Reliability Engineering and System Safety 150 (2016) 136–146

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Estimation of expected number of accidents and workforce unavailability through Bayesian population variability analysis and Markov-based model

Márcio das Chagas Moura [a,b,*], Rafael Valença Azevedo [a,b], Enrique López Droguett [c], Leandro Rego Chaves [d], Isis Didier Lins [a,b], Romulo Fernando Vilela [e], Romero Sales Filho [f]

[a] CEERMA-Center for Risk Analysis, Reliability and Environmental Modeling, Universidade Federal de Pernambuco, Av. da Arquitetura, s/n, CEP: 50.740-550, Cidade Universitária, Recife, PE, Brazil
[b] Department of Production Engineering, Universidade Federal de Pernambuco, Rua Acadêmico Hélio Ramos, s/n, Cidade Universitária, CEP: 50740-530, Recife, PE, Brazil
[c] Mechanical Engineering Department, University of Chile, Santiago, Chile
[d] Department of Statistics and Applied Mathematics, Federal University of Ceará, Avenida da Universidade, 2853 - Benfica, Fortaleza 60020-181, CE, Brazil
[e] CHESF-São Francisco's Hydroelectric Company, Rua Delmiro Gouvêia, 333 - San Martin, CEP: 50761-901, Recife, PE, Brazil
[f] Academic Unit of Garanhuns, Federal Rural University of Pernambuco, Garanhuns, Brazil

A R T I C L E I N F O

Article history:
Received 13 May 2015
Received in revised form
18 January 2016
Accepted 27 January 2016
Available online 5 February 2016

Keywords:
Occupational accidents
Bayesian variability analysis
Markov model
Expected number of accidents
Workforce unavailability

A B S T R A C T

Occupational accidents pose several negative consequences to employees, employers, environment and people surrounding the locale where the accident takes place. Some types of accidents correspond to low frequency-high consequence (long sick leaves) events, and then classical statistical approaches are ineffective in these cases because the available dataset is generally sparse and contain censored recordings. In this context, we propose a Bayesian population variability method for the estimation of the distributions of the rates of accident and recovery. Given these distributions, a Markov-based model will be used to estimate the uncertainty over the expected number of accidents and the work time loss. Thus, the use of Bayesian analysis along with the Markov approach aims at investigating future trends regarding occupational accidents in a workplace as well as enabling a better management of the labor force and prevention efforts. One application example is presented in order to validate the proposed approach; this case uses available data gathered from a hydropower company in Brazil.

1. Introduction

Occupational accidents pose several negative consequences to employees, employers, environment and people surrounding the locale where the event takes place. In this context, the development of quantitative models may aid the decision makers to determine adequate safety policies and operational procedures that in turn reduce the probability of occurrence and/or the severity of these undesired events.

Indeed, Moscoso et al. [31] assert that the analysis of the accident rate can be of great help to determine the safety procedures to be implemented. Yet, the investigation of the recovery time from accidents is also equally important in cases where the evaluation of unavailability and work time loss is of interest. In fact, significant research effort has been put on the quantitative analysis of occupational accidents regarding the assessment of rates of accident and recovery. For instance, Cawley and Homce [9], Chia et al. [11], Konstandinidou et al. [24] and Moscoso et al. [31] discuss occupational accidents in the electrical, construction, petrochemical and olive oil facilities, respectively, and assess the rates of accident and recovery for different causes. Camino et al. [7], Blanch et al. [6] and Lilley et al. [25] in turn analyzed the relationship between the severity of accidents, in terms of their recovery rate, and various organizational and individual variables such as age, gender, type of contract, time of accident, length of service in the company, company size, day of the week, among others. Carnero and Pedregal [8] and Freivalds and Johnson [18] evaluate accident rates in order to show the evolution and the profile of occupational accidents over the year, providing relevant information to detect periods where careful attention should be taken to improve safety. Furthermore, various researches have analyzed accident rates with the purpose to estimate the cost of

* Corresponding author at: Department of Production Engineering, Universidade Federal de Pernambuco, Rua Acadêmico Hélio Ramos, s/n, Cidade Universitária, CEP: 50740-530, Recife, PE, Brazil. Fax: +55 81 2126 7112.
E-mail addresses: marcio@ceerma.org, marciocmoura@gmail.com (M. das Chagas Moura).

http://dx.doi.org/10.1016/j.ress.2016.01.017
0951-8320/

accidents [21,22,37,38,40]. In this context, Yakovlev and Sobel [40] showed that profit increases with the safety investment because the relationship between the prevention expenditure and the accident rates is negative.

Initiatives such as the Workgroup Occupational Risk Model (WORM) project, which has been carried out by the Ministry of Social Affairs and Employment of the Netherlands, have also given rise to a set of quantitative models that assess occupational risks. For example, Ale et al. [3] and Ale et al. [1] developed an Occupational Risk Model (ORM) that basically consists in three stages: (i) analyzing accident data; (ii) tailoring these data into a Functional Block Diagram (FBD); and (iii) optimizing solutions for risk reduction. The first step of the ORM uses a tool called "Storybuilder" that systematically classifies and analyses reports of past accidents. Ale et al. [2] also adopted the "Storybuilder" to investigate accident reports in the construction industry in The Netherlands. Papazoglou et al. [33], Aneziris et al. [4], and Aneziris et al. [5] used FBD to evaluate the occupational risk of falling from mobile ladders, for fall from height, and for crane activities respectively. Papazoglou et al. [34] in turn presented a model based on a homogeneous Poisson stochastic process to estimate the probabilities for occupational accidents per hour of exposure and for a year of average exposure considering three possible consequences: death, permanent and recoverable injuries.

A common characteristic of most of the above-mentioned papers is that they rely on either "top-down" (national data and general statistics) or "bottom-up" (records gathered from various organizations) data-collection methods. Jallon et al. [22] state that instead of "top-down" and "bottom-up" strategies, the "local" approach, which is based on in-company data, allows the accurate assessment of occupational accidents in specific workplaces. However, "local" accident recordings are generally sparse and contain many censored records and, according to Meel et al. [28], classical statistical approaches are ineffective in this context. Therefore, the use of Bayesian methods may be more suitable for the "local" approach.

Indeed, Meel et al. [28] and Marcoulaki et al. [26] developed Bayesian approaches for the assessment of the rate of accidents. Marcoulaki et al. [27] extended these Bayesian models by considering sick leaves (time necessary for an employee to be recovered from an accident), and the associated work time loss. To this end, a prior Gamma distribution was considered to model both the rates of accident and recovery. The use of the Gamma distribution allowed an analytical solution for formulating the posterior distributions because it forms a conjugated pair with the Poisson distribution, which in turn was used in the construction of the likelihood function. From these distributions, Marcoulaki et al. [27] developed closed-form results for some occupational measures such as the accident rates, duration of recovery from an accident, and the worker unavailability. Papazoglou et al. [32] also adopted a Bayesian-based approach to assess the uncertainty in the quantification of risk rates of occupational accidents.

However, the models of Meel et al. [28], Marcoulaki et al. [27] and Papazoglou et al. [32] as well as the classical statistics based approaches have been established under the assumption that all company workers have homogeneous/identical behavior in terms of the occurrence of accidents, i.e., they assumed that rates of accident and recovery are the same for all workers. Despite that, it has been shown that due to the existence of individual factors (age, gender, experience, time, etc.) it is expected that workers have distinct rates of accident and recovery even in the situation they have similar roles in the workplace [10,36,6]. However, as pointed out by Fragola [17], the more inhomogeneous the database becomes, the tighter the uncertainty bounds due to the larger size of the aggregated population. This occurs when inhomogeneous data are aggregated as if they were homogeneous [14]. Therefore, in this case, even though the rates of accident and recovery would be better estimated in accordance with the statistical sense because the confidence intervals would be narrower, those occupational measures would be less representative of each subpopulation that composes the mixture.

Thus, it is important to analyze the variability of these rates in a population of workers in order to forecast the random behavior of the occupational accidents within a workplace. However, in a "local" perspective, we may have little or no availability of accidents data, mainly if it is desired to make a categorized analysis (for example, by type of accidents). Then, other sources of information such as data from similar facilities may be used to draw occupational measures of interest.

Therefore, this paper proposes a Bayesian method to estimate rates of accident and recovery. We consider that each worker has unique rates of accident and recovery and the Bayesian Population Variability Analysis (BPVA), also known as the first phase in a two-stage Bayesian, or hierarchical Bayes [13], is here used to estimate the variability distribution of these rates within a group of non-homogeneous workers. Unlike Marcoulaki et al. [27], where the rates of accident and recovery are considered homogeneous for different workers and conjugate distribution pairs are used, the solution here is not analytically obtainable, and thus we resort to a Markov Chain Monte Carlo (MCMC) approach to draw the non-parametric posterior distributions. Hence, we use the outputs of the BPVA, which are the distributions of the rates of accident and recovery, to feed a two-state Markov-based model that in turn will estimate the expected number of accidents and the expected unavailability of the labor force. These quantities will provide information to investigate future trends regarding occupational accidents in the workplace as well as enabling a better management of the labor force. At the best of authors' knowledge, no article has yet adopted BPVA to model occupational accidents. Indeed, the procedures here implemented for BPVA are based on the mathematical methods developed for risk and reliability analysis, such as those in Kaplan [23], Mosleh and Apostolakis [30], Pörn [35] and Droguett et al. [13]. In this work, these methods are tailored for the context of analysis of occupational accidents.

The remainder of this paper is organized as follows. Section 2 presents the theoretical background about the BPVA in the context of occupational accidents. Section 3 presents the proposed model, illustrating the implementation of the proposed BPVA-Markov procedure. Section 4 validates the model from an example application, which uses evidence from real accident reports of a hydropower company in Brazil. Finally, Section 5 provides some concluding remarks.

2. Bayesian population variability assessment

Similarly to Singpurwalla [39] claims in the context of reliability analysis, we here argue that rates of accident λ and recovery μ are expressions of our personal uncertainty about the workplace's dynamic behavior. Given that, we can associate these quantities to each individual worker $i = 1, \ldots, m$, and then consider the variability of rates over the whole population. A representation of this variability, in the form of a probability distribution, is referred to as the Population Variability Distribution (PVD) and its assessment is named Population Variability Analysis (PVA).

As in procedures of PVA [13,20,23,35], we assume that a member of a given family of a parametric distribution may describe PVD. If we have enough data for each i-th worker, it is possible to define the parameters of PVD of λ and μ directly from the dataset. In this section, we review the BPVA based on Deely and Lindley [12], Kaplan [23], Mosleh and Apostolakis [30], Pörn [35] and Droguett et al. [13].

In fact, let ρ be a random variable that defines either the accident rate λ or the recovery rate μ of a worker and $\varphi(\rho)=\varphi(\rho|\theta_1,\ldots,\theta_r)$ denotes a parametric PVD with r parameters. A probability distribution $\pi^\rho(\underline{\theta})=\pi^\rho(\theta_1,\ldots,\theta_r)$ over the parameters of the model can be used to describe the uncertainty over PVD. Then, the estimated population variability density $\hat{p}(\rho)$ is taken as

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

Therefore, the estimated $\hat{p}(\rho)$ consists of a weighted mix of distributions of the chosen model. In BPVA, the assessment of the PVD of ρ may be based on two different types of information:

- E_0^ρ: prior state of knowledge on ρ, which may be drawn from "top-down", "bottom-up" data or experts opinions;
- E_1^ρ: "local" data - exposure data from accident recordings of workers in a workplace.

Type E_0^ρ is the prior evidence that provides information about $\pi^\rho(\underline{\theta})$. Then,$\pi_0^\rho(\underline{\theta})=\pi(\underline{\theta}|E_0)$ is the prior probability distribution over the parameters $\underline{\theta}$ and $\hat{p}_0(\rho)=\hat{p}(\rho|E_0)=\int_{\underline{\theta}}\varphi(\rho|\underline{\theta})\pi_0^\rho(\underline{\theta})d\underline{\theta}$ is the prior PVD of ρ. The evidence E_1^ρ in turn includes the available data obtained from accident datasets. Therefore, the distribution of the population variability parameters based on types E_0^ρ and E_1^ρ information is developed by applying the Bayes' theorem:

$$\pi_1^\rho(\underline{\theta})=\pi(\underline{\theta}|E_0^\rho,E_1^\rho)=\frac{P(E_1^\rho|\underline{\theta},E_0^\rho)\pi_0^\rho(\underline{\theta})}{\int_{\underline{\theta}}P(E_1^\rho|\underline{\theta},E_0^\rho)\pi_0^\rho(\underline{\theta})d\underline{\theta}}, \tag{1}$$

where $P(E_1^\rho|\underline{\theta},E_0^\rho)$ is the likelihood of the evidence E_1^ρ. Then, the PVD over ρ, conditional on E_0^ρ and E_1^ρ, or the posterior PVD, is given by $\hat{p}_1(\rho)=\hat{p}(\rho|E_0^\rho,E_1^\rho)=\int_{\underline{\theta}}\varphi(\rho|\underline{\theta})\pi_1^\rho(\underline{\theta})d\underline{\theta}$.

For a workplace with m workers, the evidence E_1^ρ is the set of the exposure data of each worker ($E_{1i}^\rho, i=1,\ldots,m$). Assuming that the accidents suffered by each worker are independent, the likelihood function of the evidence E_1^ρ becomes

$$P(E_1^\rho|\underline{\theta},E_0^\rho)=\prod_{i=1}^{m}P(E_{1i}^\rho|\underline{\theta},E_0^\rho), \tag{2}$$

where $P(E_{1i}^\rho|\underline{\theta},E_0^\rho)$ is the probability of observing evidence E_{1i}^ρ for the i-th employee.

Note that the likelihood for the i-th worker can be obtained as a function of the accident measure ρ_i, which is one of the values of the random variable ρ that is in turn distributed according to $\varphi(\rho|\underline{\theta})$. Therefore, we calculate the probability of observing the evidence E_{1i} by allowing the accident measure to assume all possible values, i.e., by averaging $P(E_{1i}^\rho|\underline{\theta},E_0^\rho,\rho_i)$ over the distribution of ρ:

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

which can be replaced into Eq. (2) to obtain the likelihood function by using the whole available information.

3. BPVA and Markov-based integrated model for estimating occupational measures

3.1. Exposure data: The worker's timeline

The exposure data (type E_1^ρ evidence) is used to formulate the likelihood function of the Bayesian model described in the previous section. The likelihood model, which is proposed in Section 3.2, uses run-time data in the form of the number of events, which occurred during the exposure time. We assume all occupational accidents are fairly reported and recorded and the forms used to register them contain the accident date and the duration of the recovery time from that accident. In addition, data on the dates of admission and dismissal of each worker are required.

Indeed, the run-time data may be constructed from the timelines of the population of workers as described in Fig. 1, which shows the timelines for a workplace with $m=12$ workers over a period starting at T_S and ending at T_F, where $T_{Si}\geq T_S$ and $T_{Fi}\leq T_F$ are, respectively, the starting and end times for the i-th worker, for $i=1,\ldots,m$. Over the interval [T_{Si};T_{Fi}], the i-th worker may be involved in K_i accidents. The elapsed working time $t_{i,j}$ corresponds to the time interval between the day when the i-th worker is recovered after the (j-1)-th accident and the day of the j-th accident ($j=1,2,\ldots,K_i$); the respective recovery time is denoted by $r_{i,j}$. The time range between the recovery from the last accident, K_i, and T_{Fi} is denoted by s_i over which no accident takes place.

In Fig. 1, the starting points T_{S1}, T_{S2}, T_{S3}, T_{S6}, T_{S7}, T_{S9}, T_{S10} and T_{S11} coincide with T_S, which means these workers were hired before or at T_S, while the end times T_{F1}, T_{F2}, T_{F3}, T_{F4}, T_{F5}, T_{F7}, T_{F10} and T_{F11} overlap T_F, i.e., the workers 1, 2, 3, 4, 5, 7, 10 and 11 remained working at least by T_F. Workers 1, 2, 3, 7, 10 and 11 were observed from T_S to T_F, while the others worked for a fraction of this time. Workers 3, 5, 8 and 11 had no accidents; then, $s_3=T_{F3}-T_{S3}$, $s_5=T_{F5}-T_{S5}$, $s_8=T_{F8}-T_{S8}$ and $s_{11}=T_{F11}-T_{S11}$. The contract of worker 9 expired when he returned from the last accident, thus $s_9=0$; this situation may not be allowed in accordance with some national labor regulations that prevent workers to be fired during a specified period after their recovery from an accident.

Finally, the end time (T_F) occurs before worker 10 is back to work. Thus, the recovery time $r_{10,1}$ is said to be censored and $s_{10}=0$. Note that $T_i=\left(\sum_{j=1}^{K_i}t_{i,j}\right)+s_i=T_{Fi}-T_{Si}-R_i$ (where $R_i=\sum_{j=1}^{K_i}r_{i,j}$) represents the total time for which the i-th worker was submitted to the risks of occupational accidents, i.e., T_i is the total exposure time of the i-th worker. Thus, the run-time data of the i-th worker are the pairs $\{K_i,T_i\}$ and $\{K_{Ri},R_i\}$, which are quantities necessary for the analysis of the rates of accident and recovery respectively. Note that K_{Ri}, which is the number of recoveries of the i-th worker, may be equal to either K_i or to K_i-1 (when the last recovery time $r_{i,j}$ is censored, as in the case of worker 10). Table 1 illustrates the exposure data extracted from Fig. 1, where E_1^λ and E_1^μ correspond to type E_1^ρ evidence for the analysis of the rates of accident $\rho=\lambda$ and recovery $\rho=\mu$ respectively.

Furthermore, for some type of analysis such as work time loss, only accident with sick leave is relevant. In this case, the accidents without time loss are not considered in the timeline construction phase. For example, Table 2 displays the exposure data for workers 6 and 7 when we only consider the accidents with leave of absence. In Table 2, $K_6=0$ and $K_7=1$, while $K_6=1$ and $K_7=3$ in Table 1, given that worker 6 did not have accidents with time loss and the worker 7 suffered two accidents without time loss.

This section has provided an overview of the type E_1 evidence necessary for the likelihood function E_1^ρ construction in the BPVA. Despite its complexity, the model requires simple data likely to be available in the company database. Relevant data to use the models include only the number of workdays, the workdays lost due to recovering from occupational accidents, and the number of occupational accidents over the period of observation for each worker. The following section develops the BPVA for occupational accidents based on these data.

3.2. BPVA

3.2.1. The likelihood function

In order to perform a population variability analysis for an accident measure ρ of interest we need to specify an appropriate PVD $\varphi(\rho|\underline{\theta})$ to define the underlying variability ofρ. The specification of $\varphi(\rho|\underline{\theta})$ may be guided by the nature of the accident

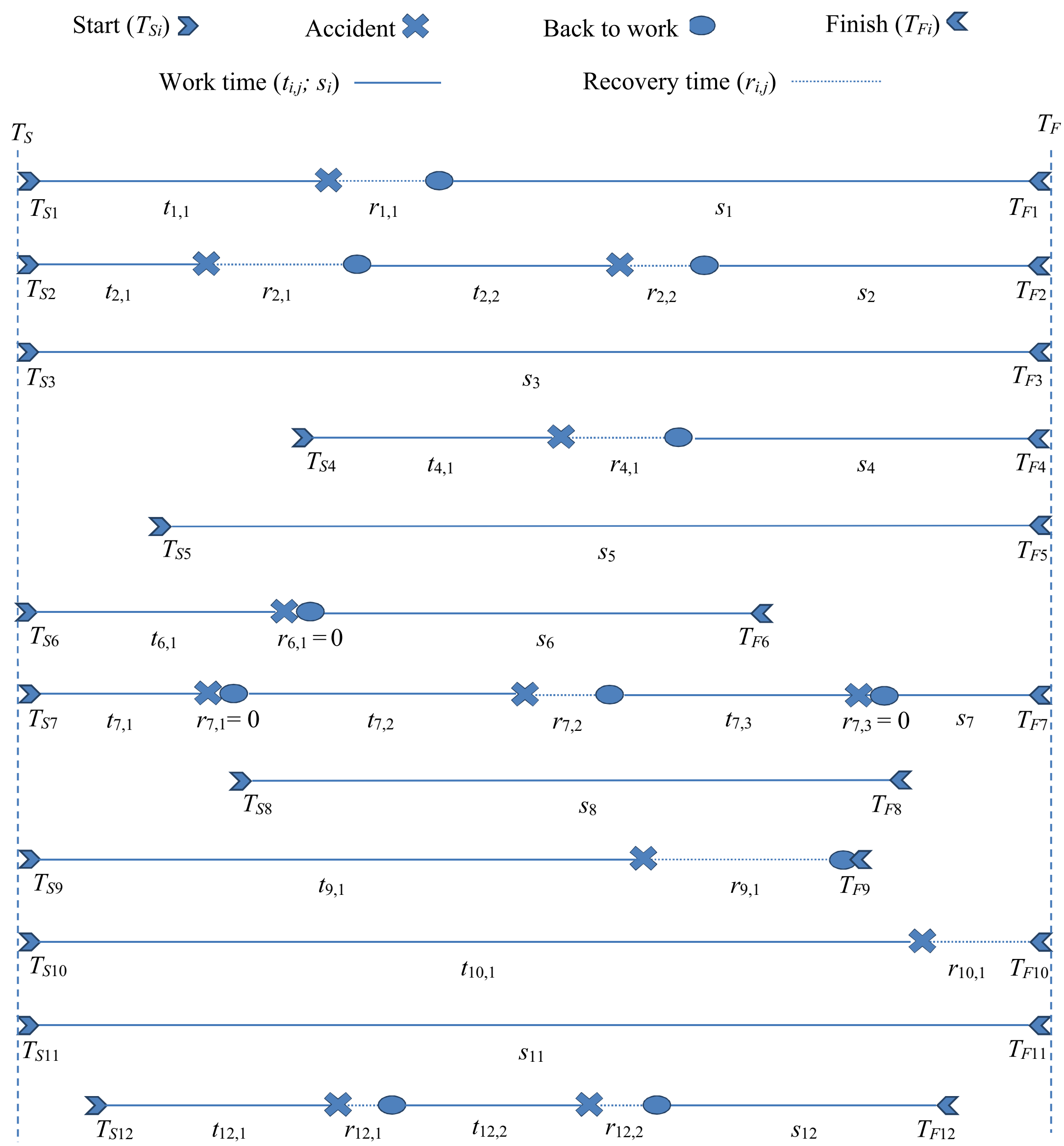

Fig. 1. Timelines for a workplace with 12 workers. Adapted from Marcoulaki et al. [27].

Table 1
Exposure data (E_1 evidence).

Worker(i)	E_1^{λ}		E_1^{μ}	
	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	$r_{12,1}+r_{12,2}$

measure. Due to conceptual similarities in the analyzed quantities, we assume that the population variability of the unknown accident measure ρ is given by a Lognormal distribution as in Droguett et al. [13]. Then,

Table 2
Run-time data of the workers 6 and 7 for accidents with time loss.

Worker (i)	E_1^{λ}		E_1^{μ}	
	K_i	T_i	K_{Ri}	R_i
6	0	$t_{6,1}+s_6$	–	–
7	1	$t_{7,1}+t_{7,2}+t_{7,3}+s_7$	1	$r_{7,2}$

$$\varphi(\rho|\underline{\theta}) = \varphi(\rho|\nu,\tau) = \frac{1}{\sqrt{2\pi}\rho\tau} e^{-\frac{1}{2}\left(\frac{\ln \rho - \nu}{\tau}\right)^2}, \quad (3)$$

where ν and τ are the mean and standard deviation of the natural logarithm of ρ.

The likelihood is a data-driven function that depends on the type of the available evidence. Let us assume that the available sources of evidence (type E_1^{ρ} evidence) are run-time data $\{(q_i, w_i), i = 1, \ldots m\}$, where $q_i = K_i$ and $w_i = T_i$ if the analysis is for the rate of accident $\rho = \lambda$ or $q_i = K_{Ri}$ and $w_i = R_i$ if the analysis is for the rate of recovery $\rho=\mu$; these quantities are obtained as discussed in the previous section. If we know the accident measure ρ_i of each worker $i=1,\ldots,m$, we can use the Poisson distribution to estimate the likelihood of observing q_i events over time w_i [13]:

$$P(q_i, w_i \mid \rho_i, E_0) = \frac{(\rho_i.w_i)^{q_i}.e^{-\rho_i.w_i}}{\Gamma(q_i+1)} \tag{4}$$

As we only know that ρ_i is one of the possible values of variable ρ which is distributed according to $\varphi(\rho \mid \nu, \tau)$, we average the likelihood given by Eq. (4) over all possible values of ρ in order to calculate the probability of the data unconditional on the unknown value of ρ:

$$P(q_i, w_i \mid \nu, \tau, E_0) = \int_\rho P(q_i, w_i \mid \rho, E_0) \cdot \varphi(\rho \mid \nu, \tau) \cdot d\rho, \tag{5}$$

Replacing (3) and (4) in to Eq. (5), we have:

$$P(E_1^\rho \mid \nu, \tau, \ E_0^\rho) = \int_\rho \frac{(\rho.w_i)^{q_i}.e^{-\rho.w_i}}{\Gamma(q_i+1)} . \frac{1}{\sqrt{2\pi}\rho\tau} e^{-\frac{1}{2}\left(\frac{\ln \rho - \nu}{\tau}\right)^2} d\rho . \tag{6}$$

transformation given as follows [13]:

$$\pi_0^\rho(\nu, \tau) = f_{LN}(\rho_{50}) . f_{LN}(EF_\rho) . \left| \frac{\partial(\rho_{50}, EF_\rho)}{\partial(\nu, \tau)} \right|$$
$$= f_{LN}(\rho_{50} \mid \alpha, \delta) . f_{LN}(EF_\rho \mid \beta, \varepsilon) . 1.645 e^\nu . e^{1.645\tau} . \tag{9}$$

where α and β are the median estimates of ρ_{50} and EF_ρ respectively, and δ and ε are the error factor estimates of ρ_{50} and EF_ρ respectively. Other prior specifications are discussed in Kaplan [23] and Pörn [35].

3.2.3. A posterior distribution specification and the variability measures

The likelihood function and prior distribution have been incorporated in a Bayesian inference procedure in which the posterior density $\pi_1^\rho(\underline{\theta}E)$ is computed for either $\rho=\lambda$ or $\rho=\mu$. If the likelihood function and prior distribution are given as in (7) and (9), then the posterior distribution of the population variability parameters in Eq. (1) can be rewritten as Eq. (10).

$$\pi_1^\rho(\nu, \tau \mid E_0^\rho, E_1^\rho) = \frac{\left[\prod_{i=1}^m \int_\rho \frac{(\rho.w_i)^{q_i}.e^{-\rho.w_i}}{\Gamma(q_i+1)} . \frac{1}{\sqrt{2\pi}\rho\tau} e^{-\frac{1}{2}\left(\frac{\ln \rho - \nu}{\tau}\right)^2} d\rho \right] \cdot f_{LN}(\rho_{50}) . f_{LN}(EF_\rho) . 1.645 e^\nu . e^{1.645\tau}}{\int_\nu \int_\tau \left[\prod_{i=1}^m \int_\rho \frac{(\rho.w_i)^{q_i}.e^{-\rho.w_i}}{\Gamma(q_i+1)} . \frac{1}{\sqrt{2\pi}\rho\tau} e^{-\frac{1}{2}\left(\frac{\ln \rho - \nu}{\tau}\right)^2} d\rho \right] \cdot f_{LN}(\rho_{50}) . f_{LN}(EF_\rho) . 1.645 e^\nu . e^{1.645\tau} . d\nu . d\tau} \tag{10}$$

where $\rho=\lambda$ or $\rho=\mu$. Then, the total likelihood is obtained by replacing (6) into Eq. (2).

$$P(E_1^\rho \mid \nu, \tau, E_0^\rho) = \prod_{i=1}^m \int_\rho \frac{(\rho.w_i)^{q_i}.e^{-\rho.w_i}}{\Gamma(q_i+1)} . \frac{1}{\sqrt{2\pi}\rho\tau} e^{-\frac{1}{2}\left(\frac{\ln \rho - \nu}{\tau}\right)^2} d\rho . \tag{7}$$

3.2.2. The prior distribution

The proposed procedure involves the specification of an informed continuous prior distribution over the parameter space $\underline{\theta}$ of the variability model. This prior state of knowledge is the type E_0^ρ evidence and it may be estimated by either "top-down" or "bottom-up" data. Then, the analyst is required to provide central value estimates and the extent of variability for the population variability parameters ($\underline{\theta}$).

Following the multiplicative error model proposed in Droguett and Mosleh [15] and Droguett and Mosleh [16], if these estimates are specified in terms of a median (central value estimate) and the error factor (extent of variability), they take the form of Lognormal distributions. Indeed, if θ_{50}^k and EF_{θ^k} are the median and the error factor of the k-th parameter θ^k $(k=1,\ldots,r)$, then the probability density over θ^k can be represented by Eq. (8):

$$f_{LN}\left(\theta^k \middle| \theta_{50}^k, EF_{\theta^k}\right) = \frac{1}{\sqrt{2\pi}.\theta_{50}^k.\frac{\ln EF_{\theta^k}}{1.645}} e^{-\frac{1}{2}\left(\frac{\ln\theta^k - \ln\theta_{50}^k}{\ln EF_{\theta^k}/1.645}\right)^2}, \tag{8}$$

where $f_{LN}(.)$ is the lognormal density function.

As ν and τ are in a natural scale, Droguett et al. [13] suggested that the prior distributions were specified over the median $(\rho_{50}=e^\nu)$ and the error factor $(EF_\rho=e^{1.645\tau})$ of the variability measure ρ instead of over the mean (ν) and standard deviation (τ). Thus, by considering that the population variability parameters are independent, the prior density over the model's parameter space $(\nu, \ \tau)$ is then found by applying the standard density

Then, as Eq. (10) is not analytically obtainable, it is estimated by using a MCMC-based method [19]. Thus, a N-sample set $S=\{(\nu_1,\tau_1),\ldots,(\nu_N,\tau_N)\}$ representing the posterior density over the parameters of the variability distribution model $\varphi(\rho \mid \nu, \tau)$ is generated from Eq. (10) and the estimated population variability density is computed as $\hat{p}(\rho)=\frac{1}{N}\sum_{k=1}^N \varphi(\rho \mid \nu_k, \tau_k)$. The corresponding mean and variance are computed as $\hat{\eta}_\rho=\frac{1}{N}\sum_{k=1}^N \eta(\nu_k, \tau_k)$ and $\hat{\sigma}_\rho{}^2=\frac{1}{N}\sum_{k=1}^N \sigma^2(\nu_k,\tau_k)$, where $\eta(\nu,\tau)$ and $\sigma^2(\nu,\tau)$ are the mean and variance of $\varphi(\rho \mid \nu, \tau)$.

Furthermore, the generated results include uncertainty bounds of the cumulative variability distribution $\hat{P}(\rho)=\int_{x=0}^\rho \hat{p}(x)dx$ in the form of z-th percentiles $\hat{P}_z(\rho)$. The z-th percentile is determined by finding the value $\hat{P}_z(\rho)$ for which a fraction $z/100$ of the samples satisfies $\int_{x=0}^\rho \varphi(x \mid \nu_k, \tau_k)dx < \hat{P}_z(\rho)$. These bounds provide the analyst a basis to assess the uncertainty associated with the estimated PVD over ρ.

3.3. Markov-based model for estimation of occupational measures

Given the percentiles over the rate of accident $\rho=\lambda$ and the rate of recovery $\rho=\mu$, a two-state continuous Markov process as shown in Fig. 2 may be used to characterize the random behavior of an accident-recovery process of a representative worker. By representative worker, we mean an employee that describes the behavior of any worker at the workplace and whose rates of accident and recovery are drawn as an average of the rates of all individuals within the population. Thus, a metric (e.g., the expected number of accidents) for the whole population of workers may be obtained by multiplying the number m of workers by the estimate obtained for the representative worker.

In Fig 2, if the worker is in first state, he/she is performing his/her function normally, whereas if he/she is in the second state he/she is recovering from an accident (on sick leave). Then, this model will be adopted in order to estimate (i) the expected number of accidents and (ii) the unavailability and work time loss.

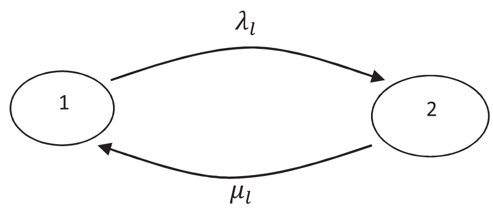

Fig. 2. Markov's diagram for the accident-recovery process of a representative worker.

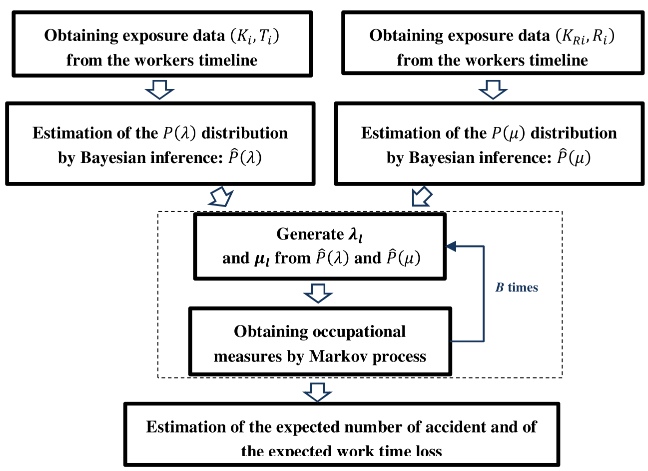

Fig. 3. Methodology: integrating BPVA and Markov models for estimating occupational measures.

To this end, consider the transitions between states 1 and 2 take place according to the rates of accident λ_l and recovery μ_l that are random variables whose cumulative variability distributions $\hat{P}(\lambda)$ and $\hat{P}(\mu)$ are estimated from the BPVA procedure described in the previous sections. The uncertainty over the rates of accident and recovery is here propagated by a Monte Carlo-based method through which it is possible to sample B values of λ_l and μ_l, $l=1,\dots,B$, and then estimate the measures of interest by solving B times the Markov process of Fig 2. Thus, we can compute the average, variance and confidence intervals over the indicators of interest.

Fig 3 illustrates the proposed methodology highlighting the integration between the BPVA and Markov models. The next three sections explain this procedure to estimate the expected number of accidents, the unavailability and work time loss.

3.3.1. Expected number of accidents

As represented in Fig 2, an occupational accident happens when a transition from state 1 to state 2 takes place. Therefore, the number of accidents of a representative worker is related to the frequency the process visits state 2. Then, the expected number of occupational accidents suffered over a period of time t is the expected number of visits to state 2 up to t, $H_l(t)$, which is given as $H_l(t)=\frac{\mu_l}{\lambda_l+\mu_l}\lambda_l t$.

Then, we generate a sample set containing B pairs (λ_l,μ_l), $l=1,\dots,B$ (B being large enough) from $\hat{P}(\lambda)$ and $\hat{P}(\mu)$ distributions, and the curve $\overline{H}(t)$ of the expected number of visits to state 2 up to t within the worker's population is estimated as $\overline{H}(t)=\frac{1}{B}\sum_{l=1}^{B} H_l(t)=\frac{t}{B}\sum_{l=1}^{B}\frac{\mu_l.\lambda_l}{\lambda_l+\mu_l}$. The expected number of occupational accidents occurred in a workplace with m workers, for which distributions of the rates of accident and recovery are estimated by $\hat{P}(\lambda)$ and $\hat{P}(\mu)$, over a period of time t is calculated as

$$E[N(t)]=m.\overline{H}(t)=\frac{m.t}{B}\sum_{l=1}^{B}\frac{\mu_l.\lambda_l}{\lambda_l+\mu_l} \tag{11}$$

3.3.2. Unavailability and work time loss

The unavailability $U(t)$ is the probability that the Markov process is in state 2 at time t. The steady-state value of $U(t)$ defined by $\overline{U}=\lim_{t\to\infty}U(t)$ represents the expected time fraction over which the worker will stay in state 2, i.e., the expected unavailable time. Thus, the U_l are formulated for B pairs (λ_l,μ_l), $l=1,\dots,B$ as $U_l=\frac{\lambda_l}{\lambda_l+\mu_l}$.

Similarly to what has been done for the expected number of accidents, we can define the average of the steady-state unavailability curve from the sample set of B pairs (λ_l,μ_l), $l=1,\dots,B$ as $\overline{U}=\frac{1}{B}\sum_{l=1}^{B}\frac{\lambda_l}{\lambda_l+\mu_l}$. Finally, the expected work time loss (or the total expected time in recovery) during a given time t is calculated as $E[L_l(t)]=t.\overline{U}_l$ [27]. Then, in a workplace with m workers, the total expected work time loss over $(0, t]$ can be estimated as

$$E[L(t)]=m.t.\overline{U}=\frac{m.t}{B}\sum_{l=1}^{B}\frac{\lambda_l}{\lambda_l+\mu_l} \tag{12}$$

3.3.3. The uncertainty over the expected estimates

The estimates of the measures of interest presented in the previous sections are point estimates in form of expected values of number of accidents ($E[N(t)]$) and work time loss ($E[L(t)]$). It is also interesting to obtain the uncertainty bounds in terms of z-th percentiles in order to provide a probability interval for the quantities$N(t)$, and $L(t)$, which may be achieved from the uncertainty bounds, $\hat{P}_z(\rho)$, defined in Section 3.2.3. Therefore, the z-th percentiles corresponding to the uncertainty about the expected number of accidents ($E[N_{z\%}(t)]$) and the expected work time loss ($E[L_{z\%}(t)]$)) may be obtained as follows:

I. Generate λ_l from $\hat{P}(\lambda)$ and μ_l from $\hat{P}(\mu)$;
II. Estimate the measures $H_l(t)$, and $L_l(t)$ as in previous sections;
III. Repeat the steps I, and II by a number of $l=1,\dots,B$ times (B being large enough). Then, we will obtain a set of B different estimates representing the distribution corresponding to the uncertainty about the estimated measures $H_l(t)$, and $L_l(t)$. Thus, the z-th percentiles of $E[N_{z\%}(t)]$ and $E[L_{z\%}(t)]$ are given from the set containing the B estimates $H_l(t)$, and $L_l(t)$ as well as the expected values $E[N(t)]$ and $E[L(t)]$ computed in previous sections.

In the next section, we illustrate and discuss the use of the BPVA-Markov model by means of one example, which applies the proposed model for a real case of a hydropower company in Brazil with the purpose of validating the model, forecasting and analyzing the workplace occupational measures.

4. Application example

4.1. Description of the problem

In this section, the BVPA-Markov procedure is applied to a real dataset obtained for a hydropower company. Run-time data were collected from 01/01/2005 to 12/31/2010 in order to construct the workers timeline (Fig 1) and the likelihood function (Eq. (6)). Then, the BVPA is applied in order to assess the population variability distributions of the rates of accidents and recovery of a population of transmission lines and electrical maintenance workers of a hydropower company. Note that our focus here is on the type of worker and not on the type of accident that this worker may suffer. In fact, all records of accidents available in the dataset for a given worker were considered. It is important to emphasize that we may also focus on different types of accidents for a same worker. In this situation, we would need to divide the dataset into categories in accordance with the type of accident (e.g. falling, slipping, stumbling, etc.).

The model is applied for evaluating the occupational measures of interest due to only the occurrence of accidents with time loss and, therefore, the run-time data were obtained as given in Table 2; then, this dataset is available in Table 3. For this population, 232 workers were analyzed (summing up 453,081 men-days of work) for which 50 accidents were recorded, involving 40 workers. From the 40 workers who had accidents, 4 suffered two accidents and 3 had three accidents. The others 33 workers had only one accident. Thus, no accident was observed for 192 out of 232 workers (see Table 3).

4.2. BVPA-Markov analysis

The prior distributions are chosen as in Eq. (9), i.e., they are formulated over the median and the error factor (PVD parameters). As "top-down", "bottom-up" and expert opinions are not here available to construct the prior distributions, the prior estimates were obtained from statistics extracted from the exposure data. We can consider the data usage in order to obtain the prior estimates under the hypothesis that the experts build their opinions based on their experience, which depends on the events that they observed, i.e., on the dataset itself.

Therefore, let a, b, c and d be the parameters of the prior distribution defined over the parameters of the accident rate PVD and e, f, g and h be the parameters of the prior distribution over the parameters of the recovery rate PVD. Indeed, consider a and e are the median estimates, b and f are the error factors of the medians, c and g are the value central estimates of the error factors, d and h are the error factors of the error factors. These quantities are here estimated as follows:

- The parameters a and e can be estimated as the 50-th percentile of the data set formed by the ratio between the number of events and the exposure time of each injured worker, i.e., $a = \text{median}\left(\frac{K_j}{T_j}\right)$ and $e = \text{median}\left(\frac{K_{Rj}}{R_j}\right)$, where $j \in \{i|\, workeriwasinjured\}$;
- According to Mosleh [29], the error factor can be written as the ratio between the 95-th percentile and the median of the distribution, thus the c and g parameters can be estimated as: $c = \frac{95 \text{ th quantile}\left(\frac{K_j}{T_j}\right)}{\text{median}\left(\frac{K_j}{T_j}\right)}$ and $g = \frac{95 \text{ th quantile}\left(\frac{K_{Rj}}{R_j}\right)}{\text{median}\left(\frac{K_{Rj}}{R_j}\right)}$, with j defined as in previous item;
- The extent of variability estimates can be interpreted as the uncertainty measure that characterizes the confidence of the analyst on the value central estimates (Section 3.2.2). Then, we chose b, f, d and h parameters were fixed in the value 5;

Then, the parameters of the prior distributions were estimated as $a = 0.00046$, $b = 5$, $c = 2.99$, $d = 5$, $e = 0.106$, $f = 5$, $g = 3.17$ and $h = 5$. After solving the BVPA model, the PVD of the rates of accident and recovery were estimated and shown in Fig 4.

Note that in Fig 4(a) (distribution for the accident rate) the uncertainty bounds found at the lowest values are wider than they are for the highest ones. This behavior can be explained by the fact that part of the distribution is highly influenced by those 192 workers for which no accident was observed, i.e., censored data. This behavior is not observed in the estimated recovery rate distribution (Fig 4(b)) because the censored data about the non-injured workers are not considered in the construction of the likelihood function of μ.

From these distributions, the Markov-based model was applied to infer the expected number of accidents with time loss (Eq. (11)) and the expected work time loss (Eq. (12)). Table 4 shows the expected number of accidents $E[N(t)]$ and the work time loss $E[N(t)]$ for $t = 365$ days, and the respective uncertainty bounds (5-th and 95-th percentiles).

Note that the censored data (no accident observed for 192 workers) implies a strong impact on the uncertainty analysis. However, this is (should be) the context possibly encountered in accident analysis since many types of accidents are rather rare events. Moreover, note that the run-time dataset gathered in Table 3 corresponds to one of the various possible samples that could be observed. Thus, the estimates given in Fig 4 and Table 4 could be different from the true measures since there still exists uncertainty over the data collected.

4.3. Sensitivity analysis

It is interesting to assess the sensitivity of the model for different prior estimates, which in this case were previously obtained from the dataset itself. Thus, we allowed the value central estimates (a, c, e and g) to vary, resulting in 8 new estimated PVDs for the accident rate and 8 new for recovery rate. Figs 5 and 6 show these estimated distributions (and the corresponding uncertainty bounds) for different combinations of a, c, e and g.

From the shape and scale of the distributions, we can verify that no significant variation is observed for the estimated distributions. The highest average of the expected PVDs of the accident rate was 0.00021 while the smallest was 0.00015. For the recovery rate, these values are 0.102 and 0.0912. Furthermore, the Markov-based model was applied using all combinations of distributions of the accident and recovery rates and the highest and the smallest expected number of accidents with time loss were 10.89 and 9.99, and the highest and the smallest expected work time loss were 264.19 and 220.57. These results show the proposed model is not too sensitive to the prior estimates. This is a rather important result especially in this case, for which "top-down", "bottom-up" and experts opinions are not available to feed the prior model.

4.4. The impact of (not) considering inhomogeneity

The BVPA model applied in this real case provides estimations for PVDs of the rates of accident and recovery. To evaluate the application of the proposed model we also predicted the measures $E[N(t)]$ and $E[L(t)]$ by using values of $\hat{\lambda}$ and $\hat{\mu}$ given by Maximum Likelihood Estimators (MLE). In this case, we assume that all company workers would have homogeneous behavior in terms of the occurring (and recovering from) accidents.

Then, the estimators of λ and μ are given as $\hat{\lambda} = \sum_{i=1}^{m} K_i / \sum_{i=1}^{m} T_i$ and $\hat{\mu} = \sum_{i=1}^{m} K_{Ri} / \sum_{i=1}^{m} R_i$, which are the MLE of the parameters of an exponential distribution. The expected number of accidents $E[N(t)]$ with time loss and the expected work time loss $E[L(t)]$ can be estimated directly from Markov model equations by $E[N(t)] = m.t \frac{\hat{\mu}.\hat{\lambda}}{\hat{\lambda}+\hat{\mu}}$ and $E[L(t)] = m.t \frac{\hat{\lambda}}{\hat{\lambda}+\hat{\mu}}$.

The uncertainty bounds of λ and μ can be calculated as $\lambda_{z\%} = \hat{\lambda} + G(z\%) \times \frac{\hat{\lambda}}{\sqrt{\sum K_i}}$ and $\mu_{z\%} = \hat{\mu} + G(z\%) \times \frac{\hat{\mu}}{\sqrt{\sum K_{Ri}}}$, where $G(z\%)$ is the z-th percentile of a variable that follows a Gaussian distribution with mean and standard deviation equal to 0 and 1,

Table 3
Run-time data used in the real case.

Worker(i)	K_i	T_i	K_{Ri}	R_i	Worker(i)	K_i	T_i	K_{Ri}	R_i	Worker(i)	K_i	T_i	K_{Ri}	R_i
1	0	2191	–	–	79	0	2191	–	–	157	0	2191	–	–
2	0	2191	–	–	80	0	2191	–	–	158	0	2191	–	–
3	0	2191	–	–	81	0	2191	–	–	159	1	2105	1	–
4	0	2191	–	–	82	1	1986	1	205	160	0	2191	–	–
5	0	2191	–	–	83	0	2191	–	–	161	3	2163	3	–
6	0	2191	–	–	84	0	2191	–	–	162	1	2158	1	–
7	0	2191	–	–	85	0	2191	–	–	163	0	2191	–	–
8	0	2191	–	–	86	0	2191	–	–	164	0	2191	–	–
9	0	57	–	–	87	0	2191	–	–	165	0	2191	–	–
10	0	2191	–	–	88	0	2191	–	–	166	1	2189	1	–
11	0	2191	–	–	89	0	2191	–	–	167	2	2172	2	–
12	0	2191	–	–	90	0	2191	–	–	168	0	2191	–	–
13	1	2176	1	15	91	0	2191	–	–	169	0	2191	–	–
14	1	2184	1	7	92	0	2191	–	–	170	2	2186	2	–
15	0	2191	–	–	93	0	2191	–	–	171	0	2191	–	–
16	0	2191	–	–	94	0	2191	–	–	172	0	2191	–	–
17	0	2191	–	–	95	0	2191	–	–	173	0	2191	–	–
18	0	2191	–	–	96	0	2191	–	–	174	0	2191	–	-
19	0	2191	–	–	97	0	2191	–	–	175	1	2188	1	3
20	0	2191	–	–	98	0	2191	–	–	176	0	2191	–	–
21	1	2177	1	14	99	0	2191	–	–	177	1	2175	1	16
22	0	2191	–	–	100	0	2191	–	–	178	0	2191	–	–
23	0	2191	–	–	101	0	2191	–	–	179	0	2191	–	–
24	0	2191	–	–	102	0	2191	–	–	180	0	2191	–	–
25	0	2191	–	–	103	0	2191	–	–	181	0	2191	–	–
26	0	2191	–	–	104	0	2191	–	–	182	0	2191	–	–
27	0	2191	–	–	105	0	2191	–	–	183	0	2191	–	–
28	0	2191	–	–	106	0	2191	–	-	184	0	2191	-	–
29	0	2191	–	–	107	1	2114	1	77	185	1	2186	1	–
30	0	2191	–	–	108	0	2191	–	–	186	0	2191	–	–
31	0	2191	–	–	109	0	2191	–	–	187	0	2191	–	–
32	0	2191	–	–	110	3	2147	3	44	188	0	2191	–	–
33	0	2191	–	–	111	1	2178	1	13	189	1	2183	1	–
34	0	2191	–	–	112	0	2191	–	–	190	0	2191	–	–
35	1	2184	1	7	113	0	2191	–	–	191	3	2174	3	17
36	0	2191	–	–	114	1	2166	1	25	192	2	2180	2	11
37	0	2191	–	–	115	0	2191	–	–	193	1	2188	1	3
38	1	2186	1	5	116	0	2191	–	–	194	0	2191	–	–
39	0	2191	-	–	117	0	2191	–	–	195	1	2014	1	12
40	1	2011	1	180	118	0	2191	–	–	196	0	1904	–	–
41	0	2191	–	–	119	0	2191	–	–	197	1	1746	1	15
42	0	2191	–	–	120	1	2176	1	15	198	0	1642	–	–
43	0	2191	–	–	121	0	2191	–	-	199	1	1633	1	9
44	0	2191	–	–	122	2	2171	2	20	200	1	1627	1	15
45	0	2191	–	–	123	0	2191	–	–	201	1	930	1	3
46	0	2191	–	–	124	0	2191	–	–	202	0	933	–	–
47	0	2191	–	–	125	0	2191	–	–	203	1	926	1	7
48	0	2191	–	–	126	1	2183	1	8	204	1	927	1	6
49	0	2191	–	–	127	0	2191	–	–	205	0	933	–	–
50	0	2191	–	–	128	0	2191	–	–	206	0	933	–	–
51	0	2191	–	–	129	0	2191	–	–	207	0	933	–	–
52	0	2191	–	–	130	0	2191	–	–	208	0	933	–	–
53	0	2191	–	–	131	0	2191	–	–	209	0	933	–	–
54	0	2191	–	–	132	0	2191	–	–	210	0	933	–	–
55	0	2191	–	–	133	0	2191	–	–	211	0	933	–	–
56	0	2191	–	–	134	1	2188	1	3	212	0	933	–	–
57	0	2191	–	–	135	0	2191	–	–	213	0	625	–	–
58	0	2191	–	–	136	0	2191	–	–	214	0	625	–	–
59	0	2191	–	–	137	0	2191	–	–	215	0	625	–	–
60	0	2191	–	–	138	0	2191	–	–	216	0	625	–	–
61	0	2191	–	–	139	0	2191	–	–	217	0	625	–	–
62	0	2191	–	–	140	0	2191	–	–	218	0	625	–	–
63	0	2191	–	–	141	0	2191	–	–	219	0	625	–	–
64	1	2171	1	20	142	0	2191	–	–	220	0	625	–	–
65	0	2191	–	–	143	0	2191	–	–	221	0	625	–	–
66	0	2191	–	–	144	0	2191	–	–	222	0	625	–	–
67	0	2191	–	–	145	0	2191	–	–	223	0	625	–	–
68	0	2191	–	–	146	0	2191	–	–	224	0	625	–	–
69	0	2191	–	–	147	0	2191	–	–	225	0	113	–	–
70	1	2176	1	15	148	0	2191	–	–	226	0	113	–	–
71	0	2191	–	–	149	1	2184	1	7	227	0	113	–	–
72	0	2191	–	–	150	0	2191	–	–	228	0	113	–	–
73	0	57	–	–	151	0	2191	–	–	229	0	113	–	–
74	0	2191	–	–	152	1	2176	1	15	230	0	113	–	–
75	0	2191	–	–	153	1	2182	1	9	231	0	113	–	–

Table 3 (*continued*)

Worker(i)	K_i	T_i	K_{Ri}	R_i	Worker(i)	K_i	T_i	K_{Ri}	R_i	Worker(i)	K_i	T_i	K_{Ri}	R_i
76	0	2191	–	–	154	0	2191	–	–	232	0	2191	–	–
77	0	2191	–	–	155	0	2191	–	–					
78	0	2191	–	–	156	0	2191	–	–					

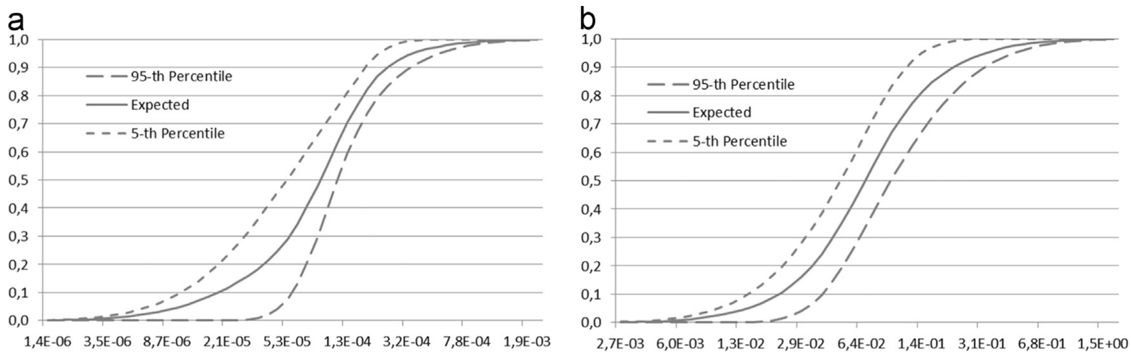

Fig. 4. Population variability cumulative distribution functions of the rates of accident (a) and recovery (b).

Table 4
Number of accidents and work time loss and estimates from proposed model for $t = 365$ days obtained by BVPA-Markov model.

Estimate	$E[N(t)]$	$E[L(t)]$
Expected	10.51	238.38
5-th percentile	6.22	82.05
95-th percentile	14.81	477.44

respectively. The uncertainty bounds of $E[N(t)]$ and $E[L(t)]$ ($E[N(t)]_{z\%}$ and $E[L(t)]_{z\%}$) can be obtained by the following procedure:

i. Generate u_1 e u_2 as random variables uniformly distributed between 0 and 100;
ii. Compute $E[N(t)]$ and $E[L(t)]$ using $\lambda_{u_1\%}$ and $\mu_{u_2\%}$ rather than $\hat{\lambda}$ and $\hat{\mu}$;
iii. Repeat the previous steps D times (D being large enough) and
iv. Make $E[N(t)]_{z\%}$ and $E[L(t)]_{z\%}$ equal to the z-th quantile of the set containing D different estimates to $E[N(t)]$ and $E[L(t)]$.

By doing that, the accident and recovery rates were estimated as $\hat{\lambda} = 0.00011$ and $\hat{\mu} = 0.0497$. The occupational measures are presented in Table 5 considering a time interval of 365 days.

First, note that, as pointed out by Fragola [17], as inhomogeneous data were aggregated as if they were homogeneous in Table 5, the confidence intervals obtained by the MLE-Markov model are narrower than those estimated by BVPA-Markov approach. However, even though in Table 5 the rates of accident and recovery are better estimated in accordance with the statistical sense, those occupational measures are less descriptive of each subpopulation (worker) that composes the mixture than those of Table 4 since inhomogeneity is not considered in this section.

Yet, the real data of the 2011 year was collected in order to compare them with predictions obtained from proposed model as well as the MLE-based model. Table 6 summarizes the accident data in that year. The proposed model presented a better result than the MLE-based model once the measures estimated from the proposed model are closer to the real data than ones estimated from the MLE-based model. Furthermore, the probability intervals obtained from the proposed model contain the real values, where as the confidence interval for the work time loss estimated from the MLE-based model does not contain the real observed value.

Also, all predicted occupational measures from MLE-based model are smaller than the ones obtained from proposed model. These results show that the MLE-based model tends to predict fewer values for the occupational measures. This behavior may be explained by the event 0 data, i.e., those workers for which no accidents were observed. The MLE model underestimates the accident rate while the proposed model is able to cope with the impact of the censored data. Thus, these results provide evidence in favor of the proposed Bayesian model as an efficient approach to represent the actual workplace.

5. Concluding remarks

The proposed model discussed in this paper is based on the Bayesian Population Variability Analysis method, which allows evaluating the population variability distributions of the rates of accidents and recovery from run-time data of workers submitted to the same occupational risks. The population variability analysis approach permits uncertainty assessment on these quantities of interest, which in turn feed a two-state Markov-based model adopted to estimate occupational measures such as the expected number of accidents and the expected work time loss distributions.

The model here developed can be informed using available databases of occupational accidents documented in the industry. Data requirement for using the proposed model are very mild from a practical standpoint, and include only the number of workdays, the workdays lost due to recovering from occupational accidents, and the number of occupational accidents over the period of observation for each worker. In addition, the model requires prior estimates about the rates of accident and recovery in form of central value and extent of variability, which can be obtained from "top-down" or "bottom-up" data. Moreover, the sensitivity analysis of the Section 4 shown that reasonable prior estimates can be obtained from the exposure dataset itself.

An example has shown that the numerical solution of the models is feasible and provides a good estimation for rates of accident and recovery. The results show that by considering the uncertainty bounds around the estimated distribution, it is in fact possible to distinguish between regions of the distribution where

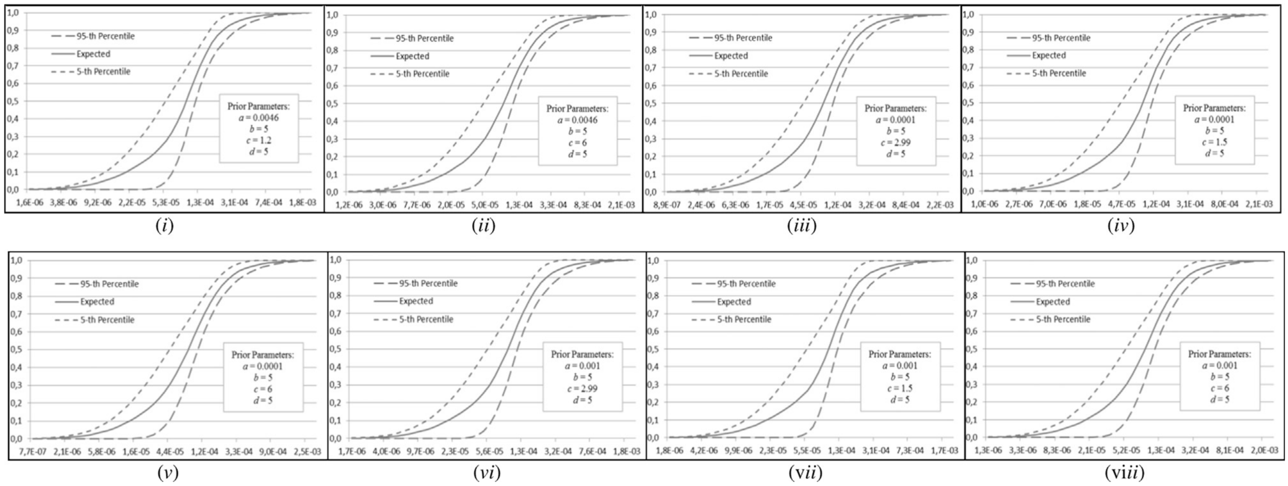

Fig. 5. Population variability distributions of the accident rate for different prior parameters.

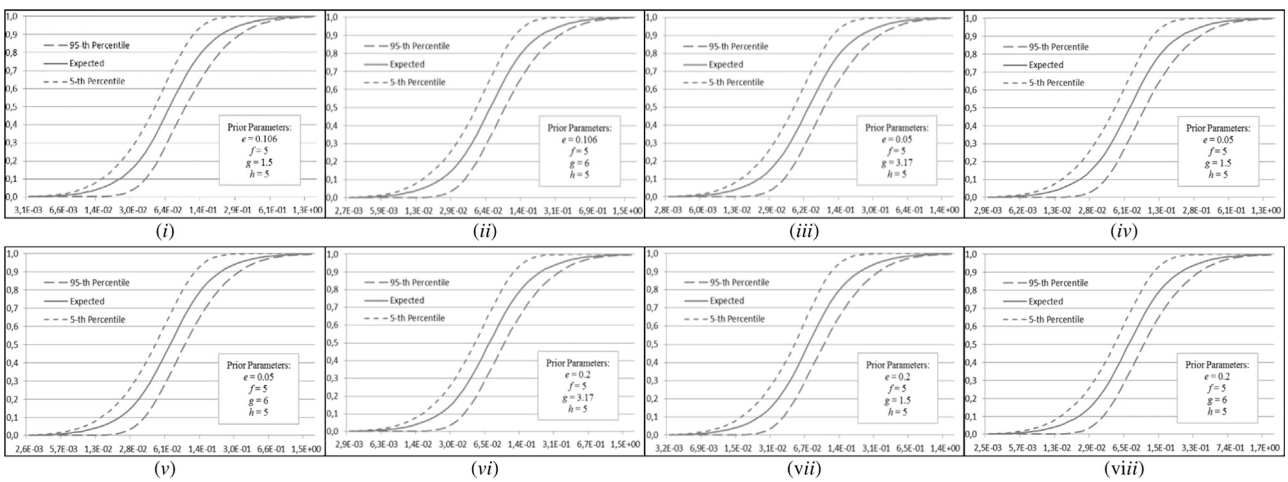

Fig. 6. Population variability distributions of the recovery rate for different prior parameters.

Table 5
Number of accidents and work time loss and estimates from MLE and Markov based model, not considering inhomogeneity.

Estimate	$E[N(t)]$	$E[L(t)]$
Expected	9.34	188.21
5-th percentile	7.17	130.94
95-th percentile	11.5	268.40

Table 6
Summary of accident data collected in 2011.

Data	Value
Total of workers	232
Accidents with time loss	11
Total work time loss	269
Total of workers for which no accident with time loss was observed	221

we believe to have a good and a poor estimate. In addition, the proposed model was compared to Maximum Likelihood Estimators. The results showed that the proposed model presented a representation of the real population whereas the maximum likelihood estimators underestimate the occupational accidents based on the estimated accident and recovery rates due to the existence of censored data. The proposed model, therefore, provides a better analysis to the accident and recovery rates and, consequently, to the estimates of frequency of accidents and of work time loss. Moreover, it is important to highlight that the approach was thought to be applied on in-company data (where the Bayesian approach is more needed) and that it is possible to apply it on regular basis, for example, to support decision making of national policies.

It is important here to point out also some limitations of the proposed model. First, the model did not assume possible trends on the rates of accidents and recovery over time, i.e., it is considered that the population variability distribution is the same during the observation period and over the future period used for prediction. Secondly, even though the proposed model only requires, as input data, the number of workdays, the workday's loss, and the number of occupational accidents over the period of observation for each worker, estimating the prior population variability parameters involves a significant degree of knowledge on probability distributions and on the variability of the rates of accidents among workers. However, in the absence of this evidence, the methodology also allows to estimate this information from the available exposure data. Finally, the application example developed in the paper did not distinguish the data into different types of accidents. Thus, the required information was obtained without discriminating the different modes of accidents (e.g., falling, slipping, stumbling, etc.). This means that the obtained results are related to the total accident frequency and number of day's loss due to any type of accidents suffered by transmission lines and electrical maintenance workers of the hydropower company considered in this paper. Nevertheless, the model is also able to analyze different accident risks separately, supporting the prioritization of preventive actions. To this end, it would be necessary to collect the required information categorized for each type of accidents.

To conclude, tailoring of this model to also consider accidents without time loss and the possibility of changing the population variability distribution over time are topics of ongoing research. While the former issue is important as a precursor to prevent the occurrence of more severe accidents, the latter may be done by proposing a model for population variability analysis of non-homogeneous rates.

Acknowledgments

The authors thank the Brazilian research-funding agency CNPq and CHESF for the financial support.

References

[1] Ale BJM, Baksteen H, Bellamy LJ, Bloemhof A, Goossens L, Hale A, Mud ML, Oh JIH, Papazoglou IA, Post J. Quantifying occupational risk: the development of an occupational risk model. Saf Sci 2008;46(2):176–85.
[2] Ale BJM, Bellamy LJ, Baksteen H, Damen M, Goossens LHJ, Hale AR, Mud M, Oh J, Papazoglou IA, Whiston JY. Accidents in the construction industry in the Netherlands: an analysis of accident reports using storybuilder. Reliab Eng Syst Saf 2008;93(10):1523–33.
[3] Ale BJM, Bellamy LJ, Papazoglou IA, Hale A, Goossens L, Post J, Baksteen H, Mud ML, Oh JIH, Bloemhoff A, Whiston JY. ORM: development of an integrated method to assess occupational risk. In: Proceedings of the international conference on probabilistic safety assessment and management, New Orleans; May 14–18, 2006.
[4] Aneziris ON, Papazoglou IA, Baksteen H, Mud M, Ale BJ, Bellamy LJ, Hale AR, Bloemhoff A, Post J, Oh J. Quantified risk assessment for fall from height. Saf Sci 2008;46(2):198–220.
[5] Aneziris ON, Papazoglou IA, Mud ML, Damen M, Kuiper J, Baksteen H, Ale BJ, Bellamy LJ, Hale AR, Bloemhoff A, Post JG, Oh J. Towards risk assessment for crane activities. Saf Sci 2008;46(6):872–84.
[6] Blanch A, Torrelles B, Salinas JA. Age and lost working days as a result of an occupational accident: a study in a shift work rotation system. Saf Sci 2009;47 (10):1359–63.
[7] Camino MA, Ritzel DO, Fontaneda I, et al. Construction industry accidents in Spain. J Saf Res 2008;39:497–507.
[8] Carnero MC, Pedregal DJ. Modelling and forecasting occupational accidents of different severity levels in Spain. Reliab Eng Syst Saf 2010;95:1134–41.
[9] Cawley JC, Homce GT. Occupational electrical injuries in the United States, 1992–1998, and recommendations for safety research. J Saf Res 2003;34:241–8.
[10] Chau N, Mur JM, Benamghar L, Siegfried C, Dangelzer JL, Français M, Jacquin R, Sourdot A. Relationship between some individual characteristics and occupational accidents in the construction industry: a case-control study on 880 victims of accidents occurred during a two-year period. J Occup Health 2002;44:131–9.
[11] Chia C, Changa T, Tingb H. Accident patterns and prevention measures for fatal occupational falls in the construction industry. Appl Ergon 2005;36:391–400.
[12] Deely JJ, Lindley DV. Bayes empirical Bayes. J Am Stat Assoc 1981;76(376):833–41.
[13] Droguett E, Groen F, Mosleh A. The combined use of data and expert estimates in population variability analysis. Reliab Eng Syst Saf 2004;83:311–21.
[14] Droguett E, Mosleh A. Bayesian methodology for model uncertainty using model performance data. Risk Anal 2008;28(5):1457–76.
[15] Droguett E, Mosleh A. Integrated treatment of model and parameter uncertainties through a Bayesian approach. J Risk Reliab 2013;227:41–54.
[16] Droguett E, Mosleh A. Bayesian treatment of model uncertainty for partially applicable models. Risk Anal 2014;34(2):252–70.
[17] Fragola JR. Reliability and risk analysis data base development: an historical perspective. Reliab Eng Syst Saf 1996;51:125–36.
[18] Freivalds A, Johnson AB. Time-series analysis of industrial accident data. J Occup Accid 1990;13:179–93.
[19] Gilks WR, Richardson WR, Spiegelhalter DJ. Markov chain Monte Carlo in practice. London: Chapman & Hall; 1996.
[20] Hora SC, Iman RL. Bayesian modeling of initiating event frequencies at nuclear power plants. Risk Anal 1990;10(1):103–9.
[21] Jallon R, Imbeau D, Warin N. A process mapping model for calculating indirect costs of workplace accidents. J Saf Res 2011;42:333–44.
[22] Jallon R, Imbeau D, Warin N. Development of an indirect-cost calculation model suitable for workplace use. J Saf Res 2011;42:149–64.
[23] Kaplan S. On a 'Two-Stage' Bayesian procedure for determining failure rates from experimental data. IEEE Trans Power Appar Syst 1983;102(1):195–9.
[24] Konstandinidou M, Nivolianitou Z, Markatos N, Kiranoudis C. Statistical analysis of incidents reported in the Greek Petrochemical Industry for the period 1997–2003. J Hazard Mater 2006;A135:1–9.
[25] Lilley R, Feyer AM, Kirk P, Gander P. A survey of forest workers in New Zealand: do hours of work, rest, and recovery play a role in accidents and injury? J Saf Res 2002;33:53–71.
[26] Marcoulaki EC, Konstandinidou M, Papazoglou IA. Dynamic failure assessment of incidents reported in the Greek Petrochemical Industry. Comput-Aided Chem Eng 2011;29:1055–9.
[27] Marcoulaki EC, Papazoglou IA, Konstandinidou M. Prediction of occupational accident statistics and work time loss distributions using Bayesian analysis. J Loss Prevent Process Ind 2012;25:467–77.
[28] Meel A, O'Neill L, Levin J, Seider W, Oktem U, Keren N. Operational risk assessment of chemical industries by exploiting incident databases. J Loss Prevent Process Ind 2007;20:113–27.
[29] Mosleh A. Bayesian modeling of expert-to-expert variability and dependence in estimating rare event frequencies. Reliab Eng Syst Saf 1992;38:1–3.
[30] Mosleh A, Apostolakis G. The development of a generic data base for failure rates. In: proceedins of the international topical meeting on probabilistic safety methods and applications, San Francisco. CA; 1985. p. 48-1–48-10.
[31] Moscoso JM, Romero JC, Canto SP. Occupational accident rate in olive oil mills. Saf Sci 2012;50:169–79.
[32] Papazoglou IA, Aneziris O, Bellamy L, Damen M, Ale BJM, Oh JIH. Uncertainty assessment in the quantification of risk rates of occupational accidents. Risk Anal 2015;35(8):1536–61.
[33] Papazoglou IA, Aneziris O, Post J, Baksteen H, Ale BJM, Oh JIH, Bellamy LJ, Mud ML, Hale A, Goossens L, Bloemhoff A. Logical models for quantification of occupational risk: falling from mobile ladders. In: Proceedings of the international conference on probabilistic safety assessment and management, New Orleans; May 14–18, 2006.
[34] Papazoglou IA, Bellamy LJ, Leidelmeijer KCM, Damen M, Bloemhoff A, Kuiper J, Ale BJM, Oh JIH. Quantification of occupational risk from accidents. In: Proceedings of the ninth international probabilistic safety assessment and management conference, Hong Kong; 18–23 May 2008.
[35] Pörn K. The two-stage Bayesian method used for the t-book application. Reliab Eng Syst Saf 1996;51(2):169–79.
[36] Pransky GS, Benjamin KL, Savageau JA, Currivan D, Fletcher K. Outcomes in work-related injuries: a comparison of older and younger workers. Am J Ind Med 2005;47:104–12.
[37] Rognstad K. Costs of occupational accidents and diseases in Norway. Eur J Operat Res 1994;75:553–66.
[38] Shalini PT. Economic cost of occupational accidents: evidence from a small island economy. Saf Sci 2009;47:973–9.
[39] Singpurwalla ND. Foundational issues in reliability and risk analysis. SIAM Rev 1988;30(2):264–82.
[40] Yakovlev P, Sobel RS. Occupational safety and profit maximization: Friends or foes? J Socio-Econ 2010;39:429–35.