

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



Estimação Probabilística da Qualidade de Serviço da Rede de Distribuição

José Pedro Telo Sanches Branco

Mestrado em Engenharia Eletrotécnica e de Computadores

Orientador: José Nuno Moura Marques Fidalgo

17 de outubro de 2023

Reconhecimentos

A realização desta dissertação conta com contribuições notáveis não apenas do orientador professor doutor José Nuno Fidalgo, mas também do engenheiro Pedro Miguel Macedo, que desempenhou um papel crucial na orientação, desenvolvimento e implementação do trabalho.

Resumo

O Sistema Elétrico Nacional (SEN) corresponde a um sistema interligado que é constituído pela produção, transporte e distribuição de energia elétrica, bem como pelas instalações elétricas dos seus clientes.

A energia elétrica é, cada vez mais, um bem essencial para as populações. Como tal, esta, deve ser fornecida com padrões de qualidade adequados e, tanto quanto possível, sem interrupções.

A Qualidade de Serviço Técnica (QST) e o cumprimento dos seus padrões é fundamental para a garantia de um nível adequado de desempenho do SEN.

É da responsabilidade do Operador da Rede de Distribuição (ORD) assegurar a Qualidade de Serviço, que engloba a continuidade de serviço e a qualidade da energia elétrica, no fornecimento de energia às populações.

O ORD, de acordo com a regulamentação vigente do setor elétrico, depara-se com a necessidade da elaboração de um Plano de Desenvolvimento e Investimento da Rede de Distribuição (PDIRD) [1], plano esse que, deve ser projetado para um horizonte temporal de cinco anos e ser revisto a cada biénio. Este plano constitui um elemento de natureza dualista, ou seja, possui duas vertentes, uma técnica e outra financeira/orçamental, que concretiza e quantifica objetivamente a obrigação genérica de planeamento das redes de distribuição. Estas vertentes encontram-se, logicamente, interligadas entre si, sendo inseparáveis na medida em que constituem premissas e conclusões, de modo que o julgamento resultante da análise técnica ditará as ações a realizar na rede.

Apesar da natureza estocástica do problema, este exercício tem sido tradicionalmente realizado com base em técnicas pontuais (i.e. não probabilísticas).

Com a realização desta dissertação, pretende-se, como objetivo principal, estimar, de um modo essencialmente probabilístico, os indicadores da QST da rede de distribuição para os próximos anos. Estes indicadores irão ser estimados com base na evolução dos próprios indicadores, registo e caracterização dos incidentes e dos investimentos passados e previstos para os próximos anos.

De forma a automatizar o processo da realização de estimações irá ser desenvolvido um programa em ambiente R com base no conceito de Spike and Slab, método de natureza probabilística, o qual permite incorporar no modelo informações *a priori* sobre os preditores. Isto constitui uma vantagem fundamental, dado que as distribuições nas entradas passam a ser traduzidas de forma consistente na distribuição associada ao resultado. Por outro lado, permite evitar a geração de modelos sem significado físico.

A abordagem adotada permitiu obter estimativas com qualidade comparável dos métodos pontuais utilizados no projeto real. Além disso, possibilitou obter, de forma automática, as margens de confiança associadas às previsões.

Abstract

The National Electricity System corresponds to an interconnected system that comprises the generation, transmission and distribution of electric energy, as well as the electrical installations of its customers.

Electricity is becoming increasingly essential for populations. As such, it must be supplied with appropriate quality standards and, whenever possible, without interruptions.

The Technical Service Quality and compliance with its standards are fundamental to ensure an adequate level of National Electricity System performance.

It is the responsibility of the Distribution Network Operator to ensure Service Quality, which includes service continuity and electrical energy quality, in supplying energy to the populations.

The Distribution Network Operator, according to the current regulations of the electricity sector, faces the need to develop a Distribution Network Development and Investment Plan, a plan that should be designed for a five-year time horizon and be reviewed biennially. This plan consists of a dualistic element, meaning it has two aspects: a technical aspect and a financial/budgetary aspect, which objectively concretizes and quantifies the generic obligation of distribution network planning. These aspects are logically interconnected since they constitute premises and conclusions, and the judgment resulting from the technical analysis will dictate the actions to be taken on the network.

Despite the stochastic nature of the problem, this exercise has traditionally been performed based on deterministic techniques (e.g. non-probabilistic).

With the completion of this dissertation, the main objective is to estimate, in a predominantly probabilistic manner, the Technical Service Quality indicators of the distribution network for the upcoming years. These indicators will be estimated based on the evolution of the indicators themselves, the record and characterization of incidents, and the investments made in the past and planned for the upcoming years.

In order to automate the estimation process, a program will be developed using the R environment based on the concept of Spike and Slab, a probabilistic method that allows incorporating a priori information about predictors into the model. This constitutes a fundamental advantage, as it ensures that the distributions in the inputs are consistently translated into the distribution associated with the outcome. Moreover, it helps avoid generating models without physical significance.

The adopted approach allowed obtaining estimates of comparable quality to the deterministic methods used in the real project. Additionally, it enabled automatically obtaining the confidence intervals associated with the predictions.

Agradecimentos

Um agradecimento especial à minha família. Agradeço à minha mãe, Maria Teresa, e ao meu pai, José Francisco. Sou-lhes grato por todo o apoio e dedicação que me concedem, bem como por tudo o que me proporcionam, desde sempre, em todos os aspetos da minha vida pessoal e estudantil. Agradeço aos meus avós pelo amor incondicional, apoio e coragem que sempre me transmitiram.

Ao professor doutor José Nuno Fidalgo e ao engenheiro Pedro Miguel Macedo por todos os ensinamentos, reparos e instruções e pelo tempo disponibilizado para me ajudar a superar este desafio.

Aos meus amigos, de infância e aos que conheci nesta faculdade, por todos os momentos que vivámos.

A todos os mencionados, o meu mais sincero obrigado.

José Pedro Branco

“Success is walking from failure to failure with no loss of enthusiasm.”

Winston Churchill

Conteúdo

1	Introdução	1
1.1	Estrutura da Dissertação	2
2	Qualidade de Serviço Técnica	3
2.1	Definição da QST na distribuição de eletricidade	3
2.2	Causas e Impactos das Interrupções no Fornecimento de Energia Elétrica	4
2.3	Enquadramento Regulamentar da QST	5
2.4	Estratégias para melhoria da QST	8
2.5	Síntese	9
3	Regressão Probabilística com Técnicas Bayesianas	11
3.1	Técnicas Bayesianas	11
3.1.1	Equação Fundamental Bayesiana	12
3.1.2	Spike and Slab	13
3.1.3	Spike and Slab em ambiente R	14
4	Metodologia	17
4.1	Dados disponíveis	17
4.2	Qualidade dos modelos	19
4.3	Regressão linear para a estimação	20
4.4	Regressão linear para a projeção	21
4.5	Abordagem adotada	21
5	Resultados	23
5.1	Testes com a função ' <i>brm</i> '	25
5.1.1	Teste com variação do número de anos de teste	26
5.1.2	Teste com variação do número de iterações	33
5.1.3	Teste com variação no prior do consumo	36
5.1.4	Teste sem a utilização da função logarítmica	40
5.1.5	Teste sem a utilização das constantes normalizadoras presentes nos logarítmicos	41
5.1.6	Teste sem a utilização de <i>priors</i>	44
5.1.7	Modelo Final	46
6	Conclusões e Trabalho Futuro	49
A	Código desenvolvido	51
A.1	Código com a função <i>lm.spike</i> e a biblioteca <i>BoomSpikeSlab</i>	51
A.2	Código com a função <i>brm</i> e a biblioteca <i>brms</i> de obtenção do modelo final	53

Referências

57

Lista de Figuras

3.1	Representação da distribuição normal com vários intervalos de confiança [2] . . .	12
3.2	Distribuições Spike e Slab [3]	14
5.1	Evolução do TIEPI	24
5.2	Fluxograma do código desenvolvido	25
5.3	Intervalo de confiança a 95% do TIEPI sem anos de teste	28
5.4	Intervalo de confiança a 95% do TIEPI com 1 ano de teste	29
5.5	Intervalo de confiança a 95% do TIEPI com 2 anos de teste	30
5.6	Comparação da evolução das previsões e projeções para diferentes anos de teste .	32
5.7	Intervalo de confiança a 95% com 2 anos de teste e 2000 iterações	33
5.8	Intervalo de confiança a 95% com 2 anos de teste e 8000 iterações	34
5.9	Comparação da evolução das previsões e projeções para diferentes valores de iterações	36
5.10	Intervalo de confiança a 95% com 2 anos de teste, 4000 iterações e alteração no prior da variação do consumo	37
5.11	Comparação da evolução das previsões e projeções para diferentes <i>priors</i> no consumo	39
5.12	Intervalo de confiança a 95% com 2 anos de teste, 4000 iterações e sem a função logarítmica	40
5.13	Evolução do índice dos 100 melhores modelos	41
5.14	Intervalo de confiança a 95% com 2 anos de teste, 4000 iterações e sem as constantes normalizadoras	42
5.15	Comparação da evolução das previsões e projeções com e sem constantes normalizadoras	43
5.16	Intervalo de confiança a 95% com 2 anos de teste, 4000 iterações e sem o uso de <i>priors</i>	44
5.17	Evolução do índice dos 100 melhores modelos	45
5.18	Distribuição gaussiana dos 100 melhores modelos do coeficiente Intercept	46
5.19	Distribuição gaussiana dos 100 melhores modelos do coeficiente Investimento . .	47
5.20	Distribuição gaussiana dos 100 melhores modelos do coeficiente Variação do Consumo	47
5.21	Distribuição gaussiana dos 100 melhores modelos do coeficiente Manutenção . .	48
5.22	Distribuição gaussiana dos 100 melhores modelos do coeficiente TIEPI	48

Lista de Tabelas

4.1	Valores históricos dos índices (Inputs e Targets)	18
4.2	Valores históricos do investimento, manutenção e variação de consumo (Inputs) .	18
5.1	Output dos melhores 10 modelos para o TIEPI (min)	23
5.2	Modelo final sem anos de teste	27
5.3	Modelo final com 1 ano de teste	29
5.4	Modelo final com 2 anos de teste	30
5.5	Comparação dos coeficientes dos modelos finais	31
5.6	Evolução da previsão do TIEPI com número diferente de anos de teste	31
5.7	Evolução da projeção do TIEPI com número diferente de anos de teste	31
5.8	Modelo final com 2000 iterações e 2 anos de teste	33
5.9	Modelo final com 8000 iterações e 2 anos de teste	34
5.10	Comparação dos coeficientes dos modelos finais com diferente número de iterações	35
5.11	Evolução da previsão do valor do TIEPI com número diferente de iterações . . .	35
5.12	Projeção dos valores do TIEPI com variação no número de iterações	35
5.13	Modelo final obtido com alteração no prior do consumo	37
5.14	Comparação dos coeficientes dos modelos finais com a alteração do prior da vari- ação do consumo	38
5.15	Evolução da previsão do valor do TIEPI com diferentes <i>prior</i> no consumo	38
5.16	Projeção dos valores do TIEPI com diferentes <i>prior</i> no consumo	38
5.17	Modelo final sem a função logarítmica	40
5.18	Modelo final sem as constantes normalizadoras	42
5.19	Coeficientes dos modelos finais com a exclusão das constantes normalizadoras .	42
5.20	Evolução da previsão do valor do TIEPI com e sem constantes normalizadoras . .	43
5.21	Projeção dos valores do TIEPI com e sem constantes normalizadoras	43
5.22	Modelo final sem <i>priors</i>	44
5.23	Média e desvio-padrão dos 100 melhores modelos	46

Abreviaturas e Símbolos

ANS	Acesso a Novos Serviços
AT	Alta Tensão
brm	<i>Bayesian Regression Model</i>
brms	<i>Bayesian Regression Models using 'stan'</i>
ERSE	Entidade Reguladora dos Serviços Energéticos
END	Energia não distribuída
EO	Eficiência Operacional
ER	Eficiência da Rede
INESC TEC	Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência
IoT	<i>Internet of Things</i>
IQF	Índice de Qualidade de Fornecimento
MAIFI	<i>Momentary Average Interruption Frequency Index</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MSE	<i>Mean Squared Error</i>
MT	Média Tensão
MWh	Mega-watt hora
NRMSE	<i>Normalized Root Mean Square Error</i>
ORD	Operador da Rede de Distribuição
PDIRD	Plano de Desenvolvimento e Investimento da Rede de Distribuição
QST	Qualidade de Serviço Técnica
RMSA	Relatório de Monitorização da Segurança de Abastecimento
RMSE	<i>Root Mean Squared Error</i>
RND	Rede Nacional de Distribuição
RQS	Regulamento da Qualidade de Serviço
RRC	Regulamento de Relações Comerciais
SA	Segurança de Abastecimento
SAIDI	<i>System Average Interruption Duration Index</i>
SAIFI	<i>System Average Interruption Frequency Index</i>
SCADA	<i>Supervisory Control and Data Acquisition</i>
SEN	Sistema Elétrico Nacional
TIEPI	Tempo de interrupção equivalente da potência instalada
TMEF	Tempo Médio Entre Falhas
TMR	Tempo Médio de Restabelecimento

Capítulo 1

Introdução

Ao longo das últimas décadas, o mundo testemunhou um aumento significativo no consumo de energia elétrica em todas as esferas da sociedade. Esse crescimento progressivo está intrinsecamente ligado ao desenvolvimento tecnológico, ao aumento populacional, à expansão industrial e ao padrão de vida cada vez mais dependente de dispositivos alimentados por este tipo de energia. Embora a eletricidade seja uma das maiores conquistas da humanidade, o aumento exponencial na sua procura traz consigo uma série de desafios, a nível do ambiente, da economia, das infraestruturas e da qualidade dos serviços, que requerem abordagens sustentáveis e inteligentes.

Uma das principais razões para o aumento no consumo de energia elétrica é a proliferação dos dispositivos eletrônicos e a evolução das tecnologias digitais. Smartphones, computadores, tablets, eletrodomésticos inteligentes e a *Internet of Things (IoT)* tornaram-se parte essencial do dia a dia de grande parte da população mundial, e essa dependência crescente tem como resultado direto um consumo elétrico cada vez maior. Além disso, várias indústrias em áreas como os transportes e serviços têm migrado para soluções elétricas, contribuindo ainda mais para o escalar da procura energética.

Assumindo a eletricidade como uma necessidade e um bem essencial na vida moderna, devemos também assumir que a continuidade e a qualidade de serviço técnica no fornecimento de eletricidade são fundamentais para garantir o bem-estar da sociedade, o funcionamento eficiente da economia e a segurança dos cidadãos. Um fornecimento de energia elétrica fiável é crucial para manter os sistemas em funcionamento, evitar interrupções prejudiciais e promover o desenvolvimento socioeconómico de forma sustentável.

Torna-se assim imperativo investir na rede de distribuição, ponte de ligação entre a rede de transporte de energia e as cargas, de modo a assegurar o contínuo e correto funcionamento da mesma.

Com isto, tem-se que o principal objetivo da realização desta dissertação é estimar, de uma forma probabilística, os indicadores da QST da rede de distribuição, de modo a definir ações de manutenção e novos investimentos.

1.1 Estrutura da Dissertação

Este documento encontra-se dividido em seis capítulos, sendo que o primeiro realiza uma introdução ao tema da dissertação.

O segundo capítulo, 2, engloba uma revisão bibliográfica que permitirá posicionar a narrativa ora dissertada relativamente a documentos da especialidade anteriores, que servem de base e de guia ao estudo que nos ocupa. Adicionalmente, visa a descrição de um de dois assuntos essenciais para o desenvolvimento desta dissertação:

- Enquadramento regulamentar da QST - descreve o conceito da QST à luz dos regulamentos, refere mecanismos de incentivo à sua melhoria e apresenta os principais indicadores usados para a caracterizar;

No capítulo 3, será abordado o segundo tema essencial ao desenvolvimento desta dissertação:

- Abordagens Bayesianas - descreve os conceitos fundamentais associados às técnicas Bayesianas, com particular ênfase na Spike and Slab.

No capítulo 4, será apresentada a descrição da metodologia adotada para realizar o estudo proposto.

Já no capítulo 5, serão apresentados e discutidos os principais resultados obtidos através dos estudos desenvolvidos.

Por fim, no capítulo 6 apresentam-se as principais conclusões retiradas do trabalho desenvolvido e expõem-se alguns trabalhos futuros.

Capítulo 2

Qualidade de Serviço Técnica

2.1 Definição da QST na distribuição de eletricidade

A QST na distribuição de eletricidade refere-se à capacidade das distribuidoras de energia elétrica em fornecer eletricidade de forma confiável, segura e eficiente aos consumidores. Para avaliar e monitorizar a qualidade desse serviço, são utilizadas diversas métricas que medem diferentes aspetos do desempenho da distribuição elétrica. Abaixo encontram-se, a título de exemplo, algumas das principais definições e métricas da QST na distribuição de eletricidade:

- Tempo Médio de Restabelecimento (TMR) : O TMR é uma métrica que mede o tempo médio necessário para restabelecer o fornecimento de energia após uma interrupção. É uma medida importante, pois reflete a eficiência e agilidade da empresa em resolver problemas e minimizar os impactos das interrupções.
- Tempo Médio Entre Falhas (TMEF) : O TMEF é uma métrica que mede o tempo médio decorrido entre as falhas no sistema de distribuição, ou seja, a média de tempo entre duas interrupções consecutivas.
- Índice de Qualidade de Fornecimento (IQF) : O IQF é uma métrica que avalia a qualidade do serviço técnico considerando não apenas as interrupções, mas também outras características do fornecimento de energia, como flutuações de tensão e variações de frequência.

É importante ressaltar que as métricas de QST na distribuição de eletricidade podem variar de acordo com as regulamentações e normas específicas de cada país ou região. As distribuidoras de energia elétrica geralmente são obrigadas a reportar essas métricas regularmente às agências reguladoras para garantir que estão a cumprir os padrões de qualidade estabelecidos e para permitir uma comparação objetiva do desempenho entre diferentes empresas do setor. A melhoria contínua dessas métricas é fundamental para garantir a satisfação dos consumidores e o bom funcionamento dos setores económicos dependentes da eletricidade.

2.2 Causas e Impactos das Interrupções no Fornecimento de Energia Elétrica

As interrupções no fornecimento de energia elétrica podem ocorrer por diversos motivos e têm impactos significativos na vida das pessoas, empresas e infraestruturas. Essas interrupções podem variar em duração e extensão, desde pequenas falhas que duram alguns segundos até grandes eventos que afetam grandes áreas e duram horas ou até mesmo dias. Abaixo estão algumas das principais causas das interrupções no fornecimento de energia elétrica:

- Falhas em equipamentos : Defeitos e falhas em equipamentos, como transformadores, disjuntores e cabos, podem levar a interrupções no fornecimento de energia;
- Eventos Climáticos Extremos : Tempestades, furacões, nevascas e ventos fortes podem causar danos à infraestrutura elétrica, derrubando postes, árvores sobre os cabos e afetando subestações;
- Manutenção programada : Algumas interrupções são planejadas para permitir a realização de trabalhos de manutenção preventiva ou corretiva na rede elétrica;
- Sobrecarga da rede : Em situações de procura muito alta, a rede pode entrar em sobrecarga e causar interrupções;
- Interrupções na Geração: Problemas em sistemas de produção ou em conexões de energia podem causar interrupções no fornecimento.

Estas interrupções provocam os mais diversos impactos nas populações, tais como:

- Prejuízos Económicos : Interrupções no fornecimento de energia podem resultar em perdas financeiras significativas para empresas, devido à paralisação de produção, perda de vendas e danos em equipamentos;
- Emergências Médicas : A falta de eletricidade pode afetar a operação de hospitais e equipamentos médicos, colocando em risco a vida de pacientes em estado crítico;
- Segurança Pública : Interrupções podem impactar a segurança pública, uma vez que os sistemas de comunicação, iluminação pública e serviços de emergência podem ser afetados;
- Impacto no Quotidiano : As interrupções podem afetar o dia a dia das pessoas, causando desconforto, impedindo o uso de eletrodomésticos e afetando a rotina.

Perante estas causas e impactos, as distribuidoras de energia elétrica em conjunto com as autoridades reguladoras trabalham para melhorar a confiabilidade da rede elétrica, implementando medidas de prevenção de falhas, modernizando as infraestruturas e respondendo de forma ágil a eventos adversos para minimizar as interrupções e os seus respetivos efeitos.

2.3 Enquadramento Regulamentar da QST

O enquadramento regulamentar da QST na distribuição de eletricidade em Portugal é estabelecido pela Entidade Reguladora dos Serviços Energéticos (ERSE). Esta é a agência reguladora responsável por supervisionar os setores de energia em Portugal, incluindo a eletricidade e o gás natural.

A ERSE define os padrões e requisitos de qualidade que as empresas de distribuição de energia elétrica devem cumprir para garantir a Qualidade do Serviço Técnico oferecido aos consumidores. Esses padrões são estabelecidos com base em diretrizes nacionais e europeias, bem como na análise das características do setor elétrico português e das necessidades da população.

A ERSE estabelece metas anuais para os indicadores de desempenho e monitoriza o cumprimento dessas metas pelas distribuidoras de energia elétrica em Portugal. As empresas de distribuição são obrigadas a relatar regularmente os seus resultados à ERSE, e o não cumprimento dos padrões de qualidade pode resultar em penalidades financeiras e outras medidas corretivas.

O enquadramento regulamentar da QST em Portugal procura garantir que os consumidores recebam um serviço seguro e de alta qualidade, incentivando as distribuidoras a investir na modernização das suas redes, na manutenção preventiva e em tecnologias avançadas para melhorar a qualidade do fornecimento de energia elétrica em todo o país. Essa abordagem visa também promover a satisfação dos consumidores e contribuir para o desenvolvimento sustentável do setor elétrico em Portugal.

Em Portugal a melhoria do vetor da QST baseia-se num conjunto de projetos de investimento, tais como, “Automação e Telecomando da Rede de MT” e “Desenvolvimento da Rede”.

O resultado destes investimentos pretende também cumprir com as exigências do Regulamento de Qualidade de Serviço do Setor Elétrico (RQS) [4].

O RQS, aprovado pela ERSE, define:

- Obrigações e níveis de qualidade de serviço a cumprir pelos operadores;
- Compensações a pagar aquando da ocorrência de incumprimentos;
- Obrigações de monitorização e prestação de informação.

Adicionalmente, o RQS contém regras para operadores de redes e comercializadores, no que respeita a:

- Continuidade de serviço: número e duração dos cortes de eletricidade;
- Qualidade da eletricidade: amplitude, frequência, forma de onda e simetria do sistema trifásico da tensão;
- Qualidade comercial: atendimento, informação, reclamações e serviços prestados aos clientes.

Respeitando o RQS, a E-REDES apura um conjunto de indicadores que caracterizam a continuidade de serviço das redes de distribuição, para os diferentes níveis de tensão. Esta continuidade de serviço é avaliada através da ocorrência de interrupções, situação em que a tensão de alimentação no ponto de entrega é inferior a 5% da tensão declarada [5].

As interrupções podem ser classificadas no que respeita à sua duração em:

- Interrupção breve: interrupção com duração igual ou superior a um segundo e inferior ou igual a três minutos;
- Interrupção longa: interrupção com duração superior a três minutos.

Estes conceitos são pertinentes porque os indicadores de QST são baseados nestas definições, como se pode constatar nas expressões 2.1 a 2.5, que se apresentam adiante.

E no que respeita ao tipo em:

- Interrupções previstas: as interrupções em que os clientes são informados com antecedência mínima estabelecida no Regulamento de Relações Comerciais (RRC);
- Interrupções acidentais: as interrupções em que o operador de rede não consegue atribuir previamente uma data para a sua ocorrência.

Posto isto, o Decreto-Lei n.º 215-b/2012 estabelece que o fornecimento de energia elétrica pode ser interrompido por:

- Razões de interesse público;
- Razões de serviço;
- Razões de segurança;
- Casos fortuitos ou casos de força maior;
- Facto imputável ao cliente;
- Acordo com o cliente.

O vetor da QST é definido por vários indicadores, no entanto, é principalmente caracterizado pelos índices SAIDI e TIEPI que são complementados pelos índices SAIFI e MAIFI.

Estes índices são definidos como [4]:

- SAIDI: Duração média das interrupções longas do sistema (sigla adotada internacionalmente a partir da designação em língua inglesa do indicador *System Average Interruption Duration Index*);

$$SAIDI = \frac{\sum(N^o clientes(PdE) * T_{interrupcao longa}(PdE))}{N^o clientes_{total}} \quad (2.1)$$

- TIEPI : Tempo de interrupção equivalente da potência instalada;

$$TIEPI = \frac{\sum (P_{instalada}(PdE) * T_{interrupcao_{longa}}(PdE))}{\sum P_{instalada}(PdE)} \quad (2.2)$$

- SAIFI: Frequência média de interrupções longas do sistema (sigla adotada internacionalmente a partir da designação em língua inglesa do indicador *System Average Interruption Frequency Index*);

$$SAIFI = \frac{N^o_{interrupcoes_{longas}}(PdE)}{Total(PdE)} \quad (2.3)$$

- MAIFI: Frequência média de interrupções breves do sistema (sigla adotada internacionalmente a partir da designação em língua inglesa do indicador *Momentary Average Interruption Frequency Index*);

$$MAIFI = \frac{N^o_{interrupcoes_{breves}}(PdE)}{Total(PdE)} \quad (2.4)$$

- END: Energia não distribuída.

$$END = \frac{TIEPI * ED}{T_{considerado}(h)} \quad (2.5)$$

Onde, ED, corresponde à energia distribuída à rede de Média Tensão (MT) do ORD, em MWh, calculada a partir da energia entregue pelo operador da rede de transporte e pelos produtores ligados à Rede Nacional de Distribuição (RND), deduzida dos consumos dos clientes ligados a Alta Tensão (AT).

Todos estes indicadores são anuais e devem respeitar os seguintes requisitos técnicos:

- Simples: devem permitir medir aspetos concretos da qualidade de forma clara e direta, sendo os processos de cálculo fáceis de entender e implementar;
- Fiáveis: os dados utilizados para o seu cálculo devem ser objetivos e a sua recolha ser de elevado grau de confiança;
- Auditáveis: deve haver possibilidade de verificar os dados utilizados no seu cálculo, bem como os próprios cálculos realizados.

Os valores associados aos índices indicados não são constantes. Estes tendem a decrescer com o aumento do investimento na rede de distribuição, através, por exemplo, da instalação de automatismo para religação após isolamento do defeito na rede. Note-se que estas medidas que têm como objetivo principal melhorar o valor de um indicador da QST, acabam por melhorar vários outros indicadores, pois estes encontram-se todos interligados. Por outro lado, tendem a aumentar com a degradação da rede e na presença de condições atmosféricas adversas.

Posto isto, pode-se afirmar que os índices da QST são muito dependentes de eventos excecionais, quer estes sejam de natureza essencialmente estocástica, tais como a ocorrência de incidentes,

ou incertos como a variação do consumo. Os incidentes relacionados com intempéries (condições climáticas excepcionais) são analisados pela ERSE; se forem considerados inevitáveis, as interrupções associadas a estes incidentes não serão contabilizadas nos indicadores de QST.

O RQS define ainda um mecanismo de incentivo à melhoria da continuidade de serviço, o qual se aplica ao ORD em MT e AT em Portugal Continental e é constituído por duas componentes, sendo que uma componente tem como objetivo promover a continuidade global de fornecimento de energia elétrica e a segunda componente tem como objetivo incentivar a melhoria do nível de continuidade de serviço dos clientes com piores índices de QST.

2.4 Estratégias para melhoria da QST

Para melhorar a qualidade de serviço técnica na distribuição de energia elétrica, as distribuidoras e empresas podem adotar diversas estratégias e ações. Essas estratégias visam garantir a confiabilidade, segurança e eficiência do fornecimento, além de proporcionar uma melhor experiência aos consumidores. Abaixo estão algumas das principais estratégias para a melhoria da QST:

- **Investimentos em infraestruturas:** Realizar investimentos contínuos na modernização da infraestrutura elétrica é fundamental para aumentar a confiabilidade da rede de distribuição. Isso inclui a substituição e atualização de equipamentos obsoletos, a instalação de novos transformadores, cabos e postes mais resistentes e a expansão da capacidade da rede para atender ao crescimento da procura;
- **Manutenção Preventiva e Preditiva:** Adotar programas de manutenção preventiva e preditiva é uma estratégia importante para identificar e corrigir problemas antes que eles se transformem em falhas graves. A realização regular de inspeções e testes em equipamentos e redes ajuda a prevenir interrupções não planeadas;
- **Manutenção Corretiva:** Atividade técnica executada após a ocorrência de uma avaria. Tem como objetivo restaurar o ativo para uma condição em que possa atuar como pretendido, quer por reparação ou substituição.
- **Automação e Monitorização Avançada:** A implementação de tecnologias de automação e monitorização avançada, como sistemas SCADA (Supervisory Control and Data Acquisition), permite um controle mais eficiente da rede elétrica em tempo real. Isto ajuda a detetar e resolver problemas rapidamente, minimizando os impactos das interrupções;
- **Integração de Energias Renováveis:** A integração de fontes de energia renovável na matriz elétrica pode contribuir para a estabilidade e a confiabilidade da rede. O uso de tecnologias de armazenamento de energia, como baterias, também pode ajudar a suavizar a variabilidade das fontes renováveis e garantir a continuidade do fornecimento.

- Incentivos e Penalizações: A utilização de sistemas de incentivos e penalidades pode motivar as empresas a atingir metas de qualidade do serviço técnico estabelecidas pelas autoridades reguladoras.

A implementação conjunta destas estratégias pode ajudar a melhorar significativamente a qualidade de serviço técnica na distribuição de energia elétrica, garantindo um fornecimento mais confiável e seguro para os consumidores e contribuindo para o desenvolvimento sustentável do setor.

2.5 Síntese

Tal como anteriormente abordado, o ORD depara-se com a necessidade de elaborar o PDIRD.

A elaboração deste plano representa um exercício exigente e complexo, dado que para além do contexto legal em que é realizado também existe um contexto socioeconómico e regulatório que influencia as opções tomadas e a consequente definição da proposta final do documento, tais como, as orientações de política energética contidas no Relatório de Monitorização da Segurança de Abastecimento mais recente (RMSA-E 2022) [6].

Posto isto, observa-se que o planeamento da rede de distribuição a médio/longo prazo constitui um exercício indispensável de modo a permitir uma gestão racional dos recursos disponíveis. Assim, a elaboração do PDIRD torna-se de cariz essencial para conjugar os diferentes objetivos num plano de investimentos racional e eficiente.

No caso de Portugal continental, o ORD é a E-Redes. Esta classifica os investimentos na RND de acordo com os seus potenciais impactos, segundo os seguintes vetores:

- Segurança de Abastecimento (SA);
- Qualidade de Serviço Técnica (QST);
- Eficiência da Rede (ER);
- Eficiência Operacional (EO);
- Acesso a Novos Serviços (ANS).

Adicionalmente, existem outros programas e investimentos que, por não se enquadrarem em nenhum dos cinco vetores estratégicos acima apresentados, são agrupados em “Outros” (*e.g.* promoção ambiental, contadores).

A caracterização dos impactos dos investimentos nestes vetores estratégicos é fundamental para a tomada de decisão sobre os cenários de investimentos mais adequados, de modo a evidenciar a racionalidade técnico-económica dos planos propostos [7].

Um dos pontos mais relevantes do PDIRD e que será objeto de estudo ao longo desta dissertação diz respeito aos investimentos previstos no vetor de QST e à estimativa de evolução dos

indicadores que o caracterizam. Estes indicadores, dependem diretamente do número e da duração das interrupções de serviço, sendo condicionados pelo tipo e estado de saúde dos ativos, pelas condições de exploração e pelas ações de agentes externos, tais como as condições atmosféricas.

Os investimentos diretos em novos ativos permitem aumentar os índices de fiabilidade e a resiliência da rede.

As ações de manutenção levadas a cabo na RND podem ser divididas em quatro tópicos [8]:

- Condicionada;
- Programas Específicos;
- Sistemáticas;
- Corretiva.

De notar que os valores relativos à manutenção corretiva foram descartados dos valores globais da manutenção, Tabela 4.2, por serem contraditórios aos índices (conforme foi testado em [7]), pois, um custo superior em manutenção corretiva significa a ocorrência de mais acidentes no ano. Ou seja, um maior valor da corretiva iria implicar um maior valor do índice.

Estes tipos de ações de manutenção contribuem igualmente para o robustecimento da rede, o que provoca melhorias nos indicadores de QST.

Para além do investimento no vetor da QST, a continuidade de serviço está também dependente de outras componentes tais como a degradação dos elementos constituintes da RND e o trânsito de potências que nela circula.

Capítulo 3

Regressão Probabilística com Técnicas Bayesianas

A origem do estudo probabilístico nos indicadores da QST deve-se, essencialmente, à natureza estocástica dos eventos associados às interrupções de serviço.

Deste modo, recorrer-se-á à técnica Spike and Slab, técnica bayesiana que utiliza probabilidades para representar partes do problema, às quais estão associadas incertezas ou probabilidades, com o objetivo de caracterizar a solução em termos de margens de confiança.

Pretende-se que este modelo seja o mais eficiente possível, uma vez que, grandes desvios do valor real provocam problemas operacionais ao ORD.

3.1 Técnicas Bayesianas

As técnicas bayesianas [9] são muito vantajosas no sentido em que permitem determinar as margens de confiança associadas aos resultados obtidos. Isto apresenta um cariz fundamental em planeamento pois, os valores previstos possuem sempre erros associados, algo que, pode ser, em certa medida, acautelado pela representação dos resultados não como um valor pontual, mas sim no formato de intervalo probabilístico.

Deste modo, torna-se relevante o cálculo do intervalo de confiança a 95% (margem típica), isto é, representar cada um dos indicadores QST como um intervalo que será válido para 95% das situações.

Neste caso, e para distribuições normais, este intervalo será dado por:

$$\mu \pm 1,96 * \sigma \quad (3.1)$$

Onde:

- μ , representa o valor médio da amostra;

- σ , representa o desvio-padrão da amostra.

Tal como se pode verificar na figura seguinte.

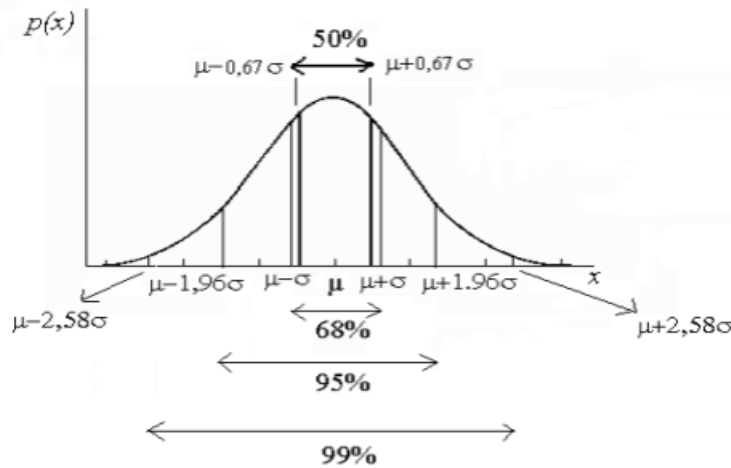


Figura 3.1: Representação da distribuição normal com vários intervalos de confiança [2]

A estimação de margens de confiança do resultado final é essencial para o ORD, ou seja, interessa não apenas antecipar os valores centrais dos indicadores de QST, mas também estimar entre que valores poderá oscilar.

3.1.1 Equação Fundamental Bayesiana

De modo a perceber o conceito de Spike and Slab é necessária a compreensão do conceito das técnicas Bayesianas.

Esta técnica, tal como referido anteriormente, recorre a probabilidades para representar partes do modelo, como por exemplo as entradas, às quais estão associadas incertezas ou probabilidades.

Contrariamente a outras técnicas, as Bayesianas, normalmente, apresentam a saída em forma de uma distribuição de probabilidades, que é representada por uma distribuição normal, tal como se apresenta na expressão seguinte:

$$y \sim N(\beta^T X, \sigma^2 I) \quad (3.2)$$

Onde:

- N , representa uma distribuição normal;
- β , representa o coeficiente do modelo;
- X , é a entrada;

- $\beta^T X$, representa a média;
- σ , representa o valor do desvio-padrão.

A equação fundamental que representa uma regressão Bayesiana pode ser dada por [3]:

$$P(\beta \setminus (y, X)) = \frac{P(y \setminus (\beta, X)) * P(\beta \setminus X)}{P(y \setminus X)} \quad (3.3)$$

Em que:

- $P(\beta \setminus X)$, é denominada como distribuição *a priori*, ou seja, representa a informação inicial dos parâmetros, sendo que não se relaciona com os dados. Usa-se, usualmente, para dar indicações de peritos ou da experiência;
- $P(y \setminus (\beta, X))$, representa a probabilidade de ocorrência dos dados, probabilidade de ocorrência da saída y com os parâmetros β e as entradas X , sendo este um indicador da qualidade da regressão com os parâmetros utilizados;
- $P(y \setminus X)$, é uma constante normalizadora. Representa a probabilidade de ter o vetor de saída y , dada a matriz de entrada X ;
- $P(\beta \setminus (y, X))$, é designada como a distribuição *a posteriori*.

O objetivo da utilização deste método é a determinação dos parâmetros β , pois são estes parâmetros que vão ajustar o modelo, dado que os dados são conhecidos.

A distribuição probabilística do resultado depende das distribuições consideradas nos dados de entrada (*priors*), da sua coerência e também da quantidade de exemplos disponíveis.

No caso de os dados de entrada serem escassos, a distribuição *a posteriori* irá ser mais espalhada.

À medida que o número de dados de entrada aumenta, a probabilidade, $P(y \setminus (\beta, X))$, sobrepõe-se à distribuição *a priori*, o que, tendo um elevado valor de dados de entrada, faz com que os parâmetros do modelo converjam para os valores do critério dos mínimos quadrados.

3.1.2 Spike and Slab

A técnica Spike and Slab [9] é uma variante do método bayesiano e corresponde a uma forma de seleção de variáveis, pois seleciona as que são verdadeiramente importantes para o modelo, descartando as que são redundantes ou mesmo irrelevantes para o mesmo.

O Spike representa a probabilidade *a priori* de um dado coeficiente do modelo possuir valor igual a zero enquanto que o Slab corresponde à distribuição *a priori* para os valores do coeficiente da regressão.

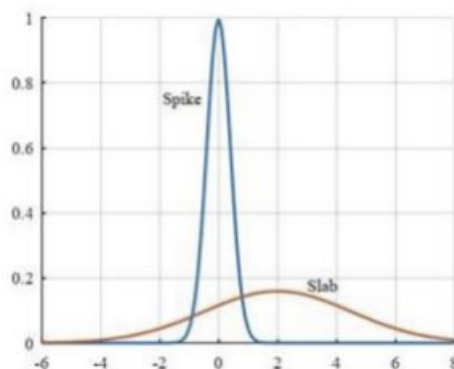


Figura 3.2: Distribuições Spike e Slab [3]

Considere-se um número N de variáveis de entrada, previamente conhecidas, associadas ao modelo. Com essas N variáveis cria-se um vetor de dimensão N , em que a cada elemento desse vetor corresponde o valor 0 ou 1. No caso de não existir qualquer tipo de informação, a atribuição do valor 0 ou 1 será obtida com recurso à distribuição de Bernoulli [10], distribuição discreta no intervalo $\{0,1\}$ que possui valor 1 com a probabilidade de sucesso p e valor 0 com a probabilidade complementar $q=1-p$. As variáveis de entrada irão ser multiplicadas por esse vetor o que leva a que as que sejam multiplicadas pelo valor 0, não constem no modelo.

No caso de a variável de entrada fazer parte do modelo é identificado um coeficiente, β , que corresponde a essa variável. Este coeficiente é calculado através de uma distribuição normal, com valor médio 0 e variância obtida através de $(X^T X)^{-1}$, em que X corresponde à matriz das variáveis de entrada.

Se a variável integrar o modelo temos um $\beta \neq 0$, já no caso de a variável não integrar o modelo temos um $\beta = 0$.

Seguidamente, com base no processo descrito anteriormente, é calculada a distribuição *a posteriori*.

Este processo é, então, repetido milhares de vezes utilizando a técnica de Markov Chain Monte Carlo [11], tendo como resultado a indicação dos coeficientes associados a cada variável e a sua introdução ou não no modelo.

3.1.3 Spike and Slab em ambiente R

De modo a proceder ao estudo do tema desta dissertação, optou-se, numa primeira instância, por recorrer ao *Package* 'BoomSpikeSlab', disponível no software R, para assim implementar o método Spike and Slab. A sua documentação e a própria biblioteca encontram-se disponíveis para *download*.

Deste modo, utilizou-se a função '*lm.spike*', presente na referida biblioteca, que implementa o método em uso nesta dissertação. Através dela é possível obter os valores dos coeficientes constituintes da regressão linear, bem como os vários tipos de erro (por exemplo, erro quadrático, RMSE e NRMSE).

Para a utilização da função anteriormente referida, serão fornecidos os dados de entrada, compostos pela evolução histórica dos valores do investimento, da variação do consumo e da manutenção, bem como, pela evolução do respetivo índice em estudo (TIEPI, SAIDI, SAIFI ou MAIFI), complementados pelo número de iterações, previamente definidas pelo utilizador, a serem implementadas pelo software, e pelos valores objetivo (Target) dos índices referidos.

Mais tarde, no início do capítulo 5, e devido a critérios e fatores que aí serão explanados, surgiu a necessidade de descartar tanto a função *'lm.spike'* como a própria biblioteca *'BoomSpikeSlab'*. Recorreu-se, portanto, à biblioteca *'brms'* (*Bayesian Regression Models using 'Stan'*), mais propriamente, à função *'brm'*. Deste modo será possível obter resultados mais credíveis.

Com o objetivo de criar uma ferramenta de estudo e análise que possa, posteriormente, ser reutilizada, criou-se um *script* em R, presente no Anexo A, facilitando, do mesmo modo, a aquisição e o respetivo tratamento dos dados obtidos.

Capítulo 4

Metodologia

4.1 Dados disponíveis

Nesta subsecção apresentam-se os dados disponíveis a ser utilizados ao longo da dissertação. Estes dados são disponibilizados pela E-Redes e podem ser organizados em 3 grandes grupos:

- Dados de investimento:
 - Valores históricos de investimento entre 2008 e 2020;
 - Cenário de investimentos;
 - Manutenções.
- Índices de qualidade:
 - TIEPI, SAIDI e SAIFI, valores anuais de 2008 a 2020;
 - MAIFI, valores anuais de 2013 a 2020.
- Caracterização da rede de distribuição:
 - Evolução da Energia Distribuída;
 - Evolução da procura;
 - Máximo trânsito de potências em cada subestação;
 - Projeção da evolução do número de consumidores;
 - Projeção da evolução da geração distribuída.

Posto isto, tem-se que os dados de entrada e de saída são dados por:

Tabela 4.1: Valores históricos dos índices (Inputs e Targets)

		Índices			
		TIEPI (min)	SAIDI (min)	SAIFI (nº)	MAIFI (nº)
Ano	2008	113,42	166,90	2,99	NA
	2009	121,44	186,28	3,13	NA
	2010	115,79	189,26	3,61	NA
	2011	75,48	125,72	2,41	NA
	2012	58,20	87,77	1,75	NA
	2013	70,00	104,61	2,00	13,00
	2014	59,90	87,70	1,80	13,10
	2015	52,60	74,40	1,60	10,22
	2016	49,99	71,36	1,68	10,35
	2017	50,45	71,40	1,55	9,06
	2018	57,59	85,00	1,77	11,57
	2019	49,13	72,22	1,74	9,28
	2020	51,12	72,67	1,75	9,33

De notar, com base na tabela anterior, que o modelo desenvolvido irá considerar, para o cálculo do índice num determinado ano, o valor desse índice no ano anterior, portanto, estes valores podem ser considerados como inputs e como targets.

Tabela 4.2: Valores históricos do investimento, manutenção e variação de consumo (Inputs)

		Investimento (€)	Variação do Consumo	Manutenção (€)
Ano	2008	36893088,63	NA	NA
	2009	30477441,08	-0,00437490	14857740,13
	2010	33492719,29	0,03826925	15938620,83
	2011	53243705,99	-0,03413802	13776859,43
	2012	65181054,24	-0,04420944	13183301,63
	2013	48524820,3	-0,02316127	13787748,41
	2014	53191915,31	-0,00162001	14819979,78
	2015	43552193,88	0,00978603	17003375,55
	2016	40863198,78	0,00903769	15979778,26
	2017	35980667,84	0,00260541	16151261,00
	2018	20559867,66	0,02716289	18110323,51
	2019	27028123,39	-0,00930198	15819535,24
	2020	28825585,73	-0,03812437	18708874,65

4.2 Qualidade dos modelos

Nesta subsecção apresentar-se-ão os critérios a utilizar de modo a verificar a qualidade dos modelos obtidos.

Numa fase inicial serão utilizados todos os modelos obtidos através do *software* adotado, em que a cada iteração é criado um modelo.

Este número de iterações e, por consequência, número de modelos obtidos são estipulados pelo utilizador, logo, é possível obter um número variado de modelos. Deste modo, pode-se afirmar que a variação do número de modelos obtidos, torna possível o estudo da sua influência sobre a criação de um modelo final robusto e eficaz para a previsão.

Tendo como base o elevado número de modelos que se podem obter, surge a necessidade de aplicar critérios de seleção a esses modelos, de modo a descartar os casos com pior desempenho e assim melhorar o resultado/modelo final.

No que respeita a esta dissertação, consideram-se 4 métricas de seleção dos modelos:

- Erro quadrático médio (MSE);
- Raiz do erro quadrático médio (RMSE);
- Raiz do erro quadrático médio normalizado (NRMSE);
- Erro percentual absoluto médio (MAPE).

De seguida, expõem-se as equações respetivas dos métodos de seleção utilizados.

$$MSE_y = \left(\frac{1}{N}\right) * \sum_{i=1}^N ((Target_i - Previsao_{i,y})^2) \quad (4.1)$$

Onde:

- N, representa o número total de amostras (anos);
- i, representa o ano;
- y, representa o respetivo modelo;
- $Target_i$, representa o valor objetivo para o ano i;
- $Previsao_{i,y}$, representa o valor estimado do índice no ano i para o modelo y.

$$RMSE_y = \sqrt{MSE_y} \quad (4.2)$$

Onde, y representa o respetivo modelo.

$$NRMSE_y = \frac{RMSE_y}{\frac{1}{N} * \sum_{i=1}^N Target_i} \quad (4.3)$$

Tal como anteriormente, i representa o ano e y o modelo.

$$MAPE_y = \left(\frac{1}{N}\right) * \sum_{i=1}^N \left(\frac{|Target_i - Previsao_{i,y}|}{|Target_i|}\right) * 100 \quad (4.4)$$

Da mesma forma que anteriormente temos que:

- N , representa o número total de amostras (anos);
- i , representa o ano;
- y , representa o respetivo modelo;
- $Target_i$, representa o valor objetivo para o ano i ;
- $Previsao_{i,y}$, representa o valor estimado do índice no ano i para o modelo y .

Após análise dos resultados intermédios, definiu-se o erro NRMSE como o critério principal de seleção dos modelos obtidos, em que, o melhor modelo seria aquele que possuísse o menor valor desse erro.

Note-se que é indiferente selecionar os valores de NRMSE, RMSE ou MSE como critério de seleção dos modelos, dado que todos correspondem ao critério dos mínimos quadrados.

Paralelamente, fixou-se que, para cada caso posteriormente exposto, os valores dos coeficientes do modelo final seriam obtidos através da média dos coeficientes dos 100 melhores modelos obtidos, ou seja, dos que apresentem os 100 erros NRMSE mais baixos.

$$Coeficiente_x = \frac{\sum_{y=1}^N Coeficiente_{y,x}}{N} \quad (4.5)$$

Onde:

- x , representa os diversos tipos de coeficientes em estudo (*Intercept*, Investimento, variação do Consumo, Manutenção e o respetivo índice);
- y , representa o modelo;
- N , representa o número total de amostras (anos).

4.3 Regressão linear para a estimação

Por forma a ser possível obter por estimação o valor do índice, com base nos valores estimados dos coeficientes (*Intercept*, Investimento, variação do Consumo, Manutenção e índice em estudo) dos modelos obtidos através do software R, desenvolveu-se a seguinte equação linear:

$$\begin{aligned}
Previsao_{i,y} = & Intercept_y + (Coeeficiente_{investimento,y} \times Input_{investimento,i-1}) \\
& + (Coeeficiente_{consumo,y} \times Input_{consumo,i}) \\
& + (Coeeficiente_{manutencao,y} \times Input_{manutencao,i-1}) \\
& + (Coeeficiente_{indice,y} \times Input_{indice,i-1})
\end{aligned} \tag{4.6}$$

Onde, tal como anteriormente:

- y , representa o número do modelo;
- i , representa o número do ano;
- $Intercept$, representa o termo independente.

4.4 Regressão linear para a projeção

Para a projeção da evolução do índice, a expressão a utilizar será a mesma que a utilizada para a estimação, apresentada na equação 4.6.

4.5 Abordagem adotada

Ao longo da elaboração deste trabalho de dissertação realizaram-se diversos tipos de ensaios, tais como:

- Realização de ensaios preliminares de modo a conhecer as vantagens e as limitações da ferramenta de regressão utilizada e aprender a ajustar os seus hiperparâmetros;
- Análise dos efeitos da normalização e da transformação das entradas na qualidade da estimação da QST;
- Impacto da divisão dos dados em conjunto de treino e de teste;
- Análise do impacto do número de iterações/modelos nos resultados da regressão;
- Análise do ajuste dos *priors* nos resultados da regressão.

Capítulo 5

Resultados

Neste trabalho de dissertação, de entre os índices da QST, deu-se especial atenção ao índice do TIEPI. Os procedimentos desenvolvidos para este índice poderão ser depois replicados para os outros indicadores.

Tal como se referiu anteriormente, inicialmente foi realizado o estudo recorrendo à função *'lm.spike'* do *Package 'BoomSpikeSlab'*.

Começou-se por considerar um número elevado de iterações, para, deste modo, obter uma maior variedade de modelos, sendo posteriormente reduzidos para um número menor pelos critérios anteriormente expostos. Dos 10000 modelos considerados inicialmente, são seleccionados os 10 melhores, ou seja, os 10 que apresentam os menores valores de NRMSE.

A Tabela 5.1 e a Figura 5.1 apresentam os valores do TIEPI obtidos com os 10 melhores modelos.

Tabela 5.1: Output dos melhores 10 modelos para o TIEPI (min)

		Ano											NRMSE
		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	
Modelo	6156	105.94949	80.44124	61.29162	59.02128	69.70989	64.44849	54.37539	53.94504	60.35261	49.22396	42.93176	11.70478
	5513	99.01368	78.31002	62.36326	59.90723	68.51656	63.19757	53.63077	53.76539	58.63498	49.28314	45.30877	12.28625
	1076	112.96001	81.32950	61.23507	60.01814	72.27686	66.43399	54.14247	53.98387	61.95108	47.07865	40.27670	12.38922
	6388	104.64601	76.77750	60.38246	60.23917	69.26287	64.10940	51.35457	50.95896	57.32825	42.20162	37.13339	12.62968
	5995	106.15697	86.38748	59.84005	54.72856	65.07788	61.53991	54.98930	51.36435	56.50526	50.74932	39.47547	12.88186
	1575	113.40517	85.48010	66.33233	63.55654	74.32468	66.48346	52.15064	53.17270	59.38079	45.57480	42.27618	12.88353
	79	102.59992	83.18525	58.59910	54.37575	63.89933	60.83211	54.35351	50.77695	55.77039	49.45285	38.66828	13.05577
	8660	104.54972	85.07532	58.49050	53.20174	63.64720	60.04447	53.68282	50.14034	55.24758	49.84279	38.64028	13.11409
	8694	102.61913	86.71371	62.92472	58.11106	65.65541	62.10285	54.67074	51.10658	54.55912	48.97604	39.81355	13.11767
	837	103.19753	86.95558	61.64840	56.07554	64.56023	60.66155	53.52292	50.02138	53.72989	48.86598	39.24269	13.16507

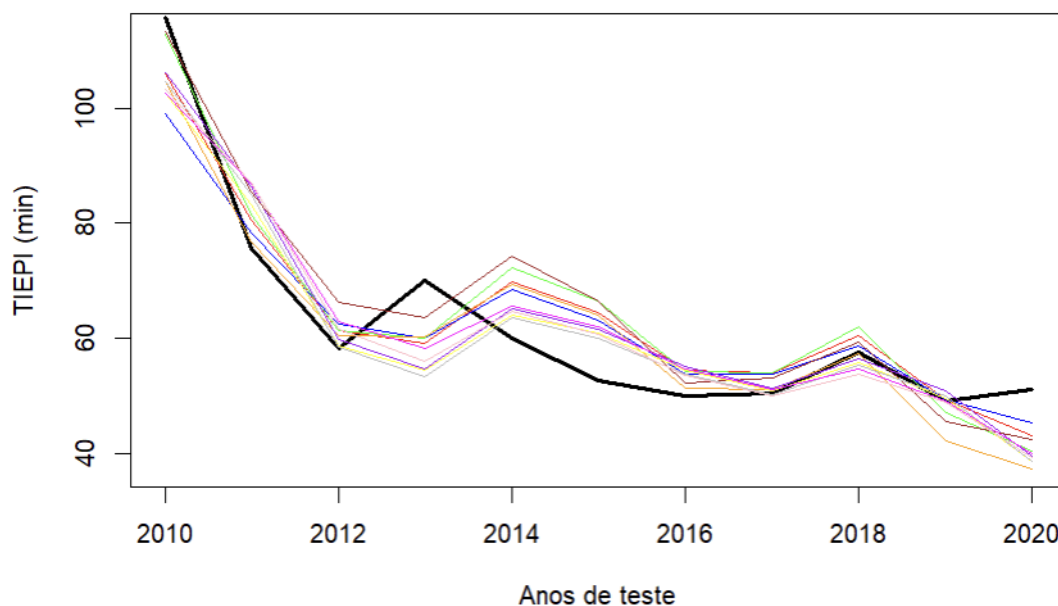


Figura 5.1: Evolução do TIEPI

Observando a figura anterior, é possível observar a preto a evolução real (Target) do TIEPI, bem como a evolução do valor do índice dos 10 melhores modelos obtidas através da função acima referida.

Contudo, e tal como referido anteriormente, após uma análise criteriosa dos resultados obtidos, chegou-se à conclusão de que estes modelos e resultados tinham de ser descartados/rejeitados pois correspondiam a modelos sem significado físico. De facto, alguns destes modelos apresentavam índices positivos para o termo de investimento e/ou da manutenção, o que significaria que o aumento do investimento e/ou da manutenção daria origem a um TIEPI mais elevado (pior QST). Uma vez que isto não faz sentido, estes modelos não são admissíveis.

Consequentemente, surgiu assim a necessidade de introduzir no modelo estas informações *a priori*, através de *priors*. Isto constitui uma das grandes vantagens deste tipo de técnica, ou seja, ser possível indicar que determinados coeficientes devem ser positivos ou negativos, ou que se encontram entre determinados limites, ou que apresentam determinada distribuição de probabilidade.

Esta necessidade obrigou à substituição da função *'lm.spike'* pelo facto de esta apenas permitir definir *priors* globais para todas as variáveis de entrada. Surgiu, deste modo, a função *'brm'* da biblioteca *'brms'* que se encontra disponível no ambiente R. Neste aspeto, a função *'brm'* é muito mais versátil, permitindo definir *priors* individuais e de diferentes tipos para cada uma das entradas do modelo.

5.1 Testes com a função 'brm'

Ao longo desta subsecção apresentam-se alguns procedimentos e resultados obtidos através dos mais variados testes realizados.

De seguida, na Figura 5.2, expõe-se uma ideia geral do código desenvolvido presente em Anexo, A.

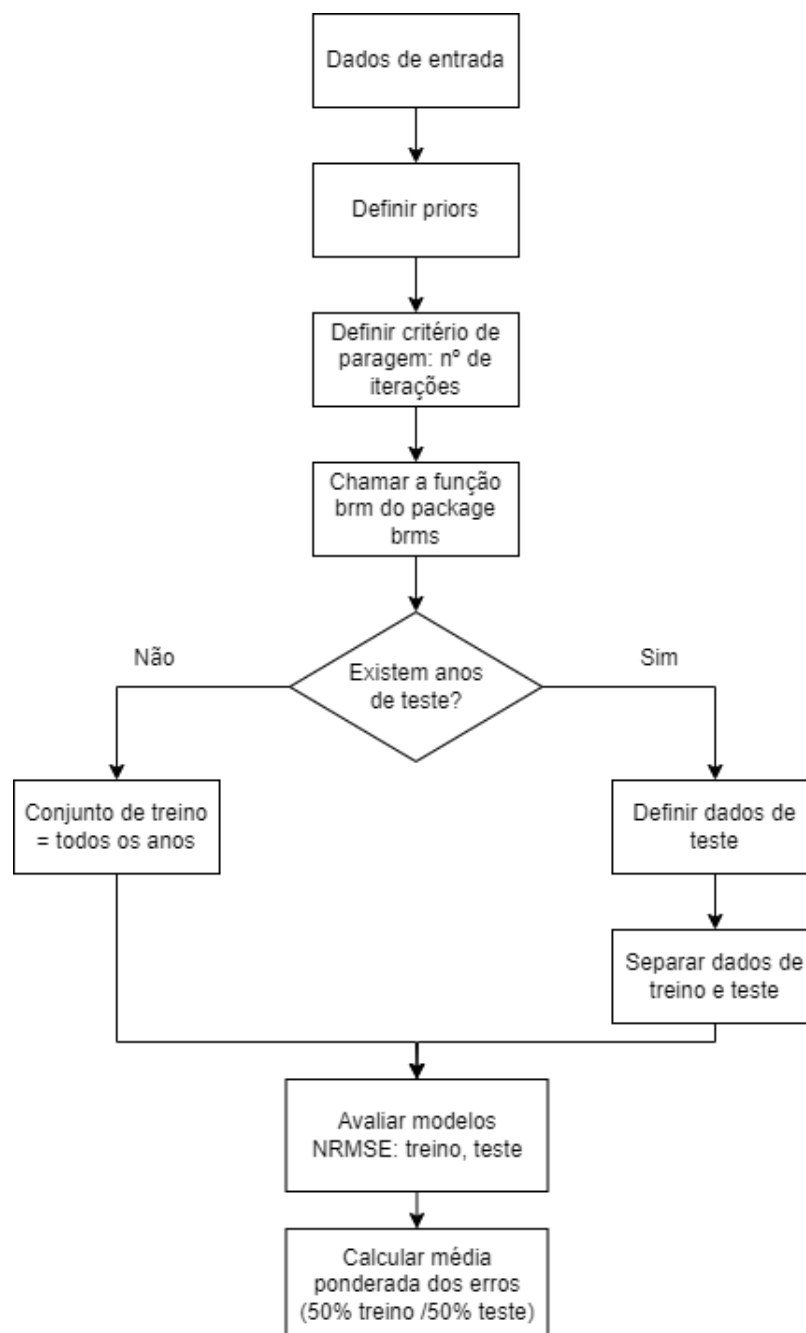


Figura 5.2: Fluxograma do código desenvolvido

5.1.1 Teste com variação do número de anos de teste

Inicialmente, de forma a facilitar a análise e compreensão dos resultados, começou-se por realizar uma normalização nas variáveis de entrada do Investimento e da Manutenção.

$$X_{1,i} = \frac{X_{1,i}}{10^6} \quad (5.1)$$

$$X_{1,i} = \log_{10}(X_{1,i} + 1) \quad (5.2)$$

$$X_{3,i} = \frac{X_{3,i}}{10^6} \quad (5.3)$$

$$X_{3,i} = \log_{10}(X_{3,i} + 7.2) \quad (5.4)$$

Onde, i , representa o ano.

De acordo com as expressões anteriores, é de notar que as variáveis x_1 e x_3 dizem respeito às variáveis de Investimento e Manutenção, respetivamente.

A primeira normalização em ambas as variáveis realizou-se simplesmente para converter os seus valores de euros para milhões de euros. Já a função logarítmica que se encontra associada a estas variáveis possui o intuito de modelizar um impacto não linear no índice, ou seja, a aplicação do primeiro milhão de euros no investimento faz reduzir o TIEPI de X minutos, mas o impacto do segundo milhão já é muito menor que X . A experiência com a E-REDES mostrou que esta transformação representa bem a realidade.

Adicionalmente, associado aos logaritmos adiciona-se um coeficiente de atenuação de forma a prevenir que, em cenários extremos, por exemplo aquando de um investimento de valor muito baixo ou nulo, evitar que esta parcela tome valores negativos. No caso do coeficiente de atenuação associado à manutenção optou-se por introduzir um valor mais elevado de modo a equilibrar os custos com investimento e manutenção, de acordo com um projeto antecessor a esta dissertação [7].

Tal como abordado anteriormente, foram definidos *priors* de modo a introduzir informações, que se têm como certas, nas variáveis.

Neste teste inicial, foram então definidos os seguintes *priors*:

- *prior* para o intercept: $N(0,100)$;
- *prior* para o investimento: $N(-10,2)$;
- *prior* para a variação do consumo: $N(0,2)$;
- *prior* para a manutenção: $N(-10,2)$;
- *prior* para o valor do índice: $N(0.8,0.1)$.

Em que, $N(m,s)$, representa a distribuição normal de média m e desvio-padrão s .

Para estes *priors* optou-se pela utilização da distribuição normal pois esta é definida por uma média e um desvio-padrão que fornecem uma grande probabilidade de os valores finais se encontrarem dentro dos limites por ela definidos.

Individualmente, atribuiu-se a distribuição $N(-10,2)$ às variáveis do Investimento e da Manutenção pelo facto de se querer que estas possuam um impacto negativo na saída. Quanto à distribuição $N(0.8,0.1)$ na variável do TIEPI deve-se pelo objetivo de que estes valores se encontrem, com mais probabilidade, entre 70% e 90% do valor do índice no ano anterior.

5.1.1.1 Teste sem nenhum ano de teste

De seguida apresentam-se os resultados obtidos para este teste inicial.

Os coeficientes do modelo final são obtidos através da média dos valores coeficientes dos 100 modelos com menor erro NRMSE de entre os 4000 modelos obtidos. Pode-se comparar a proximidade entre estes valores obtidos neste teste inicial com os obtidos em um projeto anterior a esta dissertação desenvolvido pelo INESC TEC em colaboração com a E-REDES [7].

Tabela 5.2: Modelo final sem anos de teste

MODELO FINAL	Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
	46,9358	-8,4300	1,1691	-10,9380	0,6397	15,89685

Tendo por base a tabela anterior, obtém-se a seguinte expressão para o modelo final:

$$TIEPI_i = 46.9358 + (-8.43 \times Inv_{i-1}) + (1.1691 \times Cons_i) + (-10.938 \times Man_{i-1}) + (0.6397 \times TIEPI_{i-1}) \quad (5.5)$$

Onde:

- i , corresponde ao ano;
- Inv , representa o investimento;
- $Cons$, representa a variação do consumo;
- Man , representa a manutenção;
- $TIEPI$, representa o valor do índice.

Daqui para a frente, esta notação irá ser reutilizada para expressar todas as expressões do modelo final.

Adicionalmente, apresenta-se o intervalo de confiança a 95% , obtido de acordo com o descrito na equação 3.1, tanto para o treino como para a projecção, no qual todos os dados disponíveis são

utilizados para ajuste do modelo, isto é, neste caso não são reservados exemplos apenas para teste do modelo.

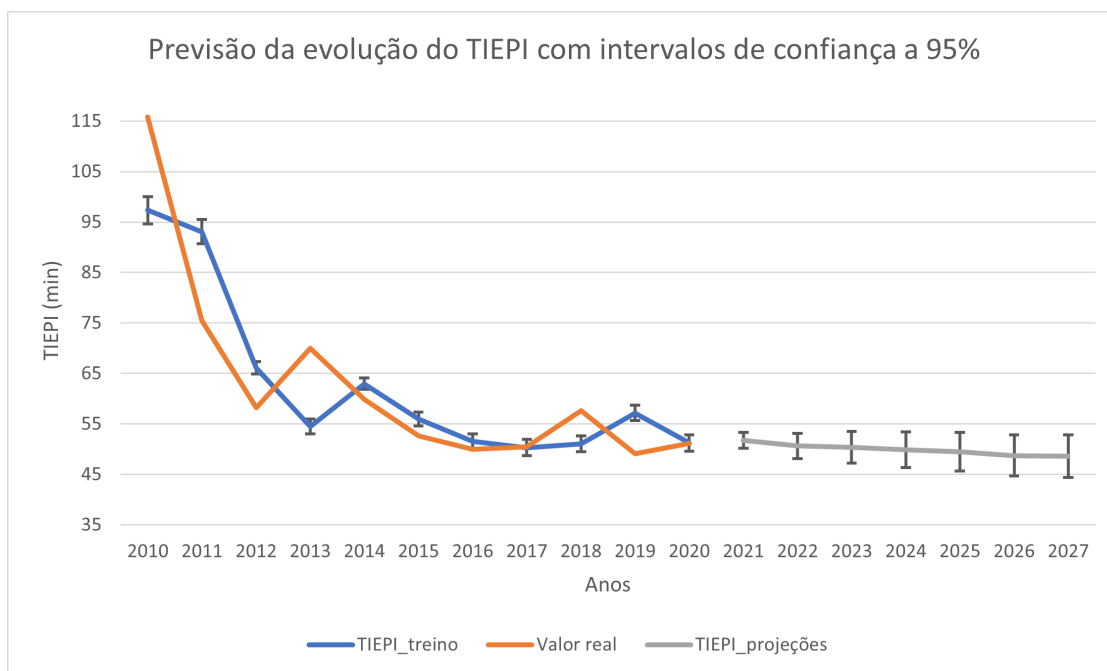


Figura 5.3: Intervalo de confiança a 95% do TIEPI sem anos de teste

A evolução do TIEPI é de natureza estocástica e dependente de eventos fortuitos (incidentes), assim como das características climáticas em cada ano.

Apesar disso, tal como se observa na Figura 5.3, o modelo obtido acompanha razoavelmente a evolução real do índice.

Note-se também que a grande dependência deste indicador face ao valor do ano anterior (64% , ver tabela 5.2), faz sentido porque significa que o estado da rede não se modifica substancialmente de um ano para o outro em termos de QST. No entanto, esta dependência do ano anterior faz com que existam "oscilações". Observem-se as transições em 2018 e 2019. Como em 2018 o valor real foi relativamente elevado, isso significa que a previsão para 2019 tenderá a ser elevada. Ou seja, quaisquer "anomalias" no valor real tendem a ser repercutidas na estimativa do ano seguinte.

Os níveis de confiança parecem algo otimistas, uma vez que não incluem a generalidade dos casos. A própria E-Redes reconhece este efeito e até costuma usar a previsão anterior do índice e não o seu valor para fazer estimativas e para estudos das margens de confiança.

5.1.1.2 Teste com 1 ano de teste

De aqui em diante, todos os testes serão realizados considerando anos de treino e anos de teste. Para estes testes, pesa-se igualmente o erro de treino (50%) e de teste (50%).

Realizando, então, o teste com as mesmas condições que as descritas imediatamente acima, considerando o ano de 2020 para teste, obtêm-se os seguintes resultados:

Tabela 5.3: Modelo final com 1 ano de teste

MODELO FINAL	Intercept	Investimento	Varição do Consumo	Manutenção	TIEPI	NRMSE
	46,2808	-9,8320	0,2415	-10,0060	0,6651	8,437976

Do mesmo modo, obtém-se a seguinte expressão para o modelo final:

$$TIEPI_i = 46.2808 + (-9.8320 \times Inv_{i-1}) + (0.2415 \times Cons_i) + (-10.006 \times Man_{i-1}) + (0.6651 \times TIEPI_{i-1}) \quad (5.6)$$

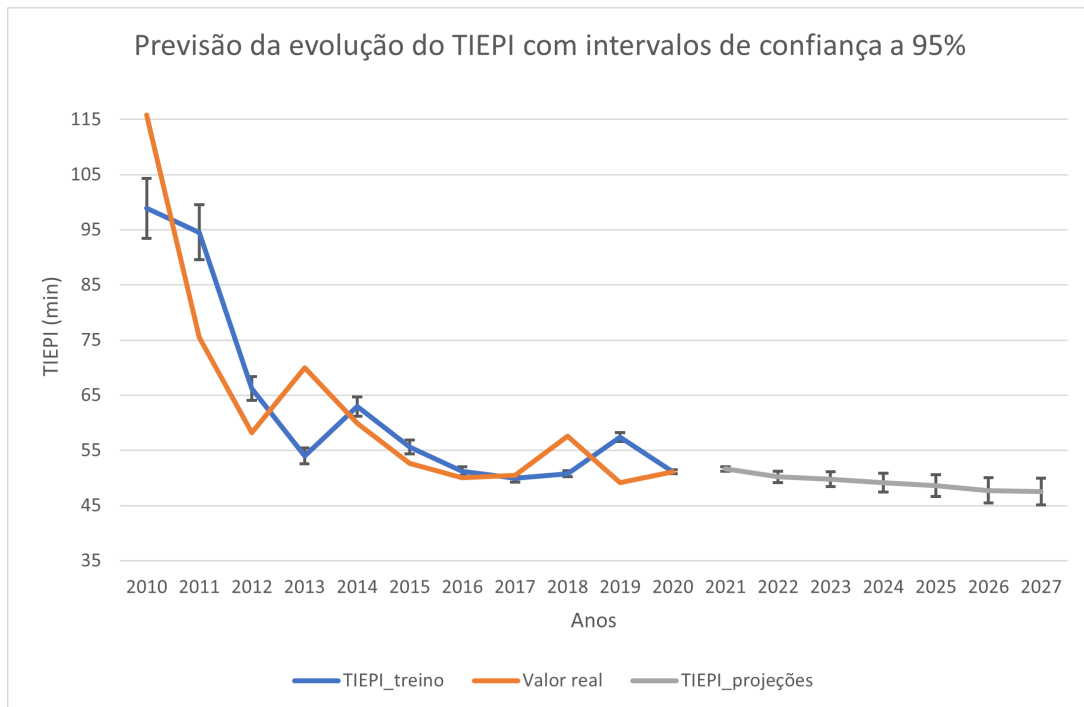


Figura 5.4: Intervalo de confiança a 95% do TIEPI com 1 ano de teste

Relativamente ao caso anterior, este modelo apresenta intervalos de confiança, que tal como anteriormente explanado são obtidos de acordo com a equação 3.1, mais estreitos. Isto pode ser devido ao facto do ano de 2020 ser um ano em que a estimativa é excelente, ou seja, com um erro muito baixo, o que faz com que as margens seguintes sejam estimadas de uma forma mais otimista.

O facto de se ter atribuído um peso de 50% ao erro de teste faz com que sejam seleccionados modelos que garantam uma aproximação quase perfeita (com erros muito próximos de zero).

5.1.1.3 Teste com 2 anos de teste

Realizando o mesmo teste considerando, neste caso, os anos de 2019 e 2020 como teste, obtêm-se os seguintes resultados:

Tabela 5.4: Modelo final com 2 anos de teste

MODELO FINAL	Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
	40,8181	-7,6618	-0,1211	-10,8051	0,6732	12,81072

Obtém-se a seguinte expressão para o modelo final:

$$TIEPI_i = 40.8181 + (-7.6618 \times Inv_{i-1}) + (-0.1211 \times Cons_i) + (-10.8051 \times Man_{i-1}) + (0.6732 \times TIEPI_{i-1}) \quad (5.7)$$

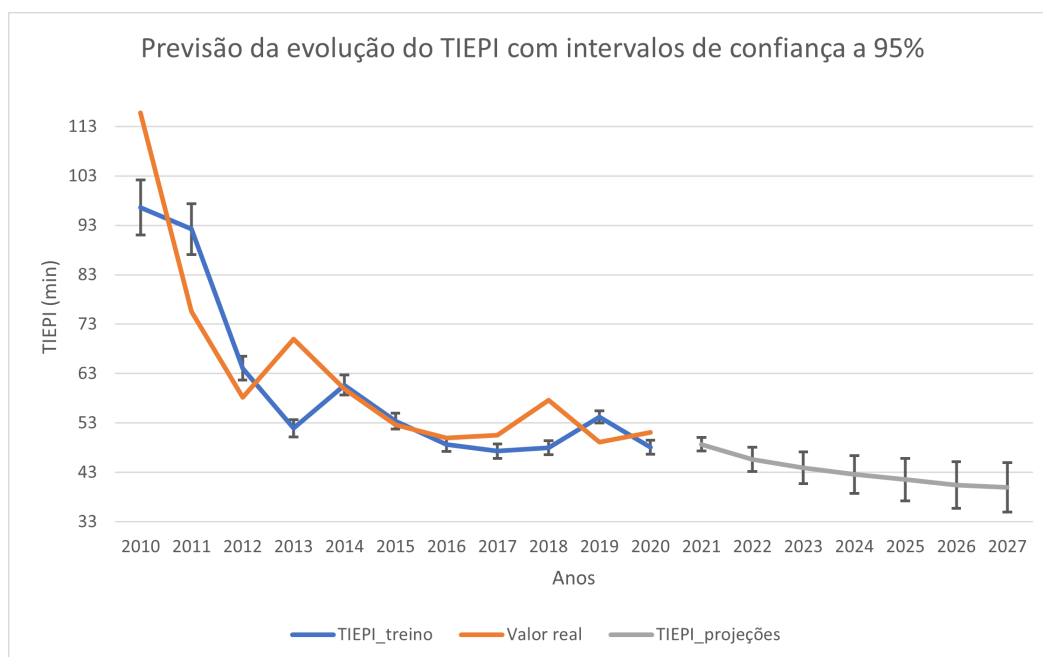


Figura 5.5: Intervalo de confiança a 95% do TIEPI com 2 anos de teste

Com 2 anos de teste, as estimativas para 2019 e 2020 apresentam um desvio mais elevado do que no caso anterior, o que leva a um aumento dos intervalos de confiança, obtidos de acordo com a equação 3.1.

5.1.1.4 Comparação dos resultados

Realizados os testes para a variação dos anos de teste, obtém-se os seguintes valores para os coeficientes dos modelos finais.

Tabela 5.5: Comparação dos coeficientes dos modelos finais

		Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
Modelos Finais	sem teste	46,9358	-8,4299921	1,16907173	-10,938026	0,63969	15,8968
	1 ano de teste	46,2808	-9,8320341	0,24146885	-10,006032	0,6651	8,43798
	2 anos de teste	40,8181	-7,6618056	-0,1211421	-10,805098	0,67319	12,8107

De seguida, expõe-se a comparação do treino e das projeções dos modelos com a variação do número de anos de teste.

Tabela 5.6: Evolução da previsão do TIEPI com número diferente de anos de teste

Ano	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Target	115,8	75,5	58,2	70	59,9	52,6	50	50,5	57,6	49,1	51,1
Sem teste	97,34109	93,08001	66,09055	54,46934	62,96554	55,96017	51,55753	50,3136	51,05557	57,17306	51,21762
1 ano teste	98,88821	94,5144	66,19464	53,98211	62,94652	55,63853	51,20862	49,92476	50,73413	57,42801	51,08619
2 ano teste	96,57133	92,24769	64,06629	51,90379	60,67244	53,34684	48,64074	47,29441	47,97916	54,20751	48,08793

Tabela 5.7: Evolução da projeção do TIEPI com número diferente de anos de teste

Ano	2021	2022	2023	2024	2025	2026	2027
Sem teste	51,7487	50,6413	50,3485	49,883151	49,4548161	48,7294704	48,5974
1 ano teste	51,6296	50,2116	49,7686	49,151122	48,6012235	47,7050902	47,4951
2 ano teste	48,6462	45,6124	43,9474	42,564332	41,5193489	40,3943357	39,9327

Com base nas Tabelas 5.6 e 5.7, obtém-se o seguinte gráfico da evolução do índice.

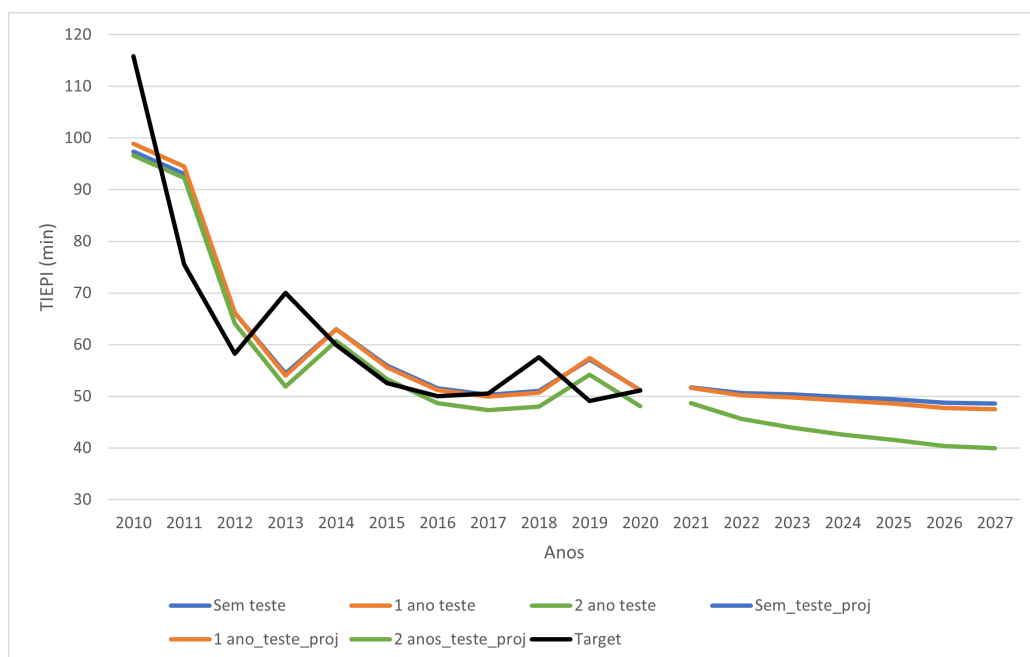


Figura 5.6: Comparação da evolução das previsões e projeções para diferentes anos de teste

Com base nos dados expostos atrás, decidiu-se proceder aos estudos posteriores utilizando o modelo com 2 anos de teste (2019 e 2020), pois, apesar de o modelo com 1 ano de teste ser aquele que apresenta um menor valor de NRMSE, o modelo com 2 anos de teste acaba por oferecer mais confiança, dado que o teste deixa de depender de uma única instância.

Adicionalmente, tal como referido anteriormente na subsecção 5.1.1.2, o facto de o modelo com um ano de teste possuir um erro muito inferior comparado com o modelo com dois anos de teste, é enviesado pela atribuição de um peso de 50% ao teste. Deste modo e para este caso (um ano de teste), ao ordenar os modelos, os cem melhores irão possuir erros de teste muito baixos. De forma a evitar este fenómeno, optou-se por utilizar a metodologia com dois anos de teste nos estudos posteriores.

5.1.2 Teste com variação do número de iterações

Nesta subsecção apresentam-se os resultados obtidos com a variação do número de iterações dos modelos.

De notar que, nesta subsecção e nas seguintes, aquando da apresentação dos intervalos de confiança, tal como anteriormente, estes são obtidos com base na equação 3.1.

5.1.2.1 Teste com 2000 iterações

Diminuindo o número de iterações de 4000 para metade, obtém-se o seguinte modelo:

Tabela 5.8: Modelo final com 2000 iterações e 2 anos de teste

MODELO FINAL	Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
	41,0092	-8,1363	-0,0639	-10,4680	0,6746	12,93597

Obtendo-se a seguinte expressão correspondente:

$$TIEPI_i = 41.0092 + (-8.1363 \times Inv_{i-1}) + (-0.0639 \times Cons_i) + (-10.4680 \times Man_{i-1}) + (0.6746 \times TIEPI_{i-1}) \quad (5.8)$$

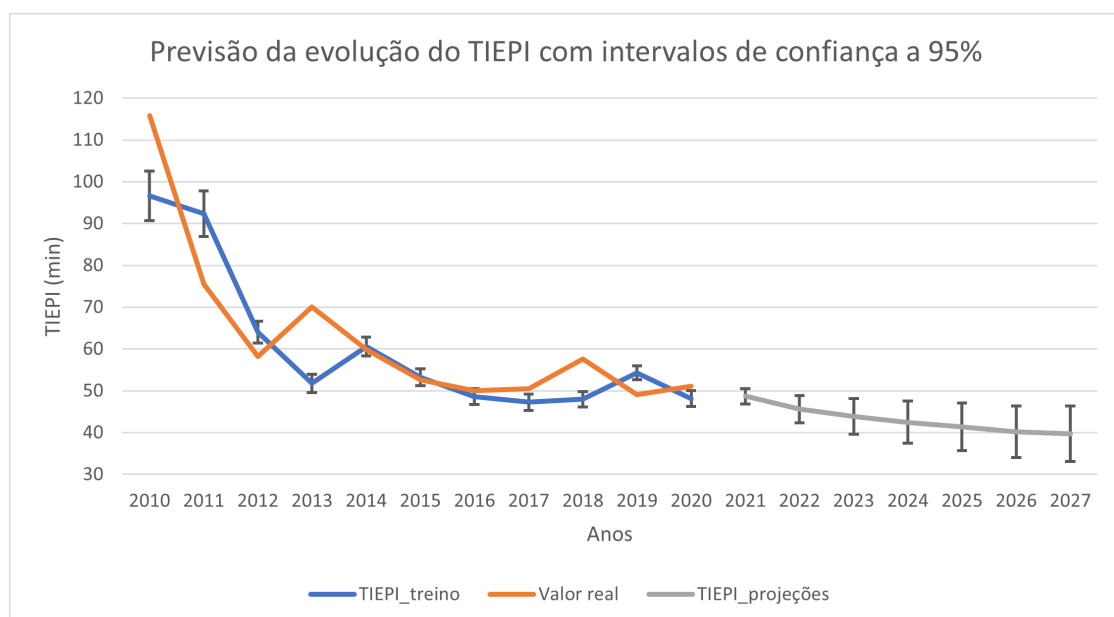


Figura 5.7: Intervalo de confiança a 95% com 2 anos de teste e 2000 iterações

Tal como anteriormente, podemos afirmar que a evolução da previsão do valor do índice acompanha, razoavelmente, a curva do valor real, exceto em anos excepcionais.

5.1.2.2 Teste com 8000 iterações

Nesta subsecção testou-se o efeito do aumento do número de iterações.

Tabela 5.9: Modelo final com 8000 iterações e 2 anos de teste

MODELO FINAL	Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
	40,4863	-7,0234	0,1814	-10,6701	0,6601	12,7143

Obtém-se a seguinte expressão para o modelo final:

$$TIEPI_i = 40.4863 + (-7.0234 \times Inv_{i-1}) + (0.1814 \times Cons_i) + (-10.6701 \times Man_{i-1}) + (0.6601 \times TIEPI_{i-1}) \quad (5.9)$$

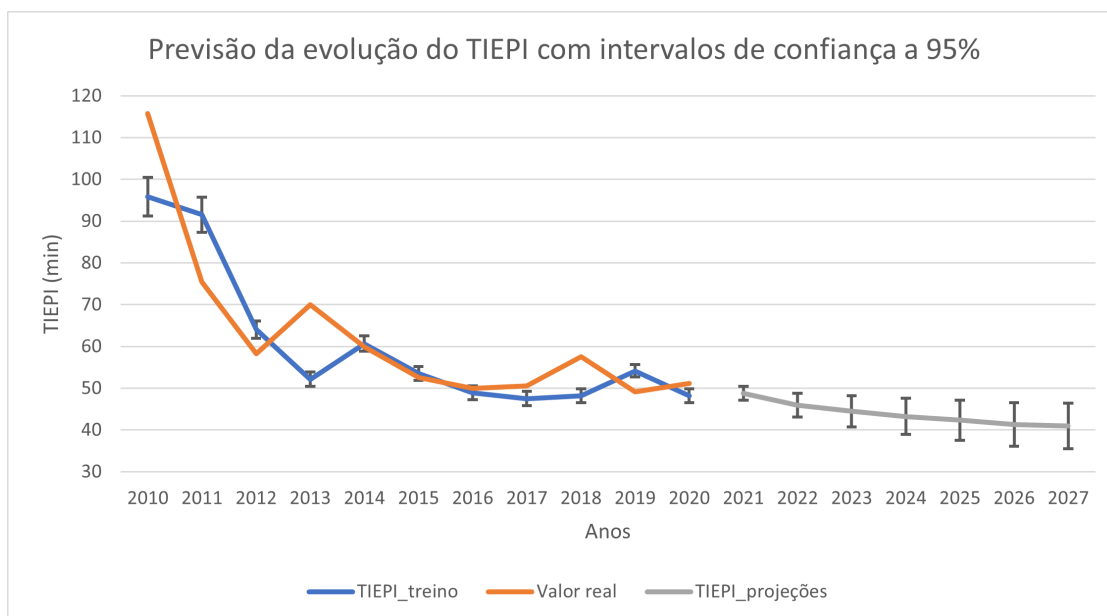


Figura 5.8: Intervalo de confiança a 95% com 2 anos de teste e 8000 iterações

Também para este caso de estudo, exceto em anos excepcionais, a curva da previsão do índice acompanha a curva do seu valor real.

5.1.2.3 Comparação dos resultados

Após a realização dos testes com a variação do número de iterações, obtém-se os seguintes valores para os coeficientes dos modelos finais.

Tabela 5.10: Comparação dos coeficientes dos modelos finais com diferente número de iterações

		Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
Modelos Finais	2000 iterações	41,0092	-8,1363042	-0,0638829	-10,468007	0,67456	12,936
	4000 iterações	40,8181	-7,6618056	-0,1211421	-10,805098	0,67319	12,847
	8000 iterações	40,4863	-7,0233673	0,1814352	-10,670097	0,66005	12,714

De seguida apresentam-se os valores obtidos para a previsão e projeção do valor do índice com a variação do número de iterações.

Tabela 5.11: Evolução da previsão do valor do TIEPI com número diferente de iterações

Ano	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Target	115,8	75,5	58,2	70	59,9	52,6	50	50,5	57,6	49,1	51,1
2000 iter	96,6723	92,3250	63,9804	51,7503	60,6003	53,2501	48,5882	47,2444	47,9578	54,3168	48,1161
4000 iter	96,5713	92,2477	64,0663	51,9038	60,6724	53,3468	48,6407	47,2944	47,9792	54,2075	48,0879
8000 iter	95,7932	91,5500	64,0150	52,1394	60,6809	53,5192	48,8600	47,5262	48,1784	54,1571	48,2071

Tabela 5.12: Projeção dos valores do TIEPI com variação no número de iterações

Ano	2021	2022	2023	2024	2025	2026	2027
2000 iter	48,6845	45,5804	43,8839	42,4573	41,3701	40,1865	39,6997
4000 iter	48,6462	45,6124	43,9474	42,5643	41,5193	40,3943	39,9327
8000 iter	48,7792	45,9799	44,4746	43,2415	42,3210	41,3292	40,9468

De acordo com as Tabelas 5.11 e 5.12, obtém-se o seguinte gráfico da evolução dos valores médios do índice.

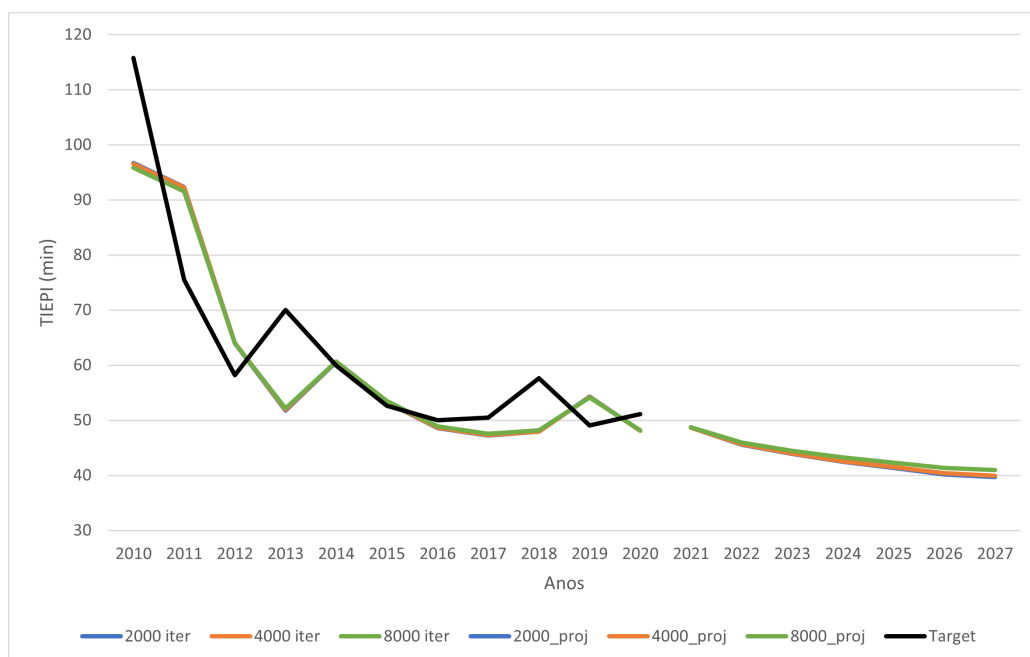


Figura 5.9: Comparação da evolução das previsões e projeções para diferentes valores de iterações

Realizando a análise dos dados expostos, podemos concluir que os resultados obtidos com 2000, 4000 e 8000 iterações não diferem muito entre si e que acompanham razoavelmente bem a curva do valor real do índice, exceto nos referidos anos excepcionais. Dado que não existem grandes diferenças, talvez seja preferível optar pelo caso mais rápido e simples (2000 iterações).

5.1.3 Teste com variação no prior do consumo

Nesta subsecção apresentam-se os resultados obtidos através da alteração dos parâmetros do *prior* da variação do consumo.

Note-se que, é de esperar que a um aumento do consumo, corresponda um aumento do TIEPI. Tipicamente, os anos com condições climáticas mais adversas (mais frio e vento) dão origem simultaneamente a um maior consumo e a uma maior quantidade de incidentes. Em alguns dos casos anteriores (por exemplo, Tabela 5.4), o coeficiente do mesmo é negativo.

Com esta modificação pretende-se garantir, com grande probabilidade, que o valor do coeficiente associado à variação do consumo seja positivo para, deste modo, fazer aumentar o valor do índice aquando do seu aumento, visto que, de uma maneira geral, o aumento do consumo origina uma maior probabilidade da ocorrência de falhas.

Esta alteração é realizada com base nos resultados obtidos a partir do teste com 4000 iterações e 2 anos de teste (2019 e 2020) presentes na Tabela 5.4 e na Figura 5.5.

Tem-se então que:

- *prior* para o intercept: $N(0,100)$;
- *prior* para o investimento: $N(-10,2)$;

- *prior* para a variação do consumo: $N(2,1)$;
- *prior* para a manutenção: $N(-10,2)$;
- *prior* para o valor do índice: $N(0.8,0.1)$.

Onde, tal como anteriormente, $N(m,s)$, representa uma distribuição normal de média m e desvio-padrão s .

Com esta alteração obtém-se o seguinte modelo final e a sua respetiva expressão:

Tabela 5.13: Modelo final obtido com alteração no prior do consumo

MODELO FINAL	Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
	40,4127	-7,4320	1,9984	-10,7571	0,6731	12,81217

$$TIEPI_i = 40.4127 + (-7.432 \times Inv_{i-1}) + (1.9984 \times Cons_i) + (-10.7571 \times Man_{i-1}) + (0.6731 \times TIEPI_{i-1}) \quad (5.10)$$

O que resulta na seguinte evolução:

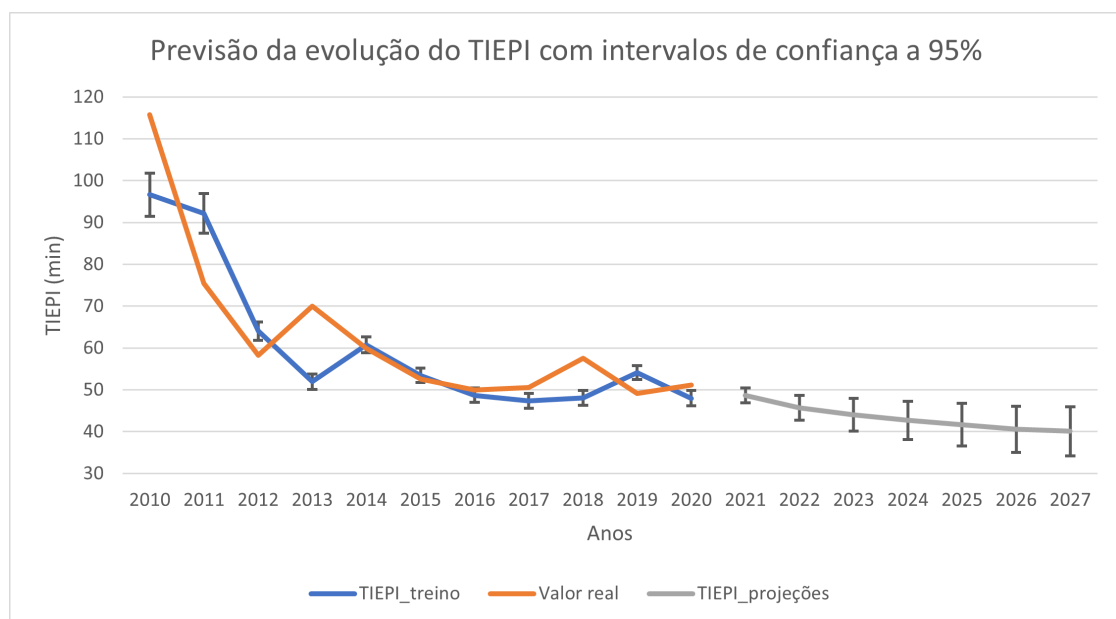


Figura 5.10: Intervalo de confiança a 95% com 2 anos de teste, 4000 iterações e alteração no prior da variação do consumo

5.1.3.1 Comparação dos resultados

Após o estudo do teste com alteração do *prior* associado à variação do consumo, podemos realizar uma comparação entre os resultados obtidos.

Tabela 5.14: Comparação dos coeficientes dos modelos finais com a alteração do prior da variação do consumo

		Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
Modelos Finais	N(0,2)	40,8181	-7,6618056	-0,1211421	-10,805098	0,67319	12,8107
	N(2,1)	40,4127	-7,4320387	1,99843067	-10,757102	0,67312	12,8122

Note-se que o novo modelo já faz sentido (coeficiente positivo para Variação do Consumo), apresentando uma estimativa do TIEPI muito semelhante.

Apresentam-se também os valores obtidos, para comparação, da previsão e projeção para cada um dos casos.

Tabela 5.15: Evolução da previsão do valor do TIEPI com diferentes *prior* no consumo

Ano	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Target	115,8	75,5	58,2	70	59,9	52,6	50	50,5	57,6	49,1	51,1
N(0,2)	96,5713	92,2477	64,066293	51,9037923	60,6724357	53,3468	48,6407	47,2944	47,9792	54,2075	48,0879
N(2,1)	96,6468	92,1803	64,0235926	51,9262128	60,7113346	53,4206	48,6959	47,329	48,0536	54,1519	47,9961

Tabela 5.16: Projeção dos valores do TIEPI com diferentes *prior* no consumo

Ano	2021	2022	2023	2024	2025	2026	2027
N(0,2)	48,6462	45,6124	43,9474	42,564332	41,5193489	40,3943357	39,9327
N(2,1)	48,6719	45,6959	44,0522	42,6954057	41,6504543	40,5468311	40,0927

Através das informações das Tabelas 5.15 e 5.16, obtém-se o seguinte gráfico da evolução dos valores médios do índice.

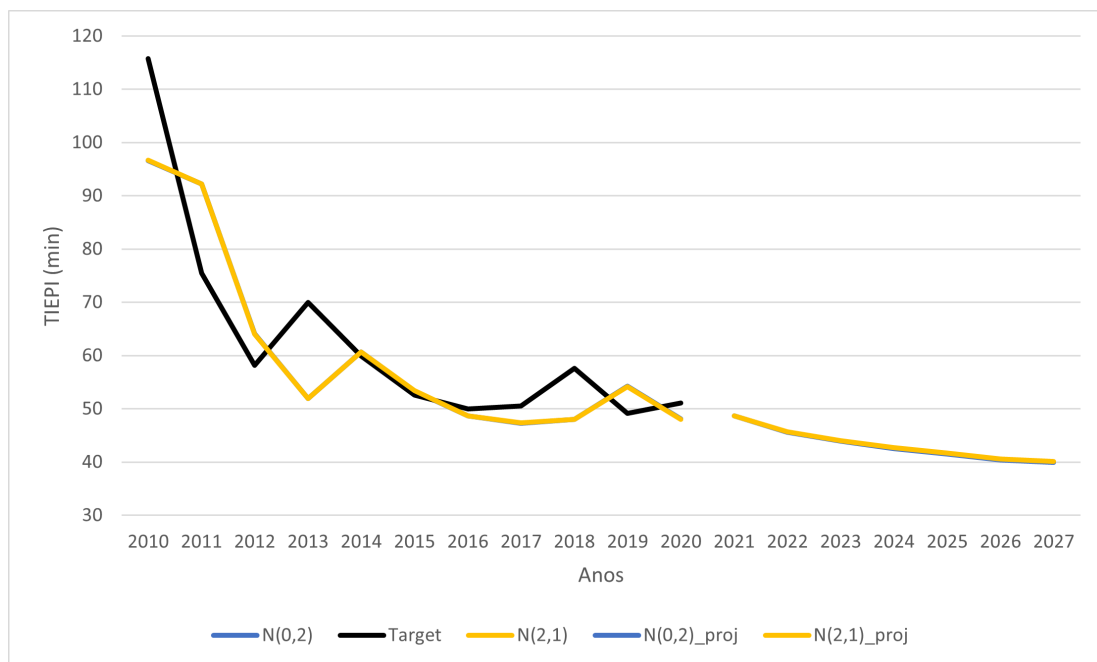


Figura 5.11: Comparação da evolução das previsões e projeções para diferentes *priors* no consumo

Analisando os dados expostos nas Tabelas 5.15 e 5.16 e na Figura 5.11, podemos observar que a variação nos parâmetros do prior do consumo não provoca alterações significativas nos resultados da previsão e/ou projeção dos valores do índice.

Contudo, de acordo com o objetivo inicial apresentado nesta subsecção, podemos afirmar que é mais provável obter resultados mais coerentes com o modelo que possui o *prior* de $N(2,1)$ na variação do consumo do que o que contém o *prior* $N(0,2)$.

5.1.4 Teste sem a utilização da função logarítmica

Para a realização deste teste tem-se por base, tal como referido anteriormente, o caso com 4000 iterações e 2 anos de teste (2019 e 2020).

Neste caso, consideram-se apenas as normalizações presentes nas equações 5.12 e 5.14, de modo a converter esses valores para milhões de euros.

Com este teste pretende-se estudar a influência da utilização da função logarítmica.

Obtém-se, então, os seguintes coeficientes para o modelo final e a sua respetiva expressão:

Tabela 5.17: Modelo final sem a função logarítmica

MODELO FINAL	Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
	145,6034	-0,5818	0,0467	-6,9214	0,6793	12,26337

$$TIEPI_i = 145.6034 + (-0.5818 \times Inv_{i-1}) + (0.0467 \times Cons_i) + (-6.9214 \times Man_{i-1}) + (0.6793 \times TIEPI_{i-1}) \quad (5.11)$$

Resultante na seguinte evolução:

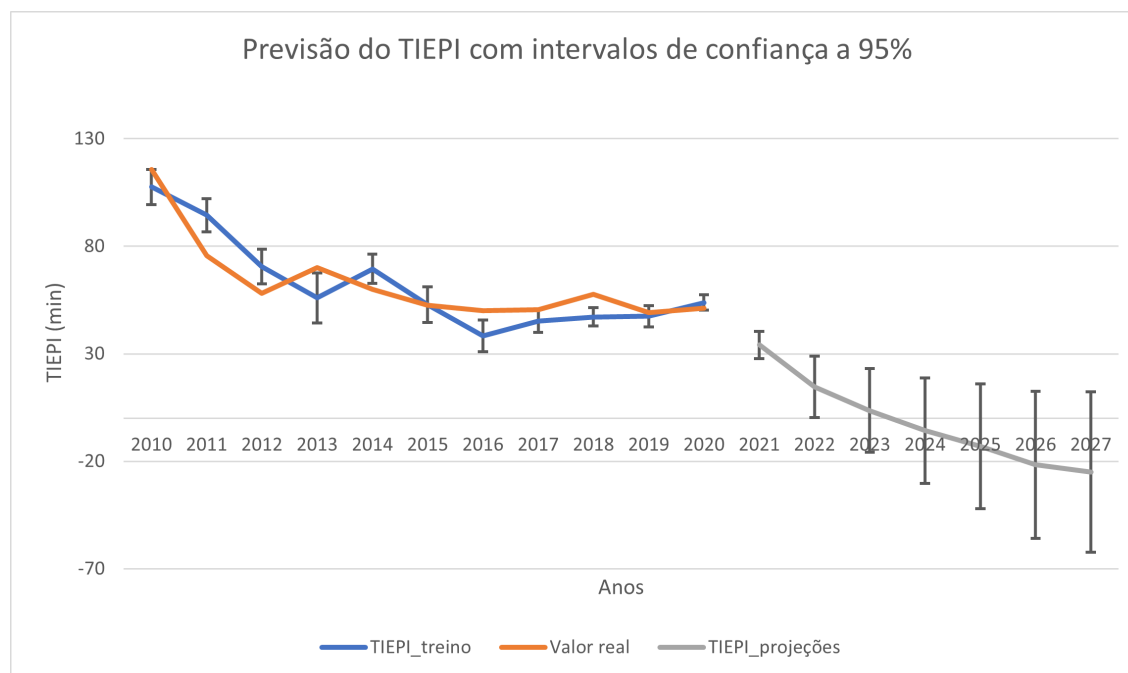


Figura 5.12: Intervalo de confiança a 95% com 2 anos de teste, 4000 iterações e sem a função logarítmica

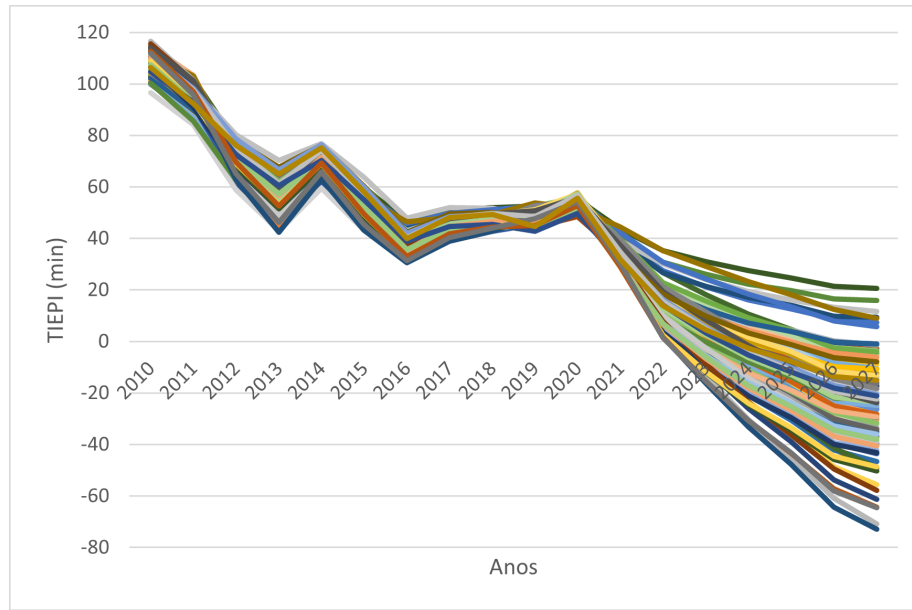


Figura 5.13: Evolução do índice dos 100 melhores modelos

5.1.4.1 Comparação dos resultados

Comparando os resultados obtidos atrás, com os presentes na Tabela 5.4 e na Figura 5.5, podemos confirmar a importância da utilização da função logarítmica, para modelizar um impacto não linear no índice, de modo a não se obter projeções do valor do índice negativas.

5.1.5 Teste sem a utilização das constantes normalizadoras presentes nos logaritmos

Este teste tem como principal objetivo analisar a importância das constantes normalizadoras definidas anteriormente de modo a atenuar o efeito logarítmico.

Posto isto, para este teste foram realizadas as seguintes normalizações:

$$X_{1,i} = \frac{X_{1,i}}{10^6} \quad (5.12)$$

$$X_{1,i} = \log_{10}(X_{1,i}) \quad (5.13)$$

$$X_{3,i} = \frac{X_{3,i}}{10^6} \quad (5.14)$$

$$X_{3,i} = \log_{10}(X_{3,i}) \quad (5.15)$$

Onde, i , representa o ano.

Com isto, obtém-se os seguintes coeficientes para o modelo final e respetiva expressão:

Tabela 5.18: Modelo final sem as constantes normalizadoras

MODELO FINAL	Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
	39,5874	-7,7111	-0,0392	-11,1692	0,6717	12,72879

$$TIEPI_i = 39.5874 + (-7.7111 \times Inv_{i-1}) + (-0.0392 \times Cons_i) + (-11.1692 \times Man_{i-1}) + (0.6717 \times TIEPI_{i-1}) \quad (5.16)$$

Obtendo-se a seguinte evolução:

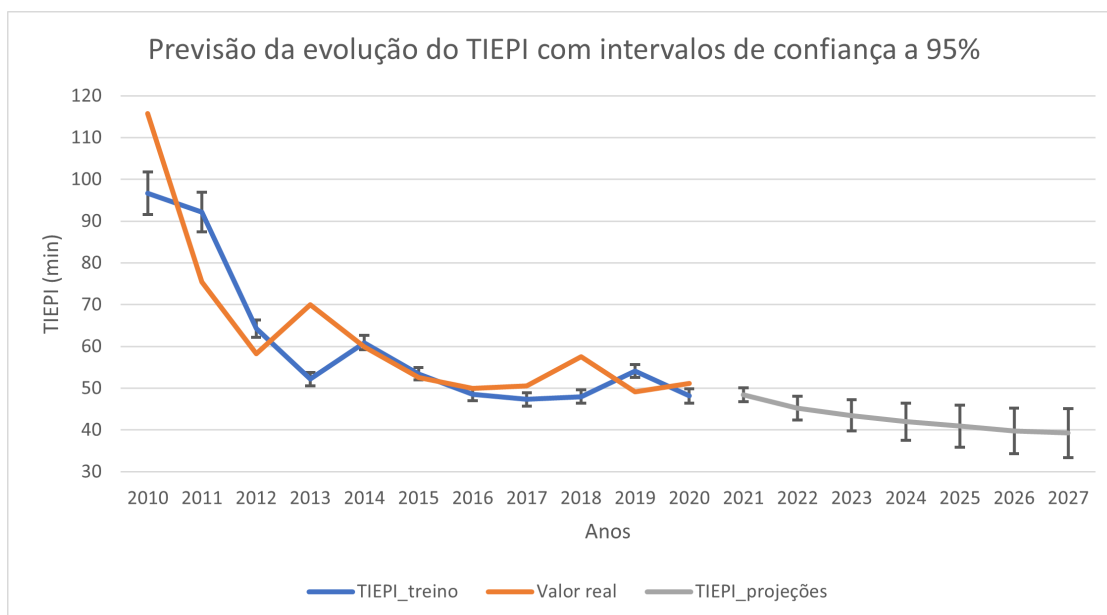


Figura 5.14: Intervalo de confiança a 95% com 2 anos de teste, 4000 iterações e sem as constantes normalizadoras

5.1.5.1 Comparação dos resultados

Após a realização deste estudo, podemos comparar os resultados obtidos com o estudo já realizado anteriormente com a inclusão das constantes normalizadoras.

Tabela 5.19: Coeficientes dos modelos finais com a exclusão das constantes normalizadoras

Modelos Finais		Intercept	Investimento	Variação do Consumo	Manutenção	TIEPI	NRMSE
	Com normalização	40,81812	-7,6618056	-0,1211421	-10,805098	0,67319	12,8107
	Sem normalização	39,58743	-7,711118	-0,0391562	-11,169154	0,67175	12,7288

De seguida, expõem-se os valores obtidos para a previsão e projeção.

Tabela 5.20: Evolução da previsão do valor do TIEPI com e sem constantes normalizadoras

Ano	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Target	115,8	75,5	58,2	70	59,9	52,6	50	50,5	57,6	49,1	51,1
Com normalização	96,5713	92,248	64,06629	51,9038	60,6724	53,3468	48,6407	47,2944	47,9792	54,2075	48,0879
Sem normalização	96,6302	92,181	64,25795	52,1855	60,8821	53,4392	48,5384	47,3	47,9824	54,099	48,157

Tabela 5.21: Projeção dos valores do TIEPI com e sem constantes normalizadoras

Ano	2021	2022	2023	2024	2025	2026	2027
Com normalização	48,646197	45,6123556	43,9474	42,564	41,51935	40,3943	39,9327
Sem normalização	48,449042	45,2443876	43,4693	42,004	40,89881	39,7222	39,2344

Através das informações presentes nas Tabelas 5.20 e 5.21, obtém-se o seguinte gráfico da evolução dos valores do índice.

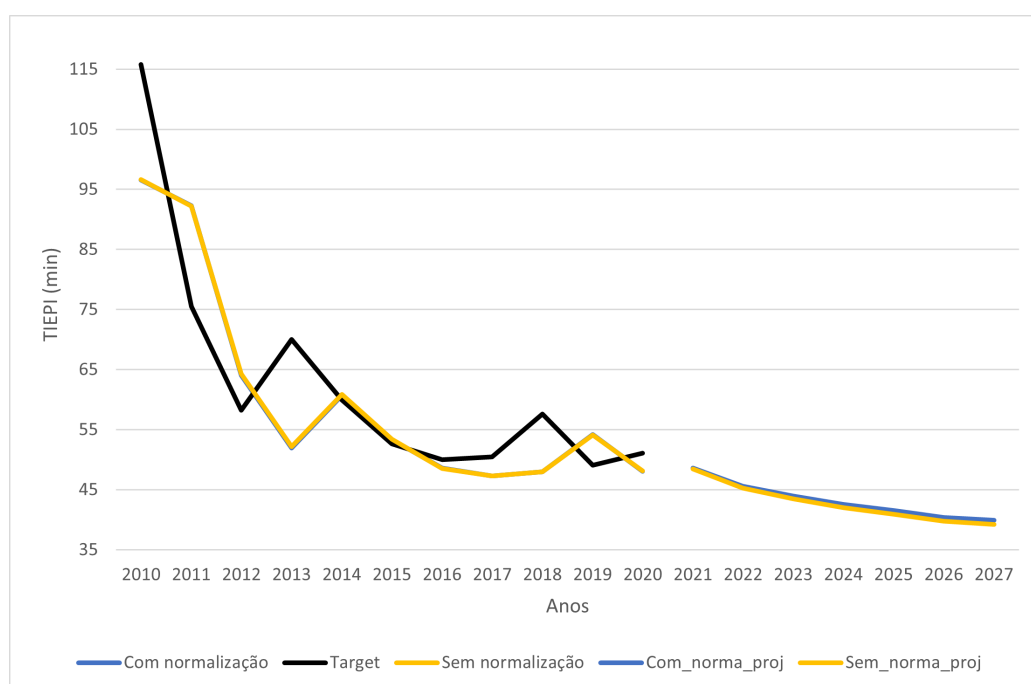


Figura 5.15: Comparação da evolução das previsões e projeções com e sem constantes normalizadoras

Com a observação dos dados expostos nas Tabelas 5.20 e 5.21 e na Figura 5.15, é possível concluir que a inclusão ou não de constantes normalizadoras nas funções logarítmicas não provoca, nestes modelos, alterações significativas nos resultados da previsão e/ou projeção dos valores do índice.

Contudo, tal como anteriormente referido e tal como se deduziu num projeto antecessor a esta dissertação [7], a inclusão de uma constante adicionada ao argumento do logaritmo tem como objetivo não excluir o caso (improvável) em que o investimento é zero, evitando, assim, a impossibilidade numérica $\log(0)$.

5.1.6 Teste sem a utilização de *priors*

Na realização deste teste não são considerados os *priors* das variáveis de entrada. Pretende-se avaliar a importância da definição destes *priors*.

Através da sua realização, obtém-se os seguintes coeficientes para o modelo final e a sua respetiva expressão:

Tabela 5.22: Modelo final sem *priors*

MODELO FINAL	Intercept	Investimento	Varição do Consumo	Manutenção	TIEPI	NRMSE
	669,8535	-53,4630	253,2515	-408,3592	0,4714	8,299919

$$TIEPI_i = 669.8535 + (-53.4630 \times Inv_{i-1}) + (253.2515 \times Cons_i) + (-408.3592 \times Man_{i-1}) + (0.4714 \times TIEPI_{i-1}) \quad (5.17)$$

Que se traduz na seguinte evolução:

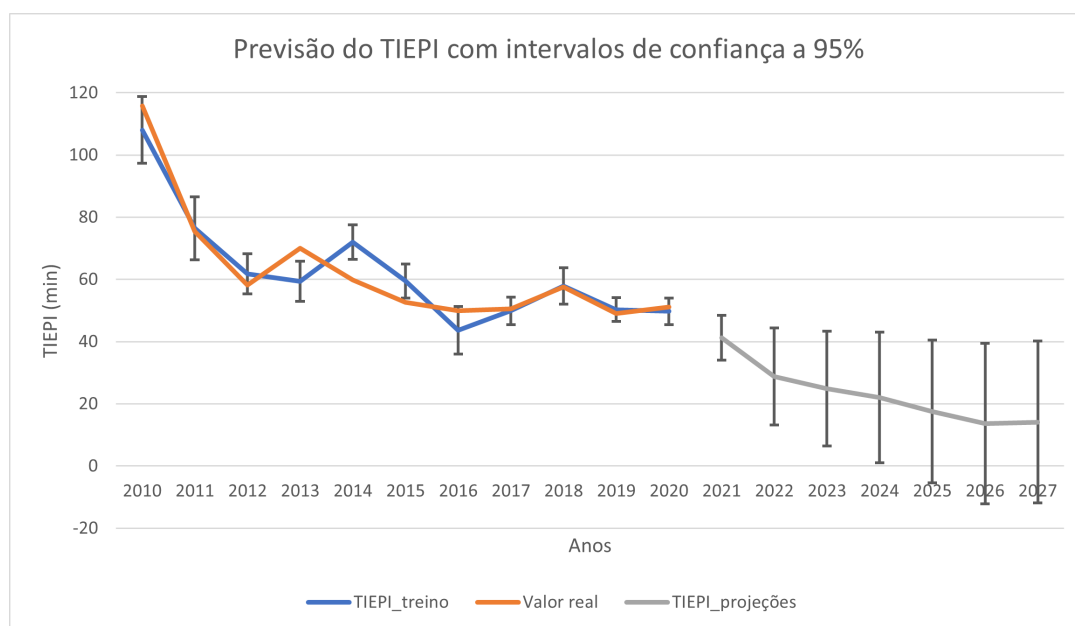


Figura 5.16: Intervalo de confiança a 95% com 2 anos de teste, 4000 iterações e sem o uso de *priors*

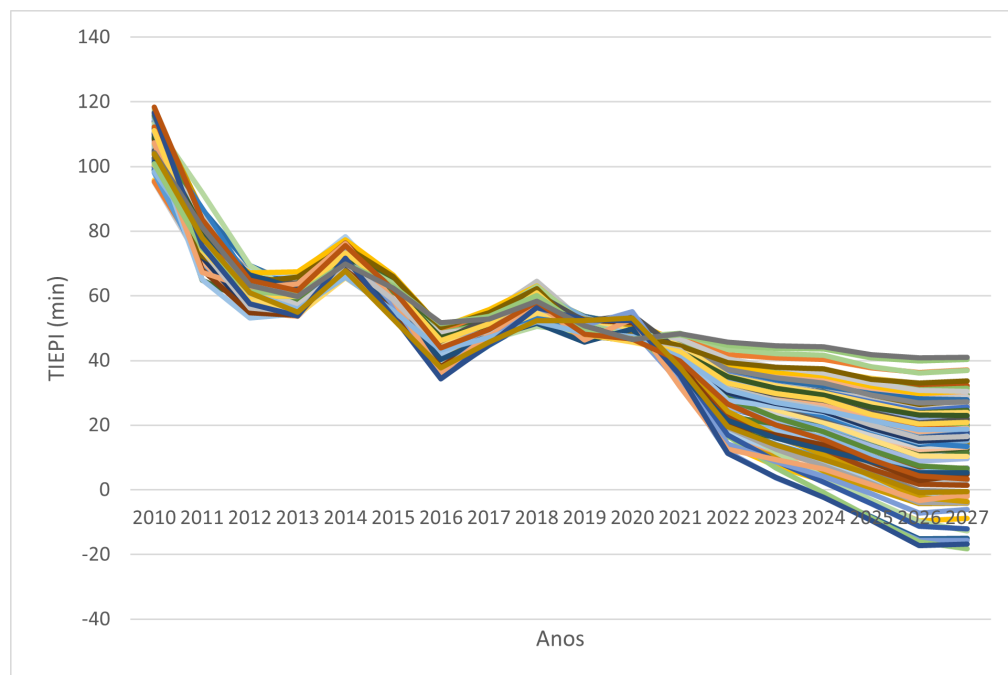


Figura 5.17: Evolução do índice dos 100 melhores modelos

5.1.6.1 Comparação dos resultados

Comparando os resultados obtidos imediatamente atrás com os presentes na Tabela 5.4 e na Figura 5.5, podemos confirmar a importância da utilização e definição de *priors* de modo a obter resultados finais coerentes e assim evitar, por exemplo, a obtenção de projeções do valor do índice negativas, tal como se verifica na Figura 5.17.

5.1.7 Modelo Final

Após a realização e análise criteriosa dos resultados obtidos a partir de todos os testes anteriormente expostos, conclui-se que o teste que melhor prevê e projeta a evolução do valor do índice TIEPI, bem como o seu intervalo de confiança a 95 % é o teste com variação no prior do consumo, apresentado na secção 5.1.3.

Os valores dos coeficientes para esse modelo final e a evolução dos valores do TIEPI com os respectivos intervalos de confiança a 95% podem ser observados na Tabela 5.13 e na Figura 5.10, respetivamente. A correspondente expressão para o modelo final apresenta-se na equação 5.10.

Como referido anteriormente, este modelo é obtido através dos 100 melhores modelos de entre os 4000 iniciais, ou seja, os 100 que possuem menor valor NRMSE.

Tabela 5.23: Média e desvio-padrão dos 100 melhores modelos

	Coeficientes				
	intercept	Investimento	Variação do Consumo	Manutenção	TIEPI
média	40,41269	-7,432038705	1,99843067	-10,75710206	0,67312
desvio-padrão	4,978343	1,508493581	1,014271815	1,787755827	0,041522

De seguida apresentam-se as distribuições gaussianas dos coeficientes dos 100 melhores modelos deste modelo final.

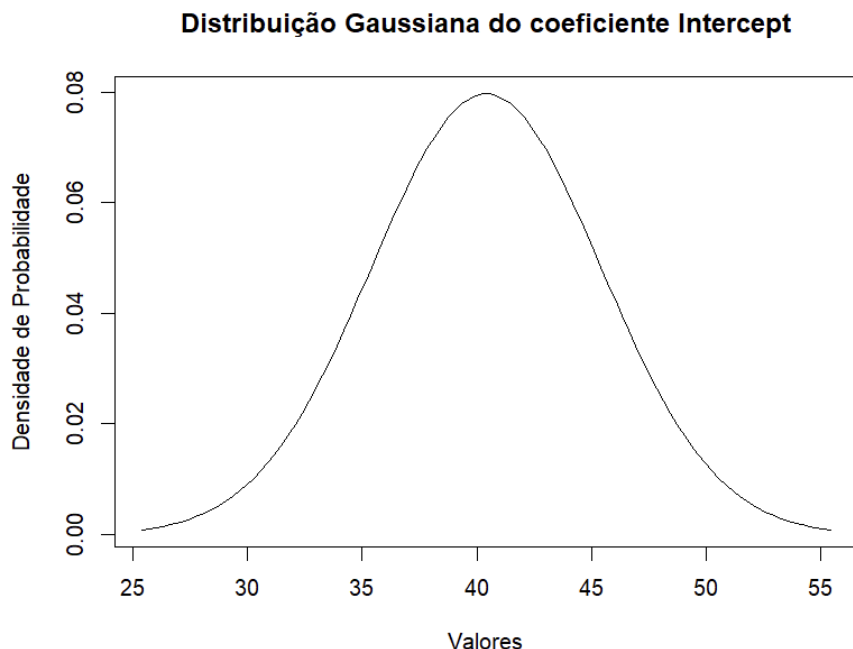


Figura 5.18: Distribuição gaussiana dos 100 melhores modelos do coeficiente Intercept

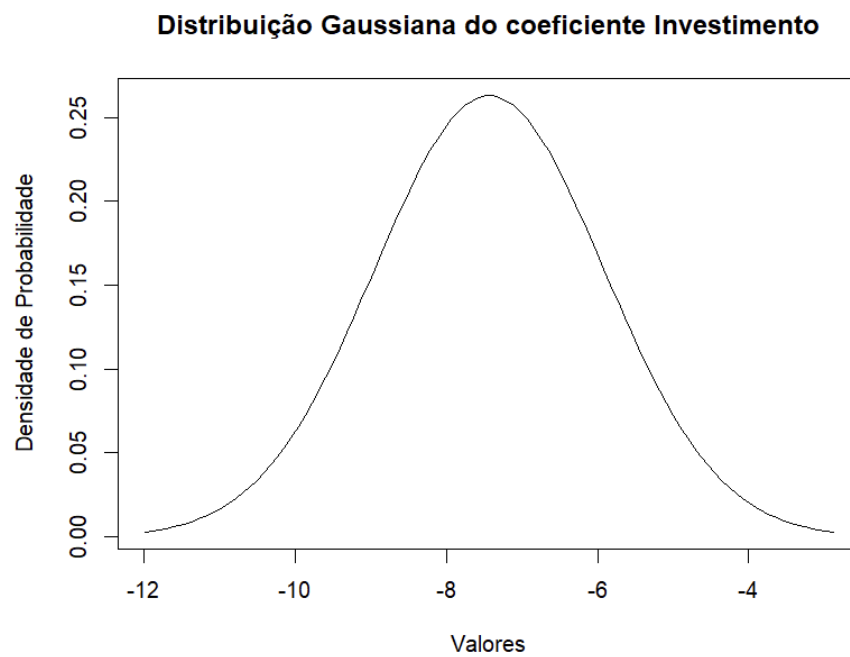


Figura 5.19: Distribuição gaussiana dos 100 melhores modelos do coeficiente Investimento

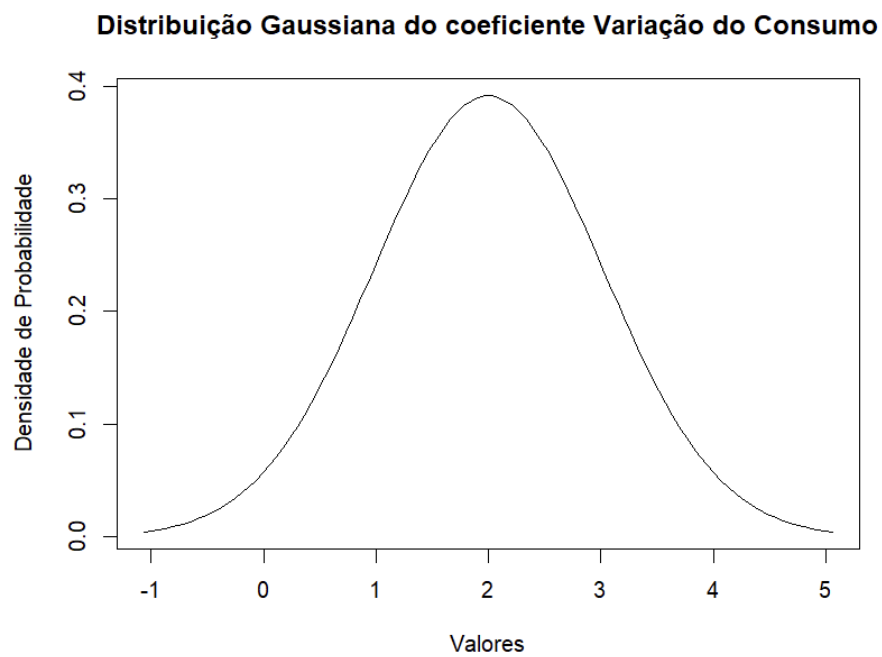


Figura 5.20: Distribuição gaussiana dos 100 melhores modelos do coeficiente Variação do Consumo

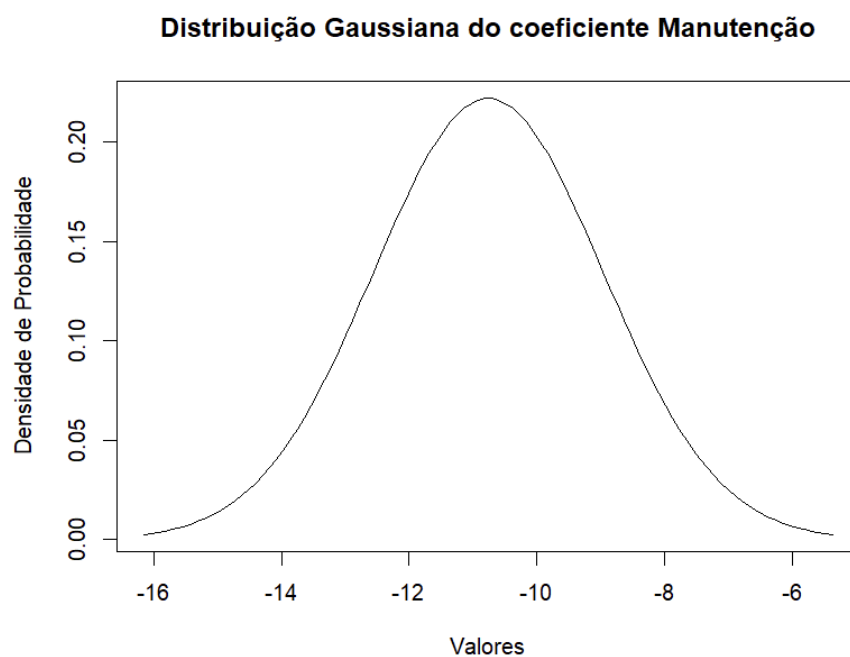


Figura 5.21: Distribuição gaussiana dos 100 melhores modelos do coeficiente Manutenção

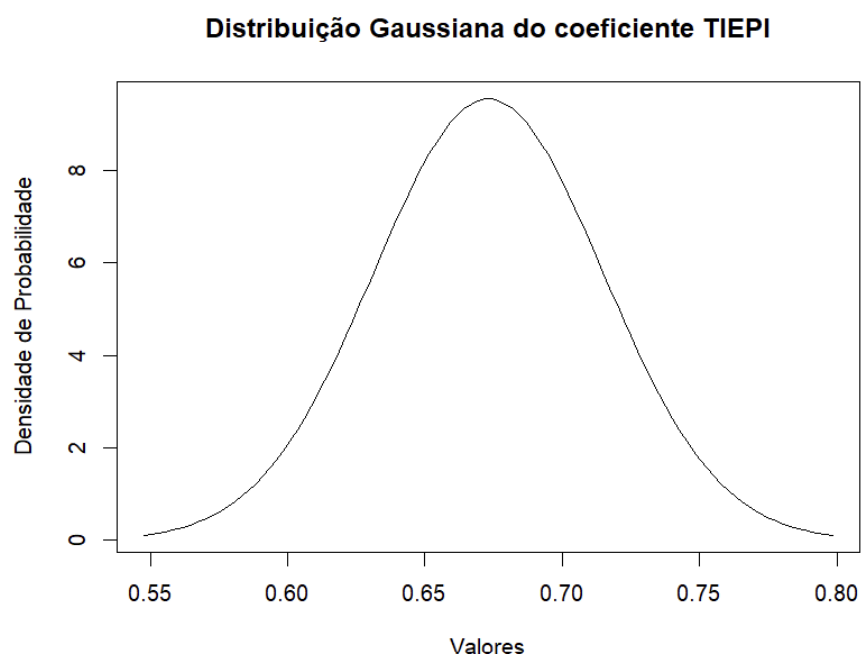


Figura 5.22: Distribuição gaussiana dos 100 melhores modelos do coeficiente TIEPI

Capítulo 6

Conclusões e Trabalho Futuro

A realização deste trabalho de dissertação possuía como principal objetivo o estudo probabilístico da QST da rede de distribuição. Decidiu-se realizar esse estudo para o TIEPI, índice mais importante, pelo que o objetivo passou a ser, principalmente, a produção de modelos para estimar este índice, algo que foi alcançado.

A QST é uma preocupação central dos operadores das redes de distribuição de energia elétrica. Com a realização desta dissertação foi possível concluir acerca da importância que o vetor da QST possui na entrega de energia elétrica às populações. Através do estudo probabilístico desse vetor é possível ao ORD priorizar e prever o impacto dos investimentos a efetuar, de modo a minimizar os tempos de interrupção de serviço e assim melhorar a qualidade de serviço da rede.

A técnica Spike and Slab mostrou-se ser robusta no que ao estudo da evolução dos índices do vetor da QST diz respeito, na medida em que permite determinar as margens de confiança associadas a um determinado resultado obtido, condição que é fundamental em planejamento, pois os valores previstos possuem sempre erros associados. Deste modo torna-se vantajoso apresentar os resultados não como um valor fixo, mas sim no formato de intervalo probabilístico.

No que ao desenvolvimento deste estudo diz respeito, foi possível observar que vários fatores podem ter influência no resultado final. Um dos aspectos mais importantes é a definição de *priors* de modo a obter resultados mais coerentes.

Um futuro desenvolvimento ao trabalho aqui exposto poderá ser a aplicação deste método aos restantes índices que compõem o vetor da QST, bem como, testar a influência do peso dos anos de treino e teste no resultado final.

Anexo A

Código desenvolvido

A.1 Código com a função *lm.spike* e a biblioteca *BoomSpikeSlab*

```
install.packages("scales")
library("scales")
install.packages("spikeslab")
library("spikeslab")
install.packages("bayesm")
library(bayesm)
library("BoomSpikeSlab")
install.packages("rstan")

setwd("C:\\Users\\zeped\\OneDrive\\Ambiente de Trabalho\\TESE")

wd <- getwd()

TARGETS <- read.csv(file.path(wd, "TARGETS_QST.csv"),
header=TRUE, sep= ";", dec=".")

rownames(TARGETS) <-
c("2008", "2009", "2010", "2011",
"2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019", "2020")

INPUTS <- read.csv(file.path(wd, "INPUTS.csv"),
header=TRUE, sep=";", dec = ".")

min<-1
max<-3
INPUTS_norm = INPUTS
```

```

for(i in 1:length(INPUTS)){
INPUTS_norm[,i] <-rescale(INPUTS[,i], to = c(min, max))
}

y=as.matrix(TARGETS$TIEPI[3:length(TARGETS$TIEPI)])

x=as.matrix(INPUTS_norm[c("Investimento_",
"Consumo","Total_s_Corretiva_","TIEPI_")])

teste = 0
nit=10000

model<-BoomSpikeSlab::lm.spike(y~x,niter=nit)
coeficientes <- model$beta
sse<-model$sse
MAPE_TREINO=0
erro2=0
NRMSE=0

for(i in 1:nrow(x)){
c <- coeficientes [,2:5]%*%x[i,]
c <- c+coeficientes[,1]
if(i==1){
y_estimado <- c
}else{
y_estimado <- cbind(y_estimado, c)
}
erro <- (abs(c-y[i,])/(y[i,]))*100
erro2 <- erro2 + ((c-y[i,])^2)
MAPE_TREINO = MAPE_TREINO + (erro/nrow(x))
}

colnames(y_estimado) <- c("2010","2011","2012",
"2013","2014","2015","2016","2017",
"2018","2019","2020")

NRMSE_TREINO=((sqrt(erro2/nrow(x)))/mean(y))*100
testel<- cbind(y_estimado, NRMSE_TREINO)
testel <- cbind(1:nrow(testel), testel)

```



```

colnames(testel) <- c("modelo", "2010", "2011", "2012",
"2013", "2014", "2015", "2016", "2017",
"2018", "2019", "2020", "NRMSE_TREINO")

teste2 <- cbind(testel, coeficientes)
testel_sorted <- testel[order(testel[, "NRMSE_TREINO"]), ]
selected_values <- apply(testel_sorted, 2, head, n=10)

SELECTED_T <- t(selected_values[, 2:12])
cores <- c("red", "blue", "green",
"orange", "purple", "brown", "yellow", "gray", "magenta", "pink", "cyan")

limite_x <- 2010:2020
plot(limite_x, TARGETS[3:nrow(TARGETS), 1], type="l", col="black",
ylim=range(selected_values[, 2:12]),
xlab="Anos de teste", ylab="TIEPI (min)", lwd=3)

for(i in 1:ncol(SELECTED_T)){
lines(limite_x, SELECTED_T[, i], col=cores[i])
}

```

A.2 Código com a função *brm* e a biblioteca *brms* de obtenção do modelo final

```

library(Metrics)
library(rstan)
library(StanHeaders)
library(brms)
library(ggplot2)

setwd("C:\\Users\\zeped\\OneDrive\\Ambiente de Trabalho\\TESE")
wd <- getwd()

TARGETS <- read.csv(file.path(wd, "TARGETS_QST.csv"),
header=TRUE, sep=";", dec=".")
rownames(TARGETS) <- c("2008", "2009", "2010", "2011", "2012",
"2013", "2014", "2015", "2016", "2017", "2018", "2019", "2020")

INPUTS <- read.csv(file.path(wd, "INPUTS.csv"), header=TRUE,

```

```

sep=";", dec = ".")
y=as.matrix(TARGETS$TIEPI[3:length(TARGETS$TIEPI)])

novos_inputs <- INPUTS[, c(1,2,3,5)]
x1 <- novos_inputs[, "Investimento_"]
x2 <- novos_inputs[, "Consumo"]
x3 <- novos_inputs[, "Total_s_Corretiva_"]
x4 <- novos_inputs[, "TIEPI_"]

test <- cbind(novos_inputs, y)
tes1 <- data.frame(test)
colnames(tes1) <- c("x1", "x2", "x3", "x4", "y")

tes1$x1 <- tes1$x1/1E6
tes1$x3 <- tes1$x3/1E6

tes1$x1 <- log10(tes1$x1+1)
tes1$x3 <- log10(tes1$x3+7.2)

prior_intercept <- prior(normal(0, 100), class=Intercept)
prior_x1 <- prior(normal(-10,2), class=b, coef=x1)
prior_x2 <- prior(normal(2,1), class=b, coef=x2)
prior_x3 <- prior(normal(-10,2), class=b, coef=x3)
prior_x4 <- prior(normal(0.8,0.1), class=b, coef=x4)

ano_teste <- 2
amostras <- nrow(tes1)
amostras_treino <- tes1[1:(amostras-ano_teste),]

modelo <- brm(y ~ x1 + x2 + x3 + x4 ,data = amostras_treino,
prior = c(prior_intercept, prior_x1, prior_x2, prior_x3, prior_x4))

fixef(modelo, probs = (c(0.05, 0.5, 0.95)))
previsão <- data.frame(predict(modelo, probs = c(0.05, 0.5, 0.95)))

erro <- (tes1$y-previsão$Estimate)
MSE <- (1/length(previsão$Estimate))*sum(((tes1$y)-(previsão$Estimate))^2)
SSE <- sum(((previsão$Estimate)-(tes1$y))^2)
RMSE <- sqrt(mean(((tes1$y-previsão$Estimate)^2)))
NRMSE <- (RMSE)/(mean(tes1$y))*100

```

```
MAPE <- (1/length(previsão$Estimate))*sum(abs((tes1$y)-
(previsão$Estimate))/abs(tes1$y))*100

#Para a ordenação e escolha dos melhores modelos
foram transportados os dados para o Excel

options(max.print = 10000)

model_df <- as.data.frame(modelo)

for(i in 1:length(model_df$b_x1)){
a<-model_df$b_Intercept[i]
b<-model_df$b_x1[i]
c<-model_df$b_x2[i]
d<-model_df$b_x3[i]
e<-model_df$b_x4[i]

coefs <- c(a,b,c,d,e)

if(i==1){
coefs_final <- coefs
}
else if(i>1){
coefs_final <- rbind(coefs_final,coefs)
}
}
coefs_final

library(clipr)
write_clip(coefs_final)
```


Referências

- [1] E-Redes. Plano de desenvolvimento e investimento da rede de distribuição, pdird-e 2020. Relatório técnico, E-Redes, 2022.
- [2] António Roque. Aula 5, estatística ii, 2008. URL: <http://sisne.org/Disciplinas/Grad/ProbEstat2/aula5.pdf>.
- [3] Hugo Francisco Rocha Costa. Planeamento de investimentos na rede de distribuição utilizando a técnica spike and slab. Tese de mestrado, MIEEC, FEUP, 2022.
- [4] Entidade Reguladora dos Serviços Energéticos (ERSE). Regulamento da qualidade de serviço. Relatório técnico, ERSE, 2023.
- [5] Entidade Reguladora dos Serviços Energéticos (ERSE). Qualidade de serviço técnica, formação ace. Relatório técnico, ERSE, 2021.
- [6] Direção Geral de Energia e Geologia. Relatório de monitorização da segurança de abastecimento do sistema elétrico nacional 2023-2040 (rmsa-e 2022). Relatório técnico, Direção Geral de Energia e Geologia, 2022.
- [7] Pedro Macedo e Nuno Fidalgo. Modelo de estimação do impacto do investimento e da manutenção na qualidade de serviço técnica. Relatório técnico, INESC TEC, 2021.
- [8] Pedro Macedo e Nuno Fidalgo. Estimação do impacto dos cenários de investimento na qualidade de serviço, na eficiência da rede, na eficiência operacional e no acesso a novos serviços. Relatório técnico, INESC TEC, 2020.
- [9] J. K. Ghosh M. DElampady T., Samanta. *An Introduction to Bayesian Analysis: Theory and Methods*. Springer, New York, 2006.
- [10] BIN DAI, SHILIN DING, e GRACE WAHBA. Multivariate bernoulli distribution. *Bernoulli*, 19(4):1465–1483, 2013. URL: <http://www.jstor.org/stable/23525760>.
- [11] Siddhartha Chib. Chapter 57 - markov chain monte carlo methods: Computation and inference. volume 5 de *Handbook of Econometrics*, páginas 3569–3649. Elsevier, 2001. URL: <https://www.sciencedirect.com/science/article/pii/S1573441201050103>, doi:[https://doi.org/10.1016/S1573-4412\(01\)05010-3](https://doi.org/10.1016/S1573-4412(01)05010-3).