Probabilistic Estimation of the Quality-of-Service Indexes in Distribution Networks

José Pedro T.S. Branco
DEEC
FEUP
Portugal
up201704179@edu.fe.up.pt

Pedro Macedo
Centre for Power and Energy Systems
INESC TEC
Portugal
pedro.m.macedo@inesctec.pt

J. Nuno Fidalgo
Centre for Power and Energy Systems
INESC TEC and FEUP
Portugal
jfidalgo@inesctec.pt

Abstract— **Ensuring reliable and high-quality electricity service is critical for consumers and Distribution System Operators (DSO). The DSO's Plan for Development and Investment in the Distribution Network (PDIDN) plays a pivotal role in enhancing network reliability and resilience while balancing technical and financial aspects. This study proposes a novel probabilistic approach for quality-of-service (QoS) estimation in distribution systems, addressing the limitations of traditional deterministic methods. Leveraging Bayesian regression, specifically the Spike and Slab technique, the model incorporates prior knowledge to improve the prediction of key QoS indicators such as SAIDI, SAIFI, and TIEPI. Using historical network data, the model demonstrates superior predictive accuracy and robustness, offering realistic confidence intervals for strategic planning. This method enables informed investments, enhances regulatory compliance, and supports renewable integration. The findings underline the potential of probabilistic modeling in advancing QoS forecasting, encouraging its application in other areas of electric network management.**

Index Terms-- **Quality-of-service, Distribution systems planning, Electricity market, Tariffs, Probabilistic modelling.**

I. Introduction

As electricity is an essential resource, ensuring reliable and high-quality service is a priority. The responsibility for QoS lies with the Distribution System Operator (DSO), who must comply with regulatory standards and regularly prepare a Plan for Development and Investment in the Distribution Network (PDIDN). This plan, which addresses both technical and financial aspects, is crucial for guiding investments and planning to enhance network reliability and resilience. PDIDN impacts electricity consumers in two ways: (a) the quality of service directly affects end-users of the distribution system, and (b) a portion of the investment costs is factored into the tariff paid by consumers.

The assessment of Quality of Service (QoS) in these systems has traditionally relied on deterministic techniques. While such methods are useful for providing baseline estimates of key reliability and service quality metrics, they often fail to capture the inherently stochastic nature of service interruptions and other random events affecting distribution network performance. This limitation can lead to over-simplified models that inadequately reflect real-world conditions, potentially resulting in suboptimal planning and decision-making.

In this context, the current research aims to improve QoS estimations by adopting a probabilistic approach. Probabilistic methods, particularly Bayesian regression, are well-suited to account for uncertainty and variability in system behavior. The proposed methodology utilizes the Spike and Slab technique, which allows for the incorporation of prior knowledge about variables [1][2]. This approach provides several key advantages: it generates more realistic confidence intervals, avoids the creation of models with limited physical significance, and enables a robust characterization of uncertainties. These attributes make the method particularly valuable for DSO tasked with planning, maintenance, and investment decisions under conditions of uncertainty.

The research employs historical data from Portuguese distribution network, including information on investment trends, maintenance activities, and consumption patterns. By processing and normalizing these datasets, the study develops a Bayesian regression model capable of accurately predicting key QoS indicators such as TIEPI, SAIDI, and SAIFI. The probabilistic nature of the model not only enhances prediction accuracy but also provides actionable insights by quantifying the uncertainty associated with each estimate.

This work makes a significant contribution by addressing critical gaps in existing approaches to QoS assessment. Traditional deterministic methods often fail to account for the complexities introduced by diverse operational conditions and stochastic variables. In contrast, the probabilistic framework developed in this study equips decision-makers with a more nuanced understanding of system behavior, supporting better-aligned investment strategies, enhanced regulatory compliance, and improved service quality for end-users.

Moreover, this study demonstrates the practical applicability of the proposed methodology through robust validation using historical data. Multiple testing configurations, including iterative training and validation scenarios, confirm the model's reliability and adaptability. The results underline the potential of probabilistic approaches to

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project LA/P/0063/2020.
DOI 10.54499/LA/P/0063/2020 | https://doi.org/10.54499/LA/P/0063/2020”).

revolutionize QoS assessment and guide the development of resilient and efficient distribution networks.

II. LITERATURE REVIEW

The scientific literature on the specific topic of this paper is limited. However, numerous studies emphasize the importance of power quality and reliability in electrical distribution systems. For instance, [3] adopts a customer-focused approach to guide investment decisions from this perspective. The monitoring and analysis of power quality indices, such as harmonics, voltage fluctuations, and imbalances, provide critical insights into system performance and help address customer issues [4]. The work reported in [5] emphasizes the key aspects of distribution system planning include load forecasting, transformer applications, and substation design. This reference also analyses the role of mathematical modelling in reliability assessments, aiming at evaluating both performance and economic considerations. The paper [6] presents a methodology to calculate a unified index for assessing power quality in power systems with distributed generation. Prettico et al. propose a data-driven model to assess the SAIDI indices in several European countries [7]. The regression performance was good, but it cannot be applied to forecast the future SAIDI as a consequence of QoS investments.

Regarding the probabilistic estimation of the QoS, articles such as [8]-[11] mainly deal with specific distribution networks (not with the whole distribution network) and local details of a particular indicator such as voltage sags or the number of interruptions. Paper [8] developed a methodology to estimate annual voltage sags and momentary interruptions by accounting for equipment sensitivity and recloser operations. Article [9] proposed a probabilistic framework for state estimation and observability assessment, emphasizing estimation accuracy. Krahl et al. [10] presented a method for deriving probability distributions of reliability indices in medium-voltage networks, supporting the evaluation of regulatory quality approaches. The focus of [11] is on power quality in networks with distributed generation, introducing probabilistic indices to address multiple disturbances concurrently.

The review of existing literature emphasizes the significance of accurately predicting and managing Quality-of-Service (QoS) indices within electrical distribution networks, particularly given their impact on investment decisions, consumer satisfaction, and regulatory compliance. Although deterministic approaches have provided valuable baseline assessments, they fail to fully capture the uncertainties inherent in system behavior, thereby limiting their practical utility. The scarcity of comprehensive probabilistic analyses and the limited attention given to integrating prior knowledge underscore a critical research gap. Consequently, the probabilistic approach proposed in this study, leveraging Bayesian regression with the Spike and Slab technique, emerges as highly pertinent. This approach not only addresses existing methodological limitations but also advances the capability of Distribution System Operators (DSOs) to make informed, reliable, and cost-effective decisions in an increasingly complex and uncertain operating environment.

III. METHODOLOGY

The research focuses on estimating the QoS indicators for electrical distribution networks using a probabilistic approach. The methodology is designed to overcome the limitations of deterministic models by incorporating Bayesian regression techniques, specifically the Spike and Slab method, to model and predict QoS indicators with enhanced accuracy and reliability.

A. Bayesian Regression

A Bayesian approach to regression relies on probability distributions rather than single point estimates. As such, the Bayesian response is expressed as a probability distribution.

$$\hat{y} \sim \mathrm{N}(\beta^T X, \sigma^2 I) \tag{1}$$

The response y is modelled as a normal (Gaussian) distribution with mean $\mu = \beta^T X$ and standard deviation σ. Here, I represents the identity matrix.

The core equation of the Bayesian regression is

$$\mathrm{P}\big(\beta \backslash (y, X)\big) = \frac{\mathrm{P}\big(y \backslash (\beta, X)\big) \times \mathrm{P}(\beta \backslash X)}{\mathrm{P}(y \backslash X)} \tag{2}$$

Where:

- $\mathrm{P}\big(\beta \backslash (y, X)\big)$ – Posterior probability distribution of the model parameters based on training data (inputs vs outputs).
- $\mathrm{P}(y \backslash \beta, X)$ – Likelihood of data, reflecting the regression quality for current parameters.
- $\mathrm{P}(\beta \backslash X)$ – prior probability of the parameters; in the case of model unawareness, a non-informative prior (uniform distribution) might be used.
- $\mathrm{P}(y \backslash X)$ – Normalization constant and represents the marginal likelihood (or evidence), i.e. the probability of the observed data given the model.

In scenarios with limited training data, the posterior distribution becomes more diffuse. As training data increases, the likelihood dominates the prior, and parameters converge to values estimated via Ordinary Least Squares (OLS).

The “Spike and slab” technique defines priors in the Bayesian linear regression. The "spike" represents the probability that a specific coefficient equals zero, while the "slab" defines the coefficient's value distribution (Fig. 1). For instance:

- Variables associated with "Spike" tend to be excluded due to a narrow distribution around zero.
- Variables linked to "Slab" have broader variance, allowing flexibility for significant coefficients.

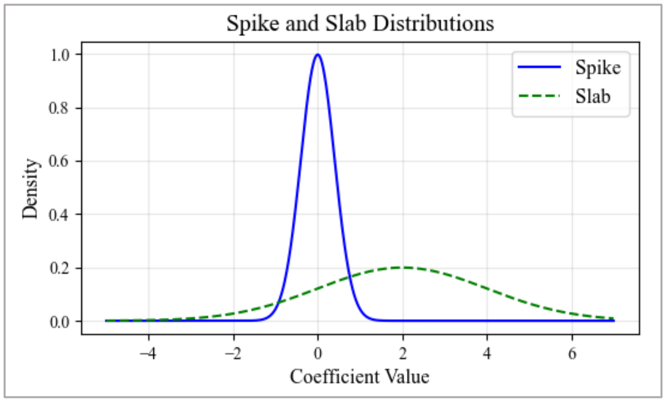

Figure 1. Spike and slab examples.

In this example, the variable associated with the Spike will most probably be removed from the model, as its coefficient is centered in zero and the distribution is quite narrow. In this example, the variable associated with the Slab is centered in 2 and has a large variance, so the search space will be wider for this variable, although centered in 2.

Since posterior evaluation for continuous variables is computationally intensive, sampling methods like Markov Chain Monte Carlo (MCMC) are used to approximate distributions. After numerous iterations (10,000–100,000 cycles), parameters are characterized as Gaussian distributions.

This probabilistic framework quantifies estimation uncertainty. For example, a predicted value might lie between 70 and 80 with a 95% confidence interval.

Modeling begins with a prior, informed by expert knowledge or initial assumptions. The Bayesian iterative process progressively integrates the available information, progressively aligning the model with observed data, validating or challenging initial beliefs. This adaptive process mirrors how humans refine their understanding over time.

B. *Key methodological steps*

1) *Data collection*

The data below was provided by the Portuguese DSO (E-Redes) in the context of an investment impact study. Except for the QoS indices, all other data is sensitive and not publicly available.

- Available historical yearly data (2008-2020) from the distribution network:
  - Investment, maintenance, and consumption.
  - QoS indicators such as TIEPI, SAIDI, and SAIFI.
- Projections for the period 2021-2027 of:
  - Consumption evolution
  - Plan of investment and maintenance expenditures (sometimes, a few scenarios are considered)

2) *Bayesian regression*

- The Spike and Slab method was employed for variable selection, enabling the model to prioritize relevant predictors while discarding insignificant ones.
- Probabilistic priors were defined for predictors (e.g., investment and maintenance) to align model outputs with realistic physical constraints.

3) *Model optimization*

- The research adopted an iterative approach using the Bayesian regression model (BRM) in the R programming environment.
- Models were evaluated based on metrics such as RMSE and NRMSE, ensuring robust predictions.

4) *Confidence Interval estimation*

- A key feature was generating 95% confidence intervals for the predictions, enhancing decision-making for planning and investment.

5) *Scenario testing*

- Various configurations were tested, including normalization of input variables and adjustments to priors.
- Models were validated against different training and testing datasets to ensure reliability.

The limited amount of data examples is challenging as it constrains the model's types and raises questions on models' generalization. That is why several splitting alternatives were considered to divide the data into train and test sets.

C. *Base model*

The present approach assumes a linear starting model, which can be modified to include non-linear characteristics. The equation (3) describes the base model.

$$Q_i = k_0 + k_{Invest} \times Invest_{i-1} + k_E \times E_i + \\ + k_M \times M_{i-1} + k_Q \times Q_{i-1} \quad (3)$$

Where:

- i – year
- Q – QoS index (e.g. TIEPI)
- $Invest$ – Investment in QoS
- E – Energy consumption (usually the percentual consumption variation from one year to the next)
- M – Maintenance budget

Note that some variables are referred to year *i* while others to year *i-1*. Some variations on this model include considering the variable absolute value, its variation with respect to the previous year or even some transformation. For instance, the investment term is often replaced by

$$k_{Invest} \times log\left(Invest_{i-1} + k_{Invest}^{aten}\right) \quad (4)$$

The logarithmic function is applied to represent the progressively higher cost of QoS improvement. For example, if the investment to reduce 1 min of TIEPI is Inv1, then the investment to reduce 2 min will be higher than 2×Inv1. The term k_{Invest}^{aten} is to prevent the case of (log(0) and to moderate the *Invest* variations impact.

D. Priors setting

Although optional in Bayesian approaches, the inclusion of priors is often crucial to obtain a good performance model, able at the same time to adhere to reality (with physical meaning). Priors can be used here to induce a set of constraints or tendencies to direct the iterative process to the right path.

A few prior examples:

- k_{Invest} should be negative – The investment is supposed to improve the QoS indices. For instance, TIEPI should decrease with higher investments.
- k_M should be negative for the same reasons of the previous coefficient. The maintenance expenditure is expected to improve the QoS performance.
- k_Q represents how the QoS index of a year (e.g. $TIEPI_i$) depends on the QoS index of the previous year (e.g. $TIEPI_{i-1}$). It is expected to be a value below but not to distant from 1. In this case, the following prior was considered a distribution N(0.8;0.1), as illustrated in Fig. 2.

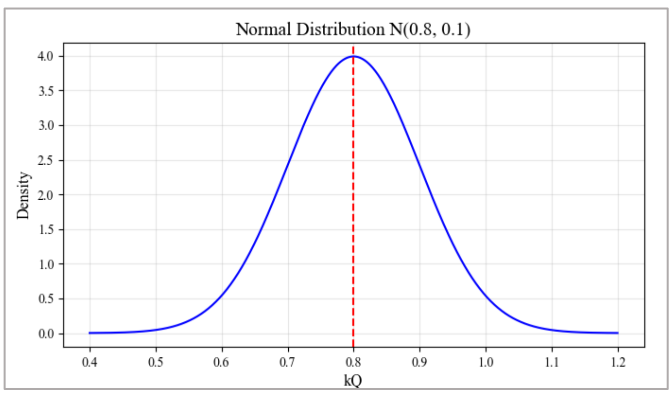

Figure 2. Prior for the coefficient k_Q

IV. RESULTS

Figs. 3 and 4 highlight the importance of incorporating priors by demonstrating the potential problems that can arise when they are omitted. Fig. 3 presents the 50 best performant models, obtained without priors' definition i.e., the models are purely data-driven. The thick black line contains the real TIEPI instances. Fig. 4 shows the final model and the corresponding 95% confidence interval (CI). Both figures show unrealistic results. On the contrary, the consideration of priors in Figs. 5 resulted in workable outcomes.

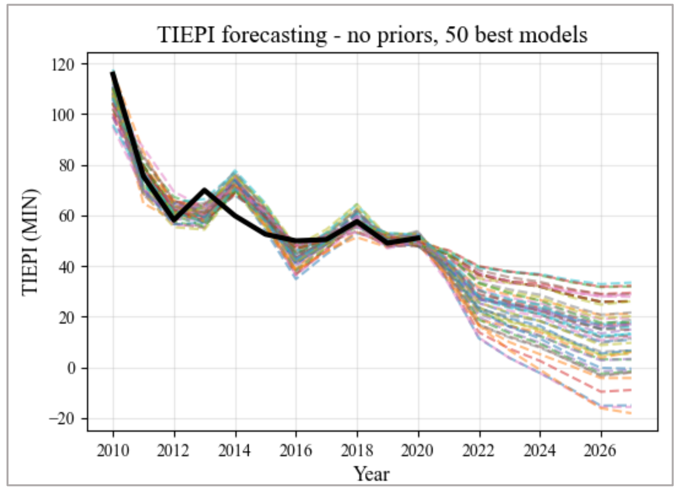

Figure 3. TIEPI estimation (multiple models) with no priors

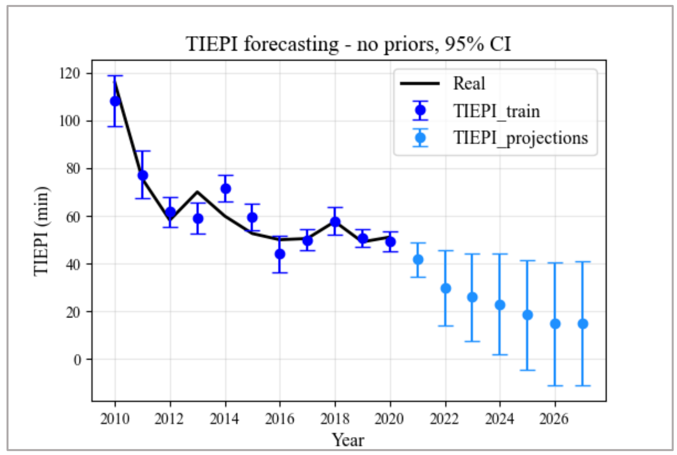

Figure 4. TIEPI estimation with no priors, 95% CI

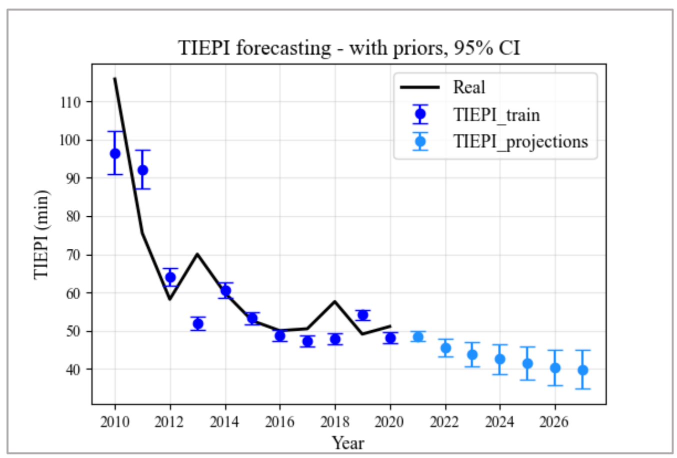

Figure 5. TIEPI estimation with priors, 95% CI

By integrating prior knowledge into the Bayesian model generation process, the results achieved improved alignment with reality. The final model was constructed by averaging the 100 best-performing individual models. Fig. 5 summarizes the performance of the final probabilistic model.

As anticipated, the error margin increases over time due to the naturally greater uncertainty associated with future predictions.

For each predicted year, the probabilistic regression generates a corresponding Probability Density Function (PDF), which not only provides the expected TIEPI value but also quantifies the prediction's uncertainty and confidence. Fig. 6 illustrates the TIEPI PDF for the years 2021 through 2027.

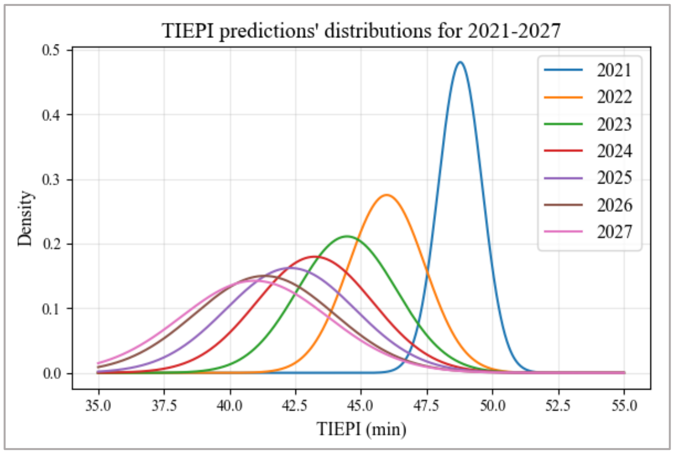

Figure 6. Estimated TIEPI PDF for 2021-2027

It is noteworthy that the remaining indices (SAIDI and SAIFI) exhibited trends consistent with those observed for the TIEPI index, both regarding their central estimates and narrower confidence intervals. This consistency indicates an improvement not only in the indices themselves but also in the reliability of their associated estimates.

V. Conclusion

This study presented a novel probabilistic approach to estimating Quality of Service (QoS) indicators in electrical distribution networks, addressing the inherent limitations of deterministic methods. By leveraging Bayesian regression with the Spike and Slab technique, the proposed methodology incorporated prior knowledge and accounted for uncertainties, resulting in enhanced prediction accuracy and reliability.

The integration of priors allowed the model to align more closely with real-world conditions, as demonstrated by improved estimations of QoS indicators such as TIEPI, SAIDI, and SAIFI. Additionally, the probabilistic framework provided realistic confidence intervals, supporting informed decision-making for investment planning and maintenance strategies.

The results highlighted the model's capability to adapt to historical data trends while accounting for the inherent uncertainties associated with the inputs as well as future uncertainties. Scenario testing and robust validation confirmed the model's applicability and reliability in diverse contexts. Moreover, this approach offers a pathway for optimizing QoS investments, ensuring regulatory compliance, and integrating renewable energy into the grid efficiently.

In conclusion, this research underscores the potential of probabilistic modeling as a transformative tool for enhancing QoS assessment in distribution networks. Future work could extend this methodology to other performance indicators and explore its application in broader contexts, including dynamic network management and real-time decision-making.

References

[1] Malsiner-Walli, G., & Wagner, H. (2016). Comparing Spike and Slab Priors for Bayesian Variable Selection. *Austrian Journal of Statistics*, *40*(4), 241–264. https://doi.org/10.17713/ajs.v40i4.215

[2] Hemant Ishwaran. J. Sunil Rao. "Spike and slab variable selection: Frequentist and Bayesian strategies." Ann. Statist. 33 (2) 730 - 773, April 2005. https://doi.org/10.1214/009053604000001147

[3] M. F. McGranaghan, "Quantifying reliability and service quality for distribution systems," 2006 IEEE Rural Electric Power Conference, Albuquerque, NM, USA, 2006, pp. 1-8, doi: 10.1109/REPCON.2006.1649040.

[4] E. A. Mertens, L. F. S. Dias, F. A. Fernandes, B. D. Bonatto, J. P. G. Abreu and H. Arango, "Evaluation and trends of power quality indices in distribution system," 2007 9th International Conference on Electrical Power Quality and Utilisation, Barcelona, Spain, 2007, pp. 1-6, doi: 10.1109/EPQU.2007.4424212.

[5] Tutan Gonen, "Electric power distribution system engineering", CRC Press, Third Edition, ISBN-13978-1482207002 2014, https://api.semanticscholar.org/CorpusID:108734607

[6] M. Sacasqui, J. Luyo and A. Delgado, "A Unified Index for Power Quality Assessment in Distributed Generation Systems Using Grey Clustering and Entropy Weight," 2018 IEEE ANDESCON, Santiago de Cali, Colombia, 2018, pp. 1-4, doi: 10.1109/ANDESCON.2018.8564631.

[7] G. Prettico, A. Marinopoulos, S. Vitiello, 2022. "Guiding electricity distribution system investments to improve service quality: A European study," Utilities Policy, Elsevier, vol. 77(C).

[8] J. C. Cebrian, N. Kagan and J. V. Milanović, "Probabilistic Estimation of Distribution Network Performance With Respect to Voltage Sags and Interruptions Considering Network Protection Setting—Part I: The Methodology," in IEEE Transactions on Power Delivery, vol. 33, no. 1, pp. 42-51, Feb. 2018, doi: 10.1109/TPWRD.2016.2633518.

[9] Brinkmann, B. and Negnevitsky, M. (2020). "Probabilistic State Estimation In Distribution Networks", In Advances in Electric Power and Energy, M.E. El-Hawary (Ed.). https://doi.org/10.1002/9781119480402.ch10

[10] S. Krahl, S. Ohrem, and H.-J. Haubrich, "Probability distributions of reliability indices of electrical distribution networks", 20th International Conference and Exhibition on Electricity Distribution (CIRED 2009), https://doi.org/10.1049/cp.2009.0714

[11] A. Bracale, P. Caramia, G. Carpinelli, A. Russo and P. Verde, "Site and System Indices for Power-Quality Characterization of Distribution Networks With Distributed Generation," in IEEE Transactions on Power Delivery, vol. 26, no. 3, pp. 1304-1316, July 2011, doi: 10.1109/TPWRD.2011.2112381.