UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS

Ciniro Aparecido Leite Nametala

Redes Neurais Atencionais aplicadas a modelagem e previsão de preços no Mercado de Eletricidade Brasileiro

São Carlos

2023

Ciniro Aparecido Leite Nametala

Redes Neurais Atencionais aplicadas a modelagem e previsão de preços no Mercado de Eletricidade Brasileiro

Tese apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, para obtenção do título de Doutor em Ciências - Programa de Pós-Graduação em Engenharia Elétrica.

Área de concentração: Sistemas Elétricos de Potência

Orientador: Prof. Dr. Benvindo Rodrigues Pereira Júnior

Trata-se da versão corrigida da tese. A versão original se encontra disponível na EESC-USP que aloja o Programa de Pós-Graduação de Engenharia Elétrica.

São Carlos

2023

FOLHA DE JULGAMENTO

Candidato: Tecnólogo **CINIRO APARECIDO LEITE NAMETALA.**

Título da tese: "Redes neurais atencionais aplicadas a modelagem e previsão de preços no mercado de eletricidade brasileiro".

Data da defesa: 14/02/2023.

Comissão Julgadora	Resultado
Prof. Dr. **Benvindo Rodrigues Pereira Junior (Orientador)** (Escola de Engenharia de São Carlos/EESC-USP)	APROVADO
Prof. Dr. **Dorel Soares Ramos** (Escola Politécnica/EP-USP)	APROVADO
Prof. Dr. **Vladimiro Henrique Barrosa Pinto de Miranda** (Instituto de Engenharia de Sistemas e Computadores/INESC-Portugal)	APROVADO
Prof. Dr. **Delberis Araujo Lima** (Pontifícia Universidade Católica do Rio de Janeiro/PUC-RJ)	APROVADO
Dr. **Júlio Augusto Druzina Massignan** (Siemens Infraestrutura e Indústria Ltda)	APROVADO

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica:
Prof. Dr. **João Bosco Augusto London Junior**

Presidente da Comissão de Pós-Graduação:
Prof. Titular **Murilo Araujo Romero**

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

Nametala, Ciniro Aparecido Leite
N174r Redes neurais atencionais aplicadas a modelagem e previsão de preços no mercado de eletricidade brasileiro / Ciniro Aparecido Leite Nametala; orientador Benvindo Rodrigues Pereira Júnior. São Carlos, 2023.

Tese (Doutorado) - Programa de Pós-Graduação em Engenharia Elétrica e Área de Concentração em Sistemas Elétricos de Potência -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2023.

1. Previsão de preços de eletricidade. 2. Preço da liquidação das diferenças. 3. Aprendizado profundo. 4. Mecanismos de atenção. 5. Análise de Séries Temporais. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

Aos guias, professores e mestres que
escolheram o altruísmo como meta de vida, dedico este trabalho.

AGRADECIMENTOS

Contribuíram diretamente para o desenvolvimento desta tese o Dr. Benvindo Rodrigues Pereira Júnior (professor orientador), Dr. Wandry Rodrigues Faria, Dr. Edilson Anselmo Correia Júnior, Dr. Adriel Magalhães de Souza, a Universidade de São Paulo e o Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais, *campus* Bambuí.

A estes, os meus mais sinceros agradecimentos.

Om bhūr bhuvaḥ svaḥ
tat savitur vareṇyaṃ
bhargo devasya dhīmahi
dhiyo yo naḥ pracodayāt Om

RESUMO

Nametala, C. A. L. **Redes Neurais Atencionais aplicadas a modelagem e previsão de preços no Mercado de Eletricidade Brasileiro**. 2023. 305p. Tese (Doutorado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2023.

Os mercados livres de energia elétrica vêm gradativamente substituindo modelos monopolistas com o objetivo de incentivar a competitividade. Entretanto, por mais liberal que seja um estado, o aprovisionamento energético é uma preocupação visto que historicamente o consumo de eletricidade apresenta correlação com crescimento econômico. Apesar desta relevância, prever variáveis relacionadas a energia elétrica é uma tarefa difícil devido às características intrínsecas desta *commodity*. Inelasticidade da demanda, condições climáticas, arcabouços regulatórios diversos e outros fatores tornam estas séries temporais voláteis, singulares e, em grande parte das vezes, não estacionárias. À revelia desta imprevisibilidade, diversos trabalhos vêm reportando sucesso na modelagem de séries como carga, geração e preço. O Brasil é atualmente um campo convidativo à pesquisa, pois desde 2004 tem promovido ações de desregulamentação. No entanto, é ao mesmo tempo um ambiente ainda pouco explorado, especialmente devido ao fato de possuir um mercado livre relativamente recente que opera apenas no atacado, possuir complexa regulação e, adicionalmente, realizar sua formação de custos de operação por meio de modelos matemáticos. Neste contexto, esta tese apresenta uma investigação multidisciplinar envolvendo conceitos em economia, técnicas de inteligência computacional e fundamentos de sistemas elétricos de potência para, no horizonte de curto prazo e com granularidade horária, realizar a modelagem e previsão do Preço de Liquidação de Diferenças (PLD). O PLD em base horária é a referência adotada em 2021 pela Câmara de Comercialização de Energia Elétrica para o mercado de dia seguinte no Brasil. Possui destaque neste documento uma análise histórica dos mercados de eletricidade nacional e mundial, uma revisão bibliográfica dos últimos 20 anos de publicações focadas em previsão energética e, quanto a metodologia, ao desenvolvimento de uma ferramenta capaz de realizar a estimação de cenários futuros no mercado livre de energia nacional para uma semana à frente. Estes cenários integram as previsões do PLD com um modelo de precificação e gestão de riscos capaz de munir o usuário com informações úteis à tomada de decisão. A principal contribuição desta pesquisa trata-se da disponibilização de um produto de inovação tecnológica que, no período de análise, além ter se mostrado mais acurado do que outras técnicas estado da arte na atividade de previsão, também leva em conta quesitos como segurança e rentabilidade ao considerar aspectos econômicos que são também de interesse dos agentes do setor elétrico.

Palavras-chave: Previsão de preços de eletricidade. Preço da liquidação das diferenças. Aprendizado profundo. Mecanismos de atenção. Análise de Séries Temporais.

ABSTRACT

Nametala, C. A. L. **Attentional Neural Networks applied to modeling and price forecasting in the Brazilian Electricity Market.**. 2023. 305p. Thesis (Doctor) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2023.

Free electricity markets are gradually replacing monopolistic models to encourage competitiveness. However, no matter how liberal a state is, energy supply is a concern as electricity consumption has historically been correlated with economic growth. Despite this relevance, predicting variables related to electricity is a difficult task due to the intrinsic characteristics of this *commodity*. Demand inelasticity, weather conditions, diverse regulatory frameworks, and other factors make these time series volatile, singular, and, in most cases, non-stationary. Despite this unpredictability, several studies have reported success in modeling series such as load, generation, and price. Brazil is currently an inviting field for research, as it has been promoting deregulation actions since 2004. However, it is at the same time an environment still little explored, especially because it has a relatively recent free market that operates only wholesale, has complex regulations, and, additionally, carries out its formation of operating costs through mathematical models. In this context, this work presents a multidisciplinary investigation involving concepts in economics, computational intelligence techniques, and fundamentals of electric power systems to, in the short-term horizon and with hourly granularity, carry out the modeling and forecast of the Differences Settlement Price (DSP). The hourly DSP is the reference adopted in 2021 by the Electric Energy Commercialization Chamber for the day-ahead market in Brazil. This thesis highlights a historical analysis of the national and world electricity markets, a bibliographic review of the last 20 years of publications focused on energy forecasting, and, in terms of methodology, the development of a tool capable of estimating future scenarios in the Brazilian free market for a week ahead. These scenarios integrate PLD forecasts with a pricing and risk management model capable of providing users with useful information for decision-making. The main contribution of this research is, therefore, the availability of a product of technological innovation that, in the period of analysis, in addition to being more accurate than other state-of-the-art techniques in the forecasting activity, also takes into account questions such as security and profitability when considering economic aspects that are also of interest to agents in the electricity sector.

Keywords: Electricity price forecasting. Brazilian electricity price. Deep learning. Attention mechanisms. Time Series Analysis.

LISTA DE ABREVIATURAS E SIGLAS

ACL	Ambiente de Contratação Livre
ACM	*Association for Computing Machinery*
ACR	Ambiente de Contratação Regulada
AE	*Autoencoder*
AGF	Aporte de Garantia Financeira
AIC	*Akaike Information Criteria*
ANEEL	Agência Nacional de Energia Elétrica
ARIMA	*Autoregressive Integrated Moving Average*
ARX	*Autoregressive with eXogenous Variables*
ASINH	*Arcsine Hiperbolic*
BFE	*Backward Feature Elimination*
BPTT	*Backpropagation Through Time*
CA	Corrente Alternada
CAE	*Convolutional Autoencoder*
CAES	*Compressed Air Energy Storage*
CatBoost	*Categorical Boosting*
CC	Corrente Contínua
CCC	Custo da carga consumida
CCE	*Continuous Cyclic Encoding*
CCEE	Câmara de Comercialização de Energia Elétrica
CCEN	Contrato de Cotas de Energia Nuclear
CCGF	Contrato de Cotas de Garantia Física
CDI	Certificado de Depósito Interbancário
CEPEL	Centro de Pesquisas de Energia Elétrica

CERN	*Conseil Européen pour la Recherche Nucléaire*
CEXC	Custo da carga excedente
CGH	Central Geradora Hidroelétrica
CGU	Central Geradora Undi-elétrica
CHESF	Companhia Hidroelétrica do São Francisco
CIC	Custo do intervalo de variação da carga programada
CIP	Custo do intervalo de variação do PLD previsto
CIV	Custo dos intervalos de variação
CMO	Custo Marginal de Operação
CMSE	Comitê de Monitoramento do Setor Elétrico
CNN	*Convolutional Neural Network*
CNPE	Conselho Nacional de Política Energética
COA	*Coyote Optimization Algorithm*
COP21	Conferência das Partes da Convenção-Quadro das Nações Unidas sobre Mudança do Clima
CPAMP	Comissão Permanente para Análise de Metodologias e Programas Computacionais do Setor Elétrico
CPL	Consumidor Parcialmente Livre
CPRO	Custo da carga programada
CPU	*Central Processing Unit*
CPUC	Custo percentual por unidade de variação da programação da carga
CPUP	Custo percentual por unidade de variação da previsão do PLD
cVaR	*Conditional Value at Risk*
DAM	*Day-ahead Market*
DBN	*Deep Bilief Network*
DCR	*Demand-to-Capacity Ratio*
DER	Diagrama Entidade-Relacionamento

DL	*Deep Learning*
DNN	*Deep Neural Network*
DTSF	*Dynamic Time Scan Forecasting*
DTW	*Dynamic Time Warping*
ECMWF	*European Centre for Medium-Range Weather Forecasts*
EER	Encargo de Energia de Reserva
EESC	Escola de Engenharia de São Carlos
ELM	*Extreme Learning Machine*
EMD	*Empirical Model Decomposition*
ENA	Energia Natural Afluente
EOL	Central Geradora Eolielétrica
EPE	Empresa de Pesquisa Energética
EPF	*Electricity Price Forecasting*
ESS	Encargo de Serviço do Sistema
EUA	Estados Unidos da América
fARX	*Fractional-order AutoRegressive Model with eXogenous Variables*
FCF	Função de Custo Futuro
FCI	Função de Custo Imediato
FCT	Função de Custo Total
FFE	*Forward Feature Elimination*
FGV	Fundação Getúlio Vargas
FPNPP	Fator percentual de negociação do PLD previsto
FPNPR	Fator percentual de negociação do PLD real
GARCH	*Generalized Autoregressive Conditional Heteroskedasticity*
GBD	*Gestor do Banco de Dados*
GBM	*Gradient Boosting Machine*

GDA	*Generalized Discriminant Analisys*
GF	Garantia Física
GNL	Gás Natural Liquefeito
GPU	*Graphics Processing Unit*
GRU	*Gated Recurrent Unit*
GW	Giacomini-White
HTTP	*Hypertext Transfer Protocol*
HTTPS	*Hypertext Transfer Protocol Secure*
IA	Inteligência Artificial
IBE	Instituto Brasileiro de Economia
IBGE	Instituto Brasileiro de Geografia e Estatística
IBOV	Índice Bovespa
ICA	*Independent Component Analysis*
ICAP	*Installed Capacity Markets*
ICE	Índice de Custo de Edificações
IDM	*Intraday Market*
IEE-BR	Indicador de Incerteza da Economia do Brasil
IGP-DI	Índice Geral de Preços de Disponibilidade Interna
INMET	Instituto Nacional de Meteorologia
IPCA	Índice Nacional de Preços ao Consumidor
IPP	*Independent Power Producer*
IQR	*Interquartile Range*
IVC	Intervalo percentual de variação da carga programada
IVP	Intervalo percentual de variação do PLD previsto
JSON	*JavaScript Object Notation*
KDE	*Kernel Density Estimation*

KNN	*K-Nearest Neighbors*
KPSS	*Kwiatkowski–Phillips–Schmidt–Shin*
LASSO	*Least Absolute Shrinkage and Selection Operator*
LDA	*Linear Discriminant Analysis*
LEAR	*LASSO Estimated Autoregressive*
LF	*Load Forecasting*
LLE	*Locally Linear Embedding*
LMP	*Local Marginal Price*
LOLP	*Loss of Load Probability*
LSTM	*Long Short-Term Memory*
LT	Linha de Transmissão
LTLF	*Long-term Load Forecasting*
M1	Módulo 1: Banco de Dados
M2	Módulo 2: Construtor de *Datasets*
M3	Módulo 3: Construtor do Previsor
M4	Módulo 4: Gerador de Cenários
MAD	*Median Absolute Deviation*
MAE	*Mean Absolute Error*
MAPE	*Mean Absolute Percentage Error*
MASE	*Mean Absolute Scaled Error*
MCLP	*Market Clearing Price*
MCP	Mercado de Curto Prazo
MCU	Mercado de Comprador Único
MIBEL	*Mercado Ibérico de Electricidade*
ML	*Machine Learning*
MLP	*Multilayer Perceptron*

MME	Média Móvel Simples
MME	Ministério de Minas e Energia
MRE	Mecanismos de Realocação de Energia
MSE	*Mean Squared Error*
MSEB	Marco Regulatório do Setor Elétrico Brasileiro
MTE	Máximo Teto Estrutural
MTLF	*Medium-term Load Forecasting*
MVC	*Model-View-Controller*
MVI	Monopólio Verticalmente Integrado
NAG	*Nesterov Accelerated Gradient*
NAN	*Neural Attention Network*
nARX	*Non-linear Autoregressive with eXogenous Variables*
NE	Nordeste
NETA	*New Energy Trade Arrangements*
NLP	*Natural Language Processing*
NNI	*Nearest Neighbour Interpolation*
NORTE	Norte
NPIT	*Normal Probability Integral Transform*
NWP	*Numeric Weather Prediction*
OCDE	Organização para Cooperação e Desenvolvimento Econômico
OHE	*One Hot Encoding*
OLS	*Ordinary Least Squares*
OMS	Organização Mundial da Saúde
ONS	Operador Nacional do Sistema
OPB	Observatório do Preço da Eletricidade do Brasil
P	Preço a pagar

PCA	*Principal Component Analisys*
PCH	Pequena Central Hidroelétrica
PDDE	Programação Dinâmica Dual Estocástica
PDE	Plano Decenal de Expansão
PDO	Programação Diária da Operação
PE	Previsão Energética
PEN	Planejamento Anual da Operação Energética
PEXC	Preço a pagar pela carga excedente
PGD	Processo Gerador de Dados
PHS	*Pumped Hydroeletric Storage*
PIB	Produto Interno Bruto
PIE	Produtor Independente de Energia
PIT	*Probability Integral Transform*
PLD	Preço de Liquidação de Diferenças
PLIM	Programação Linear Inteira-Mista
PMO	Programação Mensal da Operação
POCID	*Percentage of Change in Direction*
PPA	*Power Purchase Agreement*
PPE	Previsão de Preços de Eletricidade
PPRO	Preço a pagar pela carga programada
PUHCA	*Public Utility Holding Company Act*
QA	*Question and Answer*
RBF	*Radial Basis Function*
RE-SEB	Projeto de Reestruturação do Setor Elétrico Brasileiro
RFE	*Recursive Feature Elimination*
rMAE	*Relative Mean Absolute Error*

RMSE	*Root Mean Squared Error*
RNA	Rede Neural Artificial
RNN	*Recurrent Neural Network*
RSS	*Residual Sum of Squares*
RTM	*Real-time Market*
SAE	*Stacked Autoencoder*
SAR	Superfície de Aversão ao Risco
SARIMA	*Seasonal Autoregressive Integrated Moving Average*
SCDE	Sistema de Coleta de Dados de Energia
SDA	*Stacked Denoising Autoencoder*
SEB	Sistema Elétrico Brasileiro
SECO	Sudeste/Centro-Oeste
SEP	Sistemas Elétricos de Potência
Seq2Seq	*Sequence to Sequence*
SES	*Seasonal Exponential Smoothing*
SGBD	Sistema Gerenciador de Bancos de Dados
SGD	*Stochastic Gradient Descent*
SIN	Sistema Integrado Nacional
sMAPE	*Symmetric Mean Absolute Percentage Error*
SMBO	*Sequence Model-Based Optimization*
SMP	*System Marginal Price*
SMTP	*Simple Mail Transfer Protocol*
SOM	*Self-organizing Map*
SPF	*Solar Photovoltaic Forecasting*
SQL	*Structured Query Language*
STLF	*Short-term Load Forecasting*

SUL	Sul
SVM	*Suport Vector Machine*
SVR	*Support Vector Regression*
TDM	Teoria de *Design* de Mercados
TEO	Tarifa de Energia de Otimização
TPC	Taxa de penalização por erro da carga programada
TPP	Taxa de penalização por erro do PLD previsto
TUST	Tarifa de Uso do Sistema de Transmissão
UFV	Central Geradora Solar Fotovoltaica
UHE	Usina Hidroelétrica
USP	Universidade de São Paulo
UTE	Usina Termoelétrica
UTN	Usina Termonuclear
VAE	*Variational Autoencoder*
VaR	*Value at Risk*
VAT	Variável de atenuação
VOLL	*Value of Lost Load*
VSTLF	*Very Short-term Load Forecasting*
VT	Valor total pago pela carga consumida
WAD	*Wide and Deep*
WoS	*Web of Science*
WPF	*Wind Power Forecasting*
xc	Custo exponencial do intervalo de variação da carga programada
XGB	*XG Boosting*
XML	*Extensible Markup Language*
xp	Custo exponencial do intervalo de variação do PLD previsto

SUMÁRIO

1 INTRODUÇÃO . . . 27
1.1 Justificativa da pesquisa . . . 29
1.2 Proposta . . . 32
1.2.1 Características da pesquisa . . . 33
1.2.2 Objetivos . . . 34
1.2.3 Contribuições . . . 35
1.3 Estrutura do documento . . . 36

2 CONTEXTUALIZAÇÃO . . . 39
2.1 Economia da energia . . . 39
2.2 Mercados de eletricidade . . . 44
2.2.1 Evolução dos mercados de eletricidade . . . 45
2.2.2 Teoria de *design* de mercados de eletricidade . . . 56
2.3 Mercado de eletricidade brasileiro . . . 71
2.3.1 Panorama atual . . . 78
2.3.2 Arquitetura de mercado . . . 85

3 REVISÃO DE LITERATURA . . . 101
3.1 Aprendizado profundo . . . 101
3.1.1 Redes neurais atencionais . . . 114
3.2 Previsão energética . . . 121
3.3 Previsão de preços de eletricidade . . . 131
3.3.1 Estado da arte . . . 137
3.3.2 Previsão de preços de eletricidade com mecanismos de atenção . . . 146

4 FUNDAMENTAÇÃO TEÓRICA . . . 149
4.1 Modelagem de preços de eletricidade . . . 149
4.1.1 Definição matemática . . . 149
4.1.2 Caracterização . . . 155
4.1.3 Modelos de previsão de curto prazo . . . 165
4.2 Modelos neurais profundos . . . 170
4.2.1 Redes neurais artificiais . . . 170
4.2.2 Arquiteturas recorrentes . . . 174
4.2.3 Mecanismos de atenção . . . 180
4.2.4 Células de memória . . . 182
4.2.5 *Autoencoders* . . . 185

4.2.6 Técnicas complementares 186
4.3 Gestão de riscos 189
4.3.1 Modelos de risco VaR e CVaR 190

5 METODOLOGIA 193
5.1 Organização do sistema 193
5.2 Módulos 195
5.2.1 Módulo 1: Banco de dados 196
5.2.2 Módulo 2: Construtor de *datasets* 199
5.2.3 Módulo 3: Construtor do previsor 209
5.2.4 Módulo 4: Gerador de cenários 212

6 RESULTADOS 223
6.1 Protocolo experimental 223
6.1.1 Seleção de entradas e métodos de pré-processamento 226
6.2 Compactação das variáveis climáticas 228
6.2.1 Pré-processamento 228
6.2.2 Codificação 231
6.3 Previsão do PLD 232
6.3.1 Capacidade preditiva 235
6.3.2 Análise financeira 248

7 CONCLUSÃO 257
7.1 Limitações e propostas de continuidade 261
7.2 Produção científica 263

REFERÊNCIAS 265

APÊNDICES 297

APÊNDICE A – DIAGRAMA ENTIDADE-RELACIONAMENTO . . 299

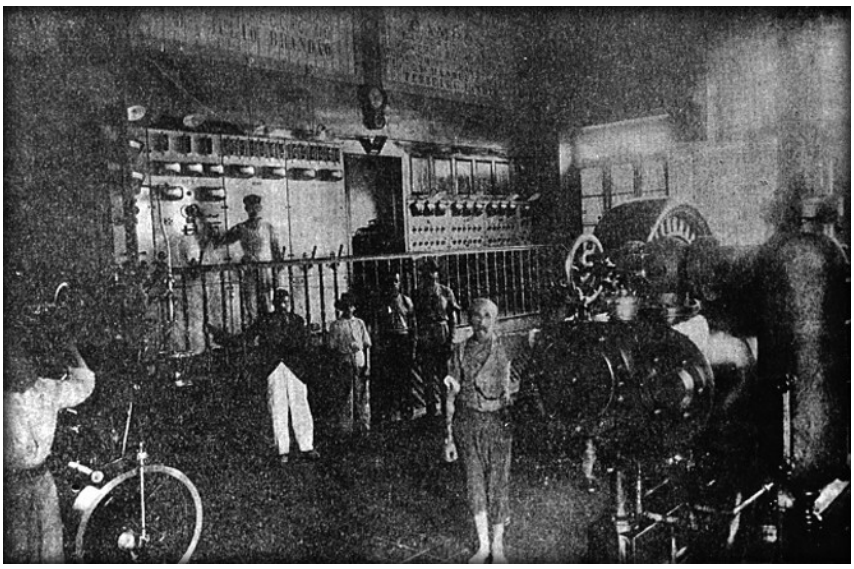
APÊNDICE B – ESPECIFICAÇÃO DAS ARQUITETURA DE RNAS . 301

APÊNDICE C – SELEÇÃO DE ENTRADAS DO CONJUNTO MISTO 303

1 INTRODUÇÃO

A atividade de previsão é parte integral do processo decisório na indústria. No caso específico da indústria de energia elétrica, esta dependência se torna ainda mais crucial, pois todo o planejamento e operação do sistema elétrico trata-se, essencialmente, de modular em tempo real a geração para atendimento da carga com as necessárias medidas para garantia de segurança (HONG *et al.*, 2020). Desde 1881, quando no mesmo ano se inaugurava a primeira central elétrica pública europeia fundada pelos irmãos Siemens e, no Brasil, a câmara municipal da cidade de Campos dos Goytacazes no Rio de Janeiro autorizava a substituição da iluminação pública a gás pela iluminação a energia elétrica (Figura 1) (MENDONÇA, 1993; LIMA, 1995) a previsão já se mostrava uma atividade necessária com a geração sendo programada com base na quantidade de lâmpadas que deveriam ser instaladas e acesas em determinados horários. Desde este momento, cidades e países se estruturaram para garantir fornecimento de energia, visto que a longo prazo foram percebidas evidências de que o acesso a eletricidade e o desenvolvimento econômico estavam correlacionados (MARTIN, 1992).

Figura 1 – Estação de força e luz em Campos dos Goytacazes no Rio de Janeiro com capacidade de 52 kV.



Fonte: Mendonça (1993).

Dado que decisões regulatórias sobre produção e uso de energia elétrica são descritas em políticas de estado, a previsão das grandezas relacionadas a eletricidade passa a ser então campo de estudo também da teoria econômica. Esta associação acontece devido ao fato de que o desenho de mercado de cada país embarca características regionais únicas que se adaptam a condições de clima, geografia e aspectos sociodemográficos. A pesquisa de Mason (1955) foi a primeira a constatar a existência de uma correlação positiva entre renda e consumo de energia *per capita* (JUNIOR, 2017). A partir deste estudo, diversos outros (MAINGUY, 1967; DARMSTADTER, 1971; JANOSI; GRAYSON, 1972; HANNESSON, 2009) se somaram corroborando esta importante condição.

Sendo assim, matrizes energéticas passaram a receber investimentos massivos ao longo da história por possuírem natureza estratégica. Como consequência, o crescimento desta área gradualmente fez com que a atividade de previsão de variáveis relacionadas a energia elétrica fosse caracterizada como um campo de estudo específico denominado de Previsão Energética (PE). A PE é, portanto, uma ferramenta não apenas relevante, como também heterogênea e complexa (HONG, 2014).

Métodos estatísticos e de aprendizado de máquinas voltados para análise e previsão de séries temporais são uma das principais ferramentas na previsão energética e, no contexto destes métodos, o uso de Redes Neurais Artificiais (RNAs) há três décadas vem sendo reportado com sucesso na literatura para variáveis como carga (KUO; HUANG, 2018a), geração convencional (DESHPANDE *et al.*, 2012; LOPES *et al.*, 2019), produção de fontes intermitentes (GENSLER *et al.*, 2016) e preços (MARCJASZ; LAGO; WERON, 2020). RNAs com arquiteturas do tipo *Sequence to Sequence* (Seq2Seq), células de memórias e com mais de duas camadas ganharam notoriedade nos últimos oito anos após a publicação das pesquisas de Bahdanau, Cho e Bengio (2015), Luong, Pham e Manning (2015) e Kim *et al.* (2017) que, no campo de processamento de linguagem natural, introduziram o conceito de mecanismos de atenção. Em dezembro de 2017, no artigo intitulado *Attention Is All You Need* (VASWANI *et al.*, 2017) de autoria de pesquisadores da Google, resultados para melhoramento na tradução automática de textos atingiram níveis de precisão até então inéditos. Este fato mostrou que os mecanismos de atenção são uma importante direção de pesquisa quando se fala de processamento de dados sequenciais. Este tipo de RNA é denominada de *Attentional Neural Network* (ANN) e, neste sentido, são aderentes à previsão em séries temporais, logo, consequentemente também a previsão de carga, geração, preço e outros subcampos da previsão energética.

Dentre estes subcampos possui destaque a Previsão de Preços de Eletricidade (PPE) ou como comumente referenciada na literatura científica, do inglês, *Electricity Price Forecasting* (EPF). A pesquisa em PPE é produto da implantação de mercados desregulados de eletricidade ao redor do mundo (WERON, 2014). As primeiras publicações datam do final da década de 90. Exatamente o momento em que os primeiros mercados de eletricidade (como os do Chile, Nova Zelândia e Inglaterra) foram estabelecidos (KAMINSKI, 1997; BASTIAN *et al.*, 1999; SZKUTA; SANABRIA; DILLON, 1999; YAO *et al.*, 2000). A modelagem e previsão de preços tornou-se essencial, dado que com a introdução da livre concorrência, novos ambientes de comercialização como bolsas de energia, balcões de leilão e câmaras de comercialização passaram a ser, na maioria das vezes, os únicos ambientes onde a negociação de eletricidade é possível e permitida. Atrelado a isso, ferramentas baseadas em RNAs (que podem ser aplicadas a PPE) transformaram-se em recursos para corretores de energia, comercializadoras, geradoras, distribuidoras e, em mercados com desenho mais avançado, até mesmo consumidores finais que já podem, inclusive, optar pelo seu próprio provedor.

Levando-se em consideração que a eletricidade é uma *commodity*, o processo de formação de preço deste insumo segue essencialmente a regra básica da teoria microeconômica que diz que o preço de um ativo-objeto em um mercado competitivo deve refletir a escassez da sua oferta para um

dado nível de demanda (MAYO, 2012). Portanto, em tese, o preço da energia elétrica deveria seguir de perto as variações da carga apresentando movimentos cíclicos intradiários e sazonais no médio e longo prazo. Entretanto, esta condição não é sempre observada nos mercados desregulados como estabelecidos hoje ao redor do mundo (ZAREIPOUR, 2012; HONG *et al.*, 2020). O preço da eletricidade sofre influência de diversos fatores externos que a tornam uma variável com alta volatilidade constatada no tempo. Dentre estes fatores de influência, muitos podem ser de natureza subjetiva como o impacto de decisões de agentes que exercem poder de mercado, regulações diversas que passam por alterações constantemente, perspectivas de investimentos futuros em capacidade vinculados a planejamentos realizados por grupos políticos, além de incentivos fiscais e tarifações muitas vezes decorrentes de posições ideológicas (TOLMASQUIM, 2015; VIANA, 2018). Adicionalmente, a própria correlação do preço com a carga cria uma condição natural de instabilidade devido as oscilações do clima, imprevisíveis no longo prazo. Visto que a eletricidade se trata de um produto estratégico e sem candidato à substituto, seria de se esperar que em um mercado totalmente livre o preço da mesma agregasse apenas características tangíveis que refletissem o seu verdadeiro valor. Dado que esta não é a situação reportada na literatura científica de PPE, então técnicas de previsão de preços tornam-se fundamentais àqueles agentes que pretendem aprimorar a sua metodologia de tomada de decisão quando comprando ou vendendo energia elétrica. Finalmente, a teoria de PPE toma por premissa que as ferramentas de previsão devem ser capazes de modelar até mesmo os fatores intangíveis, mesmo que em dadas situações estes sejam representados por modelos subjacentes que não possam ainda ser claramente entendidos ou explicados pela teoria econômica. Em outras palavras, espera-se que em casos de falta de disponibilidade de variáveis auxiliares, a série de preços em si incorpore informação suficiente para execução das atividades de modelagem computacional.

1.1 Justificativa da pesquisa

O Brasil é um campo fértil e atualmente convidativo a pesquisa em PPE. Isto se justifica pelos seguintes motivos:

a) O preço praticado no Brasil, chamado de Preço de Liquidação de Diferenças (PLD), em 2021 passou a ser divulgado diariamente em base horária. Antes era praticado um único valor vigente para toda uma semana. A alteração da granularidade modificou a forma como os agentes estabelecem seus contratos e liquidam suas posições a fim de evitar exposição ao PLD que, agora, tende a refletir de forma mais fiel o verdadeiro custo da operação do sistema.

b) O PLD é divulgado para 24 horas à frente, consequentemente, qualquer projeção para um horizonte futuro superior a 24 horas pode ser útil.

c) O Ambiente de Contratação Livre (ACL) (onde ocorre a negociação bilateral da eletricidade) vêm crescendo de forma exponencial quanto ao número de adeptos desde 2000. Em dezembro de 2020 um montante de 148,82 GW médios avaliados em R$ 11,28 bilhões foi liquidado por 10543 agentes. Isto representa 37,76% de toda a energia

transacionada no país. Dado que os requisitos para ingresso no ACL estão programados para serem extintos até 2026 é de se esperar que esse mercado cresça ainda mais (CCEE, 2020a).

d) O PLD é formado com base no Custo Marginal de Operação (CMO). O CMO é calculado via modelos matemáticos que consideram, principalmente, a otimização do custo futuro de armazenamento da água em detrimento do acionamento de usinas de custo mais alto como termoelétricas. Sendo as condições hidrológicas incertas, o PLD também tende a propagar prováveis erros de estimação incutidos no CMO. Isto é especialmente relevante, dado que com as mudanças climáticas podem ser observadas alterações nos regimes de chuvas no Brasil. O governo brasileiro vem desde 2010 investindo na diversificação da matriz energética nacional, em especial, com usinas eólicas (EPE, 2020b). A modelagem das fontes intermitentes é uma dificuldade adicional a ser incluída nos modelos de formação do preço e por isso pode demorar a ser considerada.

e) Com a base horária do PLD, existe um incentivo para que novas operações de mercado devam surgir nos próximos anos. Estas visando, principalmente, diminuir a exposição dos agentes no mercado de curto prazo. A comercialização de derivativos e instrumentos de *hedge*[1] estão entre estes prováveis produtos.

f) Levando em consideração o levantamento disponível no Capítulo 3, até onde se tem conhecimento nenhum estudo com foco na previsão do PLD horário no mercado brasileiro foi até então publicado em congressos ou periódicos indexados.

Com base nestes pontos, desenvolver pesquisa em PPE em território nacional se trata de analisar dados do mercado e de operação do sistema elétrico, selecionar variáveis representativas, eleger métodos de modelagem adequados e, fundamentado nestes estudos, desenvolver estratégias de comercialização que sejam baseadas na previsão do PLD. A previsão do PLD pode auxiliar na mitigação dos seguintes problemas atualmente observados no Sistema Elétrico Brasileiro (SEB):

a) A partir de 2008 diversos agentes do ACL começaram a adotar posturas alavancadas em suas negociações de energia executando operações à descoberto. Os agentes passaram a estimar mal valores futuros de PLD e, quando as perspectivas não se concretizaram, tiveram que lidar com valores muito mais altos do que esperavam. Estes erros levaram a processos judiciais que em abril de 2021 somaram uma dívida contabilizada de R$ 7,46 bilhões (CCEE, 2022c).

b) Com a lei de criação das cotas de garantia física, geradores apresentaram situações de sobrecontratação em momentos de baixo PLD e de subcontratação em momentos de alto PLD. Esta situação (que ainda existe) gera prejuízos aos agentes, visto que a

[1] *Hedge* no mercado financeiro trata-se de qualquer produto que tenha por objetivo proteger uma das partes negociadoras de eventuais riscos relacionados a alta variação de preços de um determinado ativo, falências, inadimplências, quebras contratuais e similares (MAYO, 2012).

liquidação contratual no mercado de curto prazo em muitas das vezes supera os 5% de margem normalmente adotada como proteção em contratos (VIANA, 2018).

c) A regulação no SEB vem sendo submetida recorrentemente a mudanças pontuais, mas que ao longo dos anos se acumularam e ainda estão se acumulando. Essas pequenas alterações são denominadas de microregulação. A microregulação é prejudicial, pois além de tornar a legislação complexa, cria efeitos difíceis de serem previstos pelos agentes (TOLMASQUIM, 2015; VIANA, 2018). Muitos destes efeitos, inclusive, não são considerados nos modelos de formação do PLD.

d) O mercado brasileiro possui diferentes sinais de preços e um deles é o PLD. No entanto, existem ainda pelo menos outros três, sendo os preços praticados pelas distribuidoras, os preços observados nos leilões e os preços na venda de cotas do Ambiente de Contratação Regulada (ACR). Esta situação afeta os agentes que precisam recorrer a fontes de informação diferentes antes de estabelecer valores contratuais. É mais uma camada de incerteza e risco que dificulta a eficiência alocativa, portanto afeta também o PLD.

Neste contexto, nesta tese é apresentada uma proposta de solução computacional capaz de auxiliar o agente comercializador brasileiro em suas negociações de energia elétrica. O produto de inovação tecnológica foi construído tendo como bases teóricas a economia da energia, conceitos em Sistemas Elétricos de Potência (SEP) e, principalmente, a aplicação de técnicas modernas de aprendizado de máquina, em especial, aquelas que englobam ANNs no mapeamento de sequências em séries temporais combinadas na forma de *ensemble* com métodos estatísticos. Valendo-se da necessidade de discutir uma temática tão abrangente, nos capítulos a seguir o leitor encontrará uma sucessão de tópicos que se conectam e, ao final, trazem luz à relevância e aos desafios relacionados a atividade de previsão, inicialmente nos mercados de eletricidade de forma geral. Na sequência é dado foco ao mercado brasileiro e suas particularidades. A narrativa continua na forma de uma revisão de literatura que cobre não somente trabalhos relacionados a PPE, como também todos os principais avanços nos subcampos da previsão energética com aplicações no Brasil e no mundo.

Em relação a este arcabouço teórico, vale destacar que durante o desenvolvimento desta pesquisa foi dada considerável atenção ao estudo bibliográfico do tema. Grande parte deste trabalho que o leitor tem em mãos reflete um grande esforço que foi, antes de tudo, empreendido no entendimento do impacto social e econômico na adoção de diferentes modelos de mercados de eletricidade ao redor do mundo. Buscou-se desenvolver uma reflexão crítica acerca da evolução histórica dos mesmos sob a ótica da experiência internacional e como os movimentos políticos relacionados ajudaram a moldar a estrutura na qual estamos inseridos hoje no Brasil. Isto foi considerado, visto que a formação do PLD é em essência um reflexo desta forma de estruturação. Executar a sua previsão sem um completo entendimento do que é esta variável e como ela impacta no mercado seria uma deficiência em um trabalho que se propõe a desenvolver uma ferramenta aplicada a um ambiente regido não somente por leis da física, mas também por princípios econômicos.

Já do ponto de vista da técnica, foi realizada uma investigação que relata os avanços da área de RNAs e suas últimas tendências, além de comparações com o estado da arte relacionado a métodos oriundos de outros campos, como da econometria e estatística. A década de 90 foi marcada por uma bolha de popularidade e promessas irrealistas que culminaram no declínio do uso de RNAs e num movimento ceticista em relação a sua aplicabilidade no meio acadêmico. Produziu-se uma análise sobre este histórico e uma argumentação sobre como após 2012 uma revolução chamada *Deep Learning* (DL) vem sendo aplicada em problemas antes considerados sem solução. O uso destes métodos e seus resultados na previsão energética e, principalmente em PPE, são explorados por meio de meta-análises de revisões bibliográficas, descrições de artigos seminais e relevantes, além de avaliações bibliométricas. O ponto principal nesta discussão são os mecanismos de atenção no contexto das ANNs. Mecanismos de atenção, por serem recentes e até então difundidos apenas junto ao processamento de linguagem natural, tem sido pouco estudados no contexto de séries temporais (CORREIA; COLOMBINI, 2022). Dentre 491 referências consultadas, apenas três fizeram uso de mecanismos de atenção em PPE, sendo uma delas publicada em 2019 (FAN *et al.*, 2019) e as outras duas em 2021 (MARSZAŁEK; BURCZYNSKI, 2021; MASHLAKOV *et al.*, 2021).

Por fim, apesar de nas páginas a seguir ser apresentado como destaque um método computacional, vale ressaltar que foi uma preocupação durante a pesquisa desenvolver um sólido levantamento bibliográfico, a catalogação da produção técnica, a organização do conhecimento e a construção de uma reflexão crítica sobre o campo de pesquisa. Este documento foi elaborado de forma a prover também à comunidade acadêmica uma fonte de referência em relação a temas relacionados aos mercados de eletricidade, principalmente o brasileiro.

1.2 Proposta

A ferramenta desenvolvida tem por finalidade a geração de cenários futuros que permitem ao usuário do sistema ter acesso a informações que visem balizar o processo de tomada de decisão em suas negociações. Neste sentido, fazem parte deste cenário futuro quatro componentes, sendo:

a) Previsão de 168 valores de PLD com início na primeira hora de segunda-feira até a última hora do domingo.

b) Determinação de um intervalo probabilístico que informa a expectativa de se observar o PLD previsto, hora a hora, neste dado intervalo.

c) Um estudo histórico de acurácia e rentabilidade da ferramenta frente a outros métodos.

d) Geração de uma análise de risco com base em métricas financeiras que levem em conta as previsões, o histórico dos valores do PLD e um método de precificação hipotético que incorpora dados reais de consumo do Sistema Integrado Nacional (SIN).

Os cenários são gerados numa dinâmica baseada em janelamento móvel semanal por meio de quatro módulos:

a) Módulo 1 – Banco de dados: Este módulo é responsável por ler arquivos providos por fontes públicas de dados *online*, interpretar as regras de texto em cada um deles, gerar uma instância de banco de dados relacional e inserir os dados dos arquivos de forma organizada neste banco.

b) Módulo 2 – Construtor de *datasets*: Este módulo é responsável por transformar os dados do banco de dados para formatos e representações computacionais interpretáveis pelos modelos quando treinados. São consideradas atividades relacionadas a tratamento de integridade, detecção e tratamento de *outliers*, estabilização de variância das séries temporais, normalização de distribuições, dimensionamento e codificação.

c) Módulo 3 – Construtor do previsor: Este módulo é responsável por mapear os conjuntos de dados de entrada e a série de saída a ser prevista. São produzidos diversos modelos distintos que prevem o PLD individualmente mas que ao final são acoplados na forma de um *ensemble*. O ensemble incorpora não somente os resultados das ANNs como também os gerados pelo preditor estatístico *Least Absolute Shrinkage and Selection Operator* (LASSO) numa forma otimizada denominada *LASSO Estimated Autoregressive* (LEAR).

d) Módulo 4 – Gerador de cenários: Este módulo determina a acurácia do modelo frente à outros preditores por meio de métricas de precisão, relaciona cenários históricos reais de consumo e preço dentro de um modelo de precificação hipotético e, por fim, estima rendimentos e possíveis perdas calculadas via metodologias de risco *Value at Risk* (VaR) e *Conditional Value at Risk* (cVaR).

Pretende-se disponibilizar no futuro no mercado real semanalmente aos agentes brasileiros por meio de uma plataforma *online* as informações geradas pela ferramenta. Neste sentido, foi criado em caráter embrionário o projeto denominado Observatório do Preço da Eletricidade do Brasil (OPEBRAS). Até o momento a marca foi registrada nas principais mídias sociais e uma página *web* provisória (OPEBRAS, 2022) foi criada. Nesta página o leitor pode consultar alguns gráficos interativos disponíveis nesta tese. Quando o gráfico possuir esta versão, a informação estará indicada na legenda. Para facilitar a visualização, adotou-se no texto o acrônimo mais curto OPB como rótulo para resultados e cenários gerados nos módulos 3 e 4, ou seja, OPB é a sigla de referência para o método desenvolvido.

1.2.1 Características da pesquisa

No que tange a metodologia científica esta pesquisa está caracterizada da seguinte forma:

a) Quanto a finalidade: É uma pesquisa aplicada, visto que propõe o estudo, desenvolvimento e utilização de um produto de inovação tecnológica em um mercado real.

b) Quanto ao objetivo: É uma pesquisa exploratória, visto que investiga condições, estruturas e variáveis relacionadas ao mercado de eletricidade brasileiro.

c) Quanto a abordagem: É uma pesquisa quantitativa, visto que todas as decisões de implementação são guiadas por métricas numéricas que expressam a qualidade de cada componente implementado. A comparação dos resultados do produto com *benchmarks* também é feita de forma numérica.

d) Quanto ao método: É uma pesquisa dedutiva, visto que emprega raciocínio lógico para realizar deduções que norteiam as conclusões a respeito de hipóteses previamente consideradas.

Neste sentido, a hipótese de pesquisa adotada neste trabalho foi:

> A qualidade da previsão para uma semana à frente do PLD horário depende do método de previsão adotado. Se o método adotado for um *ensemble* baseado em Redes Neurais Atencionais e LEAR então esta previsão torna-se mais acurada do que àquela realizada por outros métodos individualmente aplicados.

O termo "qualidade" na hipótese de pesquisa diz respeito à acurácia quantitativamente aferida por meio de métricas de precisão e métricas financeiras discutidas no Capítulo 4. Os "outros métodos individualmente aplicados" tratam-se dos *benchmarks* utilizados no Capítulo 6. A variável independente é o método, logo, a dependente é a previsão.

1.2.2 Objetivos

Por objetivo geral, como já exposto, pretendeu-se com este estudo desenvolver uma ferramenta computacional capaz de gerar cenários futuros que auxiliem o comercializador de energia elétrica nas suas tomadas de decisão. Deste propósito se originaram as seguintes questões de interesse, indo da mais abrangente para a mais específica:

a) O que é um mercado de energia elétrica?

b) Qual o papel da previsão energética?

c) Qual a relevância da previsão energética no contexto dos mercados de energia elétrica?

d) Quais são as características do mercado livre de energia elétrica no Brasil?

e) Quais variáveis influenciam os movimentos do PLD em base horária?

f) A previsão do PLD em base horária é possível de ser realizada via técnicas baseadas em aprendizado de máquina supervisionado?

g) Mecanismos de atenção contribuem para o melhoramento da acurácia na previsão de séries temporais de preço de eletricidade?

h) *Ensembles* de ANNs com métodos estatísticos pode superar a previsão de outras técnicas isoladamente aplicadas?

Visando responder estas perguntas, foram objetivos específicos deste estudo:

a) Revisar a literatura referente a mercados de energia elétrica e suas principais formas de estruturação adotadas historicamente e globalmente.

b) Revisar a literatura e a legislação que organiza e promove o estabelecimento de um mercado livre de energia elétrica no Brasil.

c) Revisar a literatura referente a previsão energética e, principalmente, aquela relacionada a PPE com o objetivo de estabelecer o atual estado da arte.

d) Identificar, analisar e avaliar trabalhos que fizeram uso de mecanismos de atenção aplicados a PPE, principalmente aqueles que porventura tenham utilizado como estudo de caso o Brasil.

e) Identificar, analisar e avaliar variáveis que influenciam os movimentos do PLD em base horária.

f) Avaliar e adaptar diferentes arquiteturas de ANNs baseadas em modelos mapeadores de sequências Seq2Seq com mecanismos de atenção para séries temporais.

g) Identificar, analisar e avaliar o uso de técnicas suplementares visando melhorar a acurácia das previsões como o uso de tratamento de dados, arquiteturas alternativas de RNAs, métodos estatísticos para previsão em séries temporais, *ensembles* e outros.

h) Eleger a arquitetura final do modelo e comparar seus resultados com *benchmarks* considerados como atual estado da arte em PPE.

1.2.3 Contribuições

As contribuições deste trabalho podem ser encaradas de três perspectivas.

1. Da perspectiva do mercado: Foi produzida uma metodologia reprodutível que pode ser utilizada por corretores de energia, comercializadoras, geradores e quaisquer outros agentes interessados em melhorar sua forma de negociação. Com base nisto, são providos como benefícios à sociedade um modelo capaz de auxiliar diretamente à eficiência das operações, dado que os agentes podem utilizar os resultados da ferramenta para se programar melhor e estabelecer contratos possivelmente mais lucrativos e realistas.

2. Da perspectiva da técnica: Foi produzida uma avaliação do uso de mecanismos de atenção no contexto da previsão de séries temporais com ANNs. Este é um estudo inovador, visto que ainda não existem muitos trabalhos que adaptaram ANNs para prever dados sequenciais. Em especial, de forma acoplada ao LEAR por meio de um *ensemble*.

3. Da perspectiva da aplicação: O estudo traz um avanço para a área de PPE ao prover uma análise detalhada de como arquiteturas Seq2Seq com atenção podem ser implementadas para prever o preço da eletricidade utilizando-se de diversas variáveis, inclusive àquelas exógenas ao sistema elétrico incluindo as de mercado e clima. Além disso os resultados destes modelos foram comparados ao estado da arte atualmente estabelecido na literatura.

Finalmente, como estudos indiretos (atualmente em processo de revisão por pares ou para serem publicados no futuro) pode-se mencionar uma análise de dados de preços e seus direcionadores no mercado nacional, a avaliação e seleção de estações meteorológicas via uso de *Autoencoder* (AE) e o impacto na previsão de preços, a previsão da margem de reserva e seu relacionamento com picos de preços, a estimação das curvas de densidade probabilísticas com base no preço horário e a produção de um banco de dados original que pode servir de *baseline* para novos projetos de pesquisa em PPE no Brasil.

1.3 Estrutura do documento

Esta tese está estruturada em sete capítulos, uma seção de referências, três apêndices, uma página *web* (OPEBRAS, 2022) e um repositório *online* com materiais suplementares (NAMETALA, 2022). A descrição do conteúdo de cada um dos capítulos seguintes à esta introdução está feita de forma detalhada logo no início de cada um deles. Sucintamente serão abordados na sequência do texto os seguintes assuntos.

No Capítulo 2 é feita uma contextualização que abrange a grande área de conhecimento da pesquisa chamada de economia da energia e como o estudo da *commodity* da eletricidade está inserido na mesma; o histórico dos mercados de eletricidade e as suas principais formas de estruturação e; o histórico, estrutura organizacional e detalhes de funcionamento do mercado de energia elétrica no Brasil.

No Capítulo 3 é apresentada uma revisão sistemática de literatura que está dividida em três partes, sendo que a primeira trata da área de DL com enfoque principal nas arquiteturas voltadas para mapeamento de sequências e mecanismos de atenção; a segunda aborda o campo da previsão energética de forma geral e; a terceira apresenta uma investigação sobre o atual estado da arte em PPE. São especulados também ao final deste capítulo quais ações podem contribuir para o avanço da área de pesquisa de PPE.

No Capítulo 4 são apresentados os fundamentos teóricos necessários ao entendimento das principais técnicas empregadas na seção de metodologia. Neste capítulo são abordados conceitos em séries temporais, uma caracterização estatística da série de preço de energia elétrica e métricas de avaliação em PPE, o arcabouço matemático que suporta os modelos neurais profundos que fazem uso de recorrência temporal com mecanismos de atenção, conceitos relacionados ao método estatístico LEAR que incorpora o *ensemble*, uma descrição dos *benchmarks* adotados e, por fim, uma introdução aos modelos de gestão de risco VaR e cVaR.

No Capítulo 5 é apresentada a metodologia. O texto se organiza na mesma sequência em que os módulos do sistema são acionados e processam os dados de forma a criar um fluxo lógico sobre como e quando cada técnica é aplicada no contexto da ferramenta.

No Capítulo 6 são apresentados e discutidos os resultados. Inicialmente o protocolo experimental é exposto por meio de informações relacionadas a configurações e detalhes sobre o período no qual o modelo foi testado. Na sequência é feita uma discussão sobre o tratamento aplicado as variáveis climáticas e, por fim, a análise da qualidade da previsão do PLD é propriamente mostrada. Esta é dividida em duas seções, sendo que a primeira aborda uma avaliação em termos de acurácia e, a segunda, em termos financeiros.

No Capítulo 7 encontram-se as conclusões. São apresentadas também algumas sugestões de continuidade para uso em trabalhos futuros, as limitações identificadas e a produção científica desenvolvida ao longo da pesquisa.

Finalmente, nos apêndices, na página *web* e no repositório *online* do projeto estão disponíveis artefatos documentais, conjuntos de dados, análises gráficas extras e, futuramente de forma exclusiva nos materiais *online*, possíveis novas produções como atualizações de projeto, artigos e códigos-fonte.

2 CONTEXTUALIZAÇÃO

Neste capítulo é apresentada uma contextualização acerca do ambiente no qual se aplica o objeto de estudo da pesquisa. Iniciando-se pela grande área da economia da energia, serão apresentados conceitos fundamentais que visam munir o leitor sobre a importância das fontes energéticas e o papel econômico que estas desempenham. Na sequência, com foco apenas na fonte de eletricidade, é apresentado um histórico e uma breve análise crítica sobre as correntes monopolistas e liberalistas, hoje, as duas principais abordagens adotadas mundialmente para negociação. A seção é complementada por uma descrição da Teoria de *Design* de Mercados, assunto que trata de um esforço de diversos autores em compilar as possíveis organizações regulatórias e operativas de mercados de eletricidade ao redor do mundo. Ao final, os conceitos tratados são analisados especificadamente sob a ótica do mercado de eletricidade brasileiro. Isto é feito levando-se em conta seus principais momentos históricos, sua atual estrutura institucional e detalhes de funcionamento.

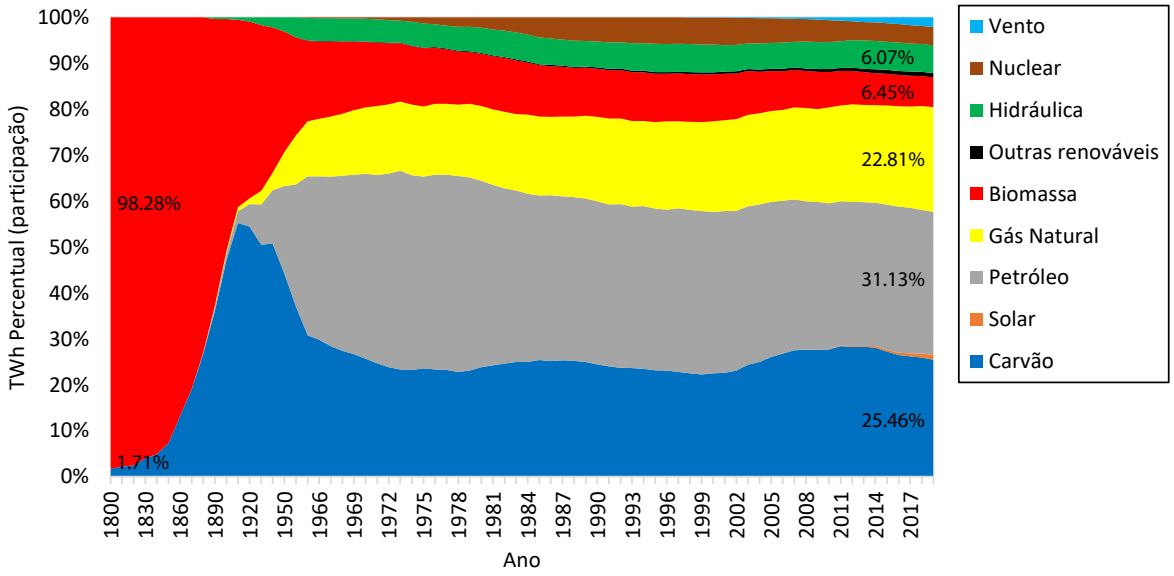
2.1 Economia da energia

O advento das máquinas a vapor no final do século XIX fez com que a sociedade se reorganizasse em torno de uma nova concepção de estrutura industrial. Essa invenção permitiu que ambientes de oficinas e salas de trabalho domiciliares se transformassem em usinas e fábricas. As primeiras máquinas a vapor eram ineficientes, sendo capazes de converter em trabalho apenas 1% do combustível consumido, no entanto, essa ineficiência pouco foi considerada. Isto ocorreu, pois estes mecanismos eram movidos a carvão mineral, até então, um combustível pouco explorado e abundante (JUNIOR, 2017; LANDES, 2005). De fato, conforme apresentado na Figura 2, a participação percentual de consumo do carvão mineral ganhou grande impulso a partir de 1900. Esse momento foi o primeiro grande marco de transição energética na história, migrando-se o que era praticamente 100% do consumo baseado em biomassa para o carvão. Esta situação despertou na sociedade a percepção de que novas fontes de energia poderiam ser exploradas com sucesso e assim, a partir de 1920, gás natural, derivados de petróleo, energia hidráulica e outras passaram a ser utilizadas em escala.

Globalmente, a indústria sempre foi o setor com maior participação no consumo de energia (IEA, 2022). Esta participação predominante frente a outros setores como residencial, comercial e transportes [1] (como visto na Figura 3) foi um dos principais motivos que, ao longo da história, promoveu a necessidade de diversificação da matriz energética na maioria dos países. Grande parte da energia na indústria é oriunda de combustíveis utilizados como fontes primárias e, apesar da energia elétrica desempenhar um papel importante neste contexto, ao contrário dos setores comercial e residencial que consomem de forma majoritária primariamente a energia elétrica, na indústria ela é

[1] Vale destacar que em países com malhas rodoviárias desenvolvidas, o setor de transportes possui tanta representatividade quanto a indústria. Pode-se destacar, por exemplo, o Brasil que em 2019 pela primeira vez na história registrou participação maior do setor de transportes (33%) do que da indústria (30%) (EPE, 2020a).

Figura 2 – Participação percentual de fontes primárias de energia no mundo.



Fonte: Adaptado de Ritchie, Roser e Rosado (2022).

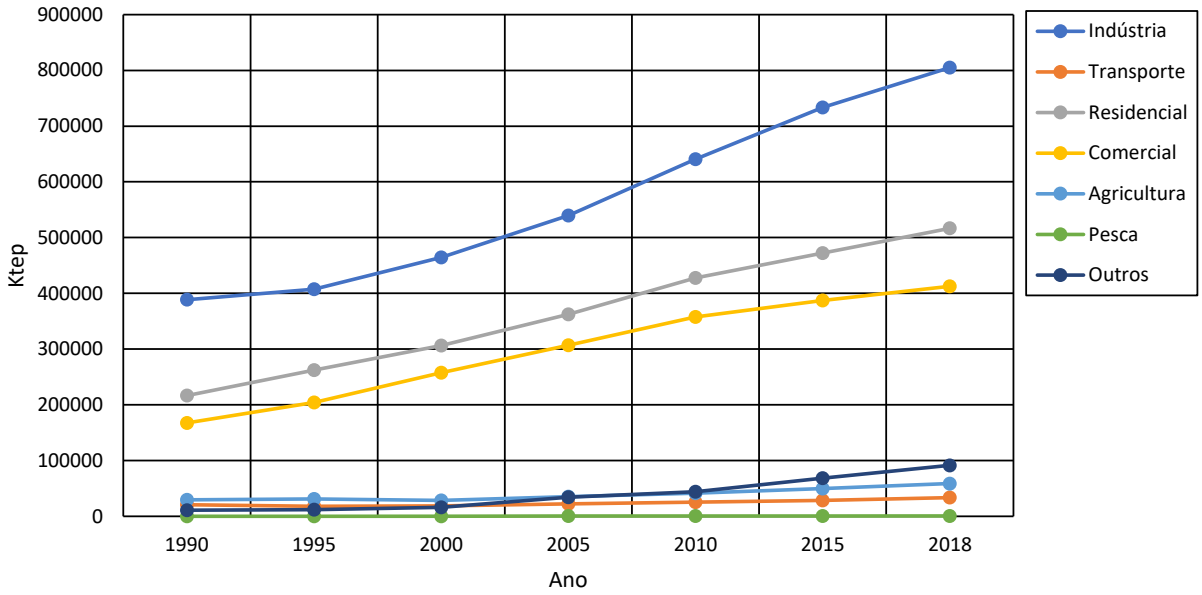
apenas uma parte do que move o setor. Como pode ser observado na Figura 4, a previsão para o consumo global industrial até 2050 (Statista Agency, 2022) ainda terá representatividade minoritária de energia elétrica. No entanto, a participação das várias fontes mostra-se equilibrada, sendo que em 2020 foi registrada como 26,3%, 26%, 23,3%, 15,5% e 8,7% respectivamente para combustíveis líquidos, carvão, gás, eletricidade e renováveis. Até 2050 a expectativa é de que a haja crescimento na adoção da eletricidade, sendo que esta será responsável por 16,2% do consumo da indústria.

Já nos setores de comércio [2] e residencial, a eletricidade é a forma base de consumo. No entanto, antes de tudo, a mesma só estará disponível após a execução de todas as etapas que compõe a cadeia da indústria de transformação. Esta garantirá a conversão de uma dada fonte primária para a energia elétrica que é secundária. Portanto, mesmo que se utilize diretamente a eletricidade, a diversificação de fontes primárias continua sendo fundamental. Neste sentido, a cadeia completa da energia caracteriza-se por ser bastante heterogênea, sendo dessa forma necessária a existência de uma área de estudo que abranja conhecimentos que tratem da exploração das mais variadas formas de combustíveis, sejam eles convencionais e/ou renováveis. Esta área de estudo é a economia da energia.

A economia da energia é descrita por Martin (1992) como tecnologias e *commodities* que devem ser distintamente tratadas de acordo com sua natureza física ou não física. Dentre as fontes não físicas está a eletricidade. Já a definição de Sweeney (2001) descreve a economia da energia como a área que estuda *commodities* energéticas e sua relação com o fornecimento, conversão,

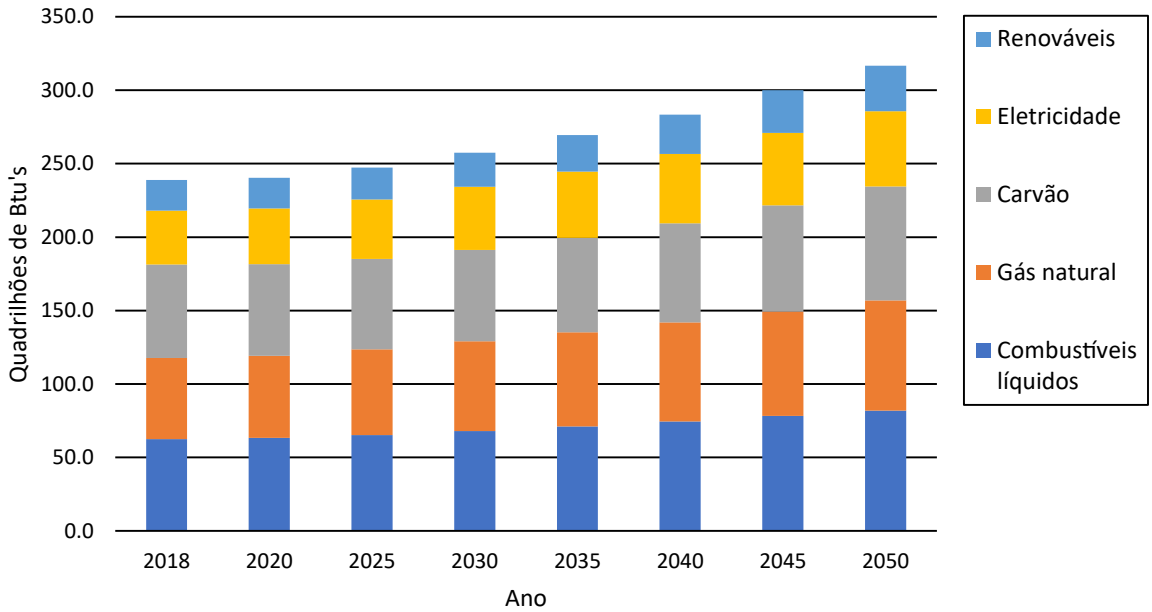
[2] O setor de comércio, além de *shoppings*, escritórios, supermercados e demais fornecedores de serviços e bens, engloba adicionalmente os serviços públicos, como escolas e hospitais.

Figura 3 – Consumo mundial total de energia por setor.



Fonte: Adaptado de IEA (2022).

Figura 4 – Previsão de consumo global industrial por fonte.

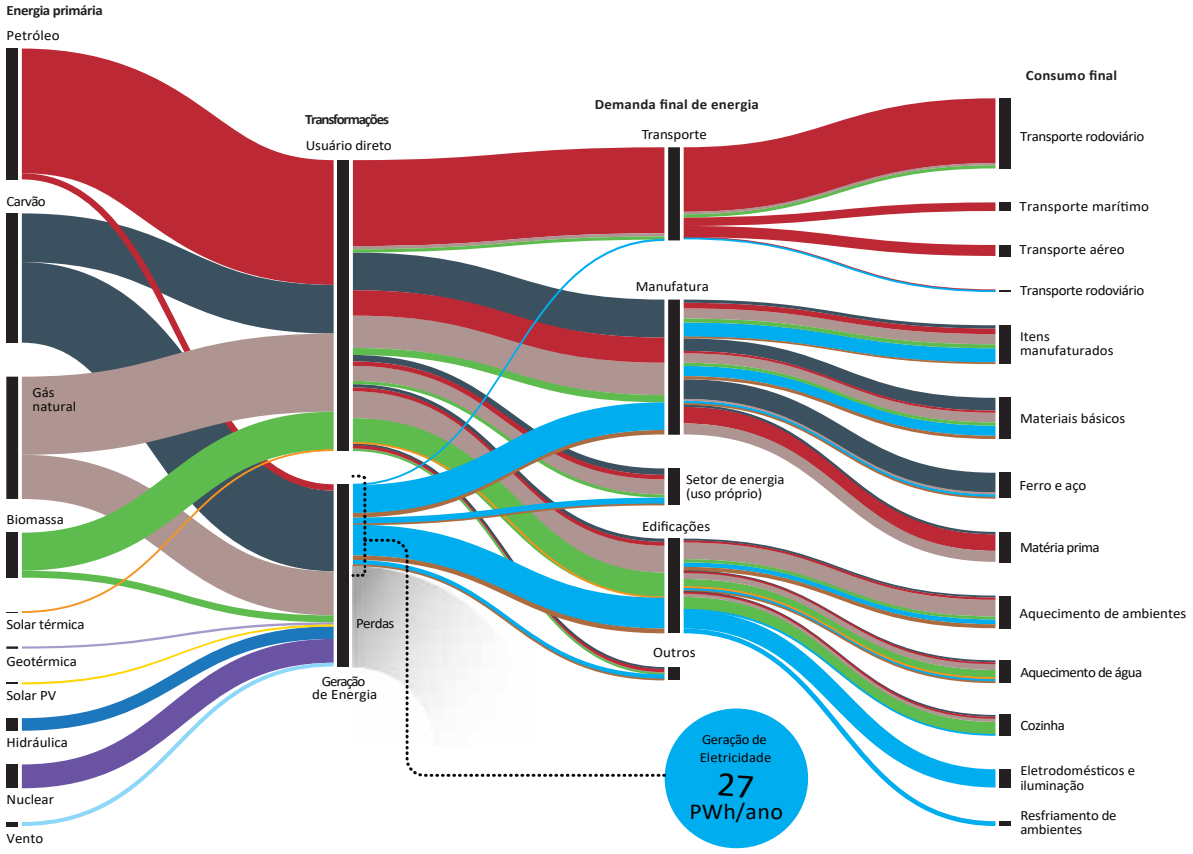


Fonte: Adaptado de Statista Agency (2022).

transporte, consumo e eliminação de resíduos, além disso, trata também das estruturas de mercado, regulações, os impactos ambientais e o uso economicamente eficiente. Por fim, Junior (2017) descreve economia da energia como sendo um "ramo da economia aplicada que tenta conjugar a análise econômica com as dimensões técnica e político-institucional que abarcam os problemas energéticos".

Todos os autores destas definições preocuparam-se em destacar a importância do tratamento das fontes como *commodities* e, nesse sentido, essa escolha mostra-se válida visto que estas são utilizadas como base para uma ampla gama de serviços e produção de bens. Na Figura 5 pode ser observada um panorama do fluxo energético da matriz global em 2018, resultado de uma pesquisa desenvolvida pela agência internacional de risco *Det Norske Veritas Group* e reportado em 2020. Como observado, a cadeia de produção existente entre uma dada fonte e o seu consumo final estão relacionadas, no entanto, possuem características técnicas diferentes entre si. Junior (2017) comenta que, apesar da aparente estrutura heterogênea que a princípio faria com que petróleo, carvão e gás fossem alocados em estudos sobre o setor extrativista, a biomassa em agricultura e, por exemplo, a eletricidade em telecomunicações, do ponto de vista da física o conhecimento técnico de cada uma destas coloca todas dentro da área de estudos energéticos. Sendo estas reguladas por fatores econômicos, então ao mesmo tempo, dentro também da área de economia. Nesse sentido, em economia da energia tanto o conhecimento técnico da cadeia energética de uma fonte, quanto o conhecimento econômico que regula a disponibilidade da mesma tornam-se igualmente importantes.

Figura 5 – Estudo do fluxo energético da matriz global em 2018.



Fonte: Adaptado de DNVGL (2020).

Dentre todas essas possibilidades de procedência, a eletricidade é a que desempenha papel mais singular, pois sendo dependente das condições de produção, transporte, distribuição e comercialização da fonte usada para gerá-la, a mesma acaba por possuir um mercado em dois níveis. O primeiro se trata do custo do seu combustível e o segundo, da sua tecnologia de geração. Neste contexto, a eletricidade é um bem diretamente influenciado pelo problema do equilíbrio entre oferta e demanda, ou seja, é por natureza um produto negociável entre partes a um dado preço. Além disso, conta ainda com mais duas particularidades, sendo (i) distribuída de forma desigual entre países, visto que é altamente dependente da disponibilidade de recursos naturais e, (ii) não existe país algum que tenha demonstrado desenvolvimento econômico sem antes prover à sua população o acesso a eletricidade (HANNESSON, 2009). Em consequência destas duas características, a ação do estado passa a ser legitimada, sendo papel dos governos elaborar políticas que reduzam as desigualdades entre regiões, com incentivos ao uso da energia de forma adaptada a cada perfil de consumidor.

O estudo da eletricidade no âmbito da economia da energia, portanto, é o campo que busca investigar os vínculos existentes entre o caráter físico da geração, transmissão, distribuição e consumo, juntamente com as leis de mercado que regem o setor. Desta perspectiva, o entendimento somente técnico da cadeia não é suficiente, sendo necessário considerar sob seu espectro também os fatores econômicos (MAYO, 2012). Alguns desses fatores são elencados a seguir segundo Junior (2017):

a) Na dimensão macroeconômica: Oferta e demanda de energia estão relacionadas a adoção de políticas de crescimento econômico, arrecadação de tributos e sustentabilidade.

b) Na dimensão microeconômica:

   – Decisões de investimentos, financiamento, consumo e formação de preços são definidos com base em critérios normalmente diferentes entre regiões.

   – O estado tem papel de definir políticas regulatórias, regime fiscal e manutenção/privatização de empresas do setor.

c) Na dimensão de políticas internacionais: A implantação e exploração da interconexão de infraestrutura física entre países e regiões sofre influência de condições geopolíticas e econômicas.

d) Na dimensão tecnológica: A concorrência existirá dado que estratégias empresariais e inovações tecnológicas são parte intrínseca da indústria energética.

e) Na dimensão ambiental: Não existe nenhuma fonte de energia que não cause algum tipo de impacto ambiental. Sendo assim, políticas de incentivo ou restrições são praticadas.

Respaldados por estes atributos, diversos governos ao redor do mundo buscaram estruturar seus setores energéticos por meio de mercados. Alguns desses adotaram modelos monopolistas, outros modelos competitivos e outros modelos mistos. Independente da corrente adotada, com o objetivo de garantir segurança de abastecimento e uso racional dos recursos naturais, em algum nível,

mesmo as nações mais economicamente liberais mantém alguma forma de aprovisionamento energético (MARTIN, 1992; JUNIOR, 2017). Neste escopo e conforme modelo adotado em cada país, os mercados de eletricidade desempenham uma função de significativa importância permitindo que a eletricidade seja transacionada entre os agentes geradores, distribuidores e consumidores. Todavia, transações em mercados de eletricidade possuem consideráveis diferenças para as formas de transações envolvendo outras *commodities*. Em razão da incapacidade de estocar energia de forma financeiramente viável e, principalmente, da necessidade de garantir simultaneidade [3] e instantaneidade [4] no intercâmbio da mesma, mercados de eletricidade foram pensados ao redor do mundo para incorporarem complexas regulações que tratam do despacho físico da energia, como também, da contabilização e liquidação financeira de seu uso. Assim, torna-se perceptível a ligação existente entre os fundamentos de SEP e a teoria econômica, assunto da qual trata a economia da energia.

Com base na experiência internacional, na subseção a seguir são apresentados conceitos sobre mercados de eletricidade, com destaque para as principais formas de organização.

2.2 Mercados de eletricidade

O uso do preço como instrumento de orientação aos agentes envolvidos na indústria de energia elétrica iniciou-se na década de 1990 [5]. Este movimento das nações em direção a um mercado competitivo não foi exclusividade dos mercados de eletricidade, na verdade, antes destes vieram os mercados de telecomunicações, petróleo e vários outros segmentos tratados no âmbito da economia da energia (SILVA, 2012). Essa corrente tomou impulso em decorrência de sinais de esgotamento (como falta de investimento em expansão da capacidade) demonstrados pelos modelos de monopólio estatal que vigoraram por 60 anos, desde 1930. A maioria dos países procurava uma solução para garantir sobrecapacidade e expansão da matriz energética a longo prazo, objetivos que só poderiam ser atingidos por meio de grandes investimentos. O caminho foi a abertura para o capital privado, uma vez que diversas crises financeiras afetaram a situação de caixa de muitos governos (JUNIOR, 2017).

A implementação das reformas no setor elétrico ao longo da história dá espaço à diversas reflexões, especialmente quanto a discussão sobre os desafios e contribuições de modelos monopolistas *versus* modelos liberalistas. Conhecer os principais estágios que fizeram parte do histórico da comercialização de energia ao redor do mundo contribui para o entendimento dos atuais desenhos de mercado modernos que hoje têm tendido a ambientes competitivos. Nesse sentido, a seguir é apresentado um resumo da experiência internacional no setor levando-se em conta, principalmente,

[3] Trata-se da relação temporal simultânea entre a geração de eletricidade e o seu consumo.

[4] Consequências de eventos na geração afetam instantaneamente a utilização de eletricidade.

[5] A década de 90 é citada na maioria das bibliografias consultadas como o início do estabelecimento dos mercados de eletricidade competitivos. Esta situação se deve a massiva quantidade de reformas que aconteceram neste período, iniciadas por Reino Unido e Noruega em 1990. No entanto, vale destacar que foi o Chile em 1982 (SILVA, 2012) que executou a primeira reestruturação de um setor elétrico no mundo, fora da década citada.

as obras de Junior (2017) e Stoft (2002). Estes autores possuem distintas visões acerca de mercados de eletricidade. O primeiro discute o papel da competitividade, mas no geral, argumenta em defesa da existência de monopólio a fim de manter a qualidade das relações que regem o setor. Já o segundo, defende a desregulamentação como caminho para eficiência e queda de preços.

2.2.1 Evolução dos mercados de eletricidade

Os primeiros passos para criação de um modelo de mercado no qual a eletricidade era o bem a ser negociado se deram após Thomas Edison, juntamente com uma equipe de 64 especialistas, testarem 6000 tipos de filamentos e, dentre todos, encontrarem um que fosse o princípio para a criação da lâmpada incandescente. Em 1879, um total de 500 lâmpadas iluminaram a cidade de Menlo Park na Califórnia. Dez dias antes Edison registrava a primeira patente de uma lâmpada (JUNIOR, 2017; TEECE, 1988).

A indústria da eletricidade nasce por meio de uma inovação tecnológica e, encaminha por consequência, o surgimento de diversas outras inovações. Graças a possibilidade de entrega da energia ser feita à distância, pela primeira vez na história as máquinas poderiam estar em qualquer lugar desde que conectadas a rede. Invenções como bonde elétrico, transformador e turbinas, em 10 anos apenas após a criação da primeira lâmpada já demonstravam que a indústria da energia revolucionaria a forma como o trabalho seria desempenhado dali em diante (JUNIOR, 2017).

A última década do século XIX foi marcada pela primeira concorrência na indústria de energia. Edison que defendia um modelo descentralizado de rede, promulgava a necessidade de implantação de sistemas com transporte em corrente contínua em baixas tensões. Já George Westinghouse e Nikola Tesla advogavam em prol de um sistema de geração centralizada, com transporte em longas distâncias por meio de corrente alternada em altas tensões. Neste contexto surgem a General Eletric e a Westinghouse. Empresas que com a venda de motores tornaram-se as duas maiores corporações do século XX. Ainda neste período, com o sucesso da implantação das primeiras linhas de transmissão de alta tensão na Europa e nos Estados Unidos da América (EUA), o modelo de corrente alternada mostrou-se mais conveniente frente ao seu concorrente e, no seu encalço, ficou demonstrada a viabilidade dos parques elétricos baseados em torres, cabos e outros meios de transmissão mecânicos.

Com o estabelecimento da eletricidade como um produto, surgiram as primeiras iniciativas no intuito de precificar a mesma. Essa precificação se deu de forma a maximizar os fatores de carga [6] e de diversidade [7] que, ao serem explorados, permitiam as empresas de energia aumentarem seus lucros com o mínimo investimento em capacidade. Essa abordagem, que visa ampliar o número de clientes para obtenção de mais rendimento levou a necessidade de padronização das características

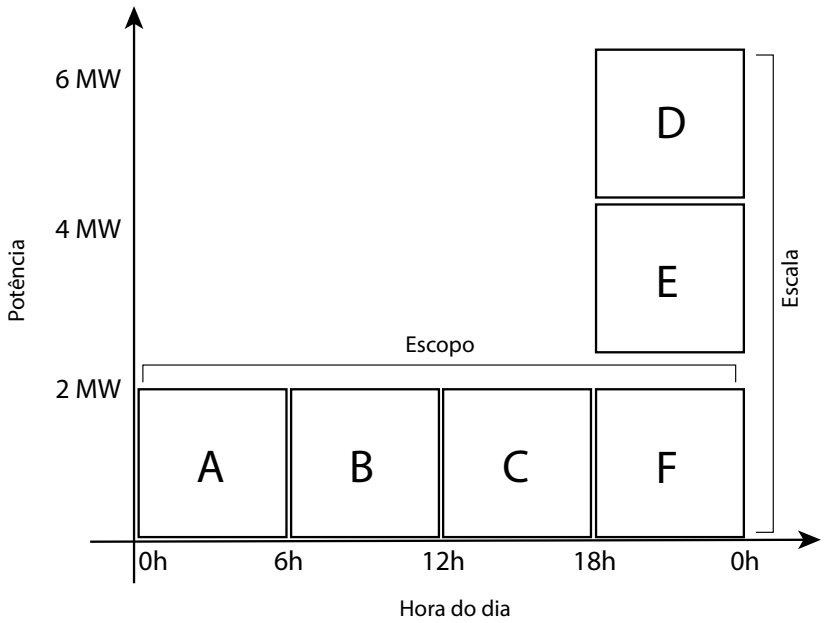
[6] Fator de carga diz respeito a razão da carga média pela capacidade instalada, em um dado período de tempo.

[7] Fator de diversidade trata da composição de um portfólio de clientes que possuam carga máxima passível de distribuição ao longo de diferentes momentos do dia, promovendo assim o atendimento da carga máxima sem a necessidade de aumento da capacidade instalada.

da energia elétrica. Assim, nos anos 20 a General Eletric e a Westinghouse passaram a produzir máquinas de lavar (1907), ar-condicionado (1911), geladeiras (1913) e outros itens domésticos com especificações técnicas padronizadas (CHANDLER, 1962; JUNIOR, 2017).

Outras duas características da indústria de energia foram importantes para a estruturação nos primórdios do setor. A reunião de consumidores com o mesmo perfil de consumo nos mesmos horários permitia manter a continuidade e regularidade da geração. Esta característica foi chamada de economia de escala e trata, portanto, de manter uma base de clientes organizada para se obter uma homogeneidade temporal. A outra é a chamada economia de escopo. Nesta os consumidores são distribuídos ao longo do dia em horários de atendimento diferentes, permitindo assim maximizar o fator de carga. Na Figura 6 é mostrada uma representação visual destes dois conceitos com cada bloco representando um cliente (JUNIOR, 2017).

Figura 6 – Economias de escopo e escala.



Fonte: Adaptado de Junior (2017).

Um dos pioneiros na exploração dos conceitos de economia de escala e escopo foi o empresário americano Samuel Insull. Entre 1903 e 1911, Insull fundou a Commonwealth Edison adquirindo 20 empresas de eletricidade nos EUA, pois percebeu que se o mercado de energia fosse fragmentado entre vários concorrentes, os investimentos em geração não valeriam o custo dada a necessidade de um sistema de transmissão e equipamentos de *backup*. Possuindo o domínio de mercado em toda a região de Chicago, a empresa de Insull conseguiu comercializar maiores quantidades de eletricidade praticando tarifas menores. Assim, visualizando uma oportunidade de negócio, o empresário sugeriu a diversos estados americanos a adoção de uma regulação governamental com o objetivo de dar exclusividade de exploração do mercado em uma dada região a uma única companhia. A sugestão foi levada em consideração pelos estados de Nova

York, Massachusetts e Wisconsin que organizaram as primeiras comissões estatais para fiscalizar a atuação das empresas de energia (MCDONALD, 1958; STOFT, 2002).

Na sequência, entre 1920 e 1940 observa-se ao redor do mundo a entrada em operação de novas linhas de transmissão que receberam nesse período grandes investimentos. O principal atributo desse empreendimento foi o aumento do nível de tensão. Isso permitiu reduzir de forma significativa as perdas e o comprimento das linhas que ultrapassaram os 300 quilômetros. Associado aos fatos do surgimento de novos usos para a energia e a construção de maiores usinas de geração, o preço do kWh caiu 65% em relação ao seu primeiro valor negociado em 1890 (JUNIOR, 2017). Neste contexto, levando-se em conta as evidências de que o preço poderia ser reduzido desde que o sistema tivesse apenas um único detentor, foram dados os primeiros passos em direção à consolidação dos modelos de mercado baseados em monopólio.

Em 1935 os EUA publicaram o *Public Utility Holding Company Act* (PUHCA) (EUA, 1935). Esta lei federal fez parte do plano de reestruturação *New Deal* e dava às comissões de governo estaduais a autoridade para desmembrar, se necessário, empresas do setor de eletricidade e, dessa forma, atuar regulando e fornecendo licenças de operação em áreas geográficas restritas. Nesse sentido, o monopólio da geração, transmissão e distribuição foi forçado via legislação, mas paradoxalmente, o direito dessa exploração foi cedido majoritariamente a empresas privadas [8]. Enquanto isso na Europa, especialmente após a segunda guerra mundial, a visão de domínio estatal sobre o setor energético também ganhava força influenciada não somente pelo cenário americano, mas adicionalmente, a obra de John Maynard Keynes neste mesmo período passou a exercer grande influência no pensamento político e econômico da época. Em contraposição ao pensamento de Adam Smith que era um popular defensor de ideias liberais, Keynes defendia a ampliação da intervenção reguladora em todo o sistema econômico (GOMES, 1998). Sua argumentação baseava-se na premissa de que a sociedade teria maior qualidade de serviços públicos, somente se, o controle desses serviços fosse agregado para permitir otimização de gerenciamento. Em sua visão, essa ação promoveria o aumento do nível de renda e uma maior quantidade de empregos (GOMES, 1998). A maioria dos países europeus acabou por instituir modelos monopolistas, mas ao contrário dos EUA que cedeu direitos as empresas privadas, na Europa essa atividade foi delegada a instituições de propriedade estatal. Finalmente, os modelos norte-americano e europeu tornaram-se a referência a ser seguida por outros países ao redor do mundo.

A experiência internacional de promoção do aprovisionamento energético se estendeu para outros segmentos com traços semelhantes aos da indústria de eletricidade. No caso, toda e qualquer indústria que operava sobre uma infraestrutura de rede. O conceito de indústria de rede trata da conexão entre um fornecedor e seus usuários através de um gerenciamento coordenado de todas as etapas existentes entre eles (MCGEE; SAMMUT-BONNICI, 2002). São exemplos destas indústrias os setores de gás, telecomunicações, água e rede ferroviária. Indústrias de rede, apesar

[8] Esse conceito ficou conhecido como *Public Utilities* e diz respeito às organizações privadas que tem o poder de fornecer serviços de interesse público sob uma regulação governamental (CLIFTON; LANTHIER; SCHRÖTER, 2011).

de diferentes, segundo Junior (2017) e Gomes (1998), possuem elementos econômicos em comum como (i) impactar na operação de outros segmentos da indústria e comércio, (ii) se expandirem por meio da economia de escala e (iii) se organizarem em torno dos meios de acoplamento entre os agentes envolvidos [9]. Sendo estes atributos elementos de interesse público, indústrias de rede passaram a ser consideradas monopólios naturais e assim ficaram até os anos 80. Desse momento em diante, o modelo começou a dar sinais de falência que não puderam mais ser ignorados e, como consequência, diversos governos iniciaram processos de reestruturação das suas indústrias de rede, incluindo a de eletricidade.

Liderada pelos EUA e países europeus capitalistas, a indústria de energia ao redor do mundo passou por um processo de "super expansão sincronizada" que durou de 1930 até o final da década de 70. São características do setor elétrico neste momento da história, o volume elevado de investimentos e a capacidade ociosa não planejada (COUTINHO; FERRAZ, 1994). Em 1973 e 1979, com a primeira e segunda crise do petróleo, a maioria dos países tiveram suas economias afetadas gerando desequilíbrios fiscais que apenas via aumento de tributação não poderiam ser contornados (GOMES, 1998). Dada a oferta de crédito internacional, muitas nações se endividaram e, nesta conjuntura, os investimentos no setor de eletricidade se tornaram inviáveis. Com a retração econômica, o setor começou os anos 80 enfraquecido, isto devido a uma combinação de aumentos de juros nos países desenvolvidos e cortes de investimentos em países em desenvolvimento. Este último motivado pela desconfiança gerada por uma moratória declarada pelo México em 1982 (GIAMBIAGI, 2016). Neste terreno, a corrente de pensamento monopolista cedeu então espaço ao retorno das ideias liberais. Prometendo a retomada do crescimento por meio do combate à inflação e austeridade fiscal, uma agenda econômica conservadora foi adotada nos EUA, Inglaterra e Alemanha (STOFT, 2002). Dentre os pontos desta agenda, estavam as primeiras reformas dos mercados de eletricidade que tinham por necessidade se reinventarem. A reforma se daria baseada em ações de desregulação e na introdução da livre concorrência.

No livro *Power System Economics* de Stoft (2002) é feita uma abrangente defesa da necessidade generalizada de eliminação da burocracia nos mercados de eletricidade. O autor comenta em sua análise que nesses mercados "a desregulação não é equivalente a uma competição perfeita" e que apesar dos mercados elétricos apresentarem ineficiências [10] os mesmos funcionam melhor quando existe competição. Stoft (2002) ainda compara o setor de energia com o setor de habitação dizendo que o setor de habitação é também uma necessidade básica do ser humano, no entanto, ao contrário do setor de energia, este é "99% desregulado". Outro ponto levantado é que quando um erro é cometido por um agente regulado, o custo deste erro é distribuído entre todos os consumidores, mas quando isso acontece com um agente desregulado, ele sozinho é quem

[9] Como exemplo, pode-se citar tubos para água e gás, linhas de transmissão para energia elétrica, antenas para telecomunicação, estradas férreas para trens e etc.

[10] Eficiência de mercado diz respeito a situação onde todos os agentes envolvidos recebem ao mesmo tempo e de forma isonômica todas as informações dos preços, podendo competir racionalmente entre si de igual para igual.

pagará pelo seu erro. Nesta linha de raciocínio, a regulamentação apresentaria dois problemas fundamentais. No primeiro, o setor regulado não possuiria capacidade de fornecer incentivos aos agentes como um mercado competitivo pode. No segundo, não existiria forma de regulamentar os próprios órgãos de regulação. O autor finaliza postulando o que chama de "Dilema do Regulador". Este dilema diz que os mercados de eletricidade verdadeiramente competitivos podem oferecer (i) a potência total do sistema ao custo marginal com aportes de investimentos e, somado a isso, (ii) sempre manter a preocupação com a minimização do custo. Já os mercados regulados podem oferecer apenas uma das duas características, mas nunca ambas ao mesmo tempo.

Já Junior (2017) prefere reduzir a questão do mau envelhecimento da regulação a um momento específico da história. Em sua visão, o mercado de eletricidade se tornou estagnado no fim dos anos 80, pois apresentava barreiras institucionais à entrada de novas empresas, direitos de propriedade com relação a infraestrutura física e uma forte orientação apenas para os mercados internos. Nesse sentido, o autor comenta que "a regulamentação que havia sido concebida para superar as imperfeições do mercado elétrico, passou a ser apontada como a origem da má alocação de recursos". Assim, o mesmo cita dois problemas deste momento sendo (i) o sobreinvestimento, já que as empresas do setor possuíam remuneração garantida, o que provocava ineficiência alocativa de recursos e; (ii) com a assimetria de informações disponíveis, os reguladores passaram a ser influenciados pelas concessionárias, não sendo mais capazes de executar um bom controle dos custos.

Ademais de ambas as visões, o fato é que diversos países na década de 90 iniciaram suas reformas rumo aos mercados desregulados. Foram ações destes movimentos segundo Silva (2012), Nery (2012) e Junior (2017):

a) Separação dos segmentos de geração, transmissão e distribuição, dando espaço assim ao surgimento do segmento de comercialização de energia.

b) Introdução de concorrência entre os agentes.

c) Separação entre os agentes de controle físico e controle econômico.

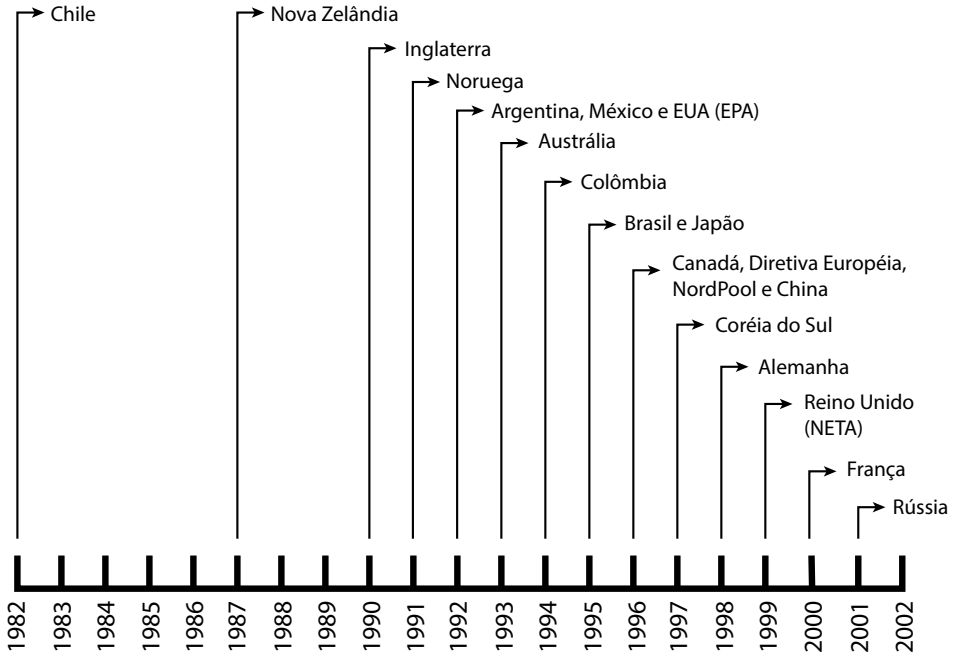
d) Privatização de empresas estatais.

Até meados do ano 2000, todas as principais economias do mundo haviam promovido mudanças estruturais em seus setores elétricos. Nesta ocasião, a título de exemplo, todos os países da Organização para Cooperação e Desenvolvimento Econômico (OCDE) tinham aberto comercialmente seus mercados para todos os grandes consumidores industriais e, em alguns casos, até residenciais. Isso representava um total de aproximadamente 500 milhões de pessoas elegíveis para escolha do seu próprio provedor de energia, ou seja, cerca de 50% da população destes países (NERY, 2012).

O Chile foi o primeiro país a implementar uma reforma. Entre 1974 e 1979 diversas empresas estatais de grande porte haviam sido privatizadas. Isso foi feito, pois uma reforma geral da economia estava sendo consolidada. A expansão desta agenda política foi o principal motivo para o qual em 1982 fosse promulgada a *Ley de Servicios Eléctricos* (CHILE, 1982). Esta lei criou

um marco regulatório com o intuito de permitir a participação de capital privado no setor. A lei ainda trouxe diversos elementos que incentivavam a participação privada, o uso racional da energia e a desburocratização (MALDONADO; MÁRQUEZ, 1995). Seguindo o exemplo do Chile, na Figura 7 pode ser vista uma linha do tempo na qual é mostrada a adesão de diversos países às reformas nos anos seguintes.

Figura 7 – Linha do tempo da primeira geração de reformas de mercados de eletricidade.



Fonte: Adaptado de Nery (2012).

Em 1990 foi formado o primeiro grande *pool* [11] energético do mundo, o qual reuniu ingleses e gauleses. Na sequência, em 1991 a Noruega criou seu próprio *pool* competitivo. Em 1996 a Suécia, a Dinamarca e a Finlândia se integraram aos noruegueses fundando um dos mercados mais sólidos do mundo até hoje, o NordPool. Ao final deste mesmo ano foi publicada uma diretiva que dava as condições necessárias para a criação de um mercado de energia unificado em todo o continente europeu, inclusive com um prazo para que todos os países se adequassem [12]. Na América do Sul, Argentina (1992), Colômbia (1994) e Brasil (1995) iniciaram também suas primeiras reformas. Em 1996, a Nova Zelândia se tornou o primeiro país do mundo a permitir negociações de energia com operações diárias por meio da criação de um mercado atacadista de adesão voluntária, resultado de uma reforma que já vinha sendo implementada desde 1987. Em 1998, a Austrália colocou em operação o seu Mercado Nacional de Eletricidade fundindo seus dois *pools*, o *Victoria Pool* e o *New South Wales Pool* (NERY, 2012).

[11] O conceito de *pool* de energia é tratado de forma ligeiramente diferente entre os autores consultados. Neste trabalho é adotada a definição de Mayo (2012) por ser a mais generalista. Este autor diz que um *pool* de energia é "um esquema em que a produção de eletricidade das diferentes usinas é agregada, visando à criação de um mercado eficiente para comercializar a energia elétrica".

[12] Situação que até o ano de 2022 ainda não se tornou uma realidade.

A primeira bolsa de energia competitiva deu início a suas operações em 1998 no mercado espanhol. Os EUA também aderiram às bolsas de eletricidade nos mercados CalPX (Califórnia), PJM (*Pennsylvania*, *New Jersey* e *Maryland*), NYPOOL (Nova York) e NEPOOL (*New England Pool*). Na Europa, Holanda (APEx), Alemanha (LPX e EEX) e França também instituíram suas bolsas a partir de 1998 (NERY, 2012).

Junto a estes movimentos de criação dos mercados foi estabelecido pela primeira vez o conceito de consumidor livre. Noruega (1991), Nova Zelândia (1994), Suécia (1996) Finlândia (1997), Califórnia (1998) e Reino Unido (1999) foram os pioneiros na abertura do mercado de energia para que seus consumidores pudessem optar individualmente pelos próprios fornecedores (NERY, 2012). No Brasil, com a reforma de 1995 e a lei nº 9.074/95 (BRASIL, 1995b) os consumidores livres também foram concebidos, no entanto, a entrada de um deles no mercado livre só foi acontecer em 1998. Uma empresa do setor químico chamada Carbocloro Indústrias Químicas (NERY, 2012).

A experiência de liberalização dos mercados de eletricidade entre 1980 e 2000 apresentou diversas características diferentes entre os países que promoveram reformas. Ao se criar regulações específicas para cada mercado os países introduziram cada qual a seu modo legislações que personificavam o seu próprio *modus operandi*, resultado de quase 100 anos de exploração da indústria de energia interna. As primeiras reformas, apesar de diferentes entre si, apresentaram problemas de implementação em comum não previstos pelos países. A maioria desses problemas estavam relacionados com a incapacidade de se encontrar o *trade-off* entre os possíveis benefícios das forças de mercado (que substituíram os antigos monopólios) e os malefícios gerados pelos recém introduzidos contratos bilaterais. Além da inexperiência com a adoção de um modelo, agora, desverticalizado (JUNIOR, 2017).

Mayo (2012) destaca quatro critérios que são requisitos para a sustentabilidade de um mercado de eletricidade eficiente. Logo, um mercado que não os promova terá problemas e tenderá a sucumbir. Todas as nações que instituíram alguma reestruturação nas suas primeiras reformas sofreram quanto a pelo menos um destes:

a) Liquidez: Um mercado líquido é aquele que permite aos compradores e vendedores transacionarem seus contratos com disponibilidade imediata, ou seja, deve sempre haver algum interessado em negociar energia sempre que um contrato for posto no mercado.

b) Transparência: Diz respeito a disponibilidade de informações de forma isonômica a todos os agentes do mercado ao mesmo tempo e com a mesma qualidade.

c) Confiança: Os agentes envolvidos devem dar credibilidade as operações que acontecem no mercado, confiar na justiça dos preços praticados e respeitar os contratos celebrados.

d) Poder de mercado: Idealmente nenhum agente deve ter meios para aumentar unilateralmente um preço acima do valor justo com a finalidade de maximizar unicamente o seu próprio lucro.

Quanto a estes requisitos e problemas para implementá-los, pode-se dar destaque ao caso da Califórnia. Em 2001, o governo do estado californiano precisou realizar uma intervenção no mercado livre de eletricidade para salvar suas distribuidoras. Exercendo poder de mercado, as usinas passaram a manipular as condições de geração e, consequentemente, os preços. Detendo uma condição privilegiada e imprevista, as mesmas negociavam a energia por um valor vinte vezes maior do que o normalmente praticado no mercado de curto prazo. Sem alternativa, as distribuidoras que precisavam da energia e estavam sujeitas a um limite máximo [13] para revenda, acabaram por entrar em colapso (GRIFFIN; PULLER, 2005).

Outro caso importante é o da Inglaterra. Um dos motivos pelos quais o governo britânico se submeteu ao processo de reforma foi o desejo de controlar o poder dos carvoeiros. As usinas baseadas em carvão representavam 75% da capacidade instalada em 1989 (NEWBERY, 2005). Após a implantação da reforma, em um primeiro momento resultados positivos foram observados. Isto ocorreu, pois as usinas passaram a poder negociar diretamente o preço do carvão. No entanto, falhas de mercado logo apareceram. Uma das mais famosas trata da formação do preço de pagamento às geradoras que recebiam de acordo com a Equação 2.1.

$$pagamento = LOLP.\left(VOLL - \begin{cases} lance, & senão\ despachada \\ SMP, & se\ despachada \end{cases}\right) \tag{2.1}$$

Loss of Load Probability (LOLP) é a probabilidade de falha do sistema que porventura inviabilizaria a geradora de realizar o fornecimento; *Value of Lost Load* (VOLL) é o valor da carga deficitária (se existisse); $lance$ é o valor que a usina poderia dar para o seu próprio kWh caso não fosse despachada e; por fim, *System Marginal Price* (SMP) trata-se do preço marginal da operação que era definido via leilão e aplicado caso uma usina fosse despachada.

O problema é que o valor do pagamento era processado um dia antes à operação real do mercado. As geradoras aproveitando-se disso, pediam altos preços pela sua energia e declaravam-se não despacháveis evitando vender ao SMP. No dia seguinte, alternavam sua posição declarando-se então despacháveis, mas como o preço era formado no dia anterior, passava-se a usar por consequência o valor artificial que haviam dado um dia antes. Dessa forma, exerciam poder de mercado lucrando acima do preço justo definido no leilão, explorando portanto uma falha no mecanismo de negociação (NEWBERY, 2005). Situações como essa fizeram com que a Inglaterra promovesse em 2001 uma segunda reforma no setor elétrico chamada de *New Energy Trade Arrangements* (NETA).

O caso da Califórnia e o da Inglaterra não foram casos isolados. Rothwell e Gómez (2003), Mayo (2012) e Viana (2018) citam outros estudos de caso ao redor do mundo. No Brasil, por exemplo, em 2005 foi instituído pela lei nº 10.848 de 15/03/2004 (Lei de Comercialização de Energia Elétrica) (BRASIL, 2004) um processo de regulação que, como nos outros países, procurou reavaliar a abertura de mercado e instituir mais mecanismos de segurança.

[13] Contratos com limite de preço superior são chamados contratos por *price cap* (MAYO, 2012).

Por fim, realizando-se uma compilação das obras consultadas, pode-se elencar como principais problemas das reformas liberalistas empreendidas entre 1980 e 2005:

a) Apesar da liberalização, inconsistências de mercado foram criadas devido ao fato de que o acesso às redes foi executado de forma restrita e discriminatória;

b) Proprietários de ativos de transmissão, por serem responsáveis por um ativo tratado como monopólio natural, se submeteram a um controle muitas vezes ineficiente sendo, em alguns casos, penalizados economicamente de forma não justificada por condições operacionais inerentes ao sistema como problemas de congestionamento e intercâmbio de energia. De outro ponto de vista, agentes com monopólio dos sistemas de transmissão que não atuavam de forma exclusiva neste segmento, quando inseridos em outras partes da cadeia (como a geração) que dependiam da transmissão, passaram a dar a si mesmos vantagens competitivas gerando desequilíbrios de mercado;

c) Mesmo com o incentivo à competitividade, diversos mercados se auto organizaram em torno de poucos agentes mantendo a concentração econômica. Esse fato possibilitou que o poder de mercado fosse exercido com a introdução de artificialidades nos preços e a criação de barreiras de entrada à novos agentes que poderiam atuar como concorrentes;

d) Na maioria das reformas, pequenos e médios consumidores simplesmente não foram incluídos ou tiveram seu acesso limitado ao mercado;

e) A experiência mostrou que os preços de mercado muitas vezes não acompanham os custos marginais da operação. Isto gera um descolamento entre a operação do mundo físico e a negociação no mundo financeiro. A falta de sinalização de um preço eficaz gerou situações passíveis de serem evitadas como sobras ou racionamento e, nesse sentido, coibiu movimentos naturais de resposta da demanda;

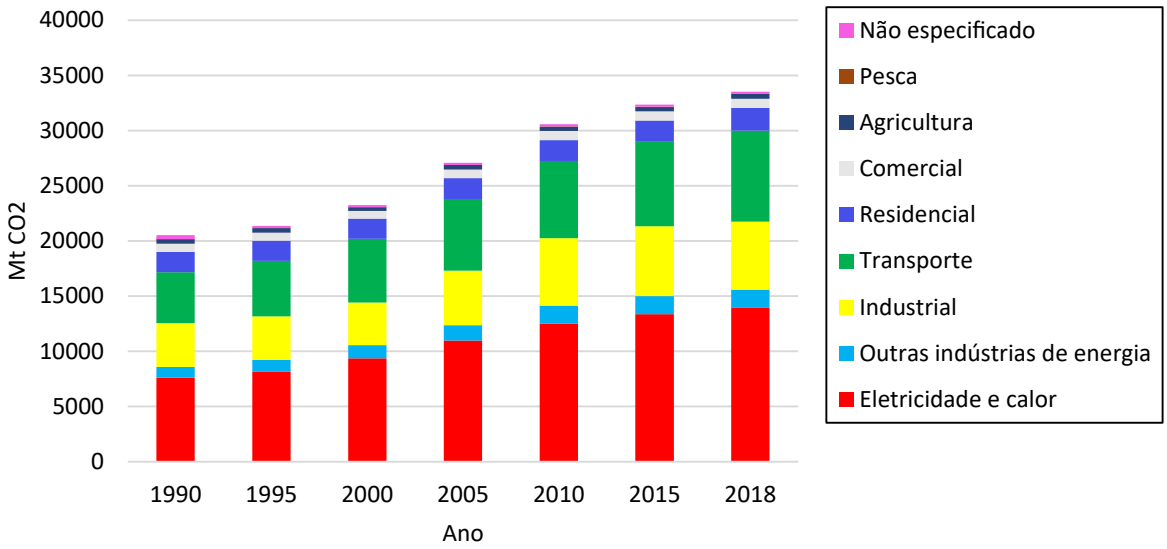
f) A defesa do mercado num ambiente puramente econômico é factível, no entanto, no mercado de energia elétrica a garantia da expansão deve ser sempre considerada dado que o crescimento econômico depende de investimentos em capacidade. Portanto, o desprezo pela visão de longo prazo ao custo de rendimentos imediatos gerou crises de energia, especialmente devido à falta de maior valorização de um planejamento da expansão.

Nos últimos 15 anos, a maioria dos países que implantaram mercados de energia competitivos vêm reestruturando seus modelos em um movimento chamado de "segunda onda" de reformas (NERY, 2012). Essa nova tentativa de desregulação e reconstrução dos mercados tem acontecido no sentido de corrigir os problemas gerados durante o período das primeiras reformas. No entanto, nesta "segunda onda" os mercados não estão tendo que lidar apenas com as lições aprendidas, mas também com outros desafios mais recentes tratados a seguir.

O setor de energia elétrica desde os anos 90 recebe críticas de ambientalistas, pois é o principal responsável pela emissão de gases de efeito estufa na atmosfera. Em 2018, essa parcela correspondeu a 42% do total de todas as emissões feitas no mundo (IEA, 2022). A Figura 8 mostra a evolução global do problema ao longo das últimas duas décadas. O Acordo de Paris (Nações Unidas, 2015) celebrado na 21º Conferência das Partes da Convenção-Quadro das Nações Unidas sobre Mudança do Clima (COP21) em 2015 e, em 2021, com 195 países signatários, atribuiu compromissos por meio de planos de ação em nível global. Dentre estes compromissos, todos os países se comprometeram a aumentar o percentual de participação de fontes renováveis em suas matrizes energéticas. Nesse sentido, a transição da geração baseada em fontes fósseis para geração limpa tem colocado o setor elétrico no centro das políticas de mitigação dos efeitos do aquecimento global. Entretanto, a inserção dessas fontes traz consigo dificuldades próprias. O principal obstáculo trata da reestruturação dos sistemas de operação para lidar com a intermitência. Fontes eólicas, solares, usinas fio d'água e outras, alteram de forma profunda a disponibilidade de energia, modificando assim a forma como é realizado hoje o balanceamento entre sobrecapacidade, armazenamento, entrega e consumo. A liquidez à pronto uso, provida pelas fontes fósseis, sempre garantiram a flexibilidade do sistema elétrico, dando possibilidade de modular a geração para acompanhar as variações da carga. Com a penetração das renováveis, o custo da geração não passa a ser mais a única preocupação dos operadores, visto que o custo da manutenção da liquidez também deverá ser considerado (JUNIOR, 2017). Outro ponto relacionado, trata-se das condições de imprevisibilidade as quais estão sujeitas as fontes renováveis. Com geração dependente de condições climáticas, o cálculo do preço da energia passa a incorporar mais uma parcela de incerteza, aumentando assim a expectativa de volatilidade. Complementarmente, existem ainda os recentes investimentos em novas tecnologias estocáveis incluindo as usinas reversíveis (*Pumped Hydroeletric Storage* (PHS)), hidrogênio, ar comprimido (*Compressed Air Energy Storage* (CAES)) e baterias. Além desses, modificam a estrutura do mercado a incorporação de centrais despacháveis de partida rápida, reserva girante, sistemas híbridos e novas tecnologias em serviços ancilares. Todos com a responsabilidade de dar mais flexibilidade ao sistema. Por fim, deve-se levar em conta que os incentivos fiscais associados a redução de custo de algumas formas de renováveis nos últimos anos transformou a microgeração distribuída em uma realidade. A demanda hoje em muitos mercados possui meios de responder a variação dos preços, seja se utilizando de produção doméstica ou injetando esta produção na rede (*prosumers*) (MAYO, 2012; JUNIOR, 2017).

Junior (2017) comenta que "a impossibilidade de armazenamento, especialmente em relação a grandes volumes, permanecerá como atributo crucial da eletricidade nas próximas décadas". Levando-se em conta esta afirmação é de se esperar que os sistemas elétricos continuem a se tornar cada vez mais complexos e diversos. Neste cenário, a necessidade de inovações tecnológicas passará a ser fundamental, visto que o uso racional da energia influenciará diretamente a qualidade da gestão das matrizes e, consequentemente, a saúde dos mercados de eletricidade. Neste contexto, tem se popularizado nos últimos anos as redes inteligentes (*smart grids*). Regulações próprias para mercados envolvendo *smart grids* já vem, inclusive, sendo propostas (MORENO, 2019).

Figura 8 – Emissão de CO2 por setor no mundo.



Fonte: Adaptado de IEA (2022).

Dado que os conceitos de economia de escala e escopo seriam impactados pela automatização. A conectividade bidirecional dos clientes implicaria num "aumento explosivo da complexidade do sistema, fruto da intensificação da interdependência gerada a partir do aumento do número de processos e da conectividade entre eles" (JUNIOR, 2017). Este fato torna necessária a adequação dos mercados para lidar com os perfis de resposta não necessariamente mais somente dos consumidores, mas também dos equipamentos inteligentes, que agora, mostram-se capazes de tomar decisões por si mesmos. Como visto, a evolução dos mercados de eletricidade, desde a criação da lâmpada incandescente até os atuais desafios impostos pelas fontes intermitentes e *smart grids* mostra que a capacidade de adaptação sempre foi necessária. Seja via um modelo monopolista ou um modelo liberal, governos ao redor do mundo sempre foram os que estiveram à frente da implementação das mudanças e reformas. Assim, como em qualquer área em evolução, o aprendizado se fez necessário em diversos momentos, exigindo ações e intervenções para manter a expansão do acesso, a isonomia entre os agentes e a qualidade do fornecimento.

A atuação econômica no contexto da indústria da energia elétrica é complexa. Essa ponderação, em geral foi observada logo nas primeiras páginas da maioria dos trabalhos utilizados como referência até aqui. Devido a isto e levando-se em conta a história do setor, a simplificação da estruturação do mercado sob a panaceia de que apenas a competição seria capaz de regular sozinha, de forma ótima toda a cadeia energética, na visão do autor desta tese é pobre e fora da realidade. Considerando que em vários momentos a liberalização de mercado foi necessária e que, também, em outros momentos a intervenção estatal foi necessária, o que fica claro é que neste tipo de mercado as regulações devem se complementar conforme o momento histórico, a estrutura legislativa de um país, o perfil de seus consumidores, a conformação geopolítica da sociedade e

tantas outras variáveis que, sendo tratadas apenas sob uma única perspectiva, seriam adequadas à resolução de alguns elementos mas, no entanto, deficitária no atendimento de outros.

Mercados de eletricidade se estruturaram ao redor do mundo em torno de alguns desenhos comuns que podem ser agrupados para estudo. A chamada Teoria de *Design* de Mercados (TDM) trata deste tema. Na próxima subseção, a mesma é apresentada por meio de uma compilação das principais práticas adotadas nos maiores mercados de eletricidade do mundo, incluindo o brasileiro.

2.2.2 Teoria de *design* de mercados de eletricidade

Mesmo com a introdução de diversas inovações nos mercados de eletricidade, ao longo de pouco mais de um século nenhum mercado até o momento pode ser considerado completo e perfeito (MAYO, 2012). O histórico discutido na subseção anterior mostra como os países evoluíram escolhendo caminhos próprios para a comercialização da energia. As diferenças entre estes caminhos refletiram-se nos aspectos dos desenhos de mercados. Seja pela propriedade dos ativos, pelo grau de centralização de poder ou pela forma como as funções foram delegadas. Alguns países nunca alteraram seu desenho de mercado, outros por sua vez passaram por reformas constantes. Essas diferentes características e políticas regulatórias mostram que não há ainda um padrão adequado para todos os contextos. No entanto, a TDM é uma área ancorada nos conceitos da microeconomia [14] que busca agrupar os atributos comuns entre a maioria dos modelos e, dessa forma, colocar em perspectiva os chamados desenhos canônicos, ou seja, os desenhos de mercado de eletricidade mais comuns. Hunt (2002) propõe quatro deles, os quais serão discutidos a seguir.

O primeiro modelo é o mercado de Monopólio Verticalmente Integrado (MVI). Neste desenho não existe o conceito de competição, visto que todos as partes envolvidas na cadeia de geração, transmissão e distribuição pertencem a um mesmo dono. As funções, portanto, são realizadas de forma integrada, centralizada e coordenada por um mesmo agente. Este agente não necessariamente é o estado. Pode ser, por exemplo, uma empresa privada ganhadora de uma concessão para administrar o setor.

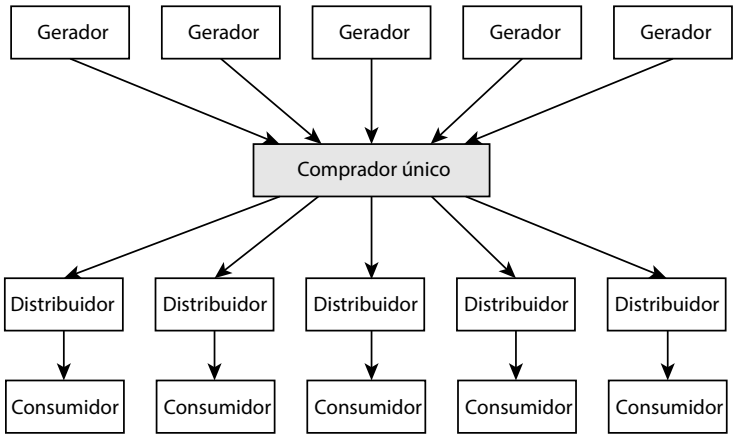
O MVI, dentre todos os quatro modelos foi o que vigorou por mais tempo na indústria mundial, sendo ainda adotado pela maioria dos países africanos e alguns países na América Latina e Ásia. Dado que em grandes economias é necessário um constante investimento em planejamento da expansão, este modelo torna-se de difícil sustentação, pois demanda flexibilidade. Segundo Viana (2018), acredita-se que em médio prazo este modelo será mantido apenas em mercados pequenos, onde o custo da estrutura de monopólio seja menor do que o custo dos ambientes competitivos. Ainda segundo este autor, com a continuidade da ascensão econômica mundial eventualmente todos os países migrarão para outros modelos.

[14] A microeconomia é definida como a análise que trata do comportamento de elementos individuais de uma economia, como a precificação de um produto ou o comportamento de um único consumidor, empresa ou segmento econômico (HERNANDEZ; OZ, 2016).

O segundo modelo é o Mercado de Comprador Único (MCU) (*single buyer*), também chamado em algumas bibliografias consultadas de mercado de aprovisionamento competitivo (*competitive procurement model*) (MAYO, 2012) ou mercado de agência central de comercialização (NERY, 2012).

Como mostrado na Figura 9, neste modelo os geradores [15] competem entre si por meio de leilões ou chamadas públicas para atender a demanda definida por um único comprador, normalmente um governo. O atendimento da demanda é realizado por menor preço e celebrado via contrato de longo prazo, este associado a vida útil das usinas. Cláusulas são estabelecidas para reajuste do preço ao longo do tempo, seja por meio de indicadores como inflação, preço de combustíveis ou situações imprevistas de operação. Neste sentido, os riscos do negócio são transferidos para o consumidor final que deverá arcar com a flutuação dos custos relacionadas à ocorrências adversas como obsolescência tecnológica, queda de demanda, aumento de competição e outros. No MCU espera-se que o agente central regule os preços de forma artificial, visto que as autorizações para venda feita pelas distribuidoras são concedidas por ele. O papel do distribuidor neste caso é o de atuar com monopólio em regiões geográficas restritas.

Figura 9 – Modelo de comprador único.



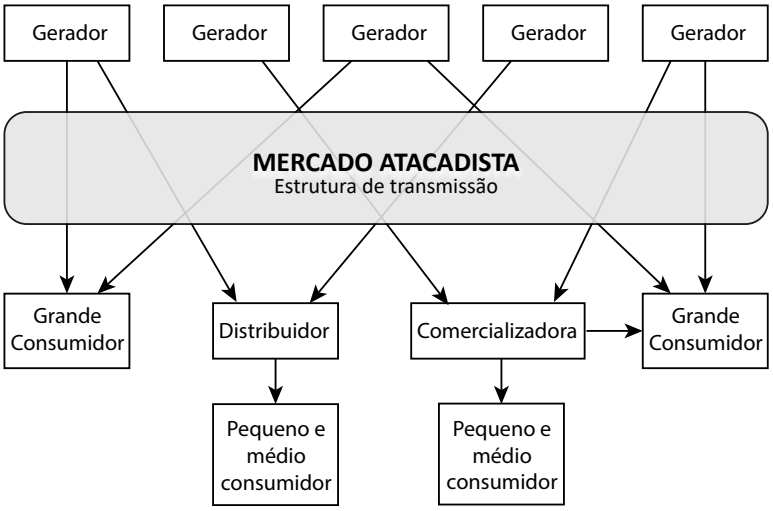
Fonte: Adaptado de Mayo (2012).

A transição do MVI para o MCU é a etapa inicial para a consolidação de um mercado competitivo. Isto ocorre, pois com a introdução das licitações e leilões de energia, uma primeira camada de disputa fica permitida. Todavia, é introduzido ao mesmo tempo um necessário estrato regulatório com objetivo de definir regras de execução destes processos. Quanto ao consumidor final, não há liberdade de exercer qualquer tipo de influência ou preferência no mercado.

[15] Hunt (2002) se refere aos geradores como *Independent Power Producers* (IPPs). Essa distinção é feita com o objetivo de generalizar o conceito para produtores de qualquer tamanho. O termo Produtores Independentes de Energia (PIEs) é o utilizado oficialmente pela ANEEL em suas regulamentações e instituído formalmente pela resolução normativa nº 674 de 11/08/2015 (BRASIL, 2015).

O terceiro modelo é o mercado de atacado e está representado na Figura 10. O modelo atacadista trata da proposta adotada pelo Chile em 1982 e seguida por EUA e diversos países europeus a partir de 1990 em suas primeiras reformas. Neste modelo, a competição entre os agentes de geração deve ser estimulada no sentido de que o preço formado a partir do custo da operação seja preterido em função dos ganhos em escala. Visando uma vantagem competitiva de longo prazo, os agentes devem estar dispostos a baixar o custo da sua energia para vencer seus concorrentes em certames envolvendo grandes quantidades, quais somente distribuidoras e grandes clientes têm condições de negociar. Portanto, no mercado de atacado não deve existir qualquer regulação sobre os valores praticados na venda da geração. Nos monopólios naturais de transmissão e distribuição as tarifas permanecem.

Figura 10 – Modelo atacadista.



Fonte: Adaptado de Hunt (2002).

Além da diminuição burocrática nas atividades de geração, torna-se necessária a separação das entidades de mercado e de operação do sistema. O operador do mercado torna-se responsável por criar, implantar e fiscalizar as regras que regem a comercialização. Já o operador do sistema deve ter a competência de garantir o despacho físico com base nas negociações consolidadas entre os agentes. Neste cenário os agentes devem ter acesso a ambientes de negociação onde os contratos podem ser celebrados, logo a existência de bolsas de energia, balcões de comércio, câmaras de licitação e leilões passam a ser um requisito. Agentes com atuação puramente financeira também devem ser incentivados. Mais pessoas negociando energia é o mesmo que maior liquidez para as operações. Nesse ambiente a figura da comercializadora surge sendo importante para gerir grandes blocos, intermediando o contato entre geradores e consumidores elegíveis. Corretores também são agentes possíveis, visto que podem atuar gerenciando juridicamente as relações entre as partes interessadas.

A última e mais importante peça de um mercado atacadista de eletricidade é a existência do chamado preço à vista (*spot price*). Também conhecido como *Market Clearing Price* (MCLP) ou

preço de referência. O mesmo trata do valor pelo qual as diferenças entre a energia física entregue e a energia negociada nos contratos serão liquidadas. A título de exemplo, pode-se ilustrar este conceito com a seguinte situação: Uma comercializadora fecha contrato com vários geradores para entrega de uma dada quantidade de energia. No entanto, no momento da operação real os consumidores que possuem contrato com a comercializadora consomem a mais (ou a menos) do que o montante negociado com os geradores. Essa diferença entre o consumido *ex-post* e o estabelecido em contrato *ex-ante* é então valorada conforme o preço à vista de mercado (ou via algum método de precificação específico que penalize o consumidor que porventura exceda os valores contratuais). A parte devedora deve, portanto, pagar a diferença ao custo do MCLP à parte credora. É o operador de mercado quem estabelece o MCLP, apura todas as diferenças e intermedia a quitação entre os agentes, além de administrar imbróglios jurídicos decorrentes de inadimplências.

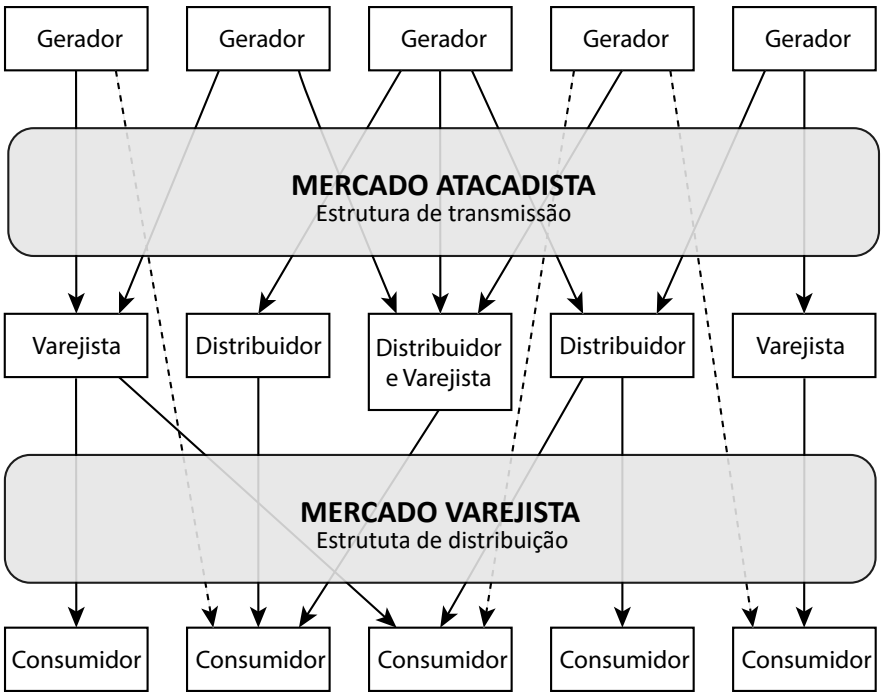
No modelo atacadista devem existir critérios que definam quais consumidores são elegíveis para participação no mercado livre. A maioria dos mercados utiliza a demanda contratada como indicador para permitir ou restringir esse acesso. Consumidores que não atendam aos critérios continuam a ser obrigados a comprar de um único distribuidor. Apesar de ser um modelo no qual a escolha do provedor de energia ainda não seja possível a todos, frente aos modelos de MVI e MCU mesmo os consumidores não elegíveis estão sujeitos a alguma sensibilidade no preço, visto que as distribuidoras podem valorar seus serviços não somente a uma única tarifa regulamentada, mas à soma do valor desta tarifa com o preço à vista. Em muitos mercados esses mecanismos são feitos com base em sinais [16] divulgados antecipadamente à prática do preço, logo pode existir alguma elasticidade da demanda.

Finalmente, o quarto modelo trata do mercado de varejo. A principal característica do mercado de varejo frente ao de atacado (como mostrado na Figura 11) é a adição de uma camada referente a comercialização ao nível da distribuição em baixa tensão. Comercializadoras de energia assumem o papel de varejistas podendo comprar em escala dos geradores e, por consequência, vender a preços mais competitivos aos consumidores. O papel dos distribuidores pode continuar a existir em regiões onde varejistas não tenham interesse em atuar e, neste caso, o estado é quem faria o atendimento. Os geradores também podem vender a energia diretamente a qualquer consumidor segundo suas próprias regras e preços.

Com a abertura do mercado para competição totalmente livre, os consumidores no mercado de varejo passam a atuar de forma ativa. Podendo portanto selecionar o seu fornecedor. Alguns dos mercados que já adotaram este modelo, inclusive, mantém *online* plataformas de contratação onde os clientes podem estudar as características de cada fornecedor e assinar planos de fornecimento. Exatamente como acontece com serviços de telefonia, internet, TV e similares. A inclusão do consumidor livre traz para a operação do sistema um aumento da elasticidade, visto que a resposta da demanda aos preços passa a existir. Nesta conjuntura, novos produtos podem ser oferecidos como instrumentos de *hegde*, consultorias de análise de consumo, venda de participação em lucros

[16] No Brasil, chamado de bandeiras.

Figura 11 – Modelo varejista.



Fonte: Adaptado de Hunt (2002).

de geração/transmissão, cooperativas, sistemas de automação de consumo e, no interesse desta tese, ferramentas inteligentes para apoio a decisão que preveem os preços.

Apesar das vantagens, a implementação deste modelo pode apresentar dificuldades, visto que toda a cadeia se torna mais complexa e a transparência das informações deve ser garantida. Além dos desafios técnicos de implementação, a instituição de um mercado de varejo adicionalmente pode enfrentar resistências políticas. Dado que historicamente os poderes executivo e legislativo usam as tarifas de energia como ferramenta de influência, a migração para este modelo pode se tornar dificultosa, visto que são mitigadas as formas de intervenção estatal (JUNIOR, 2017).

Como visto, todos os modelos apresentam suas vantagens e desvantagens e, ao longo da evolução dos mercados de eletricidade, cada nova proposta trouxe novas formas de deixar a comercialização mais sofisticada. Existem também características compartilhadas entre eles. No Quadro 1 pode ser visto um resumo dos principais atributos de cada mercado e na Figura 12 como estão distribuídos os modelos ao redor do mundo.

Dos quatro modelos de mercado, MVI e MCU tratam a eletricidade como um bem único onde, ou não existe qualquer negociação, ou, existe negociação mas esta será realizada por meio de contratos bilaterais de geração e entrega. Ao contrário, dado que nos modelos de atacado e varejo ocorre a desverticalização, naturalmente então existirá nestes uma segmentação da energia em serviços e produtos diferentes. São os principais produtos da eletricidade segundo Nery (2012) e Viana (2018):

Quadro 1 – Os quatro modelos canônicos de mercados de eletricidade.

Modelo	**Descrição**	**Vantagens**	**Desvantagens**
Monopólio verticalmente integrado	Empresa única que cuida de toda cadeia de produção de energia elétrica, incluindo transporte.	Modelo clássico e familiar para todos. Menores custos de transação. Ganhos de escala na operação conjunta da geração e transmissão.	Preço de venda maior devido a estrutura de monopólio. Ausência de escolha para o consumidor. Dificuldade de manutenção em cenário de escassez de recursos.
Comprador único	Comprador único que adquire energia elétrica de produtores independentes por meio de mecanismos competitivos.	Introdução de competição. Preços menores do que no modelo de monopólio. Mantém relativo controle regulatório.	Riscos são alocados aos consumidores. Ausência de escolha para o consumidor. Forte intervenção regulatória.
Atacado	Ambiente de competição entre vendedores e grandes compradores, incluindo distribuidoras que compram para atender consumidores residenciais.	Competição em nível de atacado. Preços menores do que no modelo de comprador único. Riscos alocados aos vendedores.	Complexidade regulatória. Complexidade institucional (operador de mercado e operador do sistema).
Varejo	Ambiente de competição entre vendedores e consumidores com livre escolha para todos os consumidores.	Competição em nível de varejo. Preços menores do que no modelo de atacado. Surgimento de novos produtos.	Perda da capacidade regulatória em influenciar preços e regras. Maior complexidade para o consumidor.

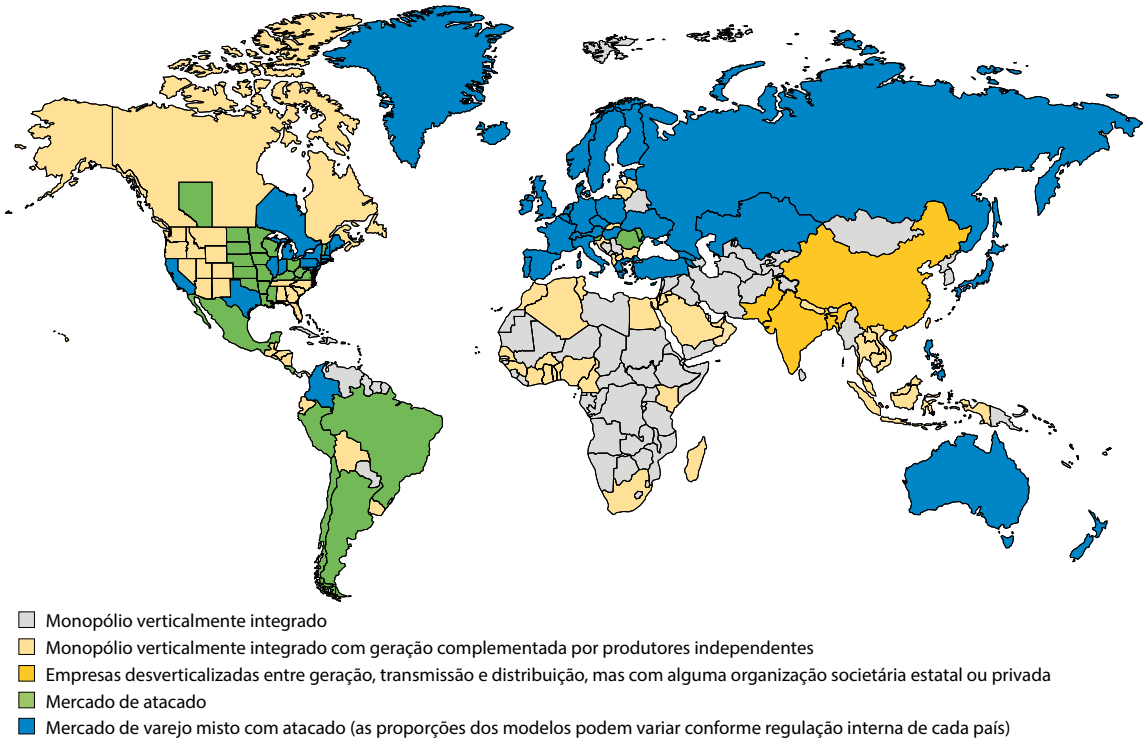
Fonte: Viana (2018).

a) Energia nova: Energia nova trata de acordos de comercialização envolvendo volumes de energia elétrica que ainda serão produzidos, normalmente relacionados a construção de novas usinas. Pode-se tratar também de contratos de longo prazo, incluindo intercâmbio entre regiões e países.

b) Capacidade: Trata da infraestrutura instalada e da capacidade dessa infraestrutura em gerar potência. É normalmente tratada como um bem público, pois aumenta a disponibilidade e a segurança do sistema. É precificada, normalmente, em unidades monetárias por ano ($/MW ano) ou por hora ($/MWh). Pode ou não incluir custos fixos.

c) Energia: Sendo a potência medida em *watts* (W) a energia trata da quantidade de potência concedida por uma dada fonte em um dado período no tempo, com precificação definida como o custo da unidade monetária por MW unitário por hora a cada hora ([$]/[MWh/h]). Sendo a energia o principal produto negociado nos mercados de eletricidade, esta está sujeita a influências externas que farão com que o preço da mesma se altere ao longo das horas do dia. Estas variações estão relacionadas com os perfis de demanda e com os limites de infraestrutura das usinas, sistemas de transporte e mecanismos de estabilização da rede.

Figura 12 – Modelos de mercados adotados mundialmente.



Fonte: Adaptado de IEA e KAPSARC (2017) e Viana (2018).

d) Reserva: Nesta categoria se incluem os produtos necessários a manutenção da confiabilidade, segurança e qualidade do sistema elétrico. Associado ao produto de energia, estes podem ser negociados sob a forma de produtos físicos (como equipamentos para controle primário de frequência ou suporte de reativos, por exemplo) ou também serviços (como equipes de manutenção 24 horas).

Os agentes envolvidos nas transações relacionadas aos produtos de energia elétrica se utilizam de ambientes de comercialização que dão legitimidade as operações. São os principais ambientes de comercialização de energia elétrica segundo Mayo (2012):

a) Balcão organizado: Ambiente onde a eletricidade é transacionada por meio de contratos não padronizados. Neste espaço pode ser acordada a entrega física da energia por meio de contratos à termo [17], como também outros instrumentos derivativos e puramente financeiros como seguros, opções [18] e *swaps* [19]. No balcão os riscos são

[17] Contrato à termo é um acordo de compra ou venda para entrega futura de energia a um preço pré-fixado, chamado preço à termo. Normalmente negociados para execução em horas de ponta, fora de ponta ou hora a hora.

[18] Opções são produtos derivativos de energia que oferecem ao seu detentor o direito, sem obrigação, de executar um contrato com preço, volume e prazo, no futuro, pagando ao vendedor um valor adicional por esta flexibilidade.

[19] *Swaps* de eletricidade são acordos que permitem aos seus detentores pagar pela energia contratada um preço fixo, independente da volatilidade durante o período contratual. Neste tipo de contrato não existe entrega física sendo que uma das partes do negócio, quando lucra, cobre os prejuízos da outra.

assumidos somente pelas partes envolvidas no contrato podendo, portanto, existirem regras discriminatórias onde determinadas empresas, por exemplo, só aceitem negociar com clientes específicos com volumes e preços definidos livremente.

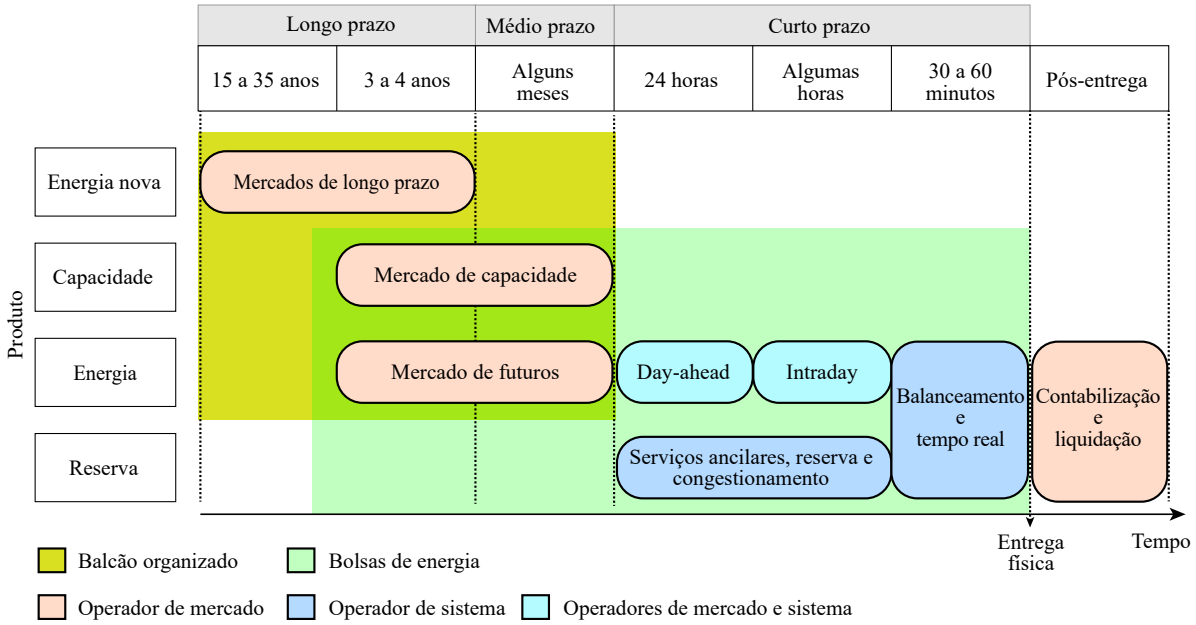
b) Bolsas de energia: Diferentemente do balcão organizado, onde as partes negociam livremente suas próprias cláusulas sem interferência do balcão no contrato, nas bolsas de energia as negociações são realizadas por meio de regras de operação. Com estas a bolsa tem papel regulatório atuando como uma contraparte. Nas bolsas de energia não existe qualquer discriminação entre vendedores e compradores, sendo inclusive aplicada como regra padrão as transações anônimas. Neste ambiente as negociações são em sua grande maioria públicas, o que garante transparência, mitigando desta forma o exercício de poder de mercado e, por consequência, evitando as artificialidades na formação dos preços. Como não há opção de escolha sobre com quem irá se negociar, passam a ser os principais diferenciais dos agentes o menor preço de venda ou o maior volume de compra. Isto incentiva a competitividade existindo, neste sentido, uma tendência de redução do preço à vista. Dessa forma, mesmo que uma negociação seja realizada por meio de um balcão, a referência de preço tenderá a ser a praticada nas bolsas de energia. Bolsas de energia normalmente são instituições privadas com fins lucrativos (VIANA, 2018) que tem permissão de atuação concedida por órgãos governamentais dos setores financeiros e de energia. Ainda segundo Mayo (2012), bolsas de energia promovem a diminuição do risco de crédito, incentivam a geração eficiente e sinalizam investimentos necessários para o aumento da capacidade.

c) Câmara de compensação (*Clearing house*): Contratos celebrados seja em balcão ou em bolsa de energia necessitam de um agente regulador que atue para garantir que as condições e acordos sejam honrados. As câmaras de compensação são a terceira parte em qualquer negócio, assumindo o papel de reduzir os riscos de liquidação por meio de monitoramento de crédito e providência de fundos de garantia. A salvaguarda das transações nestes ambientes é feita com base na exigência do depósito de uma margem financeira a ser feita pelas partes. O tamanho da margem pode variar dependendo da volatilidade do produto de energia ou do tipo de operação a ser realizada. A câmara de compensação está vinculada compulsoriamente às bolsas de energia. Já nos balcões, podem ser elegidas pelos participantes. Outro papel importante de uma câmara de compensação é a realização do ajuste diário de preços para possíveis atualizações contratuais.

A negociação dos produtos de energia nos ambientes de comercialização citados até aqui se dão por meio de prazos que acabam por criar mercados nichados e específicos. Energia nova, por exemplo, é negociada através de contratos de longo prazo entre 3 e 35 anos em balcões. Outros mercados como o de capacidade podem ter contratos para entrega entre poucos meses até 4 anos à frente, seja em bolsas ou em balcões. Essa divisão dos mercados de energia por produto e prazo é chamada de fatiamento (*slicing*) e acontece naturalmente quando estes evoluem no sentido do

incentivo à competitividade (STOFT, 2002; VIANA, 2018). Na Figura 13 são mostrados estes mercados. Segundo definições de Viana (2018) e Mayo (2012) cada um destes é explicado a seguir e, em acordo com os objetivos desta presente pesquisa, com maior detalhamento são comentados ao final os que transacionam energia.

Figura 13 – Modelo geral de mercados de energia elétrica para atacado e varejo conforme ambientes, operadores, prazos e produtos.



Fonte: O Autor.

2.2.2.1 Mercados de longo prazo

Energia nova é negociada por meio de contratos de longo prazo, os chamados *Power Purchase Agreements* (PPAs). Nestes são definidos além do preço, cláusulas de execução de projeto como cronograma, penalidades por descumprimento de prazos, condições de pagamento e responsabilidade sobre recursos a serem investidos (SIMAREMARE; HARYADI; ADITYA, 2020). Muitos países adotam neste mercado o uso de leilões. O Brasil é um caso. O ganho de um certame de PPA dá ao vencedor o direito de explorar a produção de energia ou infraestrutura de transporte, geralmente, por mais que dez anos com entrega sendo realizada entre 4 e 35 anos à frente, condicionada a investimentos na expansão ou criação de capacidade (VIANA, 2018).

2.2.2.2 Mercados de capacidade

Antes de tudo, a existência de um mercado de capacidade se justifica devido a necessidade de atração de investimentos. No entanto, quando se fala em energia elétrica, existem algumas situações que tornam difícil aos investidores enxergarem que a realização de aportes na expansão de parques de geração seja, de fato, algo rentável. Esta visão ocorre, primeiramente, devido a impossibilidade de estocagem em larga escala de energia e, também, em decorrência da inelasticidade da

demanda ao preço. Esses dois pontos podem levar a situações críticas no sistema elétrico em que o operador, propositadamente, promova períodos de racionamento. Essa condição extrema, em casos de urgente necessidade de controle de preços, pode ser aplicada como único recurso, visto que a negociação de energia não segue com perfeição as leis de oferta e demanda. Em segundo lugar, a possibilidade de controlar artificialmente os preços também pode acontecer quando o operador, visando garantir o suprimento físico regional, resolve executar o despacho das usinas fora da ordem de mérito de preço. Em terceiro lugar, o preço à vista em muitos mercados pode ser regulado via tetos (*price cap*) praticados por órgãos governamentais, neste caso, o problema existiria quando estes tetos não fossem altos o suficiente para serem considerados lucrativos. Em quarto lugar, com o constante aumento da penetração de renováveis, a intermitência dos sistemas tem aumentado as situações de incertezas além do que, nestes sistemas, o custo marginal de produção é próximo de zero. Por fim, baixos preços de combustíveis fósseis podem criar a impressão de que o preço da energia também será baixo. A combinação dessas possíveis situações pode transformar o mercado de energia em um ambiente pouco convidativo para novos negócios. Esse fenômeno é chamado de *missing money* e, quando acontece, pode comprometer a segurança de suprimento no longo prazo. Nesse contexto, surge como solução a implantação dos mercados de capacidade que tem por objetivo criar diversos mecanismos de remuneração à geração de tal forma que, aos olhos dos investidores, a construção de novas usinas (e serviços associados) tornem-se oportunidades atrativas (VIANA, 2018).

Os mercados de capacidade [20] têm contratos de médio e longo prazo transacionando produtos sob três abordagens, sendo:

a) Mercados somente de energia (*energy-only markets*): Neste se espera que o preço à vista reflita corretamente o sinal econômico, assim, em caso de escassez da oferta, o preço deveria subir e, portanto, com o aumento da remuneração pelos equipamentos de suprimento a realização de investimentos consequentemente passa a ser considerada.

b) Encargo de capacidade (*capacity payment* ou *capacity mechanism*): Nesta abordagem existe um pagamento fixo pela disponibilidade dos geradores, independentemente de estarem acionados ou não.

c) Mercados de capacidade instalada (*installed capacity markets* (ICAPs)): Nestes mercados os fornecedores devem obrigatoriamente contratar energia sobressalente junto aos geradores para os momentos de ponta de demanda a fim de suportar possíveis flutuações (MAYO, 2012).

2.2.2.3 Serviços ancilares e mercados de reserva

Os serviços ancilares e mercados de reserva [21] tratam de um conjunto de serviços pelos quais o operador do sistema ou agentes da cadeia de produção de energia devem pagar para manter

[20] Uma análise detalhada sobre mercados de capacidade pode ser vista nos trabalhos de Cramtom, Ockenfels e Stoft (2013) e Viana (2018).

[21] Uma descrição completa dos serviços ancilares e mercados de reserva pode ser vista em Mayo (2012).

a confiabilidade do fornecimento. Estes serviços devem ser precificados conforme o mercado em que estão inseridos a fim de evitar múltiplas remunerações sobre a mesma conveniência. De forma geral, serviços ancilares e mercados de reserva englobam o acompanhamento da carga e regulação da frequência (*load following and frequency regulation*), suporte de reativos para controle de tensão, reservas de contingência (*contingency reserves*) e capacidade de autorrestabelecimento das unidades de geração (*black start capacity*). Todos estes, normalmente negociados no curto prazo.

2.2.2.4 Mercados para gestão de congestionamento

Linhas de transmissão possuem capacidade limitada de tráfego elétrico. Em situações onde a rede elétrica não possa acomodar todos os fluxos físicos resultantes da programação da operação torna-se necessária a aplicação de métodos de resolução de congestionamento. A gestão do congestionamento deve, portanto, "minimizar as restrições visando à competição de mercado de um lado e de limitar os custos operacionais de outro" (MAYO, 2012). Dado que nem sempre é possível realizar o despacho de toda a energia contratada devido a congestionamentos, passa a ser papel do operador realizar a administração dos fluxos por meio das linhas disponíveis, sendo que neste caso o congestionamento pode ocorrer em qualquer ponto do sistema interligado. A forma como a gestão é feita afeta diretamente o preço da energia e pode gerar custos aos usuários do sistema.

As metodologias mais comuns para gestão de congestionamento [22] são o alívio do congestionamento (realizado *ex-post* ao ajuste do mercado) e a alocação de capacidade (realizada *ex-ante* a entrega física da energia).

No alívio de congestionamento a gestão é realizada em tempo real e se dá por meio de um rearranjo da interação entre a carga e a geração. Quando realizado sem levar em conta a situação de mercado (*redispatching*) a energia é negociada inicialmente como se não houvesse congestionamento, dessa forma garantindo um preço único para todo o sistema. Num segundo momento, caso um congestionamento ocorra, o operador altera o despacho da geração colocando as usinas fora da ordem de mérito de custo, assim redistribuindo o fluxo nas linhas. Esse redespacho onera o operador e, por isso, o prejuízo deverá ser igualmente dividido entre aos participantes pertencentes às áreas nas quais o congestionamento aconteceu. Outra forma de se realizar o alívio é levando-se em conta as condições de mercado (*counter trading* ou *buy-back*). Neste método, caso ocorra um congestionamento o operador de mercado cria um segundo mercado, paralelo. Neste segundo mercado o mesmo compra e vende eletricidade por meio de leilões com o objetivo de equilibrar a demanda. Como no redespacho, este mecanismo também acarreta custos que deverão ser repartidos entre todos os usuários afetados.

Na alocação de capacidade são utilizados leilões de capacidade de transmissão como meio de resolver congestionamentos previstos antes da operação. Esses leilões podem ser explícitos, quando ocorrem dentro de um mesmo mercado regional, ou, implícitos, quando ocorrem entre mercados regionais vizinhos. Existem ainda formas mais complexas de resolução de congestio-

[22] Em Mayo (2012) são abordados com detalhes os mercados para gestão de congestionamento.

namento que vão interferir no preço da energia e no comportamento do mercado. Algumas dessas que valem ser citadas são o apreçamento regional (adotado na união europeia), o acoplamento de mercados e o apreçamento nodal (usado nos mercados norte-americanos PJM-*Interconnection*, *New England* e Nova York) (MAYO, 2012).

2.2.2.5 Mercados de energia

A existência dos mercados de energia nova, capacidade e reserva, permitem que a energia em si seja negociada separadamente e de forma exclusiva, sem a adição de custos paralelos. Nesta categoria se enquadram o comércio dos chamados *futuros* (*forward markets*), as operações para um dia à frente (*Day-ahead Markets* (DAMs)), as operações executadas no mesmo dia (*Intraday Markets* (IDMs)), as atividades de balanceamento e as de tempo real (*Real-time Markets* (RTMs)). Estes mercados se complementam formando uma cadeia interdependente onde a energia pode ser vendida ou comprada em momentos diferentes, mas a entrega física sempre acontecerá na ponta final.

Agentes autorizados a comprar e vender a eletricidade possuem como base para suas negociações o histórico de valores do preço à vista como referência. Baseando-se nas condições de mercado passadas e, também, a partir da análise de condições futuras, estes agentes podem considerar que realizar um contrato com valores pré-determinados pode ser mais vantajoso do que se expor a um preço futuro incerto. Este contrato pode ser realizado para blocos de horas similares como horas de ponta, fora de ponta ou de carga uniforme. Pode ser elaborado também visando condições de calendário como feriados, dias úteis, estações secas e molhadas, hábitos de consumo, eventos e outros. Variando com prazo de uma ou duas semanas (médio prazo) à até 3 ou 4 anos (longo prazo) e, sendo executados em ambientes de balcão ou bolsa, é no mercado de futuros que estes contratos são celebrados. Os chamados contratos futuros de energia têm por principal finalidade proteger os agentes do preço à vista praticado no DAM, IDM e RTM. A intenção, portanto, é adotar uma posição conservadora realizando uma operação antecipada com uso de um preço conhecido (VIANA, 2018).

Contratos futuros devem ser registrados no operador de mercado para fins de compatibilização com a operação física. No entanto, não existe obrigatoriedade no seu cumprimento, sendo que neste caso o operador do sistema deverá ser apenas informado pelo de mercado. Sendo dessa forma instrumentos financeiros sujeitos à não execução, os acordos celebrados no mercado de futuros devem ser associados a outros tipos de contrato em que exista a obrigatoriedade da realização de geração e de consumo. Estes contratos associados são os acordados no DAM.

O DAM (também conhecido como mercado *D+1*) é o mercado físico diário principal da cadeia de comercialização de energia elétrica (MAYO, 2012). Neste, todos os agentes devem possuir energia assegurada para consumir ou gerar e todas as demandas ainda não contratadas devem ser atendidas. É um ambiente onde as ofertas são realizadas e liquidadas 24 horas anteriormente à hora de despacho. É considerado, portanto, parte do processo de *Unit Commitment* [23] (VIANA, 2018).

[23] *Unit Commitment* trata-se do compromisso que os agentes assumem em executar uma dada operação em um dado momento no sistema elétrico.

Para programação do despacho no dia seguinte o operador do sistema combina, consequentemente, os montantes firmados via contratos futuros com as coberturas realizadas no DAM.

Dois modelos hoje prevalecem para negociações no DAM. O primeiro trata do uso de leilões e é o mais utilizado mundialmente, se adequando tanto a atacado quanto varejo (VIANA, 2018). O segundo, diz respeito a um modelo de negociação contínua baseado no "casamento" das ordens de compra e venda submetidas à um livro de ofertas. Mesma forma de negociação adotada por bolsas de valores para transações com ações, moedas e outros ativos capitais.

Buscando-se melhorar o equilíbrio entre a geração e a previsão de demanda, após o fechamento do DAM, alguns mercados ainda permitem o ajuste de posições contratuais. Esse ajuste acontece no IDM e, como no DAM, é também realizado via leilões ou negociação contínua. O leilão no IDM acontece anteriormente a cada hora do dia e é finalizado até alguns minutos antes desta hora. O limite para submissão de lances pode variar, em geral, de 60 a 15 minutos antes, sendo que alguns mercados praticam intervalos ainda menores. O momento de fechamento do IDM é chamado de *Gate Closure* (VIANA, 2018). O operador de mercado, seja quanto ao DAM ou quanto ao IDM, deve providenciar uma forma pública de acesso a todas as negociações realizadas, pois somente com esta transparência existirá incentivo para participação dos agentes nos leilões com submissão de lances com preços que reflitam os custos reais da operação.

Um ponto que pode influenciar diretamente o desenho de mercado e está relacionado ao IDM é a forma de apreçamento, sendo esta possível de ser realizada de duas maneiras. A primeira, diz respeito ao apreçamento zonal. Nesta forma é feita uma conglomeração de várias barras e um único preço é praticado para toda a região onde as mesmas se encontram. A segunda é o apreçamento nodal. Neste existirá um preço definido para cada barra do sistema, logo os leilões não são para uma dada área mas sim para uma única barra. É uma forma que reflete melhor a realidade, mas por ser mais granular, em sistemas maiores pode trazer grande complexidade e dificuldade de interpretação das condições de mercado.

Após o IDM, são executados o RTM e o balanceamento.

No RTM, os agentes podem submeter diretamente aos operadores novas ofertas de preço que serão executadas de imediato para reduzir a carga ou aumentar a geração. Neste mercado é praticado o preço de equilíbrio [24], um valor geralmente próximo do CMO e que reflete o momento onde a oferta se iguala à demanda. O balanceamento, por fim, trata da imposição de penalidades a todos os agentes que incorrerem no descumprimento do *Unit Commitment*. Com o objetivo de mitigar desequilíbrios energéticos no sistema, os agentes geradores serão cobrados caso produzam a mais ou a menos do que o contratado e, os consumidores, caso retirem da rede mais ou menos do

[24] O preço de equilíbrio é determinado com base na agregação das curvas de oferta e demanda realizadas durante um leilão bilateral. O ponto de coincidência entre os preços é praticado para todos os agentes que tiveram lances aceitos antes de toda demanda ser contratada. Não é o foco deste trabalho discutir os tipos de leilão, no entanto, descrições sobre formação de preços, incluindo os de equilíbrio, podem ser consultadas nas obras de Stoft (2002), Silva (2012) e Mayo (2012).

que o contratado. As penalidades no balanceamento podem ser diferenciadas. Caso se tratem de uma produção em excesso o preço pago por essa energia é menor do que o CMO. Quando se trata de energia faltante, o gerador paga além do CMO. A ideia é a mesma em relação ao consumidor. O formato dessa precificação pode mudar de mercado para mercado.

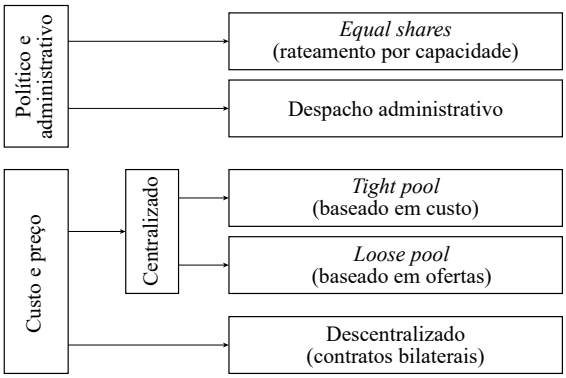
RTM e balanceamento podem ser praticados isoladamente, sem dependência de um para com o outro. Prevalece nos mercados norte-americanos, por exemplo, o uso de RTM. Já em mercados europeus como França, Reino Unido e Bélgica é adotado o balanceamento (MAYO, 2012).

Na última etapa da cadeia, após a entrega física da energia ser realizada, acontece a contabilização e a liquidação. Nesta fase, o operador de mercado inicialmente apura as diferenças contratadas e as consumidas/produzidas. Essa fase (contabilização) tem como resultado um relatório que mostra quais agentes precisam liquidar suas posições devedoras ou credoras. Na sequência, ao preço à vista, os descumprimentos são quitados (liquidação). Adicionalmente a estes valores podem ser também incluídos custos porventura ocasionados por acionamento de serviços ancilares e reservas, além das penalidades oriundas da fase de balanceamento. É importante destacar que os custos apurados na liquidação também podem ser negociados em contrato ou serem diferentes de mercado para mercado.

Neste apanhado sobre TDM em todas as bibliografias consultadas (STOFT, 2002; MAYO, 2012; NERY, 2012; SILVA, 2012; VIANA, 2018) ainda é considerado um último fator de mercado, sendo este, a forma como o despacho dos geradores é feita e, por consequência, a influência deste despacho na formação de preços.

Inicialmente, as abordagens de despacho podem ser categorizadas sob a ótica da economia de recursos que leva em conta os custos para definição da ordem de mérito, ou, sob a ótica não econômica, onde apenas decisões políticas e administrativas são consideradas. Como pode ser visualizado na Figura 14, destas categorias se subdividem cinco abordagens.

Figura 14 – Categorias e abordagens de despacho.



Fonte: O Autor.

A primeira é a *Equal Shares*, um modelo adotado atualmente pela China onde todos os custos de atendimento da demanda são divididos entre os geradores de forma proporcional a

sua capacidade de produção. A segunda, trata do despacho meramente administrativo, onde um determinado operador pode, por exemplo, querer privilegiar o uso de determinada tecnologia, favorecer uma determinada usina, atender ou não uma determinada região e assim por diante. Essas abordagens que levam em conta apenas requisitos políticos e administrativos são comuns apenas nos modelos MVI e MCU.

A terceira e quarta abordagens ocorrem apenas em modelos centralizados de despacho, onde o operador do sistema agrega todas as usinas no chamado *pool* obrigatório. Pertencendo a um *pool* obrigatório, uma dada usina só poderá realizar o despacho caso seja autorizada por uma unidade central. Essa autorização pode advir de qualquer um dos mecanismos de comercialização até aqui já apresentados. O modelo de *pool* obrigatório do tipo *tight* [25] é concebido de forma baseada nos CMOs. A ordem de mérito, neste caso, irá permitir que um gerador só produza um dado volume em um dado horário, somente se o seu custo de operação for necessariamente menor do que os dos geradores concorrentes, até o atendimento da carga. Já no modelo *pool* do tipo *loose*, os geradores realizam ofertas voluntariamente para atendimento da demanda. Não existe qualquer tipo de controle sobre os preços praticados ou os volumes. No entanto, é comum a existência de regulações que incentivem a entrada desses geradores nos leilões, logo a ordem de mérito é formada com base nos lances que se concretizam por meio das ofertas de preços.

Finalmente, a quinta abordagem é o despacho descentralizado por custo e preço. Nesta modalidade os agentes podem celebrar contratos bilaterais em balcões ou bolsas de energia, ficando o operador do sistema encarregado apenas da garantia da confiabilidade, realizando a gestão dos serviços ancilares, reserva, congestionamento e tempo real. É a abordagem mais liberalista de mercado possível, onde os agentes interagem livremente buscando sempre maximizarem suas vantagens. Não existe regulação, a não ser as contidas nas cláusulas dos contratos.

Reino Unido e diversos países da união europeia adotam o modelo de despacho bilateral. Atualmente a abordagem *loose pool* é a usada pela Austrália, Nova Zelândia, NordPool, Colômbia e alguns dos mercados canadenses e norte-americanos. EUA e Canadá mantém também mercados atacadistas que adotam modelos *tight pool*. O *tight pool* ainda é comum em países da América Latina (como o Brasil), África e Ásia. Na América Latina, África e Ásia também são encontrados mercados com abordagens de despacho político e administrativo, como o já citado caso da China que utiliza *Equal Shares*.

Nesta subseção foram apresentados os quatro modelos canônicos de mercados de energia elétrica e, no âmbito dos modelos de atacado e varejo, discutiu-se os nichos de mercado que podem ser criados quando se leva em consideração os mais comuns produtos de energia, ambientes de comercialização, tipos de operadores e prazos de negociação. O atendimento dos agentes, ainda, pode ser realizado conforme cinco abordagens de despacho diferentes. Todos estes tópicos são apresentados nas obras de Hunt (2002), Stoft (2002), Mayo (2012), Nery (2012), Silva (2012) e Viana (2018).

[25] Também encontrado com denominação de *groos pool* em alguns trabalhos (CHUA-LIANG; KIRSCHEN, 2009; DENNY; O'MAHONEY; LANNOYE, 2017; RAI; NELSON, 2020).

Levando-se em conta a variedade de combinações que podem ser feitas ao se estruturar um mercado de energia, observa-se que a complexidade existente na escolha de uma forma de comercialização para um dado país ou região não é uma tarefa trivial. Conforme visto quanto a evolução dos mercados de eletricidade, estes diversos elementos vêm sendo testados ao redor do mundo com experiências positivas e negativas relatadas. Não obstante somente a tarefa de se projetar o desenho de um mercado de energia, nos tempos atuais ainda há que se considerar diversas outras variáveis, como exemplo, a alocação de direitos de transmissão de energia (*Financial Transmission Rights* (FTRs)), intercâmbios de eletricidade entre nações e empreendimentos que visam integração de países por meio de usinas multinacionais, o crescente aumento da geração distribuída, a penetração e incentivos às fontes limpas e intermitentes, novos produtos de energia financeiros, o avanço das tecnologias de estocagem e, também, em mercados mais modernos, a realidade que vem se tornando as *smart grids*. Sendo assim, não se espera com essa subseção esgotar o assunto, mesmo porque com temática tão ampla, não somente a discussão técnica do mercado elétrico entra em pauta, mas também, questões estratégicas do ponto vista socioeconômico. A tendência vista, conforme o histórico e a evolução dos desenhos de mercados é que, gradualmente, os consumidores terão maior poder de decisão, isto promovido via maior participação, mesmo que indireta na gestão dos fluxos e seleção de provedores.

O objeto de estudo desta pesquisa se aplica neste tipo de ambiente e, de forma mais específica, ao mercado de eletricidade brasileiro. Nesse sentido, a apresentação destes conceitos se faz importante, pois o conhecimento geral de mercados elétricos via a experiência internacional provê suporte ao entendimento do desenho de mercado adotado atualmente no Brasil, inclusive das suas particularidades e mais recentes transformações. Com base nesta linha de raciocínio, na próxima subseção o leitor será apresentado as principais regulações e formas de comercialização de energia praticadas no mercado nacional. O objetivo é conduzir o entendimento até o conceito do Preço de Liquidação de Diferenças, principal valor de referência para transações de energia no Brasil e instrumento adotado em formato horário em janeiro de 2021 pela Câmara de Comercialização de Energia Elétrica para contabilização e liquidação de contratos.

2.3 Mercado de eletricidade brasileiro

A história da eletricidade no Brasil tem início em 1883 quando Dom Pedro II inaugura uma das primeiras redes de iluminação do mundo no estado do Rio de Janeiro. Em 1889 foi também no Brasil instalada a primeira usina hidroelétrica da América do Sul na cidade de Juiz de Fora em Minas Gerais. Deste ponto em diante, a arquitetura de mercado brasileiro se desenha conforme seu retrospecto. Nesse sentido, a seguir é feita uma breve apresentação do histórico do mercado de energia nacional [26]. O mesmo pode ser acompanhado por meio da linha do tempo contida nas figuras 15 e 16. A mesma esta dividida em governos, tipos de mercado e, entre 2017 e 2021, é feito um destaque relacionado ao período de aplicação do modelo proposto no Capítulo 5.

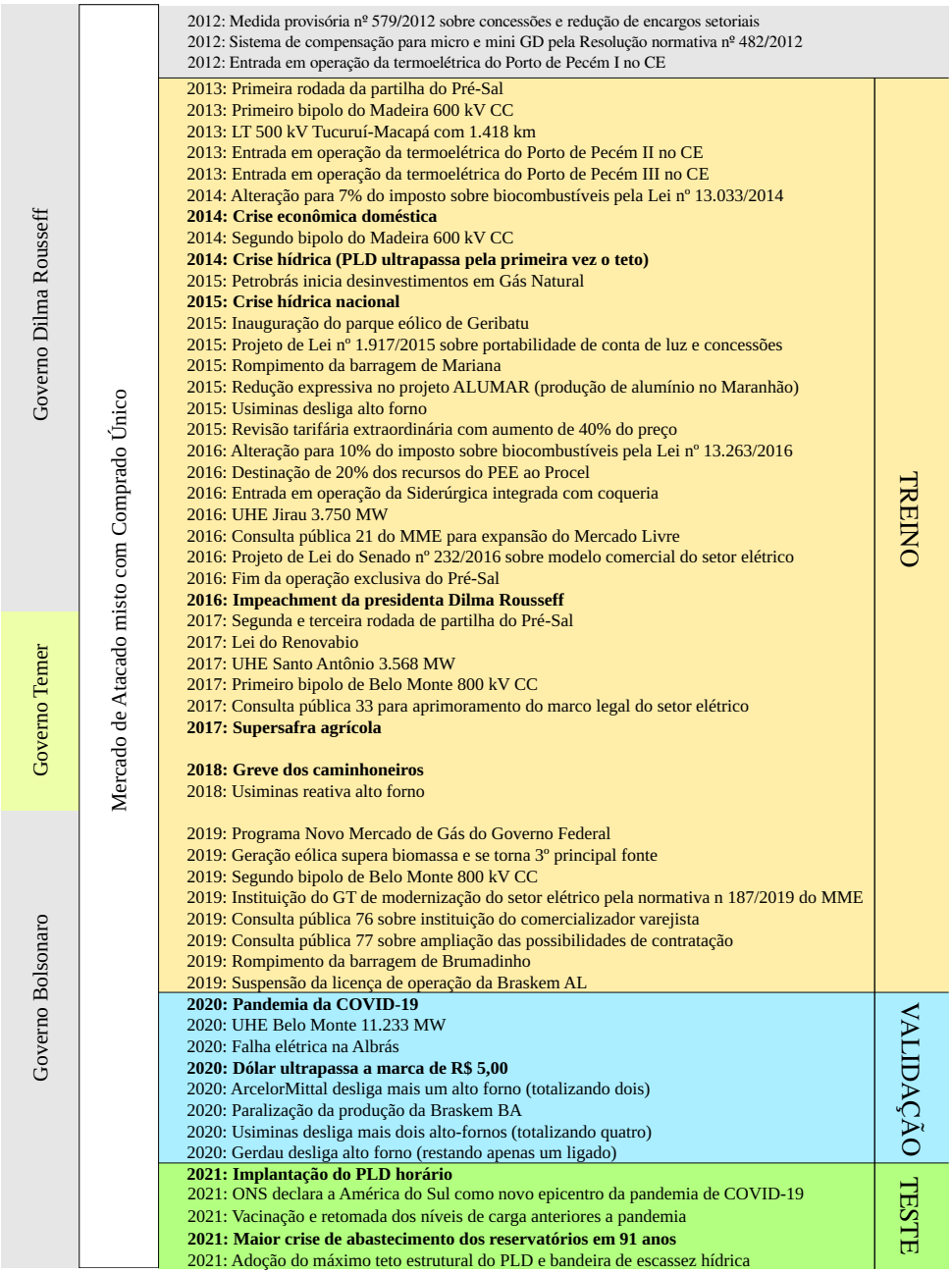
[26] Baseado no Balanço Energético Nacional (EPE, 2020b), na legislações citadas e nas obras de Tolmasquim (2015) e Junior (2017).

Figura 15 – Linha do tempo da história do mercado de eletricidade no Brasil (1983-2010).

Pré-criação do MME	Local	1883: Dom Pedro II inaugura o primeiro serviço de iluminação elétrica da América do Sul no RJ 1889: Inauguração da Usina de Marmelos (primeira central hidroelétrica da América do Sul) 1897: Criação da São Paulo Railway (primeira empresa fornecedora de energia do Brasil) 1988: Criação da Light and Power Company do Brasil 1934: Criação do Código das Águas pelo Decreto nº 24.643/1934 1939: Criação do Conselho Nacional de Águas e Energia Elétrica (CNAEE) 1945: Criação da Companhia Hidroelétrica do São Francisco (CHESF) 1946: Extensão do imposto único para a energia elétrica (primeira tarifação) 1954: Instituição do Fundo Federal de Eletrificação pela Lei nº 2.308/1954
De Juscelino Kubitschek até Itamar Franco	Mercado Verticalmente Integrado misto com Comprador Único	**1960: Criação do Ministério de Minas e Energia pela Lei nº 3.782/1960** **1961: Criação da Eletrobrás pela Lei nº 3.980-A** 1970: Interligação SE/S pela LT 230 kV Chavantes-Figueira 1972: Primeiro Plano Nacional de Desenvolvimento (I PND) 1973: Assinatura do tratado de Itaipu 1973: 1º Choque do Petróleo **1974: Criação da CEPEL** 1974: Segundo Plano Nacional de Desenvolvimento (II PND) 1975: Criação do Proálcool 1979: 2º Choque do Petróleo 1981: Interligação N/NE pela LT 500 kV V.Conde-B.Esperança 1982: Moratória do México 1984: Bipolo Itaipu 600 kV CC 1985: Programa Nacional de Conservação de Energia Elétrica (PROCEL) 1985: UTN Angra I (640 MW) 1986: Plano Cruzado 1987: Plano Bresser 1988: Política do Feijão com Arroz **1988: Fim do monopólio estatal no setor elétrico e implantação da expansão por leilões** 1989: Plano Verão 1990: Plano Collor I 1991: Plano Collor II 1991: UHE Itaipu – 18 UG (700 MW cada) 1992: Primeira usina eólica do Brasil instalada em Fernando de Noronha 1994: Plano Real 1994: Primeiro parque eólico conectado ao SIN
Governo FHC	Mercado de Atacado misto com Comprado Único	**1995: Aprovação da Lei de Concessões nº 8.987/1995** **1995: Criação do Mercado Livre de Energia pela Lei nº 9.074/1995** **1996: Criação da ANEEL pela Lei nº 9.427/1996** **1996: Projeto de reestruturação do Setor Elétrico Brasileiro** 1997: Crise cambial asiática 1998: Crise cambial russa **1998: Criação do Mercado Atacadista de Energia Nacional pela Lei nº 9.648/1998** **1998: Criação da ONS pela Lei nº 9.648/1998** 1999: Interligação N/S LT 500 kV Samambaia/Imperatriz com 1278 km 2001: Implantação do programa de eficiência energética da ANEEL pela Lei nº 10.295/2001 **2001: Crise de Abastecimento de Eletricidade no Brasil** 2001: Criação da Câmara de Gestão de Crise de Energia Elétrica pela Lei nº 2.147/2001 2001: UTN Angra II (1.350 MW) **2002: Novo marco do Mercado Atacadista de Energia pela Lei nº 10.4333/2002** 2002: Criação do Programa de Incentivo às Fontes Alternativas de Energia (Proinfa)
Governo Lula		2003: Criação do Programa Luz para Todos pelo Decreto nº 4.873/2003 **2004: Criação da CCEE, EPE e CMSE pela Lei nº 10.848/2004** 2005: Criação do selo PROCEL 2005: Descoberta do Pré-Sal na Bacia de Campos 2007: Primeiro leilão de energia eólica do ACR 2007: UHE Itaipu – 2 UGs (Conclusão da motorização, 700 MW cada) 2009: Primeiro leilão de energia de reserva, exclusivo para fontes eólicas 2009: Crise financeira internacional (subprime) 2010: EPE cria o sistema AMA para armazenamento de dados eólicos 2010: Resolução normativa n° 414/2002 da ANEEL sobre condições gerais de fornecimento

Fonte: O Autor.

Figura 16 – Linha do tempo da história do mercado de eletricidade no Brasil (2012-2021).



Fonte: O Autor.

Até 1934, com a criação do Código das Águas por meio do decreto nº 24.643/1934 (BRASIL, 1934), a demanda nacional de energia elétrica foi basicamente explorada por empresas estrangeiras do setor que fundaram suas bases nos estados do Rio de Janeiro e São Paulo. Esta foi uma fase onde o desenvolvimento do setor passava apenas pelo interesse local de políticos ou empresários, sendo considerada a eletricidade até então um bem de luxo. Com o Código das Águas todos os empreendimentos de exploração hidráulicos no país passaram a necessitar de autorização do governo federal, além do que também, com esta mesma legislação foram instituídas as primeiras tarifas de energia (TOLMASQUIM, 2015; JUNIOR, 2017).

Nos anos 40, especialmente após a segunda guerra mundial, a importação de equipamentos tornou-se dificultosa e, com uma planta gerida completamente por empresas estrangeiras, o Brasil passou a enfrentar dificuldades relacionadas a expansão da capacidade e atendimento da crescente demanda industrial. Assim, em 1945, no então governo de Getúlio Vargas foi criada a primeira companhia nacional de eletricidade, a Companhia Hidroelétrica do São Francisco (CHESF). A CHESF foi concebida para viabilizar a construção da usina hidroelétrica de Paulo Afonso, no Nordeste. Um marco histórico, pois a energia a ser produzida não seria utilizada para atendimento da indústria, mas sim, para desenvolvimento do país e atendimento a uma população pobre. Neste mesmo período foi também instituído o chamado Imposto Único. Primeira tributação sobre a eletricidade nacional. O final dos anos 40 e início dos anos 50 marca o momento de introdução de intervenção estatal no sistema elétrico brasileiro (TOLMASQUIM, 2015; JUNIOR, 2017).

Nos anos 50 as experiências de eletrificação de Minas Gerais, Rio Grande do Sul e São Paulo, até então eram os grandes casos de referência. O Governo Federal se baseou na experiência mineira para desenvolvimento da indústria nacional. Segundo Junior (2017), Minas Gerais foi escolhida como modelo devido a alguns traços marcantes, sendo (i) construção de grandes unidades de geração, (ii) opção pela tecnologia hidráulica, (iii) administração realizada por uma empresa independente, (iv) escala de projeto compatível com a indústria pesada de bens e (v) separação das atividades de geração e transmissão (a cargo do estado) e distribuição (a cargo da iniciativa privada). Associado a essas medidas foi criado o primeiro fundo federal de eletrificação (1954), um projeto de distribuição da arrecadação tributária para estados e municípios (1956), formulação de um plano nacional de eletrificação, criação da Central Elétrica de Furnas (1957), a instituição do Ministério de Minas e Energia (MME) (1960) e, em 1961, a criação da Eletrobrás. Com a Eletrobrás importantes decisões centralizadas para padronização foram tomadas, com destaque da definição em 1964 da unificação de frequência em 60 Hz pela Lei nº 4.454/1964 (BRASIL, 1964) e a inclusão de correção monetária nos ativos das concessionárias pela inflação. A marca principal deste período foi o reforço de uma visão da necessidade de desenvolvimento de uma indústria elétrica nacionalmente integrada.

Os anos 70 e 80 foram marcados pela transferência de diversos ativos da Eletrobrás para os estados e, também, do início da construção de Itaipu (com entrada em operação em 1983) e implantação do programa Proálcool (1975) em resposta a Crise do Petróleo. Outro ponto de destaque nos anos 70 foi a aquisição da concessionária Light. O que deu origem com base na divisão do patrimônio à Eletropaulo e Light Rio. No início dos anos 80, o Brasil com suas grandes dimensões territoriais passou a ter necessidade de expansão da transmissão. No entanto, neste mesmo período uma crise econômica causou impactos no setor criando endividamento de empresas de energia, racionamento de crédito e baixos investimentos. Com pouca ação do estado e diversas obras paradas, o início dos anos 90 foi marcado por um movimento em prol da reestruturação do setor elétrico brasileiro e do interesse pelo investimento de capital privado (TOLMASQUIM, 2015; JUNIOR, 2017).

O mercado de eletricidade brasileiro começou a surgir de fato a partir de 1995. Neste ano, após a aprovação da Lei de Concessões nº 8.987/1995 (BRASIL, 1995a), diversas privatizações como a Escelsa (Espírito Santo), Light Rio/CERJ (Rio de Janeiro) e Coelba (Bahia) foram executadas. Além disso foi criado o Mercado Livre de Energia pela Lei nº 9.074/1995 (BRASIL, 1995b) e a Agência Nacional de Energia Elétrica (ANEEL) por meio da Lei nº 9.427/1996 (BRASIL, 1996). Entre 1996 e 1998 foi realizado o Projeto de Reestruturação do Setor Elétrico Brasileiro (RE-SEB). Com o RE-SEB o desenho de mercado do Brasil foi instituído como um Mercado Atacadista de Energia pela Lei nº 9.648/1998 (BRASIL, 1998) e os direitos de distribuição e transmissão foram considerados monopólios naturais a serem explorados via concessão. Ainda em 1998 e via mesma lei, foi instituída a separação do operador de mercado e de sistema. Criando-se assim o Operador Nacional do Sistema (ONS).

Devido à má gestão do SEB, falta de investimentos em geração durante os anos 80 e 90 e uma crise hidrológica que afetou reservatórios, os anos 2000 tiveram início em meio a uma conturbada situação de desabastecimento que em 2001 culminou na necessidade de racionamentos. O Mercado Atacadista de Energia que acabara de ser instituído recebeu críticas e em 2004 iniciou-se a implantação do Marco Regulatório do Setor Elétrico Brasileiro (MSEB) criado pela Lei nº 10.433/2002 (BRASIL, 2002). Entre as novas diretrizes estava a Lei nº 10.848/2004 (BRASIL, 2004) que criou a Câmara de Comercialização de Energia Elétrica (CCEE), a Empresa de Pesquisa Energética (EPE) e o Comitê de Monitoramento do Setor Elétrico (CMSE). Junto aos novos órgãos administrativos do setor foram criados o ACR e o ACL. Foi estabelecido em 2004, pela primeira vez, o PLD como preço do mercado à vista para uso no ACL semanalmente. No ACR foram estabelecidos os leilões com PPAs de longo prazo (TOLMASQUIM, 2015).

Entre 2010 e 2015 o mercado de energia brasileiro foi marcado por diversas distorções de preço ocasionadas pela criação dos Contratos de Cotas de Garantia Física (CCGFs) e Contratos de Cotas de Energia Nuclear (CCENs) por meio da medida provisória nº 579/2012 (BRASIL, 2012) que afetaram grandes Usinas Hidroelétricas (UHEs) e Usinas Termonucleares (UTNs). Estes mecanismos diminuíram artificialmente os valores pagos por consumidores do ACR, pois via intervenção política foi alterado o modelo de renovação das concessões gerando rupturas contratuais entre os agentes. Uma das medidas que era praticada desde a implantação do MSEB era de que as distribuidoras deveriam estar com 100% da demanda contratada. Contudo, com o vencimento dos contratos de concessão, muitas distribuidoras não aceitaram o preço da renovação, ficando descontratadas e, as demais, ficaram contratadas mas produzindo abaixo do custo de mercado. Assim, a meta de contratação de 100% foi descumprida e grandes distribuidoras foram expostas ao PLD que atingiu o seu até então máximo valor histórico de R$ 822,00/MWh. Associado a este momento, em 2015 houve um período de estiagem prolongado, com um compromisso político do governo em diminuir o preço da eletricidade, este cenário fez com que a demanda não reagisse a escassez nos reservatórios, o que culminou na necessidade de acionamento de diversas Usinas Termoelétricas (UTEs) mais caras. O desequilíbrio do mercado causou inadimplência obrigando o governo a se utilizar de recursos do tesouro para socorrer os agentes, o que gerou um déficit público. Com o aumento do

custo de geração ocasionado pelo despacho de térmicas, uma revisão tarifária foi realizada e o preço da energia subiu 40%. Este fato levou a uma redução no consumo e, consequentemente, a uma recessão econômica (TOLMASQUIM, 2015). Esta recessão, tendo a situação do SEB como um dos seus principais motivadores, culminou em 2016 no *impeachment* da então presidenta Dilma Rousseff.

Nos últimos 5 anos, visando aplicar as lições aprendidas com a intervenção nos preços, o MME instituiu a consulta pública nº 21/2016 (MME, 2016) que trata da expansão e facilitação de entrada no ACL. Também em 2016 começou a transitar no senado o projeto de lei nº 232/2016 (BRASIL, 2016) sobre a portabilidade da conta de luz. Em 2017 foi promulgada outra consulta pública do MME, a de nº 33/2017 (MME, 2017), que diz respeito a criação de um mercado de capacidade brasileiro e extinção dos descontos para fontes incentivadas na transmissão evitando sobreinvestimento. Em 2018, o Brasil ainda enfrentou uma greve dos caminhoneiros que afetou temporariamente o preço da energia. Em 2019 foi instituído pela portaria nº 187/2019 (MME, 2019d) do MME um grupo de trabalho que, até o momento, tem estudado formas de modernizar o SEB. Junto a este grupo atua a Comissão Permanente para Análise de Metodologias e Programas Computacionais do Setor Elétrico (CPAMP). Foram também realizadas em 2019 as consultas públicas nº 76/2019 (MME, 2019b) e nº 77/2019 (MME, 2019c) que, respectivamente, tratam da formalização do papel do consumidor varejista e da ampliação das possibilidades de contratação. Por fim, estava programado a operacionalização do PLD em base horária para 2019, mas com a ocorrência da pandemia de COVID-19 que afetou o comportamento da carga e, consequentemente, dos preços, os agentes solicitaram (via consulta pública nº 71/2019 (MME, 2019a)) a postergação para a implantação em 2021. Fato que se concretizou a partir do dia 1º de janeiro deste mesmo ano.

O cenário macroeconômico brasileiro em 2021 foi marcado principalmente pela alta na inflação. A inflação no Brasil é medida pelo Índice Nacional de Preços ao Consumidor (IPCA) e ao final de 2021 foi de 10.06% (FGV, 2022a). Um valor que corresponde a quase o dobro da meta de 5.25% estipulada pelo Banco Central do Brasil e, também, aos 4.52% registrado em 2020 (Banco Central, 2022a). A alta dos preços das *commodities*, em especial do petróleo, impactaram diretamente o IPCA pois fazem parte do seu processo de formação. Por consequência, o preço dos combustíveis também aumentou elevando, por exemplo, o custo médio de abastecimento de veículos em 51% e do gás de cozinha em 40% (IPTL, 2022). Buscando-se efetuar um controle da inflação e, também, da desvalorização do Real (R$) frente ao Dólar Americano (US$) (-7.94%), o governo brasileiro iniciou um processo sistemático de incrementos na taxa de juros básica praticada no país. A chamada taxa SELIC em 2021 foi elevada de 2% para 9.25% (Banco Central, 2022b). Em relação a pandemia de COVID-19, a América do Sul foi declarada em 2021 como novo epicentro pela Organização Mundial da Saúde (OMS) (URBAN; NAKADA, 2021). O Brasil registrou no período 22 milhões de infectados e 619.056 mortes ocupando o terceiro lugar no *ranking* de países mais afetados (DONG; DU; GARDNER, 2022).

Apesar do cenário de fragilidade econômica, estados e municípios aplicaram políticas de distanciamento social para combate a COVID-19 que envolveram o fechamento de estabelecimen-

tos comerciais. Além disso, foi iniciada a aplicação de vacinas. Após um período de recesso no primeiro semestre, o PIB brasileiro apresentou alta de 4.6% no segundo semestre (FGV, 2022a). Este fato gerou uma expectativa de retomada econômica por parte do governo federal (Brasil, 2021). Esta expectativa também foi observada no consumo de energia elétrica que cresceu 4.1% em relação a 2020, retornando ao nível registrado antes do início da pandemia (CCEE, 2021). No entanto, do ponto de vista do ONS, o crescimento no consumo foi motivo de preocupação dado que em 2021 a expectativa de afluências nas bacias hidrográficas brasileiras não se confirmou (ONS, 2022a; ANA, 2022). Desta forma, os reservatórios do sistema elétrico chegaram a níveis críticos de abastecimento (ONS, 2022a). Visando contornar o problema, foram acionadas usinas térmicas fora da ordem de mérito. Esta operação foi custeada por meio de aumentos no PLD no ACL e, também, com aumento dos valores das tarifas dos consumidores do ACR, com destaque para a criação da chamada bandeira de escassez hídrica. Esta bandeira foi um mecanismo adicional adotado pelo ONS para elevar de forma excepcional os preços além dos limites até então permitidos (ONS, 2022b). Segundo Instituto Brasileiro de Estatística e Geografia (IBGE) o preço médio da energia elétrica em 2021 foi 21.21% mais alto que em 2020. Este valor representou 10.65% da variação do IPCA (IBGE, 2022; IBGE, 2022) sendo a energia elétrica, portanto, responsável direta pelo aumento da inflação.

Outras ações do ONS em relação a crise hídrica geraram um ganho de 14% em Energia Armazenada (EAR) em 2021. Dentre estas, pode-se destacar as relacionadas no Relatório Anual de Administração (ONS, 2022b) e no Protocolo de Flexibilização do Sistema Integrado Nacional (SIN) (ONS, 2021):

- Campanhas de conscientização incentivando o uso racional da energia elétrica.
- Flexibilizações no modelo de despacho e importação de energia de base térmica.
- Flexibilização nos níveis operacionais das usinas hidroelétricas de Itaipu, Furnas, Jupiá e Porto Primavera, as maiores do país.
- Flexibilização dos critérios de segurança e confiabilidade de operação passando de N-2 (que considera perdas duplas de linhas de transmissão) para N-1 (considera apenas perdas simples).
- Implantação de um programa de redução voluntária de demanda.
- Criação de grupos de trabalho para aprimoramento dos modelos computacionais que calculam o CMO.

Para os próximos anos, espera-se que o mercado se torne cada vez mais aberto aos consumidores, visto que pela portaria nº 465/2019 (MME, 2019e) do MME está previsto que gradualmente, até 2026, a demanda contratada mínima exigida para ingresso no ACL seja reduzida até não mais existir. Essa abertura, no entanto, está condicionada as condições estruturais do SEB, assunto tratado a seguir.

2.3.1 Panorama atual

O consumo total de energia do Brasil no ano de 2019 [27] foi de 546 TWh e a capacidade instalada até o final de 2020 era de 174 GW. Estes valores fazem com que o SEB figure como o nono maior mercado de eletricidade do mundo e o primeiro da América Latina (GRIEBENOW; OHARA, 2020; HOCHBERG; POUDINEH, 2021). De uma perspectiva sociodemográfica, mesmo com a desaceleração do crescimento populacional que vem acontecendo desde a década de 90, o número de habitantes continuará a aumentar à uma taxa média projetada de 0,5% a.a. até 2030 (IBGE, 2022).

Hoje o consumo por habitante residencial no Brasil é de 2581 kWh/ano. Um valor ainda modesto, visto que num contexto continental é menor do que o observado em países sul-americanos como o Chile (3880 kWh/ano), Uruguai (3085 kWh/ano), Argentina (3075 kWh/ano) e Venezuela (2719 kWh/ano) (The World Bank, 2022). No entanto, com a expectativa de crescimento da economia, projeta-se também que o consumo de eletricidade *per capita* irá evoluir com taxa média de 2,6% a.a., sendo que, segundo o último Plano Decenal de Expansão (PDE) do MME (EPE, 2020b) em 2030 o consumo total do país chegará a 762 TWh. Um valor que se realizado, colocaria o Brasil hoje em quinto lugar no mundo, atrás apenas de China (6510 TWh), EUA (3865 TWh), Índia (1230 TWh), Rússia (922 TWh) e Japão (918 TWh) (ENERDATA, 2022). Os dados apresentados na Tabela 1 sumarizam as condições atuais e, também, este possível cenário futuro projetado pela EPE para até 2030.

Tabela 1 – Indicadores do consumo de eletricidade com projeções para 2025 e 2030.

Indicador	**2019**	**2025**	**2030**	**2019-2030 (% a.a.)**
População (milhões)	211	219	225	0,5
Consumo total (TWh)	546	640	762	3,1
Autoprodução não injetada (TWh)	58	65	77	2,6
Consumo total *per capita* (kWh/hab/ano)	2581	2922	3407	2,6
Consumo por consumidor residencial (kWh/mês)	162	177	197	1,8
Número de consumidores residenciais (milhão)	73,1	81,1	87	1,6
Total de perdas no SIN (%)	19,4	19,3	18,3	-
Intensidade Energética da Economia (MWh/10^3 R$)	0,1	0,2	0,2	-

Fonte: EPE (2020a)

Para atender a este mercado consumidor, hoje o SIN conta com 10155 unidades de geração, divididas em sete categorias principais. Conforme pode ser visualizado na Tabela 2, dentre estas categorias destacam-se os empreendimentos com base hidrológica que respondem sozinhos por 62,53% da participação. Os mesmos são seguidos das termoelétricas com 24,58%, eólicas com

[27] As informações relatadas nesta seção dizem respeito, principalmente, aos dados contidos nos relatórios públicos da ANEEL (ANEEL, 2022b) e EPE (EPE, 2020a; EPE, 2020b; EPE, 2022).

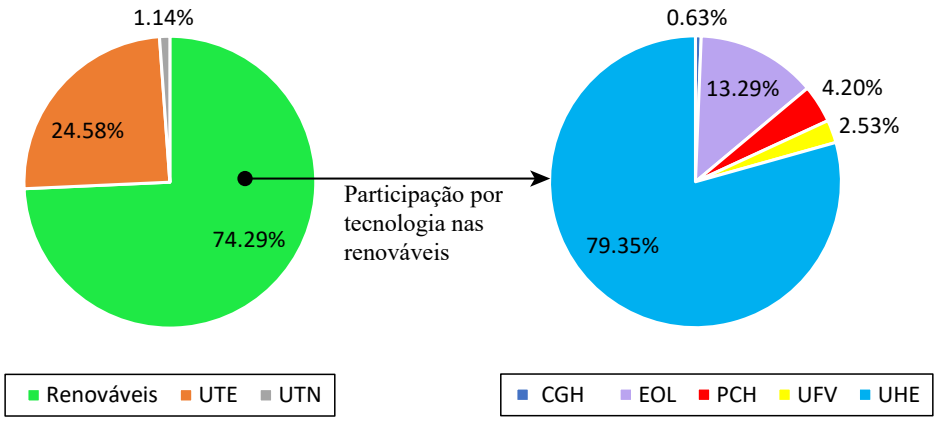
9,87%, solar fotovoltaica com 1,88% e, por fim, nucleares com 1,14%. Esses percentuais de participação refletem a característica principal da matriz energética de eletricidade brasileira: Ser fortemente baseada em fontes renováveis de energia, sendo que, neste contexto, renováveis correspondem a 74,29% contra 25,71% das não-renováveis. Na Figura 17 pode ser visualizada graficamente as proporções de cada tecnologia.

Tabela 2 – Potência fiscalizada por tipo de unidade geradora e participações.

Tipo	Potência (GW)	Unidades	Participação (%)
Central Geradora Hidroelétrica (CGH)	823,22	746	0,47
Central Geradora Eolielétrica (EOL)	17254,36	1027	9,87
Pequena Central Hidroelétrica (PCH)	5451,18	543	3,12
Central Geradora Solar Fotovoltaica (UFV)	3288,42	4440	1,88
Central Geradora Undi-elétrica (CGU)	-	-	menor que 0,1%
Usina Hidroelétrica de Energia (UHE)	103026,88	223	58,94
Usina Termoelétrica de Energia (UTE)	42956,26	3172	24,58
Usina Termonuclear (UTN)	1990,00	3	1,14
TOTAL	174790,38	10155	100,00

Fonte: ANEEL (2022b), Hochberg e Poudineh (2021)

Figura 17 – Participação de tecnologias de geração no SIN.

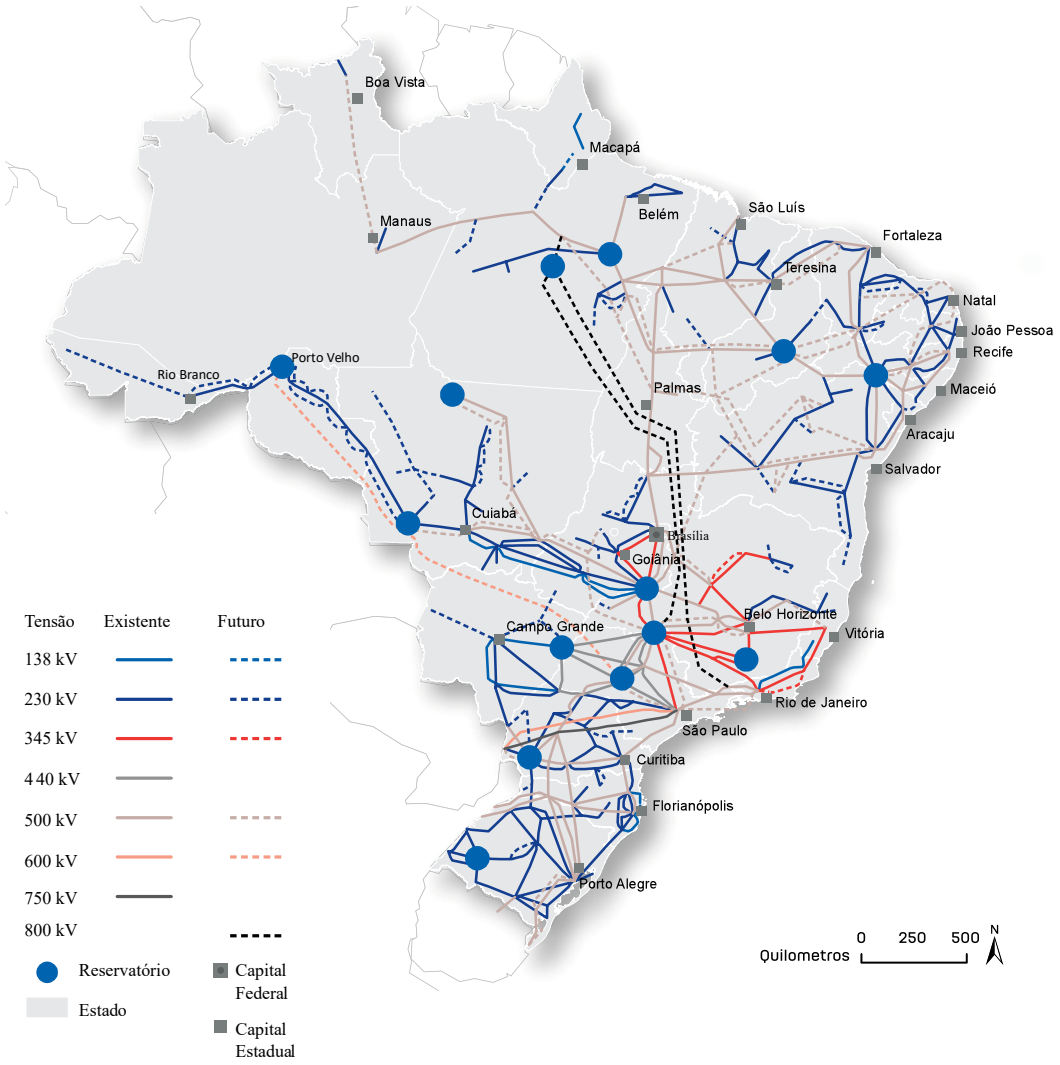


Fonte: Adaptado de ANEEL (2022b).

Outra característica marcante do SIN é a sua cobertura conectada de dimensão continental. O mesmo possui atualmente linhas de transmissão que se estendem por 145543 quilômetros compreendendo tensões entre 230 kV e 800 kV. Na Figura 18 pode ser visualizado o mapa do SIN.

Estas conexões levam eletricidade a 99% da demanda do país e dizem respeito as conexões entre os quatro subsistemas que compõe o SEB sendo estes Sudeste/Centro-Oeste (SECO), Sul (SUL), Nordeste (NE) e Norte (NORTE). As interligações possibilitam "o aproveitamento

Figura 18 – Mapa do SIN.



Fonte: Griebenow e Ohara (2020).

otimizado dos recursos energéticos disponíveis no sistema, proporcionando a gestão da complementaridade sazonal entre bacias hidrográficas e intradiária entre fontes renováveis como eólica e solar, bem como o aproveitamento dos efeitos do portfólio" (EPE, 2020a).

A interligação NORTE-SUL (que passa pelo SECO) possui capacidade atual de escoamento de 4 GW em Corrente Alternada (CA) e 8 GW em Corrente Contínua (CC). É formada por três circuitos de 500 kV e dois bipolos CC de 800 kV. A interligação NORTE-NE possui capacidade de 5,5 GW e é constituída por 8 Linhas de Transmissão (LTs) em 500 kV. Além destas, existe também uma rede de 230 kV que interliga as subestações de Peritoró e Coelho Neto no Maranhão à subestação de Teresina no Piauí. Já a interligação NE-SECO suporta 2,5 GW de capacidade. Originalmente constituída pela LT 500 kV Serra da Mesa - Rio das Éguas e pela LT 500 kV Luziânia - Rio das Éguas, a mesma em 2020 passou a complementar o novo eixo de 500 kV entre as subestações Rio das Éguas, Arinos e Pirapora 2. Essa interligação já conta com licitação executada para complementação que fará com que a capacidade da mesma seja acrescida de 5 GW

até 2023. Por fim, a interligação SUL-SECO possui capacidade máxima de 10,5 GW (dependendo do patamar de carga e do sentido do fluxo) e é formada, principalmente, por quatro trechos em 500 kV e um em 765 kV. Com implantação prevista para 2023 de três novos trechos em 525 kV e um seccionamento de linhas nas subestações de Joinville Sul e Gaspar 2, esta interligação aumentará sua capacidade em 1 GW (EPE, 2020a).

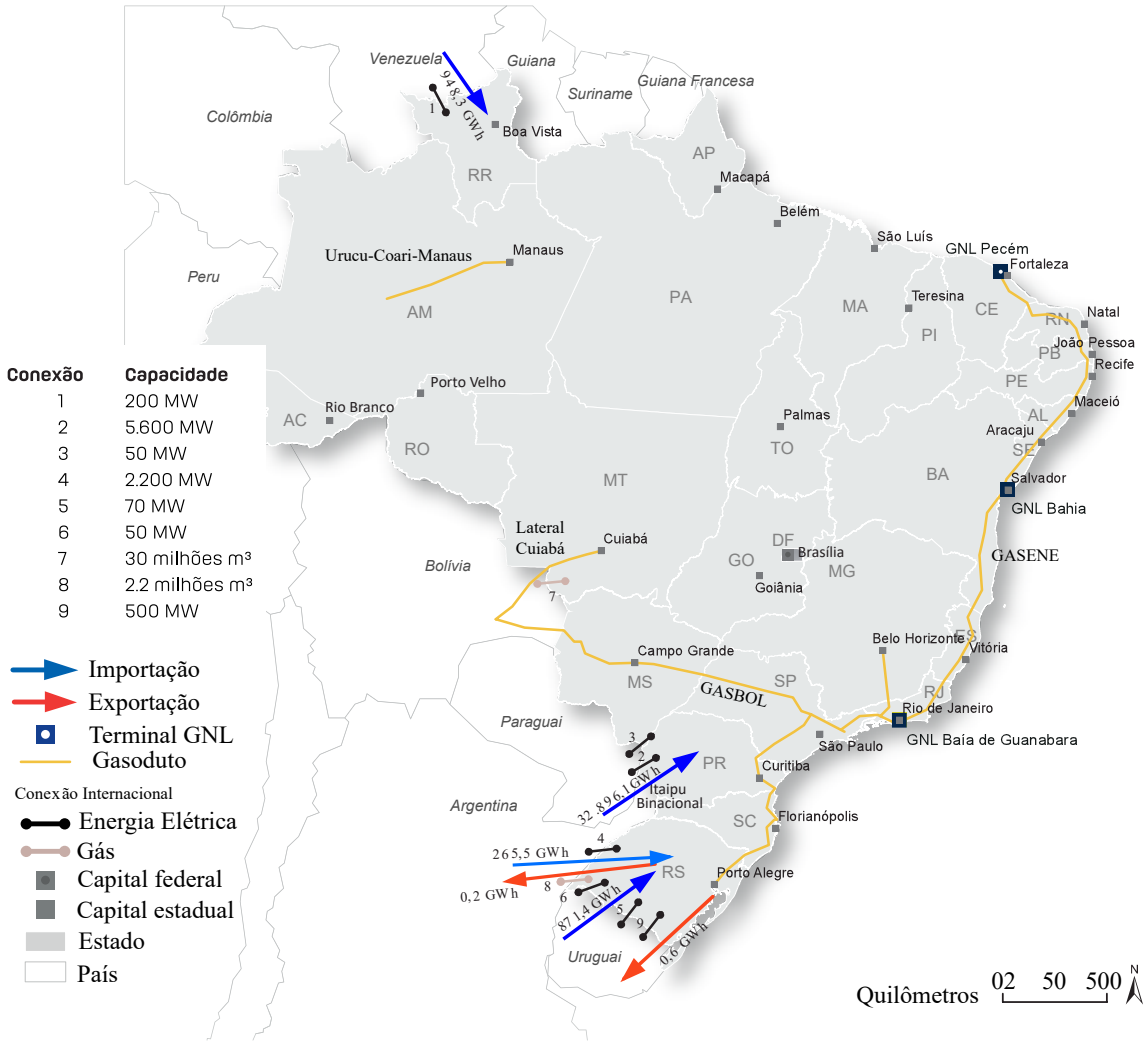
Já os outros 1% da demanda não atendidos pelo SIN dizem respeito a 270 sistemas isolados localizados principalmente nos estados do Acre, Amapá, Amazonas e Pará. Boa Vista atualmente é a única capital do Brasil que ainda não está conectada ao SIN. Em março de 2019 o governo brasileiro encerrou o contrato entre a Eletronorte e a empresa estatal venezuelana Corpoelec. O contrato mantinha possibilidade de importação por meio de uma conexão com capacidade de 200 MW. Com a finalização do acordo, o Brasil precisou realizar a contratação de nove novos geradores via leilão. Estes geradores consumiam à época do início das operações entre 700 e 1 milhão de litros de diesel por dia (GRIEBENOW; OHARA, 2020; OLIVEIRA, 2020). Para atendimento de Roraima, único estado sem conexão alguma com o SIN, foi inaugurada no início de 2022 a UTE Jaguatirica II com capacidade de 141 MW (MME, 2022b). Existe um projeto de integração entre Manaus e Boa Vista parado desde 2011 mas que em 2022 obteve licença do Ministério Público para retomada (MME, 2022a).

Além das conexões internas e os sistemas isolados, o governo brasileiro também mantém contrato de intercâmbio de energia com a Argentina (2250 MW), Uruguai (570 MW) e Paraguai (14 GW). Há que se dar destaque para a situação entre Brasil e Paraguai, visto que até 2023 toda a energia excedente produzida pela usina binacional de Itaipu e de posse paraguaia será importada de forma líquida pelo Brasil. Dada esta situação, em 2018 o país importou um total de 34980 GWh, um valor que representa 6,5% de todo o consumo. O total exportado foi de 1 GWh. Em relação a Argentina e Uruguai os intercâmbios de energia acontecem de três formas sendo (i) sem envolvimento de transações financeiras, para fins de testes ou emergência. Neste caso a eletricidade importada é devolvida; (ii) por meio de contratos bilaterais no ACL, onde os agentes podem negociar a compra de energia argentina como opção de suprimento e; (iii) exportação de excedentes em caráter excepcional em troca de quantidades iguais no futuro ou reembolso (GRIEBENOW; OHARA, 2020).

O mapa da Figura 19 mostra os intercâmbios entre Brasil e seus países vizinhos, incluindo os relacionados a Gás Natural e Gás Natural Liquefeito (GNL).

Quanto as companhias detentoras do SEB, atualmente a maior empresa do setor ainda é a Eletrobrás. Como visto na seção anterior, a Eletrobrás é uma herança do modelo de monopólio praticado no Brasil. Com economia mista e controle majoritário do governo federal, ao final de 2019, juntamente com suas subsidiárias (Amazonas GT, Eletrosul, CHESF, Eletronorte, Eletronuclear, Furnas, Cepel e Eletropar), a Eletrobrás controlava 30% da capacidade instalada de geração, 44% da transmissão, 50% das ações da binacional de Itaipu e participação na gestão da Eletronuclear (ELETROBRÁS, 2022). Desde novembro de 2019, quando um primeiro projeto foi apresentado ao congresso nacional, um processo de privatização da Eletrobrás vem sendo

Figura 19 – Mapa de intercâmbios do SIN.



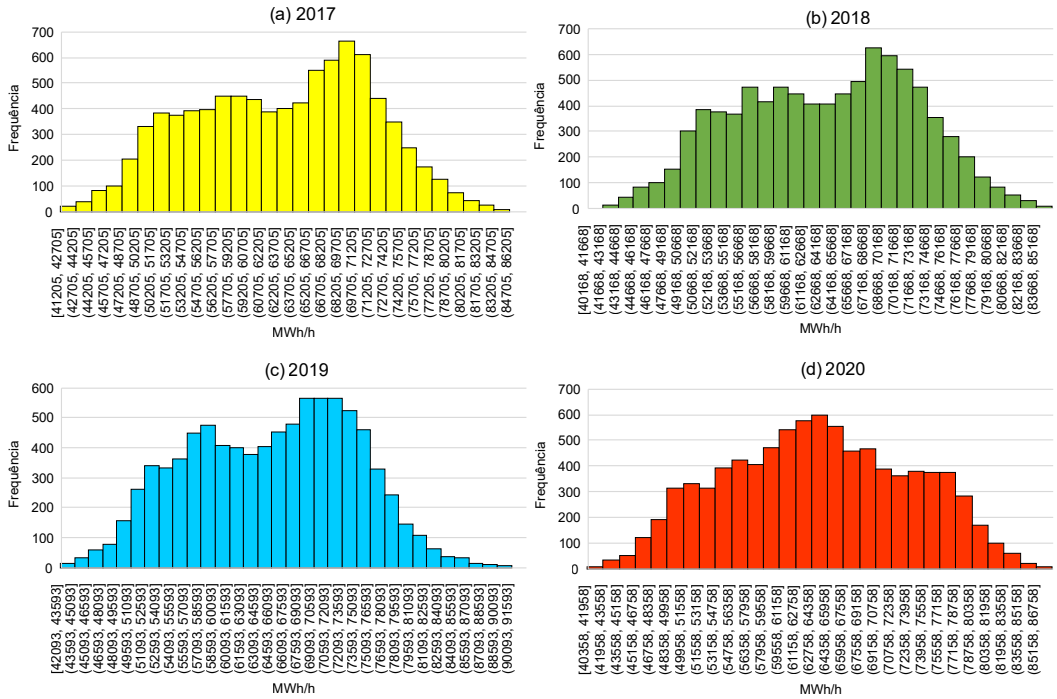
Fonte: Griebenow e Ohara (2020).

fomentado pelo governo. Com a publicação da medida provisória nº 1.031/2021 (BRASIL, 2021) diretrizes de privatização foram estabelecidas, sendo que, na prática, a privatização ocorrerá via venda de novas ações, fazendo com que o percentual acionário da união caia para menos que 50% (BRASIL, 2021). O governo brasileiro espera capitalizar valores da ordem de R$ 50 bilhões com o processo. A medida provisória ainda trata da renovação de contratos de concessão de usinas hidroelétricas para mais 30 anos. Já no setor privado, há que se dar destaque a Engie Brasil, empresa de propriedade belgo-francesa que até 2020 foi a maior companhia privada na geração com capacidade instalada de 10,21 GW e 61 usinas. A empresa adquiriu em 2019 a TAG e se tornou a detentora da mais extensa malha de transporte de gás natural do país (Engie Brasil, 2022). Ela é a maior comercializadora de energia também, respondendo por 5,5% de participação. No segmento de comercialização vale destacar que as dez maiores participações até o final de 2021 eram de empresas do setor privado. Outras empresas importantes são a Enel SP, maior distribuidora, a

CPFL Energia, empresa chinesa com maior estrutura integrada e a CTEEP, maior empresa privada do ramo de transmissão. A Tabela 3 mostra os *rankings* das maiores empresas em cada segmento.

Finalmente, cabe uma breve discussão sobre o perfil de carga atualmente do SEB. Conforme visto na Figura 20, entre 2017 e 2019, a distribuição dos valores de carga consumidos hora a hora foi semelhante, com picos variando entre 550 MWh/h e 650 MWh/h. A demanda máxima anual, conforme mostrado na Figura 21, ocorre nos meses de verão. O horário de pico do SEB acontece, desde 2008, no meio da tarde. Devido as tarifas praticadas entre 17h e 21h, grandes consumidores do ACR optaram por realizar geração própria. Em 2015, a EPE estimou uma redução de 7 GW a 9 GW (cerca de 10% da demanda) na curva de carga do SIN devido a geração distribuída no Brasil (EPE, 2015). Desde 2017, o maior pico de consumo foi de 90120 MWh/h (03:00 em 30/01/2019) e o mínimo foi de 40167,75 (07:00 de 27/05/2018). Como pode se notar também na Figura 20, a pandemia de COVID-19 teve significativo efeito na distribuição dos valores de carga, mudando o perfil de consumo, no entanto, apesar do perfil ter mudado, a média de consumo não caiu conforme pode ser visto na Tabela 4.

Figura 20 – Distribuição entre 2017 e 2020.



Fonte: O Autor.

Tabela 3 – Maiores agentes do Brasil.

Nº	Usina	Potência (GW)
1º	Belo Monte	11233,1
2º	Tucuruí	8535
3º	Itaipu (Parte Brasileira)	7000
4º	Jirau	3750
5º	Santo Antônio	3568
6º	Ilha Solteira	3444
7º	Xingó	3162
8º	Paulo Afonso IV	2462,4

(a) Usinas.

Nº	Empresa	Part. (%)	Tipo
1º	Enel	10,3	Privada
2º	CEMIG	8,1	Mista
3º	COPEL	6,5	Mista
4º	CPFL/State Grid	6,3	Privada
5º	Light	5,9	Privada
6º	COELBA	5,3	Privada
7º	Celesc	4,5	Mista
8º	CELG - Enel	3,5	Privada

(b) Distribuidoras.

Nº	Agente	Potência (GW)
1º	Norte Energia	11233,1
2º	CHESF	10347,43
3º	Furnas-Centrais	9407,83
4º	Eletronorte	8973,15
5º	Itaipu Binacional	7000
6º	Petrobrás	6426,73
7º	Rio Paraná	4995,2
8º	Engie Brasil	4392,44

(c) Agentes em capacidade.

Nº	Empresa	Part. (%)	Tipo
1º	Engie Brasil	5,5	Privada
2º	Votener	4,2	Privada
3º	CPFL	4,2	Privada
4º	BTG Pactual	4,1	Privada
5º	EDP	4,1	Privada
6º	Nova Energia	3,1	Privada
7º	Principal	2,7	Privada
8º	Boven	2,6	Privada

(d) Comercializadoras.

Nº	Empresa	Quilômetros	Tipo
1º	Eletrobrás	64833	Capital misto
2º	CTEEP	17944	Privada
3º	State Grid	11487	Estatal chinesa
4º	TAESA	10201	Privada
5º	CEMIG	9830	Capital misto
6º	COPEL	5090	Capital misto
7º	ALUPAR	3612	Privada
8º	Equatorial	3281	Privada

(e) Empresas de transmissão.

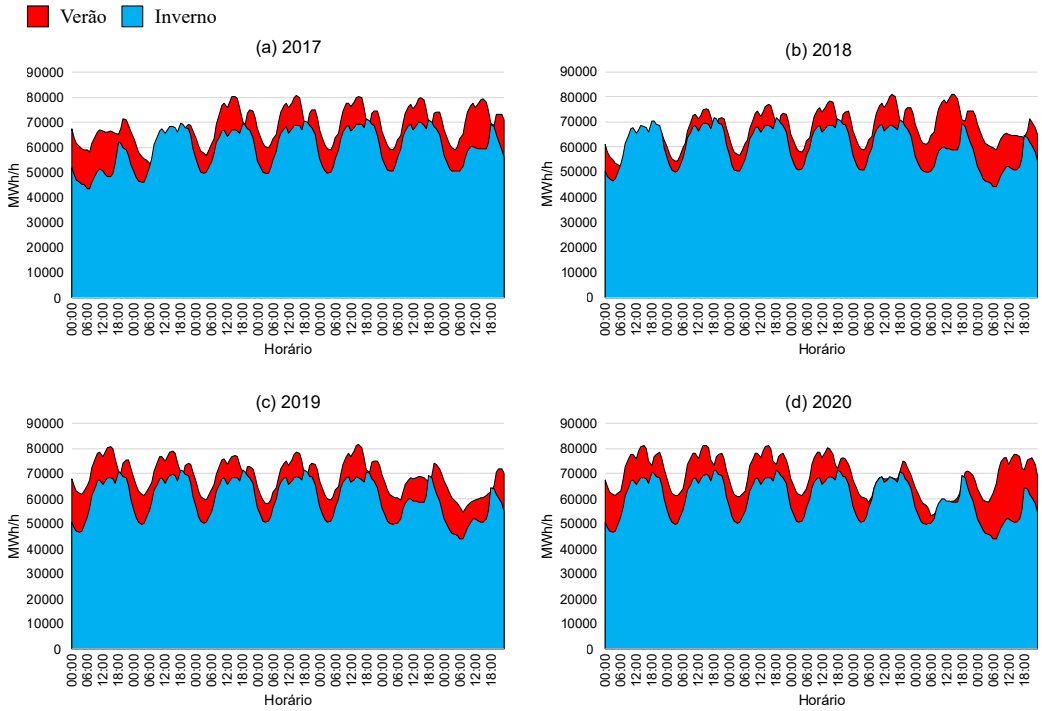
Fonte: ANEEL (2022b) e Griebenow e Ohara (2020).

Tabela 4 – Sumarização estatística dos valores de carga (MWh/h) entre 2017 e 2020.

Ano	Máximo	Mínimo	Média	Desvio padrão
2017	85305,84	41205,24	62687,05	7359,38
2018	84889,51	40167,75	63281,48	7338,65
2019	90120	42093,5	64589,27	7494,89
2020	86509,74	40358,31	63419,07	7638,11

Fonte: O Autor.

Figura 21 – Comparativo de comportamento da carga entre verão (04/02 a 10/02) e inverno (13/08 a 19/08).



Fonte: O Autor.

2.3.2 Arquitetura de mercado

O SEB está organizado com base na segunda reforma promovida em 2004 (MSEB). O arranjo está representado na Figura 22 e pode ser entendido sob uma perspectiva de quatro níveis, sendo:

a) Legislativo: Neste nível estão os órgãos responsáveis pela definição das macros diretrizes de funcionamento do SEB. Os dois órgãos mais importantes são o Congresso Nacional, responsável por elaborar e aprovar as leis de interesse do setor e a Presidência da República que, além de propor leis, emite decretos que podem definir detalhes de

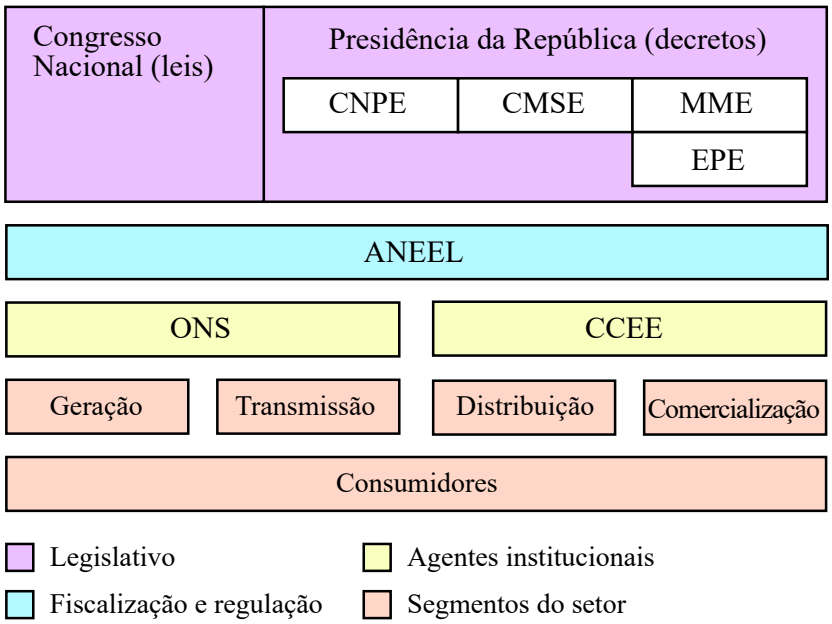
funcionamento do SEB. Neste nível ainda se encontram, no âmbito da Presidência da República, (i) o Conselho Nacional de Política Energética (CNPE), conselho destinado a elaborar as grandes diretrizes energéticas do país, especialmente para longo prazo; (ii) o CMSE, comitê que monitora a segurança do suprimento de eletricidade definindo ações operacionais que, normalmente, tornam-se novas regulamentações; (iii) o MME, ministério do governo federal responsável pela implementação das políticas energéticas, desenvolvimento do mercado, atração de investimentos e eficiência alocativa e; por fim, (iv) a EPE, empresa estatal vinculada ao MME que tem por finalidade prestar serviços na área de estudos energéticos e pesquisa, subsidiando o planejamento do setor, tanto para eletricidade, como para as demais fontes de energia.

b) Fiscalização e regulação: Neste nível encontra-se a ANEEL, uma autarquia de regime especial subordinada ao MME. Com atividades tendo início em 1997, a ANEEL tem como principais atribuições a regulação da geração, transmissão e comercialização de energia elétrica; a fiscalização das concessões, permissões e serviços de energia; implementação de políticas referentes a exploração de energia elétrica e aproveitamento de potenciais hidráulicos; estabelecimento de tarifas; gestão de divergências administrativas entre os agentes consumidores e os agentes do setor e; promoção de atividades de outorga de concessão, permissão e autorização para empreendimentos energéticos (ANEEL, 2022a).

c) Agentes institucionais: Atuando com base nas diretrizes impostas pelos órgãos de nível legislativo, fiscalização e regulação, estão dois agentes institucionais. O primeiro é o ONS, órgão responsável pela coordenação e controle da operação das instalações de geração e transmissão no SIN e sistemas isolados. É uma pessoa jurídica de direito privado, sob a forma de associação civil sem fins lucrativos que promove o acesso a rede elétrica de forma não discriminatória, a expansão do SIN, a melhoria das condições operacionais e a otimização da operação, visando sempre o menor custo e garantia de segurança do sistema (ONS, 2022). O segundo agente é a CCEE. A CCEE reúne empresas de geração, produtores independentes, autoprodutores, distribuidoras, comercializadoras, importadoras e exportadoras de energia, além de consumidores livres e especiais. É uma associação civil sem fins lucrativos que substituiu o Mercado Atacadista de Energia em 2000. Tem como principais atribuições a implantação e divulgação dos procedimentos de comercialização, gestão de contratos, registro de dados de energia gerada e consumida no país, realização de leilões, apuração de infrações, cálculo de penalidades financeiras e valoração do PLD (CCEE, 2022h).

d) Segmentos do setor: Os segmentos do setor são o último nível da cadeia sendo compostos pelos agentes participantes do mercado com geração, transmissão, distribuição, comercialização e consumo. A existência do mercado está condicionada a existência dos segmentos. Estão subordinados às regras impostas pelos agentes institucionais.

Figura 22 – Arranjo institucional do SEB.



Fonte: Adaptado de Tolmasquim (2015) e Viana (2018).

Os agentes participantes do mercado de energia brasileiro podem atuar em dois ambientes de contratação, sendo:

a) Ambiente de Contratação Regulada: São elegíveis para venda no ACR os agentes de geração e comercializadoras. Já para compra, as empresas de distribuição que devem atender os chamados consumidores cativos. Um consumidor cativo é aquele que não tem opção de escolha do seu fornecedor, sendo obrigado a consumir do distribuidor que detém direitos de fornecimento na sua região. Compradores e vendedores realizam negociações no ACR por meio de leilões de energia [28] de menor tarifa ou chamadas públicas para geração distribuída. Os contratos celebrados [29] são bilaterais com cláusulas pré-definidas pela ANEEL.

b) Ambiente de Contratação Livre: No ACL existem os consumidores livres e os especiais [30]. Um consumidor livre é aquele que pode selecionar seu fornecedor por meio de livre negociação. Já o consumidor especial é aquele que tem o direito de adquirir energia de Pequenas Centrais Hidroelétricas (PCHs) ou fontes incentivadas (eólica, biomassa ou solar) (CCEE, 2022a). De acordo com a portaria nº 514/2018 do MME (BRASIL,

[28] No ACR podem ser realizados leilões de fontes alternativas, excedentes, estruturantes, energia de reserva, energia nova, energia existente, compra e ajuste. Para maiores detalhes sobre os tipos de leilões sugere-se consultar CCEE (2022d).

[29] Detalhes sobre os tipos de contratos do ACR podem ser consultados em CCEE (2022b).

[30] Excepcionalmente, existem também os Consumidores Parcialmente Livres (CPLs). Um CPL é aquele que resolve contratar parte da sua energia junto ao ACL, no entanto, mantém uma outra parcela contratada com a distribuidora local no ACR. O mesmo possui regulação própria pela resolução normativa nº 376/2009 da ANEEL (ANEEL, 2009).

2018) são atualmente elegíveis para participar como compradores no ACL os consumidores livres com demanda contratada mínima de 2 MW e os consumidores especiais isoladamente ou em regime de comunhão [31] entre 500 kW e 2 MW. Do ponto de vista da venda, participam do ACL os geradores, produtores independentes, autoprodutores e comercializadoras. Os contratos no ACL são bilaterais e firmados livremente em balcão. Além disso, é possível realizar o repasse de montantes de energia acordados nos contratos para outros consumidores por meio das comercializadoras. Este fato faz com que a energia em si não seja negociada, mas dá aos contratos um papel de título com direitos ao portador, criando dessa forma um meio de especulação financeira.

Desde a promulgação do MSEB em 2004, a ideia principal do governo foi a implantação de um mercado atacadista nacional, nesse sentido, com a evolução do arranjo institucional observada nos últimos 15 anos pode-se dizer que houve, de fato, uma consolidação deste tipo de *design* no Brasil. Hoje a compra e venda de energia no SEB é feita em atacado, pois apesar de alguns consumidores poderem atuar livremente, a maioria ainda não. Em dezembro de 2020, o consumo contratado via ACR representou 68,23% contra 37,76% no ACL (CCEE, 2020a). Apesar do ACL ainda ser menor, o número de agentes no setor vem aumentando conforme pode ser visto na Figura 23. Até o final de 2020 o número de agentes era de 10543 e o montante total liquidado no mês de dezembro deste mesmo ano foi de R$ 11,28 bilhões, referentes a 148,82 GW médios transacionados (CCEE, 2020a).

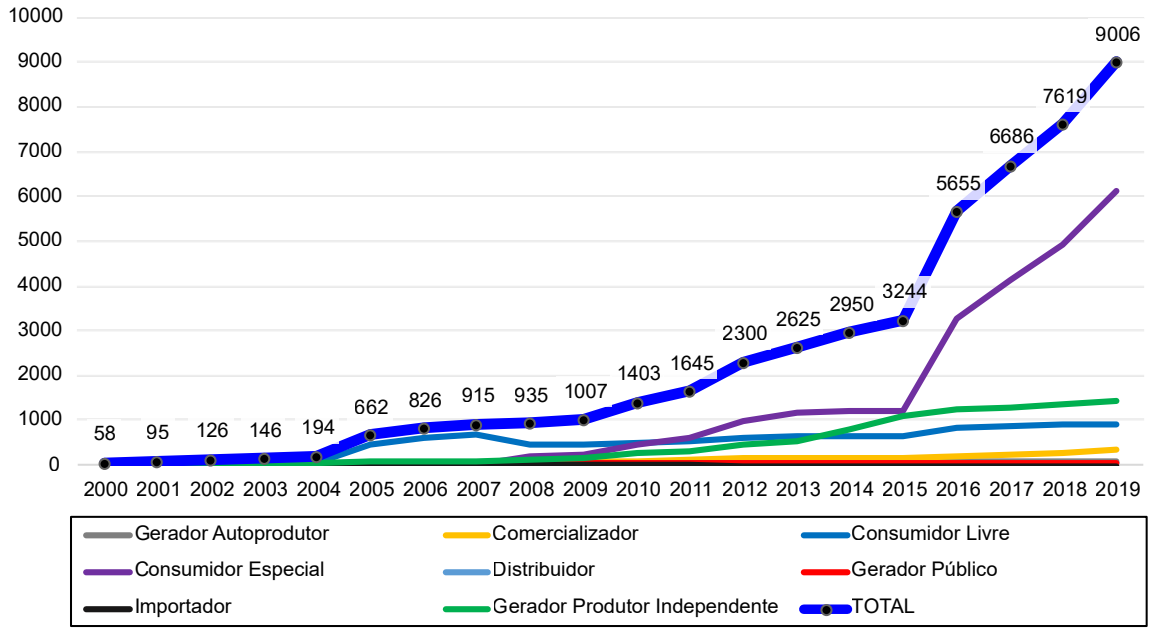
Quanto ao tipo de despacho, no Brasil pratica-se o *tight pool*. Isto ocorre, pois existe um controle centralizado de operações realizados pelo ONS que define a ordem de mérito de acionamento das usinas. Essa operação acontece de forma coordenada com a CCEE e é necessário que seja centralizada, pois as unidades de geração, além de pertencerem a múltiplos proprietários, em dadas situações compartilham o mesmo curso d'água. A atuação conjunta que define o despacho está intimamente relacionada com a forma de planejamento da operação e a dinâmica de mercado. Esses dois pontos são discutidos adiante.

2.3.2.1 Planejamento da operação

Sendo o Brasil um mercado *tight pool*, o despacho ocorre com base no mínimo custo de operação, acionando os geradores por ordem de mérito até o atendimento total da carga. De um ponto de vista básico, o mínimo custo é valorado em sistemas hidroelétricos com base nas condições hidrológicas e em sistemas termoelétricos com base no custo do combustível. Neste sentido, as usinas que apresentarem o menor custo de operação serão despachadas primeiro, considerando o horizonte de operação. São despachadas, inicialmente, as unidades de geração inflexíveis, na sequência, a ordem de mérito é posta em prática até o atendimento da demanda. Após a demanda ser atingida, uma parcela da capacidade de geração pode vir a não ser despachada,

[31] Mesmo CNPJ ou área contígua.

Figura 23 – Evolução do número de agentes no ACL.



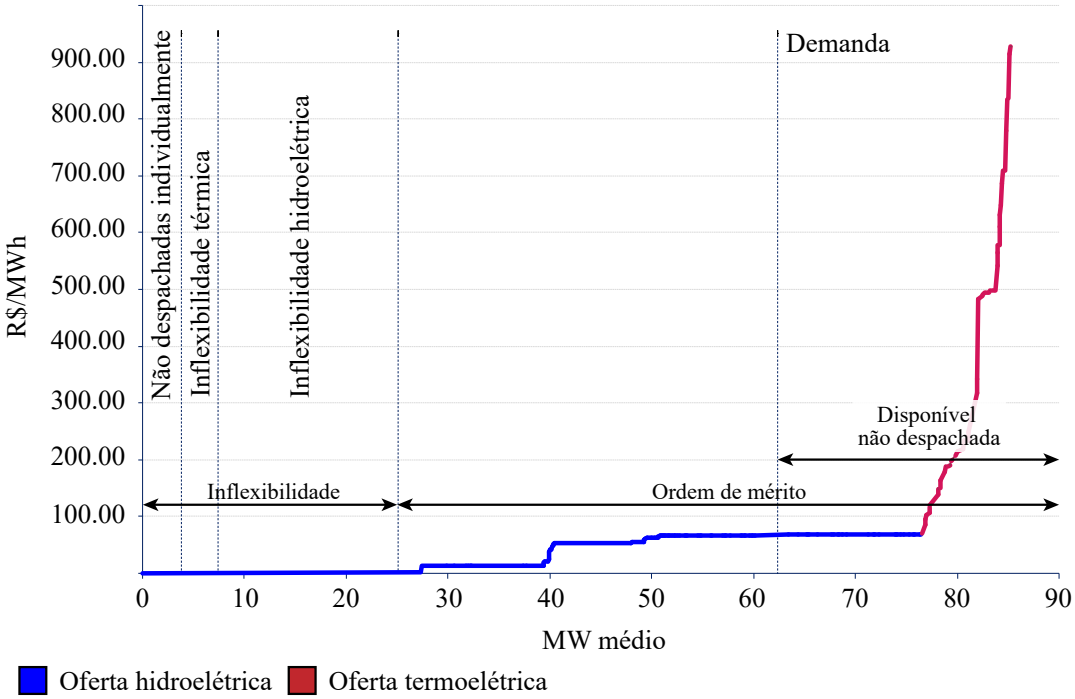
Fonte: Adaptado de CCEE (2020b).

ficando assim ociosa. Na Figura 24 pode ser visto um exemplo de curva de oferta real de energia no SIN onde houve situação de não despacho.

Sendo o Brasil um país com matriz composta (até o final de 2021) por 62,53% de geração hidráulica e 24,58% termoelétrica, a definição da ordem de mérito é quase que completamente baseada na resolução de um problema de otimização hidrotérmico. Esse problema está fundamentalmente associado com a decisão de armazenar ou não a água contida nos reservatórios. Sendo portanto acoplado no tempo, deve-se no presente tomar essa decisão considerando as possíveis consequências futuras relacionadas a (i) falta de água, em caso de não haver afluências o suficiente para manter os reservatórios em níveis seguros; ou, (ii) excesso de afluências, o que acarretaria na necessidade de realizar vertimentos e assim, por consequência, desperdiçar água. Os critérios para planejamento da operação podem ser vistos no diagrama da Figura 25.

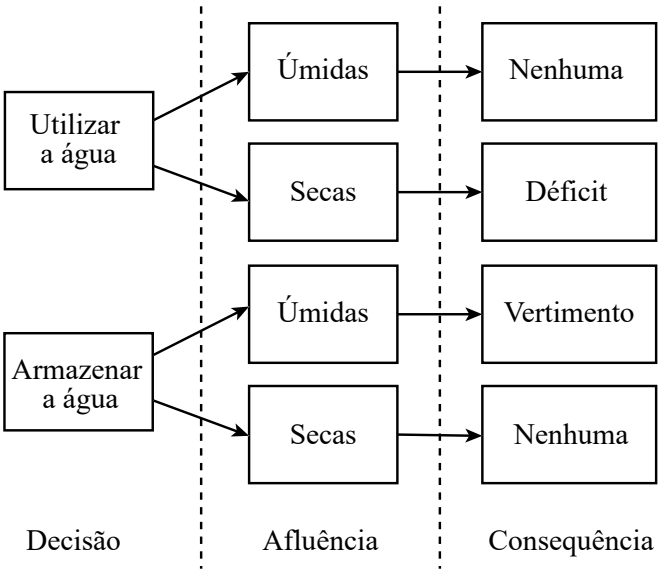
Esse relacionamento entre a decisão, a afluência e as consequência operativas em um sistema hidrotérmico, dizem respeito ao fato de que, ao se decidir economizar a água, deverão ser acionadas usinas de geração termoelétricas. Sendo estas usinas mais caras e poluentes, em caso de acionamento das mesmas no presente, o custo imediato da água passará a ser baixo. No entanto, se as reservas hidráulicas forem utilizadas com o objetivo de minimizar os custos térmicos, a incerteza relacionada a uma possível situação de seca futura levará a uma condição onde o custo da água passará a ser alto, devido a diminuição do volume disponível. Essa relação é denominada Função de Custo Imediato (FCI) e determina o valor da água com base nos custos de geração térmica. Com a mesma ideia de raciocínio, mas levando-se em conta a situação contrária, se tem

Figura 24 – Curva de oferta e despacho do SIN.



Fonte: Adaptado de CCEE (2012).

Figura 25 – Critério para planejamento da operação.

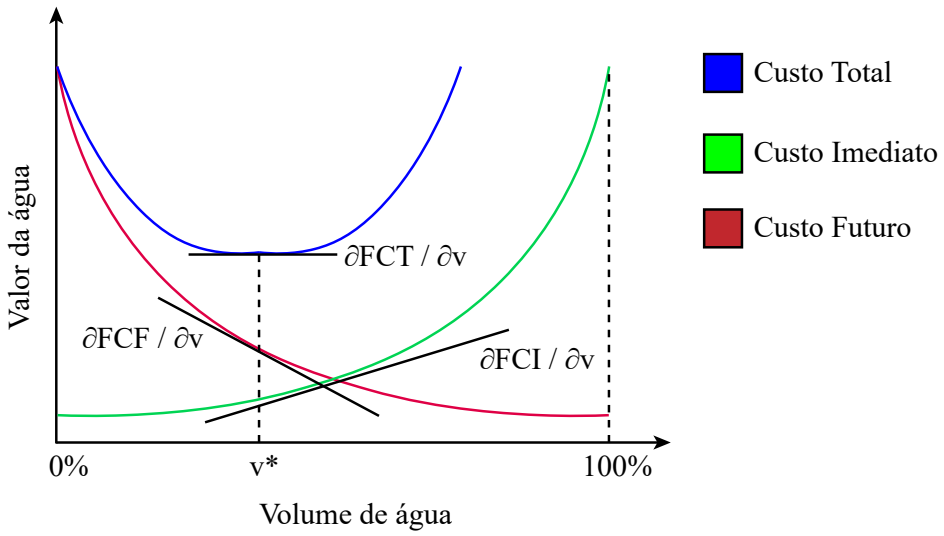


Fonte: Adaptado de Tolmasquim (2015) e Silva (2012).

a Função de Custo Futuro (FCF). A FCF está associada a diminuição do valor da água à medida que o volume armazenado aumenta, logicamente, devido a economia que leva a disponibilidade de água no futuro. A situação está representada na Figura 26.

Pode-se notar que a água armazenada possui um valor que pode ser medido tanto pela inclinação da curva FCI quanto pela curva FCF, tratando-se neste caso da derivada das funções em

Figura 26 – Minimização do custo de operação de um sistema hidrotérmico.



Fonte: Adaptado de Mayo (2012).

relação ao volume. Deste modo, o uso ótimo da água corresponde no ponto que minimiza a soma dos custos imediatos e futuros v^*. A curva de soma é chamada de Função de Custo Total (FCT) e conforme Equação 2.2, onde a derivada da mesma em relação ao volume possui valor nulo, encontra-se o ponto de otimalidade $FCI+FCF$.

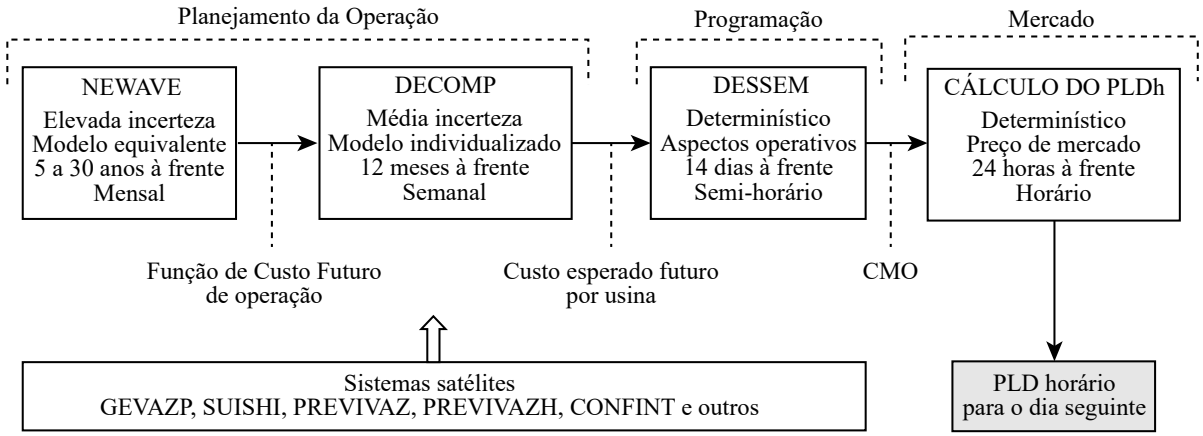
$$\frac{\partial(FCI+FCF)}{\partial v}=\frac{\partial FCI}{\partial v}+\frac{\partial FCF}{\partial v}=0\Rightarrow\frac{\partial FCI}{\partial v}=-\frac{\partial FCF}{\partial v} \tag{2.2}$$

Hoje, a definição do CMO de uma dada usina e, logo, a sua posição na ordem de despacho, acontece com base em modelos matemáticos que buscam equilibrar os benefícios presente e futuros no uso da água. O CMO é a base também para o cálculo do PLD. Neste sentido, a otimização dos recursos hidrotérmicos [32] tem influência direta na definição do preço da energia no mercado brasileiro. Os modelos estão implementados na forma de *softwares* acoplados, onde o problema de otimização do custo da água é subdividido com base em horizontes de planejamento que, na prática, dizem respeito a intervalos de tempo. A cadeia principal desses modelos (Ver Figura 27) é composta, na sequência, pelos programas NEWAVE (CEPEL, 2022d), DECOMP (CEPEL, 2022a) e DESSEM (CEPEL, 2022b; SANTOS *et al.*, 2020). Essa cadeia ainda é complementada por outros sistemas com foco mais específico como para geração eólica, solar, previsão de afluências, previsão de carga, planejamento da expansão e outros, chamados de sistemas satélites. Os mesmos são operados pela ONS e desenvolvidos pelo Centro de Pesquisas de Energia Elétrica (CEPEL). O CEPEL é um orgão vinculado a Eletrobrás que se dedica a atividades de pesquisa, desenvolvimento

[32] Um exemplo de formulação, modelagem matemática e resolução do problema de otimização hidrotérmico com foco nas características de formação de preços do Brasil pode ser visto nos trabalhos de Zucarato (2009) e Silva (2012).

tecnológico, inovação, certificação e treinamento para órgãos do setor energético. A página do CEPEL divulga um total de 49 produtos de *software* divididos em 17 categorias (CEPEL, 2022e).

Figura 27 – Cadeia de sistemas do ONS mantidos pela CEPEL.



Fonte: O Autor.

O NEWAVE (Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Longo e Médio Prazo), primeiro sistema da cadeia, determina o valor da água no tempo, em função do nível dos reservatórios, construindo-se assim a chamada política de operação. É utilizado para o planejamento e expansão do SEB com horizonte de médio (5 anos) a longo prazo (30 anos) em discretização mensal. Além de considerar toda a matriz hidrotérmica, incorpora também na sua versão mais recente as fontes renováveis intermitentes. Sua estratégia básica de operação trata do cálculo de todas as combinações possíveis de níveis de armazenamento e tendências hidrológicas. Dado o grande porte do SIN, a depender do horizonte de estudo, o problema de otimização que o mesmo resolve pode se tornar intratável. Dessa maneira, como estratégia de resolução, o NEWAVE agrega os reservatórios em unidades equivalentes e linhas de transmissão em conexões equivalentes. A técnica de otimização empregada é a Programação Dinâmica Dual Estocástica (PDDE) e foi proposta inicialmente para este problema no trabalho do Prof. Alberto Sergio Kligerman (KLIGERMAN, 1992). A modelagem foi atualizada durante um projeto de pesquisa desenvolvido entre 1993 e 1998 com resultados publicados em conferência em 2010 (KLIGERMAN; RIBEIRO, 2010). A PDDE no NEWAVE considera o histórico de afluências e possibilidades de cenários futuros gerados sinteticamente por meio de um modelo estatístico autorregressivo [33] com cálculo realizado pelo sistema satélite GEVAZP (CEPEL, 2022c). Acoplado ao sistema ainda estão os mecanismos de aversão ao risco: *Conditional Value at Risk* (CVaR), com valor condicionado aos piores cenários de afluências futuras e Superfície de Aversão ao Risco (SAR), que trata das restrições associadas ao armazenamento mínimo nos reservatórios equivalentes. O resultado da execução deste modelo é uma FCF que será utilizada, entre outras atividades, na elaboração

[33] O modelo estatístico autorregressivo é discutido em Maceira e Damázio (2006).

do PDE, na Programação Mensal da Operação (PMO) [34], no Planejamento Anual da Operação Energética (PEN) [35] e, como forma de entrada no próximo sistema da cadeia, o DECOMP.

O DECOMP (Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Curto Prazo) atua com horizonte máximo de um ano, discretização semanal no primeiro mês do horizonte e mensal nos demais. O principal objetivo do DECOMP é determinar metas de geração para cada uma das usinas hidroelétricas e térmicas, bem como os intercâmbios de energia entre os submercados do SIN. O mesmo considera uma árvore de cenários que representa as incertezas relacionadas as vazões das hidroelétricas e a geração de fontes intermitentes. Da mesma forma que o NEWAVE, o DECOMP é utilizado no PMO para minimizar o CMO de forma associada ao CVaR. O DECOMP determina o CMO levando-se em conta o custo de produção em relação aos possíveis acréscimos de carga no sistema, sendo isto por patamar [36] e submercado. A técnica aplicada no DECOMP é a programação linear, que modela características físicas das usinas individualmente, levando-se em conta restrições de despacho antecipado de até dois meses para a operação. As não linearidades são tratadas por meio de decomposição e programação dinâmica dual (MACEIRA *et al.*, 2002). São considerados no problema de otimização fatores como o tempo de viagem da água entre usinas em cascata, evaporação nos reservatórios, operação de vertedouros, deplecionamentos máximos e mínimos, limites de armazenamento, retorno e retirada de volumes relacionados a outros usos, altura de queda d'água, capacidade de turbinamento, cronogramas de operação e manutenção, inflexibilidade e restrições relacionadas aos limites de fluxo e controle de tensão (CEPEL, 2022a). Ao final do processo o DECOMP de forma similar ao NEWAVE mas no curto prazo, gera também uma FCF. A função referente a primeira semana do horizonte é inserida então no modelo DESSEM.

O DESSEM (Modelo de Despacho Hidrotérmico de Curtíssimo Prazo) trabalha de forma associada aos sistemas DECODESS e SIMHIDR. O DECODESS é responsável por converter os dados vindos do DECOMP para um formato inteligível no DESSEM. Já o SIMHIDR é um simulador da operação hidroelétrica que, com base em programação linear, otimiza o balanço hídrico das usinas levando em conta metas horárias de geração. O relacionamento entre estes sistemas pode ser visualizado na Figura 28. Operando em intervalos de 30 minutos com horizonte máximo de 14 dias, o DESSEM busca realizar uma representação da rede elétrica por meio de um modelo linearizado de fluxo CC que considera perdas. A técnica aplicada é a Programação Linear

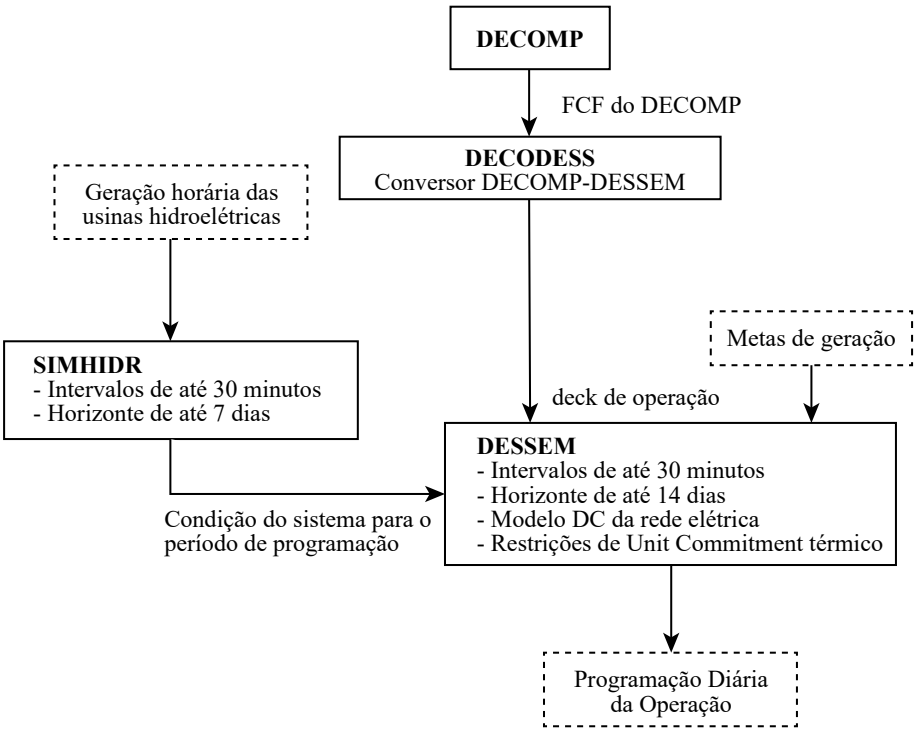
[34] Estudo elaborado mensalmente pelo ONS com participação dos agentes. Estabelece políticas de geração e intercâmbios, de modo a otimizar a operação no SIN segundo os critérios de rede estabelecidos pela ANEEL. É revisado semanalmente atualizando informações sobre o estado do sistema, condições de clima, previsões de carga e afluências (ONS, 2022b).

[35] Estudo realizado anualmente com revisões quadrimestrais. Agrega informações de entidades setoriais e dos agentes sobre oferta futura de energia e sobre previsões de consumo. Contém avaliações probabilísticas do atendimento ao mercado, em que se calculam, inclusive, riscos de ocorrência de déficits (ONS, 2022a).

[36] Os patamares de carga são determinados como pesado, na ponta de carga com duração entre 2 e 3 horas; leve, ocorrendo durante a madrugada; e médio, no restante do tempo. Os patamares são atualizados anualmente pela ANEEL e foram instituídos com base na resolução normativa nº 482 de 17 de Abril de 2012 (ANEEL, 2016).

Inteira-Mista (PLIM) com tratamento de não linearidades realizado por meio de decomposição em modelos lineares ou processos iterativos. Com o objetivo de minimizar os custos evitando-se a geração térmica, o DESSEM, como os outros sistemas da cadeia, incorpora critérios de aversão ao risco. As principais funcionalidades são, além da modelagem CC, a representação individualizada dos reservatórios (remanso e curvas de propagação da água entre usinas consecutivas), modelagem da função de produção das hidroelétricas (produtividade variável devido à altura de queda, capacidade de armazenamento, turbinamento e vertimentos), modelagem do *Unit Commitment* das termoelétricas (tempos mínimos de acionamento e desligamento, rampas de carga, custo de partida, custo não linear de geração e operação a ciclo combinado, entre outras), restrições de defluência e controle de cheias, representação de usinas elevatórias, PCHs, fontes intermitentes e possibilidade de estabelecimento de metas de geração térmica ou intercâmbio semanal entre submercados (CEPEL, 2022b). O DESSEM[37] gera como saída a chamada Programação Diária da Operação (PDO) que contém, para o dia seguinte, a ordem de mérito das usinas [38], os intercâmbios e o CMO em base semi-horária para cada um dos subsistemas.

Figura 28 – Operação integrada do modelo DESSEM.



Fonte: Adaptado de CEPEL (2022b).

O CMO é o resultado final de toda a cadeia dos modelos computacionais do ONS. Considerando que todo o planejamento da operação está condensado neste único índice, a partir dele toda a dinâmica de mercado acontece.

[37] A modelagem matemática do DESSEM pode ser consultada em Santos *et al.* (2020).

[38] Para maiores detalhes relacionados a elaboração da ordem de mérito consultar ONS (2020c).

2.3.2.2 Dinâmica de mercado

As regras de comercialização hoje praticadas no Brasil estão descritas em 26 cadernos de procedimentos (CCEE, 2022g). Os cadernos são rotulados como vermelhos e azuis. Os vermelhos trazem as regras básicas e os azuis, as suplementares. O limite de atuação do ONS vai até o momento em que os valores de CMO são publicados e, desse ponto em diante, entra em cena a CCEE que por meio dos procedimentos descritos nos cadernos realiza a conversão do CMO para os 24 valores de PLD que irão vigorar no dia seguinte no mercado brasileiro. A metodologia de cálculo do PLD horário está descrita no caderno vermelho nº 0 versão 2021.1.0 (CCEE, 2021). Diferentemente da operação, que leva em conta as restrições elétricas, no PLD é feita uma simplificação de modo a considerar a energia como disponível igualmente em todos os pontos de um dado submercado. As restrições internas nesse sentido são, portanto, ignoradas [39].

O PLD é formado com base no fluxograma da Figura 29 e esta formação está condicionada a três constantes. A primeira, trata do valor mínimo que o PLD pode assumir ($PLDmin$), já a segunda e a terceira, dizem respeito a dois valores máximos, sendo o PLD máximo horário ($PLDmaxh$) e o PLD máximo estrutural ($PLDmaxest$). Os três limites são definidos pela ANEEL e revisados anualmente com base na variação do Índice Geral de Preços de Disponibilidade Interna (IGP-DI) (FGV, 2022b) conforme resolução normativa nº 795/2017 (ANEEL, 2017).

As etapas do cálculo têm início com o recebimento na CCEE de 48 valores de CMO por submercado, correspondentes a cada meia hora do dia seguinte. Como o PLD é horário, então conforme Equação 2.3 [40] executa-se uma média simples, par a par, de forma progressiva na série até que a mesma passe a ter 24 valores. Após a transformação cada um dos valores é chamado no algoritmo de PLD intermediário ($PLDinter$). Na sequência, inicia-se um processo iterativo que, na primeira fase irá executar comparações entre o $PLDinter$ e os limites $[PLDmin, PLDmaxh]$. Caso algum limite seja extrapolado, então o valor do limite corrompido passará a ser considerado o valor do $PLDinter$. No entanto, se estiverem dentro do intervalo o processo continua. Na próxima etapa é realizada conforme Equação 2.4 [41] uma média dos 24 valores de $PLDinter$. Essa média chamada de PLD médio ($PLDmd$) é então comparada ao terceiro limite, o $PLDmaxest$. Se estiver abaixo do máximo permitido, então o PLD terá o valor até então calculado para $PLDinter$, do contrário é realizado o cálculo de um fator de ajuste ($FATOR$ [42]) conforme Equação 2.5 [43]. Este fator é utilizado para atualizar o valor do $PLDinter$ que na próxima iteração do algoritmo será novamente avaliado. O processo continua até que haja uma convergência no sentido de que todas as restrições impostas pelos limites sejam respeitadas.

[39] Esta é a regra geral, mas existem casos excepcionais em que o intercâmbio afeta o planejamento. Nestas situações o cálculo é modificado.

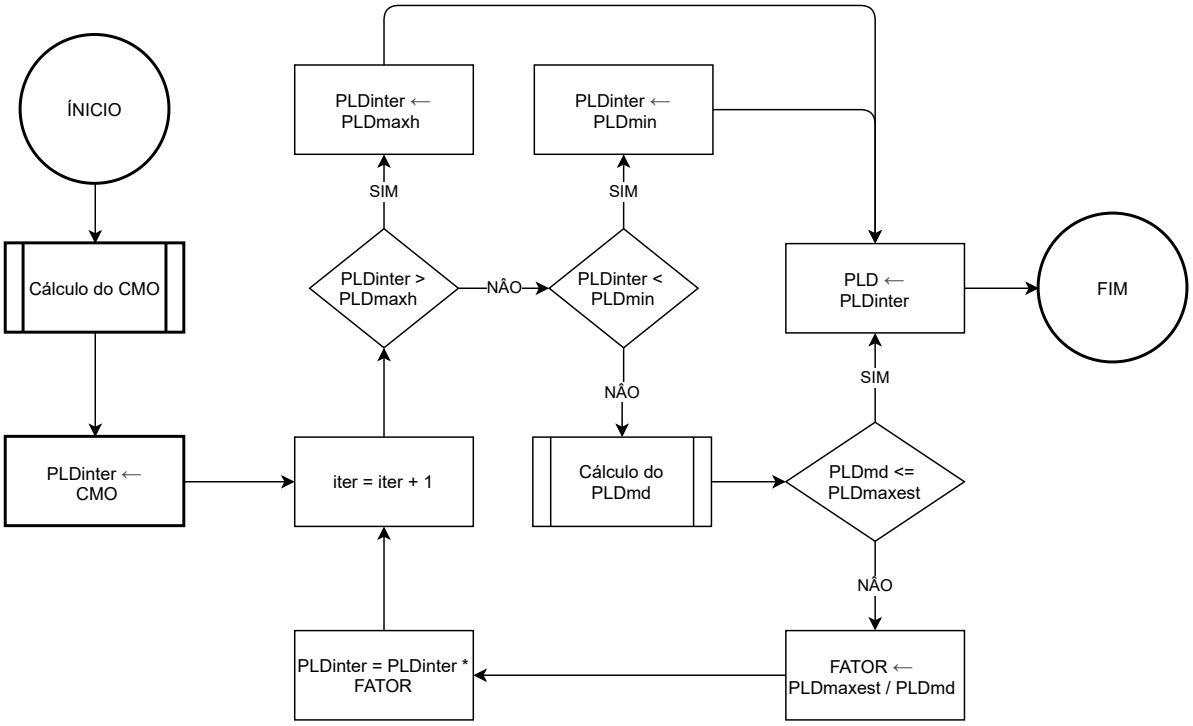
[40] Na Equação 2.3, s é o subsistema, h é o horário do dia e sh é cada semi-hora do dia.

[41] Na Equação 2.4, $iter$ é a iteração, s é o subsistema e h é a hora do dia.

[42] A título de curiosidade, o FATOR calculado pela CCEE é representado com 12 casas decimais. O valor do PLD é arredondado para cima para ter duas casas decimais ao final do processo (CCEE, 2021).

[43] Na Equação 2.5, $iter$ é a iteração, s é o subsistema, h é a hora do dia e a é o ano.

Figura 29 – Fluxograma de formação do PLD.



Fonte: Adaptado de CCEE (2021).

$$PLDinter_{s,h} = \frac{CMO_{s,sh} + CMO_{s,sh+1}}{2} \tag{2.3}$$

$$PLDmd_{s}^{iter} = \frac{\sum_{h=1}^{24} PLDinter_{s,h}^{iter}}{24} \tag{2.4}$$

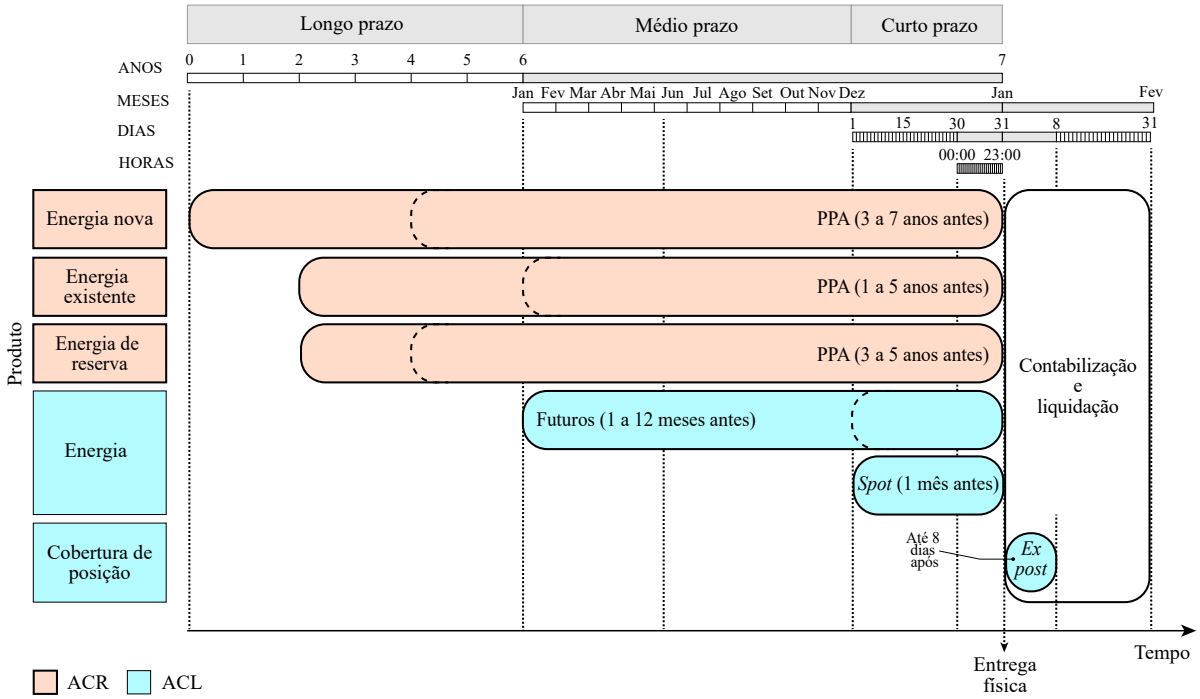
$$PLDinter_{s,h}^{iter+1} = PLDinter_{s,h}^{iter} * (\frac{PLDmaxest_{a}}{PLDmd_{s}}) \tag{2.5}$$

O uso de um preço horário no mercado brasileiro foi citado pela primeira vez na resolução nº 290/2000 da ANEEL (ANEEL, 2000). No entanto, até 1º de janeiro de 2021 o mesmo ainda não havia sido implantado. Desde a consolidação do Mercado Atacadista de Energia, o PLD sempre havia sido calculado com valor único para uma semana à frente. Em julho de 2017, por meio da consulta pública nº 33 do MME (MME, 2017) foi discutida uma nota técnica em que foi proposto que, para representar melhor a operação do sistema, os preços deveriam ser calculados com maior granularidade. Em 27 de julho de 2017, a CPAMP iniciou então um projeto piloto, chamado de operação sombra. O objetivo foi criar um ambiente de validação no qual pudessem ser verificados os impactos causados na migração do PLD semanal para o horário. Desde essa data, a CCEE disponibiliza publicamente em seu *website* (CCEE, 2022f) tanto os dados da operação sombra quanto os oficiais.

O Brasil não é único país do mundo a determinar o preço de mercado após o processamento de modelos matemáticos. Panamá, Austrália e Canadá (estado de Alberta), por exemplo, também o fazem. O Panamá executa a operação em 15 minutos e a contabilização em 30 minutos, a Austrália faz a operação em 5 minutos e contabilização em 15, já no Canadá são gerados preços por minuto com contabilização realizada a cada uma hora. Outros países que não necessariamente adotaram despacho *tight pool* mas que possuem preços horários são o Chile, Colômbia, Itália, México e os membros do NordPool (VIANA, 2018).

O PLD é o preço que serve de referência para negociação de todos os contratos no mercado brasileiro. Os contratos de energia podem ser subdivididos segundo o intervalo de tempo decorrido entre sua criação e a data da entrega física da energia. Conforme visto na Figura 30, a depender do produto negociado e do prazo, os contratos serão celebrados em ambientes e mercados diferentes.

Figura 30 – Organização de mercados de energia no Brasil.



Fonte: O Autor.

Energia nova é negociada em leilão no ACR e tem como finalidade atender aos possíveis aumentos futuros nas cargas das distribuidoras. É vendida e contratada neste mercado a energia de usinas que ainda não foram construídas. Os leilões estão divididos em duas categorias, sendo A-5 referente a usinas que entrarão em operação em até cinco anos e A-3 em três anos. Estes contratos tem duração que pode variar de 15 a 30 anos, a contar da data de início do suprimento. Com tratativas anteriores ao fechamento do acordo que podem durar até dois anos, normalmente são celebrados com antecedência variando entre três e sete anos (CCEE, 2022d).

Energia existente diz respeito a energia oriunda de usinas já construídas, que estejam em operação, com custo mais barato devido à amortização de investimentos. Possuem contratos realizados para períodos entre um e cinco anos anteriores ao suprimento, com duração mínima de um ano e máxima de 15, via leilões no ACR (CCEE, 2022d).

Também via leilões no ACR é adquirida a energia de reserva. Energia de reserva trata-se de um mecanismo no qual as usinas são contratadas com a finalidade de elevar a segurança do suprimento e só são acionadas quando necessário. A energia de reserva é contratada para todo o SIN e, devido a este fato, possui custos administrativos, financeiros, tributários e de operação, rateados igualmente entre os agentes que necessitem utilizá-la [44]. O fornecimento pode ser iniciado entre três e cinco anos anteriores a operação e os contratos possuem prazo entre 15 e 30 anos (CCEE, 2022d).

Contratos futuros são celebrados com antecedência variando entre 1 e 12 meses. Caso a energia seja negociada até um mês antes da entrega, então esta transação ocorrerá no mercado *spot*. Por fim, é permitido que ajustes para cobertura de posições sejam realizados até 8 dias após o encerramento do mês de referência. Estes são feitos no mercado *ex-ante*. Mercados futuros, *spot* e *ex-ante* são os mais influenciados pelo PLD, pois são acordados livremente entre os agentes. O comportamento integrado desses três define o chamado Mercado de Curto Prazo (MCP) e está relacionado a quatro conceitos fundamentais discutidos por Tolmasquim (2015) e Viana (2018), sendo:

a) Garantia Física (GF): Dado que o despacho é autorizado pela ONS, o gerador não possui controle próprio sobre a quantidade de energia que deverá injetar no sistema diariamente. Essa situação pode gerar descolamentos entre os valores de energia comercializada e a sua efetiva geração. Nesse sentido foi criada a GF, também chamada de energia assegurada, que trata do montante máximo qual uma dada usina possui autorização para comercializar. O cálculo da GF é realizado pela EPE e possui metodologia diferente para cada tipo de tecnologia de geração.

b) Lastro: O lastro é definido com a distribuição do montante de GFs ao longo do tempo por meio do estabelecimento dos contratos de compra. Essa distribuição deve ser realizada pelo agente anualmente e pode acontecer de duas formas, sendo a primeira denominada *flat* (onde o valor da GF é distribuído igualmente ao longo dos meses do ano) ou sazonalizada (onde a CCEE é informada pelo agente sobre a quantidade exata de cada parcela do montante será negociada para cada mês do ano).

c) Mecanismos de Realocação de Energia (MRE): O MRE é um mecanismo virtual que agrega todas as grandes hidroelétricas compulsoriamente no SEB e as PCHs de forma opcional. Seu objetivo é compartilhar os possíveis riscos hidrológicos do SIN. O MRE pode ser acionado em três situações. A primeira acontece quando uma parcela das usinas não produz toda a energia possível prevista em suas GFs, enquanto outras produzem além das suas GFs. Nestes casos, as usinas que estão com déficit compram das usinas superavitárias a um preço regulado pela ANEEL chamado de Tarifa de

[44] O encargo relacionado a este rateamento é denominado Encargo de Energia de Reserva (EER).

Energia de Otimização (TEO). A segunda acontece quando todas as usinas do sistema estão superavitárias em relação as GFs. Neste caso todas deverão comprar uma parcela do excedente na proporção de suas GFs ao valor do PLD. Por fim, a terceira situação trata de quando a GF total do sistema não é atingida, obrigando o operador do sistema a estabelecer novas GFs temporárias para as geradoras.

d) Encargo de Serviço do Sistema (ESS): Trata-se de uma cobrança referente aos serviços de manutenção, confiabilidade e estabilidade do sistema, contratados juntos aos mercados de serviços ancilares ou decorrentes de problemas com restrições de operação que levem a solicitação de desligamento ou acionamento de um gerador, fora da programação.

e) Tarifa de Uso do Sistema de Transmissão (TUST): Encargos, recalculados mensalmente, distribuídos igualmente a todos os agentes conectados às redes de alta tensão.

Levando-se em conta estes conceitos, na CCEE os agentes registram seus contratos de forma contínua. A comercialização, neste sentido, está relacionada com a venda da energia por um dado gerador a dado um comprador de interesse até o limite da sua GF, seja via leilão no ACR ou contrato bilateral no ACL. Toda a energia consumida é medida, contabilizada e liquidada com base no PLD. Sendo assim, vale destacar que devido a GF, um gerador que esteja com parte da sua produção comprometida deverá levar em conta os prazos de fornecimento dos seus contratos para não ofertar no mercado nada além do que lhe foi permitido, estando como discutido a seguir sujeito a penalidades [45]. O registro e gestão dos contratos é feito por meio de um sistema denominado clicCCEE. Concomitantemente é realizada diariamente a coleta de todos os dados de geração, consumo e perdas no SIN. Esta etapa é executada no Sistema de Coleta de Dados de Energia (SCDE).

No início de cada mês (com referência ao mês imediatamente anterior) é feita a contabilização. Primeiramente são apuradas as diferenças entre os valores consumidos a mais ou a menos e, também, as gerações realizadas acima ou abaixo do previsto nos contratos. As diferenças são valoradas ao PLD horário já incluindo as exposições referentes ao ESS e MRE, além da aplicação, caso exista, da TUST. Na sequência é feita a chamada rodada prévia. Nesta, relatórios contendo créditos e débitos devidos a cada agente são divulgados. Os agentes que possuem posições negativas devem então depositar os valores que devem em uma entidade bancária autorizada pelo Banco Central e indicada pela CCEE. Essa etapa é chamada de Aporte de Garantias Financeiras (AGFs). Após a contabilização é realizada a liquidação (CCEE, 2022e).

Na liquidação os agentes credores são pagos. No entanto, cabe ressaltar que um agente em posição credora não receberá seu devido valor de um devedor em específico, mas de todos os

[45] No Brasil, energia e lastro são vendidos em conjunto dado que todos os agentes vendedores devem apresentar lastro para a venda de energia para garantir 100% dos seus contratos; distribuidoras devem trabalhar 100% contratadas conforme a sua demanda; e consumidores participantes do ACL devem atender 100% de suas cargas por meio de geração própria ou com contratos registrados na CCEE. A negociação conjunta de energia e lastro é apontada como um problema atualmente no SEB e é discutida em Viana (2018).

agentes devedores do mercado. Quem realiza as transações é a mesma entidade bancária autorizada pela CCEE. Dessa forma, a CCEE não se torna contraparte na liquidação, sendo responsável apenas pela viabilização da compensação financeira que, por consequência, não gerará notas fiscais. A comprovação das operações deve ser feita utilizando-se relatórios de valores, divulgados mensalmente com os resultados das transações na CCEE. Agentes devedores que na etapa de AGFs não realizarem o aporte estão sujeitos a uma série de penalidades e medidas regulatórias como multas e desligamento da CCEE. A inadimplência é rateada entre todos os credores na proporção de seus créditos (CCEE, 2022e).

Por fim, terminada a liquidação é realizada a apuração das usinas de energia incentivada e lastro. Nesta última etapa é verificado se os agentes produtores de energia incentivada venderam a mesma com os descontos previstos nas regras de comercialização. Caso não o tenham feito, serão penalizados recebendo os valores referentes a sua produção com descontos. Nesta etapa é também aplicada a penalidade por insuficiência de lastro, que ocorre quando o total de energia consumida nos últimos 12 meses supera a quantidade prevista nos contratos do mesmo período. Esta e outras possíveis penalidades são apresentadas em CCEE (2013).

A dinâmica de mercado brasileira é uma consequência do planejamento de operação centralizado adotada desde 2004 no país. Ao longo dos anos sendo influenciado pela experiência dos mercados internacionais e devido também a sua natureza de alta disponibilidade hidrológica, o SEB vem evoluindo em direção a um mercado atacadista misto com varejista. A perspectiva de crescimento populacional, por si só, já seria neste ambiente um motivo para que esforços estratégicos na expansão do parque elétrico e na diversificação do portfólio energético sejam empreendidos. As recentes incorporações de novas regulações da ANEEL e MME vem de encontro a essa perspectiva, sinais disso são, entre outros, a implantação do PLD horário e a remoção de barreiras de entrada no ACL. Dado este momento, o mercado brasileiro mostra-se como um fértil campo de pesquisa. Neste âmbito, ganha importância as atividades de previsão, especialmente as relacionadas a séries de preços de eletricidade como o PLD.

No contexto de métodos modernos de previsão de séries temporais, como os que envolvem recentes técnicas em aprendizado profundo e estatístico, a área de previsão energética no próximo capítulo é explorada por meio de uma revisão sistemática de literatura.

3 REVISÃO DE LITERATURA

Este capítulo está dividido em três subseções. A primeira trata de uma revisão bibliográfica acerca da aplicação geral de *Deep Learning* [1]. Dos trabalhos seminais até as arquiteturas mais recentes, foi feita uma disposição cronológica da literatura a fim de sistematizar a evolução das contribuições. Além de situar o leitor quanto ao atual estágio da pesquisa, este início de capítulo também tem por objetivo introduzir conceitos importantes quanto ao entendimento das técnicas empregadas nos artigos discutidos, em especial, aqueles que fazem uso de mecanismos de atenção. Adicionalmente, são expostos os motivos pelos quais muitos pesquisadores ainda mantêm ceticismo quanto ao uso de Redes Neurais Artificiais. Este ceticismo teve início na década de 90. No entanto, vem sendo sistematicamente rompido desde 2010 com o surgimento de inovações e trabalhos de alto impacto. A segunda subseção apresenta o campo da previsão energética, área onde se situa o problema desta pesquisa. Inicialmente é feita uma breve análise bibliométrica da previsão de carga. A previsão de carga, por sua relevância e já centenária base literária, é investigada no sentido de identificar jargões e classificações comuns que vieram a ser adotados, de forma geral, em outros ramos da previsão em séries relacionadas a eletricidade. Na sequência, a previsão energética é explorada por meio de uma meta-análise de *surveys* publicados nos últimos 10 anos. Neste ponto é feita uma das contribuições desta pesquisa, onde buscou-se caracterizar melhor a previsão energética definindo de forma clara a sua característica interdisciplinar. Na terceira e última subseção, especificadamente a previsão de preços de eletricidade é abordada em detalhes. Com uso das principais bases de dados foram levantados artigos dos últimos 20 anos. Com foco especial em pesquisas que aplicaram modelos neurais profundos, são discutidos pontos fortes e fracos destes trabalhos. O estado da arte é analisado com uma discussão individualizada de estudos publicados entre 2016 e 2021. Neste ponto são especulados atributos que podem fazer avançar no futuro a área de pesquisa. O capítulo é finalizado com a discussão de artigos que fizeram uso de modelos neurais profundos com mecanismos de atenção para previsão de preços de eletricidade.

3.1 Aprendizado profundo

Desde a concepção dos primeiros computadores programáveis, uma das principais questões relacionadas a estes tem sido se, em algum momento, estas máquinas desenvolveriam a capacidade de demonstrar inteligência (RUSSELL; NORVIG, 2013). Ada Lovelace foi uma das primeiras pesquisadoras a tecer reflexões sobre o tema, mesmo que ainda em 1842 sequer o termo Inteligência Artificial (IA) existisse. Baseada no sucesso empreendido por Charles Babbage com a criação de uma versão rudimentar do que viria a ser o computador, Ada comentou cética que esse tipo de máquina nunca teria a pretensão de gerar nada original, se limitando apenas a fazer aquilo para o qual foi programada (FUEGI; FRANCIS, 2003; CHOLLET, 2018). Esta ideia, contudo, começou a ser reconsiderada a partir de 1943 quando McCulloch e Pitts criaram o primeiro modelo neural baseado

[1] O termo *Deep Learning* é em alguns pontos do texto referenciado em português como Aprendizado Profundo com mesmo sentido.

em *hardware* (MCCULLOCH; PITTS, 1943). Nos anos que se seguiram, durante a década de 50, a ideia de que computadores poderiam aprender foi de forma pioneira explorada por Alan Turing em seu trabalho chamado *Computing Machinery and Intelligence*. Artigo considerado um marco no campo da IA por propor o primeiro teste para avaliação de inteligência em máquinas: o Teste de Turing (TURING, 1950). Alan Turing, ainda neste trabalho contestou Ada Lovelace dizendo que computadores não só seriam capazes de aprender no futuro, como também seriam criativos.

O termo IA foi mencionado pela primeira vez pelos pesquisadores John McCarthy (Dartmouth College), Marvin Minsky (MIT), Nathaniel Rochester (IBM) e Claude Shannon (Bell Laboratories) em 1956 (DICK, 2019). Este campo de pesquisa trata do desenvolvimento de mecanismos computacionais capazes de automatizar tarefas perceptivas, cognitivas e manipulativas realizadas por organismos vivos, em especial, os seres humanos (RUSSELL; NORVIG, 2013). As primeiras aplicações de sucesso empregando IA foram aquelas nas quais especialistas determinam as regras para um dado problema elaborando modelos que os descrevem em pequena escala. Na sequência os modelos são alimentados com dados, para finalmente, apresentarem respostas para as questões quais foram programados. Essa abordagem, chamada na literatura de programação clássica ou IA simbólica é voltada para situações onde regras formais são bem estabelecidas. No entanto, para tarefas de complexa descrição, o uso de IA simbólica é inadequado e, em muitos contextos, até impossível de ser aplicado (GOODFELLOW; BENGIO; COURVILLE, 2016).

Tarefas de difícil descrição são chamadas de problemas perceptivos (CHOLLET, 2018). Os problemas perceptivos são aqueles que, em geral, são resolvidos intuitivamente por seres humanos, mas quando submetidos a computadores, exigem uma representação à nível de linguagem de máquina complexa e, por isso, são de difícil manipulação. Exemplos destes tipos de problemas são o reconhecimento de padrões em imagens, identificação de significado em expressões de fala, percepção motora, auto localização no espaço e no tempo, construções de relações semânticas entre ideias, inferência de situações que permitam planejar estratégias diante de desafios e, também, capacidade de demonstrar originalidade (LECUN; BENGIO; HINTON, 2015).

Problemas perceptivos, de forma geral, também podem não apresentar regras formais, sendo assim subjetivos. Em outros casos, as regras até podem existir, no entanto, podem fazer parte de um conjunto tão extenso de relações que o mapeamento para um modelo (matemático ou lógico) exigiria alto investimento (CHOLLET, 2018). A necessidade de modelagem de problemas perceptivos deu espaço a um ramo alternativo dentro de IA no qual, ao invés de serem modeladas as regras para a extração de respostas do modelo, o próprio modelo é aprendido por meio da apresentação de um conjunto prévio de perguntas e suas respectivas respostas. Esse ramo é o Aprendizado de Máquina, do inglês, *Machine Learning* (ML).

ML trata de um grupo de métodos que geram modelos computacionais de forma empírica a partir de dados (DOMINGOS, 2012). Segundo Mitchell (1997), "diz-se que um programa de computador aprende com a experiência E no que diz respeito a alguma classe de tarefas T e medida pelo desempenho P, se o seu desempenho nas tarefas T, conforme medido por P, melhora com a experiên-

cia E". Os esforços iniciais na construção de modelos de ML focavam em jogos de computador. Um destes primeiros exemplos foi o jogo de damas de Arthur L. Samuel (SAMUEL, 1959). Por meio de observação, este programador da IBM desenvolveu um método de análise probabilística que era capaz de ajustar as jogadas conforme mais e mais jogos eram registrados na memória do computador. A abordagem probabilística em IA é também uma área que começou a ser explorada no final da década de 50, no entanto, somente foi ganhar popularidade no início dos anos 90. O ML, apesar de intimamente relacionado com a estatística, hoje não é mais considerado uma subárea deste campo. Em ML são manipulados grandes conjuntos de dados complexos [2]. Para estes tipos de conjuntos, os métodos baseados em estatística clássica (como a bayesiana) exigiriam uma demanda de esforço computacional que em muitos casos pode ainda ser inviável tecnologicamente (CHOLLET, 2018).

Algoritmos baseados em ML são compostos de três elementos básicos: dados de entrada, dados de saída e uma métrica de avaliação da qualidade do modelo. Esses elementos se relacionam no contexto do chamado aprendizado supervisionado. Aprendizado supervisionado é a forma mais comum de ML (LECUN; BENGIO; HINTON, 2015). Essa abordagem consiste em gerar uma função h que tenha comportamento aproximado de uma função ótima f^* desconhecida, mas que se tem conhecimento de alguns pontos do domínio (dados de entrada) e a relação destes com o seu contradomínio (dados de saída). O espaço de busca que compreende todas as possíveis funções de mapeamento existentes é chamado de espaço de hipóteses. A seleção de uma função é, portanto, um processo de otimização no qual um algoritmo de ML buscará por um modelo ótimo dentro do espaço de hipóteses. Neste contexto, a determinação da qualidade do modelo estimado se dá por uma função objetivo (também chamada de função de perda) que irá avaliar a capacidade de inferência para novas amostras. Numa situação ideal, os algoritmos de aprendizado supervisionado conseguem estimar a melhor relação existente entre uma amostra desconhecida e os seus valores de contradomínio em f^* (RODRIGUES, 2018).

Algoritmos de ML superaram técnicas de IA simbólica e probabilística na resolução de problemas perceptivos na década de 90. Essas pesquisas deram evidências de que a codificação do conhecimento é uma tarefa complexa, mas que quando realizada de forma automática pode trazer avanços em diversas aplicações (GOODFELLOW; BENGIO; COURVILLE, 2016). No entanto, os algoritmos clássicos de ML dependem fortemente da representação dos dados. Nesse sentido, a escolha do conjunto correto de características que serão fornecidas como entradas a um algoritmo de ML pode determinar se o mesmo funcionará bem ou não (LECUN; BENGIO; HINTON, 2015). *Naïve Bayes* (CHEN *et al.*, 2020), Regressão Logística (ARCHER; LEMESHOW; HOSMER, 2007), *Suport Vector Machines* (SVMs) (CAMPBELL, 2002), Árvores de Decisão (KINGSFORD; SALZBERG, 2008) e até mesmo algoritmos mais recentes como Florestas Aleatórias e *Gradient Boosting Machines* (GBMs) (KONSTANTINOV; UTKIN, 2021) são técnicas de ML que dependem da escolha correta do conjunto de entradas. Levando-se em conta que para diversos problemas do mundo real pode ser difícil identificar quais características são relevantes em um dado problema

[2] No contexto de *Big Data*, campo no qual são estudadas técnicas projetadas especialmente para manipular grandes volumes de dados, alguns na ordem de *terabytes* ou maiores (KEZUNOVIC *et al.*, 2020).

e quais irão apenas introduzir ruído no modelo, a seleção destas torna-se um desafio (CHOLLET, 2018). Seleção de características, inclusive, vai além da escolha de variáveis. Essa tarefa pode envolver a inferência de informações a partir de dados não estruturados. Alguns exemplos:

a) Como identificar o que seria um pássaro em uma imagem, quando na prática uma imagem trata-se de uma matriz de *pixels*?

b) Como inferir a palavra "francês" na sequência da sentença "Fui a França e aprendi [palavra]", quando na prática a relação entre palavras é semântica?

c) Como prever uma janela futura de valores de preços de eletricidade, quando na prática o preço é uma variável que pode englobar fatores subjetivos como situação econômica e regulações?

Uma solução para estes problemas seria utilizar o ML não apenas para identificar o modelo ótimo no espaço de hipóteses, mas adicionalmente, identificar também a representação de dados ideal para uso no próprio modelo. Essa abordagem é conhecida como aprendizado de representação (LECUN; BENGIO; HINTON, 2015; GOODFELLOW; BENGIO; COURVILLE, 2016). A incorporação de aprendizado de representação a um algoritmo de ML apresenta algumas vantagens discutidas por Goodfellow, Bengio e Courville (2016). A primeira delas é relativa ao aumento de desempenho do modelo gerado, visto que esse irá se adaptar integralmente a representação inferida. Outra, é que a adaptação a novas tarefas também se torna mais rápida, dado que intervenções humanas na calibração do modelo são reduzidas. A propósito da questão da intervenção humana, ainda há que se destacar que a seleção de características feita por humanos é baseada na experiência pessoal e na forma como estes encaram o mundo. Muitas características relevantes podem se tratar de influências não observáveis no mundo físico e, nesse sentido, se feitas por humanos, seriam impossíveis de serem elaboradas (CHOLLET, 2018). Uma última vantagem é a escalabilidade para novos dados. Algoritmos que incluem aprendizado por representação podem ser alimentados com novas entradas de forma *on-line* [3] e, dado que estes podem transformar estas entradas em representações úteis, o modelo gerado passará automaticamente a separar as fontes de ruído das fontes que são relevantes para o mapeamento do domínio para o contradomínio. Algoritmos clássicos de ML não fazem esta separação sozinhos. Esse problema é chamado de Dilema da Seletividade-Invariância (LECUN; BENGIO; HINTON, 2015) e, para ser resolvido, diversos estudos (ZÖLLER; HUBER, 2021) propuseram o uso de algoritmos de ML associados a *pipelines* multiestágio onde nas primeiras etapas se inclui alguma técnica de aprendizado de representação.

Apesar de serem adequados a uma gama de problemas mais abrangente, *pipelines* multiestágio apresentam dois problemas fundamentais.

[3] Aprendizado *online* diz respeito a sistemas que mapeiam o comportamento dos dados a cada vez que uma nova observação é incluída na amostra de treino. Esta estratégia visa atualizar os modelos evitando que possíveis modificações de padrões ao longo do tempo afetem a qualidade da previsão (SILVA; SPATTI; FLAUZINO, 2016).

O primeiro ponto é que o algoritmo responsável pelo aprendizado de representação atua em separado do algoritmo de ML. Uma combinação comum é utilizar o algoritmo de ML precedido, por exemplo, de algum método de redução de dimensionalidade [4] (ESPADOTO *et al.*, 2021) ou de algum método de identificação de atributos [5] (SHEIKHPOUR *et al.*, 2017). No entanto, o problema quanto a alguns destes algoritmos de redução de dimensionalidade é que existem casos onde os mesmos podem ser insensíveis a não linearidades. Existem versões que executam dimensionamento não linear, no entanto, isto pode exigir o uso de uma função de *kernel* (ANOWAR; SADAOUI; SELIM, 2021; HOU; BEHDINAN, 2022). A função de *kernel*, por sua vez, também é elaborada a mão, logo, não será executada nenhuma busca otimizada pelo que seria a melhor solução no espaço de hipóteses, fato que pode levar a eliminação de boas características. Neste caso, o Dilema da Seletividade-Invariância seria ignorado. Já quanto aos métodos de identificação de atributos, a maior dificuldade é relativa ao esforço computacional. O treinamento de um algoritmo de ML, no geral, já tende a ser suficientemente custoso quando se utilizando grandes conjuntos de dados. Levando-se em conta que métodos de seleção de atributos trabalham com a avaliação da contribuição a cada inserção ou remoção de um novo atributo, o uso dos mesmos exige que o algoritmo de ML seja retreinado a cada modificação no conjunto de entradas. Nesse sentido, não é apenas um único algoritmo a ser parametrizado e avaliado, são pelo menos dois. E estes, inclusive, podem utilizar métricas de avaliação diferentes (HINTON, 2006; LECUN; BENGIO; HINTON, 2015).

O segundo ponto é que, em muitas das vezes, algoritmos que fazem aprendizado de representação e de ML são implementados para atuar de forma separada, não estabelecendo *feedback* entre si durante seus treinamentos individuais. Ou seja, não existe dinâmica entre a operação, por exemplo, de um PCA com SVM mesmo que estes dois estejam atuando juntos. Essa situação fica ainda mais pronunciada quando o problema exige que ao *pipeline* seja adicionada alguma etapa de transformação de dados. Casos que envolvem o uso de transformadas, filtros ou até mesmo operações de conversão logarítmica ou *z-score* são alguns exemplos (LECUN; BENGIO; HINTON, 2015; CHOLLET, 2018).

Nesse contexto, problemas com poucos dados e menores dimensões, tendem a ser bem resolvidos por *pipelines* multiestágio. No entanto, fora desse domínio e sumarizando os problemas até aqui relatados, existem situações que podem incluir:

a) Impossibilidade de formalização de regras em um modelo matemático ou lógico;

b) Grandes quantidades de dados e/ou dados não estruturados;

c) Dificuldade de identificação de características relevantes ao problema; e

d) Necessidade de uso de complexos *pipelines* difíceis de se gerenciar.

[4] *Principal Component Analisys* (PCA), *Independent Component Analysis* (ICA) ou *Generalized Discriminant Analisys* (GDA) são alguns exemplos desta classe de técnicas.

[5] *Backward Feature Elimination* (BFE) e *Forward Feature Elimination* (FFE) são exemplos desta classe de técnicas.

Nestas situações, até então, apenas uma classe de técnicas em IA tem se mostrado adequada. Essa classe é chamada de Aprendizado Profundo, do inglês, *Deep Learning* (DL).

Segundo LeCun, Bengio e Hinton (2015),

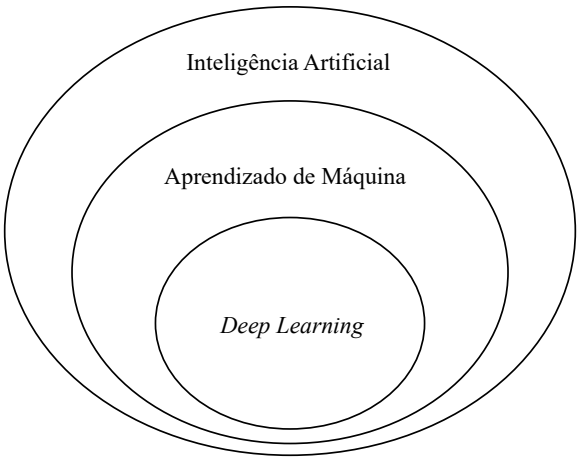
> *Deep Learning* permite que modelos computacionais compostos por camadas de processamento aprendam representações de dados com múltiplos níveis de abstração. Esses métodos melhoraram drasticamente o estado da arte (...) em diversos domínios sendo capazes de inferir uma estrutura complexa a partir de grandes conjuntos de dados simples.

Em outra definição, Goodfellow, Bengio e Courville (2016) comenta que,

> *Deep Learning* resolve o problema central do aprendizado de representação (...) permitindo que o computador construa conceitos complexos a partir de conceitos mais simples.

DL é uma subárea do ML que, por sua vez, é uma subárea de IA. Na Figura 31 pode ser observado como estes tópicos estão dispostos. DL é composto de diversas técnicas que, entre si, apresentam funcionamentos e aplicações diferentes. No entanto, todas compartilham dos mesmos elementos teóricos que vem sendo aprimorados desde a década de 40. Conforme visto na Figura 32, a atual área de DL passou por momentos de ascensão e declínio, sendo rebatizada com diferentes denominações em pelo menos três ocasiões ao longo da história. Esse movimento em ondas dá nome a três fases distintas sendo estas a Cibernética, as RNAs e, desde 2006, o DL.

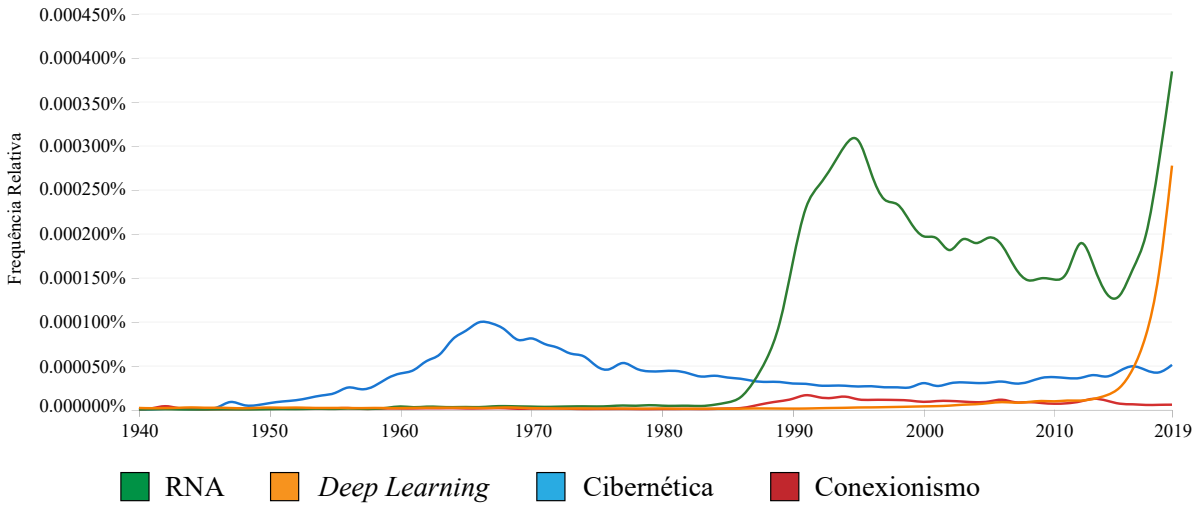
Figura 31 – Posicionamento de DL no campo de pesquisa de IA.



Fonte: Goodfellow, Bengio e Courville (2016).

A Cibernética durou de 1940 a 1960. Fez parte desta onda os primeiros antecessores de DL que nada mais eram do que modelos lineares criados a partir de uma perspectiva neurocientífica. O neurônio de McCulloch e Pitts (1943) foi a primeira arquitetura publicada e consistia de apenas uma multiplicação de um conjunto de valores $x = \{x_1, \ldots, x_n\}$ por pesos $w = \{w_1, \ldots, w_n\}$ associados

Figura 32 – Frequência relativa de citações aos vários nomes da área de DL ao longo da história.



Fonte: Adaptado de Google Books (2021).

a uma saída y e mapeados por uma função $f(x,w)=y$. É um modelo capaz de reconhecer até duas categorias de padrões mas que, no entanto, exige um ajuste manual dos pesos w. Métodos de auto-ajuste só viriam a ser propostos mais tarde por meio das arquiteturas Perceptron (ROSENBLATT, 1958) e Adaline (WIDROW; HOFF, 1960). No entanto, estas primeiras arquiteturas apresentavam ainda muitas limitações. A mais comumente discutida é a função *XOR* onde $f([0,1],w)=1$ e $f([1,0],w)=1$ mas $f([1,1],w)=0$ e $f([0,0],w)=0$. O trabalho de Minsky e Papert (1969) marcou o primeiro descenso em popularidade dos modelos neurais quando provou a incapacidade destas técnicas de atuarem em espaços não linearmente separáveis. Embora seja um período de arquiteturas rudimentares, legados importantes desta fase ainda permanecem. Um exemplo é o algoritmo *Stochastic Gradient Descent* (SGD) que hoje (em versões modificadas) é um dos algoritmos dominantes na etapa de treinamento de modelos (GOODFELLOW; BENGIO; COURVILLE, 2016).

Desde 1960, os pesquisadores ainda não haviam compreendido como as arquiteturas multicamadas poderiam ser treinadas utilizando a proposta inicial de modelo neural que combinava funções suaves, ajuste de pesos e cálculos de gradientes. A ideia de que isso poderia ser realizado foi descoberta de forma independente por vários grupos de pesquisa nas décadas de 70 e 80 (WERBOS, 1974; LECUN, 1985; RUMELHART; HINTON; WILLIAMS, 1986). Em 1986, Hinton, McClelland e Rumelhart (1986) publicaram os primeiros conceitos relacionados a representação distribuída do conhecimento. A ideia central destes autores foi demonstrar que os conjuntos que servem de entrada a um modelo neural podem ser representados por características simples do problema que, ao longo do treinamento, serão combinadas para gerar representações mais ricas. Essas representações são então utilizadas como recursos e cada camada aprende um recurso a partir das muitas entradas possíveis. Este ponto de vista foi reforçado pelos trabalhos de Rumelhart, Hinton e Williams (1986) e LeCun (1987) que publicaram a ideia do processamento distribuído e do algo-

ritmo *Backpropagation*. Baseado no conceito chave de que unidades computacionais simples podem produzir saídas de forma inteligente quando em rede, uma nova onda de pesquisas ressuscitou o campo da cibernética. Contudo agora, rebatizado de Conexionismo ou Redes Neurais Artificiais.

RNAs foram responsáveis por promover durante a década de 90 importantes avanços em muitas áreas da ciência. Entre 1980 e 2000, em uma pesquisa para o termo "*Artificial Neural Networks*"na base Scopus (SCOPUS, 2022) são retornados 15453 trabalhos em dez diferentes áreas de pesquisa, uma média de 2,11 trabalhos por dia como pode ser observado Tabela 5. A mesma pesquisa na base *Web of Science* (WoS) (Web of Science, 2022) retorna 5570 trabalhos e no Google Scholar (que incorpora livros, trabalhos acadêmicos e outras fontes) (Google Scholar, 2021) a quantidade é de 51100. Alguns exemplos de avanços deste período são a modelagem matemática de sequências com redes neurais (HOCHREITER, 1991; BENGIO; SIMARD; FRASCONI, 1994), as arquiteturas de redes recorrentes (ELMAN, 1993) e as células de memória (HOCHREITER; SCHMIDHUBER, 1997).

Tabela 5 – Produção científica entre 1980 e 2000 citando o termo "*Artificial Neural Networks*".

Área	**Quantidade**	**Produção anual**	**Produção mensal**
Engenharia	7034	351,7	29,31
Ciência da Computação	6517	325,85	27,15
Matemática	2019	100,95	8,41
Física e Astronomia	1614	80,7	6,73
Medicina	1601	80,05	6,67
Ciência dos Materiais	1506	75,3	6,28
Neurociência	1502	75,1	6,26
Bioquímica, genética e biologia molecular	1115	55,75	4,65
Química	857	42,85	3,57
Energia	763	38,15	3,18

Fonte: Scopus (2022).

Nesta fase, corporações em todo o mundo começaram seus próprios departamentos internos de IA para desenvolver o que foi convencionado chamar de sistemas especialistas (LUCAS; GAAG, 1991). Chollet (2018) comenta que “por volta de 1985 empresas americanas gastavam mais de US$ 1 bilhão por ano” com o desenvolvimento de sistemas especialistas. Em relação ao final da década de 90, Goodfellow, Bengio e Courville (2016) explica que

> (...) responsáveis por empreendimentos baseados em redes neurais e outras tecnologias de IA começaram a fazer afirmações irrealisticamente ambiciosas na busca por investimentos, no entanto, quando a pesquisa em IA não cumpriu com as expectativas irracionais, os investidores ficaram decepcionados.

Segundo Kelleher (2019), associada a questão do descredito empresarial, o consenso da comunidade científica na primeira metade dos anos 2000 sobre RNAs era de que:

a) O aprendizado de representação não era uma tarefa capaz de ser realizada por qualquer arquitetura de RNA até então existente de forma automática.

b) RNAs quando colocadas em um *pipeline* aumentavam consideravelmente a necessidade de esforço computacional. Isso numa época onde o processamento paralelo ainda era pouco acessível e a capacidade de *hardware* limitada.

c) RNAs com mais que três camadas não eram passíveis de treino. O treinamento com algoritmos baseados em gradiente descendente disponíveis era ineficaz, visto que ao retropropagar o erro através de muitas camadas, o acúmulo das multiplicações fazia com que os gradientes desaparecessem (tendendo para zero) ou explodissem (tendendo para o infinito). Esse problema foi chamado de *vanish gradients* (PASCANU; MIKOLOV; BENGIO, 2013) e é discutido em detalhes na subseção 4.2.4.

d) Algoritmos de treinamento baseados em *Backpropagation* não seriam capazes de sair de mínimos locais.

O trabalho de Hippert, Pedreira e Souza (2001) analisou 40 artigos publicados entre 1991 e 2000 que fizeram o uso de RNAs. Os autores comentam que com a grande popularidade dos métodos neurais, diversos congressos e periódicos publicaram estudos com metodologias duvidosas e pouco testadas. Os principais problemas identificados foram a utilização de conjuntos pequenos, mal uso de conceitos fundamentais relacionados a amostras de treino, validação e teste, hiperparametrização dos modelos (*overfitting* e *underfitting*) e resultados mal descritos. Ainda neste mesmo período, muitos modelos interpretáveis e concorrentes começaram a surgir. Estes se destacaram exatamente na resolução das mesmas tarefas que as RNAs vinham sendo empregadas. Máquinas de *kernel* (CORTES; VAPNIK, 1995), modelos gráficos (JORDAN, 1998) e árvores de decisão tomaram frente nas pesquisas e campeonatos de ML (CHOLLET, 2018).

Todo este contexto fez com que a área de RNAs pela segunda vez perdesse popularidade. E neste segundo descenso, um outro problema foi criado: RNAs passaram a ser tratadas com ceticismo, mesmo que durante os anos 2000 muitos avanços ainda continuassem a acontecer como as primeiras arquiteturas convolucionais (*Convolutional Neural Networks* (CNNs)) (FUKUSHIMA; MIYAKE, 1982) e redes de atraso (*Delay Networks*) (WAIBEL *et al.*, 1989). Sobre esta questão, em entrevista em maio de 2019 a *Communications of the ACM* (HOFFMANN, 2019) após ganharem o *Turing Awards* [6], os considerados "pais do *Deep Learning*" (TAPPERT, 2019) Yoshua Bengio, Geoffrey Hinton e Yann LeCun foram perguntados sobre o ceticismo com a área.

[6] Desde 1966, o *Turing Awards* é concedido anualmente pela *Association for Computing Machinery* (ACM) para pessoas que apresentaram grandes contribuições a ciência da computação. O prêmio é considerado o mais importante da área e é conhecido como "O Nobel da Computação" (ACM, 2022).

Hinton respondeu que

> Antigamente os pesquisadores envolvidos com IA fizeram grandes afirmações transformando a área em uma bolha. No entanto, Redes Neurais Artificiais vão muito além das promessas. É uma tecnologia que realmente funciona. Além disso, tem o poder de ficar automaticamente melhor conforme mais dados e mais processamento computacional são fornecidos, sem que ninguém precise escrever mais linhas de código.

Já Yann LeCun comentou que

> A ideia básica do *Deep Learning* não está desaparecendo, mas ainda é frustrante quando as pessoas perguntam.

Por fim, Bengio finaliza dizendo que

> As técnicas atuais tem muitos anos de aplicação industrial e científica pela frente (...) estamos sempre impacientes por mais, pois estamos ainda longe de uma IA de nível humano.

Após 2012, no entanto, o interesse em RNAs retornou. Este retorno foi motivado por avanços apresentados principalmente em campeonatos de ML organizados para que entusiastas, pesquisadores e estudantes resolvessem problemas específicos. Os campeonatos popularizaram técnicas de aprendizado de representação não supervisionado, vinculadas as arquiteturas com mais de dez camadas (RAINA; MADHAVAN; NG, 2009; KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Estes modelos, que vinham sendo publicados desde 2000, também resolviam o problema de *vanish gradient* por meio do treinamento separado das camadas. A técnica ficou conhecida como *greedy layer-wise pretrainning* e foi utilizada no treinamento da *Deep Bilief Network* (DBN), considerada uma das primeiras arquiteturas de DL (HINTON; OSINDERO; TEH, 2006). A estratégia foi utilizada com sucesso também em outras arquiteturas (RANZATO *et al.*, 2007). Devido a avanços dos algoritmos de otimização e do uso extensivo de GPU [7] esta técnica não foi mais necessária sendo que, novos algoritmos como Adam e RMSProp (discutidos na subseção 4.2.6.2) passaram a ser capazes de ajustar os pesos das camadas de ponta a ponta (RAINA; MADHAVAN; NG, 2009; DOGO *et al.*, 2018). Outro caso de destaque foi a adoção pelo Google nos telefones Android, em 2012, de um modelo baseado em células de memória (discutidas na subseção 4.2.4) que obteve sucesso na conversão de sinais de fala para texto (LECUN; BENGIO; HINTON, 2015).

O termo *Deep Learning* começou a ser adotado em 2014 (MONTÚFAR *et al.*, 2014; PASCANU *et al.*, 2014) com o objetivo de enfatizar que os pesquisadores agora eram capazes de

[7] *Graphics Processing Unit* (GPU) trata-se de um tipo de microprocessador especializado em processar gráficos. Sua eficiência está vinculada a uma estrutura focada em processamento paralelo. É incorporado a placas de vídeo integradas (ou não) as placas mãe de computadores (OWENS *et al.*, 2008).

treinar RNAs mais profundas do que até então havia sido possível (GOODFELLOW; BENGIO; COURVILLE, 2016). Desde então, a terceira onda desta área de IA vem sendo assim chamada.

Um ponto importante nesta nova onda é que diversas pesquisas vêm trabalhando na interpretabilidade dos modelos (ZHANG; ZHU, 2018). Além disso, os estudos de Dauphin *et al.* (2014) e Lebanon e Vishwanathan (2015) mostraram que a preocupação quanto aos mínimos locais, até então utilizada como argumento pelos céticos das RNAs, não procede. Estes autores demostraram que em arquiteturas com muitas camadas o *landscape* da função objetivo engloba um número combinatoriamente grande de pontos de sela onde o gradiente é zero e, próximo a estes pontos, a superfície da função na maioria das vezes se curva para baixo. Nos demais, os valores possuem muita proximidade com o ponto de ótimo global. Nesse sentido, mesmo que o algoritmo de otimização fique preso em um ponto de sela, este valor ainda estará contido em uma região vizinha ao ponto ótimo.

A partir de 2015, como pode ser visto na Figura 32, o uso de técnicas de DL na literatura apresentou crescimento exponencial. Além de ter também trazido em seu encalço novos trabalhos utilizando arquiteturas de RNAs clássicas. Hoje, técnicas de DL são consideradas o estado da arte para aplicação em problemas perceptivos, superando todas as outras técnicas de IA concorrentes, incluindo as clássicas de ML e modelos matemáticos ou lógicos projetados a mão (LECUN; BENGIO; HINTON, 2015; GOODFELLOW; BENGIO; COURVILLE, 2016; CHOLLET, 2018; KELLEHER, 2019).

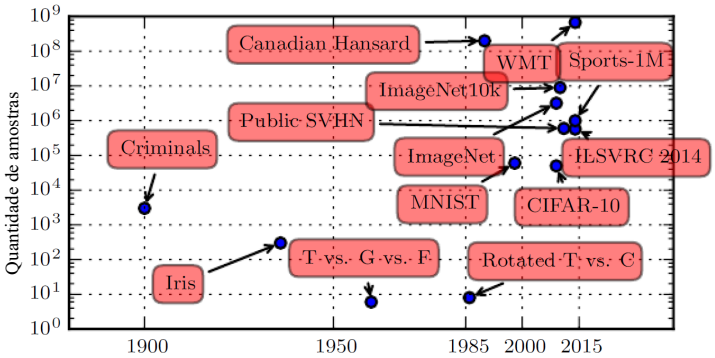
Embora as arquiteturas e os algoritmos de treinamento utilizados atualmente sejam semelhantes aos vistos na segunda onda, hoje estão disponíveis recursos que antes não existiam e que até então, não se sabia que eram fundamentais para o sucesso das RNAs. Dentre estes recursos, há que se dar destaque para a disponibilidade e quantidade de dados viabilizada pelo advento da internet. Dado que o ponto chave de um modelo de ML se baseia na estimativa estatística de uma saída, com poucos dados a tarefa de generalização se torna dificultosa ou até impossível. No entanto, com grandes quantidades de dados, essa carga é aliviada no processo de treinamento, visto que os modelos passam a estimar melhor a distribuição probabilística (GOODFELLOW; BENGIO; COURVILLE, 2016). Um exemplo neste ponto é o aumento do tamanho dos conjuntos utilizados em campeonatos de ML ao longo do tempo (ver Figura 33). Logicamente, para processar uma quantidade de dados tão grande foram necessários avanços na construção de arquiteturas progressivamente mais profundas. Algo que se tornou viável somente após 2008 quando a NVIDIA começou a comercializar GPUs (de uso geral) com preço acessível e, principalmente, lançou a plataforma CUDA [8] (CHOLLET, 2018). Em 2022 a placa de vídeo para computadores pessoais topo de linha da NVIDIA é a RTX 3090 (NVIDIA, 2022a)[9]. Esse equipamento possui 10496 núcleos operando a 1,7 GHz com núcleos projetados para operar matrizes de alta dimensionalidade (tensores) atingindo picos de até 36 *teraflops* por segundo. Este valor é 1800 vezes maior do que entrega um processador Intel Core

[8] CUDA é uma plataforma que se integra a outras linguagens de programação por meio de bibliotecas C e C++ facilitando implementações de processamento paralelo (NVIDIA, 2022b).

[9] Até o fechamento deste texto a NVIDIA havia anunciado que durante o ano de 2022 faria o lançamento de uma nova linha RTX. Dentre os produtos desta linha está inclusa a RTX 4090 (NVIDIA, 2022c).

i7 (INTEL, 2022) e 128 vezes maior do que o oferecido por uma das primeiras placas de vídeo da NVIDIA que forneceu suporte a CUDA, a GeForce 9600 GT (EVGA, 2008). O desempenho via GPU foi um dos grandes impulsos dados ao DL e substituiu o uso de *clusters* de *Central Processing Units* (CPUs) comuns em centros de pesquisa até 2010 (HUSSAIN *et al.*, 2013).

Figura 33 – Evolução do tamanho dos conjuntos de dados utilizados como *benchmarks* em campeonatos de ML.



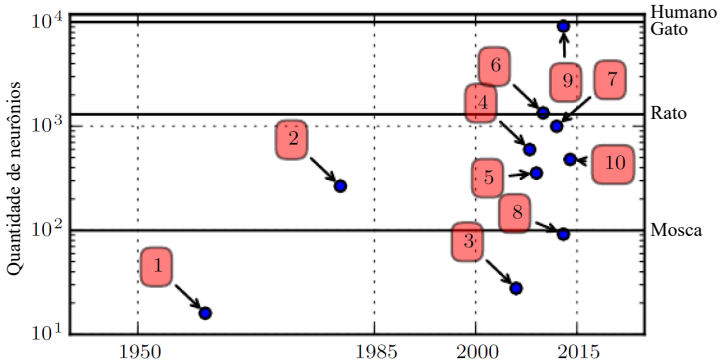
Fonte: Goodfellow, Bengio e Courville (2016).

Outra razão para o sucesso de DL é o aumento do tamanho dos modelos. Um dos legados teóricos do conexionismo é que animais se tornam inteligentes quando muitos de seus neurônios trabalham juntos. Nesse sentido, a título de curiosidade, as melhores RNAs por décadas tiveram menos conexões do que as existentes no cérebro de qualquer mamífero. RNAs tem dobrado de tamanho a cada 2,4 anos e, quanto mais aumentam, mais precisão apresentam. No entanto, a estimativa é de que um modelo de DL só atingirá a mesma quantidade de neurônios que um humano possui em 2056 (GOODFELLOW; BENGIO; COURVILLE, 2016). Portanto, não é surpreendente que RNAs que possuíam menos neurônios do que uma lombriga (ver Figura 34) ou menos conexões do que uma mosca (ver Figura 35) não fossem capazes de solucionar sofisticados problemas de IA propostos na década de 40 e 90. Apesar da comparação ser apenas ilustrativa, mesmo hoje, os maiores modelos de DL disponíveis como o Megatron da NVIDIA (8,3 bilhões de parâmetros) (SHOEYBI *et al.*, 2019), o GPT-3 da OpenAI (175 bilhões de parâmetros) (FLORIDI; CHIRIATTI, 2020) e o DeepSpeed da Microsoft (100 bilhões de parâmetros) (RASLEY *et al.*, 2020) são capazes de executar apenas tantas conexões quanto uma rã, estando ainda longe das 10^{15} conexões realizadas pelo o cérebro de um humano (DEWEERDT, 2019).

A revolução em IA ocorrida nesta terceira onda fez com que empresas ao redor do mundo investissem em novos projetos. Google, Tesla, Microsoft, Facebook, IBM, Apple, Adobe e Netflix são exemplos de empresas que estão atuando na área (CHOLLET, 2018). Muitos desses avanços vem democratizando o acesso ao DL por meio de bibliotecas *open source* como Theano, Torch, Tensorflow e Keras (PARVAT *et al.*, 2017). No âmbito da pesquisa científica, diversos avanços também vêm sendo reportados em diferentes campos de pesquisa. Desde o *Conseil Européen pour la Recherche Nucléaire* (CERN), que para identificação de partículas subatômicas no Grande Colisor

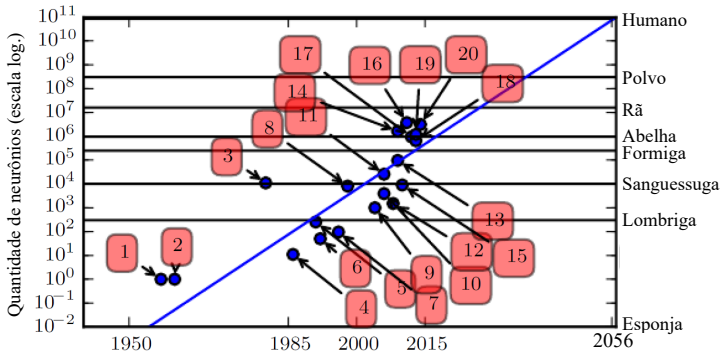
de Hádrons trocou o algoritmo do detector ATLAS, antes baseado em árvores de decisão, para uma arquitetura de DL baseada em Keras, Tensorflow e Python (ARPAIA *et al.*, 2021), até empresas farmacêuticas que vem utilizando estes modelos para projetar novas drogas (LAVECCHIA, 2019) ou centros médicos que estão usando DL para análises de imagens (PICCIALLI *et al.*, 2021).

Figura 34 – Evolução da quantidade de neurônios.



Fonte: Goodfellow, Bengio e Courville (2016).
Arquiteturas: (1) Adaptive linear element, (2) Neocognitron, (3) GPU-accelerated convolutional network, (4) Deep Boltzmann Machine, (5) Unsupervised Convolutional Network, (6) GPU-accelerated Multilayer Perceptron, (7) Distributed Autoencoder, (8) multi-GPU convolutional network, (9) COTS HPC unsupervised convolutional network e (10) GoogLeNet.

Figura 35 – Evolução do número de conexões entre neurônios.



Fonte: Goodfellow, Bengio e Courville (2016).
Arquiteturas: (1) Perceptron, (2) Adaptive linear element, (3) Neocognitron, (4) Early back-propagation network, (5) Recurrent neural network, (6) Multilayer Perceptron (7) Mean field sigmoid belief network, (8) LeNet-5, (9) Echo State network, (10) Deep belief network, (11) GPU-accelerated convolutional network, (12) Deep Boltzmann Machine, (13) GPU-accelerated deep belief network, (14) Unsupervised convolutional network, (15) GPU-accelerated multilayer perceptron, (16) OMP-1 network, (17) Distributed autoencoder, (18) Multi-GPU convolutional network, (19) COTS HPS unsupervised convolutional network e (20) GoogLeNet.

Os mais recentes avanços no campo de DL tratam da proposição de novas técnicas de treinamento mais sofisticadas, modelos maiores e, do ponto de vista da arquitetura, do uso dos chamados mecanismos de atenção. Nesse sentido, por sua importância central no contexto da metodologia adotada neste trabalho é discutida a seguir a evolução deste subcampo.

3.1.1 Redes neurais atencionais

Acompanhando a tendência de inspirações biológicas que vem motivando parte da pesquisa em IA desde a década de 40, surgiu em 2014 o conceito de atenção computacional aplicada a DL (BAHDANAU; CHO; BENGIO, 2015).

Atenção computacional é baseada no espectro do comportamento humano que trata do gerenciamento, de forma otimizada, dos recursos cognitivos para execução de tarefas que necessitam de foco seletivo em aspectos discretos da informação (GHAZANFAR; SCHROEDER, 2006; QUINN *et al.*, 2019). Dadas as limitações do cérebro, o ser humano não possui capacidade de processar de forma paralela toda a informação que o rodeia. Consequentemente, estes gargalos de processamento precisam ser tratados por meio da concentração (WANG; TAX, 2016). Exemplos deste tipo de situação podem englobar o sentido auditivo (como manter uma conversa com um ou mais interlocutores em ambientes ruidosos), visual (como observar traços sutis em expressões faciais para interpretar sentimentos e emoções), gustativo (como identificar sabores inesperados em alimentos de consumo cotidiano), olfativo (como notar perfumes específicos em ambientes com vários tipos de odores) ou tátil (como sentir marcas em linguagem braile) (CORREIA; COLOMBINI, 2022).

A atenção como comportamento humano vem sendo estudada desde o século XIX (COLOMBINI, 2014). Estas pesquisas mostraram que, na medida em que os mecanismos de atenção evoluíram para se tornarem mais especializados, a resposta aos estímulos perceptivos também se aprimorou. Fato que tornou a capacidade de ter atenção um fator crucial para sobrevivência (PARASURAMAN, 1998). Já os modelos computacionais que emulam o comportamento de atenção começaram a ser explorados somente no final da década de 80 (FRINTROP; ROME; CHRISTENSEN, 2010). Segundo Correia e Colombini (2022), trabalhos como teoria de integração de recursos de Treisman (TREISMAN; GELADE, 1980), busca guiada de Wolfe (WOLFE; CAVE; FRANZEL, 1989), arquitetura triádica (RENSINK, 2000) e modelos de atenção seletiva (VELDE; KAMPS; KLEIJ, 2004) são alguns dos primeiros trabalhos que introduziram a base teórica dos sistemas de atenção computacional. Estes sistemas, no entanto, ficaram relativamente limitados a aplicações relacionadas a imagens (SALAH; ALPAYDIN; AKARUN, 2002; WALTHER; EDGINGTON; KOCH, 2004) e robótica (SCHEIER; EGNER, 1997; FRINTROP; JENSFELT, 2008).

Redes Neurais Atencionais (do inglês, *Attentional Neural Networks* (ANNs)) foram a primeira classe de algoritmos de ML que fizeram uso dos mecanismos de atenção. *Neural Attention Networks* (NANs) surgiram no contexto de métodos para *Natural Language Processing* (NLP) onde promoveram significativos avanços. O uso de atenção nesta área colocou ANNs como o atual estado da arte em tarefas de compreensão de linguagem via máquina. Exemplos de problemas em que ANNs se destacam são análise de sentimentos, tradução de textos, sistemas de *Question and Answer* (QA) e transferência de aprendizado (GALASSI; LIPPI; TORRONI, 2020).

Uma das arquiteturas de DL desenvolvidas para tradução de texto e que foi utilizada como base na exploração das possibilidades providas pelos mecanismos de atenção é a Seq2Seq

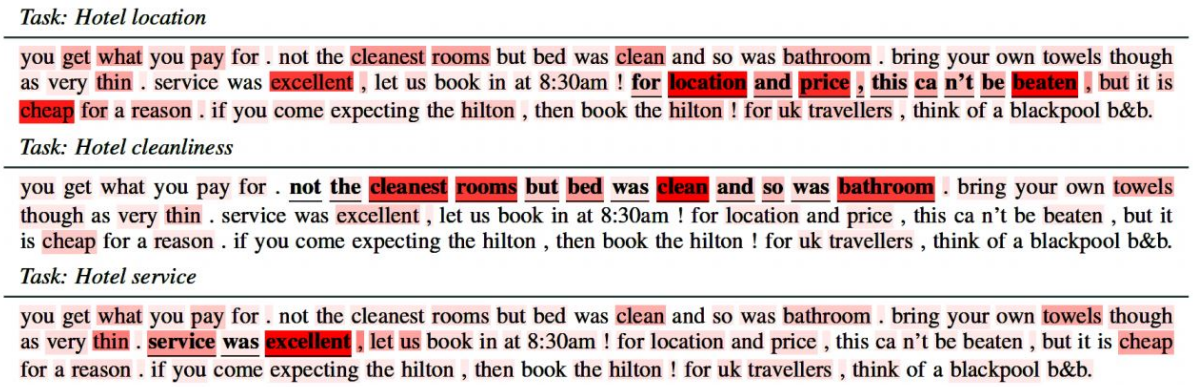
também chamada de *Encoder-Decoder* (SUTSKEVER; VINYALS; LE, 2014). Os autores desta arquitetura propuseram uma estrutura na qual uma camada neural codifica uma sentença de texto em um dado idioma para um vetor numérico de comprimento fixo, a partir do qual uma segunda camada decodifica o vetor para uma sentença traduzida em outro idioma. O vetor de comprimento fixo neste caso incorpora o que seria o significado latente da sentença que determina o seu contexto. Com base neste contexto, teoricamente a tradução para qualquer idioma seria possível, desde que o decodificador fosse treinado para isso. Na ocasião, o resultado reportado foi um recorde na qualidade de tradução do inglês para o francês (SUTSKEVER; VINYALS; LE, 2014).

A Seq2Seq isolada, no entanto, apresenta uma importante limitação. A tradução de texto não é um problema de classificação, mas sim de previsão. Nesse sentido, este tipo de arquitetura não funcionaria se armazenasse somente a relação direta entre uma sentença em um idioma A e sua correspondente em um idioma B. Para um melhor funcionamento é necessário que o histórico das sentenças seja armazenado na memória da RNA. Nesse sentido, o modelo deve ser treinado com um *corpus* de texto extenso, preferencialmente, oportunizando à rede aprender diversas nuances do idioma que, porventura, venham a codificar inclusive contextos semelhantes usando sentenças diferentes. Conjuntos de dados que viabilizam isso são livros completos, artigos, revistas, longos discursos e outros do tipo. O tratamento da memória na Seq2Seq foi explorado no trabalho de Sutskever, Vinyals e Le (2014). Para garantir que partes iniciais de um texto fossem "lembradas" pela rede durante a tradução, os autores utilizaram células de memória chamadas *Long Short-Term Memory* (LSTM). Contudo, no mesmo ano Cho *et al.* (2014) mostraram em seu trabalho que mesmo adotando células LSTM, a memória na Seq2Seq se deteriora à medida que o tamanho das frases aumenta. Utilizando este problema como motivação, Bahdanau, Cho e Bengio (2015) propuseram a incorporação de um mecanismo de atenção a LSTM Seq2Seq. Essa nova arquitetura foi batizada de RNNSearch [10]. Ao invés de codificar uma sentença completa para um vetor numérico fixo, RNNSearch codifica cada palavra em um vetor próprio que incorpora o significado da mesma no contexto das frases. O mecanismo de atenção atua identificando quando um vetor é ou não relevante. Nesse sentido, a rede torna-se capaz de escolher o subconjunto de vetores adequado para cada momento da decodificação de forma adaptativa, enquanto gera a tradução (ver Figura 36). Dado que nesta abordagem os mecanismos de atenção propagam os valores de importância de cada palavra, então as células de memória passam a ser menos exigidas. RNNSearch foi capaz de superar a LSTM Seq2Seq permitindo o uso de longas frases sem perda de memória. O sucesso da abordagem fez com que a Google e o Facebook a incorporassem em suas máquinas de tradução. Até 2017 este ainda era o modelo utilizado pelo Google Tradutor (WU *et al.*, 2016; KHAN; SARFARAZ, 2019).

Após a RNNSearch ter introduzido o primeiro uso de mecanismos de atenção em DL, diversos novos estudos buscaram se aproveitar das possibilidades desta nova estratégia. Uma sumarização desta evolução pode ser vista na Figura 37. Essa linha do tempo foi adaptada da mais recente e abrangente revisão bibliográfica sobre o tema encontrada até a finalização desta tese. A

[10] Em alguns trabalhos citada como mecanismo de atenção de Bahdanau (BAHDANAU *et al.*, 2016; BAL; SINHA, 2019; SEHOVAC; GROLINGER, 2020; YADAV; NAIK, 2021).

Figura 36 – Exemplo de visualização de atenção para uma tarefa de análise de sentimentos. As palavras são destacadas de acordo com a pontuação de atenção. As frases em negrito são as consideradas relevantes para a compreensão humana.



Fonte: Galassi, Lippi e Torroni (2020).

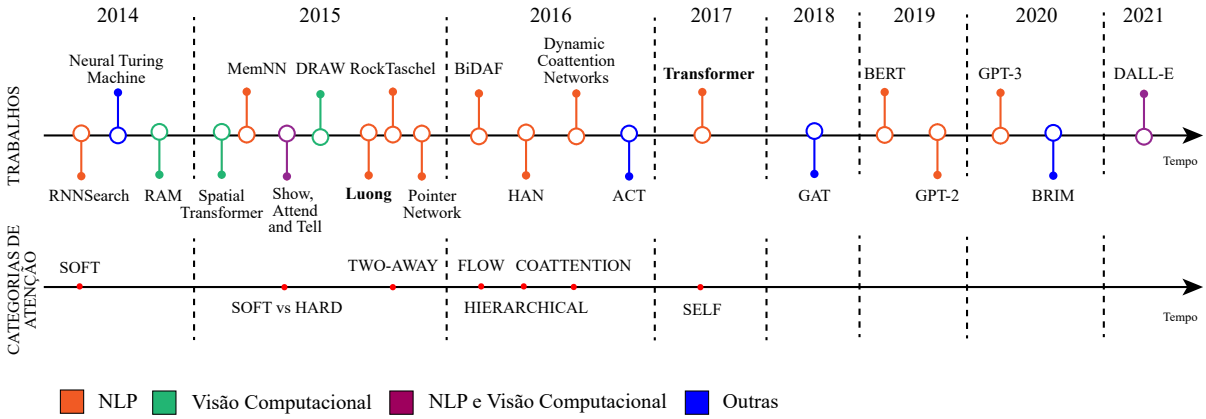
pesquisa de Correia e Colombini (2022) analisou 650 trabalhos publicados entre janeiro de 2014 e fevereiro de 2021. As autoras dão destaque para a rápida ascensão dos mecanismos de atenção citando o crescimento exponencial no número de pesquisas. Conforme visto na Figura 38, em 2015 foram publicados menos que 200 estudos, mas em 2020 essa quantidade já superava os 2500. Nos dois primeiros meses de 2021 foram 160 publicações, uma média de 2,6 artigos por dia.

Embora a comunidade científica tenha demonstrado interesse pelos mecanismos de atenção, uma situação fica nítida: Até então, grande parte do que tem se produzido possui foco apenas em NLP. Existe uma parcela menor focada em reconhecimento de imagens e, esporadicamente, alguns estudos exploraram sistemas de recomendação e teoria de grafos. De fato, noutra revisão publicada por Niu, Zhong e Yu (2021) não é relatado nenhum trabalho fora desses quatro temas. O mesmo acontece em Chaudhari *et al.* (2021). Além destas revisões, foram encontradas outras quatro, sendo que (i) a de Wang e Tax (2016) foca apenas em visão computacional, (ii) Hu (2019) e (iii) Galassi, Lippi e Torroni (2020) abordam somente NLP e (iv) Lee *et al.* (2019) explora grafos.

O destaque para o uso de DL com mecanismos de atenção em NLP foi devido a obtenção de resultados em acurácia expressivamente mais significativos do que até 2014 vinham sendo reportados com métodos tradicionais de tradução (GALASSI; LIPPI; TORRONI, 2020). Com RNNSearch o primeiro grande passo foi dado, no entanto, outras duas arquiteturas subsequentemente vieram a superar a RNNSearch. Estas foram os mecanismos de atenção de Luong (LUONG; PHAM; MANNING, 2015) e o Transformer (VASWANI *et al.*, 2017). Os autores das revisões bibliográficas citadas comentam que estes dois trabalhos foram os responsáveis por formalizar as três principais categorias de atenção em DL, sendo elas:

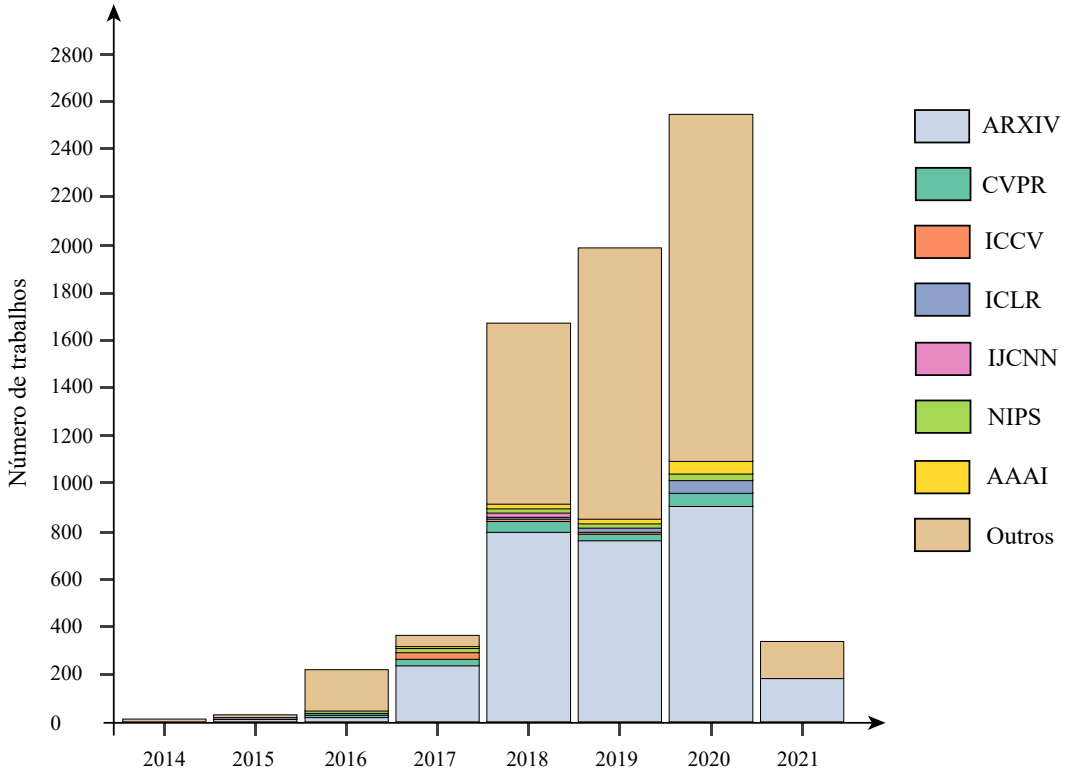
a) *Soft Attention*: Na atenção suave são atribuídos pesos de 0 a 1 para cada elemento. Dessa forma a RNA definirá quanta atenção deve ser focada em cada parte do conjunto de entrada. Esta atenção atua no contexto espacial e temporal. No espacial, as caracte-

Figura 37 – Linha do tempo das arquiteturas de DL baseadas em atenção. Em negrito, as duas arquiteturas propostas para utilização neste projeto de doutorado.



Fonte: Adaptado de Correia e Colombini (2022).

Figura 38 – Número de trabalhos publicados por ano (entre janeiro de 2014 e fevereiro de 2021) envolvendo mecanismos de atenção aplicados a DL. ARXIV é a plataforma de pré prints da Universidade de Cornell. CVPR e ICCV são congressos de visão computacional. ICLR, IJCNN, NIPS e AAAI são conferências em inteligência artificial e RNAs. “Outros” diz respeito a congressos como ICL, ACM, EMNLP, ICRA, ICPR, ACCV, CORR, ECCV, ICASSP e periódicos como IEEE Acess, Neurocomputing e demais com escopos similares.



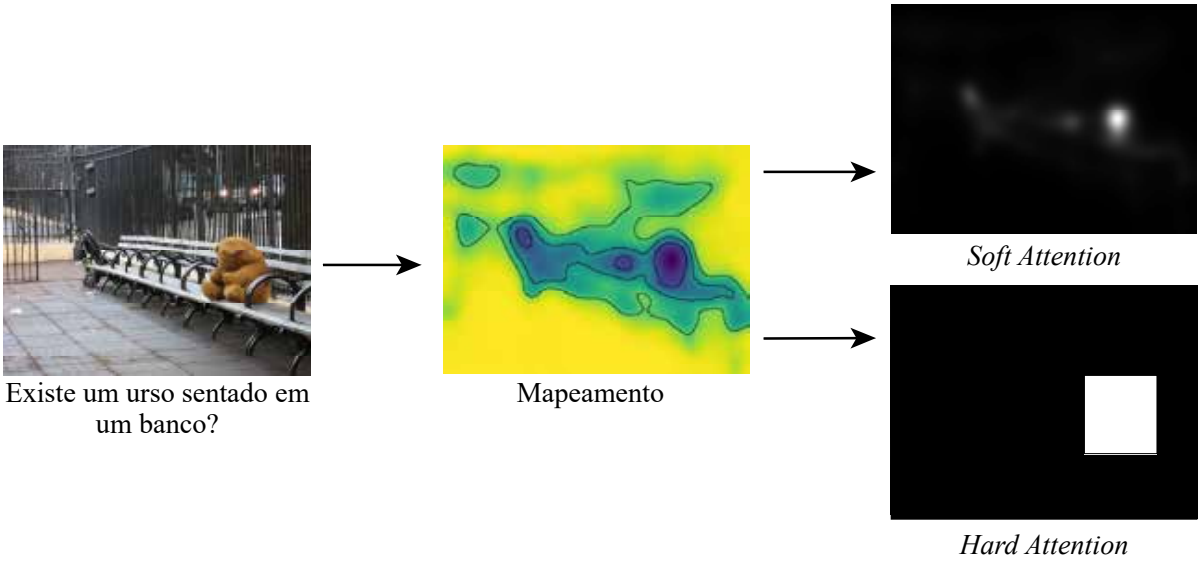
Fonte: Adaptado de Correia e Colombini (2022).

rísticas mais relevantes são separadas das menos relevantes. Já no temporal, existe um ajustamento dos pesos que leva em conta cada janela de tempo deslizante, conforme as amostras vão sendo apresentadas a rede. Isso proporciona uma avaliação diferente e adaptada para cada segmento temporal, visto que os elementos de entrada, mesmo que repetidos nas amostras, podem apresentar importâncias distintas em momentos diferentes do tempo. Dada a natureza contínua dos pesos de atenção, nesta categoria o espaço de busca é diferenciável e a resposta é determinística. Contudo, o esforço computacional também é mais alto. A Figura 39 mostra um exemplo de aplicação de um mecanismo de atenção suave em uma imagem.

b) *Hard Attention*: A atenção dura atua com o mesmo propósito da suave. A diferença é que ao invés de ponderar num intervalo contínuo, os valores são discretos sendo 0 ou 1. O esforço computacional é reduzido, no entanto (ver Figura 39) nem todas as nuances do conjunto de entrada (como o banco onde está o urso) podem ser capturadas pela atenção.

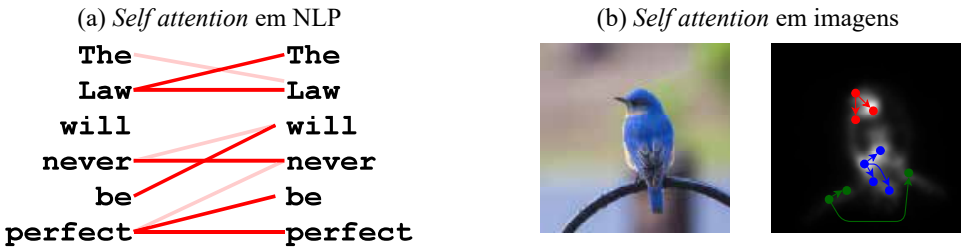
c) *Self-attention*: A auto atenção além de identificar os elementos de interesse na amostra e ponderá-los, adicionalmente investiga e pondera também a interdependência entre os elementos. Essa abordagem permite que as entradas interajam entre si. Esse cálculo é realizado por meio de matrizes paralelizáveis que não exigem qualquer tipo de recorrência. A Figura 40 mostra um exemplo de aplicação de auto atenção a uma frase e a uma imagem. Na frase são destacados em vermelho os relacionamentos entre as palavras mais importantes. Na imagem o mesmo acontece, mas relacionando pontos designados como relevantes pelo mecanismo de atenção.

Figura 39 – Exemplo de aplicação de *soft* e *hard attention* a uma imagem. A primeira categoria atuando num espaço contínuo e a segunda em um espaço discreto.



Fonte: Adaptado de Correia e Colombini (2022).

Figura 40 – Exemplo de aplicação de *self attention* em (a) uma frase e em (b) uma imagem.



Fonte: Adaptado de Correia e Colombini (2022).

Desde que o Transformer inaugurou a categoria de auto atenção em 2017, até a finalização deste texto nenhuma nova arquitetura havia sido publicada. Contudo, novos modelos sim. BERT (DEVLIN *et al.*, 2019), GPT-2 (RADFORD *et al.*, 2019), GPT-3 (FLORIDI; CHIRIATTI, 2020) e DALL-E (RAMESH *et al.*, 2021) são alguns exemplos. Estes, na prática, tratam-se das mesmas arquiteturas anteriores, no entanto, mostraram-se superiores (especialmente em reconhecimento de imagens e NLP), pois investiram massivamente no aumento do tamanho da rede. Para fins de comparação, BERT de 2018 trata-se de uma RNA com 110 milhões de parâmetros. Já DALL-E lançada em fevereiro de 2021 possui 12 bilhões.

Como visto, os mecanismos de atenção em DL possuem apenas oito anos de existência. Dada esta situação é compreensível que diversos campos de aplicação ainda não tenham feito uso desses modelos. Como se viu, apenas quatro áreas vêm sendo relatadas nas mais recentes revisões bibliográficas. Dentre os campos de pesquisa que ainda não ganharam destaque no âmbito de DL com mecanismos de atenção está a previsão em séries temporais. A revisão de Correia e Colombini (2022) inclusive aponta que esta é uma tendência futura ao comentar que

> Prever séries temporais torna-se desafiador à medida que o tamanho da série aumenta. Redes Neurais Atencionais podem contribuir significativamente para melhorar os resultados. Especificadamente, acreditamos que (...) olhar para diferentes partes da série ou diferentes faixas de frequência, pode introduzir uma nova maneira de calcular séries temporais.

Em problemas de NLP cada palavra é *tokenizada* [11] antes de ser submetida ao modelo, ou seja, todo texto e seus elementos são transformadas em vetores numéricos que representam seus respectivos significados. Desprezando-se o conceito semântico, o que sobra do ponto de vista do uso de qualquer modelo neural, por consequência, é nada mais do que uma manipulação algébrica destes vetores. Essa manipulação, passa pela otimização, que busca num espaço de hipóteses a parametrização correta que irá relacionar uma sequência de palavras na entrada com outra na saída. Prever uma frase ou palavra, portanto, é equivalente a prever um vetor numérico qualquer. Nesse sentido, qualquer série temporal pode ser modelada via arquitetura de DL que faz uso de mecanismos de atenção. Exatamente como ocorre em NLP.

[11] Técnica também conhecida como *Word2Vec* ou *Embeddings* (CHURCH, 2017).

Levando-se em consideração o período que vai de janeiro de 2014 até junho de 2021, numa pesquisa realizada na base Scopus para o conjunto de palavras-chave (*"attention" AND "deep learning"*) são retornados 10631 trabalhos. No entanto, ao especificar o campo de previsão em séries temporais com (*"attention"AND "deep learning"AND "time series" AND "forecasting"*) esse número é reduzido para apenas 220. Os gráficos da Figura 41 mostram uma análise quantitativa destes números. Como se pode ver na subfigura (a) o primeiro trabalho publicado data de 2015 e de lá pra cá o número vem mais que dobrando ano a ano.

Conforme subfigura (b) a grande maioria das publicações acontece na forma de artigos em periódicos. No gráfico da subfigura (c) (que mostra os 10 países com maior número de publicações) fica clara a vantagem dos chineses com 76,3% a mais de publicações que os americanos. Por fim, na subfigura (d) pode-se observar que 35% dos estudos tem como área de aplicação a computação, seguida pela engenharia com 17%, matemática com 11%, administração com 8%, energia com 8% e física com 5%.

Dentre estas publicações, a pesquisa retornou apenas uma realizada por brasileiros (STEFFENEL *et al.*, 2021) e o foco da mesma é em previsão atmosférica. A mesma busca feita na WoS trouxe resultados similares. São 6228 trabalhos, sem especificar a previsão de séries temporais e apenas 91, quando especificado.

Com base nesta análise pode-se chegar a três conclusões:

a) Arquiteturas de DL com mecanismos de atenção tornaram-se o estado da arte em diversos problemas. Especialmente, problemas de NLP.

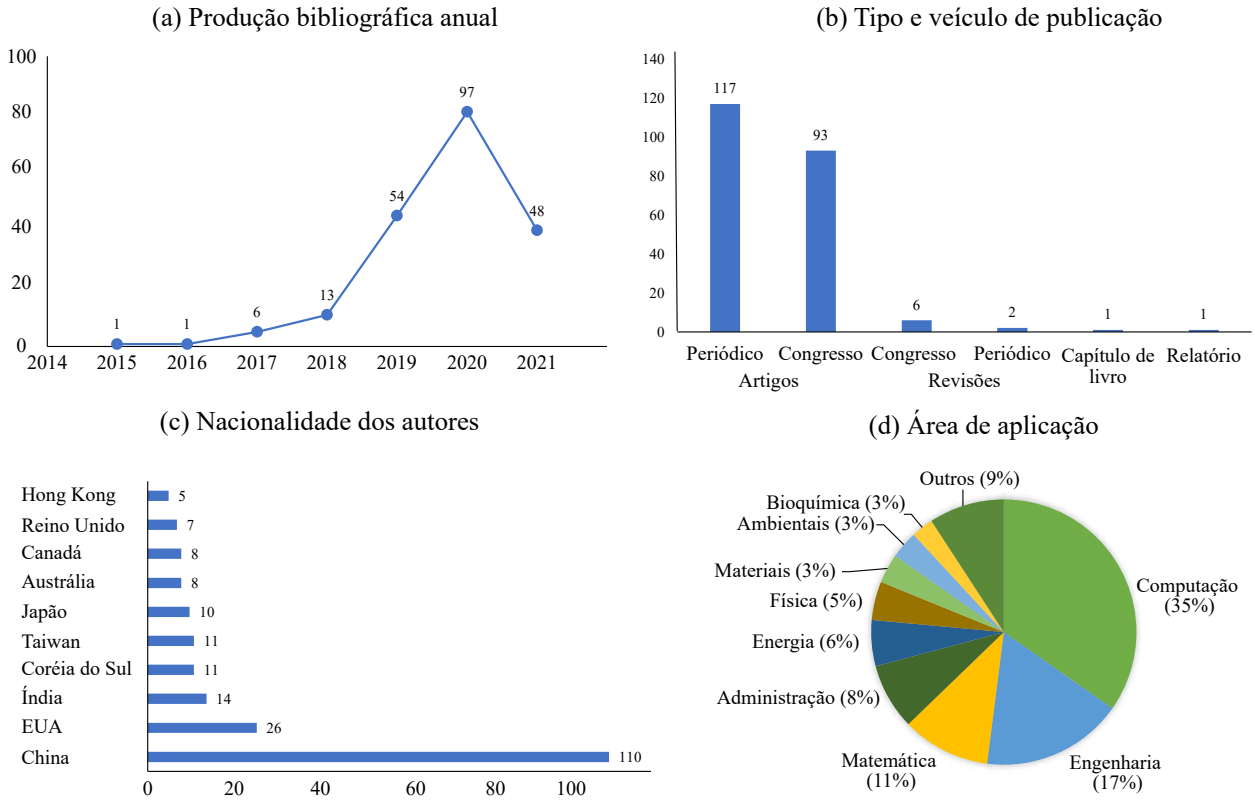
b) NLP é um problema equivalente a previsão em séries temporais.

c) DL com mecanismos de atenção em séries temporais tem sido pouco explorado até então na literatura, seja pela precocidade dos algoritmos, seja pelo foco até então dado apenas a pesquisa em NLP.

Nesse sentido se delineia a primeira meta de investigação confirmada por esta pesquisa: Os mecanismos de atenção no contexto das ANN são técnicas passíveis de uso também na previsão de séries temporais. Em relação a este ponto, nos capítulo 5 e 6 é demonstrado como o mecanismo atencional de Luong (LUONG; PHAM; MANNING, 2015) com arquitetura LSTM Seq2Seq foi capaz de ser empregado na previsão do PLD. Ademais da aplicação em si, levando-se em conta que grandezas como carga, geração e preço são dispostas temporalmente, este trabalho por consequência também apresentará nos capítulos seguintes resultados de métricas que foram aplicadas visando mensurar a capacidade preditiva desta arquitetura quando modelando a série do PLD que é dependente de carga e geração.

A previsão de carga, de geração e de preço, recentemente vem sendo explorada por meio de técnicas que permitem usufruir das possíveis relações existentes entre as mesmas. Isso, quando aferidas em um sistema unificado que agrega características de mesma natureza, para um mesmo país, uma mesma região e uma mesma população (MASHLAKOV *et al.*, 2021). É uma abordagem

Figura 41 – Análise de 220 trabalhos encontrados na base Scopus para os termos (*"attention" AND "deep learning" AND "time series" AND "forecasting"*) entre janeiro de 2014 e junho de 2021.



Fonte: Scopus (2022).

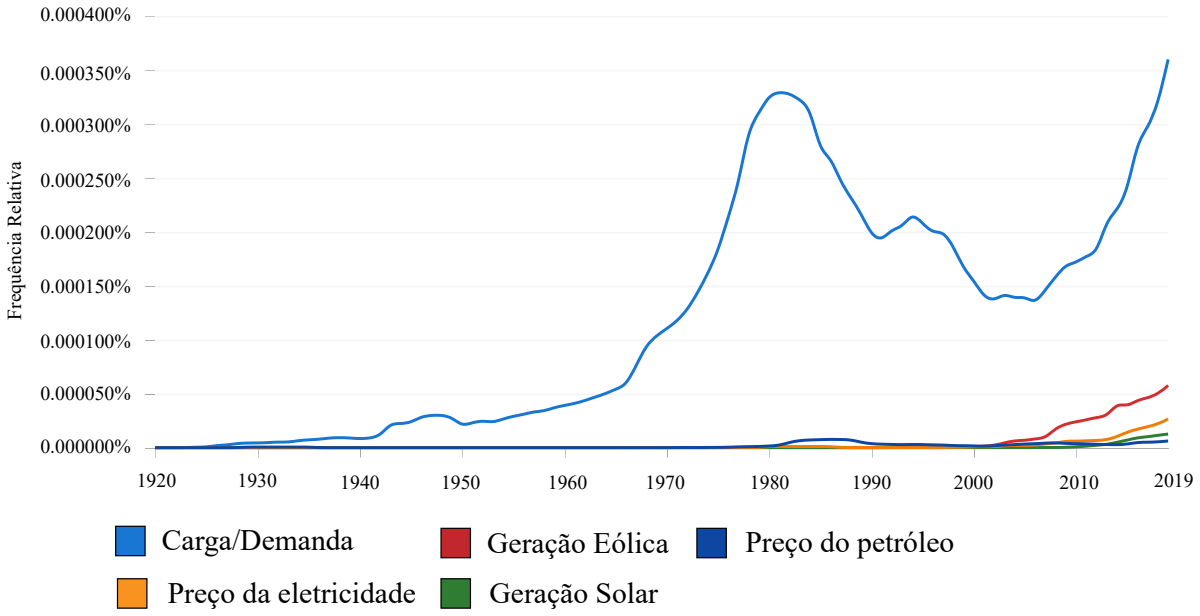
nova (HONG *et al.*, 2020; MASHLAKOV *et al.*, 2021) em que cada componente temporal não é modelada de forma separada, mas todas dentro de um mesmo modelo. Isso é especialmente importante em mercados de eletricidade, pois nesses ambientes existem regulações unificadas que determinam regras iguais de operação para toda a área de cobertura do sistema e seus usuários. O tema é chamado de previsão energética e é apresentado com foco no preço da eletricidade em detalhes na subseção a seguir.

3.2 Previsão energética

Historicamente, a modelagem de séries temporais relacionadas a energia elétrica sempre esteve majoritariamente focada na previsão de carga (kWh). De fato, as primeiras citações aos termos *Load Forecasting* e *Power Load Forecasting* aparecem na literatura já na década de 20 (Ver Figura 42). De lá pra cá, especialmente após 1970, houve um grande crescimento na área com a publicação de 20721 trabalhos e 750 patentes, segundo a base Scopus. Esse crescimento foi impulsionado pelo surgimento de métodos estatísticos e de inteligência computacional voltados para previsão (GHODDUSI; CREAMER; RAFIZADEH, 2019; HONG *et al.*, 2020). Mais de um

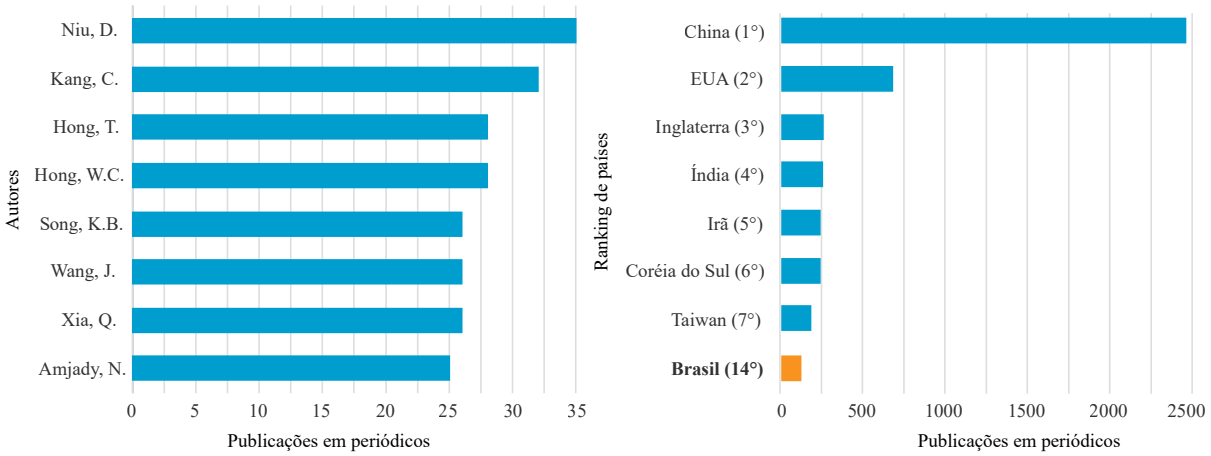
quarto destes artigos (5759) foram publicados nos últimos 20 anos. A Figura 43 mostra o *ranking* dos principais autores e países em publicações relacionadas a previsão de carga desde janeiro de 2000 até junho de 2021. O Brasil ocupa a 14º posição com 122 publicações. A Figura 44, que mostra uma análise apenas da produção brasileira, apresenta os 10 principais autores e universidades. O *ranking* é liderado pela UNESP com 18 publicações.

Figura 42 – Frequência relativa de citações literárias aos principais subdomínios da previsão energética ao longo da história.



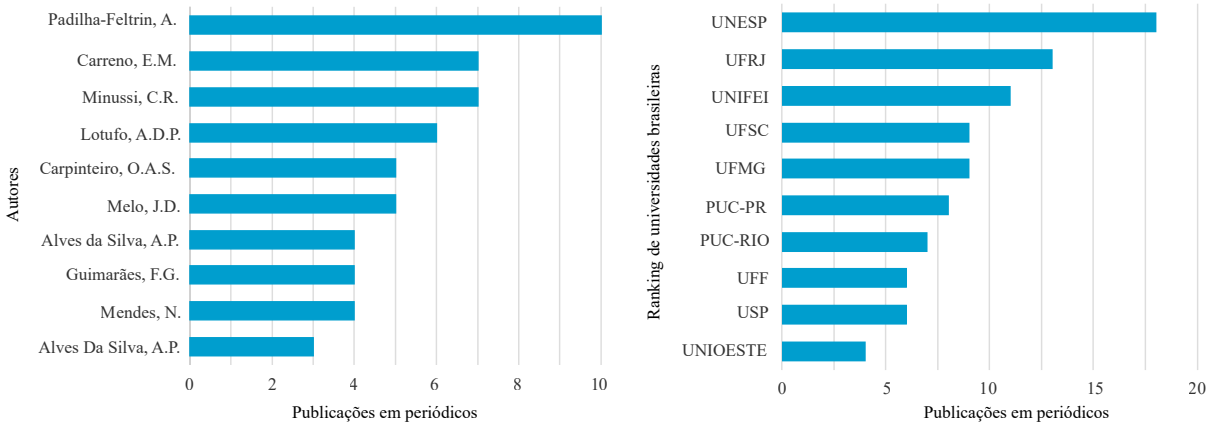
Fonte: Google Books (2021).

Figura 43 – *Ranking* de autores e países com publicações relacionadas a previsão de carga em periódicos indexados na base Scopus entre janeiro de 2000 e junho de 2021. O maior publicador é a China com 2445 artigos. O Brasil ocupa o 14º lugar com 122 publicações.



Fonte: Scopus (2022).

Figura 44 – *Ranking* de autores e universidades brasileiras com publicações relacionadas a previsão de carga em periódicos indexados na base Scopus entre janeiro de 2000 e junho de 2021. A UNESP é a primeira colocada com 18 artigos.



Fonte: Scopus (2022).

Embora a pesquisa em séries de carga já seja centenária e possua vasta literatura, ainda não há concesso sobre como classificar as milhares de abordagens propostas. Dentre estas, a mais comumente empregada é a que trata do agrupamento relacionado ao tamanho do horizonte de previsão (MISHRA *et al.*, 2020). Hong e Fan (2016) sistematizaram essa divisão por meio da definição de quatro categorias, sendo:

a) *Very Short-term Load Forecasting* (VSTLF): De segundos até um dia à frente.

b) *Short-term Load Forecasting* (STLF): De um dia até duas semanas à frente.

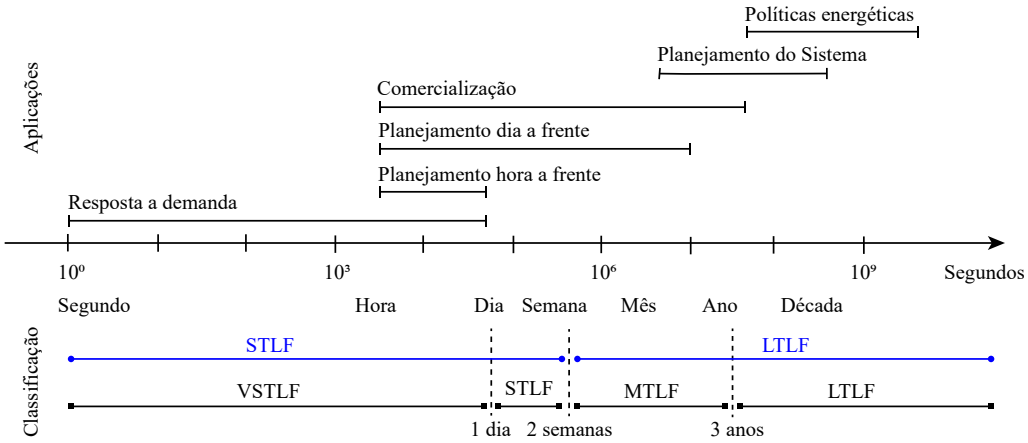
c) *Medium-term Load Forecasting* (MTLF): De duas semanas até 3 anos à frente.

d) *Long-term Load Forecasting* (LTLF): Além de 3 anos à frente.

As previsões realizadas em cada um destes horizontes servem a propósitos distintos conforme pode ser visto na Figura 45. Nesta mesma pesquisa os autores ainda propõem uma divisão mais simples baseada apenas em duas categorias, o curto (indo de segundos até 15 dias à frente) e o longo prazo (de 15 dias até décadas à frente).

Hong *et al.* (2020) estima que pelo menos 50% de toda a pesquisa empreendida em relação a séries temporais relacionadas a eletricidade vem sendo focada em previsão de carga. Isto decorre do fato de que a previsão de carga desempenha papel fundamental junto a indústria, sendo uma necessidade em diversas atividades como o planejamento e operações de sistemas de energia, programas de resposta a demanda, comercialização e definição de políticas energéticas. Previsão de carga é utilizada por entidades empresariais de diversos tipos como concessionárias, comissões reguladoras, grandes indústrias, centros comerciais e governos (HONG; FAN, 2016). Além disso, a série de carga é uma série com boas propriedades preditivas, apresentando padrões semanais, sazonalidade bem definida e pouca ocorrência de heterocedasticidade (GHELARDONI; GHIO;

Figura 45 – Classificação e aplicações da previsão de carga.



Fonte: Adaptado de Hong e Fan (2016).

ANGUITA, 2013). É uma série bem conhecida por ser correlacionada com variáveis climáticas como a temperatura. Este fato oportuniza em muitos contextos o uso de métodos simples com eficiência, como o chamado método do dia similar (GHODDUSI; CREAMER; RAFIZADEH, 2019). Nesta abordagem preditiva, o agente interessado nada mais faz do que admitir que a carga para o dia seguinte será a mesma do dia histórico mais parecido no ano anterior. Este tipo de técnica ainda hoje é utilizada (HONG, 2014), inclusive no sistema brasileiro pela ONS (ONS, 2019).

Apesar da previsão de carga possuir destaque no contexto das séries de energia, ela não é a única. Séries de energia apresentam diversos outros subdomínios. No espectro de SEP pode-se elencar as séries de geração como as renováveis eólica, solar fotovoltaica, hidroelétrica e ondas marítimas. Existem também as não renováveis como nuclear e térmicas à combustível fóssil. Todas as relacionadas a geração baseada em combustíveis ainda dão origem à outras séries, estas relacionadas ao custo de produção de fontes primárias como carvão, petróleo, diesel, gás natural e biomassa. Assim, por consequência, toda a energia negociada gera também um comportamento de mercado que dá origem, por fim, as séries de preços de eletricidade.

Com base no levantamento bibliográfico realizado para esta pesquisa, foi no trabalho de Hong (2014) que a primeira menção generalista feita a métodos de previsão aplicados a séries de energia sem distinção foi encontrada. Este autor, maior publicador de revisões bibliográficas da área (HONG, 2014; HONG, 2015; HONG; FAN, 2016; HONG; XIE; BLACK, 2019; HONG *et al.*, 2020; KEZUNOVIC *et al.*, 2020; ROSTAMI-TABAR *et al.*, 2020) utilizou o termo "*Energy Forecasting*" (traduzido livremente neste documento como Previsão Energética) dando ao mesmo a seguinte definição:

> Uma ampla gama de previsões no setor de serviços públicos, como previsão de carga de curto prazo, previsão de carga de longo prazo, previsão de carga espacial, previsão de preços de eletricidade, previsão de resposta à demanda e previsão de geração renovável e todos os insumos que compõe a cadeia de produção de energia elétrica.

Em outro artigo do mesmo autor (HONG *et al.*, 2016) este elabora melhor o termo comentando que a área vem convergindo para um campo único que trata da previsão na indústria de energia. Esta área por sua vez inclui, mas não se limita, à previsão de oferta, demanda e preço dos recursos de eletricidade, gás, água e energia renovável.

Mais tarde, em 2019, Ghoddusi, Creamer e Rafizadeh (2019) argumentaram que

> A fronteira entre a economia e a engenharia elétrica não é rígida e nem bem definida. Alguns tópicos pertencem a uma área (por exemplo, a otimização de um reator é inequivocadamente um tópico de engenharia, já a estratégia de negociação de futuros é um claro assunto de economia) mas existe uma sombra entre as áreas (por exemplo, carga de energia e previsão de preços).

Ghoddusi, Creamer e Rafizadeh (2019) concluem dizendo que deve ser adotada uma definição ampla quando lidamos com qualquer forma de variáveis relacionadas ao mercado de energia, como preços, investimentos, comportamento das empresas, perfil de consumidores e políticas públicas.

Nesse sentido, a PE apesar de ancorada em pesquisas que vem sendo desenvolvidas a décadas (como carga) é de certa forma recente, pois a visão holística de todas as séries e suas relações em um cenário macroeconômico é nova (HONG *et al.*, 2020). Além disso, esta área é por natureza um campo interdisciplinar, pois além de envolver temas em SEP e economia, deve embarcar fundamentos que tornam possível a atividade de previsão. Estes fundamentos dizem respeito a ciências como matemática, estatística e computação. Importante também destacar o papel dos recursos que influenciam o comportamento das séries de energia e que podem ser tratados pelas ciências meteorológicas, sociais, políticas, econômicas e outras. Hong *et al.* (2020) em um levantamento sobre o tema mostra, por exemplo, alguns cenários de interação entre as séries de energia discutindo a movimentação do preço da eletricidade. O autor comenta que é necessária a colaboração e estreitamento dos pesquisadores de todos os subdomínios. Nesse sentido, a pesquisa deve privilegiar times interdisciplinares compostos não só de engenheiros eletricistas, mas também por cientistas de dados, economistas e meteorologistas. Ele finaliza descrevendo algumas perspectivas para área. Nestas conclusões é feita uma suposição dizendo que, na próxima década, a fronteira do conhecimento neste campo será avançada principalmente por meio de pesquisas colaborativas.

Com base nesta análise, nesta tese é adotado e mencionado o termo PE com ampla abordagem, no sentido de não se tratar apenas de carga, mas de toda e qualquer série relacionada a energia elétrica que possa ser submetida a métodos de previsão. De forma auxiliar é proposto (pela primeira vez até onde percebeu-se pela pesquisa bibliográfica) um posicionamento para a PE entre os temas que a compõe conforme diagramado na Figura 46. Nesta imagem são cruzadas as áreas de SEP e Economia, de onde advém os fenômenos que geram as séries temporais de energia e, ao mesmo tempo, as áreas de Aprendizagem de Máquina e Análise de Séries Temporais, que fornecem o embasamento teórico para desenvolvimento e aplicação dos métodos de previsão. O sombreamento entre duplas ou trios dessas grandes áreas levam a oito temáticas distintas, sendo:

a) SEP e Economia.

   – Mercados de Energia: Como já discutido com detalhes no Capítulo 2 é uma área que envolve a aplicação de teoria econômica, como economia da energia, para tratar da negociação de ativos energéticos estudados em SEP.

b) Economia e Análise de Séries Temporais.

   – Econometria: Trata da aplicação de métodos de análise de séries temporais e estatística para a realização de modelagem e previsão em séries econômicas.

c) Análise de Séries Temporais e Aprendizado de Máquina.

   – Aprendizado de máquina para previsão: Trata da aplicação de qualquer método de aprendizado de máquina capaz de modelar sequencialmente uma série temporal e realizar previsões sobre esta.

d) Aprendizado de Máquina e SEP.

   – Modelagem e simulação: Uso de métodos de aprendizagem de máquina para modelar atividades de forma geral em SEP, não necessariamente apenas previsão em séries temporais.

e) SEP, Economia e Análise de Séries Temporais.

   – Econometria aplicada a mercados de energia: Uso de técnicas econométricas para a realização de previsão em séries relacionadas aos mercados de energia como preço de energia ou custos de tarifas.

f) Economia, Análise de Séries Temporais e Aprendizado de Máquina,

   – Finanças quantitativas: Área que trata do uso técnicas de computação inteligente e/ou estatística para realizar modelagem e previsão de séries econômicas. Envolve também a negociação automatizada de ativos financeiros como ações, contratos, cotas de fundos e moedas.

g) Análise de Séries Temporais, Aprendizado de Máquina e SEP.

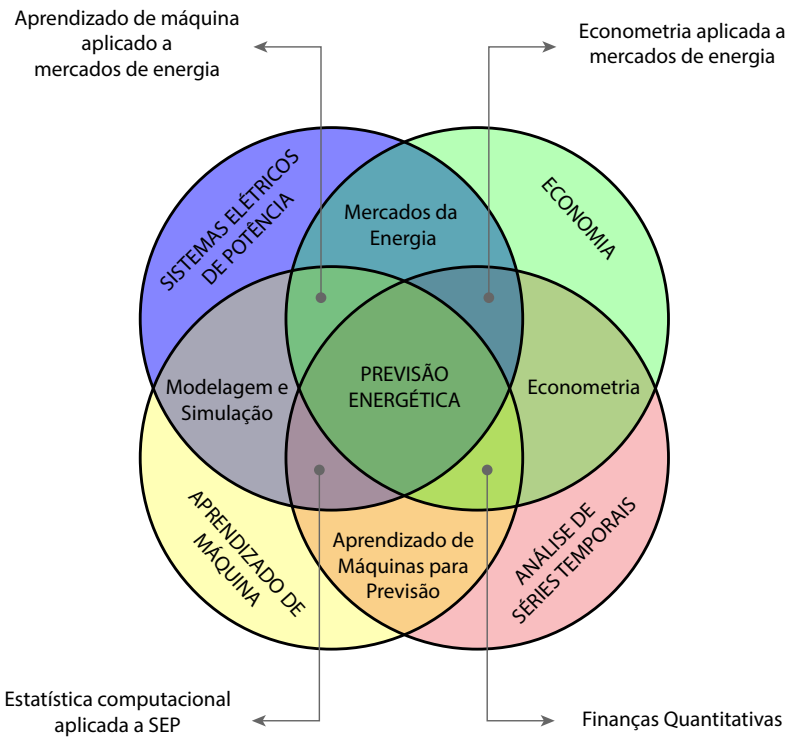
   – Estatística computacional aplicada a SEP: Uso de métodos estatísticos baseados em aprendizagem de máquina para modelar, simular e prever comportamento de fenômenos em SEP que são descritos sequencialmente.

h) Aprendizado de Máquina, SEP e Economia.

   – Aprendizado de máquina aplicado a mercados de energia: Uso de técnicas de computação inteligente para realização de previsão em séries relacionadas a mercados de energia.

Deste ponto de vista (suportado pelas também pelas definições de Hong *et al.* (2020)) a PE pode ser entendida com base em seus principais subdomínios que são quatro: Carga (do inglês, *Load Forecasting* (LF)), preço de energia (do inglês, *Electricity Price Forecasting* (EPF)), geração eólica

Figura 46 – Diagrama de conceituação da área de previsão energética.



Fonte: O Autor.

(do inglês, *Wind Power Forecasting* (WPF)) e solar (do inglês, *Solar Photovoltaic Forecasting* (SPF)). Outros subdomínios como preços de combustíveis diversos (ZHAO; LI; YU, 2017), geração fóssil (HRIBAR *et al.*, 2019), resposta a demanda (WANG *et al.*, 2020), potência reativa (BRACALE *et al.*, 2019) e variáveis específicas relacionadas a *smart grids* como carga de edifícios (ALASALI *et al.*, 2018; BERK; HOFFMANN; MÜLLER, 2018) também vem sendo exploradas mais recentemente na literatura (últimos cinco anos). No entanto, frente as quatro principais áreas, estes outros subdomínios ainda são marginais e possuem produção científica modesta. Como visto na Figura 42, a previsão solar e eólica só começou a ter destaque a partir de 2000 com a introdução das fontes renováveis às matrizes de diversos países que começaram, como discutido no Capítulo 2, a buscar diversificação de portfólio. A produção científica relacionada a previsão de preços de eletricidade, também como mostrado no Capítulo 2, coincide com a maturidade dos mercados desregulados que só vieram a ganhar força a partir de 2000. A previsão de preços de petróleo também é um dos temas, mas que teve foco maior somente durante a década de 80.

A primeira revisão bibliográfica no âmbito da PE foi focada em previsão de carga espacial de longo prazo (WILLIS; NORTHCOTE-GREEN, 1983). A carga espacial trata da modelagem de regiões geográficas específicas. Além de levar em conta o histórico de consumo, também envolve planejamentos governamentais relativos a concessões de terras e projetos de expansão de energia. Quanto a STLF, a primeira revisão foi a de Gross e Galiana (1987) que abordou questões práticas do campo na década de 80 levantando e discutindo desafios da previsão de carga para até 168 horas

à frente. O autor aborda a influência de fatores econômicos e climáticos e faz uma classificação dos modelos existentes. Entre 1990 e 2000, durante a segunda onda das RNAs muitos trabalhos passaram a envolver meteorologistas (HONG *et al.*, 2020). A revisão de Hippert, Pedreira e Souza (2001) foi a primeira a tratar do uso de RNAs para previsão de carga. Os autores dedicam grande parte do artigo a realização de críticas a diversos estudos que não fizeram o uso correto da técnica e acabaram por gerar resultados duvidosos. Neste trabalho são apontadas questões de ordem prática que devem ser observadas antes de se aplicar métodos de inteligência artificial. Na década de 2000, a revisão com maior número de citações é a de Hong e Fan (2016). Esta revisão, pela primeira vez, abordou a previsão de carga sob a ótica probabilística, onde ao invés de se prever um ponto, o método de previsão entrega uma curva de densidade. Essa revisão é especialmente importante, pois aborda um tema que mais tarde foi amplamente adotado em vários outros subdomínios. A previsão de carga probabilística foi também tema do campeonato de previsão de carga GEFCom 2014 (HONG *et al.*, 2016). À medida que medidores inteligentes vêm sendo adotados e as *smart grids* se tornam realidade, a previsão de carga industrial ou em edifícios começa a florescer. A revisão de Zhao e Magoulès (2012) avalia fatores específicos neste subdomínio como características de edifícios, iluminação, fatores ambientais e tipos de estruturas físicas prediais.

Quanto a preços de energia, pode-se dar destaque a três revisões de literatura. No trabalho de Zareipour (2012) são discutidas as características das séries de preços levando-se em conta diversos mercados. Na sequência, os autores descrevem diretrizes sobre como construir um sistema para previsão de preços de curto prazo. São elencadas atividades como pré-processamento, seleção de variáveis, escolha da técnica de predição e ajustes dos modelos adotados. Neste estudo são abordados apenas métodos estatísticos. Outra revisão relacionada a preços, essa a mais citada segundo as bases WoS e Scopus é a de Weron (2014). O autor avalia 304 artigos da WoS e 497 da Scopus publicados até 2013, dos quais 342 são de periódicos. É comentado que, com raros casos isolados, a literatura relacionada a previsão de preços de energia é praticamente inexistente antes de 2000, com algum avanço ocorrendo apenas a partir de 2006. São apresentadas métricas de avaliação para previsão pontual e, também, é descrita uma taxonomia dos métodos. Mais recentemente, dois trabalhos de Jesus Lago (LAGO; De Ridder; De Schutter, 2018; LAGO *et al.*, 2021) revisitaram a revisão de Weron e a atualizaram com base nos últimos avanços (2014-2020), especialmente os relacionados a *Deep Learning*. As conclusões destes dois estudos são similares ao dizer que, apesar dos significativos avanços com uso de DL, essa área ainda vem sendo pouco explorada em previsão de preços de energia. Os principais motivos sugeridos são a dificuldade técnica de implementação, o alto custo computacional e o ceticismo dos pesquisadores decorrentes do mau uso das RNAs no passado. Em Lago *et al.* (2021) são propostas diretrizes para publicação de artigos em EPF. Os autores citam vários artigos problemáticos e apontam soluções para essas deficiências. Por fim, pode-se citar também a revisão de Nowotarski e Weron (2018) que aborda a EPF do ponto de vista probabilístico.

No que se refere a previsão de geração eólica, a primeira revisão encontrada foi publicada em 2010 por Soman *et al.* (2010). Neste trabalho os autores discutem métodos que fazem inicialmente a previsão da velocidade do vento e, na sequência, a usam para estimar a capacidade de

geração. O artigo também discorre sobre a importância da previsão climática numérica (*Numeric Weather Prediction* (NWP)), métodos estatísticos, RNAs e técnicas híbridas sendo aplicadas a diferentes horizontes de previsão. Como na previsão de carga, neste estudo é feita também uma categorização entre curtíssimo, curto, médio e longo prazo. A revisão de Costa *et al.* (2008) aborda apenas a WPF no curto prazo realizando um levantamento de 30 anos. É analisado o estado da arte antes de 1990, durante a década de 90 e após a década de 90. São apresentadas as experiências do Canadá, Noruega, Finlândia, Alemanha, México, Holanda, Dinamarca e Suécia. Já o *survey* realizado por Zhang, Wang e Wang (2014) aborda a WPF probabilística. O estudo comenta a análise de incerteza relacionada ao vento e o impacto das previsões no processo decisório relacionada a programação e operação de sistemas, além de preço. São avaliados de forma quantitativa estudos que levam em conta risco e diversos cenários climáticos possíveis. Os autores discutem ainda os principais métodos de previsão eólica probabilística, com destaque para os *ensembles* que visam agregar diversas previsões em uma. A questão da incerteza relacionada ao vento é especificadamente discutida no artigo de Pinson (2013). Este autor aborda esta característica no contexto das previsões pontuais, probabilística e espaço-temporais. É feita uma análise física do comportamento do vento e como a meteorologia tem papel crucial no desenvolvimento de modelos para prever a geração eólica. Por fim, pode-se citar a revisão de Sweeney *et al.* (2020) que traça um paralelo entre previsão eólica e solar. É abordado neste estudo como novos produtos de energia podem ser comercializados e o impacto desses na tomada de decisão do consumidor. Nesse contexto, os autores comentam que produtos futuros de energia precisarão incluir informações probabilísticas e a entrega desses deverá ser feita de forma personalizada ao consumidor final. Indo além, é comentado também que novas tecnologias de transações deverão ser adotadas nos mercados de energia como o uso de *blockchain* [12].

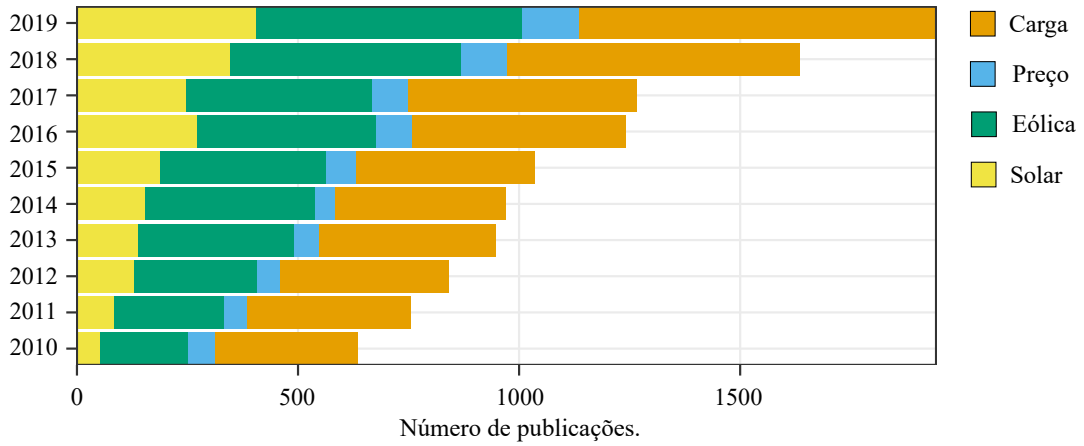
Em relação ao subdomínio da previsão de geração solar fotovoltaica, uma das primeiras e a mais citada revisão é a de Inman, Pedro e Coimbra (2013). Neste trabalho os autores exploram as variáveis que viabilizam a SPF como irradiância, massas de ar e condições de céu aberto. É definida uma taxonomia para os métodos discutindo modelos regressivos e autorregressivos, além de técnicas baseadas em RNAs. São abordadas ainda fontes de coleta de dados como medições em solo, via satélite e outras feitas por agências meteorológicas. A análise é conduzida com base em janelas futuras que vão de 30 a 120 minutos. No estudo de Yang *et al.* (2018) são analisados 1000 trabalhos utilizando uma abordagem bibliométrica baseada em mineração de texto no período de 2013 a 2017. Destes artigos, foram filtrados 249. Este trabalho foi realizado pelos editores do periódico *Solar Energy*, a revista da área com maior fator de impacto. A principal contribuição desta revisão é a sistematização de diversos conceitos em SPF que são necessários para pessoas que desejam ingressar no campo. Complementarmente a este estudo, Yang também propôs o conjunto de diretrizes ROPES (do inglês, *reproducible, operational, probabilistic or physically based ensemble and skill*) (YANG, 2019). Estas diretrizes funcionam como uma sequência de atividades para produção de pesquisas em SPF. O trabalho de Meer, Widén e Munkhammar (2018)

[12] *Blockchain* é uma tecnologia de registro de informações que permite o compartilhamento e a segurança de transações por meio de uma rede de membros descentralizada (YANG *et al.*, 2021).

explora a SPF probabilística. Por fim, na publicação de Yang *et al.* (2020) (acompanhado de 18 coautores, incluindo Tao Hong) é feita uma tentativa de criar uma nomenclatura e diretriz definitiva para SPF. Isto é feito fornecendo um *framework* geral para pesquisas em previsão solar de forma que todos os trabalhos que utilizem estas práticas possam ser comparados entre si.

Baseados nestes principais *surveys* acerca dos quatro principais subdomínios da PE, Hong *et al.* (2020) realizaram um agrupamento por subdomínio e, na sequência, calcularam a razão do número de artigos em cada subdomínio pelo total de publicações, por revista. Dessa forma os autores argumentam que foi possível identificar os periódicos de mais alta relevância em PE. Identificados os periódicos, foi então levantada toda a produção bibliográfica gerada entre 2010 e 2019. No gráfico da Figura 47 é mostrado o percentual de publicações por revista e na Figura 48 é exibido o número de pesquisas ao longo dos anos. Conforme pode ser observado, todos os quatro subdomínios vêm crescendo sendo que na última década todas as áreas, no mínimo, dobraram de tamanho.

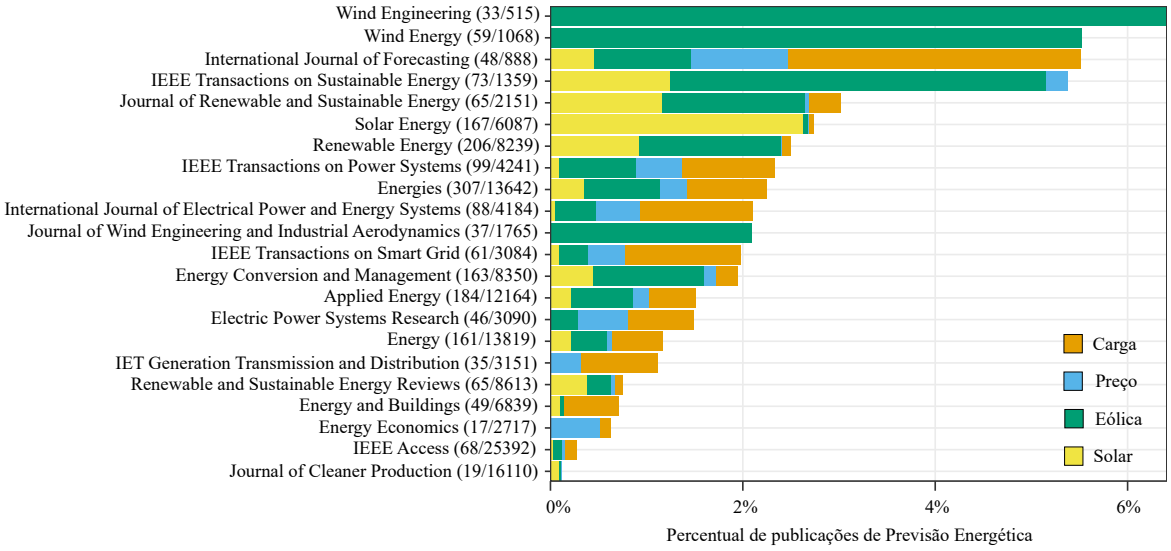
Figura 47 – Percentual de publicações sobre previsão energética nos 22 principais periódicos área entre 2010 e 2019.



Fonte: Hong *et al.* (2020).

O crescimento da área de PE pode ainda ser analisado segundo a maturidade da qualidade preditiva dos modelos existentes em termos de acurácia. Na Figura 49 a previsão de carga dividida em curto prazo (indicada por STLF) e longo prazo (indicada por LTLF), a previsão eólica (indicada por WPF), a previsão solar (indicada por SPF) e previsão de preços de eletricidade (indicada por EPF) são dispostas em quatro quadrantes. Os quadrantes são definidos com base nas abordagens de previsão pontual e probabilística. Dado que a temperatura é um fator fortemente correlacionado à carga e que previsões de temperatura são precisas no curto prazo, então STLF foi posto como subdomínio que apresenta maior maturidade em previsões pontuais. A previsão de geração eólica, sendo dependente da previsão da velocidade do vento, por sua natureza vem sendo desenvolvida de forma acoplada a diversos modelos probabilísticos de previsão meteorológica de longo prazo. Fato que também ocorre em LTLF que conta com modelos de precipitação anuais. Nesse sentido, as áreas de WPF e LTLF são as mais maduras em previsões probabilísticas. A penetração de energia solar não foi significativa até a última década, devido a esta situação SPF é ainda imatura

Figura 48 – Total de publicações sobre previsão energética nos 22 principais periódicos da área entre 2010 e 2019.



Fonte: Hong *et al.* (2020).

tanto em pontuais quanto probabilísticas. Já EPF, dada a dificuldade de modelar incertezas de mercado, ainda é imatura do ponto de vista probabilístico. No entanto, menos que SPF por não ser tão recente quando esta. No espectro da previsão pontual, EPF apresenta relativa maturidade, pois bons resultados no curto prazo podem ser obtidos devido a correlação existente com a carga, o que não ocorre no longo prazo (HONG *et al.*, 2016).

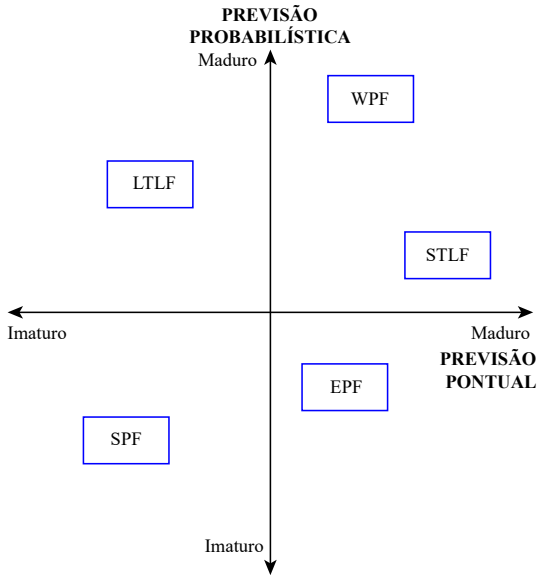
Sendo EPF o subdomínio de PE onde se enquadra a proposta desta tese, na subseção a seguir são discutidos detalhes que levam a definição do atual estado da arte neste tema.

3.3 Previsão de preços de eletricidade

O área de EPF tem por objetivo prever os preços futuros à vista nos mercados de eletricidade seja em cenário pontual ou probabilístico, contudo, esta tarefa não é trivial (LAGO *et al.*, 2021).

Antes de se desenvolver e aplicar qualquer modelo em EPF, há que se considerar a diversidade de regulamentos que regem mercados específicos ao redor do mundo. Fato que exige que todo e qualquer método seja adaptado não somente no tempo, mas também ao local de uso (LAGO *et al.*, 2021). Em segundo lugar, os mercados de eletricidade comercializam um insumo único, mas que no entanto possui características muito particulares. A energia elétrica não é armazenável, fato que torna todo o sistema elétrico dependente de um constante equilíbrio entre produção e consumo. Dado que o preço ainda sofre influência das mais diversas variáveis, erros de previsão nesse contexto são particularmente destrutivos, especialmente para concessionárias (ZAREIPOUR; CANIZARES; BHATTACHARYA, 2010). Como visto no Capítulo 2 em relação a Crise da Califórnia, uma vez que geralmente estes agentes não podem repassar seus custos aos consumidores, situações

Figura 49 – Quadrante de maturidade das principais áreas de previsão energética. LTLF (Previsão de Carga de Longo Prazo), STLF (Previsão de Carga de Curto Prazo), WPF (Previsão de Geração Eólica), SPF (Previsão de Geração Solar) e EPF (Previsão de Preço de Eletricidade).



Fonte: Hong *et al.* (2016).

de excesso ou subcontratação exigirão a compensação dos déficits nos mercados de balanceamento e tempo real. Esses ambientes tipicamente expõe os agentes a preços mais altos, logo podem levar os mesmos a consideráveis perdas financeiras. Um terceiro ponto é que preços de eletricidade podem ser até duas vezes mais voláteis do que os de qualquer outra *commodity* (WERON, 2014). Situação que força os comercializadores a se protegerem não apenas contra o risco relacionado ao volume negociado, mas também contra a movimentação (às vezes imprevisível) no curto e médio prazo. Preços são conhecidos por apresentar comportamento volátil, sendo assim, a investigação dos motivos desta volatidade podem fornecer indicações sobre como prevê-lo (MORETIN; TOLOI, 2006; GUJARATI; PORTER, 2011). No caso da energia elétrica, algumas destas indicações podem se tratar da inelasticidade da demanda, a correlação positiva (mas não linear) entre demanda e preço, condições climáticas, custo de combustíveis, programação de interrupções, calendário, congestionamento de sistema, capacidade de reserva e tecnologias de geração (MAYO, 2012; SILVA, 2012).

A prática de EPF, apesar de apresentar desafios, tornou-se um tópico de grande interesse de gestores de portfólio de energia interessados em aperfeiçoar seu planejamento de curto e médio prazo (WERON, 2014). Um gerador, uma empresa de serviços públicos ou um grande consumidor industrial que tem capacidade de prever preços com um razoável nível de precisão, pode ajustar suas estratégias e otimizar entradas em contratos, licitações, leilões e até mesmo seu cronograma de produção ou consumo (WERON, 2014; MARCJASZ; LAGO; WERON, 2020). Nesse sentido, uma variedade de métodos e ideias vem sendo publicadas demonstrando sucesso na modelagem das séries de preços de eletricidade.

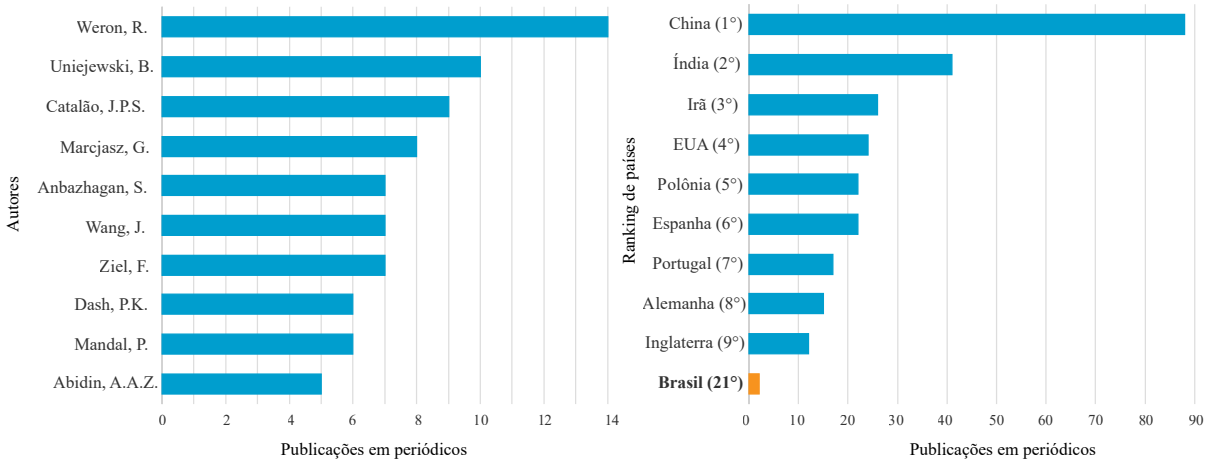
Uma das primeiras discussões sobre modelagem de preços de eletricidade é feita no livro *The US Power Market* (KAMINSKI, 1997). Embora neste livro o assunto sobre previsão não seja o foco, os autores foram um dos primeiros a discutir a série de preço de eletricidade e o impacto de sua movimentação em preços de contratos e derivativos. Os autores comentam que modelos de preços de eletricidade devem incorporar alguma forma de volatilidade variável no tempo. Siddiqi e Kader (1994) discutem a aplicação do computador como ferramenta auxiliar no gerenciamento de riscos por meio da avaliação de retornos financeiros de contratos. Os autores introduzem um *software* chamado POWERCOACH que visa auxiliar analistas nas transações envolvendo energia elétrica. Em Bastian *et al.* (1999) um modelo de simulação multiagente denominado MAPS é apresentado para o mercado americano PJM realizando previsões unitárias de cinco minutos à frente. Szkuta, Sanabria e Dillon (1999) fazem uma das primeiras aplicações de RNAs para previsão no curto prazo. Os autores usam uma *Multilayer Perceptron* (MLP) de saída única treinada com *Backpropagation* no California ISO. A RNA prevê o preço diário para um dia à frente e foi aplicada em apenas uma semana do ano 1997. Os trabalhos Yao *et al.* (2000) e Kim, Yu e Song (2002) foram um dos primeiros a fazer uso de transformações (*wavelet*) antes da aplicação do método preditor. Ambos estudos usaram dados do mercado inglês. Em Hong e Hsiao (2002) é explorada pela primeira vez o uso de *Recurrent Neural Networks* (RNNs) e modelagem *fuzzy* em EPF. O estudo foi desenvolvido para o mercado americano PJM e utilizou dados de abril a maio de 1999. A previsão em EPF usando *ensembles* é hoje reconhecidamente uma das melhores práticas (WERON, 2014; HONG *et al.*, 2020; LAGO *et al.*, 2021). Uma combinação de previsões feitas por RNAs com arquiteturas *Radial Basis Function* (RBF) e MLP demonstrou que comitês de decisão são capazes de apresentar performance superior a métodos individualizados. Neste estudo foram utilizadas 23 variações de entradas para a RBF e 55 para a MLP para prever de maio a outubro de 2002 o preço no mercado de *New England* (GUO; LUH, 2004). Finalmente, quanto aos trabalhos seminais da área, pode-se dar destaque a primeira modelagem de volatilidade em EPF realizada para o período entre abril de 1998 e agosto de 2000 (KNITTEL; ROBERTS, 2005). Foram avaliadas as propriedades estocásticas do preço depois da reestruturação do mercado californiano. Os autores comentam que métodos estatísticos clássicos tiveram pouco uso prático devido a fortes ciclos determinísticos intradiário, semanais e efeitos sazonais de longa escala no preço.

Como pode-se notar, a literatura em EPF entre 2000 e 2005 basicamente se concentra nas primeiras experimentações de métodos preditivos em vários mercados diferentes. Conforme mostrado na Figura 42, a partir deste ponto a pesquisa ganha força. De 2006 em diante até junho de 2021, numa busca pelo termo "*Electricity Price Forecasting*" junto as bases Scopus e WoS são retornados, respectivamente, 636 e 440 trabalhos. Uma quantidade que pode ser considerada pequena, visto que representa somente 5,1% e 10,5% em relação a previsão de carga que no mesmo período retorna 12448 e 4168 artigos. Levando-se em conta apenas a base Scopus, são 276 publicações em conferências e 317 em periódicos. As veiculadas em periódicos estão categorizadas na Figura 50.

Rafal Weron, professor de ciências da administração da Universidade de Ciência e Tecnologia de Wroclaw na Polônia, além de ser o autor com maior número de publicações (14) é também

o mais citado da área. Sua revisão de 2014 (WERON, 2014) é o trabalho com maior quantidade de citações (680). A China lidera o *ranking* de países com 72 publicações. O Brasil possui apenas 2 publicações em periódicos e 3 em conferências. Estes estudos nacionais são individualmente discutidos a seguir.

Figura 50 – *Ranking* de autores e países com publicações relacionadas a previsão de preço de eletricidade em periódicos indexados na base Scopus entre janeiro de 2006 e junho de 2021. O maior publicador é a China com 72 artigos. O Brasil ocupa o 21º lugar com apenas 2 publicações.



Fonte: Scopus (2022).

Filho, Affonso e Oliveira (2009) desenvolveram um modelo baseado em árvores de decisão com pré-processamento de dados baseado em clusterização. Os autores utilizaram a biblioteca WEKA (HALL *et al.*, 2009) e o modelo foi aplicado aos quatro submercados brasileiros no período de maio de 2003 a dezembro de 2008. O estudo diz respeito a previsões de PLD mensal, logo foram utilizadas apenas 55 observações, quantidade que para uma série temporal de preços de energia é considerada pequena segundo limites mínimos propostos em Weron (2014), Hong *et al.* (2020) e Lago *et al.* (2021). As entradas incluíram preços, energia armazenada, geração hídrica, geração térmica e carga histórica, além de mês e estação. O treinamento ocorreu em 90% dos dados com validação cruzada e o teste nos outros 10%. Os autores sugerem que árvores de decisão podem ser utilizadas para previsão no curto prazo.

Coelho e Santos (2011) desenvolveram um estudo aplicado ao mercado de eletricidade espanhol para realizar previsões para 24, 48 e 72 horas à frente. A avaliação foi feita em termos de *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) e outras métricas [13]. No trabalho os autores utilizaram uma rede com arquitetura RBF associada a um modelo *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH). A parametrização do GARCH foi feita via método de verossimilhança e da RBF com algoritmo genético. Foram

[13] As métricas comumente utilizadas para avaliação de previsão de modelos em EPF são discutidas na subseção 4.1.3.1.

utilizados 80% dos dados para treinamento e 20% para teste. Não foi informada a forma de validação e o conjunto de dados compreende apenas o último trimestre de 2008. Nesta situação, como comenta Lago *et al.* (2021) não é possível avaliar a capacidade preditiva em todas a situações que um ano inteiro contém. A principal falha deste estudo, contudo, foi a utilização de *benchmarks* regressivos com apenas um termo. Com termo único, a modelagem da série (mesmo considerando volatilidade condicional) gera modelos sempre lineares. Dado que a RBF e o GARCH podem modelar não linearidades, então a comparação se torna injusta.

Em novo estudo de Filho, Affonso e Oliveira (2015) os autores exploraram o uso de árvores de decisão, além de uma RNA. As mesmas foram aplicadas ao submercado da região norte do Brasil. Foram feitas também combinações na forma de *ensemble* de pares desses modelos. Considerando o PLD semanal foi classificada a magnitude do preço segundo patamares muito baixos, baixos, médios, altos e muito altos. Neste sentido, foi previso um valor único de classificação para uma semana à frente. O período utilizado foi de agosto de 2002 a dezembro de 2013, contendo 500 observações. Foram utilizadas como variáveis de entrada a carga média, energia armazenada e afluência dos reservatórios. Os autores utilizaram o *software* IBM SPSS para importar os modelos, contudo, não comentam qual parametrização ou arquitetura de RNA foi usada. Não é informado também quais períodos utilizados para treinamento, validação ou teste.

Ribeiro *et al.* (2020) utilizaram como conjunto de dados o preço mensal de eletricidade brasileiro com 289 observações de abril de 1996 a dezembro de 2019 em séries do setor industrial e comercial. Os dados foram divididos em 70% para treino e 30% para testes com treinamento via validação cruzada. Foi utilizado o *Coyote Optimization Algorithm* (COA) para parametrizar um *Empirical Model Decomposition* (EMD) responsável por selecionar os recursos de entrada que foram preços, demanda comercial, demanda industrial, geração hidráulica, térmica e nuclear. Os métodos utilizados foram *Extreme Learning Machine* (ELM), GBM, processo gaussiano e *Support Vector Regression* (SVR) no contexto de um *ensemble*. Para avaliação foram adotadas as métricas *Symmetric Mean Absolute Percentage Error* (sMAPE), RMSE, R^2 e outras. A janela de previsão utilizada foi de um a três valores a frente e a otimização dos modelos foi baseada em *grid-search*. Os autores comentam que o método proposto foi capaz de superar os modelos individualizados, *Autoregressive Integrated Moving Average* (ARIMA) e *naïve*.

Por fim, Gontijo, Costa e Santis (2021) aplicaram *Dynamic Time Scan Forecasting* (DTSF) para identificar janelas passadas no PLD semanal similares e, com base nessas, estimar o preço futuro nos quatros submercados do Brasil. Foram utilizadas 701 observações referentes a janeiro de 2006 a maio de 2019. Adicionalmente, foram utilizadas como entradas os patamares leve, médio e pesado da carga histórica. Os *benchmarks* utilizados foram técnicas como *naïve* e suavizações *Seasonal Exponential Smoothing* (SES), *Damped* e Holt-Winter. Como métricas foram utilizadas sMAPE e *Mean Absolute Scaled Error* (MASE). É relatado apenas o caso da região sudeste com patamar de carga pesada. O DTSF, segundo os autores, superou os concorrentes na primavera, no entanto, nas outras estações não.

Conforme pode-se notar, além da literatura de trabalhos indexados com filiações brasileiras ser escassa, apenas uma destas publicações foi veiculada em um dos periódicos listados na Figura 48. Adicionalmente, todos esses estudos apresentam problemas como (i) conjunto de dados com poucas observações, (ii) conjuntos de dados que não extrapolam, no mínimo, um ano por completo, (iii) faltam informações sobre como os dados foram divididos para treinamento, teste e validação, (iv) não está claro como os métodos foram parametrizados e escolhidos e (v) são feitas comparações com *benchmarks* não otimizados [14]. Essas deficiências não são exclusividade de trabalhos nacionais, as mesmas vem sendo sistematicamente relatadas em revisões bibliográficas não só no subdomínio da EPF, mas também em toda a área de PE e previsão no geral (HONG *et al.*, 2020). Nowotarski e Weron (2018) e Lago *et al.* (2021) comentam que o progresso no campo é difícil de acompanhar, pois falta rigor estatístico às análises e, por causa disso, comparações entre métodos ainda são difíceis de se realizar. Estes autores apontam as seguintes falhas:

a) Existem vários estudos que comparam métodos de ML com estatística. No entanto, as conclusões são contraditórias. Isto ocorre, pois quando o viés do trabalho é relacionado a estatística, os autores recorrem a *benchmarks* de ML muito simples e, em contrapartida, quando os autores promulgam um método baseado em ML, então técnicas simples de estatística são usadas para comparar.

b) Os períodos de testes são muito curtos. Em vários casos limitados a apenas uma semana. Em Xiao *et al.* (2017), por exemplo, o período de avaliação se limita a somente um dia. Esta situação é especialmente preocupante em EPF, pois tende a ignorar o comportamento dos preços em dias especiais como feriados ou diferentes estações climáticas do ano.

c) As pesquisas são dificilmente reprodutíveis, pois não publicam dados e código usados. Além disso, muitos trabalhos (como alguns dos nacionais citados) não especificam com exatidão a divisão dos conjuntos de dados, não indicam as entradas dos modelos, o período de aplicação ou não informam em qual mercado o método está sendo testado.

d) No caso de estudos que usam métodos híbridos, a contribuição individualizada de cada um dos métodos não é informada. Dessa forma, não há como saber se a hibridização em si traz alguma contribuição.

e) Por fim, muitos trabalhos evitam comparações com métodos recentes ou alguns clássicos bem estabelecidos.

Para mitigar estas deficiências, estes mesmos autores propõem três ações.

[14] Foi também realizada uma pesquisa para os termos (*"previsão" AND "PLD horário"*) no Google Scholar. Dentre os 34 resultados retornados, todos tratavam-se de monografias de graduação ou dissertações de mestrado explorando o PLD em escala semanal ou mensal (mesmo com a especificação do termo *"horário"*). A maioria desses estudos explora a avaliação do impacto financeiro na mudança para precificação horária em contextos variados como parques eólicos e contratos de hidroelétricas reversíveis.

Em primeiro lugar, deve-se utilizar um ambiente de programação popular que possua caixas de ferramentas adequadas que recebem constantes atualizações. Estas devem ser suportadas por empresas ou associações com reconhecida *expertise* na área e, em especial, aquelas que mantém repositórios públicos com código fonte aberto. Em Lago *et al.* (2021) são recomendados os pacotes *open source* Keras e Tensorflow para uso em linguagem de programação Python, Java ou R. É comentado também por estes autores que

> Novos métodos vêm sendo comparados utilizando-se de modelagens baseadas em redes neurais da caixa de ferramentas do MATLAB®. Uma plataforma que nessa área há muitos anos se mantém desatualizada e não inclui muitos dos desenvolvimentos cruciais à EPF, como por exemplo, funções de ativações de última geração e até mesmo métodos de otimização mais recentes.

Em segundo lugar, dados e códigos-fonte devem ser disponibilizados publicamente para promover reprodutibilidade. Quanto aos dados, deve-se ainda descrever a forma de coleta indicando a fonte de extração. Os conjuntos devem compreender longos períodos para treino com máxima resolução possível. Em Weron (2014) são recomendadas resoluções horárias ou menores, sendo utilizados pelo menos de um ano para uma verificação de qualidade no conjunto de teste. O conjunto de teste deve estar completamente separado das amostras de treino e validação, sendo preferencialmente a janela de validação o último segmento da amostra de treino, evitando-se dessa forma, inclusive, o uso de validação cruzada.

Em terceiro lugar, segundo Lago *et al.* (2021), devem ser ignorados trabalhos problemáticos, publicados em veículos de baixa relevância ou aqueles que fazem uso de métodos ultrapassados, dando preferência a utilização de referências que tratem de pesquisas atualmente consideradas recentes e relevantes em EPF. Estas pesquisas que representam o atual estado da arte são discutidas a seguir.

3.3.1 Estado da arte

Ademais da dificuldade de se comparar estudos em EPF, as mais recentes revisões bibliográficas consultadas (ZIEL; WERON, 2018; NOWOTARSKI; WERON, 2018; LAGO; De Ridder; De Schutter, 2018; NARAJEWSKI; ZIEL, 2020; LAGO *et al.*, 2021) concordam ao afirmar que o estado da arte atualmente está compreendido especificadamente sob duas abordagens, sendo, métodos estatísticos e métodos de DL. Métodos híbridos também são mencionados de forma complementar. A seguir é feita uma meta análise com base na seleção de trabalhos levantados por Lago *et al.* (2021) que vai de janeiro de 2016 a dezembro de 2019. Quando necessário, as mesmas palavras chave foram utilizadas para expandir a pesquisa e incluir também estudos publicados em 2020 e 2021.

3.3.1.1 Métodos estatísticos

Métodos estatísticos para previsão em séries temporais estimam uma variável dependente por meio de uma combinação linear ou não linear de variáveis explanatórias (independentes) com uso de termos que, no contexto do método, são chamados de coeficientes regressores. Em EPF, considerando que para uma hora h e um dia d existe um respectivo preço $p_{d,h}$, então um método de previsão irá estimar coeficientes $\theta_h = [\theta_{h,0}, \theta_{h,1}, \ldots, \theta_{h,n}]$ para uma determina hora h que serão associados as entradas $\mathbf{X}_{d,h} = \left[X^1_{d,h}, \ldots, X^n_{d,h}\right]^\top$ na forma da Equação 3.1, onde $\varepsilon_{d,h}$ diz respeito a um termo de erro.

$$p_{d,h} = \theta_h \mathbf{X}_{d,h} + \varepsilon_{d,h} \tag{3.1}$$

Neste contexto, métodos estatísticos podem ser univariados ou multivariados. A abordagem multivariada tem se mostrado mais precisa, pois agrega uma maior quantidade de variáveis explanatórias que, se bem escolhidas, podem refletir melhor o comportamento da variável dependente (CHRISTENSEN; HURN; LINDSAY, 2012). Outra característica a ser modelada é o tamanho do horizonte de previsão, sendo possível um (*one-step*) (CARLÀ *et al.*, 2016; PAPACHARALAMPOUS; TYRALIS; KOUTSOYIANNIS, 2018) ou mais passos (*multi-step*) à frente (Kalate Ahani; SALARI; SHADMAN, 2019).

Um dos desafios em se estimar um modelo previsor estatístico é determinar a quantidade de coeficientes. Modelos com poucos termos tendem a não modelar o comportamento do preço, enquanto que modelos com muitos tendem a se superajustar aos dados incorrendo em problema de falta generalização. Uma das técnicas mais conhecidas de estimação é o Mínimos Quadrados Ordinários (*Ordinary Least Squares* (OLS)) que minimiza a soma residual dos quadrados entre valores preditos e reais (*Residual Sum of Squares* (RSS)). A definição de termos em OLS pode ser automatizada por meio de regularização (BICKEL *et al.*, 2006). A regularização pode ser aplicada com três abordagens, sendo LASSO (TIBSHIRANI, 1996), *Elastic Net* (ZOU; HASTIE, 2005) ou *Ridge* (HASTIE, 2020). Estas técnicas automatizam a seleção de recursos num modelo multivariado por terem a capacidade de zerar coeficientes θ_h associados a entradas específicas do conjunto $\mathbf{X}_{d,h}$ (algumas dessas, muitas vezes redundantes entre si). Com a aplicação de penalidades, modelos complexos são desprezados em detrimento dos mais simples. Nesse sentido, a regularização associada a um método de regressão como OLS diminui a quantidade de parâmetros ao mesmo tempo em que minimiza o RSS. Esta é uma abordagem autorregressiva chamada de LEAR. LEAR é uma técnica que não demanda grande esforço computacional e tem sido reportada com destaque em trabalhos de EPF que optaram pela via estatística (ZIEL; STEINERT; HUSMANN, 2015; ZIEL, 2016; UNIEJEWSKI; NOWOTARSKI; WERON, 2016; UNIEJEWSKI; WERON, 2018; LAGO; De Ridder; De Schutter, 2018; ZIEL; WERON, 2018; UNIEJEWSKI; MARCJASZ; WERON, 2019; NARAJEWSKI; ZIEL, 2020). LEAR também pode ser aplicado com sucesso em EPF probabilística (MUNIAIN; ZIEL, 2020; UNIEJEWSKI; WERON, 2021).

Séries de preços de eletricidade tendem a ser heterocedásticas [15] (GUJARATI; PORTER, 2011; SANDHU; FANG; GUAN, 2016). Devido a esta característica alguns trabalhos vêm optando por incrementar seus modelos com técnicas de pré-processamento para normalização. O uso destas práticas cria melhores condições estatísticas em séries temporais de natureza altamente volátil. Uma das formas de normalização comumente utilizadas é a logtransformação, no entanto, em EPF a mesma não é sempre aplicável dado que preços negativos podem existir. Técnicas de tratamento de *outliers* são também uma alternativa (KIEU; YANG; JENSEN, 2018). Métodos estabilizadores de variância como transformação pela função inversa hiperbólica do seno (*asinh*), Boxcox, Polinomial, Log espelhado e *Probability Integral Transform* (PIT) são também empregados (SCHNEIDER, 2011; DIAZ; PLANAS, 2016; UNIEJEWSKI; WERON, 2018; ZIEL; WERON, 2018). O trabalho de Uniejewski, Weron e Ziel (2018) examinou os impactos desses métodos em EPF e concluíram que PIT com ajuste gaussiano (N-PIT) foi o método que apresentou melhores resultados, no entanto, sem significativas diferenças para os demais.

Séries de preços de eletricidade podem apresentar comportamentos sazonais (MAYO, 2012). Métodos estatísticos clássicos como os preconizados pela metodologia Box-Jenkins (BOX; JENKINS, 1970) tratam sazonalidades por meio de técnicas como *Seasonal Autoregressive Integrated Moving Average* (SARIMA). Contudo, a sazonalidade pode ser um componente de difícil detecção, atuando como um modelo subjacente. Segundo Lago *et al.* (2021) ainda não está claro se existe alguma contribuição da modelagem decomposta das séries de preços com foco na previsão desagregada da sazonalidade. Apesar disso, alguns estudos associaram a modelagem de sazonalidade a métodos autorregressivos com regularização e estabilização de variância. Todos fizeram uso de modelagens de longo prazo e concluem dizendo que houve melhora nas previsões (NOWOTARSKI; TOMCZYK; WERON, 2013; LISI; PELAGATTI, 2018; MARCJASZ; UNIEJEWSKI; WERON, 2019).

Janelas de tempo de tamanhos diferentes impactam diretamente no comportamento preditivo de métodos estatísticos. Enquanto um modelo calibrado em uma janela curta pode responder rapidamente a movimentos inéditos numa série, um modelo calibrado em janela longa pode demorar a se ajustar pois é influenciado por observações históricas de longo prazo. Neste contexto, Weron (2014) recomenda o uso de *ensembles* de preditores submetidos a janelas diversas ou mesmo preditores diferentes entre si. Em Hubicka, Marcjasz e Weron (2019) e Marcjasz, Serafin e Weron (2018) foram combinados modelos estatísticos de janelas curtas e longas. Os autores concluíram que o uso de média aritmética simples junto as previsões dos modelos treinados em janelas de tamanhos diferentes melhoraram significativamente os resultados das previsões. Em Lago *et al.* (2021) o uso de média aritmética simples entre preditores diferentes, mesmo usando janelas de mesmo tamanho, também melhorou significativamente os resultados frente a previsões feitas por métodos individualmente. Outra forma de abordar a questão é treinar os modelos em janelas passadas similares. Estas janelas podem ser identificadas por meio de algoritmos de clusterização (MARCOS *et al.*, 2020).

[15] Heterocedasticidade, em suma, trata-se da variância que varia no tempo (MORETIN; TOLOI, 2006).

Do ponto de vista da volatilidade como um termo adicional ao modelo de predição, em econometria é difundido o uso de GARCH. Esta técnica, apesar de modelar as volatilidades condicionais, não tem demonstrado sucesso em EPF (LAGO; De Ridder; De Schutter, 2018; LAGO *et al.*, 2021).

Levando-se em conta os trabalhos relatados até aqui, tanto o estudo de Lago *et al.* (2021) como outros (ZIEL; WERON, 2018; NARAJEWSKI; ZIEL, 2020) afirmam que LEAR associada a técnicas de tratamento de dados são atualmente o estado da arte em EPF no escopo dos métodos estatísticos.

3.3.1.2 Métodos de aprendizado profundo

Entre janeiro de 2016 e junho de 2021 foram publicados 35 trabalhos sobre EPF com uso de DL. O primeiro artigo desta série foi o de Wang, Zhang e Chen (2017) que propôs o uso de um *Stacked Denoising Autoencoder* (SDA) aplicado aos mercados de Nebraska, Arkansas, Louisiana, Texas e Indiana. Foram geradas previsões uma e 24 horas à frente. O modelo foi comparado com MLP, SVM, LASSO e regressão multivariada adaptativa com *splines*. Os autores relataram sucesso do método. No entanto, o trabalho considerou apenas três meses de teste. O segundo trabalho publicado foi o de Lago, De Ridder e De Schutter (2018) que avaliou quatro modelos de DL incluindo *Deep Neural Network* (DNN), LSTM, *Gated Recurrent Unit* (GRU) e CNN contra 23 *benchmarks* envolvendo redes neurais clássicas como MLP e RBF, métodos de ML como SVR e *XG Boosting* (XGB), técnicas estatísticas como *Autoregressive with eXogenous Variables* (ARX), *Fractional-order AutoRegressive Model with eXogenous Variables* (fARX), ARIMA e dois *softwares* comerciais. O estudo foi aplicado ao mercado belga compreendendo um ano completo de dezembro de 2015 a novembro de 2016. Os autores utilizaram LASSO junto a fARX, ou seja, compararam DL com métodos estatísticos. Nas conclusões são relatados importantes contribuições dando destaque para a superioridade de modelos envolvendo LSTM, GRU e DNN que venceram todos os concorrentes quando submetidas ao teste de significância de Diebold-Mariano (DIEBOLD, 2015). Este trabalho também mostrou que mesmo uma MLP otimizada pôde superar técnicas combinadas em *pipelines* multiestágio como implementações de *Self-organizing Map* (SOM) e SVR, SVR e ARIMA, ARIMA e GARCH ou outras abordagens no estado da arte de ML (para problemas não perceptivos) como o XGB.

Os demais trabalhos de 2018 tiveram foco na análise de RNNs profundas (UGURLU; OKSUZ; TAS, 2018; ZHU *et al.*, 2018), hibridizações entre CNN e LSTM (KUO; HUANG, 2018b; UGURLU *et al.*, 2018; XIE; XU; TAN, 2018; ZHANG; CHEEMA; SRINIVASAN, 2018) ou uso de DNN (CHINNATHAMBI *et al.*, 2018). Em Ugurlu, Oksuz e Tas (2018) uma arquitetura composta de células GRU em três camadas foi utilizada para realizar previsões 24 horas à frente no mercado turco. O horizonte de previsão foi de 365 dias com janela deslizante de um dia. Foram utilizadas como entradas o preço histórico, demanda prevista e temperatura. O método se mostrou superior. No entanto, foi comparado apenas a versões simples de outros métodos de DL como

LSTM e CNN isoladas com camada única, além de ARIMA. Uma combinação de CNN com LSTM foi realizada no estudo de Kuo e Huang (2018a). Apesar dos bons resultados, estes autores utilizaram um período de testes de apenas um mês. O estudo de Zhang, Cheema e Srinivasan (2018) foi conduzido no mercado alemão. O método trata de uma combinação de CNN e LSTM que foi comparada apenas contra um ARIMA e um GARCH não otimizados. Chinnathambi *et al.* (2018) aplicaram uma DNN ao mercado *Mercado Ibérico de Electricidade* (MIBEL). Apesar de construírem um modelo com 17 entradas e compararem o efeito da inclusão de mais camadas na rede, o estudo se limita a apenas um único dia de testes. Jiang e Hu (2018) aplicaram LSTMs para previsão uma hora à frente no mercado australiano e de Singapura. O método é comparado a três redes neurais MLP e um ARIMA. São usadas apenas quatro semanas de dados. Em Zhu *et al.* (2018) também é proposto um modelo baseado em LSTM. Foram utilizados dados dos mercados PJM e *New England*. O estudo mostra o desempenho do modelo contra uma árvore de decisão e uma SVM sendo testado entre outubro a dezembro de 2009.

Os estudos de 2019 mantiveram o mesmo foco de 2018. Em Zhou *et al.* (2019) foi proposta uma estrutura heterogênea que combina células LSTM com parametrização feita pelo algoritmo *Sequence Model-Based Optimization* (SMBO). O modelo prevê uma hora à frente e foi aplicado ao mercado PJM. São feitas comparações com outras cinco técnicas. No entanto, nesse estudo não é especificada nenhuma forma validação e o conjunto de testes é de apenas quatro semanas estando compreendido entre fevereiro e abril de 2018. No trabalho de Zahid *et al.* (2019) é proposto o uso de um modelo baseado em CNN. O mesmo foi comparado a XGB, árvores de decisão, *Recursive Feature Elimination* (RFE) e florestas aleatórias. O foco deste estudo é na previsão horária do preço e da carga conjuntamente. Foram utilizados dados do mercado *New England*. Os autores relatam melhoria de performance do seu modelo. Contudo, são utilizados apenas dois meses para treinamento e uma semana para testes. A combinação de uma DNN mista com LSTM foi proposta para previsão horária de preço e carga no contexto de *smart cities* no estudo de Mujeeb *et al.* (2019). Os mercados usados foram os de Nova York e *New England*. Foram avaliadas previsões de 24, 168 e 744 horas à frente. O período de testes compreendeu 3 meses. As comparações foram feitas contra três métodos de ML. Um estudo comparativo desenvolvido por Atef e Eltawil (2019) testou uma DNN contra SVR no mercado dinamarquês. Não é feito uso de conjunto de validação. Foram utilizados 90% dos dados para treino e 10% para testes. Na pesquisa de Luo e Weng (2019) foi abordado um modelo supervisionado com dois estágios aplicado ao mercado do Texas. O estudo realiza previsão horária de preços por barra no primeiro estágio e depois inclui dados de geração eólica. O modelo é comparado a uma regressão polinomial, a SVR e uma MLP treinada com *Backpropagation*. O período de testes inclui apenas janeiro de 2016. Em novo estudo de Mujeeb e Javaid (2019) foram utilizados como período de testes oito meses de dados dos mercados de *New England* e PJM. Os autores propõem o uso de duas arquiteturas de DL trabalhando simultaneamente para prever carga e preço. Modelos *Autoencoders* juntamente com LSTMs foram comparados contra ELM, *Non-linear Autoregressive with eXogenous Variables* (nARX) e variações desses otimizadas via Evolução Diferencial. As entradas foram horário, temperatura, velocidade do vento, preço histórico

e carga histórica. Os autores concluem relatando superioridade do modelo proposto. Em Chen *et al.* (2019) são utilizadas LSTMs bidirecionais tendo como entrada apenas a série histórica de preços do mercado francês. O período de teste compreende um ano indo de setembro de 2017 a setembro de 2018. Os conjuntos de treino, validação e teste são bem definidos. Os resultados são comparados com oito *benchmarks* envolvendo variações de LSTM, DNN e *naïve*. O modelo bidirecional foi o que apresentou melhor performance. Por fim, Afrasiabi *et al.* (2019) realizou uma comparação entre CNN-GRU, CNN-LSTM, CNN 2D, GRU, LSTM, ARIMA, *K-Nearest Neighbors* (KNN) e MLP aplicados no contexto de *smart grids*. O trabalho trata de previsões 24 à frente considerando a cidade de Londres e usa dados de fevereiro de 2012 a janeiro de 2013. Contudo, não foi definido qualquer conjunto de validação. A amostra foi separada com 70% para treino e 30% para testes. Foram previstas as variáveis preço, velocidade do vento, radiação solar e carga residencial. Dentre estas, preço foi a que apresentou piores valores de acurácia em termos de MAPE.

Em relação a 2020, Jahangir *et al.* (2020) propuseram um modelo híbrido envolvendo *Stacked Autoencoders* (SAEs) para redução de dimensionalidade e uma DNN para previsão de preço horário 24 horas à frente no mercado de Ontário. Os dados compreenderam períodos indo de janeiro de 2016 a dezembro de 2017. Foram consideradas 15 entradas. Sete sendo dados de geração histórica nuclear, hídrica, eólica, solar, biogás, gás natural e total. Três programações de despacho previstas para 1, 2 e 3 horas à frente. Temperatura, temperatura do orvalho e umidade e, for fim, demanda histórica agregada do país e do submercado. O treinamento e validação foram realizados em um ano e, sem recalibração, todo o ano seguinte foi utilizado com teste. Em comparações contra LSTM, duas variações de SAE e uma SVM, o modelo proposto foi o que apresentou melhores resultados com valores de MAPE variando entre 24,07% e 32,38% em quatro configurações diferentes. Hong, Taylar e Fajardo (2020b) utilizaram dados espaço-temporais no contexto de uma CNN para realizar previsões 24 horas à frente. A rede foi otimizada se utilizando um algoritmo genético e aplicada ao mercado PJM. Um diferencial deste estudo é que as análises foram realizadas em conjuntos separados para cada estação do ano e por zonas, ou seja, a previsão neste caso foi do *Local Marginal Price* (LMP). Os dados compreendem o período de dezembro de 2017 a novembro de 2018 com 20% da amostra (dois meses e meio, 14 dias em cada estação do ano) destinados para testes. Os autores relatam valores de MAPE próximos a 5% que se sobressaem aos *benchmarks* LSTM (5,8%), SVM (11,13%), KNN (10,27%), MLP (13,54%), ARIMA (8,18%) e CNN sem otimização (15,6%). Estes autores propuseram ainda uma expansão do mesmo estudo considerando apenas comparações entre diversas estruturas possíveis para a CNN (HONG; TAYLAR; FAJARDO, 2020a). Em Zhang, Li e Ma (2020) é feita uma hibridização entre arquiteturas DBN, LSTM e CNN para a realização de previsão probabilística. A conversão do preço pontual para intervalo foi feita por meio de regressão quantílica. O *pipeline* utilizou seleção de características via LASSO, normalização, tratamento de *outliers* e detecção de anomalias. Foram utilizados dados de junho de 2018 a dezembro de 2019. O conjunto foi dividido em treino, validação e testes. No entanto, o conjunto de testes compreendeu apenas os sete últimos dias de cada estação do ano. Num estudo envolvendo LSTMs bidirecionais para previsão e

Categorical Boosting (CatBoost), Zhang, Fleyeh e Bales (2020) testaram este modelo contra MLP, SVR, florestas aleatórias, ARIMA e versões isoladas de LSTM e GRU. O mercado NordPool foi o utilizado com dados do ano de 2018. Separados por estações do ano, foram treinados modelos diferentes com 80% dos dados envolvendo treinamento e 20% testes (dois meses e meio). Os autores relatam superioridade do método proposto com MAPE variando entre 5,1% e 7,74%. Em Marcjasz (2020) é proposta uma DNN com topologia otimizada via *grid-search* contra métodos LEAR, ARX e um *ensemble* dos três preditores. Dentre os estudos consultados este é o único que (ao mesmo tempo) utiliza conjuntos de treinamento, validação e teste sequenciais, extrapola um ano de avaliação usando dois anos para testes e envolve três séries de preços diferentes (PJM, NordPool, dados da GEFCom2014). As janelas históricas utilizadas foram de 364, 728 e 1092 horas para previsão 24 horas à frente. Foram aplicados pré-processamentos para estabilização de variância e regularização. A falha deste trabalho é que os resultados são relatados em termos de erros absolutos, o que inviabiliza qualquer tipo de comparação entre os mercados. No entanto, é reportado que a transformação *asinh* em conjunto com *ensemble* de DNNs foi o melhor método entre os testados.

Finalmente, dois artigos haviam sido publicados até o fechamento dos trabalhos relacionados a esta pesquisa bibliográfica. Deng, Liu e Zhu (2021) propuseram uma CNN com LSTM para previsão em resolução intra-hora de cinco minutos, com dados extraídos de medidores inteligentes vinculados a fontes renováveis em *smart grids*. O modelo também prevê picos de preços e foi aplicado ao mercado de Ontário. Os autores concluíram dizendo que o modelo proposto supera uma versão básica de CNN e um modelo ARX. Toubeau *et al.* (2021) avaliaram previsões feitas por uma LSTM bidirecional para 24 horas à frente levando-se em contas preços nodais calculados para a base *IEEE European Low Voltage Test Feeder*. Foram utilizadas como entradas a geração solar e a carga histórica, além de variáveis de calendário como hora e dia da semana. Foram feitas previsões pontuais e probabilísticas com uso de regressão quantílica. Os autores relataram que a LSTM bidirecional superou uma LSTM e uma DNN isoladas, além de GBM.

Com visto, a produção relacionada a DL em EPF tem focado fortemente na experimentação de combinações de arquiteturas diferentes, associadas ou não, a técnicas de pré-processamento, normalização, regularização e pós-processamento. Contudo, é difícil estabelecer uma comparação entre esses trabalhos devido a diferenças metodológicas como:

a) Períodos e mercados distintos.

b) Adoção de métricas distintas.

c) Metodologias de treino distintas. Algumas incluindo longas janelas de calibração, outras curtas. Algumas também considerando janelas deslizantes, outros janelas incrementais com retreinamento.

d) Horizontes de previsão distintos. Alguns trabalhos preveem uma hora à frente, outros, 24 horas à frente e, em alguns poucos casos, mais que 24 horas.

Devido a essas características, estabelecer um atual estado da arte de DL em EPF é uma tarefa complexa. Mesmo assim, alguns trabalhos tentaram estabelecer esta fronteira por meio de protocolos experimentais rígidos com uso de testes de significância estatística, variados métodos e períodos maiores que um ano. Lago, De Ridder e De Schutter (2018) fizeram um destes estudos. Os autores concluíram que uma DNN otimizada com algum algoritmo de hiperparametrização (incluindo otimização da topologia), de forma geral, pode ter desempenho tão bom quanto uma RNN com células LSTM ou GRU puras. Neste sentido, a princípio, parece que modelos baseados nestas três arquiteturas são no momento as que vem apresentando melhores resultados. Corroborando esta suposição, o estudo comparativo de Lago *et al.* (2021) aplicou quatro DNNs parametrizadas via estimador de árvore estruturada Parzen contra um método LEAR (estado da arte estatístico). Foram considerados neste trabalho também *ensembles* desses métodos. Foram usados dados de cinco mercados diferentes (NordPool, PJM, Bélgica, França e Dinamarca) e dois anos de testes. Os conjuntos de treino e validação foram separados com a janela de validação sendo o último segmento do conjunto de treino. Os resultados foram reportados por meio de testes estatísticos de diferenças e métricas de erro relativo dimensionados. Os resultados mostraram que em todos os cenários os modelos baseados em DNN foram capazes de superar o método LEAR. Curiosamente, a combinação das DNNs na forma de um *ensemble* foi também capaz de superar o melhor caso individualizado da melhor DNN. Isto sugere, portanto, que um *ensemble* de DNNs, LSTMs ou GRUs otimizadas para propósito geral, só deve ser superado por alguma arquitetura que seja:

a) Baseada em DL associada a algum mecanismo mais sofisticado ainda não testado ou pouco explorado: Como os mecanismos de atenção.

b) Parametrizada por meio de otimização: Como com uso de estimadores de árvore Parzen ou até mesmo uma busca simples como ocorre via *grid-search*.

c) Capaz de agregar previsões na forma de *ensembles*: Como médias de previsões para janelas de tamanhos diferentes ou ainda combinando preditores com características diferentes (como àqueles baseados em DL e estatísticos).

d) Projetada para funcionar especificadamente em um dado mercado incluindo entradas e tratamento personalizado de características da série temporal de aplicação: Como no mercado brasileiro.

Alguns estudos ainda sugerem etapas complementares. Não foi feita uma discussão individualizada destes por se tratarem de modelos híbridos. Neste caso, preferiu-se dispor no Quadro 2 uma sumarização das técnicas reportadas por categoria, seguidas das respectivas referências.

Quadro 2 – Relação de trabalhos que empregaram técnicas híbridas para melhoramento de qualidade preditiva em séries de preços de eletricidade.

Categoria	Técnica	Trabalhos
Tratamento da série de preços	Estabilização de variância	(UNIEJEWSKI; WERON; ZIEL, 2018).
Avaliação financeira	Impacto financeiro de erro	(ZAREIPOUR; CANIZARES; BHATTACHARYA, 2010).
Avaliação de performance	Métricas de acurácia para previsão	(HYNDMAN; KOEHLER, 2006).
Detecção de anomalias	Detecção de anomalia climática	(SOBHANI; HONG; MARTIN, 2020).
	Análise espectral singular	(VARSHNEY; SHARMA; KUMAR, 2016) e (XIAO *et al.*, 2017).
	Detecção de picos de preços	(CHRISTENSEN; HURN; LINDSAY, 2012) e (YU *et al.*, 2019b).
Seleção de entradas do modelo	Seleção de estações climáticas	(HONG; WANG; WHITE, 2015).
	Métodos de seleção de recursos	(XIE; HONG, 2018).
	Análise de variáveis exógenas	(KHATOON *et al.*, 2014).
	Análise de correlação	(HONG *et al.*, 2015), (KIM, 2015), (BENTO *et al.*, 2018), (KHAJEH *et al.*, 2018) e (BISOI; DASH; DAS, 2020) .
	Informação mútua	(ABEDINIA *et al.*, 2015), (EBRAHIMIAN *et al.*, 2018), (GAO *et al.*, 2018), (KHAJEH *et al.*, 2018), (GAO *et al.*, 2019) e (POURDARYAEI *et al.*, 2019).
	Árvores de classificação	(NAZ *et al.*, 2019).
	Relief-F	(NAZ *et al.*, 2019).
Clusterização da série de preços	*K-Means*	(ITABA; MORI, 2018) e (GHAYEKHLOO *et al.*, 2019).
	Fuzzy	(ITABA; MORI, 2017) e (GAO *et al.*, 2018).
	Teoria dos jogos	(GHAYEKHLOO *et al.*, 2019).
	Mapas auto-organizáveis	(GHOFRANI *et al.*, 2017), (NAZAR *et al.*, 2018) e (GHAYEKHLOO *et al.*, 2019).
Decomposição da série de preços	*Wavelets*	(ZHANG; TAN; LI, 2015), (OLAMAEE *et al.*, 2016), (YANG; CE; LIAN, 2017), (SINGH; MOHANTY; DEVSHUKLA, 2017), (PETER; RAGLEND, 2017), (NAZAR *et al.*, 2018), (BENTO *et al.*, 2018), (ANAMIKA; PEESAPATI; KUMAR, 2018), (GAO *et al.*, 2018) e (CHANG; ZHANG; CHEN, 2019).
	Empirical Mode Decomposition	(HONG *et al.*, 2015) e (ZHANG *et al.*, 2019).
	Variational Mode Decomposition	(LAHMIRI, 2017) e (VICTOIRE *et al.*, 2018).
	Hilbert-Huang	(KURBATSKY *et al.*, 2014).
Otimização de hiperparâmetros dos modelos	Enxame de Partículas	(YANG; CE; LIAN, 2017), (LAHMIRI, 2017), (ITABA; MORI, 2017), (ANAMIKA; PEESAPATI; KUMAR, 2018), (EBRAHIMIAN *et al.*, 2018) e (POURDARYAEI *et al.*, 2019).
	Heurísticas de busca	(KIM, 2015), (XIAO *et al.*, 2017) e (POURDARYAEI *et al.*, 2019).
	Outras metaheurísticas evolucionárias	(OLAMAEE *et al.*, 2016), (BENTO *et al.*, 2018), (VICTOIRE *et al.*, 2018) e (BISOI; DASH; DAS, 2020).

Fonte: O Autor e Lago *et al.* (2021).

3.3.2 Previsão de preços de eletricidade com mecanismos de atenção

Até aqui foram revisados trabalhos indo de uma abordagem ampla envolvendo DL e PE, até a aplicação do estado da arte unicamente ao subdomínio da EPF. Esta análise, como abordado nos quatro últimos pontos levantados no final da subseção anterior, indica que mecanismos de atenção e *ensembles* envolvendo preditores estatísticos e de DL são uma alternativa válida para um possível avanço na fronteira do conhecimento. Entretanto, levando-se em conta o período que vai de janeiro de 1994 a junho de 2021, poucas pesquisas até então se dedicaram a isso e, no caso do Brasil, nenhuma foi encontrada .

Uma busca com a chave (*"deep learning"AND "electricity price forecasting"AND "attention mechanism"*) nas bases Scopus e WoS não retornou nenhum resultado. Procedendo-se com a mesma pesquisa no Google Scholar foram obtidas 74 referências, sendo que 64,86% destas são de 2020 ou 2021. Desprezando-se trabalhos de graduação e artigos em *magazines*, passaram pelo filtro 24 trabalhos (entre dissertações, teses, artigos de conferências e de periódicos). Após análise individual de título, palavras-chave e resumo de todos esses, entre trabalhos de previsão de produção científica (FAHIM *et al.*, 2021), vendas (LI *et al.*, 2021), terremotos (BANNA *et al.*, 2021), carga (DING *et al.*, 2021), geração solar (PENG *et al.*, 2021), preço de petróleo e gás natural (NIU; XU; LIU, 2021) e outros não relacionados ao tema, três apenas fizeram aplicação em EPF.

O artigo de Fan *et al.* (2019) é uma publicação de conferência na qual foi proposto um modelo multi-horizonte de previsão probabilística sem variáveis exógenas. A arquitetura empregada foi baseada em uma Seq2Seq LSTM bidirecional com mecanismo de atenção. Para avaliação foram utilizados dois conjuntos de dados. O primeiro trata-se de uma série de vendas e, o segundo, diz respeito a série de preços de eletricidade do campeonato GEFCom 2014 (HONG *et al.*, 2016). Em relação ao *dataset* de preços de eletricidade, os autores realizaram uma divisão de três anos de dados horários em 12 subséries. Nestas foram treinados 12 modelos e as previsões foram coletadas para 99 quantis. O quantil médio das 12 sequências foi reportado em termos de *Mean Squared Error* (MSE) como resultado final. Os *benchmarks* empregados foram *naïve*, GBM, uma RNN e quatro variações do modelo proposto (duas com uso de atenção e duas sem). As conclusões do trabalho informam que todos os modelos que envolveram uso de atenção superaram os métodos concorrentes.

O trabalho de Marszałek e Burczynski (2021) foi veiculado no periódico *Open Acess Journal of Smart Environments and Green Computing*. Os autores aplicaram um modelo Seq2Seq LSTM com mecanismo de atenção associado a um algoritmo *K-Means* para prever 24 horas à frente. O modelo usa como entrada apenas o preço histórico. O *K-Means* foi usado para agrupar partes da série e o mecanismo de atenção foi treinado *cluster* a *cluster* de forma a criar um vetor de contexto específico para cada parte da série temporal. Foram utilizados dados do mercado NordPool com período de testes de um ano compreendendo janeiro a dezembro de 2019. Neste estudo os conjuntos de treino e validação foram corretamente determinados indo de janeiro de 2013 a dezembro de 2018. A parametrização foi feita via *grid-search* e os resultados foram comparados a variações do mesmo modelo sem o uso de mecanismos de atenção, sem o uso de *K-Means* ou

sem ambos. Os resultados, em termos de MAPE, indicaram que o modelo com clusterização e atenção foi o melhor (13,32%), seguido pelo que fez uso somente de atenção (13,48%) e, por último, o modelo LSTM simples (13,87%). Os autores também relataram que resultados mais precisos foram obtidos durante o outono e o inverno.

A pesquisa de Mashlakov *et al.* (2021) foi publicada na *Applied Energy* e diz respeito a previsão simultânea de geração eólica, solar, carga e preço para horizontes intradiários e 24 horas à frente. O estudo abrangeu avaliações de previsões pontuais e probabilísticas em resolução horária realizadas com quatro modelos estado da arte em DL (alguns proprietários), sendo que em dois desses, mecanismos de atenção foram empregados. Foram utilizadas 183 variáveis com séries de todos os países do continente europeu na forma de um modelo equivalente unificado. Os dados foram de janeiro de 2015 a fevereiro de 2017. Destas 183 variáveis, 59 eram relacionadas a carga, 31 a preços, 57 a geração eólica e 36 a geração solar. As séries fazem parte do *dataset* público *Open Power System Data* (OPSD, 2022). O conjunto foi dividido em 70% para treino, 15% para validação e 15% para testes. A otimização dos hiperparâmetros foi realizada por meio de um estimador de árvore Parzen e as janelas históricas de entrada no modelo foram de 168 horas. Os autores comentam que os quatro modelos foram treinados com uso de 4 GPUs NVIDIA Tesla V100-SXM2 com 32 GB de memória cada. A avaliação dos resultados foi feita com base em teste de significância estatística. Os autores ainda realizaram uma análise na qual consideraram dados apenas de mercados específicos e, também, avaliaram o impacto da inclusão de todos os mercados juntos (modelo global). Os resultados mostraram que o DeepAR (modelo projetado pela Amazon) (SALINAS *et al.*, 2020) foi superior a todos os outros, inclusive os dois que envolveram uso de atenção. Por fim, os autores concluem o artigo dizendo que quando o *dataset* europeu unificado foi utilizado, todas os modelos apresentaram melhor desempenho do que quando usados em mercados específicos.

4 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo é apresentada a base teórica necessária ao entendimento das técnicas empregadas na metodologia. O capítulo está dividido em três subseções. A primeira trata de conceitos fundamentais em análise de séries temporais, importantes não somente à compreensão do comportamento da série de preços de eletricidade em si, como também das métricas de avaliação e dos *benchmarks* adotados. Sob a ótica destes fundamentos, ainda é apresentada uma breve investigação do comportamento dos preços em quatro mercados, incluindo o do Brasil. Na segunda subseção são discutidos os conceitos matemáticos que dão suporte às arquiteturas neurais que compõe o modelo preditor desenvolvido. São apresentadas as características dos modelos profundos que fazem uso de recorrência temporal, as células de memória, os mecanismos de atenção e, por fim, os *autoencoders*. Levando-se em conta a necessidade de traduzir as previsões em potenciais lucros ou prejuízos, na última subseção é feita uma introdução aos métodos de gestão de risco *Value at Risk* e *Conditional Value at Risk*.

4.1 Modelagem de preços de eletricidade

A série temporal de preços possui características específicas que a tornam diferente de outras séries econômicas. Alta curtose, baixa assimetria e valores extremos ocasionais são alguns desses atributos (BUENO, 2011). Tratando-se especificadamente de séries de preços de eletricidade outras características ainda podem ser observadas como efeitos sazonais e linhas de tendência. Diversas técnicas podem ser aplicadas na modelagem e previsão da série de preços de eletricidade desde que a mesma seja compreendida do ponto de vista matemático. O arcabouço necessário a essa compreensão é fornecido pela teoria de Análise de Séries Temporais, discutida a seguir.

4.1.1 Definição matemática

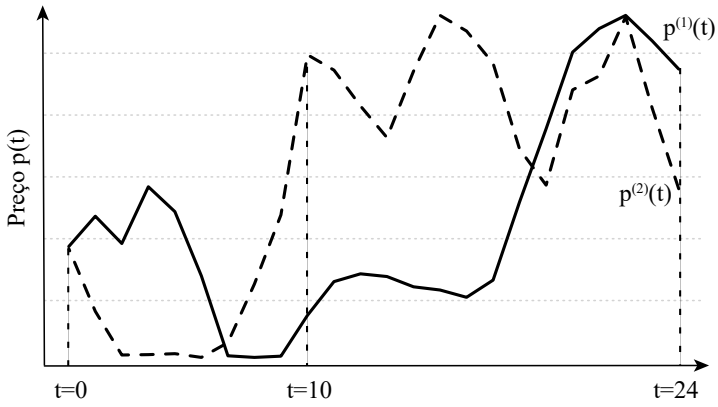
Segundo definição de Moretin e Toloi (2006) um conjunto qualquer de observações dispostas de forma ordenada no tempo é uma série temporal. Séries temporais são compostas de observações de caráter contínuo ou discreto. Séries discretas são aquelas em que um intervalo de tempo equidistante Δt é previamente estabelecido para a realização de amostragem de uma dada observação no tempo t. O intervalo em questão, quando tratado do ponto de vista de preços de eletricidade, deve ser definido segundo a granularidade máxima provida pelo processo que gera o preço. Sendo assim, no intervalo $[0,T]$ para uma série p tem-se que a discretização $N = \frac{T}{\Delta t}$ é aquela em que Δt é um valor unitário considerando o mínimo intervalo possível sendo por exemplo em mercados de eletricidade de cinco minutos, trinta minutos, uma hora ou outros.

A menos que o objetivo seja investigar o comportamento do preço em um momento específico do tempo, amostragens com Δt diferente do mínimo intervalo não irão refletir o Processo Gerador de Dados (PGD) da série p. Isto ocorre, pois o valor do preço discretizado em N não é resultado

dos valores defasados da própria série até o instante t. Neste caso, torna-se necessária a aplicação de algum método de agregação como uma média dos k preços existentes no intervalo $[t-k,t]$.

Considerando-se $p(t)$ o preço no instante t dada uma discretização $N=24$ para um $\Delta t = 1$, são tomadas como exemplo as duas séries $p^{(1)}(t)$ e $p^{(2)}(t)$ da Figura 51. Ambas as curvas são possibilidades de trajetórias do PGD para a série $p(t)$. Nesse sentido, assumindo que o PGD é regido por uma lei probabilística, então as curvas são funções amostrais de uma população P que contém todas as possíveis curvas de preços $p^{(j)}(t)$ determinadas segundo um processo estocástico. Sendo assim, para cada $p^{(j)}(t)$ existirá uma variável aleatória com uma certa distribuição de probabilidades associada. Existem situações onde a análise simultânea de diversas trajetórias de um dado PGD é possível. No entanto, em casos envolvendo variáveis econômicas (como o preço da eletricidade), normalmente apenas uma trajetória poderá ser observada (MORETIN; TOLOI, 2006).

Figura 51 – Trajetórias do processo de geração de preços $p(t)$.



Fonte: O Autor.

Considerando o parâmetro t como o tempo e $p(t)$ um conjunto de valores, então o conjunto amostral em questão trata-se de um vetor unidimensional para um total de dimensões $m = 1$, e univariado para um total de variáveis $n = 1$. Levando-se em consideração que o PGD de um processo estocástico de preço de eletricidade pode ser composto também de variáveis exógenas e que cada uma dessas pode ser multidimensional, então esta definição pode ser expandida para ser descrita em alguns casos como um conjunto $\mathbf{p}(\mathbf{t})$. Neste, a condição multidimensional é dada para $m > 1$, conforme Equação 4.1, e multivariada para $n > 1$, conforme Equação 4.2.

$$\mathbf{t} = [t_1, t_2, \ldots, t_m] \tag{4.1}$$

$$\mathbf{p}(\mathbf{t}) = [p_1(\mathbf{t}), p_2(\mathbf{t}), p_n(\mathbf{t})] \tag{4.2}$$

São possíveis interesses na modelagem de uma série temporal de preços de eletricidade:

a) Investigar o mecanismo do PGD para identificar, por exemplo, quais são os prováveis fatores de influência como temperatura, custo de combustíveis, precipitação e outros.

b) Realizar previsões de valores futuros como preços em horizontes de curto, médio ou longo prazo.

c) Descrever o comportamento da série para detectar a existência de sazonalidade, tendência, ciclos e outros fenômenos subjacentes como características de comportamento próprias de inverno e verão.

d) Procurar intervalos de periodicidade para identificar momentos de incerteza ou anomalia como mudança de perfil em feriados e dias especiais.

A descrição formal desta modelagem pode ser feita por meio de um sistema dinâmico caracterizado por um conjunto de entradas $X(t)$, uma série de saídas $Z(t)$ e uma função de transferência $v(t)$ conforme Equação 4.3. Com base nesta equação, os interesses listados acima podem ser formalizados para (a) estimação da função de transferência $v(t)$, (b) realização de previsões de $Z(t)$, (c) estudo por meio de simulações do sistema com o conjunto de entradas $X(t)$ e (d) ajustamento do modelo que relaciona $X(t)$ e $v(t)$ para controlar a saída $Z(t)$ (AGUIRRE, 2015).

$$Z(t) = \sum_{\tau=0}^{\infty} v(\tau) X(t-\tau) \tag{4.3}$$

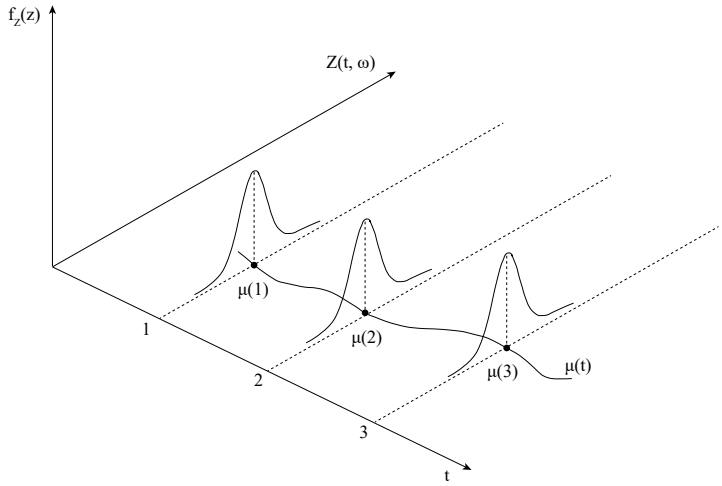
Tomando-se como base essa descrição e considerando T um conjunto factível de valores no tempo, então um processo estocástico para cada $t \in T$ diz respeito a uma família $Z = \{Z(t), t \in T\}$, tal que $Z(t)$ por definição é uma variável aleatória determinada segundo um espaço de probabilidades $\omega \in \Omega$, logo, $Z(t)$ pode ser representado segundo uma função de dois argumentos conforme apresentado na Equação 4.4 e diagramado na Figura 52.

$$Z(t,\omega) \; \forall \, t \in T, \omega \in \Omega \tag{4.4}$$

Nesta figura pode ser observado que, para cada $t \in T$, $Z(t,\omega)$ segue uma função de densidade de probabilidade (f.d.p.) com média $\mu(t)$. Nesse sentido, a trajetória do processo com base na realização de um valor a cada instante t, segundo esta f.d.p., gera uma série temporal. Todas as possíveis trajetórias calculadas segundo uma média aritmética $\mu(t)$ são o *ensemble* do PGD em $\omega \in \Omega$ para todo instante t. Portanto, em $Z = \{Z(t), t \in T\}$ se observa um espaço de estados com cada $Z(t)$ sendo um dos estados possíveis. Para um conjunto finito $T = \{1, 2, ..., N\}$ o processo é discreto, já para T no intervalo dos reais é contínuo. Como o preço da eletricidade varia em um intervalo contínuo, então para este caso o vetor $p(\mathbf{t})$ discutido anteriormente trata-se de um sistema dinâmico de processo estocástico contínuo, passível de discretização.

Sendo $p(t)$ um processo estocástico a ser descrito por um modelo regido sob uma função de transferência $v(t)$ conforme Equação 4.3, então a modelagem deste sistema deve considerar a possibilidade de existência ou não de estacionariedade. Considerando o processo da Equação 4.5 composto de distribuições finito-dimensionais para todo $n \geq 1$, então, o processo estocástico $Z = \{Z(t), t \in T\}$ será estritamente estacionário se todas as distribuições $\omega \in \Omega$ forem invariantes

Figura 52 – Interpretação de um processo estocástico gerado segundo uma família de variáveis aleatórias.



Fonte: Moretin e Toloi (2006).

no tempo com mesma média $\mu(t)$ e mesma variância $V(t)$ conforme equações 4.6 e 4.7, sendo que τ na Equação 4.6 é um deslocamento discretizado por Δt.

$$F(z_1, \dots, z_n; t_1, \dots, t_n) = P\{Z(t_1) \leq z_1, \dots, Z(t_n) \leq z_n\} \tag{4.5}$$

$$F(z_1, \dots, z_n; t_1+\tau, \dots, t_n+\tau) = F(z_1, \dots, z_n; t_1, \dots, t_n) \tag{4.6}$$

$$\mu(t) = \mu, V(t) = \sigma^2 \tag{4.7}$$

O mesmo processo será fracamente estacionário se, além das condições descritas na Equação 4.7 for também averiguado que, independentemente do valor de defasagem k (sempre múltiplo de τ) a autocovariância $\gamma(Z(t), Z(t+k))$ for constante, conforme descrito na Equação 4.8.

$$\gamma(Z(t), Z(t+k)) = Cov\{[Z(t) - \mu(t)][Z(t+k) - \mu(t)]\} \tag{4.8}$$

Séries de preços de eletricidade (e outras séries econômicas) são dependentes de condições de mercado que, de forma geral, passam a apresentar um determinado comportamento em decorrência de um evento. Este comportamento, por sua vez, tenderá a se manter até que um novo evento ocorra. O novo evento, então, fará com que o comportamento da série se modifique. Nesse sentido, mesmo que preços de eletricidade não apresentem estacionariedade, agrupamentos estacionários relacionados a eventos podem ser observados. Esta condição é chamada de processo não estacionário homogêneo (GUJARATI; PORTER, 2011).

A suposição inicial sobre uma série de preços de eletricidade é que, considerando um ruído branco ε_t conforme apresentado na Equação 4.9, existirá uma tendência determinística relacionada ao tempo t aplicada a média e variância conforme Equação 4.10 se o movimento da série for um passeio aleatório como o visto na Equação 4.11.

$$\varepsilon_t \sim i.i.d.(0,\sigma_\varepsilon^2) \tag{4.9}$$

$$E(Z_t) = t\mu_\varepsilon, Var(Z_t) = t\sigma_\varepsilon^2 \tag{4.10}$$

$$p(t) = \mu + p(t-1) + \sigma\varepsilon_t \tag{4.11}$$

Considerando que a variância determina a volatilidade da série e que, tanto a variância quanto a média mudam conforme eventos de mercado, então, preços de eletricidade estão sob condição natural de heterocedasticidade. Sendo assim, a ideia na estimação de um sistema dinâmico $Z(t)$ conforme proposto na Equação 4.3 trata da aplicação de técnicas que transformem a série temporal numa série estacionária, dessa forma, permitindo a sua modelagem. Modelos desta espécie estão divididos entre paramétricos e não-paramétricos. Modelos paramétricos possuem um número de parâmetros finitos com análise feita no domínio do tempo, enquanto que os não-paramétricos são aplicados no domínio das frequências (MORETIN; TOLOI, 2006).

Considerando-se a série temporal representada conforme Equação 4.12 onde $f(t)$ ocorre em função do tempo t e a_t diz respeito a um ruído estacionário independente de $f(t)$, o modelo temporal pode ser estimado separadamente do ruído por meio da modelagem exclusiva de $f(t)$. Neste caso, $f(t)$ pode ser um modelo que considera, por exemplo, (i) apenas o ruído conforme Equação 4.13 onde Q é constante, (ii) uma tendência linear determinística β dependente do tempo t segundo Equação 4.14, (iii) uma regressão linear conforme Equação 4.15, (iv) uma curva de crescimento não linear conforme Equação 4.16 ou (v) um polinômio conforme Equação 4.17.

$$Z_t = f(t) + a_t, \forall\, t = \{1, 2, ..., N\} \mid a_t \sim i.i.d.\, \mathcal{N}(0,\sigma_a^2) \tag{4.12}$$

$$f(t) = Q \tag{4.13}$$

$$f(t) = \alpha + \beta t \tag{4.14}$$

$$f(t) = \alpha + \beta x_t \tag{4.15}$$

$$f(t) = log(\alpha + \beta t) \tag{4.16}$$

$$f(t) = \beta_0 + \beta_1 t + \cdots + \beta_m t^m \tag{4.17}$$

Todos estes modelos, no entanto, trabalham com a hipótese de que os erros não são correlacionados. Fato incomum de se observar em uma série de preços de eletricidade, visto que o ruído associado (relacionado a ocorrência de eventos) tende a influenciar a evolução dos valores da série. Neste sentido, Box e Jenkins (1970) propuseram uma classe de modelos mais adequada a este problema denominada ARIMA. ARIMA é capaz de modelar processos lineares estacionários, não estacionários homogêneos e de memória longa (MORETIN; TOLOI, 2006).

Na Equação 4.18, a_t é um ruído branco, $\mu = E(Z_t)$ e ψ_k são os termos que descrevem a dependência temporal dos erros até k defasagens. Para uma condição de não estacionariedade se assume que a aplicação de d diferenças à série temporal é capaz deixá-la estacionária. A descrição do modelo ARIMA é feita segundo o número de parâmetros associados, sendo p relacionado à componente autorregressiva (AR); sendo q relacionado à componente de médias móveis (MA); e sendo d relacionado à quantidade de diferenças (I). A nomenclatura utilizada é, portanto, $ARIMA(p, d, q)$ e sua formulação é a apresentada na Equação 4.19 onde θ é uma constante e α a ponderação da componente AR.

$$Z_t - \mu = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots = \sum_{k=0}^{\infty} \psi_k a_{t-k},\ \psi_0 = 1 \tag{4.18}$$

$$Z(t) = \theta + \sum_{i=1}^{p} \alpha_i (Z(t-1) - \mu) + \sum_{j=1}^{q} \psi_j a_{t-1} \tag{4.19}$$

Considerando que uma série diferenciada com pelo menos um termo $I(1)$ é estacionária, então, por consequência, esta será um conjunto de retornos R_t. Retornos (por exemplo, logtransformados) são recomendados neste tipo de análise, pois na maioria dos casos tendem a apresentar além da estacionariedade, outras características benéficas a modelagem como ergodicidade [1] e normalidade [2] (MORETIN; TOLOI, 2006; BUENO, 2011; GUJARATI; PORTER, 2011).

Se $R_t \sim i.i.d.\ \mathcal{N}(0, \sigma_t^2)$, então na distribuição normal ou em qualquer outra distribuição de R_t pode-se medir a mesma em termos de assimetria e curtose, sendo que para uma amostra X_t com média amostral $\hat{\mu}$ e variância amostral $\hat{\sigma}^2$, respectivamente, assimetria e curtose podem ser quantificadas segundo as equações 4.20 e 4.21. Retornos de preços tendem a possuir distribuições leptocúrticas e baixa assimetria (MORETIN; TOLOI, 2006).

[1] Um processo temporal é considerado ergódico se, na média, a média amostral convergir, em probabilidade, para a média verdadeira do processo, portanto, neste caso, torna-se possível estimar a média e a variância de uma série temporal a partir de uma única trajetória (MORETIN; TOLOI, 2006).

[2] Quando a distribuição dos dados possui aderência a uma distribuição normal ideal.

$$\hat{A}(X) = \frac{1}{(N-1)\hat{\sigma}^3} \sum_{t=1}^{N} (X_t - \hat{\mu})^3 \tag{4.20}$$

$$\hat{K}(X) = \frac{1}{(N-1)\hat{\sigma}^4} \sum_{t=1}^{N} (X_t - \hat{\mu})^4 \tag{4.21}$$

Apesar dos retornos R_t servirem de recurso na modelagem de uma dada série Z_t, o seu uso desconsidera possíveis efeitos de tendência T_t e sazonalidade S_t, muito comuns em preços de eletricidade em decorrência, por exemplo, de crescimento econômico (para tendência) e estações do ano (para sazonalidade). Neste sentido, abordagens aditivas ou multiplicativas para decomposição da série Z_t de forma associada ao termo de ruído a_t, conforme apresentado nas equações 4.22 e 4.23, podem ser mais adequadas. É possível estimar separadamente as componentes T_t e S_t ou, ainda, modelar diretamente a série Z_t com prévia remoção das componentes T_t e S_t. Um exemplo de decomposição do PLD de 2021 pode ser visualizado na Figura 58.

$$Z_t = T_t + S_t + a_t \tag{4.22}$$

$$Z_t = T_t S_t a_t \tag{4.23}$$

A mesma representação apresentada na Equação 4.19 pode ser reescrita de forma genérica para considerar uma modelagem ARIMA que leva em conta separadamente a componente sazonal S_t conforme Equação 4.24, onde um valor de retorno é calculado. O modelo denominado $SARIMA_{[f]}(p, d, q, P, D, Q)$ exige a remoção de tendência para obtenção de estacionariedade nas duas componentes e a predefinição de uma constante f. A constante f diz respeito à frequência com que a sazonalidade é verificada. Em preços de eletricidade o valor de frequência f mais adequado é aquele que estabelece um intervalo de análise anual para observações mensais ou semanal para observações diárias/horárias (WERON, 2014).

$$\Delta Z_t = \theta + \sum_{i=1}^{p} \alpha_i \Delta_{t-i}^{d} + \sum_{j=1}^{q} \psi_j a_{t-j}^{d} + \sum_{k=1}^{P} \Lambda_k \Delta_{t-k}^{D} + \sum_{l=1}^{Q} \varphi_l a_{t-l}^{D} \tag{4.24}$$

4.1.2 Caracterização

Devido à algumas das características já discutidas no Capítulo 2 como inelasticidade da demanda, impossibilidade de modificação rápida na capacidade do sistema e necessidade de instantaneidade no consumo, o preço da eletricidade esta suscetível a grandes variações no curto prazo. Existem também casos em que essas variações poderão ser observadas no longo prazo, devido, por exemplo, a possível correlação entre preços e carga (MAYO, 2012). Sob a ótica da formalização dos conceitos apresentados na subseção anterior, uma análise mais aprofundada da

série de preços de eletricidade pode dar evidências de características comuns que, segundo diversos autores (ZAREIPOUR, 2012; WERON, 2014; SANDHU; FANG; GUAN, 2016; LAGO; De Ridder; De Schutter, 2018) são independentes de mercado. Com esse objetivo foram selecionadas 4008 observações horárias indo da primeira à última hora no período de 01/01/2021 a 16/06/2021 para preços e carga nos mercados brasileiro (subsistema SECO), PJM (média agregada de todos os submercados), Canadá (região de Ontário) e NordPool (mercado sueco). Esta análise visa complementar esta fundamentação teórica ao fornecer ao leitor uma visão empírica dos atributos específicos da série temporal de preços de eletricidade discutidos na subseção anterior.

Inicialmente foi realizada uma sumarização dos dados de preços que pode ser visualizada na Tabela 6 [3]. Nesta tabela, a primeira diferença vista na coluna de mínimos é que os mercados sueco e canadense por adotarem bolsas e leilões permitem preços negativos. Com a energia comprometida em contrato, em alguns momentos os provedores precisam liquidar suas posições. Nesse sentido, preferem pagar pelo consumo da sua produção a ficarem expostos ao preço à vista. Outro ponto a ser destacado é a distância existente entre os valores de média e mediana nos mercados canadense e PJM. Essa situação, acompanhada dos altos valores de desvio padrão e preço máximo (em relação a Brasil e Suécia) mostra como estes mercados estão mais suscetíveis a picos de preços extremos. Picos de preços extremos são resultado da forte interligação entre regiões [4] que, separadamente, praticam regras de mercado diferentes e, nesse sentindo, sofrem influência de fatores como valor de combustíveis e restrições de transmissão.

Tabela 6 – Sumarização dos preços (R$/MWh) em quatro mercados em 2021.

Mercado	Mínimo	Mediana	Média	Máximo	Desvio Padrão	Assimetria	Curtose
Nord Pool (Suécia)	-0,83	23,6	24,11	117,96	10,28	1,25	9,48
Canadá (Ontário)	-16,15	58,5	74	6833,88	136,44	32,68	1524,63
PJM (Agregado)	60,01	120,12	145,62	2744,39	107,52	8,77	144,22
Brasil (SECO)	49,77	173,61	183,7	424,56	70,82	0,16	2,26

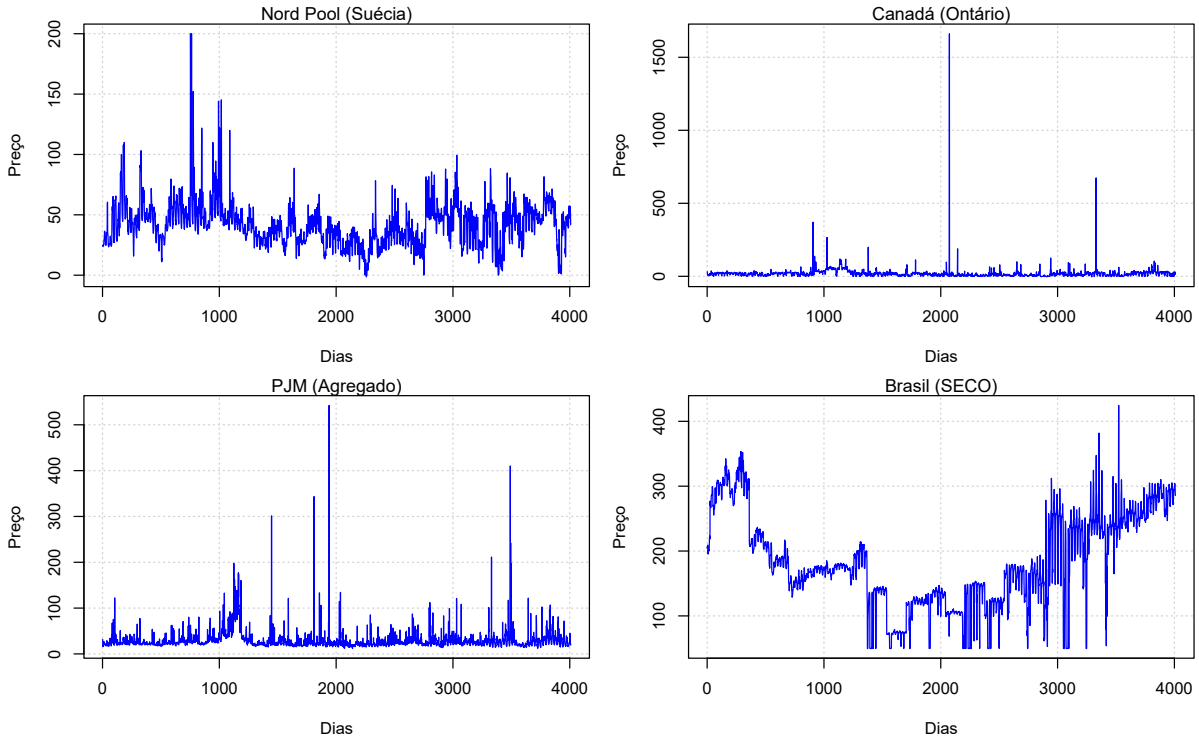
Fonte: O Autor.

[3] Para facilitar a comparação, as moedas vigentes em cada mercado tiveram seus valores convertidos para Real (R$).

[4] Os dois mercados são interligados pelo acordo *North American Electric Reliability Corporation Interconnection* (NERC) a outros grandes *grids* americanos e a região de Quebec. A região de Ontário, ao contrário do Canadá como um todo, possui apenas 22% de produção hidroelétrica, situação parecida com o PJM que tem 34% da sua capacidade instalada baseada em carvão. Isso faz com que importações e exportações sejam recorrentes para manutenção da competividade e flexibilidade operativa dos sistemas (VIANA, 2018).

Conforme observado na Figura 53, os picos de preços extremos são recorrentes nos quatro mercados. Essa característica comum faz com que a média fique deslocada do ponto de maior densidade da f.d.p., portanto, a mediana é quem deve ser preferencialmente considerada (ZAREIPOUR, 2012). Com base nas medianas descritas na Tabela 6, pode-se notar que o Brasil é o país que cobrou mais caro pelo MWh em 2021. Suécia, Ontário e PJM cobraram, respectivamente, 86,87%, 59,71% e 20,72% a menos. A análise pode ser ainda complementada pelos histogramas dos preços normalizados [5] da Figura 54. Como visto, não existe nenhuma disposição das frequências que claramente mostre alguma forma de comportamento padrão. Mesmo no mercado sueco, onde por inspeção visual pode ser observada uma f.d.p. com média mais próxima da mediana e aparente variância homogênea, ainda assim, devido a longa cauda da direita não se pode afirmar que exista aderência a uma gaussiana. Para verificar esta hipótese foi aplicado um teste de Shapiro-Wilk (MONTGOMERY, 2019) que retornou p-valores indicando rejeição de normalidade em todos os mercados. Já os valores de assimetria e curtose, também constantes na Tabela 6, corroboram a afirmação de Moretin e Toloi (2006) que diz que séries de preços, apesar de simétricas, tendem a ser leptocúrticas. A alta curtose é clara nos mercados de Ontário e PJM. Já no Brasil e Suécia a mesma não é tão pronunciada.

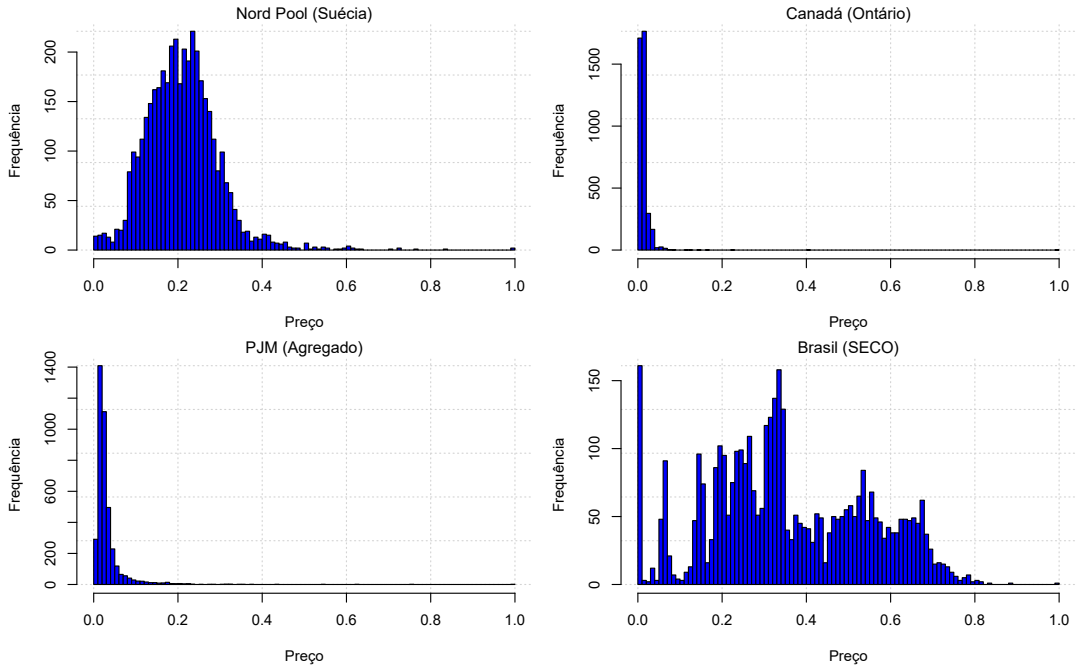
Figura 53 – Preços da eletricidade em 2021 em quatro mercados.



Fonte: O Autor.

[5] Toda normalização discutida neste capítulo diz respeito a conversão dos valores para o intervalo [0,1] conforme $y_i = (x_i - x_{min})/(x_{max} - x_{min})$. Isto é feito para facilitar a comparação de valores.

Figura 54 – Distribuição dos preços em quatro mercados em 2021.



Fonte: O Autor.

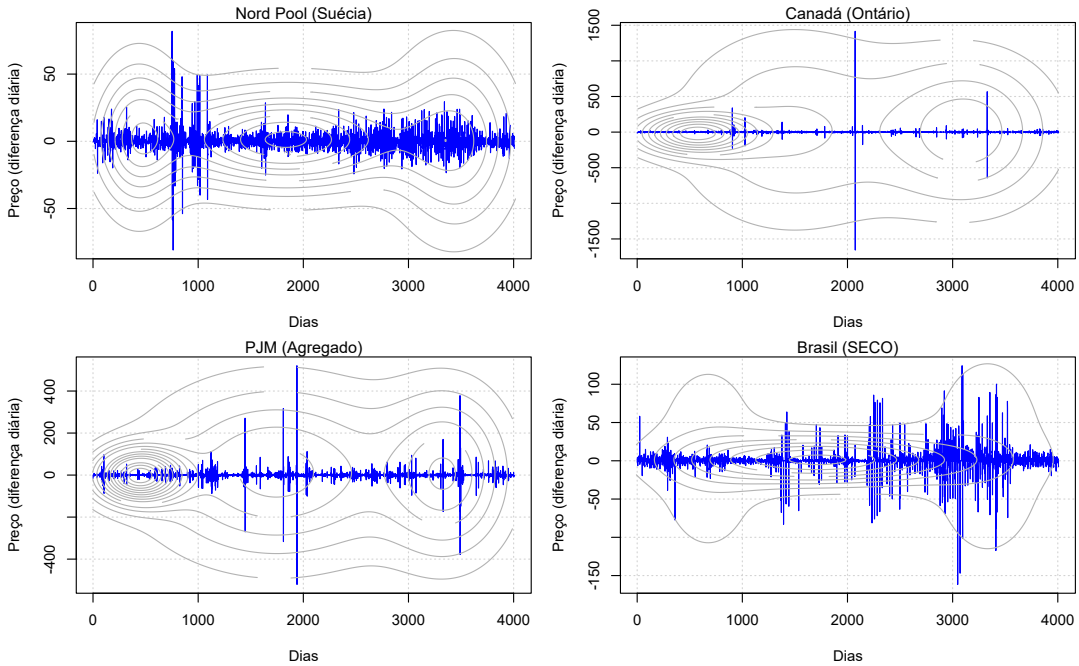
Quanto a possibilidade de mudança de comportamento dos preços ao longo do tempo foi feita uma análise por meio do teste de raiz unitária *Kwiatkowski–Phillips–Schmidt–Shin* (KPSS) (GUJARATI; PORTER, 2011). Os resultados dos p-valores foram menores que 0,05 em todos os mercados. A hipótese nula, portanto, é rejeitada indicando a presença de heterocedasticidade. Visando investigar e avaliar a relação deste provável comportamento com efeitos sazonais, são dispostos na Figura 55 os retornos agrupados por meio de estimativa de densidade [6].

Como pode-se notar, a clusterização indica a presença de grupos com densidades diferentes em todos os casos, ou seja, momentos onde o comportamento dos retornos é similar no mesmo grupo, mas diferente entre grupos. Nos mercados sueco e PJM foram identificados três agrupamentos. Já no canadense são vistos dois grupos e no brasileiro, apenas um. Apesar do mercado brasileiro possuir apenas um grupo é possível perceber uma concentração de retornos maiores no início e no final da série, fato que cria uma área de densidade estimada mais estreita entre os horários de índice 1500 e 2500. Voltando-se especificadamente para os preços do Brasil apresentados na Figura 53 ficam claros dois momentos, um de queda e outro de alta. Embora estas tendências sejam nítidas nessa figura, a volatilidade não é visualmente observável, no gráfico de retornos com o agrupamento, contudo, é mostrada essa condição.

Levando-se em consideração que a demanda de mercado é uma das variáveis que exerce maior influência no preço de um produto e que, no caso da eletricidade, essa demanda se traduz em carga prevista. Há que se considerar que a influência climática que afeta a carga, por consequência, pode afetar também o preço. Na Figura 56 são mostradas as séries horárias de carga e preços

[6] O agrupamento foi realizado por meio do algoritmo de clusterização de Fraley e Raftery (2002)

Figura 55 – Retornos dos preços agrupados em quatro mercados em 2021.

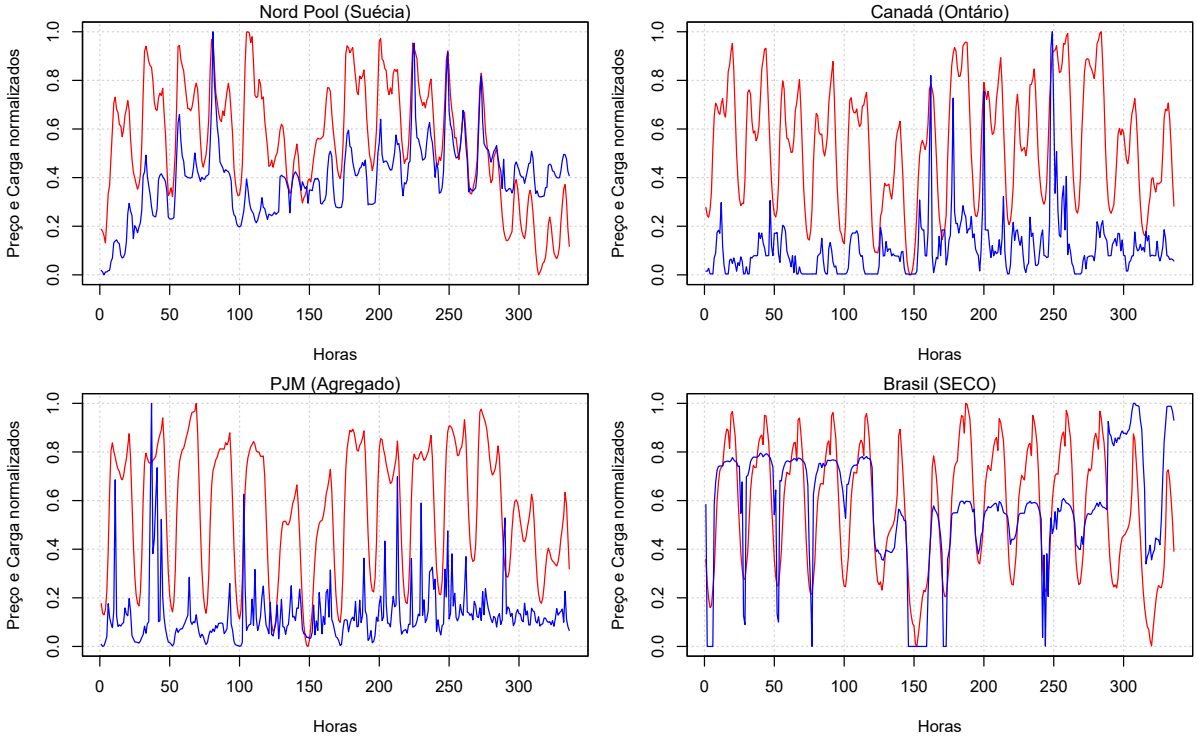


Fonte: O Autor.

no período de 05/04 a 18/04. Os valores para um teste de correlação de Pearson neste período resultaram em 0,4007 no NordPool, 0,4053 em Ontário, 0,3690 no PJM e 0,5606 no Brasil. Como visto, todos os mercados apresentam uma correlação positiva notória, especialmente o brasileiro. No entanto, ao se considerar a série completa de dados, esses valores mudam para 0,4499, 0,2973, 0,3778 e 0,1449, respectivamente. Logo, percebe-se que os preços no mercado brasileiro são os menos correlacionados com a carga, contudo, ao se considerar semanas isoladas, existem momentos em que a correlação no Brasil é a mais relevante entre todas.

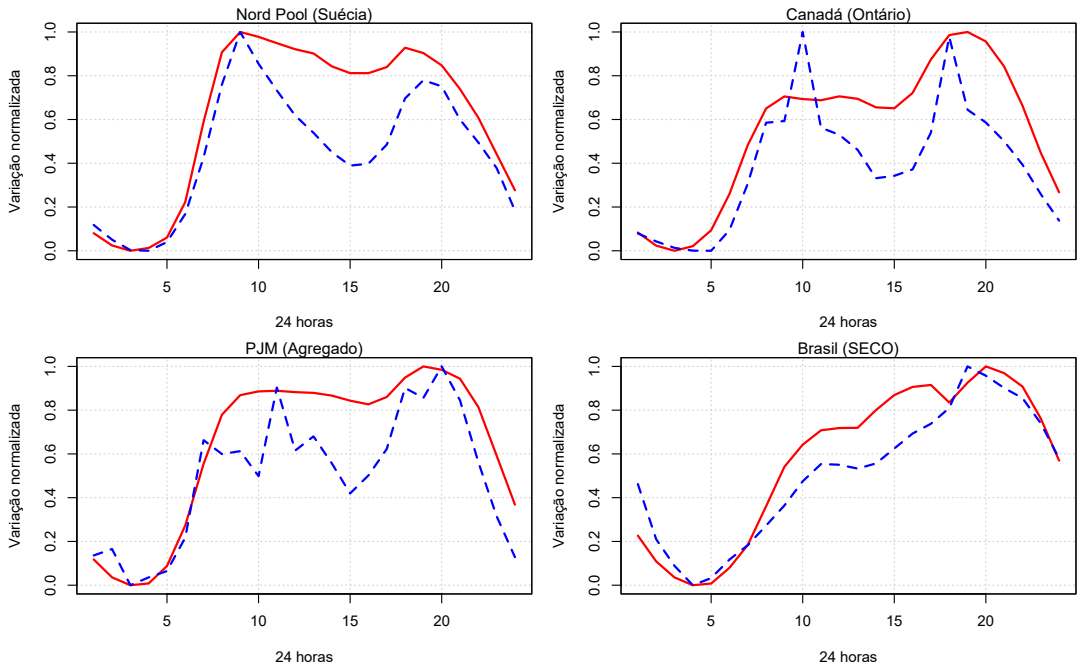
Apesar da correlação ser um indicativo de acoplamento entre as séries, esse índice pode dar uma ideia falsa do relacionamento entre duas variáveis. Dado que as séries podem ser decompostas em tendência, sazonalidade e ruído como nas equações 4.22 e 4.23, em alguns casos, é possível que apenas a componente sazonal dos preços esteja correlacionada com a carga. Com base nisso, na Figura 57 é mostrada uma comparação entre a média dos valores de carga (hora a hora) de todas as segundas-feiras no período analisado, frente ao bloco sazonal diário S_t isolado dos preços. A série S_t foi estimada via decomposição aditiva conforme Equação 4.22 e está ilustrada para o caso do Brasil na Figura 58. Pode-se notar nesta abordagem que a componente sazonal do preço (em azul) apresenta valores normalizados muito próximos à média das cargas diárias (em vermelho) em todos os mercados. A relação mais evidente é a vista no Brasil. Isto acontece, pois no Brasil é observado somente um horário de pico por volta de 19:00 horas e não dois como nos demais. Esta análise foi feita também para os outros dias da semana e os resultados são semelhantes aos da segunda-feira. Isto mostra que os preços são influenciados pela carga, especialmente em relação à sazonalidade.

Figura 56 – Preço (em azul) e carga (em vermelho) normalizados entre 05/04/2021 e 18/04/2021 para quatro mercados.



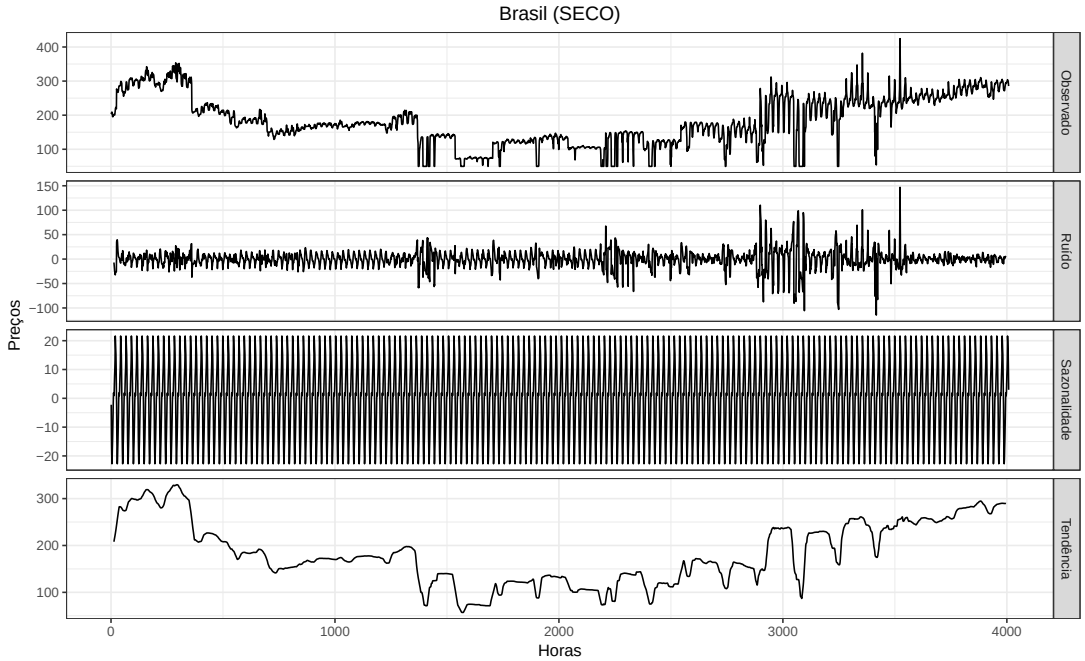
Fonte: O Autor.

Figura 57 – Análise de correlação entre a componente sazonal do preço (em azul - decomposição aditiva com frequência de 24 horas) e a carga média nas segundas-feiras (em vermelho) em 2021 para quatro mercados.



Fonte: O Autor.

Figura 58 – Decomposição aditiva com frequência de 24 horas para o PLD horário em 2021 no subsistema SECO no mercado brasileiro.



Fonte: O Autor.

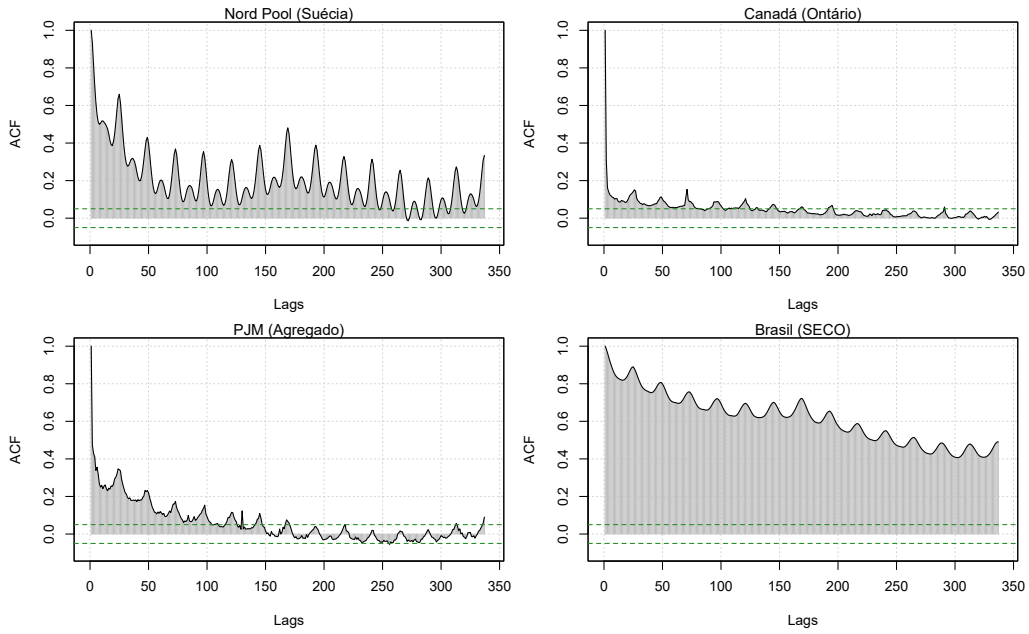
Embora a constatação de influência da carga nos preços seja importante, é necessária a averiguação também da influência de valores históricos da própria série de preços. A autocorrelação avalia a existência de autossimilaridade entre os valores de uma série e seus valores passados com base em uma defasagem escolhida. Ao se considerar, particularmente, um processo estocástico descrito conforme Equação 4.11 que apresenta autocovariância como tratado na Equação 4.8, então a autocorrelação pode ser descrita como na Equação 4.25. Nesta, k é a quantidade de defasagens e γ_t é a autocovariância.

$$\rho_{kt} = \frac{\gamma_{kt}}{\gamma_{0t}} \tag{4.25}$$

Na Figura 59 são mostrados os valores de autocorrelação indo de duas até 336 defasagens (14 dias). Nestes gráficos pode ser notada a existência de uma sazonalidade diária e, também, semanal com comportamento cíclico do preço. As linhas pontilhadas em verde informam o limiar de significância estatística estabelecido por um $\alpha = 0{,}05$. Pode-se perceber que a autocorrelação diminui ao longo dos dias em todos os mercados, deixando de ser significativa aproximadamente ao final de uma semana no PJM e em Ontário. Ao contrário, NordPool e Brasil, mesmo após duas semanas, continuam a apresentar dependências históricas estatisticamente significantes. Esta análise mostra que, nestes dois mercados especificadamente, os preços passados tendem a influenciar os preços futuros até o máximo horizonte considerado como curto prazo [7]. Este é um indicativo de que técnicas que consideram longas janelas históricas devem ser consideradas.

[7] Como estabelecido por Hong e Fan (2016).

Figura 59 – Função de autocorrelação aplicada com *lags* de duas a 366 horas para quatro mercados em 2021.



Fonte: O Autor.

Outra análise de interesse feita com base apenas nos preços é a que considera a volatilidade média. Este estudo informa o nível da variação estatística da série temporal dado um conjunto de janelas móveis. A volatilidade média pode ser calculada conforme a Equação 4.26 onde N é a quantidade de janelas, σ_j diz respeito ao desvio padrão dos retornos normalizados na janela móvel j e k é o tamanho da janela. A volatilidade (com janelas de 168 horas) para o mercado NordPool é de 0,4525, para Ontário de 0,0831, para o PJM de 0,2186 e para o Brasil de 0,4619. Neste sentido, apesar dos mercados de Ontário e PJM serem os que apresentam valores de picos mais extremos, estes foram ao mesmo tempo os com menor volatilidade média. Isso acontece, pois mesmo os valores de pico sendo altos em relação à média, a frequência com que ocorrem é baixa. Já os mercados do Brasil e NordPool, mais uma vez apresentam características semelhantes indicando presença de alta volatilidade.

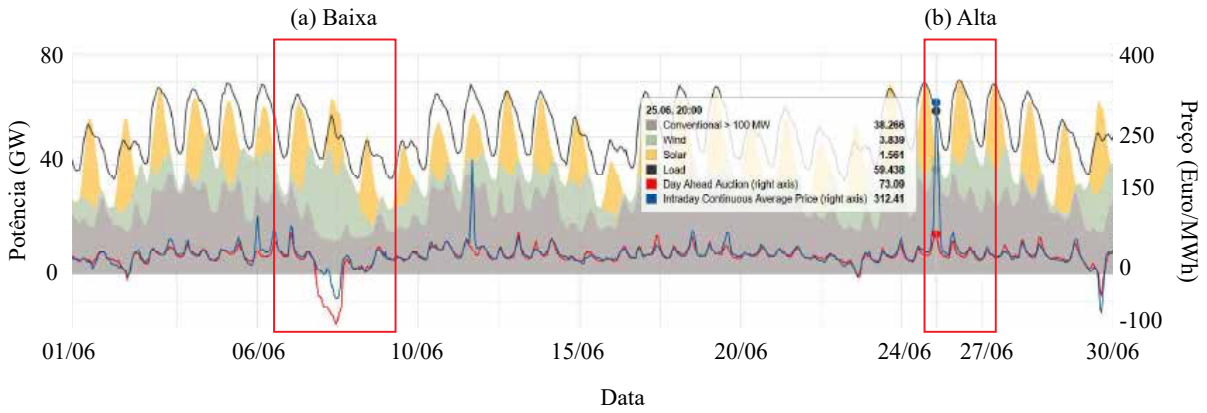
$$volatilidade = \frac{1}{N}\sum_{j=1}^{N}\hat{\sigma}_j\sqrt{k} \tag{4.26}$$

Por fim, pode-se caracterizar a série de preços de eletricidade segundo a influência das fontes de geração. Na Figura 60, adaptada do trabalho de Hong *et al.* (2020) desenvolvido para o mercado alemão, pode-se notar que a escassez ou abundância de oferta de potência pode impactar diretamente no preço. No quadro (a) é mostrado um momento onde entre os dias 06/06 e 10/06 houve um aumento da geração eólica. Esta situação conduziu à diminuição da oferta de geração convencional que, por consequência, fez com que o preço caísse para valores negativos. A situação inversa pode ser observada entre os dias 24/06 e 27/06 no quadro (b).

Este estudo realizado para o mercado alemão apresenta comportamento semelhante quando repetido para o mercado brasileiro. Na Figura 61 é mostrado no eixo da esquerda as cinco principais fontes de geração existentes no Brasil e suas respectivas produções no período analisado. No eixo da direita o preço do subsistema SECO está relacionado em preto. Tomando-se como guia as linhas pontilhadas que descrevem uma curva de regressão polinomial de grau 4 para geração hídrica (azul), térmica (cinza) e preço (vermelho), pode-se perceber que o preço possui uma tendência de movimentação inversa à oferta de geração hídrica. Já quanto a geração térmica, existe uma tendência de movimentação conjunta. Sendo o portfólio brasileiro composto de 58,94% de participação hidroelétrica e 24,58% de térmica, não é inesperada essa relação visto que a produção térmica, além de possuir custo mais alto, somente é utilizada quando a produção hidroelétrica (e de outras fontes alternativas) não são suficientes para a manutenção da operação do sistema (SILVA, 2012).

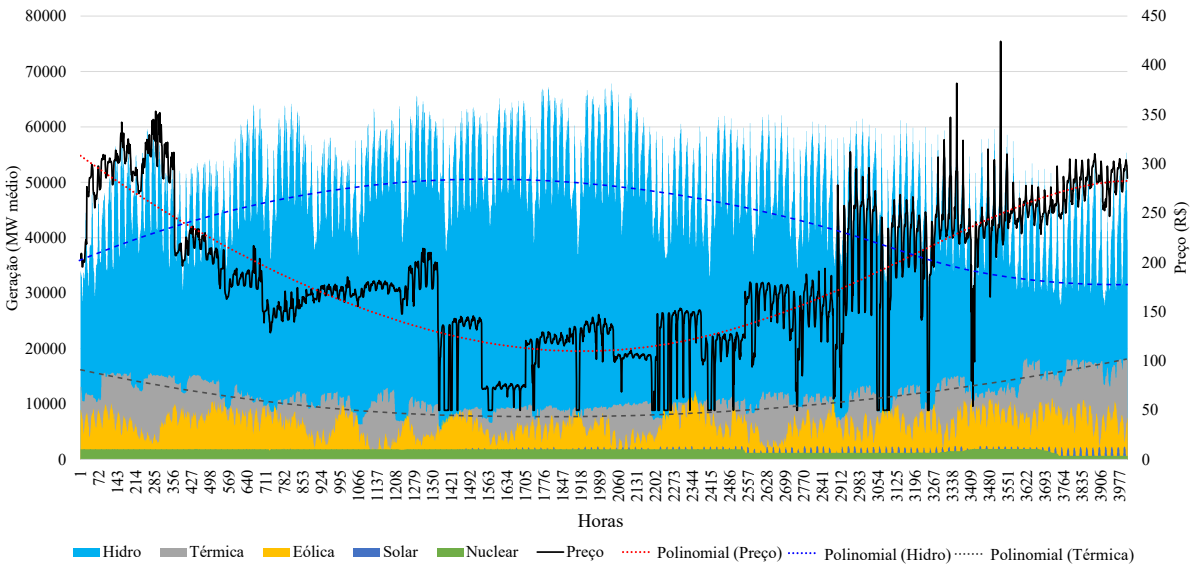
Na matriz de correlações da Figura 62 podem ser observados os índices de correlação de Pearson para cada um dos pares que relacionam preço, carga, geração hídrica, térmica, eólica, solar e nuclear. Neste gráfico só são informados os valores para os quais existe significância estatística na correlação aferida via um teste T pareado com nível de significância de 0,05. De acordo com estes valores, a relação entre preço, geração hídrica e térmica é, de fato, válida. Como pode-se perceber existe uma alta correlação positiva (0,81) entre o preço e a geração térmica. Já entre preço e geração hídrica a correlação é inversa (-0,18). Vale destacar ainda uma correlação positiva (0,87) entre carga e geração hídrica. Este último valor pode ser considerado esperado, visto que a programação do despacho é baseada na previsão da carga. Por fim, existe uma correlação inversa entre a geração hídrica e, respectivamente, eólica (-0,43) e térmica (-0,23). Fato que mostra que a modelagem da otimização do custo futuro da água feita pela cadeia de sistemas da ONS tende, realmente, a preterir uma fonte pela outra.

Figura 60 – Impacto das fontes de geração no preço no mercado alemão em julho de 2019.



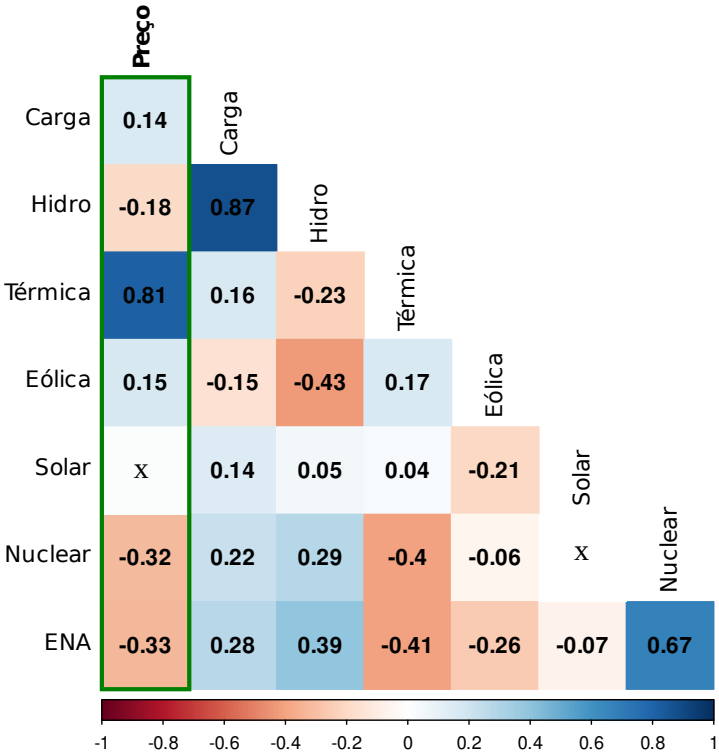
Fonte: Hong *et al.* (2020).

Figura 61 – Impacto das fontes de geração do SIN no preço do subsistema SECO no mercado brasileiro em 2021.



Fonte: O Autor.

Figura 62 – Matriz de correlações das fontes de geração vs. preço no mercado brasileiro em 2021. Correlação calculada via coeficiente de Pearson. O valor indica o índice de correlação e o "x" indica a não existência de significância estatística na correlação, conforme teste T pareado com $\alpha = 0{,}05$.

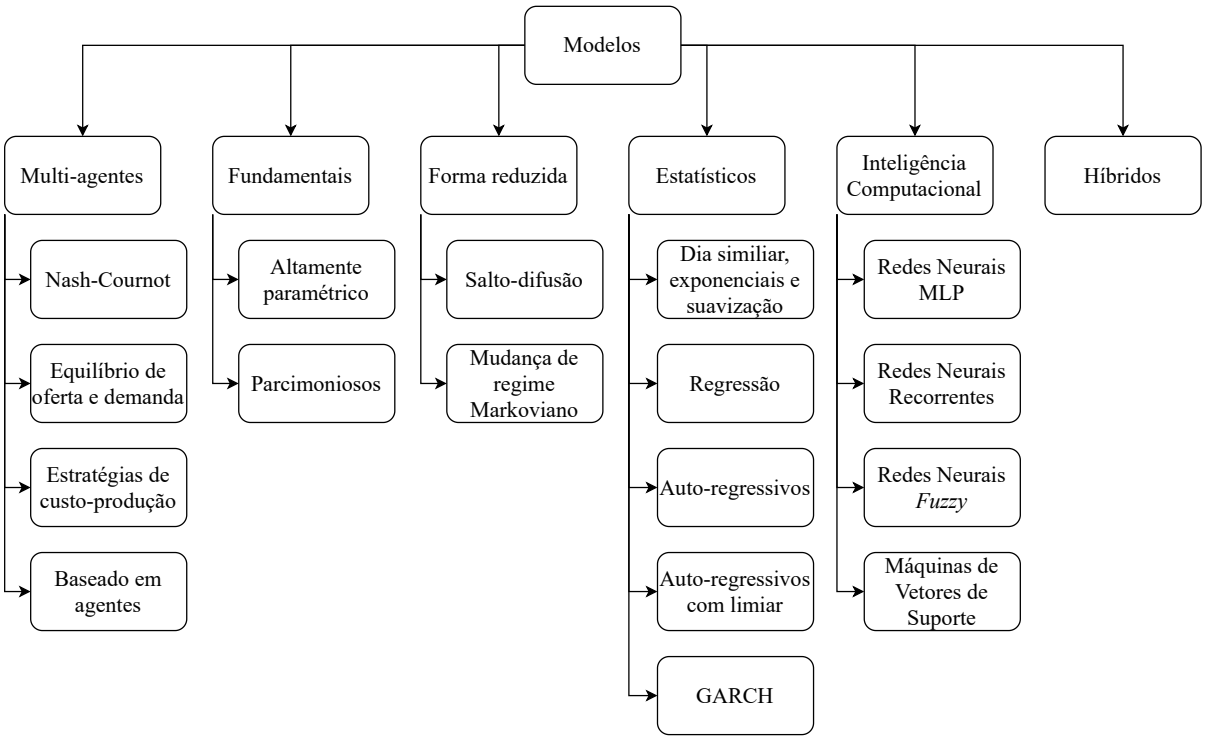


Fonte: O Autor.

4.1.3 Modelos de previsão de curto prazo

A série de preços de eletricidade, independentemente do mercado, pode ser modelada segundo diversas técnicas. O trabalho de Weron (2014) sistematizou uma taxonomia com foco no curto prazo que pode ser visualizada na Figura 63.

Figura 63 – Taxonomia das principais técnicas de modelagem de preços de eletricidade no curto prazo.



Fonte: Weron (2014).

A primeira destas classificações trata das técnicas que envolvem o uso de agentes heterogêneos com objetivo de simular o funcionamento do sistema elétrico (unidades de geração, transmissão, comercializadoras e consumidores) por meio de interações. O resultado da simulação é o preço de equilíbrio entre a oferta e a demanda. Fazem parte deste grupo técnicas que envolvem teoria dos jogos, funções de equilíbrio (como a de Nash-Cournot) e estratégias que relacionam custo e produção.

A segunda classificação diz respeito aos métodos estruturais fundamentais. Estes são aqueles que descrevem a dinâmica dos preços modelando os impactos dos fatores físicos e econômicos que interferem no preço da eletricidade. Neste grupo estão as técnicas que realizam estimação de parâmetros e modelos estruturais parcimoniosos.

O terceiro grupo de técnicas trata dos modelos de forma reduzida que tem por objetivo identificar as propriedades estatísticas dos preços para replicar suas condições futuras de comportamento com base em distribuições de dados e correlações. São normalmente utilizadas fora da

atividade de previsão, mais especificadamente para análise de risco. São técnicas desta classe os modelos de salto-difusão e os que envolvem análise de mudança de regime markoviano.

A quarta classificação trata das técnicas estatísticas para modelagem e previsão da série temporal de preços, como algumas das abordagens já discutidas na subseção 4.1.1. Trata da aplicação de fundamentos de econometria e regressão estatística para viabilizar uma interpretação física das componentes que formam o preço da eletricidade e, dessa forma, dar suporte a concepção de cenários prospectivos. Se enquadram nesta categoria os modelos de dia similar, exponenciais, suavizações, regressão, autorregressivos e de estimação de volatilidade condicional.

A quinta categoria trata das técnicas de inteligência computacional. As RNAs discutidas de forma geral no Capítulo 3 são as principais. Além de RNAs estão contidos também neste grupo os métodos *fuzzy*, SVM e outras abordagens de aprendizado de máquina que modelam e realizam previsões.

A última categoria trata das técnicas que fazem uso de dois ou mais modelos dos outros grupos e, portanto, são híbridas. Os seis grupos apresentam pontos fortes e fracos como os mostrados no Quadro 3.

A despeito da técnica selecionada para modelagem do preço, todas ao final do processo irão produzir valores futuros estimados. A qualidade da técnica, portanto, está associada ao quão cada previsão representa a realidade. Modelos de previsão, como os apresentados na Figura 63, nem sempre são capazes de se auto adaptar. Como visto na subseção 4.1.2, a série de preços de eletricidade apresenta comportamentos que variam, sendo esta não estacionária. Por este motivo, os melhores modelos são aqueles que se ajustam no tempo modificando seu comportamento conforme novas informações são fornecidas ou, de alguma forma, são recalibrados para refletir o comportamento mais recente do mercado.

A avaliação da qualidade de modelos de previsão pode ser aferida com base em métricas e testes de significância estatística. Estas podem ser usadas para auxiliar na recalibração e também identificar técnicas adequadas para o uso. Existem métricas e testes próprios para avaliação de previsão pontual ou probabilística. Para este trabalho são de interesse as que tratam de previsão pontual.

4.1.3.1 Avaliação de modelos de previsão pontual

As métricas mais utilizadas para avaliação de modelos no campo de EPF tem sido *Mean Absolute Error* (MAE), RMSE e MAPE (LAGO *et al.*, 2021). Dentre estas, MAPE está relacionada em praticamente todos os trabalhos citados no Capítulo 3. MAE, RMSE e MAPE são descritos nas equações 4.27, 4.28 e 4.29 onde $p_{d,h}$ é o preço real e $\hat{p}_{d,h}$ é o preço previsto para o dia d na hora h.

$$MAE = \frac{1}{24N_d}\sum_{d=1}^{N_d}\sum_{h=1}^{24}|p_{d,h} - \hat{p}_{d,h}| \tag{4.27}$$

Quadro 3 – Pontos fortes e fracos das técnicas de modelagem de preços de eletricidade no curto prazo.

Categoria	**Vantagens**	**Desvantagens**
Multiagentes	Flexibilidade, dado que o interessado pode incluir, remover e alterar tanto o agente quanto o comportamento que o mesmo irá desempenhar no sistema.	Complexidade na identificação de agentes relevantes e estratégias que os mesmos utilizam no mercado. Dificuldade e risco na modelagem manual dos comportamentos e interações entre os agentes.
Fundamentais	Descrevem bem condições de mercado por envolverem o uso de teoria econômica. São adequados para o médio e longo prazo.	Pouca aplicação prática no curto prazo, pois, fazem uso de dados em resolução mensal ou trimestral. Sensíveis a violação de suposições relacionadas a modelos subjacentes. Modelos muito detalhados exigem mais esforço no ajuste de parâmetros. Dependem do uso de alguma teoria econômica.
Forma reduzida	Ideais para identificar as características estatísticas do preço por meio da criação de um modelo simplificado do seu comportamento. Bons resultados na análise de risco e modelagem de preços de derivativos.	Pouca precisão na previsão do preço futuro, pois seu foco é na descrição das características e não na estimação.
Estatísticos	Baixo custo computacional. Facilidade de modelagem. Interpretabilidade. Independência de modelos econômicos.	Dependem da validação de premissas estatísticas. As séries de preços precisam ser tratadas para se tornarem estacionárias. Não apresentam bom desempenho na presença de alta volatilidade, não linearidades ou ruídos.
Inteligência computacional	Robustez a ruídos. Estado da arte para modelar séries de preços com entradas multivariadas e não lineares. Não dependem de qualquer suposição estatística ou teoria econômica, pois aprendem com os dados.	Dependem da disponibilidade de grandes quantidades de dados. Exigem parametrização. Para haver interpretabilidade do modelo, arquiteturas e técnicas auxiliares devem ser adotadas. Alto custo computacional.

Fonte: Adaptado de Weron (2014).

$$RMSE = \sqrt{\frac{1}{24N_d}\sum_{d=1}^{N_d}\sum_{h=1}^{24}(p_{d,h} - \hat{p}_{d,h})^2} \tag{4.28}$$

$$MAPE = \frac{1}{24N_d}\sum_{d=1}^{N_d}\sum_{h=1}^{24}\frac{|p_{d,h} - \hat{p}_{d,h}|}{p_{d,h}} \tag{4.29}$$

Apesar destas métricas serem as mais comuns, erros absolutos como os propostos via MAE e RMSE são difíceis de serem comparados, pois as magnitudes dos preços mudam de mercado para mercado. RMSE ainda possui um problema adicional. Dado que RMSE é uma métrica de escala não linear, a avaliação do impacto financeiro é prejudicada, visto que lucros e prejuízos são linearmente dependentes do preço absoluto. MAPE resolve alguns destes pontos por se tratar de uma métrica relativa, contudo, quando os preços estão próximos de zero o valor de MAPE tende a se tornar muito grande. Nesse contexto, MAPE deixa de ser informativo em períodos de preços baixos. Se o modelo está sendo avaliado em um mercado onde os valores são cotados em centavos, por exemplo, o uso do MAPE pode conduzir a interpretações erradas. Uma avaliação empírica destes problemas foi realizada por Hyndman e Koehler (2006) e Lago *et al.* (2021).

Uma forma de ajustar o MAPE é controlar o mesmo com base na razão do erro. Isso pode ser feito por meio do sMAPE conforme Equação 4.30. O sMAPE estabiliza os valores distribuindo os mesmos segundo uma gaussiana. O problema neste caso é que a média e a variância da gaussiana não serão iguais para diferentes mercados. Logo, o problema de escala visto junto a MAE e RMSE permanece. Portanto, sMAPE é a métrica ideal para comparar efeitos de diferentes preditores quando estes são submetidos todos a um mesmo mercado. No entanto, entre mercados diferentes outras métricas devem ser consideradas.

$$sMAPE = \frac{1}{24N_d}\sum_{d=1}^{N_d}\sum_{h=1}^{24} 2\frac{|p_{d,h} - \hat{p}_{d,h}|}{|p_{d,h}| + |\hat{p}_{d,h}|} \tag{4.30}$$

Alguns trabalhos (HYNDMAN; KOEHLER, 2006; LUO; WENG, 2019; LAGO *et al.*, 2021) propõe e fazem o uso de métricas dimensionadas voltadas para mercados diferentes. Estas levam em conta a razão do erro do modelo no conjunto de testes pelo erro médio de uma previsão ingênua [8] no conjunto de treinamento. A métrica em questão é o MASE formulado na Equação 4.31, onde p_i^{in} refere-se a previsão ingênua, n é a quantidade de observações no conjunto de treinamento e N é a quantidade de observações no conjunto de testes. Um MASE menor que um informa que o modelo tem qualidade superior à da previsão ingênua.

$$MASE = \frac{1}{N}\sum_{k=1}^{N} \frac{|p_k - \hat{p}_k|}{\frac{1}{n-1}\sum_{i=2}^{n}\left|p_i^{in} - p_{i-1}^{in}\right|} \tag{4.31}$$

Embora as métricas dimensionadas sejam livres de escala, elas ainda não são completamente adequadas visto que são calculadas com base numa janela histórica determinada pelo uso do conjunto de treinamento. Nesse sentido, o denominador do MASE (que na prática é um fator de escala) irá mudar de série para série. A situação também é válida para modelos que fazem uso de janelas deslizantes. A cada novo passo de previsão, com a entrada e retirada de valores da janela,

[8] Previsão ingênua é aquela na qual se assume que o próximo valor futuro será igual ao último valor conhecido.

o valor de escala muda. Este problema fica ainda mais claro quando são utilizadas técnicas de validação cruzada na qual vários conjuntos de treinamento diferentes existem.

Nesta lógica, as melhores métricas para avaliação de modelos em séries temporais oriundas de mercados diferentes são aquelas em que os erros sejam dimensionados, mas que ao mesmo tempo este dimensionamento seja independente de uma janela histórica fixa. O *Relative Mean Absolute Error* (rMAE) reproduz esse tipo abordagem quando aplica uma razão entre o MAE da previsão do modelo pelo MAE de uma previsão *naïve*. A diferença do rMAE para o MASE é que, ao invés de considerar o uso do conjunto de treinamento no denominador, este considera o conjunto de testes. A formulação é a apresentada na Equação 4.32, onde $p_{d,h}^{naive}$ é uma previsão *naïve*.

$$rMAE = \frac{\frac{1}{24N_d}\sum_{d=1}^{N_d}\sum_{h=1}^{24}|p_{d,h} - \hat{p}_{d,h}|}{\frac{1}{24N_d}\sum_{d=1}^{N_d}\sum_{h=1}^{24}\left|p_{d,h} - \hat{p}_{d,h}^{naive}\right|} \tag{4.32}$$

A regra associada ao *naïve* $p_{d,h}^{naive}$ deve ser previamente determinada e, para especificadamente preços de eletricidade, segundo Lago *et al.* (2021) podem ser considerados três tipos.

O primeiro, descrito na Equação 4.33, trata de considerar que o valor de preço na hora seguinte será igual ao último valor conhecido da mesma hora no dia anterior. O segundo, descrito na Equação 4.34, diz respeito a assumir que o próximo valor será igual ao último valor conhecido da mesma hora, do mesmo dia, na semana anterior. Por fim, como visto na Equação 4.35, no terceiro tipo considera-se que o próximo valor será igual ao último valor conhecido, da mesma hora, no dia anterior, se o dia corrente for terça, quarta, quinta ou sexta. Para sábado, domingo ou segunda-feira é considerado o valor da mesma hora, do mesmo dia, na semana anterior.

Neste sentido, portanto, rMAE associado a alguma regra de *naïve* é a métrica mais qualificada para avaliação de preditores quando submetidos a diferentes mercados.

$$\hat{p}_{d,h}^{naive,1} = p_{d-1,h} \tag{4.33}$$

$$\hat{p}_{d,h}^{naive,2} = p_{d-7,h} \tag{4.34}$$

$$\hat{p}_{d,h}^{naive,3} = \begin{cases} p_{d-1,h}, \, se \; d = terça, quarta, quinta \; ou \; sexta \\ p_{d-7,h}, \, se \; d = sábado, domingo \; ou \; segunda \end{cases} \tag{4.35}$$

Em relação a tendência pode-se utilizar o *Percentage of Change in Direction* (POCID). Esta métrica avalia a capacidade de um determinado modelo preditor acompanhar corretamente a mudança da direção nos movimentos da variável avaliada. Sua formulação é a apresentada na Equação 4.36 sendo que $\hat{p}_i$ diz respeito ao preço previsto e p_i ao preço real. O resultado do POCID é um valor entre 0 e 1. Quanto mais próximo de um melhor é a qualidade da previsão.

$$POCID = \frac{1}{n}\sum_{i=1}^{n} a_i \text{ em que } \begin{cases} a_i = 1, \text{ se } (\hat{p}_i - \hat{p}_{i-1})(p_i - p_{i-1}) > 0 \\ a_i = 0, \text{ se } (\hat{p}_i - \hat{p}_{i-1})(p_i - p_{i-1}) \leq 0 \end{cases} \tag{4.36}$$

Já quanto a variabilidade, o coeficiente de explicação R^2 pode ser utilizado para indicar o quanto os valores previstos estão próximos de uma linha de regressão ajustada entre as respostas observadas. O R^2 diz respeito a razão da variação explicada pela variação total, sendo portanto uma medida com resposta percentual onde 0% indica nenhuma explicação do modelo e 100% indica que o modelo explica toda a variabilidade dos dados de resposta ao redor de sua média.

Por fim, outra importante abordagem para avaliação de modelos de previsão (pouco vista nos trabalhos listados no Capítulo 3) são os testes de significância estatística que avaliam a diferença via intervalos de confiança. A teoria de Análise e Planejamento de Experimentos (MONTGOMERY, 2019) fornece ferramentas úteis que, para o caso particular de EPF são adequadas, visto que ao invés de simplesmente fornecer uma medida, informam de forma complementar se a qualidade de um modelo é ou não decorrente de aleatoriedade. Weron (2014) comenta que testes estatísticos são de suma importância no contexto de EPF, no entanto, poucos autores vem fazendo uso dos mesmos. Lago *et al.* (2021) argumenta também que esta é uma deficiência que precisa ser corrigida na área de pesquisa, pois sem rigor estatístico não é possível afirmar se um dado método é ou não superior a outros.

4.2 Modelos neurais profundos

Arquiteturas de DL focadas no tratamento de informações sequenciais compartilham de alguns dos conceitos clássicos das primeiras RNAs. RNAs clássicas podem ser modificadas para incluir recorrência de informação. Esta é uma alteração que leva ao aprimoramento da capacidade de armazenar dados históricos. Além da recorrência, o uso de células de memória permite expandir esta habilidade promovendo a recuperação de padrões, teoricamente, sem degradação no tempo. Entretanto, com o armazenamento contínuo de padrões, uma rede recorrente tende a ter a performance piorada devido ao acúmulo de informações muitas vezes inúteis. Mecanismos de atenção e *autoencoders* tem a proposta de auxiliar na mitigação deste problema. A fundamentação matemática que suporta cada uma destas técnicas é descrita a seguir conforme as obras de , Haykin (2008), Silva, Spatti e Flauzino (2016) e Goodfellow, Bengio e Courville (2016).

4.2.1 Redes neurais artificiais

A estrutura básica de uma RNA trata de unidades de processamento vinculadas umas às outras por meio de conexões ponderadas. As unidades são dispostas em camadas que mapeiam valores de entrada para valores de saída por meio de funções matemáticas. Considerando-se I a quantidade de unidades numa dada camada l e x um vetor, como visto na Equação 4.37, os valores em x são multiplicados pelos pesos $w_{ji}^{(l)}$ que denotam uma conexão da unidade i para a unidade j somados a um termo de viés $b_i^{(l)}$ na unidade i da camada l. Conforme Equação 4.38, o resultado a_h

é então submetido a uma função θ_h que irá gerar uma saída I_h. Esta função é denominada função de ativação e as unidades individualizadas que realizam o processamento são denominadas neurônios.

$$a_h = \sum_{i=1}^{I} w_{ji}^{(l)} x_i + b_i^{(l)} \tag{4.37}$$

$$I_h = \theta_h(a_h) \tag{4.38}$$

Funções de ativação comuns são as de formato sigmoidal, com destaque especial para a tangente hiperbólica e a logística sigmoide como descritas, respectivamente, nas equações 4.39 e 4.40. Não existe superioridade de uma sobre a outra, isto fica demonstrado pela transformação linear mostrada na Equação 4.41 que descreve a equivalência existente entre elas. O que muda é apenas o intervalo de saída que, para alguns casos pode ser mais adequado se ficar compreendido entre 0 e 1 ou -1 e 1. A predileção pelo uso destas funções é devido ao fato de serem diferenciáveis e não lineares. Estes dois atributos permitem o uso de algoritmos de descida do gradiente que fazem uso das derivadas descritas nas equações 4.42 e 4.43.

$$tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \tag{4.39}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{4.40}$$

$$tanh(x) = 2\sigma(2x) - 1 \tag{4.41}$$

$$\frac{\partial tanh(x)}{\partial x} = 1 - tanh(x)^2 \tag{4.42}$$

$$\frac{\partial \sigma(x)}{\partial x} = \sigma(x)(1 - \sigma(x)) \tag{4.43}$$

Com o cálculo dos valores de ativação das unidades de entrada concluídos, os mesmos são então propagados de forma a se repetir o processo para todas as camadas conectadas seguintes. A última camada pode ser composta de um ou mais neurônios, estes que gerarão uma saída y. Para fins de convenção, a camada que recebe o vetor de entrada é chamada de camada de entrada, a última camada que gera a saída y é chamada de camada de saída e as outras possivelmente existentes entre as duas são chamadas de camadas ocultas.

A saída y de uma RNA pode ser modificada conforme o problema a ser resolvido. Arquiteturas com múltiplas camadas que não fazem uso de informações históricas (apenas mapeamento de uma entrada independente para uma saída independente) são normalmente aplicadas a problemas de classificação. Nesse sentido, para uma tarefa de classificação binária que tem como função de

ativação na camada de saída uma sigmoide, pode-se determinar a probabilidade de ocorrência de uma classe z tal uma entrada x conforme uma saída y segundo a Equação 4.44. Este modelo é conhecido como *logit*. Para duas classes ou mais o modelo é conhecido como *logit* multinomial e espera uma quantidade k de neurônios na saída igual ao número de classes. Cada saída é normalizada para obtenção das probabilidades por classe. Essa função de normalização é chamada de *softmax* e pode ser vista na Equação 4.45.

$$p(z|x) = y^z(1-y)^{1-z} \tag{4.44}$$

$$p(C_k|x) = y_k = \frac{e^{a_k}}{\sum_{j=1}^{K} e^{a_j}} \; \forall \, k \in \{1,2,...,K\} \tag{4.45}$$

Levando-se em consideração o interesse em se avaliar com qual qualidade uma RNA corretamente classifica uma saída, de forma análoga, qualquer outra função pode ser escolhida. Neste sentido, o interesse está em quantificar a diferença existente entre uma saída estimada $\hat{y}$ e seu valor real y. De modo a generalizar, esta diferença será dada por uma função de perda (também chamada de função de erro ou função objetivo) que no contexto de máxima verossimilhança pode ser descrita como $\mathcal{L}(y,\hat{y})$. $\mathcal{L}(y,\hat{y})$ pode ser utilizada para calcular valores de gradientes descendentes das derivadas das funções de ativação em todas as camadas de uma RNA. Deste modo, os pesos $w_{ji}^{(l)}$ que conectam os neurônios i e j em uma dada camada l podem ser ajustados para gerar saídas $\hat{y}$ que se aproximem (ou acertem exatamente) os valores alvo em y. Este processo é denominado treinamento.

O algoritmo mais famoso de descida de gradiente para treinamento de RNAs é o *Backpropagation* (RUMELHART; HINTON; WILLIAMS, 1986). Seu uso junto a um modelo contendo pelo menos uma camada oculta dá nome a uma arquitetura conhecida como Perceptron de Múltiplas Camadas ou MLP. MLP é a base de diversas outras arquiteturas de RNAs (HAYKIN, 2008). Em uma MLP o treinamento via *Backpropagation* é feito com base na apresentação de exemplos e retropropagação dos erros calculados segundo a função de perda $\mathcal{L}(y,\hat{y})$. Trata-se basicamente da aplicação de uma regra da cadeia por repetidas vezes, conforme descrito por Haykin (2008):

1. Inicialização: Defina um conjunto de pesos iniciais $w_{ji}^{(l)}$ via algum método de inicialização. Com média zero e intervalo pré-definido, a variância para sorteio dos pesos segundo uma distribuição uniforme deve considerar as regiões onde as funções de ativação não apresentam saturação. Para métodos de inicialização de pesos, sugere-se consultar Sousa (2016).

2. Apresentação de exemplos de treinamento: Cada iteração do algoritmo é denominada época. Em cada uma destas "apresente" à RNA um conjunto de exemplos de entradas, estes denominados *batches*. Os *batches* podem ser ordenados quando a sequência das informações importa. A cada apresentação, execute a computação das saídas $\hat{y}$ com propagação dos valores à frente e, na sequência, a retropropagação dos erros baseada na função $\mathcal{L}(y,\hat{y})$. A etapa

de propagação é chamada de fase *forward* e está descrita no passo 3. A de retropropagação se chama fase *backward* e está descrita no passo 4.

3. Fase *forward*: Conforme Equação 4.46, onde $y_i^{(l-1)}(n)$ é a saída do neurônio i da camada anterior $(l-1)$, na época n, e $w_{ij}^{(l)}(n)$ é o peso do neurônio j da camada l conectada ao neurônio i da camada $(l-1)$, camada a camada, calcule o valor de ativação local $v_j^{(l)}(n)$ dado um exemplo de treinamento descrito pelo par de entradas e respectivas saídas esperadas $[\mathbf{x}(n),\mathbf{y}(n)]$. Na primeira época considere $y_0^{(l-1)}(n) = +1$ e $w_{j0}^{(l)}(n) = b_j^{(l)}(n)$. Considerando que θ_j é uma função sigmoide aplicada ao sinal de saída do neurônio j na camada l, então, o resultado será conforme a Equação 4.47 se calculado em uma camada oculta; igual a Equação 4.48 se calculado na camada $l = 0$ e; igual a Equação 4.49 se calculado na camada $l = L$.

$$v_j^{(l)}(n) = \sum_{i=0}^{m_0} w_{ji}^{(l)}(n) y_i^{(l-1)}(n) \; \forall \, l \in \{1,2,...,L\} \tag{4.46}$$

$$y_j^{(l)} = \theta_j(v_j(n)) \tag{4.47}$$

$$y_j^{(0)}(n) = x_j(n) \tag{4.48}$$

$$y_j^{(L)}(n) = \hat{y}(n) \tag{4.49}$$

4. Fase *backward*: Calcule os gradientes locais $\delta_j^{(l)}$ conforme Equação 4.50 para a camada de saída e conforme Equação 4.51 para as camadas ocultas . Finalmente, atualize os pesos em todas as camadas conforme Equação 4.52 onde η é um parâmetro denominado taxa de aprendizagem que pode acelerar ou retardar a convergência da otimização e α é a chamada constante de momento que pode ser modificada para regularizar os pesos. Este processo de atualização é chamado de regra delta.

$$\delta_j^{(l)} = \mathcal{L}(y,\hat{y})\theta_j'(v_j^{(L)}(n)) \tag{4.50}$$

$$\delta_j^{(l)} = \theta_j'(v_j^{(l)}(n))' \sum_k \delta_k^{(l+1)}(n) w_{kj}^{(l+1)}(n) \tag{4.51}$$

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha \left[w_{ji}^{(l)}(n-1) \right] + \eta \delta_j^{(l)} y_i^{(l-1)}(n) \tag{4.52}$$

5. Iteração: Execute os passos 3 e 4 continuamente, época a época, até que sejam finalizados todos os exemplos contidos nos *batches* do conjunto de treinamento. O algoritmo deve parar conforme algum critério, sendo alguns possíveis a tolerância ao erro, o máximo de épocas ou a estagnação da convergência.

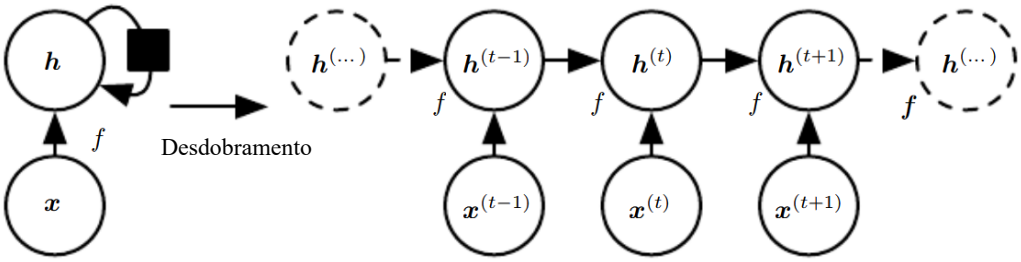
4.2.2 Arquiteturas recorrentes

A proposta de uma MLP é a de mapear a relação em pares de entradas e saídas $[\mathbf{x}(n),\mathbf{y}(n)]$. No entanto, quando valores passados $\mathbf{x}(n-k)$ tem influência no valor de saída $\mathbf{y}(n)$, a MLP deixa de ser uma arquitetura conveniente, pois não compartilha uma matriz de pesos $w_{ij}^{(l)}$ para todos os instantes possíveis (n). Algumas modificações na MLP podem ser feitas para tratar dados temporais. Uma dessas estratégias é definir uma camada de entrada que tenha um neurônio para cada observação histórica e uma camada de saída que tenha um neurônio para cada valor a ser previsto. Contudo, neste tipo de abordagem o mapeamento ainda será entre um par $[\mathbf{x}(n),\mathbf{y}(n)]$, dado que no cálculo desse mapeamento não são considerados todos valores passados de $\mathbf{x}$. Em Graves (2012) é demonstrado como MLPs são superadas por uma outra classe de RNAs projetada especialmente para modelar sequências. Essas as denominadas RNNs.

RNNs compartilham os mesmos parâmetros para todos os instantes de tempo em uma série temporal considerando todo o histórico de dados. Isto é feito, pois se houvessem parâmetros separados para cada instante, uma RNN treinada em um dado conjunto não seria capaz de generalizar o aprendizado e, por consequência, não obteria sucesso num conjunto fora da amostra (GOODFELLOW; BENGIO; COURVILLE, 2016). Outro ponto importante que diferencia uma RNN de uma MLP é que, com parâmetros compartilhados, as entradas de uma RNN não necessariamente precisam sempre corresponder a uma mesma informação. Numa MLP, por exemplo, a codificação das frases "Viajei para o Brasil em 2022" e "Em 2022 viajei para o Brasil" são tratadas como vetores independentes e a resposta desejada para "Em que ano viajei?" seria processada de forma diferente na estrutura da rede. Isso implicaria na necessidade de construir uma MLP que incorporasse todas as regras de um dado idioma para estimar a resposta correta para cada combinação de palavras. Uma situação que não ocorre numa RNN.

Considerando uma sequência que contém vetores $\mathbf{x}^{(t)}$ com índice de passo de tempo t variando de 1 a τ, uma RNN pode ser representada como um grafo computacional que, desdobrado no tempo, exibe uma quantidade de recorrências $\mathbf{h}^{(t)}$ correspondentes a uma cadeia de eventos de tamanho τ processados por uma função f. Na Figura 64 (à esquerda) pode ser visto um diagrama para esta situação sem desdobramento e (à direita) com desdobramento no tempo.

Figura 64 – Desdobramento de um grafo computacional recorrente.



Fonte: Goodfellow, Bengio e Courville (2016).

A recorrência é calculada conforme demonstrado nas equações 4.53 e 4.54. O estado $\mathbf{h}^{(t)}$ atua como um acumulador que processa passo a passo todas as entradas $\mathbf{x}^{(t)}$ segundo uma função $g^{(t)}$. O estado $\mathbf{h}^{(t-1)}$ é então associado por meio de um conjunto de parâmetros θ e uma função f à próxima entrada corrente $\mathbf{x}^{(\mathbf{t})}$ da sequência. Dessa forma é gerado um novo estado $\mathbf{h}^{(t)}$ que será passado adiante. O processo se repete até o final da cadeia.

$$h^{(t)} = g^{(t)}\left(x^{(t)}, x^{(t-1)}, x^{(t-2)}, \ldots, x^{(2)}, x^{(1)}\right) \tag{4.53}$$

$$h^{(t)} = f\left(h^{(t-1)}, x^{(t)}, \theta\right) \tag{4.54}$$

Com base nesta ideia, uma RNN pode ser projetada segundo três arquiteturas básicas.

A primeira trata da propagação do estado oculto $\mathbf{h}^{(t)}$ com geração de uma saída $\mathbf{o}^{(t)}$ para cada passo de tempo t até o último instante τ conforme visto na Figura 65. Na fase *forward* esta arquitetura processa um valor de saída $\mathbf{a}^{(\mathbf{t})}$ para cada neurônio na camada oculta realizando um somatório dos estados anteriores $\mathbf{h}^{(t-1)}$, da entrada corrente $\mathbf{x}^{(\mathbf{t})}$ e de um viés $\mathbf{b}$. Os estados e as entradas são, respectivamente, multiplicados pelas matrizes de pesos $\mathbf{W}$ e $\mathbf{U}$. Ambas são compartilhadas para todos os instantes de tempo t conforme visto na Equação 4.55. A saída $\mathbf{a}^{(\mathbf{t})}$ é então utilizada para gerar o estado corrente $\mathbf{h}^{(t)}$ conforme a função de ativação descrita na Equação 4.56. O estado corrente $\mathbf{h}^{(t)}$ também é ponderado por uma matriz compartilhada $\mathbf{V}$ que será somada a um viés $\mathbf{c}$. Como visto nas equações 4.57 e 4.58, o resultado desta operação irá gerar um valor $\mathbf{o}^{(t)}$ que será utilizado para estimar a saída $\hat{\mathbf{y}}^{(\mathbf{t})}$. Neste exemplo a saída foi estimada segundo uma função *softmax* que, ao ser comparada com a saída esperada $\mathbf{y}^{(t)}$ segundo uma função $L^{(t)}$, informa o erro da estimação. O erro, finalmente, é retropropagado na fase *backward* permitindo o treinamento da RNN por meio de um algoritmo de otimização dos pesos, como o *Backpropagation* já discutido.

$$\mathbf{a}^{(t)} = \mathbf{b} + \mathbf{W}\mathbf{h}^{(t-1)} + \mathbf{U}\mathbf{x}^{(t)} \tag{4.55}$$

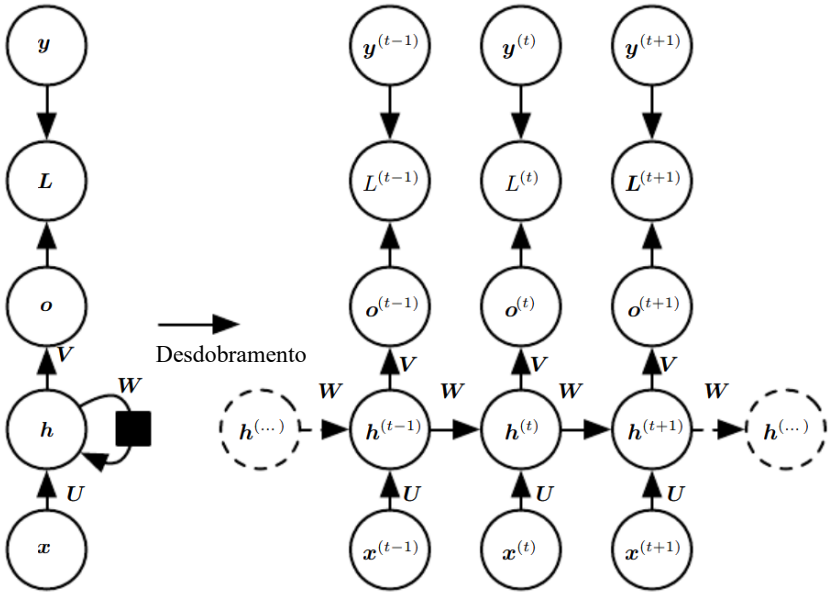
$$\mathbf{h}^{(t)} = tanh\left(\mathbf{a}^{(t)}\right) \tag{4.56}$$

$$\mathbf{o}^{(t)} = \mathbf{c} + \mathbf{V}\mathbf{h}^{(t)} \tag{4.57}$$

$$\hat{\mathbf{y}}^{(t)} = softmax(\mathbf{o}^{(t)}) \tag{4.58}$$

Um problema observado nesta arquitetura é o custo do cálculo dos gradientes via retropropagação do erro dado pela função de perda. O tempo de execução é $O(\tau)$ e não pode ser paralelizado, visto que a saída depende do último estado oculto calculado. A segunda arquitetura

Figura 65 – RNN retroalimentada pelo estado $\mathbf{h}^{(t)}$ com saída a cada passo de tempo.



Fonte: Goodfellow, Bengio e Courville (2016).

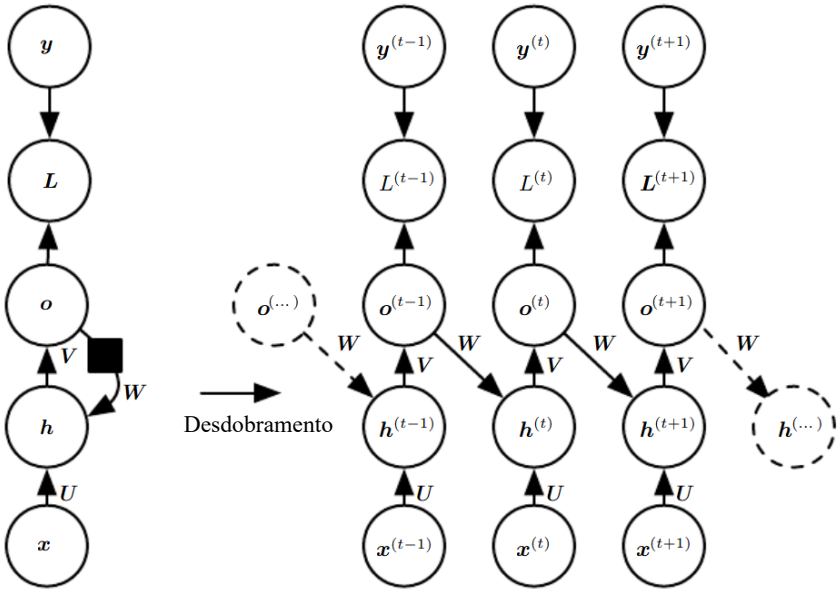
(ilustrada na Figura 66) é uma alternativa que visa corrigir esse problema. Nesta abordagem ao invés de propagar o estado oculto $\mathbf{h}^{(t)}$, o que é propagado é a saída da camada oculta $\mathbf{o}^{(t)}$. Apesar de permitir a paralelização e dessa forma diminuir o custo computacional, o uso deste procedimento afeta negativamente a capacidade preditiva da RNN. Isto ocorre, pois as unidades de saída são treinadas para se ajustar a função de perda $L^{(t)}$ e não para acumular os padrões históricos gerados pelas camadas ocultas. Outra opção seria propagar a própria saída esperada $\mathbf{y}^{(t)}$, um método chamado de aprendizado forçado. O aprendizado forçado exige que a RNN aprenda padrões históricos por meio das saídas, logo só é aplicável quando $\mathbf{y}^{(t)}$ possui alta similaridade entre os conjuntos de treino e teste. Situação que na prática é pouco observada em problemas do mundo real (GOODFELLOW; BENGIO; COURVILLE, 2016).

As arquiteturas da Figura 65 e Figura 66 exigem que a sequência de entrada tenha o mesmo tamanho da sequência de saída. Uma modificação nestas estruturas que tem por objetivo resumir todo o vetor de entradas em apenas um valor é a ilustrada na Figura 67. Neste caso, o estado oculto $\mathbf{h}^{(t)}$ também é propagado, entretanto, somente o último estado é utilizado para calcular os gradientes que servirão para ajustar os pesos no algoritmo de treinamento.

O algoritmo de treinamento *Backpropagation* pode ser portado de uma MLP para uma RNN sem modificação. O ajuste dos pesos é feito de acordo com a profundida máxima da recorrência, ou seja, o gradiente descendente é aplicado para cada passo de tempo no grafo desdobrado. Essa operação é denominada *Backpropagation Through Time* (BPTT).

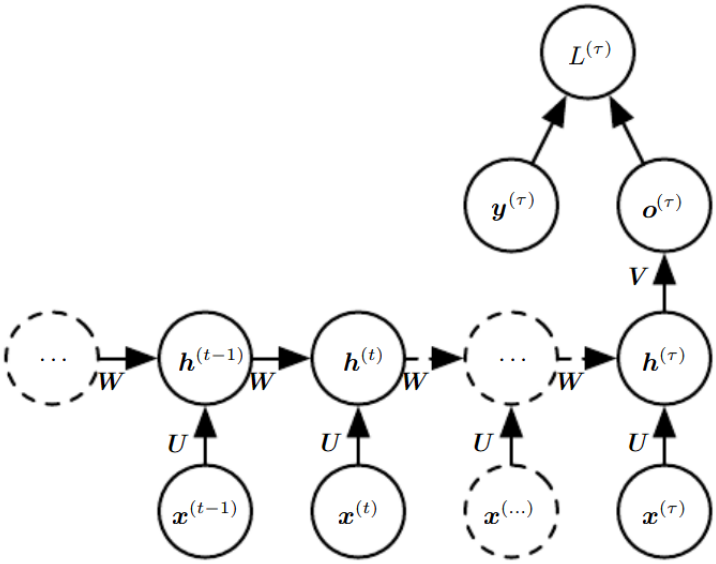
No BPTT, inicialmente é calculado o gradiente imediatamente anterior a função de perda L conforme Equação 4.59. Considerando-se o exemplo da Equação 4.55 e que a função de perda é uma *softmax* como na Equação 4.58, então o gradiente descendente $\nabla_{o^{(t)}} L$ da saída da camada

Figura 66 – RNN retroalimentada pela saída $\mathbf{o}^{(t)}$ com saída a cada passo de tempo.



Fonte: Goodfellow, Bengio e Courville (2016).

Figura 67 – RNN retroalimentada pelo estado $\mathbf{h}^{(t)}$ com saída única.



Fonte: Goodfellow, Bengio e Courville (2016).

oculta em relação ao erro é dado pela saída estimada $\hat{y}_i^{(t)}$ como exibido na Equação 4.60. Visto que na retropropagação o primeiro cálculo é feito com base na última etapa de tempo τ, então $\mathbf{h}^{(\tau)}$ é obtido por meio do gradiente descendente de $\mathbf{o}^{(\tau)}$ como na Equação 4.61. Deste ponto em diante iterações devem ser feitas indo de $t=(\tau-1)$ até $t=1$ sempre levando-se em consideração que $\mathbf{h}^{(t)}$ terá como predecessores $\mathbf{o}^{(t)}$ e $\mathbf{h}^{(\mathbf{t-1})}$ conforme Equação 4.62, onde as derivadas parciais são equivalentes a matriz diagonal $diag\left(1-\left(h^{(t+1)}\right)^2\right)$ contendo os elementos $1-\left(h_i^{(t+1)}\right)^2$. Este cálculo trata, portanto, do uso da matriz jacobiana (com função de ativação igual a tangente hiperbólica) associada com um neurônio i da camada oculta no tempo $(t+1)$ como na Equação 4.63.

$$\frac{\partial L}{\partial L^{(t)}} = 1 \tag{4.59}$$

$$(\nabla_{o^{(t)}} L)_i = \frac{\partial L}{\partial o_i^{(t)}} = \frac{\partial L}{\partial L^{(t)}} \frac{\partial L^{(t)}}{\partial o_i^{(t)}} = \hat{y}_i^{(t)} - 1_{i=y^{(t)}} \tag{4.60}$$

$$\nabla_{h^{(\tau)}} L = \mathbf{V}^\top \nabla_{o^{(\tau)}} L \tag{4.61}$$

$$\nabla_{\mathbf{h}^{(t)}} L = \left(\frac{\partial \mathbf{h}^{(t+1)}}{\partial \mathbf{h}^{(t)}} \right)^\top (\nabla_{\mathbf{h}^{(t+1)}} L) + \left(\frac{\partial \mathbf{o}^{(t)}}{\partial \mathbf{h}^{(t)}} \right)^\top (\nabla_{\mathbf{o}^{(t)}} L) \tag{4.62}$$

$$\mathbf{W}^\top diag\left(1 - \left(\mathbf{h}^{(t+1)}\right)^2\right)(\nabla_{\mathbf{h}^{(t+1)}} L) + \mathbf{V}^\top (\nabla_{\mathbf{o}^{(t)}} L) \tag{4.63}$$

Calculados os gradientes internos $L^{(t)}$, $\mathbf{o}^{(t)}$ e $\mathbf{h}^{(t)}$ de cada um dos nós no grafo desdobrado, o próximo passo é calcular os gradientes relacionados aos parâmetros compartilhados $\mathbf{c}$, $\mathbf{b}$, $\mathbf{V}$, $\mathbf{W}$ e $\mathbf{U}$. A operação segue a mesma dinâmica e, respectivamente para cada um desses, está descrita nas equações 4.64, 4.65, 4.66, 4.67 e 4.68. Em relação a $\mathbf{x}^{(t)}$ não é necessário o cálculo de nenhum gradiente, visto que para a sequência de entrada não existem termos predecessores no grafo desdobrado.

$$\nabla_c L = \sum_t \left(\frac{\partial \mathbf{o}^{(t)}}{\partial \mathbf{c}} \right)^\top \nabla_{\mathbf{o}^{(t)}} L = \sum_t \nabla_{\mathbf{o}^{(t)}} L \tag{4.64}$$

$$\nabla_b L = \sum_t \left(\frac{\partial \mathbf{h}^{(t)}}{\partial \mathbf{b}^{(t)}} \right)^\top \nabla_{\mathbf{h}^{(t)}} L = \sum_t diag\left(1 - \left(\mathbf{h}^{(t)}\right)^2\right) \nabla_{\mathbf{h}^{(t)}} L \tag{4.65}$$

$$\nabla_{\mathbf{V}} L = \sum_t \sum_i \left(\frac{\partial L}{\partial o_i^{(t)}} \right) \nabla_{\mathbf{V}^{(t)}} o_i^{(t)} = \sum_t (\nabla_{\mathbf{o}^{(t)}} L) \mathbf{h}^{(t)^\top} \tag{4.66}$$

$$\nabla_{\mathbf{W}} L = \sum_t \sum_i \left(\frac{\partial L}{\partial h_i^{(t)}} \right) \nabla_{\mathbf{W}^{(t)}} h_i^{(t)} = \sum_t diag\left(1 - \left(\mathbf{h}^{(t)}\right)^2\right) (\nabla_{\mathbf{h}^{(t)}} L) \mathbf{h}^{(t-1)^\top} \tag{4.67}$$

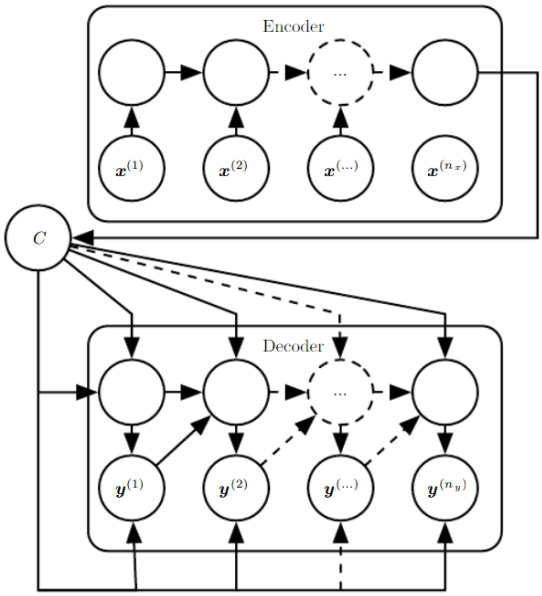
$$\nabla_{\mathbf{U}} L = \sum_t \sum_i \left(\frac{\partial L}{\partial h_i^{(t)}} \right) \nabla_{\mathbf{U}^{(t)}} h_i^{(t)} = \sum_t diag\left(1 - \left(\mathbf{h}^{(t)}\right)^2\right) (\nabla_{\mathbf{h}^{(t)}} L) \mathbf{x}^{(t)^\top} \tag{4.68}$$

As arquiteturas mostradas até aqui estão restritas a duas abordagens, sendo (i) utilizar um tamanho de sequência fixo e idêntico tanto para a entrada quanto para a saída (Figura 65 e Figura 66) e (ii) utilizar um tamanho de sequência fixo de entrada para se gerar um valor único na saída (Figura 67). No entanto, existem situações onde pode ser necessário mapear sequências

de comprimentos diferentes. A tradução da frase em português "Você fala inglês?" para o idioma Hindi (falado na Índia) é "*Kya aap angrejee bolate hain?*". Este é um bom exemplo, pois fica claro que será necessário introduzir no modelo uma sentença com três palavras e convertê-la para uma outra com cinco. A mesma situação acontece quando se deseja utilizar uma janela histórica com comprimento de um mês como base para previsão uma semana à frente. A arquitetura que provê esta maleabilidade foi independentemente introduzida por dois trabalhos, sendo chamada de *Encoder-Decoder* por Cho *et al.* (2014) e Seq2Seq por Sutskever, Vinyals e Le (2014). Neste documento a mesma é referenciada como Seq2Seq.

Uma Seq2Seq trata do uso acoplado de duas RNNs e está ilustrada na Figura 68. A primeira RNN denominada codificadora, lê e processa uma sequência de entrada $\mathbf{x}^{(t)}$ gerando um vetor de saída de comprimento fixo denominado vetor de contexto C. O vetor de contexto tem o papel de cifrar a sequência de entrada (com representação muitas vezes não numérica, como no caso da tradução de texto) para um vetor numérico intermediário que armazenará o significado latente de $\mathbf{x}^{(t)}$ num dado instante de tempo t. Além do vetor de contexto C, o último estado oculto $\mathbf{h}_{\mathbf{enc}}^{(\tau)}$ da codificadora também é armazenado. Na sequência, uma segunda RNN denominada decodificadora, terá seu estado oculto inicial setado com o estado oculto final da codificadora, ou seja, $\mathbf{h}_{dec}^{(0)} \leftarrow \mathbf{h}_{enc}^{(\tau)}$. Então, o vetor C irá ser convertido para sequência de saída $\mathbf{y}^{(t)}$. As duas RNNs são treinadas em conjunto de modo que seja maximizada a probabilidade da codificadora gerar um vetor de contexto representativo da sequência de entrada e, ao mesmo tempo, da decodificadora "entender" o vetor de contexto para convertê-lo apropriadamente para uma sequência de saída esperada.

Figura 68 – Arquitetura Seq2Seq ou *Encoder-Decoder*.



Fonte: Goodfellow, Bengio e Courville (2016).

Embora a arquitetura Seq2Seq faça parte do atual estado da arte no mapeamento de sequências longas (GOODFELLOW; BENGIO; COURVILLE, 2016) esta apresenta ainda três limitações:

a) Acúmulo de padrões históricos: Teoricamente uma Seq2Seq pode armazenar todos os padrões apresentados durante a fase de treinamento. Entretanto, o acúmulo desses padrões pode criar combinações não desejadas no momento da previsão.

b) Desaparecimento de gradientes: Treinar uma Seq2Seq utilizando BPTT é possível apenas em pequenas topologias e para sequências curtas. Dado que sequências longas exigem automaticamente o uso de várias recorrências, então o processo de treinamento com BPTT se torna impraticável, pois em redes profundas os gradientes explodem ou desaparecem após sucessivas multiplicações.

c) Perda de informação: O vetor de contexto é limitado no armazenamento de informação, pois não é seu papel guardar dados e sim representá-los.

Essas três limitações são resolvidas na Seq2Seq com a implementação dos mecanismos de atenção, células de memória e *autoencoders*. Aprimoramentos explicados a seguir.

4.2.3 Mecanismos de atenção

O principal objetivo de um mecanismo de atenção é o de introduzir na Seq2Seq uma estrutura interna que pondere a relação entre a saída da codificadora e a saída da decodificadora de forma a permitir o aprendizado de parâmetros que informem a qualidade do vetor de contexto, a cada passo de tempo (BAHDANAU; CHO; BENGIO, 2015). Na prática, um mecanismo de atenção age como um "detector de correlações". Com o detector treinado, padrões distantes na rede podem ser recuperados. A ideia básica está ilustrada na Figura 69.

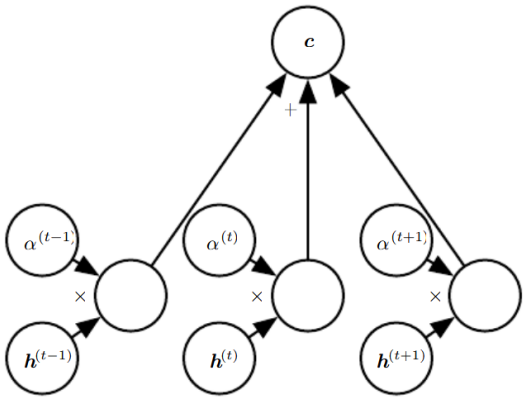
Nesta imagem, o vetor de contexto c é resultado de uma média ponderada entre os estados $\mathbf{h}^{(t)}$ e um vetor de pesos definido por $\alpha^{(t)}$ chamado de vetor de *scores* ou coeficientes de atenção. O vetor $\alpha^{(t)}$ é gerado segundo uma função *softmax* interna que irá informar qual importância possui a saída de um dado neurônio em uma dada camada oculta da codificadora. Se durante o processo de treinamento o valor de $\alpha^{(t)}$ associado a um estado $\mathbf{h}^{(t)}$ for definido como próximo de 1, então o estado oculto será propagado da codificadora para a decodificadora. Do contrário, quanto mais próximo de zero, menos o sinal será aproveitado.

Como descrito nas equações 4.69 e 4.70, a forma mais simples de se gerar um vetor de contexto atencional $\mathbf{c}$ é por meio da chamada atenção básica. Neste método os coeficientes de atenção α_i são calculados segundo uma função *softmax* que irá determinar os *scores* levando-se em conta uma função de similaridade ξ. Esta função relaciona cada vetor de entrada $\mathbf{v}_i$ com sua representação $\mathbf{u}$, ambos previamente definidos (LUONG; PHAM; MANNING, 2015).

$$\alpha_i = softmax(\xi(\mathbf{u}, \mathbf{v}_i)) \tag{4.69}$$

$$\mathbf{c} = \sum_i \alpha_i \mathbf{v}_i \tag{4.70}$$

Figura 69 – Geração de um vetor de contexto com base em um mecanismo de atenção.



Fonte: Goodfellow, Bengio e Courville (2016).

A atenção básica trata do uso de um vetor de coeficientes de atenção α_i unidimensional, contudo, essa operação pode ser generalizada visando o aumento de dimensionalidade. Uma estrutura bidimensional irá caracterizar uma matriz de atenção. Já para dimensões acima de três, as estruturas tratarão de tensores de atenção. A vantagem deste procedimento frente a atenção básica é que aspectos múltiplos que relacionam o vetor de entrada e sua representação podem ser focados separadamente pelo mecanismo de atenção. Na Equação 4.71 pode ser visualizada uma descrição do vetor de contexto calculado via atenção multidimensional. Neste exemplo, cada matriz W_n é aplicada à função de similaridade separadamente de forma a transpor o resultado de um espaço vetorial para outro maior. Cada função de similaridade será, portanto, única. A entrada da *softmax* irá receber a concatenação da saída de todas as funções gerando dessa forma o vetor $\mathbf{c}$.

$$\mathbf{c} = softmax(\xi(\mathbf{u},\mathbf{v}_i,W_1),\xi(\mathbf{u},\mathbf{v}_i,W_2),...,\xi(\mathbf{u},\mathbf{v}_i,W_n)) \tag{4.71}$$

Outra abordagem é a chamada atenção múltipla. Nesta forma, como pode ser observado nas equações 4.72 e 4.73, ao invés de um único vetor de contexto são utilizados vários. Neste caso, a principal vantagem é que um vetor de contexto $\mathbf{c}_j$ poderá ter sua correlação quantificada em relação a todos os outros vetores de contexto. Sendo assim, a sequência de saída $\mathbf{u}_j$ terá sua similaridade calculada frente a todas as entradas $\mathbf{v}_i$, que quando submetidas a *softmax* irão gerar *scores* de atenção α_{ji} que representam a correlação multidimensional entre todos os valores do conjunto de entradas.

$$\alpha_{ji} = softmax(\xi(\mathbf{u}_j,\mathbf{v}_i)) \tag{4.72}$$

$$\mathbf{c}_j = \sum_i \alpha_{ji}\mathbf{v}_i \tag{4.73}$$

A atenção múltipla pode ser estendida por meio da chamada auto atenção. Nesta variação o mecanismo de atenção é aplicado juntamente com alguma técnica de geração de representações

densas, a mais conhecida é a chamada *embedding* (LEVY; GOLDBERG, 2014). Nesta abordagem, o mecanismo de atenção foca em partes diferentes da mesma sequência de entrada, a depender da função de perda e do momento em que a mesma é apresentada a rede. Neste sentido, a correlação é estimada entre os próprios valores contidos em u_j. Os coeficientes de atenção α_{ji} são calculados por meio de uma combinação entre as representações densas do vetor u_j e das entradas v_i gerando uma nova representação densa que será o vetor c_j. Assim, o vetor de contexto expressará não somente a informação que representa uma dada entrada, como também a combinação de todos os valores da sequência ponderados pela similaridade existente entre eles.

4.2.4 Células de memória

RNNs, apesar de apresentarem uma estrutura adequada para o tratamento de sequências, tem sua qualidade degradada conforme o tamanho da sequência de entrada aumenta. O desafio de se armazenar dependências de longo prazo com RNAs foi relatado pela primeira vez no início da década de 90 (BENGIO; SIMARD; FRASCONI, 1994).

Sendo a RNA um grafo computacional, ao se utilizar uma topologia com camadas profundas, qualquer valor que flua pelo grafo será repetidamente multiplicado por uma matriz de pesos $\mathbf{W}$. Considerando t a quantidade de multiplicações, então este procedimento será equivalente a $\mathbf{W}^t$. Ao ser tomada a decomposição da matriz $\mathbf{W}^t$ como descrita na Equação 4.74 pode-se observar que se os autovalores em λ estiverem acima de um, então o resultado da operação tenderá para infinito. Do contrário, se abaixo de um, então tenderá para zero. Qualquer uma dessas situações inviabiliza a otimização dos parâmetros da rede, pois ambas interferem na direção da descida dos gradientes.

$$\mathbf{W}^t = \left(\mathbf{V} diag(\lambda) \mathbf{V}^{-1}\right)^t = \mathbf{V} diag(\lambda)^t \mathbf{V}^{-1} \tag{4.74}$$

No caso específico das RNNs este problema é ainda mais grave, pois além das camadas profundas, as entradas são submetidas adicionalmente as multiplicações relacionadas as recorrências. A Equação 4.74 pode ser adaptada para uma RNN de forma a ficar como representada na Equação 4.75 onde todas os estados recorrentes são propagados via multiplicação de $\mathbf{W}^t$ com o primeiro estado $\mathbf{h}^{(0)}$. Ao se levar em conta que neste caso o resultado da decomposição da matriz $\mathbf{W}$ é como o descrito na Equação 4.76 sendo $\mathbf{Q}$ ortogonal, então o estado $\mathbf{h}^{(t)}$ apresentará também o problema de explosão ou desaparecimento de gradientes devido aos autovalores acima ou abaixo de um em $\mathbf{\Lambda}^t$, como observado na Equação 4.77.

$$\mathbf{h}^{(t)} = \mathbf{W}^\top \mathbf{h}^{(t-1)} = \left(\mathbf{W}^t\right)^\top \mathbf{h}^{(0)} \tag{4.75}$$

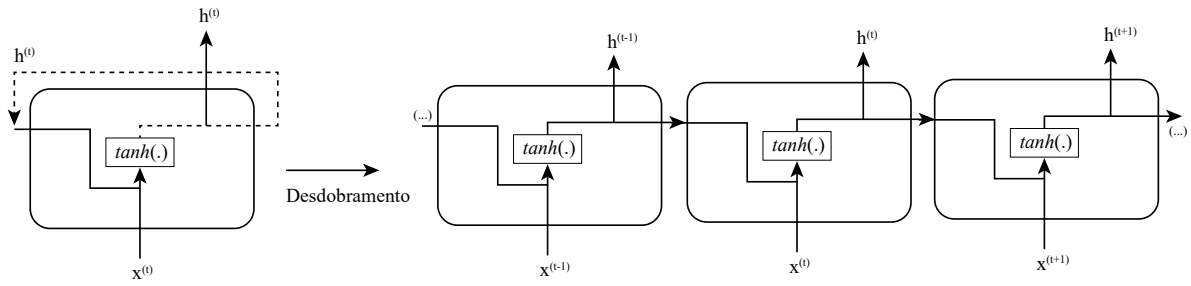
$$\mathbf{W} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top \tag{4.76}$$

$$\mathbf{h}^{(t)} = \mathbf{Q}^\top \mathbf{\Lambda}^t \mathbf{Q} \mathbf{h}^{(0)} \tag{4.77}$$

Os trabalhos de Hochreiter e Schmidhuber (1997) e Chung *et al.* (2014) propuseram soluções para a otimização de arquiteturas que tratam as dependências de longo prazo. A contribuição consiste em modificar o fluxo do gradiente para fazer com que as matrizes de peso sejam condicionadas ao contexto. Esta abordagem foi chamada de célula de memória e, entre as células de memória existentes, destacam-se a LSTM e a GRU.

Uma célula LSTM ou GRU difere de uma RNN convencional, pois ao invés de modelar a recorrência apenas por meio de uma única unidade não linear, ambas possuem uma recorrência interna controlada por funções não lineares que combinam, alteram e propagam o estado oculto em conjunto com a entrada. Estas funções são chamadas de "portões". Para fins de comparação, na Figura 70 pode ser observado o esquema de desdobramento temporal de uma RNN convencional e, na Figura 71, o de uma RNN com célula LSTM.

Figura 70 – Esquema de recorrência e desdobramento de uma RNN.



Fonte: O Autor.

Uma célula LSTM possui duas saídas. A primeira saída é chamada de estado da célula $c^{(t)}$ e a segunda, como numa RNN convencional, é o estado oculto $h^{(t)}$. Essas saídas são processadas por quatro portões que, de forma análoga a um circuito, modulam a entrada por meio de funções de ativação, termos de viés e pesos. Considerando um vetor de entradas $x_j^{(t)}$ e um vetor de estado oculto anterior $h_j^{(t-1)}$, todos os portões possuem suas próprias matrizes de pesos que ponderam estes vetores sendo estas $U_{i,j}^k$ e $W_{i,j}^k$. Nestas matrizes, i é o índice da célula LSTM, j é índice da entrada e k indica o portão sendo:

a) Esquecimento: Definido por $f_i^{(t)}$ com $k=f$, conforme a Equação 4.78.

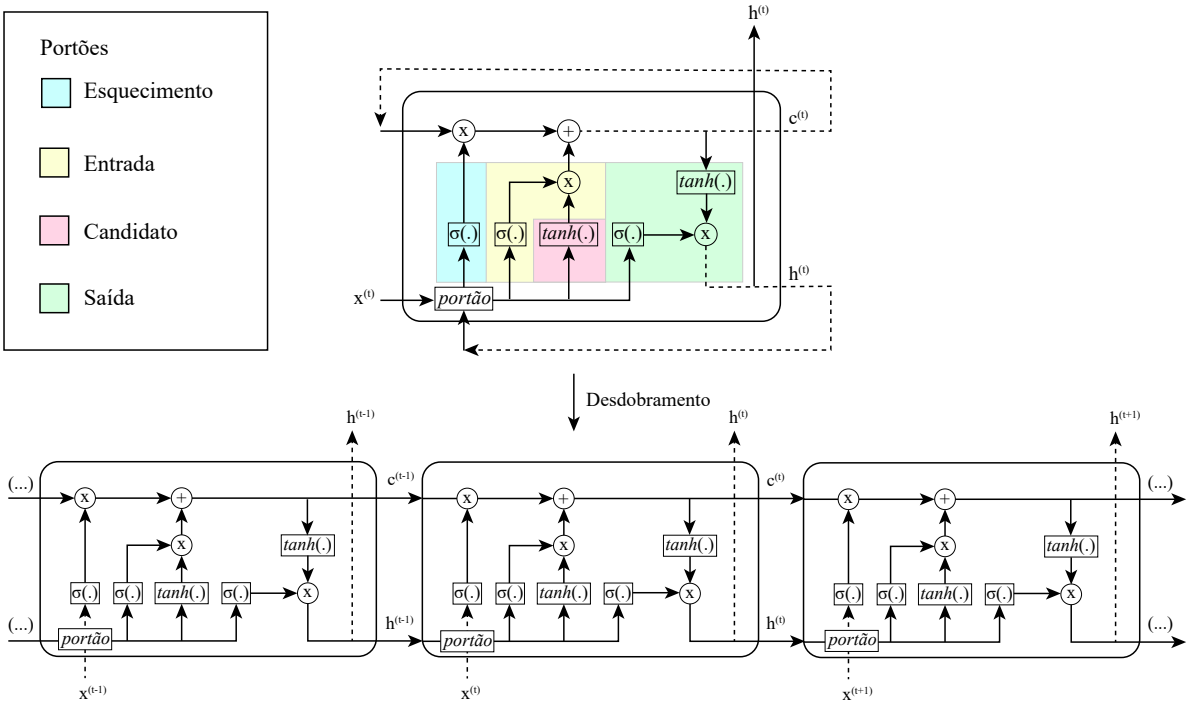
b) Entrada: Definido por $p_i^{(t)}$ com $k=p$, conforme Equação 4.79.

c) Candidato: Definido por $d_i^{(t)}$ com $k=d$, conforme Equação 4.80.

d) Saída: Definido por $o_i^{(t)}$ com $k=o$, conforme Equação 4.81.

$$f_i^{(t)} = \sigma\left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)}\right) \tag{4.78}$$

Figura 71 – Esquema de recorrência e desdobramento de uma RNN com célula de memória LSTM.



Fonte: Adaptado de Yu *et al.* (2019a).

$$p_i^{(t)} = \sigma\left(b_i^p + \sum_j U_{i,j}^p x_j^{(t)} + \sum_j W_{i,j}^p h_j^{(t-1)}\right) \tag{4.79}$$

$$d_i^{(t)} = tanh\left(b_i^d + \sum_j U_{i,j}^d x_j^{(t)} + \sum_j W_{i,j}^d h_j^{(t-1)}\right) \tag{4.80}$$

$$o_i^{(t)} = \sigma\left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)}\right) \tag{4.81}$$

A ideia central é que o estado da célula $c_i^{(t)}$ seja capaz de armazenar e propagar para frente a informação do instante $(t-1)$. A LSTM pode decidir se este estado será realmente propagado ou, do contrário, ignorado. Esta tarefa é realizada pelo portão de esquecimento que pondera $c_i^{(t-1)}$ por meio da saída submetida a uma função logística sigmoide no intervalo entre 0 e 1. A próxima etapa é somar o estado ponderado da célula anterior $c_i^{(t-1)}$ com o valor de entrada modulado. O valor de entrada é gerado pelos portões de entrada e candidato. O portão de entrada, como o de esquecimento, gera uma saída entre 0 e 1 que pode propagar ou ignorar o resultado do portão candidato. O portão candidato, por meio de uma função tangente hiperbólica, é capaz de considerar valores negativos, logo possui a capacidade de inibir um sinal ao invés de simplesmente bloqueá-lo. Como descrito na Equação 4.82, a combinação destes três portões gera uma atualização no estado da célula $c_i^{(t)}$ que, de forma recorrente, irá ser utilizado no instante $(t+1)$.

$$c_i^{(t)} = f_i^{(t)} c_i^{(t-1)} + p_i^{(t)} d_i^{(t)} \tag{4.82}$$

O último passo é a geração da saída. O estado corrente da célula $c_i^{(t)}$ já atualizado é submetido a uma função tangente hiperbólica. A resposta é então multiplicada pelo valor resultante do portão de saída. Como pode ser observado na Equação 4.83, finalmente, o estado oculto $h_i^{(t)}$ então é gerado.

$$h_i^{(t)} = o_i^{(t)} tanh\left(c_i^{(t)}\right) \tag{4.83}$$

Uma célula GRU atua da mesma maneira que uma LSTM. No entanto, difere desta ao considerar um único portão para controlar simultaneamente as tarefas de esquecimento e atualização do estado oculto $c_i^{(t)}$. A Equação 4.84 descreve a forma de geração do estado oculto $h_i^{(t)}$. As equações 4.85 e 4.86, respectivamente, descrevem os portões de atualização e redefinição que fazem essas substituições.

$$h_i^{(t)} = u_i^{(t-1)} h_i^{(t-1)} + \left(1 - u_i^{(t-1)}\right) \sigma\left(b_i + \sum_j U_{i,j}^d x_j^{(t)} + \sum_j W_{i,j} r_j^{(t-1)} h_j^{(t-1)}\right) \tag{4.84}$$

$$u_i^{(t)} = \sigma\left(b_i^u + \sum_j U_{i,j}^u x_j^{(t)} + \sum_j W_{i,j}^u h_j^{(t)}\right) \tag{4.85}$$

$$r_i^{(t)} = \sigma\left(b_i^r + \sum_j U_{i,j}^r x_j^{(t)} + \sum_j W_{i,j}^r h_j^{(t)}\right) \tag{4.86}$$

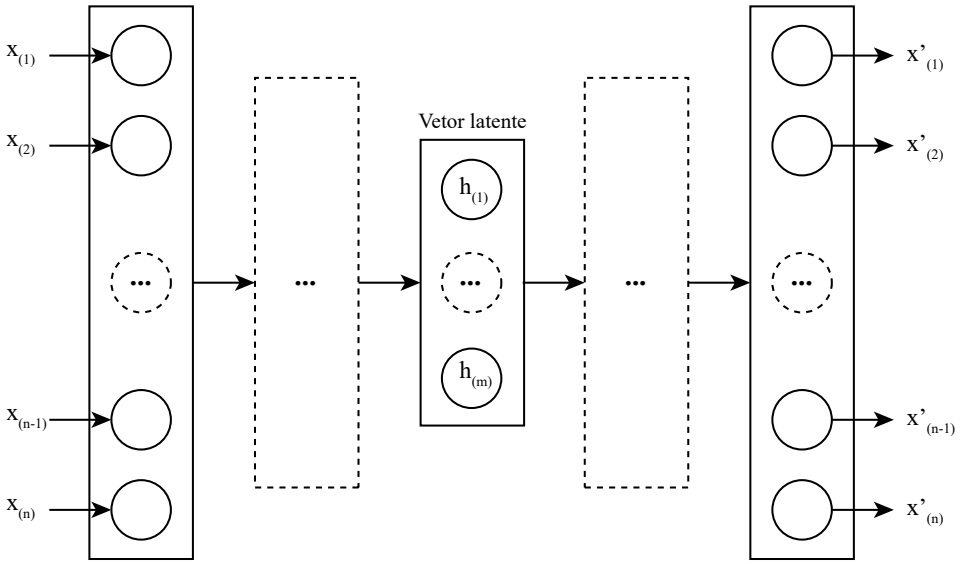
LSTM e GRU são utilizadas para a mesma finalidade. A escolha de uma, de outra ou de utilizá-las em conjunto depende de experimentação. A adequação tende a ser definida conforme o problema de aplicação. Em teoria a célula LSTM modela mais parâmetros que a GRU, por isso, teria maior capacidade de armazenar dependências de longo prazo. De outro ponto de vista, a GRU apresenta maior facilidade de convergência e menor custo computacional. Para um estudo comparativo entre as duas, sugere-se consultar Yang, Yu e Zhou (2020). O trabalho de Yu *et al.* (2019a) traz uma revisão de algumas variações possíveis para GRU como *Minimal Gated Unit* e, também para LSTM, como LSTM bidirecionais, multidimensionais, baseada em grafos, em árvores, em tabelas e outras.

4.2.5 *Autoencoders*

Um AE é uma RNA do tipo *feedforward* que tenta copiar os valores do vetor de entrada na camada de saída. No entanto, durante este processo uma modificação de dimensionalidade é forçada nas camadas internas. Essa propriedade é conhecida como gargalo (*bottleneck*) e o vetor mais interno da rede é denominado vetor latente (GOODFELLOW; BENGIO; COURVILLE, 2016).

Como visto na Figura 72, a rede se assemelha a uma Seq2Seq por possuir também uma codificadora e uma decodificadora. A codificadora tem o papel de reduzir (ou aumentar) um vetor de entrada de tamanho n para o tamanho m no vetor latente. Na sequência, a decodificadora reconstrói o vetor de entrada a partir do vetor latente. A operação, portanto, tem o objetivo de inferir uma função de codificação $h(x)$ e uma função de decodificação $g(h(x)) = x'$.

Figura 72 – Esquema da arquitetura *autoencoder*.



Fonte: O Autor.

Como o modelo é forçado por meio da função de treinamento a priorizar quais aspectos da entrada devem ser copiados, um AE aprenderá a codificar no vetor latente apenas propriedades úteis da informação. Dessa forma, pode-se desacoplar a decodificadora do AE e, posteriormente, incluir a codificadora como entrada em um processo externo ou outra RNA.

A arquitetura AE possui capacidade de executar uma modificação de dimensionalidade não linear que pode se tornar mais acurada conforme a profundidade do modelo e quantidade de dados aumenta. Como outras arquiteturas de DL, um AE é robusto também a ruídos, nesse sentido, pode apresentar resultados superiores a outras técnicas que tem a mesma finalidade como PCA, *Linear Discriminant Analysis* (LDA), *Locally Linear Embedding* (LLE), *Isomap* e outras (WANG; YAO; ZHAO, 2016; FOURNIER; ALOISE, 2019; LUO *et al.*, 2019). AEs podem apresentar variações como *Sparse* AE, *Denoising* AE e *Variational* AE. Para mais detalhes sobre estas variações segure-se consultar Goodfellow, Bengio e Courville (2016).

4.2.6 Técnicas complementares

A definição de uma arquitetura de DL adequada para um dado problema é dependente de um processo de experimentação (LAGO; De Ridder; De Schutter, 2018). Apesar de existirem hiper heurísticas que visam auxiliar na seleção de parâmetros, alguns pontos cruciais como a escolha de uma arquitetura (e sua posterior topologia) são, antes de tudo, tarefas de *design*. Dada a extensão do

tema e já introduzidos os conceitos fundamentais das arquiteturas utilizadas, a seguir são descritas de forma breve algumas técnicas complementares importantes no treinamento de modelos neurais profundos. Algumas das quais, utilizadas na metodologia discutida no capítulo seguinte.

4.2.6.1 Inicialização

Métodos de inicialização de parâmetros podem afetar a qualidade da otimização em modelos profundos. Isto é particularmente importante no que tange a velocidade de convergência. Uma inicialização que parte de pesos bem determinados acelera o processo de treinamento e tende a construir um modelo mais generalista (GLOROT; BENGIO, 2010).

Métodos comuns de inicialização incluem determinar pesos de forma aleatória dentro de um intervalo segundo alguma distribuição, como normal ou uniforme. Outra abordagem é iniciar uma RNA com pesos todos iguais a um ou zero. Outras técnicas mais recentes como as propostas por Glorot e Bengio (2010), He *et al.* (2015) e Kalmbauer *et al.* (2017) analisam a distribuição dos dados para definir parâmetros iniciais. Estes parâmetros são então utilizados para gerar os pesos com base nas funções de ativação, na quantidade de neurônios e em limites de variação.

4.2.6.2 Otimização

Células de memória resolvem o problema dos gradientes que explodem ou desaparecem. No entanto, dada a profundidade das redes no contexto de DL, técnicas auxiliares e algoritmos mais modernos substituíram o uso de *Backpropagation*. Estes vieram não somente para acelerar o processo de treinamento, como também para aumentar a probabilidade de se obter uma rede capaz de produzir respostas de qualidade sem perda de generalização. Como já discutido no Capítulo 3, desde 2012 com a introdução do *pre-trainning*, trabalhos como o de Martens e Sutskever (2011) focados em redes recorrentes, detectaram que em RNNs as segundas derivadas desaparecem ao mesmo tempo que as primeiras. Nesse sentido, dado que a segunda derivada apresenta uma diminuição que segue a mesma proporção da primeira, então métodos de segunda ordem passaram a ser alvos de pesquisa por, teoricamente, apresentarem superioridade (GOODFELLOW; BENGIO; COURVILLE, 2016). No entanto, métodos de segunda ordem possuem limitações como alto custo computacional, necessidade de um grande *batch* e uma tendência para serem atraídos para pontos de sela (SUTSKEVER, 2012). Com o objetivo de tornar o uso dos métodos de segunda ordem viáveis outras técnicas foram desenvolvidas. Dentre estas destacam-se o *clipping* de gradientes e o *momentum* de Nesterov.

Clipping diz respeito a realização de um corte no valor do gradiente, reposicionando este valor em um dado limite se, durante o processo de otimização, o valor deste limite for violado. O corte é feito na norma do gradiente antes da atualização do parâmetro conforme descrito na Equação 4.87 onde v é o limite e g é o valor do gradiente. O *clipping* evita a explosão dos gradientes devido as recorrências. É útil também se resultados computacionalmente inválidos como *Inf* (infinito) e *NaN* (não numérico) forem gerados. Apesar de alterar o gradiente, a correção

empregada pelo *clipping* não afeta negativamente a direção da otimização, visto que na época seguinte o gradiente é sempre atualizado (PASCANU; MIKOLOV; BENGIO, 2013).

$$se\ \|g\| > v,\ g \leftarrow \frac{gv}{\|g\|} \tag{4.87}$$

Momentum de Nesterov ou *Nesterov Accelerated Gradient* (NAG) pode ser considerado uma evolução do termo de *momentum* clássico (QIAN, 1999), pois apresenta garantias teóricas de convergência em funções convexas. O trabalho de Sutskever *et al.* (2013) demonstra empiricamente esta superioridade. A diferença principal entre ambos é que no NAG são estimados parâmetros provisórios antes de se definir, de fato, o valor do gradiente. Isto é chamado de gradiente de antecipação (do inglês, *lookahead gradient*). Dado que o método usa parâmetros provisórios, então o NAG possui capacidade de responder mais rapidamente as mudanças de direção do gradiente no decorrer da otimização. Essa propriedade garante um comportamento mais estável com variações mais sutis (SUTSKEVER *et al.*, 2013).

Tanto *clipping* quanto NAG quando aplicados junto a algoritmos de primeira ordem contribuem para que os mesmos apresentem desempenho similar aos de segunda ordem (SUTSKEVER *et al.*, 2013). Este fato fez com que o *pre-trainning* e outras abordagens mais antigas fossem substituídas em RNNs, especialmente as envolvendo LSTMs (GOODFELLOW; BENGIO; COURVILLE, 2016). Um destes algoritmos de primeira ordem é o Adam (KINGMA; BA, 2014) que, quando adaptado para incluir o NAG passa a ser denominado Nadam (DOZAT, 2016).

O principal avanço do otimizador Adam em relação aos algoritmos anteriores é o uso de taxas de aprendizado individuais adaptáveis para cada um dos parâmetros da rede, ao invés de apenas uma única taxa para a rede inteira. Esta abordagem trata o problema da esparsidade, geralmente observado junto a gradientes em redes profundas. A adaptação dos parâmetros é ajustada por meio de médias móveis dos gradientes e dos quadrados dos gradientes. As médias móveis são ainda ponderadas temporalmente a fim de executar um controle do decaimento exponencial das mesmas. Essa ponderação é realizada via dois hiperparâmetros específicos. A regra de atualização do algoritmo envolve também o controle de limites, impedindo que os termos de ponderação descaracterizem as médias móveis e, consequentemente, os valores dos gradientes (KINGMA; BA, 2014). Em Ruder (2016) são comentadas algumas das possíveis variações do Adam.

4.2.6.3 Generalização

Um problema conhecido em RNAs é o super ou sobreajuste. O mesmo acontece quando a parametrização do modelo se torna tão específica para um determinado conjunto de treinamento que, quando aplicado a dados fora da amostra, a qualidade dos resultados fica prejudicada, visto que não há generalização. Neste sentido, pode-se dizer que a rede não "aprende" o problema, ela o "memoriza". A forma mais simples de se evitar este problema é por meio de uma parada prematura do treinamento (do inglês, *early stopping*). Embora seja útil, esta ação sozinha não

elimina o problema e, por isso, técnicas chamadas de regularizadoras foram propostas. A ideia da regularização é restringir parte dos parâmetros do modelo durante a otimização para que o ajuste dos pesos mude de forma a favorecer o fluxo da informação para possíveis partes da RNA que, porventura, não estavam sendo utilizadas.

Uma das formas de se aplicar regularização é adicionar ou multiplicar o resultado das funções de perda a valores de penalidades. As penalidades, então, são aplicadas aos pesos com altos valores de forma a zerá-los. Duas técnicas que tratam desta abordagem são a regularização L1 e L2. Outra forma consiste em desativar temporariamente e de forma aleatória, segundo uma taxa de probabilidade pré-especificada (i) um ou mais neurônios, (ii) uma ou mais conexões entre neurônios ou (iii) neurônios e conexões ao mesmo tempo. A desativação de neurônios é chamada de *dropout* e a de conexões de *drop connect* (MORADI; BERANGI; MINAEI, 2020).

Por fim, técnicas de normalização internas (não as utilizadas para pré-processamento de dados) podem ser empreendidas a fim de evitar sobreajuste. Ações como normalizar gradientes camada a camada (após saída das funções de ativação) (BA; KIROS; HINTON, 2016) ou normalizar gradientes por *batch* (IOFFE; SZEGEDY, 2015) mantendo a média centrada em zero e a variância em um, são alternativas.

4.3 Gestão de riscos

A expectativa do preço da eletricidade é utilizada para determinar montantes transacionados via contratos. No caso do Brasil, uma estimativa futura acertada do PLD pode ser utilizada para estabelecer um contrato lucrativo para o agente interessado. Como visto no Capítulo 2, operações das distribuidoras e consumidores livres no Brasil devem estar totalmente lastreadas, respectivamente, em contratos e garantia física. A determinação da garantia física para cada gerador depende do que é estabelecido pelo MME. O cálculo, em termos gerais, considera a capacidade histórica de produção simulada junto a cenários de condições climáticas desfavoráveis. Neste contexto, os termos estabelecidos em contrato se tornam parâmetros sensíveis, pois condições climáticas são incertas, consequentemente, o PLD também. Se o PLD é mal estimado, então possíveis violações as regras de lastreamento poderão ocorrer, o que por sua vez gerará ao agente não apenas prejuízos, como também penalidades junto a CCEE.

Métodos de previsão, por mais sofisticados que sejam, estão suscetíveis a erros. Dessa maneira, a previsão de valores futuros passa a ser tão importante quanto a forma como estes valores serão utilizados para estabelecer contratos. Mesmo que um preditor possua capacidade de gerar previsões acertadas em, por exemplo, 99% das ocasiões, todo sucesso da operação pode ser perdido em apenas 1% de previsões erradas. Sendo assim, torna-se importante realizar uma associação dos métodos de previsão com técnicas focadas na criação de estratégias de negociação inteligentes que quantifiquem o grau de risco.

Uma estratégia de negociação inteligente é aquela capaz de maximizar o lucro de uma

operação ao mesmo tempo em que minimiza as chances de perdas (CAETANO, 2017). Essa área de estudo é chamada de gestão de riscos e é relativamente nova no contexto de mercados de eletricidade (MAYO, 2012). Risco trata-se da possibilidade de ocorrência de um evento potencialmente danoso ao resultado financeiro de um negócio. Em mercados de eletricidade estes eventos podem estar relacionados a alta volatilidade, picos extremos de preços, preços dos combustíveis, inadimplência, previsões imprecisas de demanda, falhas de sistema, falta de liquidez, mudanças de leis e outros (MAYO, 2012).

O risco só pode ser mitigado em uma operação financeira se houver alguma forma de mensuração. Neste sentido, além das métricas descritas na subseção 4.1.3.1, o cenário que envolve as previsões do PLD pode ser quantificado financeiramente com base no risco, a fim de se gerar uma medida de incerteza. A incerteza é de crucial importância, pois pode informar qual é a exposição ao risco de um dado agente ao considerar o histórico de preços e as previsões. A performance histórica de um preditor pode ser calculada em termos financeiros com base nos seus retornos ajustados dentro do modelo de precificação hipotético utilizado em um dado mercado. Este ajustamento pode ainda ser melhor elaborado levando-se em conta possíveis cláusulas contratuais que especifiquem a proteção da operação por meio, por exemplo, de intervalos onde as previsões de carga e preço podem variar caso não estejam exatamente corretas. Já o risco pode ser estimado conforme modelos de risco. Dentre estes modelos destacam-se o VaR e o CVaR.

4.3.1 Modelos de risco VaR e CVaR

O VaR é uma medida estatística utilizada para estimar perdas potenciais de um dado método de investimento. VaR informa a perda máxima esperada conforme um conjunto histórico de retornos financeiros com um nível de significância α para um horizonte de tempo futuro. O cálculo do VaR pode ser realizado segundo as abordagens paramétrica e não paramétrica. Na paramétrica, se admite que os retornos seguem uma distribuição normal. Já na não paramétrica, dado que os retornos não são normalmente distribuídos, utiliza-se a própria distribuição do histórico dos retornos ou outra proveniente de um método de simulação como Monte Carlo (CAETANO, 2017; LIMA, 2018). No contexto desta tese interessa apenas a versão não paramétrica com dados não simulados.

Dado um conjunto de retornos percentuais calculados como na Equação 4.88, uma descrição genérica do VaR é feita na Equação 4.89. Nesta a probabilidade P de um retorno Δx no instante t ser menor ou igual ao estimado pelo VaR deve ser igual a α. Nesse sentido, o $VaR_{(\alpha)}$ para um dado conjunto de preços X pode ser definido como na Equação 4.90 onde $q^{\alpha}(X)$ é o maior quantil de α considerando os limites descritos como na Equação 4.91.

$$\Delta x = \frac{\left(x^{(t)} - x^{(t-1)}\right)}{x^{(t-1)}} \tag{4.88}$$

$$P\left\{\Delta x^{(t)} \leq VaR^{(t)}\right\} = \alpha \tag{4.89}$$

$$VaR_{(\alpha)}(X) = -q^{\alpha}(X) \tag{4.90}$$

$$q^{\alpha}(X) = inf\{x : P(X \leq x) > \alpha\} = sup\{x : P(X < x) \leq \alpha\} \tag{4.91}$$

Levando-se em conta estas definições, o cálculo do VaR para um dado α pode ser realizado conforme a Equação 4.92 onde $\hat{\sigma}_{\Delta x}$ é o desvio padrão amostral dos retornos e $z(\alpha)$ é o limiar definido pelo maior quantil de α na distribuição normal (*z-score*) descrito como na Equação 4.93. O VaR no instante t é dado com base na média dos retornos $\mu_{\Delta x^{(t)}}$ e no último valor de preço conhecido $x^{(t)}$ como descrito no Equação 4.94.

$$VaR_{(\alpha)} = z(\alpha)\hat{\sigma}_{\Delta x} \tag{4.92}$$

$$P(Z \geq \alpha) = \alpha,\ Z \sim i.i.d.\ \mathcal{N}(0,1) \tag{4.93}$$

$$VaR^{(t)}_{(\alpha)}(X) = -x^{(t)}\left(VaR_{(\alpha)} + \mu_{\Delta x^{(t)}}\right) \tag{4.94}$$

A interpretação prática do VaR trata de considerar que o resultado do cálculo da equação 4.94 corresponde a perda máxima possível com $(1-\alpha)\%$ de chance de ocorrência. Valores superiores ao VaR teriam uma chance menor que $(1-\alpha)\%$ de acontecer.

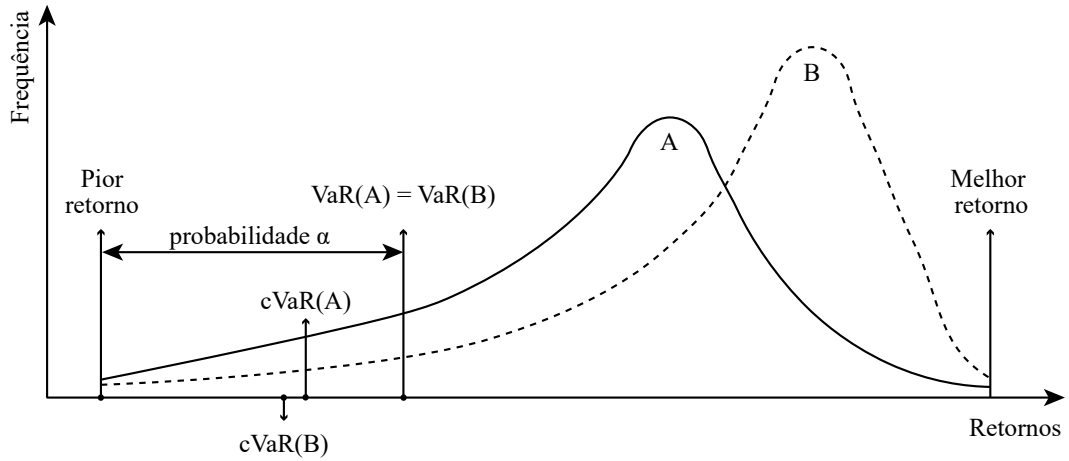
A versão não paramétrica do VaR trata-se da estimação não dependente de premissas e leva em conta apenas a distribuição histórica dos retornos. Neste caso admite-se que o comportamento futuro dos preços não sofrerá mudanças. O cálculo é baseado no ordenamento decrescente dos valores obtidos como na Equação 4.88, ou seja, da máxima perda para o máximo ganho. O VaR, neste sentido, corresponderá a menor perda observada no conjunto das $\alpha\%$ maiores perdas. O uso do modelo não se altera mesmo quando não existem lucros para serem avaliados. Este é o caso da compra de eletricidade onde são considerados apenas os gastos de consumo.

O VaR trata da perda limite esperada com um nível de significância α. No entanto, este valor não informa nada sobre o provável valor de perda se este limite descrito pelo VaR for excedido. Em outras palavras, torna-se interessante determinar qual a perda provável levando-se em consideração todas as possibilidades de perdas contidas em $\alpha\%$ dos piores retornos. Essa métrica é chamada de CVaR.

O CVaR diz respeito a média de todos os retornos que excedem o valor do VaR, logo informa um retorno médio considerando-se os piores retornos históricos. O CVaR é mais específico do que o VaR. Como exemplificado na Figura 73, para duas distribuições diferentes A e B com mesmos valores de pior e melhor retorno, o VaR assume o mesmo valor para um dado α. No entanto, o cVaR será diferente, pois o valor médio na cauda de piores retornos difere entre as

distribuições. Dessa forma, o CVaR tende a retratar melhor a realidade apesar de não ter exatamente o mesmo propósito do VaR .

Figura 73 – Comparação de VaR e CVaR para duas hipotéticas distribuições de retornos.



Fonte: O Autor.

5 METODOLOGIA

Este capítulo apresenta a metodologia empregada no desenvolvimento de um método computacional voltado para o mercado de eletricidade brasileiro capaz de realizar previsões semanais do PLD horário levando-se em conta o atual estado da arte com uso de um *ensemble* de preditores baseados em ANNs e LEAR incluindo gestão de risco. A seguir o leitor irá encontrar duas subseções. Na primeira é apresentada uma macro visão que trata da organização do sistema como um todo e seus quatro módulos. É apresentado o fluxo da informação entre estes módulos de forma a deixar claro o papel que cada um possui e como se integram. Na segunda subseção são expostos detalhes das atividades internas de cada módulo separadamente. São discutidas as agências provedoras de dados, as variáveis de entrada selecionadas, o processamento destas entradas, as partes que compõe a arquitetura neural do modelo preditor e, por fim, a forma de construção dos cenários futuros que convertem as previsões do PLD em termos financeiros por meio de um modelo de precificação hipotético.

Ao adotar um ponto de vista de produto, o método proposto ao longo deste e do próximo capítulo será chamado de sistema OPB. Esta é uma referência ao nome do projeto Observatório do Preço da Energia do Brasil (OPB) como já citado no Capítulo 1.

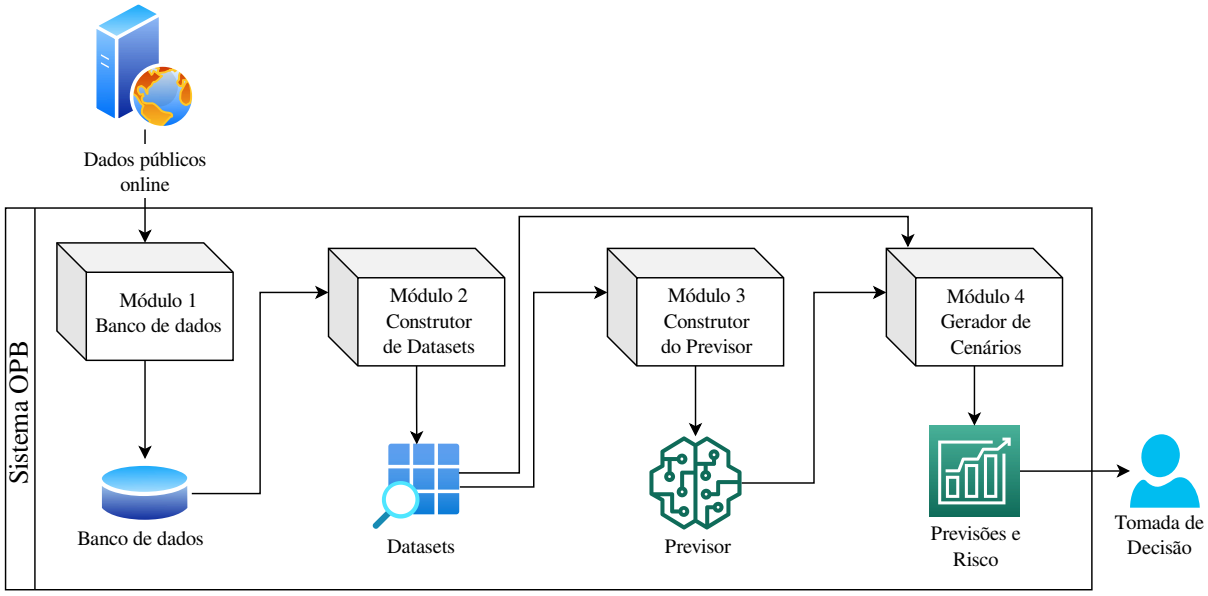
5.1 Organização do sistema

O sistema OPB está fundamentado na utilização de dados públicos disponibilizados *online* por agências governamentais. Estes dados são processados segundo fluxograma de operação da Figura 74. Durante este processamento diversas avaliações e consequentes decisões conduzem a produção de artefatos específicos como resultado de fases distintas. O último artefato da cadeia (resultado final de todo processo) trata-se de um cenário previsto que deve ser utilizado por um operador humano. É importante salientar, portanto, que não faz parte do escopo do sistema OPB a automação da fase comercial. O foco do sistema trata-se da geração de informações que servem de suporte a tomada de decisão por parte do agente comercializador.

O primeiro módulo chamado de Módulo 1: Banco de Dados (M1) é responsável por ler os dados retirados das fontes *online* no formato de origem e convertê-los para inserção em um banco de dados relacional. O M1 é no contexto do sistema um componente capaz de gerar de forma autônoma um banco de dados completo do zero, contendo de forma estruturada todas as informações providas pelas agências selecionadas. Este módulo também é capaz de atualizar os dados quando novos arquivos das agências são disponibilizados.

Após a construção (ou atualização) do banco de dados realizada pelo primeiro módulo, atua o segundo módulo chamado de Módulo 2: Construtor de *Datasets* (M2). O papel do M2 é selecionar partes específicas dos dados contidos no banco, executar pré-processamentos diversos e gerar conjuntos prontos que servirão de base para o treinamento, validação e testes dos modelos preditores. Neste contexto, as etapas de pré-processamento e codificação destacam-se como mais importantes, pois visam aplicar separadamente aos conjuntos de dados diversas transformações

Figura 74 – Fluxograma de operação.



Fonte: O Autor.

com o objetivo de facilitar a otimização no módulo seguinte e, por consequência, favorecer o aumento de confiabilidade dos cenários previstos úteis ao agente comercializador.

O terceiro módulo é responsável por acionar o M2 quando necessário requisitando conjuntos de dados específicos para um dado intervalo no tempo. É chamado de Módulo 3: Construtor do Previsor (M3). A questão do tempo é importante no M3, pois a construção dos conjuntos de dados deverá sempre levar em conta a técnica de janelamento móvel, discutida mais adiante. Na sequência, estes *datasets* devem ser lidos, interpretados e usados como base na fase de geração de modelos de previsão intermediários que, após avaliação, são combinados na forma de um *ensemble*. Dentro da dinâmica de operação, o *ensemble* é o previsor final utilizado como base para geração dos cenários no módulo seguinte.

Finalmente, o previsor gerado serve de insumo ao último módulo da cadeia chamado de Módulo 4: Gerador de Cenários (M4). O M4 (como o M3) também aciona o segundo módulo requisitando a construção de *datasets*. A diferença é que ao invés de solicitar a criação de vários conjuntos de dados para treinamento e validação, o M4 solicita ao M2 apenas a última janela de dados imediatamente anterior ao ponto de previsão. Esta janela é o suficiente para realizar a estimação de um cenário futuro que compreende um horizonte de 168 horas. O cenário é composto não somente do PLD previsto de forma pontual, como também de valores que podem ser usados para estimar uma curva de densidade probabilística, de métricas de incerteza relacionadas aos intervalos de variação e indicativos de risco financeiro que levam em conta o histórico dos preços até o final da janela de previsão. É exatamente este cenário que é entregue ao agente comercializador como um relatório para auxiliar nas suas operações de negociação de eletricidade.

5.2 Módulos

Conforme a organização do sistema apresentada na subseção anterior, a seguir cada módulo é discutido individualmente de forma a demonstrar como é feita a geração dos quatro artefatos principais da operação: Banco de dados, *datasets*, modelo previsor e cenários. Antes no entanto cabe expor algumas decisões de implementação relacionadas à tecnologias e ferramentas.

O sistema OPB foi implementado parte em Python versão 3.7.6 (PYTHON, 2022) e parte em R 4.1.2 (R Core Team, 2022)[1]. Para o banco de dados foi utilizado o PostgreSQL versão 10 (POSTGRESQL, 2022). Todas estas ferramentas são *open source* e interoperáveis entre sistemas Windows, Linux e MacOS. A plataforma de desenvolvimento utilizada foi um Intel Core i9-10900 com RAM de 32 GB e GPU dedicada NVIDIA RTX 3080 TI operando com distribuição Linux Pop!_OS 22.04 (System 76, 2022).

Quanto ao PostgreSQL, o motivo pela escolha de um Sistema Gerenciador de Bancos de Dados (SGBD) (e não o uso direto de arquivos) é pelo fato de que cada agência faz a disponibilização das informações em formato diferente, quase sempre sendo conjuntos de arquivos de texto não estruturados, sem uso de qualquer padrão internacional (como *JavaScript Object Notation* (JSON) ou *Extensible Markup Language* (XML)). Outro fator é a quantidade de informação. O uso do SGBD além de promover padronização, facilita também o processo de consulta visto que fornece vantagens como suporte a *Structured Query Language* (SQL), organização de dados, mitigação de redundâncias, facilidade de compartilhamento de informações entre sistemas, escalabilidade e, em caso de geração de uma futura e provável *build* [2] do sistema, a existência de um SGBD facilitaria todo o processo de migração para um servidor de bancos de dados descentralizado contendo os necessários itens de segurança de acesso.

Em relação ao *software*, todos os módulos fazem uso de uma camada de leitura de dados. Essa camada contém todo o código-fonte necessário para abstração das tratativas que exigem controle de acesso ao PostgreSQL. Essa camada está inserida dentro do módulo M1 que foi construído com Python em paradigma orientado a objetos seguindo o padrão arquitetural de código-fonte *Model-View-Controller* (MVC) como discutido em Gamma *et al.* (2000). As parcelas relacionadas a modelagem das arquiteturas neurais e geração das previsões contidas nos módulos M2 e M3 foi implementada também em Python em formato de *notebooks* da plataforma Jupyter 6.4.3 (Jupyter Team, 2022). Por fim, a análise estatística dos dados e a geração dos modelos de risco do módulo M4 foi construída em R. A expectativa é que no futuro a arquitetura de código-fonte elaborada para o módulo M1 seja a base para a unificação de todos os módulos dentro de um mesmo projeto de *software* a ser feita por meio de refatoração do código-fonte dos outros três módulos.

---

[1] O uso destas diferentes tecnologias foram decisões de implementação do autor desta tese levando-se em conta a sua experiência prévia com as linguagens e a disponibilidade de bibliotecas contendo funções específicas a depender da necessidade. Python e R estão entre as 10 linguagens de programação mais populares do mundo em 2022 segundo a última versão do *ranking* IEEE Spectrum (IEEE, 2022).

[2] Versão finalizada do *software* disponibilizada para uso em um ambiente de produção profissional.

5.2.1 Módulo 1: Banco de dados

Antes da implementação do primeiro módulo, uma etapa preliminar foi realizada buscando-se identificar e selecionar agências que fossem capazes de prover variáveis úteis às fases que compõe a metodologia. Manteve-se o foco em três categorias de dados:

a) Energia: Qualquer informação relacionada ao histórico de operação do SEB e a comercialização de energia elétrica no ACL.

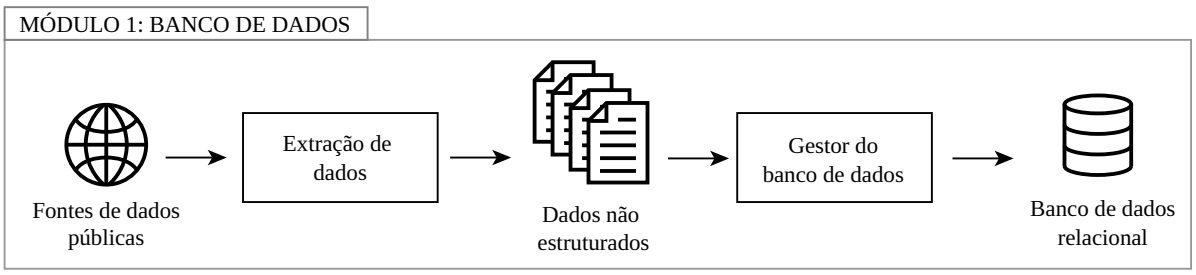
b) Economia: Dados relacionados as condições econômicas do Brasil.

c) Clima: Dados históricos de variáveis relacionadas as condições horárias de tempo.

Na seleção das agências levou-se em consideração critérios subjetivos como sugestões recebidas de outros pesquisadores e/ou profissionais da área e, também, objetivos como frequência de citações em outros trabalhos, se a agência era ou não governamental, se tratava-se de uma associação de comercializadores/geradores/consumidores ou mesmo empresas privadas detentoras das maiores participações no ACL. Já as variáveis foram selecionadas considerando-se principalmente a frequência de uso observada nos trabalhos constantes na revisão bibliográfica (Capítulo 3) e, além disso, quais destes dados as agências brasileiras identificadas eram capazes de prover. Outros critérios serviram de eliminação ao longo do processo como preferir àqueles em formato horário, acesso *online*, autorização de uso para terceiros, frequência de atualização feita pela agência publicadora, granularidade que contemplasse no mínimo cada um dos subsistemas do SIN individualmente e outros.

O uso destas variáveis acontece inicialmente no módulo M1 que está diagramado conforme mostrado na Figura 75. A transformação dos dados desde a fonte até o banco é realizada em duas etapas. A primeira se trata da extração manual dos arquivos nas fontes selecionadas[3] e, a segunda, do processo automatizado realizado por um submódulo chamado *Gestor do Banco de Dados* (GBD).

Figura 75 – Atividades do Módulo 1: Banco de Dados.



Fonte: O Autor.

[3] Foi tentada a implementação de um submódulo *Web Crawler* capaz de realizar os procedimentos de acesso e extração dos dados das agências automaticamente. No entanto, durante a pesquisa percebeu-se que as agências constantemente mudam a forma de acesso aos arquivos. Dada a necessidade de constante esforço na manutenção destas regras optou-se pela não utilização do *Web Crawler*.

Na etapa de extração dos dados são consideradas as fontes listadas a seguir. Todas estas dispõe seus dados publicamente nos portais indicados em cada uma das respectivas referências:

a) CCEE (CCEE, 2022f).

b) Instituto Brasileiro de Economia (IBE) (FGV, 2022c).

c) Instituto Brasileiro de Geografia e Estatística (IBGE) (IBGE, 2022).

d) Bolsa de valores B3 (B3, 2022).

e) Sistema SINtegre do ONS (ONS, 2022c).

f) Instituto Nacional de Meteorologia (INMET) (INMET, 2022).

Alguns dos portais só liberam a requisição após várias horas com um *link* de acesso enviado por *e-mail*. Os arquivos extraídos são mantidos em seu formato original exatamente como as agências os proveriam para qualquer cidadão que realizasse o *download*. Estes são então armazenados localmente. Após esta etapa é acionado o submódulo GBD.

O GBD tem o papel de criar uma instância de banco de dados do zero caso uma não exista. No entanto, caso já exista, então o que é feito é uma atualização do banco de dados existente com os dados mais recentes. O GBD é programado para interpretar as regras de formatação estabelecidas por cada agência em cada um dos seus arquivos. A leitura é realizada e, na sequência, os dados são organizados em comandos SQL que, por sua vez, são executados no SGBD. A partir da conclusão desta operação os dados ficam disponíveis *offline*.

As variáveis selecionadas podem ser divididas em variáveis do sistema elétrico, econômicas e climáticas. As mesmas estão descritas nas tabelas 7, 8 e 9. O sistema OPB utiliza somente dados em frequência horária, nesse sentido, variáveis com frequências de leituras diferentes tem seus valores repetidos tantas vezes quanto necessário até a obtenção desta discretização. Todos os dados inseridos no banco e, consequentemente, utilizados pelo sistema OPB são reais. Não foram consideradas nesta pesquisa quaisquer informações de origem simulada. Além disso, 31 das 33 variáveis tratam-se de valores históricos. As exceções dizem respeito aos dois CMOs mostrados na Tabela 7. O CMO relacionado ao PMO trata-se de uma projeção divulgada pelo ONS mensalmente (ONS, 2022b). Essa projeção normalmente é corrigida e divulgada toda semana. Foi utilizada apenas a versão publicada no início de cada mês, sem considerar as correções. O segundo CMO relativo ao dia seguinte, apesar de existir no banco de dados não é utilizado em nenhum momento no sistema OPB. Isto ocorre dado que o interesse no âmbito desta metodologia é o mapeamento de uma semana histórica (anterior ao ponto de previsão) para uma semana futura não conhecida. Ambas neste caso completas contendo os sete dias e não apenas um.

Tabela 7 – Variáveis do sistema elétrico.

Variável	**Unidade**	**Frequência**
PLD	R$	Hora
Carga	MWh/h	Hora
Custo Marginal de Operação - Programação semanal	R$	Semana
Custo Marginal de Operação - Dia seguinte	R$	Hora
Geração Hidráulica	MW médio	Hora
Geração Térmica	MW médio	Hora
Geração Eólica	MW médio	Hora
Geração Solar	MW médio	Hora
Geração Nuclear	MW médio	Hora
Energia Armazenada	GWh	Dia
Energia Natural Afluente (ENA)	MW médio	Dia
Média móvel de longo prazo da ENA	%	Dia
Intercâmbio entre subsistemas	MW médio	Hora

Fonte: O Autor.

Tabela 8 – Variáveis econômicas.

Variável	**Unidade**	**Frequência**
Vendas de óleo combustível	R$	Mês
Vendas de gás de petróleo liquefeito	R$	Mês
Vendas de diesel	R$	Mês
Produção de petróleo	Barril	Mês
Produção de gás natural	m^3	Mês
Preço do barril de petróleo tipo Brent	US$/barril	Dia
Índice Bovespa (IBOV)	Pontos	Dia
IGP-DI	Pontos	Mês
Índice de Custo de Edificações (ICE)	Pontos	Mês
Indicador de Incerteza da Economia do Brasil (IEE-BR)	Pontos	Mês
Produto Interno Bruto (PIB)	R$	Mês
Taxa de juros básica	%	Mês
Valor do câmbio (R$ para US$)	R$	Mês

Fonte: O Autor.

Tabela 9 – Variáveis climáticas.

Variável	**Unidade**	**Frequência**
Precipitação total	mm	Hora
Pressão atmosférica	mB	Semana
Radiação global	W/m²	Hora
Temperatura do ar (bulbo seco)	°C	Hora
Umidade relativa do ar	%	Hora
Direção do vento	graus	Hora
Velocidade do vento	m/s	Hora

Fonte: O Autor.

Geograficamente as leituras das variáveis do sistema elétrico são feitas separadamente para cada um dos quatro subsistemas do SIN. Esta regra possui uma exceção para os intercâmbios de energia que podem conter adicionalmente dados referentes as transferências entre o SIN e países vizinhos conectados. Em casos específicos que exijam totalizações são executados somatórios envolvendo os subsistemas de interesse (para identificação da geração total, por exemplo). Já as variáveis econômicas dizem respeito sempre a valores relativos a todo o território nacional. Por fim, as variáveis climáticas dizem respeito a leituras históricas de cada uma das estações meteorológicas de solo mantidas pelo INMET (não existem previsões climáticas). O banco de dados eventualmente pode conter valores faltantes visto que algumas das fontes em dados momentos não fornecem leituras como esperado. Isso ocorre principalmente em relação às estações meteorológicas que podem falhar por falta de energia ou manutenção. As correções de problemas como este são feitas em tempo real no módulo M2. Para detalhes adicionais sobre o banco de dados recomenda-se consultar o Apêndice A.

5.2.2 Módulo 2: Construtor de *datasets*

Conforme apresentado no diagrama da Figura 76, os dados vindos do banco de dados são no M2 manipulados em duas etapas, sendo estas o pré-processamento e a codificação dos *datasets*.

5.2.2.1 Pré-processamento

No pré-processamento são executadas sequencialmente as seguintes atividades:

a) Tratamento de integridade.

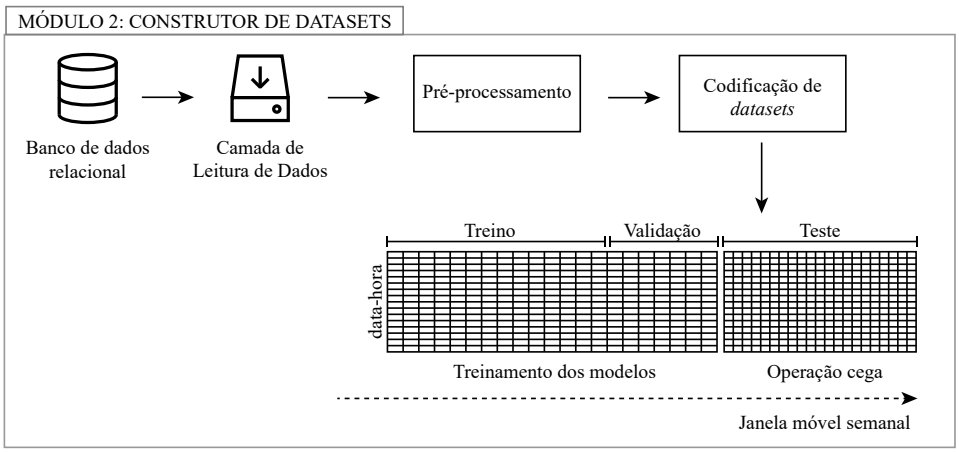
b) Tratamento de *outliers*.

c) Estabilização de variância.

d) Normalização.

e) Dimensionamento.

Figura 76 – Atividades do Módulo 2: Construtor de Datasets.



Fonte: O Autor.

5.2.2.1.1 Tratamento de integridade

Os dados extraídos das agências em muitas das vezes podem apresentar falhas. Neste sentido, a primeira etapa do pré-processamento trata-se da realização de uma conferência de integridade na qual busca-se averiguar a existência de possíveis dados faltantes para todas as horas, todas as datas e em todas as variáveis no período selecionado. Além da falta de dados, é identificado também conteúdo computacional inválido como *inf*, *null* ou *NaN*.

Na sequência os dados não íntegros detectados são substituídos de duas possíveis maneiras a depender da fonte e quantidade de dados faltantes, sendo:

a) Estações meteorológicas do INMET: Muitas das estações meteorológicas automáticas do INMET em alguns momentos passam por situações de instabilidade decorrentes de erros de *software*, de instrumentos de medição, paralização por manutenção, falta de energia ou outros. Quando isto ocorre, normalmente vê-se períodos de várias horas sem a realização de leituras. Neste sentido, no contexto da metodologia, estações que apresentam menos que 95% de confiabilidade, ou seja, que tenham mais que 5% de dados não íntegros, são descartadas do estudo. As demais que passam por este filtro tem suas leituras não íntegras substituídas por meio de um algoritmo *Nearest Neighbour Interpolation* (NNI) (CHAPRA; CANALE, 2016). Neste método determinístico para cada ponto faltante é calculada a distância euclidiana para o ponto vizinho válido mais próximo. O valor contido neste ponto vizinho é então inserido na posição onde estava o dado não íntegro.

b) Demais fontes: Ao contrário das estações do INMET que quando apresentam erros tendem a fazê-lo num intervalo sequencial de várias horas, nas demais fontes os erros detectados, de forma geral, tratam-se de poucos pontos específicos no tempo. Nesse sentido, preferiu-se utilizar estratégias individualizadas para cada variável:

- PLD: Substituição pelo valor do mesmo dia e hora da semana anterior.
- CMO: Substituição pelo valor da projeção semanal divulgada pela ONS.
- Carga, dados de geração e intercâmbio: Se detectado apenas um valor faltante, então substituição realizada com a média entre os valores das duas horas existentes imediatamente anterior e posterior ao dado faltante. Já para intervalos com 24 horas ou mais, substituição feita com valores dos mesmos dias e horas da semana anterior.
- Variáveis econômicas: Substituição realizada com último valor conhecido.

5.2.2.1.2 Tratamento de *outliers*

Outlier é qualquer observação com valor que foge ao padrão da amostra de dados. A detecção deste tipo de observação neste estudo é feita por meio da técnica *Interquartile Range* (IQR) (VINUTHA; POORNIMA; SAGAR, 2018). Neste método deve ser identificado o valor na amostra que corresponde a um dado quantil superior q_s e outro inferior q_i pré-especificados. Isto é feito em relação a mediana. A diferença entre estes dois valores é o IQR conforme demonstrado na Equação 5.1. A partir do IQR são calculados os limites L_s (superior) e L_i (inferior) conforme descrito nas equações 5.2 e 5.3 onde m é uma margem. Todos os valores que extrapolam estes limites são considerados *outliers*.

Na Figura 77 é ilustrada a aplicação do IQR em uma distribuição normal com margem de 1,5 e quantis 0,25 (inferior) e 0,75 (superior). Como pode-se notar, neste caso são considerados não *outliers* as observações contidas no intervalo entre -2,698 e +2,698 desvios padrões a partir da mediana da distribuição. Para este exemplo, 99,3% das observações não serão consideradas *outliers*. Como o método se ajusta a distribuição dos dados (devido aos valores estabelecidos pelos quantis) o percentual de observações que serão consideradas *outliers* irá mudar conforme a densidade da amostra.

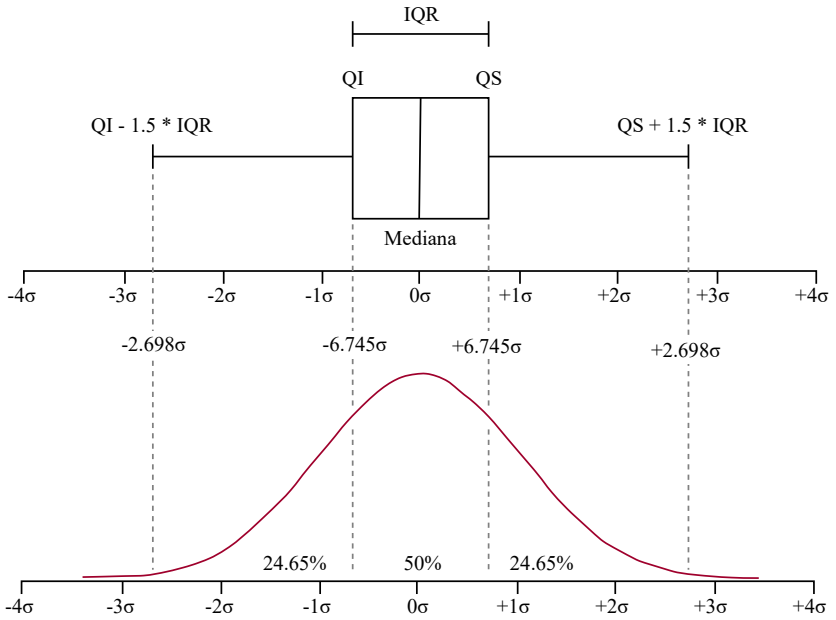
$$IQR = q_s - q_i \tag{5.1}$$

$$L_s = q_s + m * IQR \tag{5.2}$$

$$L_i = q_i - m * IQR \tag{5.3}$$

Outliers detectados são substituídos via NNI se possuírem vizinhos válidos. No caso do *outlier* não possuir vizinhos (ao estar na primeira ou última posição da série temporal, por exemplo) então a substituição ocorre conforme uma média aritmética simples dos limites do IQR, sendo esta $(L_s + L_i)/2$.

Figura 77 – IQR com quantis 0,25 e 0,75 em uma distribuição normal com margem de 1,5.



Fonte: O Autor.

5.2.2.1.3 Estabilização de variância

Mesmo com a substituição dos *outliers*, uma dada série temporal pode apresentar heterocedasticidade. A heterocedasticidade pode ser suavizada na série original e após a previsão ser realizada, então ela é adicionada de volta. O fato de realizar essa operação de forma separada do treinamento dos modelos elimina a necessidade de se aprender a inferir em qual medida a média e a variância mudam com o tempo. Isto em alguns casos pode facilitar a otimização. Na prática, a estabilização de variância é composta de um processo de conversão e, nas previsões, é aplicada a desconversão. O M2 é capaz de realizar estabilização de variância baseada nas seguintes técnicas.

a) Retornos: Comumente utilizada em métodos econométricos como os da classe ARIMA e GARCH como discutido na subseção 4.1.1. Neste caso, a desconversão dos retornos previstos para a série em escala original depende de uma soma cumulativa a partir do último preço conhecido. É uma técnica adequada para o curto prazo, como para 24 horas à frente. No entanto, quando o horizonte de previsão é maior pode ocorrer um acúmulo dos erros. Nesse caso, a série prevista tende a apresentar uma tendência de alta ou de baixa, o que prejudica a qualidade dos resultados.

b) Transformação logarítmica: Útil para homogeneizar variâncias, no entanto, tem pouca aplicação prática no contexto dos preços de eletricidade devido a alta volatilidade observada. A principal vantagem é a capacidade de corrigir distribuições assimétricas. Neste sentido, o uso combinado desta transformação seguida da aplicação de diferenças para obtenção de retornos pode amenizar a heterocedasticidade. Dessa forma, o que se observa é uma convergência para uma distribuição normal (NAMETALA *et al.*, 2016).

c) Função *Arcsine Hiperbolic* (ASINH): Uma técnica conveniente para horizontes de previsão maiores definida como na Equação 5.4 onde $p_{d,h}$ é o valor do preço p no dia d e hora h. Essa transformação, apesar de modificar a variância, mantém características estatísticas da distribuição. Por exemplo, preservando a unimodalidade. Os trabalhos de Schneider (2011), Ziel e Weron (2018) fizeram uso de ASINH discutindo vantagens desta transformação. Em Uniejewski, Weron e Ziel (2018) é comentado que esta transformação pode funcionar bem com RNAs, no entanto, depende de testes prévios. A operação oposta ao processo que visa a desconversão é $p_{d,h} = sinh\left(p_{d,h}^{novo}\right)$.

$$p_{d,h}^{novo} = asinh(p_{d,h}) \equiv log(p_{d,h} + \sqrt{p_{d,h}^2 + 1}) \tag{5.4}$$

5.2.2.1.4 Normalização

Visando controlar a assimetria e curtose das distribuições dos dados pode ser realizado o processo de normalização.

A normalização no M2 pode ser acionada para uso como demonstrado na Equação 5.5. Os termos a e b normalmente empregados em estatística são respectivamente a média do conjunto amostral e o desvio padrão. Versão chamada de transformação *z-score* (CHEADLE *et al.*, 2003). No entanto, como discutido na subseção 4.1.2, devido aos picos vistos em séries de preços de eletricidade, a mediana deve ser preferencialmente considerada. A normalização pela mediana em preços de eletricidade para uso em RNAs foi também avaliada no estudo de Uniejewski, Weron e Ziel (2018) e a superioridade do método foi constatada quando realizada aplicação prévia de estabilização de variância ASINH ou *Normal Probability Integral Transform* (NPIT).

A normalização por mediana é feita com base em um fator de escala denominado *Median Absolute Deviation* (MAD) que é dado conforme Equação 5.6. Na Equação 5.6, a é a mediana do conjunto de preços P e b é determinado segundo Equação 5.7 onde k é uma constante definida com base na distribuição dos dados. Como espera-se que com a estabilização de variância as amostras estejam aproximadamente distribuídas segundo uma normal, então deve-se usar $k = 1/z_{0,75} \approx 1{,}4856$, onde $z_{0,75}$ corresponde ao quantil de 75% de forma a estar de acordo com o intervalo adotado quando da aplicação de IQR para detecção de *outliers*.

$$p_{d,h}^{novo} = \frac{(p_{d,h} - a)}{b} \tag{5.5}$$

$$MAD = mediana(|p_{d,h} - mediana(P)|) \tag{5.6}$$

$$b = k \cdot MAD \tag{5.7}$$

5.2.2.1.5 Dimensionamento

O último passo do pré-processamento é o dimensionamento. Nesta etapa os valores são ajustados para ficarem contidos entre 0 e 1 conforme Equação 5.8. Isto é feito, pois no contexto das RNAs a magnitude de todas as variáveis processadas deve ser encarada sob mesma escala. Em outras palavras, zero deve ser definido para valores mínimos e um para valores máximos em todas as entradas contínuas. Esta etapa é também importante, pois pode ser usada para posicionar os valores dentro da região dinâmica de operação de certas funções de ativação (SILVA; SPATTI; FLAUZINO, 2016). Este dimensionamento pode ser feito identificando na série temporal os valores mínimos e máximos históricos. No entanto, isto gera um problema quando se usa janelamento móvel. Com a entrada de novos valores a cada janela, pode ocorrer de novos mínimos e máximos substituírem os antigos gerando deste modo uma distorção entre os conjuntos. Neste sentido, foram adotados valores mínimos e máximos fixos já que o intervalo de variação de cada uma das entradas usadas no sistema OPB é bem compreendido. Pode-se ilustrar com a temperatura. Neste caso não é necessário identificar valores mínimos e máximos históricos quando pode-se optar por usar um mínimo de -50 e máximo de 50. A regra adotada foi a de subtrair do mínimo histórico e somar ao máximo histórico uma quantidade referente a um terço do valor máximo em cada série identificados nos conjuntos de treino/validação. Dessa forma definindo p_{min} e p_{max} como descritos na Equação 5.8. A atualização de p_{min} e p_{max} só acontece no caso de surgir algum novo valor que extrapole um desses limites.

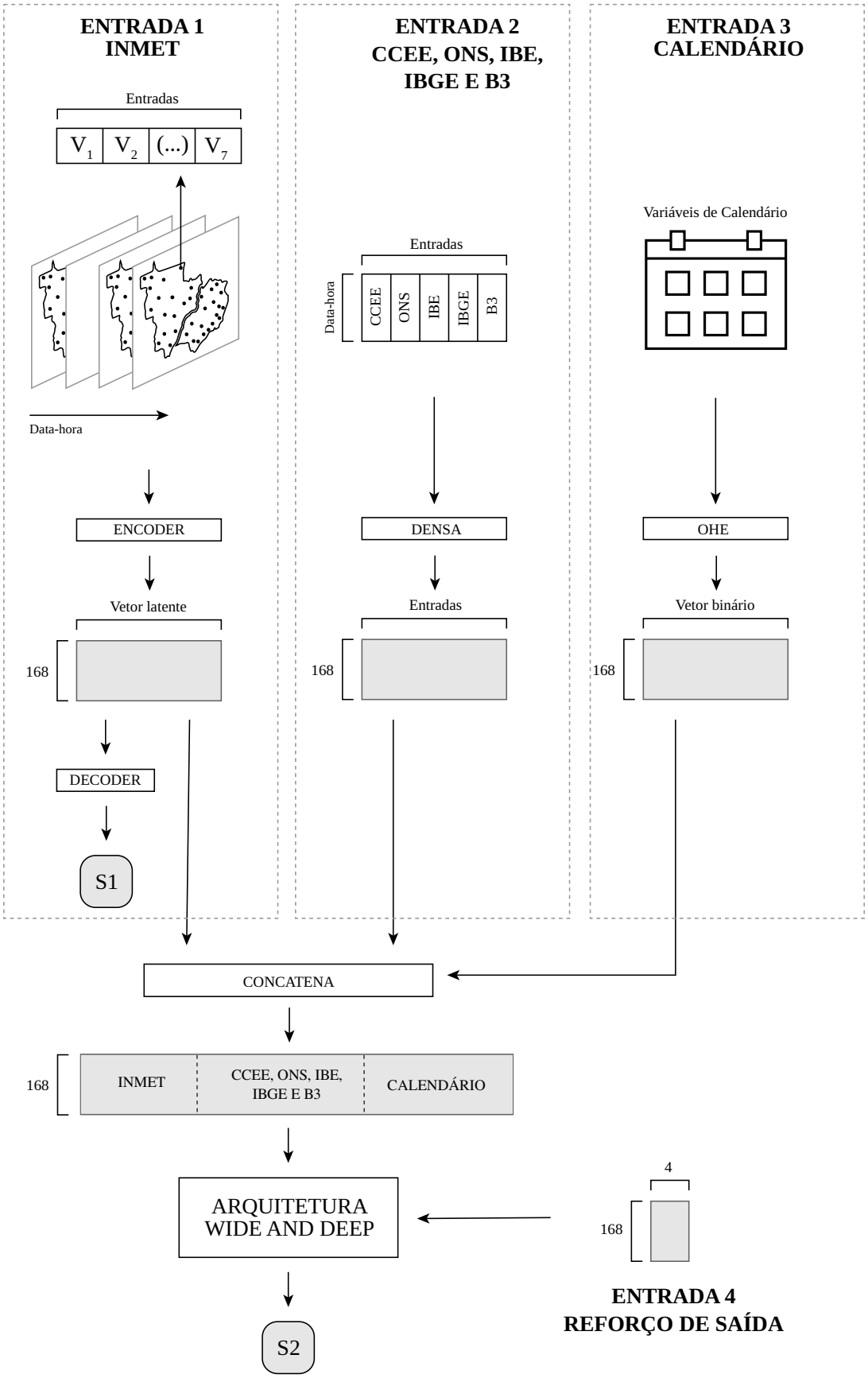
$$p_{d,h}^{novo} = \frac{(p_{d,h} - p_{min})}{(p_{max} - p_{min})} \tag{5.8}$$

5.2.2.2 Codificação

Os conjuntos de treinamento, validação e testes utilizados junto aos modelos preditores são preparados dentro do M2 seguindo procedimentos diferentes, a depender do formato como são disponibilizados. Garantido que estes dados já foram pré-processados e estão em condição íntegra, então, na próxima fase é feita uma codificação para uma representação computacional que irá viabilizar as etapas de treinamento e previsão nos módulos M3 e M4. Cabe neste ponto deixar claro que tanto o pré-processamento quanto a codificação no sistema OPB são realizadas separadamente entre os conjuntos de treinamento/validação e testes [4]. Inclusive com realização de atualização do pré-processamento a cada novo janelamento móvel requisitado ao M2 quando da operação no conjunto de testes. A arquitetura utilizada espera quatro entradas de dados. Três destas entradas são concatenadas para se tornarem uma. Já a quarta é inserida no modelo em um ponto avançado das camadas da arquitetura da RNA. A seguir é discutido o processamento de cada uma das quatro entradas e como estas alimentam o modelo. Sugere-se ao leitor que utilize o diagrama da Figura 78 para acompanhar a descrição destes procedimentos.

[4] O problema do tratamento completo do conjunto de dados é denominado *data linkage*. Como discutido no Capítulo 3 é um erro observado em diversos trabalhos que envolvem RNAs.

Figura 78 – Métodos de codificação das variáveis de entrada com base nas fontes.



Fonte: O Autor.

5.2.2.2.1 Entradas 1: Estações meteorológicas do INMET

O INMET fornece os dados climáticos das estações meteorológicas com valores distribuídos geograficamente. Neste sentido, os dados possuem duas discretizações, sendo uma relacionada ao tempo (intervalo em horas) e outra ao espaço (latitude e longitude). São utilizadas apenas as leituras das estações automáticas de solo. As seleção das estações depende do subsistema para o qual o PLD será previsto. No caso deste trabalho são de interesse as 195 estações localizadas nas regiões Sudeste e Centro-Oeste (Subsistema SECO).

Conforme visto na Tabela 9 a cada hora do dia as 195 estações do INMET selecionadas realizam a leitura de 7 variáveis. Neste sentido, em um ano são gerados aproximadamente 12 milhões de valores [5]. Este é um problema que exige técnicas de *Big Data*, visto que certos modelos (como mais adiante apresentado no Capítulo 6) são treinados em períodos de até 8 anos. Nesse sentido (devido a massiva quantidade de informação manipulada) é necessária uma estratégia de codificação que seja capaz de selecionar dados relevantes e reduzi-los a uma dimensionalidade que permita o treinamento com tempo computacional viável.

A estrutura utilizada para este propósito é o AE como apresentado na subseção 4.2.5. No M2 é adotada a arquitetura SAE. Com o SAE a matriz de dados geográficos para cada intervalo de tempo é "achatada" para um vetor unidimensional. Este vetor por sua vez é então encriptado para um vetor latente. O processo de encriptação e reconstrução da matriz original no SAE é feito via uma sequência de camadas densas. Para fins de avaliação de qualidade da encriptação do vetor latente no M2 são testados vetores com tamanhos de 64 e 128 posições envolvendo as diferentes técnicas de pré-processamento discutidas na subseção anterior. A configuração que apresenta nesta avaliação individualizada as melhores taxas de reconstrução é a escolhida. O AE é então agregado diretamente na arquitetura final do modelo preditor, já treinado. Na Figura 78 isto está representado pelas saídas $S1$ e $S2$. A saída $S1$ é a saída do AE que força o aprendizado da redução de dimensionalidade no vetor latente. Já a saída $S2$ diz respeito a previsão de fato dos 168 valores futuros do PLD. Essa abordagem arquitetural é chamada de RNA Siamesa. RNAs siamesas foram testadas pela primeira vez no contexto de redes de atraso no tempo (BROMLEY *et al.*, 1994) demonstrando bons resultados em sequências temporais, no entanto, devido à falta de poder computacional da época foram pouco utilizadas. As mesmas vêm sendo redescobertas desde 2017 (DEY *et al.*, 2017; DONG; SHEN, 2018; GHOSH *et al.*, 2021).

Por fim, considerando o horizonte de previsão adotado, então o AE para cada janela móvel deve gerar 168 vetores latentes, sendo que cada vetor condensa numericamente o que seriam as condições de clima históricas fornecidas pelo INMET para cada uma das 168 horas anteriores ao ponto de previsão. $S1$ é portanto uma matriz com as dimensões *(quantidade de estações)*$\times$*(quantidade de variáveis)* e a saída que continua no restante da arquitetura é outra matriz encriptada com dimensões de 168×64 ou 168×128.

[5] 11957400.

5.2.2.2.2 Entrada 2: CCEE, ONS, IBE, IBGE e B3

Os dados da CCEE, ONS, IBE, IBGE e B3 são séries univariadas com discretização já em formato horário conforme correção realizada no módulo M1. Neste sentido, os dados da entrada 2 são diretamente convertidos para uma matriz de dimensões $168\times$*(quantidade de variáveis)* por meio de uma única camada densa. As variáveis consideradas podem ser selecionadas livremente dentre as 13 disponíveis na Tabela 8 e as 13 da Tabela 7, com exceção apenas para CMO do dia seguinte. No janelamento móvel são utilizados os valores da semana anterior ao ponto de previsão.

5.2.2.2.3 Entrada 3: Calendário futuro

Na entrada 3 são inseridas as condições de calendário da semana de previsão. Neste sentido podem ser incluídas as seguintes:

a) Hora do dia.

b) Dia da semana.

c) Dia do mês.

d) Mês do ano.

e) Semana do ano.

f) Trimestre do ano.

g) Feriados e datas especiais.

h) Horário de verão.

Hora do dia, dias da semana e do mês, mês do ano, semana do ano e trimestre são contabilizadas com base na configuração do ano corrente. Nesse sentido, dia do mês e semana do ano podem ter valores máximos diferentes a depender do mês e de anos bissextos.

Quanto a feriados e datas especiais são considerados todos os feriados nacionais e estaduais. No caso de feriados estaduais são levados em conta apenas aqueles referentes aos estados que fazem parte do subsistema para o qual o estudo de caso é realizado. O modelo também discrimina as situações de existência ou não de horário de verão, dia normal (onde não há feriado ou data especial), dia anterior e dia posterior ao feriado, além de feriado ou dia especial que ocorra em dia da semana, sábado ou domingo. Neste sentido, é feita uma codificação numérica com valores inteiros identificadores para cada hora e data do calendário da semana de previsão com base nestas condições. Os valores inteiros são processados antes da entrada no *dataset* segundo método denominado *One Hot Encoding* (OHE).

OHE consiste em elaborar um vetor binário que indique através do *bit* 1 qual é a situação corrente para um dado momento da previsão. A título de exemplo, um vetor OHE que indique os dias da semana deve ser composto de sete posições. Para representar uma terça-feira o *bit* 1 deve estar setado na terceira posição do vetor (como em $OHE_{semana} = \{0,0,1,0,0,0,0\}$ com a semana

tendo início no domingo). Do ponto de vista da compreensão da sequência de valores, o OHE gera uma representação próxima da realidade visto que, para o exemplo dos dias da semana, caso fosse usados valores sequenciais como de 1 a 7 com domingo sendo o dia 1, então do ponto de vista numérico o valor 1 estaria distante do valor 7. Isto na prática não seria uma representação adequada dado que domingo é um dia sequente a segunda feira, enquanto que 1 não é sequente a 7. A mesma ideia se aplica a qualquer variável com padrão cíclico de calendário utilizada. Com o uso de OHE o dimensionamento entre 0 e 1 necessário a transformação dos dados também é feito naturalmente.

As variáveis de calendário podem ser escolhidas na etapa de treinamento do modelo. No caso de todas serem utilizadas então a matriz gerada possuirá a dimensão de 168×222. Sendo, uma posição para horário de verão, 24 posições para horas do dia, 7 posições para dias da semana, 31 para dias do mês, 12 para meses do ano, 53 para semanas do ano, 4 para trimestres do ano, uma para horário de verão e 90 para feriados e dias especiais. A codificação OHE, portanto, pode gerar uma matriz esparsa de zeros se muitas variáveis diferentes forem utilizadas para representar as condições de calendário ao mesmo tempo.

5.2.2.2.4 Entrada 4: Reforço de saída

A entrada 4 é fixa e diz respeito a quatro variáveis:

a) PLD no subsistema SECO na última semana anterior ao ponto de previsão.

b) Carga no subsistema SECO na última semana anterior ao ponto de previsão.

c) Geração térmica em todo o SIN na última semana anterior ao ponto de previsão.

d) Valor do CMO projetado pela ONS para a semana de previsão repetido 168 vezes.

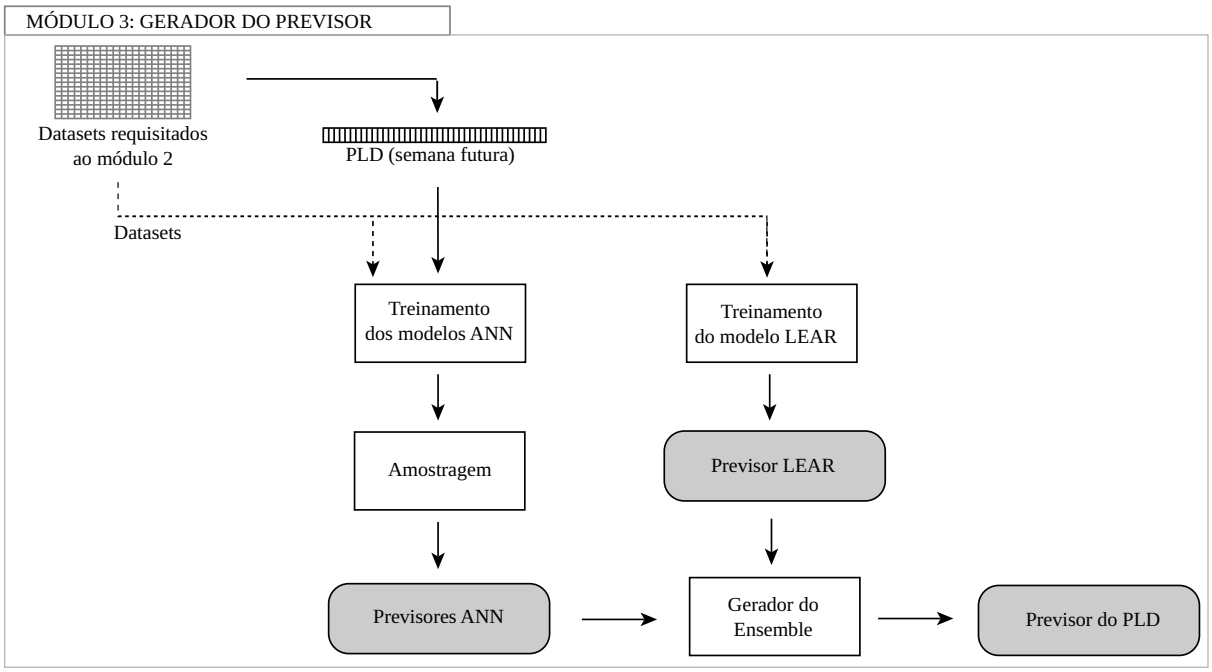
Os dados referentes a entrada 4 correspondem a uma matriz 168×4. Essa é inserida próxima as camadas de saída do modelo. Esta abordagem em modelagem profunda de RNAs é chamada de *Wide and Deep* (WAD) e foi introduzida por pesquisadores do Google em 2016 (CHENG, 2016) que a utilizaram como algoritmo de recomendação para filmes no produto Google Play (CHENG *et al.*, 2016). A ideia principal do procedimento é garantir que o modelo não dependa unicamente das células de memória para recuperar um dado padrão histórico. Visto que algumas informações passadas são facilmente recuperáveis e exercem influência direta na saída do modelo, então estas podem ser resgatadas e diretamente inseridas nas camadas próximas a camada de saída. O WAD tende a aprimorar a capacidade da rede em construir detalhes na série temporal que, devido as longas recorrências, porventura podem ser ignorados devido a suavização gerada pela regressão que ocorre nos dados (CHENG, 2016).

A última etapa na preparação dos *datasets* trata-se da divisão correta dos conjuntos de treino, validação ou testes, conforme solicitados pelos módulos M3 ou M4. Os módulos devem informar qual o tamanho da janela móvel, além das datas iniciais e finais de cada conjunto. É feito então um "fatiamento" dos dados já pré-processados e codificados separadamente gerando-se assim as janelas que são salvas para uso posterior.

5.2.3 Módulo 3: Construtor do previsor

No M3 são treinados modelos de duas classes diferentes, sendo uma delas baseada em ANNs e outra baseada em LEAR conforme observado na Figura 79.

Figura 79 – Atividades do Módulo 3: Gerador do Previsor.



Fonte: O Autor.

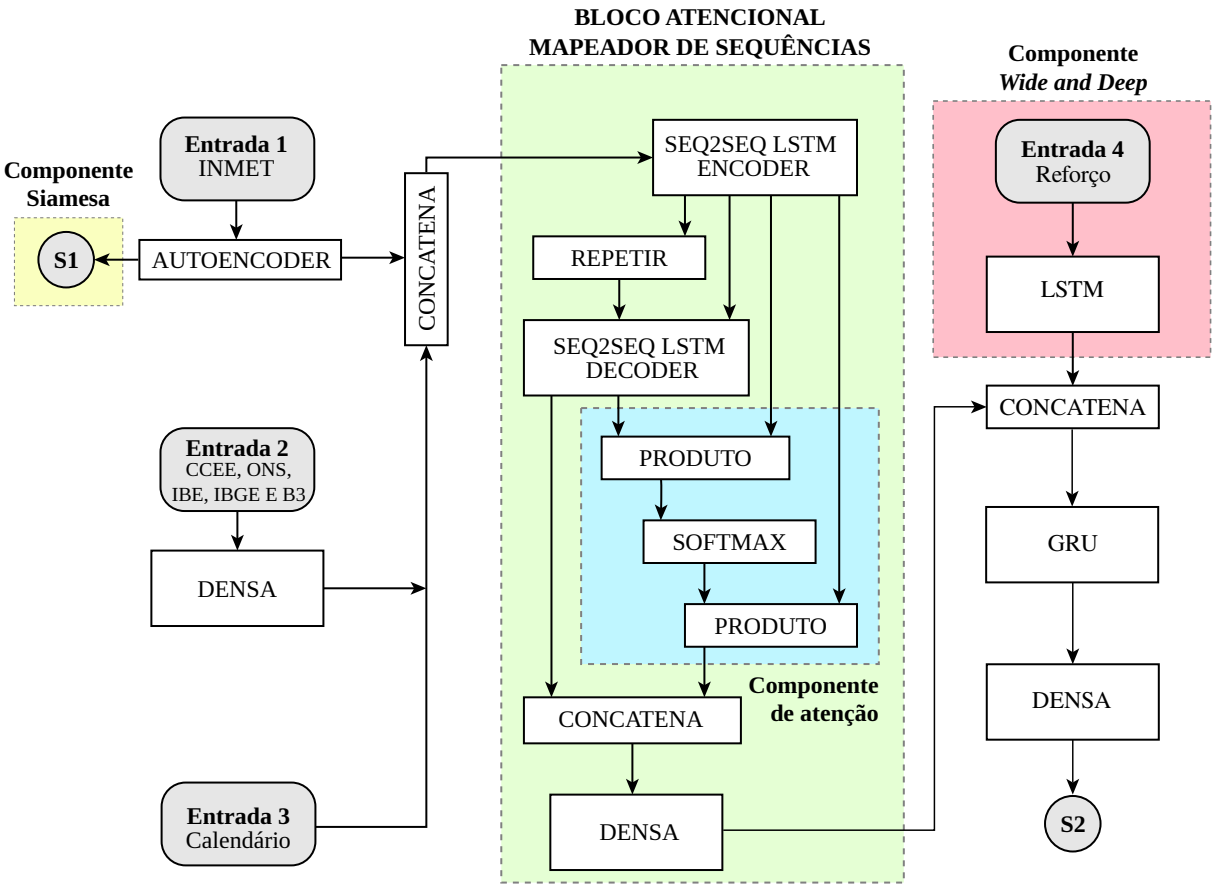
Em relação as ANNs, o treinamento trata de aplicar uma otimização à estrutura da RNA visando encontrar o conjunto de pesos ideal que combine as variáveis de entrada (codificadas conforme *datasets* gerados pelo módulo M2) separadas em grupos de janelas móveis com a finalidade de mapear corretamente os valores futuros do PLD. Estes modelos são estocásticos devido a aleatoriedade existente na inicialização dos pesos da rede. Neste sentido, mais de um modelo precisa ser treinado até que estatisticamente uma quantidade específica de modelos sejam suficientes para representar a real capacidade preditiva da ANN quando amostrados. A quantidade desses modelos a serem treinados no M3 é estimada com base nas características estatísticas observadas nos erros de previsão de uma amostra inicial contendo 30 modelos aplicados ao conjunto de validação. Essa avaliação é feita com cálculo de tamanho amostral baseado em teste T [6]. O treinamento de todos os modelos ANN é feito com inicialização de pesos que se distribuem segundo uma normal (método *Lecun*) via aplicação de um algoritmo de otimização Nadam com uso de *clipping* de gradientes conforme apresentado na subseção 4.2.6. Como a saída trata-se de uma

[6] A convergência assintótica dos erros para uma distribuição normal foi constatada com base em teste de normalidade de Shapiro-Wilk para 100, 400 e 800 modelos com $\alpha = 0{,}95$. O cálculo do tamanho amostral com teste T é *one sided* e foi configurado com $\alpha = 0{,}95$, $\delta^* = 0{,}1$, potência de $0{,}7$ e com uso do desvio padrão da amostra.

série temporal univariada, então a função de perda escolhida é o MSE. Para evitar sobreajustamento também se utiliza o método de parada precoce com monitoramento de estagnação de convergência.

A arquitetura dos modelos ANN é baseada em cinco componentes principais e pode ser observada na Figura 80. O centro da arquitetura trata-se de um bloco mapeador de sequências responsável por realizar as previsões. Este bloco é composto de uma Seq2Seq com células de memória LSTM (*encoder* e *decoder* do bloco mapeador) acopladas a um mecanismo de atenção suave como proposto por Luong, Pham e Manning (2015). A saída da função *softmax* neste bloco de atenção é o vetor de *scores* atencionais. O bloco mapeador de sequências é ainda complementado pelas componentes siamesa e WAD. Com uso da componente siamesa, o AE utilizado para processamento das variáveis climáticas gera duas saídas. Uma saída é propagada até o final da arquitetura contribuindo para a geração da previsão do PLD (*encoder* do AE), enquanto que a outra é utilizada no cálculo da função de perda que minimiza o erro na reconstrução das informações das estações meteorológicas do INMET (*decoder* do AE). Já a componente WAD trata-se de uma segunda RNA constituída de células LSTM que recebem a entrada de reforço. Esta é concatenada à saída do bloco mapeador de sequências antes de um último processamento ser realizado por meio de uma camada temporal baseada em células GRU e um ajuste para saída em formato 168×1 feito por uma camada densa.

Figura 80 – Arquitetura siamesa WAD Seq2Seq com mecanismo de atenção.



Fonte: O Autor.

Os modelos ANN são treinados no conjunto de treino e avaliados no conjunto de validação. O conjunto de validação trata-se do segmento final de tempo correspondente ao último ano de dados. A parametrização da arquitetura é estimada via otimização baseada em *grid-search* onde os parâmetros vão sendo alterados até que a melhor configuração dentre todas as possíveis seja encontrada. A configuração encontrada via *grid-search* é usada igualmente em todos os modelos e com esta fixada, os mesmos são então retreinados nos conjuntos de treino e validação, desta vez concatenados em um único conjunto. Importante ressaltar que o conjunto de testes nesta fase nem sequer é considerado. O mesmo fica separado da base de dados gerada pelo módulo M2. O mesmo não é usado para fins de pré-processamento, codificação ou qualquer forma de treinamento para evitar enviesamento dos resultados. Por fim, é gerada a quantidade de modelos indicada conforme resposta do cálculo de tamanho amostral [7].

Já o modelo baseado em LEAR trata-se de um regressor estatístico linear com quantidade de parâmetros otimizada com uso de regularizador baseado em LASSO [8] conforme apresentado na subseção 3.3.1.1. O regressor é estimado no mesmo conjunto de treino e validação concatenado no qual as ANNs são treinadas, inclusive levando em conta os dados pré-processados pelo módulo M2 compostos exatamente das mesmas variáveis selecionadas para a entrada na RNA (inclusive com os vetores latentes gerados pelo AE). O LEAR é determinístico, portanto, apenas um único modelo é gerado.

Com todos os modelos prontos é criada uma função específica que permite o acesso a geração das previsões do PLD feita por estes. Na prática, esta função é considerada o "Previsor do PLD". A mesma pode ser instanciada para geração de previsões pontuais ou probabilísticas:

a) Caso seja requisitada uma previsão pontual: A função calcula inicialmente uma média aritmética de todas as previsões dos modelos baseados em ANN. Na sequência é calculada uma nova média entre a previsão do modelo LEAR e a média dos modelos ANN. Consequentemente, a previsão pontual é um *ensemble* baseado em média simples que retorna um único valor de PLD por hora.

b) Caso seja requisitada uma previsão probabilística: A função realiza uma média aritmética entre cada um dos modelos ANN com o modelo LEAR. Esse conjunto de valores é retornado sendo que possuirá tantos valores de PLD por hora quanto modelos ANN gerados.

A saída do M3, por fim, trata-se de um previsor que recebe uma entrada multivariada e gera uma saída univariada para uma semana à frente em resolução horária.

[7] Para detalhes adicionais sobre a especificação das arquiteturas AE e ANN sugere-se consultar o Apêndice B.

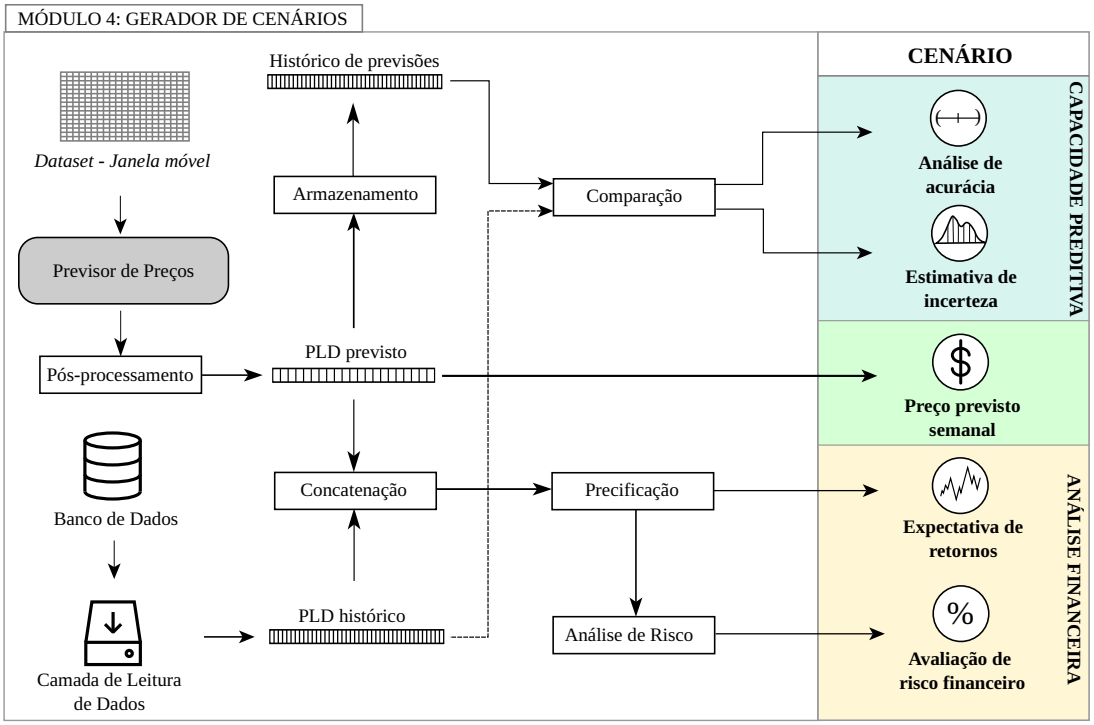
[8] O LASSO foi configurado para atuar com norma mista L1 e L2 em modo multi-tarefas devido ao formato multivariado das entradas. Foi utilizada uma quantidade máxima de iterações igual a 1000, tolerância de $1e^{-4}$ e cinco conjuntos para determinação da parametrização usando validação cruzada.

5.2.4 Módulo 4: Gerador de cenários

Considerando que o previsor do PLD treinado no módulo M3 está pronto para uso e que um determinado conjunto de dados pode ser requisitado com devido pré-processamento ao módulo M2, então aciona-se o módulo M4 para construção de um cenário conforme diagramado na Figura 81.

A construção de um cenário se dá baseada em uma janela que é requisitada ao módulo M2. Para fins de pré-processamento são considerados todos os dados disponíveis e anteriores ao ponto de previsão. Contudo, após este pré-processamento realizado entra no previsor do PLD como *dataset* apenas o último segmento com dimensão de *(168 horas)*$\times$*(quantidade de variáveis)*. Esta é a janela móvel. Cabe ressaltar que apesar desta janela conter em sua maioria dados históricos (como exposto nas subseções 5.2.1 e 5.2.2) os dados de CMO e condições de calendário tratam-se de valores esperados para a semana seguinte. Outro ponto é que como os dados são convertidos no módulo M2 para uma escala diferente da original para serem usados no treino dos modelos, quando a previsão é realizada os valores são gerados também nesta escala diferente. Sendo assim, no módulo M4 é necessária a realização de uma desconversão dos valores em acordo com a técnica de conversão selecionada no módulo M2. Esta etapa é chamada de pós-processamento. O PLD previsto desconvertido para reais (R$/MWh) ainda é armazenado em um histórico de previsões e, também, concatenado aos dados reais do PLD histórico. O PLD previsto para semana seguinte, o histórico de previsões e a série concatenada *[PLD real histórico, PLD previsto para semana à frente]* são então utilizados na produção de quatro indicadores. Estes divididos em duas categorias de análise discutidas a seguir.

Figura 81 – Atividades do Módulo 4: Gerador de Cenários.



Fonte: O Autor.

5.2.4.1 Capacidade preditiva

A capacidade preditiva do sistema OPB é avaliada com base na comparação do histórico das previsões realizadas com o PLD real no mesmo período. Essa comparação é feita em termos de acurácia e também de incerteza.

A análise da acurácia é feita utilizando-se as métricas sMAPE, rMAE, R^2 e POCID. Estas podem ser calculadas para períodos específicos de tempo onde a previsão já tenha sido realizada. Opcionalmente, pode-se também incluir resultados de outros modelos de previsão para fins de comparação. Estão inclusos no sistema OPB os seguintes *benchmarks*:

- Simulação de contrato *flat*: Considera que o PLD futuro terá o mesmo valor do mesmo mês (*flat* mensal) ou mesmo trimestre (*flat* trimestral) do ano anterior.
- Modelos de DL puros: RNAs contendo quatro camadas e treinadas com parametrização definida conforme metodologia discutida na subseção 5.2.3. As camadas de entrada e saída são as mesmas usadas na arquitetura ANN do sistema OPB, contudo as duas camadas ocultas são ambas de um mesmo tipo [9]. São consideradas:
  - DNN: RNA com camadas ocultas do tipo densa (MLP profunda).
  - LSTM: RNA com camadas ocultas contendo células de memória do tipo LSTM.
  - GRU: RNA com camadas ocultas contendo células de memória do tipo GRU.
- Modelos econométricos:
  - *Naive* (Controle): É a previsão básica usada como grupo-controle. Diz respeito ao método do dia similar (ONS, 2019) que considera que o valor da semana seguinte será igual ao da última semana conhecida.
  - SARIMA: SARIMA é a versão do ARIMA que considera sazonalidade como discutido na subseção 4.1.1. É utilizado um modelo otimizado no período de interesse por meio de minimização do *Akaike Information Criteria* (AIC) com avaliação das quantidades necessárias dos termos p, q, P e Q variando entre 0 e 5 (cada) e de diferenças d e D entre 0 e 3.
- Modelos de ML baseados em *ensembles*: Estes modelos consistem do uso de diversos algoritmos diferentes de ML que tem seus resultados agregados via aplicação de alguma técnica baseada em árvores de decisão ou outras a depender do objetivo relacionado ao problema. De forma geral, o método identifica pontos do *dataset* onde um ou outro algoritmo supera os demais e pondera cada um dos algoritmos criando uma combinação dos melhores para conjuntos específicos de dados (KONSTANTINOV; UTKIN, 2021). A maioria destes

[9] As especificações detalhadas destas RNAs podem ser encontradas no material suplementar da tese em Nametala (2022).

algoritmos são recentes e alguns incluídos nestes modelos são *RandomForest*, *ExtraTree*, *AdaBoost*, *GradientBoosting*, *Bagging* e outros. Como apresentado nas seções 3.1 e 3.3.1.2, muitos trabalhos vem utilizando estes métodos que são considerados parte do estado da arte em diversos campos de aplicação, incluindo regressão em séries temporais (CAMPAGNER; CIUCCI; CABITZA, 2023). São considerados dois *benchmarks* deste tipo:

  - CatBoost: CatBoost é uma biblioteca livre mantida pela empresa Yandex definida como "um algoritmo de *gradient boosting* baseado em árvores de decisão (...) usado para pesquisa, sistemas de recomendação, assistente pessoal, carros autônomos, previsão do tempo e outras tarefas". CatBoost é utilizado atualmente em empresas e instituições como o CERN, Cloudflare e Careem taxi (YANDEX, 2022) [10].
  - XGBoost: XGBoost é também uma biblioteca livre inicialmente desenvolvida por Tianqui Chen e que hoje está disponível nos principais *frameworks* de ML em linguagens como Python, R e Java. (CAMPAGNER; CIUCCI; CABITZA, 2023).

- Modelos isolados:
  - Arquitetura ANN isolada: Arquitetura ANN usada no sistema OPB isoladamente, sem considerar o *ensemble*. Este *benchmark* é avaliado segundo a melhor ANN, a pior ANN e a média de todas as ANNs.
  - Modelo LEAR isolado: LEAR usado no sistema OPB isoladamente, sem considerar o *ensemble*.
  - Casos do sistema OPB: Melhor e pior caso considerando as previsões probabilísticas.

Além da comparação usando as métricas de acurácia pode-se gerar gráficos que comparam as previsões do sistema OPB com os dados reais diretamente, incluindo a identificação dos melhores e piores períodos dentro histórico de previsões.

Quanto à estimativa da incerteza são consideradas duas abordagens, sendo (i) o percentual de permanência do preço real dentro de um intervalo de confiança calculado a partir das previsões probabilísticas e (ii) o percentual de permanência do preço pontual predito dentro de um intervalo definido segundo bandas de *bollinger* calculadas a partir do preço real.

O intervalo de confiança é estimado pelo método do percentil ao levar em consideração a distribuição dos valores previstos para uma dada hora i e os limites definidos conforme as equações 5.9 e 5.10. O cálculo é dado conforme Equação 5.11 onde um dado limite é subtraído (caso do limite inferior) ou acrescido (caso do limite superior) do produto entre o desvio padrão σ do conjunto de limites (inferior ou superior) e uma margem m. Para o intervalo de confiança foi utilizado $\alpha = 0{,}95$ e para a margem um $m = 1{,}5$.

[10] CatBoost foi configurado com 500 iterações, taxa de aprendizagem de $0{,}1$ e três níveis de profundidade máxima da árvore. O modelo é otimizado junto a função de erro RMSE.

Já os limites inferior e superior das bandas de *bollinger* são calculados conforme Equação 5.12 para cada hora i com base no valor do desvio padrão σ subtraído ou acrescido de uma Média Móvel Simples (MMS) entre dois intervalos de tempo separados por uma distancia j. O tamanho da banda foi definido em duas vezes o desvio padrão ($d=2$) e a MMS possui 48 períodos ($j=48$).

$$lim_{inf} = \frac{100 \cdot (1-\alpha)}{2} \tag{5.9}$$

$$lim_{sup} = \frac{100 \cdot (1-(1-\alpha))}{2} \tag{5.10}$$

$$IC_i = lim_i \pm \sigma_{lim_i} \cdot m \tag{5.11}$$

$$BB_i = MMS_{[i-j,i]} \pm d \cdot \sigma_{[i-j,i]} \tag{5.12}$$

Os valores dos parâmetros utilizados no cálculo do intervalo de confiança e das bandas de *bollinger* são, respectivamente, os encontrados como padrão em pacotes que fornecem os mesmos como funções em linguagens de programação como na função *percentile* da biblioteca Numpy (NUMFOCUS, 2022) ou em ferramentas de *home broker* para negociação em bolsas de valores como o MetaTrader (METAQUOTES, 2022).

5.2.4.2 Análise financeira

Conforme visto na seção 2.3, contratos de eletricidade possuem duas variáveis principais: A carga programada para ser consumida no período em que o contrato vigorar e um preço previsto a ser pago por esta carga. A regulação destas duas variáveis dependem de acordos visando estabelecer regras no caso da existência de consumos excedentes ou no caso do preço projetado estar distante do preço real. Outro ponto é que sendo o contrato um instrumento de compromisso, o uso do mesmo apresenta vantagens para ambas as contrapartes, pois (i) os preços contratuais tendem a ser mais vantajosos para o consumidor que tem espaço para negociar o pagamento de valores inferiores ao valor real (evitando exposição) e (ii) para o produtor que aceita esses preços menores visto que adiciona à sua base de geração um cliente permanente que garante segurança de receitas futuras. Neste sentido, a existência de um intervalo de variação se mostra um mecanismo válido que pode atender ao consumidor e ao produtor, pois clientes que desejam estabelecer contrato com um produtor vão obter vantagem caso sejam capazes de negociar bons custos de aquisição da energia em relação ao preço real. Mas ao mesmo tempo, para garantir estes valores os mesmos devem estar dispostos a se proteger de eventuais erros de previsão pagando por um intervalo de variação que contemple momentos onde a carga executada seja maior ou menor do que a programada ou o preço real fique fora de uma margem inferior ou superior em relação ao preço previsto. Neste raciocínio, o modelo deve regular o

risco que um produtor assume ao precificar o custo do intervalo de variação que um consumidor por ventura venha a requisitar. Por consequência, um consumidor que tenha capacidade de programar corretamente a sua carga, prever bem preços futuros e negociar contratos, pode se dar ao luxo de não pagar por grandes intervalos minimizando desta forma seus custos de aquisição. Ao passo que para o produtor torna-se vantajoso ter em contrato este tipo de cláusula por dois motivos principais. O primeiro se deve ao fato de que clientes com programações acuradas geram menores custos relacionados a mudanças no cronograma de geração. Já o segundo diz respeito a cobertura financeira recebida quando previsões mal feitas passam a ser custeadas pelos próprios clientes. Em qualquer dos dois pontos de vista, seja do consumidor ou do produtor, neste contexto o modelo de precificação cria regras objetivas que valoram o risco assumido pelas partes conforme contrato. Uma forma de definir estas regras está descrita nos algoritmos 1 e 2 e é uma das contribuições deste trabalho.

Neste ponto vale ressaltar que este modelo de precificação é hipotético e serve apenas como base para definir uma espécie de valoração em termos financeiros das previsões executadas. É sabido que diversos modelos de precificação são utilizados de formas diversas em mercados ao redor do mundo e que existe vasta literatura especializada sobre o tema. Não sendo este o ponto principal do que trata esta pesquisa, cabe deixar claro ao leitor que os algoritmos apresentados a seguir não refletem o que existe hoje no mercado brasileiro e, para seus devidos fins, é empregado como um adendo a este protocolo experimental, sendo desta forma apenas um ponto de partida para futuros estudos que queiram envolver formalmente a etapa de precificação.

O Algoritmo 1 trata da valoração do custo da carga consumida por hora com projeções de carga e preço para uma semana à frente. O mesmo recebe como entrada as seguintes variáveis:

- PLD previsto: Valor de PLD previsto pelo consumidor via algum mecanismo adotado como uso de preços históricos, preço fixo estabelecido em contrato para períodos específicos ou àqueles produzidos via modelos previsores como o sistema OPB e outros.

- PLD real: O PLD à vista divulgado pela CCEE.

- Carga programada: Valor da carga programada para ser consumida por meio de uso do perfil de carga histórico ou projeções de consumo geradas por modelos de previsão de carga.

- Carga executada: A carga efetivamente consumida.

- Taxa de penalização por erro do PLD previsto (TPP): Uma taxa a ser paga pelo consumidor ao produtor caso o PLD real extrapole os limites definidos em um intervalo de variação percentual calculado a partir do PLD previsto.

- Taxa de penalização por erro da carga programada (TPC): Uma taxa a ser paga pelo consumidor ao produtor caso a carga executada extrapole os limites definidos em um intervalo de variação percentual calculado a partir da carga programada.

- Fator percentual de negociação do PLD previsto (FPNPP): Percentual fixo do PLD previsto a ser pago definido em contrato.
- Fator percentual de negociação do PLD real (FPNPR): Percentual fixo do PLD real a ser pago definido em contrato.

A primeira etapa (linhas 20 e 21) trata de ajustar os valores previstos e reais do PLD para serem usados no decorrer do algoritmo. Este procedimento personifica por meio das variáveis FPNPP e FPNPR a capacidade de negociar que uma determinada parte eventualmente possa ter sobre a outra. O interesse do produtor (visando máximo lucro) seria que o consumidor pagasse sempre o valor cheio, neste caso ambas as variáveis teriam valor igual a um. No entanto, se isto acontecesse seria de se esperar que o consumidor buscasse valores percentuais mais baixos com outros produtores. Outro ponto é que em cenários envolvendo contratos mais longos, pode ser que o produtor aceite receber percentuais mais baixos. Esse mecanismo, portanto, estabelece uma forma de quantificar a competitividade onde consumidores dispostos a aceitar valores de FPNPP e FPNPR mais próximos de um e produtores dispostos a aceitar valores de FPNPP e FPNPR mais próximos de zero, sejam aqueles preferidos no mercado.

Entre as linhas 23 e 56 estão descritos quatro casos baseados nas possibilidades da carga executada e do PLD real estarem dentro ou fora dos intervalos calculados a partir das previsões de carga e PLD feitas pelo consumidor (o custo do intervalo é discutido mais adiante). São eles:

a) Carga dentro e PLD dentro (linhas 23 a 26): Neste caso como não existe nenhuma extrapolação dos limites, então não existem aplicações de taxas e o custo da unidade de carga consumida é igual ao valor do PLD previsto ajustado.

b) Carga dentro e PLD fora (linhas 27 a 30): Nesta situação onde existe uma extrapolação de limite relacionada ao PLD, a carga consumida é paga ao valor do PLD real ajustado acrescido de uma taxa relativa ao erro da previsão do preço.

c) Carga fora e PLD dentro (linhas 31 a 43):

   – Caso a carga executada supere o programado (linhas 32 a 38): Neste caso a carga programada é paga ao valor do PLD previsto ajustado. No entanto, a carga excedente terá o valor do PLD real ajustado acrescido de uma taxa relativa ao erro da previsão da carga.

   – Caso a carga executada seja inferior ao programado (linhas 39 a 42): Neste caso, o valor da carga será o PLD previsto ajustado acrescido da taxa relativa ao erro da previsão da carga.

d) Carga fora e PLD fora (linhas 44 a 56):

   – Caso a carga executada supere o programado (linhas 45 a 51): Nesta situação toda a carga é valorada ao custo do PLD real ajustado. Além disso, as taxas relativas a

extrapolação de limites do PLD e da carga são, respectivamente, somadas a carga programada e a carga excedente separadamente.

– Caso a carga programada seja inferior ao programado (linhas 52 a 55): Se a carga consumida for inferior ao programado, então toda a carga terá o custo do PLD real ajustado acrescido de ambas as taxas.

O custo final da eletricidade em um dado período trata do somatório de custos gerado pela aplicação do Algoritmo 1 a cada um dos valores horários de carga consumida, PLD previsto, PLD real e constantes que devem estar estabelecidas em contrato sendo estas as taxas por erro de previsão (TPP e TPC) e os fatores de negociação (FPNPP e FPNPR).

O Algoritmo 2 incorpora a regra que define o custo dos intervalos nos quais podem variar o PLD real e a carga executada. Como visto no Algoritmo 1, com o uso de intervalos pequenos se aumenta a necessidade de confiabilidade no modelo preditor adotado, pois preditores ruins ao gerar valores distantes da realidade irão fazer com que a exposição ao PLD real ajustado aumente e também que mais taxas sejam aplicadas. Portando, vale a pena para o consumidor pagar por margens de variação maiores que garantam maior proteção a eventuais flutuações. Contudo, dentro deste modelo de precificação o custo da unidade percentual de margem tem crescimento exponencial. O crescimento exponencial é justificado, pois do ponto de vista do produtor o risco de vender ao PLD predito ajustado também precisa ser levado em conta. São entradas do Algoritmo 2:

- Valor total pago pela carga consumida (VT): Somatório de todos os custos horários gerados pela aplicação do Algoritmo 1 em um dado período de tempo.
- Intervalo percentual de variação do PLD previsto (IVP): Valor percentual relativo ao tamanho da margem calculada sob o PLD previsto na qual o PLD real pode flutuar.
- Intervalo percentual de variação da carga programada (IVC): Valor percentual relativo ao tamanho da margem calculada sob a carga programada na qual a carga executada pode flutuar.
- Custo exponencial do intervalo de variação do PLD previsto (xp): Expoente aplicado ao IVP que define quão mais caro fica o intervalo do PLD conforme o seu tamanho aumenta.
- Custo exponencial do intervalo de variação da carga programada (xc): Expoente aplicado ao IVC que define quão mais caro fica o intervalo da carga conforme o seu tamanho aumenta.
- Variável de atenuação (VAT): Variável de ajuste que pode ser parametrizada para controle numérico do comportamento do algoritmo promovendo maior ou menor impacto no custo do intervalo a depender da magnitude dos valores de carga.

Inicialmente o custo total dos intervalos é setado como zero na linha 10. Na sequência são calculados nas linhas 12 e 13 o custo do uso do intervalo para cada unidade, respectivamente, de PLD e de carga. Nas linhas 14 e 15 já com o custo unitário definido, são feitos separadamente os

Algoritmo 1 Custo da carga por hora

```
Entradas:
PLD previsto
PLD real
Carga programada
Carga executada
Taxa de penalização por erro da carga programada(TPC)
Taxa de penalização por erro do PLD previsto (TPP)
Fator percentual de negociação do PLD previsto (FPNPP)
Fator percentual de negociação do PLD real (FPNPR)

INÍCIO
Preço a pagar (P) ← 0
Preço a pagar pela carga programada (PPRO) ← 0
Preço a pagar pela carga excedente (PEXC) ← 0

Custo da carga consumida (CCC) ← 0
Custo da carga programada (CPRO) ← 0
Custo da carga excedente (CEXC) ← 0

PLD previsto ajustado (PP) ← PLD previsto · FPNPP
PLD real ajustado (PR) ← PLD real · FPNPR

Se Carga dentro e PLD dentro então
    P ← PP
    CCC ← Carga executada · P
Fim Se
Se Carga dentro e PLD fora então
    P ← PR + TPP
    CCC ← Carga executada · P
Fim Se
Se Carga fora e PLD dentro então
    Se Carga acima do programado então
        PPRO ← PP
        PEXC ← PR + TPC
        CPRO ← Carga programada · PPRO
        CEXC ← (Carga executada − Carga programada) · PEXC
        CCC ← CPRO + CEXC
    Fim Se
    Se Carga abaixo do programadado então
        P ← PP + TPC
        CCC ← Carga executada · P
    Fim Se
Fim Se
Se Carga fora e PLD fora então
    Se Carga acima do programado então
        PPRO ← PR + TPP
        PEXC ← PR + TPC
        CPRO ← Carga programada · PPRO
        CEXC ← (Carga executada − Carga programada) · PEXC
        CCC ← CPRO + CEXC
    Fim Se
    Se Carga abaixo do programadado então
        P ← PR + TPP + TPC
        CCC ← Carga executada · P
    Fim Se
Fim Se

Retorna CCC

FIM
```

Algoritmo 2 Custo dos intervalos de variação

1: **Entradas:**
2: Valor total pago pela carga consumida (VT)
3: Intervalo percentual de variação do PLD previsto (IVP)
4: Intervalo percentual de variação da carga programada (IVC)
5: Custo exponencial do intervalo de variação do PLD previsto (xp)
6: Custo exponencial do intervalo de variação da carga programada (xc)
7: Variável de atenuação (VAT)
8:
9: **INÍCIO**
10: Custo dos intervalos de variação (CIV) $\leftarrow$ 0
11:
12: Custo percentual por unidade de variação da previsão do PLD (CPUP) $\leftarrow$ $(\text{IVP}/\text{VAT})^{xp}$
13: Custo percentual por unidade de variação da programação da carga (CPUC) $\leftarrow$ $(\text{IVC}/\text{VAT})^{xc}$
14: Custo do intervalo de variação do PLD previsto (CIP) $\leftarrow$ CPUP $\cdot$ VT
15: Custo do intervalo de variação da carga programada (CIC) $\leftarrow$ CPUC $\cdot$ VT
16: CIV $\leftarrow$ CIP $+$ CIC
17:
18: **Retorna** CIV
19:
20: **FIM**

produtos entre estes custos e o somatório pago pela carga consumida definido de hora em hora no Algoritmo 1. Na linha 16 os custo dos intervalos de preço e carga são somados em um único. Este custo final é o preço para os tamanhos de margem em cada intervalo escolhidos pelo consumidor e estabelecidos em contrato.

O custo dos intervalos (CIV) gerado como saída no Algoritmo 2 deve ser somado ao total pago pela carga consumida (VT) que diz respeito ao somatório das saídas horárias (CCC) calculadas no Algoritmo 1. Desta forma, o custo final que o consumidor deve pagar ao produtor é dado por $CIV + \sum_{i=1}^{t} CCC_i$ onde t é o total de horas. Conforme se aumenta a margem adquirida, se aumenta também exponencialmente o custo do intervalo de forma que em um dado limite o preço da sua aquisição deixa fazer sentido, visto que não compensaria mais a proteção. O que se nota é que existe um *trade-off* entre a exposição ao risco e o custo da proteção comprada. Bons preditores nesta situação são cruciais, pois além de dispensarem a compra de intervalos grandes, mesmo em situações de maior exposição erram menos e por isso geram menos taxas. O caso ótimo dentro deste modelo de precificação é, portanto, aquele que contém os tamanhos de intervalo ideais para PLD e carga quando se aplica o melhor preditor gerando-se desta forma os menores custos possíveis.

Este modelo de precificação ao ser acionado no módulo M4 gera uma série temporal que contém os custos de aquisição hora a hora desde o momento no qual o sistema OPB começou a ser utilizado. Esta série chamada de retornos cumulativos pode ser também calculada para cada um dos *benchmarks* mostrados na subseção 5.2.4.1. Dessa forma é possível gerar um estudo no qual testam-se vários tamanhos diferentes de intervalo para PLD e carga a fim de se identificar a combinação na qual os custos com proteção e exposição se balanceiam. Finalmente, identificados os melhores tamanhos de intervalo, então são realizados os cálculos relativos ao VaR e CVaR conforme apresentado na seção 4.3. Com o VaR e CVaR (para o sistema OPB e *benchmarks*) são estimadas os custos prováveis levando-se em conta o histórico de previsões.

A última componente de cenário gerada pelo módulo M4 é a análise estatística de diferen-

ças entre os custos cumulativos gerados via sistema OPB frente a todos os *benchmarks* quando da aplicação do caso ideal de proteção ao risco. Nesta análise, inicialmente é feito um teste de Kruskal-Wallis para indicação de existência de diferenças entre grupos. Posteriormente, caso a hipótese nula do teste seja rejeitada, então aplica-se o teste *posthoc* de Nemenyi para aferição da magnitude das diferenças[11]. São gerados dois estudos, sendo uma matriz de significância estatística e um gráfico de comparação de intervalos de confiança. Desta forma é possível averiguar se, estatisticamente, existem evidências de que um previsor foi superior ao outro do ponto de vista financeiro.

[11] Dado que os conjuntos amostrais usados são independentes, de tamanho igual e não se distribuem segundo uma normal (confirmado via teste de Shapiro-Wilk) então optou-se pelos testes não paramétricos indicados. Ambos os testes foram configurados com nível de significância $\alpha = 0{,}05$.

6 RESULTADOS

Este capítulo está dividido em três subseções. Na primeira subseção é apresentado o protocolo experimental. São expostas informações sobre os períodos de tempo correspondentes a cada conjunto de dados, características destes períodos e a dinâmica de avaliação. Além disso, são abordados os testes prévios aplicados visando identificar o conjunto de variáveis de entradas e os métodos de pré-processamento a serem usados junto ao módulo M2. Na segunda subseção é feita uma análise individualizada dos resultados apresentados pelo AE na redução de dimensionalidade das variáveis climáticas. São discutidos também alguns dos efeitos que cada uma das etapas tiveram na seleção das estações de solo do INMET e nas variáveis que estas provêm. A última subseção trata especificadamente das previsões do PLD. O sistema OPB durante esta subseção é comparado com 16 *benchmarks* e analisado individualmente. Estas análises se dão numa primeira discussão do ponto de vista de acurácia e, num segundo momento, do ponto de vista financeiro.

6.1 Protocolo experimental

Os resultados tratados a seguir dizem respeito a previsão do PLD no subsistema SECO levando-se em consideração o período que vai de 07/01/2013 (primeira segunda-feira de 2013) a 26/12/2021 (último domingo de 2021) (9 anos). Dentro deste, o intervalo que vai de 07/01/2013 a 05/01/2020 (primeiros 7 anos) corresponde aos dados utilizados para treino dos modelos. Neste conjunto, todo o período de 7 anos foi utilizado no treinamento do modelo AE. Já para o modelo Seq2Seq foram utilizados 4 anos, iniciando em 02/01/2017. Os dados sequentes indo de 06/01/2020 a 03/01/2021 (1 ano) foram utilizados para validação. O restante, referente a 2021 (04/01/2021 a 26/12/2021 - 1 ano) diz respeito ao conjunto fora da amostra, ou seja, dados que nunca foram apresentados ou utilizados de qualquer forma em nenhuma das fases de treino, validação, otimização ou demais ajustes do sistema OPB ou *benchmarks*. 2021 é o ano utilizado para a chamada "operação cega" do sistema OPB, como se de fato estivesse sendo aplicado ao contexto real. Os conjuntos de treino e validação agregam juntos 61296 horas (2554 dias) sendo que o último segmento de 7968 horas em 2020 (332 dias) é o de validação. O conjunto de testes possui 8544 horas dispostas em 356 dias. Treino, validação e testes, respectivamente, correspondem as parcelas de 76,36%, 11.41% e 12.23% de todo o *dataset* conforme descrito na Tabela 10.

Quanto a estes períodos, como pode ser observado na Figura 16 comentada na seção 2.3, o intervalo que compreende os conjuntos de treino, validação e testes apresentam momentos importantes nos últimos 10 anos da linha histórica do mercado de energia elétrica do Brasil. No conjunto de treino, além de incrementos na capacidade instalada feita via entrada em operação de duas termoelétricas, segundo bipolo do Madeira, parque eólico de Geribatu e UHE de Jirau, foi também neste período que se estabeleceu a crise hídrica nacional de 2015 que gerou a primeira necessidade de revisão do teto máximo do PLD. Já em 2019 houve o acionamento do segundo bipolo de Belo Monte e, adicionalmente, observou-se crescimento da penetração de renováveis, principalmente

Tabela 10 – Divisão dos conjuntos de dados. As datas de início tratam-se sempre da primeira segunda-feira no ano e a data de fim de um domingo.

Conjunto	Início	Fim	Horas	Dias	Parcela
Completo	07/01/2013	26/12/2021	69840	2910	100.00%
Treino e Validação	07/01/2013	03/01/2021	61296	2554	87.77%
Treino	07/01/2013	05/01/2020	53328	2222	76.36%
Validação	06/01/2020	03/01/2021	7968	332	11.41%
Teste	04/01/2021	26/12/2021	8544	356	12.23%

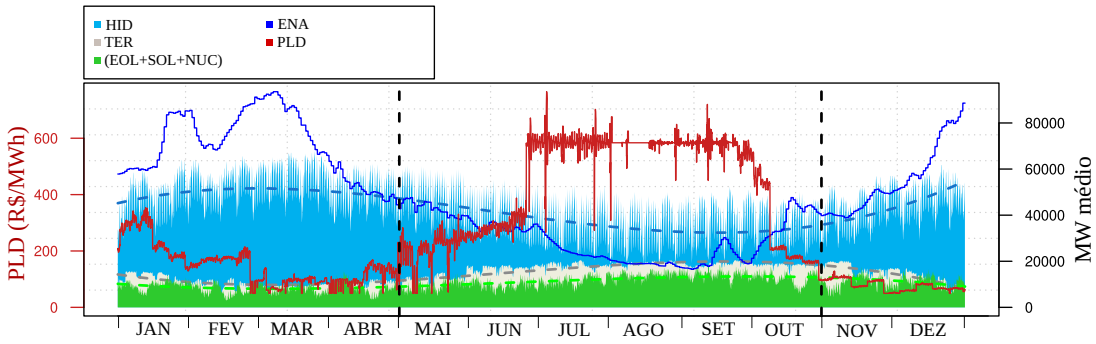
Fonte: O Autor.

a eólica. Do ponto de vista político se destaca no intervalo do conjunto de treino a crise econômica doméstica de 2014, em 2016, o *impeachment* da então presidenta Dilma Rousseff e, em 2018, da greve dos caminhoneiros. Já no conjunto de validação possui relevância a pandemia de COVID-19 que modificou o perfil da carga e os horários de pico de consumo em 2020 (MOREIRA *et al.*, 2021; GETIRANA; LIBONATI; CATALDI, 2021). Neste ano o dólar ultrapassou a marca dos cinco reais. Com a retração econômica houveram diversas paralisações de impacto no SIN envolvendo grandes consumidores como a Albrás, ArcelorMittal, Usiminas, Gerdau e Braskem. Todas estas empresas e diversas outras do setor de energia apresentaram significativas quedas na bolsa de valores em 2020. Finalmente, no conjunto de testes que diz respeito ao ano de 2021, o ponto mais importante para fins deste trabalho foi a adoção oficial do PLD horário pela CCEE. Além disso, apesar de ser observada a retomada dos níveis de consumo de eletricidade que foram reflexo da vacinação para COVID-19 e encerramento de campanhas de distanciamento social, o Brasil enfrentou ainda a maior crise hídrica registrada nos últimos 91 anos. A resposta adotada pelo governo frente ao desabastecimento dos reservatórios foi o acionamento de usinas termelétricas de alto custo, fora da ordem de mérito. Como pode ser observado na Figura 82, esta ação fez com que o PLD apresentasse volatilidade e alta a partir de maio. A alta atingiu o limite em julho quando o Máximo Teto Estrutural (MTE) no valor de R$ 583,88/MWh foi adotado. O MTE vigorou por 57 dias de 26/06/2021 a 23/08/2021 e só deixou de ser aplicado a partir de outubro quando a precipitação esperada começou de fato a ocorrer e a Energia Natural Afluente (ENA) nos reservatório indicou recuperação. Ainda neste período foi praticada uma bandeira de custo excepcional no valor de R$ 14,20 por 100 kWh consumidos denominada bandeira de escassez hídrica. Ela vigorou até abril de 2022 (ONS, 2022b; BANGOLIN *et al.*, 2021; NAUMANN *et al.*, 2021; ONS, 2022a; ANA, 2022).

Os resultados discutidos neste capítulo dizem respeito a aplicação do sistema OPB sem retreino semanal, mas com uso de janelamento móvel. Os modelos foram criados com base em três passos:

(i) Um modelo experimental base foi treinado em 7 anos de dados (conjunto de treino) e avaliado no ano seguinte (conjunto de validação). Este processo foi repetido até que testes complementares (escolha de variáveis de entradas e métodos de pré-processamento) e a otimização via *grid-search* retornassem a parametrização a ser utilizada.

Figura 82 – PLD, ENA, gerações hídrica (HID), térmica (TER) e alternativas eólica, solar e nuclear somadas (EOL+SOL+NUC) em 2021. Os valores referem-se a todo o SIN com exceção do PLD específico do subsistema SECO. O período da crise hídrica está destacado entre as linhas verticais pontilhadas.



Fonte: O Autor.

(ii) Com os parâmetros indicados, então 30 modelos com estes parâmetros fixos foram treinados novamente em 7 anos de dados (conjuntos de treino) e os erros de suas previsões foram aferidos no ano seguinte (conjunto de validação). Se o comportamento médio dos 30 modelos (curvas de erro ao longo das épocas) se mantivesse o mesmo em relação ao modelo experimental base, então essa parametrização era escolhida e seguia-se para o passo (iii), senão, o passo (i) novamente era executado.

(iii) Na sequência, os erros dos 30 modelos válidos foram utilizados para cálculo do tamanho amostral que indicou a necessidade de 209 modelos ANN [1]. Os 209 modelos foram então gerados com novo treino realizado desta vez em 8 anos de dados (conjuntos de treino e validação concatenados). As 209 ANNs e um modelo LEAR são os que compõe o *ensemble* no sistema OPB diretamente aplicado de forma cega, semana a semana, no ano de 2021.

Quanto a alguns dos *benchmarks* utilizados para comparação citados adiante, cabem as seguintes considerações:

- SARIMA: Foi aplicado à série de retornos log-transformados do PLD conforme indicações de Box e Jenkins (1970), Moretin e Toloi (2006), Bueno (2011), Gujarati e Porter (2011). Realizou-se reajuste dos termos regressores semanalmente via otimização do AIC, ou seja, há retreino neste *benchmark*.

- Modelos de DL puros: Usam exatamente o mesmo conjunto de dados e dinâmica de parametrização do sistema OPB, portanto não há retreino e a janela é semanal. Os resultados apresentados dizem respeito as respectivas médias das previsões realizadas por 30 modelos de DNN, 30 de LSTM e 30 de GRU.

[1] Os resultados completos (p-valor e estatística-teste) relativos ao cálculo do tamanho amostral podem ser consultados em Nametala (2022) (pasta *tamanho-amostral*).

- Modelos de ML baseados em *ensembles*: Não há retreino mas recebem dados da última janela semanal. Tanto para CatBoost quanto para XGBoost é gerada via otimização uma árvore de decisão (*ensemble*) que combina mais de um algoritmo de ML[2]. Os dados de entrada são os mesmos do sistema OPB.

Como pode-se perceber, com exceção do SARIMA (que usa somente a série do preço) todos os outros *benchmarks* que exigem treino recebem como entrada exatamente os mesmos conjuntos, ou seja, o que é trocado para fins de comparação é unicamente o modelo preditor.

Levando-se em conta os trabalhos apresentados no Capítulo 3, em apenas três deles foi observado o uso de um período de previsão igual ou maior que o proposto neste estudo, ou seja, superior a um ano. Em nenhum destes artigos, contudo, o retreino do modelo (janela a janela) foi dispensado. Há ainda que se considerar que alguns dos trabalhos listados na revisão bibliográfica fizeram também o uso de previsões do tempo. O retreino a cada janelamento móvel e o uso de previsões de tempo não está incluso no sistema OPB que, com exceção das variáveis de calendário e CMOs, usa apenas dados históricos.

6.1.1 Seleção de entradas e métodos de pré-processamento

Os resultados a seguir dizem respeito sempre ao sMAPE médio de 30 modelos. A escolha das entradas e métodos de pré-processamento foi o primeiro experimento desenvolvido. Neste sentido, foi utilizado um *set* de parâmetros manualmente definidos[3] sem aplicação ainda da otimização *grid-search* que foi realizada posteriormente à esta fase. As tabelas com os erros podem ser consultadas em Nametala (2022) (pasta *testes-entradas*).

Quanto ao AE foram testadas 18 possibilidades de combinações de entradas e métodos de pré-processamento envolvendo as seguintes características:

- Tamanho do vetor latente:
  - 64 posições.
  - 128 posições.
- Método de pré-processamento:
  - Nenhum tratamento de *outliers*.
  - IQR padrão (conforme subseção 5.2.2.1.2).
  - IQR com margens estabelecidas manualmente [4].

[2] Os algoritmos incorporados a estes modelos e a técnica de otimização das árvores de decisão podem ser consultados em suas respectivas referências (YANDEX, 2022; DMLC, 2022).

[3] Baseado nos resultados obtidos para a qualificação.

[4] Baseado nos resultados obtidos para a qualificação. As margens em questão são mais largas e removem menos *outliers* que o IQR padrão, sendo elas, 2 para pressão e umidade, 2,2 para radiação, 3 para temperatura, 4,5 para velocidade do vento e 7 para precipitação.

- Entradas:
  - Precipitação e temperatura.
  - Precipitação, temperatura, radiação e umidade.
  - Precipitação, temperatura, radiação, umidade, pressão e velocidade do vento.

A combinação que apresentou menor erro na reconstrução das matrizes foi a envolvendo vetor latente com 128 posições, tratamento IQR para os *outliers* e uso das variáveis precipitação, temperatura, radiação e umidade.

O AE escolhido foi então acoplado ao modelo Seq2Seq que recebeu inicialmente como conjunto de entrada na parcela *deep* da arquitetura todas as variáveis das tabelas 7 e 8 (com exceção do CMO para o dia seguinte), além de todas as variáveis de calendário apresentadas na subseção 5.2.2.2.3. Aplicadas à estas entradas foram testadas 4 possibilidades de pré-processamento, sendo:

- Nenhum tratamento de *outliers* e nenhum tratamento de variância.
- Nenhum tratamento de *outliers* e tratamento de variância ASINH.
- IQR padrão e nenhum tratamento de variância.
- IQR padrão e tratamento de variância ASINH.

Dentre estes o que apresentou menor sMAPE médio foi o que combinou IQR padrão para tratamento de *outliers* e nenhum método para tratamento de variância. Após isto, foram testados 5 conjuntos de entradas, sendo:

- Conjunto simples: PLD do subsistema SECO apenas [5].
- Conjunto das variáveis do sistema elétrico: PLD do subsistema SECO e todas as variáveis do sistema elétrico dispostas na Tabela 7 (com exceção do CMO para o dia seguinte).
- Conjunto das variáveis econômicas: PLD do subsistema SECO e todas as variáveis econômicas dispostas na Tabela 8.
- Conjunto misto de variáveis[6]: PLD, CMO, ENA e carga do subsistema SECO; gerações hídrica, térmica e alternativas (eólica, solar e nuclear somadas) em todo território nacional e; valor do câmbio (R$ para US$) e vendas de óleo combustível.
- Conjunto completo: Todas as variáveis das tabelas 7 e 8.

[5] Visando obter uma comparação relacionada ao modelo mais simples possível, especificadamente neste caso a entrada *wide* também foi setada para receber somente o PLD do subsistema SECO ao invés das entradas de reforço discutidas na subseção 5.2.2.2.4.

[6] Os motivos pelos quais foram selecionadas estas variáveis para compor o conjunto misto é explicado no Apêndice C.

O conjunto misto de variáveis foi o que apresentou menor sMAPE médio. Finalmente, foram testadas combinações de variáveis relacionadas ao calendário da semana de previsão, sendo as possíveis:

- Hora.
- Dia da semana.
- Hora e dia da semana.
- Hora e feriado.
- Hora, dia da semana e semana.
- Hora, dia da semana e trimestre.
- Hora, dia da semana, dia e mês.
- Hora, dia da semana, feriado e horário de verão.
- Hora, dia da semana, semana, dia, mês, trimestre, feriado e horário de verão.

A combinação com menor sMAPE médio foi a que contém apenas hora e feriado.

6.2 Compactação das variáveis climáticas

Após feita a requisição de todas as matrizes geográficas providas pelas estações do INMET ao M1, inicialmente no M2 é realizado o pré-processamento e codificação dos dados, conforme visto na Figura 78 onde se indica a entrada 1. A seguir os resultados destas duas etapas são discutidos.

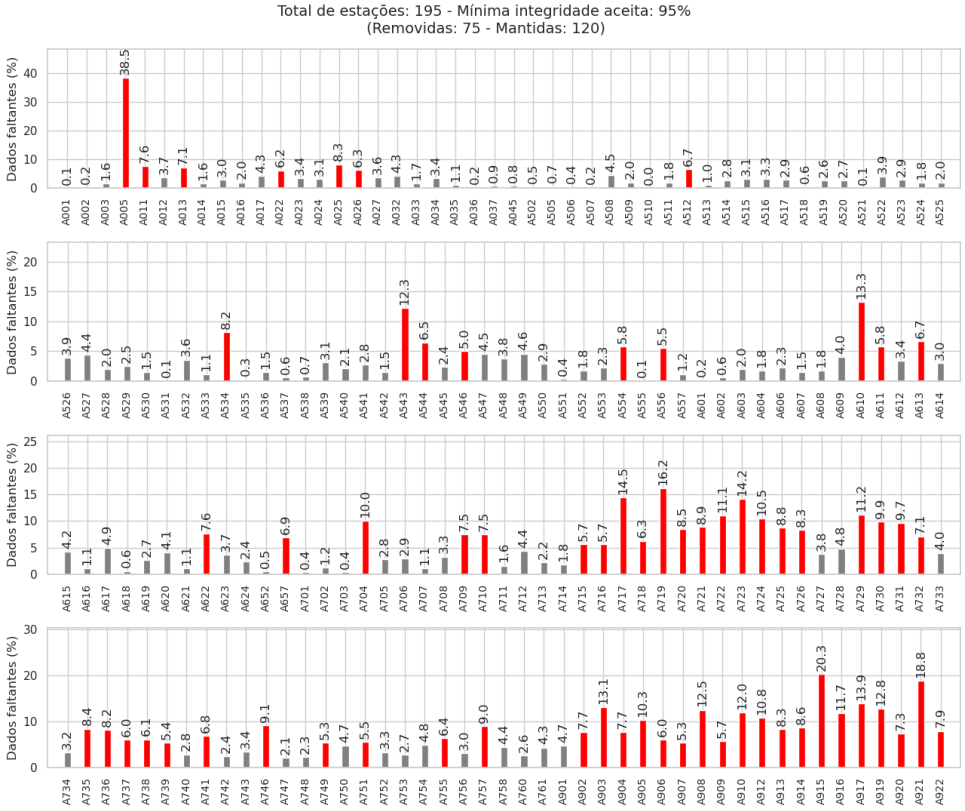
6.2.1 Pré-processamento

Na primeira etapa [7] são selecionadas apenas as estações que possuem no máximo 5% de problemas de integridade nos conjuntos de treino e validação. Foram consideradas como entradas as leituras horárias de precipitação, temperatura, radiação e umidade. Neste sentido, levando-se em conta as 195 estações existentes no subsistema SECO, foram extraídos do banco de dados para análise um total de aproximadamente 47,8 milhões de valores. O percentual de integridade aferido por estação é o mostrado na Figura 83. Os percentuais analisados indicaram a necessidade de remoção do estudo de 75 estações. As 120 restantes são as que provêm a cobertura geográfica observada na Figura 84.

O percentual de dados não íntegros das 120 estações filtradas foi tratado via pré-processamento no M2. A estação A607 apresentou, por exemplo, 1,5% de não integridade, em números absolutos

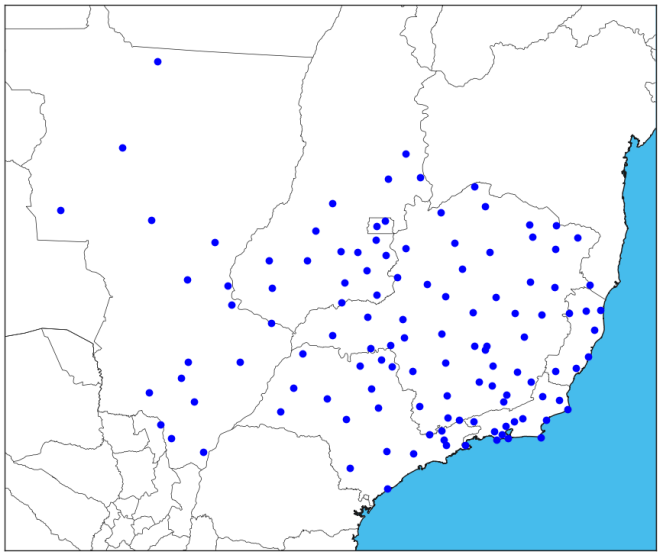
[7] As estações do INMET possuem códigos iniciados pela letra "A" seguidos de uma sequência numérica. A estação de Campos dos Goytacazes (que será utilizada como exemplo em algumas demonstrações feitas nesta subseção) possui código A607.

Figura 83 – Análise da integridade de dados das estações meteorológicas de solo do INMET.



Fonte: O Autor.

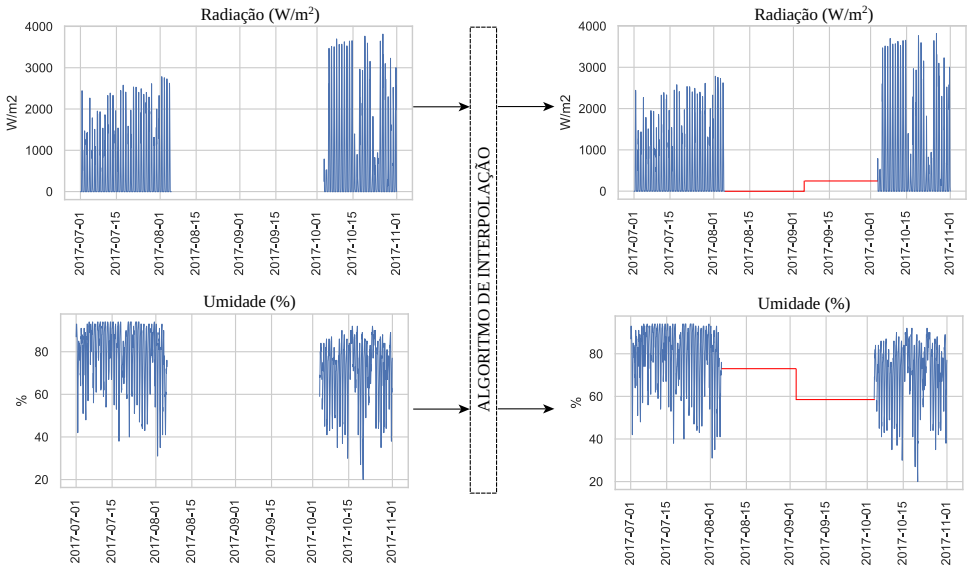
Figura 84 – Cobertura geográfica das estações filtradas pela análise de integridade.



Fonte: O Autor.

são 245154 leituras não realizadas apenas nesta estação. A substituição de dados não íntegros via aplicação do NNI no segundo semestre de 2017 (onde a maioria destes erros foram observados) cria o efeito na série temporal como mostrado na Figura 85.

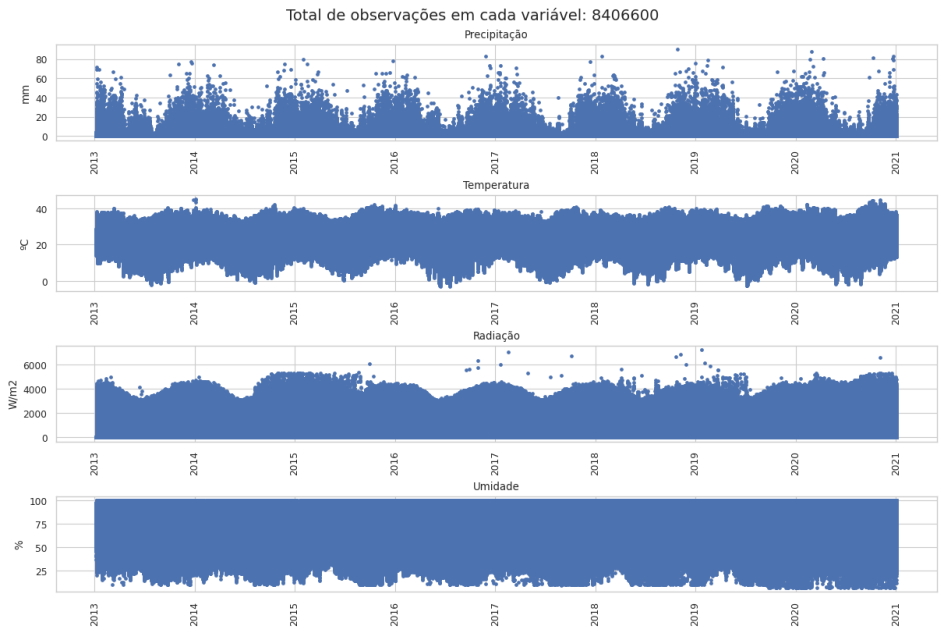
Figura 85 – Efeito da aplicação do tratamento de integridade aos dados faltantes de radiação e umidade da estação A607 de Campos dos Goytacazes entre julho e novembro de 2017.



Fonte: O Autor.

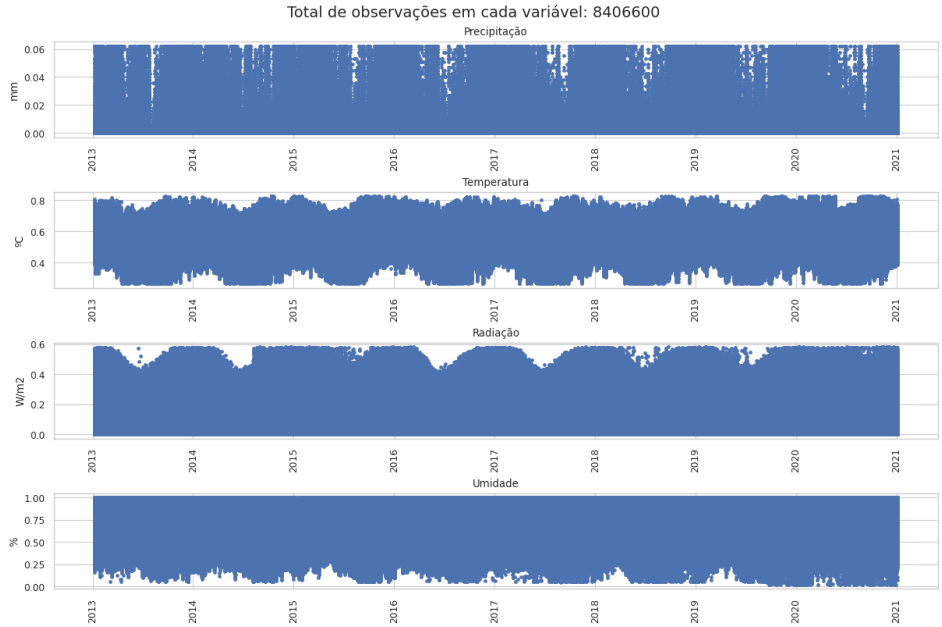
Finalizado o tratamento de integridade, foi realizada então a aplicação das rotinas de condicionamento de dados. Quanto ao tratamento de *outliers* foram feitas 71157 substituições utilizando-se IQR. O efeito dos tratamentos pode ser observado nas figuras 86 e 87. Como visto, todas as variáveis apresentam ciclos alternados que correspondem a inverno e verão. Isso é especialmente evidente em precipitação e temperatura que juntas denotam a existência de períodos úmidos associados ao clima mais quente no sudeste e centro-oeste do Brasil.

Figura 86 – Variáveis de entrada no AE antes da aplicação do pré-processamento.



Fonte: O Autor.

Figura 87 – Variáveis de entrada no AE após a aplicação do pré-processamento.



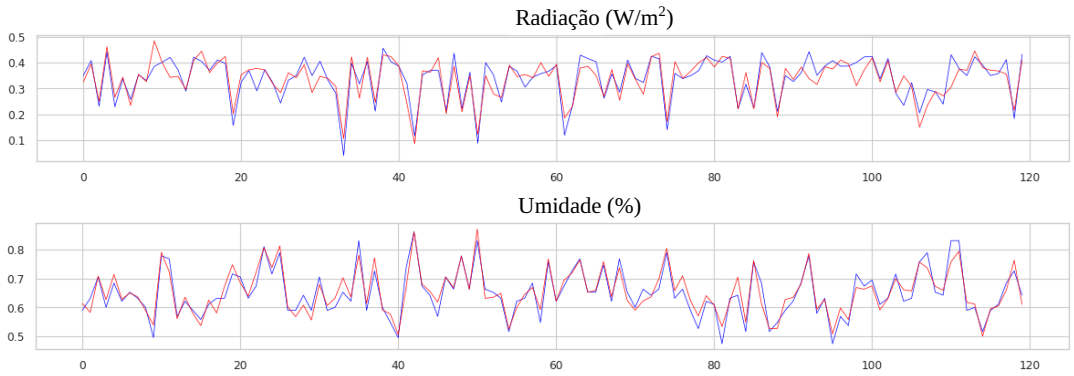
Fonte: O Autor.

Até o momento, a ONS e a CCEE não dispõe (ou não divulgam) a existência de um sistema previsor do PLD. No entanto, existe um sistema para previsão de carga denominado PrevCarga. Segundo seu manual de uso (ONS, 2020a), o PrevCarga é baseado em um modelo híbrido de RNA e SVR. O mesmo faz uso da temperatura como variável climática. Para resolver o problema das várias medições em diferentes locais emprega uma abordagem chamada no manual de "temperatura equivalente". Nesta são usados valores ponderados de temperatura de alguns aeródromos localizados em grandes centros urbanos por subsistema (ONS, 2020b). Apesar de ser uma abordagem válida, ela não considera a estação meteorológica e sim um aeroporto. Outro ponto é que diversas regiões marginais aos aeroportos (porventura importantes no cálculo) são simplesmente desprezadas. No subsistema SECO, por exemplo, são consideradas apenas os localizados em seis cidades. O AE empregado nesta pesquisa possui a vantagem de aprender, inclusive no tempo, quais estações e leituras são relevantes, além de inferir a melhor forma de agregá-las. O trabalho de Hong, Wang e White (2015) foi o único encontrado que trata especificadamente da seleção de estações climáticas. Aplicado ao mercado americano, neste artigo os autores propõem uma abordagem baseada em regressão linear e, como no sistema do ONS, considera apenas a temperatura. Outra diferença é que no trabalho de Hong, Wang e White (2015) são realizadas previsões 24 horas à frente e não as 168 que aqui são propostas.

6.2.2 Codificação

O resultado da fase de pré-processamento são 61296 matrizes geográficas com dimensões 120×4 correspondentes a cada hora no período de 2013 a 2020. O AE empregado para redução de dimensionalidade converte cada uma destas matrizes num vetor latente de 128 posições e depois reconstrói o mesmo para as dimensões originais.

Figura 88 – Comparativo entre série real (em azul) e série reconstruída pelo AE (em vermelho) para as leituras das variáveis radiação e umidade da estação A607 de Campos dos Goytacazes numa semana aleatoriamente selecionada dentro do *dataset*.



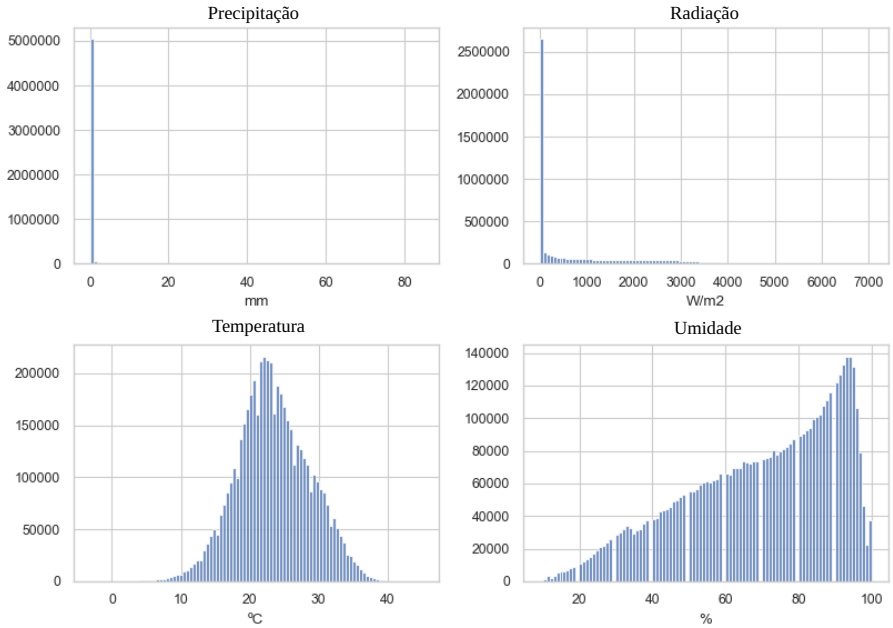
Fonte: O Autor.

O modelo obtido após otimização dos pesos foi capaz de reconstruir todas as matrizes dos 8 anos de dados com MSE de 2.2×10^{-5}. A título de exemplo, na Figura 88 é mostrado um momento aleatoriamente selecionado no *dataset* para radiação e umidade na estação A607. Como pode-se notar, o AE é capaz de reconstruir os dados a partir do vetor latente. É importante também destacar que, enquanto as distribuições dos dados sem nenhum tratamento envolvem 33,6 milhões de observações com características estatísticas diferentes em cada variável, após todo o processo, a distribuição dos vetores latentes que incorporam 99,9% da informação numa compressão correspondente a 26,66% do tamanho original, apresenta características estatísticas mais desejáveis ao treinamento dos modelos. Como exemplo, pode-se citar a aproximação com uma distribuição normal com uma média próxima de zero. O efeito deste processo de tratamento de dados nas distribuições pode ser observado nas figuras 89 e 90. Por fim, curiosamente pode-se observar na Figura 90 a existência de uma barra com valores acumulados entre 0,98 e 1. Esta concentração provavelmente é decorrente dos dados de precipitação e radiação que em suas distribuições originais também acumulam valores numa mesma classe.

6.3 Previsão do PLD

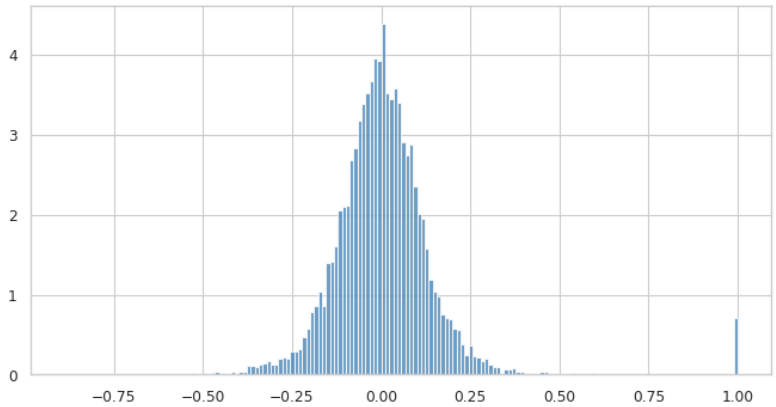
A arquitetura Seq2Seq foi aplicada com o objetivo de modelar a série do PLD que no período de aplicação utilizado apresentou comportamento como mostrado na Figura 91. As distribuições dos preços (ano a ano) são mostradas na Figura 92. Como pode-se observar, o PLD apresentou um comportamento similar em 2017 e 2018 com preços mais altos sendo praticados no mês de agosto. Esse comportamento foi alterado sendo que pode ser observado um deslocamento progressivo nos anos de 2019 e 2020 da curva de alta nos preços mais para o último trimestre do ano. Apesar destas características, de forma geral o que se vê em todos os anos é a prática de preços mais baixos no início e final de cada ano. Esta curva é condizente com a sazonalidade das chuvas e temperaturas no Brasil. Há ainda que se destacar que, apesar do comportamento da série

Figura 89 – Distribuições dos dados das estações meteorológicas do INMET antes do tratamento com o AE.



Fonte: O Autor.

Figura 90 – Distribuição dos valores dos vetores latentes gerados após o tratamento de dados dos dados das estações meteorológicas do INMET feito via AE.



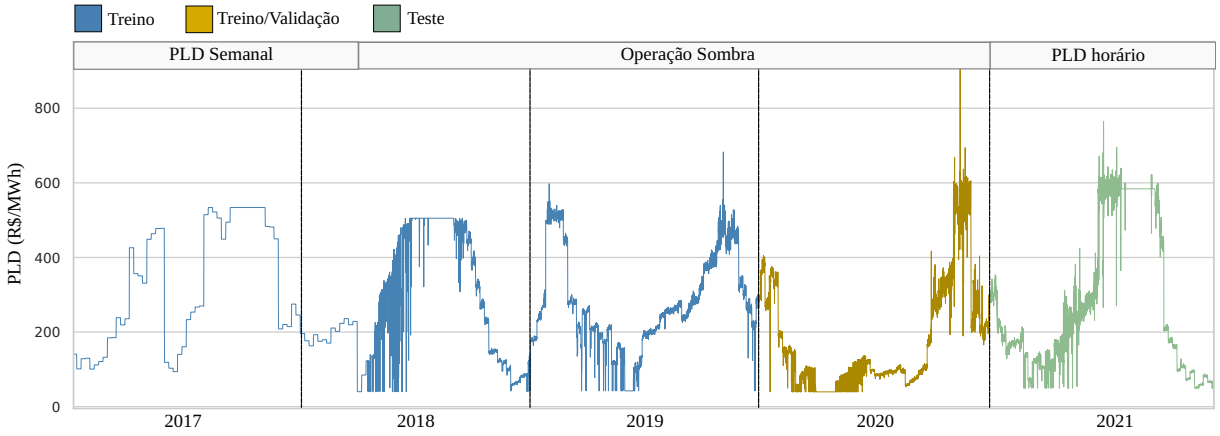
Fonte: O Autor.

no período de teste ser aparentemente mais similar com 2019 e 2020, no que tange as distribuições existe maior semelhança desta com os anos de 2017 e 2018, pois como visto na Figura 92, o pico dos preços está concentrando na cauda mais à direita das distribuições e não à esquerda.

No período de treino e validação foram consideradas 364 janelas móveis. Na Figura 93 pode ser observada a evolução dos erros ao longo das épocas[8] comparativamente entre os conjuntos de treino e validação. No treinamento o MSE foi de $1{,}4 \times 10^{-4}$. Já na validação, como esperado, há um aumento do erro para $2{,}5 \times 10^{-4}$. Após definição dos parâmetros e com o retreino no conjunto

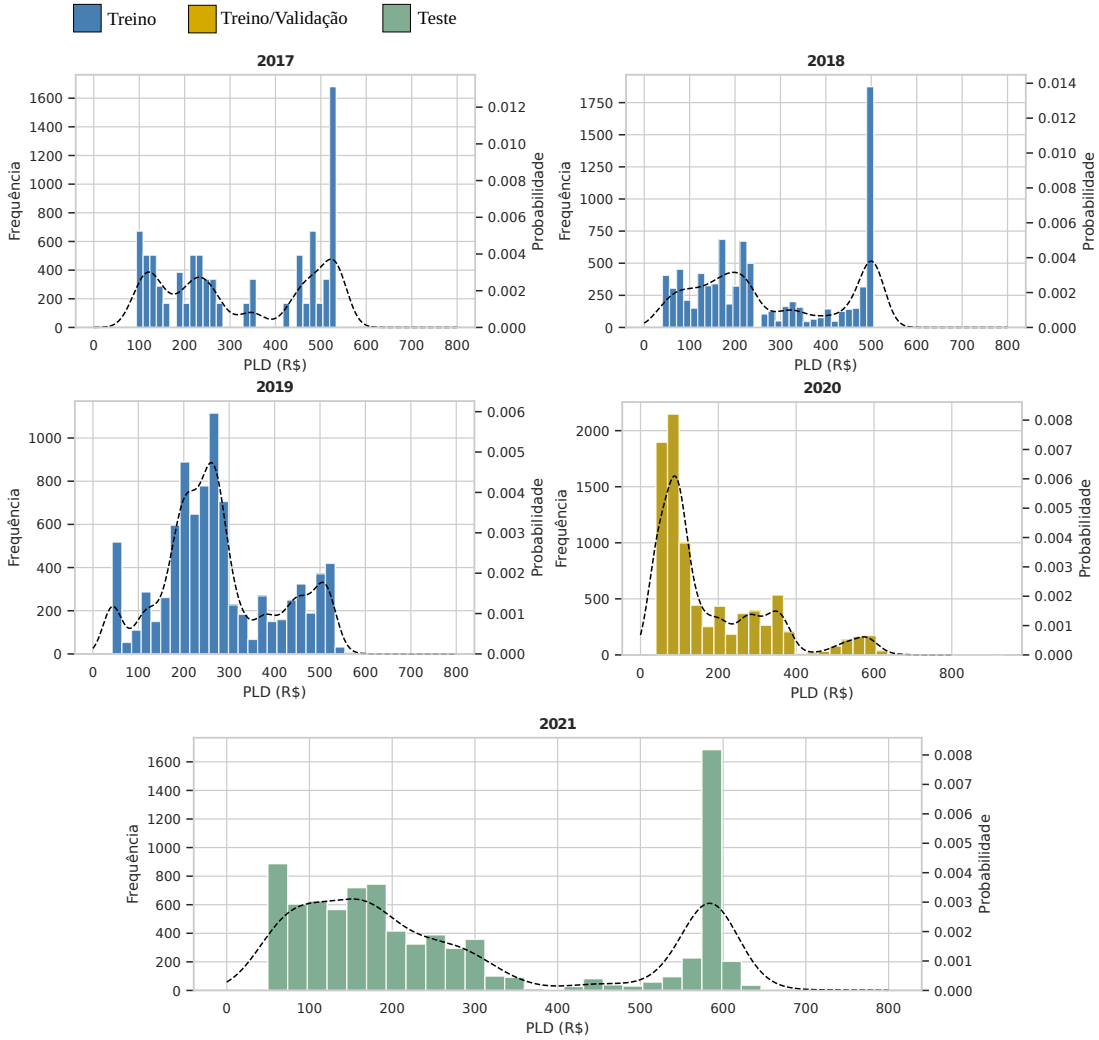
[8] Para o melhor caso entre 30 modelos ANN gerados.

Figura 91 – Período de 02/01/2017 a 26/12/2021 utilizado para treinamento, validação e testes.



Fonte: O Autor.

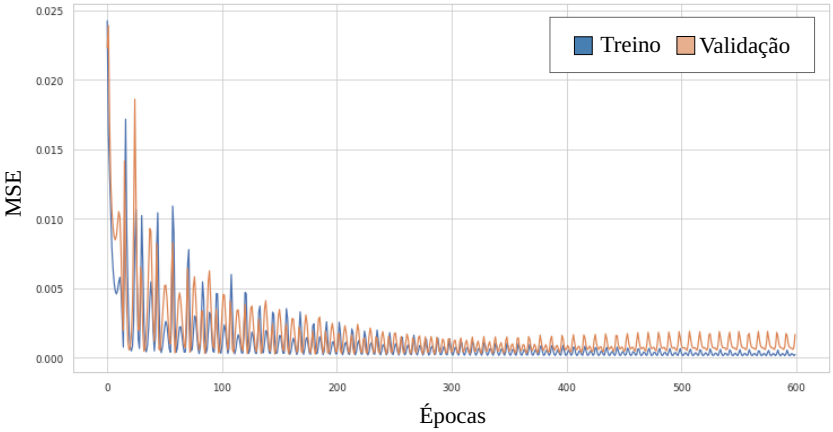
Figura 92 – Distribuição anual dos valores de carga.



Fonte: O Autor.

único, foi gerado então um cenário futuro no módulo M4 para as 51 janelas existentes no conjunto de teste de 2021. Estes resultados são discutidos a seguir contemplando duas partes, sendo, a capacidade preditiva e a avaliação financeira.

Figura 93 – Evolução dos erros baseados em MSE ao longo de 600 épocas.



Fonte: O Autor.

6.3.1 Capacidade preditiva

A capacidade preditiva nesta subseção é avaliada primeiramente comparando-se os resultados do sistema OPB em relação os *benchmarks*. Na sequência é feita uma análise individualizada do sistema OPB.

6.3.1.1 Comparativo de modelos

Na Tabela 11 são mostrados os resultados das métricas sMAPE, rMAE, R^2 e POCID para todos os modelos previsores durante o ano de 2021. Os destacados em amarelo e negrito são os três melhores resultados em cada métrica. Como discutido na subseção 4.1.3.1 cada uma destas possui propósito específico, contudo o sMAPE é a principal pois é a que mais se adequa para uso em um mesmo mercado (no caso, subsistema SECO do mercado brasileiro) visto que é livre de escalas e considera ao mesmo tempo a qualidade da regressão, o deslocamento temporal e a proximidade também com os pontos atípicos. Já o rMAE é uma métrica que tem por principal vantagem a inclusão no seu cálculo dos resultados de um modelo *Naive* que está indicado na Tabela 11 como Controle. Os resultados do modelo Controle tratam-se de uma referência, visto que informam o que aconteceria se os valores do PLD real da semana anterior fossem simplesmente considerados como as previsões para a próxima semana. O rMAE mostra o ganho (valores menores do que um) ou perda (valores maiores do que um) em relação ao Controle. No entanto, os resultados de rMAE devem ser observados com cuidado, pois esta métrica privilegia modelos que apenas regridem os dados numa curva suave ao não considerar a morfologia da série. Esta observação é importante, pois em preços de eletricidade os detalhes que desenham a série são aqueles que indicam certos padrões específicos como horários de pico e perfis de dias da semana. O R^2, por sua vez, mostra o acoplamento das previsões

aos dados reais ao levar em consideração a dispersão das previsões no entorno da curva de regressão ideal gerada pelos dados reais, logo considera a cobertura das duas distribuições (como visto na Figura 97). Por fim, POCID indica a capacidade do modelo preditor em gerar uma série prevista que se movimenta na mesma direção da real, sem ponderar no entanto a magnitude desta variação.

Na Tabela 11 pode-se observar inicialmente que o sistema OPB é o único que em todas as métricas aparece entre os três melhores. Quanto a sMAPE, mesmo ao considerar o pior caso, ele novamente é superior a todos os outros modelos. Cabe destacar que nesta métrica, a melhor ANN e o LEAR individualmente não superam o sistema OPB e apresentam resultados, inclusive, muito próximos entre si. No entanto, quando combinados pelo *ensemble* são capazes de baixar o sMAPE em uma unidade. Isso se dá, pois o modelo LEAR tende a criar uma curva de regressão muito suave centrada sobre a média, no entanto, não é capaz de modelar a variância temporal. Ao contrário do LEAR, os modelos ANN modelam a variância ao tentar reconstruir os pontos atípicos e, por isso, podem fugir em certos momentos da média. A combinação dos dois modelos, portanto, é uma escolha acertada pois incorpora o melhor dos dois ao "trazer" para a média correta os valores gerados pela ANN que o LEAR sozinho não é capaz de modelar. Isto fica mais claro ao se perceber que em rMAE o LEAR mostra-se superior a melhor ANN e ao próprio sistema OPB, enquanto em R^2 todos apresentam valores próximos. No entanto, isoladamente, como o LEAR não modela os *outliers* e a ANN pode errar a média, em relação a POCID ambos perdem até mesmo para modelo Controle. O modelo Controle, inclusive, tende a apresentar bons resultados em POCID, pois a direção dos movimentos do PLD de forma geral se repete ao longo do tempo em decorrência da sazonalidade do ano e dos perfis de consumo estáveis. Neste contexto, o sMAPE privilegia os resultados dos modelos que tendem a ser bons em todas as outras métricas, por este motivo o sistema OPB que está entre os três melhores em todos os casos aparece em destaque na coluna do sMAPE.

O ano de 2021 foi um ano no qual o Brasil foi acometido pela maior crise de desabastecimento hídrico dos reservatórios do SIN em toda sua história. Neste sentido, a análise completa do período dá um indicativo da capacidade dos modelos previsores de se adaptarem à mudanças de padrões no preço. No entanto, devido a esta situação torna-se necessária também uma análise separada dos erros dentro e fora da crise hídrica visando avaliar a capacidade preditiva dos modelos em momentos críticos e adversos, como também naqueles em que uma situação de aparente normalidade existe. Como visto com mais detalhes na Figura 95, os movimentos atípicos nos preços tiveram início em maio com uma crescente alta no PLD acompanhada de volatilidade que durou até o final de junho. Na última semana de junho ocorreu um aumento abrupto no preço com magnitude próxima de 50%. A média do PLD foi então estabelecida próxima ao MTE e assim se manteve variando com a presença de pontos atípicos durante todo o mês de julho. Em agosto o MTE foi estipulado de forma fixa e isto durou até o início de outubro quando, finalmente, o PLD voltou a cair. Na Tabela 12 são mostrados os erros dos modelos previsores durante a crise hídrica e na Tabela 13, fora dela.

Durante a crise hídrica, como visto na Tabela 12, o modelo LEAR foi o que apresentou melhores resultados seguido do sistema OPB em sMAPE, rMAE e R^2. Em POCID os melhores

Tabela 11 – Erros em 2021. Negrito e amarelo destacam os três melhores modelos em cada métrica.

Modelo	sMAPE	rMAE	R2	POCID
OPB Melhor	**11.9626**	**0.539**	**0.972**	0.7732
OPB	**12.124**	**0.5337**	**0.9721**	**0.7907**
OPB Pior	**12.3674**	0.5602	0.9708	0.7597
ANN Melhor	13.3083	0.6785	0.9578	0.7685
LEAR	13.3107	**0.5047**	**0.9749**	0.764
ANN Média	13.4587	0.6595	0.9599	0.7866
ANN Pior	14.7107	0.7324	0.9552	0.7447
XGBoost	16.4549	0.7652	0.9523	0.6288
Controle	22.0532	1	0.897	**0.8744**
CatBoost	22.4543	0.9898	0.9307	**0.8272**
SARIMA	23.826	1.0648	0.8947	0.6235
GRU	32.9846	1.3878	0.8868	0.6627
DNN	37.0719	1.4912	0.8732	0.6267
LSTM	37.8171	1.5754	0.8512	0.6985

Fonte: O Autor.

Tabela 12 – Erros durante o período da crise de 2021. Negrito e amarelo destacam os três melhores modelos em cada métrica.

Modelo	sMAPE	rMAE	R2	POCID
LEAR	**9.8171**	**0.5327**	**0.9542**	0.8257
OPB Melhor	**10.541**	**0.6484**	**0.944**	0.8286
OPB	**10.6347**	**0.6386**	**0.9446**	0.8345
OPB Pior	11.008	0.6787	0.9412	0.828
ANN Melhor	12.1035	0.8383	0.9148	0.8345
ANN Média	12.2082	0.8096	0.9195	**0.84**
XGBoost	12.7503	0.8527	0.9111	0.6924
ANN Pior	13.0991	0.9047	0.9082	0.8275
Controle	14.3975	1	0.8017	**0.9121**
CatBoost	16.2723	1.1197	0.8691	**0.8549**
GRU	17.7299	1.1775	0.8578	0.7072
SARIMA	18.3948	1.2598	0.7902	0.672
LSTM	19.4147	1.3029	0.8102	0.7495
DNN	19.7343	1.2619	0.8467	0.706

Fonte: O Autor.

modelos foram o Controle, CatBoost e a ANN média. Estes resultados são esperados, pois como o LEAR não modela os *outliers*, na presença de grande variância como visto durante a crise hídrica os modelos que o fazem tenderão também a apresentar maiores erros. A situação, no entanto, se inverte ao observar os resultados da Tabela 13 que apresenta os erros fora da crise. Como observado, os resultados do sistema OPB são superiores ao LEAR em todas as métricas e, inclusive, mesmo nos casos de rMAE onde o LEAR tem uma tendência a ter vantagem. Além disso, fora da crise

Tabela 13 – Erros fora do período da crise de 2021. Negrito e amarelo destacam os três melhores modelos em cada métrica.

Modelo	sMAPE	rMAE	R2	POCID
OPB Melhor	**13.4749**	**0.3962**	**0.9092**	0.7137
OPB	**13.7067**	**0.3968**	**0.9065**	**0.7436**
OPB Pior	**13.8116**	**0.4054**	**0.9097**	0.6867
ANN Melhor	14.5942	0.4701	0.8689	0.698
ANN Média	14.7926	0.4638	0.872	0.7296
ANN Pior	16.4287	0.5078	0.8725	0.6563
LEAR	17.0213	0.4681	0.8844	0.698
XGBoost	20.3841	0.6507	0.7949	0.5616
CatBoost	29.024	0.8201	0.7161	**0.7978**
SARIMA	29.5778	0.8097	0.6539	0.5724
Controle	30.1924	1	0.605	**0.8347**
GRU	49.1978	1.6619	-0.0227	0.6151
DNN	55.492	1.7896	-0.1928	0.5421
LSTM	57.3751	1.9306	-0.3234	0.6443

Fonte: O Autor.

hídrica todos os modelos ANN isoladamente são superiores ao LEAR. Essa situação, mais uma vez, indica a necessidade do uso do *ensemble* para que se crie um favorecimento a adaptabilidade do modelo, visto que momentos de crise (seja hídrica ou não) podem ocorrer de forma inesperada e se manter por tempo indeterminado.

Na Tabela 14 é mostrada uma contagem de ocorrências de cada um dos modelos entre os três melhores nas tabelas 11, 12 e 13. Como se nota, o sistema OPB é aquele que aparece em primeira posição. Sendo, portanto, deste ponto de vista o melhor modelo dentre todos os testados. Ainda nesta tabela cabe ressaltar os piores modelos, sendo, todas as RNAs puras, o modelo econométrico otimizado SARIMA, o XGBoost e dois dos casos isolados do modelo ANN. Todos esses sem nenhuma ocorrência. Os modelos Controle, CatBoost e ANN média constam na tabela por seus resultados em POCID e LEAR por seus resultados em rMAE e R^2.

Por fim, na Figura 94 pode ser visualizada a vantagem ou desvantagem percentual de um modelo sobre o outro ao considerar apenas a principal métrica do estudo, o sMAPE. Nesta análise a observação fundamental trata-se mais uma vez da vantagem apresentada pelo uso do sistema OPB em relação a aplicação individualizada do LEAR ou de qualquer caso do modelo ANN. O ganho do sistema OPB em relação ao LEAR foi de 4.59% e em relação a melhor ANN de 4.58%. Este resultado é interessante dado a proximidade dos valores percentuais de ganho, pois mostra que o uso destes modelos isolados de certa forma se mostrou equivalente ao se considerar todo o ano de 2021. Outro ponto de destaque é que o sistema OPB possuiu diferença de apenas 0.62% para o seu melhor caso e de 0.94% para o seu pior caso (considerando as previsões probabilísticas). Neste sentido, em 2021 foi mais vantajoso usar o sistema OPB mesmo incorrendo no pior caso,

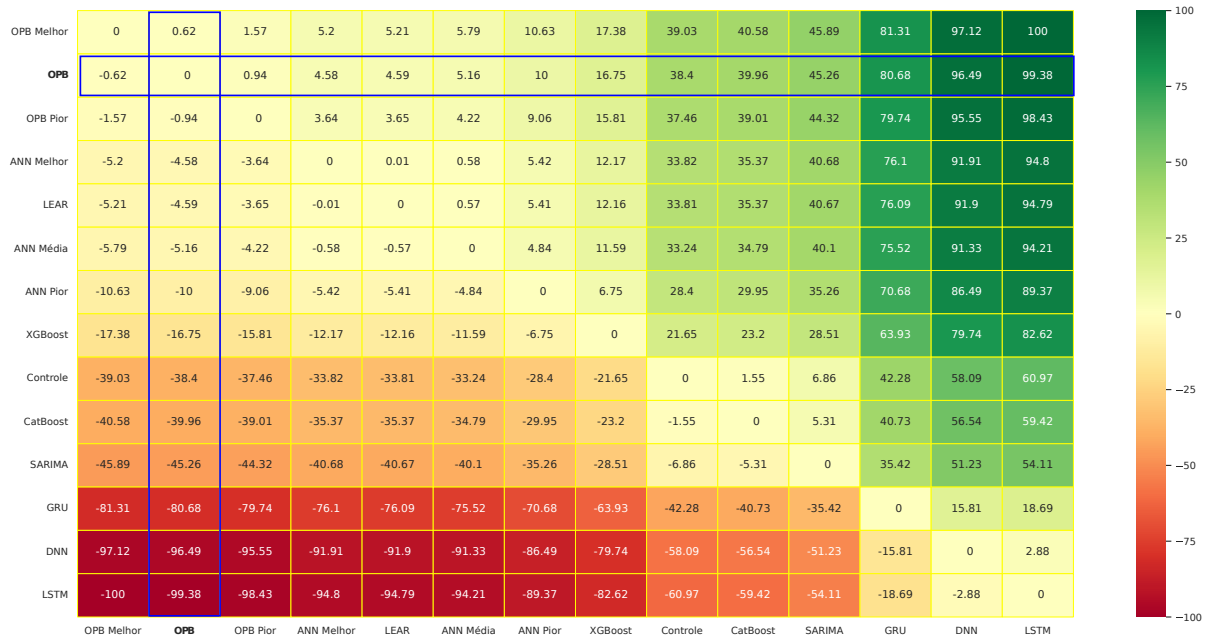
Tabela 14 – Contabilização de ocorrências de um dado modelo entre os três melhores por métrica nas tabelas 11, 12 e 13. Negrito e amarelo destaca o melhor.

Modelo	**Ocorrências**
OPB	**11**
OPB Melhor	9
LEAR	5
OPB Pior	4
Controle	3
CatBoost	3
ANN Média	1
ANN Melhor	0
ANN Pior	0
XGBoost	0
GRU	0
SARIMA	0
LSTM	0
DNN	0

Fonte: O Autor.

do que seria ter adotado qualquer um dos modelos concorrentes.

Figura 94 – Comparativo percentual de vantagem ou desvantagem percentual de um método sobre o outro em relação a sMAPE.

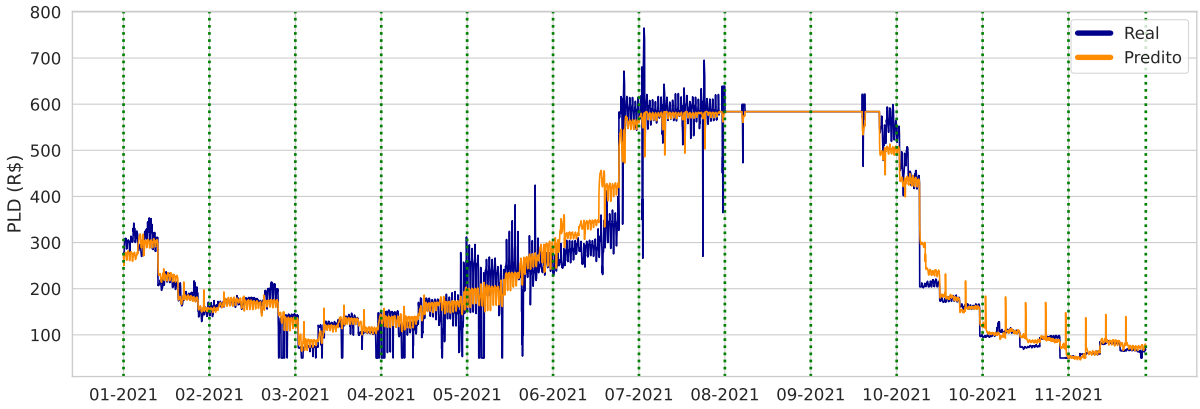


Fonte: O Autor.

6.3.1.2 Análise individual do sistema OPB

Na Figura 95 podem ser visualizadas as previsões do PLD horário geradas (semana a semana) em laranja contra os valores reais em azul. Sugere-se também ao leitor acompanhar a versão interativa deste e de outros gráficos dispostos adiante em OPEBRAS (2022).

Figura 95 – PLD previsto pelo sistema OPB vs. PLD real em 2021. Uma versão interativa deste gráfico contendo todos os preditores pode ser acessada em OPEBRAS (2022).

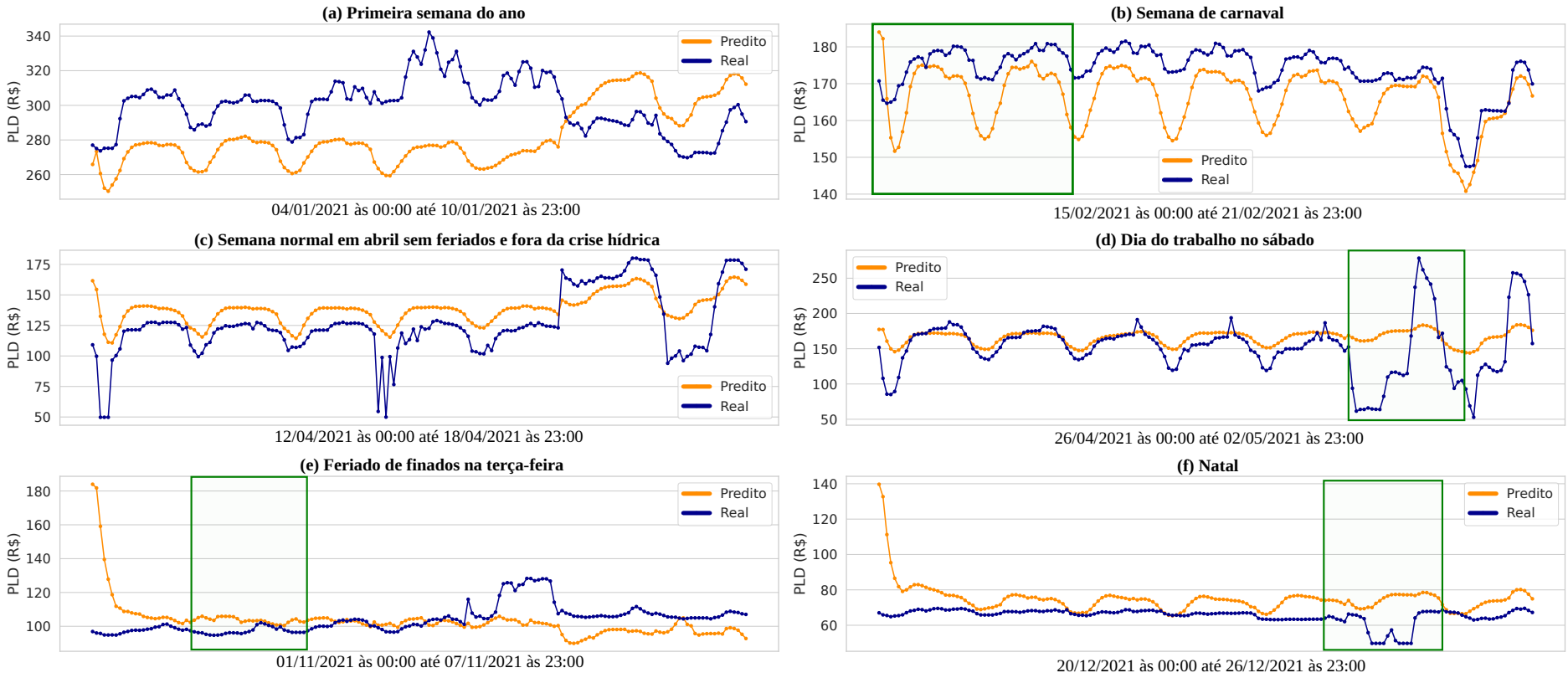


Fonte: O Autor.

Como se nota visualmente, as previsões geradas pelo sistema OPB acompanharam os valores reais do PLD de forma geral durante todo o ano de 2021. Observa-se, no entanto, um descolamento de valores em meio ao início da crise hídrica em junho. A análise visual neste ponto é importante, pois todas as métricas que informam capacidade preditiva irão indicar prejuízo nos resultados neste ponto, entretanto, este descolamento é benéfico. Isso se dá, pois a série predita adiantou o movimento dos preços ao indicar que haveria uma aceleração da alta durante todo o mês de junho, fato que realmente se concretizou. O adiantamento desta informação ao agente de mercado poderia, por exemplo, ter sido de suma importância para ajustamento de contratos visto que na sequência o PLD subiu abruptamente para um valor quase 50% maior. Essa capacidade do sistema OPB também pôde ser observada no final de setembro e início de outubro, durante a saída da crise hídrica. Outra situação de destaque é a previsão correta do MTE associada à alguns dos valores *outliers* que aconteceram neste período. Isto mostra que o sistema OPB, mesmo não possuindo uma modelagem específica para pontos atípicos, foi capaz de capturar parte deles.

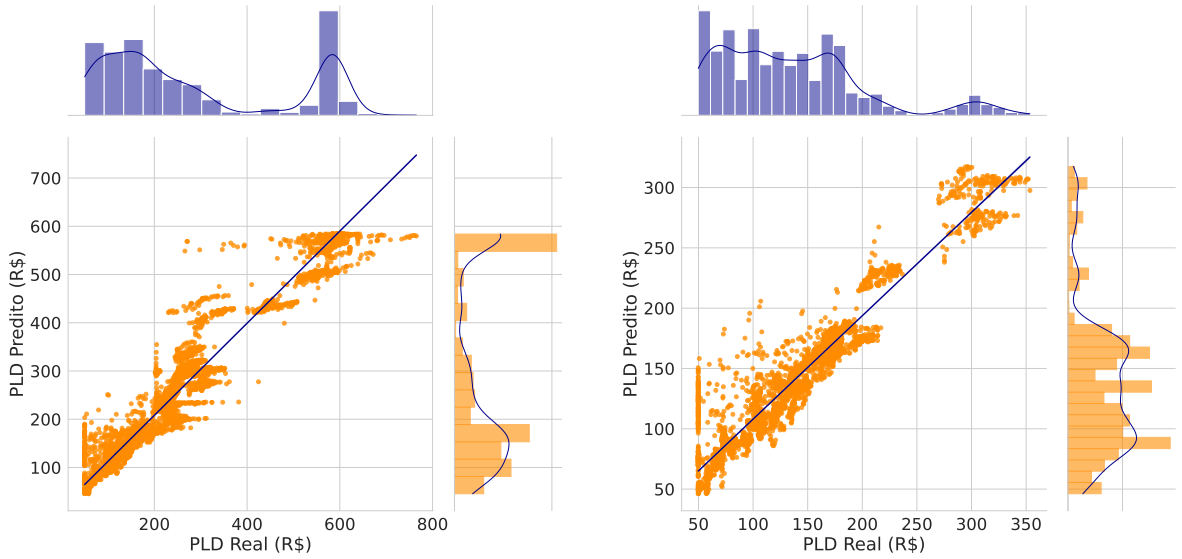
Os gráficos comparativos da Figura 96 dão uma visualização mais aproximada em semanas específicas de 2021. Pode-se perceber que o sistema OPB, apesar de receber como entradas indicativos de feriados e dias especiais na semana de previsão, ainda assim apresentou dificuldades em prever corretamente estes momentos. Esse fato fica mais nítido ao se comparar o desempenho de uma semana normal de abril (mostrada na subfigura *(c)*) e as demais semanas. Com exceção do carnaval, em todas as outras o sistema em certa medida ignorou a existência do feriado. Isso pode indicar a necessidade de uso de um modelo complementar modelado especificadamente para este tipo de data.

Figura 96 – Semanas de interesse em 2021 com quadros verdes indicando as datas dos feriados.



Fonte: O Autor.

Figura 97 – Distribuições e comparativo com a regressão do PLD predito vs. PLD real.



(a) Todo período.

(b) Fora da crise hídrica.

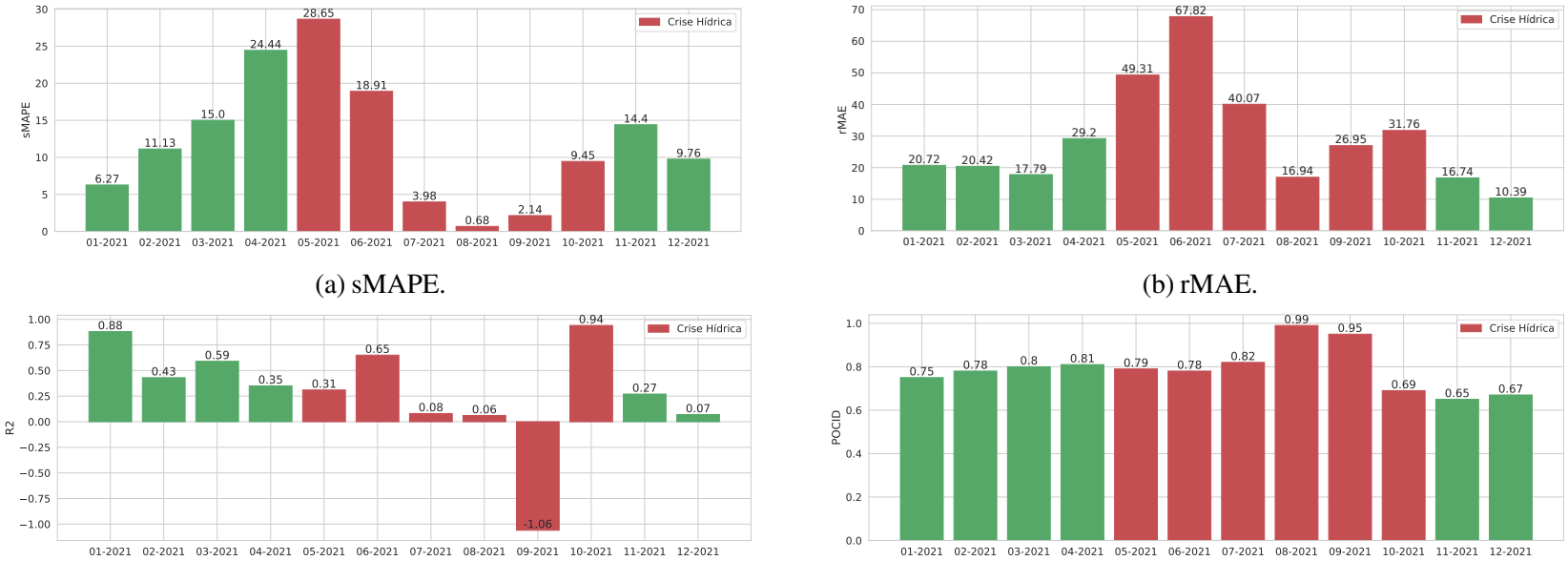
Fonte: O Autor.

Em relação a capacidade preditiva do modelo especificadamente fora da crise hídrica, pode-se notar um melhoramento frente a todo o período como visto na Figura 97. Neste gráfico é possível visualizar um melhor ajustamento dos dados previstos em laranja com a curva de regressão ideal calculada a partir dos dados reais em azul. O modelo apresentou maior facilidade em prever preços nas classes [R$ 49,77 (PLD mínimo) - R$ 280,00] e [R$ 450,00 - R$ 765,11] no período completo como visto na subfigura *(a)*. No entanto, a dispersão dos valores diminuiu na subfigura *(b)* mostrando que mesmo fora destas duas classes observa-se maior capacidade preditiva em períodos de normalidade. Na subfigura *(b)* também pode ser notada maior semelhança entre as distribuições do que na subfigura *(a)*.

Levando em consideração estas observações, nos gráficos da Figura 98 podem ser observados os erros mensais médios para sMAPE, rMAE, R^2 e POCID. Percebe-se que em sMAPE [9] e rMAE existe um aumento considerável nos erros durante a crise hídrica (indicada pelas barras em vermelho). Durante o período em que vigora o MTE esse valor cai para níveis próximos de zero dado que o modelo preditor se mantém gerando apenas um único valor. O coeficiente explicativo R^2 apresenta comportamento parecido com sMAPE e rMAE, no entanto, apenas na saída da crise hídrica em setembro e outubro. Já quanto a POCID o que se observa é um valor relativamente constante entre 0,75 e 0,85 entre janeiro e julho, com aumento nos meses de agosto e setembro devido ao contínuo movimento de baixa do PLD observado nestes meses. Os piores resultados quanto a POCID estão concentrados no último trimestre do ano, visto que preços atípicos foram praticados especialmente nos finais de semana.

[9] Em OPEBRAS (2022) pode ser consultada uma visualização interativa do erro sMAPE hora a hora durante 2021 com todos os *benchmarks* incluídos.

Figura 98 – Erros mensais do sistema OPB em 2021. Barras verdes indicam o período fora da crise hídrica e vermelhas, dentro.



(a) sMAPE.

(b) rMAE.

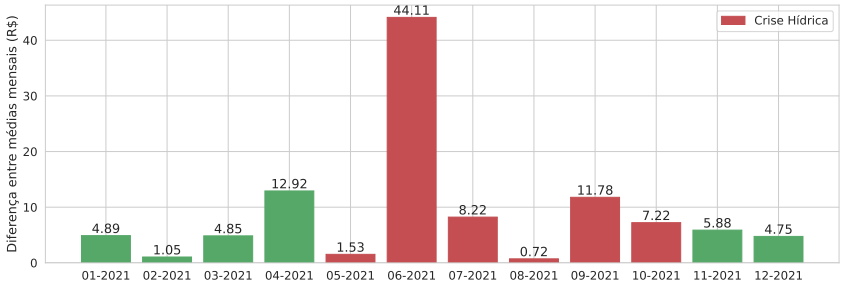
(c) R2.

(d) POCID.

Fonte: O Autor.

Uma análise sintetizada deste comportamento pode ser vista na Figura 99. Neste gráfico é possível observar a diferença entre o PLD médio do mês previsto e o real. Ao considerar a escala de reais (R$/MWh) pode-se perceber, como mostrado também com auxílio da Tabela 15, que o erro médio ao longo de todos os meses apresentado pelo sistema OPB foi de R$ 8,99 e a mediana de R$ 5,38 com desvio padrão de R$ 11,21. Considerando-se que em 2021 o PLD variou entre o mínimo de R$ 49,77 e dois máximos sendo, o estrutural no valor de R$ 583,88 e o horário no valor de R$ 1141,85, pode-se afirmar que o erro médio aferido representou 1,54% do MTE e 0,78% do horário. Esses valores são baixos em relação a alguns dos erros observados hora a hora. Em outras palavras, o sistema tem capacidade preditiva horária que pode apresentar oscilação na escala mais discretizada, no entanto, continua a apresentar bons resultados se avaliado em termos de valores médios.

Figura 99 – Diferença (em R$) entre o PLD médio predito e PLD médio real.



Fonte: O Autor.

Tabela 15 – Sumarização estatística dos erros médios mensais em escala de reais (R$).

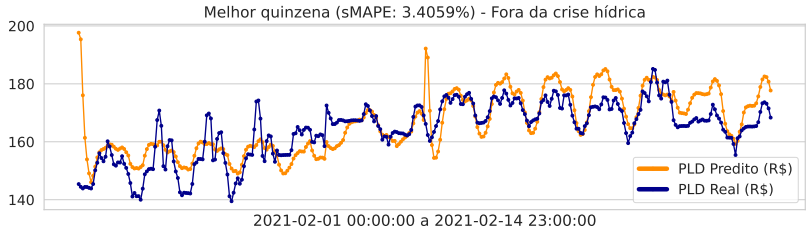
Métrica	Todo período	Fora da crise	Dentro da crise
Média do erro (R$/MWh)	8,99	5,72	12,26
Mediana do erro (R$/MWh)	5,38	4,87	7,72
Desvio padrão do erro (R$/MWh)	11,21	3,56	14,75

Fonte: O Autor.

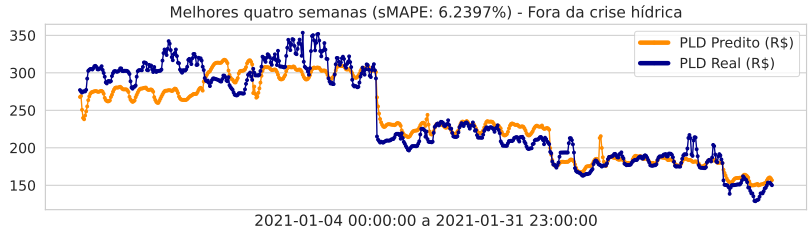
A análise por períodos pode ainda ser complementada pelos gráficos da Figura 100 que mostram as melhores previsões em janelas de 2 e 4 semanas em 2021 quanto ao sMAPE fora da crise hídrica [10]. A melhor quinzena e as melhores quatro semanas apresentaram, respectivamente, valores de sMAPE de 3,4% e 6,23%. Nos piores períodos estes valores sobem para 31,8% e 24,31%. O resultado do pior período de quatro semanas é diluído num espaço de tempo maior. As melhores semanas de previsão aconteceram no mês de janeiro e as piores no mês de março quando uma semana com várias ocorrências do PLD mínimo aconteceu. Em relação a estes resultados, vale relembrar que muitos dos trabalhos citados no Capítulo 3 tem períodos de avaliação inferiores a quatro semanas. Como visto, os erros podem variar substancialmente em intervalos pequenos, por este motivo trabalhos como os de Weron (2014), Nowotarski e Weron (2018) e Lago *et al.* (2021) sugerem que estudos com intervalos inferiores a um ano devem ser encarados com cautela.

[10] Esta avaliação é feita fora da crise hídrica, pois no período do MTE os valores são próximos de zero.

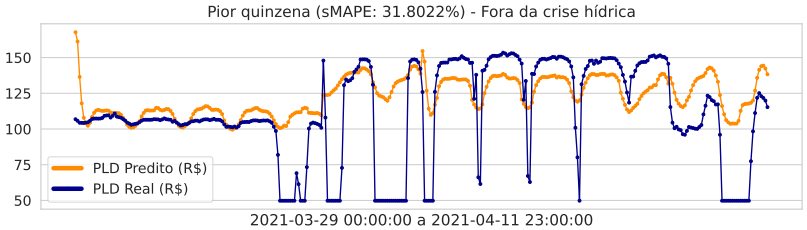
Figura 100 – Melhores e piores períodos de previsão fora da crise hídrica.



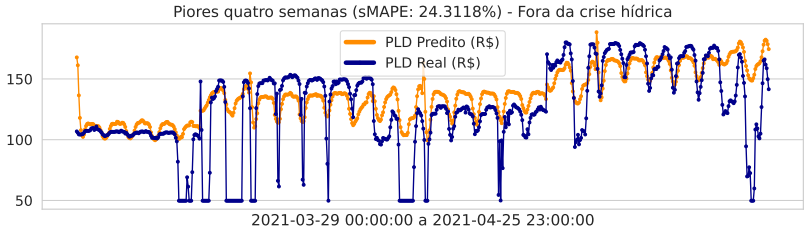
(a) Período de 2 semanas com melhor previsão fora da crise hídrica.



(b) Período de 4 semanas com melhor previsão fora da crise hídrica.



(c) Período de 2 semanas com pior previsão fora da crise hídrica.



(d) Período de 4 semanas com pior previsão fora da crise hídrica.

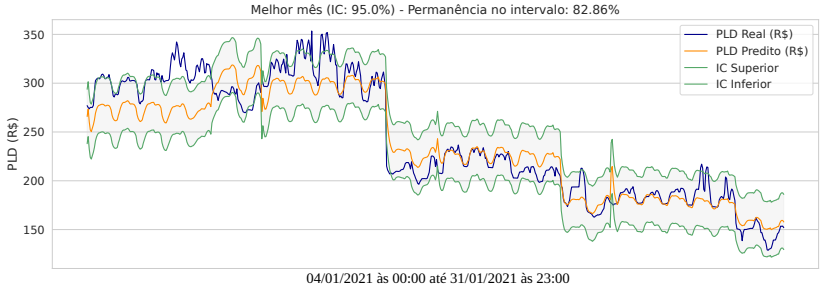
Fonte: O Autor.

Finalmente, em relação a capacidade preditiva do modelo pode-se ainda considerar em um último estudo a incerteza que determina o quão confiável é a previsão ao serem gerados valores de PLD que se mantenham dentro de um intervalo de variação proposto. Similarmente ao visto na Figura 100, estes intervalos em questão podem ser observados para melhores e piores períodos de 4 semanas nos gráficos da Figura 101.

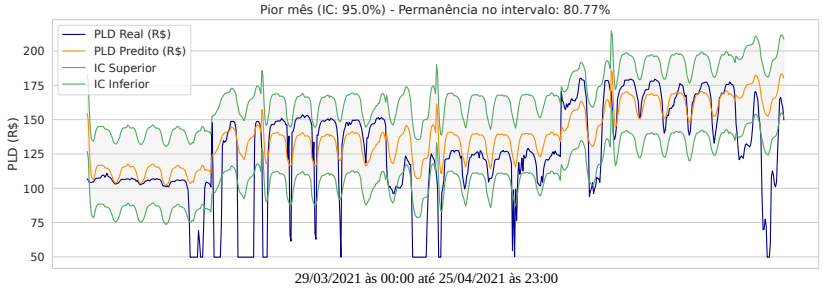
Nas subfiguras *(a)* e *(b)* pode ser visualizado, respectivamente, que o PLD real se manteve variando dentro de um intervalo de confiança de 95% calculado a partir do PLD previsto em 82,86% e 80,77% do tempo levando-se em conta o melhor e o pior mês. Os intervalos de melhor e pior permanência dentro do intervalo de confiança aconteceram em janeiro e março, coincidindo desta forma com os melhores e piores períodos de previsão mostrados na Figura 100. Estes mesmo períodos também são observados quando o cálculo do intervalo é baseado numa banda de *bollinger* com média móvel de 48 horas e dois desvios padrões calculados a partir do PLD real indicada nas subfiguras *(c)* e *(d)*. Nesta forma de avaliação alternativa ao intervalo de confiança pode-se perceber que o PLD previsto se manteve dentro das bandas no melhor e pior mês, respectivamente, por 90,01% e 81,7% do tempo. No caso das bandas de *bollinger* com o aumento da variância observado no PLD real aumenta-se consequentemente as dimensões do intervalo, por este motivo existe maior probabilidade do preço previsto ficar dentro das bandas invertendo as datas de ambos os períodos em relação ao visto para os intervalo de confiança, o que de fato aconteceu.

Ao se considerar todo o ano de 2021, o PLD real se manteve dentro dos intervalos de confiança por 77,32% do tempo, enquanto que o PLD previsto se manteve dentro das bandas de *bollinger* por 74,98%. Apesar das duas métricas informarem situações distintas calculadas a partir das duas fontes de preço separadamente, nota-se que os resultados do tempo de permanência são semelhantes com menos que 3% de diferença entre si.

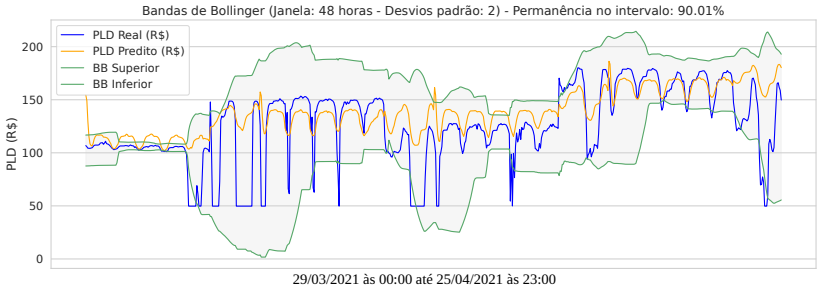
Figura 101 – Melhores e piores períodos de permanência de 4 semanas.



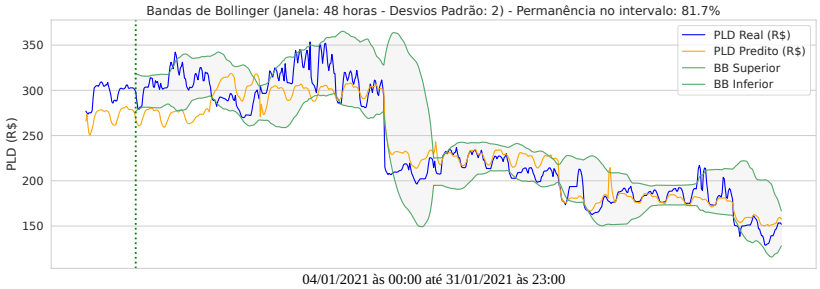
(a) Melhor permanência do PLD real dentro do IC.



(b) Pior permanência do PLD real dentro do IC.



(c) Melhor permanência do PLD previsto dentro das BB.



(d) Pior permanência do PLD previsto dentro das BB.

Fonte: O Autor.

6.3.2 Análise financeira

A análise financeira é baseada nos custos de aquisição hora a hora gerados pelo uso das previsões do PLD calculadas com o sistema OPB e *benchmarks* segundo os algoritmos 1 e 2 discutidos na subseção 5.2.4.2. Em relação aos intervalos de variação do preço e da carga, foram testados cenários de maior ou menor risco envolvendo todas as combinações para margens de 1%, 5%, 10%, 20%, 35% e 50%. A carga considerada como programada foi a curva de 2020 do subsistema SECO e a executada, a curva de 2021. Os valores de carga foram multiplicados por 10^{-4} para facilitar a visualização dos valores nos gráficos e tabelas. Vale lembrar neste ponto, mais uma vez, que este modelo de precificação é hipotético e não retrata as condições do sistema brasileiro. As demais entradas dos algoritmos foram as seguintes:

- Taxa de penalização por erro da carga programada(TPC) = R$ 39,68 $\times$ 1,5
- Taxa de penalização por erro do PLD previsto (TPP) = R$ 39,68 $\times$ 2,5
- Fator percentual de negociação do PLD previsto (FPNPP) = 0,5
- Fator percentual de negociação do PLD real (FPNPR) = 0,8
- Custo exponencial do intervalo de variação do PLD previsto (xp) = 1,3
- Custo exponencial do intervalo de variação da carga programada (xc) = 1,3
- Variável de atenuação (VAT) = 70

O valor de R$ 39,68 utilizado no cálculo de TPC e TPP é referente ao PLD mínimo de 2020. Os parâmetros restantes foram manualmente calibrados a fim de criar uma situação semelhante ao que seria um caso real onde o custo dos intervalos deixa de valer a pena se forem muito largos mas, ao mesmo tempo, se forem muito pequenos, então o erro relacionado às previsões acabaria gerando taxas que justificariam a aquisição das margens. Para fazer sentido, o valor do PLD previsto ajustado precisa ser menor do que o do PLD real ajustado (para bonificar clientes que realizam corretamente sua programação) e, por sua vez, ambos os PLDs ajustados precisam ser inferiores ao PLD real não ajustado (para criar uma situação de vantagem para clientes e fornecedores no uso dos contratos). Como já exposto no capítulo anterior, o objetivo é a identificação de uma combinação ideal de margens que balanceie o custo de se expor ao PLD real (dada as possíveis previsões erradas de preço e carga) e o custo de compra da proteção fornecida pelos intervalos.

O gráfico da Figura 102 mostra os custos de aquisição da energia utilizando-se todos os modelos preditores em todos os cenários de proteção. A primeira coluna diz respeito a uma exposição totalmente sem contrato, onde o custo da carga consumida é valorada diretamente ao PLD real sem qualquer tipo de ajuste praticado na hora do consumo. Nesse sentido, o valor é igual para todos os casos sendo que se o consumidor não tivesse contrato, não dependeria consequentemente de previsões e por isso teria pago o montante de R$ 946,97 em 2021.

Figura 102 – Custos de aquisição absolutos de cada método em relação ao intervalo de variação utilizado para preço e carga.

	Sem contrato	Preço: 1% - Carga: 1%	Preço: 1% - Carga: 5%	Preço: 1% - Carga: 10%	Preço: 1% - Carga: 20%	Preço: 1% - Carga: 35%	Preço: 1% - Carga: 50%	Preço: 5% - Carga: 1%	Preço: 5% - Carga: 5%	Preço: 5% - Carga: 10%	Preço: 5% - Carga: 20%	Preço: 5% - Carga: 35%	Preço: 5% - Carga: 50%	Preço: 10% - Carga: 1%	Preço: 10% - Carga: 5%	Preço: 10% - Carga: 10%	Preço: 10% - Carga: 20%	Preço: 10% - Carga: 35%	Preço: 10% - Carga: 50%	Preço: 20% - Carga: 1%	**Preço: 20% - Carga: 5%**	Preço: 20% - Carga: 10%	Preço: 20% - Carga: 20%	Preço: 20% - Carga: 35%	Preço: 20% - Carga: 50%	Preço: 35% - Carga: 1%	Preço: 35% - Carga: 5%	Preço: 35% - Carga: 10%	Preço: 35% - Carga: 20%	Preço: 35% - Carga: 35%	Preço: 35% - Carga: 50%	Preço: 50% - Carga: 1%	Preço: 50% - Carga: 5%	Preço: 50% - Carga: 10%	Preço: 50% - Carga: 20%	Preço: 50% - Carga: 35%	Preço: 50% - Carga: 50%
OPB Melhor	-946.973	-949.09	-957.85	-988.59	-1090.54	-1285.3	-1503.83	-836.81	-840.23	-862.01	-942.22	-1102.9	-1286.52	-752.48	-752.06	-765.22	-827.54	-956.53	-1110.64	-706.04	-700.19	-705.04	-745.38	-842.46	-968.47	-784.48	-772.51	-772.06	-801.82	-887.46	-1003.26	-906.95	-890.45	-886.13	-909.92	-990.01	-1104.47
OPB	-946.973	-945.44	-953.88	-984.57	-1085.94	-1280.07	-1497.63	-831	-834.59	-855.42	-934.51	-1094.2	-1276.05	-751.84	-751.68	-764.81	-827.08	-955.46	-1109.14	-706.68	-700.81	-705.64	-746.22	-843.11	-969	-783.93	-772.11	-771.63	-800.8	-887.04	-1001.62	-908.91	-892.24	-888.23	-911.7	-992.15	-1107.63
OPB Pior	-946.973	-964.57	-973.64	-1005.62	-1109.79	-1308.37	-1530.96	-860.72	-864.99	-888.51	-971.93	-1138.55	-1328.51	-749.98	-749.19	-762.94	-824.06	-952.96	-1105.98	-712.83	-707.16	-712.04	-754.26	-853.22	-980.03	-783.39	-771.53	-770.41	-800.67	-886.05	-1001.28	-905.47	-889.08	-884.03	-908.1	-987.88	-1102.03
LEAR	-946.973	-944.84	-953.01	-983.89	-1085.52	-1278.35	-1495.97	-821.82	-825.12	-846.43	-924.56	-1081.13	-1261.98	-742.93	-742.02	-755.15	-815.22	-941.12	-1092.64	-713.99	-707.72	-713.17	-753.43	-852.26	-978.72	-794.59	-783.65	-782.87	-813.68	-900.4	-1018.76	-916.48	-900.47	-895.93	-919.68	-1002.53	-1118.33
ANN Média	-946.973	-965.04	-973.85	-1005.81	-1110.24	-1308.64	-1531.01	-866.08	-870.46	-894.07	-978.6	-1146.18	-1337.21	-805.77	-807.57	-823.71	-893.98	-1036.12	-1201.66	-739.4	-733.44	-740.19	-786.07	-891.01	-1023.19	-789.88	-779.01	-777.96	-809.06	-896.24	-1012.76	-905.75	-889.38	-883.69	-908.54	-988.04	-1102.38
XGBoost	-946.973	-971.43	-980.47	-1013.13	-1118.18	-1317.71	-1541.48	-925.98	-932.32	-961.6	-1054.22	-1233.67	-1439.62	-857.49	-860.62	-882.02	-956.64	-1109.29	-1288.29	-748.74	-743.95	-750.89	-796.38	-903.96	-1037.92	-803.66	-792.49	-792.59	-824.31	-914.88	-1035.09	-918.2	-902.21	-897.63	-921.9	-1006.07	-1121.75
ANN Melhor	-946.973	-964.23	-973.2	-1004.9	-1109.28	-1307.94	-1529.95	-875.04	-879.73	-903.68	-989.58	-1159.52	-1353.28	-813.07	-814.81	-832.04	-903.27	-1047.51	-1215.9	-751.83	-746.6	-753.36	-801.57	-909.24	-1044.56	-788.93	-777.67	-777.02	-807.96	-895.36	-1011.79	-909.17	-893.06	-887.78	-913.61	-993.63	-1108.92
ANN Pior	-946.973	-972.24	-981.56	-1014.07	-1119.58	-1319.88	-1544.37	-916.93	-922.54	-950.57	-1043.2	-1223.26	-1427.31	-835.95	-838.25	-857.35	-931.69	-1081.18	-1255.3	-761.95	-757.01	-765.38	-815.17	-926.4	-1063.29	-802.22	-791.1	-791.06	-824.2	-912.46	-1031.91	-915.22	-898.8	-893.34	-919.45	-1000.41	-1116.02
CatBoost	-946.973	-977.24	-986.59	-1019.69	-1126.32	-1326.69	-1552.2	-949.68	-956.93	-987.62	-1084.36	-1269.5	-1482.04	-905.57	-910.14	-934.3	-1019.22	-1182.45	-1373.93	-827.1	-824.12	-835.63	-894.64	-1020.93	-1173.59	-834.92	-824.68	-826.48	-862.7	-958.13	-1083.04	-937.64	-922.18	-918.28	-945.73	-1031.47	-1150.98
Controle	-946.973	-953.96	-962.86	-993.32	-1095.42	-1291.33	-1510.95	-900.93	-906.91	-932.09	-1019.58	-1195.62	-1395.85	-874.87	-878.47	-898.74	-974.64	-1132.82	-1316.5	-845.99	-843.37	-855.54	-911.88	-1041.11	-1196.8	-915.5	-907.65	-913.13	-958.77	-1070.81	-1212.17	-994.18	-979.91	-978.72	-1011.72	-1107.87	-1235.11
SARIMA	-946.973	-969.95	-979.25	-1011.68	-1116.69	-1315.85	-1539.38	-916.11	-922.58	-950.13	-1040.85	-1218.51	-1421.97	-874.72	-878.41	-900.04	-979.3	-1136.25	-1318.78	-846.72	-844.4	-857.63	-917.62	-1046.23	-1202.23	-863.49	-854.15	-857.41	-897.37	-998.2	-1128.88	-946.4	-930.82	-927.59	-956.11	-1044.46	-1164.51
GRU	-946.973	-981.41	-991.17	-1024.3	-1130.95	-1333.4	-1560.16	-968.27	-976.22	-1007.43	-1106.26	-1299.13	-1516.49	-951.01	-956.78	-982.21	-1072.03	-1251.26	-1453.88	-959.16	-959.81	-978.47	-1054.37	-1214.61	-1396.48	-1012.11	-1006.11	-1014.78	-1075.66	-1212.85	-1373	-1098.66	-1086.88	-1088.46	-1136.15	-1257.36	-1403.15
DNN	-946.973	-983.22	-992.83	-1026.14	-1133.1	-1335.96	-1563.21	-973.7	-981.72	-1013.14	-1112.95	-1307.1	-1525.69	-956.36	-962.48	-988.8	-1078.67	-1259.15	-1462.94	-961.03	-961.7	-980.56	-1056.38	-1216.24	-1398.23	-1020.56	-1014.95	-1024.43	-1085.45	-1222.99	-1384.54	-1107.52	-1096.43	-1098.44	-1145.03	-1266.87	-1413.55
LSTM	-946.973	-981	-990.58	-1023.71	-1130.58	-1333.1	-1559.8	-965.79	-973.73	-1004.53	-1103.19	-1296.03	-1512.77	-954.12	-959.87	-985.89	-1076.12	-1256.04	-1459.27	-972.81	-973.67	-992.83	-1070.85	-1232.94	-1418.01	-1046.3	-1040.96	-1052.01	-1116.97	-1261.11	-1428.27	-1146.86	-1135.98	-1139.47	-1193.44	-1323.46	-1477.85
Flat trimestral	-946.973	-1114.85	-1129.18	-1171.34	-1298.7	-1531.8	-1792.1	-1124.38	-1137.76	-1178.67	-1302.56	-1530.35	-1786.55	-1165.35	-1178.37	-1217.75	-1340.11	-1566.47	-1819.37	-1259.84	-1269.39	-1305.68	-1423.24	-1643.22	-1890.16	-1426.02	-1430.67	-1463.38	-1574.61	-1782.89	-2019.7	-1631.45	-1632.45	-1661.11	-1768.65	-1968.88	-2199.21
Flat mensal	-946.973	-1114.12	-1128.37	-1170.46	-1297.84	-1530.72	-1791.04	-1130.94	-1144.5	-1185.93	-1310.92	-1540.04	-1798.14	-1168.7	-1181.92	-1221.42	-1344.71	-1571.52	-1825.54	-1272.95	-1283.09	-1320.26	-1440.36	-1663.51	-1913.83	-1455.29	-1460.38	-1494.14	-1609.51	-1825.77	-2068.39	-1663.42	-1665.48	-1693.98	-1806.45	-2017.02	-2252.88

Fonte: O Autor.

Quanto aos modelos preditores dispostos no eixo vertical na Figura 102, observa-se que o uso de bons preditores minimiza o custo com as taxas relacionadas a possíveis erros na previsão dos preços. Dado que o cenário de carga é o mesmo para todos, então o que se vê em intervalos pequenos (por exemplo, com apenas 1% de proteção no erro do preço e 1% na estimativa da carga) é que baixas proteções podem ter um custo final tão caro quanto com o uso de alta proteção, especialmente se um preditor ruim estiver sendo utilizado. Baixa ou alta proteção em demasia, inclusive, podem impactar tanto no preço que operar sem contrato nestas situações seria mais vantajoso. O cenário ideal visto neste estudo é aquele que indica o uso de um intervalo de variação de 20% para o preço e 5% para a carga utilizando-se o sistema OPB, pois conforme visto entre as linhas destacadas em verde claro é nesta combinação que os menores custos de aquisição foram praticados.

Uma visualização alternativa dos custos de aquisição da Figura 102 é a mostrada na Figura 103 onde o valor da operação sem contrato é setado como 1 e os demais cenários são exibidos como diferenças percentuais em relação a esta referência. Como visto, a operação com o sistema OPB no melhor cenário de proteção é 26% mais vantajosa do que a operação sem contrato. Essa é a maior vantagem entre todas. Com o uso do sistema OPB, mesmo que fosse adquirido um intervalo pequeno de 1% para preço e carga, ainda assim o valor final a ser pago seria no mínimo o mesmo da operação sem contrato. Os custos com o sistema OPB são inferiores a operação sem contrato, inclusive, para a maioria dos cenários de proteção. No entanto, cabe ressaltar que ao se escolher de forma errada um dado intervalo, nem mesmo o melhor preditor seria capaz de reverter a situação de prejuízo no contexto da operação 100% exposta (apesar desta não ser a realidade em nenhum mercado de energia). Isso fica claro, por exemplo, para o caso envolvendo baixa proteção de preço com 1% e alta proteção para a carga com 50%. Este é o pior cenário possível com uso do sistema OPB e nesta situação o resultado foi 58% pior do que a referência. Em relação aos demais *benchmarks*, pode-se notar que o uso de qualquer forma de contrato do tipo *flat* mostrou-se inferior a alternativa de adotar qualquer preditor. Em outras palavras, usar um dado preditor mesmo que este seja ruim ainda assim seria mais vantajoso do que adotar um contrato *flat*. Em relação a operação sem contrato, no entanto, o uso de preditores ruins também pode ser problemático como visto com a aplicação das RNAs puras que em nenhum dos cenários mostraram-se superiores. Os demais preditores, em cenários vantajosos de proteção como o de 20% para preço e 5% para carga apresentaram em sua maioria custos de aquisição melhores do que a referência com economias variando entre 11% e 25%.

Os custos de aquisição hora a hora ao longo de 2021 com todos os preditores para o cenário sem contrato e para o cenário com melhor proteção podem ser visualizados de forma interativa em dois gráficos disponíveis em OPEBRAS (2022).

Figura 103 – Custos de aquisição percentuais de cada método em relação ao intervalo de variação utilizado para preço e carga. Custos baseados em economia ou prejuízo em relação a operação sem contrato exposta ao PLD real conforme primeira coluna.

	Sem contrato	Preço: 1% - Carga: 1%	Preço: 1% - Carga: 5%	Preço: 1% - Carga: 10%	Preço: 1% - Carga: 20%	Preço: 1% - Carga: 35%	Preço: 1% - Carga: 50%	Preço: 5% - Carga: 1%	Preço: 5% - Carga: 5%	Preço: 5% - Carga: 10%	Preço: 5% - Carga: 20%	Preço: 5% - Carga: 35%	Preço: 5% - Carga: 50%	Preço: 10% - Carga: 1%	Preço: 10% - Carga: 5%	Preço: 10% - Carga: 10%	Preço: 10% - Carga: 20%	Preço: 10% - Carga: 35%	Preço: 10% - Carga: 50%	Preço: 20% - Carga: 1%	**Preço: 20% - Carga: 5%**	Preço: 20% - Carga: 10%	Preço: 20% - Carga: 20%	Preço: 20% - Carga: 35%	Preço: 20% - Carga: 50%	Preço: 35% - Carga: 1%	Preço: 35% - Carga: 5%	Preço: 35% - Carga: 10%	Preço: 35% - Carga: 20%	Preço: 35% - Carga: 35%	Preço: 35% - Carga: 50%	Preço: 50% - Carga: 1%	Preço: 50% - Carga: 5%	Preço: 50% - Carga: 10%	Preço: 50% - Carga: 20%	Preço: 50% - Carga: 35%	Preço: 50% - Carga: 50%
OPB Melhor	1	1	1.01	1.04	1.15	1.36	1.59	0.88	0.89	0.91	0.99	1.16	1.36	0.79	0.79	0.81	0.87	1.01	1.17	0.75	0.74	0.74	0.79	0.89	1.02	0.83	0.82	0.82	0.85	0.94	1.06	0.96	0.94	0.94	0.96	1.05	1.17
OPB	1	1	1.01	1.04	1.15	1.35	1.58	0.88	0.88	0.9	0.99	1.16	1.35	0.79	0.79	0.81	0.87	1.01	1.17	0.75	0.74	0.75	0.79	0.89	1.02	0.83	0.82	0.81	0.85	0.94	1.06	0.96	0.94	0.94	0.96	1.05	1.17
OPB Pior	1	1.02	1.03	1.06	1.17	1.38	1.62	0.91	0.91	0.94	1.03	1.2	1.4	0.79	0.79	0.81	0.87	1.01	1.17	0.75	0.75	0.75	0.8	0.9	1.03	0.83	0.81	0.81	0.85	0.94	1.06	0.96	0.94	0.93	0.96	1.04	1.16
LEAR	1	1	1.01	1.04	1.15	1.35	1.58	0.87	0.87	0.89	0.98	1.14	1.33	0.78	0.78	0.8	0.86	0.99	1.15	0.75	0.75	0.75	0.8	0.9	1.03	0.84	0.83	0.83	0.86	0.95	1.08	0.97	0.95	0.95	0.97	1.06	1.18
ANN Média	1	1.02	1.03	1.06	1.17	1.38	1.62	0.91	0.92	0.94	1.03	1.21	1.41	0.85	0.85	0.87	0.94	1.09	1.27	0.78	0.77	0.78	0.83	0.94	1.08	0.83	0.82	0.82	0.85	0.95	1.07	0.96	0.94	0.93	0.96	1.04	1.16
XGBoost	1	1.03	1.04	1.07	1.18	1.39	1.63	0.98	0.98	1.02	1.11	1.3	1.52	0.91	0.91	0.93	1.01	1.17	1.36	0.79	0.79	0.79	0.84	0.95	1.1	0.85	0.84	0.84	0.87	0.97	1.09	0.97	0.95	0.95	0.97	1.06	1.18
ANN Melhor	1	1.02	1.03	1.06	1.17	1.38	1.62	0.92	0.93	0.95	1.04	1.22	1.43	0.86	0.86	0.88	0.95	1.11	1.28	0.79	0.79	0.8	0.85	0.96	1.1	0.83	0.82	0.82	0.85	0.95	1.07	0.96	0.94	0.94	0.96	1.05	1.17
ANN Pior	1	1.03	1.04	1.07	1.18	1.39	1.63	0.97	0.97	1	1.1	1.29	1.51	0.88	0.89	0.91	0.98	1.14	1.33	0.8	0.8	0.81	0.86	0.98	1.12	0.85	0.84	0.84	0.87	0.96	1.09	0.97	0.95	0.94	0.97	1.06	1.18
CatBoost	1	1.03	1.04	1.08	1.19	1.4	1.64	1	1.01	1.04	1.15	1.34	1.57	0.96	0.96	0.99	1.08	1.25	1.45	0.87	0.87	0.88	0.94	1.08	1.24	0.88	0.87	0.87	0.91	1.01	1.14	0.99	0.97	0.97	1	1.09	1.22
Controle	1	1.01	1.02	1.05	1.16	1.36	1.6	0.95	0.96	0.98	1.08	1.26	1.47	0.92	0.93	0.95	1.03	1.2	1.39	0.89	0.89	0.9	0.96	1.1	1.26	0.97	0.96	0.96	1.01	1.13	1.28	1.05	1.03	1.03	1.07	1.17	1.3
SARIMA	1	1.02	1.03	1.07	1.18	1.39	1.63	0.97	0.97	1	1.1	1.29	1.5	0.92	0.93	0.95	1.03	1.2	1.39	0.89	0.89	0.91	0.97	1.1	1.27	0.91	0.9	0.91	0.95	1.05	1.19	1	0.98	0.98	1.01	1.1	1.23
GRU	1	1.04	1.05	1.08	1.19	1.41	1.65	1.02	1.03	1.06	1.17	1.37	1.6	1	1.01	1.04	1.13	1.32	1.54	1.01	1.01	1.03	1.11	1.28	1.47	1.07	1.06	1.07	1.14	1.28	1.45	1.16	1.15	1.15	1.2	1.33	1.48
DNN	1	1.04	1.05	1.08	1.2	1.41	1.65	1.03	1.04	1.07	1.18	1.38	1.61	1.01	1.02	1.04	1.14	1.33	1.54	1.01	1.02	1.04	1.12	1.28	1.48	1.08	1.07	1.08	1.15	1.29	1.46	1.17	1.16	1.16	1.21	1.34	1.49
LSTM	1	1.04	1.05	1.08	1.19	1.41	1.65	1.02	1.03	1.06	1.16	1.37	1.6	1.01	1.01	1.04	1.14	1.33	1.54	1.03	1.03	1.05	1.13	1.3	1.5	1.1	1.1	1.11	1.18	1.33	1.51	1.21	1.2	1.2	1.26	1.4	1.56
Flat trimestral	1	1.18	1.19	1.24	1.37	1.62	1.89	1.19	1.2	1.24	1.38	1.62	1.89	1.23	1.24	1.29	1.42	1.65	1.92	1.33	1.34	1.38	1.5	1.74	2	1.51	1.51	1.55	1.66	1.88	2.13	1.72	1.72	1.75	1.87	2.08	2.32
Flat mensal	1	1.18	1.19	1.24	1.37	1.62	1.89	1.19	1.21	1.25	1.38	1.63	1.9	1.23	1.25	1.29	1.42	1.66	1.93	1.34	1.35	1.39	1.52	1.76	2.02	1.54	1.54	1.58	1.7	1.93	2.18	1.76	1.76	1.79	1.91	2.13	2.38

Fonte: O Autor.

A análise referente ao montante anual dos custos de aquisição pode ser ainda estratificada para uma discretização horária e, dessa forma, serem calculados os custos de aquisição esperados com uso do VaR e CVaR. Os valores da Tabela 16 dizem respeito ao desempenho de cada preditor quando da aplicação do melhor cenário para montante mínimo com intervalo de 20% para preço e 5% para carga. O nível de significância α utilizado foi de 5%. Nota-se que os custos relativos a distribuição dos dados no ponto de corte do VaR são mais vantajosos com uso do sistema OPB, com dois casos do modelo ANN, LEAR e XGBoost. Em relação ao custo esperado na média dos menores valores definidos pelo CVaR vê-se novamente o sistema OPB e o modelo LEAR com vantagem. Neste estudo, o sistema OPB apresentou uma expectativa de custo de aquisição horário 54,83% e 56,25% menor do que obteria o usuário com o pior *benchmark*, respectivamente, em VaR e CVaR.

Tabela 16 – VaR e CVaR com o melhor intervalo de variação para preço (20%) e carga (5%). Negrito e amarelo destacam os melhores modelos.

Modelo	**VaR (5%)**	**Modelo**	**cVaR (5%)**
OPB Melhor	**-0.17**	OPB Melhor	**-0.18**
OPB Pior	**-0.17**	OPB Pior	**-0.18**
OPB	**-0.17**	OPB	**-0.18**
ANN Pior	**-0.17**	LEAR	**-0.18**
ANN Média	**-0.17**	ANN Pior	-0.19
LEAR	**-0.17**	ANN Média	-0.19
XGBoost	**-0.17**	XGBoost	-0.2
ANN Melhor	-0.18	Controle	-0.21
Controle	-0.18	CatBoost	-0.21
CatBoost	-0.18	ANN Melhor	-0.22
GRU	-0.19	GRU	-0.23
LSTM	-0.19	LSTM	-0.23
DNN	-0.19	DNN	-0.24
SARIMA	-0.21	SARIMA	-0.26
Flat trimestral	-0.31	Flat trimestral	-0.32
Flat mensal	-0.31	Flat mensal	-0.32

Fonte: O Autor.

Esta análise pode ainda ser complementada com a avaliação dos valores de VaR e CVaR para cada caso de proteção com uso do sistema OPB como visto na Tabela 17. É interessante inicialmente perceber que, apesar de muito próximos, os menores custos de aquisição esperados do VaR não necessariamente coincidem com o caso de melhor proteção. Como visto nesta tabela e também nos gráficos da Figura 104, a proteção envolvendo intervalos de preço com 10% de margem foi ligeiramente mais vantajosa do que aquela com 20%. Isso acontece, pois as distribuição dos custos e o ponto de corte do VaR não estão diretamente relacionados com o montante total sendo que, nesta situação, a classe com maior frequência tem mais relevância. Assim, a escolha do melhor intervalo de proteção pode considerar também não somente o montante, como também

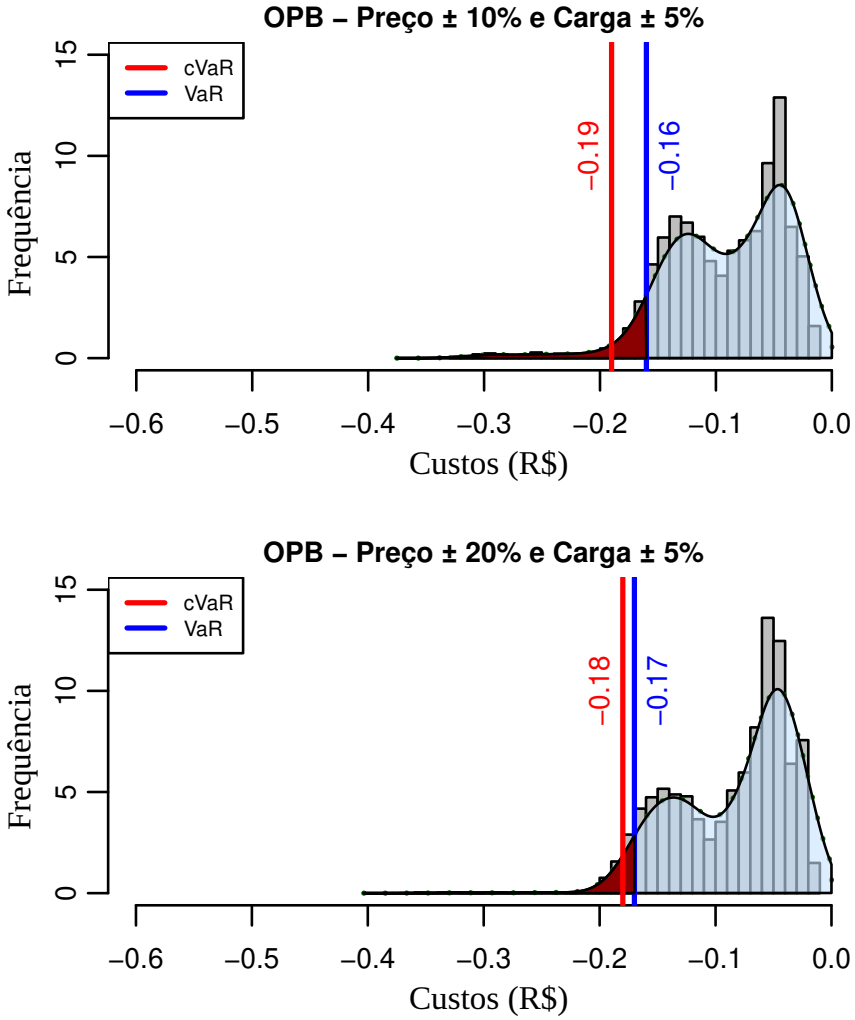
a expectativa de custo horário que neste caso poderia conduzir à decisão de se adotar um intervalo para preço um pouco menor.

Tabela 17 – VaR e CVaR para todas combinações de intervalos de variação de preço e carga com previsões feitas pelo sistema OPB. Negrito e amarelo destacam as melhores combinações.

Intervalos	**VaR (5%)**	**cVaR (5%)**
Preço ± 1% - Carga ± 1%	-0.22	-0.24
Preço ± 1% - Carga ± 5%	-0.23	-0.25
Preço ± 1% - Carga ± 10%	-0.24	-0.26
Preço ± 1% - Carga ± 20%	-0.26	-0.29
Preço ± 1% - Carga ± 35%	-0.31	-0.34
Preço ± 1% - Carga ± 50%	-0.36	-0.4
Preço ± 5% - Carga ± 1%	-0.18	-0.23
Preço ± 5% - Carga ± 5%	-0.18	-0.23
Preço ± 5% - Carga ± 10%	-0.19	-0.24
Preço ± 5% - Carga ± 20%	-0.21	-0.26
Preço ± 5% - Carga ± 35%	-0.25	-0.31
Preço ± 5% - Carga ± 50%	-0.29	-0.37
Preço ± 10% - Carga ± 1%	**-0.16**	-0.19
Preço ± 10% - Carga ± 5%	**-0.16**	-0.19
Preço ± 10% - Carga ± 10%	-0.17	-0.21
Preço ± 10% - Carga ± 20%	-0.18	-0.22
Preço ± 10% - Carga ± 35%	-0.21	-0.26
Preço ± 10% - Carga ± 50%	-0.25	-0.32
Preço ± 20% - Carga ± 1%	-0.17	**-0.18**
Preço ± 20% - Carga ± 5%	-0.17	**-0.18**
Preço ± 20% - Carga ± 10%	-0.17	**-0.18**
Preço ± 20% - Carga ± 20%	-0.18	-0.19
Preço ± 20% - Carga ± 35%	-0.21	-0.23
Preço ± 20% - Carga ± 50%	-0.24	-0.26
Preço ± 35% - Carga ± 1%	-0.19	-0.2
Preço ± 35% - Carga ± 5%	-0.19	-0.2
Preço ± 35% - Carga ± 10%	-0.19	-0.2
Preço ± 35% - Carga ± 20%	-0.2	-0.21
Preço ± 35% - Carga ± 35%	-0.23	-0.24
Preço ± 35% - Carga ± 50%	-0.26	-0.27
Preço ± 50% - Carga ± 1%	-0.23	-0.24
Preço ± 50% - Carga ± 5%	-0.22	-0.23
Preço ± 50% - Carga ± 10%	-0.22	-0.23
Preço ± 50% - Carga ± 20%	-0.23	-0.24
Preço ± 50% - Carga ± 35%	-0.26	-0.27
Preço ± 50% - Carga ± 50%	-0.29	-0.3

Fonte: O Autor.

Figura 104 – Comparativo dos valores de VaR e CVaR com proteção de 10% e 20% no preço.



Fonte: O Autor.

6.3.2.1 Comparação estatística

Finalmente, nesta análise é conduzido um último estudo no qual buscou-se realizar uma comparação entre intervalos de confiança derivados com teste não paramétrico de Nemenyi[11] conforme discutido na subseção 5.2.4.2. Foram considerados como conjuntos amostrais cada uma das séries de custos de aquisição hora a hora com cada preditor ao longo de 2021. O cenário de proteção adotado mais uma vez foi aquele que gerou os menores montantes totais com 20% para preço e 5% para carga.

Na Tabela 18 pode ser visualizada uma sumarização estatística de cada conjunto. Nesta tabela percebe-se que o sistema OPB é aquele com menor custo médio horário e que o modelos ANN e LEAR isolados não são capazes de superar a combinação dos dois na forma do *ensemble*. Em relação a mediana todos os preditores de forma geral apresentaram valores próximos com custo de aquisição se repetindo mais sob o valor de 0,07. Como o LEAR trata-se de uma curva

[11] Após aplicação de um teste de Kruskal-Wallis para detecção de diferenças que retornou p-valor inferior a 0,05 indicando rejeição da hipótese nula.

de regressão, então como esperado seu desvio padrão foi o menor entre todos. O custo mínimo foi visto junto ao modelo Controle e o máximo é idêntico para todos, com exceção do *flat* trimestral.

Tabela 18 – Sumarização estatística dos retornos dos modelos. Amarelo destaca o sistema OPB. Cinza destaca os modelos isolados LEAR e ANN média. Negrito indica os melhores resultados em cada coluna.

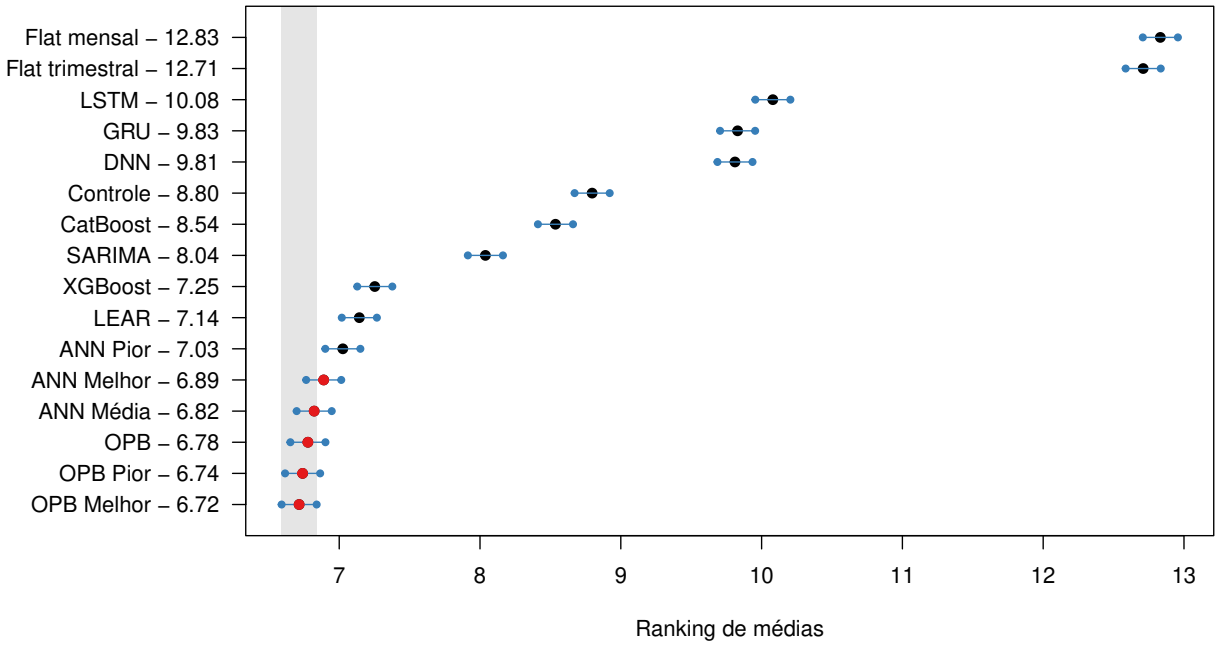
Modelo	**Média**	**Mediana**	**DP**	**Mínimo**	**Máximo**
OPB Melhor	**-0.0817**	**-0.07**	0.0494	-0.36	**-0.01**
OPB	-0.0818	**-0.07**	0.0492	-0.36	**-0.01**
OPB Pior	-0.0825	**-0.07**	0.0497	-0.36	**-0.01**
LEAR	-0.0826	**-0.07**	**0.0476**	-0.36	**-0.01**
ANN Média	-0.0856	**-0.07**	0.0539	-0.37	**-0.01**
XGBoost	-0.0868	**-0.07**	0.0494	-0.36	**-0.01**
ANN Melhor	-0.0871	**-0.07**	0.0562	-0.37	**-0.01**
ANN Pior	-0.0884	-0.08	0.0516	-0.36	**-0.01**
CatBoost	-0.0962	-0.09	0.0491	-0.36	**-0.01**
Controle	-0.0984	-0.09	0.0513	**-0.33**	**-0.01**
SARIMA	-0.0986	-0.08	0.0603	-0.36	**-0.01**
GRU	-0.112	-0.11	0.0491	-0.37	**-0.01**
DNN	-0.1122	-0.11	0.0484	-0.37	**-0.01**
LSTM	-0.1136	-0.11	0.0485	-0.36	**-0.01**
Flat trimestral	-0.1482	-0.12	0.0846	-0.37	-0.02
Flat mensal	-0.1498	-0.12	0.0829	-0.37	**-0.01**

Fonte: O Autor.

O gráfico da Figura 105 mostra a comparação dos intervalos de confiança com sobreposição, nesta imagem aqueles que apresentam intercessão não são estatisticamente significantes entre si. Uma visualização alternativa também pode ser vista na matriz de significância da Figura 106 onde o bloco branco indica superioridade do modelo no eixo vertical sobre o do eixo horizontal.

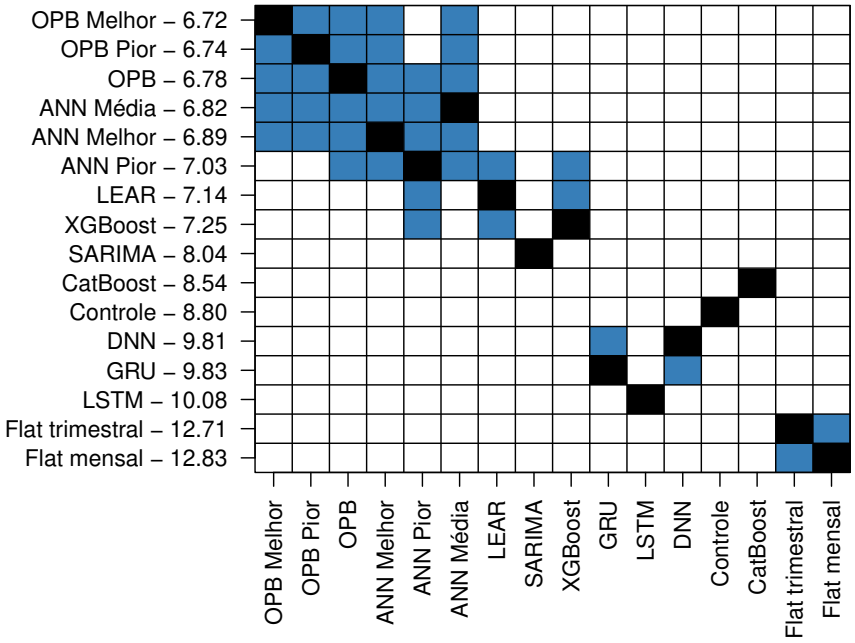
Como visto, estatisticamente não é possível afirmar que existem evidências de que o melhor caso gerado pelo sistema OPB é diferente do seu caso médio e nem do seu pior caso. O mesmo acontece em relação ao melhor caso e a média dos modelos ANN. Em outras palavras, na ponta dos melhores modelos estes cinco preditores apresentaram a mesma capacidade de economia financeira. No entanto, são superiores aos demais. Em relação ao LEAR isoladamente, o mesmo mostrou-se igual ao pior caso do modelo ANN e ao XGBoost. O SARIMA é inferior a todos os modelos citados até aqui e superior aos outros, aparecendo de forma isolada no *ranking* de médias. A situação é a mesma para o modelo Controle e o CatBoost que entre si são iguais. Por fim, RNAs puras e contratos *flat* são inferiores a todos os *benchmarks* testados mas não diferentes entre si em seus grupos.

Figura 105 – Comparação dos intervalos de confiança calculados pelo teste de Nemenyi.



Fonte: O Autor.

Figura 106 – Matriz de significância estatística dos modelos com teste de Nemenyi. O bloco branco indica superioridade do modelo do eixo vertical sobre o do eixo horizontal.



Fonte: O Autor.

7 CONCLUSÃO

Neste trabalho foi avaliada a aplicação de um *ensemble* baseado em dois modelos preditores estado da arte na previsão de preços de eletricidade no mercado brasileiro. A pesquisa teve foco em duas frentes de trabalho, sendo a primeira aquela que trata do desenvolvimento de um sólido levantamento bibliográfico sobre o campo da pesquisa e, a segunda, do desenvolvimento de um produto de inovação tecnológica capaz de realizar a previsão do PLD horário no subsistema SECO.

O levantamento bibliográfico cobriu inicialmente uma etapa de contextualização contendo conceitos básicos relacionados a grande área de pesquisa denominada economia da energia. Na sequência foram exploradas as estruturas dos mercados de eletricidade ao redor do mundo por meio do histórico que abarca toda a experiência internacional desde os primórdios da indústria de eletricidade até tópicos atuais como redes inteligentes, penetração de renováveis e fontes alternativas. Com foco específico no mercado brasileiro, ainda durante a contextualização foi realizado um estudo do atual panorama do SIN buscando-se apresentar suas principais características e, adicionalmente, como o mercado brasileiro ao longo dos anos veio se organizando no sentido de cada vez mais se tornar um ambiente de negociação competitivo. Levando-se em conta também o caráter político e os principais eventos históricos discutidos, pode-se concluir que há uma tendência do mercado brasileiro rumo a um ambiente que favoreça de forma geral às operações desreguladas. Já houveram, no entanto, tentativas de implantação deste tipo de modelo em variados momentos de reforma desde a década de 30 e, mesmo que existam relatórios e posicionamentos abertos de agentes a favor da liberalização, o que se vê de fato é que historicamente quando isto aconteceu foram necessárias intervenções estatais para conter vícios como poder de mercado e outras ações que criaram artificialidades nos preços. Essa dicotomia teve como principal resultado a transformação do mercado brasileiro em um ambiente complexo devido a massiva microrregulação que ano após ano modifica os processos de operação e tem impacto muitas vezes imprevisíveis na saúde do mercado elétrico. No entanto, entende-se também que em muitas vezes a microrregulação é a alternativa que existe para suportar as complexidades do mercado, ainda que esta tenha um custo.

Este ponto apresentado fica mais claro, por exemplo, no recente movimento relacionado a um dos objetivos estabelecidos com a promulgação da portaria MME No. 33/2017 (MME, 2017) que era de que a adoção do PLD horário no Brasil permitiria maior acoplamento entre o custo da operação física do sistema e o preço da energia praticado. Desta forma, criando um mecanismo de sinalização econômica confiável superior ao preço semanal que, historicamente, não foi capaz de acompanhar a diversificação da matriz energética criando anomalias e gerando prejuízos especialmente nos horários de pico de consumo. Visto que todos os custos do desacoplamento entre CMO e PLD são rateados entre os agentes do mercado na forma de encargos, pode-se afirmar que a implantação da granularidade horária ao ser testada pela ocorrência da crise hídrica em 2021 não foi suficientemente capaz de resolver este problema. Com intervenções por parte das agências reguladoras, houveram momentos onde o descolamento entre CMO e PLD médio foi de até 330%. Essa

situação tende a incentivar a celebração de contratos de longo prazo, dado que nestes contratos os riscos são mitigados e a exposição ao PLD é reduzida. No entanto, cabe ressaltar que contratos de longo prazo diminuem a relevância do MCP ignorando dessa forma as preferências do consumidor. Esta situação prejudica o estabelecimento de um mercado sustentável, pois indica a existência de artificialidade no preço (deve-se lembrar que em muitos casos os próprios agentes participam do processo de regulação, não sendo este um mecanismo coercitivo necessariamente). A artificialidade distorce a visão dos agentes na tomada de decisão tanto no curto prazo para minimização de custos, por exemplo, quanto no longo prazo, como no contexto de investimentos em expansão da capacidade. Neste sentido, conclui-se que neste primeiro ano de adoção do PLD horário pôde-se ver um avanço nas políticas regulatórias de preço da energia no Brasil, entretanto, se a intenção for de fato o estabelecimento de um mercado livre sustentável, então a formação do preço precisará avançar ainda mais a fim de diminuir as inconsistências entre a operação física do sistema e o PLD. Caso este objetivo seja atingido, espera-se que as solicitações feitas ao governo pelos agentes de mercado durante a operação sombra[1] possam finalmente serem atendidas de modo a favorecer a transparência, liquidez, redução de encargos, gestão de riscos e, principalmente, o surgimento de novos negócios, produtos e serviços como, no interesse desta pesquisa, de ferramentas como o sistema OPB.

Em relação a revisão de literatura foram levantados trabalhos que discutem a evolução do campo das RNAs desde seus primórdios até a atual área de DL no âmbito da previsão de preços de eletricidade. Nesta etapa foi realizada uma discussão acerca da aplicabilidade das RNAs modernas em séries temporais, com foco especial naquelas que fazem uso de mecanismos de atenção. Quanto a esta análise o que se observou é que existe uma crescente adoção dos modelos atencionais. No entanto, ainda é um campo pouco explorado. As duas últimas arquiteturas deste tipo foram publicadas em 2014 e 2017, ambas voltadas para processamento de linguagem natural e imagens. Com base no levantamento realizado, o que se viu é que a previsão em séries temporais com Redes Neurais Atencionais está começando a ser explorada apenas nos últimos anos e tem se mostrado uma promissora direção de pesquisa.

Na sequência a área ampla da previsão energética foi abordada por meio de uma análise bibliométrica e meta-análise de *surveys* publicados nos últimos 10 anos. Neste ponto, conclui-se que muitos dos jargões e termos utilizados atualmente neste tipo de pesquisa são ainda relativamente confusos, sendo que denominações diferentes são utilizadas por diversos autores. Estudos recentes inclusive propuseram taxonomias específicas. No entanto, ainda estão limitados a escopos fixos, como o de previsão de carga. Nesse sentido, visando preencher esta lacuna foi proposto um posicionamento específico para a previsão energética que, sendo uma área interdisciplinar, embarca ao mesmo tempo fundamentos de economia, computação, engenharia elétrica e estatística. A falta deste posicionamento até o momento provavelmente foi motivada pela relativa novidade que são mercados totalmente liberalizados e, também, devido a dificuldade na disponibilização de dados públicos que somente nos últimos anos vem se tornando realidade em decorrência da evolução

[1] Descritas em (EVELINA *et al.*, 2018)

dos mecanismos de acesso à internet, do incentivo à transparência em atividades financeiras, a evolução dos instrumentos de medição, ao desenvolvimento de técnicas de *big data* e outros.

Identificou-se também que trabalhos de previsão de preços de eletricidade são menos numerosos dos que os vistos em outras áreas da previsão energética como previsão de carga e eólica. Os que usam o mercado brasileiro como estudo de caso são praticamente inexistentes. No escopo do levantamento realizado entre 1994 e 2021 não foi encontrada nenhuma abordagem entre todas as referências que contemplasse simultaneamente pontos importantes empreendidos nesta tese como a previsão com discretização horária, previsão para 168 horas à frente, o uso combinado dos métodos preditores ANN e LEAR, aplicação no mercado nacional e o tamanho de um ano adotado como intervalo de teste. Cabe ressaltar ainda que mesmo nos trabalhos que utilizaram períodos de um ano não foi visto nenhum que envolvesse momentos de instabilidade, como os verificadas no Brasil em 2021 decorrentes da maior crise hídrica no SIN dos últimos 91 anos e a implantação oficial do PLD horário.

Uma última e importante conclusão acerca da revisão bibliográfica é a notória existência de problemas em estudos relacionados à área de previsão energética que, sistematicamente, vem aplicando ao longo dos anos modelos preditores que necessitam de treinamento (como RNAs) de forma errada. Muitos trabalhos relatam metodologias onde os autores treinam modelos nos mesmos conjuntos em que os testam, apresentam resultados em períodos inferiores muitas vezes a um semestre (alguns com apenas um mês ou um dia), não fazem uso de qualquer tipo de análise estatística na avaliação dos resultados e outros. Essa situação, inclusive, tem criado espaço para a publicação de artigos focados na disponibilização de *datasets* públicos prontos para uso e roteiros que devem ser minimamente seguidos em novas pesquisas. Adicionalmente, a metodologia de testes utilizada aqui é encorajada a ser seguida, pois antes de tudo, baseia-se nestes roteiros mais recentemente publicados.

Quanto ao sistema OPB foi realizada uma comparação dos seus resultados contra 16 *benchmarks* em termos de acurácia e financeiros. Estes *benchmarks* são em sua maioria estado da arte em diversas áreas de aplicação, inclusive na previsão de séries temporais. Outros são, no mínimo, amplamente utilizados como o método do dia similar adotado em algumas situações pela ONS (ONS, 2019) e os contratos *flat* adotados comumente por comercializadoras e consumidores livres no Brasil. Durante 356 dias (8544 horas) com uso de dados reais, o sistema OPB foi submetido a previsão semanal do PLD horário no subsistema SECO a partir de dados históricos do sistema elétrico, informações econômicas, medições climáticas das estações meteorológicas do INMET e condições futuras de calendário. O modelo foi treinado e otimizado em 8 anos de dados desde 2013 sendo composto por uma parcela redutora de dimensionalidade com arquitetura AE, esta capaz de comprimir os dados a 26,66% do tamanho original mantendo 99,99% da informação. A parcela previsora foi composta de uma combinação de dois modelos, sendo o primeiro, uma arquitetura ANN Seq2Seq WAD que se mostrou especializada na modelagem de padrões finos da série temporal como àqueles que determinam horários de pico, pontos atípicos e sazonalidade e, a segunda, um modelo regressor LEAR que foi capaz de capturar com precisão as características

estatísticas da série como média e mediana. A combinação dos dois num *ensemble* mostrou-se uma estratégia acertada, pois os pontos fortes de um modelo compensaram as deficiências do outro.

Quanto a capacidade preditiva, o sistema OPB foi superior a todos os *benchmarks* em grande parte das análises empreendidas, especialmente naquelas envolvendo a métrica sMAPE que para um mesmo mercado e conjuntos com mesmo tamanho, é a mais indicada. O sistema obteve sMAPE de 3,4%, 6,23% e 11,96%, respectivamente, na melhor quinzena, no melhor mês e no ano inteiro. A pior quinzena e o pior mês tiveram sMAPE de 31,8% e 24,31%. Estes valores colocam o sistema OPB como o melhor preditor em todo o período, o melhor fora da crise hídrica e o segundo melhor dentro dela. Em relação a sua capacidade individual de previsão, destaca-se o momento em que o mesmo foi capaz de indicar com antecedência o movimento de alta de 50% no valor do PLD em junho e a saída da crise hídrica em outubro, além da previsão correta do MTE na maior parte do tempo em que este vigorou. Quanto ao movimento dos preços, o sistema OPB manteve-se com boas taxas de acerto de direção, sendo que o POCID mensal médio foi de 79%. Quanto ao erro mensal médio destaca-se a conversão desta métrica para escala de reais (R$/MWh) que indicou os valores de R$ 8,99, R$ 5,72 e R$ 12,26, respectivamente, para todo ano de 2021, para o período fora da crise e dentro da crise. O erro mensal médio em 2021 representou apenas 1,54% do MTE. Por fim, pode-se destacar que o PLD real se manteve por 77,32% do tempo dentro de um intervalo de confiança calculado a partir do PLD previsto e, simultaneamente, o PLD previsto se manteve por 74,98% do tempo dentro de um intervalo de bandas de *bollinger* calculado sob o PLD real.

Em relação a análise financeira, conclui-se que o sistema OPB foi capaz de gerar valores mais econômicos de custos de aquisição do que aqueles gerados pelos *benchmarks* sob as mesmas condições de experimentação quando considerando-se a curva de carga de 2020 do subsistema SECO como a programação de consumo e a curva de 2021 como o consumo efetivamente realizado. Neste ponto cabe destacar a proposta de algoritmo de precificação do custo da energia incluída no sistema OPB que leva em consideração, principalmente, a capacidade de criação um cenário que simule as possíveis vantagens envolvendo situações contratuais *vs.* sem contrato. O algoritmo também estima o custo e o efeito da proteção ao risco garantido por intervalos de variação para prováveis erros na previsão do preço ou na estimativa da carga a ser consumida. Neste contexto, foi possível perceber que com o uso do sistema OPB garantiu-se 26% de economia em relação a uma operação sem contrato quando utilizando-se o melhor cenário de proteção para preço e carga. Esse percentual é o mais expressivo entre todos os *benchmarks* e todos os cenários testados. Esse resultado impactou também na minimização dos custos unitários de aquisição calculados via VaR e CVaR, sendo que neste estudo, o sistema OPB apresentou uma expectativa de custo de aquisição horário 54,83% e 56,25% menor do que obteria o usuário se utilizasse o pior *benchmark*, respectivamente, em VaR e CVaR. Por fim, foi aplicado um teste estatístico para comparação de diferenças utilizando-se como conjunto amostral os custos de aquisição hora a hora. O teste indicou que com um intervalo de confiança de 95% existem evidências para considerar o sistema OPB (incluindo o seu melhor e pior caso isolados), o melhor caso do modelo ANN e a média de todos os modelos ANN como iguais entre si, mas estatisticamente diferentes e superiores a todos os demais preditores.

Neste contexto, a principal conclusão relacionada a capacidade preditiva do sistema OPB é que os modelos ANN e LEAR não são superiores entre si, mas quando combinados na forma de um *ensemble* baseado em média, apresentam capacidade de previsão do PLD com performance superior àquela vista quando utilizados individualmente e, também, aos demais *benchmarks*. Em outras palavras, torna-se oportuno resgatar a hipótese da pesquisa que diz:

> A qualidade da previsão para uma semana à frente do PLD horário depende do método de previsão adotado. Se o método adotado for um *ensemble* baseado em Redes Neurais Atencionais e LEAR então esta previsão torna-se mais acurada do que àquela realizada por outros métodos individualmente aplicados.

Assim, dentro das condições de experimentação adotadas, conclui-se que o método como variável independente exerce influência na qualidade das previsões e que, a previsão como variável dependente, possui mais qualidade quando utilizando-se o sistema aqui proposto. Tanto da perspectiva de acurácia como financeira.

7.1 Limitações e propostas de continuidade

A principal limitação do sistema OPB é a dependência de informações públicas e, por consequência, estar diretamente vinculado a divulgação de dados feitas por agências terceiras. Hoje, algumas destas dispõe parte das suas operações *online* devido a obrigatoriedade jurídica imposta pelo programa de transparência e dados abertos do governo federal. Contudo, a maioria delas é independente e não governamental. Logo, a possibilidade do acesso tornar-se restrita a qualquer momento é real e deve ser considerada. Neste caso, deve-se prever esforço em prováveis e necessárias modificações no modelo, além de um *roadmap* de desenvolvimento constante visando a construção de versões alternativas. Este é um problema relevante e possivelmente difícil de ser contornado, mas que não é exclusivo desta pesquisa sendo comum em trabalhos que fazem modelagem de informações econômicas.

É também limitação desta pesquisa a dependência de um mercado que não pratique um preço artificial, ou seja, onde não exista a distribuição isonômica da informação a todos os agentes de modo a gerar uma série temporal que siga um passeio aleatório. Hoje o mercado brasileiro é um mercado que sofre intervenções de agências centralizadas de coordenação e, no nível atual, o sistema OPB foi capaz de realizar as previsões. No entanto, possíveis modificações regulatórias podem alterar as atuais condições e, dessa forma, por mais que a ferramenta (ou qualquer outra técnica de modelagem) seja capaz de se adaptar aos novos padrões, é necessário antes de tudo que padrões existam. Do contrário, incorreria numa situação de passeio aleatório onde a atividade de previsão se torna impossível conforme preconizado pela mais básica teoria de sistemas dinâmicos.

Finalmente, apesar de transparência e isonomia serem condições fundamentais para saúde dos mercados financeiros, muitos dos mercados elétricos que evoluíram para o varejo hoje adotaram

a estratégia de cobrar pelos dados como aconteceu, por exemplo, com o NordPool que foi um dos primeiros a disponibilizar uma plataforma *online* mas que a partir de 2020 exige o pagamento de uma mensalidade para total acesso as informações de mais de dez países europeus. Dado que o desenvolvimento do sistema OPB foi todo realizado no contexto de um programa de pós-graduação gratuito, sem investimentos externos e com uso eventual de recursos do seu próprio desenvolvedor na aquisição de equipamentos (como processadores e placas GPU de alta performance) hoje é limitação da pesquisa o acesso a recursos de custeio que viabilizariam sua continuidade.

Outras limitações com conseguintes propostas são:

a) Levando-se em consideração que o sistema OPB em muitos momentos não foi capaz de modelar com precisão os valores do PLD em feriados, sugere-se que esta limitação seja corrigida por meio do desenvolvimento de um modelo auxiliar desenvolvido especificadamente para este tipo de data.

b) Dado que dentro do período da crise hídrica o sistema OPB não apresentou resultados tão bons quanto fora dela, pode-se empreender o desenvolvimento de um modelo específico treinado apenas em períodos históricos com características atípicas. Este modelo pode ser acionado quando um monitor de anomalias indicar modificações bruscas nos padrões estatísticos de comportamento do PLD.

Podem ser ainda empreendidas no futuro as seguintes ações de aprimoramento:

a) Aproveitamento da capacidade que o sistema OPB demonstrou em adiantar certos movimentos e acoplar ao mesmo um módulo gerador de linhas de tendência que indiquem direções de movimentos futuros sem considerar necessariamente a importância de valores recentes.

b) Avaliação da qualidade com a inclusão de variáveis de previsão do tempo que hoje não são utilizadas. As previsões do tempo poderiam ainda ser utilizadas para aprimorar a previsão probabilística gerando-se modelos que recebam pequenas variações em cada entrada relativa a clima futuro.

c) Elaborar um esquema de infraestrutura que permita a construção automatizada de *builds* dos modelos, fornecendo assim em tempo real as mais recentes previsões baseados em modelos retreinados semanalmente.

d) Averiguação de possível aumento na qualidade preditiva utilizando-se modelos que tenham como primeiro dia da previsão os dias da semana diferentes de segunda-feira. Idealmente trabalhando-se com sete modelos, um para cada dia da semana e dessa forma provendo modelos específicos dia a dia que podem aproveitar como entrada o CMO para o dia seguinte publicado pela ONS que já consta no banco de dados do módulo M1 mas que até o momento não foi utilizado.

e) Investigação da qualidade preditiva em janelas móveis com tamanhos diferentes como com um dia, dois dias, duas semanas e até um mês à frente.

f) Investigação da aplicação do sistema OPB em outros subsistemas do SIN.

g) Uso do módulo M3 para previsão também de carga e geração, garantindo dessa forma um cenário previsto no módulo M4 mais completo. Internamente estas duas variáveis já são previstas, neste sentido, as previsões de carga e geração podem ser extraídas e usadas para se calcular prováveis picos de preço decorrentes de desequilíbrios na margem de reserva. Esse padrão em relação a margem de reserva foi verificado durante o desenvolvimento desta pesquisa e realmente existe no subsistema SECO.

7.2 Produção científica

Os trabalhos listados a seguir tratam-se de produções científicas desenvolvidas entre janeiro de 2019 e junho de 2022 durante o programa de doutorado.

- **NAMETALA, CINIRO A. L.**; FARIA, WANDRY R.; PEREIRA JÚNIOR, B. R.. Analysis of hourly price granularity implementation in the Brazilian deregulated electricity contracting environment. Utilities Policy, 2023.
- **NAMETALA, CINIRO A. L.**; FARIA, WANDRY R.; PEREIRA JÚNIOR, B. R.. On the performance of the Bayesian Optimization Algorithm with combined scenarios of Search Algorithms and Scoring Metrics. Genetic Programming and Evolvable Machines, 2022.
- **NAMETALA, CINIRO A. L.**; SOUZA, ADRIEL MAGALHÃES; PEREIRA JÚNIOR, B. R.; DA SILVA, ERALDO JANNONE. A simulator based on artificial neural networks and NSGA-II for prediction and optimization of the grinding process of superalloys with high performance grinding wheels. CIRP Journal of Manufacturing Science and Technology, 2021.
- FARIA, RODRIGUES WANDRY; **NAMETALA, CINIRO A. L.**; PEREIRA JÚNIOR, B. R.. Cost-Effectiveness Enhancement in Distribution Networks Protection System Planning. IEEE Transactions on Power Delivery, 2021.
- **NAMETALA, CINIRO A. L.**; FARIA, WANDRY R.; PEREIRA JÚNIOR, B. R.. A Volatility Index for Bilateral Electricity Contracting Auctions. IEEE Latin America Transactions, 2020.
- FARIA, WANDRY R.; MARTINS, DANIELLA DE B.; **NAMETALA, CINIRO A.L.**; PEREIRA JÚNIOR, B. R.. Protection system planning for distribution networks: A probabilistic approach. Electric Power Systems Research, 2020.
- FARIA, WANDRY R.; MARTINS, DANIELLA DE B.; **NAMETALA, CINIRO A. L.**; PEREIRA JÚNIOR, B. R.. Protection System Planning for Distribution Networks: a Probabilistic Approach. 21st Power Systems Computation Conference, 2020.

REFERÊNCIAS

ABEDINIA, O. *et al.* Electricity price forecast using Combinatorial Neural Network trained by a new stochastic search method. **Energy Conversion and Management**, v. 105, p. 642–654, nov 2015. ISSN 01968904. DOI: 10.1016/j.enconman.2015.08.025.

ACM. **A.M. Turing Awards**. 2022. Disponível em www.amturing.acm.org. Acesso em: 13/10/2022.

AFRASIABI, M. *et al.* Multi-agent microgrid energy management based on deep learning forecaster. **Energy**, v. 186, p. 115873, nov 2019. ISSN 03605442. DOI: 10.1016/j.energy.2019.115873.

AGUIRRE, L. A. **Introdução à Identificação de Sistemas - Técnicas Lineares e Não Lineares: Teoria e Aplicação**. 2. ed. [*S.l.: s.n.*]: Editora UFMG, 2015. ISBN 9788542300796.

ALASALI, F. *et al.* Day-ahead industrial load forecasting for electric RTG cranes. **Journal of Modern Power Systems and Clean Energy**, v. 6, n. 2, p. 223–234, mar 2018. ISSN 2196-5625. DOI: 10.1007/s40565-018-0394-4.

ANA. **Agência Nacional de Águas - Monitor das Secas**. 2022. Disponível em www.monitordesecas.ana.gov.br. Acesso em: 13/10/2022.

ANAMIKA; PEESAPATI, R.; KUMAR, N. Electricity Price Forecasting and Classification Through Wavelet–Dynamic Weighted PSO–FFNN Approach. **IEEE Systems Journal**, v. 12, n. 4, p. 3075–3084, dec 2018. ISSN 1932-8184. DOI: 10.1109/JSYST.2017.2717446.

ANEEL. **Resolução nº 290 de 3 de Agosto de 2000**. 2000. Disponível em https://bit.ly/3Td8KSg. Acesso em: 13/10/2022.

ANEEL. **Resolução Normativa nº 376 de 25 de Agosto de 2009**. 2009. Disponível em https://bit.ly/3CU5qGx. Acesso em: 13/10/2022.

ANEEL. **Consolidação da previsão de carga para a elaboração do Programa Mensal da Operação Energética**. 2016. Disponível em https://bit.ly/3SYVv8e. Acesso em: 13/10/2022.

ANEEL. **Resolução Normativa nº 795 de 5 de Dezembro de 2017**. 2017. Disponível em https://bit.ly/3MsM6mU. Acesso em: 13/10/2022.

ANEEL. **A ANEEL - Institucional**. 2022. Disponível em https://bit.ly/3Cy5RES. Acesso em: 13/10/2022.

ANEEL. **Sistema de Informações de Geração da ANEEL - SIGA**. 2022. Disponível em https://bit.ly/3rSNjdE. Acesso em: 13/10/2022.

ANOWAR, F.; SADAOUI, S.; SELIM, B. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). **Computer Science Review**, v. 40, 2021. DOI: 10.1016/j.cosrev.2021.100378.

ARCHER, K. J.; LEMESHOW, S.; HOSMER, D. W. Goodness-of-fit tests for logistic regression models when data are collected using a complex sampling design. **Computational Statistics & Data Analysis**, v. 51, n. 9, p. 4450–4464, may 2007. ISSN 01679473. DOI: 10.1016/j.csda.2006.07.006.

ARPAIA, P. *et al.* Machine learning for beam dynamics studies at the CERN Large Hadron Collider. **Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment**, v. 985, p. 164652, jan 2021. ISSN 01689002. DOI: 10.1016/j.nima.2020.164652.

ATEF, S.; ELTAWIL, A. B. A Comparative Study Using Deep Learning and Support Vector Regression for Electricity Price Forecasting in Smart Grids. *In*: **2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA)**. [*S.l.: s.n.*]: IEEE, 2019. p. 603–607. ISBN 978-1-7281-0851-3. DOI: 10.1109/IEA.2019.8715213.

B3. **Ibovespa B3**. 2022. Disponível em https://bit.ly/3VAaLdw. Acesso em: 13/10/2022.

BA, J. L.; KIROS, J. R.; HINTON, G. E. Layer Normalization. jul 2016. DOI: 10.48550/arXiv.1607.06450.

BAHDANAU, D.; CHO, K.; BENGIO, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *In*: **3th International Conference on Learning Representations**. San Diego, USA: [*S.l.: s.n.*], 2015. DOI: 10.48550/arXiv.1409.0473.

BAHDANAU, D. *et al.* End-to-end attention-based large vocabulary speech recognition. *In*: **2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [*S.l.: s.n.*]: IEEE, 2016. p. 4945–4949. ISBN 978-1-4799-9988-0. DOI: 10.1109/ICASSP.2016.7472618.

BAL, R.; SINHA, S. Modelling Bahdanau Attention using Election methods aided by Q-Learning. nov 2019. DOI: 10.48550/arXiv.1911.03853.

Banco Central. **Decomposição da inflação de 2021**. 2022. Disponível em https://bit.ly/3SWAv1O. Acesso em: 13/10/2022.

Banco Central. **Taxa de juros básicas - Histórico**. 2022. Disponível em https://bit.ly/3EHmdhg. Acesso em: 13/10/2022.

BANGOLIN, D. *et al.* **Crise energética: impactos e oportunidades ao agronegócio brasileiro**. 2021. Disponível em https://tinyurl.com/2p87mc4h. Acesso em: 02/11/2022.

BANNA, M. H. A. *et al.* Attention-Based Bi-Directional Long-Short Term Memory Network for Earthquake Prediction. **IEEE Access**, v. 9, p. 56589–56603, 2021. ISSN 2169-3536. DOI: 10.1109/ACCESS.2021.3071400.

BASTIAN, J. *et al.* Forecasting energy prices in a competitive market. **IEEE Computer Applications in Power**, v. 12, n. 3, p. 40–45, jul 1999. ISSN 08950156. DOI: 10.1109/67.773811.

BENGIO, Y.; SIMARD, P.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. **IEEE Transactions on Neural Networks**, v. 5, n. 2, p. 157–166, mar 1994. ISSN 1045-9227. DOI: 10.1109/72.279181.

BENTO, P. *et al.* A bat optimized neural network and wavelet transform approach for short-term price forecasting. **Applied Energy**, Elsevier, v. 210, 2018. ISSN 03062619. DOI: 10.1016/j.apenergy.2017.10.058.

BERK, K.; HOFFMANN, A.; MÜLLER, A. Probabilistic forecasting of industrial electricity load with regime switching behavior. **International Journal of Forecasting**, v. 34, n. 2, p. 147–162, apr 2018. ISSN 01692070. DOI: 10.1016/j.ijforecast.2017.09.006.

BICKEL, P. J. *et al.* Regularization in statistics. **Test**, v. 15, n. 2, p. 271–344, sep 2006. ISSN 1133-0686. DOI: 10.1007/BF02607055.

BISOI, R.; DASH, P. K.; DAS, P. P. Short-term electricity price forecasting and classification in smart grids using optimized multikernel extreme learning machine. **Neural Computing and Applications**, v. 32, n. 5, p. 1457–1480, mar 2020. ISSN 0941-0643. DOI: 10.1007/s00521-018-3652-5.

BOX, G. E. P.; JENKINS, G. **Time Series Analysis Forecasting and Control**. [*S.l.: s.n.*]: John Wiley & Sons, 1970. DOI: . ISBN 9780470272848.

BRACALE, A. *et al.* Short-term industrial reactive power forecasting. **International Journal of Electrical Power & Energy Systems**, v. 107, p. 177–185, may 2019. ISSN 01420615. DOI: 10.1016/j.ijepes.2018.11.022.

BRASIL. **Decreto nº 24.643 de 10 de Julho de 1934**. 1934. Disponível em https://bit.ly/3RYMfQ3. Acesso em: 13/10/2022.

BRASIL. **Lei nº 4.454 de 6 de Novembro de 1964**. 1964. Disponível em https://bit.ly/3g58RRA. Acesso em: 13/10/2022.

BRASIL. **Lei nº 8.987 de 13 de Fevereiro de 1995**. 1995. Disponível em https://bit.ly/3VnyDRg. Acesso em: 13/10/2022.

BRASIL. **Lei nº 9.074 de 7 de Julho de 1995**. 1995. Disponível em https://bit.ly/3CygaZA. Acesso em: 13/10/2022.

BRASIL. **Lei nº 9.427 de 26 de Dezembro de 1996**. 1996. Disponível em https://bit.ly/3CU5oOZ. Acesso em: 13/10/2022.

BRASIL. **Lei nº 9.648 de 27 de Maio de 1998**. 1998. Disponível em https://bit.ly/3rSEuRe. Acesso em: 13/10/2022.

BRASIL. **Lei nº 10.433 de 24 de Abril de 2002**. 2002. Disponível em https://bit.ly/3T0MDiv. Acesso em: 13/10/2022.

BRASIL. **Lei nº 10.848 de 15 de Março de 2004**. 2004. 2 p. Disponível em https://bit.ly/3EAlXkh. Acesso em: 13/10/2022.

BRASIL. **Medida Provisória nº 579 de 11 de Setembro de 2012**. 2012. Disponível em https://bit.ly/3TlrWxx. Acesso em: 13/10/2022.

BRASIL. **Resolução Normativa de 11 de Agosto de 2015 da Agência Nacional de Energia Elétrica - ANEEL**. 2015. Disponível em https://bit.ly/3CRygHE. Acesso em: 13/10/2022.

BRASIL. **Projeto de Lei nº 232 de 2016**. 2016. Disponível em https://bit.ly/3MqqAPK. Acesso em: 13/10/2022.

BRASIL. **Portaria nº 514 de 27 de Dezembro de 2018**. 2018. Disponível em https://bit.ly/3rPC85y. Acesso em: 13/10/2022.

Brasil. **IPEA - Resultado do PIB em 2021 e Perspectivas**. 2021. Disponível em https://bit.ly/3VqCtcA. Acesso em: 13/10/2022.

BRASIL. **Medida Provisória Nº 1.031 de 23 de Fevereiro de 2021**. 2021. Disponível em https://bit.ly/3CWkZxz. Acesso em: 13/10/2022.

BROMLEY, J. *et al.* Signature Verification using a "Siamese"Time Delay Neural Network. *In*: **Advances in Neural Information Processing Systems - NeurIPS Proceedings**. Denver, Colorado, EUA: [*S.l.: s.n.*], 1994. p. 737–744. DOI: dl.acm.org/doi/10.5555/2987189.2987282.

BUENO, R. d. L. d. S. **Econometria de Séries Temporais**. São Paulo: Cengage Learning, 2011. ISBN 9788522111572.

CAETANO, M. A. L. **Análise de Risco em Aplicações Financeiras**. São Paulo: Blucher, 2017. 109–124 p. ISBN 9788521211440.

CAMPAGNER, A.; CIUCCI, D.; CABITZA, F. Aggregation models in ensemble learning: A large-scale comparison. **Information Fusion**, v. 90, p. 241–252, 2023. DOI: 10.1016/j.inffus.2022.09.015.

CAMPBELL, C. Kernel methods: a survey of current techniques. **Neurocomputing**, v. 48, n. 1-4, p. 63–84, oct 2002. ISSN 09252312. DOI: 10.1016/S0925-2312(01)00643-9.

CARLÀ, T. *et al.* A statistical-based approach for determining the intensity of unrest phases at Stromboli volcano (Southern Italy) using one-step-ahead forecasts of displacement time series. **Natural Hazards**, v. 84, n. 1, p. 669–683, oct 2016. ISSN 0921-030X. DOI: 10.1007/s11069-016-2451-5.

CCEE. **InfoPLD nº 22 de Fevereiro de 2012**. São Paulo: [*S.l.: s.n.*], 2012. Disponível em https://bit.ly/3igD3YS. Acesso em: 13/10/2022.

CCEE. **Caderno Azul Versão 2013.1.0 de Penalidades de Energia**. São Paulo: [*S.l.: s.n.*], 2013. 52 p. Disponível em https://bit.ly/3A6b1Wa. Acesso em: 13/10/2022.

CCEE. **InfoMercado Mensal nº 162 de Dezembro de 2020**. São Paulo: [*S.l.: s.n.*], 2020. Disponível em https://bit.ly/3VoC0HO. Acesso em: 13/10/2022.

CCEE. **Relatório Anual de Administração 2019**. 2020. Disponível em https://bit.ly/3EEZ6UK. Acesso em: 13/10/2022.

CCEE. **Caderno Vermelho Versão 2021.1.0 do Preço da Liquidação das Diferenças**. São Paulo: Câmara de Comercialização de Energia Elétrica, 2021. 26 p. Disponível em https://bit.ly/3jju2NO. Acesso em: 13/10/2022.

CCEE. **Consumo de energia elétrica cresce 4.1% em 2021, aponta CCEE**. 2021. Disponível em https://bit.ly/3rSN2Yk. Acesso em: 13/10/2022.

CCEE. **Adesão**. 2022. Disponível em https://bit.ly/3VqHge4. Acesso em: 13/10/2022.

CCEE. **Contratos CCEE**. 2022. Disponível em https://bit.ly/3yFnUYW. Acesso em: 13/10/2022.

CCEE. **Informações sobre a Liquidação Financeira do MCP**. 2022. Disponível em https://bit.ly/3MrsRdi. Acesso em: 13/10/2022.

CCEE. **Leilões**. 2022. Disponível em https://bit.ly/3T4gTZJ. Acesso em: 13/10/2022.

CCEE. **Liquidação**. 2022. Disponível em https://bit.ly/3CQRIUS. Acesso em: 13/10/2022.

CCEE. **Painel de Preço**. 2022. Disponível em https://bit.ly/3yCjoKL. Acesso em: 13/10/2022.

CCEE. **Regras de Comercialização**. 2022. Disponível em https://bit.ly/3VnJwT8. Acesso em: 13/10/2022.

CCEE. **Sobre nós**. 2022. Disponível em https://bit.ly/3g0qGBs. Acesso em: 13/10/2022.

CEPEL. **DECOMP - Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Curto Prazo**. 2022. Disponível em https://bit.ly/3EE2yyF. Acesso em: 13/10/2022.

CEPEL. **DESSEM - Modelo de Despacho Hidrotérmico de Curto Prazo**. 2022. Disponível em https://bit.ly/3EDYkqR. Acesso em: 13/10/2022.

CEPEL. **GEVAZP – Modelo de Geração de Cenários de Energias e Vazões**. 2022. Disponível em https://bit.ly/3VqYYxM. Acesso em: 13/10/2022.

CEPEL. **NEWAVE - Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Longo e Médio Prazo**. 2022. Disponível em https://bit.ly/3MuHNYj. Acesso em: 13/10/2022.

CEPEL. **Produtos**. 2022. Disponível em https://bit.ly/3VlWTDs. Acesso em: 13/10/2022.

CHANDLER, A. D. **Strategy and Structure: Chapters in the History of the American Industrial Enterprise**. Cambridge: Beard Books, 1962. ISBN 158798198X.

CHANG, Z.; ZHANG, Y.; CHEN, W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. **Energy**, v. 187, p. 115804, nov 2019. ISSN 03605442. DOI: 10.1016/j.energy.2019.07.134.

CHAPRA, S. C.; CANALE, R. P. **Métodos Numéricos para Engenharia**. Porto Alegre: McGraw-Hill e Bookman, 2016. ISBN 9788580555687.

CHAUDHARI, S. *et al.* An Attentive Survey of Attention Models. *In*: **ACM Transactions on Intelligent Systems and Technology**. [*S.l.: s.n.*]: ACM, 2021. v. 12, p. 1–32. DOI: 10.1145/3465055.

CHEADLE, C. *et al.* Analysis of Microarray Data Using Z Score Transformation. **The Journal of Molecular Diagnostics**, v. 5, n. 2, p. 73–81, may 2003. ISSN 15251578. DOI: 10.1016/S1525-1578(10)60455-2.

CHEN, S. *et al.* A novel selective naïve Bayes algorithm. **Knowledge-Based Systems**, v. 192, 2020. ISSN 09507051. DOI: 10.1016/j.knosys.2019.105361.

CHEN, Y. *et al.* BRIM: An Accurate Electricity Spot Price Prediction Scheme-Based Bidirectional Recurrent Neural Network and Integrated Market. **Energies**, v. 12, n. 12, p. 2241, jun 2019. ISSN 1996-1073. DOI: 10.3390/en12122241.

CHENG, H.-T. **Wide & Deep Learning: Better Together with TensorFlow**. 2016. Disponível em https://bit.ly/3rUkP3s. Acesso em: 13/10/2022.

CHENG, H.-T. *et al.* Wide and Deep Learning for Recommender Systems. *In*: **Proceedings of the 1st Workshop on Deep Learning for Recommender Systems**. New York, NY, USA: ACM, 2016. p. 7–10. ISBN 9781450347952. DOI: dl.acm.org/doi/10.1145/2988450.2988454.

CHILE. **Lei nº 4 de 22 de Junho 1982**. 1982. Disponível em https://bit.ly/3EF7lQG. Acesso em: 13/10/2022.

CHINNATHAMBI, R. A. *et al.* Deep Neural Networks (DNN) for Day-Ahead Electricity Price Markets. *In*: **2018 IEEE Electrical Power and Energy Conference (EPEC)**. [*S.l.: s.n.*]: IEEE, 2018. p. 1–6. ISBN 978-1-5386-5419-4. DOI: 10.1109/EPEC.2018.8598327.

CHO, K. *et al.* Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *In*: **Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2014)**. Doha: [*S.l.: s.n.*], 2014. p. 1724–1734. DOI: 10.3115/v1/D14-1179.

CHOLLET, F. **Deep Learning with Python**. Nova York: Manning Publications Co., 2018. 4–5 p. ISBN 9781617294433.

CHRISTENSEN, T.; HURN, A.; LINDSAY, K. Forecasting spikes in electricity prices. **International Journal of Forecasting**, v. 28, n. 2, p. 400–411, apr 2012. ISSN 01692070. DOI: 10.1016/j.ijforecast.2011.02.019.

CHUA-LIANG, S.; KIRSCHEN, D. Quantifying the Effect of Demand Response on Electricity Markets. **IEEE Transactions on Power Systems**, v. 24, n. 3, p. 1199–1207, aug 2009. ISSN 0885-8950. DOI: 10.1109/TPWRS.2009.2023259.

CHUNG, J. *et al.* Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *In*: **NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems**. [*S.l.: s.n.*], 2014. DOI: 10.48550/arXiv.1412.3555.

CHURCH, K. Word2Vec. **Natural Language Engineering**, v. 23, n. 1, p. 155–162, jan 2017. ISSN 1351-3249. DOI: .

CLIFTON, J.; LANTHIER, P.; SCHRÖTER, H. Regulating and deregulating the public utilities 1830–2010. **Business History**, v. 53, n. 5, p. 659–672, aug 2011. ISSN 0007-6791. DOI: 10.1080/00076791.2011.599592.

COELHO, L. d. S.; SANTOS, A. A. A RBF neural network model with GARCH errors: Application to electricity price forecasting. **Electric Power Systems Research**, v. 81, n. 1, p. 74–83, jan 2011. ISSN 03787796. DOI: 10.1016/j.epsr.2010.07.015.

COLOMBINI, E. L. **An Attentional Model for Intelligent Robotics Agents**. 2014. Tese (Doutorado) — Instituto Tecnológico da Aeronáutica (ITA), 2014. Disponível em https://bit.ly/3g2phKE. Acesso em: 13/10/2022.

CORREIA, A. d. S.; COLOMBINI, E. L. Attention, please! A survey of Neural Attention Models in Deep Learning. **Artificial Intelligence Review**, 2022. DOI: 10.1007/s10462-022-10148-x.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, v. 20, n. 3, p. 273–297, sep 1995. ISSN 0885-6125. DOI: 10.1007/BF00994018.

COSTA, A. *et al.* A review on the young history of the wind power short-term prediction. **Renewable and Sustainable Energy Reviews**, v. 12, n. 6, p. 1725–1744, aug 2008. ISSN 13640321. DOI: 10.1016/j.rser.2007.01.015.

COUTINHO, L. G.; FERRAZ, J. C. **Estudo da Competitividade na Indústria Brasileira**. São Paulo: Papirus e UNICAMP, 1994. Disponível em https://bit.ly/3TdYJEo. Acesso em: 13/10/2022.

CRAMTOM, P.; OCKENFELS, A.; STOFT, S. Capacity market fundamentals. **Economics of Energy & Environmental Policy**, v. 2, n. 2, p. 27–46, 2013. Disponível em https://bit.ly/3VrEZPC. Acesso em: 13/10/2022.

DARMSTADTER, J. **Energy in the world economy: a statistical review of trends in output, trade, and consumption**. London: Johns Hopkins University Press, 1971.

DAUPHIN, Y. N. *et al.* Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *In*: **Proceedings of the 27th International Conference on Neural Information Processing Systems**. [*S.l.: s.n.*], 2014. p. 2933–2941. Disponível em https://stanford.io/3VtAvbd. Acesso em: 13/10/2022.

DENG, Z.; LIU, C.; ZHU, Z. Inter-hours rolling scheduling of behind-the-meter storage operating systems using electricity price forecasting based on deep convolutional neural network. **International Journal of Electrical Power & Energy Systems**, v. 125, p. 106499, feb 2021. ISSN 01420615. DOI: 10.1016/j.ijepes.2020.106499.

DENNY, E.; O'MAHONEY, A.; LANNOYE, E. Modelling the impact of wind generation on electricity market prices in Ireland: An econometric versus unit commitment approach. **Renewable Energy**, v. 104, p. 109–119, apr 2017. ISSN 09601481. DOI: 10.1016/j.renene.2016.11.003.

DESHPANDE, P. *et al.* Thermal power plant analysis using artificial neural network. *In*: **2012 Nirma University International Conference on Engineering (NUiCONE)**. [*S.l.: s.n.*]: IEEE, 2012. p. 1–6. ISBN 978-1-4673-1719-1. DOI: 10.1109/NUICONE.2012.6493290.

DEVLIN, J. *et al.* BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019. DOI: 10.18653/v1/N19-1423.

DEWEERDT, S. How to map the brain. **Nature**, v. 571, n. 7766, p. S6–S8, jul 2019. ISSN 0028-0836. DOI: 10.1038/d41586-019-02208-0.

DEY, S. *et al.* SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification. **CoRR**, jul 2017. DOI: 10.48550/arXiv.1707.02131.

DIAZ, G.; PLANAS, E. A Note on the Normalization of Spanish Electricity Spot Prices. **IEEE Transactions on Power Systems**, v. 31, n. 3, p. 2499–2500, may 2016. ISSN 0885-8950. DOI: 10.1109/TPWRS.2015.2449757.

DICK, S. Artificial Intelligence. **Harvard Data Science Review**, 2019. DOI: 10.1162/99608f92.92fe150c.

DIEBOLD, F. X. Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests. **Journal of Business & Economic Statistics**, v. 33, n. 1, p. 1–1, jan 2015. ISSN 0735-0015. DOI: 10.1080/07350015.2014.983236.

DING, Z. *et al.* Evolutionary double attention-based long short-term memory model for building energy prediction: Case study of a green building. **Applied Energy**, v. 288, p. 116660, apr 2021. ISSN 03062619. DOI: 10.1016/j.apenergy.2021.116660.

DMLC. **DMLC XGBoost - Referência para API Python**. 2022. Disponível em https://tinyurl.com/2fs84sd6. Acesso em: 02/11/2022.

DNVGL. Energy Transition Outlook 2020 - A global and regional forecast to 2050. **Dnv Gl Energy Transition Outlook**, 2020. Disponível em https://bit.ly/3RZydgY. Acesso em: 13/10/2022.

DOGO, E. M. *et al.* A Comparative Analysis of Gradient Descent-Based Optimization Algorithms on Convolutional Neural Networks. *In*: **2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)**. [*S.l.: s.n.*]: IEEE, 2018. p. 92–99. ISBN 978-1-5386-7709-4. DOI: 10.1109/CTEMS.2018.8769211.

DOMINGOS, P. A few useful things to know about machine learning. **Communications of the ACM**, v. 55, n. 10, p. 78–87, oct 2012. ISSN 0001-0782. DOI: 10.1145/2347736.2347755.

DONG, E.; DU, H.; GARDNER, L. An interactive web-based dashboard to track covid-19 in real time. **The Lancet**, 2022. DOI: 10.1016/S1473-3099(20)30120-1.

DONG, X.; SHEN, J. Triplet Loss in Siamese Network for Object Tracking. *In*: . [*S.l.: s.n.*], 2018. p. 472–488. DOI: 10.1007/978-3-030-01261-828.

DOZAT, T. Incorporating Nesterov Momentum into Adam. *In*: **Proceedings of International Conference on Learning Representations**. San Juan: [*S.l.: s.n.*], 2016. Disponível em https://bit.ly/3RSOJzu. Acesso em: 13/10/2022.

EBRAHIMIAN, H. *et al.* The price prediction for the energy market based on a new method. **Economic Research-Ekonomska Istraživanja**, v. 31, n. 1, p. 313–337, jan 2018. ISSN 1331-677X. DOI: 10.1080/1331677X.2018.1429291.

ELETROBRÁS. **Institucional**. 2022. Disponível em https://bit.ly/3yFFRq9. Acesso em: 13/10/2022.

ELMAN, J. L. Learning and development in neural networks: the importance of starting small. **Cognition**, v. 48, n. 1, p. 71–99, jul 1993. ISSN 00100277. DOI: 10.1016/0010-0277(93)90058-4.

ENERDATA. **World Power consumption, Electricity consumption**. 2022. Disponível em https://bit.ly/3SYOZhN. Acesso em: 13/10/2022.

Engie Brasil. **Sobre a ENGIE**. 2022. Disponível em https://bit.ly/3S0pU4P. Acesso em: 13/10/2022.

EPE. **Estimativa de capacidade instalada de geração distribuída no SIN: Aplicações no horário de ponta**. Brasília: Ministério de Minas e Energia, 2015. 25 p. Disponível em https://bit.ly/3D4pTbX. Acesso em: 13/10/2022.

EPE. Balanço Energético Nacional 2020. **Empresa de Pesquisa Energética**, 2020. Disponível em https://bit.ly/3Te7uOM. Acesso em: 13/10/2022.

EPE. **Plano Decenal de Expansão de Energia**. [*S.l.: s.n.*]: Ministério de Minas e Energia, 2020. 447 p. Disponível em https://bit.ly/3rQUpiK. Acesso em: 13/10/2022.

EPE. **BEN Interativo**. 2022. Disponível em https://bit.ly/3g16T4S. Acesso em: 13/10/2022.

ESPADOTO, M. *et al.* Toward a Quantitative Survey of Dimension Reduction Techniques. **IEEE Transactions on Visualization and Computer Graphics**, v. 27, n. 3, p. 2153–2173, mar 2021. ISSN 1077-2626. DOI: 10.1109/TVCG.2019.2944182.

EUA. Public Utilities Holding Company Act. *In*: **74th Congress Session**. [*S.l.: s.n.*]: Library of Congress, 1935. cap. 687, p. 60. Disponível em https://bit.ly/3TfkvIF. Acesso em: 13/10/2022.

EVELINA, N. *et al.* A Adoção do Preço Horário na Operação do Mercado Brasileiro - Questões e Soluções. *In*: **Proceedings of XIV Symposium of Specialists in Electric Operational and Expansion Planning**. Recife, Brasil: CIGRE-Brasil, 2018. Disponível em https://tinyurl.com/rty4x66v. Acesso em: 14/11/2022.

EVGA. **GeForce 9600 GT**. 2008. Disponível em https://bit.ly/3VtNM3x. Acesso em: 13/10/2022.

FAHIM, A. *et al.* Hybrid LSTM Self-Attention Mechanism Model for Forecasting the Reform of Scientific Research in Morocco. **Computational Intelligence and Neuroscience**, v. 2021, p. 1–14, may 2021. ISSN 1687-5273. DOI: 10.1155/2021/6689204.

FAN, C. *et al.* Multi-Horizon Time Series Forecasting with Temporal Attention Learning. *In*: **Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**. New York, NY, USA: ACM, 2019. p. 2527–2535. ISBN 9781450362016. DOI: 10.1145/3292500.3330662.

FGV. **Fundação Getúlio Vargas - Dados**. 2022. Disponível em https://bit.ly/3exxDcH. Acesso em: 13/10/2022.

FGV. **Metodologia do Índice Geral de Preços - Disponibilidade Interna**. São Paulo, 2022. Disponível em https://bit.ly/3VtCVH8. Acesso em: 13/10/2022.

FGV. **Portal de consulta ao Índice Geral de Preços - Disponibilidade Interna**. São Paulo: [*S.l.: s.n.*], 2022. Disponível em https://bit.ly/3rTa7Kr. Acesso em: 13/10/2022.

FILHO, J. C. R.; AFFONSO, C. M.; OLIVEIRA, R. C. L. Pricing analysis in the Brazilian energy market: A decision tree approach. *In*: **2009 IEEE Bucharest PowerTech**. [*S.l.: s.n.*]: IEEE, 2009. p. 1–6. ISBN 978-1-4244-2234-0. DOI: 10.1109/PTC.2009.5282272.

FILHO, J. C. R.; AFFONSO, C. M.; OLIVEIRA, R. C. L. de. Energy price classification in North Brazilian market using decision tree. *In*: **2015 12th International Conference on the European Energy Market (EEM)**. [*S.l.: s.n.*]: IEEE, 2015. p. 1–5. ISBN 978-1-4673-6692-2. DOI: 10.1109/EEM.2015.7216629.

FLORIDI, L.; CHIRIATTI, M. GPT-3: Its Nature, Scope, Limits, and Consequences. **Minds and Machines**, v. 30, n. 4, p. 681–694, dec 2020. ISSN 0924-6495. DOI: 10.1007/s11023-020-09548-1.

FOURNIER, Q.; ALOISE, D. Empirical Comparison between Autoencoders and Traditional Dimensionality Reduction Methods. *In*: **2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)**. [*S.l.: s.n.*]: IEEE, 2019. p. 211–214. ISBN 978-1-7281-1488-0. DOI: 10.1109/AIKE.2019.00044.

FRALEY, C.; RAFTERY, A. E. Model-Based Clustering, Discriminant Analysis, and Density Estimation. **Journal of the American Statistical Association**, v. 97, n. 458, p. 611–631, jun 2002. ISSN 0162-1459. DOI: 10.1198/016214502760047131.

FRIENDLY, M. Corrgrams: Exploratory displays for correlation matrices. **The American Statistician**, v. 56, n. 4, p. 316–324, 2002. Disponível em 10.1198/000313002533.

FRINTROP, S.; JENSFELT, P. Attentional Landmarks and Active Gaze Control for Visual SLAM. **IEEE Transactions on Robotics**, v. 24, n. 5, p. 1054–1065, oct 2008. ISSN 1552-3098. DOI: 10.1109/TRO.2008.2004977.

FRINTROP, S.; ROME, E.; CHRISTENSEN, H. I. Computational visual attention systems and their cognitive foundations. **ACM Transactions on Applied Perception**, v. 7, n. 1, p. 1–39, jan 2010. ISSN 1544-3558. DOI: 10.1145/1658349.1658355.

FUEGI, J.; FRANCIS, J. Lovelace & Babbage and the creation of the 1843 'notes'. **IEEE Annals of the History of Computing**, v. 25, n. 4, p. 16–26, oct 2003. ISSN 1058-6180. DOI: 10.1145/2810201.

FUKUSHIMA, K.; MIYAKE, S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. **Pattern Recognition**, v. 15, n. 6, p. 455–469, jan 1982. ISSN 00313203. DOI: 10.1016/0031-3203(82)90024-3.

GALASSI, A.; LIPPI, M.; TORRONI, P. Attention in Natural Language Processing. **IEEE Transactions on Neural Networks and Learning Systems**, p. 1–18, 2020. ISSN 2162-237X. DOI: 10.1109/TNNLS.2020.3019893.

GAMMA, E. *et al.* **Padrões de Projetos: Soluções Reutilizáveis de Software Orientados a Objetos**. Porto Alegre: Bookman, 2000. ISBN 8573076100.

GAO, W. *et al.* Different states of multi-block based forecast engine for price and load prediction. **International Journal of Electrical Power & Energy Systems**, v. 104, p. 423–435, jan 2019. ISSN 01420615. DOI: 10.1016/j.ijepes.2018.07.014.

GAO, W. *et al.* Combination of fuzzy based on a meta-heuristic algorithm to predict electricity price in an electricity markets. **Chemical Engineering Research and Design**, v. 131, p. 333–345, mar 2018. ISSN 02638762. DOI: 10.1016/j.cherd.2017.09.021.

GENSLER, A. *et al.* Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM Neural Networks. *In*: **2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)**. [*S.l.: s.n.*]: IEEE, 2016. p. 002858–002865. ISBN 978-1-5090-1897-0. DOI: 10.1109/SMC.2016.7844673.

GETIRANA, A.; LIBONATI, R.; CATALDI, M. Brazil is in water crisis — it needs a drought plan. **Nature**, v. 600, p. 218–220, 2021. Disponível em 10.1038/d41586-021-03625-w.

GHAYEKHLOO, M. *et al.* A combination approach based on a novel data clustering method and Bayesian recurrent neural network for day-ahead price forecasting of electricity markets. **Electric Power Systems Research**, v. 168, p. 184–199, mar 2019. ISSN 03787796. DOI: 10.1016/j.epsr.2018.11.021.

GHAZANFAR, A.; SCHROEDER, C. Is neocortex essentially multisensory? **Trends in Cognitive Sciences**, v. 10, n. 6, p. 278–285, jun 2006. ISSN 13646613. DOI: 10.1016/j.tics.2006.04.008.

GHELARDONI, L.; GHIO, A.; ANGUITA, D. Energy Load Forecasting Using Empirical Mode Decomposition and Support Vector Regression. **IEEE Transactions on Smart Grid**, v. 4, n. 1, p. 549–556, mar 2013. ISSN 1949-3053. DOI: 10.1109/TSG.2012.2235089.

GHODDUSI, H.; CREAMER, G. G.; RAFIZADEH, N. Machine learning in energy economics and finance: A review. **Energy Economics**, Elsevier B.V., v. 81, p. 709–727, 2019. ISSN 0140-9883. DOI: 10.1016/j.eneco.2019.05.006.

GHOFRANI, M. *et al.* A new day-ahead hourly electricity price forecasting framework. *In*: **2017 North American Power Symposium (NAPS)**. [*S.l.: s.n.*]: IEEE, 2017. p. 1–6. ISBN 978-1-5386-2699-3. DOI: 10.1109/NAPS.2017.8107269.

GHOSH, S. *et al.* A novel spatio-temporal Siamese network for 3D signature recognition. **Pattern Recognition Letters**, v. 144, p. 13–20, apr 2021. ISSN 01678655. DOI: 10.1016/j.patrec.2021.01.012.

GIAMBIAGI, F. **Economia brasileira contemporânea: 1945-2015**. Rio de Janeiro: GEN Atlas, 2016. ISBN 853526793X.

GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. *In*: **Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics**. Sardinia, Itália: PMLR, 2010. p. 249–256. Disponível em https://bit.ly/3er98y9. Acesso em: 14/10/2022.

GOMES, A. A. d. C. **A reestruturação das indústrias de rede: Uma avaliação do setor elétrico brasileiro**. 1998. Tese (Doutorado) — Universidade Federal de Santa Catarina, 1998. Disponível em https://bit.ly/3S329sS. Acesso em: 14/10/2022.

GONTIJO, T. S.; COSTA, M. A.; SANTIS, R. B. de. Electricity price forecasting on electricity spot market: a case study based on the Brazilian Difference Settlement Price. **E3S Web of Conferences**, v. 239, p. 00002, feb 2021. ISSN 2267-1242. DOI: 10.1051/e3sconf/202123900002.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT Press, 2016. 767 p. ISBN 9780262035613.

Google Books. **Google Ngram Viewer**. 2021. Disponível em https://bit.ly/3Mz8D1e. Acesso em: 14/10/2022.

Google Scholar. **Base Google Scholar**. 2021. Disponível em https://bit.ly/3yI5KWt. Acesso em: 14/10/2022.

GRAVES, A. Supervised Sequence Labelling with Recurrent Neural Networks. *In*: **Studies in Computational Intelligence**. [*S.l.: s.n.*]: Springer, 2012. p. 4–38. Disponível em https://bit.ly/3CXVRX5. Acesso em: 14/10/2022.

GRIEBENOW, C.; OHARA, A. **Panorama do Setor Elétrico Brasileiro**. Rio de Janeiro: Instituto E+ Transição Energética, 2020. 85 p. Disponível em: https://bit.ly/3xneEoC.

GRIFFIN, J. M.; PULLER, S. L. Lessons from the california Electricity Crisis. *In*: **Electricity Deregulation: Choices and Challenges**. Chicago: University of Chicago Press, 2005. cap. 1, p. 145–182. ISBN 0226308561.

GROSS, G.; GALIANA, F. Short-term load forecasting. **Proceedings of the IEEE**, v. 75, n. 12, p. 1558–1573, 1987. ISSN 0018-9219. DOI: 10.1109/PROC.1987.13927.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. Nova York: McGraw-Hill, 2011.

GUO, J.-J.; LUH, P. Improving Market Clearing Price Prediction by Using a Committee Machine of Neural Networks. **IEEE Transactions on Power Systems**, v. 19, n. 4, p. 1867–1876, nov 2004. ISSN 0885-8950. DOI: 10.1109/TPWRS.2004.837759.

HAHSLER, M.; HORNIK, K.; BUCHTA, C. Getting things in order: An introduction to the r package seriation. **Journal of Statistical Software**, v. 25, n. 3, p. 1–34, 2008. Disponível em 10.18637/jss.v025.i03.

HALL, M. *et al.* The WEKA data mining software: An update. **ACM SIGKDD Explorations Newsletter**, v. 11, n. 1, p. 10–18, nov 2009. ISSN 1931-0145. DOI: 10.1145/1656274.1656278.

HANNESSON, R. Energy and GDP growth. **International Journal of Energy Sector Management**, v. 3, n. 2, p. 157–170, jun 2009. ISSN 1750-6220. DOI: 10.1108/17506220910970560.

HASTIE, T. Ridge Regularization: An Essential Concept in Data Science. **Technometrics**, v. 62, n. 4, p. 426–433, oct 2020. ISSN 0040-1706. DOI: 10.1080/00401706.2020.1791959.

HAYKIN, S. **Redes Neurais Artificiais: Princípios e Prática**. São Paulo: Bookman, 2008. 685–710 p. ISBN 9788573077186.

HE, K. *et al.* Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *In*: **Proceedings of the IEEE International Conference on Computer Vision**. Santiago, Chile: [*S.l.: s.n.*], 2015. p. 1026–1034. DOI: 10.1109/ICCV.2015.123.

HERNANDEZ, V.; OZ, A. Methodology and Strategies for Companies in the Process of Internationalization. **Journal of Global Economics**, v. 04, n. 04, 2016. ISSN 23754389. DOI: 10.4172/2375-4389.1000224.

HINTON, G. E. Reducing the Dimensionality of Data with Neural Networks. **Science**, v. 313, n. 5786, p. 504–507, jul 2006. ISSN 0036-8075. DOI: 10.1126/science.1127647.

HINTON, G. E.; MCCLELLAND, J.; RUMELHART, D. **Distributed representations**. Cambridge: MIT Press, 1986. v. 1. 77–109 p.

HINTON, G. E.; OSINDERO, S.; TEH, Y.-W. A Fast Learning Algorithm for Deep Belief Nets. **Neural Computation**, v. 18, n. 7, p. 1527–1554, jul 2006. ISSN 0899-7667. DOI: 10.1162/neco.2006.18.7.1527.

HIPPERT, H.; PEDREIRA, C.; SOUZA, R. Neural networks for short-term load forecasting: a review and evaluation. **IEEE Transactions on Power Systems**, v. 16, n. 1, p. 44–55, 2001. ISSN 08858950. DOI: 10.1109/59.910780.

HOCHBERG, M.; POUDINEH, R. The brazilian electricity market architecture: An analysis of instruments and misalignments. **Utilities Policy**, v. 72, 2021. DOI: 10.1016/j.jup.2021.101267.

HOCHREITER, S. **Untersuchungen zu dynamischen neuronalen Netzen**. 1991. Tese (Doutorado) — T.U. München, 1991. Disponível em https://bit.ly/3er3AUk. Acesso em: 14/10/2022.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, nov 1997. ISSN 0899-7667. DOI: 10.1162/neco.1997.9.8.1735.

HOFFMANN, L. Reaching new heights with artificial neural networks. **Communications of the ACM**, v. 62, n. 6, p. 96, may 2019. ISSN 0001-0782. DOI: 10.1145/3324011.

HONG, T. **Energy forecasting: Past, present and future**. [*S.l.*], 2014. v. 32, 43–48 p. Disponível em https://bit.ly/3Mu9RuZ. Acesso em: 14/10/2022.

HONG, T. Crystal Ball Lessons in Predictive Analytics. **EnergyBiz, Vol. 12, no. 2**, p. 35–37, 2015. Disponível em https://bit.ly/3CyxU7l. Acesso em: 14/10/2022.

HONG, T.; FAN, S. Probabilistic electric load forecasting: A tutorial review. **International Journal of Forecasting**, Elsevier B.V., v. 32, n. 3, p. 914–938, jul 2016. ISSN 01692070. DOI: 10.1016/j.ijforecast.2015.11.011.

HONG, T. *et al.* Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. **International Journal of Forecasting**, Elsevier B.V., v. 32, n. 3, p. 896–913, 2016. ISSN 01692070. DOI: 10.1016/j.ijforecast.2016.02.001.

HONG, T. *et al.* Energy Forecasting: A Review and Outlook. **IEEE Open Access Journal of Power and Energy**, v. 7, n. October, p. 376–388, 2020. ISSN 2687-7910. DOI: 10.1109/OAJPE.2020.3029979.

HONG, T.; WANG, P.; WHITE, L. Weather station selection for electric load forecasting. **International Journal of Forecasting**, Elsevier B.V., v. 31, n. 2, p. 286–295, apr 2015. ISSN 01692070. DOI: 10.1016/j.ijforecast.2014.07.001.

HONG, T.; XIE, J.; BLACK, J. Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting. **International Journal of Forecasting**, Elsevier B.V., v. 35, n. 4, p. 1389–1399, 2019. ISSN 01692070. DOI: 10.1016/j.ijforecast.2019.02.006.

HONG, Y.-Y.; HSIAO, C.-Y. Locational marginal price forecasting in deregulated electricity markets using artificial intelligence. **IEE Proceedings - Generation, Transmission and Distribution**, v. 149, n. 5, p. 621, 2002. ISSN 13502360. DOI: 10.1049/ip-gtd:20020371.

HONG, Y.-Y. *et al.* Short-term LMP forecasting using an artificial neural network incorporating empirical mode decomposition. **International Transactions on Electrical Energy Systems**, v. 25, n. 9, p. 1952–1964, sep 2015. ISSN 20507038. DOI: 10.1002/etep.1949.

HONG, Y.-Y.; TAYLAR, J. V.; FAJARDO, A. C. Locational marginal price forecasting in a day-ahead power market using spatiotemporal deep learning network. **Sustainable Energy, Grids and Networks**, v. 24, p. 100406, dec 2020. ISSN 23524677. DOI: 10.1016/j.segan.2020.100406.

HONG, Y.-Y.; TAYLAR, J. V.; FAJARDO, A. C. Locational Marginal Price Forecasting Using Deep Learning Network Optimized by Mapping-Based Genetic Algorithm. **IEEE Access**, v. 8, p. 91975–91988, 2020. ISSN 2169-3536. DOI: 10.1109/ACCESS.2020.2994444.

HOU, C. K. J.; BEHDINAN, K. Dimensionality Reduction in Surrogate Modeling: A Review of Combined Methods. **Data Science and Engineering**, 2022. DOI: 10.1007/s41019-022-00193-5.

HRIBAR, R. *et al.* A comparison of models for forecasting the residential natural gas demand of an urban area. **Energy**, v. 167, p. 511–522, jan 2019. ISSN 03605442. DOI: 10.1016/j.energy.2018.10.175.

HU, D. An introductory survey on attention mechanisms in nlp problems. *In*: **Proceedings of SAI Intelligent Systems Conference**. [*S.l.: s.n.*], 2019. p. 432–448. DOI: 10.48550/arXiv.1811.05544.

HUBICKA, K.; MARCJASZ, G.; WERON, R. A Note on Averaging Day-Ahead Electricity Price Forecasts Across Calibration Windows. **IEEE Transactions on Sustainable Energy**, v. 10, n. 1, p. 321–323, jan 2019. ISSN 1949-3029. DOI: 10.1109/TSTE.2018.2869557.

HUNT, S. **Making competition work in electricity**. Nova York: John Wiley & Sons, 2002. ISBN 0471220981.

HUSSAIN, H. *et al.* A survey on resource allocation in high performance distributed computing systems. **Parallel Computing**, v. 39, n. 11, p. 709–736, nov 2013. ISSN 01678191. DOI: 10.1016/j.parco.2013.09.009.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, v. 22, n. 4, p. 679–688, oct 2006. ISSN 01692070. DOI: 10.1016/j.ijforecast.2006.03.001.

IBGE. **Índice Nacional de Preços ao Consumidor Amplo**. 2022. Disponível em https://bit.ly/3Vss7c8. Acesso em: 14/10/2022.

IBGE. **Inflação sobe 0,73% em dezembro e fecha 2021 com alta de 10,06%**. 2022. Disponível em https://bit.ly/3RXktDv. Acesso em: 14/10/2022.

IBGE. **Projeção da população do Brasil e das unidades da federação**. 2022. Disponível em https://bit.ly/3MEkyen. Acesso em: 14/10/2022.

IBGE. **Sistema Nacional de Índices de Preços ao Consumidor - SNIPC**. 2022. Disponível em https://bit.ly/3ELMgUz. Acesso em: 14/10/2022.

IEA. **World Energy Balances: Data and Statistics**. [*S.l.: s.n.*]: International Energy Agency, 2022. 258 p. Disponível em https://bit.ly/3CXMgiS. Acesso em: 14/10/2022.

IEA; KAPSARC. **Transitioning to Liberalized Energy Markets**. Riyadh, Arábia Saudita: King Abdullah Petroleum Studies and Research Center (KAPSARC), 2017. 6–7 p. Disponível em https://bit.ly/3yklH2X. Acesso em: 14/10/2022.

IEEE. **IEEE Spectrum Top Programming Languages 2022**. 2022. Disponível em https://tinyurl.com/bdfhxv4x. Acesso em: 28/10/2022.

INMAN, R. H.; PEDRO, H. T.; COIMBRA, C. F. Solar forecasting methods for renewable energy integration. **Progress in Energy and Combustion Science**, v. 39, n. 6, p. 535–576, dec 2013. ISSN 03601285. DOI: 10.1016/j.pecs.2013.06.002.

INMET. **Banco de Dados Meteorológicos do INMET**. 2022. Disponível em https://bit.ly/3VvOe1t. Acesso em: 14/10/2022.

INTEL. **Intel® Core™ i7-11700F Processor**. 2022. Disponível em https://intel.ly/3rWV7uW. Acesso em: 14/10/2022.

IOFFE, S.; SZEGEDY, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *In*: **Proceedings of 32nd International Conference on Machine Learning**. Lille, França: [*S.l.: s.n.*], 2015. p. 448–456. DOI: dl.acm.org/doi/10.5555/3045118.3045167.

IPTL. **Índice de Preços Ticket Log**. 2022. Disponível em https://bit.ly/3TjCEo6. Acesso em: 14/10/2022.

ITABA, S.; MORI, H. A Fuzzy-Preconditioned GRBFN Model for Electricity Price Forecasting. **Procedia Computer Science**, v. 114, p. 441–448, 2017. ISSN 18770509. DOI: 10.1016/j.procs.2017.09.010.

ITABA, S.; MORI, H. An Electricity Price Forecasting Model with Fuzzy Clustering Preconditioned ANN. **Electrical Engineering in Japan**, v. 204, n. 3, p. 10–20, aug 2018. ISSN 04247760. DOI: 10.1002/eej.23094.

JAHANGIR, H. *et al.* A Novel Electricity Price Forecasting Approach Based on Dimension Reduction Strategy and Rough Artificial Neural Networks. **IEEE Transactions on Industrial Informatics**, v. 16, n. 4, p. 2369–2381, apr 2020. ISSN 1551-3203. DOI: 10.1109/TII.2019.2933009.

JANOSI, P. E. de; GRAYSON, L. E. Patterns of energy consumption and economic growth and structure. **The Journal of Development Studies**, v. 8, n. 2, p. 241–249, jan 1972. ISSN 0022-0388. DOI: 10.1080/00220387208421399.

JIANG, L.; HU, G. Day-Ahead Price Forecasting for Electricity Market using Long-Short Term Memory Recurrent Neural Network. *In*: **2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)**. [*S.l.: s.n.*]: IEEE, 2018. p. 949–954. ISBN 978-1-5386-9582-1. DOI: 10.1109/ICARCV.2018.8581235.

JORDAN, M. I. **Learning in Graphical Models**. Dordrecht: Springer, 1998. DOI: 10.1007/978-94-011-5014-9.

JUNIOR, H. Q. P. **Economia Da Energia**. Rio de Janeiro: Elsevier, 2017. ISBN 9788535284607.

Jupyter Team. **Jupyter Project**. 2022. Disponível em https://tinyurl.com/2cn2cb3a. Acesso em: 28/10/2022.

Kalate Ahani, I.; SALARI, M.; SHADMAN, A. Statistical models for multi-step-ahead forecasting of fine particulate matter in urban areas. **Atmospheric Pollution Research**, v. 10, n. 3, p. 689–700, may 2019. ISSN 13091042. DOI: 10.1016/j.apr.2018.11.006.

KALMBAUER, G. *et al.* **Self-Normalizing Neural Networks**. [*S.l.: s.n.*]: Curran Associates, 2017. DOI: dl.acm.org/doi/10.5555/3294771.3294864.

KAMINSKI, V. The Challenge of Pricing and Risk Managing Electricity Derivates. *In*: **The US Power Market**. London: Risk Publications, 1997.

KELLEHER, J. D. **Deep Learning**. Cambridge: MIT Press, 2019. 280 p.

KEZUNOVIC, M. *et al.* Big data analytics for future electricity grids. **Electric Power Systems Research**, Elsevier, v. 189, n. August, p. 106788, 2020. ISSN 03787796. DOI: 10.1016/j.epsr.2020.106788.

KHAJEH, M. G. *et al.* Electricity price forecasting using neural networks with an improved iterative training algorithm. **International Journal of Ambient Energy**, v. 39, n. 2, p. 147–158, feb 2018. ISSN 0143-0750. DOI: 10.1080/01430750.2016.1269674.

KHAN, A.; SARFARAZ, A. RNN-LSTM-GRU based language transformation. **Soft Computing**, v. 23, n. 24, p. 13007–13024, dec 2019. ISSN 1432-7643. DOI: 10.1007/s00500-019-04281-z.

KHATOON, S. *et al.* Effects of various factors on electric load forecasting: An overview. *In*: **2014 6th IEEE Power India International Conference (PIICON)**. [*S.l.: s.n.*]: IEEE, 2014. p. 1–5. ISBN 978-1-4799-6041-5. DOI: 10.1109/POWERI.2014.7117763.

KIEU, T.; YANG, B.; JENSEN, C. S. Outlier Detection for Multidimensional Time Series Using Deep Neural Networks. *In*: **2018 19th IEEE International Conference on Mobile Data Management (MDM)**. [*S.l.: s.n.*]: IEEE, 2018. p. 125–134. ISBN 978-1-5386-4133-0. DOI: 10.1109/MDM.2018.00029.

KIM, C.-i.; YU, I.-K.; SONG, Y. Prediction of system marginal price of electricity using wavelet transform analysis. **Energy Conversion and Management**, v. 43, n. 14, p. 1839–1851, sep 2002. ISSN 01968904. DOI: 10.1016/S0196-8904(01)00127-3.

KIM, M. K. Short-term price forecasting of Nordic power market by combination Levenberg–Marquardt and Cuckoo search algorithms. **IET Generation, Transmission & Distribution**, v. 9, n. 13, p. 1553–1563, oct 2015. ISSN 1751-8695. DOI: 10.1049/iet-gtd.2014.0957.

KIM, Y. *et al.* Structured Attention Networks. *In*: **5th International Conference on Learning Representations**. Toulon, France: [*S.l.: s.n.*], 2017. Disponível em https://bit.ly/3S7l1GV. Acesso em: 14/10/2022.

KINGMA, D. P.; BA, J. Adam: A Method for Stochastic Optimization. dec 2014. Disponível em https://bit.ly/3MztxNT. Acesso em: 14/10/2022.

KINGSFORD, C.; SALZBERG, S. L. What are decision trees? **Nature Biotechnology**, v. 26, n. 9, p. 1011–1013, sep 2008. ISSN 1087-0156. DOI: 10.1038/nbt0908-1011.

KLIGERMAN, A. S. **Operação otima de subsistemas hidrotermicos interligados utilizando programação dinamica estocastica dual**. 1992. 72 p. Tese (Doutorado) — Universidade Estadual de Campinas, 1992. Disponível em https://bit.ly/3g1Qg93. Acesso em: 14/10/2022.

KLIGERMAN, A. S.; RIBEIRO, C. C. A bi-criteria decision support system for the Brazilian hydrothermal operation planning. *In*: **25th Mini-EURO Conference on Uncertainty and Robustness in Planning and Decision Making**. Coimbra, Portugal: [*S.l.: s.n.*], 2010. p. 1–7. Disponível em https://bit.ly/3MzLdc4. Acesso em: 14/10/2022.

KNITTEL, C. R.; ROBERTS, M. R. An empirical examination of restructured electricity prices. **Energy Economics**, v. 27, n. 5, p. 791–817, 2005. ISSN 01409883. DOI: 10.1016/j.eneco.2004.11.005.

KONSTANTINOV, A. V.; UTKIN, L. V. Interpretable machine learning with an ensemble of gradient boosting machines. **Knowledge-Based Systems**, v. 222, p. 106993, jun 2021. ISSN 09507051. DOI: 10.1016/j.knosys.2021.106993.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet classification with deep convolutional neural networks. *In*: **Advances in Neural Information Processing Systems**. [*S.l.: s.n.*]: NIPS, 2012. v. 25, p. 1097–1105. DOI: 10.1145/3065386.

KUO, P.-H.; HUANG, C.-J. A High Precision Artificial Neural Networks Model for Short-Term Energy Load Forecasting. **Energies**, v. 11, n. 1, p. 213, jan 2018. ISSN 1996-1073. DOI: 10.3390/en11010213.

KUO, P.-H.; HUANG, C.-J. An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks. **Sustainability**, v. 10, n. 4, p. 1280, apr 2018. ISSN 2071-1050. DOI: 10.3390/su10041280.

KURBATSKY, V. G. *et al.* Forecasting nonstationary time series based on Hilbert-Huang transform and machine learning. **Automation and Remote Control**, v. 75, n. 5, p. 922–934, may 2014. ISSN 0005-1179. DOI: 10.1134/S0005117914050105.

LAGO, J.; De Ridder, F.; De Schutter, B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. **Applied Energy**, Elsevier, v. 221, p. 386–405, jul 2018. ISSN 03062619. DOI: 10.1016/j.apenergy.2018.02.069.

LAGO, J. *et al.* Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. **Applied Energy**, Elsevier, v. 293, 2021. DOI: 10.1016/j.apenergy.2021.116983.

LAHMIRI, S. Comparing Variational and Empirical Mode Decomposition in Forecasting Day-Ahead Energy Prices. **IEEE Systems Journal**, v. 11, n. 3, p. 1907–1910, sep 2017. ISSN 1932-8184. DOI: 10.1109/JSYST.2015.2487339.

LANDES, D. S. **Prometeu Desacorrentado: Transformação Tecnológica e Desenvolvimento Industrial na Europa Ocidental de 1750 até os dias de hoje**. 2. ed. Rio de Janeiro: Editora Nova Fronteira, 2005. ISBN 9788535215106.

LAVECCHIA, A. Deep learning in drug discovery: opportunities, challenges and future prospects. **Drug Discovery Today**, v. 24, n. 10, p. 2017–2032, oct 2019. ISSN 13596446. DOI: 10.1016/j.drudis.2019.07.006.

LEBANON, G.; VISHWANATHAN, S. The Loss Surfaces of Multilayer Networks. *In*: **Proceedings of Machine Learning Research**. San Diego, USA: [*S.l.: s.n.*], 2015. p. 192–204. Disponível em https://bit.ly/3s1vvwU. Acesso em: 14/10/2022.

LECUN, Y. Une procedure d'apprentissage ponr reseau a seuil asymetrique. **Cognitiva 85: a la Frontière de I'Intelligence Artificielle, des Sciences de la Connaissance et des Neurosciences**, p. 599–604, 1985. Disponível em https://bit.ly/3CZYSGz. Acesso em: 14/10/2022.

LECUN, Y. **Modèles connexionistes de l'apprentissage**. 1987. Tese (Doutorado) — Université de Paris, 1987. Disponível em https://bit.ly/3MyHQlU. Acesso em: 14/10/2022.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, may 2015. ISSN 0028-0836. DOI: 10.1038/nature14539.

LEE, J. B. *et al.* Attention Models in Graphs: A survey. **ACM Transactions on Knowledge Discovery from Data**, v. 13, n. 6, p. 1–25, dec 2019. ISSN 1556-4681. DOI: 10.1145/3363574.

LEVY, O.; GOLDBERG, Y. Dependency-Based Word Embeddings. *In*: **Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014. p. 302–308. DOI: 10.3115/v1/P14-2050.

LI, Y. jiang *et al.* Clothing Sale Forecasting by a Composite GRU-Prophet Model With an Attention Mechanism. **IEEE Transactions on Industrial Informatics**, p. 1–1, 2021. ISSN 1551-3203. DOI: 10.1109/TII.2021.3057922.

LIMA, F. G. **Análise de Risco**. São Paulo: Atlas, 2018. ISBN 9788597016444.

LIMA, J. L. **A Eletrobrás e a história do setor de energia elétrica no Brasil**. Rio de Janeiro: Centro da Memória da Eletricidade, 1995. Disponível em https://bit.ly/3CBUSdR. ISBN 8585147334. Acesso em: 14/10/2022.

LISI, F.; PELAGATTI, M. M. Component estimation for electricity market data: Deterministic or stochastic? **Energy Economics**, v. 74, p. 13–37, aug 2018. ISSN 01409883. DOI: 10.1016/j.eneco.2018.05.027.

LOPES, M. N. G. *et al.* Artificial neural networks approaches for predicting the potential for hydropower generation: a case study for Amazon region. **Journal of Intelligent & Fuzzy Systems**, v. 36, n. 6, p. 5757–5772, jun 2019. ISSN 10641246. DOI: 10.3233/JIFS-181604.

LUCAS, P. J. F.; GAAG, L. C. van der. **Principles of Expert Systems**. Amsterdan: Addison-Wesley, 1991.

LUO, S.; WENG, Y. A two-stage supervised learning approach for electricity price forecasting by leveraging different data sources. **Applied Energy**, v. 242, p. 1497–1512, may 2019. ISSN 03062619. DOI: 10.1016/j.apenergy.2019.03.129.

LUO, X. *et al.* Discriminant autoencoder for feature extraction in fault diagnosis. **Chemometrics and Intelligent Laboratory Systems**, v. 192, p. 103814, sep 2019. ISSN 01697439. DOI: 10.1016/j.chemolab.2019.103814.

LUONG, M.-T.; PHAM, H.; MANNING, C. D. Effective Approaches to Attention-based Neural Machine Translation. *In*: **Conference on Empirical Methods in Natural Language Processing**. Lisboa, Portugal: [*S.l.: s.n.*], 2015. DOI: 10.18653/v1/D15-1166.

MACEIRA, M. *et al.* Chain of Optimization Models for Setting the Energy Dispatch and Spot Price in the Brazilian System. *In*: **14th Power Systems Computation Conference**. Sevilla, Itália: [*S.l.: s.n.*], 2002. Disponível em https://bit.ly/3eAxlC3. Acesso em: 14/10/2022.

MACEIRA, M. E. P.; DAMÁZIO, J. M. The use of the PAR(p) model in the stochastic dual dynamic programming optimization scheme used in the operation planning of the brazilian hydropwer system. **Probability in the Engineering and Informational Sciences**, v. 20, n. 1, p. 143–156, jan 2006. ISSN 0269-9648. DOI: 10.1017/S0269964806060098.

MAINGUY, Y. L'économie de l'énergie. *In*: **Annales de Géographie**. Paris: Coll. Finances et économie, 1967.

MALDONADO, P.; MÁRQUEZ, M. **Reestructuración energética y desarrollo sustentable: El caso del sector eléctrico chileno**. Santiago: Naciones Unidas Comisión Económica para América Latina y el Caribe (CEPAL), 1995. 77 p. Disponível em https://bit.ly/3yHWyBo. Acesso em: 14/10/2022.

MARCJASZ, G. Forecasting Electricity Prices Using Deep Neural Networks: A Robust Hyper-Parameter Selection Scheme. **Energies**, v. 13, n. 18, p. 4605, sep 2020. ISSN 1996-1073. DOI: 10.3390/en13184605.

MARCJASZ, G.; LAGO, J.; WERON, R. Neural networks in day-ahead electricity price forecasting: single vs. multiple outputs. **Energy Conversion and Management**, 2020. DOI: 10.48550/arXiv.2008.08006.

MARCJASZ, G.; SERAFIN, T.; WERON, R. Selection of Calibration Windows for Day-Ahead Electricity Price Forecasting. **Energies**, v. 11, n. 9, p. 2364, sep 2018. ISSN 1996-1073. DOI: 10.3390/en11092364.

MARCJASZ, G.; UNIEJEWSKI, B.; WERON, R. On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks. **International Journal of Forecasting**, Elsevier, v. 35, n. 4, 2019. ISSN 01692070. DOI: 10.1016/j.ijforecast.2017.11.009.

MARCOS, R. A. de *et al.* Short-Term Electricity Price Forecasting with Recurrent Regimes and Structural Breaks. **Energies**, v. 13, n. 20, p. 5452, oct 2020. ISSN 1996-1073. DOI: 10.3390/en13205452.

MARSZAŁEK, A.; BURCZYNSKI, T. Forecasting day-ahead spot electricity prices using deep neural networks with attention mechanism. **Journal of Smart Environments and Green Computing**, 2021. DOI: 10.20517/jsegc.2021.02.

MARTENS, J.; SUTSKEVER, I. Learning recurrent neural networks wth Hessian-free optimization. *In*: **Proceedings of the 28th International Conference on Machine Learning (ICML 2011)**. Washington: Omnipress, 2011. DOI: dl.acm.org/doi/10.5555/3104482.3104612.

MARTIN, J.-M. **A economia mundial da energia**. Paris: Editora UNESP e Armand Colin Editeur, 1992. ISBN 8571390312.

MASHLAKOV, A. *et al.* Assessing the performance of deep learning models for multivariate probabilistic energy forecasting. **Applied Energy**, v. 285, p. 116405, mar 2021. ISSN 03062619. DOI: 10.1016/j.apenergy.2020.116405.

MASON, E. S. **Energy Requirements and Economic Growth**. Michigan: National Planning Association, 1955. 51 p.

MAYO, R. **Mercados de Eletricidade**. Rio de Janeiro: Synergia, 2012. 207 p. ISBN 9788561325787.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, dec 1943. ISSN 0007-4985. DOI: 10.1007/BF02478259.

MCDONALD, F. Samuel Insull and the Movement for State Utility Regulatory Commissions. **Business History Review**, v. 32, n. 3, p. 241–254, jul 1958. ISSN 0007-6805. DOI: 10.2307/3111741.

MCGEE, J.; SAMMUT-BONNICI, T. Network Industries in the New Economy. **European Business Journal**, v. 14, p. 116–133, 2002. Disponível em https://bit.ly/3CsfVPK. Acesso em: 14/10/2022.

MEER, D. van der; WIDÉN, J.; MUNKHAMMAR, J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. **Renewable and Sustainable Energy Reviews**, v. 81, p. 1484–1512, jan 2018. ISSN 13640321. DOI: 10.1016/j.rser.2017.05.212.

MENDONÇA, A. M. **História da Eletricidade em Campos**. Campos, Rio de Janeiro: Editora Lar Cristão, 1993.

METAQUOTES. **MetaTrader 4**. 2022. Disponível em https://metatrader4.com. Acesso em: 28/10/2022.

MINSKY, M. L.; PAPERT, S. A. **Perceptrons**. Cambridge: MIT Press, 1969.

MISHRA, M. *et al.* Deep learning in electrical utility industry: A comprehensive review of a decade of research. **Engineering Applications of Artificial Intelligence**, v. 96, n. August, p. 104000, nov 2020. ISSN 09521976. DOI: 10.1016/j.engappai.2020.104000.

MITCHELL, T. M. **Machine Learning**. Nova York: McGraw-Hill, 1997. 432 p. ISBN 0070428077.

MME. **Consulta pública nº 21 de 5 de Outubro de 2016**. 2016. Disponível em https://bit.ly/3gcKlxW. Acesso em: 14/10/2022.

MME. **Consulta pública nº 33 de 5 de Julho de 2017**. 2017. Disponível em https://bit.ly/3evy4EB. Acesso em: 14/10/2022.

MME. **Consulta pública nº 71 de 30 de Abril de 2019**. 2019. Disponível em https://bit.ly/3D05692. Acesso em: 14/10/2022.

MME. **Consulta pública nº 76 de 8 de Agosto de 2019**. 2019. Disponível em https://bit.ly/3erQE0t. Acesso em: 14/10/2022.

MME. **Consulta pública nº 77 de 9 de Agosto de 2019**. 2019. Disponível em https://bit.ly/3g2OhkV. Acesso em: 14/10/2022.

MME. **Portaria nº 187 de 4 de Abril de 2019**. 2019. Disponível em https://bit.ly/3g4QHzA. Acesso em: 14/10/2022.

MME. **Portaria nº 465 de 12 de Dezembro de 2019**. 2019. Disponível em https://bit.ly/3MvOLMC. Acesso em: 14/10/2022.

MME. **Acordo judicial permite a retomada das obras da linha de transmissão entre Manaus e Boa Vista**. 2022. Disponível em https://bit.ly/3RWA32k. Acesso em: 14/10/2022.

MME. **Usina termelétrica Jaguatirica II entra em operação em Roraima**. 2022. Disponível em https://bit.ly/3Tq7s6S. Acesso em: 14/10/2022.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. 9º. ed. [*S.l.: s.n.*]: John Wiley & Sons, 2019.

MONTÚFAR, G. *et al.* On the number of linear regions of deep neural networks. *In*: **Proceedings of the 27th International Conference on Neural Information Processing Systems**. [*S.l.: s.n.*], 2014. DOI: dl.acm.org/doi/10.5555/2969033.2969153.

MORADI, R.; BERANGI, R.; MINAEI, B. A survey of regularization strategies for deep models. **Artificial Intelligence Review**, v. 53, n. 6, p. 3947–3986, aug 2020. ISSN 0269-2821. DOI: 10.1007/s10462-019-09784-7.

MOREIRA, F. de S. *et al.* Future scenarios for the development of the desalination industry in contexts of water scarcity: A brazilian case study. **Technological Forecasting and Social Change**, v. 167, 2021. Disponível em 10.1016/j.techfore.2021.120727.

MORENO, N. d. A. **Smart Grids: Modelagem regulatória de infraestruturas**. Rio de Janeiro: Synergia Editora, 2019. ISBN 9786550100117.

MORETIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. São Paulo: Editora Blucher, 2006. ISBN 9788521203896.

MUJEEB, S.; JAVAID, N. ESAENARX and DE-RELM: Novel schemes for big data predictive analytics of electricity load and price. **Sustainable Cities and Society**, v. 51, p. 101642, nov 2019. ISSN 22106707. DOI: 10.1016/j.scs.2019.101642.

MUJEEB, S. *et al.* Deep Long Short-Term Memory: A New Price and Load Forecasting Scheme for Big Data in Smart Cities. **Sustainability**, v. 11, n. 4, p. 987, feb 2019. ISSN 2071-1050. DOI: 10.3390/su11040987.

MUNIAIN, P.; ZIEL, F. Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices. **International Journal of Forecasting**, v. 36, n. 4, p. 1193–1210, oct 2020. ISSN 01692070. DOI: 10.1016/j.ijforecast.2019.11.006.

Nações Unidas. **The Paris Agreement**. 2015. 27 p. Disponível em https://bit.ly/3CxyzpK. Acesso em: 14/10/2022.

NAMETALA, C. A. L. **Material suplementar - Redes neurais atencionais aplicadas a modelagem e previsão de preços no mercado de eletricidade brasileiro**. 2022. Disponível em https://github.com/ciniro/phd. Acesso em: 14/10/2022.

NAMETALA, C. A. L. *et al.* Uma estratégia automatizada de investimento por meio de redes neurais artificiais e preditores econométricos. *In*: **Anais do Simpósio Brasileiro de Sistemas de Informação (SBSI)**. [*S.l.: s.n.*]: Sociedade Brasileira de Computação, 2016. p. 152–159. DOI: 10.5753/sbsi.2016.5957.

NARAJEWSKI, M.; ZIEL, F. Ensemble forecasting for intraday electricity prices: Simulating trajectories. **Applied Energy**, v. 279, p. 115801, dec 2020. ISSN 03062619. DOI: 10.1016/j.apenergy.2020.115801.

NAUMANN, G. *et al.* **O episódio de seca extrema de 2019-2021 na bacia de La Plata**. 2021. Disponível em https://tinyurl.com/yw7xhn73. Acesso em: 02/11/2022.

NAZ, A. *et al.* Short-Term Electric Load and Price Forecasting Using Enhanced Extreme Learning Machine Optimization in Smart Grids. **Energies**, v. 12, n. 5, p. 866, mar 2019. ISSN 1996-1073. DOI: 10.3390/en12050866.

NAZAR, M. S. *et al.* Hybrid model using three-stage algorithm for simultaneous load and price forecasting. **Electric Power Systems Research**, v. 165, p. 214–228, dec 2018. ISSN 03787796. DOI: 10.1016/j.epsr.2018.09.004.

NERY, E. **Mercados e Regulação de Energia Elétrica**. Rio de Janeiro: Editora Interciência, 2012. ISBN 9788571932791.

NEWBERY, D. Electricity liberalisation in Britain: The quest for a satisfactory wholesale market design. **The Energy Journal**, v. 26, n. Special Issue: European Electricity Liberalisation, p. 43–70, 2005. Disponível em https://bit.ly/3rU1I9J. Acesso em: 14/10/2022.

NIU, H.; XU, K.; LIU, C. A decomposition-ensemble model with regrouping method and attention-based gated recurrent unit network for energy price prediction. **Energy**, v. 231, p. 120941, sep 2021. ISSN 03605442. DOI: 10.1016/j.energy.2021.120941.

NIU, Z.; ZHONG, G.; YU, H. A review on the attention mechanism of deep learning. **Neurocomputing**, v. 452, p. 48–62, sep 2021. ISSN 09252312. DOI: 10.1016/j.neucom.2021.03.091.

NOWOTARSKI, J.; TOMCZYK, J.; WERON, R. Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. **Energy Economics**, v. 39, p. 13–27, sep 2013. ISSN 01409883. DOI: 10.1016/j.eneco.2013.04.004.

NOWOTARSKI, J.; WERON, R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. **Renewable and Sustainable Energy Reviews**, v. 81, n. June 2017, p. 1548–1568, jan 2018. ISSN 13640321. DOI: 10.1016/j.rser.2017.05.234.

NUMFOCUS. **Numpy**. 2022. Disponível em https://numpy.org. Acesso em: 28/10/2022.

NVIDIA. **About CUDA | NVIDIA Developer**. 2022. Disponível em https://bit.ly/3T1Mnj3. Acesso em: 14/10/2022.

NVIDIA. **RTX 3090**. 2022. Disponível em https://bit.ly/3MyRKUC. Acesso em: 14/10/2022.

NVIDIA. **RTX 4090**. 2022. Disponível em https://tinyurl.com/3ru5erkp. Acesso em: 18/10/2022.

OLAMAEE, J. *et al.* Day-ahead price forecasting based on hybrid prediction model. **Complexity**, v. 21, n. S2, p. 156–164, nov 2016. ISSN 10762787. DOI: 10.1002/cplx.21792.

OLIVEIRA, V. **Sem fornecimento da Venezuela, custo para manter energia em RR chega a R$ 1,6 bilhão em um ano**. 2020. Disponível em http://glo.bo/3yGJMmL. Acesso em: 14/10/2022.

ONS. **Nota Técnica nº 114/2019-SRG/ANEEL**. 2019. Disponível em https://tinyurl.com/3e8db7s9. Acesso em: 14/10/2022.

ONS. **Manual do Sistema PrevCarga DESSEM**. 2020. Disponível em https://tinyurl.com/2p8bp5wj. Acesso em: 14/10/2022.

ONS. **Metodologia de Cálculo da Temperatura Equivalente para Previsão de Carga**. 2020. Disponível em https://tinyurl.com/yu4bs7x9. Acesso em: 14/10/2022.

ONS. **Procedimentos de Rede**. 2020. Disponível em https://tinyurl.com/bdz8a7jr. Acesso em: 14/10/2022.

ONS. **Adoção de critérios de flexibilização para suprimento eletroenergético no SIN**. [*S.l.*], 2021. Disponível em https://tinyurl.com/yer5nx9b. Acesso em: 14/10/2022.

ONS. **O que é ONS**. 2022. Disponível em https://tinyurl.com/mvnk4y58. Acesso em: 14/10/2022.

ONS. **Operador Nacional do Sistema - Histórico da operação**. 2022. Disponível em https://tinyurl.com/rvsubk2r. Acesso em: 14/10/2022.

ONS. **Operador Nacional do Sistema - Relatório Anual 2021**. 2022. Disponível em https://tinyurl.com/2jf989jh. Acesso em: 14/10/2022.

ONS. **Planejamento Anual da Operação Energética**. 2022. Disponível em https://tinyurl.com/25rsnz4v. Acesso em: 14/10/2022.

ONS. **Programação Mensal da Operação**. 2022. Disponível em https://tinyurl.com/2z9ypntw. Acesso em: 14/10/2022.

ONS. **Sistema SINtegre**. 2022. Disponível em https://tinyurl.com/yc48ycup. Acesso em: 14/10/2022.

OPEBRAS. **Projeto OPEBRAS - Observatório do Preço da Eletricidade do Brasil**. 2022. Disponível em https://tinyurl.com/4yehm28p. Acesso em: 14/10/2022.

OPSD. **Open Power System Data – A platform for open data of the European power system.** 2022. Disponível em https://tinyurl.com/2wudd7nk. Acesso em: 14/10/2022.

OWENS, J. *et al.* GPU Computing. **Proceedings of the IEEE**, v. 96, n. 5, p. 879–899, may 2008. ISSN 0018-9219. DOI: 10.1109/JPROC.2008.917757.

PAPACHARALAMPOUS, G.; TYRALIS, H.; KOUTSOYIANNIS, D. One-step ahead forecasting of geophysical processes within a purely statistical framework. **Geoscience Letters**, v. 5, n. 1, p. 12, dec 2018. ISSN 2196-4092. DOI: 10.1186/s40562-018-0111-1.

PARASURAMAN, R. **The Attentive Brain**. [*S.l.: s.n.*]: MIT Press, 1998. ISBN 0262161729.

PARVAT, A. *et al.* A survey of deep-learning frameworks. *In*: **2017 International Conference on Inventive Systems and Control (ICISC)**. [*S.l.: s.n.*]: IEEE, 2017. p. 1–7. ISBN 978-1-5090-4715-4. DOI: 10.1109/ICISC.2017.8068684.

PASCANU, R. *et al.* How to construct deep recurrent neural networks. *In*: **Proceedings of the 27th International Conference on Neural Information Processing Systems**. [*S.l.: s.n.*], 2014. DOI: 10.48550/arXiv.1312.6026.

PASCANU, R.; MIKOLOV, T.; BENGIO, Y. On the difficulty of training recurrent neural networks. *In*: **International conference on machine learning**. [*S.l.: s.n.*]: PMLR, 2013. p. 1310–1318. Disponível em https://tinyurl.com/5n85fdjm. Acesso em: 14/10/2022.

PENG, T. *et al.* An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting. **Energy**, v. 221, p. 119887, apr 2021. ISSN 03605442. DOI: 10.1016/j.energy.2021.119887.

PETER, S. E.; RAGLEND, I. J. Sequential wavelet-ANN with embedded ANN-PSO hybrid electricity price forecasting model for Indian energy exchange. **Neural Computing and Applications**, v. 28, n. 8, p. 2277–2292, aug 2017. ISSN 0941-0643. DOI: 10.1007/s00521-015-2141-3.

PICCIALLI, F. *et al.* A survey on deep learning in medicine: Why, how and when? **Information Fusion**, v. 66, p. 111–137, feb 2021. ISSN 15662535. DOI: 10.1016/j.inffus.2020.09.006.

PINSON, P. Wind Energy: Forecasting Challenges for Its Operational Management. **Statistical Science**, v. 28, n. 4, nov 2013. ISSN 0883-4237. DOI: 10.1214/13-STS445.

POSTGRESQL. **PostgreSQL**. 2022. Disponível em https://tinyurl.com/3ddyka4j. Acesso em: 14/10/2022.

POURDARYAEI, A. *et al.* Short-Term Electricity Price Forecasting via Hybrid Backtracking Search Algorithm and ANFIS Approach. **IEEE Access**, v. 7, p. 77674–77691, 2019. ISSN 2169-3536. DOI: 10.1109/ACCESS.2019.2922420.

PYTHON. **Portal do Python**. 2022. Disponível em https://tinyurl.com/2p8d39du. Acesso em: 14/10/2022.

QIAN, N. On the momentum term in gradient descent learning algorithms. **Neural Networks**, v. 12, n. 1, p. 145–151, jan 1999. ISSN 08936080. DOI: 10.1016/S0893-6080(98)00116-6.

QUINN, J. *et al.* **Dive into Deep Learning: Tools for Engagement**. Thousand Oaks, California: SAGE Company, 2019.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2022. Disponível em https://tinyurl.com/46r4xmrp. Acesso em: 28/10/2022.

RADFORD, A. *et al.* Language Models are Unsupervised Multitask Learners. **OpenAI blog**, 2019. Disponível em https://tinyurl.com/bdzf685c. Acesso em: 14/10/2022.

RAI, A.; NELSON, T. Australia's National Electricity Market after Twenty Years. **Australian Economic Review**, v. 53, n. 2, p. 165–182, jun 2020. ISSN 0004-9018. DOI: 10.1111/1467-8462.12359.

RAINA, R.; MADHAVAN, A.; NG, A. Y. Large-scale deep unsupervised learning using graphics processors. *In*: **Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09**. New York, New York, USA: ACM Press, 2009. p. 1–8. ISBN 9781605585161. DOI: 10.1145/1553374.1553486.

RAMESH, A. *et al.* Zero-Shot Text-to-Image Generation. feb 2021. Disponível em https://tinyurl.com/2xtb7b5t. Acesso em: 14/10/2022.

RANZATO, M. *et al.* Efficient learning of sparse representations with an energy-based model. *In*: **Proceedings of the 19th International Conference on Neural Information Processing Systems**. [*S.l.: s.n.*]: NIPS, 2007. p. 1137–1144. Disponível em https://tinyurl.com/46cfmava. Acesso em: 14/10/2022.

RASLEY, J. *et al.* DeepSpeed: System Optimizations Enable Training Deep Learning Models with Over 100 Billion Parameters. *In*: **Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**. New York, NY, USA: ACM, 2020. p. 3505–3506. ISBN 9781450379984. DOI: 10.1145/3394486.3406703.

RENSINK, R. A. The Dynamic Representation of Scenes. **Visual Cognition**, v. 7, n. 1-3, p. 17–42, jan 2000. ISSN 1350-6285. DOI: 10.1080/135062800394667.

RIBEIRO, M. *et al.* Electricity Price Forecasting Based on Self-Adaptive Decomposition and Heterogeneous Ensemble Learning. **Energies**, v. 13, n. 19, p. 5190, oct 2020. ISSN 1996-1073. DOI: 10.3390/en13195190.

RITCHIE, H.; ROSER, M.; ROSADO, P. Energy. **Our World in Data**, 2022. Disponível em https://tinyurl.com/3cd64aue. Acesso em: 14/10/2022.

RODRIGUES, V. **Como máquinas aprendem**. Florença: [*S.l.: s.n.*], 2018. 661 p. ISBN 9781724099648.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386–408, 1958. ISSN 1939-1471. DOI: 10.1037/h0042519.

ROSTAMI-TABAR, B. *et al.* Forecasting for Social Good. *In*: **27th IIF Workshop: Forecasting for Social Good**. Bordeaux, France: [*S.l.: s.n.*], 2020. DOI: 10.1016/j.ijforecast.2021.02.010.

ROTHWELL, G.; GÓMEZ, T. **Electricity Economics: Regulation and Deregulation**. [*S.l.: s.n.*]: Wiley-IEEE Press, 2003. ISBN 9780470544495.

RUDER, S. An overview of gradient descent optimization algorithms. sep 2016. DOI: 10.48550/arXiv.1609.04747.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, oct 1986. ISSN 0028-0836. DOI: 10.1038/323533a0.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial**. 3º. ed. Rio de Janeiro: GEN LTC, 2013. 1016 p. ISBN 8535237011.

SALAH, A.; ALPAYDIN, E.; AKARUN, L. A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 3, p. 420–425, mar 2002. ISSN 01628828. DOI: 10.1109/34.990146.

SALINAS, D. *et al.* DeepAR: Probabilistic forecasting with autoregressive recurrent networks. **International Journal of Forecasting**, v. 36, n. 3, p. 1181–1191, jul 2020. ISSN 01692070. DOI: 10.1016/j.ijforecast.2019.07.001.

SAMUEL, A. L. Some Studies in Machine Learning Using the Game of Checkers. **IBM Journal of Research and Development**, v. 3, n. 3, p. 210–229, jul 1959. ISSN 0018-8646. DOI: 10.1147/rd.33.0210.

SANDHU, H.; FANG, L.; GUAN, L. Forecasting day-ahead price spikes for the Ontario electricity market. **Electric Power Systems Research**, Elsevier, v. 141, 2016. ISSN 03787796. DOI: 10.1016/j.epsr.2016.08.005.

SANTOS, T. *et al.* Hourly pricing and day-ahead dispatch setting in Brazil: The dessem model. **Electric Power Systems Research**, v. 189, 2020. ISSN 03787796. DOI: 10.1016/j.epsr.2020.106709.

SCHEIER, C.; EGNER, S. Visual attention in a mobile robot. *In*: **ISIE 97 Proceeding of the IEEE International Symposium on Industrial Electronics**. [*S.l.: s.n.*]: IEEE, 1997. v. 1, p. SS48–SS52. ISBN 0-7803-3936-3. DOI: 10.1109/ISIE.1997.651734.

SCHNEIDER, S. Power spot price models with negative prices. **The Journal of Energy Markets**, v. 4, n. 4, p. 77–102, dec 2011. ISSN 17563607. DOI: 10.21314/JEM.2011.079.

SCOPUS. **Base Scopus**. 2022. Disponível em https://tinyurl.com/mr9canpk. Acesso em: 14/10/2022.

SEHOVAC, L.; GROLINGER, K. Deep Learning for Load Forecasting: Sequence to Sequence Recurrent Neural Networks With Attention. **IEEE Access**, v. 8, p. 36411–36426, 2020. ISSN 2169-3536. DOI: 10.1109/ACCESS.2020.2975738.

SHEIKHPOUR, R. *et al.* A Survey on semi-supervised feature selection methods. **Pattern Recognition**, v. 64, p. 141–158, apr 2017. ISSN 00313203. DOI: 10.1016/j.patcog.2016.11.003.

SHOEYBI, M. *et al.* Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism. **arXiv preprint**, sep 2019. DOI: 10.48550/arXiv.1909.08053.

SIDDIQI, R.; KADER, A. POWERCOACH-an electricity trading advisor. **IEEE Computer Applications in Power**, v. 7, n. 3, p. 41–46, jul 1994. ISSN 0895-0156. DOI: 10.1109/67.294169.

SILVA, E. L. da. **Formação de preços em mercados de energia elétrica**. 2º. ed. Florianópolis, SC: Edição do autor, 2012. ISBN 9788591392605.

SILVA, I. N. da; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais para Engenharia e Ciências Aplicadas**. 2º. ed. São Paulo: Artiliber, 2016. 54 p.

SIMAREMARE, A. A.; HARYADI, F. N.; ADITYA, I. Techno-financial evaluation of wind power plant power purchase agreement (PPA) in Indonesia. *In*: . [*S.l.: s.n.*], 2020. DOI: 10.1063/5.0013640.

SINGH, N.; MOHANTY, S.; DEVSHUKLA, R. Short term electricity price forecast based on environmentally adapted generalized neuron. **Energy**, Elsevier, v. 125, 2017. ISSN 03605442. DOI: 10.1016/j.energy.2017.02.094.

SOBHANI, M.; HONG, T.; MARTIN, C. Temperature anomaly detection for electric load forecasting. **International Journal of Forecasting**, Elsevier B.V., v. 36, n. 2, p. 324–333, 2020. ISSN 01692070. DOI: 10.1016/j.ijforecast.2019.04.022.

SOMAN, S. S. *et al.* A review of wind power and wind speed forecasting methods with different time horizons. *In*: **North American Power Symposium 2010**. [*S.l.: s.n.*]: IEEE, 2010. p. 1–8. ISBN 978-1-4244-8046-3. DOI: 10.1109/NAPS.2010.5619586.

SOUSA, C. A. R. de. An overview on weight initialization methods for feedforward neural networks. *In*: **2016 International Joint Conference on Neural Networks (IJCNN)**. [*S.l.: s.n.*]: IEEE, 2016. p. 52–59. ISBN 978-1-5090-0620-5. DOI: 10.1109/IJCNN.2016.7727180.

Statista Agency. **Forecasted industrial energy consumption worldwide from 2018 to 2050**. 2022. Disponível em https://tinyurl.com/57t3ymdd. Acesso em: 14/10/2022.

STEFFENEL, L. A. *et al.* Forecasting upper atmospheric scalars advection using deep learning: An O3 experiment. **Machine Learning**, mar 2021. ISSN 0885-6125. DOI: 10.1007/s10994-020-05944-x.

STOFT, S. **Power System Economics**. New York: John Wiley & Sons, 2002. ISBN 0471150401.

SUTSKEVER, I. **Training Recurrent Neural Networks**. 2012. Tese (Doutorado) — University of Toronto, 2012. Disponível em https://tinyurl.com/4ry6kzdw. Acesso em: 15/10/2022.

SUTSKEVER, I. *et al.* On the importance of initialization and momentum in deep learning. *In*: MCALLESTER, S. D.; DAVID (ed.). **Proceedings of the 30th International Conference on Machine Learning**. Atlanta: PMLR, 2013. p. 1139–1147. Disponível em https://tinyurl.com/2p9yjcr5. Acesso em: 15/10/2022.

SUTSKEVER, I.; VINYALS, O.; LE, Q. V. Sequence to sequence learning with neural networks. *In*: **NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems**. [*S.l.: s.n.*], 2014. p. 3104–3112. Disponível em https://tinyurl.com/3cedtw2w. Acesso em: 15/10/2022.

SWEENEY, C. *et al.* The future of forecasting for renewable energy. **WIREs Energy and Environment**, v. 9, n. 2, mar 2020. ISSN 2041-8396. DOI: 10.1002/wene.365.

SWEENEY, J. Energy Economics. *In*: **International Encyclopedia of the Social & Behavioral Sciences**. [*S.l.: s.n.*]: Elsevier, 2001. p. 4513–4520. DOI: 10.1016/B0-08-043076-7/04174-7.

System 76. **PopOS**. 2022. Disponível em https://tinyurl.com/4m3ujs2y. Acesso em: 28/10/2022.

SZKUTA, B.; SANABRIA, L.; DILLON, T. Electricity price short-term forecasting using artificial neural networks. **IEEE Transactions on Power Systems**, v. 14, n. 3, p. 851–857, 1999. ISSN 08858950. DOI: 10.1109/59.780895.

TAPPERT, C. C. Who Is the Father of Deep Learning? *In*: **2019 International Conference on Computational Science and Computational Intelligence (CSCI)**. [*S.l.: s.n.*]: IEEE, 2019. p. 343–348. ISBN 978-1-7281-5584-5. DOI: 10.1109/CSCI49370.2019.00067.

TEECE, D. J. **Technological Change and the Nature of the Firm**. Londres: Francis Pinter, 1988. 256–281 p.

The World Bank. **Electric power consumption (kWh per capita)**. 2022. Disponível em https://tinyurl.com/3x279jws. Acesso em: 15/10/2022.

TIBSHIRANI, R. Regression Shrinkage and Selection Via the Lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 58, n. 1, p. 267–288, jan 1996. ISSN 00359246. DOI: 10.1111/j.2517-6161.1996.tb02080.x.

TOLMASQUIM, M. T. **Novo Modelo do Setor Elétrico Brasileiro**. Rio de Janeiro: Synergia Editora, 2015. ISBN 9788568483060.

TOUBEAU, J.-F. *et al.* Capturing Spatio-Temporal Dependencies in the Probabilistic Forecasting of Distribution Locational Marginal Prices. **IEEE Transactions on Smart Grid**, v. 12, n. 3, p. 2663–2674, may 2021. ISSN 1949-3053. DOI: 10.1109/TSG.2020.3047863.

TREISMAN, A. M.; GELADE, G. A feature-integration theory of attention. **Cognitive Psychology**, v. 12, n. 1, p. 97–136, jan 1980. ISSN 00100285. DOI: 10.1016/0010-0285(80)90005-5.

TURING, A. M. Computing Machinery and Intelligence. *In*: **Parsing the Turing Test**. Dordrecht: Springer Netherlands, 1950. p. 23–65. DOI: 10.1007/978-1-4020-6710-5_3.

UGURLU, U.; OKSUZ, I.; TAS, O. Electricity Price Forecasting Using Recurrent Neural Networks. **Energies**, v. 11, n. 5, p. 1255, may 2018. ISSN 1996-1073. DOI: 10.3390/en11051255.

UGURLU, U. *et al.* The Financial Effect of the Electricity Price Forecasts' Inaccuracy on a Hydro-Based Generation Company. **Energies**, v. 11, n. 8, p. 2093, aug 2018. ISSN 1996-1073. DOI: 10.3390/en11082093.

UNIEJEWSKI, B.; MARCJASZ, G.; WERON, R. Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO. **International Journal of Forecasting**, v. 35, n. 4, p. 1533–1547, oct 2019. ISSN 01692070. DOI: 10.1016/j.ijforecast.2019.02.001.

UNIEJEWSKI, B.; NOWOTARSKI, J.; WERON, R. Automated Variable Selection and Shrinkage for Day-Ahead Electricity Price Forecasting. **Energies**, v. 9, n. 8, p. 621, aug 2016. ISSN 1996-1073. DOI: 10.3390/en9080621.

UNIEJEWSKI, B.; WERON, R. Efficient Forecasting of Electricity Spot Prices with Expert and LASSO Models. **Energies**, v. 11, n. 8, p. 2039, aug 2018. ISSN 1996-1073. DOI: 10.3390/en11082039.

UNIEJEWSKI, B.; WERON, R. Regularized quantile regression averaging for probabilistic electricity price forecasting. **Energy Economics**, v. 95, p. 105121, mar 2021. ISSN 01409883. DOI: 10.1016/j.eneco.2021.105121.

UNIEJEWSKI, B.; WERON, R.; ZIEL, F. Variance Stabilizing Transformations for Electricity Spot Price Forecasting. **IEEE Transactions on Power Systems**, v. 33, n. 2, p. 2219–2229, mar 2018. ISSN 0885-8950. DOI: 10.1109/TPWRS.2017.2734563.

URBAN, R. C.; NAKADA, L. Y. K. Covid-19 pandemic: Solid waste and environmental impacts in brazil. **Science of The Total Environment**, v. 755, 2021. DOI: 10.1016/j.scitotenv.2020.142471.

VARSHNEY, H.; SHARMA, A.; KUMAR, R. A hybrid approach to price forecasting incorporating exogenous variables for a day ahead electricity Market. *In*: **2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)**. [*S.l.: s.n.*]: IEEE, 2016. p. 1–6. ISBN 978-1-4673-8587-9. DOI: 10.1109/ICPEICES.2016.7853355.

VASWANI, A. *et al.* Attention Is All You Need. *In*: **31st Conference on Neural Information Processing Systems (NIPS 2017)**. Long Beach, CA, USA: [*S.l.: s.n.*], 2017. DOI: dl.acm.org/doi/10.5555/3295222.3295349.

VELDE, F. van der; KAMPS, M. de; KLEIJ, G. T. CLAM: Closed-loop attention model for visual search. **Neurocomputing**, v. 58-60, p. 607–612, jun 2004. ISSN 09252312. DOI: 10.1016/j.neucom.2004.01.102.

VIANA, A. G. **Leilões como mecanismo alocativo para um novo desenho de mercado no Brasil**. 2018. Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo, 2018. DOI: 10.11606/T.3.2018.tde-06042018-082743.

VICTOIRE, A. A. *et al.* Two-Stage Machine Learning Framework for Simultaneous Forecasting of Price-Load in the Smart Grid. *In*: **2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)**. [*S.l.: s.n.*]: IEEE, 2018. p. 1081–1086. ISBN 978-1-5386-6805-4. DOI: 10.1109/ICMLA.2018.00176.

VINUTHA, H. P.; POORNIMA, B.; SAGAR, B. M. Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset. *In*: . [*S.l.: s.n.*], 2018. p. 511–518. DOI: 10.1007/978-981-10-7563-6_53.

WAIBEL, A. *et al.* Phoneme recognition using time-delay neural networks. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 37, n. 3, p. 328–339, mar 1989. ISSN 00963518. DOI: 10.1109/29.21701.

WALTHER, D.; EDGINGTON, D.; KOCH, C. Detection and tracking of objects in underwater video. *In*: **Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.** [*S.l.: s.n.*]: IEEE, 2004. v. 1, p. 544–549. ISBN 0-7695-2158-4. DOI: 10.1109/CVPR.2004.1315079.

WANG, F.; TAX, D. M. J. Survey on the attention based RNN model and its applications in computer vision. jan 2016. DOI: 10.48550/arXiv.1601.06823.

WANG, F. *et al.* Smart Households' Aggregated Capacity Forecasting for Load Aggregators Under Incentive-Based Demand Response Programs. **IEEE Transactions on Industry Applications**, v. 56, n. 2, p. 1086–1097, mar 2020. ISSN 0093-9994. DOI: 10.1109/TIA.2020.2966426.

WANG, L.; ZHANG, Z.; CHEN, J. Short-Term Electricity Price Forecasting With Stacked Denoising Autoencoders. **IEEE Transactions on Power Systems**, v. 32, n. 4, p. 2673–2681, jul 2017. ISSN 0885-8950. DOI: 10.1109/TPWRS.2016.2628873.

WANG, Y.; YAO, H.; ZHAO, S. Auto-encoder based dimensionality reduction. **Neurocomputing**, v. 184, p. 232–242, apr 2016. ISSN 09252312. DOI: 10.1016/j.neucom.2015.08.104.

Web of Science. **Base Web of Science**. 2022. Disponível em https://tinyurl.com/2p8wdys4. Acesso em: 15/10/2022.

WERBOS, P. **Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences**. 1974. Tese (Doutorado) — Harvard University, 1974. Disponível em https://tinyurl.com/37f94kb7. Acesso em: 15/10/2022.

WERON, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. **International Journal of Forecasting**, Elsevier B.V., v. 30, n. 4, p. 1030–1081, oct 2014. DOI: 10.1016/j.ijforecast.2014.08.008.

WIDROW, B.; HOFF, M. E. Adaptive switching circuits. *In*: IRE. **IRE Wescon Convention Record**. New York, 1960. p. 96–104. DOI: dl.acm.org/doi/10.5555/65669.104390.

WILLIS, H.; NORTHCOTE-GREEN, J. Spatial electric load forecasting: A tutorial review. **Proceedings of the IEEE**, v. 71, n. 2, p. 232–253, 1983. ISSN 0018-9219. DOI: 10.1109/PROC.1983.12562.

WOLFE, J. M.; CAVE, K. R.; FRANZEL, S. L. Guided search: An alternative to the feature integration model for visual search. **Journal of Experimental Psychology: Human Perception and Performance**, v. 15, n. 3, p. 419–433, 1989. ISSN 1939-1277. DOI: 10.1037/0096-1523.15.3.419.

WU, Y. *et al.* Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. 2016. DOI: 10.48550/arXiv.1609.08144.

XIAO, L. *et al.* Research and application of a hybrid wavelet neural network model with the improved cuckoo search algorithm for electrical power system forecasting. **Applied Energy**, v. 198, p. 203–222, jul 2017. ISSN 03062619. DOI: 10.1016/j.apenergy.2017.04.039.

XIE, J.; HONG, T. Variable Selection Methods for Probabilistic Load Forecasting: Empirical Evidence from Seven States of the United States. **IEEE Transactions on Smart Grid**, IEEE, v. 9, n. 6, p. 6039–6046, nov 2018. ISSN 1949-3053. DOI: 10.1109/TSG.2017.2702751.

XIE, X.; XU, W.; TAN, H. The Day-Ahead Electricity Price Forecasting Based on Stacked CNN and LSTM. *In*: . [*S.l.: s.n.*], 2018. p. 216–230. DOI: https://doi.org/10.1007/978-3-030-02698-1_-19.

YADAV, N.; NAIK, D. Generating Short Video Description using Deep-LSTM and Attention Mechanism. *In*: **2021 6th International Conference for Convergence in Technology (I2CT)**. [*S.l.: s.n.*]: IEEE, 2021. p. 1–6. ISBN 978-1-7281-8876-8. DOI: 10.1109/I2CT51068.2021.9417907.

YANDEX. **CatBoost**. 2022. Disponível em https://catboost.ai. Acesso em: 28/10/2022.

YANG, D. A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES). **Journal of Renewable and Sustainable Energy**, v. 11, n. 2, p. 022701, mar 2019. ISSN 1941-7012. DOI: 10.1063/1.5087462.

YANG, D. *et al.* Verification of deterministic solar forecasts. **Solar Energy**, v. 210, p. 20–37, nov 2020. ISSN 0038092X. DOI: 10.1016/j.solener.2020.04.019.

YANG, D. *et al.* History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. **Solar Energy**, v. 168, p. 60–101, jul 2018. ISSN 0038092X. DOI: 10.1016/j.solener.2017.11.023.

YANG, J. *et al.* A Proof-of-Stake public blockchain based pricing scheme for peer-to-peer energy trading. **Applied Energy**, v. 298, p. 117154, sep 2021. ISSN 03062619. DOI: 10.1016/j.apenergy.2021.117154.

YANG, S.; YU, X.; ZHOU, Y. LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example. *In*: **2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)**. [*S.l.: s.n.*]: IEEE, 2020. p. 98–101. ISBN 978-1-7281-8149-3. DOI: 10.1109/IWECAI50956.2020.00027.

YANG, Z.; CE, L.; LIAN, L. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. **Applied Energy**, v. 190, p. 291–305, mar 2017. ISSN 03062619. DOI: 10.1016/j.apenergy.2016.12.130.

YAO, S. J. *et al.* Prediction of system marginal prices by wavelet transform and neural networks. **Electric Machines and Power Systems**, v. 28, n. 10, p. 983–993, 2000. ISSN 15210502. DOI: 10.1080/07313560050129855.

YU, Y. *et al.* A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. **Neural Computation**, v. 31, n. 7, p. 1235–1270, jul 2019. ISSN 0899-7667. DOI: 10.1162/neco_a_01199.

YU, Z. *et al.* Deep Learning for Daily Peak Load Forecasting–A Novel Gated Recurrent Neural Network Combining Dynamic Time Warping. **IEEE Access**, IEEE, v. 7, p. 17184–17194, 2019. ISSN 2169-3536. DOI: 10.1109/ACCESS.2019.2895604.

ZAHID, M. *et al.* Electricity Price and Load Forecasting using Enhanced Convolutional Neural Network and Enhanced Support Vector Regression in Smart Grids. **Electronics**, v. 8, n. 2, p. 122, jan 2019. ISSN 2079-9292. DOI: 10.3390/electronics8020122.

ZAREIPOUR, H. Short-Term Electricity Market Prices: A Review of Characteristics and Forecasting Methods. *In*: . [*S.l.: s.n.*], 2012. p. 89–121. DOI: 10.1007/978-3-642-23193-3_4.

ZAREIPOUR, H.; CANIZARES, C.; BHATTACHARYA, K. Economic Impact of Electricity Market Price Forecasting Errors: A Demand-Side Analysis. **IEEE Transactions on Power Systems**, v. 25, n. 1, p. 254–262, feb 2010. ISSN 0885-8950. DOI: 10.1109/TPWRS.2009.2030380.

ZHANG, F.; FLEYEH, H.; BALES, C. A hybrid model based on bidirectional long short-term memory neural network and Catboost for short-term electricity spot price forecasting. **Journal of the Operational Research Society**, p. 1–25, nov 2020. ISSN 0160-5682. DOI: 10.1080/01605682.2020.1843976.

ZHANG, J.; TAN, Z.; LI, C. A Novel Hybrid Forecasting Method Using GRNN Combined With Wavelet Transform and a GARCH Model. **Energy Sources, Part B: Economics, Planning, and Policy**, v. 10, n. 4, p. 418–426, oct 2015. ISSN 1556-7249. DOI: 10.1080/15567249.2011.557685.

ZHANG, J.-l. *et al.* Electrical Power and Energy Systems Forecasting day-ahead electricity prices using a new integrated model. **Electrical Power and Energy Systems**, Elsevier, v. 105, n. August 2018, p. 541–548, 2019. ISSN 0142-0615. DOI: 10.1016/j.ijepes.2018.08.025.

ZHANG, Q.-s.; ZHU, S.-c. Visual interpretability for deep learning: a survey. **Frontiers of Information Technology & Electronic Engineering**, v. 19, n. 1, p. 27–39, jan 2018. ISSN 2095-9184. DOI: 10.1631/FITEE.1700808.

ZHANG, R.; LI, G.; MA, Z. A Deep Learning Based Hybrid Framework for Day-Ahead Electricity Price Forecasting. **IEEE Access**, v. 8, p. 143423–143436, 2020. ISSN 2169-3536. DOI: 10.1109/ACCESS.2020.3014241.

ZHANG, W.; CHEEMA, F.; SRINIVASAN, D. Forecasting of Electricity Prices Using Deep Learning Networks. *In*: **2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)**. [*S.l.: s.n.*]: IEEE, 2018. p. 451–456. ISBN 978-1-5386-5685-3. DOI: 10.1109/APPEEC.2018.8566313.

ZHANG, Y.; WANG, J.; WANG, X. Review on probabilistic forecasting of wind power generation. **Renewable and Sustainable Energy Reviews**, v. 32, p. 255–270, apr 2014. ISSN 13640321. DOI: 10.1016/j.rser.2014.01.033.

ZHAO, H.-x.; MAGOULÈS, F. A review on the prediction of building energy consumption. **Renewable and Sustainable Energy Reviews**, v. 16, n. 6, p. 3586–3592, aug 2012. ISSN 13640321. DOI: 10.1016/j.rser.2012.02.049.

ZHAO, Y.; LI, J.; YU, L. A deep learning ensemble approach for crude oil price forecasting. **Energy Economics**, v. 66, p. 9–16, aug 2017. ISSN 01409883. DOI: 10.1016/j.eneco.2017.05.023.

ZHOU, S. *et al.* An Optimized Heterogeneous Structure LSTM Network for Electricity Price Forecasting. **IEEE Access**, v. 7, p. 108161–108173, 2019. ISSN 2169-3536. DOI: 10.1109/ACCESS.2019.2932999.

ZHU, Y. *et al.* Power Market Price Forecasting via Deep Learning. *In*: **IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society**. [*S.l.: s.n.*]: IEEE, 2018. p. 4935–4939. ISBN 978-1-5090-6684-1. DOI: 10.1109/IECON.2018.8591581.

ZIEL, F. Forecasting Electricity Spot Prices Using Lasso: On Capturing the Autoregressive Intraday Structure. **IEEE Transactions on Power Systems**, v. 31, n. 6, p. 4977–4987, nov 2016. ISSN 0885-8950. DOI: 10.1109/TPWRS.2016.2521545.

ZIEL, F.; STEINERT, R.; HUSMANN, S. Forecasting day ahead electricity spot prices: The impact of the EXAA to other European electricity markets. **Energy Economics**, v. 51, p. 430–444, sep 2015. ISSN 01409883. DOI: 10.1016/j.eneco.2015.08.005.

ZIEL, F.; WERON, R. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. **Energy Economics**, v. 70, p. 396–420, feb 2018. ISSN 01409883. DOI: 10.1016/j.eneco.2017.12.016.

ZÖLLER, M.-A.; HUBER, M. F. Benchmark and Survey of Automated Machine Learning Frameworks. **Journal of Artificial Intelligence Research**, v. 70, p. 409–472, jan 2021. ISSN 1076-9757. DOI: 10.1613/jair.1.11854.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 67, n. 2, p. 301–320, apr 2005. ISSN 1369-7412. DOI: 10.1111/j.1467-9868.2005.00503.x.

ZUCARATO, A. N. **Mecanismos de capacidade em sistemas de energia elétrica com predominância de geração hidrelétrica**. 2009. 116 p. Tese (Doutorado) — Universidade Federal de Santa Catarina, 2009. Disponível em https://tinyurl.com/4rkanskz. Acesso em: 15/10/2022.

APÊNDICES

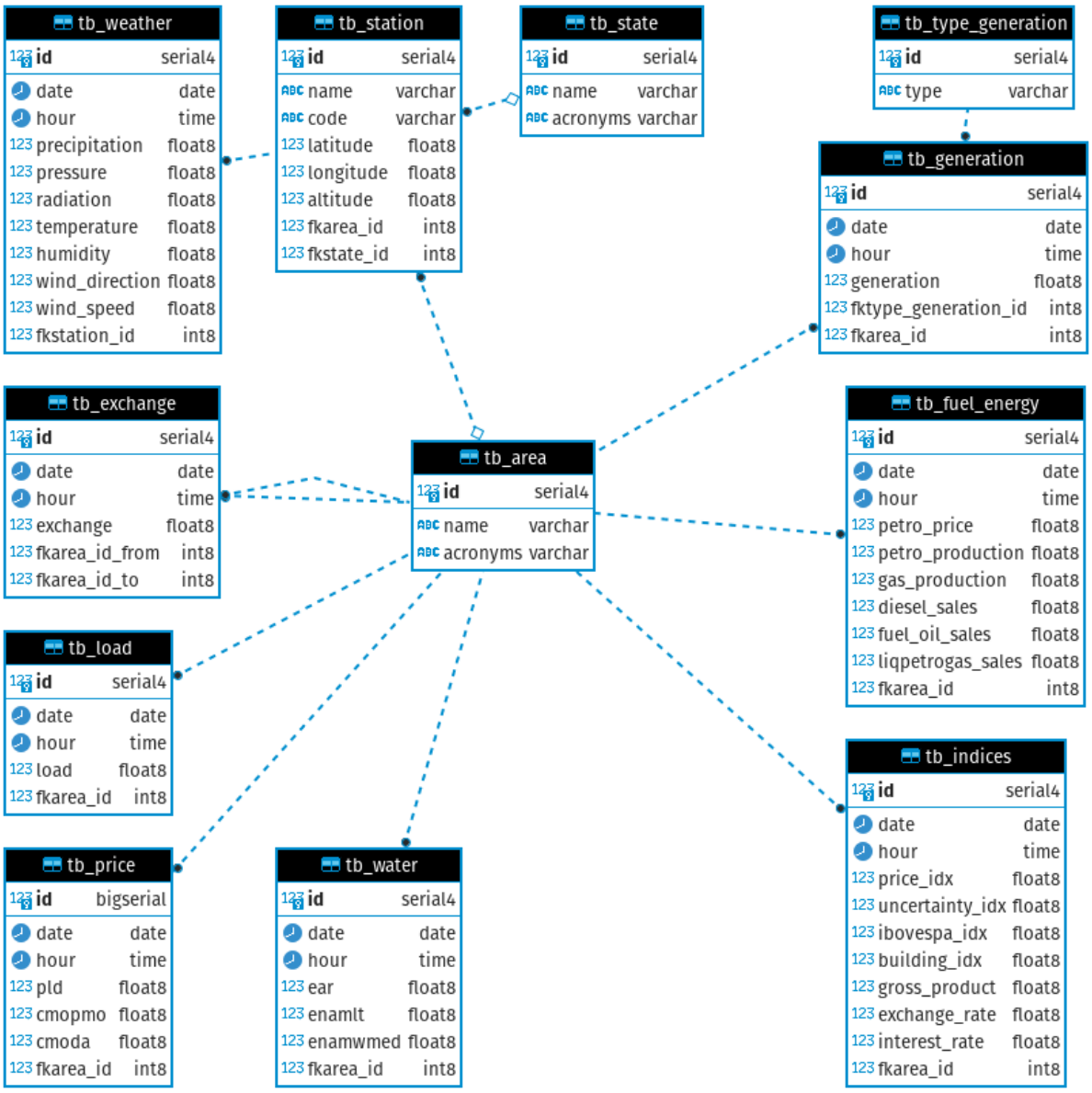
APÊNDICE A – DIAGRAMA ENTIDADE-RELACIONAMENTO

No Diagrama Entidade-Relacionamento (DER) da Figura 107 é mostrado como os dados extraídos das agências públicas são formatados no contexto do banco de dados relacional produzido pelo sistema OPB (ver subseção 5.2.1). O banco em questão contém 9 anos de registros em granularidade horária que vão de 07/01/2013 às 00:00 até 02/01/2022 às 23:00. Compõe o mesmo um total de 12 entidades, sendo:

- *tb_area*: Armazena o nome e a sigla de cada uma das áreas relacionadas com o SIN. Inclui-se aqui não somente os quatro subsistemas como também os países com os quais o SIN faz intercâmbio e o território nacional como um todo.
- *tb_state*: Armazena o nome e a sigla de todos os estados do Brasil.
- *tb_station*: Armazena o nome, código, coordenada geográfica (latitude, longitude e altitude), estado e a qual subsistema do SIN uma estação meteorológica do INMET está vinculada.
- *tb_weather*: Armazena leituras horárias de precipitação, pressão atmosférica, radiação solar, temperatura, umidade do ar, direção e velocidade do vento para cada uma das estações listadas na tabela *tb_station*.
- *tb_type_generation*: Descreve tipos de fontes de geração de eletricidade como hidráulica, térmica, eólica, solar e nuclear.
- *tb_generation*: Armazena o total gerado por cada tipo de fonte listada na tabela *tb_type_-generation* e por subsistema vinculado ao SIN listados na tabela *tb_area*.
- *tb_water*: Armazena dados horários relacionados aos reservatórios como energia armazenada total, média móvel de longo prazo da energia natural afluente e energia natural afluente por subsistema vinculado ao SIN listados na tabela *tb_area*.
- *tb_price*: Armazena dados horários de PLD e CMOs descritos no PMO e para o dia seguinte por subsistema vinculado ao SIN listados na tabela *tb_area*.
- *tb_load*: Armazena dados horários de carga por subsistema vinculado ao SIN listados na tabela *tb_area*.
- *tb_indices*: Armazena dados horários de IGP-DI, IEE-BR, IBOV, ICE, PIB, valor do câmbio (R$ para US$) e taxa de juros básica no território brasileiro.
- *tb_fuel_energy*: Armazena dados horários de preço do petróleo, produção de petróleo, produção de gás natural, vendas de diesel, vendas de óleo combustível e vendas de gás de petróleo liquefeito no território brasileiro.

- *tb_exchange*: Armazena a quantidade total horária de intercâmbios de eletricidade realizados entre os subsistemas do SIN e entre o SIN e países vizinhos conectados.

Figura 107 – Diagrama Entidade-Relacionamento do Sistema OPB.



Fonte: O Autor.

No momento de fechamento deste texto um artigo para submissão no periódico *Data in Brief* se encontra em processo de redação. Esta é uma revista com escopo específico para publicação de conjuntos de dados. Em caso de aceite do trabalho, este banco de dados ficará disponível para *download* em alguns dos principais repositórios indexados pelas bases científicas de consulta.

APÊNDICE B – ESPECIFICAÇÃO DAS ARQUITETURA DE RNAS

A arquitetura SAE usada no AE é uma estrutura de RNA convencional composta por uma sequência de camadas densas conforme apresentado em Goodfellow, Bengio e Courville (2016). Já a arquitetura ANN foi definida com base nos trabalhos de Sutskever, Vinyals e Le (2014), Cho *et al.* (2014), Bahdanau, Cho e Bengio (2015), Luong, Pham e Manning (2015), Cheng *et al.* (2016), Cheng (2016), Wu *et al.* (2016), Dey *et al.* (2017), Dong e Shen (2018) e Ghosh *et al.* (2021).

Durante a primeira etapa desta pesquisa (desenvolvida para qualificação) foram testadas diversas combinações das arquiteturas propostas por estes autores no contexto da previsão de carga em acordo com as evidências de relacionamento com o preço observadas conforme apresentado na subseção 4.1.2. Estes modelos foram capazes de prever a carga sem retreino e com janelamento móvel durante todo o ano de 2018 apresentando MAPE inferior a 5%. Na segunda etapa da pesquisa estas arquiteturas foram aprimoradas (conforme mostrado na Figura 80) para previsão não mais da carga, mas sim do PLD. Em relação a esta atividade alguns pontos podem ser elencados:

- As componentes que compõe os blocos da arquitetura (topologia) foram todas mantidas na nova versão, com exceção de um segundo AE que esperava entradas das agências de previsão meteorológica. Este foi removido, pois os dados de previsão destas agências não puderam ser adquiridos.
- Foi incluída uma camada GRU adicional antes da última camada densa que formata a saída para uma matriz 168×1. Isto foi feito visando trazer para o modelo uma capacidade de acúmulo de padrões temporais mais diversa, excedendo o provido pelas camadas LSTM.
- A componente WAD foi modificada para receber entradas de reforço que apresentaram maior facilidade de uso em relação ao uso de semanas históricas mais similares identificadas via *Dynamic Time Warping* (DTW).
- A parametrização da arquitetura (funções de ativação, número de neurônios, algoritmos de inicialização de pesos e etc.) atualmente é otimizada via *grid-search* utilizando-se opções providas pela biblioteca *Keras* do Python no contexto do que é necessário para operação em cada bloco.
- A Seq2Seq atual é composta de aproximadamente sete milhões e meio de parâmetros [1]. Todos treináveis.
- O AE atual é composto de aproximadamente um milhão de parâmetros [2]. Todos treináveis.

A arquitetura final do AE é inicializada com pesos estimados segundo distribuição normal definida via método *Glorot*, todas as funções de ativação são logísticas sigmoides e entre

[1] 7395193
[2] 1172384

as camadas densas é incluída uma camada de normalização (*LayerNormalization*). A saída do *encoder* do AE está portanto, entre 0 e 1. Esta operação foi realizada visto que as entradas da Seq2Seq são também formatadas neste intervalo e nesta nova versão existe uma conexão entre a saída do *encoder* do AE e as entradas na Seq2Seq por meio de uma camada de concatenação. Já a arquitetura da Seq2Seq é inicializada com pesos estimados segundo distribuição normal definida via método *Lecun*, todas as funções de ativação são tangentes hiperbólicas e não há normalização entre as camadas. Ambas as arquiteturas não fazem uso de *dropout*.

A parametrização final de treinamento (definida via *grid-search*) do AE utilizou algoritmo Nadam com taxa de aprendizagem de 5×10^{-5}, β_1 de 0,9, β_2 de 0,999, ϵ de 1×10^{-7} e *clipping* de norma 1. O número de épocas foi de 3000, o tamanho de *batch* de 32 e 300 épocas foram consideradas como paciência máxima num critério de parada baseado em estagnação de convergência. Para treinamento foi utilizada a métrica MSE. Já na Seq2Seq foi utilizado um algoritmo Nadam com taxa de aprendizagem de 3×10^{-5}, β_1 de 0,9, β_2 de 0,999, ϵ de 1×10^{-7} e *clipping* igual a 1. O número de épocas adotado foi de 600, o tamanho de *batch* de 64 e a quantidade de épocas máxima em caso de estagnação da convergência foi 60. MSE foi a métrica selecionada como função de perda.

A arquitetura foi construída utilizando-se as bibliotecas Keras 2.7 e Tensorflow GPU 2.5. Keras foi utilizado em todo projeto da RNA com duas exceções que precisaram ser programadas diretamente em Tensorflow, sendo (i) a conexão entre a saída do *encoder* do AE à entrada da Seq2Seq e (ii) o mecanismo de atenção suave.

Para visualizar um exemplo com especificações detalhadas (para fins de reprodução) sugere-se consultar o material suplementar disponível em Nametala (2022) (pasta *arquiteturas-rna*). Neste exemplo, o AE utilizou uma entrada considerando 120 estações do INMET e 4 variáveis climáticas, enquanto que a Seq2Seq recebeu 187 entradas.

APÊNDICE C – SELEÇÃO DE ENTRADAS DO CONJUNTO MISTO

A definição do conjunto misto de entradas utilizado no sistema OPB se deu com base na avaliação de 13 variáveis do sistema elétrico e 13 variáveis macroeconômicas (apresentadas na subseção 5.2.1) no ano de 2020 (conjunto de validação). Na Tabela 19 pode ser vista a identificação das variáveis, suas siglas, escopo, unidade e o campo correspondente utilizado no banco de dados discutido no Apêndice A.

Tabela 19 – Descrição das variáveis utilizadas nos testes de correlação.

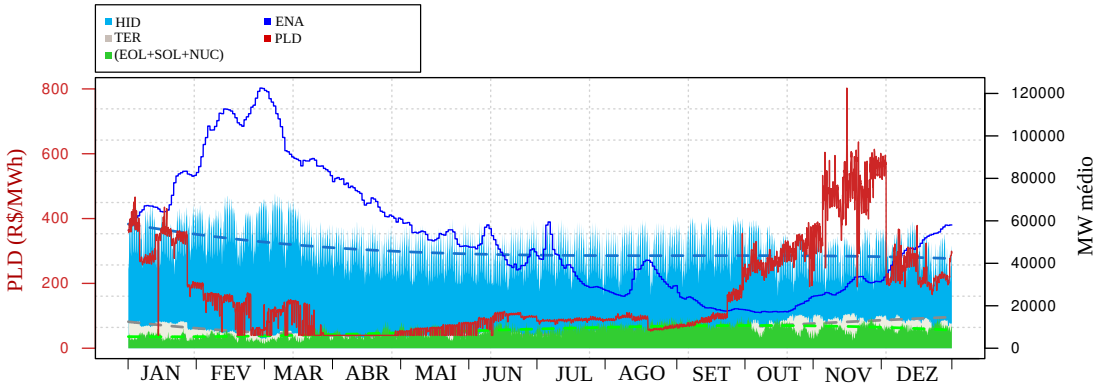
SIGLA	DESCRIÇÃO	ESCOPO	UNIDADE	CAMPO
PLD	Preço da Liquidação das Diferenças	SECO	R$	pldseco
CMOP	Custo Marginal de Operação - PMO	SECO	R$	cmopmoseco
CMOD	Custo Marginal de Operação - Dia Seguinte	SECO	R$	cmodaseco
LOAD	Carga	SECO	MWh/h	loadseco
GHYD	Geração Hidráulica	BRASIL	MW médio	genhydro
GTHE	Geração Térmica	BRASIL	MW médio	gentherm
GWIN	Geração Eólica	BRASIL	MW médio	genwind
GSOL	Geração Solar	BRASIL	MW médio	gensolar
GNUC	Geração Nuclear	BRASIL	MW médio	gennuc
EARM	Energia Armazenada	SECO	GWh	earseco
ENAM	Energia Natural Afluente	SECO	MW médio	enamwmedseco
ENAL	Média de Longo Prazo da Energia Natural Afluente	SECO	%	enamltseco
INNS	Intercâmbio NORTE → SECO	BRASIL	MW médio	internorte_seco
INSS	Intercâmbio SECO → SUL	BRASIL	MW médio	interseco_sul
FUOS	Vendas de Combustíveis: Óleo Combustível	BRASIL	R$	fuel_oil_sales
LPSL	Vendas de Combustíveis: GLP	BRASIL	R$	liqpetrogas_sales
DISL	Vendas de Combustíveis: Óleo Diesel	BRASIL	R$	diesel_sales
PEPD	Produção de Petróleo	BRASIL	Barris	petro_production
GASP	Produção de Gás Natural	BRASIL	m³	gas_production
PEPC	Cotação Diária do Barril de Petróleo Tipo Brent	BRASIL	US$/barril	petro_price
IBOV	Índice Ibovespa – Fechamento	BRASIL	Pontos	ibovespa_idx
IGPX	Índice Geral de Preços – IGP-DI	BRASIL	Pontos	price_idx
BUIX	Índice de Custo de Edificações – Média Geral	BRASIL	Pontos	building_idx
UNCX	Indicador de Incerteza da Economia do Brasil – IIE-BR	BRASIL	Pontos	uncertainty_idx
GRSP	PIB a Preços Correntes	BRASIL	R$	gross_product
INTR	Taxa de Juros de Longo Prazo	BRASIL	% a.a.	interest_rate
EXCX	Taxa de Câmbio: Real – Dólar	BRASIL	R$	exchange_rate

Fonte: O Autor.

A análise se concentrou em identificar a similaridade da série do PLD com as outras séries. Para isso foi aplicado um teste de correlação de Pearson par a par. Os valores da estatística-teste são dispostos num correlograma onde são assinalados os pares que não possuem significância estatística $\alpha = 0.05$. O teste de significância segue uma distribuição T com intervalo de confiança assintótico dado com base numa transformada Z de Fisher. A análise foi implementada com auxílio do pacote *corrplot* na geração do correlograma e do *cor.test* para o teste de significância, ambos para linguagem R (FRIENDLY, 2002; HAHSLER; HORNIK; BUCHTA, 2008).

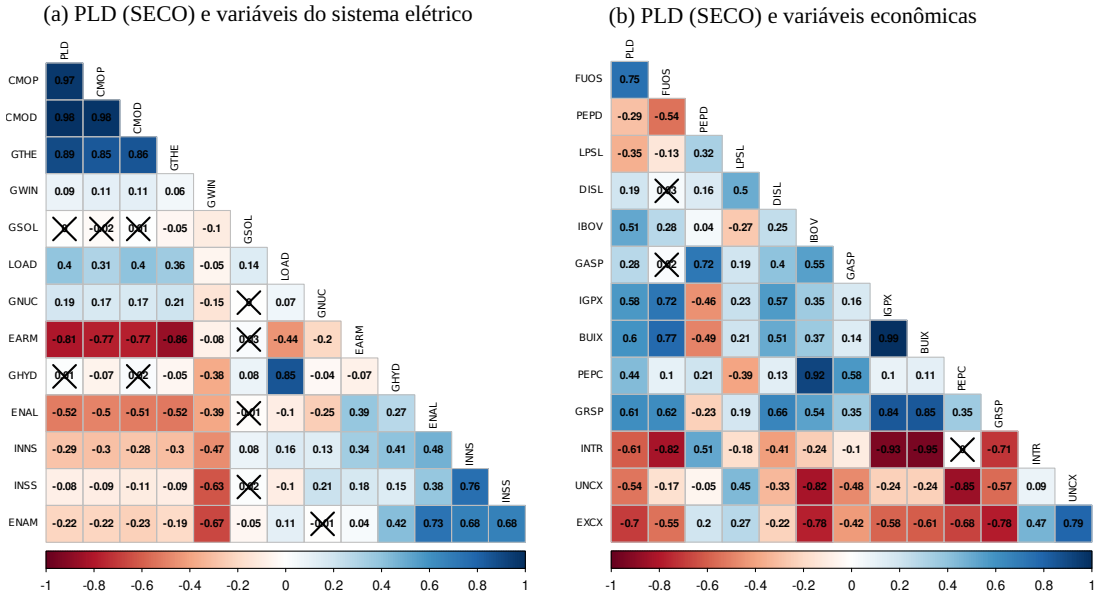
Como pode ser observado na Figura 108 durante o ano de 2020 a movimentação do PLD precedeu a da ENA. A geração hídrica se manteve estável (com baixa queda) desde o meio do ano (como pode ser observado pela linha pontilhada azul que denota uma regressão polinomial nesta curva). No entanto, com a ENA chegando aos menores níveis em setembro, observa-se também um aumento da geração térmica no mesmo período. O PLD se movimentou no sentido do uso da geração térmica. No correlograma (a) da Figura 109 pode-se perceber que as variáveis do sistema elétrico que apresentaram maior correlação proporcional e inversa com o preço foram, respectivamente, o CMO do PMO e a ENA. Já no correlograma (b) das variáveis econômicas a maior correlação proporcional foi com vendas de óleo combustível e a inversa com taxa de câmbio.

Figura 108 – PLD, ENA, gerações hídrica (HID), térmica (TER) e alternativas eólica, solar e nuclear somadas (EOL+SOL+NUC) em 2020. Os valores referem-se a todo o SIN com exceção do PLD específico do subsistema SECO.



Fonte: O Autor.

Figura 109 – Correlograma do PLD do subsistema SECO com variáveis do sistema elétrico e econômicas em 2020.



Fonte: O Autor.

Com base nisto, as variáveis utilizadas no conjunto misto foram selecionadas pelos seguintes motivos:

- PLD do subsistema SECO: É a variável que se quer prever.
- Carga no subsistema SECO: É variável que indica o consumo no subsistema SECO.
- CMO do PMO e ENA do subsistema SECO: Foram as duas variáveis do sistema elétrico mais correlacionadas com o PLD do subsistema SECO em 2020.
- Valor do câmbio (R$ para US$) e vendas de óleo combustível: Foram as duas variáveis econômicas mais correlacionadas com o PLD do subsistema SECO em 2020.
- Gerações hídrica, térmica e alternativas (eólica, solar e nuclear somadas) em todo território nacional: Apesar de estarem em escala diferente e por isso não existir proporcionalidade em relação as magnitudes dos valores, estas variáveis (especialmente a geração térmica) apresentaram movimentos que precederam os do PLD em 2020.