



FRANCISCO NUNES JANUÁRIO DA SILVA

Licenciado em Engenharia Eletrotécnica e de Computadores

**AVALIAÇÃO DE SITUAÇÕES DE
SEPARAÇÃO DE MERCADO NO MIBEL E
DETERMINAÇÃO DE CAUSAS POSSÍVEIS**

MESTRADO EM ENGENHARIA ELETROTÉCNICA E DE COMPUTADORES

Universidade NOVA de Lisboa

Novembro, 2023



AVALIAÇÃO DE SITUAÇÕES DE SEPARAÇÃO DE MERCADO NO MIBEL E DETERMINAÇÃO DE CAUSAS POSSÍVEIS

FRANCISCO NUNES JANUÁRIO DA SILVA

Licenciado em Engenharia Eletrotécnica e de Computadores

Orientador: Nuno Manuel Ortega Amaro

Professor Auxiliar, FCT-NOVA

Júri:

Presidente: Daniel de Matos Silvestre

Professor Auxiliar, FCT-NOVA

Arguente: Francisco Alexandre Ganho da Silva Reis

REN Serviços SA

Orientador: Nuno Manuel Ortega Amaro

Professor Auxiliar, FCT-NOVA

Avaliação de situações de separação de mercado no MIBEL e determinação de causas possíveis

Copyright © Francisco Nunes Januário da Silva, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

AGRADECIMENTOS

Agradeço à minha família que durante estes 5 anos, mesmo estando longe de casa, fez de tudo para me apoiar. Sem eles não seria possível ter uma experiência académica que considero privilegiada, e por isso, estou-lhes eternamente grato.

Aos meus amigos e colegas que tornaram uma experiência de aprendizagem supostamente inconspícua, em algo do qual irei certamente sentir saudade. Foi com eles que senti não apenas a solidariedade académica, mas também uma diversão constante que serviu de contraste agradável à seriedade deste curso, algo que só eles me poderiam dar. Para eles tenho um enorme respeito e espero que a minha presença os tenha feito sentir da mesma forma.

Quero também agradecer ao meu orientador, professor Nuno Amaro, pelo seu apoio e orientação no desenvolvimento desta dissertação, e que sempre demonstrou disponibilidade e paciência face às minhas dúvidas. A ele, e aos vários professores desta instituição que se cruzaram com o meu percurso académico, expresso uma enorme gratidão.

*«Doubt is not a pleasant condition,
but certainty is absurd. » (Voltaire)*

RESUMO

O Mercado Ibérico de Energia Elétrica, MIBEL, permite a integração dos sistemas elétricos português e espanhol num mercado de eletricidade comum. Devido às limitações físicas das interligações entre as zonas de mercado destes dois países, nem sempre é possível realizar a totalidade das trocas determinadas em mercado, pois a capacidade de interligação é superada. Nestas situações, é aplicado o mecanismo de gestão conjunta das interligações entre Portugal e Espanha, conhecido como separação de mercado, que resulta no cálculo dos preços de eletricidade com base nas curvas de procura e oferta específicas a cada zona de mercado, resultando em preços diferentes. O principal objetivo desta dissertação é o de identificar as principais causas de separação de mercado considerando o contexto atual do MIBEL.

É feita uma análise inicial da evolução de separação de mercado e são identificados alguns padrões de ocorrência para o período de 2017 a 2023. Foram também estudados os graus de correlação entre as variáveis de estudo e a diferença de preços entre Portugal e Espanha em momentos de separação de mercado.

A separação de mercado é modelada como uma variável binária alvo, que conjuntamente com os dados de produção de energia elétrica discriminados por tecnologia, consumos nacionais e capacidades de interligação, são utilizados na construção de três modelos explicativos distintos: Regressão Logística, Random Forest e Rede Neuronal Artificial. Utilizam-se técnicas de Explainable Machine Learning, sempre que adequado, como forma de extrair conhecimento acerca dos determinantes da separação de mercado destes modelos.

Por fim são analisados os resultados obtidos e são feitas algumas recomendações para a redução da separação de mercado tendo em conta os futuros desenvolvimentos dos sistemas elétricos português e espanhol.

ABSTRACT

The Iberian Electricity Market, MIBEL, allows the Portuguese and Spanish electricity systems to be integrated into a common electricity market. Due to the physical limitations of the interconnections between the market areas of these two countries, it is not always possible to realise all the exchanges determined in the market, as the interconnection capacity is exceeded. In these situations, the mechanism for joint management of the interconnections between Portugal and Spain, known as market splitting, is applied, which results in electricity prices being calculated on the basis of demand and supply curves specific to each market area, resulting in different prices. The main objective of this dissertation is to identify the main causes of market splitting in the current context of MIBEL.

An initial analysis of the evolution of market splitting is made and some patterns of occurrence are identified for the period of 2017 to 2023. The degrees of correlation between the study variables and the price difference between Portugal and Spain at times of market splitting were also studied.

Market splitting is modelled as a binary target variable, which together with electricity generation data broken down by technology, national demand and interconnection capacities, is used to build three different explanatory models: Logistic Regression, Random Forest and Artificial Neural Network. Where appropriate, Explainable Machine Learning techniques are used to extract knowledge about the determinants of market splitting from these models.

Finally, the results obtained are analysed and some recommendations are made for reducing market splitting, taking into account future developments in the Portuguese and Spanish electricity systems.

ÍNDICE

Índice de Figuras	x
Índice de Tabelas	xii
Siglas	xiv
1 Introdução	1
1.1 Enquadramento	1
1.2 Motivação e Objetivos	2
1.3 Estrutura do documento	3
2 Revisão da Literatura	4
2.1 Setor Elétrico	4
2.1.1 Modelos Organizacionais	4
2.1.2 Diretivas europeias	6
2.2 Mercados de eletricidade	7
2.2.1 Bolsas de eletricidade (pool)	7
2.2.2 Acoplamento de mercados	10
2.3 O caso de Portugal e Espanha	11
2.3.1 Reforma do setor elétrico português	11
2.3.2 Reforma do setor elétrico espanhol	12
2.3.3 Mercado Ibérico de Eletricidade	13
2.4 Desenvolvimento de fontes de energia renováveis na Península Ibérica	17
2.5 Separação de mercado e convergência de preços nos mercados europeus de eletricidade	18
2.6 Explainable Machine Learning	20
3 Tratamento de dados e análise estatística da separação de mercado	23
3.1 Recolha e tratamento de dados	23
3.2 Evolução da separação de mercado	29

3.3	Análise de padrões de ocorrência de separação de mercado	33
3.4	Análise de correlação	37
3.4.1	Coeficientes de correlação	37
3.4.2	Correlação entre variáveis de estudo e spreads de preços de eletricidade	38
4	Metodologia	42
4.1	Divisão do conjunto de dados	42
4.2	Regressão Logística	43
4.2.1	Introdução teórica	43
4.2.2	Implementação	44
4.3	Random Forest	45
4.3.1	Introdução teórica	45
4.3.2	Implementação	46
4.4	Rede Neuronal Artificial	46
4.4.1	Introdução teórica	47
4.4.2	Implementação	50
4.5	Métricas de avaliação dos modelos	51
4.6	Técnicas de Explainable Machine Learning	54
5	Resultados	57
5.1	Avaliação dos modelos	57
5.1.1	Regressão Logística	57
5.1.2	Random Forest	59
5.1.3	Rede Neuronal Artificial	59
5.2	Coeficientes estimados e importâncias	60
5.2.1	Regressão Logística	60
5.2.2	Random Forest	62
5.2.3	Rede Neuronal Artificial	63
5.3	Partial Dependence Plots	64
5.3.1	Capacidade de transferência disponível nas interligações	65
5.3.2	Consumo	66
5.3.3	Produção eólica	67
5.3.4	Produção hídrica	68
5.3.5	Produção solar	69
5.3.6	Produção nuclear	70
5.3.7	Produção com Ciclo Combinado de Gás Natural	71
5.3.8	Produção com Gás Natural	72
5.3.9	Produção com Carvão	73
5.4	Aplicação do algoritmo LIME na Rede Neuronal	74
5.5	Comparação entre modelos	76

6	Discussão e conclusões	77
6.1	Discussão geral dos resultados	77
6.2	Soluções para a redução das ocorrências de separação de mercado	81
6.2.1	Aumento das capacidades de transferência disponíveis	82
6.2.2	Equilíbrio dos mixes energéticos	82
6.3	Trabalho futuro	83
6.4	Conclusões	84
	Bibliografia	85

ÍNDICE DE FIGURAS

2.1	<i>Modelo de competição no mercado grossista, adaptado de [7]</i>	5
2.2	<i>Modelo de competição no mercado retalhista, adaptado de [7]</i>	6
2.3	Pool assimétrica e simétrica, adaptado de [10]	9
2.4	Preço de equilíbrio de mercado, adaptado de [11]	9
2.5	Estrutura do Sistema Elétrico Nacional após a reforma de 1995, retirado de [14]	12
2.6	Esquema organizacional do Operador de Mercado Ibérico, retirado de [9] .	14
2.7	Localização das interligações entre Portugal e Espanha, retirado de [20] . . .	16
2.8	Previsão da capacidade de interligação para fins comerciais em 2021, retirado de [18]	17
2.9	Troca interpretabilidade-exatidão hipotética para os modelos utilizados . .	22
3.1	Exemplo da transformação dos dados recolhidos	25
3.2	Séries de tempo para a produção eólica no período de 2017 a 2023	27
3.3	Séries de tempo para a produção solar no período de 2017 a 2023	28
3.4	Séries de tempo para as capacidades de interligação no período de 2017 a 2023	29
3.5	Evolução anual da separação de mercado	30
3.6	Contagem das horas em separação de mercado num mês rolante	31
3.7	Histograma dos spreads	32
3.8	Evolução do país importador durante a separação de mercado	33
3.9	Mapa de calor de todas as ocorrências de separação de mercado (2017 a 2022) .	34
3.10	Mapa de calor das ocorrências de separação de mercado no sentido Espanha-Portugal (2017 a 2022)	34
3.11	Mapa de calor das ocorrências de separação de mercado no sentido Portugal-Espanha (2017 a 2022)	35
3.12	Perfis de carga médios para fevereiro de 2017	36
3.13	Mapa de calor de todas as ocorrências de separação de mercado (2022)	37
3.14	Gráficos de dispersão de algumas variáveis versus os spreads de preços de eletricidade	40
4.1	Arquitetura de um Multilayer Perceptron	47

4.2	Estrutura de um neurónio	48
4.3	Função de ativação ReLU	49
4.4	Função de ativação sigmoide	49
4.5	Curvas de treino e validação	51
4.6	Matriz de confusão	52
5.1	Diminuição média na impuridade por variável do modelo de Random Forest	63
5.2	Partial Dependence Plots da probabilidade de separação de mercado e da capacidade de transferência disponível nas interligações	65
5.3	Partial Dependence Plots da probabilidade de separação de mercado e dos consumos nacionais	66
5.4	Partial Dependence Plots da probabilidade de separação de mercado e das produções eólicas	67
5.5	Partial Dependence Plots da probabilidade de separação de mercado e das produções hídricas	68
5.6	Partial Dependence Plots da probabilidade de separação de mercado e das produções solares	69
5.7	Partial Dependence Plots da probabilidade de separação de mercado e da produção nuclear espanhola	70
5.8	Partial Dependence Plots da probabilidade de separação de mercado e produção em ciclo combinado de gás natural	71
5.9	Partial Dependence Plots da probabilidade de separação de mercado e da produção com gás natural	72
5.10	Partial Dependence Plots da probabilidade de separação de mercado e produção com carvão	73
5.11	Explicações LIME	75
5.12	Resultado da previsão dos modelos para o período de 2022 e início de 2023 na forma de contagem de horas em separação de mercado num mês rolante	76
6.1	Séries de tempo da agregação das contribuições associadas às variáveis de baixo custo marginal portuguesas (eólica, solar e hídrica)	80

ÍNDICE DE TABELAS

2.1	Exemplos de bolsas de eletricidade, retirado de [8]	8
3.1	Variáveis de estudo	26
3.2	Coefficientes de correlação	39
4.1	Hiperparâmetros otimizados do algoritmo Random Forest implementado	46
4.2	Hiperparâmetros otimizados da Rede Neuronal implementada	51
5.1	Métricas de avaliação do modelo Logit para a separação de mercado no sentido Espanha-Portugal	58
5.2	Métricas de avaliação do modelo Logit para a separação de mercado no sentido Portugal-Espanha	58
5.3	Métricas de avaliação do modelo Random Forest	59
5.4	Métricas de avaliação do modelo de Rede Neuronal	60
5.5	Coefficientes estimados para o modelo Logit da separação de mercado no sentido Espanha-Portugal	61
5.6	Coefficientes estimados para o modelo Logit da separação de mercado no sentido Portugal-Espanha	62
5.7	Importâncias calculadas para o modelo de Rede Neuronal	64
5.8	Avaliação dos modelos para o período de 2022 e início de 2023	76

SIGLAS

AAC	Already Allocated Capacity (<i>p. 10</i>)
ATC	Available Transfer Capacity (<i>pp. 10, 19, 24, 28, 44, 54, 79</i>)
CACM	Capacity Allocation and Congestion Management (<i>pp. 15, 16</i>)
CAE	Contratos de Aquisição de Energia (<i>p. 4</i>)
CMVM	Comissão do Mercado de Valores Mobiliários (<i>p. 13</i>)
CNE	Comisión Nacional de Energía (<i>p. 13</i>)
CNMV	Comisión Nacional del Mercado de Valores (<i>p. 13</i>)
CNSE	Comisión Nacional del Sistema Eléctrico (<i>pp. 12, 13</i>)
CUR	Comercializador de Último Recurso (<i>p. 12</i>)
CWE	Central West Europe (<i>pp. 20, 31, 33, 43, 45</i>)
DGEG	Direção-Geral de Energia e Geologia (<i>p. 18</i>)
DLR	Dynamic Line Rating (<i>pp. 82, 84</i>)
EDP	Energias de Portugal (<i>p. 11</i>)
ENTSO-E	European Network of Transmission System Operators for Electricity (<i>p. 83</i>)
EPEX	European Power Exchange (<i>pp. 10, 11</i>)
ERSE	Entidade Reguladora dos Serviços Energéticos (<i>pp. 11–13</i>)
EUPHEMIA	EU Pan-European Hybrid Electricity Market Integration Algorithm (<i>p. 14</i>)
FBMC	Flow-Based Market Coupling (<i>p. 10</i>)
IA	Inteligência Artificial (<i>p. 20</i>)
LIME	Local Interpretable Model-Agnostic Explanations (<i>pp. 55, 56, 74</i>)
LOSEN	Ley Orgánica del Sector Eléctrico Nacional (<i>p. 12</i>)

MIBEL	Mercado Ibérico da Energia Elétrica (<i>pp. 1–3, 10, 11, 13–15, 19, 31, 42, 43, 56, 77, 84</i>)
MIE	Mercado Interno de Energia (<i>pp. 1, 6, 7, 10, 13, 14</i>)
ML	Machine Learning (<i>pp. 20, 21, 42, 46, 47, 50, 54, 55, 78, 83, 84</i>)
MLE	Marco Legal y Estable (<i>p. 12</i>)
NEMO	Nominated Electricity Market Operator (<i>pp. 10, 14</i>)
NTC	Net Transfer Capacity (<i>p. 10</i>)
OMI	Operador de Mercado Ibérico (<i>pp. 1, 14</i>)
OMIE	OMI pólo espanhol (<i>pp. 1, 14, 15, 23</i>)
OMIP	OMI pólo português (<i>pp. 1, 14</i>)
ORD	Operador da Rede de Distribuição (<i>p. 6</i>)
ORT	Operador da Rede de Transporte (<i>pp. 1, 5–7, 10, 13, 15, 24</i>)
PCR	Price Coupling of Regions (<i>pp. 10, 11, 14</i>)
PDP	Partial Dependence Plot (<i>pp. 54–56, 64–73, 81</i>)
PNEC	Plano Nacional de Energia e Clima (<i>pp. 17, 18</i>)
PRE	Produção em Regime Especial (<i>p. 11</i>)
REE	Red Eléctrica de España (<i>pp. 13, 15, 25</i>)
REN	Rede Eléctrica Nacional (<i>pp. 11, 15, 24, 25</i>)
RES-E	eletricidade gerada via fontes de energia renováveis (<i>pp. 2, 18–20</i>)
RNA	Rede Neuronal Artificial (<i>pp. 21, 46, 47, 56</i>)
SDAC	Single Day-ahead Coupling (<i>p. 14</i>)
SEI	Sistema Eléctrico Independente (<i>p. 11</i>)
SEN	Sistema Eléctrico Nacional (<i>pp. 11, 12</i>)
SENV	Sistema Eléctrico Não Vinculado (<i>pp. 11, 12</i>)
SEP	Sistema Eléctrico de Serviço Público (<i>p. 11</i>)
SHAP	SHapley Additive exPlanations (<i>p. 55</i>)
SIDC	Single Intraday Coupling (<i>p. 15</i>)
SLR	Seasonal Line Rating (<i>p. 82</i>)
TSO	Transmission System Operator (<i>p. 6</i>)
UE	União Europeia (<i>pp. 2, 6, 17, 18, 81</i>)
XAI	Explainable Artificial Intelligence (<i>p. 21</i>)
XML	Explainable Machine Learning (<i>pp. 3, 21, 42, 54, 55, 74, 79, 84</i>)

INTRODUÇÃO

1.1 Enquadramento

O estabelecimento do [Mercado Ibérico da Energia Elétrica \(MIBEL\)](#), que entrou em pleno funcionamento a 1 de julho de 2007, promoveu a integração dos sistemas elétricos português e espanhol, na forma de um mercado comum ao nível ibérico. Este foi fruto de um esforço cooperativo entre os governos destes países, que no âmbito dos potenciais benefícios para o consumidor final, estabeleceram um mercado livre e competitivo na dimensão da península Ibérica, constituindo assim, tal como o primeiro mercado regional europeu transnacional que o antecedeu, NordPool, mais um passo na construção do [Mercado Interno de Energia \(MIE\)](#).

Com a consolidação deste mercado de eletricidade, foi necessário recorrer à harmonização de um conjunto de procedimentos, regras e condições económicas/técnicas entre os dois países. Foi então definido o [Operador de Mercado Ibérico \(OMI\)](#) responsável pela gestão dos mercados organizados do [MIBEL](#), estruturado num [OMI pólo português \(OMIP\)](#) e [OMI pólo espanhol \(OMIE\)](#). O [OMIP](#) é a entidade responsável pelo mercado a prazo, enquanto o [OMIE](#) é o operador de mercado designado, responsável pela administração dos mercados diário e intradiário. Associado aos mercados geridos pelo [OMIE](#) existe um mecanismo de gestão conjunta das interligações entre Portugal e Espanha, nomeadamente a [Separação de Mercados \(Market Splitting\)](#). Este é um mecanismo de atribuição da capacidade de interligação a curto prazo, ou seja, num horizonte diário e intradiário, que tem em conta o total de ofertas de compra e venda apresentadas em mercado, tal como a capacidade disponível de interligação, comunicada pelo respetivo [Operador da Rede de Transporte \(ORT\)](#). Este mecanismo estabelece dois resultados possíveis para o mercado diário:

- A capacidade de interligação não é superada – Haverá um programa de interligação que corresponde ao encontro de ofertas de compra e venda, e existe um único preço de equilíbrio para as duas zonas do mercado;
- A capacidade de interligação é superada – A diferença de preços resultante do encontro de ofertas de compra e venda resulta num fluxo físico de energia que

excede a capacidade disponível no mesmo sentido. O mercado é separado em duas zonas, e são calculados dois preços diferentes. Nesta situação diz-se que se está em regime de market splitting; O trabalho proposto pretende analisar as situações em que se verifica um regime de market splitting, com o objetivo de identificar os seus principais determinantes e as soluções que evitam a sua ocorrência.

1.2 Motivação e Objetivos

Os sistemas elétricos europeus estão a passar por mudanças profundas e contínuas. Com a penetração cada vez maior da [eletricidade gerada via fontes de energia renováveis \(RES-E\)](#), consequência do aumento das capacidades instaladas destas tecnologias, surgem novos desafios tanto ao nível técnico como de design dos mercados de eletricidade [1]. O “2030 Climate and Energy framework”, o sucessor do pacote legislativo da [União Europeia \(UE\)](#) para o clima e energia referente ao ano de 2020, estabeleceu um conjunto de metas e medidas no âmbito de tornar a economia e sistemas de energia da [UE](#) mais competitivos, seguros e, tendo por base o Acordo de Paris, mais sustentáveis [2]. Este foi discutido e aprovado pelo Conselho Europeu no ano de 2014, resultando em algumas metas ambiciosas para o ano de 2030, como 32% do consumo de energia ser de origem renovável, e a renovação do objetivo de 10% do pico de consumo como capacidade de interligação entre sistemas elétricos para 15%. O mais recente pacote legislativo da [UE](#), Fit-for-55, intensifica estas medidas, algo que se verifica notavelmente no aumento da meta de 32% de energia de fontes renováveis no mix energético para 40% até 2030, uma das medidas adotadas como forma de alcançar a meta de redução na emissão de gases com efeito de estufa em 55% [3].

No sentido de corresponder às políticas e metas estabelecidas ao nível europeu nas últimas duas décadas, alguns estados-membros da [UE](#) criaram mecanismos atrativos de suporte e financiamento para o desenvolvimento e implementação de tecnologias de produção de energia via fontes renováveis [1]. Países como Portugal e Espanha viram aumentar a proporção de [RES-E](#) face às restantes tecnologias. No caso português, é possível verificar um aumento de cerca de 40% na potência instalada do tipo renovável face ao ano de 2012, sendo a tecnologia que mais cresceu, em termos relativos, a solar fotovoltaica [4]. Paralelamente a estes desenvolvimentos, o comportamento da separação de mercado observado no [MIBEL](#) manteve-se em linha com a tendência de redução de ocorrência ao nível anual, identificada em estudos prévios como [5]. Concretamente, a incidência de separação de mercado passou de representar uma média de 7,5% das horas observadas (de 2011 a 2015) para 4,8% (de 2017 a 2021) [6].

Tendo em conta o contexto atual do [MIBEL](#), consequência de políticas ao nível europeu e do crescimento nas vertentes de produção renovável e na capacidade de interligação, a tese pretende reavaliar os determinantes de separação de mercado com base num inquérito bibliográfico justaposto ao estudo a realizar com dados de 2017 a 2021. Posteriormente, e com base nas conclusões obtidas nos passos anteriores, pretende-se avaliar as soluções

que permitiram/permitem a redução da ocorrência de separação de mercado, aplicado ao [MIBEL](#). Dados os objetivos descritos, deve-se considerar as seguintes tarefas:

- Elaborar um inquérito bibliográfico acerca dos determinantes de separação de mercado, integração de mercados e congestionamentos nas interligações;
- Efetuar uma análise estatística dos dados referentes ao [MIBEL](#);
- Criar modelos explicativos, aplicados ao contexto recente do [MIBEL](#), que permitam a análise da influência de cada uma das variáveis relevantes;

1.3 Estrutura do documento

Este documento está estruturado em secções que ilustram as diferentes dimensões do trabalho realizado. No início, **Capítulo 1**, do qual esta descrição faz parte, foi feita um pequeno enquadramento da separação de mercado no contexto do [MIBEL](#) seguido da introdução da motivação e objetivos para este trabalho. De seguida, no **Capítulo 2**, foi feita uma revisão da literatura à volta do tópico de separação de mercado, nomeadamente uma descrição do mercado de eletricidade e da sua evolução no contexto europeu e da península ibérica, os desenvolvimentos da produção renovável na região, a separação de mercado ou divergência de preços, e o tópico de [Explainable Machine Learning \(XML\)](#) que está relacionado com as metodologias utilizadas em estudos deste tipo. No **Capítulo 3** é apresentada uma descrição do processo de recolha e tratamento dos dados utilizados neste estudo e é feita uma análise estatística, em várias dimensões, da separação de mercado no contexto atual. Seguidamente, no **Capítulo 4**, são apresentados os algoritmos que foram utilizados neste estudo, tal como as técnicas que foram aplicadas na sua avaliação e extração de conhecimento. No **Capítulo 5**, são expostos e apreciados os resultados dos modelos desenvolvidos e é interpretado o conhecimento extraído dos mesmos, na forma de medidas de importância e gráficos de dependência parcial. Por fim, no **Capítulo 6**, é feita uma discussão geral dos resultados obtidos, consideram-se algumas soluções para a redução de separação de mercado e são dadas algumas sugestões para trabalho futuro e conclusões.

REVISÃO DA LITERATURA

2.1 Setor Elétrico

2.1.1 Modelos Organizacionais

O setor elétrico, nas suas variadas interpretações, é constituído por um conjunto de atividades essenciais e que podem ser categorizadas em dois tipos. O primeiro diz respeito às atividades físicas da indústria, sendo estas a produção da eletricidade, o seu transporte, distribuição e a necessária operação do sistema. O segundo está associado às funções de mercado como a participação e operação dos mercados grossista e retalhista [7]. O conceito de competição no setor geralmente aplica-se apenas às atividades de produção de eletricidade e de mercado, visto que as restantes atividades físicas são incompatíveis com tal noção. As atividades de transporte e distribuição, devido às suas necessidades de investimento, manutenção e operação, incorrem a modelos de monopólio natural, que segundo a teoria económica, são mais eficientes e viáveis do que a abertura à concorrência e a consequente implicação de estabelecimento de diferentes redes e operadores [7].

Até à década de 70, o modelo organizacional tradicional do setor elétrico, adotado na altura de uma forma quase universal, foi praticamente incontestado. Este, na sua forma genérica, consistia num monopólio verticalmente integrado. Todas as funções da indústria eram agrupadas e reguladas por uma única entidade. Este modelo de monopólio, embora permitisse aos poderes públicos implementar políticas de natureza estratégica e social, como a uniformidade tarifária, não permitia competição dentro do setor [8].

Uma evolução do modelo descrito vem na forma do modelo de comprador único. Neste, um único agente - normalmente a empresa concessionária incumbente ou o operador da rede de transporte, caso esteja autorizado - está encarregue de comprar energia elétrica aos vários produtores, sendo que estes podem ser propriedade do próprio comprador. O modelo introduz uma componente de competição, embora muito limitada, através de concursos para atribuição de direitos de construção e operação de centros produtores. Os produtores e o comprador único estabelecem [Contratos de Aquisição de Energia \(CAE\)](#) de longa duração (dezenas de anos), que protegem os primeiros dos vários riscos associados ao mercado. Neste caso, o comprador único continua a deter o monopólio das redes e dos

clientes finais, e não permite o acesso à rede por terceiros [8].

Um terceiro modelo – Figura 2.1 – permite a competição entre os produtores ao nível de um mercado grossista. Os produtores oferecem a sua energia numa bolsa de eletricidade (pool) na qual os distribuidores fazem propostas de compra, também existindo a possibilidade de estabelecerem acordos bilaterais de venda direta. Alguns consumidores, denominados de consumidores elegíveis, que consomem em excesso de um determinado patamar, estão habilitados a adquirir energia no mercado grossista ou contactar diretamente com os produtores. Este modelo exige a autonomização da rede de transporte no plano de gestão, ou preferencialmente, no plano acionista. Se se tratar do segundo caso, o **ORT** está integrado na empresa proprietária da rede de transporte [8]. É de se notar que a atividade de gestão da bolsa exige a criação de um operador de mercado, que no caso de obrigatoriedade na participação na bolsa, pode ser o próprio operador do sistema. A possibilidade de contratos bilaterais também implica a existência de um operador dedicado.

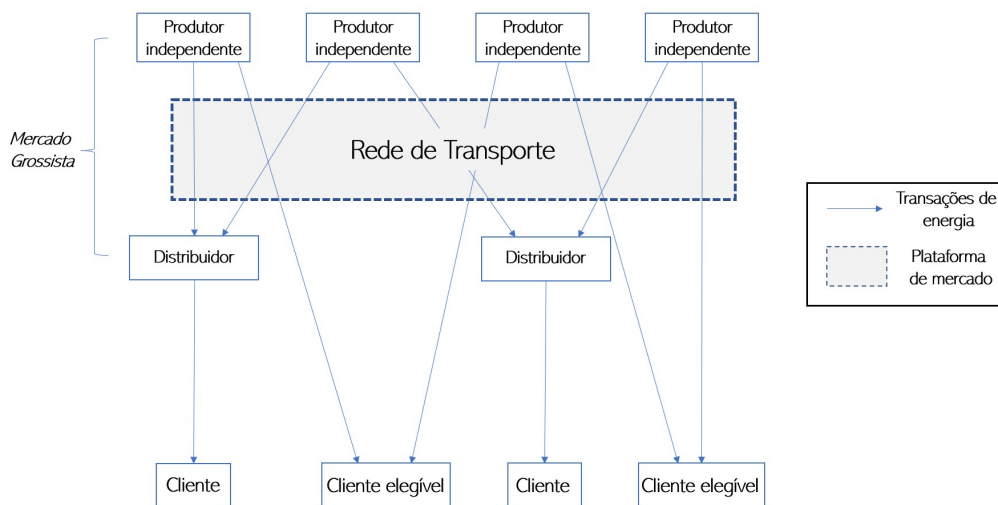


Figura 2.1: Modelo de competição no mercado grossista, adaptado de [7]

O modelo de competição no retalho, representado na Figura 2.2, é uma evolução do modelo anterior, no qual o consumidor é livre de escolher o seu fornecedor, algo que requer acesso livre às redes de transporte e distribuição. À semelhança do modelo de competição no retalho, o fornecedor pode ser um produtor, no caso de consumidores de grande escala, ou um comercializador, que participa nos mercados retalhista e grossista. É então estabelecido um operador da rede de distribuição, com características semelhantes às do **ORT**, que é independente da atividade de comercialização [8].

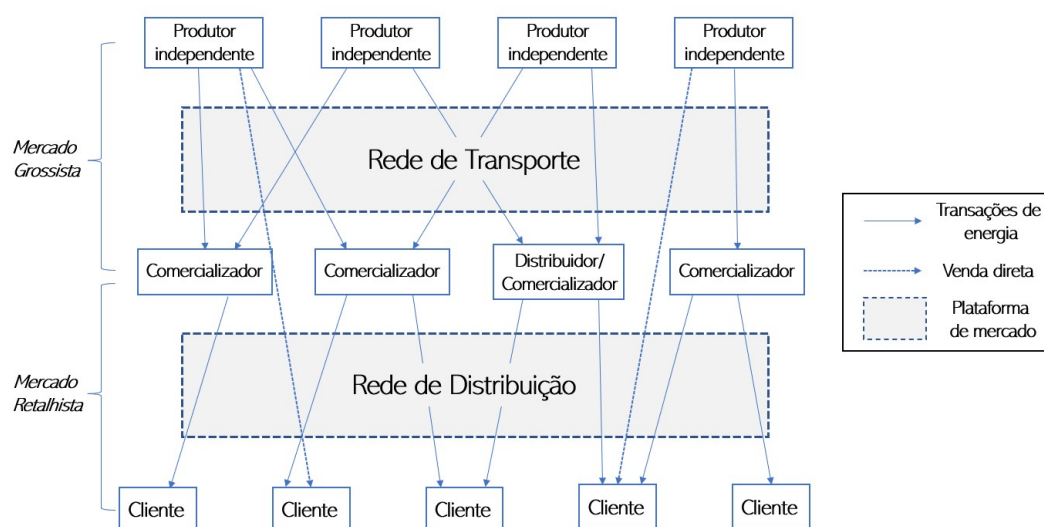


Figura 2.2: Modelo de competição no mercado retalhista, adaptado de [7]

2.1.2 Diretivas europeias

No âmbito de estabelecer um mercado de eletricidade único ao nível europeu, a Comissão Europeia criou diretivas que reestruturaram os setores elétricos dos Estados-membros da UE e definiram as suas respetivas trajetórias de evolução. A primeira contribuição significativa na criação do MIE deu-se em 1996 com a publicação da Diretiva 96/92/CE. Esta promoveu a liberalização e harmonização dos setores elétricos dos Estados-membros europeus através do estabelecimento de uma organização para o setor elétrico, semelhante à do modelo de competição no mercado grossista abordado na secção anterior. No entanto, foi dada a escolha de optar por um modelo de comprador único, que poderia ser mais conveniente para certos Estados-membro com certas limitações, tais que impedissem uma transição completa para o modelo de competição no mercado. Foi ainda estipulado que estes podem impor obrigações de serviço público às empresas do setor, em específico, nas vertentes de segurança e proteção do ambiente [8].

Referente à atividade de transporte, a Diretiva tornou mandatário o estabelecimento do **ORT (Transmission System Operator (TSO))** na terminologia europeia), responsável pela exploração, manutenção e expansão da rede de transporte, incluindo a criação de novas interligações com outras redes. A responsabilidade do despacho das instalações de produção também cai sobre o **ORT**, que por sua vez deve atuar de forma imparcial, evitando medidas que prejudiquem os utilizadores da rede ou favoreçam os seus acionistas. A diretiva não exigiu uma separação total no plano acionista do **ORT**, no entanto, estipulou que o mesmo deve ser independente de todas as outras atividades do setor, pelo menos ao nível do plano de gestão. Ao nível da distribuição também foi necessária a criação do **Operador da Rede de Distribuição (ORD)**, com características funcionais semelhantes à do **ORT** [8].

Em 2001 foi apresentada uma proposta de alteração à Diretiva 96/92/CE com objetivo de acelerar a criação do MIE, que resultou na publicação em 2003 da Diretiva 2003/54/CE [8]. À semelhança da diretiva anterior, esta aponta para a implementação de um novo modelo organizacional do setor elétrico, nomeadamente, o modelo de competição no retalho [7]. Nela também são claramente definidas as atividades de produção, transporte, distribuição e comercialização de eletricidade. O ORT é responsável pelo despacho dos centro produtores, determinação da capacidade das interligações e pelos mecanismos de ajuste entre a produção e o consumo, no entanto, é agora obrigado a ser independente das entidades associadas à produção, distribuição e comercialização. Esta independência diz respeito ao nível legal, organizacional e decisional, pelo que o ORT pode ainda pertencer a uma empresa verticalmente integrada. São também estipulados os critérios de existência de tarifas públicas de acesso às redes por parte de terceiros [8].

Esta diretiva definiu um planeamento de abertura faseada dos mercados envolvidos. Em 1 de julho de 2004, os consumidores comerciais e industriais, obtiveram a liberdade de escolher os seus fornecedores e, a partir de 1 de julho de 2007, todos os consumidores passaram a ter esse direito. É obrigatória a criação de entidades reguladoras independentes, que podem fixar ou aprovar os termos e condições referentes à ligação e acesso às redes (tarifas de transporte e distribuição), definir mecanismos de gestão de congestionamentos na respetiva rede nacional e interligações internacionais, e ainda, criar mecanismos de ajuste [8].

2.2 Mercados de eletricidade

2.2.1 Bolsas de eletricidade (pool)

Dado que a eletricidade não poder ser armazenada, no sentido convencional, esta também não pode ser transacionada num mercado spot da mesma forma que outras mercadorias indiferenciadas (commodities). Num mercado spot real, o despacho do bem transacionado é imediato visto que se recorre a reservas armazenadas, que por sua vez são repostas em tempo real pelos produtores. As transações nos mercados de eletricidade estão então sujeitas a contratos que antecedem a entrega física, geralmente na questão de um dia a minutos antes, e que seguem previsões de consumo. Desta forma, surgem desequilíbrios entre os valores contratados, baseados em previsões, e os valores observados de produção e consumo. Existe então uma necessidade de ajuste através de procedimentos que possibilitam uma componente adicional de competitividade, por exemplo, os mercados de balanço [8].

A bolsa de eletricidade, ou pool, corresponde a uma aproximação da definição de mercado spot. As regras que definem o seu funcionamento podem variar em certos aspetos, como por exemplo, a obrigatoriedade de participação. Os preços de compra e as quantidades a fornecer são, na grande maioria dos casos, fixados ex-ante através de

leilões competitivos baseados na carga programada. Os preços finais de venda são fixados tendo em conta os preços de compra, custos de serviço de sistema, perdas e outros custos associados à conservação de energia ou apoio às energias renováveis. Como alternativa, é possível fixar os preços em ex-post, ou seja, a partir do fornecimento e procura efetivamente verificados. Neste caso, o preço de aquisição da pool coincide com o preço de venda [8]. As propostas de venda podem ser simples, representadas por pares quantidade-preço e sem qualquer dependência temporal, ou seja, não dependem de propostas emitidas pela mesma entidade para outros intervalos de tempo. Este tipo de proposta é notavelmente simples, mas ignora algumas das restrições presentes no lado do produtor. As propostas de venda também podem ser complexas, se o mercado em que se inserem assim permitir, contendo estes parâmetros adicionais relacionados com as condições de operação dos grupos geradores, nomeadamente, valores mínimos de produção, rampas de subida e descida e o requisito de remuneração mínima [9].

A componente de propostas de oferta de compra, considerando uma fixação de preço ex-ante, gera dois modelos possíveis para a bolsa de eletricidade, sendo estes a pool assimétrica e simétrica. Numa pool assimétrica apenas são feitas ofertas de venda de energia elétrica, pelo que os comercializadores e consumidores elegíveis revelam as suas necessidades de consumo esperadas para os respetivos intervalos em negociação [10]. Neste caso a procura é considerada inelástica em relação ao preço, e a respetiva característica será representada por uma reta vertical [8], como pode ser verificado na Figura 2.3. O operador de mercado aceita as ofertas de venda, partindo da mais barata, até alcançar a quantidade de energia fixa estabelecida pela procura. Neste caso, o preço de encontro entre a oferta e a procura é definido pela última oferta a ser aceite [10]. No caso simétrico, para além das ofertas de venda também existem as ofertas de procura. O operador de mercado ordena as ofertas de venda por preço ascendente, tal como no caso assimétrico, e as ofertas de compra por preço descendente. O preço de equilíbrio é então determinado pelo cruzamento das curvas de oferta e compra. Todas as ofertas de compra com preço superior ao preço de equilíbrio, tal como todas as ofertas de venda com preço inferior, são aceites [10]. A Tabela 2.1 contém alguns exemplos de bolsas de eletricidade e as suas respetivas características.

Bolsa	Participação	Ofertas dos compradores	Ofertas simples	Fixação dos preços	Remuneração da disponibilidade	Despacho integrado
NETA	Obrigatória	Não	Não	Ex-ante	Sim	Não
NordPool	Voluntária	Sim	Sim	Ex-ante	Não	Não
Espanha	Voluntária	Sim	Não	Ex-ante	Sim	Não
Austrália	Obrigatória	Sim	Sim	Ex-post	Não	Em parte
N.Zelândia	Voluntária	Sim	Sim	Ex-post	Não	Sim
Califórnia	Voluntária	Sim	Sim	Ex-ante	Não	Não

Tabela 2.1: Exemplos de bolsas de eletricidade, retirado de [8]

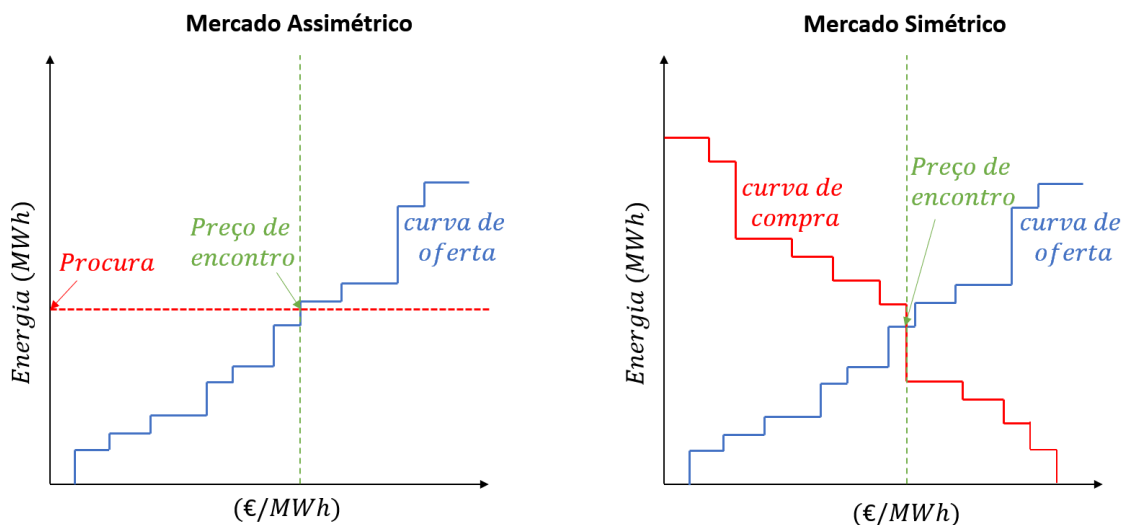


Figura 2.3: Pool assimétrica e simétrica, adaptado de [10]

No caso de existir viabilidade técnica do despacho, os agentes produtores são remunerados ao preço de encontro do mercado, gerando assim um excedente dos produtores, que consiste na diferença entre a receita obtida na venda do produto e o custo (variável) incorrido na respetiva produção – admitindo que o preço de venda corresponde ao custo marginal de produção. De uma forma análoga é gerado um excedente dos consumidores, sendo este a diferença entre o benefício resultante da utilização do produto e a quantia gasta na aquisição da energia – admitindo que o preço de compra corresponde ao benefício marginal do consumo [8]. A diferença destes excedentes resulta num excedente social, que por sua vez corresponde à área total sombreada na Figura 2.4. Note-se que o preço de equilíbrio formado no mercado corresponde à maximização deste excedente social.

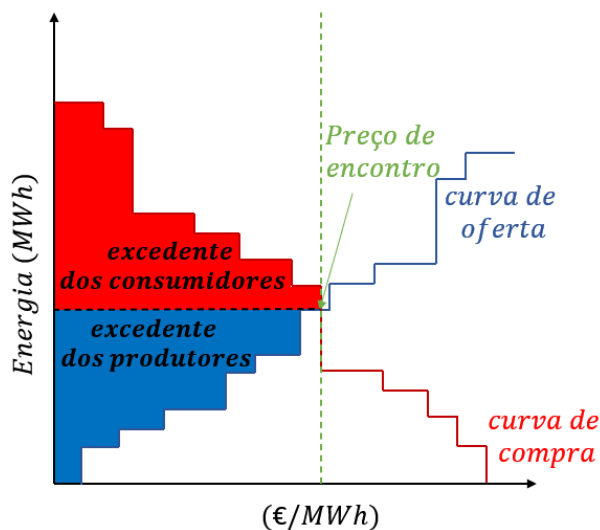


Figura 2.4: Preço de equilíbrio de mercado, adaptado de [11]

2.2.2 Acoplamento de mercados

É possível acoplar vários mercados de eletricidade e integrá-los num só mercado baseado numa bolsa. Os méritos e dificuldades apresentadas por esta decisão são imensos, no entanto, é essencial à ambição europeia de construir o MIE [5]. Nas condições ótimas de funcionamento, sem restrições ou congestionamentos nas interligações entre as áreas do mercado, o preço de equilíbrio é idêntico em todas as zonas. Isto dá-se pois não existe distinção entre os sistemas elétricos. No caso em que o fluxo de energia gerado pelas transações no mercado excede a capacidade da interligação entre as zonas no mesmo sentido, devido a diferenças entre preços das propostas de diferentes zonas, geram-se preços diferentes para as zonas envolvidas no congestionamento. A alocação da capacidade da interligação é feita através de leilões, preferencialmente implícitos nas propostas de mercado [12].

Em [12] faz-se a distinção entre os dois modelos principais para descrever a implementação de leilões implícitos de alocação de capacidade de interligação. O primeiro, market splitting, refere-se aos mercados com duas ou mais áreas, mas geridos por um único operador de mercado, enquanto o segundo, market coupling, refere-se aos mercados em que existe cooperação entre vários operadores de mercado referentes a diferentes áreas (Nominated Electricity Market Operator (NEMO) no contexto europeu). Ambos os modelos utilizam o mesmo esquema de cálculo da capacidade de interligação disponível:

$$ATC = NTC - AAC \quad (2.1)$$

O Net Transfer Capacity (NTC), representa a capacidade máxima que pode ser oferecida ao mercado, que é comunicada pelos ORTs, enquanto o Already Allocated Capacity (AAC) representa a capacidade já alocada correspondente a outros leilões ou contratos. A Available Transfer Capacity (ATC) resultante é utilizada para efeitos de atribuição de capacidade de interligação [12].

Outro modelo que descreve os leilões implícitos de alocação de capacidade de interligação é o Flow-Based Market Coupling (FBMC). Incorpora o uso de fatores de distribuição de transmissão de energia (PTDFs), que refletem as alterações nos fluxos físicos induzidos por mudanças na produção entre regiões [12].

A iniciativa denominada de Price Coupling of Regions (PCR) foi lançada no Fórum Regulatório de Florença em 2009 por três bolsas de eletricidade: NordPool, European Power Exchange (EPEX) e MIBEL [5]. Esta é baseada em três aspetos essenciais [13]:

- O algoritmo comum fornece uma determinação justa e transparente dos preços de eletricidade do dia seguinte e uma posição líquida de uma área de licitação em toda a Europa. O algoritmo é desenvolvido respeitando as características específicas dos vários mercados de energia na Europa e as restrições da rede elétrica. Ele otimiza o bem-estar geral e aumenta a transparência;

- O processo PCR baseia-se na partilha descentralizada de dados, resultando numa operação robusta e resiliente;
- O serviço PCR Matcher e Broker permite a troca de ordens anónimas e restrições da rede elétrica entre as bolsas de eletricidade para calcular os preços das zonas de licitação, outros preços de referência e posições líquidas de todas as áreas de licitação incluídas.

Em 2014, o acoplamento de mercado entre o NordPool, EPEX e MIBEL foi alcançado [5].

2.3 O caso de Portugal e Espanha

2.3.1 Reforma do setor elétrico português

O setor elétrico português iniciou um processo de abertura em 1988 com a promulgação do Decreto-Lei n.º 189/88, que permitiu a produção independente de eletricidade via fontes renováveis ou cogeração. Em 1994 foi criada a Rede Elétrica Nacional (REN), subsidiária da Energias de Portugal (EDP), que foi associada à atividade de gestão da rede de transporte, simbolizando um passo inicial na separação das atividades ligadas ao setor elétrico. Mais tarde, em 1995, foi publicado um novo pacote legislativo, contendo os Decretos-Lei n.ºs 182/95 a 188/95, que estabeleceu um novo modelo organizacional para o Sistema Elétrico Nacional (SEN), e que em conjunto com a Diretiva Europeia 96/92/CE, impulsionou a liberalização do setor. Este modelo é constituído pelo Sistema Elétrico de Serviço Público (SEP), que corresponde ao modelo de comprador único abordado na Secção 2.1.1, e o Sistema Elétrico Independente (SEI) [8]. No âmbito do SEP, a produção, transporte e distribuição de energia são realizadas dentro de um quadro de prestação de serviço público, e estão associadas à obrigatoriedade de fornecer energia elétrica com padrões adequados de qualidade e uniformidade tarifária [14]. A produção afeta a este sistema está dependente de um planeamento centralizado, pelo que novos centros eletroprodutores adquiriam licenças através de concursos públicos.

O SEI compreende duas componentes, sendo uma delas o Sistema Elétrico Não Vinculado (SENV), onde existe a possibilidade de integrar produtores e consumidores elegíveis através de contratos bilaterais físicos e livremente negociáveis, afetos ao pagamento de tarifas reguladas que lhes dão a possibilidade de acesso à rede pública. Também incluído no SEI existe a Produção em Regime Especial (PRE), que engloba os produtores independentes que utilizam fontes renováveis, até 10 MW de potência instalada se for mini-hídrica, ou a cogeração. A regulação independente do setor elétrico deu início em 1997 com a publicação dos estatutos da Entidade Reguladora do Setor Energético, agora conhecida como Entidade Reguladora dos Serviços Energéticos (ERSE), entidade já prevista na legislação do setor de 1995 e na Diretiva Europeia 92/96/CE (1996) [8].

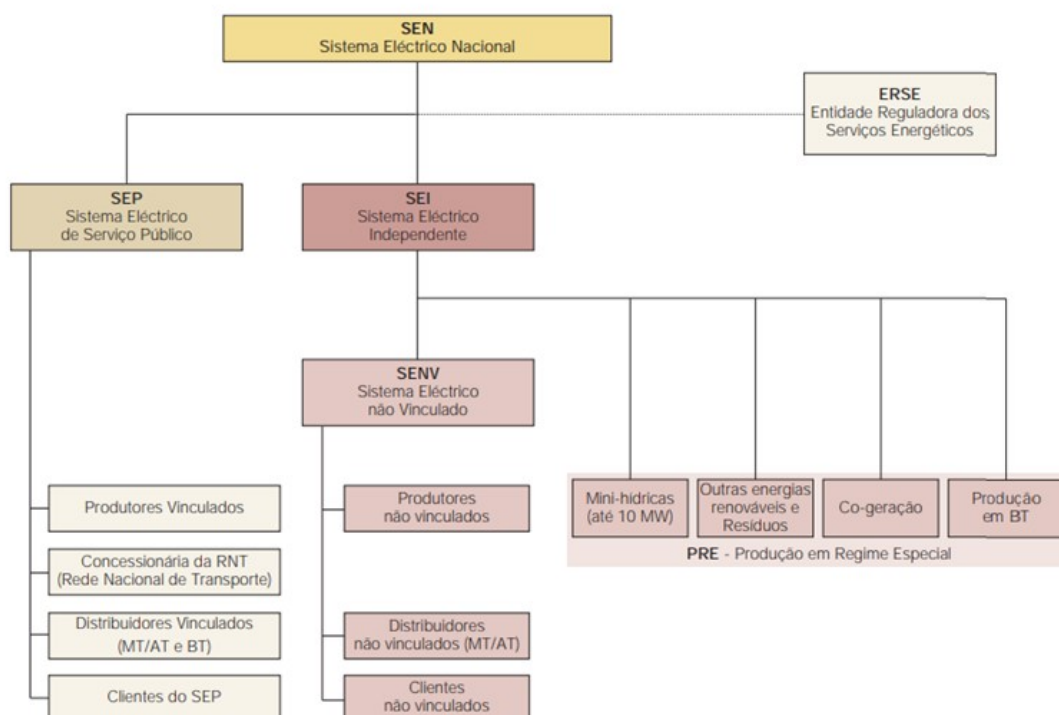


Figura 2.5: Estrutura do Sistema Eléctrico Nacional após a reforma de 1995, retirado de [14]

Em 2006, com a publicação do Decreto-Lei no 29/2006, o SEN foi novamente reestruturado no âmbito da abertura total do mercado retalhista, de modo a compatibilizar o setor elétrico com o princípios e políticas presentes na Diretiva Europeia 2003/54/CE [8]. Foram então separadas as atividades de distribuição e comercialização. A primeira ficou a cargo da EDP-Distribuição, enquanto que a segunda passou a ser desempenhada pelos agentes que anteriormente operavam no SENV e pelo Comercializador de Último Recurso (CUR), representado pela EDP Serviço Universal, e que fornecia os clientes mediante tarifas fixadas pela ERSE.

2.3.2 Reforma do setor elétrico espanhol

O Real Decreto 1538/1987 foi publicado em 11 de dezembro de 1987 e tinha como objetivo consolidar um novo modelo regulatório para o setor elétrico. Este modelo, conhecido como Marco Legal y Estable (MLE), foi fundamental para a recuperação económica do setor elétrico, implementando um planeamento centralizado de investimentos de longo prazo e uma tarifa nacional uniforme [9].

Em 1994, a Ley Orgánica del Sector Eléctrico Nacional (LOSEN) foi publicada para adaptar a legislação às diretivas europeias e liberalizar o setor elétrico. A LOSEN introduziu a competitividade no setor e criou a Comisión Nacional del Sistema Eléctrico (CNSE), "a entidade reguladora dos sistemas energéticos, cujos objetivos incluem garantir

a objetividade e transparência no funcionamento do setor"[9].

A Ley 54/1997 foi aprovada em 1997 e estabeleceu uma nova estrutura para o setor elétrico, criando duas entidades: o Operador de Sistema, "responsável pela continuidade e segurança do abastecimento de energia elétrica", e o Operador de Mercado, "responsável pela gestão das ofertas de compra e venda de energia elétrica". A [CNSE](#) continuou exercendo a regulação e supervisão do setor. Em 2000, um decreto do Ministério da Economia estabeleceu que as atividades de transporte e distribuição de energia elétrica deveriam ser desempenhadas por empresas cujo objetivo exclusivo seria o transporte e a distribuição de energia elétrica. A [Red Eléctrica de España \(REE\)](#) é responsável pelo transporte de energia e também atua como Operador de Sistema e [ORT](#). A Ley 17/2007, publicada em 2007, estabeleceu a [REE](#) como o "único Operador de Sistema e concessionária da rede de transporte, em exclusividade"[9].

2.3.3 Mercado Ibérico de Eletricidade

2.3.3.1 Conceção e Estrutura

A Diretiva 96/92/CE delineou normas comuns para o desenvolvimento do [MIE](#). A dificuldade inerente à proposta de um único mercado fez com que as conversações se orientassem para a alternativa de um sistema que interligue os mercados individuais de cada Estado-membro. Nesse sentido, os Estados-membros procuraram desenvolver mercados regionais transnacionais, como por exemplo, o NordPool ou o [MIBEL](#) [15].

Em novembro de 2001 deu-se a celebração do Protocolo de colaboração entre as Administrações espanhola e portuguesa no âmbito de criação do MIBEL. Este foi resultado das conversações e estudos que começaram em 1998 e que perspetivavam um mercado comum ao nível ibérico possibilitado por um esforço de convergência física, económica e legal por parte dos dois países. Os méritos associados ao estabelecimento de um mercado comum como este são os benefícios para os consumidores de ambos os países e um quadro de acesso livre em condições de objetividade, transparência e igualdade, qualidades atrativas e alinhadas com as ambições europeias [16].

Em 2004, foi assinado o Acordo de Santiago de Compostela, que desenhou um quadro estável que permitiu aos operadores de ambos os sistemas elétricos nacionais desenvolver as suas respetivas atividades em toda a Península. Deste também resultou a criação do Conselho de Reguladores, cujas responsabilidades, de entre várias, dizem respeito ao acompanhamento e desenvolvimento do [MIBEL](#) [16]. O Conselho é constituído pelas entidades [Comisión Nacional de Energía \(CNE\)](#), [ERSE](#), [Comisión Nacional del Mercado de Valores \(CNMV\)](#) e [Comissão do Mercado de Valores Mobiliários \(CMVM\)](#) [16].

Só em 2007, após verificada uma harmonização das várias condições, o mercado entrou em pleno funcionamento, tornando-se no segundo mercado regional europeu transnacional a ser constituído, sucedendo o NordPool [8].

O [MIBEL](#) está centrado numa bolsa de eletricidade, do tipo pool simétrica, gerida pelo

OMI, o qual está dividido em dois pólos, português e espanhol. O pólo espanhol OMIE gere os mercados diário e intradiário, enquanto o pólo português OMIP está associado ao mercado a prazo. Também existe a possibilidade de estabelecer contratos bilaterais [8]. Na Figura 2.6 está representada a estrutura organizacional do operador de mercado.

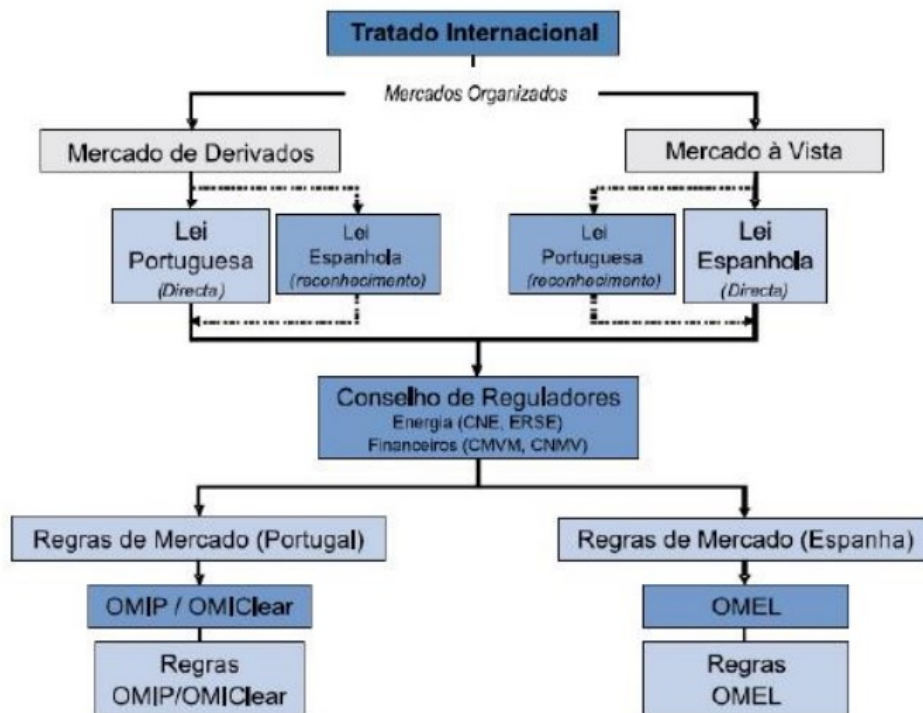


Figura 2.6: Esquema organizacional do Operador de Mercado Ibérico, retirado de [9]

2.3.3.2 Mercado Diário

O mercado diário tem como objetivo realizar transações de energia elétrica para o dia seguinte com base em propostas de compra e venda provenientes dos agentes de mercado. Estas propostas são do tipo simples, par quantidade-preço, pelo que o mercado segue o mesmo algoritmo de maximização do excedente social, observado no Capítulo 2.2.1. Ou seja, todos os produtores são remunerados ao preço marginal que corresponde ao preço de encontro. Este mercado encontra-se acoplado à europa desde 2014, através do projeto PCR, no âmbito da realização do MIE, conhecido como *Single Day-ahead Coupling (SDAC)* [17].

A sessão diária do mercado ocorre todos os dias do ano às 12:00 CET (Hora da Europa Central), durante a qual são estabelecidos os preços e volumes de energia elétrica para toda a Europa para o dia seguinte. Todas as propostas de compra e venda são comunicadas ao operador de mercado OMIE, o NEMO designado para o MIBEL, e são expostas ao algoritmo comum a todos os membros do PCR, *EU Pan-European Hybrid Electricity Market Integration Algorithm (EUPHEMIA)*. As ofertas de compra e venda são aceites

segundo uma ordem de mérito económico e em função da capacidade de interligação disponível entre as zonas de preço. Como foi visto no Capítulo 2.2.2, se em qualquer hora a capacidade de interligação entre as duas zonas for suficiente para permitir o fluxo de energia sem congestionamento, o preço para essa hora vai ser igual para ambas as zonas. Caso aconteça o contrário, ou seja, ocorra um congestionamento na interligação entre as zonas, o algoritmo resolve dois preços distintos para as zonas em questão [17].

Os resultados do mercado diário são depois analisados quanto à sua viabilidade no ponto de vista físico. Os respetivos **ORTs**, a **REN** e a **REE**, efetuam uma gestão das restrições técnicas do sistema, resultando num programa diário viável [17].

2.3.3.3 Mercado Intradário

O mercado intradiário é uma componente importante na dinâmica dos mercados de eletricidade, pois este permite ajustar os programas resultantes do mercado diário através de ofertas de compra e venda consoante as necessidades que surgem em tempo real. Os mercados intradiários encontram-se estruturados em seis sessões de leilões no âmbito do **MIBEL** e num mercado contínuo transfronteiriço europeu. À semelhança do mercado diário, depois de concretizadas as respetivas sessões de mercado, os resultados são comunicados aos **ORTs** que programam os seus processos de balanço. Os mercados diários seguem o modelo marginalista e o de acoplamento de mercados para as várias fronteiras que gerem [17].

O mercado intradiário contínuo, conhecido como acoplamento único intradiário, **Single Intraday Coupling (SIDC)** no plano europeu, apresenta a possibilidade de gerir os desbalanços de energia com duas diferenças essenciais quando comparado ao mercado intradiário de leilões. Em primeiro lugar, é permitido o ajuste até uma hora antes da entrega física da energia. Em segundo, os agentes do mercado podem beneficiar da liquidez disponível nos outros mercados da Europa, sempre que exista capacidade nas interligações entre as zonas [17].

2.3.3.4 Caracterização das interligações

Para efeitos de participação no mercado são consideradas nove interligações operacionais entre Portugal e Espanha, representadas na Figura 2.7. O trânsito de potência máximo nestas linhas é regido pelos limites térmicos das mesmas, valores determinados pelos **ORTs** com base em fatores como a temperatura ambiente. É então possível verificar variações destes valores máximos ao longo do ano, tendo em especial atenção o mínimo verificado nos meses de verão. Tal como foi referido na secção 2.2.2, os **ORTs** comunicam ao operador de mercado diário e intradiário, **OMIE**, os valores de capacidade de transporte nas interligações, sendo estes calculados segundo uma metodologia submetida ao regulamento europeu de **Capacity Allocation and Congestion Management (CACM)** [18]. No âmbito do mercado diário, o cálculo é efetuado a D-2, ou seja, dois dias antes dos trânsitos de energia e na véspera da sessão de mercado diário. A metodologia consiste na

determinação da mais alta capacidade segura, baseada na variação dos valores de carga e consumo numa simulação da rede, sendo verificados os cenários de contingência que se enquadrem no CACM. É utilizada uma resolução de 50 MW e é aplicado o método de Newton-Raphson no cálculo do trânsito de potência [19].

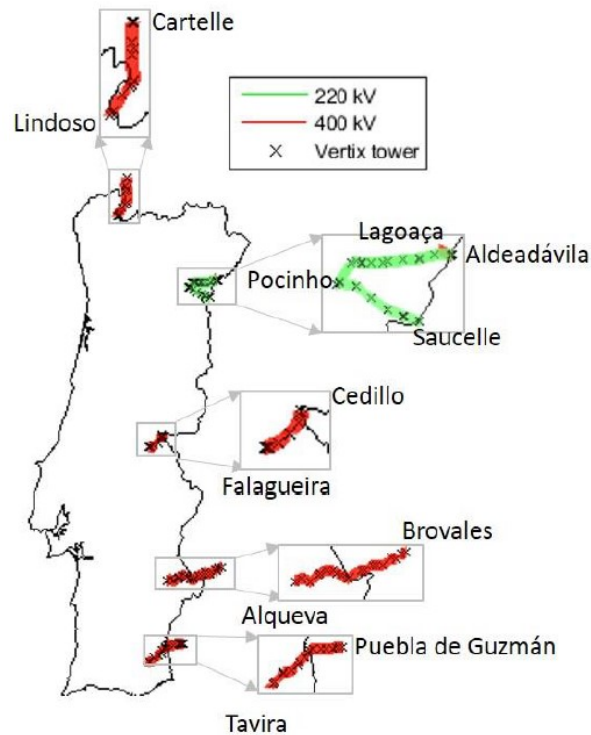


Figura 2.7: Localização das interligações entre Portugal e Espanha, retirado de [20]

Tendo em conta as restrições associadas às redes de transporte portuguesa e espanhola, foram elaboradas previsões para a capacidade de interligação para o ano de 2021, como mostra a Figura 2.8. Note-se que estes valores não têm em conta as eventuais restrições impostas pelo parque eletroprodutor. A volatilidade e intermitência da produção de energia eólica, é um exemplo de fator condicionante destas capacidades [18].

2.4. DESENVOLVIMENTO DE FONTES DE ENERGIA RENOVÁVEIS NA PENINSULA IBÉRICA

Sentido das trocas	Portugal → Espanha				Espanha → Portugal			
	Ponta e Cheia		Vazio		Ponta e Cheia		Vazio	
Plano de indisponibilidades programadas de elementos da rede	Inativo	Ativo	Inativo	Ativo	Inativo	Ativo	Inativo	Ativo
Janeiro		-		-		-		-
Fevereiro	2970	-	2790	-	2070	-	2160	-
Março		-		-		1890		1890
Abril	2070	1710	2430	2340	2700	-	2880	2790
Maio		1080		1080		1530		1260
Junho	2430	1440	3200	1710	3060	1260	3420	2520
Julho		2200		2800		-		3330
Agosto	2430	1890	3200	2800	3060	2610	3420	2700
Setembro	3150	2700	3780	2600	2790	-	3420	2700
Outubro		2800		3000		-		3400
Novembro	2700	-	2250	-	2070	-	2700	-
Dezembro		-		-		-		-

Figura 2.8: Previsão da capacidade de interligação para fins comerciais em 2021, retirado de [18]

2.4 Desenvolvimento de fontes de energia renováveis na Península Ibérica

A crescente consciencialização da crise das alterações climáticas levou a uma intensificação das medidas tomadas para combater e adaptar-se aos seus impactos. A UE, enquanto ator global importante, tem estado na vanguarda destes esforços, demonstrando um forte empenho na luta contra as alterações climáticas e na promoção da sustentabilidade através da criação de políticas ecológicas e do estabelecimento de objetivos cada vez mais ambiciosos. Em 2014, foi introduzido um novo quadro (framework) relativo ao clima e à energia para o período de 2020 a 2030, que procura alargar o compromisso da UE em matéria de resposta à crise climática para além do quadro ativo de 2020. Este quadro definiu 4 metas importantes, que já sofreram algumas revisões e reformulações, para serem atingidas no prazo considerado, sendo elas, no momento em que escrevemos: uma quota de 32% de energias renováveis na produção total; uma melhoria de 32,5% na eficiência energética; uma redução de 55% das emissões de gases com efeito de estufa; e uma capacidade de interligação capaz de transportar 15% da eletricidade produzida no país [21].

Como parte do pacote Clean Energy for all Europeans, cada Estado-Membro da UE foi obrigado a desenvolver um Plano Nacional de Energia e Clima (PNEC) para o período de 2021 a 2030 [22]. Tanto Portugal como Espanha, em consonância com a estratégia de longo prazo da UE para 2050, têm a ambição de alcançar a neutralidade carbónica até

2050. Assim, os objetivos definidos nos respetivos PNECs correspondem, na sua maioria, às ambições da UE ou, em alguns casos, ultrapassam-nas. No caso de Portugal, o país tem a ambição de atingir uma quota de 49% de energias renováveis no consumo final de energia até 2030, enquanto a Espanha segue um objetivo de 48%, segundo as versões revistas de ambos os PNECs.

Em Portugal, a produção anual de eletricidade a partir de fontes renováveis aumentou ligeiramente em relação aos valores de 2014. A Direção-Geral de Energia e Geologia (DGEG) disponibiliza uma perspetiva sobre o estado do desenvolvimento renovável em Portugal através dos seus relatórios mensais publicados [23]. Utilizando a metodologia estabelecida pela Diretiva 2009/28/CE, foi possível verificar que em 2022 a produção anual de RES-E foi de aproximadamente 61% da produção total de eletricidade, o que é um pouco superior aos 52% de 2014. A dependência de Portugal da produção de energia hidroelétrica, que sempre representou uma grande parte da produção anual total devido à sua grande capacidade instalada, parece ter diminuído ligeiramente, em parte graças ao desenvolvimento de outras fontes renováveis como a energia eólica e solar. No final de janeiro de 2023, a capacidade instalada das tecnologias hidroelétrica e eólica representava 80% (47% e 33%, respetivamente) da capacidade total instalada de fontes de energia renovável, enquanto a energia solar, em comparação, ainda está subdesenvolvida, mas a crescer significativamente. O mesmo não se pode dizer de Espanha, onde a energia solar tem uma forte presença na sua produção, consistindo em cerca de 10% da produção total de eletricidade do país em 2022. No mesmo ano, a Espanha produziu 42% da sua eletricidade a partir de fontes renováveis [24]. É de notar que, ao contrário de Portugal, a Espanha desenvolveu a energia nuclear, que é também uma tecnologia de baixo custo marginal com a vantagem de ser despachável.

Esta região tem capacidade para atingir uma elevada penetração de energia eólica, o que faz com que a energia eólica seja a tecnologia com maior capacidade instalada em Espanha (30 MW), ficando em segundo lugar no caso de Portugal (5,6 MW), atrás da energia hídrica (8,1 MW). A Península Ibérica tem também um dos níveis de irradiação solar mais elevados da Europa, o que faz da energia solar fotovoltaica a tecnologia com maior potencial de crescimento na região. Os PNEC de ambos os países estão conscientes deste facto e prevêem um crescimento para cerca de 20 GW e 76 GW de capacidade instalada de energia solar até 2030 em Portugal e Espanha, respetivamente [25, 26].

2.5 Separação de mercado e convergência de preços nos mercados europeus de eletricidade

O número de estudos sobre a divergência de preços, ou convergência de preços, é reduzido no contexto da separação de mercados. No entanto, foram recolhidas algumas referências sobre este tema. Os determinantes da separação de mercados no mercado ibérico de eletricidade foram já estudados por [5] considerando um período de tempo

2.5. SEPARAÇÃO DE MERCADO E CONVERGÊNCIA DE PREÇOS NOS MERCADOS EUROPEUS DE ELETRICIDADE

entre 2008 e 2012. Este estudo utiliza dados horários de produção de eletricidade, consumo e capacidade de transferência disponível ATC das interligações para ambas as áreas de licitação, ou zonas de mercado, do MIBEL (Portugal e Espanha) como variáveis exógenas para descrever a ocorrência de separação de mercado, modelada como uma variável endógena binária, através da estimação de modelos logit e não paramétricos. O estudo encontrou evidências que confirmam o aumento da probabilidade de ocorrência de separação de mercado com eletricidade de baixo custo marginal mais elevada numa das áreas de licitação, especialmente no caso em que existe uma assimetria de dimensão dos sistemas eléctricos das áreas de licitação. Da mesma forma, um estudo realizado pelos mesmos autores [27] também estima a probabilidade de separação do mercado entre as zonas de mercado dinamarquesas do Nord Pool, considerando a elevada penetração da energia eólica na região. Outro estudo [28] analisa o mercado sueco de eletricidade, que consiste em quatro zonas de mercado separadas que também fazem parte do mercado integrado de energia nórdico, o Nord Pool. Neste estudo, é utilizado um índice denominado "MSplit" para avaliar o grau de convergência dos preços entre as zonas de licitação adjacentes. O índice mede a proporção de horas num mês rolante durante as quais se verifica a separação do mercado.

Outros estudos, embora não mencionem a separação de mercado, analisam a diferença de preços entre as áreas de licitação, por vezes designadas zonas, de esquemas de separação de mercado. Por exemplo, [29] analisa as diferenças de preços zonais no mercado grossista de eletricidade italiano. A utilização de um modelo reg-ARFIMA confirma a influência significativa que os congestionamentos das interligações e as tecnologias de produção têm na dinâmica dos preços. O grau de integração dos mercados nacionais norueguês e sueco na Nord Pool é abordado por [30]. O estudo conclui que, no período considerado, os mercados grossistas estavam efetivamente integrados, enquanto as diferenças nos preços de retalho e nas margens comerciais, bem como a legislação nacional relativa ao mercado, condicionavam a concorrência a nível retalhista. Os efeitos a longo prazo da implementação da separação de mercado, referindo à divisão em múltiplas zonas de preços de uma única área de licitação ou das fronteiras nacionais, são também um tema de interesse na procura de configurações zonais ótimas. Esta divisão pode ser observada nas zonas de mercado da, por exemplo, Suécia, Dinamarca, Noruega e Itália, e permite uma melhor gestão dos congestionamentos, uma vez que tem em conta as limitações internas do sistema de transporte [31].

O impacto a longo prazo de uma potencial separação de mercado na Alemanha é estudado em [32], que conclui que os preços da eletricidade para o dia seguinte, a gestão dos congestionamentos e o bem-estar (welfare) social são fortemente influenciados por essa divisão. Concluem também que uma configuração ótima da zona de ofertas pode ser incompatível com a evolução do mercado, como a expansão das RES-E, a extensão da rede e o investimento planeado na produção, tornando-a desatualizada e em necessidade de uma reavaliação regular, como recomendado pelos autores. Outro estudo [33] também analisa os efeitos da separação de mercado numa perspetiva de longo prazo e conclui que

a introdução de mais zonas de preços pode diminuir o bem-estar devido a um possível sobre-investimento na capacidade de produção que é limitada por congestionamentos intra-zonais, que são invisíveis para o mercado. O estudo apoia, mais uma vez, a necessidade de uma análise pormenorizada que tenha em conta o investimento potencial e a evolução a longo prazo antes de se decidir sobre uma separação do mercado.

A alegação de escassez não se aplica à literatura sobre a convergência de preços em mercados acoplados, que é um tema vasto e muito procurado. Uma vez que não tem qualquer diferença prática em relação ao contexto da separação de mercados, os conhecimentos e as metodologias são aplicáveis no domínio da separação de mercado. A região da Europa Central Ocidental [Central West Europe \(CWE\)](#) tem sido objeto de muitos estudos sobre convergência de preços e integração global do mercado. No âmbito da implementação do Flow-Based Market Coupling na região [CWE](#), [34] emprega um modelo de Random Forest para extrair conhecimento sobre o congestionamento das interligações, especificamente o impacto de variáveis selecionadas, tais como trocas comerciais programadas e previsão de carga nos mercados acoplados. Uma das variáveis identificadas permitiu reconhecer um determinado cenário de congestionamento da interligação causado pela elevada produção de eletricidade a partir de fontes renováveis na Alemanha, quando comparada com a França, provocando um fluxo excessivo de eletricidade incompatível com as capacidades de transporte observadas, causando a divergência de preços no mercado diário. Um estudo mais recente [35] apoia e reforça estas conclusões anteriores, incorporando explicitamente dados de previsão da produção de [RES-E](#).

2.6 Explainable Machine Learning

Dada a presença de modelos explicativos de [Machine Learning \(ML\)](#) na literatura à volta do tópico de separação de mercado e convergência de preço, decidiu-se levar este estudo no mesmo sentido. Desse modo, é necessário aprofundar a compreensão dos algoritmos a utilizar e de como estes permitem a extração de conhecimento de um conjunto de dados.

A evolução dos métodos de [ML](#), e de [Inteligência Artificial \(IA\)](#) de um modo geral, contribuíram de forma significativa para a área de estatísticas e ciência de dados. Face à crescente capacidade de extração e criação de dados nas várias indústrias modernas, e a inerente complexidade nos padrões e estruturas neles verificada, tornou-se mais frequente a opção por metodologias de [ML](#) invés de modelos estatísticos tradicionais em aplicações de, para citar algumas, previsão e classificação. Esta troca, especialmente motivada pelo desempenho superior destes métodos, dada a elevada capacidade de modelar relações não-lineares complexas entre variáveis, apresentou no entanto algumas fragilidades. A fraca capacidade de interpretação de alguns modelos de [ML](#) e da explicação das suas respetivas previsões, devido à complexidade crescente dos modelos, criou uma barreira para a integração correta em áreas de elevado risco como a medicina e assistência médica [36].

Deste modo, a área de **XML**, ou **Explainable Artificial Intelligence (XAI)**, viu um aumento significativo de interesse e pesquisa em anos recentes. O seu principal objetivo é o de criar explicações compreensíveis pelo humano do porquê de um modelo de **ML** fazer uma dada previsão, ou de como este chega à mesma, sem necessariamente adicionar interpretabilidade ao modelo em si [37]. Em [37] também é feita a distinção entre interpretabilidade e capacidade explicativa no contexto dos modelos de caixa-negra, com o primeiro sendo a compreensão direta de como o modelo funciona, e por sua vez como gera uma dada previsão, enquanto o segundo refere-se a uma compreensão via representação ou analogia do modelo, muitas vezes alcançado através do auxílio de modelos interpretáveis que replicam o seu comportamento. A comparação de diferentes métodos permite compreender a importância desta distinção. Um modelo de Regressão Logística pode ser facilmente interpretado por inspeção dos coeficientes estimados, e por sua vez podem ser explicadas as respectivas previsões, enquanto um modelo de **Rede Neuronal Artificial (RNA)**, devido à maior complexidade do algoritmo, não permite a extração de conhecimento por inspeção direta, ou seja, o grau de interpretação oferecido por um dado modelo está limitado ao domínio do humanamente compreensível [38], o que leva ao recurso a técnicas capazes de explicar as suas previsões. É possível descrever interpretação e capacidade explicativa como termos similares, em que o grau de interpretação inerente a um dado modelo, sem recurso a técnicas de **XML**, equivale à sua capacidade explicativa [39, 40].

Um dos tópicos de debate do **XML** é a potencial proporcionalidade inversa (trade-off) entre o desempenho e capacidade explicativa nos modelos de **ML**. Vários autores introduzem variações deste conceito, mas que de um modo geral têm em comum o estabelecimento de uma relação de proporcionalidade inversa entre uma quantificação da complexidade do algoritmo, ou uma métrica de desempenho, e o potencial analítico dos seus resultados, seja este caracterizado como interpretabilidade [41], capacidade explicativa [39] ou até mesmo transparência do algoritmo [42]. A Figura 2.9 contém uma representação do posicionamento dos algoritmos selecionados para o estudo em questão, ao longo dos eixos de interpretabilidade inerente ao algoritmo e exatidão, de uma forma coerente com a relação previamente descrita. Embora não seja o foco deste estudo, a distinção clara da capacidade dos algoritmos escolhidos é essencial na ponderação dos resultados obtidos.

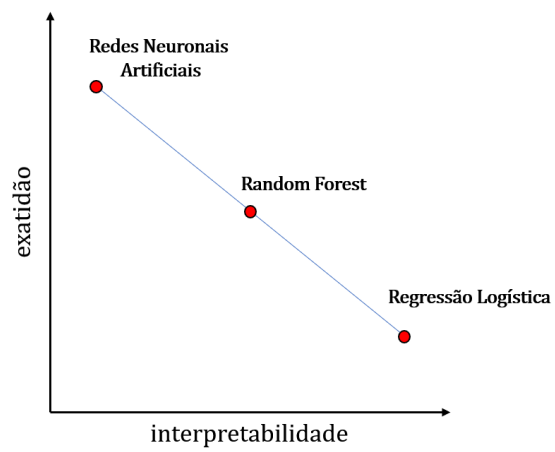


Figura 2.9: Troca interpretabilidade-exatidão hipotética para os modelos utilizados

TRATAMENTO DE DADOS E ANÁLISE ESTATÍSTICA DA SEPARAÇÃO DE MERCADO

Neste capítulo é feita uma descrição do procedimento de recolha e tratamento de dados, sendo estes utilizados numa análise estatística à separação de mercado e, eventualmente, no desenvolvimento de modelos explicativos. Optou-se pela construção de scripts executáveis na linguagem de programação Python para o processamento dos dados recolhidos e criação dos gráficos aqui apresentados.

3.1 Recolha e tratamento de dados

Alinhado com os objetivos definidos para este estudo, dados relevantes acerca do mercado de eletricidade ibérico, e das produções e consumos dos sistemas elétricos envolvidos, foram coletados e processados, de forma a permitir a construção de modelos explicativos da separação de mercado. Considerou-se um período entre **1 de janeiro de 2018** e **30 de abril de 2023**, visto que os dados disponíveis estavam limitados a este intervalo.

Começou-se pela recolha dos dados referentes aos preços de eletricidade, de ambas as zonas de mercado (portuguesa e espanhola), definidos em mercado diário para cada uma das 24 horas do dia seguinte, disponibilizados pela plataforma do [OMIE](#) [43]. Visto que o objetivo deste estudo é o de analisar os determinantes da separação de mercado, é necessário considerar uma variável que possa caracterizar a sua ocorrência. Como já foi referido no Capítulo 2.2.2, a separação de mercado leva a preços diferentes entre as zonas de mercado. Esta diferença de preços, conhecida como **spread**, entre a zona espanhola e portuguesa pode ser expressa da seguinte forma para uma dada hora h :

$$Spread(h) = Preço_{PT}(h) - Preço_{ES}(h) \quad (3.1)$$

Por inspeção dos valores de spread calculados, é possível construir uma variável binária representativa da ocorrência de separação de mercado: OSM . É possível encontrar na literatura acerca deste tópico, e outros semelhantes, a caracterização do congestionamento

de interligações ou ocorrência da separação de mercado, e da convergência de preços, como uma variável binária [5, 28, 35]. Esta nova variável pode ser calculada da seguinte forma:

$$OSM(h) = \begin{cases} 0 & , se Spread(h) = 0 \\ 1 & , se Spread(h) \neq 0 \end{cases} \quad (3.2)$$

Numa dada hora h , a variável OSM toma o valor de 0 caso a diferença de preços entre as zonas de mercado seja nula, ou seja, existe um único preço na totalidade do mercado. Caso contrário, regista-se a ocorrência de separação de mercados. Também se consideraram variações desta variável representativas dos dois cenários possíveis: estabelecimento de um fluxo de eletricidade nas interligações no sentido Espanha-Portugal, $OSM_{ES \rightarrow PT}$, ou Portugal-Espanha, $OSM_{PT \rightarrow ES}$. Isto é facilmente verificável via análise da variável $Spread$. Caso esta seja menor que 0, e considerando a definição da variável na equação 3.1, significa que o preço estabelecido na zona de mercado espanhola é superior ao da portuguesa, implicando que ocorreu separação de mercado devido ao estabelecimento de um fluxo de eletricidade no sentido Portugal-Espanha incompatível com a capacidade da interligação no mesmo sentido, e vice-versa. Estas variáveis podem ser descritas como:

$$OSM_{ES \rightarrow PT}(h) = \begin{cases} 0 & , se Spread(h) \leq 0 \\ 1 & , se Spread(h) > 0 \end{cases} \quad (3.3)$$

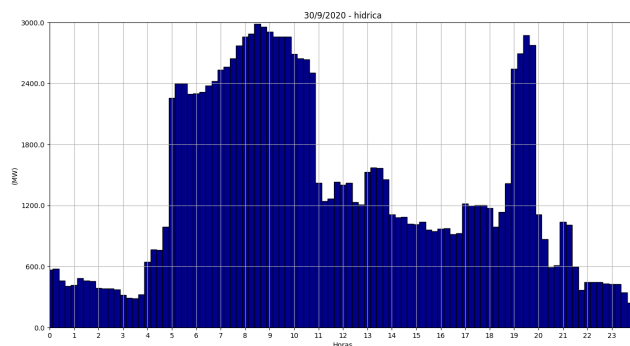
e

$$OSM_{PT \rightarrow ES}(h) = \begin{cases} 0 & , se Spread(h) \geq 0 \\ 1 & , se Spread(h) < 0 \end{cases} \quad (3.4)$$

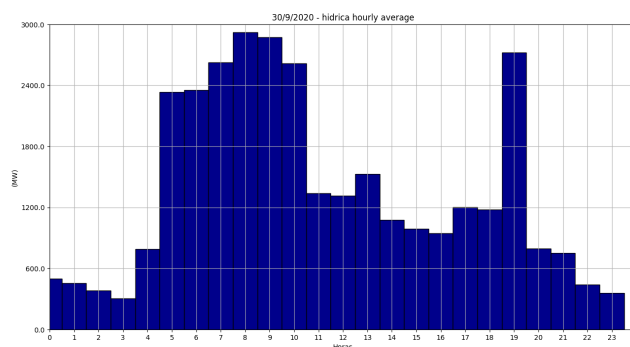
Da mesma plataforma também se recolheu os dados referentes às capacidades de transferência das interligações, $ATCs$, para ambos os sentidos e considerando as restrições impostas pelo mercado diário [44].

Os dados referentes às produções e consumos dos sistemas elétricos português e espanhol foram extraídos das plataformas dos respetivos $ORTs$. No caso português, recorreu-se ao Data Hub da **REN** [6], plataforma que agrega e disponibiliza dados acerca do setor energético, incluindo o mercado diário. Daqui foi possível recolher os dados referentes à produção nacional de eletricidade discriminada por tecnologia utilizada, tal como os valores de consumo e de bombagem hidroelétrica. As tecnologias de produção de eletricidade, seguindo as denominações atribuídas pela plataforma, são: **Gás Natural, Solar, Hídrica, Eólica, Outra Térmica, Biomassa, Carvão** e **Ondas**. Estes dados apresentam uma resolução de 15 minutos, algo que é incompatível com os dados referentes aos preços de eletricidade e $ATCs$ que vêm na resolução horária comum aos mercados de eletricidade. Foi então necessário fazer a média das produções e consumos por hora de modo a resolver esta incompatibilidade. A Figura 3.1 demonstra, para um dia exemplo, o resultado desta

alteração.



(a) Original



(b) Processado

Figura 3.1: Exemplo da transformação dos dados recolhidos

Os dados correspondentes ao sistema elétrico espanhol foram obtidos na plataforma da REE [45]. À semelhança da sua congénere portuguesa, esta fornece os valores de produção discriminados por tecnologia e consumo nacional, no entanto, utiliza uma resolução de 10 minutos para os dados. Este último aspeto foi corrigido, mais uma vez, recorrendo ao cálculo da média por hora destes valores de produção e consumo. Existem algumas diferenças no leque de tecnologias discriminadas nos dados, visto que a produção espanhola apresenta algumas alternativas não disponíveis em Portugal, como por exemplo a tecnologia nuclear. As tecnologias de produção de eletricidade discriminadas nos dados, mantendo as denominações da plataforma, são: **Ciclo Combinado de Gás Natural, Solar Fotovoltaica, Solar Térmica, Hídrica, Eólica, Cogeração e Resíduos, Carvão, Nuclear e Térmica Renovável**. Todos os dados referentes à produção e consumo estão representados em MW.

Uma das principais preocupações na agregação de todos os dados descritos até agora foi o da sua sincronização no tempo. Note-se que os dados fornecidos pela REN encontram-se no Tempo Médio de Greenwich (GMT ou UTC+0), enquanto os dados fornecidos pela REE e os referentes ao mercado diário encontram-se no fuso horário da europa central (UTC+1).

CAPÍTULO 3. TRATAMENTO DE DADOS E ANÁLISE ESTATÍSTICA DA SEPARAÇÃO DE MERCADO

Decidiu-se então considerar um período equivalente mas que começa uma hora antes para os dados portugueses, ou seja, que começa em 31 de dezembro de 2016 às 23 horas (UTC+0). Isto permitiu sincronizar com os restantes dados que começam à meia-noite de 1 de janeiro de 2017 (UTC+1). Todos os dados foram reunidos em ficheiros Excel e CSV de modo a facilitar a sua análise e manipulação com recurso ao módulo Python: **pandas** [46].

A Tabela 3.1 faz um resumo de todas as variáveis consideradas tal como algumas estatísticas associadas às mesmas como: **média, desvio padrão, mínimo, máximo e percentis 25, 50 (mediana) e 75**. Note-se que se considera um período superior a 5 anos, o que implica algumas mudanças nas variáveis ao longo do tempo, em especial devido a alterações nas capacidades instaladas das tecnologias de produção. Um exemplo concreto verifica-se no término da utilização de carvão na produção de eletricidade em Portugal durante o período considerado, o que leva à exclusão desta variável de grande parte deste estudo.

Variável	Média	Desvio padrão	Mínimo	25%	50%	75%	Máximo	Unidade
<i>Preço_{ES}</i>	79,06	62,8	0	42,8	55,1	90,43	700	€/MWh
<i>Preço_{PT}</i>	79,28	62,7	0	42,96	55,19	90,92	651	€/MWh
<i>ATC_{ES-PT}</i>	2911,42	966,5	0	2115	2835	3645	5559	MW
<i>ATC_{PT-ES}</i>	3073,35	597,24	100	2700	3100	3510	4635	MW
<i>Consumo_{ES}</i>	28047,43	4515,05	16539,51	24372,17	28005,34	31465,5	41828,82	MW
<i>Consumo_{PT}</i>	5849,04	975,1	3250,02	5061,91	5864,46	6529,26	9827,8	MW
<i>Eólica_{PT}</i>	1447,7	1072,56	1,73	572,51	1117,05	2133,96	4759,53	MW
<i>Solar_{PT}</i>	168,27	266,89	0	0	4,07	278,3	1727,09	MW
<i>Hídrica_{PT}</i>	1199,92	1294,12	0	63,61	781,58	1920,75	6527,94	MW
<i>Biomassa_{PT}</i>	350,82	42,1	137,75	323,19	348,4	384,71	457,93	MW
<i>Gás_{PT}</i>	434,03	136,55	101,91	331,91	414,78	567,83	709,88	MW
<i>Gás_{CCPT}</i>	1352,63	903,56	0	584,1	1357,7	2089,2	3695,1	MW
<i>Carvão_{PT}</i>	593,49	709,61	0	0	100,4	1399,2	1762,7	MW
<i>Outra_Térmica_{PT}</i>	27,15	7,61	1,2	23,14	27,92	31,57	47,1	MW
<i>Eólica_{ES}</i>	6179,01	3619,56	143	3335,92	5477,17	8394,99	20686	MW
<i>Solar_{PV_{ES}}</i>	1766,01	2705	0	15,99	105,84	2703,74	14760,16	MW
<i>Solar_Térmica_{ES}</i>	532,1	657,25	0	12,68	209,17	821,5	2207,83	MW
<i>Hídrica_{ES}</i>	2880,55	2140,2	0	1264,84	2524,16	4151,67	11656,33	MW
<i>Gás_{CC_{ES}}</i>	4719,78	3151,51	313,51	2221,17	3859,5	6561,34	17656,16	MW
<i>Nuclear_{ES}</i>	6317,67	847,13	3616,67	5944	6811,83	7044,84	7137,33	MW
<i>Carvão_{ES}</i>	1963,66	2109,99	0	500,67	921,18	2988,01	8726,66	MW
<i>Térmica_Renovável_{ES}</i>	475,11	81,1	0	422,32	459,17	540,48	674,83	MW
<i>Cogeração_Resto_{ES}</i>	3226,14	675,8	264,83	3049,83	3461,98	3681,67	4261,5	MW

Tabela 3.1: Variáveis de estudo

Foram também consideradas transformações das variáveis de produção de eletricidade com base numa expansão do conceito de penetração de energia eólica às restantes tecnologias [47]. Esta consiste na representação dos valores produzidos por cada tecnologia como a fração do consumo nacional que satisfazem. Ao longo do estudo, estas variáveis vão ser referidas como contribuições, e são calculadas como nos seguintes exemplos para as produções eólica portuguesa e nuclear espanhola:

$$Eólica_{PT-contribuição}(t) = \frac{Eólica_{PT}(t)}{Consumo_{PT}(t)} \quad (3.5)$$

e

$$Nuclear_{ES-contribuição}(t) = \frac{Nuclear_{ES}(t)}{Consumo_{ES}(t)} \quad (3.6)$$

De particular interesse para este estudo são as fontes de energia renováveis que, de todas as tecnologias, são as que verificam um maior aumento nas capacidades instaladas ao longo do período considerado. As Figuras 3.2a e 3.2b mostram a série de tempo da geração eólica em Portugal e Espanha, respetivamente. Por inspeção das mesmas é possível identificar um ligeiro aumento na produção espanhola ao longo dos anos, enquanto a produção portuguesa não verifica qualquer alteração significativa. Através da aplicação de uma média móvel de 3 meses podemos observar um padrão sazonal, em que a produção eólica é máxima no início/fim do ano e mínima a meio do ano.

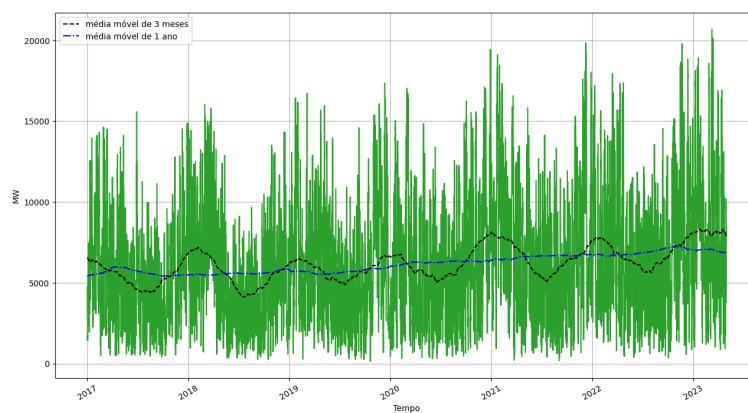
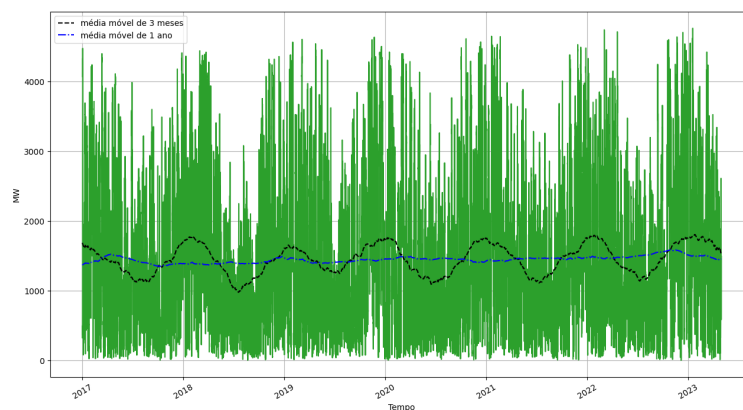
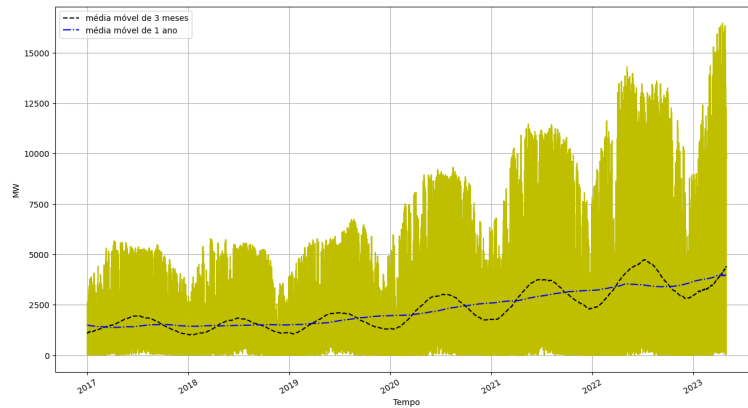
(a) $Eólica_{ES}$ (b) $Eólica_{PT}$

Figura 3.2: Séries de tempo para a produção eólica no período de 2017 a 2023

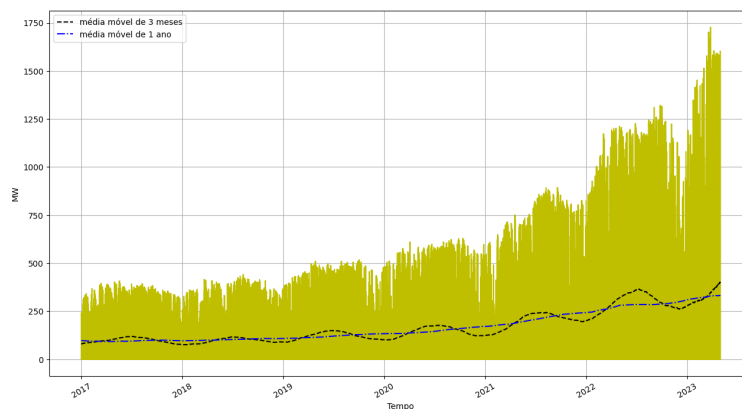
No que toca à produção de eletricidade via energia solar, verifica-se um aumento substancial tanto em Portugal como em Espanha ao longo destes anos. Espanha tem uma capacidade instalada considerável enquanto que em Portugal é vista como subdesenvolvida. As Figuras 3.3a e 3.3b têm as séries de tempo para a produção solar de ambos os países, que ao aplicar novamente uma média móvel de 3 meses, demonstram um padrão

CAPÍTULO 3. TRATAMENTO DE DADOS E ANÁLISE ESTATÍSTICA DA SEPARAÇÃO DE MERCADO

sazonal inverso ao verificado para a produção eólica. Note-se que para a produção solar espanhola considera-se a soma da solar fotovoltaica e térmica, representadas na Tabela 3.1.



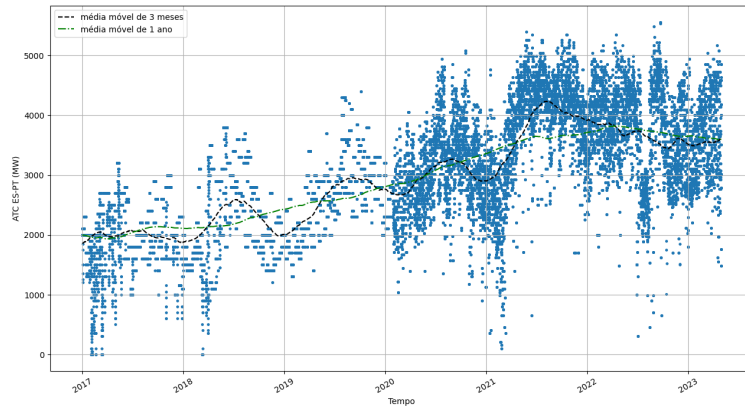
(a) *Solar_{ES}*



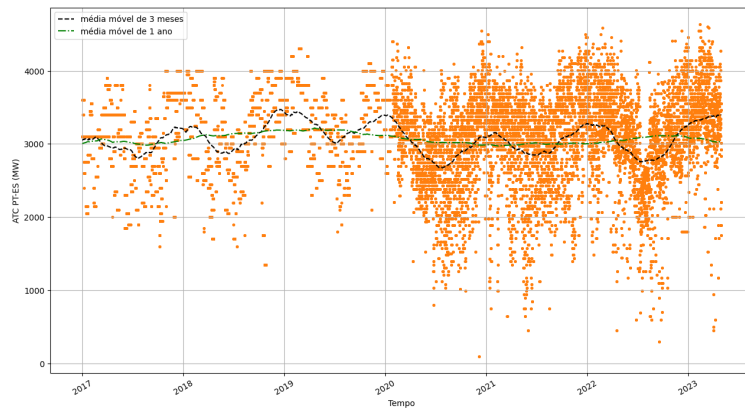
(b) *Solar_{PT}*

Figura 3.3: Séries de tempo para a produção solar no período de 2017 a 2023

As capacidades de interligação também são variáveis importantes para o estudo em questão. Nas Figuras 3.4a e 3.4b encontram-se as séries de tempo das capacidades de interligação para ambos os sentidos (Espanha-Portugal e Portugal-Espanha). Por análise destes gráficos conclui-se que a capacidade de interligação no sentido Espanha-Portugal tem aumentado (em média) ao longo dos anos, enquanto que a capacidade no sentido contrário praticamente não variou. Isto significa que para os dados referentes a 2017 e 2018 existiam maiores restrições na ATC de Espanha para Portugal. É também possível observar um comportamento interessante a partir de 2020, em que os dados parecem ter uma maior variação. Acredita-se que este fenómeno esteja associado ao cálculo da capacidade de interligação, que a 29 de janeiro de 2020 passou a ser feito pela entidade CORESO. Esta alteração não aparenta ter afetado o valor médio das capacidades de interligação.



(a) ATC_{ES-PT}



(b) ATC_{PT-ES}

Figura 3.4: Séries de tempo para as capacidades de interligação no período de 2017 a 2023

3.2 Evolução da separação de mercado

Com base nas variáveis de ocorrência de separação de mercado descritas no capítulo anterior, é feita uma análise da evolução da separação de mercado ao longo do período considerado. A Figura 3.5 mostra o número de horas, e a percentagem de tempo correspondente, em que ocorreu separação de mercado para cada ano do período considerado. É possível verificar uma tendência de diminuição da sua ocorrência, pelo que passa de 6,7% das horas em 2017 para apenas 2,9% em 2022. De se notar que de 2021 para 2022 há um ligeiro aumento, possivelmente indicando o início da inversão desta tendência ou até da sua estagnação.

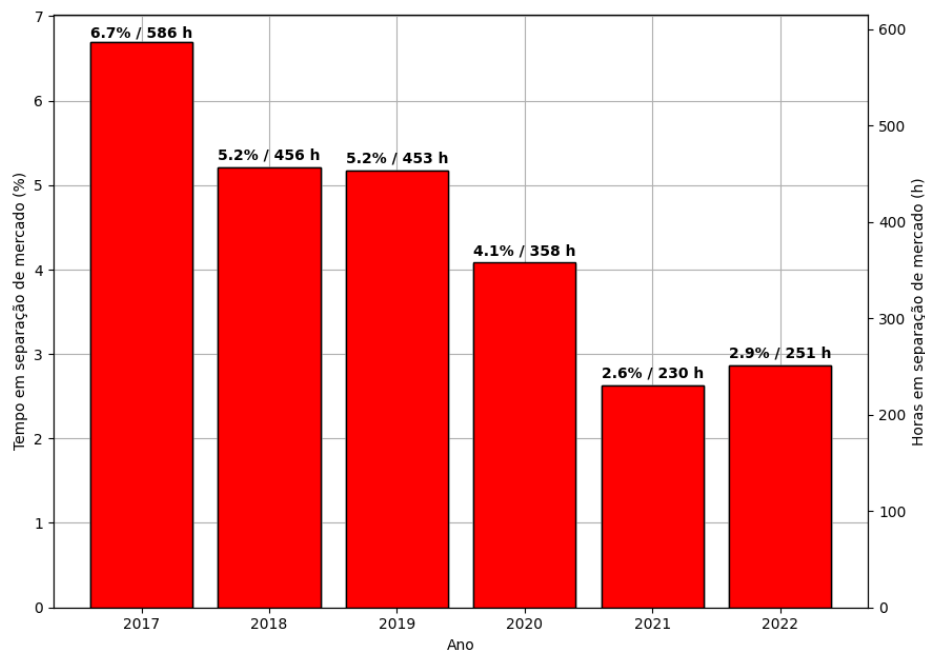


Figura 3.5: Evolução anual da separação de mercado

Um método de análise da separação de mercado, utilizado em [5, 27, 28], consiste na contagem de ocorrências num mês rolante, que pode ser expressa da seguinte forma:

$$f(h) = \sum_{i=0}^{720} OSM(h - i) \quad (3.7)$$

Por outras palavras, em cada hora h considerada, são contadas as ocorrências de separação de mercado nas 720 horas (cerca de 1 mês) que a antecedem mais a própria. Ao aplicar este método ao período considerado para o estudo resulta o gráfico da Figura 3.6.

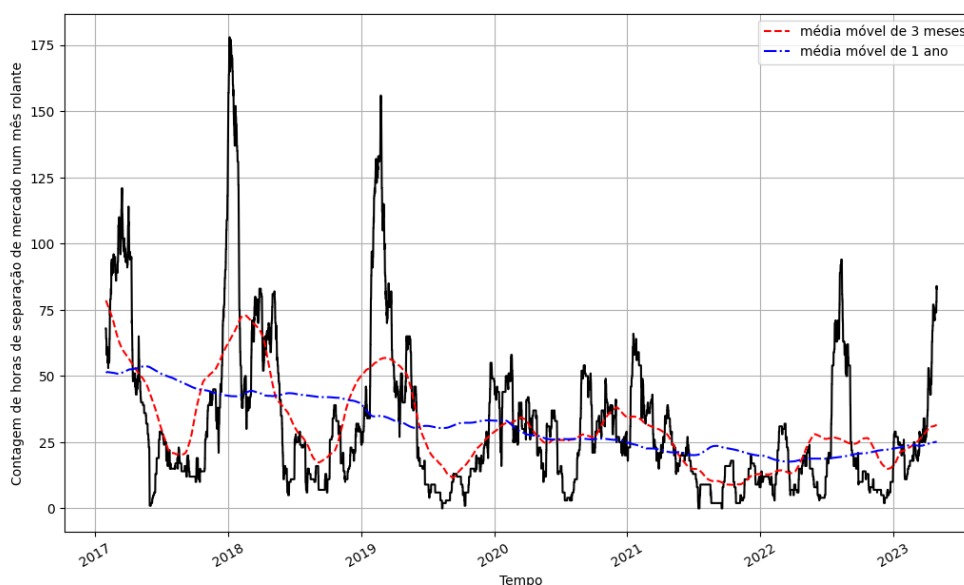


Figura 3.6: Contagem das horas em separação de mercado num mês rolante

A aplicação de médias móveis de 3 meses e de 1 ano sobre o gráfico resultante permitiram capturar um possível padrão de sazonalidade e a tendência decrescente da ocorrência de separação de mercado, respetivamente. De 2017 a 2020 é possível verificar aumentos de ocorrência no início/fim dos anos e a formação de um “vale” a meio dos mesmos. Isto poderá estar relacionado com a produção eólica, considerado um dos principais fatores no congestionamento das interligações [5, 27, 35] e que partilha o padrão descrito. A partir de 2021 não se observa este comportamento de separação de mercado, que pelo contrário demonstra dois picos relevantes a meio de 2022 e em abril de 2023.

Deve se ter em conta que, ao contrário de outros mercados onde a convergência de preços é o cenário menos comum (i.e. *CWE*), a ocorrência de separação de mercado no *MIBEL* é pouco frequente e, como se constata nos gráficos anteriores, tem diminuído ao longo dos anos. De acordo com a literatura, a análise da convergência de preços, ou de forma equivalente o estudo da ocorrência de separação de mercado, são considerados como bons índices de avaliação do grau de integração dos mercados observados [28, 34, 48]. Em [28] considera-se que menos de 180 horas (25% do período de 1 mês) na contagem de ocorrências num mês rolante indica um bom nível de integração. Por inspeção da Figura 3.6 percebe-se que, para o período considerado, excecionalmente em dezembro de 2018 são atingidas as 180 horas, o que é um sinal de ótima integração do mercado ibérico. A distribuição dos spreads calculados segundo a equação 3.1, representado na Figura 3.7, demonstram que a grande maioria das ocorrências de separação de mercado resultam em diferenças muito reduzidas nos preços de equilíbrio das zonas de mercado separadas, o que significa que a separação de mercado não tem um grande impacto nos preços português e espanhol para o período considerado, como se pode verificar nos

valores médios destes preços registados na Tabela 3.1.

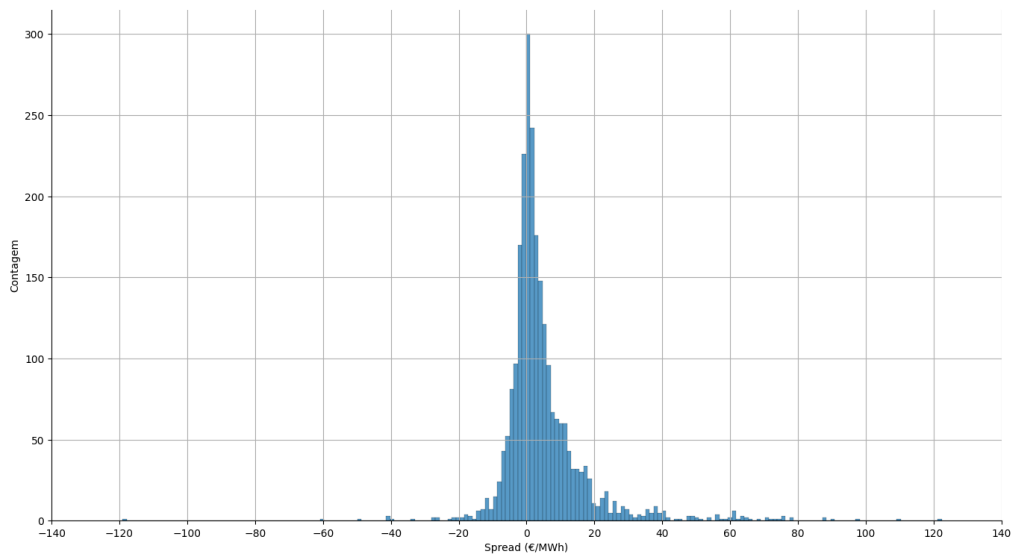


Figura 3.7: Histograma dos spreads

A análise da evolução da separação de mercado pode ser expandida no aspeto de como esta ocorre. Como foi referido no capítulo anterior, a separação de mercado é causada pelo congestionamento das interligações entre duas zonas de mercado, a portuguesa e espanhola no caso em estudo, causada por um fluxo de eletricidade que pode ter dois sentidos possíveis. Considere-se então o conceito de país importador e exportador durante ocorrências de separação de mercado, em que o importador verifica um preço de equilíbrio de mercado diferente e superior ao do exportador. A Figura 3.8 demonstra a evolução da zona de mercado importadora em momentos de separação de mercado. Com exceção dos anos de 2019 e 2022, verifica-se uma diminuição da percentagem de horas em que Portugal representa a zona de mercado importadora face ao número de horas total em separação de mercado, efetivamente atingindo um equilíbrio com Espanha em 2021.

3.3. ANÁLISE DE PADRÕES DE OCORRÊNCIA DE SEPARAÇÃO DE MERCADO

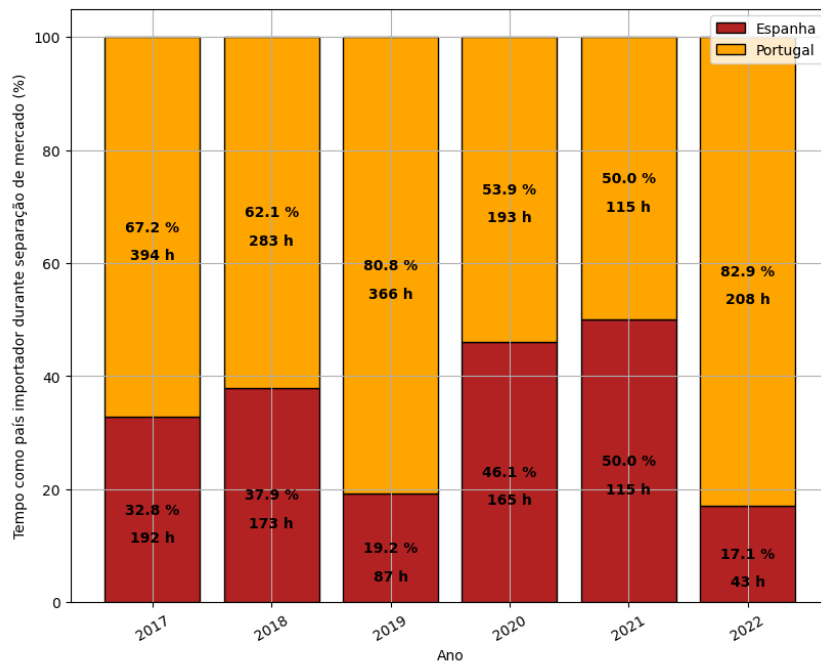


Figura 3.8: Evolução do país importador durante a separação de mercado

3.3 Análise de padrões de ocorrência de separação de mercado

Foi efetuada uma análise mais detalhada dos possíveis padrões de ocorrência de separação de mercado ao longo do ano e nas 24 horas do dia. Para isso, recorreu-se a uma técnica de visualização de dados, em concreto o **mapa de calor**, na representação da contagem de ocorrências, inspirado na metodologia utilizada em [34] para estudar a convergência de preços no **CWE**. Considerou-se inicialmente o período de **2017 a 2022** para três abordagens diferentes: todas as ocorrências de separação de mercado na Figura 3.9; apenas as associadas a um fluxo de eletricidade no sentido Espanha-Portugal na Figura 3.10; e apenas as associadas ao fluxo no sentido contrário (zona de mercado espanhola como importadora), Figura 3.11. Isto permitiu identificar padrões exclusivos a cada um destes últimos dois cenários.

CAPÍTULO 3. TRATAMENTO DE DADOS E ANÁLISE ESTATÍSTICA DA SEPARAÇÃO DE MERCADO

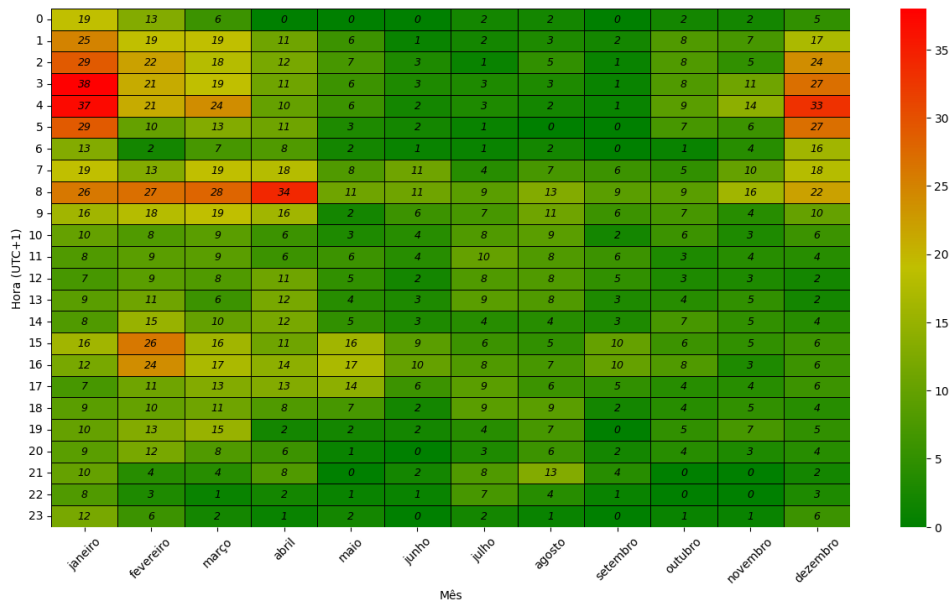


Figura 3.9: Mapa de calor de todas ocorrências de separação de mercado (2017 a 2022)

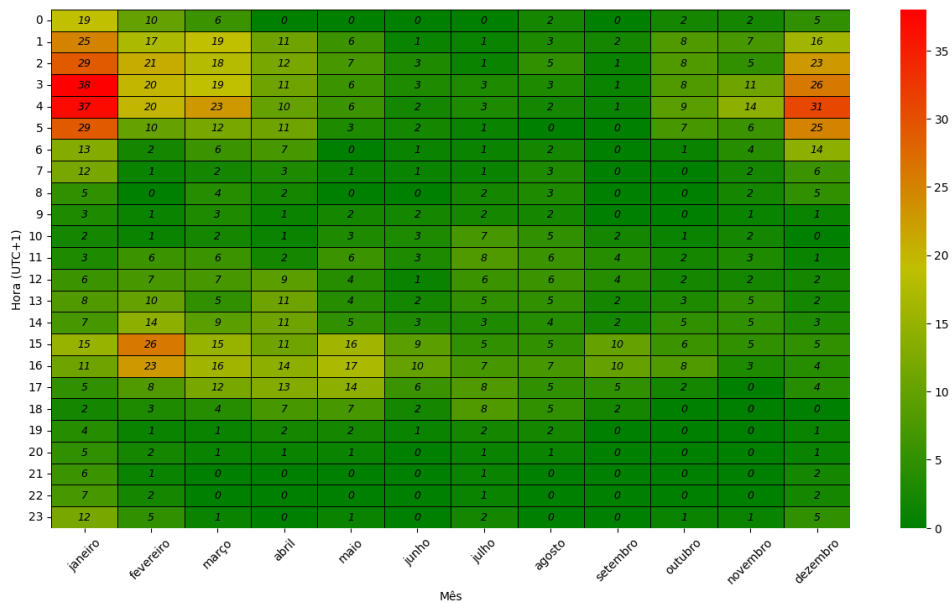


Figura 3.10: Mapa de calor das ocorrências de separação de mercado no sentido Espanha-Portugal (2017 a 2022)

3.3. ANÁLISE DE PADRÕES DE OCORRÊNCIA DE SEPARAÇÃO DE MERCADO

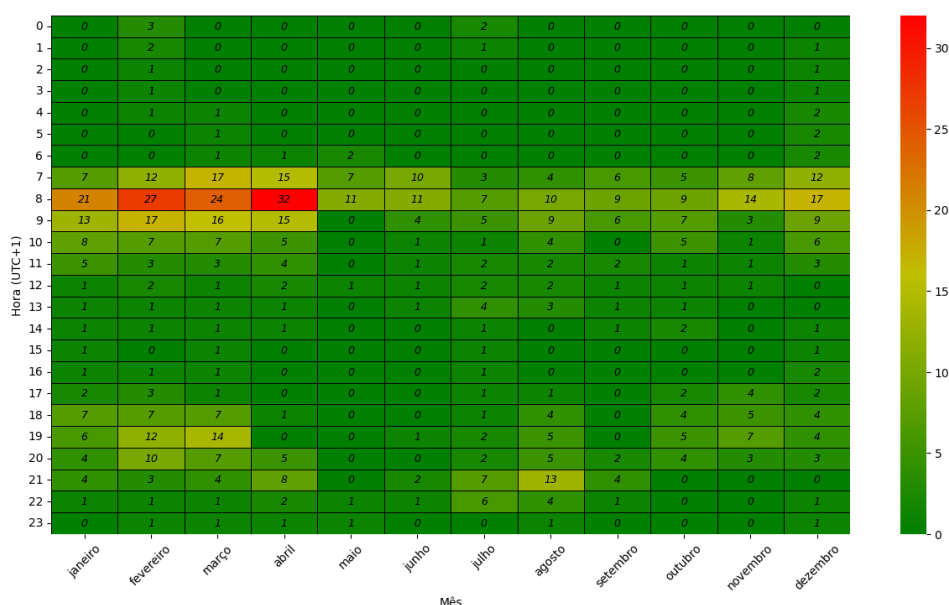


Figura 3.11: Mapa de calor das ocorrências de separação de mercado no sentido Portugal-Espanha (2017 a 2022)

É possível constatar na Figura 3.9 que, para o período considerado, o instante com maior frequência de ocorrências é às 4 da manhã (UTC+1) durante o mês de janeiro. Pode-se também notar que as células na proximidade desta, nomeadamente as que dizem respeito aos meses de dezembro, janeiro, fevereiro e março entre a meia-noite e 6 da manhã, também registaram um elevado número de ocorrências. Este padrão também surge no mapa de calor referente às ocorrências nas quais a zona de mercado portuguesa é importadora, Figura 3.10, sendo possível associar as mesmas. Este fenómeno poderá possivelmente ser explicado pela elevada disponibilidade de eletricidade de baixo custo marginal a estas horas e em especial nesta altura do ano, nomeadamente a produção eólica. A diferença significativa no tamanho dos mercados, acoplado ao facto de os consumos nacionais serem baixos para as horas em questão, poderá levar a que, seguindo uma lógica de ordem de mérito económico, a procura nessas horas seja satisfeita exclusivamente por ofertas associadas a tecnologias de muito baixo custo marginal, em que a zona de mercado espanhola detém a maioria dessas ofertas, provocando um fluxo de eletricidade neste sentido particular.

Por inspeção da Figura 3.11 podemos verificar que o maior número de ocorrências de separação de mercado no sentido Portugal-Espanha acontece às 8h da manhã (UTC+1) em quase todos os meses do ano, atingindo um máximo de 32 ocorrências no mês de abril na hora em questão. Este padrão poderá ser explicado pela diferença horária entre Portugal e Espanha e como este aspeto afeta os respetivos diagramas de carga. O período entre as 8 horas e 9 horas da manhã representa uma subida gradual nos perfis de carga de ambos os países, e dado que Portugal utiliza um fuso horário que está 1 hora atrasado

CAPÍTULO 3. TRATAMENTO DE DADOS E ANÁLISE ESTATÍSTICA DA SEPARAÇÃO DE MERCADO

em relação ao fuso horário espanhol, temos então um possível desequilíbrio entre os consumos nacionais face às respetivas produções, gerando-se assim um fluxo no sentido do sistema elétrico com uma maior necessidade relativa. A Figura 3.12 mostra os perfis de carga médios de Portugal e Espanha para o mês de fevereiro em 2017, ano que demonstra de forma acentuada o padrão identificado.

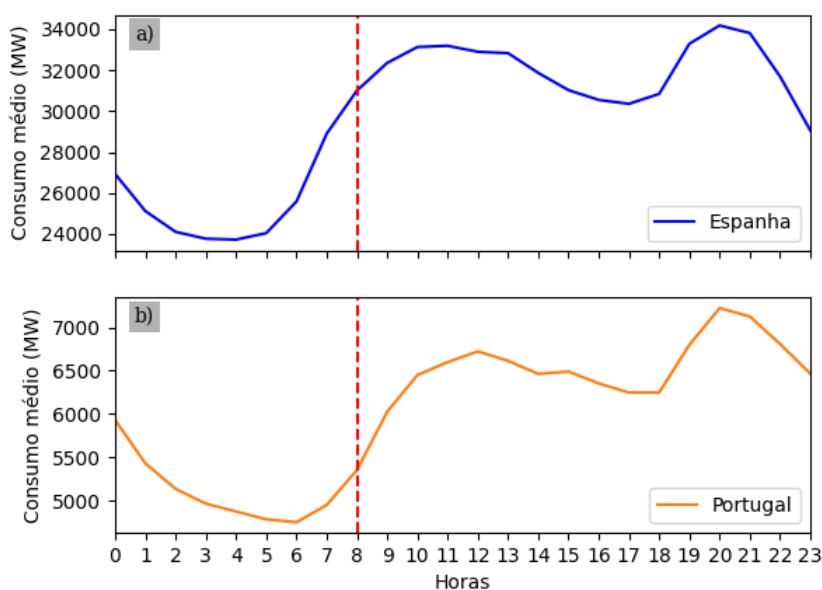


Figura 3.12: Perfis de carga médios para fevereiro de 2017

Como foi referido, os perfis têm formas muito semelhantes, embora estejam em escalas diferentes, e é possível identificar um desfasamento de 1 hora entre eles, especialmente no período das 7 até às 10 horas da manhã. Note-se que este padrão de ocorrência de separação de mercado tem proeminência nos meses mais frios (novembro a abril) que geralmente implicam maiores consumos e conseqüentemente uma maior diferença no valor absoluto do consumo entre os períodos de vazio e ponta, o que poderá acentuar o efeito que este desfasamento nos perfis de carga tem no mercado diário.

Este tipo de análise foi também empregada no estudo dos padrões de cada ano individual, na qual foi possível constatar que os padrões aqui descritos podem tomar diferentes níveis de intensidade ao longo dos anos, algo consistente com as observações da diminuição da ocorrência da separação de mercado e nas alterações na zona de mercado importadora. Em particular, o ano de 2022 demonstra de uma forma fraca os padrões identificados, mas mostra uma forte incidência da separação de mercado para o período das 10 às 18 horas nos meses de julho e agosto, menos comum nos restantes anos. Uma possível explicação para este tipo de padrão está relacionada com a produção solar espanhola, visto que existe uma maior capacidade instalada face ao consumo nacional quando comparada à portuguesa. Neste período do dia, e para os meses de julho e agosto, verifica-se uma

elevada produção solar, o que significa uma maior quantidade de ofertas de baixo custo marginal na zona de mercado espanhola, esgotando assim a capacidade de interligação. A Figura 3.13 contém o mapa de calor referente ao ano de 2022.

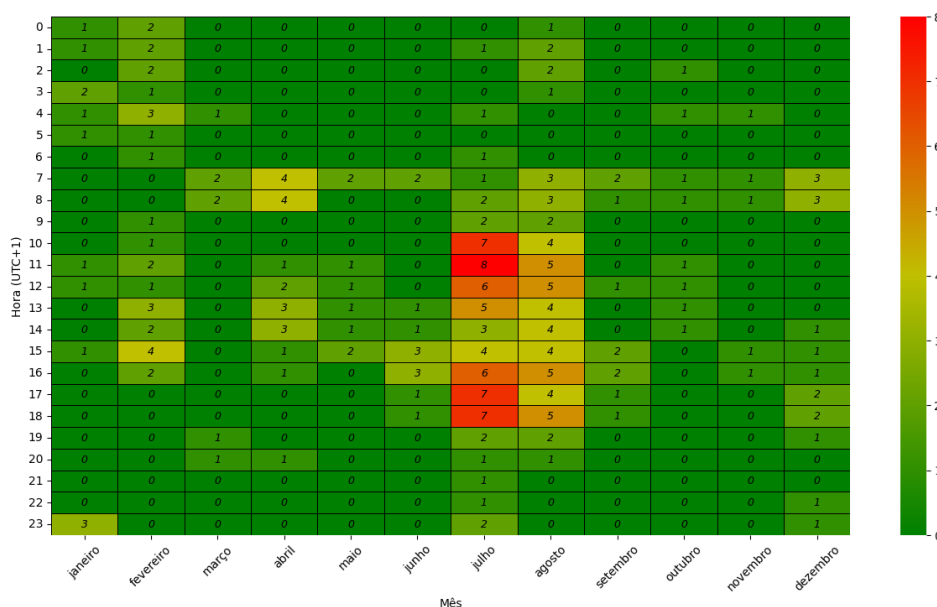


Figura 3.13: Mapa de calor de todas as ocorrências de separação de mercado (2022)

3.4 Análise de correlação

No âmbito de obter uma melhor compreensão das relações entre as variáveis que visam explicar a separação de mercado e a diferença de preços que esta produz, decidi-se efetuar uma análise de correlação. A correlação é uma ferramenta estatística que permite estabelecer uma associação linear ou não-linear entre duas variáveis. A medição da correlação é feita através de coeficientes de correlação, que de uma forma quantitativa numa escala de -1 a 1, representa a força da associação entre as variáveis observadas. Quando o coeficiente é nulo não existe correlação e as variáveis são independentes, enquanto para um valor de 1 ou -1 representa uma correlação perfeita e indica, respetivamente, uma associação positiva ou negativa. Note-se que a presença de correlação não é suficiente para inferir a presença de uma relação causal entre as variáveis, pelo que deve-se ter alguma cautela nas conclusões retiradas neste tipo de análise.

3.4.1 Coeficientes de correlação

Existem vários coeficientes de correlação, sendo um dos mais comuns o coeficiente de correlação de Pearson. Este coeficiente é sensível apenas a relações lineares e pode

ser obtido pela divisão entre a covariância de duas variáveis, X e Y , com o produto dos respetivos desvios padrões, σ_X e σ_Y , como da seguinte forma:

$$r_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \quad (3.8)$$

No caso da relação entre as variáveis não ser linear é necessário recorrer a outros coeficientes. O coeficiente de Spearman é uma medida não paramétrica de correlação de postos, geralmente identificada por ρ (rho). Esta utiliza, em vez do valor real das variáveis, os valores correspondentes à ordenação das suas observações, ou postos. No caso de todos os n postos serem inteiros e diferentes, ou seja, n é correspondente ao número de observações, utiliza-se a seguinte fórmula no seu cálculo:

$$\rho_{X,Y} = 1 - \frac{6 \sum_{i=0}^n d_i^2}{n(n^2 - 1)} \quad (3.9)$$

Na equação 3.9, d_i corresponde à subtração entre o posto da variável X , numa dada instância i , com o posto da variável Y , na mesma instância. Esta correlação é sensível a relações monótonas, podendo estas ser lineares.

Uma alternativa ao coeficiente de Spearman é o coeficiente de correlação τ de Kendall. À semelhança do último coeficiente, este também é uma medida de correlação de postos não paramétrica, sendo geralmente descrita como uma medida de concordância entre dois conjuntos de classificações. Para uma dada observação i , ao qual corresponde o par (x_i, y_i) , são contados os pares de entre todas as observações em que x_i e y_i concordam em termos de relação de classificação. Considere-se o par da observação j , (x_j, y_j) . Este é contado como sendo um par concordante se $x_i > x_j$ e $y_i > y_j$ ou $x_i < x_j$ e $y_i < y_j$, discordante caso contrário. No caso em que $x_i = x_j$ ou $y_i = y_j$ o par não é considerado na contagem. O coeficiente pode então ser calculado da seguinte forma:

$$\tau_{X,Y} = \frac{(n^\circ \text{ de pares concordantes}) - (n^\circ \text{ de pares discordantes})}{(n^\circ \text{ total de pares possíveis})} \quad (3.10)$$

3.4.2 Correlação entre variáveis de estudo e spreads de preços de eletricidade

Estes três coeficientes de correlação foram utilizados na análise da relação entre a diferença de preços durante separação de mercado, definido como Spread, e as produções de eletricidade, consumos e ATCs. Isto permite observar possíveis associações entre uma medida de “intensidade” de separação de mercado e as variáveis que a procuram explicar. Algumas das combinações com pelo menos um dos coeficientes, em módulo, acima de 0,3, considerada uma correlação moderada [49], encontram-se representadas nas Figuras abaixo. A Tabela 3.2 contém todos os coeficientes calculados que respeitam o mesmo critério (>0,3 em módulo). Os coeficientes foram calculados recorrendo ao módulo Python: Scipy [50].

Deve-se notar que as correlações aqui estudadas dizem respeito a associações entre a diferença de preços, em momentos de separação de mercado, e as variáveis consideradas,

pelo que não são indicativas das causas de separação de mercado, mas sim dos fatores que influenciam o aumento ou diminuição da diferença de preços observada. Como foi referido previamente, podemos olhar para a diferença de preços como uma medida da “intensidade” da separação de mercado, em que diferenças próximas de 0 representam cenários com pequenos desequilíbrios entre as zonas de mercado, enquanto diferenças elevadas poderão implicar falhas significativas.

Variável	r (Pearson)	ρ (Spearman)	τ (Kendall)
<i>ATC_{PT-ES}</i>	0,095	0,308	0,210
<i>Hídrica_{PT}</i>	-0,430	-0,659	-0,477
<i>Solar_{PT}</i>	0,459	0,128	0,096
<i>Gás_CC_{PT}</i>	-0,252	-0,490	-0,345
<i>Gás_{PT}</i>	-0,419	-0,491	-0,349
<i>Consumo_{ES}</i>	-0,299	-0,508	-0,336
<i>Solar_PV_{ES}</i>	0,456	0,110	0,074
<i>Hídrica_{ES}</i>	-0,476	-0,731	-0,543
<i>Gás_CC_{ES}</i>	-0,138	-0,411	-0,281
<i>Cogeração_Resto_{ES}</i>	-0,415	-0,375	-0,260
<i>Hídrica_{PT-contribuição}</i>	-0,436	-0,660	-0,478
<i>Solar_{PT-contribuição}</i>	0,477	0,133	0,100
<i>Gás_CC_{PT-contribuição}</i>	-0,239	-0,487	-0,344
<i>Gás_{PT-contribuição}</i>	-0,394	-0,452	-0,317
<i>Eólica_{ES-contribuição}</i>	0,153	0,361	0,246
<i>Solar_PV_{ES-contribuição}</i>	0,480	0,140	0,097
<i>Solar_Térmica_{ES-contribuição}</i>	0,308	0,277	0,190
<i>Hídrica_{ES-contribuição}</i>	-0,483	-0,721	-0,535
<i>Gás_CC_{ES-contribuição}</i>	-0,082	-0,332	-0,228
<i>Nuclear_{ES-contribuição}</i>	0,328	0,548	0,373

Tabela 3.2: Coeficientes de correlação

Os sinais dos coeficientes de correlação indicam se a correlação de cada variável com o spread de preços é positiva ou negativa. Coeficientes negativos indicam que o aumento do valor da variável tende a levar à diminuição dos spreads, ou seja as variáveis discordam, enquanto que coeficientes positivos indicam que o aumento do valor da variável tende a levar ao aumento dos spreads, as variáveis concordam.

CAPÍTULO 3. TRATAMENTO DE DADOS E ANÁLISE ESTATÍSTICA DA SEPARAÇÃO DE MERCADO

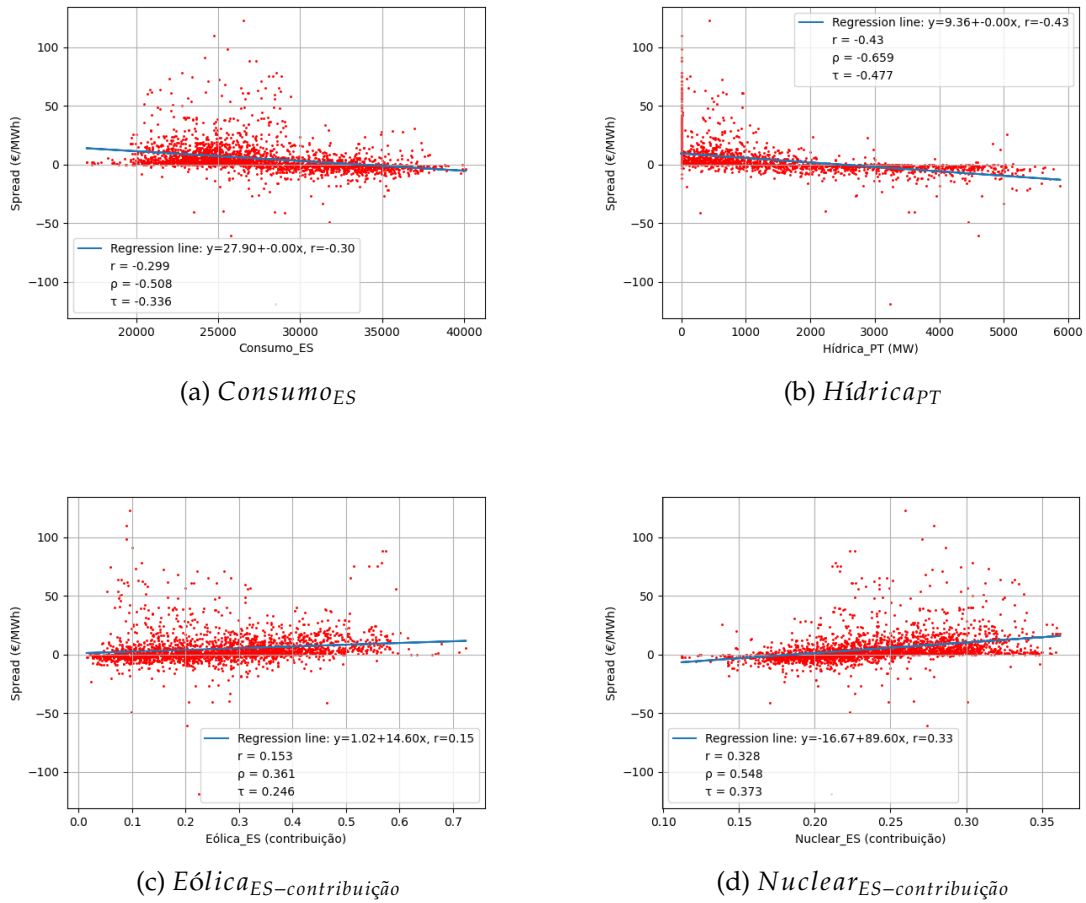


Figura 3.14: Gráficos de dispersão de algumas variáveis versus os spreads de preços de eletricidade

A análise dos gráficos de dispersão das variáveis de estudo versus os spreads calculados permitem retirar algumas ilações acerca do sentido para o qual ocorre de separação de mercado. A Figura 3.14a mostra o gráfico de dispersão do spread versus o consumo espanhol, onde se pode visualizar uma correlação negativa, ou seja, o aumento do consumo leva a uma diminuição dos spreads calculados. Para valores baixos de consumo temos spreads de valor positivo, que segundo a equação 3.1 do Capítulo 3.1, representa uma separação de mercado na qual a zona de mercado portuguesa é a importadora, enquanto que para elevados valores do consumo observa-se uma inversão de sentido (zona de mercado espanhola como importadora). Na Figura 3.14b, referente à produção hídrica portuguesa, verifica-se uma relação semelhante, em que um baixo nível de produção está associado a spreads positivos, e vice-versa.

É identificada uma correlação positiva na Figura 3.14c, gráfico de dispersão da contribuição da produção eólica espanhola para o consumo nacional. Algumas variáveis apenas demonstraram coeficientes de correlação acima de 0,3 nas suas versões de contribuição. Constata-se no gráfico, e no valor de sinal positivo dos coeficientes de correlação, que com o aumento da fração de consumo satisfeita pela produção eólica também aumenta o spread calculado. Para baixas frações os spreads são maioritariamente negativos, o que

está associado a separação de mercado no sentido Portugal-Espanha, enquanto frações elevadas até 0,7 coincidem principalmente com separação de mercado no sentido oposto. A produção eólica é de baixo custo marginal e uma parte essencial do mix energético espanhol, o que significa que fraca produção poderá levar à necessidade de importar. O contrário acontece quando há uma forte produção, que em conjunto com outras tecnologias de baixo custo marginal, como a solar e nuclear (tecnologia inexistente em Portugal), permitem por vezes satisfazer toda a procura do mercado, levando a uma separação de mercado no sentido Espanha-Portugal caso as ofertas provenientes da zona portuguesa não consigam competir (baixa produção renovável). O gráfico de dispersão da Figura 3.14d, referente à produção nuclear, partilha da mesma lógica.

Os gráficos de dispersão associados às tecnologias solares demonstram uma forma peculiar que poderá, inadvertidamente, levar a maiores coeficientes de correlação. Visto que para grandes períodos do dia estas tecnologias têm uma produção nula, o gráfico de dispersão mostra uma grande concentração de pontos para o valor 0 de produção e ao longo do eixo dos spreads.

METODOLOGIA

Devido à complexidade deste tipo de estudo, a análise estatística, por si só, é insuficiente na tarefa de caracterização adequada das causas de separação de mercado no MIBEL. Por este motivo, alguns estudos recorrem ao uso de modelos explicativos para extrair conhecimento dos dados referentes ao mercado de eletricidade. Consideram-se três modelos distintos, distribuídos ao longo de uma escala representativa da troca entre interpretabilidade e exatidão, um conceito adotado pela literatura relacionada com ML e o ramo que o visa tornar mais interpretável, XML, no entanto ainda muito debatido no que toca a esta relação de proporcionalidade inversa. As metodologias adotadas são, por uma hipotética ordem crescente de complexidade e decrescente na facilidade de interpretação:

- Regressão Logística;
- Random Forest;
- Rede Neuronal Artificial.

A escolha destes métodos é feita com base em recomendações da literatura que aborda temas semelhantes, no caso dos dois primeiros, e no potencial como classificador, no caso do terceiro. Decidiu-se utilizar mais que uma metodologia de modo a evitar a dependência nos resultados de um único modelo. Note-se que o objetivo do estudo não é o de comparar as metodologias, mas sim de obter várias explicações complementares acerca dos fatores determinantes da separação de mercado suportadas pelos modelos explicativos construídos.

Neste Capítulo, serão introduzidos ao nível teórico os modelos escolhidos, tal como as técnicas que serão utilizadas nas suas respetivas interpretações.

4.1 Divisão do conjunto de dados

Antes de se efetuar a implementação dos modelos propostos, teve-se um cuidado especial na separação dos dados utilizados na construção dos mesmos e nas suas sucessivas testagens. Os dados foram estratificados de modo a que as amostras de treino e teste

tenham a mesma distribuição de casos de ocorrência de separação de mercado. Isto foi feito recorrendo ao módulo Python: **scikit-learn** [51] em particular a sua função `train_test_split`. Considerou-se para o treino e teste dos modelos o período de 2017 a 2021, pelo que ficaram de parte os dados referentes a 2022 e início de 2023 para uma comparação póstuma de todos modelos utilizados. Escolheu-se um rácio de **80%** dos dados para treino, que corresponde a **35059 observações**, e **20%** para a testagem, que corresponde a **8765 observações**, valores considerados comuns em datasets largos e que neste caso em particular permitiu ter o equivalente a um ano de dados na amostra de teste. Não foram discriminados quais os anos a utilizar no conjunto de treino ou no de teste pois a estratificação trata-se de uma amostragem aleatória.

4.2 Regressão Logística

A **regressão logística**, ou modelo **Logit**, e similares como a regressão probit e tobit, são modelos estatísticos de probabilidade direta que podem ser utilizados em classificação binária. O modelo Logit em particular, modela a probabilidade de ocorrer um dado evento através da representação das log-odds do mesmo como a combinação linear de variáveis independentes [52]. Este modelo já foi utilizado no estudo da separação de mercado no **MIBEL** para o período de 2008 a 2012 em [5], enquanto [48] e [35] são exemplos da aplicação de modelos probit em tópicos relacionados como a análise dos determinantes do congestionamento das interligações e convergência de preços entre zonas de mercado, respetivamente, para a zona **CWE**.

4.2.1 Introdução teórica

De modo a possibilitar a modelação da probabilidade de uma dada variável Y pertencer a uma de duas classes (classificação binária) é necessário que a resposta do modelo esteja contida entre 0 e 1. Na regressão logística, faz-se uso da função logística, que pode ser descrita da seguinte maneira [53]:

$$p(X) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}} \quad (4.1)$$

Alguma manipulação da equação 4.1 permite chegar a:

$$\frac{p(X)}{1 - p(X)} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k} \quad (4.2)$$

Do lado esquerdo da equação 4.2 temos uma quantidade definida como as odds, que pode tomar qualquer valor entre 0 e ∞ . Os valores próximos de 0 e ∞ desta quantidade são indicativos, respetivamente, de uma baixa e elevada probabilidade de pertencer à classe positiva. Ao aplicar o logaritmo em ambos os lados da equação 4.2 chegamos a:

$$\log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (4.3)$$

O lado esquerdo da equação 4.3 chama-se log-odds da variável de resposta, mais conhecido por logit, que pode ser modelado como a combinação linear das k variáveis de entrada. Chegamos então a um modelo linear generalizado com uma função de ligação logit, o que permite considerar uma distribuição binomial da variável de resposta. Os coeficientes β são estimados através de algoritmos de máxima verosimilhança, que de uma forma sucinta, fazem ajustes sucessivos destes coeficientes de modo a que $p(X)$ gere valores próximos de 0 e 1 para as respetivas instâncias classificadas desse modo. Este raciocínio pode ser formalizado através de uma função de verosimilhança, $\mathcal{L}(\beta)$, cujo correto ajuste dos coeficientes β representa a sua maximização:

$$\mathcal{L}(\beta) = \prod_{i:y_i=1} p(X_i) \prod_{i':y'_{i'}=1} (1 - p(X_{i'})) \quad (4.4)$$

4.2.2 Implementação

Tendo em conta as capacidades reduzidas deste tipo de modelo e face à complexidade da separação de mercado que pode ocorrer para dois cenários distintos, decidiu-se construir dois modelos separados. Um vai modelar as ocorrências de separação de mercado em que a zona de mercado portuguesa é a importadora, OSM_{ES-PT} , enquanto o outro trata do cenário oposto, OSM_{PT-ES} , permitindo assim reduzir a complexidade. Recorreu-se a um processo semelhante ao aplicado em [35] para determinar as variáveis relevantes ao modelo baseado nos valores-p associados a cada.

As variáveis consideradas inicialmente foram todas as produções de eletricidade na forma de contribuições, consumos nacionais e ATCs. Estas últimas são de particular interesse pois representam a limitação na importação ou exportação, pelo que se espera que tenham a maior relevância para o modelo. Foi considerado apenas o ATC pertinente a cada cenário na lista inicial de variáveis do respetivo modelo, sendo o outro ATC omitido. Esta escolha advém do facto de que a separação de mercado implica um fluxo de eletricidade num dado sentido, pelo que a capacidade de transferência disponível para o sentido contrário é irrelevante no mesmo contexto.

Foi necessário padronizar os dados de treino e teste referentes às variáveis de entrada do modelo, por forma a obter coeficientes que permitam comparar as variáveis explicativas quanto às magnitudes relativas dos seus efeitos no modelo [54]. Isto é feito através do redimensionamento das variáveis de modo a que a média e desvio padrão sejam 0 e 1, respetivamente, algo que pode ser facilmente aplicado aos dados em questão com recurso à classe **StandardScaler** do módulo Python: **scikit-learn** [51].

Os modelos de regressão logística foram implementados recorrendo ao módulo Python: **statsmodels** [55] em específico a classe **Logit**. Foi necessário adicionar uma constante unitária aos dados referentes às variáveis exógenas para que o modelo possa ter uma ordenada na origem diferente de 0, sendo esta interpretada como sendo as log-odds da saída do modelo quando todas as variáveis, ou previsores, encontram-se no seu valor de

referência. A estimativa dos coeficientes por máxima verosimilhança é feita através do método de **Newton-Raphson** integrado no módulo Python utilizado.

4.3 Random Forest

A **Random Forest** é um método de **ensemble learning** (agregação) que consiste na construção de várias **árvores de decisão** (estimadores) durante o treino da mesma. Num contexto de classificação, este algoritmo faz as suas previsões através da determinação da classe mais votada pelas árvores de decisão. Este tipo de modelo já foi utilizado no estudo da convergência de preços no **CWE** [34], onde foi implementado como um modelo explicativo de uma variável binária de convergência (ou não) dos preços de eletricidade de diferentes zonas de mercado.

4.3.1 Introdução teórica

Dado que este algoritmo é composto por múltiplas árvores de decisão, é importante ter alguma compreensão de como estas funcionam. A árvore de decisão, neste caso árvore de classificação, prevê que uma dada observação pertence à classe mais comum do conjunto de treino na região a que a mesma pertence [53]. Isto é feito através de uma separação binária recursiva dos dados de treino, perante um certo critério, que resulta na atribuição de uma classe, ou probabilidade de pertencer a esta, a subconjuntos dos dados. Um dos algoritmos de construção de árvores de decisão é o **CART** (Classification and Regression Trees) [56], que foi utilizado neste estudo.

Uma descrição simplificada do algoritmo de CART começa pela definição de um “**nó de raiz**” (root node), que corresponde ao conjunto de dados de treino completo, e que sendo por enquanto o único nó existente é também considerado um **nó terminal**. Enquanto o conjunto de dados num nó terminal for de tamanho razoável e houver a presença de instâncias pertencentes a diferentes classes, é feita uma partição dos dados em duas partes (split) com base na **regra de separação**, criando assim dois novos nós terminais. Isto é conhecido como um algoritmo de indução de cima para baixo, considerada uma das estratégias mais comuns na construção de árvores de decisão [57].

As regras de separação utilizadas em cada split são escolhidas segundo um critério que decide qual a variável e que condição é que levam à melhor separação possível. No caso do algoritmo CART é o **índice de Gini** que pode ser expresso da seguinte forma:

$$G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk}) \quad (4.5)$$

em que \hat{p}_{mk} representa a proporção de observações do conjunto de treino pertencentes a uma região m que são da k -ésima classe, que num contexto de classificação binária pode ser uma de duas: 0 ($k = 1$) ou 1 ($k = 2$). Temos então que este índice é uma medida da variância total nas K classes e que é vista como uma quantificação da pureza de

um nó, ou seja, um valor pequeno indica que o nó contém maioritariamente observações pertencentes a uma única classe [53]. Existem critérios alternativos, como a entropia, que também são utilizados em outros algoritmos populares.

Devido ao hábito de sobreajuste (overfitting) das árvores de decisão aos dados de treino, consideram-se técnicas de agrupamento (ensemble) que permitem corrigir este problema, nomeadamente a **Random Forest** [58]. Este algoritmo constroi múltiplas árvores de decisão, recorrendo a uma amostragem por bagging e seleção aleatória de características, por forma a criar árvores diferentes. Nas previsões, a classe mais votada pelas árvores é considerada a classe prevista pela Random Forest.

4.3.2 Implementação

A implementação do algoritmo Random Forest foi feita com recurso à classe **RandomForestClassifier** do módulo Python: scikit-learn [51]. Esta classe permite a escolha do critério determinante das árvores de decisão, os estimadores, do qual se escolheu a **impureza de Gini**. O modelo foi otimizado quanto ao número de estimadores e ao peso ideal da classe minoritária através da sucessiva testagem do modelo com a variação destes parâmetros. Os valores ótimos, tal como os intervalos de variação estão representados na Tabela 4.1.

Hiperparâmetro	Intervalo de otimização	Valor ótimo
Número de estimadores	[10; 200]	50
Peso da classe positiva	[1; 21]	4

Tabela 4.1: Hiperparâmetros otimizados do algoritmo Random Forest implementado

Note-se que o valor máximo no intervalo de otimização do peso da classe positiva poder ser visto como o peso, arredondado, para o qual a classe minoritária (positiva neste caso) é replicada, implicitamente, até atingir o mesmo número de observações da classe maioritária.

4.4 Rede Neuronal Artificial

Tal como a Random Forest, a **RNA** insere-se no ramo do **ML** e da aprendizagem supervisionada. Esta metodologia é comumente utilizada em tarefas de previsão dos preços de eletricidade [59, 60], carga [61, 62] e até, por exemplo, da necessidade de serviços auxiliares [63]. Também possui a capacidade de atuar como um classificador, tendo sido usada na previsão de congestionamento das linhas de transmissão, modelado como uma variável binária, na singular zona de mercado alemã em [64]. No entanto, ao contrário dos últimos dois métodos apresentados, e de que se tenha conhecimento, as **RNAs** não foram utilizadas, até à data, em estudos compreensivos dos fatores afetos à separação de mercado ou convergência de preços, pelo que a preferência no seu uso advém da sua implementação em outros estudos relacionados, e da vasta literatura que facilita a sua compreensão e implementação.

4.4.1 Introdução teórica

As RNAs têm como inspiração a estrutura e funcionalidade das redes de neurónios biológicos encontrados nos cérebros dos animais. Estas são geralmente descritas como a interconexão de unidades de computação básicas, também conhecidas como neurónios, que através da atualização sucessiva dos pesos associados a cada interconexão, são capazes de aprender a realizar uma dada tarefa ou função, imitando assim o processo de aprendizagem do cérebro. Um elemento comum a todas as arquiteturas de RNAs é a existência de uma única camada de entrada e de saída, em que a primeira serve de ponto de entrada de dados, onde não é feita nenhuma computação, e a segunda que gera os sinais de saída da rede [65].

A arquitetura proposta para este estudo é uma Multilayer Feedforward Network, em específico uma variação conhecida como Multilayer Perceptron, diferenciada pelo facto de todos os neurónios de uma dada camada estarem completamente interconectados aos neurónios de camadas adjacentes [65]. O termo Multilayer, ou multi-camadas, alude ao facto de existirem mais camadas para além da entrada e saída, que, sendo impossível observá-las por nenhuma dessas duas extremidades, recebem o nome de hidden layers, ou camadas “ocultas”. Esta diferença importante serve como base dos algoritmos de Deep Learning, um subconjunto do ML. A arquitetura básica desta rede pode ser observada na Figura 4.1.

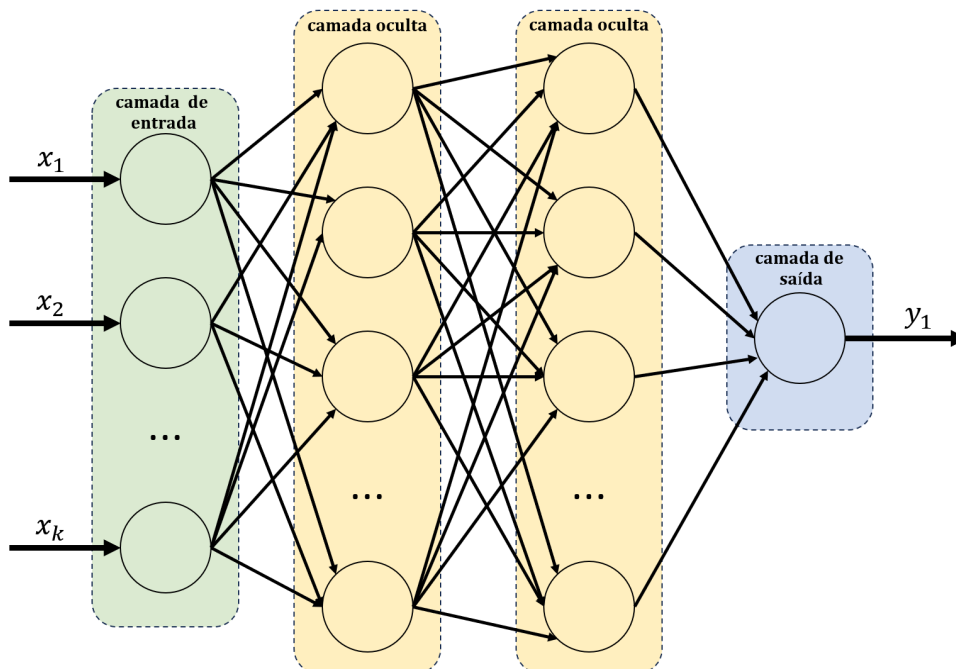


Figura 4.1: Arquitetura de um Multilayer Perceptron

O neurónio artificial, representado na Figura 4.2, pode ser descrito da seguinte forma:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (4.6)$$

e

$$y_k = \varphi(u_k + b_k) \quad (4.7)$$

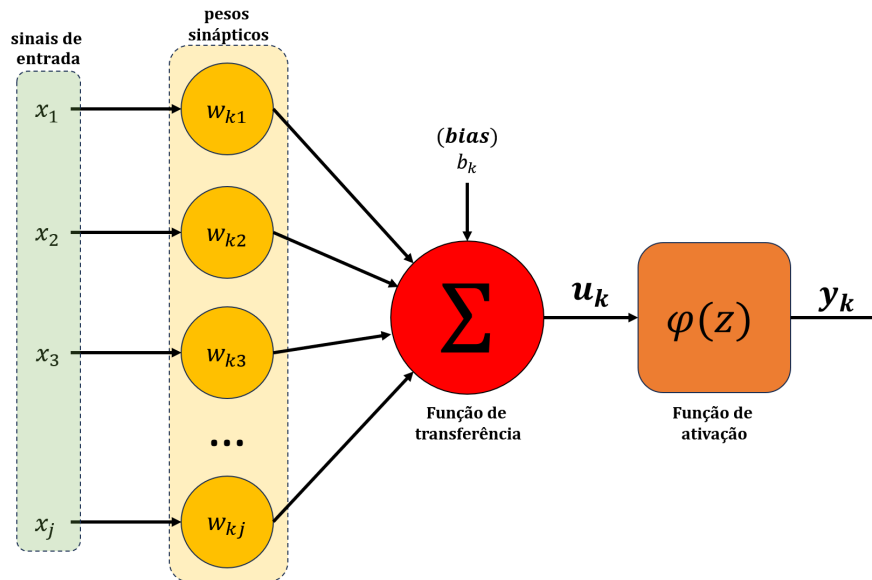


Figura 4.2: Estrutura de um neurónio

A equação 4.6 descreve um dado neurónio k como a soma dos seus sinais de entrada x_j multiplicados pelos respetivos pesos sinápticos w_{kj} . Já a equação 4.7 representa o sinal de saída desse mesmo neurónio, y_k , que pode ser descrito como a resposta de uma função de ativação, $\varphi(z)$, aquando tem como valor de entrada a combinação linear de u_k e um bias b_k [65]. Existem várias possibilidades na escolha de uma função de ativação, mas tendo em conta a função pretendida do modelo, foram utilizadas duas funções distintas. Com a exceção do neurónio final da camada de saída, a função utilizada foi a Unidade Linear Retificada (ReLU), que pode ser descrita por:

$$\varphi(z) = ReLU(z) = \begin{cases} 0 & , z \leq 0 \\ z & , z > 0 \end{cases} \quad (4.8)$$

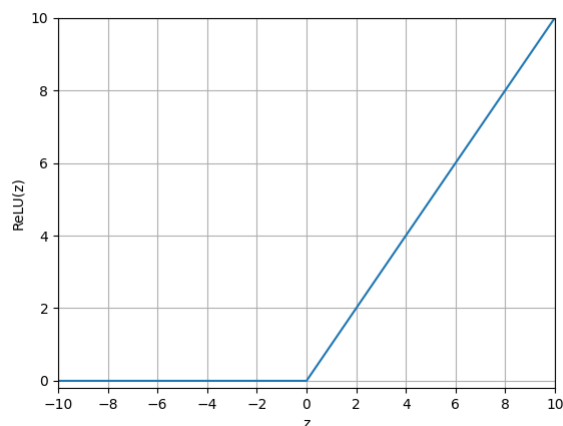


Figura 4.3: Função de ativação ReLU

No caso do singular neurónio na camada de saída, foi utilizada a função sigmoide. Esta função de ativação é particularmente vantajosa para o propósito de classificação binária pois o seu intervalo de resposta encontra-se entre 0 e 1. É possível descrevê-la da seguinte forma:

$$\varphi(z) = \sigma(z) = \frac{1}{1 + e^{-z}} \quad (4.9)$$

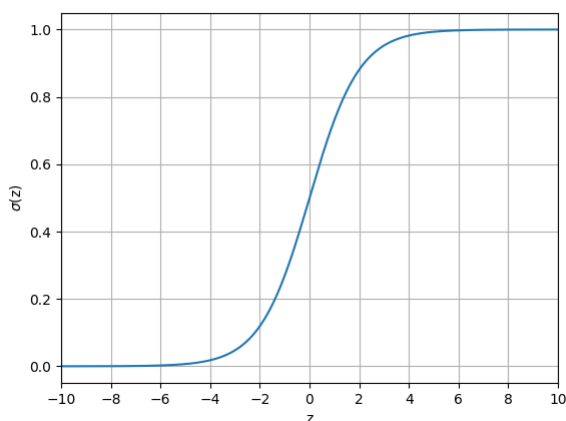


Figura 4.4: Função de ativação sigmoide

O treino da rede neuronal é feito através de um algoritmo de retropropagação (backpropagation). De uma forma simples, trata-se do ajuste dos pesos sinápticos baseado numa perda (loss) calculada entre o valor esperado e a saída do modelo. Existem várias escolhas possíveis no que toca à função representativa das perdas (loss function), sendo escolhida a função de “binary cross-entropy”, dado adequar-se ao propósito de classificação binária. Esta função, também conhecida por log loss, pode ser descrita da seguinte forma genérica:

$$L_{bce} = -\frac{1}{M} \sum_{m=1}^M [y_m \times \log(h_{\theta}(x_m)) + (1 - y_m) \times \log(1 - h_{\theta}(x_m))] \quad (4.10)$$

Onde M representa o número de amostras utilizadas no treino da rede para uma dada epoch, x_m os dados fornecidos à rede referentes à amostra m , y_m o valor real da variável alvo para essa mesma amostra m , e h_θ a rede neuronal com os pesos θ , resultantes do último ajuste [66].

4.4.2 Implementação

A implementação da rede neuronal artificial foi feita recorrendo ao módulo Python: **TensorFlow** [67], em específico recorrendo ao seu API de alto nível: **Keras**. Utilizou-se a classe **Sequential** para criar uma pilha linear de camadas. Foram adicionadas camadas do tipo **Dense**, que correspondem a camadas completamente conectadas, por forma a criar um **Multilayer Perceptron**.

Consideraram-se todas variáveis dos **consumos nacionais**, **ATCs** e produção de eletricidade por tecnologia com exceção de: **Biomassa**, **Carvão** e **Outra térmica** do lado português e **Térmica Renovável** e **Cogeração Resto** do lado espanhol. A exclusão destas variáveis advém do facto de serem consideradas estatisticamente irrelevantes para o caso em estudo, terem sido excluídas por elevados valores-p dos modelos Logit ou então por apresentarem algumas dificuldades na integração com os restantes dados, como acontece com a produção termoelétrica a carvão portuguesa que passa a ter valores nulos em grande parte dos dados de treino e teste.

Foi necessário criar um conjunto de dados de validação para medir o desempenho da generalização do modelo face ao seu treino. Para isso, foi preciso voltar a dividir o conjunto de dados de treino em duas partes. Escolheu-se um rácio de 75% para os dados de treino (**26294 observações**) e 25% (**8765 observações**) para os de validação, também estratificados, sendo que o conjunto de teste descrito no Capítulo 4.1 ficou inalterado. Os dados foram **normalizados** segundo recomendações para a implementação deste tipo de algoritmo.

Foi empregue um algoritmo de **otimização bayesiana** no ajuste (finetuning) dos hiperparâmetros da rede neuronal. O interesse neste tipo de otimização reside em encontrar o mínimo, ou o máximo, como é o caso, de uma função de caixa negra desconhecida. Isto é conseguido através da construção de um modelo probabilístico da função, que é utilizado para determinar sucessivamente com que valores dos seus parâmetros deve avaliar a função, o que é bastante vantajoso quando as avaliações da função são dispendiosas, como acontece com a maioria dos modelos de **ML** [68]. A função em questão é o cálculo da **área sob a curva precisão-revocação** (AUCPR) da classificação do modelo no conjunto de dados de validação, a métrica de validação recomendada quando na presença de conjuntos de dados fortemente desequilibrados [69].

Os hiperparâmetros ajustados foram: **número de neurónios nas camadas ocultas; número de camadas ocultas; taxa de aprendizagem; número de epochs de treino; batch size** e o **peso da classe positiva**. Este último é utilizado para endereçar o desequilíbrio entre as classes, pelo que se procura um peso adequado que valorize a classe positiva

(minoritária) acima da negativa (maioritária). Foi utilizado o módulo Python: **Bayesian Optimization** [70], que simplificou significativamente o processo de otimização. Foi ainda efetuado um ajuste manual de alguns dos hiperparâmetros, nomeadamente o número de epochs, onde se verificaram algumas pequenas melhorias no desempenho do modelo. Os parâmetros finais podem ser verificados na Tabela 4.2.

Hiperparâmetro	Intervalo de otimização	Valor ótimo
Neurónios nas camadas ocultas	[10; 30]	25
Camadas ocultas	[1; 3]	2
Taxa de aprendizagem	[0,001; 0,1]	0,084
Epochs	[200; 1000]	348
Batch size	[32; 2048]	1271
Peso da classe positiva	[1; 21]	7,4

Tabela 4.2: Hiperparâmetros otimizados da Rede Neuronal implementada

As Figuras 4.5a e 4.5b correspondem às curvas de treino da rede neuronal para a loss e a área sob a curva precisão-revocação, respetivamente, em que a curva azul representa o treino e a laranja a validação.

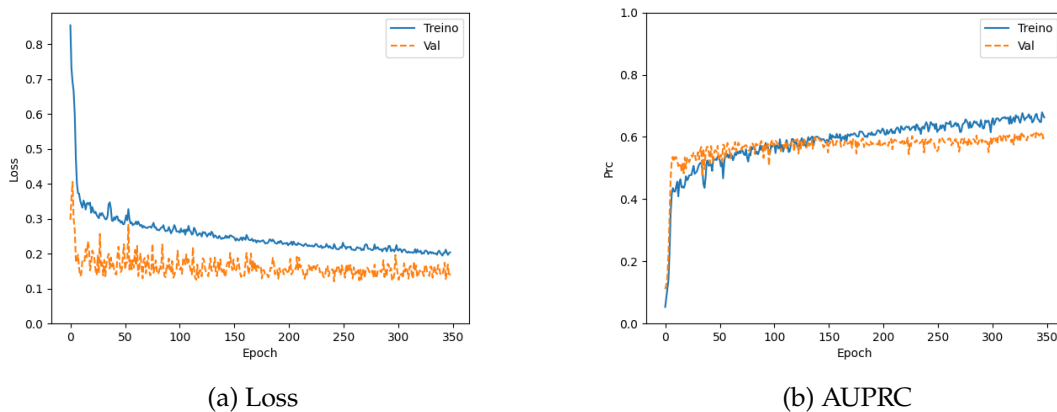


Figura 4.5: Curvas de treino e validação

4.5 Métricas de avaliação dos modelos

De modo a avaliar os modelos explicativos desenvolvidos, e tendo em conta que estes servem um propósito de classificação binária quanto à ocorrência de separação de mercado, são utilizadas algumas métricas apropriadas na apreciação da qualidade dos modelos. Algumas destas métricas são comuns a todos os modelos enquanto outras são utilizadas exclusivamente em certos métodos.

Em problemas de classificação binária é comum recorrer ao conceito de matriz de confusão como uma medida representativa das previsões feitas pelo modelo e de como

estas se relacionam com os valores reais. Isto é feito através da classificação de todas as previsões feitas pelo modelo segundo 4 grupos distintos [71]:

- **TN** (True Negative ou Verdadeiro Negativo) – Diz respeito às previsões negativas, ou seja, cuja classificação binária é 0, em que o valor real tem a mesma classificação;
- **FN** (False Negative ou Falso Negativo) – Trata-se das previsões negativas que contrariam os valores reais, sendo estes últimos positivos;
- **TP** (True Positive ou Verdadeiro Positivo) – Refere-se às previsões positivas, de classificação binária 1, em que o valor real é coincidente com esta classificação;
- **FP** (False Positive ou Falso Positivo) – Inclui as previsões positivas cujo valor real tem uma classificação contrária (negativo).

		VALORES REAIS	
		POSITIVO (1)	NEGATIVO (0)
VALORES PREVISTOS	POSITIVO (1)	TP	FP
	NEGATIVO (0)	FN	TN

Figura 4.6: Matriz de confusão

Note-se que qualquer previsão insere-se num único grupo, pelo que a soma de todos os elementos dos 4 grupos resulta na totalidade de previsões feitas pelos modelos. Um esquema da matriz de confusão encontra-se representado na Figura 4.6. Com base nestas classificações é possível construir métricas descritivas do desempenho dos modelos que retratam certos aspetos. Uma das mais utilizadas é a **exatidão**, que representa o rácio do número de classificações corretas (TN e TP) com o número de classificações total, que pode ser formalizado através da seguinte expressão:

$$exatidão = \frac{TN + TP}{TN + TP + FN + FP} \quad (4.11)$$

Esta métrica, embora expresse a capacidade do modelo de classificar corretamente dentro da amostra fornecida [72], pode ser enganadora quando utilizada em conjuntos de dados com um desequilíbrio severo entre as quantidades de casos positivos e negativos

[71], como é o caso para o estudo em questão. É então comum recorrer-se a outras métricas que permitem especificar uma dada componente das previsões. A **precisão** e a **revocação**, também conhecida como **sensibilidade**, são medidas que informam acerca do rácio de previsões positivas cujo valor real é positivo e o rácio de todos os valores reais positivos que são corretamente classificados pelo modelo como positivos, respetivamente [73]. Estas podem ser expressas, com base nos 4 grupos de previsões, da seguinte forma:

$$precisão = \frac{TP}{TP + FP} \quad (4.12)$$

e

$$revocação = \frac{TP}{TP + FN} \quad (4.13)$$

Estas métricas são particularmente importantes quando se valoriza a previsão positiva acima da negativa. Para o estudo em questão, a ocorrência de separação de mercado, que é caracterizada como sendo o caso positivo (valor 1 binário), tem uma maior importância que o caso negativo correspondente à convergência de preços marginais (ausência de separação de mercado). Uma apreciação da qualidade do modelo na previsão de casos negativos é geralmente feita através da métrica de **especificidade**. Esta trata-se do rácio de todos os valores reais negativos que foram corretamente classificados pelo modelo como sendo negativos [73], e que pode ser expressa como:

$$especificidade = \frac{TN}{TN + FP} \quad (4.14)$$

Outra métrica popular para casos de desequilíbrios entre as classes positiva e negativa é o **F1-score**. Esta corresponde à média harmónica entre a precisão e revocação e que é dada pela seguinte expressão:

$$F_1 = 2 \frac{precisão \times revocação}{precisão + revocação} \quad (4.15)$$

Esta medida permite avaliar o desempenho do modelo no reconhecimento dos casos positivos e na minimização de falsas classificações, sendo utilizado na utilização simultânea da precisão e revocação, especialmente em conjuntos de dados com desequilíbrios.

Dado que a saída dos modelos não é binária, mas sim um valor contido entre 0 e 1, é necessário definir um **threshold**, ou limite, que permita distinguir a classificação positiva da negativa. Geralmente este valor limite é 0,5, o que significa que todos os valores acima ou iguais a 0,5 são classificados como positivos enquanto os abaixo são classificados como negativos. Nos casos em que os modelos são aplicados a conjuntos de dados que demonstram desequilíbrios entre o número de casos positivos e negativos, como acontece neste estudo, é frequente recorrer a uma variação do threshold dentro de um intervalo significativo. Isto deve-se ao facto de os classificadores ficarem enviesados no sentido da classe predominante da amostra de treino fornecida [74]. Como tal, vão ser representados os valores das métricas referidas acima para diferentes thresholds contidos num intervalo de **0,2 a 0,8**.

Por recomendação da literatura [5, 35], os modelos Logit vão ser também avaliados consoante uma métrica conhecida como pseudo- R^2 de McFadden, $R^2_{McFadden}$ [75]. Esta métrica é considerada por ser adequada a modelos estimados por metodologias de verosimilhança, como acontece na estimação dos coeficientes do modelo de regressão logística. Esta pode ser descrita por:

$$R^2_{McFadden} = 1 - \frac{\log(\mathcal{L}_{modelo})}{\log(\mathcal{L}_{nulo})} \quad (4.16)$$

em que \mathcal{L}_{modelo} representa a verosimilhança maximizada do modelo ajustado aos dados de treino, enquanto \mathcal{L}_{nulo} representa a verosimilhança maximizada do modelo ajustado sem variáveis de entrada [35]. Os valores de pseudo- R^2 servem de analogia ao R^2 tradicional nos casos em que este não pode ser aplicado como uma medida do quão bem um modelo estatístico se adequa ao conjunto de observações [76]. Segundo o próprio McFadden em [75], pode se considerar um excelente ajuste do modelo aos dados quando o pseudo- R^2 proposto, $R^2_{McFadden}$, toma um valor entre 0,2 e 0,4.

4.6 Técnicas de Explainable Machine Learning

Para extrair conhecimento sobre a separação de mercados no mercado ibérico de eletricidade utilizando métodos de ML, é necessário recorrer a técnicas XML de modo a possibilitar a interpretação e explicação dos modelos explicativos desenvolvidos. Um dos métodos preferidos em XML é o Partial Dependence Plot (PDP), que pode representar a dependência marginal que um conjunto de entradas tem na saída do modelo [77]. Trata-se de um método de visualização em que é representada uma função de dependência parcial, de uma ou duas entradas no máximo devido à sua limitação a uma visualização de baixa dimensão. Para um modelo de ML de caixa negra, a função de dependência parcial pode ser estimada por:

$$\bar{f}_s(X_s) = \frac{1}{N} \sum_{i=1}^N f(X_s, X_c^{(i)}) \quad (4.17)$$

Esta expressão introduz os vetores de características X_s e $X_c^{(i)}$, sendo o primeiro as características de interesse, enquanto o segundo representa a instância i das restantes características dos dados de treino. Uma vez que N é o número de instâncias, \bar{f}_s calcula a média marginal da função f , a saída do modelo, que pode servir como uma descrição do efeito que o subconjunto de características X_s escolhido tem na resposta do modelo. Se X_s tiver mais do que uma variável, é possível observar a interação entre ambas as variáveis e o resultado, utilizando um gráfico tridimensional (gráfico 3D). É de salientar o pressuposto do método de que não existe correlação entre as características para as quais é calculada a dependência parcial e as restantes características, o que é extremamente raro quando se consideram variáveis de produção de eletricidade, consumo e ATC. Este facto pode levar à inclusão de pontos de dados muito pouco comuns, ou mesmo impossíveis, no PDP. É também importante interpretar o PDP resultante como o efeito da variável X_s depois de

contabilizar o efeito médio das variáveis X_c na saída do modelo [78, 79].

Outro tipo de técnica utilizada no XML é a **importância das características** (feature importance), ou análise da importância das variáveis. Consiste na atribuição de uma pontuação a cada variável de entrada relativamente à sua contribuição para as previsões do modelo. A importância das características é muitas vezes utilizada na fase de desenvolvimento de um modelo, uma vez que pode ser útil na seleção de variáveis, bem como na redução da dimensionalidade. Uma medida simples de importância de variáveis baseada nos PDPs é apresentada em [80]. A importância da variável é dada através de uma medida do quão "plana" é a função de dependência parcial para uma determinada variável, algo que é facilmente obtido através do cálculo do desvio padrão da amostra:

$$I(X_s) = \sqrt{\frac{1}{k-1} \sum_{i=1}^k \left[\bar{f}_s(X_s^{(k)}) - \frac{1}{k} \sum_{i=1}^k \bar{f}_s(X_s^{(k)}) \right]^2} \quad (4.18)$$

em que k representa o número de valores únicos para a característica X_s . Note-se que, uma vez que se trata de uma medida baseada no próprio modelo, a sua afinação e treino adequado poderão afetar os resultados obtidos.

Tanto os PDPs como a técnica de importância de características apresentada são considerados **explicabilidade global**, o que significa que o conhecimento deles extraído relaciona-se com todas as instâncias de dados e com o funcionamento global do modelo. Por outro lado, a **explicabilidade local** oferece uma visão sobre a razão pela qual um modelo prevê de uma determinada forma para uma instância escolhida, acrescentando potencialmente um grau de confiança à previsão, desde que a explicação seja sensata para o seu observador. As explicações globais podem ser vulneráveis à chamada **troca fidelidade-interpretabilidade**, conceito abordado no Capítulo 2.6, o que significa que quanto mais interpretável é um modelo explicativo, menos fiel é à complexidade do modelo original de caixa-negra [81]. Este problema pode ser atenuado através da explicação local, que pode atingir um elevado grau de interpretabilidade, mantendo a fidelidade à complexidade do modelo de caixa-negra a um nível local: na proximidade da instância observada.

Dois métodos populares de interpretação local são os **valores de SHapley Additive exPlanations (SHAP)** [82], que utilizam o conceito de valores de Shapley [83] para medir a importância das características, e as **Local Interpretable Model-Agnostic Explanations (LIME)** [84]. Ambos os métodos são agnósticos em relação ao modelo, o que significa que, à semelhança dos PDPs e da medida de importância apresentada, podem ser aplicados a qualquer classificador de ML sem qualquer conhecimento prévio do seu funcionamento interno, o que os torna desejáveis na análise de modelos e arquiteturas menos comuns, bem como para fazer comparações com as mesmas métricas. O LIME é uma técnica de explicação que envolve a criação de um modelo substituto que aproxima, localmente, as previsões dos modelos de caixa-negra. Isto é feito através da geração de um novo conjunto

de dados baseado na perturbação de uma dada instância, cuja previsão se pretende explicar, e depois treinando um modelo interpretável nos dados gerados e na respetiva saída do modelo de caixa-negra [78, 84].

Neste estudo, os **PDPs** serão aplicados como método de visualização dos efeitos das variáveis exógenas nas saídas dos vários modelos. A importância das características baseada em **PDP** e o **LIME** serão utilizados na análise do modelo **RNA** explicativo da ocorrência de separação de mercados no **MIBEL**. As **LIME** serão implementadas usando o módulo Python: **lime** que usa modelos de regressão linear ponderada como substituto local do modelo original, regressão Ridge em específico, que é totalmente interpretável através dos valores dos seus pesos estimados.

RESULTADOS

5.1 Avaliação dos modelos

São agora apresentados os valores calculados para as métricas **exatidão**, **especificidade**, **precisão**, **revocação** e **f1-score**, considerando uma variação do threshold de 0,2 a 0,8, para todos os modelos desenvolvidos. Dado o desequilíbrio severo entre o número de ocorrências de separação de mercado e de convergência de preços, dá-se uma atenção particular ao threshold para o qual o f1-score é máximo, pois este representa um equilíbrio entre a precisão e revocação, métricas estas que estão associadas à classe positiva, ocorrência de separação de mercado, e variam de forma oposta com o aumento do threshold.

5.1.1 Regressão Logística

As Tabelas 5.1 e 5.2 contêm os valores calculados para as métricas de avaliação de ambos os modelos de regressão logística. Dado considerarem dois cenários opostos, separação de mercado no sentido Espanha-Portugal e Portugal-Espanha, que têm um número de observações de classe positiva diferentes, apresentam algumas diferenças. Para o primeiro modelo (Espanha-Portugal), Tabela 5.1, verifica-se um f1-score máximo de **54,55%** para um threshold de **0,3**. A especificidade é consideravelmente elevada ao longo de toda a variação do threshold, visto que representa a classe maioritária (negativa). Quanto à precisão e revocação, estas são baixas na perspetiva de um conjunto de dados equilibrado, o que não é o caso para os dados utilizados, visto que o número de horas em separação de mercado com Portugal como importador, utilizado no primeiro modelo, representa apenas **3,1%** da amostra de treino considerada, sendo que este aspeto piora para o número de horas de separação de mercado com Espanha como importadora, que representa apenas **1,7%**. Podemos então considerar estes resultados como razoáveis para o contexto observado.

Threshold	exatidão (%)	especificidade (%)	precisão (%)	revocação (%)	f1-score (%)
0,2	96,24	97,08	43,12	69,63	53,26
0,25	96,71	97,81	47,46	62,22	53,85
0,3	97,09	98,38	52,58	56,67	54,55
0,35	97,23	98,72	55,51	50,37	52,82
0,4	97,20	98,96	56,22	41,85	47,98
0,45	97,18	99,15	56,89	35,19	43,48
0,5	97,27	99,36	61,15	31,48	41,56
0,55	97,27	99,53	63,96	26,30	37,27
0,6	97,23	99,62	64,84	21,85	32,69
0,65	97,19	99,75	68,18	16,67	26,79
0,7	97,17	99,85	72,92	12,96	22,01
0,75	97,14	99,92	78,79	9,63	17,16
0,8	97,05	99,94	76,19	5,93	11,00

Tabela 5.1: Métricas de avaliação do modelo Logit para a separação de mercado no sentido Espanha-Portugal

Para o segundo modelo (Portugal-Espanha), Tabela 5.2, verifica-se um f1-score máximo de **50,76%** para um threshold de **0,35**. Comparativamente ao primeiro modelo, apresenta piores resultados nas métricas relacionadas com a classe positiva (precisão e revocação), diferença que poderá ser atribuída ao facto de se tratar de um cenário oposto que possui um desequilíbrio diferente nas classes relativamente ao modelo anterior.

Threshold	exatidão (%)	especificidade (%)	precisão (%)	revocação (%)	f1-score (%)
0,2	97,91	98,55	41,31	60,27	49,03
0,25	98,19	98,92	46,24	54,79	50,16
0,3	98,38	99,21	51,43	49,32	50,35
0,35	98,52	99,41	56,78	45,89	50,76
0,4	98,54	99,57	59,78	37,67	46,22
0,45	98,52	99,63	60,00	32,88	42,48
0,5	98,59	99,73	66,18	30,82	42,06
0,55	98,59	99,81	70,37	26,03	38,00
0,6	98,57	99,87	74,42	21,92	33,86
0,65	98,54	99,90	75,00	18,49	29,67
0,7	98,51	99,92	75,86	15,07	25,14
0,75	98,53	99,95	84,00	14,38	24,56
0,8	98,45	99,97	81,25	8,90	16,05

Tabela 5.2: Métricas de avaliação do modelo Logit para a separação de mercado no sentido Portugal-Espanha

Quanto aos pseudo- R^2 de McFadden, o primeiro modelo (Espanha-Portugal) obteve um valor de **0,524**, enquanto o segundo (Portugal-Espanha) teve **0,560**. Em ambos os casos o valor encontra-se fora do intervalo indicativo de um excelente ajuste (0,2 a 0,4). A omissão das variáveis referentes à produção em ciclo combinado de gás natural portuguesa, $Gás_CCPT-contribuição$, e produção termoelétrica a carvão espanhola, $CarvãoES-contribuição$, reduziram o valor do pseudo- R^2 para 0,389 (modelo 1) e 0,413 (modelo 2), no entanto

pioraram consideravelmente as restantes métricas.

5.1.2 Random Forest

A Tabela 5.3 faz o sumário da avaliação do modelo de Random Forest desenvolvido. Encontra-se o valor de F1-score máximo de **70,04%** para um threshold de **0,35**, o que resulta numa precisão de **74,06%** e revocação de **66,43%**. Estes valores são, em regra geral e sem ter em conta o desequilíbrio severo presente nos dados, bons. Ao contrário dos modelos logit, o Random Forest tem em conta a totalidade de ocorrências de separação de mercado e não uma discriminação consoante o sentido da separação, o que por uma parte significa um maior número de observações de classe positiva, mas por outra representa uma maior complexidade nas relações entre as variáveis e saída do modelo.

Threshold	exatidão (%)	especificidade (%)	precisão (%)	revocação (%)	f1-score (%)
0,2	95,71	96,49	53,27	80,10	63,98
0,25	96,71	97,83	63,14	74,34	68,28
0,3	97,08	98,35	68,42	71,70	70,02
0,35	97,30	98,84	74,06	66,43	70,04
0,4	97,33	99,04	76,68	63,07	69,21
0,45	97,17	99,33	80,07	53,96	64,47
0,5	97,11	99,54	84,17	48,44	61,49
0,55	96,93	99,70	87,37	41,49	56,26
0,6	96,68	99,74	87,06	35,49	50,43
0,65	96,45	99,82	88,97	29,02	43,76
0,7	96,31	99,88	91,23	24,94	39,17
0,75	96,03	99,94	93,67	17,75	29,84
0,8	95,82	99,95	93,22	13,19	23,11

Tabela 5.3: Métricas de avaliação do modelo Random Forest

5.1.3 Rede Neuronal Artificial

Por fim temos na Tabela 5.4 os resultados referentes ao modelo de rede neuronal desenvolvido. É possível encontrar o F1-score máximo de **60,68%** para um threshold de **0,7**. Este valor, sendo inferior ao registado para a Random Forest, contraria a ordenação hipotética imposta inicialmente, que colocava a rede neuronal numa posição superior de desempenho mas inferior na capacidade explicativa. Mais uma vez, estes resultados são vistos como razoáveis dentro do contexto considerado.

Threshold	exatidão (%)	especificidade (%)	precisão (%)	revocação (%)	f1-score (%)
0,2	90,15	89,93	31,96	94,72	47,79
0,25	90,95	90,82	33,74	93,53	49,59
0,3	91,71	91,67	35,65	92,33	51,44
0,35	92,21	92,25	37,06	91,37	52,73
0,4	92,76	92,90	38,74	89,93	54,15
0,45	93,22	93,50	40,26	87,77	55,20
0,5	93,70	94,06	42,12	86,57	56,67
0,55	94,12	94,67	43,81	83,21	57,40
0,6	94,58	95,29	46,02	80,34	58,52
0,65	94,98	95,87	48,28	77,22	59,41
0,7	95,40	96,44	51,15	74,58	60,68
0,75	95,54	96,96	52,43	67,15	58,89
0,8	95,93	97,66	56,67	61,15	58,82

Tabela 5.4: Métricas de avaliação do modelo de Rede Neuronal

5.2 Coeficientes estimados e importâncias

Neste Capítulo são apresentados os valores dos coeficientes estimados, no caso dos modelos de regressão logística, e medidas de importância calculadas. Isto permitirá determinar quais as variáveis com maior impacto nas previsões, e por sua vez, quais os fatores determinantes da separação de mercado, segundo os modelos.

5.2.1 Regressão Logística

No caso dos modelos de regressão logística, a importância de uma dada variável é dada pela inspeção direta dos coeficientes estimados, visto que estes representam a alteração nas log-odds quando a variável tem um valor unitário. Os coeficientes estimados para ambos os modelos encontram-se representados nas Tabelas 5.5 e 5.6, ordenados de ordem decrescente do seu valor absoluto. Note-se que todas as variáveis restantes do processo descrito no Capítulo 4.2.2, para ambos os modelos, são significantes ao nível de 1%, ou seja, os valores-p associados são menores que 0,01.

Variáveis	Coefficientes
<i>Constante</i> (β_0)	8,4782
ATC_{ES-PT}	-2,985
$Eólica_{PT-contribuição}$	-2,599
$Gás_CC_{PT-contribuição}$	-2,475
$Hídrica_{PT-contribuição}$	-1,961
$Consumo_{PT}$	1,222
$Carvão_{ES-contribuição}$	-0,9131
$Eólica_{ES-contribuição}$	0,829
$Consumo_{ES}$	-0,703
$Hídrica_{ES-contribuição}$	-0,580
$Solar_{ES-contribuição}$	0,302

Tabela 5.5: Coeficientes estimados para o modelo Logit da separação de mercado no sentido Espanha-Portugal

Por inspeção da Tabela 5.5, referente aos coeficientes estimados do modelo Logit para a ocorrência de separação de mercado no sentido Espanha-Portugal, podemos verificar que, como era expectável, a capacidade de transferência disponível, ATC_{ES-PT} , tem a maior contribuição para a saída do modelo com um coeficiente de **-2,985**. A interpretação do sinal negativo dos coeficientes é a de que o aumento do valor das variáveis associadas diminui a probabilidade de ocorrência de separação de mercado, e vice-versa. Com exceção das variáveis $Consumo_{PT}$, $Eólica_{ES-contribuição}$ e $Solar_{ES-contribuição}$, todas as variáveis têm um coeficiente de sinal negativo. É possível notar que os coeficientes associados às variáveis de produção portuguesas têm uma maior relevância para o modelo do que as homólogas espanholas, em especial a eólica portuguesa.

Um aspeto interessante a notar é o facto de os sinais referentes aos coeficientes associadas à produção de eletricidade com carvão e hídrica em Espanha são negativos. A produção termoelétrica a carvão encontra-se em posições mais altas da ordem de mérito económico (elevados custos marginais) e é por isso maioritariamente utilizada em situações de pico de carga, o que significa que, quando a zona de mercado adjacente tem disponíveis mais ofertas de menor custo marginal, estabelece-se um fluxo de eletricidade no sentido do mercado que necessita das tecnologias de alto custo marginal para satisfazer a sua procura local, aumentando assim a probabilidade de separação de mercado.

Variáveis	Coefficientes
$Constante(\beta_0)$	-10,124
$Eólica_{PT-contribuição}$	3,108
$Gás_CC_{PT-contribuição}$	2,549
$Hídrica_{PT-contribuição}$	2,14
ATC_{PT-ES}	-1,676
$Consumo_{ES}$	1,342
$Carvão_{ES-contribuição}$	1,024
$Hídrica_{ES-contribuição}$	0,961
$Consumo_{PT}$	-0,483
$Eólica_{ES-contribuição}$	-0,462
$Solar_{ES-contribuição}$	-0,353

Tabela 5.6: Coeficientes estimados para o modelo Logit da separação de mercado no sentido Portugal-Espanha

É possível verificar algumas diferenças significativas entre os coeficientes estimados para os dois cenários. Na Tabela 5.6 podemos verificar os coeficientes referentes ao modelo da separação de mercado no sentido Portugal-Espanha. Podemos averiguar da mesma que o coeficiente de maior valor absoluto é o que está associado à produção eólica portuguesa, $Eólica_{PT-contribuição}$, com **3,108**. Os sinais dos coeficientes estão invertidos quando comparados aos do cenário oposto, com a exceção do associado à variável ATC_{PT-ES} visto que esta substituiu a equivalente no sentido contrário. Esta diferença suporta a decisão de separar a modelação da ocorrência de separação de mercado para os dois cenários opostos, visto que o modelo de regressão logística seria incapaz de mapear uma relação desta complexidade caso se considerasse uma vista global das ocorrências.

Todos os coeficientes, exceto os associados às variáveis $Consumo_{PT}$, $Eólica_{ES-contribuição}$, $Solar_{ES-contribuição}$ e ATC_{PT-ES} (por motivos já explicados) têm coeficientes de sinal positivo. Mais uma vez, as variáveis associadas à produção de eletricidade portuguesas têm coeficientes de maior valor absoluto comparativamente aos da produção espanhola. Deve-se também notar que os posicionamentos na ordenação por ordem decrescente dos coeficientes referentes aos consumos nacionais trocaram entre si. Isto significa que para o primeiro cenário (OSM_{ES-PT}) o consumo português tinha uma maior contribuição para a saída do modelo do que o consumo espanhol, enquanto que para o segundo cenário OSM_{PT-ES} acontece o inverso. É importante referir que estas duas variáveis têm sempre coeficientes com sinais opostos, e que o coeficiente associado ao consumo da zona de mercado importadora é positivo para ambos os cenários.

5.2.2 Random Forest

A “importância de Gini”, descrita no Capítulo 4.3.1, é aqui utilizada como um indicador geral da relevância das variáveis de estudo para o modelo. A Figura 5.1 contém as medidas de diminuição média na impuridade, ou impuridade de Gini, para cada uma das variáveis

consideradas. Destacam-se, por ordem decrescente de importância para o modelo, as variáveis de produção eólica portuguesa, produção hídrica espanhola, capacidade de interligação no sentido Espanha-Portugal, produção com ciclo combinado de gás natural portuguesa e capacidade de interligação no sentido Portugal-Espanha.

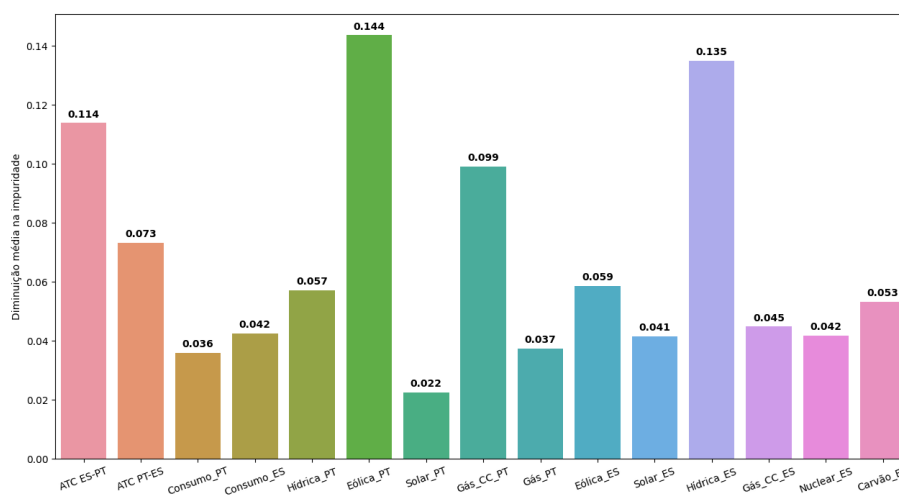


Figura 5.1: Diminuição média na impureza por variável do modelo de Random Forest

5.2.3 Rede Neuronal Artificial

Visto que este modelo não tem uma medida de importância de variáveis própria, recorre-se à medida de importância baseada nos PDPs que é descrita no Capítulo 4.6 e pela equação 4.18. A Tabela 5.7 contém as importâncias calculadas segundo este método, ordenadas de forma decrescente. Distinguem-se as variáveis de capacidade de interligação no sentido Espanha-Portugal, produção hídrica portuguesa, capacidade de interligação no sentido Portugal-Espanha, consumo português, produção eólica portuguesa, produção com gás em ciclo combinado portuguesa e produção eólica espanhola.

Variáveis	Importância baseada nos PDPs
ATC_{ES-PT}	0,1764
$Hídrica_{PT-contribuição}$	0,1067
ATC_{PT-ES}	0,1065
$Consumo_{PT}$	0,1059
$Eólica_{PT-contribuição}$	0,1020
$Gás_CC_{PT-contribuição}$	0,0736
$Eólica_{ES-contribuição}$	0,0566
$Nuclear_{ES-contribuição}$	0,0156
$Gás_{PT-contribuição}$	0,0154
$Consumo_{ES}$	0,0143
$Hídrica_{ES-contribuição}$	0,0131
$Carvão_{ES-contribuição}$	0,0068
$Solar_{ES-contribuição}$	0,0066
$Solar_{PT-contribuição}$	0,0030

Tabela 5.7: Importâncias calculadas para o modelo de Rede Neuronal

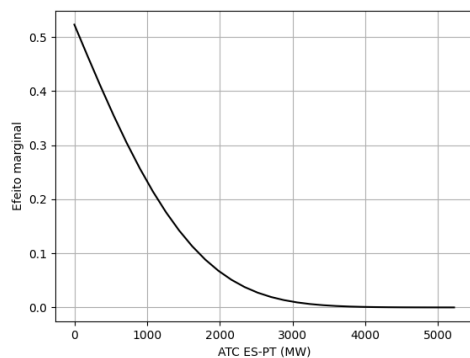
Tal como nos coeficientes estimados para os modelos Logit, verifica-se uma maior relevância das variáveis de produção portuguesas em comparação às espanholas, pelo que as importâncias calculadas para as produções hídrica e eólica portuguesas constituem as duas variáveis de produção com maior impacto no modelo. De notar também a elevada importância atribuída ao consumo português em comparação ao consumo espanhol. Isto poderá ser explicado pelo maior número de ocorrências de separação de mercado no sentido Espanha-Portugal do que no sentido contrário, dando assim uma maior importância ao consumo português.

É possível verificar que as duas variáveis de produção espanholas com as maiores importâncias são a produção eólica (0,0566) e nuclear (0,0156), ambas de baixo custo marginal, e poderão estar associadas às situações de separação de mercado que ocorrem quando a zona de mercado portuguesa tem uma baixa produção eólica e hídrica.

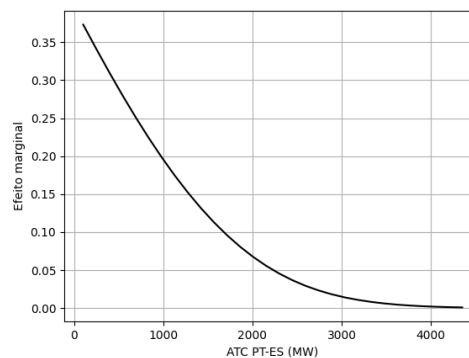
5.3 Partial Dependence Plots

São agora demonstrados os PDPs referentes ao impacto das variáveis exógenas nos modelos desenvolvidos. Para variáveis associadas a uma dada zona de mercado que tenham homólogas na outra opta-se por uma visualização 3D da função parcial de ambas as variáveis, caso contrário, utiliza-se um gráfico 2D na representação. Para certas variáveis, como as produções eólicas, é vantajoso utilizar PDPs tridimensionais visto que estas tecnologias encontram-se na base da ordem de mérito económico, o que permite visualizar o impacto dos desequilíbrios entre estas produções na probabilidade de ocorrência de separação de mercado. O mesmo se aplica às variáveis de consumo, em que as suas diferenças influenciam diretamente os fluxos de eletricidade entre as zonas de mercado.

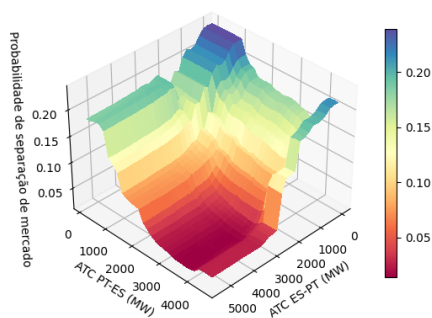
5.3.1 Capacidade de transferência disponível nas interligações



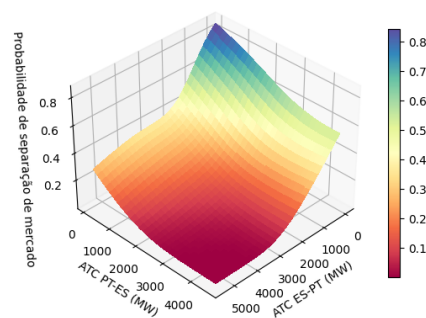
(a) Logit ES-PT



(b) Logit PT-ES



(c) Random Forest



(d) Rede Neuronal Artificial

Figura 5.2: Partial Dependence Plots da probabilidade de separação de mercado e da capacidade de transferência disponível nas interligações

Dado que os modelos de regressão logística apenas utilizaram uma das capacidades de transferência disponíveis, nomeadamente a que dizia respeito a cada cenário, os seus PDPs são de uma só variável. Estes, Figura 5.2a e Figura 5.2b, têm uma forma semelhante e mostram que a probabilidade de separação de mercado diminui com o aumento da capacidade, e vice-versa. Um comportamento semelhante pode ser verificado nos PDPs referentes à Random Forest, Figura 5.2c, e Rede Neuronal, Figura 5.2d, que têm formas similares e acrescentam algum detalhe quanto à suposta interação entre estas duas variáveis. Encontra-se uma probabilidade de separação de mercado máxima quando ambas as capacidades de interligação são mínimas. Isto pode ser explicado pelo facto de que as interligações ficam vulneráveis a congestionamentos de ambos os sentidos, o que aumenta a probabilidade. Como era esperado, a probabilidade é mínima para quando ambas as capacidades são máximas.

Note-se que as funções parciais foram calculadas segundo um intervalo dos valores reais que as variáveis já tomaram, o que poderá não ser representativo da realidade visto

que os extremos deste intervalo são por vezes casos únicos. No caso dos PDPs de duas variáveis existem pontos que são impossíveis ou muito pouco prováveis, pelo que deve-se ter algum cuidado na análise destes gráficos.

5.3.2 Consumo

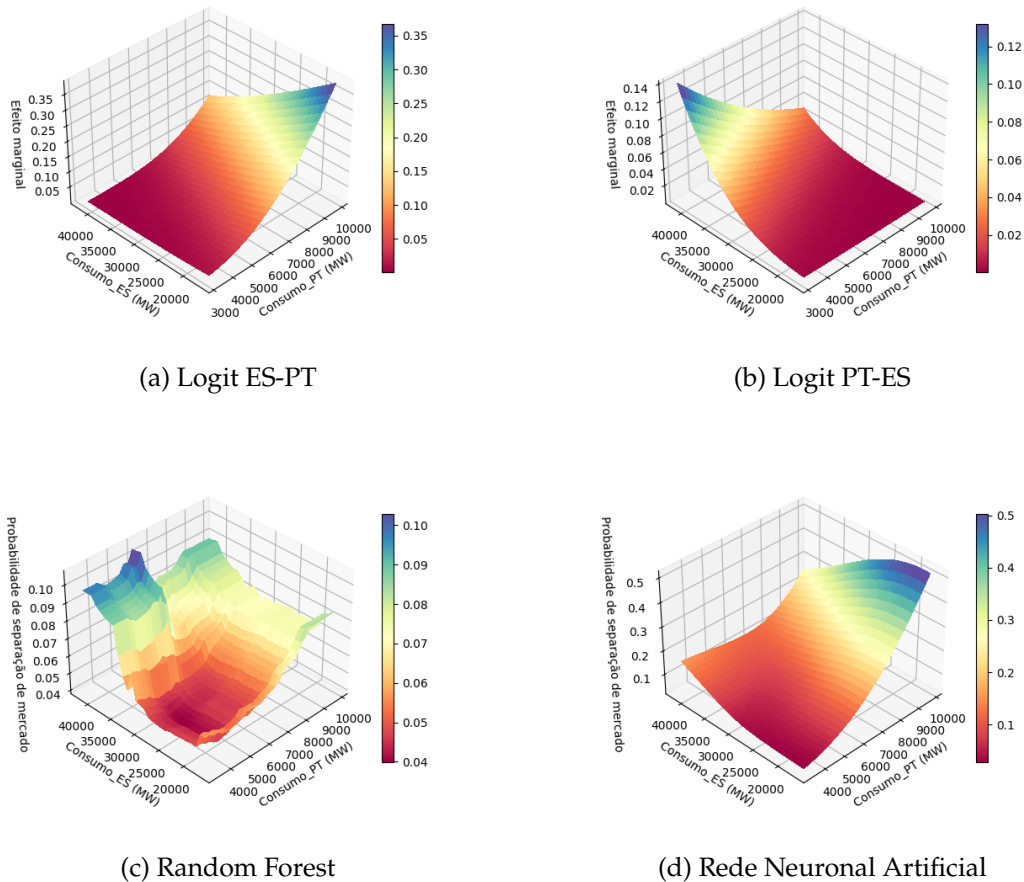


Figura 5.3: Partial Dependence Plots da probabilidade de separação de mercado e dos consumos nacionais

Quanto aos consumos verifica-se que, para ambos os modelos de regressão logística, Figura 5.3a e Figura 5.3b, enquanto o aumento de um dos consumos aumenta a probabilidade de separação de mercado o aumento do outro diminui a mesma. No modelo referente à separação de mercado no sentido Espanha-Portugal o aumento do consumo português aumenta a probabilidade enquanto o espanhol diminui. Há uma inversão do efeito destas variáveis quando se considera o modelo que retrata a separação de mercado no sentido Portugal-Espanha. Como já foi referido, os consumos nacionais podem ser vistos como a procura do mercado, que quando aumenta numa dada zona de mercado particular, relativamente às zonas adjacentes, vai causar um fluxo de eletricidade nesse sentido, aumentado por sua vez a probabilidade de separação de mercado.

O PDP referente à Rede Neuronal, Figura 5.3d, segue um comportamento compatível

com os demonstrados pelos modelos de regressão logística, com especial ênfase no consumo português. Este atinge um máximo global para um consumo português máximo e consumo espanhol mínimo, e um máximo local no caso oposto. Algo semelhante acontece para o PDP da Random Forest, Figura 5.3c, no entanto o consumo espanhol é mais valorizado neste.

5.3.3 Produção eólica

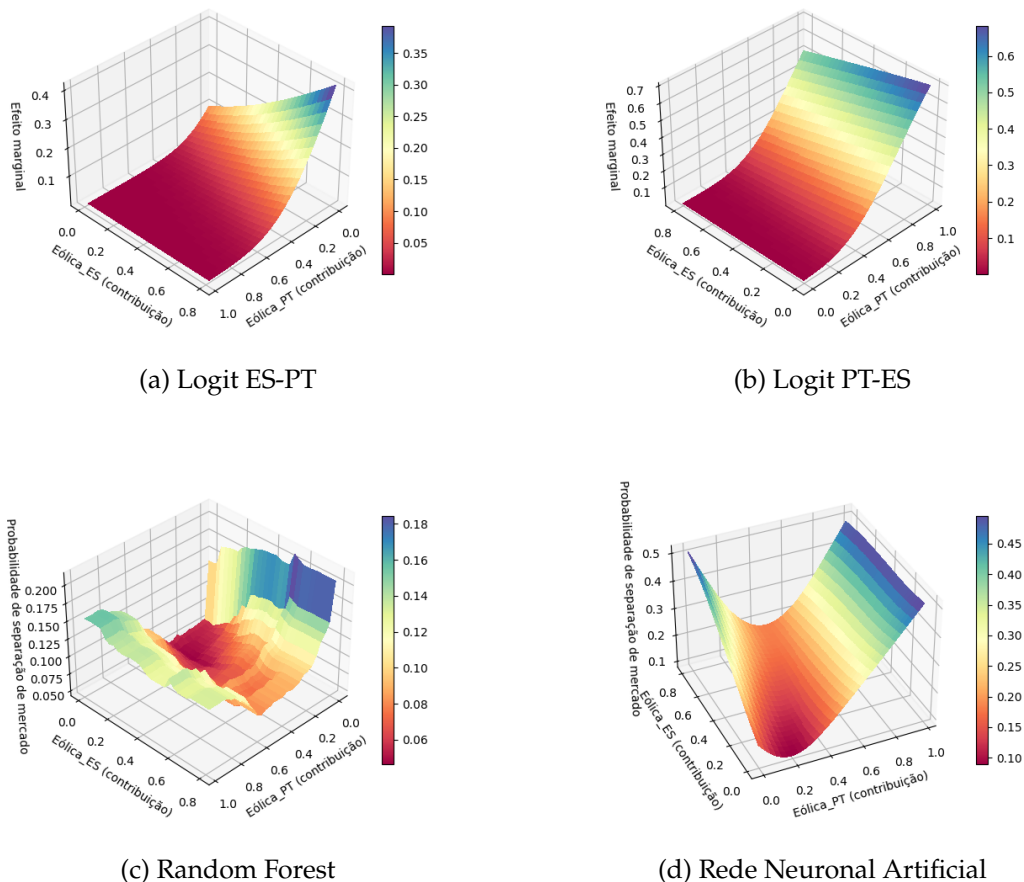


Figura 5.4: Partial Dependence Plots da probabilidade de separação de mercado e das produções eólicas

Para a produção eólica existe uma inversão no efeito das variáveis quando comparadas as respostas dos dois modelos de regressão logística. No primeiro cenário (ES-PT), Figura 5.4a, a probabilidade de separação de mercado atinge um máximo quando a contribuição da produção eólica portuguesa para o consumo nacional é mínima e a homóloga espanhola é máxima. Isto representa um cenário em que a zona de mercado espanhola tem uma grande quantidade de ofertas de eletricidade de baixo custo marginal, relativamente ao seu consumo, associadas à produção eólica, enquanto que a zona de mercado portuguesa não tem acesso ao mesmo tipo de ofertas devido a uma baixa produção eólica, provocando um fluxo no sentido Espanha-Portugal, que por sua vez aumenta a probabilidade de

separação de mercado. O inverso acontece para o segundo cenário (PT-ES), Figura 5.4b em que o máximo é atingido para uma elevada contribuição da produção eólica para o consumo nacional enquanto a homóloga espanhola é mínima. A explicação atribuída ao primeiro modelo também pode ser aqui aplicado, invertendo apenas o sentido da separação de mercado.

A junção dos PDPs dos modelos de regressão logística coincide com o comportamento observado nos PDPs referentes à Random Forest, Figura 5.4c, e Rede Neuronal, Figura 5.4d. Verifica-se um aspeto interessante na forma de “vale” nestes PDPs. Para valores semelhantes de contribuição da produção eólica para os respetivos consumos nacionais, a probabilidade de separação de mercado é mínima. Este comportamento é fácil de visualizar no gráfico associado à Rede Neuronal, e poderá ser um indicador de que o equilíbrio entre as produções eólicas de ambos os países resultam numa diminuição dos fluxos de eletricidade nas interligações, considerando os efeitos médios das restantes variáveis.

5.3.4 Produção hídrica

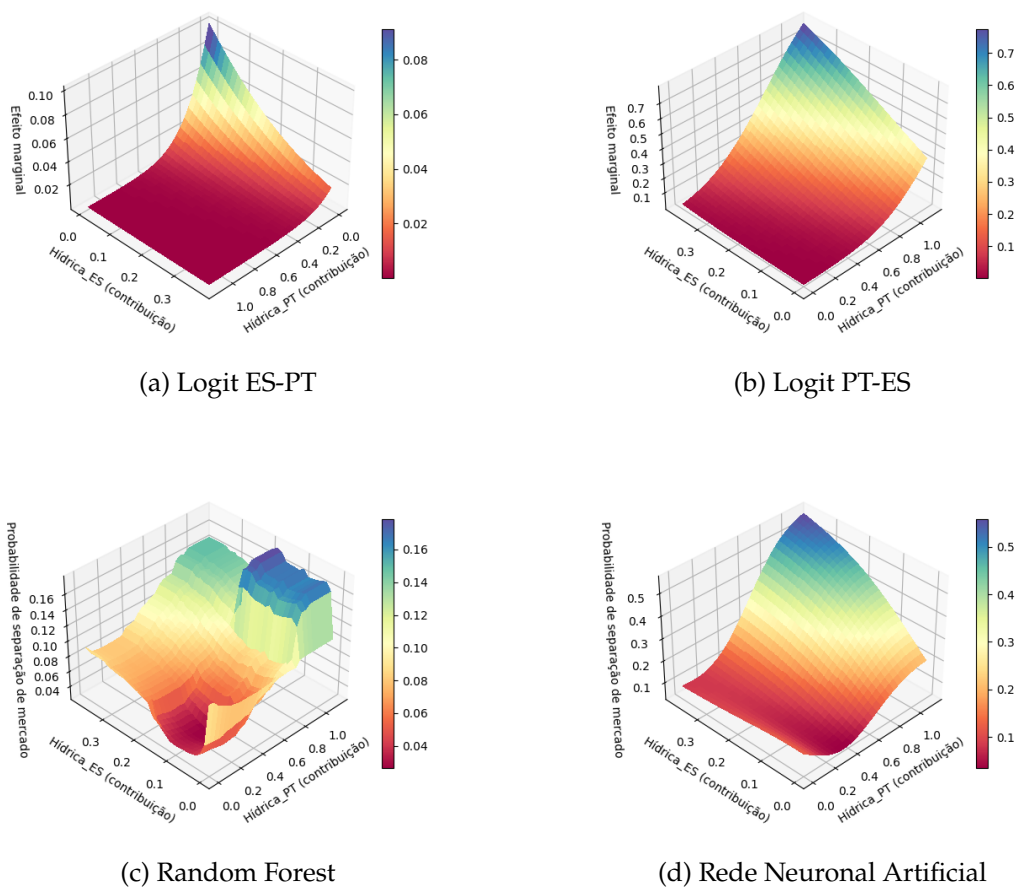
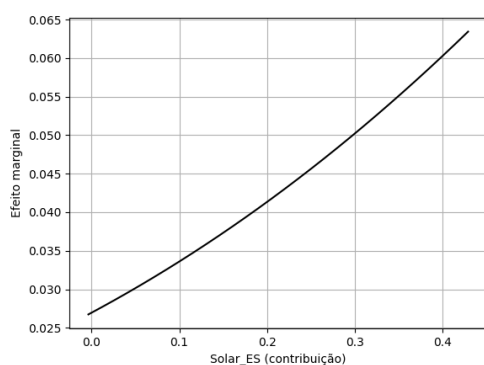


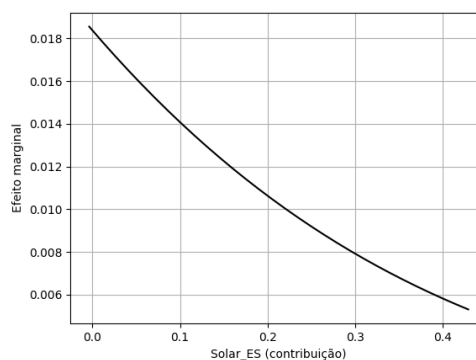
Figura 5.5: Partial Dependence Plot da probabilidade de separação de mercado es das produções hídricas

Por inspeção dos PDPs dos modelos de regressão logística, Figura 5.5a e Figura 5.5b, verifica-se que para produções hídricas portuguesa e espanhola mínimas, a probabilidade de separação de mercado no sentido Espanha-Portugal é máxima, enquanto que para produções máximas a probabilidade de separação de mercado no sentido inverso é máxima. Tal como foi possível verificar nos coeficientes estimados destes modelos, ambas as produções hídricas têm o mesmo sinal. O PDP associado à Rede Neuronal, Figura 5.5d, toma uma forma semelhante à do modelo de regressão logística para o sentido Portugal-Espanha, sem se verificar um aumento de probabilidade significativo para os valores de produção mínimos, como acontece para o modelo de regressão logística para o sentido Espanha-Portugal. Já o PDP do modelo de Random Forest, Figura 5.5c, toma uma forma consideravelmente diferente da dos restantes PDPs. Este apresenta uma zona de probabilidade máxima para elevados valores de produção hídrica portuguesa e baixos valores de produção hídrica espanhola.

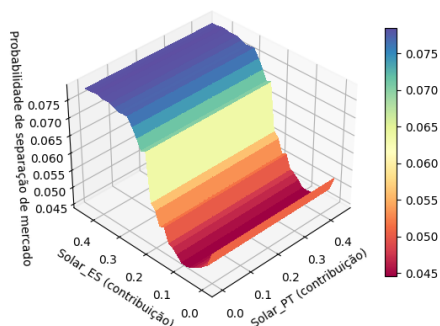
5.3.5 Produção solar



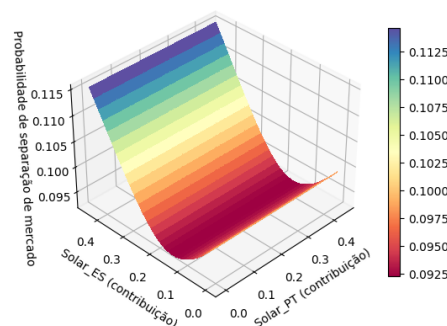
(a) Logit ES-PT



(b) Logit PT-ES



(c) Random Forest



(d) Rede Neuronal Artificial

Figura 5.6: Partial Dependence Plots da probabilidade de separação de mercado e das produções solares

Dado que a produção solar portuguesa foi excluída no processo de construção dos

modelos de regressão logística (por ter um valor-p acima de 0,01), são feitos apenas os PDPs da contribuição para o consumo da produção solar espanhola. Estes, Figura 5.6a e Figura 5.6b, demonstram um comportamento coerente com as expectativas, em que o aumento da produção solar espanhola contribui para uma separação de mercado no sentido Espanha-Portugal, embora sem grande influência na probabilidade. Acontece o contrário para o modelo logit referente à separação de mercado no sentido Portugal-Espanha, em que a probabilidade diminui com o aumento da produção solar.

Para os modelos de Random Forest, Figura 5.6c, e Rede Neuronal, Figura 5.6d, foi possível incluir a produção solar portuguesa nos PDPs visto que esta foi utilizada nestes. O comportamento destes PDPs é semelhante, e são coerentes com os gráficos referentes aos modelos de regressão logística. A produção solar portuguesa tem um efeito desprezável pelo que não influencia a probabilidade de ocorrência de separação de mercado nestes modelos.

5.3.6 Produção nuclear

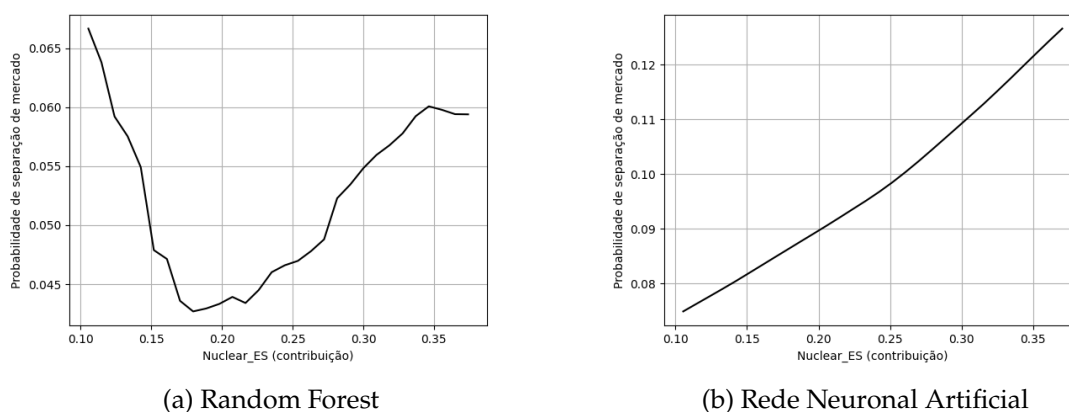
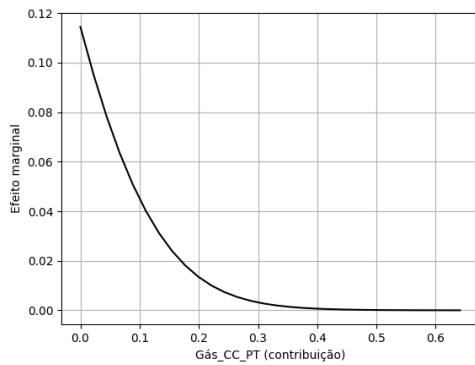


Figura 5.7: Partial Dependence Plots da probabilidade de separação de mercado e produção nuclear espanhola

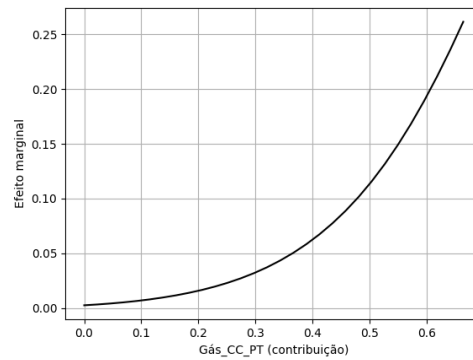
Como a produção nuclear só existe em Espanha, e devido a ter sido excluída dos modelos de regressão logística, são apenas apresentados os PDPs referentes aos modelos de Random Forest, Figura 5.7a, e Rede Neuronal, Figura 5.7b. Estes apresentam formas diferentes e uma fraca interação com a probabilidade de separação de mercado, coincidente com as importâncias calculadas. Enquanto o PDP associado ao modelo de Rede Neuronal demonstra uma relação quase linear entre a produção nuclear e a probabilidade, o PDP associado ao modelo de Random Forest tem um comportamento em forma de parábola. A produção nuclear é de baixo custo marginal e situa-se acima apenas das produções eólica e solar na ordem de mérito económico, o que significa que poderá contribuir para possíveis desequilíbrios na disponibilidade de energia de baixo custo marginal entre as zonas de mercado, favorecendo

o lado espanhol que por sua vez aumenta a probabilidade de separação de mercado no sentido Espanha-Portugal.

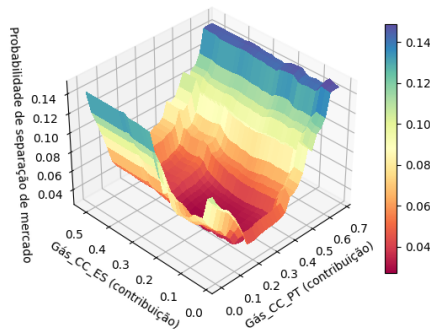
5.3.7 Produção com Ciclo Combinado de Gás Natural



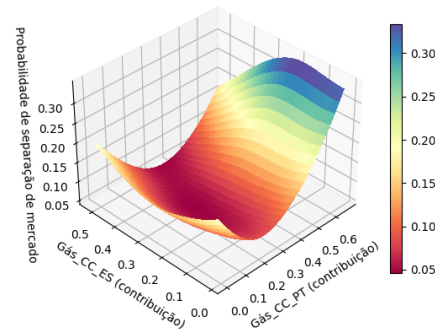
(a) Logit ES-PT



(b) Logit PT-ES



(c) Random Forest



(d) Rede Neuronal Artificial

Figura 5.8: Partial Dependence Plots da probabilidade de separação de mercado e produção em ciclo combinado de gás natural

Como a produção em ciclo combinado de gás natural espanhola foi excluída dos modelos Logit, por ter um valor-p menor que 0,1, os PDPs referentes a estes modelos dizem respeito a apenas a produção portuguesa. Por inspeção dos gráficos podemos observar que têm formas que são coerentes entre si. No caso dos modelos de regressão logística, Figura 5.8a e Figura 5.8b, verifica-se que o aumento da produção portuguesa diminui a probabilidade de separação de mercado, para o modelo que considera as ocorrências no sentido Espanha-Portugal, e aumenta para o modelo do sentido contrário.

Nos PDPs associados à Random Forest, Figura 5.8c, e Rede Neuronal, Figura 5.8d, podemos constatar que a produção espanhola não impacta a saída dos modelos de forma significativa. Verifica-se também a forma de "vale" consistente para os dois modelos e que é coincidente com a junção dos PDPs dos modelos Logit. Para valores mínimos e máximos de contribuição da produção para os respetivos consumos nacionais temos

máximos da probabilidade de separação de mercado, enquanto o mínimo surge quando, independentemente do valor da produção espanhola, a produção portuguesa está entre os 0,2 e 0,3 de contribuição para o consumo nacional.

5.3.8 Produção com Gás Natural

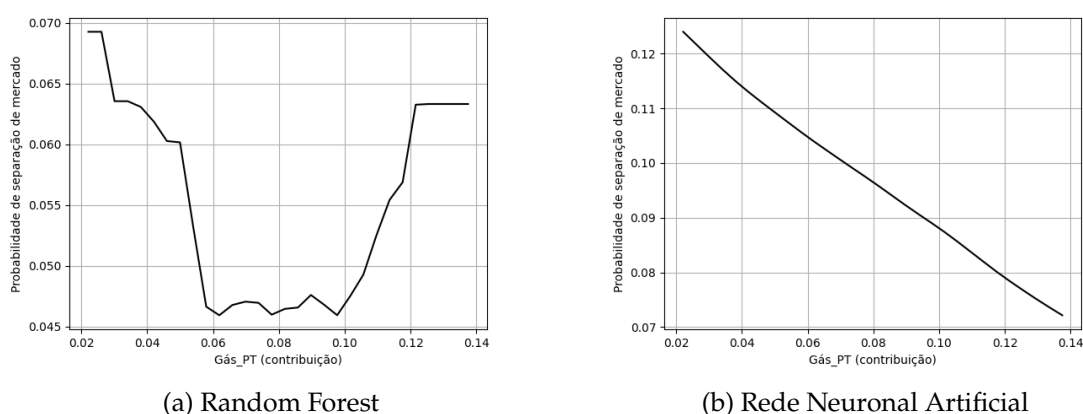


Figura 5.9: Partial Dependence Plots da probabilidade de separação de mercado e produção com gás natural

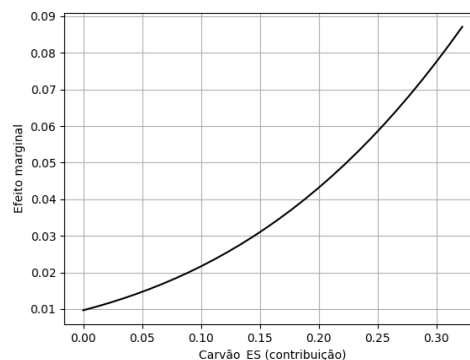
Tal como para a variável anterior, a variável de produção com Gás Natural portuguesa também foi excluída dos modelos Logit, pelo que apenas existem os PDPs referentes ao modelo de Random Forest, Figura 5.9a, e Rede Neuronal, Figura 5.9b. O PDP associado ao modelo de Rede Neuronal mostra que a probabilidade de separação de mercado varia linearmente com a contribuição desta variável, em concreto, diminui quando a produção aumenta. Já o PDP do modelo de Random Forest mostra novamente um comportamento de "vale", em que a probabilidade de separação de mercado é elevada quando a produção é mínima e máxima.

Dado que esta variável corresponde a uma tecnologia de elevado custo marginal, quanto menor for a sua contribuição para o consumo maior será a presença de produção de baixo custo marginal, resultando assim numa maior probabilidade de separação de mercado. Neste caso, quando a produção com Gás natural portuguesa é mínima espera-se que a zona de mercado portuguesa tenha uma elevada disponibilidade de ofertas de baixo custo marginal, aumentando assim a probabilidade de separação de mercado no sentido Portugal-Espanha. Quando a produção é máxima acontece o inverso, sendo esta lógica coerente com a forma do PDP do modelo de Random Forest.

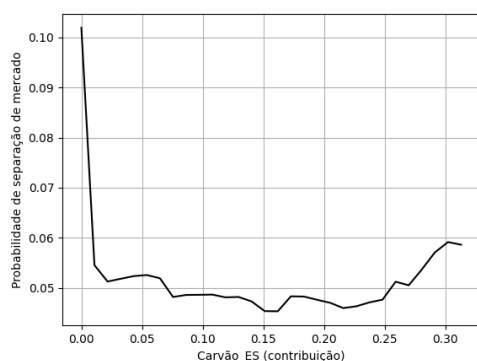
5.3.9 Produção com Carvão



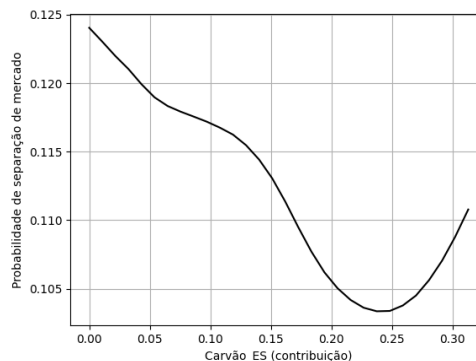
(a) Logit ES-PT



(b) Logit PT-ES



(c) Random Forest



(d) Rede Neuronal Artificial

Figura 5.10: Partial Dependence Plots da probabilidade de separação de mercado e produção com carvão

Visto que a produção a carvão portuguesa foi excluída de todos os modelos, por motivos que já foram descritos, os PDPs contêm exclusivamente a produção espanhola. No caso dos modelos Logit, Figura 5.10a e Figura 5.10b, verifica-se que o aumento da contribuição da produção com Carvão para o consumo espanhol reduz a probabilidade de separação de mercado no sentido Espanha-Portugal, e aumento no sentido Portugal-Espanha. Para os modelos de Random Forest, Figura 5.10c, e Rede Neuronal, Figura 5.10d, verifica-se que a probabilidade de separação de mercado é máxima quando a produção é mínima.

É possível aplicar uma lógica idêntica à utilizada na interpretação dos PDPs referentes à produção com Gás Natural.

5.4 Aplicação do algoritmo LIME na Rede Neuronal

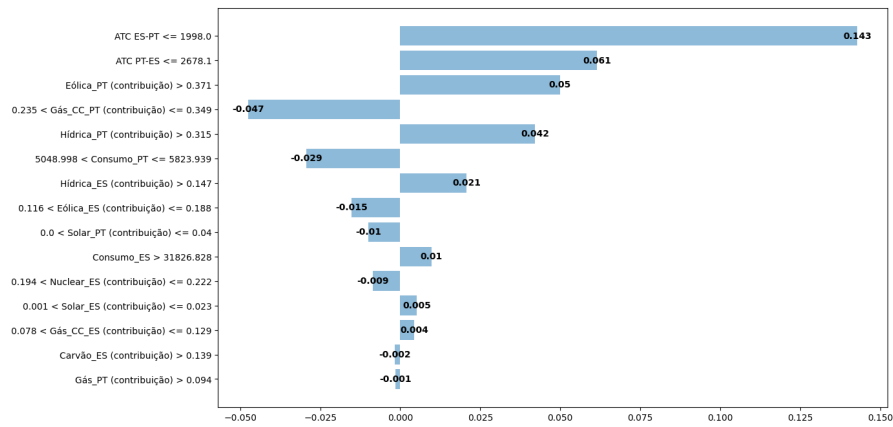
Neste capítulo foi aplicada a técnica de XML, LIME, na rede neuronal desenvolvida para dias exemplo associados aos padrões encontrados no Capítulo 3.3, por forma a compreender quais os aspetos valorizados pelo modelo na determinação de uma ocorrência de separação de mercado. As Figuras 5.11a, 5.11b e 5.11c contêm as explicações fornecidas pelo algoritmo em representações gráficas.

Na Figura 5.11a encontramos a explicação fornecida pelo algoritmo para o dia **20 de fevereiro de 2017 às 8h**, uma observação que corresponde a uma ocorrência de separação de mercado no sentido Portugal-Espanha e que pertence a um dos padrões identificados. Verifica-se que a variável com maior influência é a capacidade de interligação no sentido Espanha-Portugal, que na realidade é irrelevante para o caso em questão. Isto poderá ser uma das fraquezas de não modelar a separação de mercado de forma separada para os dois cenários opostos como foi feito para os modelos Logit, visto que as ocorrências no sentido Espanha-Portugal são mais comuns. No entanto, a segunda variável identificada pelo modelo é a capacidade de interligação no sentido Portugal-Espanha, que por ser menor que 2678,1 MW contribui positivamente para a separação de mercado. Em terceiro lugar, e de impacto positivo na separação de mercado por ser maior que 0,371, aparece a contribuição da produção eólica portuguesa para o consumo nacional. Tal como tinha sido hipotizado no Capítulo 3.3, uma maior disponibilidade de ofertas de baixo custo marginal relativamente à procura desta zona de mercado (hídrica e eólica portuguesas), e tendo em conta o desfasamento nos perfis de consumo dos dois países, causa para este cenário particular um fluxo no sentido Portugal-Espanha que esgota a capacidade de interligação, causando assim a separação de mercado.

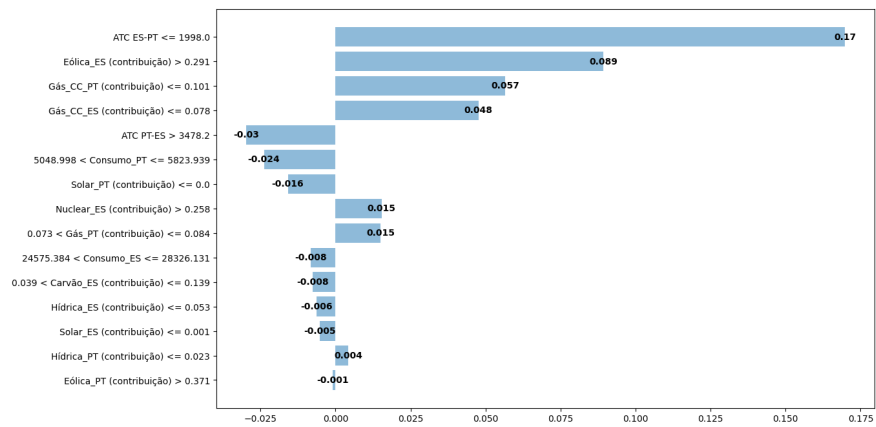
A segunda observação exemplo escolhida é **16 de janeiro de 2018 às 2h**, Figura 5.11b, que corresponde numa separação de mercado no sentido Espanha-Portugal e que está associada a um dos padrões de ocorrência identificados. Tal como no exemplo anterior, a capacidade de interligação no sentido Espanha-Portugal tem a maior relevância na explicação. Em segundo lugar surge a contribuição da produção eólica para o consumo espanhol, que sendo maior que 0,291 tem um impacto positivo na separação de mercado.

Por último considera-se a observação correspondente a **16 de julho de 2022 às 15h**, Figura 5.11c, que trata-se de uma ocorrência de separação de mercado no sentido Espanha-Portugal e que faz parte de um padrão com presença maioritária no ano de 2022. Na explicação, o consumo português é considerado como o maior fator contribuinte para a separação de mercado, que devido a ser superior a 6469,7 MW, valor elevado, tem um impacto positivo. Como a capacidade de interligação no sentido Espanha-Portugal é razoável, pois não se verifica nenhuma restrição significativa, temos que esta variável tem um impacto negativo. De notar que a solar espanhola é a única tecnologia da zona de mercado espanhola com um impacto positivo na separação de mercado, algo que é coerente com a hipótese considerada no Capítulo 3.3, tal como a produção hídrica em Portugal, devido a ter uma contribuição menor que 0,023.

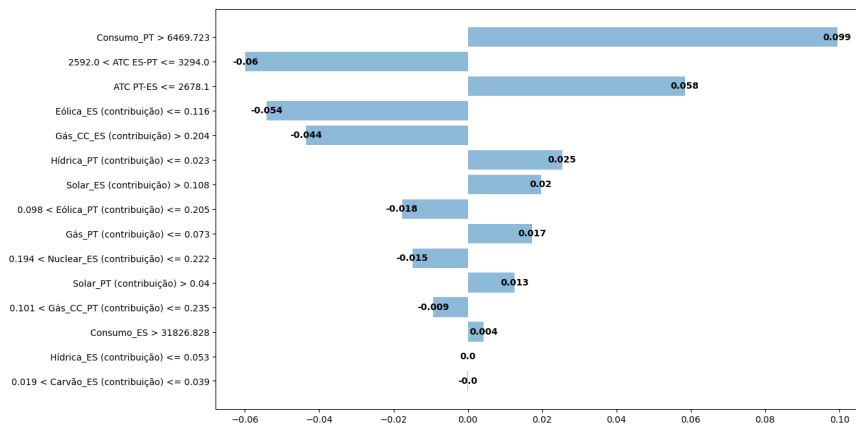
5.4. APLICAÇÃO DO ALGORITMO LIME NA REDE NEURONAL



(a) 20/02/2017 às 8h



(b) 16/01/2018 às 2h



(c) 16/07/2022 às 15h

Figura 5.11: Explicações LIME

5.5 Comparação entre modelos

Foram aplicados os dados referentes ao período de 2022 e início de 2023 nos modelos desenvolvidos e foram registadas as respetivas previsões. A Figura 5.12 mostra o resultado das previsões de cada modelo na forma de contagem de ocorrências de separação de mercado num mês rolante, e a Tabela 5.8 mostra a avaliação dos modelos para este período. O modelo Logit representado trata-se da aplicação de um OU lógico entre as previsões dos dois modelos desenvolvidos.

Deve-se notar que, tal como foi identificado no Capítulo 3.3, o período de 2022 e 2023 contem um padrão de ocorrência de separação de mercado diferente dos padrões observados nos anos anteriores, o que significa que os dados utilizados na construção dos modelos poderão não ser representativos desta nova amostra. Por inspeção da Figura 5.12, o modelo de Random Forest parece ser o que melhor segue os valores reais.

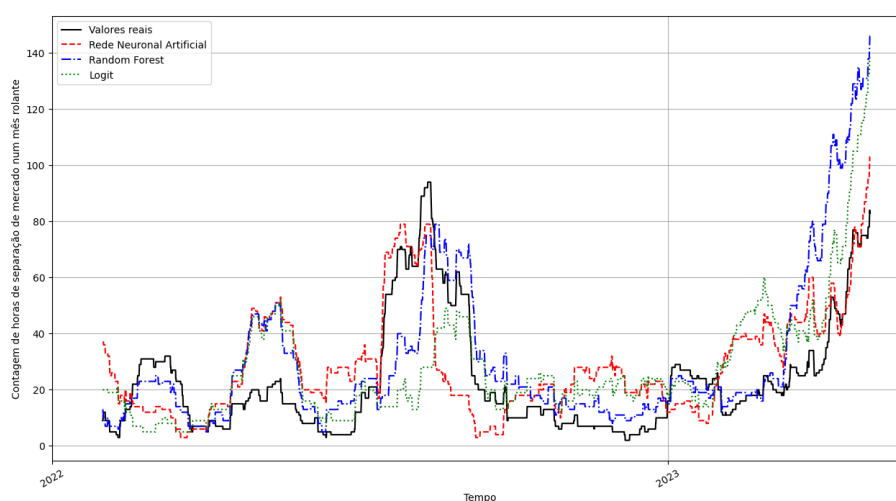


Figura 5.12: Resultado da previsão dos modelos para o período de 2022 e início de 2023 na forma de contagem de horas em separação de mercado num mês rolante

Modelo	exatidão (%)	especificidade (%)	precisão (%)	revocação (%)	f1-score
Logit ES-PT*	97,44	98,81	55,15	50,46	52,70
Logit PT-ES*	99,37	99,69	51,35	50,67	51,01
Random Forest	95,15	96,83	35,39	48,27	40,84
Rede Neuronal	95,22	96,96	35,71	47,03	40,60

*- Foram utilizados os dados de ocorrência de separação de mercado para o respetivo cenário

Tabela 5.8: Avaliação dos modelos para o período de 2022 e início de 2023

DISCUSSÃO E CONCLUSÕES

Neste capítulo vai ser feita uma discussão geral dos resultados obtidos, tal como uma revisão das possíveis soluções para a redução da separação de mercado. Também são feitas algumas considerações acerca de possíveis alterações ou novas abordagens para futuros estudos deste tema.

6.1 Discussão geral dos resultados

Ao longo deste trabalho foi feita uma análise multifacetada da ocorrência de separação de mercado no [MIBEL](#). Um primeiro passo consistiu na verificação da sua evolução no período considerado, onde se concluiu que a separação de mercado tem diminuído numa perspetiva anual, passando de representar 6,7% do tempo em 2017 para apenas 2,6% em 2021, com um aumento ligeiro para 2,9% em 2022 que poderá ser indicativo do início da reversão desta tendência decrescente.

Noutra perspetiva, feita através de uma contagem em mês rolante das ocorrências, verifica-se a presença de uma certa tendência sazonal para os anos de 2017 a 2020, esta que poderá estar associada à maior disponibilidade de energia de baixo custo marginal, nomeadamente a referente à produção eólica, e aos maiores perfis de consumo. Nos períodos seguintes não é possível confirmar esta tendência, havendo até a presença de "picos" a meio de 2022 e 2023.

Foram também identificados alguns padrões de ocorrência de separação de mercado e o sentido específico para os quais estes acontecem. Um primeiro padrão coincide com o período entre a meia-noite e 5 horas da manhã de novembro a abril, que acontece particularmente para o sentido de Espanha-Portugal, reconhecidos com recurso aos mapas de calor apresentados no [Capítulo 3.3](#). Justificou-se este padrão com a elevada disponibilidade de energia de baixo custo marginal a estas horas em relação aos baixos consumos nacionais.

Um outro padrão foi identificado para as 8 horas da manhã (UTC+1), presente ao longo de todo o ano mas especialmente intenso para o mês de abril, que regista 32 ocorrências para o período considerado (2017 a 2022). Este foi associado às diferenças entre os perfis de consumo, potencialmente causadas pelo desfazamento de 1 hora nos fusos horários

português e espanhol, em que é possível notar um maior consumo espanhol quando comparado ao português relativamente aos seus respectivos valores mínimos e máximos.

Foram então calculados os coeficientes de correlação de Pearson, r , Spearman, ρ , e Kendall, τ , das variáveis de estudo com os spreads de preços calculados. Esta diferença de preços, conhecida como spread e que foi utilizada na construção das variáveis binárias indicativas da ocorrência de separação de mercado, pode ser vista como uma medida de intensidade da separação de mercado. Isto quer dizer que, quanto maior o valor absoluto do spread, maior será o suposto desequilíbrio entre as porções de procura e oferta entre as zonas de mercado envolvidas.

Desta análise de correlação foi possível constatar quais as variáveis que influenciam os spreads de preços e em que sentido. Coeficientes de sinal negativo indicam que à medida que aumenta a variável em questão, diminui o spread. Para coeficientes de sinal positivo verifica-se o oposto, em que o aumento da variável leva ao aumento dos spreads. De se notar que spreads positivos são indicativos de uma separação de mercado no sentido Espanha-Portugal, enquanto que spreads negativos dizem respeito ao sentido contrário, tal como é definido pela equação 3.1. É então comum observar nos gráficos de dispersão associados às correlações identificadas, que a variação da variável provoca uma transição de spreads positivos para spreads negativos, ou vice-versa.

Um aspeto interessante a notar é o de que de entre as produções de eletricidade de uma dada zona de mercado verifica-se coeficientes com sinais negativos e positivos. No caso espanhol, as variáveis associadas à produção de eletricidade de baixo custo marginal, como a eólica, solar e nuclear, têm associados coeficientes positivos, ou seja, o seu aumento está correlacionado com o aumento do spread, que por sua vez está relacionado com a separação de mercado no sentido Espanha-Portugal. As tecnologias de elevado custo marginal, como as termoelétricas a carvão e ciclo combinado de gás natural, têm coeficientes de sinal negativo, sendo que as suas presenças no mix energético espanhol poderão indicar a necessidade acrescida de importar. No caso português todos os coeficientes relacionados com as variáveis de produção, de entre as consideradas relevantes na análise de correlação (ter pelo menos um coeficiente $> 0,3$), têm um sinal negativo.

De entre os coeficientes de correlação calculados destacam-se, para os coeficientes de sinal negativo, os das produções hídricas de ambos os países, consumo espanhol e produções de eletricidade com gás natural portuguesas (contando com o ciclo combinado), enquanto que com sinal positivo apenas a produção nuclear espanhola teve valores razoáveis para as três medidas de correlação aplicadas.

Uma segunda parte deste estudo passou pela elaboração de modelos explicativos de ML para as ocorrências de separação de mercado. Decidiu-se utilizar três métodos diferentes: Regressão Logística, Random Forest e Redes Neurais, que por essa mesma ordem representam uma hipotética troca entre interpretabilidade e complexidade do modelo. Isto permitiria encontrar possíveis consensos entre os modelos tal como evitar a dependência num só modelo. Enquanto a Regressão Logística e Random Forest têm presença na literatura à volta da análise de separação de mercado ou convergências de preços [5,

34], as Redes Neurais ainda não tinham sido aplicadas a este contexto específico.

Na avaliação dos modelos desenvolvidos verificou-se que o modelo com o melhor desempenho foi a Random Forest com uma exatidão de 97,30% e f1-score de 70,04% na amostra de teste e para um threshold de 0,35. Quanto à Rede Neuronal verificou-se uma exatidão de 95,40% e f1-score de 60,68% para um threshold de 0,7. Os algoritmos com o pior desempenho foram os dois modelos Logit, embora tivessem a vantagem de cada um modelar um dos dois cenários distintos da separação de mercado: congestionamento da interligação no sentido Espanha-Portugal e Portugal-Espanha. O primeiro teve uma exatidão de 97,09% e um f1-score de 54,55% para um threshold de 0,3, enquanto o segundo teve uma exatidão de 98,52% e um f1-score de 50,76% para um threshold de 0,35. É de se notar que para o estudo em questão, e devido ao desequilíbrio severo entre o número de ocorrências de separação de mercado e de convergência de preços, o f1-score é consideravelmente mais importante que a exatidão, visto que este representa a média harmónica da precisão e revocação, métricas associadas à classe positiva (separação de mercado).

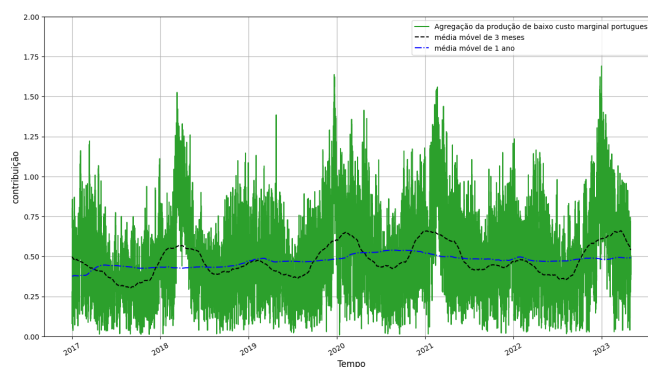
A aplicação de técnicas de XML e a inspeção direta dos coeficientes no caso dos modelos de Regressão Logística desenvolvidos, permitiram obter medidas de importâncias para as variáveis consideradas neste estudo. Embora não seja possível comparar os valores obtidos por tratarem-se de medidas de importância diferentes ou por considerarem dois cenários distintos (modelos Logit), podemos no entanto comparar as importâncias de uma dada variável quanto à posição em que surgem nas ordenações da importâncias de cada modelo.

As variáveis de maior importância comum aos modelos construídos são: **as capacidades de transferência disponíveis (ATCs)** em especial no sentido Espanha-Portugal; **produção eólica, hídrica e com ciclo combinado de gás natural portuguesas**. Existem algumas diferenças nas importâncias calculadas, que poderão estar associadas à utilização de diferentes medidas ou potenciais diferenças no treino dos algoritmos ou ajuste dos dados. Por exemplo, o modelo de Random Forest identifica a produção hídrica espanhola como relevante, enquanto os restantes modelos não. O mesmo acontece quando os modelos de Rede Neuronal e Logit atribuem importâncias altas aos consumos nacionais, enquanto o Random Forest não o faz.

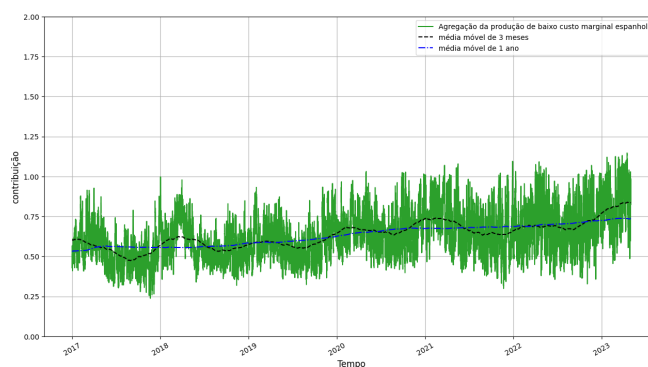
A importância elevada atribuída à produção eólica portuguesa pode ser explicada pelo facto de esta ser uma das principais fontes de energia de baixo custo marginal na zona de mercado portuguesa. Visto que em Portugal a produção solar está subdesenvolvida, a produção hídrica é dependente do ciclo hidrológico e serve outras dimensões do mercado, e o facto de não existir produção nuclear, levam a que a produção eólica seja uma componente essencial do mix energético português. Temos então que baixa produção eólica em Portugal leva ao aumento da probabilidade de separação de mercado no sentido Espanha-Portugal, pois existe uma maior quantidade de energia de baixo custo marginal na zona de mercado espanhola, enquanto que para elevada produção eólica, em momentos de baixa disponibilidade de renováveis em Espanha relativamente ao consumo nacional, aumenta

a probabilidade de separação de mercado no sentido Portugal-Espanha. A interação com os consumos nacionais também poderá ser um aspeto importante na relevância desta variável, tal como foi identificado no Capítulo 3.3 para certos padrões de ocorrência de separação de mercado.

De um modo geral, a relevância das variáveis de produção de baixo custo marginal portuguesas para todos os modelos, pode ser atribuída a uma diferença nos parques eletroprodutores dos dois países. As Figuras 6.1a e 6.1b mostram a série de tempo da agregação das contribuições para o consumo das variáveis de baixo custo marginal (eólica, solar, hídrica e nuclear) de ambos os países. Nestas é possível verificar que para Portugal, a disponibilidade de energia de baixo custo marginal varia significativamente mais que em Espanha, e que em média, a espanhola é superior à portuguesa. Também se verificou que a contribuição da produção de baixo custo marginal para o consumo tem aumentado consideravelmente em Espanha, possivelmente devido ao desenvolvimento recente da produção solar, enquanto que Portugal manteve-se relativamente constante.



(a) Portugal



(b) Espanha

Figura 6.1: Séries de tempo da agregação das contribuições associadas às variáveis de baixo custo marginal (eólica, solar, hídrica e nuclear)

A diferença de variabilidade entre os dois países pode ser explicada pela **diferença nas**

capacidades instaladas de produção de energia elétrica. Embora Espanha tenha uma capacidade instalada de eólica e hídrica inferior a Portugal, relativamente ao consumo nacional, esta tem acesso à produção nuclear (não existente em Portugal) e a uma produção solar mais desenvolvida, o que resulta numa maior disponibilidade de energia de baixo custo marginal em mais cenários, que por sua vez reduz a variabilidade. No caso português verifica-se momentos de muito elevada e muito baixa disponibilidade, uma consequência da dependência nas elevadas capacidades instaladas de eólica e hídrica, o que, face à consistência encontrada no lado espanhol, influencia a dinâmica de separação de mercado.

Foram também aplicados PDPs por forma a analisar os efeitos marginais de uma dada variável tendo em conta o efeito médio das restantes variáveis. Aqui foi possível observar alguns comportamentos interessantes, como no PDP das produções eólicas portuguesa e espanhola, em que se verifica que a probabilidade de separação de mercado é mínima quando as contribuições das produções para os respetivos consumos são semelhantes. Algo parecido acontece para o consumo, em que se verifica uma probabilidade mínima quando ambos os consumos são baixos e máxima quando existe um desequilíbrio óbvio.

Para todos os modelos verifica-se um impacto reduzido das produções solares na probabilidade de separação de mercado. Isto poderá ser um dos principais fatores de possíveis alterações na dinâmica de separação de mercado, dadas as ambições de ambos os países para esta tecnologia. Também se deve ter em conta que o impacto da produção solar vai além das ofertas no mercado diário, pelo que também altera os perfis de consumo através da conhecida "duck curve" da Califórnia [85]. Isto deve-se ao aumento significativo de instalações fotovoltaicas para auto-consumo, que durante certas horas do dia, eliminam ou reduzem substancialmente os consumos, diminuindo assim a procura na zona de mercado portuguesa.

6.2 Soluções para a redução das ocorrências de separação de mercado

Como já foi referido, as zonas de mercado portuguesa e espanhola estão consideravelmente bem integradas e as ocorrências de separação de mercado não têm um impacto significativo no preços marginais médios em cada país. No entanto, o aumento de potências instaladas de tecnologias renováveis, no âmbito de corresponder às metas definidas para 2030 pela UE, poderá alterar a dinâmica da separação de mercado. Pode-se então fazer uma apreciação das possíveis formas de reduzir as ocorrências. Considerando as observações feitas neste estudo e tendo em conta a literatura à volta deste tópico e similares, identificam-se duas maneiras distintas e complementares que poderão auxiliar na redução de separação de mercado:

- Aumento das capacidades de transferência disponíveis;
- Equilíbrio dos mixes energéticos;

Neste capítulo serão abordadas estas duas metodologias fundamentadas nos resultados obtidos.

6.2.1 Aumento das capacidades de transferência disponíveis

Segundo as importâncias calculadas (coeficientes no caso da regressão logística) a capacidade de interligação é um dos principais fatores determinantes da separação de mercado. Uma solução simples, mas que poderá não ser vantajosa face à potencial redução do spread médio ponderado dos preços de eletricidade, é o investimento em novas interligações. Atualmente, a interconexão entre Portugal e Espanha está definida por nove linhas de alta tensão, com uma prevista décima linha de 400 kV a entrar em serviço em 2024, permitindo uma capacidade mínima sustentável de 3000 MW, para propósitos comerciais, em ambos os sentidos [86].

Uma possível alternativa à construção de mais interligações, quando não há qualquer vantagem económica, passa pela alteração do modelo de equilíbrio térmico que determina as capacidades sazonais das linhas de transmissão, sendo este tradicionalmente um modelo de "Steady-State" que utiliza valores de referência sazonais como a velocidade do vento e irradiância ([Seasonal Line Rating \(SLR\)](#)), por um modelo dinâmico que calcula as capacidades de interligação segundo parâmetros atualizados para dados meteorológicos "near real-time" ([Dynamic Line Rating \(DLR\)](#)), como propõe [87]. Em [87] considera-se um modelo numérico de previsão do tempo que, quando determinadas as horas em separação de mercado, recalcula os parâmetros a utilizar no modelo de equilíbrio térmico, o que permite fazer a verificação de capacidade de interligação "extra". Se a capacidade determinada em [DLR](#) superar a capacidade em [SLR](#), utiliza-se a nova capacidade, que por sua vez poderá ser suficiente para evitar a separação de mercado, e mesmo que tal não seja possível, esta metodologia permite reduzir os spreads entre os preços de eletricidade. Para o ano de 2018, considerando um cenário em que tanto Portugal e Espanha aplicam a metodologia de [DLR](#), e são atualizados os equipamentos da rede de transmissão, [87] estima uma redução de 57% nas horas em separação de mercado, passando de 456 horas para 260 horas.

6.2.2 Equilíbrio dos mixes energéticos

Como foi possível verificar nos resultados obtidos nos capítulos anteriores, o desequilíbrio na disponibilidade de energia de baixo custo marginal entre as zonas de mercado, causada por diferentes níveis de produção de energia renovável de tecnologias não despacháveis e voláteis, tal como possíveis diferenças nos perfis de consumo, são identificados como contribuidores significantes para a ocorrência de separação de mercado, coerente com as conclusões obtidas em [5, 34]. Em [48] é utilizada uma medida da similaridade dos mixes de geração de energia elétrica (**GMSI**) entre os mercados da França e Alemanha, que determina que a convergência de preços entre estes acontece nas horas em que as estruturas de geração são semelhantes. Isto deve-se ao facto de que uma quantidade

semelhante de ofertas nas mesmas posições da ordem de mérito económico resulta na redução significativa dos fluxos entre os mercados, permitindo assim a convergência de preços, exceto quando na presença de diferenças significativas nos perfis de consumo.

Deve se ter em conta que existem benefícios na existência de diferenças nos mixes de produção, visto que permitem um melhor aproveitamento da energia renovável e a consequente redução dos preços de eletricidade, no entanto vêm ao custo de maiores investimentos nas capacidades de interligação [48]. Dadas as possíveis futuras dissimilaridades entre as capacidades instaladas de tecnologias de baixo custo marginal nas zonas de mercado portuguesa e espanhola, relativamente aos respetivos consumos nacionais, será do interesse de ambos os países, e a ao nível europeu, a determinação de medidas que permitam uma evolução simétrica dos mercados interligados, como propõe [34]. Isto, conjuntamente com um aumento ponderado das capacidades de interligação, poderá permitir sustentar futuras alterações na dinâmica de separação de mercado, causado pelo aumento substancial de tecnologias como a solar, e manter os mercados português e espanhol bem integrados. A diferença nos tamanhos dos sistemas elétricos também deve ser considerada no estudo de um potencial desenvolvimento simétrico.

6.3 Trabalho futuro

O tema abordado neste estudo é complexo e suscetível a alterações causadas pelo desenvolvimento dos sistemas elétricos, pelo que se podem fazer algumas recomendações quanto à direção de futuras investigações da dinâmica de separação de mercado no MIBEL:

- Utilização de outros modelos de [ML](#) e outras arquiteturas de redes neuronais, em especial as adequadas a classificação binária em séries de tempo;
- Inclusão de variáveis que representem o mercado diário e que façam uma discriminação por várias tecnologias (não disponíveis durante a realização deste estudo);
- Utilização de modelos das redes elétricas compatíveis com a expansão e desenvolvimento das redes, que poderão representar uma alternativa à abordagem com modelos explicativos e permitir a previsão da dinâmica de ocorrência de separação de mercado para anos futuros recorrendo a, por exemplo, dados referentes aos cenários desenvolvidos pelo [European Network of Transmission System Operators for Electricity \(ENTSO-E\)](#) para o ano de 2030;

Tendo em conta o desenvolvimento planeado das fontes de energia renovável na península ibérica, que foi abordado no Capítulo 2.4, e os potenciais reforços das interligações, espera-se que este tema seja revisitado com o intuito de verificar as alterações na dinâmica de ocorrência de separação de mercado.

6.4 Conclusões

Esta dissertação teve como principal objetivo a análise do contexto atual da separação de mercado no MIBEL e determinação das suas causas. Para tal foi necessário fazer uma análise dos dados referentes ao período de 2017 a 2022 e construir modelos explicativos para extrair conhecimento acerca das ocorrências de separação de mercado.

Foi através da análise da diferença de preços entre a zona portuguesa e espanhola que se construiu uma variável binária representativa da ocorrência de separação de mercado, com a qual se identificaram padrões.

A implementação de algoritmos de ML acompanhados de técnicas de XML permitiram concluir que a produção eólica portuguesa, as capacidades de transferência disponíveis, desequilíbrios na disponibilidade de eletricidade de baixo custo marginal e consumos nacionais, tal como a diferença na dimensão dos sistemas elétricos português e espanhol são fatores determinantes da separação de mercado.

A quantidade reduzida de ocorrências de separação de mercado no período considerado, embora represente um ótimo sinal de integração das zonas de mercado portuguesa e espanhola, foi um fator que influenciou a qualidade dos modelos e a capacidade de identificar as verdadeiras causas. Futuros desenvolvimentos, especialmente nas capacidades instaladas das produções solares de ambos os países, visto que não foram identificadas, explicitamente, como relevantes para os modelos, poderão alterar significativamente a dinâmica de separação de mercado.

Consideram-se algumas recomendações para a redução da separação de mercado, ou pelo menos a manutenção do elevado nível de integração verificado, que consistiram no aumento das capacidades de interligação, tendo em conta um método menos dispendioso comparativamente à construção de novas interligações, o DLR, e o desenvolvimento simétrico das tecnologias de produção renovável em ambas as zonas de mercado.

Como foi ilustrado nos capítulos anteriores, o setor elétrico encontra-se num processo constante de mudança. Desafios como a correta integração das energias renováveis nos mercados de eletricidade, os compromissos de harmonização tomados em prol da edificação de um Mercado Interno de Energia, tal como as metas ambiciosas definidas pela UE, representam a complexidade do setor. Compreender as várias dinâmicas dos mercados de eletricidade, especialmente como as tecnologias do futuro os poderão afetar, poderá servir de guia para um melhor design de mercado e na construção de políticas e medidas capazes de fazer face às necessidades energéticas do futuro.

BIBLIOGRAFIA

- [1] Nuno Carvalho Figueiredo e Patrícia Pereira Da Silva. «The price of wind power generation in Iberia and the merit-order effect». Em: *International Journal of Sustainable Energy Planning and Management* 15 (2018). ISSN: 22462929. DOI: [10.5278/ijsepm.2018.15.4](https://doi.org/10.5278/ijsepm.2018.15.4).
- [2] European Commission. «2030 Climate Energy Framework». Em: *Report* (October 2014). ISSN: 0717-6163.
- [3] Consilium. *Fit for 55 - The EU's plan for a green transition*. URL: <https://www.consilium.europa.eu/pt/policies/green-deal/fit-for-55-the-eu-plan-for-a-green-transition/>.
- [4] *Estatísticas rápidas das renováveis - dezembro 2021*. URL: <https://www.dgeg.gov.pt/pt/destaques/estatisticas-rapidas-das-renovaveis-dezembro-2021/>.
- [5] Nuno Carvalho Figueiredo, Patrícia Pereira da Silva e Pedro A. Cerqueira. «Evaluating the market splitting determinants: Evidence from the Iberian spot electricity prices». Em: *Energy Policy* 85 (2015). ISSN: 03014215. DOI: [10.1016/j.enpol.2015.06.013](https://doi.org/10.1016/j.enpol.2015.06.013).
- [6] *REN DATA HUB*. URL: <https://datahub.ren.pt/>.
- [7] S Hunt e Joan S. Hunt. «Making competition work in electricity». Em: *Immunological Reviews* 213 (8 2002). ISSN: 0941-293X. DOI: [10.1007/s00347-007-1607-9](https://doi.org/10.1007/s00347-007-1607-9).
- [8] José Pedro Sucena Paiva. *Redes de Energia Elétrica - Uma Análise Sistémica*.
- [9] João Manuel da Silva Araújo. «Análise dos Resultados do Mercado Ibérico de Eletricidade no ano de 2017». 2018. URL: <https://hdl.handle.net/10216/113398>.
- [10] Isabel Praça et al. *Mascem: A Multiagent System that Simulates Competitive Electricity Markets*. 2003. DOI: [10.1109/MIS.2003.1249170](https://doi.org/10.1109/MIS.2003.1249170).
- [11] Yogesh K Bichpuriya e S. A. Soman. «Electric Power Exchanges : A Review». Em: 2010.

- [12] Alexander Weber, Dietmar Graeber e Andreas Semmig. «Market Coupling and the CWE Project». Em: *Zeitschrift für Energiewirtschaft* 34 (4 2010). ISSN: 0343-5377. DOI: 10.1007/s12398-010-0033-x.
- [13] NordPool. *Price Coupling of Regions (PCR)*. URL: <https://www.nordpoolgroup.com/en/the-power-market/Day-ahead-market/Price-coupling-of-regions/>.
- [14] ERSE. *Caracterização do Sector Eléctrico, Portugal Continental*. 2001. URL: <https://paginas.fe.up.pt/~mam/Sector2001x.pdf>.
- [15] DGEG. *Mercados e mecanismos de capacidade*. URL: <https://www.dgeg.gov.pt/pt/areas-transversais/mercados-e-mecanismos-de-capacidade/mibel/>.
- [16] Conselho de Reguladores do MIBEL. *Descrição do Funcionamento do MIBEL*.
- [17] OMIE. *Mercado da electricidade*. URL: <https://www.omie.es/pt/mercado-de-electricidad>.
- [18] REN. *Caracterização das Interligações a 2021-12-31*. 2022. URL: <https://mercado.ren.pt/PT/Electr/AcessoRedes/AcessoRNT/CaractRNT/BibRelAnoInter/CIRel2021.pdf>.
- [19] (EU) 2015/1222 of 24 July 2015. *South West Europe TSOs common capacity calculation methodology for the day-ahead and intraday market timeframe in accordance with Article 21 of Commission Regulation*. 2022. URL: https://mercado.ren.pt/PT/Electr/ActServ/ProjInter/NCs/CACM/BibCACM/SWE%5C%20CCM%5C%20Amendment_January2022_clean-v3.pdf.
- [20] RD Nester e LNEG. *OPTIGRID project Deliverable 4.2 - Development of Case Study C - Market splitting scenarios in MIBEL*.
- [21] EU. *2030 climate energy framework*. URL: https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2030-climate-energy-framework_en.
- [22] EU. *Clean energy for all Europeans package*. URL: https://energy.ec.europa.eu/topics/energy-strategy/clean-energy-all-europeans-package_en.
- [23] DGEG. *Estatísticas rápidas das renováveis, janeiro 2023*. URL: <https://www.dgeg.gov.pt/pt/destaques/estatisticas-rapidas-das-renovaveis-janeiro-2023/>.
- [24] REE. *Informe del Sistema Eléctrico - Informe resumen de energías renovables*. 2023. URL: https://www.sistemaelectrico-ree.es/sites/default/files/2023-03/Informe_Renovables_2022.pdf.
- [25] *Plano Nacional de Energia e Clima 2030*. 2019. URL: https://www.dgeg.gov.pt/media/vedhi5t1/pnec-pt_template-final-vers%C3%A3o-final_30_06_2023.pdf.
- [26] *Plan Nacional Integrado de Energía y Clima (PNIEC) 2021-2030*. 2020. URL: <https://energia.gob.es/en-us/participacion/paginas/detalleparticipacionpublica.aspx?k=607>.

- [27] Nuno Carvalho Figueiredo, Patrícia Pereira da Silva e Pedro A. Cerqueira. «It is windy in Denmark: Does market integration suffer?» Em: *Energy* 115 (2016-11), pp. 1385–1399. ISSN: 03605442. DOI: [10.1016/j.energy.2016.05.038](https://doi.org/10.1016/j.energy.2016.05.038).
- [28] Nuno Carvalho Figueiredo, Patricia Pereira da Silva e Pedro Cerqueira. «Market Splitting and the Swedish electricity market». Em: *IEEE*, 2016-06, pp. 1–7. ISBN: 978-1-5090-1298-5. DOI: [10.1109/EEM.2016.7521189](https://doi.org/10.1109/EEM.2016.7521189).
- [29] A. Gianfreda e L. Grossi. «Zonal price analysis of the Italian wholesale electricity market». Em: *IEEE*, 2009-05, pp. 1–6. ISBN: 978-1-4244-4455-7. DOI: [10.1109/EEM.2009.5207198](https://doi.org/10.1109/EEM.2009.5207198).
- [30] Eirik S. Amundsen e Lars R. Bergman. «Integration of multiple national markets for electricity: The case of Norway and Sweden». Em: *Energy Policy* (2007-06). DOI: [10.1016/j.enpol.2006.12.014](https://doi.org/10.1016/j.enpol.2006.12.014). URL: <https://doi.org/10.1016/j.enpol.2006.12.014>.
- [31] Pär Holmberg e Ewa Lazarczyk. «Congestion management in electricity networks: nodal, zonal and discriminatory pricing». Em: *Social Science Research Network* (2012-01). DOI: [10.2139/ssrn.2055655](https://doi.org/10.2139/ssrn.2055655). URL: <https://doi.org/10.2139/ssrn.2055655>.
- [32] Christoph Fraunholz et al. «On the long-term efficiency of market splitting in Germany». Em: *Energy Policy* 149 (2021-02), p. 111833. DOI: [10.1016/j.enpol.2020.111833](https://doi.org/10.1016/j.enpol.2020.111833). URL: <https://doi.org/10.1016/j.enpol.2020.111833>.
- [33] Veronika Grimm et al. «On the long run effects of market splitting: Why more price zones might decrease welfare». Em: *Energy Policy* 94 (2016-07), pp. 453–467. DOI: [10.1016/j.enpol.2015.11.010](https://doi.org/10.1016/j.enpol.2015.11.010). URL: <https://doi.org/10.1016/j.enpol.2015.11.010>.
- [34] Yago Saez et al. «Integration in the European electricity market: A machine learning-based convergence analysis for the Central Western Europe region». Em: *Energy Policy* 132 (2019). ISSN: 03014215. DOI: [10.1016/j.enpol.2019.06.004](https://doi.org/10.1016/j.enpol.2019.06.004).
- [35] Luis Corona, Asuncion Mochon e Yago Saez. «Electricity market integration and impact of renewable energy sources in the Central Western Europe region: Evolution since the implementation of the Flow-Based Market Coupling mechanism». Em: *Energy Reports* 8 (2022). ISSN: 23524847. DOI: [10.1016/j.egy.2021.12.077](https://doi.org/10.1016/j.egy.2021.12.077).
- [36] Alfredo Vellido. «The importance of interpretability and visualization in machine learning for applications in medicine and health care». Em: *Neural Computing and Applications* 32.24 (2019-02), pp. 18069–18083. DOI: [10.1007/s00521-019-04051-w](https://doi.org/10.1007/s00521-019-04051-w). URL: <https://doi.org/10.1007/s00521-019-04051-w>.
- [37] Jeremy Petch, Shuang Di e William H. Nelson. «Opening the black box: The promise and limitations of Explainable machine learning in Cardiology». Em: *Canadian Journal of Cardiology* 38.2 (2022-02), pp. 204–213. DOI: [10.1016/j.cjca.2021.09.004](https://doi.org/10.1016/j.cjca.2021.09.004). URL: <https://doi.org/10.1016/j.cjca.2021.09.004>.

- [38] Grégoire Montavon, Wojciech Samek e Klaus Robert Müller. «Methods for interpreting and understanding deep neural networks». Em: *Digital Signal Processing* 73 (2018-02), pp. 1–15. DOI: [10.1016/j.dsp.2017.10.011](https://doi.org/10.1016/j.dsp.2017.10.011). URL: <https://doi.org/10.1016/j.dsp.2017.10.011>.
- [39] Lukas-Valentin Herm et al. «Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability». Em: *International Journal of Information Management* 69 (2023-04), p. 102538. DOI: [10.1016/j.ijinfomgt.2022.102538](https://doi.org/10.1016/j.ijinfomgt.2022.102538). URL: <https://doi.org/10.1016/j.ijinfomgt.2022.102538>.
- [40] Amina Adadi e Mohammed Berrada. «Peeking Inside the Black-Box: A survey on Explainable Artificial Intelligence (XAI)». Em: *IEEE Access* 6 (2018-01), pp. 52138–52160. DOI: [10.1109/access.2018.2870052](https://doi.org/10.1109/access.2018.2870052). URL: <https://doi.org/10.1109/access.2018.2870052>.
- [41] Alejandro Barredo Arrieta et al. «Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI». Em: *Information Fusion* 58 (2020-06), pp. 82–115. DOI: [10.1016/j.inffus.2019.12.012](https://doi.org/10.1016/j.inffus.2019.12.012). URL: <https://doi.org/10.1016/j.inffus.2019.12.012>.
- [42] Vaishak Belle e Ioannis Papantonis. «Principles and practice of explainable machine learning». Em: *Frontiers in big data* 4 (2021-07). DOI: [10.3389/fdata.2021.688969](https://doi.org/10.3389/fdata.2021.688969). URL: <https://doi.org/10.3389/fdata.2021.688969>.
- [43] OMIE. *Preço horário do mercado diário*. URL: <https://www.omie.es/pt/market-results/daily/daily-market/daily-hourly-price>.
- [44] OMIE. *Capacidade e ocupação das interconexões após as restrições do mercado diário*. URL: https://www.omie.es/pt/file-access-list?parents%5B0%5D=/&parents%5B1%5D=%20Mercado%20Di%C3%A1rio&parents%5B2%5D=6.%20Capacidades&dir=Capacidade%20e%20ocupação%20das%20interconexões%20após%20as%20restrições%20do%20mercado%20diário&readdir=capacidad_inter_pvp.
- [45] REE. *Seguimiento de la demanda de energía eléctrica*. URL: <https://demanda.ree.es/visiona/peninsula/demandaqh/total>.
- [46] The pandas development team. *pandas-dev/pandas: Pandas*. Versão latest. 2020-02. DOI: [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134). URL: <https://doi.org/10.5281/zenodo.3509134>.
- [47] Tryggvi Jónsson, Pierre Pinson e Henrik Madsen. «On the market impact of wind energy forecasts». Em: *Energy Economics* 32 (2 2010-03), pp. 313–320. ISSN: 01409883. DOI: [10.1016/j.eneco.2009.10.018](https://doi.org/10.1016/j.eneco.2009.10.018).

- [48] Klaus Gugler e Adhurim Haxhimusa. «Market integration and technology mix: Evidence from the German and French electricity markets». Em: *Energy Policy* 126 (2019-03), pp. 30–46. ISSN: 03014215. DOI: [10.1016/j.enpol.2018.10.014](https://doi.org/10.1016/j.enpol.2018.10.014).
- [49] Bruce Ratner. «The correlation coefficient: Its values range between +1/1, or do they?» Em: *Journal of Targeting, Measurement and Analysis for Marketing* 17 (2 2009-06), pp. 139–142. ISSN: 1479-1862. DOI: [10.1057/jt.2009.5](https://doi.org/10.1057/jt.2009.5).
- [50] Pauli Virtanen et al. «SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python». Em: *Nature Methods* 17 (2020), pp. 261–272. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [51] F. Pedregosa et al. «Scikit-learn: Machine Learning in Python». Em: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [52] Juliana Tolles e William J. Meurer. «Logistic Regression». Em: *JAMA* 316 (5 2016-08), p. 533. ISSN: 0098-7484. DOI: [10.1001/jama.2016.7653](https://doi.org/10.1001/jama.2016.7653).
- [53] Gareth James et al. *An Introduction to Statistical Learning*. Springer International Publishing, 2023. ISBN: 978-3-031-38746-3. DOI: [10.1007/978-3-031-38747-0](https://doi.org/10.1007/978-3-031-38747-0).
- [54] Christy Lleras. «Path Analysis». Em: Elsevier, 2005, pp. 25–30. DOI: [10.1016/B0-12-369398-5/00483-7](https://doi.org/10.1016/B0-12-369398-5/00483-7).
- [55] Skipper Seabold e Josef Perktold. «statsmodels: Econometric and statistical modeling with python». Em: *9th Python in Science Conference*. 2010.
- [56] Leo Breiman et al. *Classification and regression trees*. Chapman e Hall/CRC, 1984-01.
- [57] «CART: Classification and Regression Trees». Em: Chapman e Hall/CRC, 2009-04, pp. 193–216. DOI: [10.1201/9781420089653-17](https://doi.org/10.1201/9781420089653-17).
- [58] Leo Breiman. «Random Forests». Em: *Machine Learning* 45 (1 2001), pp. 5–32. ISSN: 08856125. DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [59] Ioannis P. Panapakidis e Athanasios S. Dagoumas. «Day-ahead electricity price forecasting via the application of artificial neural network based models». Em: *Applied Energy* 172 (2016-06), pp. 132–151. ISSN: 03062619. DOI: [10.1016/j.apenergy.2016.03.089](https://doi.org/10.1016/j.apenergy.2016.03.089).
- [60] Gao Gao, Kwoklun Lo e Fulin Fan. «Comparison of ARIMA and ANN Models Used in Electricity Price Forecasting for Power Market». Em: *Energy and Power Engineering* 09 (04 2017), pp. 120–126. ISSN: 1949-243X. DOI: [10.4236/epe.2017.94B015](https://doi.org/10.4236/epe.2017.94B015).
- [61] Chafak Tarmanini et al. «Short term load forecasting based on ARIMA and ANN approaches». Em: *Energy Reports* 9 (2023-05), pp. 550–557. ISSN: 23524847. DOI: [10.1016/j.egyr.2023.01.060](https://doi.org/10.1016/j.egyr.2023.01.060).
- [62] Arjun Baliyan, Kumar Gaurav e Sudhansu Kumar Mishra. «A Review of Short Term Load Forecasting using Artificial Neural Network Models». Em: *Procedia Computer Science* 48 (2015), pp. 121–125. ISSN: 18770509. DOI: [10.1016/j.procs.2015.04.160](https://doi.org/10.1016/j.procs.2015.04.160).

- [63] Pedro Faria et al. «ANN based day-ahead ancillary services forecast for electricity market simulation». Em: *Melecon 2010 - 2010 15th IEEE Mediterranean Electrotechnical Conference*. 2010, pp. 1159–1164. DOI: [10.1109/MELCON.2010.5476367](https://doi.org/10.1109/MELCON.2010.5476367).
- [64] Philipp Staudt et al. «Predicting Transmission Line Congestion in Energy Systems with a High Share of Renewables». Em: *IEEE*, 2019-06, pp. 1–6. ISBN: 978-1-5386-4722-6. DOI: [10.1109/PTC.2019.8810527](https://doi.org/10.1109/PTC.2019.8810527).
- [65] Simon S. Haykin. *Neural networks and learning machines*. Third. Upper Saddle River, NJ: Pearson Education, 2009.
- [66] Yaoshiang Ho e Samuel Wookey. «The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling». Em: *IEEE Access* 8 (2020), pp. 4806–4813. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2019.2962617](https://doi.org/10.1109/ACCESS.2019.2962617).
- [67] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: <https://www.tensorflow.org/>.
- [68] Jasper Snoek, Hugo Larochelle e Ryan P. Adams. «Practical Bayesian Optimization of Machine Learning Algorithms». Em: (2012-06).
- [69] Helen R. Sofaer, Jennifer A. Hoeting e Catherine S. Jarnevich. «The area under the precision-recall curve as a performance metric for rare binary events». Em: *Methods in Ecology and Evolution* 10 (4 2019-04), pp. 565–577. ISSN: 2041-210X. DOI: [10.1111/2041-210X.13140](https://doi.org/10.1111/2041-210X.13140).
- [70] Fernando Nogueira. *Bayesian Optimization: Open source constrained global optimization tool for Python*. 2014–. URL: <https://github.com/fmfn/BayesianOptimization>.
- [71] Ajay Kulkarni, Deri Chong e Feras A. Batarseh. «Foundations of data imbalance and solutions for a data democracy». Em: Elsevier, 2020, pp. 83–106. DOI: [10.1016/B978-0-12-818366-3.00005-8](https://doi.org/10.1016/B978-0-12-818366-3.00005-8).
- [72] Gurol Canbek et al. «Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights». Em: *IEEE*, 2017-10, pp. 821–826. ISBN: 978-1-5386-0930-9. DOI: [10.1109/UBMK.2017.8093539](https://doi.org/10.1109/UBMK.2017.8093539).
- [73] Peter Bruce, Andrew Bruce e Peter Gedeck. *Practical Statistics for Data Scientists: 50+ essential concepts using R and Python*. 2020-04. URL: <https://openlibrary.telkomuniversity.ac.id/home/catalog/id/163886/slug/practical-statistics-for-data-scientists-50-essential-concepts-using-r-and-python-2-e-.html>.
- [74] Vimalraj S Spelmen e R Porkodi. «A Review on Handling Imbalanced Data». Em: *IEEE*, 2018-03, pp. 1–11. ISBN: 978-1-5386-3702-9. DOI: [10.1109/ICCTCT.2018.8551020](https://doi.org/10.1109/ICCTCT.2018.8551020).
- [75] Daniel Mcfadden. *Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments*. 1977.

- [76] Jacob Cohen et al. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Psychology Press, 2014-04. ISBN: 9781135468255. DOI: [10.4324/9781410606266](https://doi.org/10.4324/9781410606266).
- [77] Jerome H. Friedman. «Greedy function approximation: A gradient boosting machine.» Em: *The Annals of Statistics* 29 (5 2001-10). ISSN: 0090-5364. DOI: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451).
- [78] Christoph Molnar. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. 2^a ed. 2022. URL: <https://christophm.github.io/interpretable-ml-book>.
- [79] Trevor Hastie, Robert Tibshirani e Jerome H. Friedman. *The elements of statistical learning*. 2009-01. DOI: [10.1007/978-0-387-84858-7](https://doi.org/10.1007/978-0-387-84858-7). URL: <https://doi.org/10.1007/978-0-387-84858-7>.
- [80] Brandon Greenwell, Bradley C. Boehmke e Andrew J. McCarthy. «A simple and effective Model-Based variable importance measure». Em: *arXiv (Cornell University)* (2018-05). URL: <https://arxiv.org/pdf/1805.04755.pdf>.
- [81] Ilse Van Der Linden, Hinda Haned e Evangelos Kanoulas. «Global Aggregations of Local Explanations for Black Box models». Em: *arXiv (Cornell University)* (2019-07). DOI: [10.48550/arxiv.1907.03039](https://doi.org/10.48550/arxiv.1907.03039). URL: <http://arxiv.org/abs/1907.03039>.
- [82] Scott Lundberg e Su-In Lee. «A unified approach to interpreting model predictions». Em: *Neural Information Processing Systems* 30 (2017-12), pp. 4768–4777. URL: <https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
- [83] Lloyd S. Shapley. 17. *A value for N-Person games*. 1953-12, pp. 307–318. DOI: [10.1515/9781400881970-018](https://doi.org/10.1515/9781400881970-018). URL: <https://doi.org/10.1515/9781400881970-018>.
- [84] Marco Tulio Ribeiro, Sameer Singh e Carlos Guestrin. «"Why should I trust you?": explaining the predictions of any classifier». Em: *arXiv (Cornell University)* (2016-02). DOI: [10.48550/arxiv.1602.04938](https://doi.org/10.48550/arxiv.1602.04938). URL: <http://arxiv.org/abs/1602.04938>.
- [85] California ISO. *Demand Response And Energy Efficiency Roadmap*. 2017. URL: file:///C:/Users/franc/Downloads/TN217999_20170608T151647_Demand_Response_And_Energy_Efficiency_Roadmap.pdf.
- [86] REN. *Capacidades Indicativas de Interligação para Fins Comerciais para 2023*. 2022. URL: <https://mercado.ren.pt/PT/Electr/Explora%C3%A7%C3%A3odosistema/Interlig/Previsao/BibRelAno/Cap2023.pdf>.
- [87] Hugo Algarvio et al. «Increase cross-border capacity to reduce market splitting of day-ahead electricity markets - A dynamic line rating approach». Em: *IEEE*, 2022-04, pp. 1–5. ISBN: 978-1-6654-4329-6. DOI: [10.1109/TD43745.2022.9816949](https://doi.org/10.1109/TD43745.2022.9816949).



