.2.3 Variability Measures

The likelihood functions and prior distributions have been incorporated in a Bayesian inference procedure in which the posterior density $\pi_1(\underline{\theta} \mid E)$ is computed. The Bayesian inference is performed by using a Markov Chain Monte Carlo Method (MCMC), which allows samples to be generated from a continuous unnormalized density [14]. The MCMC method, which is frequently applied to Bayesian Inference problems [15], results in a m-samples set $S = \{\underline{\theta}_1,\ldots,\underline{\theta}_m\}$, representing the posterior density over the parameters of the variability distribution model $\varphi(\rho \mid \underline{\theta})$. Given S, the estimated population variability density is computed as

$$\hat{p}(\rho) = \frac{1}{m}\sum_{i=1}^{m} \varphi\left(\rho|\underline{\theta}_i\right) \tag{16}$$

The corresponding mean and variance are computed as

$$\hat{\mu}_\rho = \frac{1}{m}\sum_{i=1}^{m} \mu(\underline{\theta}_i) \tag{17}$$

$$\hat{\sigma}_\rho{}^2 = \frac{1}{m}\sum_{i=1}^{m} \sigma^2(\underline{\theta}_i) \tag{18}$$

where $\mu(\underline{\theta})$ and $\sigma^2(\underline{\theta})$ are the mean and variance of $\varphi(\rho \mid \underline{\theta})$.

Furthermore, the generated results include uncertainty bounds of the cumulative variability density $P(\rho) = \int_{x=0}^{\rho} p(x)dx$ in the form of z^{th} percentiles $P_z(\rho)$, defined as $\Pr(P(\rho) < P_z) = z/100$. The z^{th} percentiles are determined by finding the value $\hat{P}_z(\rho)$ for which for a fraction z/100 of the samples $\underline{\theta_i} \in$ S:

$$\int_{x=0}^{\rho} \varphi\left(x|\underline{\theta}_i\right).dx < \hat{P}_z(\rho) \tag{19}$$

These bounds provide the analyst with a basis to assess the uncertainty associated with the estimated population variability distribution related to accidents or recovery rates.

4.3 Estimation of the Work Time Loss Distribution by Markov-based model

As discussed in Section 3 it is assumed that each employee has a unique accident rate even that they have functions similar or be allocated in the same occupational environment as the other employees. This assumption is due to the existence of idiosyncrasies that make every worker particular.

The random behavior, which leads to occupational accidents, and the corresponding recovery of the worker, can be represented by a Markov-based stochastic process as shown in Figure 2. Workers who are in the first state are performing their function normally (they are available for work), while those in the second state are recovering from an accident (on medical leave). Random mechanisms, which lead the worker i to an accident, occur with rate λ_i, so that if the worker was available (State 1) he becomes unavailable (state 2). Another random mechanism recovers the victim to the available status with rate μ_i. It is assumed that the two rates, λ_i and μ_i, are independent.

If we know the accident rate λ_i and the recovery rate μ_i we can determine the unavailability curve of the worker by applying some method of solution of the Markov diagram (numerical methods, analytical treatment, simulation, etc.). However, we only know that λ_i and μ_i are one of possibly many values of the variables λ and μ, distributed as $p(\lambda)$ and $p(\mu)$ respectively. $p(\lambda)$ and $p(\mu)$ are estimated by $\hat{p}_1(\lambda \mid E_{0\lambda},E_{1\lambda})$ and $\hat{p}_1(\mu \mid E_{0\mu},E_{1\mu})$ from the Bayesian model presented in Section 4.2.

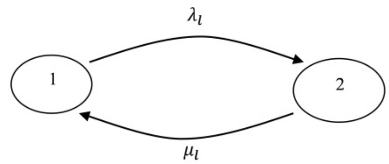

Figure 2 – Markov diagram of the stochastic process.

Of course, if workers have different accident and recovery rates, they have different unavailability curves. We can simulate the stochastic process linked to accidental behavior of a worker by generating the values λ_i and μ_i from $\hat{p}_1(\lambda \mid E_{0\lambda}, E_{1\lambda})$ and $\hat{p}_1(\mu \mid E_{0\mu}, E_{1\mu})$. By doing this, we obtain the unavailability curve by solving the Markov process. However, we can't determine the worker for which this result is valid. Repeating this process N times (N must be large enough to ensure the representativeness of the simulation), we obtain an N-samples set where each sample contains one curve of unavailability. Thus, we estimate the average unavailable curve ($\widehat{U}(t)$) as

$$\widehat{U}(t) = \frac{1}{N}\sum_{i=1}^{N} U_i(t) \qquad (20)$$

where $U_i(t)$ corresponds to the curve of the unavailability resulting from the i-th iteration of Markov process.

It is acceptable to assume that $\widehat{U}(t)$ represents the fraction of man-hour loss due to occupational accidents. Also, it is possible to define the uncertainty associated with this estimate by the z^{th} percentiles of the set $\{U_i(t), \ldots, U_N(t)\}$.

In this paper, we use the numerical treatment based on transition frequency densities by Moura & Dorguett [8] to infer unavailability curves. The Population Variability and Markov-based integrated model can then be resumed by Figure 3.

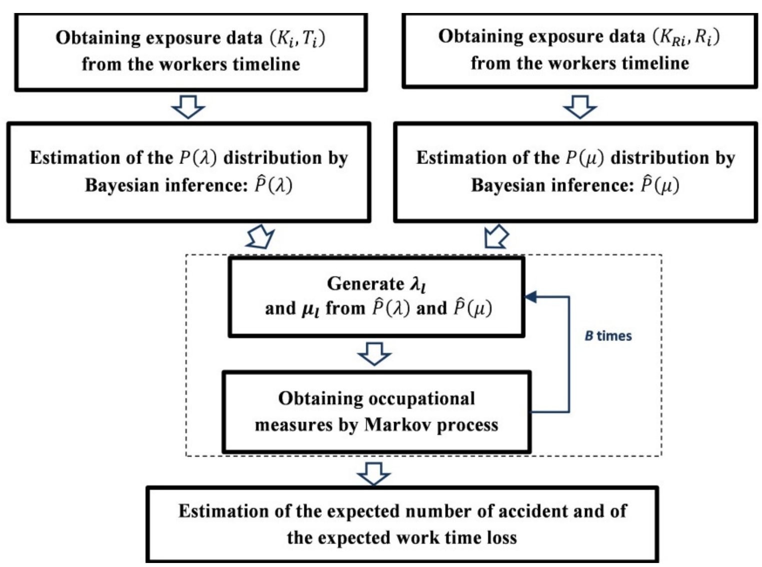

Figure 3 – Proposed Model.

5. EXAMPLE APLICATION

In this section, we illustrate and discuss the use of the Bayesian Population Variability analysis and Markov-based models by means of an example. We start by supposing that we are interested in assessing the average distribution of work time loss due to occupational accidents for workers of a hydroelectric power company in Brazil. Runtime data were collected from the timeline of operation employees between 01/01/2005 and 09/31/2012 in order to construct the likelihood function.

As discussed in Section 3.1 the Bayesian Population Variability Analysis must be used in a workers population subject to similar risk of accidents. So, in this section, the model is applied to employees with the same job and workplace. The administrators located in the operation sector of the company were analyzed. These workers suffered mainly two types of accidents: (1) accidents in commuting and (2) falls. It is expected that one worker has different accident and recovery rates for each type of accident. Therefore, the model was applied separately for analyzing the work time loss distributions due to accidents in commuting and falls.

5.1 Results

57 workers were analyzed in a total of 157,784 men-days of work. 27 accidents were recorded in the period, of which 13 are accidents in commuting and 14 are falls. Upon solving the Bayesian population variability and Markov-based integrated model, the expected unavailability of the population of workers has a 5th and 95th percentiles of 0.00017 and 0.0043 due accidents in commuting an of 0.00025 and 0.008 due falls, respectively, with a mean of 0.0013 for accidents in commuting and of 0.0022 for falls.

If $T_L(t)$, $T_L^{5\%}(t)$ and $T_L^{95\%}(t)$ correspond to the mean and the 5th and 95th percentiles distributions of expected work time lost by one worker of population then the mean and the 5th and 95th percentiles distributions of expected work time lost in the population of workers ($T_P(t)$, $T_P^{5\%}(t)$ and $T_P^{95\%}(t)$) can be obtained by multiplying the $T_L(t)$, $T_L^{5\%}(t)$ and $T_L^{95\%}(t)$ by the number of workers in population. This means that the expected work time loss by workers located in the operation sector in one year is $0.46 \times 57 = 26.35$ men-days due to accidents in commuting and $0.81 \times 57 = 46.33$ men-days due to falls, amounting 72.68 men-days loss per year among the workers located in operation sector of the hydroelectric power company. The corresponding 5th and 95th percentiles are $0.063 \times 57 = 3.60$ and $1.58 \times 57 = 90.10$ for accidents in commuting and $0.09 \times 57 = 5.10$ and $2.9 \times 57 = 165.81$ for falls. Figure 4 and Figure 5 illustrate the expected work time loss curves for accidents in commuting and for falls.

5.1.1 Validation

To validate the application of the model we generate the data from the results of the Bayesian population variability model in order to compare them with the actual data. The Bayesian model applied in the real case provides estimations for population variability distributions of accidents and recoveries rates. From these distributions we generate the accidents and recoveries data of workers. It is expected that these data can represent the accident-recovery process of this population of workers.

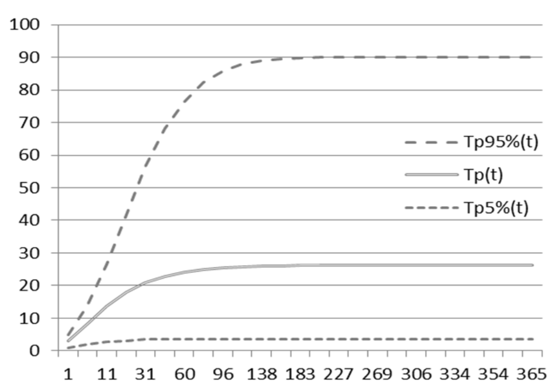

Figure 4 – Work time loss distributions for accidents in commuting.

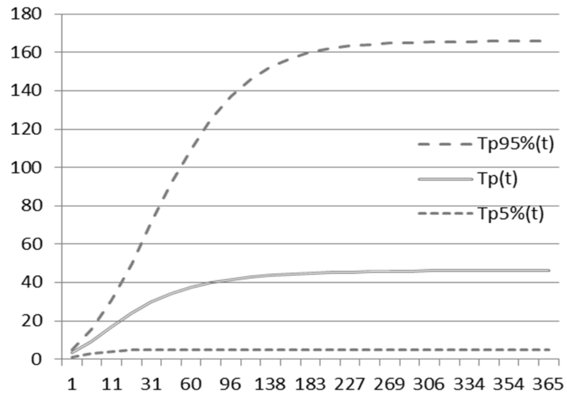

Figure 5 – Work time loss distributions for falls.

These data were compared with the actual data by application of the statistic Kruskal hypothesis test [16] under the null hypothesis that the data follow the same distribution. Three tests were realized for each type of accident: (i) test on time to accident, (ii) test on time to recovery and (iii) test on unavailability, amounting six (6) hypothesis tests. Table 2 shows the *p*-values of each test.

All *p*-values are greater than 0.1, then we can say that there is no evidence to reject the null hypothesis. Furthermore, we can accept the null hypothesis to be true i.e. the simulated and actual data come from the same distribution since the *p*-values in Table are large. This result validates the proposed model in this paper because it proves the efficiency of the model to represent the real population.

Table 1 – *p*-values of the hypothesis tests.

Test	accidents in commuting	falls
Time to accident	0.1871	0.125
Time to recovery	0.2606	0.1755
Unavailable	0.2606	0.1705

6. CONCLUDING REMARKS

The model presented in this paper represents an extension of the Bayesian population variability assessment method in accidents analysis. A Markov-based model is used for estimation of the expected work time loss distributions due to occupational accidents. The Bayesian population variability assessment method allows the evaluating of population variability of accidents and recovery rates based on exposure data of workers submitted to same occupational risks and the Markov-based model is used to derive the worker unavailability statistics to predict the amount of time that workers will be recovering from accidents and therefore won't be available to perform the job they are paid for. The use of the population variability analysis allows to assess the uncertainty presented on the results of the Markov-based model.

The models developed here can be informed using available databases of occupational accidents documented in the industries. Sufficient statistics to use the models include only the number of workdays, the workdays loss due to recovering from occupational accidents, and the number of occupational accidents over the period of observation for each worker.

The implementation of the extension used in the examples in Section 5 has shown that the numerical solution of the models is feasible and provides a good estimation for the expected work time loss distribution due to occupational accidents. Also, the comparison of simulated data, from the population variability distributions, with real data showed that the estimated population variability distributions of the accidents and recoveries rates can be used to represent the actual situation.

7. REFERÊNCIAS

[1] MARCOULAKI E.C. et al, “Prediction of occupational accident statistics and work time loss distributions using Bayesian analysis”, *Journal of Loss Prevention in the Process Industries*, 25: 467-477, 2012.

[2] MARCOULAKI, E.C., KONSTANDINIDOU, M. & PAPAZOGLOU, I.A., “Dynamic failure assessment of incidents reported in the Greek Petrochemical Industry”, *Computer-Aided Chemical Engineering*, 29: 1055-1059, 2011.

[3] MEEL, A. et al, “Operational risk assessment of chemical industries by exploiting incident databases”, *Journal of Loss Prevention in the Process Industries*, 20: 113-127, 2007.

[4] BLANCH, A., TORRELLES, B. & SALINAS, J.A., “Age and lost working days as a result of an occupational accident: a study in a shift work rotation system”, *Safety Science*, 47 (10): 1359-1363, 2009.

[5] BABU, B.V. & NAYAK, A.N., “Treatment costs and work time loss due to episodic adenolymphangitis in lymphatic filariasis patients in rural communities of Orissa, India”, *Tropical Medicine & International Healt*, 8:1102-1109, 2003.

[6] HORA, S.C. & IMAN, R.L., “Bayesian modeling of initiating event frequencies at nuclear power plants”, *Risk Analysis*, 10 (1): 103-109, 1990.

[7] MEEL, A. & SEIDER, W.D., “Plant-specific dynamic failure assessment using Bayesian theory”, *Chemical Engineering Science*, 61: 7036–7056, 2006.

[8] MOURA, M.C. & DROGUETT, E.L., “Mathematical formulation and numerical treatment based on transition frequency densities and quadrature methods for non-homogeneous semi-Markov processes”, *Reliability Engineering &System Safety*, 94: 342-349, 2009.

[9] SINGPURWALLA, N.D., “Foundational Issues in Reliability and Risk Analysis”, *SIAM Review*, 30 (2): 264-282, 1988.

[10] KAPLAN, S. 1983, “On a ‘Two-Stage’ Bayesian Procedure for Determining Failure Rates from Experimental Data”, *IEEE Transactions on Power Apparatus and Systems*, 102 (1):195-199, 1983.

[11] DEELY, J.J. & LINDLEY, D.V. “Bayes Empirical Bayes”, *Journal of the American Statistical Association*, 76 (376): 833-841, 1981.

[12] DROGUETT, E. et al., “The Combined Use of Data and Expert Estimates in Population Variability Analysis”, *Reliability Engineering and System Safety*, 83: 311-321, 2004.

[13] Pickard, Lowe and Garrick, Inc., “Seabrook Station Risk Management and Emergency Planning Study,” prepared for Public Service Company of New Hampshire and Yankee Atomic Electric Company, PLG-0432, December 1985.

[14] CHIBAND, S. & GREENBERG, E., “Understanding the Metropolis-Hastings Algorithm”, *The American Statistician*, 49 (4): 327-335, 1995.

[15] GILKS, W.R., RICHARDSON, W.R. & SPIEGELHALTER, D.J., *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London (1996).

[16] KRUSKAL, W.H. & WALLIS, W.A., “Use of Ranks in One-Criterion Variance Analysis”, *Journal of the American Statistical Association*, 47 (260): 583-621, 1952.