

HYPERPARAMETER OPTIMIZATION OF BAYESIAN PRIOR DISTRIBUTION THROUGH EQUIPMENT DEVELOPMENT IN O&G INDUSTRY: A LITERATURE REVIEW

Lavinia Maria Mendes Araújo^{a,b}
lavinia.mendes@ufpe.br

Isis Didier Lins^{a,b}
isis.lins@ufpe.br

Caio Bezerra Souto Maior^{a,b,c}
caio.maior@ufpe.br

Márcio das Chagas Moura^{a,b}
marcio.cmoura@ufpe.br

^aCentro de Estudos e Ensaios em Risco e Modelagem Ambiental (CEERMA)

^bUniversidade Federal de Pernambuco (UFPE)

Av. da Arquitetura, s/n - Cidade Universitária, Recife - PE, 50740-550

^cCentro de Tecnologia (UFPE)

Av. Marielle Franco, s/n - Km 59 - Nova, PE, 55014-900

ABSTRACT

The creation of new technologies poses many technical obstacles in the Oil and Gas (O&G) industry, including disruptive developments. The process of estimating reliability in the design process is not trivial and includes several factors of uncertainty, especially in Research and Development (R&D). Thus, the Bayesian methodology has been widely used in these cases where a broad background of information is not available. The maximum entropy method (ME) is a technique used for estimating the parameters of a Bayesian prior probability distributions of failure events. Therefore, this paper aims to perform a systematic literature review to identify the main optimization methods used for solving ME in the Bayesian context and in the O&G industry. Among the gaps identified, it can be seen that the O&G context is a vast application area to be explored. Thus, opening a broad possibility of studies of optimization methods to be tested on this problematic.

KEYWORDS. Bayesian Prior Distribution. Oil and Gas. Reliability.
EN&PG – OR in the Area of Energy, Oil and Gas; MH – Metaheuristics.

1. Introduction

In the development phase of technologies, estimating reliability is critical for ensuring predictability in the life of equipment and technologies installed in oil wells. However, data, especially for technologies being developed in the Oil and Gas (O&G) companies, may be limited and/or expensive to acquire. Expert opinion and general databases are often available sources of knowledge [Maior et al. 2022]. Bayesian inference is increasingly used to combine multiple types of information and to evaluate the reliability of new techniques, as it allows updating prior knowledge as new information is obtained [Cai et al. 2012], [Afshari et al. 2022], [Shafiee; Elusakin; Enjema 2020], [Wu et al. 2015], [Yaez; Joglar; Modarres 2002].

According to Bayesian methodology, the phase of obtaining prior knowledge is crucial for effective reliability estimation. Data generally comes from similar equipment or expert knowledge in the initial stages. Documents such as FMECA (Failure Modes, Effects, and Criticality Analysis) can be used as a basis for the detection and sequencing of failure events. Each major event can be assigned to the top of a fault tree (FT), which combines failures of basic components. For the calculation of the prior distribution, different approaches can be used [Fernández & Pérez-González 2012], [Wang et al. 2018], [Zhong et al. 2022]. [Maior et al. 2022], for example, combine expert opinion with the Method of Moments (MM) to discrete probability models and with the Maximum Entropy (ME) method to continuous distributions. The ME seeks to maximize a measure of entropy to estimate the parameters of the reliability model represented by continuous probability distributions (e.g., Weibull). The related optimization problem is not simple to solve analytically, as it entails objective function and constraints involving integrals with no closed expressions.

Therefore, this paper this article is motivated by the research question: "how to obtain optimal parameter estimates for the a priori distributions of failure events to provide a successful Bayesian model for reliability estimation of new equipment in the O&G industry?". To this end, a systematic literature review will be conducted to investigate optimization methods have already been used in this context.

The remainder of this paper is organized as follows. Section 2 provides a brief theoretical background about the Bayesian approach and the ME method. Section 3 addresses the research method when conducting the literature review. Section 4 presents the main results and gaps found. Finally, Section 5 presents the conclusive aspects.

2. Theoretical Background

2.1. The Bayesian Approach

The Bayesian approach is a probability formulation that calculates the possibility of an event happening based on the knowledge related to the event. As new evidence becomes available, the method updates the original estimates for the set of parameters of a probability distribution, θ . The parameter of interest of a reliability function, such as the alfa and beta parameters of Weibull distribution, can be represented as θ . The Bayes' theorem (Eq. 1) is used to update the prior distribution with evidence [Bhardwaj; Teixeira; Soares 2022], [Shafiee; Elusakin; Enjema 2020].

$$\pi_1(\theta) = \pi_0(\theta) \times \frac{L(E|\theta)}{\int \pi_0 \times L(E|\theta) d\theta} \quad (1)$$

The prior distribution ($\pi_0(\theta)$) is defined using the information available. To truly reflect the first ideas, prior distributions must be carefully chosen. Because likelihood functions can be weak when there is minimal data, choosing a non-informative prior distribution can result in a biased posterior distribution with significant uncertainty. As a result, it is critical to compile all accessible data in order to construct the prior distributions. From the view of Bayesian estimating inference, expert opinion is a source of evidence. Experts can either provide direct parameter

estimations of past distributions or present important values from these distributions, such as expected failure probability or dependability, or a lower or upper percentile [Maior et al. 2022].

2.2. Maximum Entropy

After selecting a technology to be developed, from a reliability standpoint, the collection of information begins to estimate the updates of the failure events in each of the stages of project development. In this way, it is possible to compare the level of maturity in parallel to the decrease in the probability of failure. Before the development stages of the project itself, it is necessary to estimate the priori distribution of the failure events that make up the FT. This estimation can be based on expert opinions and data from previous equipment [Maior et al. 2022].

In FT, the basic events have their relative influence on the top events according to certain weights. These weights allow the construction of relationships between the top and bottom events used to estimate the reliability of equipment [Kabir; Sadiq; Tesfamariam 2016], [Ong et al. 2022].

The system's reliability can be represented by a multilevel reliability model (MRM), which comprises unique probabilistic models to characterize the failure of basic components, due to the complexity of analyzing a multilevel hierarchical structure [Maior et al. 2022].

Parametric reliability distributions, such as Bernoulli, Exponential, Weibull, and others, can be used as the MRM. They can also be based on physics, such as Arrhenius-Weibull. The problem of this research project is based on the FT logic gates, where each event “ i ” has parameters (θ_i) that represent the probability of occurrence of the relevant failure mechanism, according to [Maior et al. 2022].

The reliability function of the system as a whole is a function of (θ_i), as follows:

$$f[R_i(t|\theta_i)], \forall i \quad (2)$$

This equation represents the various paths that can lead to equipment failure, such as during installation or throughout the course of a mission. [Maior et al. 2022] dealt with two types of methods depending on the distribution. The MM was used for the discrete distribution cases (success/failure event) and the ME for the continuous distribution cases.

The maximum entropy method is the core of this work for parameter optimization. The occurrence of an event “ i ” is described by a continuous distribution. This method involves maximizing the entropy measure H (Eq. 3), where $\pi(\theta_i)$ is the probability density function (PDF) for the parameters of event i and θ_i is the parameter space of θ_i . The constraints (Eq. 4 - Eq. 6) are related to the percentiles of the reliability function, $R(t)$. Here, we consider that the expected value and the 5th and 95% percentiles of each event i must equal the expected value, 5th and 95th percentiles of the reliability R_t of the top event weighted (exponentially) by its specific contribution [Lei et al. 2010], [Maior et al. 2022], [Wang et al. 2018].

$$\text{Maximize } H = \int_{\Theta} -\pi(\theta_i) * \log[\pi(\theta_i)] d\theta_i \quad (3)$$

$$E[R_i(t|\theta_i)] = E[R_T(t)^{w'_i}] \quad (4)$$

$$P_5[R_i(t|\theta_i)] = P_5[R_T(t)^{w'_i}] \quad (5)$$

$$P_{95}[R_i(t|\theta_i)] = P_{95}[R_T(t)^{w'_i}] \quad (6)$$

Then, for each event, solve Eq. 3 subject to constraints 4 to 6 to define the model parameters. It is worth noting that analytically solving such an optimization problem is not straightforward, and it gets more difficult as the number of hyperparameters grows. Thus, to improve classical methods and efficiently obtain parameters' estimates, it is necessary to investigate different optimization methods and identify possible modifications, such as adding

local search processes or quantum increments. From this, the probability of failure will be updated at each development milestone. These updates are beyond the scope of this article.

4. Materials and Methods

A useful method for assessing the status of knowledge in regard to a particular issue is a systematic literature review. It differs from a traditional review in that it is accomplished using a repeatable, scientific, and transparent procedure, eliminating the possibility of bias and lacking in-depth analysis [Forcina et al. 2020]. For conducting systematic literature reviews, various methods are utilized [Xiao and Watson 2019]. In this paper the review is conducted based in three main steps: (1) planning the review; (2) gathering and choosing articles; and, (3) reporting).

In the step (1) Web of Science (WoS) was the database chosen. It combines numerous catalogues into a single search engine [Wang and Waltman 2016]. We chose this base because it is the oldest, most popular, and reliable database of research publications and citations in the world [Birkle et al. 2020]. However, there are other options that could be used such as Scopus and IEEE. The second step consists on (2) gathering and choosing articles. The language of publication was one criterion for filtering the articles. Only English-language publications were chosen. The investigation was undertaken from 1945 to 2022, that is the WoS default. Only publications from high-impact journals were selected when the type of articles was filtered.

Also, in (2), the main keywords were also assumed to conduct the study. A large group of more generic keywords were combined. These terms are: "Maximum entropy", "Prior" and "Bayesian". In this group, 203 articles were found. Restricting the scope of the search for our main focus, we added in separate searches "Reliability", "Oil and Gas", "R&D", and "Equipment", which returned 13, 1, 3 and 4 articles, respectively. This restricted our analysis to 21 papers. Other combinations were tested, but did not return significant results. Finally, after a thorough reading of the articles, our sample was reduced to a number of 5 papers that are somehow connected with our research question. Table 1 summarizes the combination of keywords, as well as, the number of papers resulting from each of the filtering performed.

Table 1 - Keywords Combination on Step (2)

STEP (2) PROCEDURE	Number of papers
Filter 1: "Maximum entropy" AND "Prior" AND "Bayesian"	203
Filter 2: "Maximum entropy" AND "Prior" AND "Bayesian" AND ("Reliability" OR "Equipment" OR "R&D")	19
Filter 3: Complete Reading	5

Search: Authors (2022)

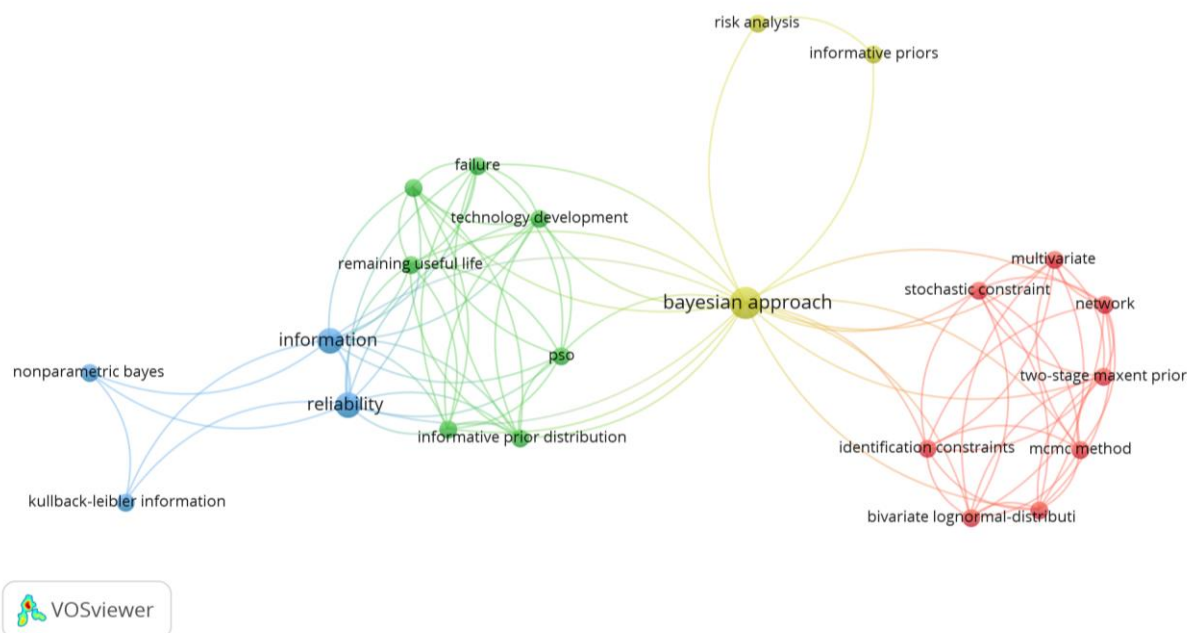
Finally, step (3) consists in the presentation of the results obtained through the review, amongst which, descriptive analyses, discussions and gaps found. These are presented in this article.

4. Results

The VOSviewer®, a free freeware program, was used to do the descriptive analysis of the papers. Network maps can be used to link the keywords and important terms found in the article portfolio [Jan and Ludo 2010]. As a result, Figure 1 provides an overview of the main aspects of the publications. We can identify that, as expected, the "Bayesian approach" is at the center of the keywords, being the main point that correlates the articles. "Information" and "Reliability" are also terms that have larger spheres, being, therefore, the focus of the articles. The cluster in green presents terms quite congruent with our research theme. That is, it is related to the development of equipment. In this network we can already highlight an optimization method, which is the Particle

Swarm Optimization (PSO). We can also see that an analyzed distribution in this sample is the lognormal (red cluster).

Figure 1 – Articles Keywords co-occurrence network visualization.



Search: Authors (2022).

Table 2 below shows a summary of the papers found. One can see that most of the papers have no description of the optimization method used. Note that the works do not focus on the optimization method itself, but on the modeling of the method to obtain the Bayesian a priori distribution. Furthermore, applications for some specific equipment can be seen, such as an equipment form an automobile plant [Kim 2017], a low-voltage switchgear [Wang et al. 2018], and an oil well packer [Maier et al. 2022]. However, there are more theoretical and generalist applications [Guikema 2007], [Mazzuchi et al., 2000]. Among the continuous distributions popular in the reliability field, there are the Weibull, Log-normal, and Exponential distributions for parameter estimation.

Table 2 – Papers resume

Paper	Optimization Method	Application	Distribution
[Maier et al. 2022]	PSO	Oil Well Equipment	Weibull, Arrhenius-Weibull, Exponential
[Kim 2017]	No description	Equipment from Automobile plant	Log-normal
[Guikema 2007]	Fmincon from MATLAB	Generic Application	Binomial
[Wang et al. 2018]	No description	Low-voltage switchgear	Exponential, Normal, Weibull
[Mazzuchi et al., 2000]	No description	Generic Application	Weibull, Exponential, Pareto, Gamma, Log-normal

Search: Authors (2022).

Only one study focused on the area of equipment development in the O&G industry. [Maior et al. 2022] presented a procedure for defining the Bayesian prior distribution $\pi_0(\theta|E)$ based on experts' knowledge and generic data that does not require direct elicitation of parameters to define informative distribution for Fault Trees (FT) basic events (component level), but instead relies on expert opinion (the system level). The proposed methodology is appropriate for equipment that is constantly in use. For the optimization of the hyperparameters he used the metaheuristic PSO. The drawbacks of the PSO method are that it is easy to fall into a local optimum in a high-dimensional space and that the iterative process has a low convergence rate [BHATTACHARYEE, 2021].

As a main gap, it is perceived that the O&G area needs more studies that focus on alternatives for the optimization of these parameters from the maximum entropy method. In terms of processing, depending on the distribution, this process can take a considerably large computing time. Thus, justifying the search for alternatives that are more efficient than the PSO.

5. Conclusions

This paper aimed to perform a literature review to identify the main optimization methods used for solving ME in the Bayesian context and in the O&G industry. Given the discussions and gaps found, it is possible to identify a research agenda to be carried out in a practical way to the O&G industry:

- to investigate different metaheuristic methods to be applied in this context, evolutionary algorithms such as Genetic Algorithm and Differential Evolution;
- to apply quantum-inspired algorithms that can speed up the process of parameter estimation;
- to mix different methods and techniques of local search that improve the traditional methods;
- to test the parameter estimates of different probability distributions of different equipment.

In conclusion, there are a variety of research options. Notably, extending this research area will make it possible for managers, reliability engineers, maintenance experts, and other decision-makers to handle these challenges using more computationally effective techniques. We highlight that our study's limitation is that not all approaches found in the literature were covered in detail. In order to learn more about them, one must consult the references. Additionally, it is essential to broaden the search to include other databases like Scopus. Despite the aforementioned drawbacks, this study can fill in the gaps that have already been discovered, produce new ideas for reviews and empirical studies, and offer pertinent information about the issue.

Acknowledgements

The authors thank the Brazilian National Agency for Research (CNPq) and the Humans Resource Program (PRH 38.1) managed by Brazilian Agency for Petroleum, Natural Gas and Biofuels (ANP) and Brazilian Funding Authority for Studies and Projects (FINEP) for the financial support through research grants. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

- Bhardwaj, U., Teixeira, A. P., & Soares, C. G. (2022). Bayesian framework for reliability prediction of subsea processing systems accounting for influencing factors uncertainty. *Reliability Engineering and System Safety*, 218(PA), 108143.
<https://doi.org/10.1016/j.ress.2021.108143>

- Bhattacharyee, N. (2021). Development of a blended particle swarm optimization to optimize mission design life of a series – parallel reliable system with time dependent component reliabilities in imprecise environments. *Soft Computing*, 25(17), 11745–11761. <https://doi.org/10.1007/s00500-021-05942-8>
- Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1), 363–376. https://doi.org/10.1162/qss_a_00018
- Cai, B., Liu, Y., Liu, Z., Tian, X., Dong, X., & Yu, S. (2012). Using Bayesian networks in reliability evaluation for subsea blowout preventer control system. *Reliability Engineering and System Safety*, 108, 32–41. <https://doi.org/10.1016/j.res.2012.07.006>
- Fernández, A. J., & Pérez-González, C. J. (2012). Generalized beta prior models on fraction defective in reliability test planning. *Journal of Computational and Applied Mathematics*, 236(13), 3147–3159. <https://doi.org/10.1016/j.cam.2012.02.012>
- Forcina, A., Silvestri, L., Di Bona, G., & Silvestri, A. (2020). Reliability allocation methods: A systematic literature review. *Quality and Reliability Engineering International*, 36(6), 2085–2107. <https://doi.org/10.1002/qre.2675>
- Guikema, S. D. (2007). Formulating informative, data-based priors for failure probability estimation in reliability analysis. *Reliability Engineering and System Safety*, 92(4), 490–502. <https://doi.org/10.1016/j.res.2006.01.002>
- Jan, N., & Ludo, V. E. (2010). Software survey : VOSviewer , a computer program for bibliometric mapping. *Scientometrics*, 84, 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Kabir, G., Sadiq, R., & Tesfamariam, S. (2016). A fuzzy Bayesian belief network for safety assessment of oil and gas pipelines. *Structure and Infrastructure Engineering*, 12(8), 874–889. <https://doi.org/10.1080/15732479.2015.1053093>
- Kim, H. J. (2017). Bayesian hierarchical scale mixtures of log-normal models for inference in reliability with stochastic constraint. *Entropy*, 19(6). <https://doi.org/10.3390/e19060274>
- Lei, X., Jin, M., & Wang, Q. (2010). Application of bayesian decision theory based on prior information in the multi-objective optimization problem. *International Journal of Computational Intelligence Systems*, 3(December), 31–42. <https://doi.org/10.1080/18756891.2010.9727751>
- Maior, C. B. S., Macêdo, J. B., Lins, I. D., Moura, M. C., Azevedo, R. V., Santana, J. M. M., da Silva, M. F., & Magalhães, M. V. C. (2022). Bayesian prior distribution based on generic data and experts' opinion: A case study in the O&G industry. *Journal of Petroleum Science and Engineering*, 210(October 2021). <https://doi.org/10.1016/j.petrol.2021.109891>
- Mazzuchi, T. A., Soofi, E. S., & Soyer, R. (2000). Computation of maximum entropy Dirichlet for modeling lifetime data. *Computational Statistics and Data Analysis*, 32(3–4), 361–378. [https://doi.org/10.1016/S0167-9473\(99\)00090-0](https://doi.org/10.1016/S0167-9473(99)00090-0)

- Mohd Nizam Ong, N. A. F., Sadiq, M. A., Md Said, M. S., Jomaas, G., Mohd Tohir, M. Z., & Kristensen, J. S. (2022). Fault tree analysis of fires on rooftops with photovoltaic systems. *Journal of Building Engineering*, 46(November 2021), 103752. <https://doi.org/10.1016/j.job.2021.103752>
- Saraygord Afshari, S., Enayatollahi, F., Xu, X., & Liang, X. (2022). Machine learning-based methods in structural reliability analysis: A review. *Reliability Engineering and System Safety*, 219(August 2021), 108223. <https://doi.org/10.1016/j.ress.2021.108223>
- Shafiee, M., Elusakin, T., & Enjema, E. (2020). Subsea blowout preventer (BOP): Design, reliability, testing, deployment, and operation and maintenance challenges. *Journal of Loss Prevention in the Process Industries*, 66(July 2019), 104170. <https://doi.org/10.1016/j.jlp.2020.104170>
- Wang, J. Q., Zhang, Z. G., Wang, C. H., & Wang, L. (2018). A Maximum Entropy Multisource Information Fusion Method to Evaluate the MTBF of Low-Voltage Switchgear. *Discrete Dynamics in Nature and Society*, 2018. <https://doi.org/10.1155/2018/2746871>
- Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, 10, 347–364.
- Wu, W. S., Yang, C. F., Chang, J. C., Château, P. A., & Chang, Y. C. (2015). Risk assessment by integrating interpretive structural modeling and Bayesian network, case of offshore pipeline project. *Reliability Engineering and System Safety*, 142, 515–524. <https://doi.org/10.1016/j.ress.2015.06.013>
- Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723971>
- Yaez, M., Joglar, F., & Modarres, M. (2002). Generalized renewal process for analysis of repairable systems with limited failure experience. *Reliability Engineering and System Safety*, 77(2), 167–180. [https://doi.org/10.1016/S0951-8320\(02\)00044-3](https://doi.org/10.1016/S0951-8320(02)00044-3)
- Zhong, S., Ye, X., Wang, M., Mei, X., Song, D., & Wang, W. (2022). An Integrated Framework for Mapping Nationwide Daily Temperature in China. *Advances in Meteorology*, 2022, 1–15. <https://doi.org/10.1155/2022/9895576>