Advances in Reliability, Safety and Security, Part 8 – Kołowrocki, Magryta-Mut (eds.)
© 2024 Polish Safety and Reliability Association, Gdynia, ISBN 978-83-68136-20-3 (printed), ISBN 978-83-68136-07-4 (electronic)

Advances in Reliability, Safety and Security

ESREL 2024
Monograph Book Series

Optimization Of Testing Strategy In Safety Systems Using Hierarchical Bayesian Models

Marcello Dantas Gomes Júnior[a], Pauli Adriano de Almada Garcia[b]

[a]*Petróleo Brasileiro SA, Rio de Janeiro, Brazil*
[b]*Universidade Federal Fluminense, Volta Redonda, Brazil*

Abstract

The search for optimal maintenance policies and periodic testing of Safety systems is recurrent in research. In the face of scarce and/or non-homogeneous data, approaches based on Bayesian inference are proposed to estimate the parameters of the models. This work, whose objective is to optimize the testing strategy of safety system components, makes use of hierarchical Bayesian models to make inferences about the failure behavior of safety devices that are installed on an oil and natural gas production platform. Evolutionary computation provides more efficient and more successful approaches, where the balance between the cost of the maintenance strategy and the probability of failure in demand provides a Pareto boundary. To assist in the choice of the optimal point, this work proposes the use of clustering. By means of an application case related to safety components, the application of the approach is exemplified and its effectiveness for the situation presented is shown.

Keywords: pareto frontier filtering, evolutionary optimization, hierarchical bayesian model, reliability of safety systems

1. Introduction

The models that delineate failure behaviors in reliability engineering are essentially grounded in field data, controlled experiments, and the expertise of experts. Consideration of the uncertainties inherent in these models and data is crucial, as highlighted by (Park et al., 2010). In contexts where monitoring and safety systems need to ensure reliability over mission time, reliability estimation during the operational cycle is performed through field data and expert opinions, supported by appropriate probabilistic models (Jiang et al., 2020). The selection of these models, through inferences, brings with it uncertainties, especially when dealing with non-homogeneous data from various sources. In this sense, the Bayesian approach has been employed as a valuable tool, (Droguett et al., 2004; Hamada et al., 2008; Kelly and Smith, 2011). Safety devices, subject to hidden failures that only manifest themselves during equipment demand or during testing (Taghipour and Banjevic, 2011), require periodic testing to ensure adequate levels of availability (Wang and Pham, 2011). However, the search for a compromise solution in the quantity of tests, balancing cost, and operational availability, is a challenge addressed through optimization (Etemadi and Fotuhi-Firuzabad, 2012; Hokstad et al., 1995; Wang and Pham, 2011).

Evolutionary methods for multi-objective optimization have been the focus of research over the past decade (Coit and Zio., 2019). Such methods return several viable and unmastered solutions, forming a Pareto Frontier (Deb et al., 2002; Deb and Jain, 2014; Khalili-Damghani et al., 2014; Raquel and Naval, 2005). However, choosing a solution among the many offered by the Pareto Frontier is a crucial decision for the decision-maker. Given the potential extension of the frontier, the use of filters to reduce the PF solutions initially proposed has been the subject of study, (Bal and Satoglu, 2019; Tavana et al., 2016; Wang and Rangaiah, 2017).

The selection of an optimal solution to a problem of optimization in the reliability of Safety systems, when confronted with an extensive Pareto frontier, represents a significant challenge for the decision-maker. The multifaceted nature of the Pareto frontier, which offers a diverse range of unmastered solutions, imposes additional complexity on decision-making. The difficulty lies in the abundance of viable alternatives, each representing a trade-off between different goals. The decision-maker is faced with the intricate task of weighing trade-offs between conflicting parameters, such as the cost of the maintenance strategy and the probability of demand failure. In this context, the choice of solution must meet the specific requirements of the system, highlighting the need for effective filtering methods or decision-making strategies that allow for an informed and efficient choice in the face of this diversity of alternatives at the Pareto frontier. A comprehensive review of relevant work on Pareto Frontier treatment was presented by (Petchrompo et al., 2022), including several reliability engineering works.

Therefore, this work proposes a hierarchical Bayesian model to model the reliability of safety devices using non-homogeneous data populations, application of multiobjective optimization to obtain a set of optimal solutions and, finally, the clustering of the Pareto Frontier, to support decision making.

2. Hierarchical Bayesian Models

The two-stage Hierarchical Bayesian Model (HBM) is used to capture the uncertainties of the estimation of the parameters of interest and the source-to-source variability (Pickard, 1983; Yu et al., 2017). Just like the Maxim Likelihood, in the Bayesian approach, HBM models the uncertainty of the sampled data in the priori of the parameter, which is sampled from hyperparameters θ in a non-informative distribution (Yu et al., 2017). In Figure 1, α and β are hyperparameters sampled by a non-informative distribution $P(A, B)$, which allows a wide range of values for α and β, which leaves $p(\theta|\alpha,\beta)$ quite generic to capture source-to-source variability. In the upgrade phase, the Priori has no great influence on its posteriori. So, the Bayesian update is completely dependent on the data (Gelman, 2006). Thus, a posteriori captures the uncertainty of the data and more accurately reflects its true value. Using the uniform for non-informative distribution brings the problem of having to work in a break, which brings bias and invariability in case of reparameterization. To satisfy Jeffrey's rule, a fuzzy range can be used such as (Whistled and Essebbar, 2003; Datta and Ghosh, 1996; Gelman et al., 2004; Hamada et al., 2008; Kelly and Smith, 2011), ***Priori*** to model data uncertainty, as proposed in (Yu et al., 2017). The ***Priori*** can be updated using new data to generate a ***Posteriori*** of the data of interest. There are three steps.

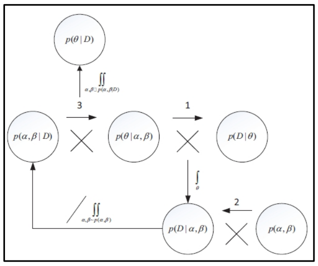

Fig. 1. Stages of the construction and updating of the Hierarchical Bayesian Model.

In the first step, the likelihood of the hyperparameters is obtained by calculating the expected value (or hope) of the likelihood of the parameters of interest:

$$l(D|\alpha,\beta) = \int l(D|\theta) \cdot p(\theta|\alpha,\beta)\partial\theta \quad (1)$$

In the second step, the a posteriori distribution of the hyperparameters is obtained, using Bayes' theorem:

$$p(\alpha,\beta|D) = \frac{p(\alpha,\beta)\cdot l(D|\alpha,\beta)}{\iint p(\alpha,\beta)\cdot l(D|\alpha,\beta)\partial\alpha\partial\beta} \quad (2)$$

The numerical integration of the denominator has been done based on simulations of Markov Chain Monte Carlo (MCMC), (Cronin et al., 2010; Link et al., 2002; Qin et al., 2005; Yu et al., 2017), for α and β sampling and joint distribution $p(\alpha,\beta)$. Then the integrals are approximated by the sample mean of $l(D|\alpha,\beta)$.

In the third step, by marginalizing the parameters α and β, the a posteriori distribution of θ, given the observed data, can be obtained. It is also known as an informative posteriori of θ. The double integral can be approximated with MCMC, (Cronin et al., 2010; Link et al., 2002; Qin et al., 2005; Yu et al., 2017), by the sample mean of $p(\theta|\alpha,\beta)$.

$$p(\theta|D) = \iint p(\theta|\alpha,\beta) \cdot p(\alpha,\beta|D)\partial\alpha\partial\beta \quad (3)$$

In this way, the update with the arrival of new data can be done by:

$$p(\theta|D^t, D) = \frac{p(\theta|D) \cdot p(D^t|\theta)}{\int p(\theta|D) \cdot p(D^t|\theta) \partial\theta} \quad (4)$$

Where $p(D^t|\theta)$ is captures the new data every time interval t.

As seen, Figure 1 synthesizes the construction of the Hierarchical Bayesian Model (HBM). Step 1 multiplies the likelihood of the parameter θ and Priori first-stage, integrating in θ, which produces the relationship between the data from the sources and the parameters α and β of the Priori first-stage, i.e., the likelihood between the data and the parameters of the Priori first-stage stage. In step 2, according to Figure 1, the distribution of the parameters α and β is multiplied by the likelihood between the data and the parameters of the Priori first-stage, the parameters α and β are marginalized by applying Bayes' theorem and and integrating the denominator in α and β using MCMC. In the third step, as Figure 1**Błąd! Nie można odnaleźć źródła odwołania.**, relate to the distribution of Posteriori to α and β from the ***Priori*** of the first stage to produce the distribution to Posteriori of the parameter θ.

3. Probability of failure on demand modeling with a Hierarchical Bayesian Model

Safety systems typically remain in standby mode and act on demand to mitigate or eliminate risks associated with unwanted events. Therefore, such systems and their components are subject to hidden failures, (Eisinger and Oliveira, 2021; Hokstad et al., 1995; Moubray, 1997; Taghipour and Banjevic, 2011; J. Wang et al., 2017). More broadly in the literature, the failure modes of safety systems can be classified into four types:

- Dangerous failures revealed (λ_{DD})
- Hidden Dangerous Failures (λ_{DU})
- Secure failures revealed (λ_{SD})
- Hidden Safe Loopholes (λ_{SU})

Dangerous failures lead to the loss of the safety function. This means that the risk arising from an unwanted event will not be mitigated and the respective safety barrier will fail. Safe failures, in the first instance, bring operational disruption and eventual production losses, but do not affect the safety function. Revealed failures are evident as soon as they occur or are detected via self-diagnostics that some safety devices have, (Ponte Junior, 2015). Hidden failures become apparent only when the safety function is demanded, since the safety device remains in standby mode, (Moubray, 1997; Ponte Junior, 2015). In the latter case, the safety function may be required either in a real situation or in a test. Failure rates ($\lambda_{DD}, \lambda_{DU}, \lambda_{SD}, \lambda_{SU}$) of safety devices can be determined based on field data, databases, expert experience, technical documentation from manufacturers, and more. These rates are measured in the number of failures per unit of time.

In this work, the interest lies in the hidden dangerous failures (λ_{DU}). Hidden failures are discovered only in real-world demand situations or in testing, addressing the frequency of component failure, periodic component testing, and unavailability due to eventual repairs.

Typically, the reliability of safety devices is modeled, according to Equation 5, using constant failure rate, (Kumamoto, 2007; Rausand, 2014; Stamatelatos and Dezfuli, 2011).

$$R(t) = e^{-\lambda_{DU} \cdot t} \quad (5)$$

The metric of interest in safety system modeling is the PFD, which is obtained by calculating the steady-state availability of the safety function. The PFD depends on the failure rate of the components and the periodicity of testing. According to (Lewis et al., 1994; Rausand, 2014; Smith, 2011), without loss of generality, it is considered a non-repairable safety device, with negligible testing time. In this case, availability equals reliability:

$$A(t) = R(t) = e^{-\lambda_{DU} \cdot t} \quad (6)$$

For m periodical test intervals T_o, we have:

$$A^*(\infty) = \lim_{m\to\infty} \frac{1}{mT_o} \int_0^{mT_o} A(t)dt = \frac{1}{T_o} \cdot \int_0^{T_o} A(t)dt \quad (7)$$

In this modeling, it would be possible to achieve an availability level very close to 1 (or PFD very close to zero) just by reducing the test interval. This is not reasonable in practice, as it would require an impractical amount of work from the maintenance team. In addition, some devices are out of operation during testing (Ponte Junior, 2015), which would leave such devices unavailable most of the time due to excessive testing. Then two parcels are included in the template. The first is related to the time t_{off} in which the safety device is unavailable due to the test, while the second is related to the average repair time, $\frac{1}{\mu}$, of the safety device if a failure is found during the test. The formulation that best represents availability is as per Equation 8.

$$A^* = \frac{1}{T_o} \cdot \int_0^{T_o} R(t)dt - \frac{t_{off}}{T_o} - \frac{\lambda_{DU}}{\mu} \quad (8)$$

Safety systems are arrangements of safety devices or equipment. Then it is necessary to model the availability of these systems, based on the individual availability of the devices or equipment. The equation for the availability of a system follows the same construction logic as the equation for the reliability of the system, (Lewis et al., 1994). In both cases, the topology of the system will reflect equally in the mathematical modeling. For example, for n serial devices, the availability of the system as a function of the test periodicity is as follows: Equation 9:

$$A(t) = A_1(t) \cdot A_2(t) \cdot \ldots \cdot A_n(t) \tag{9}$$

And in the case of parallel items, we have Equation 10:

$$A(t) = 1 - \prod_{i=1}^{n}\left(1 - A_i(t)\right) \tag{10}$$

There are other arrangements of safety devices or equipment besides the serial and parallel arrangement. It is common to design Safety instrumented systems with redundancies and voting. The most common application of voting logic is two out of three or 2oo3 (2 *out-of-3*), (Rausand, 2014). In this case, two devices need to report the abnormality in the variant of interest for the system's safety function to be triggered. The availability of the system is obtained by applying the Binomial as a function of the test policy and failure rates of its components, and its probabilistic complement corresponds to the PFD. Therefore, it is important to determine the best test intervals for safety system devices.

According to (Kelly and Smith, 2011; Hamada et al., 2008; Babaleye et al., 2019), the most common application for Hierarchical Bayesian Models is in the use of several similar sources of data, where the interest lies in modeling source-to-source variability.

In the methodological procedure proposed in this work, the Binomial distribution is applied to model the behavior of each data source, where the number of failures X_i follows a Binomial with n_i data (demands) and parameter p_i (probability of failure). To have a hierarchical model for the parameter p, a first-stage priori distribution is specified. In the specific case, to work with a conjugate, you choose the distribution $Beta(\alpha, \beta)$. However, it is not compulsory to work with a $Beta(\alpha, \beta)$ conjugate priori. It is therefore necessary to infer the parameters (α, β) of the first-stage priori. Such an inference is carried out by a second-stage priori distribution, or hyperpriori. In the area of Safety and reliability analysis, the use of two-stage models is predominant, although there are no limitations for further stages, (Babaleye et al., 2019; Hamada et al., 2008; Kelly and Smith, 2011; Yu et al., 2017).

To have a population variability updated by the different sources, the parameters can be inferred by a diffuse distribution known as the Jeffrey distribution, (Whistled and Essebbar, 2003; Datta and Ghosh, 1996; Gelman et al., 2004; Hamada et al., 2008; Kelly and Smith, 2011), which plays the role of the second-stage priori. It is not imperative to use Jeffrey, as it will depend on the application. The goal of α and β parameter inference is to have up-to-date population variability for different data sources. The second-stage Priori is also important to establish dependence on the parameters α and β. With the observed data, a posteriori will reflect this dependency. In turn, the Binomial parameter pi_i is inferred by the first-stage Priori distribution, which is a Beta distribution with the parameters α and β. The index i corresponds to each of the data sources, and the value N corresponds to the total number of data sources. Knowing the future number of demands and the probability of a failure given demand, in each data source, it is possible to make predictions about the number of failures in the next demands.

The failure probabilities of data sources are conditionally independent given the values of the α and β parameters. The posteriori predictive distribution for p represents source-to-source variability. The predictive distribution of p will be given by the mean of the a posteriori distribution of p, conditioned by the Beta distribution of parameters α and β, weighted by the a posteriori distribution for α and β, according to Equation 11.

$$\pi(p_i|\tilde{x},\tilde{n}) = \int_0^1 \int_0^1 \ldots \int_0^1 \{\int \int [\prod_{i=1}^{N} \pi_1(p_i|\alpha,\beta)] \pi_2(\alpha,\beta|\tilde{x},\tilde{n}) \partial\alpha\partial\beta\} \partial p_1 \partial p_2 \ldots \partial p_{n-1} \tag{11}$$

Where:
$\tilde{x}$ is the vector of the number of failures for each data source.
$\tilde{n}$ is the number of tests performed on each data source.
p_i is the predictive binomial parameter of each data source.

This corresponds to:

$$\pi(p_i|\tilde{x},\tilde{n}) = \int \int \pi_1(p_i|\tilde{x},\tilde{n},\alpha,\beta) \pi_2(\alpha,\beta|\tilde{x},\tilde{n}) \partial\alpha\partial\beta \tag{12}$$

Here we have the p distribution of variability from source to source. This is the marginal distribution a posteriori. From this distribution, it is possible to estimate the number of failures in the next demands in each of the sources. The following is the a posteriori predictive distribution by Equation 13:

$$\pi_{pred}(p^*|\tilde{p}) = \int \int \pi_1(p^*|\alpha,\beta,\tilde{x},\tilde{n}) \pi_2(\alpha,\beta|\tilde{x},\tilde{n}) \partial\alpha\partial\beta \tag{13}$$

Where p^* is the parameter of the mediated predictive binomial distribution, i.e., global.

To estimate how many x_n failures will occur in the set of sources in the next n_n demands, the predictive distribution is obtained, according to Equation 14:

$$\pi_{pred}(x_n|p, n_n) = \int_0^1 L(x_n|p)\pi_1(p|\tilde{n}, \tilde{x})dp \quad (14)$$

It is possible to update the Hierarchical Bayesian Model with the arrival of new data. If conjugate priori is used, there is no need to infer the parameters of this priori with the MCMC. The update is made from the first-stage priori parameters, according to Equations 3 and 4.

The output of the Hierarchical Bayesian Model is the failure rate for each data source and the overall failure rate for all data sources. These rates will feed the reliability model in the next step to obtain the system's PFD. The predictions of the number of failures made by the Bayesian model will be compared with the field data. According to (Montgomery and Runger, 2014), Pearson's correlation coefficient below $|0{,}5|$ usually indicates a weak correlation, if it is greater than $|0{,}8|$, it usually indicates a strong correlation. This modeling must be able to handle non-homogeneous data and deal with data scarcity.

4. Case study

In oil and gas drilling, production, processing, or refining activities, it is possible to have the accidental presence of flammable, asphyxiating or toxic gases in the process plant environments. Production systems are designed and maintained in such a way as to avoid the presence of these gases. But, if this presence occurs, the gases must be detected by fire and gas detection devices designed for this. Some detector models have electrochemical sensory cells that respond to the stimulus of the presence of gas, transforming the chemical response into a direct current electronic signal of 4 to 20 mA. To perform this function reliably, the detectors need to be periodically tested and calibrated. Tests are necessary to know if the detectors are responding reliably to the actual concentration of the gas, if they are reporting a presence greater than the real one or less than the real one, in parts per million (ppm). There is an acceptable range of response of the detector to the stimulus of the presence of the gas. If the detector responds above the actual gas concentration, it is a sure failure. On the other hand, if the detector responds below the actual gas concentration, it is a dangerous failure. The loss of calibration is considered a hidden failure, since its occurrence is not evident to the operators, and is therefore perceived only in tests, or in a situation of real demand, and it is a protective function, according to (Moubray, 1997). After each test, the detector is compulsorily recalibrated and returns to operating in the acceptable response range, close to the ideal response. As a premise of this work, the recalibration action of the detector is considered "*as good as new*", that is, the calibration of the detector returns to the initial level of reliability. In this case study, the premise of perfect testing is adopted, i.e., the occurrence of calibration loss, the test will certainly detect this failure.

The detectors are installed throughout the entire process plant, in different areas. Each area is subject to a different environmental condition, which interferes with the loss of calibration of the detectors, as has been observed in professional experience of the authors. Therefore, depending on the area in which the detector is installed, it will have a different dangerous hidden failure rate, λ_{DU}. Each area of the plant studied will have a calibration loss (hidden fault) modeling for their respective detectors. In each area, the detectors are grouped into 1ooN or 2ooN voting, depending on the design requirement. Each voting is designed to detect the presence of a gas cloud. The calculation of the probability of demand failure (PFD) will be done at the voting level, since the vote must inform the presence of the gas and its failure leaves the protection function unavailable, that is, it does not confirm the presence of the gas. Another important point is the level of difficulty of access to the detector for interventions, such as testing and calibration. Depending on the height from the ground, access can be easy, medium, or difficult. This impacts the testing and calibration time, therefore, the time in which the detector is unavailable, as well as the preparation time to perform the service and the demobilization of the team after completion. The sequence of steps of this work is illustrated in the Figure 2.

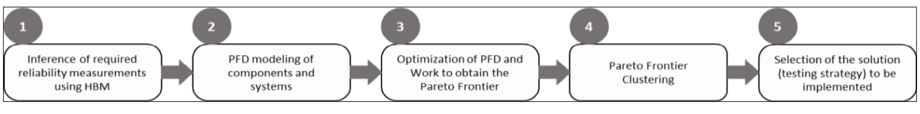

Fig. 2. Methodological procedure of this work.

According to Table 1, the field data were divided into three samples, according to the time window in which they were obtained:

Table 1. Field data samples used to build the Hierarchical Bayesian Models.

Area	First sample		Second sample		Third sample	
	Tests	Failures	Tests	Failures	Tests	Failures
1	79	6	27	7	24	8
2	18	3	8	4	4	1
3	42	9	22	6	14	7
4	105	29	51	16	31	11
5	6	1	3	1	3	1
6	71	4	16	5	8	1
7	635	81	247	31	174	27
8	231	51	90	24	54	14
Total	1.187	181	464	94	312	70

HBM was applied to the first sample, updated with the second sample. The failure prediction of the first HBM was purchased with the field data of the second sample, as Figure 3. Then, the predictions of the updated HBM model were compared with the field data of the third sample, as Figure 3.

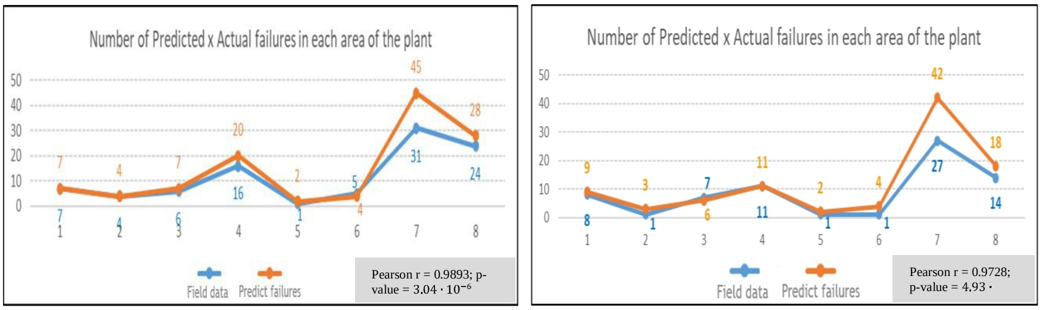

Fig. 3. (a) comparison between the number of failures predicted by the initial HBM and the number of failures observed in the field; (b) comparison between the number of failures predicted by the updated HBM and the number of failures observed in the field.

As HBM is updated with the addition of new data, the individual population metrics converge to the global metrics, due to the reduction of source-to-source variability, including the first-stage parameters and a priori. The $\alpha\beta$Table 2 shows this behavior for the failure rate. The mean value and standard deviation of the failure rates are evidenced, as well as the failure rates at the credibility threshold of 97.5%.

Table 2. Detector failure rates for each plant area in the initial and upgraded HBMs.

Failure rate (1/hours)	Initial HBM			First update of HBM			Second update of HBM		
	Mean	Standard deviation	97.5% (cred)	Mean	Standard deviation	97.5% (cred)	Mean	Standard deviation	97.5% (cred)
λ_1	4.84E-5	1.60E-5	8.33E-5	1.05E-4	1.04E-5	1.26E-4	1.11E-4	8.44E-6	1.28E-4
λ_2	8.22E-5	3.24E-5	1.59E-4	1.06E-4	1.07E-5	1.28E-4	1.09E-4	8.46E-6	1.26E-4
λ_3	7.49E-5	2.39E-5	1.29E-4	1.05E-4	1.04E-5	1.27E-4	1.12E-4	8.54E-6	1.29E-4
λ_4	1.32E-4	2.55E-5	1.88E-4	1.10E-4	1.04E-5	1.31E-4	1.12E-4	8.42E-6	1.29E-4
λ_5	8.21E-5	4.11E-5	1.82E-4	1.04E-4	1.06E-5	1.25E-4	1.09E-4	8.49E-6	1.26E-4
λ_6	4.23E-5	1.62E-5	7.79E-5	1.05E-4	1.05E-5	1.27E-4	1.08E-4	8.42E-6	1.25E-4
λ_7	6.40E-5	6.95E-6	7.83E-5	8.91E-5	7.84E-6	1.05E-4	1.03E-4	7.42E-6	1.18E-4
λ_8	1.11E-4	1.55E-5	1.43E-5	1.09E-4	1.01E-5	1.30E-4	1.11E-4	8.28E-6	1.28E-4
λ_{Global}	8.30E-5	5.5E-5	2.09E-4	1.03E-4	1.05E-5	1.25E-4	1.09E-4	8.47E-6	1.26E-4

The next step is to obtain the Pareto Boundary. The Multiobjective Particle Swarm Optimization based on Crowning Distance (MOPSO-CD) will be used, as per (Raquel and Naval, 2005). The PFD of each of the 16 detector votes (*koon*) will be minimized, Equation 15, and the total work required to perform the tests and calibrations, Equation 16. In this way, there are 16 objective functions in total. The decision variables are the T_{o_i} periodicities in which the detectors in each i area will be tested and calibrated.

$$PFD_j = 1 - \sum_k^n \left[\left(\frac{n!}{k!(n-k)!} \right) \left(\frac{1}{T_{o_i}} \int_0^{T_{o_i}} e^{-\lambda_7 t} dt - \frac{T_{off}}{T_{o_i}} \right)^k \left(1 - \left(\frac{1}{T_{o_i}} \int_0^{T_{o_i}} e^{-\lambda_7 t} dt - \frac{T_{off}}{T_{o_i}} \right) \right)^{n-k} \right] \quad (15)$$

$\forall\ i = 1, \dots, 8$ and $\forall\ j = 1, \dots, 16$

$$Work_{total} = 8760 \sum_{i=1}^{8} \left(\frac{N_i \cdot T_{w_i}}{T_{o_1}} \right) \quad (16)$$

The MOPSO-CD parameters used in this multi-objective optimization are shown in Table 3:

Table 3. Parameters used in MOPSO-CD for optimization.

Objective Functions: 16	Search space limits (hours): 720 and 8760
Decision Variables: 8	Mutation Probability: 0.5
Population: 100 individuals	Inertia Weight: 0.4
Generations: 1000	Acceleration coefficient C_1: 2
External solution repository size: 100	Acceleration coefficient C_2: 2

After the application of MOPSO-CD, a Pareto Frontier of 16 objectives and 100 non-mastered solutions was obtained. However, as previously mentioned, it is not feasible for decision-makers to analyse the Pareto Frontier and choose the best solution. To facilitate the choice of the most appropriate solution to implement the detector testing and calibration strategy, the Pareto Frontier will have its solutions clustered.

The division into clusters in k-means occurs by minimizing the sum of squares, that is, by the Euclidean distance from the points to the centroid of each of its clusters, as (MacQueen, 1967). Among the diversity-based Pareto boundary filtering methods, according to (Petchrompo et al., 2022), k-means is by far the most widely used. This fact is confirmed in (Lal and Datta, 2019; Mahdavian et al., 2017; Sato et al., 2019; Taboada et al., 2007; Taboada and Coit, 2007). K-medoids shows up as an alternative to k-means, as (Reynolds et al., 2006; Schubert and Rousseeuw, 2019; Schubert and Rousseeuw 2021), where, after finding a set of k medoids, k clusters are constructed by assigning each observation to then nearest medoid. The k-means has centroids that do not necessarily coincide with any of the points in the cluster, while the k-medoids indicate which point is the medoid of each cluster, (Petchrompo et al., 2022).

In the present case study, for clustering via k-medois, the solutions of the Pareto Frontier are the observations, and the objectives are the analysis variables. Of the 16 total objectives, 15 of them correspond to the PFD of each vote, which varies between 0 and 1. The other objective is the Annual Work in man-hours and does not vary from 0 to 1. Thus, the first step was to divide all the Annual Work values by the highest Annual Work value of the respective Pareto Frontier, to vary between 0 and 1 as well. Another way, often found in the literature, is to standardize the variables by applying the ZScore, making the variables mean 0 (zero) and standard deviation 1 (one). Next, the Elbow test is done to find out the optimal number of k clusters, (Abdulhafedh, 2021; Kodinariya and Makwana, 2013; Sinaga and Yang, 2020; Syakur et al., 2018; Wu, 2012), in case the decision-maker does not want to choose a priori the number k of clusters. Pareto Frontier solutions are grouped into clusters and medoids are identified. To verify which objectives did or did not influence the definition of at least one *cluster*, the analysis of variance test, better known as ANOVA, is applied.

The result was the division of VT into three clusters: a bolder cluster, which prioritizes job reduction, a more conservative cluster, which prioritizes safety, and a more moderate cluster, which establishes a more interesting trade-off between safety and work demanded.

The Table 4 4 shows the result of the ANOVA test applied to the clusters. If the null hypothesis is rejected, it means that that criterion was relevant in the partitioning of the clusters. If the null hypothesis is not rejected, it means that that criterion was not relevant in the partitioning of the clusters. The higher the value of the F statistic, the greater the contribution of that criterion to the formation of clusters.

Table 4. Results of the ANOVA test performed on Pareto Frontier clusters.

Area	Voting	Statistic F	P-value	Hypothesis
1	2oo3	6.27	2.76E-03	Alternative: PFD contributes to forming a cluster
1	1oo1	6.30	2.69E-03	Alternative: PFD contributes to forming a cluster
2	2oo2	11.30	3.88E-05	Alternative: PFD contributes to forming a cluster
3	2oo2	11.29	3.91E-05	Alternative: PFD contributes to forming a cluster
3	2oo4	13.43	7.08E-06	Alternative: PFD contributes to forming a cluster
4	2oo7	10.75	6.08E-05	Alternative: PFD contributes to forming a cluster
4	2oo8	11.11	4.52E-05	Alternative: PFD contributes to forming a cluster
5	1oo1	2.86	6.19E-02	Null: PFD does not contribute to cluster formation
6	2oo4	0.84	4.35E-01	Null: PFD does not contribute to cluster formation
7	1oo3	110.90	<2.00E-16	Alternative: PFD contributes to forming a cluster
7	2oo2	284.40	<2.00E-16	Alternative: PFD contributes to forming a cluster
7	2oo2	283.30	<2.00E-16	Alternative: PFD contributes to forming a cluster
8	2oo2	2.88	6.12E-02	Null: PFD does not contribute to cluster formation
8	2oo2	2.88	6.11E-02	Null: PFD does not contribute to cluster formation
8	2oo2	2.88	6.11E-02	Null: PFD does not contribute to cluster formation
Work		322.7	<2.00E-16	Alternative: Work contributes to forming a cluster

K-medoids provides k best solutions, corresponding to k Clusters. That is, the decision-maker can analyse only the centroid solution of each cluster, instead of analysing the entire PF. The Table 5 shows the optimal solutions that are centroids of the three clusters:

Table 5. Centroid solutions of the three clusters of the Pareto Frontier

Area	Voting	Objective	Conservative cluster	Bold cluster	Moderate cluster
1	*2oo3*		7.39E-2	7.34E-2	7.30E-2
1	*1oo1*		1.67E-1	1.66E-1	1.66E-1
2	*2oo2*		3.84E-1	3.14E-1	3.59E-1
3	*2oo2*		3.91E-1	3.19E-1	3.65E-1
3	*2oo4*		8.92E-4	2.26E-3	8.56E-4
4	*2oo7*		1.25E-4	1.72E-5	7.20E-5
4	*2oo8*		2.40E-5	2.42E-6	1.27E-5
5	*1oo1*	PFD	2.81E-1	2.88E-1	2.83E-1
6	*2oo4*		4.84E-3	2.29E-3	7.19E-3
7	*1oo3*		1.25E-4	4.70E-3	3.33E-4
7	*2oo2*		9.86E-2	3.07E-1	1.35E-1
7	*2oo2*		1.01E-1	3.08E-1	1.36E-1
8	*2oo2*		4.30E-1	4.62E-1	4.10E-1
8	*2oo2*		4.31E-1	4.62E-1	4.10E-1
8	*2oo2*		4.31E-1	4.62E-1	4.11E-1
		Annual Work (man-hours)	3321	1409	2571

The choice of the most suitable solution among the three of the Table 5 it is at the discretion of the decision-makers, respecting the Safety Integrity Level (SIL) limits for each PFD, where it exists.

5. Final thoughts and recommendations

The general objective was achieved, as a methodological procedure was proposed to optimize the testing strategy of a Safety system with hundreds of components, divided into areas and votes.

In relation to the Hierarchical Bayesian Model approach, already established in the literature, to infer the failure behavior of the safety systems under study, the modeling proved to be concise and scalable for the failure behavior of the components of a Safety system. The Hierarchical Bayesian Model was updated as new field information arrived, dealing with non-homogeneous data, without losing the ability to predict. It should be noted that, in the comparisons between the predictions made by the model before and after the first update and the validation samples, the correlation remained high. These results ratified the efficacy of the Bayesian model in the treatment of non-homogeneous data, uncertainties, and source-to-source variability, enabling high prediction capacity, with reliable representation of what occurs in the field. In terms of gain in scale, the use of Hierarchical Bayesian Modeling can be adopted in other types of systems subject to hidden failures and periodic testing. Additionally, as an improvement in the inference step of the failure parameters of the model, a homogeneity test can be applied between the various data sources, to verify the real need to apply Hierarchical Bayesian Modeling.

The goal of supporting decision-making in various scenarios was also achieved. The clustering of non-mastered solutions produced by k-medoids assured the decision-maker that the preferred solutions bring a very balanced compromise between safety and work, prioritizing and ensuring safety. The decision-maker will have no difficulty in analysing and deciding on the best solution. Additionally, the decision-maker does not have to bring information a *priori* to the treatment of the Pareto boundary, mitigating the degree of subjectivity of the decision-making process.

A limitation to the case study model to be listed is the failure to consider the aging of safety system devices, when modeling their reliability and PFD with a constant failure rate in an exponential model. With the use of approaches such as the Non-homogeneous Poisson Process (NHPP), it is possible to model the aging failure behavior of safety devices. The application of Markov chains can assist in this process. It is emphasized that this limitation does not impose restrictions or impact the proposed methodological procedure, which is the objective of this study.

References

Abdulhafedh, A. 2021. Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation. Journal of City and Development, 3(1), 12–30.

Assoudou, S., and Essebbar, B. 2003. A Bayesian Model for Markov Chains via Jeffrey's Prior. Communications in Statistics - Theory and Methods, 32(11), 2163–2184.

Babaleye, A. O., Kurt, R. E., and Khan, F. 2019. Hierarchical Bayesian model for failure analysis of offshore wells during decommissioning and abandonment processes. Process Safety and Environmental Protection, 131, 307–319.

Bal, A., and Satoglu, S. I. 2019. The use of data envelopment analysis in evaluating pareto optimal solutions of the sustainable supply chain models. Procedia Manufacturing, 33, 485–492.

Cronin, B., Stevenson, I. H., Sur, M., and Körding, K. P. 2010. Hierarchical bayesian modeling and Markov chain Monte Carlo sampling for tuning-curve analysis. Journal of Neurophysiology, 103(1), 591–602.

Datta, G. S., and Ghosh, M. 1996. On the invariance of noninformative priors. In The Annals of Statistics (Vol. 24, Issue 1), 141-159.

Deb, K., and Jain, H. 2014. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints. IEEE Transactions on Evolutionary Computation, 18(4), 577–601.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2), 182–197.

Eisinger, S., and Oliveira, L. F. 2021. Evaluating the safety integrity of safety systems for all values of the demand rate. Reliability Engineering and System Safety, 210, 107457.

Etemadi, A. H., and Fotuhi-Firuzabad, M. 2012. Design and routine test optimization of modern protection systems with reliability and economic constraints. IEEE Transactions on Power Delivery, 27(1), 271–278.

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). In Bayesian Analysis (Vol. 1, Issue 3).

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2004. Bayesian Data Analysis. Chapman and Hall/CRC, New York

Hamada, M. S., Wilson, A. G., Reese, C. S., and Martz, H. F. 2008. Bayesian Reliability. Springer, New York.

Hokstad, P., Flotten, P, Holmstrom, S., Mckenna, F, and Onshus, and T. 1995. A reliability model for optimization of test schemes for fire and gas detectors. In Reliability Engineering and System Safety (Vol. 47), 15-25

Jiang, Q., Gao, D. ce, Zhong, L., Guo, S., and Xiao, A. 2020. Quantitative sensitivity and reliability analysis of sensor networks for well kick detection based on dynamic Bayesian networks and Markov chain. Journal of Loss Prevention in the Process Industries, 66, 104180

Kelly, D., and Smith, C. 2011. Bayesian Inference for Probabilistic Risk Assessment. Springer, London.

Khalili-Damghani, K., Abtahi, A. R., and Tavana, M. 2014. A decision support system for solving multi-objective redundancy allocation problems. Quality and Reliability Engineering International, 30(8), 1249–1262.

Kodinariya, T. M., and Makwana, P. R. 2013. Review on determining number of Cluster in K-Means Clustering. International Journal of Advance Research in Computer Science and Management Studies, 1(6), 90-95

Kumamoto, H. 2007. Satisfying Safety Goals by Probabilistic Risk Assessment (1st ed.). Springer-Verlag London Limited, Kyoto

Lal, A., and Datta, B. 2019. Optimal Groundwater-Use Strategy for Saltwater Intrusion Management in a Pacific Island Country. Journal of Water Resources Planning and Management, 145(9).

Lewis, E. E., Wiley, J., York, N., Brisbane, C., and Singapore, T. 1994. Introduction to Reliability Engineering, John Wiley and Sons , Inc, New York.

Link, W. A., Cam, E., Nichols, J. D., and Cooch, E. G. 2002. Of Bugs and Birds: Markov Chain Monte Carlo for Hierarchical Modeling in Wildlife. The Journal of Wildlife Management (Vol. 66, Issue 2), 277-291

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. In L. Lecam and J. Neyman (Eds.), In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 281–297.

Mahdavian, M., Sudeng, S., and Wattanapongsakorn, N. 2017. Multi-objective optimization and decision making for greenhouse climate control system considering user preference and data clustering. Cluster Computing, 20(1), 835–853.

Montgomery, D. C., and Runger, G. C. 2014. Applied Statistics and Probability for Engineers (6th ed.). John Wiley and Sons, Inc, Tucson.

Moubray, J. 1997. Reliability Centered Maintenance (Second). Industrial Press, London.

Park, I., Amarchinta, H. K., and Grandhi, R. V. 2010. A Bayesian approach for quantification of model uncertainty. Reliability Engineering and System Safety, 95(7), 777–785.

Petchrompo, S., Coit, D. W., Brintrup, A., Wannakrairot, A., and Parlikad, A. K. 2022. A review of Pareto pruning methods for multi-objective optimization. Computers and Industrial Engineering, 167.

Pickard, S. K. 1983. "Two-stage" bayesian procedure for determining failure rates from experiential data. In IEEE Transactions on Power Apparatus and Systems (Vol. 102, Issue 1), 195-202.

Ponte Junior, G. P. 2015. Risk Management in the Oil and Natural Gas Industry (1st ed., Vol. 1). Elsevier, Rio de Janeiro.

Qin, X., Ivan, J. N., Ravishanker; Nalini, and Liu, J. 2005. Hierarchical Bayesian Estimation of Safety Performance Functions for Two-Lane Highways Using Markov Chain Monte Carlo Modeling. Journal of Transportation Engineering, 345-351.

Carlo R. Raquel and Prospero C. Naval. 2005. An effective use of crowding distance in multiobjective particle swarm optimization. In Proceedings of the 7th annual conference on Genetic and evolutionary computation (GECCO '05). Association for Computing Machinery, New York, NY, USA, 257–264.

Rausand, M. 2014. Reliability of safety-critical systems: theory and application (1st ed.). John Wiley and Sons, Inc, Hoboken, New Jersey.

Reynolds, A. P., Richards, G., de La Iglesia, B., and Rayward-Smith, V. J. 2006. Clustering rules: A comparison of partitioning and hierarchical clustering algorithms. Journal of Mathematical Modelling and Algorithms, 5(4), 475–504.

Sato, Y., Izui, K., Yamada, T., and Nishiwaki, S. 2019. Data mining based on clustering and association rule analysis for knowledge discovery in multiobjective topology optimization. Expert Systems with Applications, 119, 247–261.

Schubert, E., and Rousseeuw, P. J. 2019. Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms, 171–187.

Schubert, E., and Rousseeuw, P. J. 2021. Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. Information Systems, 101.

Sinaga, K. P., and Yang, M.-S. 2020. Unsupervised K-Means Clustering Algorithm. IEEE Access, 8, 80716–80727.

Smith, D. J. 2011. Reliability, Maintainability and Risk: Practical methods for engineers (8th ed., Vol. 1). Elsevier, Boston.

Stamatelatos, M., and Dezfuli, H. 2011. Probabilistic Risk Assessment Procedures Guide for NASA Managers, Create Space Independent Publishing Platform, Hanover.

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., and Satoto, B. D. 2018. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. IOP Conference Series: Materials Science and Engineering, 336, 012017.

Taboada, H. A., Baheranwala, F., Coit, D. W., and Wattanapongsakorn, N. 2007. Practical solutions for multi-objective optimization: An application to system reliability design problems. Reliability Engineering and System Safety, 92(3), 314–322.

Taboada, H. A., and Coit, D. W. 2007. Data Clustering of Solutions for Multiple Objective System Reliability Optimization Problems. Quality Technology and Quantitative Management, 4(2), 191–210.

Taghipour, S., and Banjevic, D. 2011. Periodic inspection optimization models for a repairable system subject to hidden failures. IEEE Transactions on Reliability, 60(1), 275–285.

Tavana, M., Li, Z., Mobin, M., Komaki, M., and Teymourian, E. 2016. Multi-objective control chart design optimization using NSGA-III and MOPSO enhanced with DEA and TOPSIS. Expert Systems with Applications, 50, 17–39.

Wang, J., Zhang, Q., and Jiang, W. 2017. Optimization of calibration intervals for automatic test equipment. Measurement: Journal of the International Measurement Confederation, 103, 87–92.

Wang, Y., and Pham, H. 2011. A multi-objective optimization of imperfect preventive maintenance policy for dependent competing risk systems with hidden failure. IEEE Transactions on Reliability, 60(4), 770–781.

Wang, Z., and Rangaiah, G. P. 2017. Application and Analysis of Methods for Selecting an Optimal Solution from the Pareto-Optimal Front obtained by Multiobjective Optimization. Industrial and Engineering Chemistry Research, 56(2), 560–574.

Wu, J. 2012. Advances in K-means Clustering. Springer, Berlin.

Yu, H., Khan, F., and Veitch, B. 2017. A Flexible Hierarchical Bayesian Modeling Technique for Risk Analysis of Major Accidents. Risk Analysis, 37(9), 1668–1682.