

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL

Guilherme Guilhermino Neto

Previsão de Carga Elétrica em Curto Prazo Utilizando Combinação de
Estações Meteorológicas

Juiz de Fora

2022

Guilherme Guilhermino Neto

**Previsão de Carga Elétrica em Curto Prazo Utilizando Combinação de
Estações Meteorológicas**

Tese apresentada ao Programa de Pós-graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Modelagem Computacional. Área de concentração: Modelagem computacional.

Orientador: Henrique Steinherz Hippert, D.Sc.

Juiz de Fora

2022

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Guilhermino Neto, Guilherme.

Previsão de Carga Elétrica em Curto Prazo Utilizando Combinação de
Estações Meteorológicas / Guilherme Guilhermino Neto. – 2022.

184 f. : il.

Orientador: Henrique Steinherz Hippert

Tese (Doutorado) – Universidade Federal de Juiz de Fora, Faculdade de
Engenharia. Programa de Pós-graduação em Modelagem Computacional,
2022.

1. Previsão de carga elétrica. 2. Combinação de estações meteorológicas.
3. Redes neurais. I. Hippert, Henrique Steinherz, orient. II. Título.

Guilherme Guilhermino Neto

Previsão de Carga Elétrica em Curto Prazo Utilizando Combinação de Estações Meteorológicas

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor(a) em Modelagem Computacional. Área de concentração: Modelagem Computacional

Aprovada em 07 de fevereiro de 2021.

BANCA EXAMINADORA

Prof. Dr. Henrique Steinherz Hippert - Orientador
Universidade Federal de Juiz de Fora

Prof. Dr. Carlos Cristiano Hasenclever Borges
Universidade Federal de Juiz de Fora

Prof. Dr. Bernardo Martins Rocha
Universidade Federal de Juiz de Fora

Prof. Dr. Fabrizzio Condé de Oliveira
Universidade Federal Fluminense

Prof. Dr. Francisco Augusto Lima Manfrini
Instituto Federal do Sudeste de Minas Gerais

Juiz de Fora, 21/01/2022.



Documento assinado eletronicamente por **Bernardo Martins Rocha, Professor(a)**, em 07/02/2022, às 20:00, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Henrique Steinherz Hippert, Professor(a)**, em 07/02/2022, às 20:00, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Francisco Augusto Lima Manfrini, Usuário Externo**, em 07/02/2022, às 20:00, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Carlos Cristiano Hasenclever Borges, Professor(a)**, em 07/02/2022, às 20:02, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Fabrízio Condé de Oliveira, Usuário Externo**, em 09/02/2022, às 09:52, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **0653601** e o código CRC **223C3677**.

Para Maria das Graças, (*in memoriam*).

AGRADECIMENTOS

Citando George R.R. Martin, “*The last one was a bitch. This one were three bitches and a bastard.*”. Portanto, não poderia deixar de agradecer às pessoas e entidades listadas a seguir.

À Micheli, por ter cedido um tempo considerável de nossas vidas a esta tese.

Ao gato Pepe Mujica, companheiro de todas as horas.

Aos meus pais e irmãos, pelo apoio.

Ao meu orientador, professor Henrique Steinherz Hippert, pela parceria e a confiança (mais uma vez).

Ao professor Fabrízio Condé de Oliveira que, dez anos atrás, apresentou-me à Modelagem Computacional e às Redes Neurais (o que definiu minha vida profissional).

Ao professor Carlos Cristiano Hasenclever Borges, cujas contribuições para esta linha de pesquisa ao longo dos anos são inestimáveis.

Aos professores Bernardo Martins Rocha e Leonardo Goliatt da Fonseca, pela condução dos trâmites para que este trabalho viesse à luz.

Aos três últimos e ao professor Francisco Augusto Lima Manfrini, por terem feito parte da banca examinadora e trazido grandes colaborações ao texto.

Aos amigos do Willcafé, pelo combustível de qualidade, o espaço para escrita e a amizade cultivada.

A todos os membros do projeto de extensão Nutrilac, na figura dos professores Luiz Maurílio Maciel, Priscila Capriles, Almira França e Amália Chaves, pela oportunidade de participar da gestação e do desenvolvimento de um produto do *campus* para a comunidade.

Aos colegas do grupo EMEP, na figura do líder e amigo Lauro Chagas e Sá, pelo aprendizado em pesquisa científica.

Aos colegas de PGMC, em especial Janaína Silva e João Alves.

Ao amigo Alonso Alvarez do LNCC, pela força, e na esperança de que o arrefecimento da pandemia de COVID-19 nos possibilite ainda mais colaborações *in loco*.

Aos amigos do Ifes, Renato Miranda, Éllen Pessoa, Marcelo Pandolfi e César Xavier, sempre próximos (mesmo quando não fisicamente).

À Universidade Federal de Juiz de Fora e ao Programa de Pós-graduação em Modelagem Computacional, por acreditarem neste trabalho quando ainda era um projeto.

Ao Instituto Federal do Espírito Santo (Ifes), em particular ao *campus* Linhares, que possibilitou-me cursar o Doutorado afastado de minhas atividades docentes.

"Terror é não saber o amanhã." (José Mojica Marins, o Zé do Caixão).

RESUMO

Previsões de carga, em base horária, são necessárias para a operação diária de empresas da Cadeia de Suprimentos do Setor Elétrico (CSSE) e vários participantes envolvidos no mercado de energia. Como a oferta deve ocorrer idealmente de forma concomitante com a demanda, por conta da impossibilidade de se estocar eletricidade em grandes volumes, a má previsão leva ao desequilíbrio do sistema, causando desde o aumento do custo financeiro da operação devido a transações em cima da hora, até falhas em equipamentos. No atual contexto, muitas empresas da CSSE têm atualizado seus equipamentos para se tornarem smart grids. Com estas tecnologias, é possível monitorar a carga em diversos pontos do sistema, passando por áreas geograficamente diversas. Some-se isto ao fato de que aspectos climáticos estão entre aqueles que mais influenciam o consumo de energia elétrica, emerge um novo desafio: como selecionar e combinar dados das estações meteorológicas disponíveis, para prever a carga de um território específico? Neste trabalho, utilizamos uma heurística para ranquear as estações meteorológicas e propomos dois novos métodos de combinação, que levam em conta aspectos de ordem prática como fatores geográficos. Para sete conjunto de dados, correspondentes a zonas de carga do operador do estado de Nova Iorque, EUA, testamos os métodos propostos para compor as entradas de redes neurais artificiais. Também testamos a combinação pela média aritmética simples, como costuma se encontrar na literatura. Analisamos a sensibilidade das redes a ruídos nos dados exógenos e comparamos os resultados com os de benchmarks univariados de séries temporais, baseados apenas no histórico das cargas. Verificamos que, para a maioria das zonas de carga em que a combinação de duas ou mais estações se faz necessária para abarcar as características climáticas, os dois métodos de combinações propostos têm melhor desempenho, inclusive com sugestão de robustez a ruído nos dados exógenos. Também notamos que levar em conta características de ordem prática, como as geográficas, nos modelos de previsão, parece trazer benefícios. Além disso, os resultados mostram que há se atentar para a qualidade dos dados de fontes secundárias, visto que a partir de certo nível de ruído, passa a ser interessante trabalhar apenas com os dados de carga, para certos conjuntos de dados.

Palavras-chave: Previsão de carga elétrica. Combinação de estações meteorológicas. Redes neurais.

ABSTRACT

Hourly-based load forecasts are necessary for the daily scheduling of the Electric Power Supply Chain (EPSC), as for many players involved on the operation of energy markets. Because of the impossibility of storing large volumes of electricity, it is desirable for the supply to meet its demand on real time, which leads bad forecasts to unbalance the whole system, causing from raises on the financial costs of the operations to equipment failure.

Lately, many companies in the EPSC are updating their technologies in order to become smart grids. With smart grid apparatuses, it is possible to use sensing to monitor the load in multiple points of the system, crossing many distinct geographical areas. As the demanda of electrical energy is known to be weather-dependant, a new challenge emerges: how to select and combine data from many available weather stations in order to feed load forecasting models for a specific territory?

On this work, we use an heuristics to rank weather stations and introduce two new methods for combining them, taking into account practical aspects such as geographical factors.

For seven datasets, corresponding to load zones from the independente operator system of the state of New York, USA, we test the proposed methods to compose the inputs for neural networks. We also try a combination by the simple average, as it is frequent on the literature. We carry on a sensitivity analysis for the networks injecting noise to the exogenous data and compare the results to univariate time series benchmarks, which depend only on the historical load.

We verify that, for most of the load zones for which the combination of two or more weather stations is necessary in order to consider its weather characteristics, the two introduced combination methods have the best median performance, and also appear to be robust to noise in exogenous data. We also note that forecasting models may benefit from taking into account practical order factors, such as geographical characteristics of the load zone, in the forecasting models. Besides, results show that it is important to be aware of the quality of secondary data, since from a certain noise level it may be of best practice to work only with load data, for some of the experimented datasets.

Keywords: Load forecasting. Weather station combination. Neural networks.

SUMÁRIO

1	INTRODUÇÃO	15
1.1	A CADEIA DE SUPRIMENTOS DO SETOR ELÉTRICO (CSSE) . . .	16
1.2	NOVOS DESAFIOS PARA A PREVISÃO DE CARGA ELÉTRICA . .	19
1.2.1	Mudanças climáticas	19
1.2.2	Cidades inteligentes	21
1.3	COMO SE PREVÊ A CARGA ELÉTRICA?	23
1.3.1	Horizontes de previsão	23
1.3.2	Características dos dados de carga elétrica	24
<i>1.3.2.1</i>	<i>Dias atípicos</i>	<i>24</i>
<i>1.3.2.2</i>	<i>Fatores sazonais</i>	<i>25</i>
<i>1.3.2.3</i>	<i>Fatores relacionados ao clima</i>	<i>27</i>
1.4	O PROBLEMA DA SELEÇÃO DE ESTAÇÕES METEOROLÓGICAS	29
1.5	OBJETIVOS DA TESE	31
1.6	ORGANIZAÇÃO DA TESE	32
2	COLETA, PRÉ-PROCESSAMENTO E ANÁLISE DOS DADOS DE CARGA E TEMPERATURA	34
2.1	COLETA DOS DADOS	34
2.1.1	Dados de carga elétrica	34
2.1.2	Dados de temperatura	36
2.2	PRÉ-PROCESSAMENTO DOS DADOS	36
2.2.1	Pré-processamento dos dados de carga	37
2.2.2	Pré-processamento dos dados de temperatura	38
2.3	ANÁLISE EXPLORATÓRIA DOS DADOS	39
2.3.1	Geografia Física do Estado de Nova Iorque	39
2.3.2	A análise	41
2.4	RESUMO DO CAPÍTULO	46
3	INTRODUÇÃO À MODELAGEM E PREVISÃO DE CARGA ELÉTRICA	47
3.1	MODELAGEM DA CARGA ELÉTRICA POR DECOMPOSIÇÃO DA CURVA DE CARGA	47
3.2	MODELOS ESTATÍSTICOS E MATEMÁTICOS PARA PREVISÃO DE CARGA ELÉTRICA	48
3.3	RECONHECIMENTO DE PADRÕES A PARTIR DE TÉCNICAS DE APRENDIZADO DE MÁQUINA	52
3.4	Modelagem e previsão de carga elétrica atualmente	57
4	MÉTODOS DE PREVISÃO: SÉRIES TEMPORAIS	59
4.1	AMORTECIMENTO EXPONENCIAL	59

4.1.1	Determinação das constantes do Amortecimento Exponencial Utilizando Otimização Numérica	62
4.2	MODELOS DE BOX & JENKINS	68
4.2.1	Identificação de um modelo de Box & Jenkins: método de Hyndman-Khandakar	70
4.2.2	Estimação dos parâmetros de um modelo de Box & Jenkins	73
4.3	<i>Benchmark</i> ingênuo: o previsor <i>Naive</i> sazonal	74
5	MÉTODOS DE PREVISÃO: REGRESSÃO	76
5.1	INTRODUÇÃO À REGRESSÃO LINEAR	76
5.2	ESTIMAÇÃO DOS COEFICIENTES DE UM MODELO DE REGRESSÃO LINEAR	79
5.2.1	Equações normais	79
5.2.2	Otimização numérica para estimação dos coeficientes de regressão	80
5.3	UM PREVISOR <i>NAIVE</i> PARA CARGAS ELÉTRICAS BASEADO EM REGRESSÃO LINEAR	82
6	MÉTODOS DE PREVISÃO: REDES NEURAIS	84
6.1	REDES NEURAIS <i>FEEDFORWARD</i> (FFNN)	85
6.1.1	Notação para modelagem matemática de FFNNs	86
6.1.2	Treinamento de uma FFNN	87
6.2	PRÉ-PROCESSAMENTO DE DADOS PARA REDES NEURAIS	90
7	SELEÇÃO E COMBINAÇÃO DE ESTAÇÕES METEOROLÓGICAS	91
7.1	UMA HEURÍSTICA PARA COMBINAÇÃO DE ESTAÇÕES METEOROLÓGICAS	91
7.2	MAIS ABORDAGENS PARA COMBINAÇÃO DAS TEMPERATURAS	92
7.3	POSSIBILIDADES DE APRIMORAMENTO PARA A SELEÇÃO E COMBINAÇÃO DE ESTAÇÕES METEOROLÓGICAS	93
7.4	NOVAS ABORDAGENS PARA COMBINAÇÃO DE ESTAÇÕES METEOROLÓGICAS	94
7.4.1	Método C1: Ponderação pela distância da estação meteorológica até o centro geográfico da zona de carga	94
7.4.2	Método C2: Ponderação pela correlação não-linear entre cargas e temperaturas	95
8	SIMULAÇÕES E RESULTADOS	98
8.1	ROTEIRO DAS SIMULAÇÕES	98
8.1.1	Heurística para seleção e combinação de estações meteorológicas	98
8.1.2	Métodos de séries temporais	98
8.1.3	Redes neurais <i>Feedforward</i> (FFNN)	99
8.1.3.1	Ajuste dos hiperparâmetros das FFNNs	100

8.1.4	Divisão dos dados	101
8.1.5	Avaliação do desempenho dos modelos	102
8.1.6	Recursos computacionais	103
8.1.7	Regressão linear múltipla: considerações sobre o modelo TVB	103
8.2	RESULTADOS	105
8.2.1	Zona de carga A - WEST	105
8.2.1.1	<i>Benchmarks</i> de séries temporais	105
8.2.1.2	FFNN e ranqueamento das estações meteorológicas	106
8.2.1.3	Combinação das estações meteorológicas	111
8.2.2	Zona de carga B - GENESE	114
8.2.2.1	<i>Benchmarks</i> de séries temporais	114
8.2.2.2	FFNN e ranqueamento das estações meteorológicas	115
8.2.2.3	Combinação das estações meteorológicas	118
8.2.3	Zona de carga C - CENTRL	121
8.2.3.1	<i>Benchmarks</i> de séries temporais	121
8.2.3.2	FFNN e ranqueamento das estações meteorológicas	122
8.2.3.3	Combinação das estações meteorológicas	125
8.2.4	Zona de carga D - NORTH	128
8.2.4.1	<i>Benchmarks</i> de séries temporais	128
8.2.4.2	FFNN e ranqueamento das estações meteorológicas	129
8.2.4.3	Combinação das estações meteorológicas	132
8.2.5	Zona de carga E - MHKVL	135
8.2.5.1	<i>Benchmarks</i> de séries temporais	135
8.2.5.2	FFNN e ranqueamento das estações meteorológicas	136
8.2.5.3	Combinação das estações meteorológicas	139
8.2.6	Zona de carga F - CAPITL	142
8.2.6.1	<i>Benchmarks</i> de séries temporais	142
8.2.6.2	FFNN e ranqueamento das estações meteorológicas	143
8.2.6.3	Combinação das estações meteorológicas	146
8.2.7	Zona de carga G - HUDVL	149
8.2.7.1	<i>Benchmarks</i> de séries temporais	149
8.2.7.2	FFNN e ranqueamento das estações meteorológicas	150
8.2.7.3	Combinação das estações meteorológicas	153
8.3	RESUMO DOS RESULTADOS	155
9	CONCLUSÕES	156
9.1	TRABALHOS FUTUROS	158
	REFERÊNCIAS	160
	APÊNDICE A – Séries temporais: conceitos e notação	170

APÊNDICE B – Mais gráficos e estatísticas para análise exploratória dos dados.	173
--	-----

1 INTRODUÇÃO

A eletricidade não é um produto criado pelo ser humano. Antes que Thomas Alva Edison pudesse inventar a lâmpada, a energia elétrica já se manifestava na natureza através dos raios, dos mecanismos de defesa de alguns animais, do funcionamento do nosso sistema nervoso, além de outras formas. No entanto, a partir do momento em que o ser humano descobre como produzir, transportar e utilizar a eletricidade para fazer funcionar máquinas e aparelhos, passamos a experimentar uma revolução social, econômica e tecnológica que influenciaria para sempre o curso da História.

Foram três as principais aplicações que impulsionaram o uso da eletricidade ao redor do mundo: a luz elétrica, o desenvolvimento e uso do motor elétrico, e a “portabilidade” da eletricidade, com o desenvolvimento das baterias e a aplicação aos transportes [1]. O uso comercial destes três serviços desenvolveu-se a partir do século XIX. Nessa época, ter uma casa iluminada por lâmpadas ainda era símbolo de *status* social. Em pouco tempo, porém, o desenvolvimento tecnológico levou à redução de custos, possibilitando que pessoas da classe trabalhadora também tivessem acesso à eletricidade, eliminando a restrição dia-noite. Isso reduziu o riscos do trabalho, o que, em termos econômicos, se traduziu em maior produtividade e aumento dos lucros [1]. À medida que ampliar circuitos pré-existentes ligados às usinas torna-se uma operação de mais baixo custo, a introdução de novos consumidores passa a criar economias de escala. Assim plantam-se as sementes para a formação da cadeia de suprimentos do sistema elétrico (CSSE).

A gestão da CSSE tem uma peculiaridade diretamente relacionada com a natureza de seu produto final: como não é possível manter grandes estoques de eletricidade para uso posterior¹, a demanda deve ser atendida em tempo real. Por exemplo, se um cliente residencial quiser utilizar o chuveiro elétrico para um banho quente, a eletricidade necessária para aquecer a água deverá estar disponível no momento em que ele ligar o aparelho, sem que seja necessário avisar com antecedência. A obrigatoriedade de disponibilizar a energia no local e na quantidade desejada pelo cliente torna a previsão de demanda uma das atividades mais importantes para as entidades que compõem a CSSE.

Durante anos, os mercados de energia elétrica ao redor do mundo seguiram estrutura de monopólios verticalmente integrados, com administração estatal. Antes dos anos 90, quando houve uma onda de reestruturação do setor elétrico, a Indústria se encontrava em uma zona de conforto: utilizando métodos de séries temporais e Econometria, as empresas estatais conseguiam prever as necessidades dos consumidores comerciais, industriais e residenciais com acurácia suficiente para fazer o planejamento até o médio prazo (*i.e.* de duas semanas até três anos à frente) [3]. Caso a demanda excedesse a oferta, os sistemas

¹Em sistemas hidrelétricos é possível armazenar energia potencial por meio de reservatórios de água, mas isso é diferente de guardar energia elétrica. Com as tecnologias atuais, o armazenamento, em grandes quantidades, teria custos proibitivos [2].

operavam em regime de sobrecapacidade. Por isso, não-raro havia produção em excesso e o excedente não-estocável convertia-se em desperdício.

A economia dos mercados de eletricidade passou a receber críticas devido à baixa eficiência, e muitas entidades passaram a advogar em favor da liberalização. Sobre a argumentação dos agentes favoráveis à mudança:

Apesar de pequenas diferenças, a motivação para a liberalização dos setores elétricos ao redor do mundo partilhou de razões políticas e ideológicas. Em particular, de uma forte crença de que o sucesso da liberalização de outras indústrias poderia ser replicada pelo setor elétrico, e da “necessidade” de se quebrar (ou desagregar) as estruturas de monopólio verticalmente integrado que tradicionalmente administraram a geração, transmissão e distribuição. ([4], p.1)

O cerne da mudança estaria na intenção de promover ganhos, impulsionar a inovação tecnológica e estimular o investimento eficiente no longo prazo [4]. Seguindo tal premissa, os mercados de eletricidade passaram a se reorganizar a partir de 1986, a começar pelo Chile, seguido de Grã-Bretanha, Noruega, Suécia, Finlândia, Dinamarca, Austrália, Nova Zelândia e Estados Unidos [4].

As mudanças imediatamente fizeram com que a previsão de demanda passasse a exercer ainda mais influência sobre os lucros, o *market share* e o *valuation* das empresas [3]. No mercado atacadista, formato almejado pela maioria dos sistemas que optaram pela quebra do monopólio, empresas interconectadas compram e vendem livremente a energia elétrica com base nas expectativas do setor [2]. A negociação é feita em um mercado de curto prazo onde o preço (em muitos casos recalculado em base horária) é função da diferença entre a oferta declarada pelos geradores e a demanda dos consumidores no momento [2]. Logo, caso uma empresa erre a previsão, haverá desequilíbrio entre oferta e demanda, influenciando o preço da energia para transações e, portanto, os custos das operações.

1.1 A CADEIA DE SUPRIMENTOS DO SETOR ELÉTRICO (CSSE)

Um sistema elétrico (também chamado de *grid*, devido à estrutura de rede interligada, similar a uma “grade”) é definido como “um conjunto de usinas, cargas e redes de transmissão e distribuição, conectadas como um circuito elétrico” [1]. O circuito é tipicamente “grande”, no sentido de que é possível que haja usinas localizadas em regiões geográficas diferentes e que atendem a outras localidades. Consideremos, por exemplo, a CSSE brasileira, o Sistema Interligado Nacional (SIN). O SIN é dividido em quatro subsistemas (Sul, Sudeste/Centro-Oeste, Nordeste e Norte), cuja capacidade instalada de geração é composta, principalmente, por hidrelétricas distribuídas ao longo de dezesseis

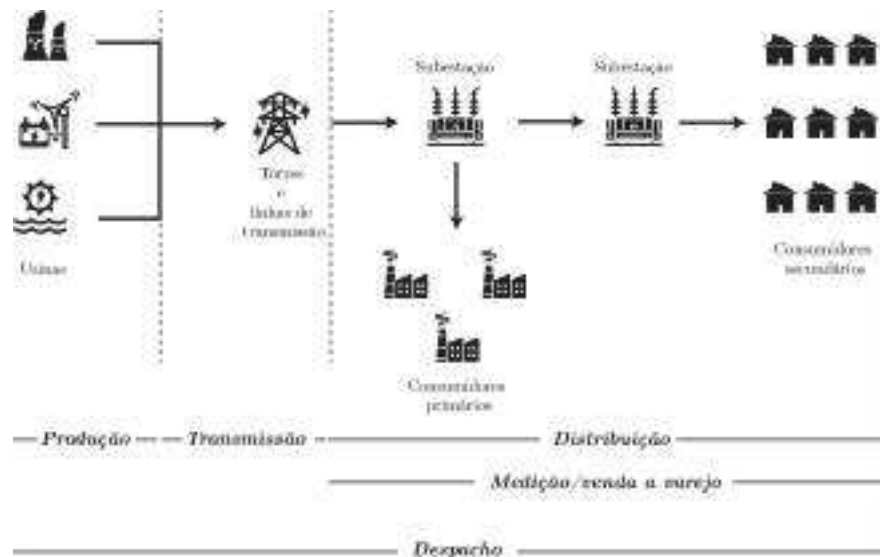
bacias hidrográficas. Os subsistemas são interconectados por linhas de transmissão, de modo que a eletricidade possa ser transferida entre eles, aproveitando a diversidade dos regimes hidrológicos das bacias como parte da estratégia de produção [5].

A **carga** é peça-chave para nosso trabalho. Definimos carga como a potência, em um ponto do sistema, mensurada em múltiplos ou submúltiplos de watts (W), consumida por um componente para desempenhar uma função.

A rede de negócios que opera a *grid* de modo a provisionar eletricidade aos clientes é o que denominamos CSSE. As atividades da CSSE podem ser divididas em produção, transmissão, distribuição, medição/venda a varejo, e despacho [1].

Na Figura 1.1, mostramos um esquema de um sistema elétrico, simplificado mas suficiente para identificar as atividades de uma CSSE.

Figura 1.1 - Fluxograma para a CSSE



A produção é a etapa de geração da eletricidade, é feita em usinas que operam tecnologias de diversas naturezas, como hidrelétrica, solar, eólica, nuclear *etc.* Podem coexistir com as usinas entidades que produzem a própria energia, mas sem injetá-la no sistema (a chamada *auto-geração*), como é o caso de algumas empresas do ramo industrial.

A transmissão consiste na movimentação da eletricidade das usinas até subestações próximas dos clientes. Por meio de linhas de transmissão conectadas às usinas, a eletricidade é levada à região de consumo.

A rede de transmissão conecta-se à de distribuição, que é o conjunto de transformadores e linhas que entregam a eletricidade aos consumidores finais. A conexão é feita por meio das subestações, cujo principal papel é reduzir a voltagem para que a eletricidade seja própria para consumo. No caso dos consumidores primários (*i.e.* industriais), a tensão é reduzida de alta para média. Já no caso dos consumidores secundários (residenciais e

comerciais), este procedimento pode ser feito outras vezes por meio de subestações locais, a fim de que a tensão seja reduzida de média para baixa ².

A medição e a venda da eletricidade para o consumidor final estão diretamente relacionadas à distribuição, e é por isso que tais atividades costumam ser feitas pela mesma empresa. Durante muitos anos, a medição foi ³ realizada conferindo periodicamente um medidor instalado no local de consumo. De tempos em tempos, normalmente em intervalos calculados de forma que seja possível que a empresa colete informações de todos os clientes, o pessoal contratado registrava o consumo mostrado no aparelho, calculava a diferença para a última medição, e tarifava o valor resultante. Além de implicar em custos com pessoal, este método não permite conhecer o padrão de consumo ao longo do intervalo entre as medições. Nos últimos anos, empresas das CSSE têm investido na instalação de medidores mais modernos, os *smart meters*, capazes de calcular e armazenar o consumo em curtos períodos de tempo (ou até mesmo mostrá-lo em tempo real). Alguns *smart meters* podem agrupar informações de consumo por dia da semana e por hora, permitindo que o próprio usuário conheça melhor seu perfil e trace estratégias para racionalizar o uso da energia, além de possibilitar que os demais participantes do mercado entendam o perfil dos usuários.

Ao final do fluxo, temos as atividades de **despacho de carga**. O despacho de carga é a seleção ótima de usinas e cargas, sob restrições de transmissão e restrições operacionais [1]. As empresas responsáveis pelo despacho devem, seguindo algum critério ótimo (como a minimização dos custos), programar quais usinas devem produzir e quais regiões devem ser servidas, para que o fluxo de energia na *grid* esteja balanceado. Vários parâmetros são levados em conta para elaborar a agenda de despacho, como a disponibilidade das usinas, topologia da *grid*, disponibilidade das linhas e, principalmente a previsão de carga, tanto em quantidade quanto em localização na *grid* [1].

A operação da *grid* é feita por uma ou mais empresas, que chamamos **Operador do Sistema (OS)**. A essência da operação da *grid* está no balanceamento: se a carga que deve ser distribuída não for atendida imediatamente pela geração, o sistema estará desbalanceado [1]. Nesse caso, o OS deve lançar mão de algum mecanismo compensatório, já que a eletricidade injetada na *grid* deve coincidir com a que será retirada dela. Quando isso ocorre, é preferível interromper a retirada, procedimento denominado **alívio de carga** [1]. O alívio de carga traz uma série de prejuízos, a começar pelo impedimento súbito de o consumidor utilizar a carga com que esperava contar. Em termos práticos, as consequências são as diversas, desde a queima de aparelhos eletrodomésticos até acidentes de trânsito (causados pelo desligamento de semáforos), além de falhas em sistemas de controles em hospitais, *datacenters*, etc.

²O sistema de distribuição brasileiro opera em níveis de alta tensão (superior a 69 kV e inferior a 230 kV), média tensão (superior a 1 kV e inferior a 69 kV) e baixa tensão (igual ou inferior a 1 kV) [6]

³E, em vários lugares do mundo, ainda é.

1.2 NOVOS DESAFIOS PARA A PREVISÃO DE CARGA ELÉTRICA

A influência da previsão de carga sobre a formação de preços, o custo operacional, a qualidade do abastecimento e a confiabilidade do sistema motivam há décadas a pesquisa acadêmica na área. Além destes temas de interesse, mudanças no cenário global nos últimos anos têm trazido novos desafios.

Nesta seção, discorreremos sobre duas mudanças que julgamos cruciais para os próximos anos: **mudanças climáticas** e **projetos de cidades inteligentes**.

Antes, vale comentar que o tema também encontra eco nos **Objetivos de Desenvolvimento Sustentável** da Organização das Nações Unidas, a ONU [7]. Em particular, destacamos que os resultados podem contribuir para atingir os seguintes objetivos:

(i) **Objetivo 7: Energia limpa e acessível - Assegurar o acesso confiável, sustentável, moderno e a preço acessível à energia para todas e todos.**

Previsões de carga mais acuradas equilibram a oferta e a demanda, contribuindo para manter o nível dos preços, evitando desperdício de recursos naturais e garante que não haverá falta. Também a pesquisa em aparatos da CSSE leva ao aprimoramento das redes, com possibilidade de ampliar o acesso e reduzir custos no longo prazo.

(ii) **Objetivo 11: Cidades e comunidades sustentáveis - Tornar as cidades e comunidades mais inclusivas, seguras, resilientes e sustentáveis.** Previsões de carga mais acuradas contribuem para redução do impacto ambiental nas cidades e comunidades, já que suprir a demanda de energia faz com que evite-se recorrer, em situação de emergência, a formas de geração como as termoelétricas, que liberam gases de efeito estufa.

1.2.1 Mudanças climáticas

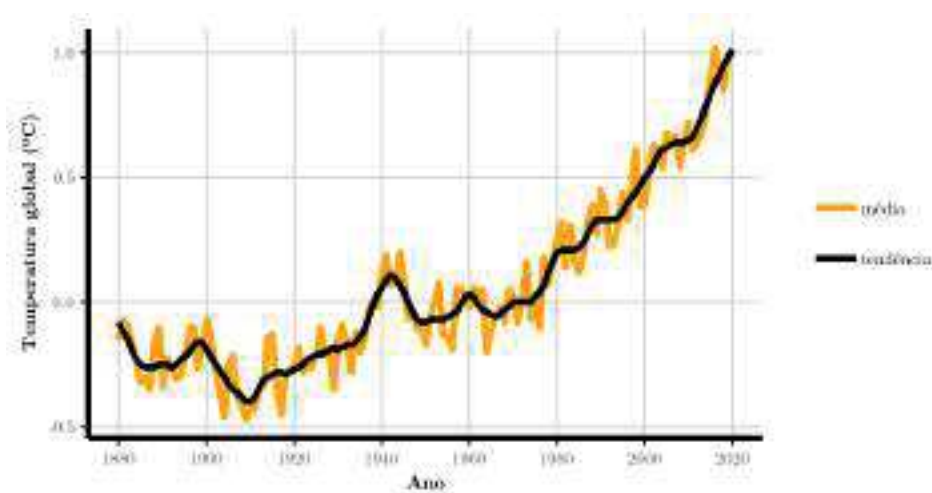
Existem registros de mudanças climáticas ao longo de toda a história da Terra. A maioria dessas mudanças são atribuídas a variações muito pequenas na órbita do planeta, que alteram a quantidade de energia solar que recebemos [8]. Porém, nos últimos anos, a Terra tem experimentado um aquecimento progressivo que evolui com rapidez sem precedentes.

Ao investigar as razões das mudanças, o Painel Intergovernamental Sobre Mudanças Climáticas (*Intergovernmental Panel on Climate Change*, IPCC⁴), concluiu que há mais de 95% de probabilidade de que a principal causa deste aquecimento seja a atividade humana nos últimos 50 anos, o que está registrado no Quinto Relatório de Avaliação do IPCC [10]. Este relatório chama atenção para o aumento na emissão de gases-estufa, como dióxido de carbono, metano e óxido nitroso, decorrente de atividades industriais e do agronegócio.

⁴O IPCC é um grupo de mais de 1300 especialistas de diversos países ao redor do mundo, supervisionados pelas Nações Unidas [9].

Diversas evidências, obtidas por registros de satélites que orbitam a Terra e outras tecnologias, corroboram a tese das mudanças climáticas. Dados coletados ao longo de anos mostram, como sinais das mudanças, o aumento da temperatura global, o aquecimento dos oceanos, o encolhimento de camadas de gelo, a retração de glaciares, a redução da camada de neve, o aumento do nível do mar, a observação de eventos extremos (como precipitação além do esperado) e a acidificação dos oceanos [8].

Figura 1.2 - Temperaturas globais anuais médias e curva de tendência (obtida pelo método LOESS).



Fonte: Dados de [12]

Podemos ter uma ideia da mudança nas temperaturas observando o gráfico da Figura 1.2. Nesta figura, acompanhamos a evolução da temperatura média anual global desde 1880. A curva de tendência, obtida pelo método LOESS [11], reforça o aumento das médias anuais globais, em especial nos últimos 50 anos.

O clima é um dos principais fatores que influenciam o comportamento da carga elétrica, o que discutiremos com maior profundidade em outras seções da tese. O impacto da variabilidade climática na curva de carga elétrica faz com que seja necessário pensar em como incluir os efeitos das mudanças climáticas nos modelos de previsão [14].

Embora esta tese tenha como foco as previsões de curto prazo, é importante dizer que é imperativo que o impacto das mudanças climáticas seja levado em conta também para as previsões de médio e longo prazo. A principal razão disso está na assinatura de acordos entre países com medidas para mitigar os efeitos das mudanças climáticas no longo prazo. Tais acordos contêm cláusulas que afetam diretamente a forma como se produz e se consome a energia, o que significa que devemos esperar mudanças nos padrões na curva de carga para os próximos anos.

Como exemplo de acordo internacional sobre mudanças climáticas, podemos citar a Diretiva 2018/2001/EU [13], firmada por 29 países europeus e renovada no ano de

2018. Esta diretiva fixa metas para o aumento do uso de energias renováveis nos países participantes até o ano de 2030. As gerações eólica e solar são realidade em muitos países; em especial, tem-se observado a tendência de sua auto-geração em fazendas [14]. Como o cálculo da demanda líquida da CSSE é dado pela previsão de carga menos a previsão da auto-geração, devemos pensar em como incluir esta última componente nos modelos.

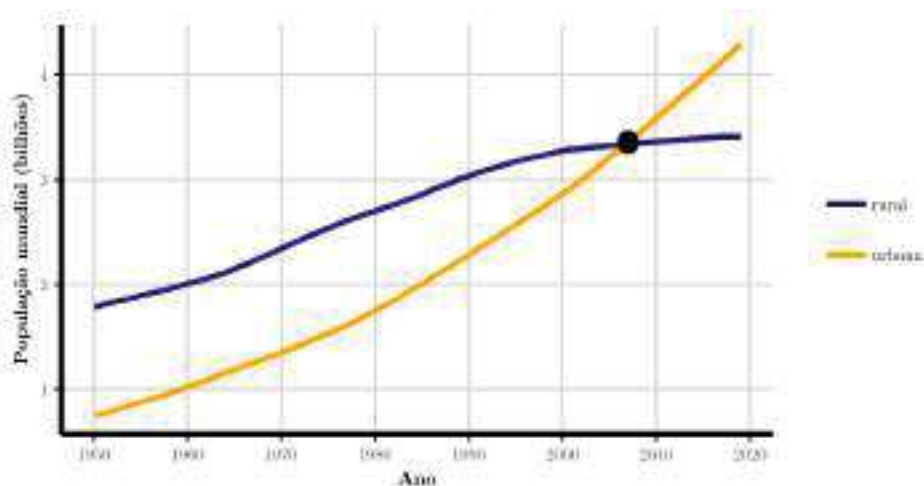
A previsão da carga de auto-geração solar, particularmente, tem outros pormenores. Muitos painéis solares não possuem medidores e, além disso, a penetração de painéis no mercado no longo prazo depende de incentivos fiscais dos governos, o que é difícil antever [14].

1.2.2 Cidades inteligentes

Como a reconfiguração do espaço urbano impacta diretamente o planejamento das CSSE, não é possível desvincular as discussões sobre o mercado de energia do fato de que cidades ao redor do mundo têm se reorganizado para se tornarem **Cidades inteligentes**⁵.

Durante a maior parte da História, o ser humano viveu em pequenas comunidades rurais. Foi nos séculos (e principalmente nas décadas) mais recentes que uma migração em massa do campo para as cidades passou a ocorrer no mundo inteiro.

Figura 1.3 - População mundial entre 1950 e 2019



Fonte: Dados de [15].

A Figura 1.3 mostra a evolução das populações mundiais, urbana e rural, desde 1950. Do gráfico, depreendemos que, desde 1950, a população mundial urbana persiste no que parece ser uma tendência de crescimento (possivelmente exponencial). Em contrapartida, a população mundial rural, que manteve-se em crescimento mais moderado (possivelmente linear) desde a década de 1950, parece caminhar para uma constante. Hoje, 55% da

⁵O termo *smart cities* também é muito usado, mas preferimos adotar o equivalente em Português.

população mundial vive nas cidades, e projeções dão conta de que esta proporção deva ir a 68% até 2050 [15].

Apesar dos pontos positivos que atraem os migrantes para a zona urbana, a vida nas cidades costuma remeter a uma série de aspectos negativos, como a poluição do ar, o tráfego carregado, a violência, a alta densidade demográfica, o déficit habitacional e a poluição sonora. Um projeto de uma cidade inteligente tem como principal ideia utilizar a inovação tecnológica para amenizar estes problemas e introduzir vantagens competitivas.

Uma cidade inteligente é aquela que “aplica tecnologias digitais sistematicamente para reduzir o uso de insumos, melhorar a qualidade de vida das pessoas, e aumentar a competitividade da economia regional de forma sustentável” [16]. Projetar uma cidade inteligente demanda soluções inteligentes para infraestrutura, energia, habitação, serviços, e segurança, baseadas em tecnologia integrada de sensores, conectividade, análise de dados, e processos que agreguem valor de forma independente [16].

Quanto aos impactos nas CSSE, os projetos de Cidades Inteligentes, demandam a substituição das redes de eletricidade antigas por *smart grids*. Uma *smart grid* compreende infraestrutura, sistemas eletrônicos e sistemas de computador conectados por aparelhos de comunicação rápida e protocolos padronizados, consequência do aumento da demanda, somado ao desgaste da infraestrutura das redes antigas e o desenvolvimento de tecnologias de baixa emissão de carbono, como veículos elétricos e energias renováveis [1]. Estes novos arranjos representam um avanço em relação aos sistemas tradicionais, porque o alto grau de automação permite até mesmo monitorar a carga e conhecer sua intensidade em tempo real, por sensores instalados em diversos pontos da *grid*.

Um aspecto que deve ser levado em conta é que uma das principais medidas de incentivo à energia limpa e melhoria na mobilidade urbana (o que não apenas são ideais de uma Cidade Inteligente, mas também constam em acordos pela mudança climática) está na popularização do acesso aos veículos elétricos. Os picos usuais de 3 a 5 KW por dia nos lares europeus foram amplamente ultrapassados após a introdução dos veículos elétricos [16]. Se pensarmos que projetos de cidades inteligentes devem levar em conta a instalação de pontos para carregamento dos veículos ao longo do território urbano, há que se incorporar o efeito da demanda de energia dos veículos elétricos nos modelos de previsão.

Para planejamento estratégico das CSSE no âmbito das cidades inteligentes, alguns fatores exigem atenção. Em particular, cerca de 70% da demanda mundial de energia está nos centros urbanos; isto implica a existência de regiões com alta densidade energética, o que é um dificultador para as metas de descentralização da produção e de autogeração de energia dos países europeus [16].

Devemos também destacar que, na essência da gestão de uma cidade inteligente, está a coleta de dados a todo instante, com alto nível de detalhes. O conceito da **Internet**

das Coisas (IoT, do inglês *Internet of Things*) está intimamente ligado a isso. A Internet das Coisas (IOT) compreende a relação entre as pessoas e produtos, serviços e lugares (as “coisas”) através da tecnologia, como sensores instalados em casas, roupas e redes de energia, além de smartphones e tablets. Estes dispositivos conectam-se à Internet e produzem dados. A mudança na forma como gerenciamos cadeias de suprimento, então, muda, com a possibilidade de monitorar as atividades em maior nível de detalhes. Tal abordagem melhora a significativamente a coleta de dados (e conseqüentemente a precisão), mas obviamente impacta a demanda de energia, já que é o monitoramento é feito por aparelhos eletrônicos e o armazenamento depende do funcionamento ininterrupto de *datacenters*.

1.3 COMO SE PREVÊ A CARGA ELÉTRICA?

A previsão de carga elétrica é tradicionalmente feita pelo entendimento dos padrões da carga, a partir de dados históricos, e da extrapolação destes padrões para o futuro (algumas vezes levando em conta fatores exógenos).

Antes de escolher um método de previsão, é necessário definir o horizonte de previsão e conhecer as características dos dados de carga.

1.3.1 Horizontes de previsão

A previsão de carga elétrica é classificada, de acordo com o horizonte de planejamento, em quatro grupos: **curtíssimo prazo**, **curto prazo**, **médio prazo** e **longo prazo**. Podemos estabelecer horizontes de até um dia para as previsões de curtíssimo prazo, um dia até duas semanas para as de curto prazo, duas semanas até três anos para médio prazo e a partir de três anos para longo prazo [14].

Previsões em diferentes horizontes servem a diferentes propósitos dentro da CSSE. As previsões de curtíssimo prazo, que chegam a ser feitas para alguns minutos à frente, são a base de procedimentos automatizados, como a manutenção de alguns parâmetros de eficiência no deslocamento da carga entre subestações [17]. Já as previsões de médio prazo (decisão em nível tático), subsidiam decisões que não podem ser feitas com tanta urgência, como a compra de maquinário e combustível. Finalmente, no longo prazo, previsões podem fundamentar decisões como a construção e a desativação de usinas, que levam tempo maior, dependem de mais capital e têm impacto mais duradouro.

Apesar de a gestão da CSSE depender de previsões para todos os horizontes, a liberalização e a emergência de mercados competitivos fizeram com que as previsões em curto prazo, especialmente as feitas em base horária com até 36 horas de antecedência, passassem a receber atenção especial [3]. Como a oferta e a demanda flutuam diariamente, e os preços podem subir até dez vezes em questão de horas, antecipar-se à demanda no curto prazo interessa a todos os *players* na CSSE, razão pela qual estas previsões são feitas

e revistas todos os dias [4]. As previsões de carga elétrica em curto prazo, em particular as previsões de perfil diário (*i.e.* as 24 cargas horárias do dia) são o foco desta tese.

1.3.2 Características dos dados de carga elétrica

Mesmo que os padrões de consumo variem de região para região, a carga elétrica tem características que se repetem quando falamos de cargas de consumidores comerciais e residenciais ⁶.

Os principais fatores que influenciam o comportamento da carga elétrica comercial e residencial são **dias atípicos**, **fatores sazonais** e **fatores dependentes do clima** [4]. Nesta seção, explicamos cada fator através de exemplos com dois conjuntos de dados.

O primeiro conjunto de dados contém as cargas (MW ⁷) a cada meia hora, para os anos 1997, 1998 e janeiro de 1999 para uma região da Eslováquia [22]. Estes dados são da *Eastern Slovakian Electricity Corporation* e foram disponibilizados no ano de 2001, no contexto de uma competição promovida pela *European Network on Intelligent Technologies* (EUNITE). O objetivo da competição era desenvolver aplicações de tecnologias inteligentes à previsão de carga elétrica. Na data em que escrevemos este texto, a página da competição na Internet já não pode mais ser acessada em seu endereço original, mas uma cópia ainda pode ser visualizada através do *Internet Archive* [23], iniciativa que arquiva páginas da Internet periodicamente e as disponibiliza em uma plataforma.

O segundo conjunto contém dados horários de carga (MW) e temperatura ambiente (originalmente em °F, mas que convertemos para °C), observadas de 01/01/1982 a 12/10/1992, para região de Puget Sound, no estado de Washington, nos Estados Unidos. Na data em que escrevemos, os dados ainda estão disponíveis na página do professor Mohamed A. El-Sharkawi [24], na Universidade de Washington.

1.3.2.1 Dias atípicos

Caso o comportamento da carga em certo dia seja muito diferente do que acontece nos demais, as divergências poderão ser absorvidas pelo modelo e repassadas para as previsões, comprometendo sua qualidade. Padrões diferentes normalmente ocorrem em dias de feriado, dias de eventos como finais de campeonatos de esportes, dias de manifestação popular (como greves), e outros.

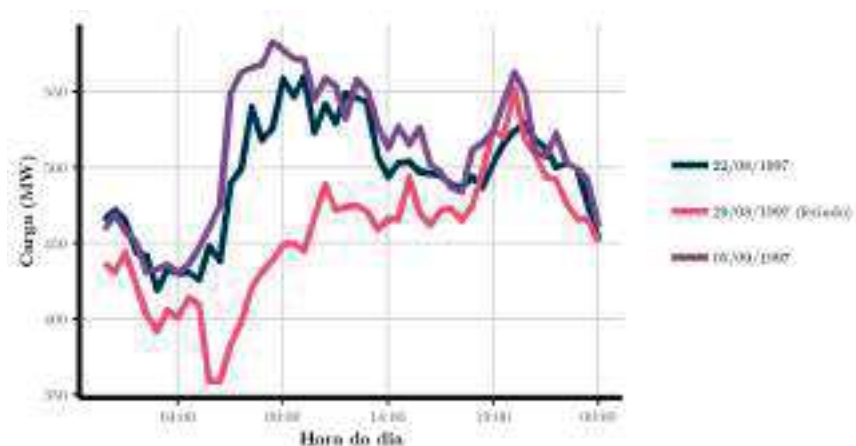
No gráfico da Figura 1.4, observamos, para os dados da Eslováquia, as cargas horárias de três quartas-feiras seguidas, 22/08, 29/08 (feriado do Dia Nacional da Revolta da Eslováquia) e 05/09, do ano de 1997.

Nas três quartas-feiras, o perfil diário da carga obedece a um padrão gráfico semelhante. Contudo, o nível das cargas no feriado de 29/08 é mais baixo, o que podemos

⁶Consumidores industriais fogem do foco desta tese.

⁷Grandeza física de potência, equivalente a 10^6 Watts.

Figura 1.4 - Cargas horárias para três quartas-feiras seguidas (escolhidas por conveniência), sendo 29/08/1997 um feriado (dados da Eslováquia)



Fonte: Dados de [22].

deduzir que ocorra devido ao fechamento do comércio. Caso os dados desse dia fossem utilizados como entradas de um modelo de previsão, é provável que os resultados fossem prejudicados, já que se descreveria o padrão de cargas futuras com base em um comportamento não-convencional.

É aconselhável averiguar e tratar cargas anômalas para adequá-las à distribuição dos dados. Uma estratégia para que estas cargas não prejudiquem as previsões é substituí-las por outros valores que não provoquem grandes perturbações nas previsões dos dias regulares e prever o dia anômalo em separado. Para mais formas de lidar com dias especiais, a literatura traz uma gama de métodos específicos [25].

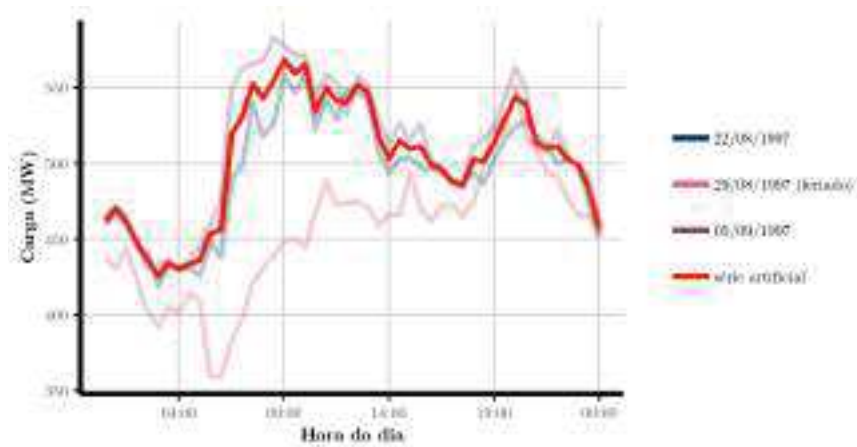
Na Figura 1.5, mostramos um exemplo de uma série de valores imputados usando a média aritmética simples entre as cargas horárias dos dias 22/08 e 05/09 (as duas quartas-feiras adjacentes ao feriado). Este novo perfil parece mais próximo dos demais do que aquele originalmente observado no dia 22/08.

1.3.2.2 Fatores sazonais

Fatores periódicos, como o dia da semana, a hora do dia e a estação do ano também influenciam o padrão da carga. Podemos ver isto na Figura 1.6, que exhibe as cargas horárias de duas semanas de inverno (13/01/1997 a 26/01/1997) e duas de verão (07/07/1997 a 20/07/1997), para os dados da Eslováquia.

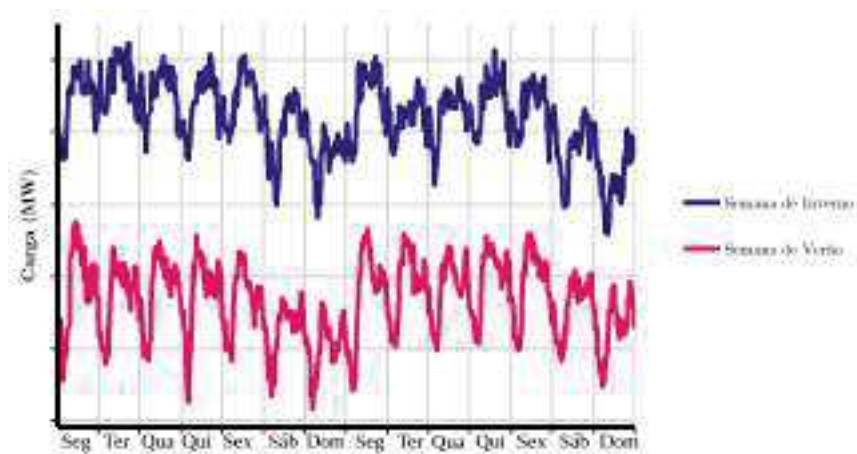
Na Figura 1.6, repete-se a cada dia um padrão de cargas mais altas nas horas mais próximas ao meio do dia e mais baixas durante a madrugada. Vemos também que a estrutura das cargas de dia de semana e de finais de semana é diferente: o nível médio cai no final de semana, provavelmente devido ao fechamento do comércio. Tais características

Figura 1.5 - Cargas horárias para três quartas-feiras (escolhidas por conveniência), e série imputada usando as médias dos dias 22/08 e 05/09 (dados da Eslováquia)



Fonte: Dados de [22].

Figura 1.6 - Cargas horárias para duas semanas de inverno e duas de verão (escolhidas por conveniência; dados da Eslováquia)



Fonte: Dados de [22].

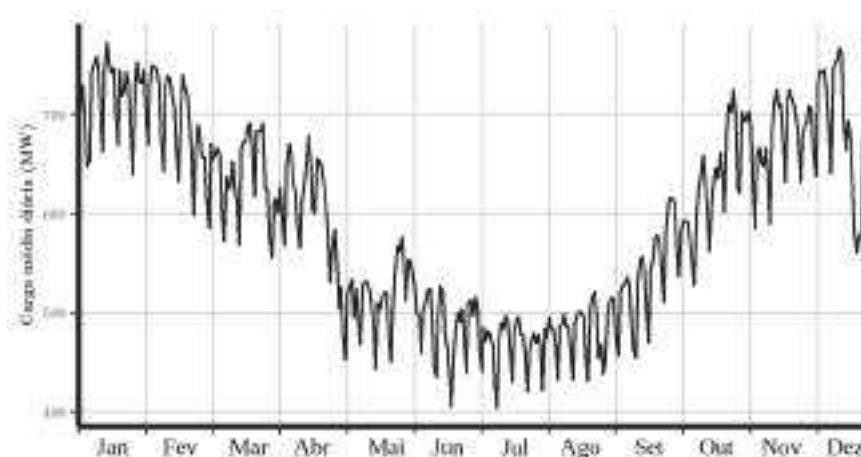
são comuns ao inverno e ao verão, mas o nível da carga no inverno é consideravelmente mais alto.

A diferença entre as estações se dá ao fato de que os aparelhos aquecedores de ar, ligados no inverno, consomem mais energia que os aparelhos de ar condicionado e de ventilação, ligados no verão: um aparelho de ar condicionado consome cerca de um décimo da quantidade de energia consumida por um aquecedor [26]. As circunstâncias em que operam os aparelhos também tem influência: enquanto o ar condicionado normalmente tem que superar uma diferença de temperatura de cerca de 15°F (cerca de 8°C), o aquecedor

lida com diferenças da ordem de 40°F (cerca de 22°C), o que demanda mais carga [26].

Podemos observar o efeito da estação do ano de forma mais clara na Figura 1.7, que traz o gráfico das cargas médias diárias ao longo do ano de 1997, para os dados da Eslováquia.

Figura 1.7 - Cargas médias diárias para no ano de 1997 (escolhido por conveniência; dados da Eslováquia)



Fonte: Dados de [22].

1.3.2.3 Fatores relacionados ao clima

Variáveis relacionadas ao clima (no Inglês chamadas de *weather-dependent variables*), especialmente aquelas que exercem maior influência sobre a sensação térmica, como a temperatura ambiente e a umidade relativa do ar, costumam ser correlacionadas com a carga. Um dos principais motivos é, mais uma vez, o uso de aparelhos de ar condicionado e aquecedores de ar.

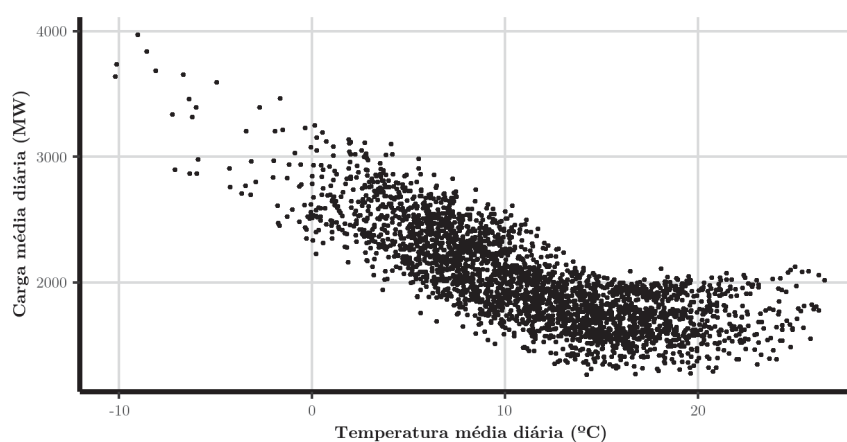
Na Figura 1.8 temos o diagrama de dispersão de carga média diária *versus* temperatura ambiente média diária, para os dados de Puget Sound.

O diagrama da Figura 1.8 indica que existe correlação não-linear entre carga e temperatura. Cargas mais altas são observadas quando as temperaturas são mais baixas (efeito dos aquecedores de ar). À medida que a temperatura aumenta, a carga cai, até que se chegue a um ponto (após os 20°C) em que a carga parece voltar a subir, ainda que relativamente pouco (efeito do ar condicionado).

Podemos chamar os modelos de previsão que levam em conta apenas o histórico de carga de **modelos univariados**, e os modelos que, adicionalmente, consideram os efeitos do clima, de **modelos multivariados**.

Embora modelos multivariados sejam amplamente utilizados e tenham sido reportados casos longevos de sucesso de aplicação industrial [4], agregar variáveis exógenas

Figura 1.8 - Cargas médias diárias *versus* temperaturas médias diárias (dados de Puget Sound)



Fonte: Dados de [24].

ao modelo pode acarretar a dependência de fontes secundárias de dados (como estações meteorológicas), levando ao aumento dos custos ou, minimamente, introduzir um novo fator de risco à operação. A preferência por um modelo univariado ou multivariado permanece controversa entre alguns autores.

Em um estudo em que se comparam modelos univariados com outros baseados em variáveis climáticas, os métodos univariados mostram-se mais acurados para previsões de até 5 horas à frente [27]. Com base nestes resultados, o autor do estudo afirma que variáveis como a temperatura ambiente tendem a mudar pouco no curto prazo, e, logo, tem pouca influência sobre a carga. Toda a variabilidade causada pela temperatura estaria implícita nos próprios dados de carga, afinal, alguns fatores sazonais que afetam a carga existiriam justamente por conta das variações na temperatura. Estes resultados, porém, são vistos com cautela por [14] e [28]. Sobre o trabalho mencionado, um crítico escreve que “um especialista em nível mundial em amortecimento exponencial desenvolveu cuidadosamente vários modelos de amortecimento exponencial”, mas “o artigo menciona brevemente como os métodos baseados na meteorologia funcionam, informação que não é suficiente para julgar quão acurados são estes métodos” [28]. Conclui o crítico que é difícil fazer estudos comparativos de forma justa, porque a qualidade do modelo depende muito da competência do profissional que o desenvolve e aplica: usando a mesma técnica, profissionais diferentes podem obter resultados diferentes, a depender de sua experiência com a técnica em questão.

No âmbito do Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora (PGMC - UFJF), dois trabalhos [29, 30] comparam modelos univariados e multivariados, com testes sobre bases de dados do Rio de Janeiro, da Eslováquia e de Washington. Os autores verificaram que, para previsões horárias para

até 24 horas à frente, a acurácia média dos modelos multivariados parece ligeiramente superior, mas indicam que os resultados mais interessantes parecem vir da combinação das previsões dos dois tipos de método. Acreditamos que o tema ainda possa ser explorado.

Faz-se, neste ponto, importante destacar que, nesta linha de pesquisa, o PGMC-UFJF produziu outros trabalhos ao longo dos anos, tendo investigado desde o uso de modelos clássicos de séries temporais até os de computação intensiva (como redes neurais e *gradient boosting*), passando por aplicações a gestão de contratos de energia. Para conhecer estes trabalhos, consultar as referências [31, 32, 33, 34].

Existem mais fatores que podem influenciar a carga, como os relacionados à Economia, mas estes fogem do escopo deste trabalho.

1.4 O PROBLEMA DA SELEÇÃO DE ESTAÇÕES METEOROLÓGICAS

Elaborar um modelo de previsão baseado em variáveis relacionadas ao clima requer acesso e armazenamento constantes de observações e previsões de variáveis relacionadas ao clima, para, respectivamente, construir modelos e fazer previsões de carga. Previsões meteorológicas podem ser produzidas ou adquiridas de terceiros, enquanto as observações de variáveis relacionadas ao clima costumam ser obtidas de estações meteorológicas.

Uma estação meteorológica coleta dados relacionados ao tempo e ao meio-ambiente do local onde está instalada, usando diferentes sensores, como termômetros, barômetros, *etc* [35]. Uma estação pode medir temperatura, pressão atmosférica, precipitação, umidade, velocidade do vento, entre outras variáveis.

Dentro do território de uma CSSE, pode haver mais de uma estação meteorológica. Daí emerge um problema cada vez mais relevante no atual contexto do mercado de eletricidade: a seleção e combinação de estações meteorológicas. Negligenciado por muitos anos, este problema passou a receber atenção da comunidade de previsão de carga devido a aspectos inerentes aos atuais mercados de energia [20, 21].

Voltemos às *smart grids*. Neste tipo de arranjo, a informação é coletada em tempo real em alto nível de detalhes, o que cria novas possibilidades para os participantes do mercado de energia e faz crescer a importância da **previsão hierárquica de carga**. A previsão hierárquica de carga é uma classe de problemas que envolve prever em vários níveis da *grid*, desde a geração até a residência, para vários horizontes, desde alguns minutos até vários anos à frente [14].

Dissemos que podem-se encontrar várias estações meteorológicas pelo território de uma *grid*. Ao prever a carga de um ponto em determinada localização, de qual estação devemos prospectar os dados?

O problema da seleção de estações meteorológicas ganhou destaque na literatura a partir da primeira edição da GEFCom (*Global Energy Forecasting Competition*), em

2012, competição internacional que proporcionou aos participantes a oportunidade de se engajar na proposição de soluções para a previsão de demanda de energia. Uma das categorias da GEFCom 2012 foi sobre a previsão hierárquica de carga. Nesta categoria, o desafio era prever cargas horárias para 20 zonas de um sistema americano, tanto em nível zonal (20 séries de carga) quanto em nível do sistema como um todo (a soma das 20 cargas zonais). Os organizadores forneceram cargas horárias de quatro anos e meio e o histórico de temperaturas para o período. Oito semanas não-consecutivas de dados de carga foram propositalmente removidas dos dados. A tarefa consistia em prever as oito semanas faltantes (o que é chamado *backcast*; algo como “prever para trás”) e a semana subsequente aos 4 anos e meio de carga. As temperaturas reais dos próximos dias não foram fornecidas - a tarefa de prevê-las também ficava por conta dos competidores. A ideia dos organizadores de propor uma tarefa de previsão em zonas foi simular a previsão de carga em curto prazo no contexto de uma *smart grid*, onde idealmente tem-se acesso à informação em diversos níveis. Para maiores informações sobre a competição, o leitor pode consultar [20], cujo texto fundamenta a escrita deste parágrafo.

Um dos desafios da GEFCom 2012 foi lidar com padrões de consumo diferentes para as diferentes zonas de carga. Os dados de temperatura continham medições feitas por 11 estações meteorológicas espalhadas pelo território do sistema, mas nem a localização das zonas de carga e nem a das estações foram fornecidas. Cabia aos competidores traçar uma estratégia para seleção das estações meteorológicas.

Os resultados mostram que os três primeiros colocados da competição resolveram o problema, respectivamente, das seguintes maneiras⁸:

- (i) Construção de 11 modelos de previsão diferentes para cada uma das 11 estações meteorológicas;
- (ii) Inclusão das temperaturas de todas as estações no modelo de previsão;
- (iii) Utilização um procedimento *stepwise* para escolher a estação que minimizasse o erro de previsão na amostra de teste.

O problema apareceu mais uma vez na edição seguinte da GEFCom, em 2014, que também trouxe uma categoria de previsão hierárquica de carga. Dessa vez, a tarefa dos participantes era prever os quantis das cargas horárias para um sistema americano, para um mês. Inicialmente, foram fornecidos 69 meses de dados de carga e 11 meses de dados meteorológicos em base horária. Ao longo de 16 semanas, os organizadores tornavam disponíveis, semanalmente, um mês adicional de dados de carga e meteorológicos.

Foram oferecidos dados de 25 estações meteorológicas, mais uma vez sem identificar suas regiões geográficas. Os organizadores comunicaram a expectativa de que os compe-

⁸Mais detalhes em [21].

tidores desenvolvessem algoritmos avançados para a seleção de estações meteorológicas [21].

As soluções dos três primeiros colocados na competição foram, respectivamente:

- (i) Uso de dados de quatro estações, escolhidas com base em validação cruzada generalizada;
- (ii) Uso de dados de três estações, escolhidas com base em validação cruzada generalizada, seguida de uma heurística para refinamento;
- (iii) Escolha das estações que melhor descrevem as zonas cargas quando ajustado um modelo de regressão cúbica.

Dos quatro próximos colocados, o seguinte utilizou 11 estações (sem detalhar como fez a escolha) e os demais, curiosamente, não utilizaram nenhum método de seleção: ou incluíram os dados de todas as estações no modelo ou criaram uma estação virtual a partir da média dos dados de todas as estações. Percebe-se, portanto, que não houve engajamento de todos os participantes mencionados no sentido daquilo que era esperado pelos organizadores. É também notável que os três primeiros colocados dedicaram alguma atenção ao tema.

Depois destas competições, foi publicada, por um dos organizadores das GEFCom, uma heurística para determinar quantas e quais estações serão utilizadas [36]. Os autores usaram esta heurística para combinar estações de forma a compor os dados de entrada de um modelo de regressão linear (o chamado *Tao's Vanilla Benchmark*) e fizeram testes para dados de carga de um sistema da Carolina do Norte (EUA), em nível zonal e total, e de temperatura de 25 estações meteorológicas.

Outra publicação trouxe mais tentativas de combinar os dados das estações meteorológicas, após o uso da heurística de seleção, usando médias ponderadas, para os mesmos dados da Carolina do Norte, e para os dados da GEFCom 2012 [37]. O modelo de previsão foi, novamente, o *Tao's Naive Benchmark*.

Estes dois trabalhos serão analisados com mais profundidade no Capítulo 7 desta tese.

1.5 OBJETIVOS DA TESE

Apesar de sua importância, a seleção e combinação de estações meteorológicas permanece pouco explorado na literatura de previsão de carga elétrica. Por conta disso, escolhemos este problema como foco desta tese.

Até o momento, os métodos sugeridos para combinar os dados de diferentes estações guardam pouca ou nenhuma relação com fatores de ordem prática. Em [36], todas as

estações escolhidas recebem peso igual, e em [37] e [38], pesos são calculados apenas com base na resolução de problemas de minimização, novamente ignorando fatores práticos. Por exemplo, seria razoável atribuir maior peso a estações meteorológicas mais próximas da zona de carga ou cujos dados tivessem maior correlação com os dados de carga.

Isto posto, os objetivos da tese de Doutorado são:

- (i) Introduzir novas formas de combinar as melhores estações meteorológicas, a fim de melhorar a qualidade (na forma da acurácia) das previsões;
- (ii) Comparar o desempenho destas novas combinações com os de métodos estabelecidos por [36].

Não são objetivos desta tese escolher o melhor método de previsão ou definir quais variáveis de entrada levam às melhores previsões, embora experimentemos diferentes modelos com diferentes variáveis.

1.6 ORGANIZAÇÃO DA TESE

Esta tese está organizada em dez capítulos e dois apêndices.

Neste primeiro capítulo, fizemos uma introdução aos sistemas elétricos e ao funcionamento da cadeia de suprimentos do setor elétrico, comentamos os atuais desafios da área, sobre o problema da previsão de carga elétrica e nos aprofundamos no problema da seleção de estações meteorológicas. Discutimos a relevância do problema no contexto dos atuais mercados de energia e apresentamos os objetivos da tese de Doutorado.

No Capítulo 2, apresentamos as bases de dados com que trabalhamos nesta tese. Comentamos o processo de limpeza dos dados e fazemos a análise exploratória dos dados que foram utilizados para nossos estudos.

No Capítulo 3, fazemos um histórico sobre a modelagem e a previsão de carga elétrica e estabelecemos os fundamentos para escolha dos modelos e métodos.

No Capítulo 4, apresentamos métodos de previsão de séries temporais, que descrevem a carga em função de seus valores passados.

No Capítulo 5, apresentamos a regressão linear múltipla aplicada à modelagem e previsão de carga elétrica, como forma de incorporar o efeito de variáveis relacionadas ao clima.

No Capítulo 6, apresentamos as redes neurais como método para prever a carga com base em suas relações não-lineares com a temperatura. Comentamos os principais pontos de atenção para sua implementação e uso.

No Capítulo 7, apresentamos uma heurística para seleção e combinação de estações meteorológicas. Revisamos o que se encontra, até este momento, na literatura e propomos novos métodos para combinar as estações.

No Capítulo 8, descrevemos as simulações que realizamos, e seus resultados.

No Capítulo 9, trazemos nossas considerações finais e sugestões para trabalhos futuros.

No Apêndice A, revisamos conceitos e terminologia de Séries Temporais que utilizamos ao longo da tese.

No Apêndice B, mostramos mais gráficos e estatísticas descritivas para os dados de carga e temperatura, que optamos por manter em separado do Capítulo 2 para que este ficasse mais conciso.

2 COLETA, PRÉ-PROCESSAMENTO E ANÁLISE DOS DADOS DE CARGA E TEMPERATURA

O conjunto de dados que utilizamos nesta tese contém observações de carga elétrica NYISO e dados de temperatura da Administração Nacional Oceânica e Atmosférica (*National Oceanic and Atmospheric Administration*, NOAA).

Neste capítulo mostramos como desenvolvemos a coleta, o pré-processamento e análise exploratória do material.

2.1 COLETA DOS DADOS

2.1.1 Dados de carga elétrica

O Operador Independente do Sistema de Nova Iorque (*New York Independent System Operator, Inc.*, NYISO) é uma organização sem fins lucrativos que coordena a distribuição de energia e monitora a confiabilidade do sistema elétrico do estado de Nova Iorque [39].

O NYISO divide o território do estado em onze zonas de carga, conforme o mapa da Figura 2.1. Os rótulos das zonas guardam relação com a região do estado em que se localizam ¹, mas as fronteiras não necessariamente coincidem com as fronteiras do mapa civil do estado [41].

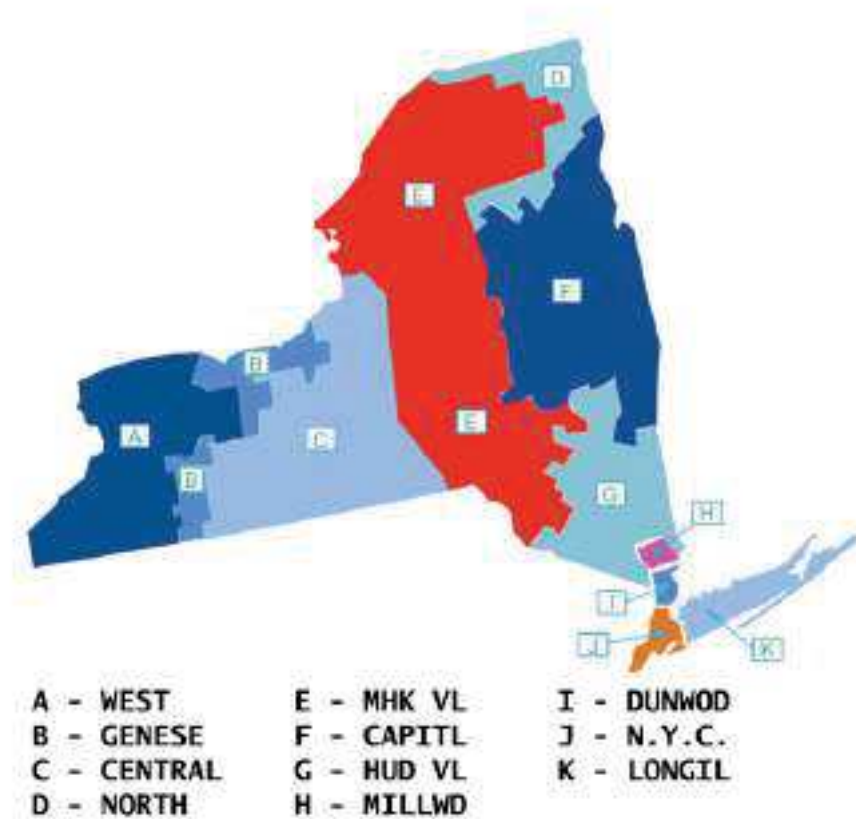
Existem operadores e agências reguladoras, em outras regiões do mundo, que abrem seus dados ao público assim como o NYISO. Por exemplo, na plataforma *web* da Rede Europeia de Sistemas de Transmissão de Energia (*European Network of Transmission System Operators for Electricity*, ENTSO-E) encontramos dados de 42 operadores de 35 países da Europa [42]. Estes dados não são interessantes para nosso trabalho, todavia: como ocorre para grande parte dos operadores no mundo, os dados da ENTSO-E se referem à **carga total** dos países. Conhecer a demanda do sistema inteiro pode ser útil para previsões de médio e longo prazo, mas pouco ajuda no curto prazo, quando a maior parte das decisões são tomadas em nível local (despacho de carga para um bairro de uma cidade, por exemplo). Como nosso foco está nas previsões de curto prazo, optamos pelos dados do NYISO, que estão disponíveis para um sistema estadual, de forma desagregada.

O NYISO alimenta, em tempo real, sua plataforma *web* com dados de carga [43]. Para esta tese, coletamos os dados de cargas horárias, em MW, para os anos de 2015 a 2018, totalizando quatro anos cheios², ou 35.064 observações para cada zona de carga. Períodos

¹A - WEST: oeste, B - GENESE: condado de Genesee, C - CENTRL : central, D - NORTH: norte, E - MHKVL: Vale do Mohawk, F - CAPITL: região da capital Albany, G - HUDVL: Vale do Hudson, H - MILLWD: Millwood, I - DUNWOD: Dunwoodie, J - NYC: cidade de Nova Iorque, K - LONGIL: Long Island.

²A experiência do NYISO mostra que dados de três a quatros anos mais recentes são suficientes para modelar a carga [44].

Figura 2.1 - Zonas de carga do NYISO



Fonte: [40]

posteriores, não incluímos por duas razões: primeiramente, começamos as simulações no final do ano de 2018 e, incluir mais dados demandaria refazer boa parte do trabalho e poderia limitar o tempo para experimentar mais. Segundo, a partir de 2019 os dados teriam o viés da pandemia de COVID-19, e, por isso, certamente demandariam uma análise em separado.

Apesar de ser de nossa vontade, não trabalhamos com dados do Brasil porque não temos fácil acesso. Dentre os critérios de transparência para manter as concessões da rede elétrica do país, não existe a obrigação de que as empresas disponibilizem dados históricos de carga [45]. Com isso, as concessionárias optam por manter proprietários os dados da distribuição em tempo real.

Dentre as zonas de carga do NYISO que mostramos na Figura 2.1, não trabalhamos com as zonas H, I, J e K porque não dispomos das coordenadas geográficas das fronteiras destas zonas. Estas coordenadas são importantes para os métodos de combinações de estações meteorológicas que propomos no Capítulo 7.

As coordenadas das fronteiras das zonas A até G estão disponíveis, nesta data,

na plataforma *web* colaborativa ArcGis, nos endereços das referências bibliográficas de números [46] a [53].

2.1.2 Dados de temperatura

A Administração Nacional Oceânica e Atmosférica (*National Oceanic and Atmospheric Administration*, NOAA) é uma agência do governo americano dedicada ao monitoramento climático. Sua missão é “entender e prever mudanças no clima, tempo, oceanos e costas, compartilhar este conhecimento e informação com terceiros, e conservar e gerenciar ecossistemas marinhos e costais e recursos” [54].

A NOAA disponibiliza, dentro do projeto NCEI (*National Centers for Environmental Information*, a base de dados ISD (*Integrated Surface Database*) [55].

O ISD contém observações mundiais de mais de 35 mil estações meteorológicas de superfície. São dados como temperatura, ponto de orvalho, qualidade do ar, pressão atmosférica, precipitação e velocidade do vento.

Para o estado de Nova Iorque, procuramos pelas 16 estações meteorológicas que o NYISO usa em seus modelos de previsão de carga. Por mais que não seja possível comparar o desempenho de nossos modelos com os do NYISO³, consideramos razoável partir das estações cujo impacto para as cargas do sistema já é cancelado pela própria administradora. Destas estações, a única cujos dados não estão disponíveis no ISD é a estação localizada no Aeroporto MacArthur, de Long Island (estação de código ISP⁴).

Na Tabela 2.1 listamos os códigos IATA, os nomes e as coordenadas geográficas das estações cujos dados utilizamos. Na Figura 2.2 mostramos, no mapa de zonas de carga do NYISO, onde se localizam estas estações.

Os dados das estações meteorológicas são creditados a [57].

2.2 PRÉ-PROCESSAMENTO DOS DADOS

Os registros a que tivemos acesso são de dados brutos, isto é, não foram submetidos a nenhum tipo de tratamento. Nesses casos, é crucial verificar se há valores faltantes, ruídos, duplicatas ou pontos discrepantes e, se for o caso, fazer ajustes. O objetivo, nesta fase, é evitar que anomalias nos registros prejudiquem as análises devido a distorções na distribuição dos dados.

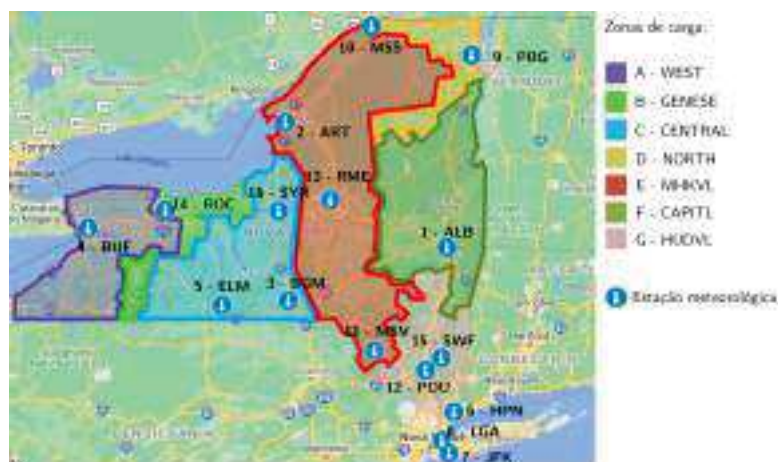
³Algumas informações necessárias para reproducibilidade não são abertas pela organização.

⁴As 16 estações meteorológicas estão localizadas em aeroportos internacionais, razão por que é habitual referir-se a elas por meio do seu código IATA (*International Air Transport Association Identifier*). Existem outros protocolos, mas o IATA é mais utilizado para a climatologia [56].

Tabela 2.1 – Informações sobre as estações meteorológicas.

Código	Estação	Coordenadas
ALB	Aeroporto Internacional de Albany	42°45'N 73°48'W
ART	Aeroporto Internacional de Watertown	44°00'N 76°01'W
BGM	Aeroporto Internacional de Binghamton	42°13'N 75°59'W
BUF	Aeroporto Internacional Buffalo Niagara	42°56'N 78°44'W
ELM	Aeroporto Regional Elmira Corning	42°09'N 76°53'W
HPN	Aeroporto do Condado de Westchester	41°04'N 73°42'W
JFK	Aeroporto Internacional John F. Kennedy	40°39'N 73°47'W
LGA	Aeroporto Internacional de LaGuardia	40°46'N 73°54'W
PBG	Aeroporto Internacional de Plattsburgh	44°39'N 73°28'W
MSS	Aeroporto Internacional de Massena	44°56'N 74°51'W
MSV	Aeroporto Internacional do Condado de Sullivan	41°42'N 74°47'W
POU	Aeroporto Regional de Hudson Valley	41°37'N 73°53'W
RME	Aeroporto Internacional de Griffiss	43°14'N 75°24'W
ROC	Aeroporto Internacional Frederick Douglass	43°07'N 77°40'W
SWF	Aeroporto Internacional de Stewart	41°30'N 74°06'W
SYR	Aeroporto Internacional de Syracuse	43°07'N 76°07'W

Figura 2.2 - Estações meteorológicas e zonas de carga do NYISO (escala 1:8,48)



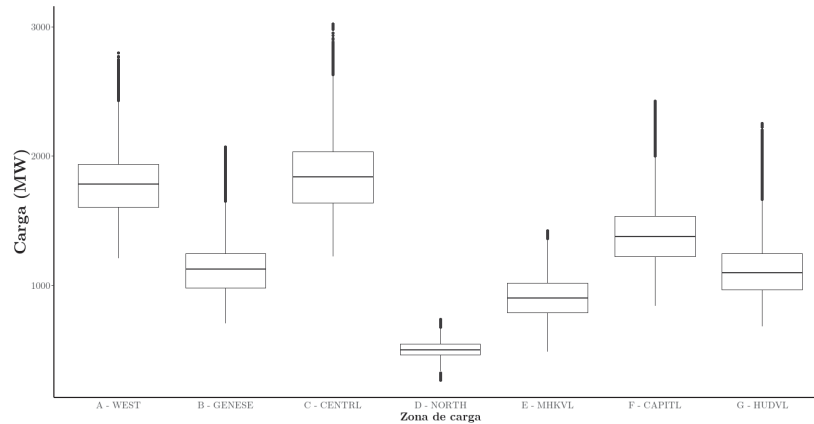
2.2.1 Pré-processamento dos dados de carga

Na base de dados de carga elétrica, os horários do registro estão no fuso GMT-4, em que se localiza o estado de Nova Iorque. Nos anos de 2015 a 2018, o governo adotou a política de *daylight saving time*, análoga ao nosso horário de verão: durante os meses mais quentes, o relógio é adiantado em uma hora, de modo que escureça mais tarde e, assim, promova-se economia de energia. Como a sazonalidade dos dados é importante, removemos o efeito da mudança de relógio. Onde havia adiantamento de uma hora e, portanto, “perda” de uma observação, fizemos a imputação do valor “perdido” pela média aritmética simples das observações anterior e posterior. Já onde havia atraso de uma hora e, logo, aparecimento de uma “duplicatada” no mesmo horário, substituímos as duas

observações por um ponto apenas, resultado da média aritmética simples do par.

No boxplot da Figura 2.3, vemos posição e dispersão dos dados, além de pontos discrepantes.

Figura 2.3 - Boxplot das cargas das zonas A até G do NYISO.



Cargas muito abaixo do usual costumam acontecer em dias de feriado, quando o consumo cai. Para nossa base de dados, somente a zona D - NORTH apresenta valores discrepantes na parte inferior do boxplot, e nenhuma dessas observações foi feita em dia de feriado. Possivelmente, a razão de estas cargas aparentemente não serem afetadas em feriados está na maneira como os EUA administram seus feriados: salvo poucas exceções, feriados são deslocados para as segundas-feiras [58], o que evita uma quebra súbita no padrão da carga no meio da semana. Também devemos levar em conta que paralisação ou não das atividades é de jurisdição local e, por isso, estados e municípios podem optar por não fazê-lo [58].

Valores discrepantes na parte superior do boxplot ocorrem para todas as zonas de carga. Verificamos que, tanto estes valores, quanto os valores extremos na parte inferior do boxplot na zona D - NORTH são observações feitas em dias em que as temperaturas nas cidades mais populosas da zona de carga foram extraordinariamente altas ou baixas [59]. Como os valores não parecem decorrer de erro de registro ou de outros efeitos exógenos para os quais não dispomos de dados, optamos por deixar estes valores como estão.

Na base de dados de carga, não encontramos valores faltantes, ruído ou duplicatas.

2.2.2 Pré-processamento dos dados de temperatura

Não detectamos ruído, duplicatas ou valores discrepantes que pudessem vir de erros de registro, nos dados de temperatura. Verificamos, porém, a existência de alguns valores faltantes. Mostramos quantos são, para cada estação, na Tabela 2.2.

Tabela 2.2 – Número de valores faltantes para cada estação meteorológica.

Estação	Número de valores faltantes
1 - ALB	7
2 - ART	97
3 - BGM	111
4 - BUF	22
5 - ELM	113
6 - HPN	97
7 - JFK	3
8 - LGA	9
9 - MSS	48
10 - MSV	2080
11 - PBG	219
12 - POU	108
13 - RME	9
14 - ROC	21
15 - SWF	1680
16 - SYR	22

No caso das estações ELM, MSV, PBG, POU e SWF, boa parte dos faltantes aparecem **de forma consecutiva**. Como as interpolações de sequências de valores demandam um trabalho mais minucioso (e que acabaria fugindo do nosso escopo), optamos por não trabalhar com estas estações. Para os valores nas demais estações, geramos valores sintéticos por meio de uma interpolação linear.

2.3 ANÁLISE EXPLORATÓRIA DOS DADOS

Antes da análise exploratória, faremos uma breve digressão.

O comportamento da carga elétrica está intimamente ligado ao clima, e o clima de uma região é determinado, dentre outros fatores, pela geografia física. Por isso, apresentamos a seguir as características das principais regiões fisiográficas do estado de Nova Iorque. Estas informações serão importantes para a análise exploratória dos dados.

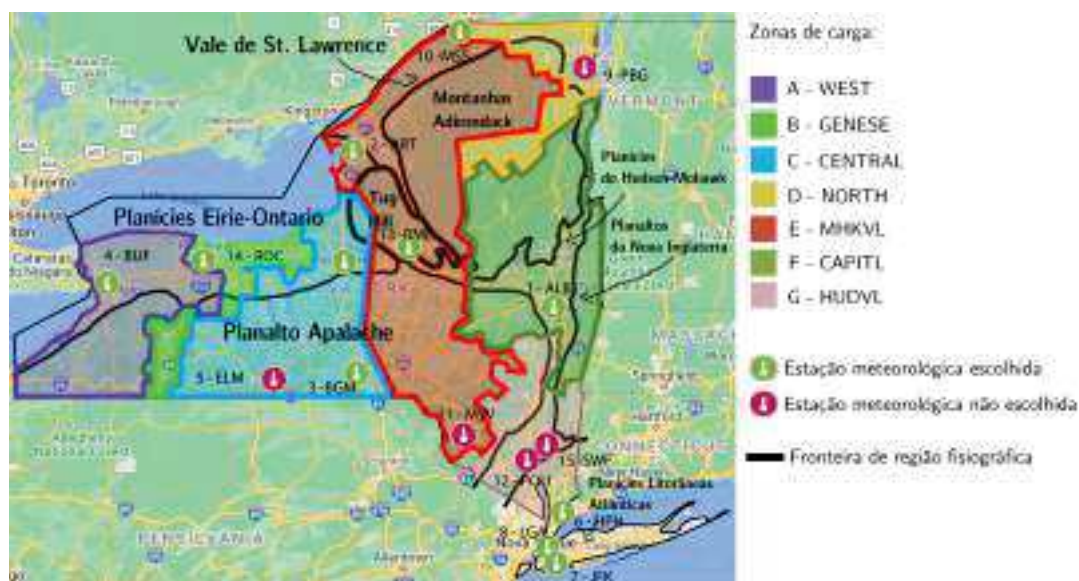
2.3.1 Geografia Física do Estado de Nova Iorque

Trazemos, na Figura 2.4, um mapa com as fronteiras das zonas de carga, as localizações das estações meteorológicas (com aquelas que escolhemos utilizar identificadas) e os limites das regiões fisiográficas⁵.

As temperaturas médias no estado de Nova Iorque variam, em julho, de 18 °C em Indian Lake (nas Montanhas Adirondack) a 25 °C na cidade de Nova Iorque; e, em janeiro, de -10 °C na Reserva Stillwater (também nas Montanhas Adirondack) a 0,5 °C

⁵Fisiográfico é o adjetivo relativo à geografia física (ou fisiografia)

Figura 2.4 - zonas de carga, estações meteorológicas e regiões fisiográficas do Estado de Nova Iorque (escala 1:8,48).



em Long Island [60]. O clima não é homogêneo ao longo do território, e existem diferenças consideráveis entre a cidade de Nova Iorque e as regiões interioranas, como na capital Albany, em Buffalo, Rochester e Syracuse.

A maior região do estado é o planalto dos Apalaches, que se estende desde o Vale do Rio Hudson, a leste, até as fronteiras oeste e sul do estado. Nesta região, encontramos as Montanhas Catskill (com picos de até 1200 metros, onde se praticam principalmente atividades de lazer), os Lagos Finger e a Bacia Fluvial Delaware (onde se praticam agricultura e pecuária) [60]. A Figura 2.4 mostra que apenas a zona de carga F não possui território nos Apalaches. Há três estações meteorológicas nesta região, ELM, BGM e MSV; mas a primeira e a última enquadram-se no caso de muitos valores faltantes em sequência. Representando esta região, ficaremos, portanto, com BGM (que se localiza entre as duas outras, o que pode ser uma vantagem).

A norte dos Apalaches e oeste do Vale do Mohawk, as planícies Eirie-Ontario se estendem ao longo das margens dos Grandes Lagos, próximo da fronteira com o Canadá. Como uma bacia hidrográfica, os Grandes Lagos absorvem e liberam calor com facilidade, promovendo equilíbrio e fazendo com que poucas temperaturas extremas sejam observadas [61]. Ao longo das planícies, encontramos partes dos territórios de quatro zonas de carga, A, B, C e E, bem como quatro estações meteorológicas. Apesar das temperaturas equilibradas por conta dos lagos, certas partes das planícies Eirie-Ontario recebem grandes quantidades da chamada “neve por efeito de lago” [62], como os arredores de Syracuse (em média 2900 mm de neve), próximos do Lago Ontario, e a região de Buffalo (em média 2400 mm de neve), perto do Lago Erie. Felizmente, tanto Syracuse quanto Buffalo abrigam estações meteorológicas (SYR e BUF), o que nos pode ser útil para mapear o possível efeito de

tantos fenômenos climáticos sobre a carga.

A nordeste, no planalto Adirondack, estão as montanhas mais altas do estado. A localização na zona de transição entre as florestas temperadas dos Apalaches e a floresta boreal do Canadá faz com que a região seja abrigo de grande biodiversidade [63]. Três zonas de carga, D, E e F, atendem a região. Não há, porém, estações meteorológicas no território (embora PBG e MSS estejam próximas da fronteira, foram excluídas da análise, como explicamos).

A norte do Adirondack, margeando a fronteira com o Canadá, se alarga o Vale de St. Lawrence, uma planície com lagos. As zonas de carga D e E passam por ali, e MSS e PBG são as estações mais próximas. Devemos ter cautela, todavia, ao tentar associar as temperaturas medidas destas estações com as cargas das zonas D e E, porque parte destas zonas também está no Adirondack, onde a altitude pode ser mais que o dobro do que se observa no Vale de St. Lawrence [64].

Entre as Planícies Eirie-Ontario e o oeste do Adirondack, está o planalto Tug Hill. Esta é uma das regiões menos habitadas por conta do solo pobre e excesso de neve no inverno [60]. As zonas de carga C e E, atendem a região, onde está a estação meteorológica RME.

Partindo da cidade de Nova Iorque e seguindo na direção norte até Albany, a planície do Hudson-Mohwak acompanha o Rio Hudson. A leste do Rio Hudson, está o planalto da Nova Inglaterra, que cruza para Massachusetts e Connecticut. No Hudson-Mohawk estão as zonas E e F, e na Nova Inglaterra a zona E. Nestas regiões, encontramos três estações meteorológicas; uma na parte Norte (a qual escolhemos) e duas na parte Sul (as quais não escolhemos).

No sudoeste encontra-se a Planície Costeira do Atlântico, que vai de Massachusetts até a Flórida, abrangendo Long Island e Staten Island. Não trabalhamos com as zonas de carga que atendem a esta região, mas vamos experimentar incluir as temperaturas das estações HPN, LGA, JFK nos modelos de outras zonas de carga, para investigar como se eles comportam ao ser alimentados dados de estações de regiões diferentes.

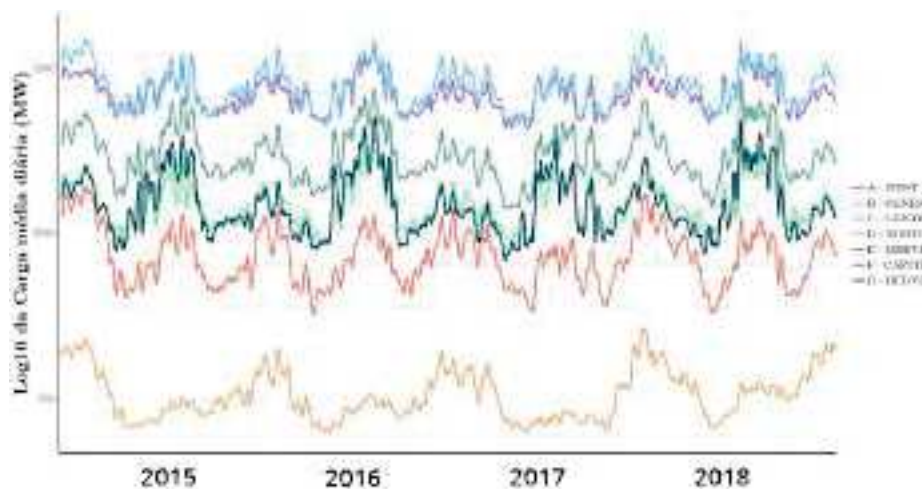
2.3.2 A análise

Começamos a análise investigando a sazonalidade anual dos dados de carga.

Na Figura 2.5, mostramos gráficos de linha para as cargas médias diárias de cada zona. Os dados foram suavizados utilizando médias móveis⁶, e foi aplicada uma transformação logarítmica de base dez no eixo y , para reduzir a distância entre os maiores valores de carga e facilitar a visualização.

⁶O tipo de média móvel escolhido foi uma média móvel simples de janela q , em que um valor é substituído pela média dos q anteriores. A janela foi escolhida por conta da sazonalidade de 7 dias para as médias diárias.

Figura 2.5 - Cargas médias diárias para as zonas A a G do NYISO, anos de 2015 a 2018.



Verificamos que um padrão sazonal similar ocorre em todas as zonas de carga, com valores mais altos próximo dos meses de outono e inverno e mais baixos próximos dos meses de primavera e verão ⁷. Acreditamos que a diferença se dê por conta do uso de aquecedores de ar e água nos meses em que ocorrem temperaturas nos extremos mais baixos. Reforça esta suposição o fato de que, segundo a última pesquisa do Governo dos EUA sobre o uso de energia, 73% do uso da energia dos lares do estado de Nova Iorque corresponde aos aquecedores de água e de ar [66]. Recordamos, do Capítulo 1, que aparelhos aquecedores são menos eficientes do ponto de vista energético do que resfriadores de ar, por isso tamanha diferença entre meses mais quentes e mais frios.

Outra informação importante que a Figura 2.5 nos traz é a diferença no nível médio das cargas entre as zonas, em especial para a zona D - NORTH. Em busca de uma razão para isto, procuramos de imediato fazer um levantamento do número de habitantes de cada zona, imaginando que D - NORTH pudesse ter menos habitantes. Como não há uma lista com as cidades dentro de cada zona de carga, nós a construímos a partir da localização das cidades e dos mapas das zonas. Depois, fizemos o levantamento da população de cada cidade [67] e somamos para obter o número de habitantes por zona de carga. Mostramos os resultados na Tabela 2.3, onde também exibimos medidas descritivas para os dados de carga.

⁷No hemisfério Norte, geralmente, o inverno começa nos dias 21 ou 22 de Dezembro (solstício de inverno), a primavera em 20 ou 21 de Março (equinócio de primavera), o verão em 20 ou 21 de Junho (solstício de verão) e o outono em 22 ou 23 de Setembro (equinócio de outono) [65]

Tabela 2.3 – População e medidas descritivas por zona de carga.

Zona	População (milhões de hab.)	Carga (GW)				
		Mínima	Máxima	Mediana	Média	Desvio padrão
A - WEST	1,79	1,21	2,80	1,78	1,79	0,25
B - GENESE	0,86	0,70	2,07	1,12	1,12	0,21
C - CENTRL	1,57	1,22	3,02	1,84	1,54	0,29
D - NORTH	0,15	0,27	0,74	0,50	0,33	0,06
E - MHKVL	1,00	0,49	1,42	0,90	0,89	0,16
F - CAPITL	1,23	0,84	2,42	1,38	1,31	0,25
G - HUDVL	1,87	0,68	2,25	1,10	1,31	0,25

As zonas A e C, que têm médias das cargas próximas, são a segunda e a terceira áreas mais populosas. A próxima zona com maior média das cargas é a quarta mais populosa, F. As zonas seguintes em população, E e B, também têm cargas próximas. Chama atenção que a zona D, como imaginávamos, é a menos populosa, com um número de habitantes consideravelmente menor que o das demais zonas. Isto reforça a suspeita de que a carga mais baixa, consequência de menor demanda, ocorra devido à menor população.

A zona G parece ser a única exceção a este padrão. Boa parte da zona localiza-se no Vale do Hudson-Mohawk, onde é sabido que as mudanças climáticas têm aumentado as temperaturas de maneira mais rápida que as médias locais e a global [68]. Como as temperaturas mais altas demandam menos uso de aquecedores de ar, é possível que o aumento nas temperaturas sejam uma das razões para este comportamento da carga.

Em adição à sazonalidade anual, os dados de carga também apresentam outras variações periódicas. Por conveniência e concisão, a partir deste ponto, os gráficos deste capítulo mostrarão sempre dados da zona A. Mais gráficos e estatísticas descritivas, mostramos no Apêndice B.

Na Figura 2.6, mostramos as cargas de duas semanas típicas de inverno e outras duas de verão (03/08 a 16/08 de 2015 para o verão; 02/02/2015 a 15/02/2015 para o inverno; amostradas por conveniência). As cargas são mais altas nos dias de semana e mais baixas nos finais de semana, o que possivelmente se deve ao cessar de boa parte da atividade comercial aos sábados e domingos. Também é perceptível que há um fator sazonal de periodicidade diária: as cargas sempre são mais baixas nas madrugadas. O nível médio é mais alto no inverno devido à sazonalidade anual, como já discutimos.

Para entender a relação entre os dados de carga e temperatura, analisamos o diagrama de dispersão entre cargas médias e temperaturas médias diárias (escolhemos a estação BUF, que está dentro da zona de carga), exposto na Figura 2.7.

Figura 2.6 – Cargas de uma semana de inverno e outra de verão, amostradas por conveniência (zona de carga A - WEST).

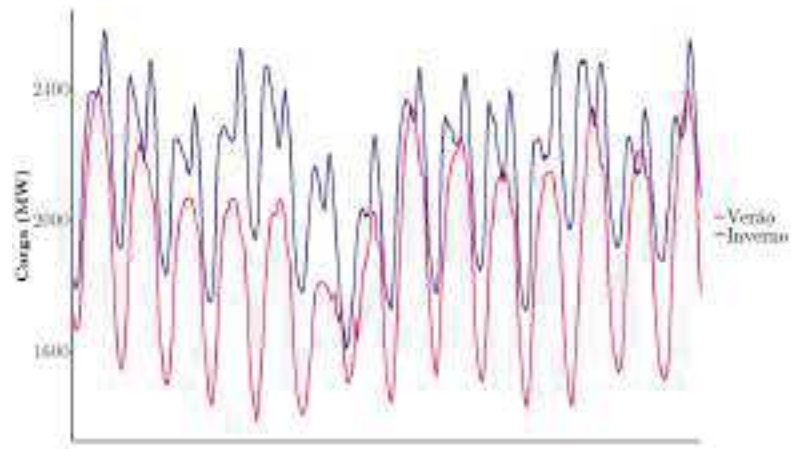
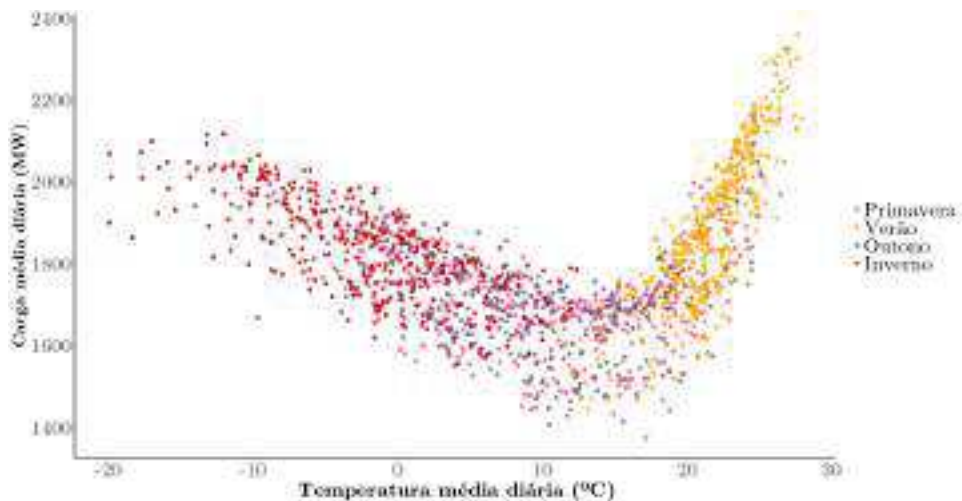


Figura 2.7 – Cargas *versus* temperaturas, médias diárias, para estação meteorológica BUF e zona de carga A - WEST).

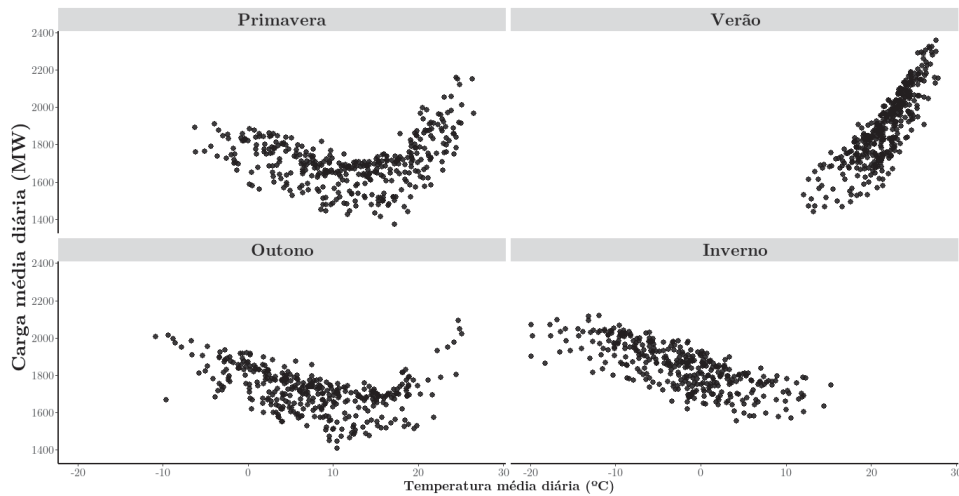


O gráfico indica relação não-linear entre as variáveis, com cargas mais altas à medida que as temperaturas se aproximam do inverno (pontos em vermelho) e do verão (pontos em amarelo). As observações de outono e primavera (pontos respectivamente em lilás e rosa) encontram-se em uma espécie de “zona intermediária”, onde há temperaturas e cargas próximas das duas outras estações.

Para tentar entender ainda mais as relações entre carga e temperatura em cada estação do ano, construímos os diagramas em separado, na Figura 2.8.

No verão, existe forte correlação linear positiva, com coeficiente de correlação linear de Pearson para a amostra igual a 0,83. As cargas são mais altas conforme aumentam as temperaturas. Como não são observadas temperaturas abaixo dos 7,25 °C, pouco se nota o efeito do uso de aquecedores de ar.

Figura 2.8 – Cargas *versus* temperaturas, médias diárias, para estação meteorológica BUF e zona de carga A - WEST); diagramas separados por estação do ano



Já no inverno, o gráfico implica correlação não-linear. De modo geral, as cargas médias diárias mais altas são associadas às temperaturas médias diárias mais baixas. No entanto, existem duas zonas, próximas dos intervalos $[-20\text{ °C}, -10\text{ °C}]$ e $[5\text{ °C}, 15\text{ °C}]$ onde a carga parece pouco afetada pela temperatura. Acreditamos que estas faixas sejam como zonas de conforto para o consumidor, e que mudanças de temperatura dentro dessas faixas sejam menos perceptíveis, a ponto alterar pouco o consumo de aquecedores ou resfriadores.

Nos meses de primavera e outono, o comportamento parece um misto entre o que ocorre no verão e no inverno. Apesar de as temperaturas média e mediana (Tabela 2.4) serem um pouco mais altas do que no inverno, ainda nos concentramos em uma faixa térmica onde se sente o efeito do uso dos aquecedores de ar. Diferentemente do inverno, no entanto, também observamos temperaturas além dessa faixa, fazendo com que algo do efeito do uso do ar condicionado começa a aparecer - e, logo, um formato linear como o do verão também seja parte dos diagramas. Isto é mais pronunciado na primavera, que está mais próxima do verão, do que no outono, mais próximo do inverno.

Tabela 2.4 – Medidas descritivas, em °C, para temperaturas da estação meteorológica BUF

	Primavera	Verão	Outono	Inverno
Média	12,07	21,6	7,79	-2,22
Mediana	12,25	21,75	7,65	-1,75
Mínimo	-8,95	7,25	-12,85	-24,45
Máximo	32,25	33,35	32,25	21,16
Amplitude	41,2	26,1	45,1	45,61
Desvio padrão	8,35	4,42	7,63	7,23

Em razão das diferenças observadas, consideramos importante levar em conta, nos modelos de previsão, a existência de quatro estações do ano.

2.4 RESUMO DO CAPÍTULO

Neste capítulo, apresentamos o conjunto de dados utilizado, que consiste em observações horárias de quatro anos de cargas elétricas de sete zonas de carga do estado de Nova Iorque, e de temperaturas observadas em 16 estações meteorológicas da região.

Falamos sobre o processo de coleta e também sobre o pré-processamento dos dados. Para as cargas elétricas, fizemos somente ajustes para desconsiderar as mudanças de relógio por conta do horário de verão. Já para as temperaturas, foi preciso imputar dados onde havia valores faltantes. No caso de algumas estações meteorológicas, havia muitos valores faltantes em sequência, o que nos fez optar por não trabalhar com estas estações.

Concluído o pré-processamento, fizemos a análise exploratória, através da qual verificamos que:

- Os dados de carga apresentam sazonalidades anual, semanal e diária. Geralmente os valores mais altos ocorrem nos meses mais frios do ano. Dentro da semana, as cargas mais altas ocorrem de segunda e sexta e, dentro de um dia, fora das madrugadas.
- O padrão sazonal é similar para todas as zonas de carga, mas o nível médio varia. A zona D - NORTH, que é consideravelmente menos populosa que as demais, apresenta um nível muito menor que das demais.
- A carga apresenta correlação com a temperatura. Temperaturas extremas, tanto para baixo quanto para cima, estão associadas a cargas mais altas.
- A forma como a carga se correlaciona com a temperatura é diferente para cada uma das estações do ano.

Os itens acima servirão de base para a modelagem e escolha dos métodos de previsão, que serão discutidos nos próximos capítulos.

3 INTRODUÇÃO À MODELAGEM E PREVISÃO DE CARGA ELÉTRICA

A literatura em modelagem e previsão de carga elétrica atravessa décadas de pesquisa, discussões e experiências na Indústria e na Academia.

Faremos uma revisão de como estes métodos foram desenvolvidos e aplicados na indústria ao longo dos anos e apresentaremos os métodos mais utilizados hoje.

Não pretendemos, neste capítulo, entrar nos detalhes de cada método, mas fazer uma reflexão sobre por que optar por determinadas ferramentas no contexto da previsão de carga.

Daremos enfoque de maior profundidade aos métodos nos capítulos 4, 5, e 6.

3.1 MODELAGEM DA CARGA ELÉTRICA POR DECOMPOSIÇÃO DA CURVA DE CARGA

Na Indústria, na década de 1980, a abordagem mais utilizada para modelar a carga elétrica era a **decomposição da curva de carga** [18]. Neste método, a curva da carga elétrica ao longo do tempo, a qual denotamos por Y , é decomposta em três partes: base (B), dependente do clima (W) e aleatória (ε).

O raciocínio por trás da decomposição da curva de carga é o seguinte: existe uma carga base no entorno da qual ocorre uma flutuação, para baixo ou para cima, devido a fatores relacionados ao clima e a fatores aleatórios.

Para um modelo aditivo, a carga no instante t é descrita pela seguinte equação:

$$Y_t = B_t + W_t + \varepsilon_t \quad (3.1)$$

B_t é o valor da carga base no instante t ¹. Esta componente corresponde à parte da carga influenciada pelo seu comportamento histórico, levando em conta a tendência e a sazonalidade, por exemplo.

W_t corresponde à influência das variáveis dependentes do clima sobre a carga no instante t . Temperatura ambiente, umidade relativa do ar, velocidade do vento, nebulosidade e outras características do clima local podem ser consideradas.

ε_t (também chamado de *erro de previsão*) é uma variável aleatória que representa os efeitos sobre a carga causados por fatores que não podem ser capturados pelo modelo, devido à baixa previsibilidade. Por mais que conheçamos os fatores que influenciam o comportamento da carga elétrica, acontecimentos que não foram antevistos sempre afetam qualquer modelo de previsão.

Em muitos casos, os fatores excepcionais de que falamos têm impacto passageiro na demanda de energia. Uma repentina greve de uma classe industrial que paralise a

¹A periodicidade de t pode ser em horas do dia, dias da semana ou outra, a depender da aplicação

atividades das fábricas por um dia pode aumentar o erro de previsão, já que a demanda será mais baixa, mas o prejuízo pode ser minorado com a venda do excedente para outras localidades interligadas à *grid*, por exemplo.

Em outros casos, o problema pode ser mais complexo. Pensamos na pandemia da COVID-19, decretada pela Organização Mundial da Saúde no dia 03 de Março de 2020. A pandemia tem causado impacto significativo na saúde pública, com 184.572.371 casos confirmados e 3.997.640 por COVID-19 até o dia 08 de Julho de 2021 [70]. Somente no Brasil, até a referida data foram contabilizados 18.855.015 casos e 526.892 mortes [71].

Para conter o vírus, países lançaram mão de políticas que vão desde recomendações para ficar em casa até *lockdown* forçado. Atividades econômicas e sociais foram afetadas e os hábitos diários das pessoas mudaram por completo, o que impactou diretamente a demanda de energia elétrica.

Pesquisadores confrontaram a demanda de energia entre os anos de 2017 e 2020 com o período desde a decretação da pandemia para vários países, como China, Índia, Itália, Alemanha e Estados Unidos [72]. Dados mostram que o fato de mais pessoas trabalharem / ficarem em casa provocou uma queda geral na demanda, e a mudança em horários de pico, devido a mudanças de hábito dos consumidores como postergação do horário de acordar e maior uso de serviços de *streaming* à noite. Para se ter a dimensão do impacto no mercado de energia, na cidade de Nova Iorque o teletrabalho fez com que grandes prédios comerciais não precisassem mais ligar aparelhos climatizadores de ar e nem elevadores, o que reduziu a demanda média das 8 da manhã em 18% [72].

Especialmente no início da pandemia, quando medidas restritivas eram alteradas em base diária, com diferenças em nível nacional e estadual (devido a peculiaridades regionais), os fatores não antevistos tomaram uma dimensão desafiadora, de difícil controle e que exige acompanhamento e revisão dos modelos de previsão quase em tempo real.

Independentemente do caso, com um modelo de previsão devemos tentar minimizar a parcela correspondente ao erro, que aqui denotamos por ε , buscando prever as outras parcelas da forma mais acurada possível. A importância de uma boa previsão, no caso da carga elétrica, deve-se ao que discutimos no primeiro capítulo: nos mercados de energia, os custos dos erros são altos.

3.2 MODELOS ESTATÍSTICOS E MATEMÁTICOS PARA PREVISÃO DE CARGA ELÉTRICA

A forma tradicionalmente utilizada para descrever as componentes B_t e W_t na Indústria, nos anos 1980, foi a utilização de modelos estatísticos e matemáticos.

Para a carga base B_t , podemos recorrer a técnicas de séries temporais, como médias móveis, amortecimento exponencial e modelos auto-regressivos de Box & Jenkins. Estas

técnicas fundamentam-se na correlação serial dos dados para modelar padrões históricos, como tendência e sazonalidade [69]. Também podem ser usadas funções periódicas, como senóides.

A componente dependente do clima, W_t , descreve a relação entre carga e variáveis como a temperatura ambiente. Esta relação costuma ser não-linear e, por isso, uma saída pode ser usar polinômios de grau 2 ou maior, por exemplo.

Para comentar aspectos que emergem da prática da decomposição da curva de carga com modelos estatísticos e matemáticos, vamos recorrer a um exemplo.

Vamos supor que queiramos prever a carga de uma região às 21 horas do dia de amanhã, a partir de dados históricos de carga horária até as 21 horas de hoje. Vamos chamar o instante atual (*i.e.*, 21 horas de hoje) de instante t . Vamos supor que os dados indiquem a existência de sazonalidade diária, com cargas de mesma hora do dia apresentando comportamento similar para todos os dias da semana. Partindo destes pressupostos, poderíamos descrever a carga base B_t como um polinômio auto-regressivo que combina linearmente as cargas observadas às 21 horas nos últimos sete dias, com cada dia recebendo um peso diferente α_k , onde k é um número inteiro que serve de índice para o dia.

O polinômio para B_t ficaria como na equação a seguir:

$$B_t = \sum_{k=1}^7 \alpha_k Y_{(t-k \times 24)} \quad (3.2)$$

Já a componente dependente do clima, poderíamos escrever como função da temperatura ambiente² T_t . Vamos supor que a inspeção dos dados tenha sugerido que a relação entre carga e temperatura seja não-linear e optemos por utilizar um polinômio de grau 2.

W_t ficaria, então, expressa como na Equação (3.3):

$$W_t = \beta_0 + \beta_1 T_t + \beta_2 T_t^2 \quad (3.3)$$

β_0 (o intercepto), β_1 e β_2 são os coeficientes do modelo (de regressão).

A modelo completo para a decomposição da curva de carga seria, então, o seguinte:

$$Y_t = \underbrace{\sum_{k=1}^7 \alpha_k Y_{t-k \times 24}}_{\text{componente base}} + \underbrace{\beta_0 + \beta_1 T_t + \beta_2 T_t^2}_{\text{componente meteorológica}} + \underbrace{\varepsilon_t}_{\text{erro}} \quad (3.4)$$

Para fazer a previsão da carga às 21 horas do dia de amanhã, partindo das 21 horas do

²Como, no momento de fazer a previsão, não dispomos da temperatura observada no instante futuro, o usual é trabalhar com uma previsão desta temperatura, que pode ser obtida de fontes externas ou elaborada em separado.

dia de hoje, $\hat{Y}_{t+24|t}$, usaríamos a Equação (3.5):

$$\begin{aligned}\hat{Y}_{t+24|t} = & \alpha_1 Y_t + \alpha_2 Y_{t-24} + \alpha_3 Y_{t-48} + \alpha_4 Y_{t-72} \\ & + \alpha_5 Y_{t-96} + \alpha_6 Y_{t-120} + \alpha_7 Y_{t-144} \\ & + \beta_0 + \beta_1 T_{t+24} + \beta_2 T_{t+24}^2\end{aligned}\tag{3.5}$$

Quando conhecermos a carga observada às 21 horas de amanhã, poderemos calcular o erro de previsão, que seria dado pela diferença entre os valores previstos e observados, ou seja, $\varepsilon_{t+24} = \hat{Y}_{t+24|t} - Y_{t+24}$.

O entendimento da Equação (3.5) é: a carga das 21 horas de amanhã será prevista com base na sazonalidade diária, expressa pela combinação das cargas no mesmo horário dos sete dias anteriores, e na influência da temperatura, como um polinômio cujos regressores são as temperaturas previstas para o horário. Necessitamos de um histórico de sete observações de carga e uma previsão de temperatura para fazer a previsão da carga.

Este exemplo mostra a essência da decomposição da curva de carga: dividimos a carga em componentes menores, para as quais estabelecemos modelos que nos pareçam pertinentes para os dados. Estabelecer formas funcionais explicitamente faz com que o modelo seja de interpretação simples. Além disso, é mais fácil reajustar partes de um modelo quando sabemos qual a sua forma funcional (principalmente quando esta forma é relativamente simples).

Porém, uma limitação desta abordagem costuma aparecer logo que analisamos o comportamento dos dados: as formas funcionais das relações entre a carga e as cargas passadas, particularmente aquela entre carga e variáveis dependentes do clima, não costumam ser tão triviais quanto um polinômio de segundo grau.

Isto pode se complicar ainda mais. Vimos no Capítulo 2 que os dados do NYISO trazem relações diferentes entre carga e temperatura, a depender do corte na base de dados: para cada estação do ano, a forma como a temperatura afeta a carga muda drasticamente. Poderia ser necessário, então, estabelecer formas funcionais diferentes para cada estação do ano.

Vamos analisar, da literatura, um exemplo prático do quão difícil pode ser escolher as formas funcionais para a carga elétrica.

O trabalho [73] vem do contexto de uma competição onde equipes de pesquisadores foram convidadas a desenvolver modelos de previsão para dados da concessionária de energia de Puget Sound, em Washington, EUA (os mesmos dados a que nos referimos em alguns exemplos do Capítulo 1). A empresa fazia, às 8:00, previsões horárias, começando à meia-noite do dia atual até a meia noite do outro dia. Às sextas-feiras, os engenheiros também faziam previsões para todo o final de semana e para a segunda-feira. Antes dos feriados, a previsão era feita para até o próximo dia útil. No momento da previsão, o profissional responsável dispunha de previsões da temperatura para o próximo dia, além

dos históricos de carga e temperatura.

Os autores optaram por usar 24 modelos de decomposição da curva de carga, um para cada hora do dia. A motivação para isso foi de que tentar estimar 24 modelos mais simples seria preferível a estimar um que fosse muito complicado³.

O modelo geral para todas as horas é o seguinte:

$$Y_t = B_t + W_t + D_t + \varepsilon_t \quad (3.6)$$

Na Equação (3.6), B é a carga base, W é a porção dependente do clima, D é uma componente que engloba variáveis determinísticas, como dia da semana, ou mês do ano (modeladas como variáveis indicadoras binárias) e ε_t é o erro.

Várias abordagens, como *splines* variantes no tempo, modelos não-paramétricos e regressão dinâmica, foram testadas para as componentes da curva de carga. Ao final, o melhor modelo foi uma regressão linear múltipla com seguintes variáveis explicativas:

- Carga às 8:00 a.m. do dia anterior;
- Tendência;
- 7 variáveis indicadoras binárias para dia da semana;
- 12 variáveis indicadoras binárias para mês do ano;
- Variável indicadora para dia depois de feriado;
- Temperatura prevista;
- Temperatura prevista, ao quadrado;
- Temperatura máxima prevista para o dia;
- Temperatura máxima prevista para o dia, ao quadrado;
- Temperatura máxima observada no último dia;
- Temperatura máxima observada no dia anterior, ao quadrado;
- Interação da segunda-feira com a carga às 8:00 a.m. do dia anterior;
- Interação de dia depois do feriado com a carga às 8:00 a.m. do dia anterior;

Foi estimado um modelo para cada hora do dia, usando o método dos mínimos quadrados ordinários (MQO), acrescido de uma estrutura autoregressiva de Cochrane-Orcutt para os erros de quinta ordem ⁴.

³Os 24 modelos horários finais, somados, têm 393 coeficientes!

⁴Nesse caso, os modelos são estimados pelos MQO e, depois, introduzidos resíduos defasados até quinta ordem à equação, para então estimar novamente o modelo pelos MQO [74].

Os autores justificam que, apesar de os modelos terem 31 coeficientes, eles não são, de forma alguma, superparametrizados, já que diversas variáveis são ortogonais entre si.

Cada um dos 24 modelos foi investigado em detalhes. Mesmo que, de início, parta-se das mesmas variáveis, os modelos finais para cada hora são diferentes, porque cada parte da curva de carga tem comportamento diferente e é afetada à sua maneira pelas variáveis exógenas. Por exemplo, em certas horas do dia, como na madrugada, a temperatura não tem tanta influência sobre a carga.

O seguinte trecho, retirado das conclusões dos autores, ilustra a complexidade advinda do quão específico pode ser o comportamento da carga a depender da hora do dia (e, logo, a vantagem em se usar 24 modelos):

É particularmente interessante examinar o efeito de uma variável através das horas. Por exemplo, o coeficiente da temperatura para a HORA1 É -2,86 enquanto para HORA8 é -25,03, ou quase dez vezes maior. É esperado que o efeito da carga de aquecimento seja muito mais importante às 8 da manhã do que à meia-noite. Presumivelmente, vários consumidores possuem termostatos atrasados que são relativamente indiferentes a temperaturas externas até o alvorecer. A impressão geral a partir do exame dos coeficientes é que eles variam através das horas de maneira um tanto sistemática. Esta observação confirma nossa conclusão precoce de que construir um modelo separado para cada hora era bastante superior a supor que o mesmo modelo poderia ser usado para todas as horas. ([73], p.165)

Este estudo de caso mostra que, apesar das vantagens da decomposição da curva de carga e da escolha prévia de formas funcionais para a carga, esta estratégia pode tornar-se bastante complicada quando os padrões encontrados nos dados são complexos. Além de ser difícil definir as formas funcionais (e isto demandar tempo), todo o procedimento é pouco automatizado; o que é uma desvantagem para as atividades de curto prazo da CSSE, como o despacho de carga.

Haveria procedimentos que levassem a modelos mais específicos para os dados e com maior possibilidade de automação?

3.3 RECONHECIMENTO DE PADRÕES A PARTIR DE TÉCNICAS DE APRENDIZADO DE MÁQUINA

O **Aprendizado de Máquina** (do inglês *machine learning*) é uma sub-área da Ciência da Computação que desenvolve e utiliza algoritmos capazes de “aprender”⁵, por meio de uma base de dados, o padrão de comportamento de algum fenômeno [75]. Um dos interesses do aprendizado de máquina é usar o padrão aprendido para fazer previsões.

⁵Não pensar, aqui, que os algoritmos sejam autoconscientes. Longe disto; na verdade, são pouco capazes de operar fora de contextos razoavelmente específicos.

Começemos tratando das **redes neurais**, um dos métodos mais utilizados para previsão de carga elétrica. Redes neurais conseguem mapear as mais específicas relações entre dados sem que seja necessário escolher uma forma funcional explícita. Basta construir a arquitetura da rede, definindo as entradas (as variáveis que influenciam a carga), as saídas (as cargas que queremos prever) e escolhendo alguns hiperparâmetros, e algoritmos de otimização se encarregarão de construir o mapa funcional entre os dados de entrada e saída de maneira que o usuário não precise se envolver diretamente; uma ideia que parece perfeita para a automação de sistemas.

Algumas dessas técnicas já existiam na literatura na década de 1980, ainda que em formas mais primitivas⁶. O problema é que, naquela época, o custo computacional tornava a implementação inviável. Alguns dos autores citados na compilação de artigos [18] enfatizam até mesmo a necessidade de se apagarem os dados mais antigos do histórico para economizar espaço na memória do computador (isso para métodos de séries temporais, que têm custo mais baixo).

Com a evolução da capacidade de processamento e armazenamento dos computadores pessoais, surgiram (e proliferaram-se), na década de 1990, casos de sucesso com o uso de redes neurais para previsão de carga.

Vamos comentar o primeiro trabalho de maior repercussão e, até hoje, um dos mais impressionantes devido ao alcance comercial do sistema desenvolvido [77] .

Liderada pelo professor A. Khotanzad⁷, uma equipe desenvolveu um *software* de previsão de carga para a indústria de energia elétrica, o ANNSTLF (*Artificial Neural Network Short-Term Load Forecaster*).

No *back-end* do programa funcionam redes neurais do tipo *feedforward* (FFNN, do inglês *feedforward neural network*⁸). O sistema foi adquirido e utilizado com sucesso por 32 empresas de energia dos EUA e do Canadá, consistindo na primeira solução escalável para previsão de carga de que se tem notícia. No artigo, os autores mostram resultados obtidos na operação de dez dessas empresas.

Em sua primeira versão, o programa opera 38 FFNNs agrupadas em 3 módulos, para modelar separadamente padrões semanais, diários e horários. As saídas destes módulos são combinadas para chegar à previsão final.

Os módulos semanal e diário consistem em 7 FFNNs cada, sendo uma rede por dia da semana. A diferença entre estes módulos é que o primeiro tem por objetivo modelar

⁶Para um abrangente histórico sobre redes neurais até o ano de 1990, consultar [76]

⁷O professor emérito da Southern Methodist University Alizera Khotanzad faleceu em 2019, aos 62 anos [78]. Sua contribuição para a área de previsão de carga elétrica será, decerto, lembrada como inestimável.

⁸Na verdade, os autores usam a nomenclatura *Multilayer Perceptron* (MLP), ou, traduzindo, *Perceptron* Multicamadas. Esta nomenclatura foi bastante usada até anos 2000. Com o aumento na popularidade das redes recorrentes (RNNs), boa parte dos autores parece preferir, hoje, a nomenclatura FFNN, para diferenciar estas redes daquelas onde há ciclos no fluxo.

o comportamento da carga para cada dia da semana, enquanto o segundo serve para entender como a carga de um dia se relaciona com a do dia anterior.

As entradas para o módulo semanal são:

- 24 cargas horárias do mesmo dia previsto, na última semana;
- 24 temperaturas horárias no mesmo dia que o previsto, na semana anterior;
- 24 previsões de temperatura para o dia que previsto.

As saídas são as 24 cargas horárias para o dia previsto.

Para o módulo diário, as entradas são:

- 24 cargas horárias do dia anterior ao previsto;
- 24 temperaturas horárias do dia anterior ao previsto;
- 24 previsões de temperatura para o dia que previsto.

As saídas do módulos diário são as mesmas do módulo semanal: as 24 cargas horárias para o dia previsto.

Finalmente, o módulo horário tem o objetivo de capturar os perfis da carga para cada hora do dia em separado, consistindo, portanto, em 24 FFNNs.

Cada rede possui as seguintes entradas:

- Carga observada ontem na mesma hora;
- Temperatura observada ontem na mesma hora;
- Umidade relativa do ar ontem na mesma hora (opcional);
- Temperatura dois dias atrás na mesma hora;
- Umidade relativa do ar dois dias atrás na mesma hora (opcional);
- Previsão da temperatura;
- Previsão da umidade relativa do ar (opcional);
- Índice do dia da semana (0,1 para o domingo; 0,2 para a segunda-feira, *etc.*).

A saída é a previsão de carga para a hora desejada.

A previsão final é dada por uma combinação linear das saídas dos três módulos.

O ANNSTLF traz também uma função para previsão de temperatura e umidade.

A solução foi comercializada como uma aplicação em interface gráfica para PC compatível com IBM, para sistemas operacionais MS-DOS, MS-Windows (3.1, 95 ou NT). Os requisitos mínimos para operação do programa eram dois a três anos de dados históricos de carga e temperatura para treinar o pacote, e cargas horárias do dia anterior e previsões do tempo para dias futuros para uso *on-line* da ferramenta.

Os desenvolvedores relatam que acompanharam os clientes de 1992 até 1997, quando o artigo foi publicado, tendo recebido notícia de vários resultados positivos, com previsões satisfatoriamente acuradas [77]. Bastava fornecer o histórico de dados e o programa se encarregava de tudo.

Por terem conseguido transformar em solução comercial uma das primeiras implementações noticadas de redes neurais para previsão de carga, os desenvolvedores⁹ produziram, sem dúvida, uma contribuição muito importante para o campo. Mas vamos comentar alguns desafios para usar o ANNSTLF no dia a dia.

Primeiramente, o ANNSTLF é uma solução comercial, com código-fonte proprietário. Como qualquer produto desse tipo, o usuário depende de um serviço de atendimento ao cliente que o dê suporte sempre que houver algum problema na operação. No caso da previsão de carga em curto prazo, que lida com oferta e demanda que ocorrem de forma simultânea, seria necessário que a disponibilidade para resolução de problemas funcionasse de forma *on-line* sem interrupções (ou que o sistema fosse, idealmente, à prova de falhas).

Quando falamos em resolução de problemas, não nos referimos somente a *bugs* no código-fonte. Para explicar, citamos um trecho de comentário dos desenvolvedores:

Uma das vantagens do ANNSTLF é sua portabilidade. A estrutura de entrada-saída dos *engines* de previsão permanece a mesma para uma empresa. A única customização que deve ser feita é **encontrar o número apropriado de nós ocultos para os MLPs** do ANNSTLF. ([77], p.845)

O trecho destacado nos leva a crer que ou o usuário deveria ter conhecimento prévio em redes neurais para operar o sistema ou o comprador do ANNSTLF ficaria obrigado a contratar o suporte.

Vamos além do caso específico do ANNSTLF para falar sobre o que nos interessa mais: o uso de um sistema que constrói modelos e faz previsões a partir do reconhecimento de padrões no dados, em comparação com o uso de um sistema cujo modelo foi previamente definido.

O principal atrativo das técnicas de aprendizado de máquina é que, a princípio,

⁹A empresa por trás do ANNSTLF, que hoje atende pelo nome de Enverus Trading & Risk, continua produzindo soluções para o mercado de energia e tem em seu portfólio companhias como Shell, Hydro Québec e Luminant [79].

basta dispor dos dados e informar as entradas e saídas¹⁰. Mas um processamento acontece entre as entradas e as saídas, o que envolve composição de (possivelmente muitas) funções não-lineares e otimização não-linear e seus detalhes (como número de iterações, taxas de aprendizado, gradientes, *etc.*).

Toda essa complexidade faz com que seja difícil utilizar um sistema desses no dia a dia sem que seja como uma caixa-preta. Isto despe as soluções de duas grandes qualidades dos modelos estatísticos e matemáticos que são pré-definidos: interpretabilidade e autonomia do usuário.

É relativamente simples examinar as relações entre as entradas e saídas em um modelo de regressão pré-definido, por exemplo. Para o modelo de regressão de [73], vimos que os coeficientes estimados para o modelo permitem tirar conclusões de forma direta, como aquela de que às 8 a.m. a carga é mais dependente da temperatura do que à 1 a.m..

Do comprometimento da interpretabilidade, decorre a menor comunicabilidade: passa a ser necessário um esforço para convencer os *players* da CSSE de que um modelo de previsão caixa-preta é uma boa ideia.

Mas não só a interpretabilidade fica comprometida, como também a própria autonomia do usuário sobre o modelo. No caso de uma rede neural, as relações construídas pela composição de funções são tão complexas e têm tantos pormenores que torna-se desafiador descobrir quais pontos da rede trazem informações e quais não.

Por último, mas não menos importante, um ponto crucial relativo ao uso não só das redes neurais, como de qualquer outro método de aprendizado de máquina, é mencionado por uma revisão de literatura de grande repercussão [19]. Esta revisão traz a análise de dezenas de artigos sobre redes neurais para previsão de carga elétrica, nas principais publicações científicas da área, até o ano 2000. Os autores alertam que, apesar de os artigos sempre apresentarem bons resultados no que diz respeito à acurácia, há muito pouco cuidado na validação dos resultados. A maioria das aplicações são feitas com poucas bases de dados¹¹. Ora, se as redes neurais são capazes de capturar os padrões mais específicos dos dados que fornecemos, isto não comprometeria a sua capacidade de generalização para novos dados?

Apesar destes problemas, os métodos de aprendizado de máquina possuem grandes vantagens, como a possibilidade de alcançar maior acurácia a partir da identificação de relações complexas (que dificilmente conseguiríamos mapear por conta própria) entre a carga e as variáveis que a impactam, além da maior facilidade de uso para automação de sistemas. Há, porém, que se ter especial atenção para a validação dos resultados, e adotar estratégias para garantir que os sistemas não se especializem demais nos dados já vistos e

¹⁰Sim, também é necessário ajustar hiperparâmetros, mas até mesmo isso pode ser terceirizado para uma biblioteca em Python ou R nos dias de hoje.

¹¹Em vários casos, com somente uma.

tenham boa capacidade de generalização para dados novos.

3.4 Modelagem e previsão de carga elétrica atualmente

Até o momento falamos de duas grandes abordagens para previsão de carga elétrica: o uso de modelos estatísticos e matemáticos pré-definidos e os métodos de aprendizado de máquina. A primeira foi o padrão da indústria na década de 1980. A segunda popularizou-se a partir da década de 1990, com a evolução do computador pessoal.

Nos últimos anos, a evolução da tecnologia trouxe ainda mais desafios para a previsão de carga elétrica. As CSSE têm feito, gradualmente, a substituição dos aparatos antigos por *smart grids*, abrindo para a necessidade da previsão hierárquica de carga, por exemplo. Isto requer processamento de ainda mais informações e a elaboração de previsões de carga para em níveis diversos da *grid*.

Felizmente, passamos da época em que o computador pessoal tinha limitações severas na capacidade de processamento e armazenamento e também já experimentamos durante anos com técnicas de aprendizado de máquina. Com isso, hoje entendemos melhor as vantagens e desvantagens das abordagens e desenvolvemos técnicas para contorná-las.

Hoje é possível armazenar e consultar grandes bancos de dados de maneira remota, e recorrer à computação em nuvem para executar programas utilizando várias unidades de processamento de alto desempenho em paralelo. Também temos a vantagem poder escrever códigos usando poderosas e confiáveis bibliotecas de código aberto, libertando-nos a dependência obrigatória de soluções proprietárias.

Tudo isso nos traz a um período particularmente interessante para a previsão de carga: uma situação em que são muitos os desafios, mas, mais do que nunca, temos as condições de experimentar com uma vasta gama de ferramentas, do amortecimento exponencial ao *Deep Learning*¹², e até mesmo combinar o melhor de cada uma das técnicas que seja mais adequada aos dados de que dispomos.

Não há consenso sobre um método que funcione melhor para todas as situações de previsão de carga elétrica, já que as características dos dados podem variar.

Uma revisão de literatura de 2016 divide os métodos que mais têm sido utilizados ao longo dos últimos anos nas seguintes categorias: regressão linear múltipla, modelos semi-paramétricos, modelos de amortecimento exponencial, modelos de Box & Jenkins, redes neurais artificiais, regressão fuzzy, máquinas de vetor suporte, *gradient boosting* [14].

Devido à característica dos dados de carga e temperatura que apresentamos no Capítulo 2, escolhemos as redes neurais como principal método de previsão para esta tese. Garantindo uma cautelosa validação, acreditamos que o método seja capaz de mapear as

¹²Aqui não temos uma ordem de qualidade das ferramentas, mas a sequência das mais antigas para as mais recentes.

não-linearidades na correlação entre carga e temperatura sem comprometer a generalização.

Como *benchmarks*¹³, experimentamos métodos estatísticos: um método de amortecimento exponencial, modelos de Box & Jenkins e um previsor ingênuo. Estes *benchmarks* podem indicar, para nossos dados, se parece haver diferença entre considerar ou não o efeito de variáveis meteorológicas para as previsões de curto prazo.

Também faremos *benchmark* com um modelo de regressão linear múltipla, sugerido na literatura como *benchmark naive* para previsão de cargas.

¹³Nesta tese, tratamos por *benchmark* a comparação de métodos em um problema, independentemente de quais sejam. Já por *benchmark* ingênuo (ou *naive*), entendemos o método mais simples de que se tem referência (ou seja, se uma proposta for pior que o *benchmark naive*, deverá ser revisada ou descartada).

4 MÉTODOS DE PREVISÃO: SÉRIES TEMPORAIS

Métodos para análise e previsão de séries temporais são aqueles que utilizam técnicas estatísticas para detectar padrões em dados históricos, com o objetivo de extrapolar esses padrões para o futuro.

Os métodos de séries temporais com que optamos por trabalhar não consideram o efeito de variáveis exógenas, como as meteorológicas.

Como a carga elétrica é afetada pela temperatura, poderíamos pensar que usar estes métodos significa negligenciar parte importante do problema. No entanto, há relatos de casos de sucesso de previsão de carga com o uso destes métodos, como [25, 27, 81]. Autores, como [27], atribuem o sucesso à possibilidade de os efeitos das mudanças no clima não serem perceptíveis no curto prazo. Como dificilmente estas mudanças são bruscas, a mudança no comportamento do consumidor levaria mais tempo do que o horizonte para o qual queremos prever a carga (24 horas à frente).

Consideramos importante utilizar métodos de séries temporais como *benchmarks*, porque demandam menos dados, não dependem de fontes secundárias, são computacionalmente mais baratos e sua interpretação é direta.

Escolhemos o **amortecimento exponencial** e os **métodos de Box & Jenkins**, que são citados dentre os métodos mais importantes para a previsão de carga na revisão [14]. Também experimentamos um previsor ingênuo.

No Apêndice A, trazemos uma introdução à notação e aos conceitos que utilizamos ao longo do texto.

4.1 AMORTECIMENTO EXPONENCIAL

Robert Goodell Brown foi um dos nomes mais importantes para a pavimentação do que se tornaria a Gestão da Cadeia de Suprimentos. Quando servia à Marinha Americana durante a Segunda Guerra Mundial, Brown trabalhou no *Operations Evaluation Group*, onde desenvolveu o método de previsão do **Amortecimento Exponencial**, que aplicou a controle de estoques [82]. Algum tempo depois, trabalhando para a empresa de consultoria Arthur D. Little, Brown publicizou o método, primeiro em um congresso e depois em um relatório de consultoria [83], enfatizando sua aplicação na Indústria.

A equação de previsão para o método que hoje é conhecido como **Amortecimento Exponencial de Brown** (ou **Amortecimento Exponencial Simples**) é a seguinte ¹:

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}, \quad (4.1)$$

¹Curiosamente, a versão apresentada em [83] trazia também equações para a tendência e a variância, embora estas normalmente não sejam mencionadas na literatura quando se fala do método proposto por Brown.

No contexto da previsão de carga elétrica, no amortecimento exponencial de Brown a previsão da carga para o instante $t + 1$ feita no instante t , denotada por $\hat{y}_{t+1|t}$, é função da carga observada e da previsão mais recentes, y_{t-1} e $\hat{y}_{t|t-1}$, respectivamente. A observação mais recente recebe peso α (a chamada *constante de amortecimento*), e a última previsão é corrigida pelo fator $(1 - \alpha)$. Impomos a restrição $0 \leq \alpha \leq 1$, e inicializamos o método fazendo $\hat{y}_{1|0} = y_0$ (portanto, é necessária uma observação para inicializar o método).

Alternativamente, podemos expandir a Equação (4.1) e escrever o método como uma média ponderada de todas observações passadas, em que observações mais antigas recebem pesos cada vez menores, que decaem exponencialmente:

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots, \quad (4.2)$$

A escolha da constante de amortecimento é fundamental para a qualidade das previsões. Se α for muito pequeno, o período base será longo e mudanças repentinas na demanda poderão ser perdidas. Em contrapartida, se α for muito grande, o período base será curto, e, mesmo que o método capture mudanças repentinas, o fará com certo atraso.

Na época da publicação do método, utilizar algoritmos de otimização numérica para encontrar α era incipientes ou inviáveis. Como guia, [83] fornecem uma tabela com valores da constante de acordo com a importância que se quer dar aos períodos mais recentes (Figura 4.1).

Figura 4.1 - Tabela com recomendações para escolha da constante de amortecimento, para o amortecimento exponencial de Brown.

TABLE 1

VALUES OF THE SMOOTHING CONSTANT α THAT GIVE 50% AND 30% OF THE TOTAL WEIGHT TO DEMAND IN THE MOST RECENT n PERIODS

Base Number of Periods n	Smoothing Constant	
	50%	30%
2	0.500	0.393
3	0.585	0.217
4	0.633	0.150
6	0.719	0.100
12	0.775	0.056
18	0.790	0.038
24	0.795	0.028
30	0.797	0.018
32	0.797	0.013

Fonte: [83].

Com a capacidade de processamento dos computadores, podemos utilizar métodos numéricos para encontrar um valor α que minimize alguma função de perda, como veremos na próxima subseção.

Depois de Brown, Charles C. Holt propôs uma extensão do amortecimento exponencial para séries com tendência linear [84]. Esta forma do amortecimento exponencial ficaria conhecida como **amortecimento exponencial de Holt**, ou **método de Holt para tendência linear**.

As equações para a versão aditiva do método de Holt são as seguintes:

$$\text{Previsão} \quad \hat{y}_{t+k|t} = \ell_t + kb_t \quad (4.3a)$$

$$\text{Nível} \quad \ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (4.3b)$$

$$\text{Tendência} \quad b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}, \quad (4.3c)$$

O método decompõe a série em nível e tendência. No conjunto de equações 4.3, ℓ_t é o nível no instante t , b_t é a tendência da série no instante t . α , $0 \leq \alpha \leq 1$ é a constante de amortecimento para o nível e β , $0 \leq \beta \leq 1$ é a constante de amortecimento para a tendência. A previsão para k passos à frente será a soma do nível com a tendência multiplicada por k .

Existem várias formas de se inicializar o método. Pode-se, por exemplo, inicializar o nível como uma média dos primeiros valores da série e obter a tendência subtraindo esta média da observação, ou mesmo ajustar uma reta de regressão linear para os primeiros valores [69].

Peter Winters, estudante de Holt, estendeu o método para séries sazonais. O **amortecimento exponencial de Holt-Winters**, em sua versão aditiva, segue o conjunto de equações adiante:

$$\text{Previsão} \quad \hat{y}_{t+k|t} = \ell_t + kb_t + s_{(t-m+k)} \quad (4.4a)$$

$$\text{Nível} \quad \ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (4.4b)$$

$$\text{Tendência} \quad b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \quad (4.4c)$$

$$\text{Sazonalidade} \quad s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}, \quad (4.4d)$$

Em Holt-Winters, acrescenta-se uma equação para o fator sazonal s_t com constante de amortecimento γ . m é o tamanho do período sazonal.

A inicialização para os fatores sazonais pode ser feita calculando a média das observações durante o primeiro período sazonal e subtraindo o nível médio. Isto quer dizer que necessitamos, no mínimo, de uma quantidade de valores do tamanho do período sazonal para inicialização. Por exemplo, se os dados são horários e têm sazonalidade diária, considerando 7 dias da semana necessitamos de 168 observações para inicialização.

Durante muito tempo os métodos de amortecimento exponencial foram difundidos entre profissionais da logística, especialmente para controle de estoque, como na concepção original de Brown.

Décadas depois, James W. Taylor, em aplicação à previsão de carga elétrica em curto prazo, propõe uma nova atualização do amortecimento exponencial, à qual chama de *Amortecimento Exponencial de Holt-Winters-Taylor* [80], que vamos chamar de HWT. Como os dados de carga elétrica têm mais de uma sazonalidade, o método de Holt-Winters é adaptado para acomodar dois fatores sazonais, um diário e outro semanal.

O HWT é descrito pelo seguinte conjunto de equações:

$$\text{Previsão} \quad \hat{y}_{t+k|t} = l_t + d_{t-s_1+k} + w_{t-s_2+k} + \phi^k e_t \quad (4.5a)$$

$$\text{Erro} \quad e_t = (y_t - (l_{t-1} + d_{t-s_1} + w_{t-s_2})) \quad (4.5b)$$

$$\text{Nível} \quad l_t = l_{t-1} + \alpha e_t \quad (4.5c)$$

$$\text{Sazonalidade diária} \quad d_t = d_{t-s_1} + \delta e_t \quad (4.5d)$$

$$\text{Sazonalidade semanal} \quad w_t = w_{t-s_2} + \omega e_t \quad (4.5e)$$

onde l_t , d_t e w_t são, respectivamente, o nível da carga, o fator sazonal diário e o fator sazonal semanal. α, δ, ω são constantes de amortecimento para as referidas componentes, e ϕ , com $0 \leq \phi \leq 1$ é uma constante para a correção da autocorrelação dos erros de primeira ordem.

Interessante notar que o método não traz a equação de tendência que, segundo o autor, não melhorou as previsões quando inserida. De certa forma, isto não surpreende, já que dados de carga elétrica têm nível médio usualmente constante no curto prazo.

4.1.1 Determinação das constantes do Amortecimento Exponencial Utilizando Otimização Numérica

Dentre as infinitas possibilidades de valores para as constantes do amortecimento exponencial, temos interesse no conjunto que leve às previsões mais acuradas.

Uma abordagem possível é procurar pelos valores que minimizam uma função de perda baseada nos erros de previsão, como o **erro médio quadrático** (ou MSE, do inglês *mean squared error*). O MSE é particularmente interessante por ser uma função com características desejáveis para a convergência de métodos de otimização não-linear, como ser uma função diferenciável, e convexa em suas entradas e parâmetros ^{2,3}.

²Para uma prova da convexidade do MSE no contexto de problemas de regressão, ver [85].

³Apesar de o MSE ser uma função convexa em suas entradas e parâmetros, é importante dizer que em certos casos, como em redes neurais, o acréscimo de muitas não-linearidades podem torná-lo uma função não-convexa (o que não é o caso).

O MSE é dado pela Equação (4.6):

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_t - y_t)^2 \quad (4.6)$$

Nessa equação, \hat{y}_t e y_t são, respectivamente, os valores previstos e observados no instante t , e N é o número de previsões para as quais queremos calcular o MSE.

Para encontrar os valores das constantes do HWT que minimizam o MSE, escolhemos um método de otimização *quasi*-Newton, o **L-BFGS**.

Para explicar como funciona o L-BFGS, usaremos alguns conceitos de otimização numérica. Para maior aprofundamento, recomendamos [86].

Um procedimento possível para encontrar um ponto \mathbf{x}^* que minimiza uma função $f(\mathbf{x})$ é partir de um ponto inicial \mathbf{x}_0 e fazer uma busca iterativa na direção contrária do gradiente da função. Como o gradiente corresponde à direção de maior incremento (ou de *subida mais íngreme*), a direção contrária será a de *descida mais íngreme*, ou seja, a direção em que $f(\mathbf{x})$ decresce mais rápido.

Este raciocínio nos leva ao **algoritmo de descida de gradiente**, cuja regra de atualização para \mathbf{x} a cada iteração k é a seguinte:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \eta \nabla f(\mathbf{x}_k) \quad (4.7)$$

Na Equação (4.7), \mathbf{x} é ajustado caminhando proporcionalmente à direção contrária do gradiente. A taxa η é a **taxa de aprendizado**, um número real positivo que dá o tom de quão “agressivo” será o ajuste. A escolha deste parâmetro é importante, já que uma taxa muito alta pode causar a divergência do algoritmo, ao passo que uma taxa muito baixa pode tornar a convergência muito lenta.

Os passos do algoritmo de descida de gradiente estão descritos no pseudocódigo a seguir:

Algoritmo 1 Descida de gradiente

Forneça \mathbf{x}_0 , forneça ϵ , faça $k \leftarrow 0$;

enquanto $\|\nabla f(\mathbf{x}_k)\| \geq \epsilon$ **faça**

$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k - \eta \nabla f(\mathbf{x}_k)$;
 $k \leftarrow k + 1$;

fim

retorna \mathbf{x}_k

Pensando geometricamente, em cada iteração o algoritmo faz uma aproximação linear da função de perda no ponto \mathbf{x}_k e “salta” em direção ao seu mínimo, fazendo outras aproximações lineares, ponto a ponto.

A convergência do algoritmo de descida de gradiente é garantida para uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convexa, diferenciável e de gradiente Lipschitz contínuo com constante L positiva⁴.

O algoritmo é de fácil implementação e computacionalmente barato. No entanto, pode ser inconvenientemente lento para problemas que não sejam fortemente convexos [88].

Outra desvantagem é que, como o método utiliza informação do gradiente (informações de primeira ordem) para encontrar o mínimo, cada iteração trabalha apenas com informação local.

O *Método de Newton*, que, inicialmente, é proposto para resolver sistemas de equações não-lineares, pode ser generalizado como um método de descida mais íngreme, e usado para minimizar funções convexas, com convergência mais rápida que o algoritmo de descida de gradiente [86] e uso de mais informação a cada iteração.

A ideia do método de Newton é buscar soluções \mathbf{x} para as quais $f(\mathbf{x}) = 0$, para alguma função $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Pensando em uma dimensão, o método parte de um ponto arbitrário \mathbf{x}_0 e iterativamente tenta encontrar o ponto em que a função se iguala a zero, usando retas tangentes como guia. Se o palpite inicial \mathbf{x}_0 for suficientemente próximo da raiz de $f(\mathbf{x})$, a reta tangente ao ponto $(\mathbf{x}_0, f(\mathbf{x}_0))$ cortará o eixo \mathbf{x} em um ponto mais próximo da raiz do que \mathbf{x}_0 .

Tratando do caso unidimensional, a reta tangente da função f em um ponto x_k tem a forma $f(x_k) + f'(x_k)(x - x_k)$.

Igualando a expressão a zero, encontramos a raiz desta reta, dada pela Equação (4.8):

$$x = x_k - f(x_k)f'(x_k)^{-1} \quad (4.8)$$

Estendendo para o caso multidimensional, temos:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - f(\mathbf{x}_k)J_f(\mathbf{x}_k)^{-1} \quad (4.9)$$

onde $J_f(\mathbf{x})$ é o **Jacobiano** da função (*i.e.* a matriz de derivadas parciais).

Mas nosso objetivo é minimizar uma função convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$, ou seja, encontrar a solução x para a qual $\nabla f(\mathbf{x}) = 0$. O Jacobiano do gradiente $\nabla f(\mathbf{x}) = 0$ é a matriz **Hessiana** H_f (a matriz de derivadas parciais de segunda ordem). Então, a equação de atualização para o método de Newton generalizado para problemas de minimização de funções convexas fica:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \nabla f(\mathbf{x}_k)H_f(\mathbf{x}_k)^{-1} \quad (4.10)$$

O Algoritmo 2 mostra o pseudocódigo do método completo:

⁴Isto é, $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2$ para quaisquer x, y [88].

Algoritmo 2 Método de Newton para minimização de funções

Forneça \mathbf{x}_0 , forneça ϵ , faça $k \leftarrow 0$;

enquanto $\|\nabla f(\mathbf{x}_k)\| \geq \epsilon$ **faça**

$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k - \nabla f(\mathbf{x}_k) H_f(\mathbf{x}_k)^{-1}$;

$k \leftarrow k + 1$;

fim

retorna \mathbf{x}^k

A informação de segunda ordem trazida pela Hessiana faz com que o algoritmo convirja muito mais rápido que a descida de gradiente. Com um palpite inicial \mathbf{x}_0 suficientemente próximo da vizinhança do ponto ótimo, a convergência é garantida e usualmente mais rápida que no método de descida de gradiente [89]. Porém, as características do problema que se tenta resolver requerem cuidado. Se os autovalores da Hessiana não forem não-negativos, a aproximação quadrática local é um parabolóide com concavidade voltada para baixo. Daí, a iteração pode levar a um máximo local ou a um ponto-sela. Pela mesma razão, o método é bastante sensível às condições iniciais (i.e., à escolha de \mathbf{x}_0). Além disso,

“Métodos que usam a direção de Newton têm uma alta taxa de convergência local, tipicamente quadrática. Depois que uma vizinhança da solução é atingida, a convergência para alta acurácia costuma acontecer em poucas iterações. A principal desvantagem da direção de Newton é a necessidade da Hessiana. O cálculo explícito dessa matriz de derivadas segundas pode às vezes ser pesado, sujeito a erros, e um processo caro.” ([86], p.23)

Pensando no ganho de eficiência trazido pela informação das derivadas de segunda ordem, mas tentando evitar o custo computacional do cálculo da Hessiana, a família de métodos *quasi*-Newton traz alternativas que usam aproximações da inversa da matriz Hessiana no lugar do cálculo explícito, conseguindo preservar a alta taxa de convergência sem um alto custo.

A ideia de um método *quasi*-Newton é aproximar a Hessiana por uma matriz positiva-definida B , que é atualizada a cada iteração. Dessa forma, o uso de informações da iteração passada, em vez de novo cálculo da matriz, reduzirá o custo computacional.

Para que o método funcione, a atualização B_k a cada iteração deve ser feita de maneira que, em iterações sucessivas, os pontos \mathbf{x}_k e \mathbf{x}_{k+1} , e os gradientes $\nabla f(\mathbf{x}_k)$ e $\nabla f(\mathbf{x}_{k+1})$ contendam informação da Hessiana, isto é, que atendam à seguinte condição:

$$\nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k) \approx H_f(\mathbf{x}_{k+1})(\mathbf{x}_{k+1} - \mathbf{x}_k) \quad (4.11)$$

Sejam $\Delta \mathbf{x}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ e $\mathbf{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$, devemos escolher B_{k+1} que satisfaça:

$$B_{k+1} \Delta \mathbf{x}_k = \mathbf{y}_k \quad (4.12)$$

A Equação (4.12) é chamada *condição quasi-Newton*. Os métodos *quasi-Newton* se diferenciam na forma com que fazem a aproximação B para a Hessiana.

Um dos métodos *quasi-Newton* mais populares é o *BFGS*. Faremos uma derivação resumida do BFGS, baseada em [90], para uma ideia geral sobre o método. Para uma derivação mais aprofundada, ver [86].

Para desenvolver uma regra de atualização para B_{k+1} , temos que impor as restrições para a aproximação da Hessiana. Como nosso objetivo é minimizar uma função convexa, é interessante que a Hessiana seja simétrica positiva-definida para garantir convergência do algoritmo. Além disso, queremos que B_{k+1} seja suficientemente próxima de B_k .

Resumindo estas condições, temos que resolver o seguinte problema:

$$\begin{aligned} \min_{B_{k+1}} \quad & \|B_{k+1} - B_k\| \\ \text{sujeito a} \quad & B_{k+1}^T = B_{k+1}; \\ & B_{k+1} \Delta \mathbf{x}_k = \mathbf{y}_k \end{aligned} \quad (4.13)$$

que podemos reescrever:

$$\begin{aligned} \min_{B_{k+1}^{-1}} \quad & \|B_{k+1}^{-1} - B_k^{-1}\| \\ \text{sujeito a} \quad & B_{k+1}^{-1 T} = B_{k+1}^{-1}; \\ & \Delta x_k = B_{k+1}^{-1} \mathbf{y}_k \end{aligned} \quad (4.14)$$

No método BFGS, a norma utilizada é a norma de Frobenius, o que torna o problema equivalente a atualizar a Hessiana somando duas matrizes U e V simétricas de posto igual a um ⁵:

$$B_{k+1} = B_k + U_k + V_k \quad (4.15)$$

Para que B_{k+1} permaneça simétrica e positiva-definida, as matrizes U e V devem ter a forma $U = a\mathbf{u}\mathbf{u}^T$ e $V = b\mathbf{v}\mathbf{v}^T$, onde a e b são constantes e \mathbf{u} e \mathbf{v} são vetores linearmente independentes e não-zero.

Então, temos:

$$B_{k+1} = B_k + a\mathbf{u}\mathbf{u}^T + b\mathbf{v}\mathbf{v}^T \quad (4.16)$$

Impondo a condição *quasi-Newton*, reescrevemos:

$$B_{k+1} \Delta \mathbf{x}_k = B_k \Delta \mathbf{x}_k + a\mathbf{u}\mathbf{u}^T \Delta \mathbf{x}_k + b\mathbf{v}\mathbf{v}^T \Delta \mathbf{x}_k = \mathbf{y}_k \quad (4.17)$$

Fazendo $\mathbf{u} = \mathbf{y}_k$ e $\mathbf{x} = B_k \Delta \mathbf{x}_k$, temos:

$$B_k \Delta \mathbf{x}_k + a\mathbf{y}_k \mathbf{y}_k^T \Delta \mathbf{x}_k + bB_k \Delta \mathbf{x}_k \Delta \mathbf{x}_k^T B_k^T \Delta \mathbf{x}_k = \mathbf{y}_k \quad (4.18)$$

⁵Este passo é detalhado na dedução do BFGS feita em [86].

Isolando a e b , chegamos a:

$$a = \frac{1}{\mathbf{y}_k^T \Delta \mathbf{x}_k}, b = -\frac{1}{\Delta \mathbf{x}_k^T B_k \Delta \mathbf{x}_k} \quad (4.19)$$

Substituindo na Equação (4.15), chegamos à equação de atualização da Hessiana para o BFGS:

$$B_{k+1} = B_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \Delta \mathbf{x}_k} - \frac{B_k \Delta \mathbf{x}_k \Delta \mathbf{x}_k^T B_k}{\Delta \mathbf{x}_k^T B_k \Delta \mathbf{x}_k} \quad (4.20)$$

Para a implementação, recordemos, do Método de Newton, na Equação (4.9), que precisamos da inversa da Hessiana aproximada B . Para inverter a matriz da Equação (4.20), aplicamos a fórmula de Sherman-Morrisson, chegando a:

$$B_{k+1}^{-1} = \left(I - \frac{\Delta \mathbf{x}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \Delta \mathbf{x}_k} \right) B_k^{-1} \left(I - \frac{\mathbf{y}_k \Delta \mathbf{x}_k^T}{\mathbf{y}_k^T \Delta \mathbf{x}_k} \right) + \frac{\Delta \mathbf{x}_k \Delta \mathbf{x}_k^T}{\mathbf{y}_k^T \Delta \mathbf{x}_k} \quad (4.21)$$

A Equação (4.21) é, finalmente, a atualização para a Hessiana aproximada, invertida, do BFGS.

O algoritmo fica, então:

Algoritmo 3 BFGS para minimização de funções

Forneça \mathbf{x}_0 , forneça B_0^{-1} , forneça ϵ , faça $k \leftarrow 0$;

enquanto $\|\nabla f(\mathbf{x}_k)\| \geq \epsilon$ **faça**

 Calcule a direção de busca $p_k = -B_k \nabla f(\mathbf{x}_k)$

 Faça $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k - \eta_k p_k$ onde η_k é calculada por uma busca linear;

 Defina $\Delta \mathbf{x}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ e $\mathbf{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$;

 Calcule B_{k+1} usando a Equação (4.21);

fim

retorna \mathbf{x}_k

Comentamos alguns pormenores do Algoritmo 3. A inicialização de B_0^{-1} pode ser feita escolhendo a matriz identidade, por exemplo. Sobre a taxa de aprendizado, vale notar que acrescentamos um índice k . No BFGS, a taxa de aprendizado é adaptativa e muda a cada iteração. A atualização é feita por uma busca linear, atendendo às condições de Wolfe. Para maior profundidade neste procedimento, ver [86].

O BFGS é um algoritmo bastante eficiente para otimização numérica, dentre alguns possíveis, frente aos seus predecessores. A convergência é superlinear⁶ e, como as atualizações não requerem a inversão da matriz Hessiana, a complexidade computacional é $\mathcal{O}(n^2)$, em comparação com $\mathcal{O}(n^3)$ no Método de Newton [86].

⁶Dizemos que uma seqüência x_t converge para x_∞ de maneira superlinear se $\lim_{n \rightarrow \infty} \frac{\|x_{n+1} - x_\infty\|}{\|x_n - x_\infty\|} = 0$ [91].

Terminamos esta seção homenageando os desenvolvedores do algoritmo, cujas iniciais dão seu nome: os matemáticos Charles George Broyden, Roger Fletcher, Donald Goldfarb e David Shanno são os sujeitos retratados, em momento de descontração, na Figura 4.2.

Figura 4.2 - Broyden, Fletcher, Goldfarb e Shanno (BFGS), em momento de descontração.



Fonte: [92].

4.2 MODELOS DE BOX & JENKINS

A abordagem desenvolvida por George E.P. Box e Gwilym Jenkins para análise e previsão de séries temporais, também conhecida por **modelos de Box & Jenkins** [93] ou *modelos ARIMA*, fundamenta-se em descrever o valor de uma série temporal em um dado instante como combinação linear de seus valores no passado, podendo levar em conta também o histórico dos erros de previsão.

Usaremos os modelos conforme definidos originalmente por Box & Jenkins [93].

Um **modelo auto-regressivo de ordem p** , ou $AR(p)$ (AR do inglês *Auto-Regressive*), é escrito da forma a seguir:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (4.22)$$

A Equação (4.22) é uma regressão múltipla dos valores defasados da variável y a partir do instante t . c é uma constante e ε_t é ruído branco (uma série temporal sem autocorrelação).

Alternativamente, podemos construir um modelo auto-regressivo usando os valores ε , defasados a partir de t . O **modelo de médias móveis de ordem q** , ou $MA(q)$ (do

inglês *Moving Average*), fica da seguinte forma:

$$y_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q} \quad (4.23)$$

Se combinarmos os dois modelos, teremos um modelo ARMA(p, q), como este:

$$y_t = c + \phi_1y_{t-1} + \dots + \phi_py_{t-p} + \dots + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q} \quad (4.24)$$

Em notação mais concisa, utilizando o operador retardo ⁷ para todo $t > 1$, temos:

$$\phi_p(B)y_t = \theta_q(B)\varepsilon_t \quad (4.25)$$

No caso de séries temporais não-estacionárias, podemos torná-las estacionárias diferenciando os dados. Seja $\{y'_t\}$ a série diferenciada (que pode ser diferenciada mais de uma vez), um modelo ARIMA (do inglês *auto-regressive integrated moving average* tem a forma geral a seguir:

$$y_t = c + \phi_1y'_{t-1} + \dots + \phi_py'_{t-p} + \dots + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q} \quad (4.26)$$

Em notação de operador retardo, sendo d a ordem da diferenciação, temos:

$$\phi_p(B)(1 - B)^d y_t = \theta_q(B)\varepsilon_t \quad (4.27)$$

Um modelo ARMA pode ser adaptado para dados com sazonalidade. Consideremos um período sazonal de tamanho s . Um modelo ARMA periódico, ou SARMA(p, q) \times (P, Q) _{s} (do inglês *Seasonal ARMA*) é escrito como:

$$\phi_p(B)\Phi_P(B^s)(y_t - c) = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (4.28)$$

Nesta equação, c é uma constante; ε_t é ruído branco; B é o operador retardo; e ϕ_p , Φ_P , θ_q e Θ_Q são funções polinomiais de ordens p , P , q e Q , respectivamente.

Também podemos elaborar modelos SARMA para dados com diferenciação, além de diferenciação sazonal, como o seguinte:

$$\phi_p(B)\Phi_P(B^s)(1 - B)^d(1 - B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (4.29)$$

Nesta equação, acrescentamos d , o ordem da diferenciação regular, e D , a ordem da diferenciação sazonal. s é a periodicidade da sazonalidade.

Modelos de Box & Jenkins foram muito utilizados para prever a carga elétrica na indústria na década de 1980 [18], mas ainda hoje encontramos casos de sucesso. Talvez o de maior destaque seja um modelo utilizado no sistema de transmissão espanhol (*Red Eléctrica de España*) há duas décadas, recentemente aprimorado por [94].

⁷O operador retardo B atua sobre uma série temporal de modo a retornar o elemento anterior, isto é, $Bx_t = x_{t-1}$. O operador também pode ser elevado a potências inteiras, de modo que $B^k x_t = x_{t-k}$

O modelo de [94] tem como base uma regressão linear múltipla, com variáveis explicativas baseadas na meteorologia e na ocorrência de dias especiais, somados a uma componente que é função das cargas passadas. Esta componente é modelada como um SARIMA, adaptado para acomodar mais de uma sazonalidade.

Os autores, em [94], analisam e prevêm cada hora do dia em separado, ficando, portanto, com 24 modelos. Um modelo básico de duas sazonalidades (diária e semanal) para a hora h do dia d , a partir do instante t , é o seguinte:

$$y_{h,d|t} = \beta_h^T \mathbf{x}_d + z_{d,h} \quad (4.30a)$$

$$\phi_h(B) \nabla_{D_1}^{s_1} \nabla_{D_2}^{s_2} z_{h,d|t} = \theta_h(B^{s_1}) \Theta_h(B^{s_2}) \varepsilon_{h,d|t} \quad (4.30b)$$

A Eq. (4.30a) é uma regressão linear múltipla, onde o vetor \mathbf{x}_d de variáveis explicativas é diferente para cada dia d , mas é o mesmo para todas as horas de um mesmo dia. As variáveis explicativas podem ser relacionadas ao clima ou a dias especiais, como dissemos.

A componente $z_{h,d|t}$ segue o modelo SARIMA da Eq. (4.30b), onde ∇_{s_1} e ∇_{s_2} são operadores de diferenças sazonais, com períodos s_1 e s_2 (não são feitas diferenciações regulares), ϕ_h , θ_h , Θ_h são funções polinomiais e $\varepsilon_{h,d|t}$ é ruído branco.

Sobre a escolha desta estrutura específica para a Eq. (4.30b), os autores dizem que

...é justificável dos pontos de vista empírico, teórico, e computacional. Primeiro, partindo de uma perspectiva empírica, a metodologia padrão de identificação, que é baseada na análise das funções simples e parcial de autocorrelação das 24 séries diárias, sugere duas componentes de médias móveis do lado direito e um polinômio auto-regressivo e as diferenças que aparecem do lado esquerdo. Segundo, o modelo teórico inclui as três componentes necessárias para explicar as principais características que influenciam a dinâmica da carga elétrica: a componente auto-regressiva considera os efeitos de curto prazo, enquanto a componente de médias móveis coleta os ciclos diário e semanal. Finalmente, do ponto de vista computacional, um modelo com esta estrutura pode ser estimado de forma muito eficiente. ([86], p.23).

Os autores mencionam uma “metodologia padrão de identificação”. Esta metodologia será descrita nas próximas seções, bem como as formas de se escolher a ordem de um modelo de Box & Jenkins e estimar seus parâmetros.

4.2.1 Identificação de um modelo de Box & Jenkins: método de Hyndman-Khandakar

A chamada **metodologia de Box & Jenkins** para construir modelos de auto-regressão [93] consiste nos três passos seguintes:

- **Identificação do modelo:** por meio de gráficos e medidas de autocorrelação, escolhemos o tipo de modelo e a ordem, isto é, p , d , q e, no caso de modelos periódicos, P , D e Q ;
- **Estimação de parâmetros:** encontramos os parâmetros do modelo (ϕ, θ , etc.) através de um método de otimização adequado;
- **Análise de resíduos:** analisamos as autocorrelações na série de resíduos, para verificar se o modelo foi capaz de capturar toda a correlação serial (sobrando, pois, apenas a porção aleatória).

O problema desse procedimento é que ele já é suficientemente trabalhoso para uma série temporal. Quando lidamos com várias, como é o nosso caso, é imperativo procurar por alguma maneira confiável de automatizar algumas partes do processo.

Para sustentar a razão pela qual vários autores tentaram propor formas de automatizar a construção de modelos de Box & Jenkins, [95] ilustram o tamanho do potencial problema através de um exemplo:

Suponha que tenhamos dados sazonais e consideremos modelos $ARIMA(p, d, q)(P, D, Q)_m$ onde p e q possam assumir valores de 0 a 3, e P e Q possam assumir valores de 0 a 1. Quando $c = 0$, existe um total de 288 modelos, e quando $c \neq 0$, há um total de 192 modelos possíveis, somando 480 modelos. Se os valores p, d, q, P, D, Q podem variar ainda mais o número de modelos possíveis cresce rapidamente. Consequentemente, não é viável simplesmente ajustar cada potencial modelo e escolher aquele com o menor AIC⁸. ([95], p. 10).

Hyndman e Kandakhar propõem, então, uma estratégia *step-wise* para percorrer o espaço de modelos de maneira eficiente, a fim de chegar ao modelo com menor AIC [95]. Este é o chamado *Método Hyndman-Kandakhar*, o qual utilizamos na implementação dos modelos de Box & Jenkins nesta tese.

O AIC (**Critério de Informação de Akaike**, do inglês **Akaike's Information Criterion**) é uma medida muito utilizada para a seleção de modelos de séries temporais porque é uma medida que não depende de dados *out-of-sample*, como o MSE [95].

O AIC é dado por:

$$AIC = 2k + 2 \log(L) \quad (4.31)$$

Na Eq. (4.1), k é o número de parâmetros do modelo e L é o valor máximo da função de verossimilhança.

⁸Definimos esta medida algumas linhas adiante.

No caso de modelos de Box & Jenkins, temos:

$$AIC = 2(p + q + P + Q + k) - 2\log(L) \quad (4.32)$$

Na Eq. (4.32), $k = 1$ se $c \neq 0$, e $k = 0$ caso contrário, e L é o máximo da função de verossimilhança para o modelo ajustado para os dados diferenciados, $(1 - B^s)^D(1 - B)^D y_t$

O método é dividido em três passos. O **Passo 1** é determinar a ordem da diferenciação, d , e da diferenciação sazonal, D , o que pode ser feito utilizando, respectivamente, os testes KPSS [96] e Canova-Hansen [97].

Considerando s o tamanho do período sazonal, o **Passo 2** é tentar quatro modelos possíveis:

- ARIMA(2, d , 2) se $s = 1$, e ARIMA(2, d , 2)(1, D , 1) se $s > 1$;
- ARIMA(0, d , 0) se $s = 1$ e ARIMA(0, d , 0)(0, D , 0) se $s > 1$;
- ARIMA(1, d , 0) se $s = 1$ e ARIMA(1, d , 0)(1, D , 0) se $s > 1$;
- ARIMA(0, d , 1) se $s = 1$ e ARIMA(0, d , 1)(0, D , 1) se $s > 1$.

Se $d + D \leq 1$, ajustamos os modelos com $c \neq 0$; caso contrário, com $c = 0$. O modelo com menor AIC torna-se o “modelo atual”.

No **Passo 3**, testamos as seguintes variações no modelo atual:

- Variamos p , q , P , ou Q , um de cada vez, em ± 1 em relação ao modelo atual;
- Variamos p e q , em ± 1 em relação ao modelo atual;
- Variamos P e Q , em ± 1 em relação ao modelo atual;
- Incluímos a constante no modelo atual se $c = 0$ ou excluímos se $c \neq 0$.

Sempre que um modelo com menor AIC é encontrado, este torna-se o modelo atual. O processo é repetido até que não se possa mais encontrar um modelo próximo do modelo atual com AIC menor.

Para garantir convergência, [95] listam algumas restrições para os modelos ajustados:

- Os valores de p e q não devem exceder algum limite superior específico (por padrão, 5);
- Os valores de P e Q não devem exceder algum limite superior específico (por padrão, 2);

- Qualquer modelo próximo de ser não-inversível ou não-causal (quando as raízes de $\Phi(B)\phi(B)$ e/ou $\Theta(B)\theta(B)$ forem menores que 1,001 em módulo) deve ser rejeitado;
- Se, durante a otimização dos parâmetros, houver algum erro, o modelo deve ser rejeitado (qualquer modelo que seja muito difícil de estimar não deve ser considerado bom para os dados).

Garantindo que o espaço de modelos seja finito, o método convergirá para algum modelo válido.

4.2.2 Estimação dos parâmetros de um modelo de Box & Jenkins

Para estimar os parâmetros em um problema de regressão, o primeiro método a que normalmente recorreremos é o **método dos mínimos quadrados ordinários** (MQO), que consiste em encontrar β para $\mathbf{y} = \beta X$ que minimize $\mathbf{y} - X\beta$.

Apesar de os modelos de Box & Jenkins serem uma família de modelos de (auto) regressão, há um detalhe a ser considerado. O método dos MQO utiliza a matriz de regressores X para estimar os coeficientes, fazendo $\beta = X^T X - X^T \mathbf{y}$; o que só é possível quando todos regressores são observáveis. Quando $q > 0$, um modelo ARIMA terá variáveis não-observáveis: os erros ε defasados.

Por conta disso, para estimação dos parâmetros de modelos de Box & Jenkins costuma-se utilizar a *estimação por máxima verossimilhança*. Vejamos como isso pode ser feito, em um roteiro baseado em [98]. Descreveremos o processo para modelos ARMA por razões didáticas, mas é possível estender para casos sazonais e/ou com diferenciação.

Seja $\boldsymbol{\theta} = [\phi_1, \dots, \phi_p, \dots, \theta_q, \sigma^2]^T$ o vetor de parâmetros populacionais que gostaríamos de estimar.

Suponha que dispomos da amostra de observações $\mathbf{y} = (y_1, \dots, y_N)$, de tamanho N .

Seja a função de distribuição de probabilidade conjunta dos dados denotada por $f(y_t, y_{t-1}, \dots, y_1; \boldsymbol{\theta})$.

A função de verossimilhança será esta função, condicionada à observação da amostra \mathbf{y} :

$$L(\boldsymbol{\theta} | \mathbf{y}) = f(y_t, y_{t-1}, \dots, y_1; \boldsymbol{\theta}) \quad (4.33)$$

O *estimador de máxima verossimilhança* (EMV) para os parâmetros será:

$$\hat{\boldsymbol{\theta}}_{EMV} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} L(\boldsymbol{\theta} | \mathbf{y}); \boldsymbol{\theta} \in \Theta \quad (4.34)$$

onde Θ é o espaço de parâmetros.

Por conta de aspectos computacionais, é usual trabalhar com o logaritmo natural da função de verossimilhança ⁹. Para fins de otimização, isto não altera a solução do

⁹Não fazê-lo poderia levar a um produtório com potencial de *underflow*.

problema, porque o logaritmo é uma função monotônica, então o conjunto de parâmetros que maximiza a função de verossimilhança é o mesmo que maximiza seu logaritmo natural.

Vamos denotar $\ln L(\boldsymbol{\theta} \mid \mathbf{y}) = l(\boldsymbol{\theta} \mid \mathbf{y})$. Se assumirmos que l é contínua e que a derivada de $l(\boldsymbol{\theta} \mid \mathbf{y})$ em relação a $\boldsymbol{\theta}$ existe, o EMV $\hat{\boldsymbol{\theta}}$ será a solução de:

$$l'(\boldsymbol{\theta} \mid \mathbf{y}) = \mathbf{0} \quad (4.35)$$

Suponhamos que os dados \mathbf{y} tenham sido extraídos de um processo ARMA(p, q) com distribuição Gaussiana de média zero. A função de verossimilhança de \mathbf{y} será dada por:

$$L(\boldsymbol{\theta} \mid \mathbf{y}) = (2\pi)^{-N/2} |\Gamma(\boldsymbol{\theta})|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{y}^T \Gamma(\boldsymbol{\theta})^{-1} \mathbf{y} \right\} \quad (4.36)$$

onde $\Gamma(\boldsymbol{\theta}) = E(\mathbf{y}\mathbf{y}^T)$ é a matriz $N \times N$ de variância-covariância de \mathbf{y} dados os parâmetros.

Aplicando o logaritmo natural, chegamos à log-verossimilhança, $l(\boldsymbol{\theta} \mid \mathbf{y})$:

$$l(\boldsymbol{\theta} \mid \mathbf{y}) = -\frac{1}{2} [N \ln(2\pi) + \ln |\Gamma(\boldsymbol{\theta})| + \mathbf{y}^T \Gamma(\boldsymbol{\theta})^{-1} \mathbf{y}] \quad (4.37)$$

Para encontrar os EMV para os parâmetros do modelo ARMA, basta, então, resolver a equação para o modelo desejado, AR(p), MA(q) ou ARMA(p, q). Esta resolução passa por um problema de otimização não-linear que pode, inclusive, ser resolvido utilizando um dos métodos *quasi-Newton* abordados na seção anterior.

Resumindo o problema da estimação de parâmetros para modelos ARMA: como não é possível utilizar o método dos MQO para estimar os parâmetros de alguns modelos, é consenso entre a comunidade de séries temporais trabalhar com o método da máxima verossimilhança, cujos estimadores associados têm características desejáveis. Supomos que a função de verossimilhança seja contínua e diferenciável, e que os dados da série temporal venham de uma distribuição Gaussiana. Então, resolvemos o problema de otimização que é encontrar o conjunto de parâmetros do modelo que maximiza a função de verossimilhança Gaussiana, isto é, os parâmetros que maximizam a probabilidade¹⁰ de se observarem aqueles dados, dado o modelo proposto.

Não faremos a derivação dos EMV para os modelos ARMA porque são inúmeros os modelos possíveis, e as derivações são longas.

Felizmente, existem rotinas computacionais confiáveis implementadas há vários anos e já amplamente testadas e difundidas para resolver este problema. Listamos as rotinas que utilizamos no Capítulo 7.

4.3 *Benchmark* ingênuo: o previsor ***Naive*** sazonal

Um **previsor ingênuo** (*naive*) sazonal tem a seguinte forma:

$$\hat{y}_{t+h|t} = L_{t+h-s} \quad (4.38)$$

¹⁰Supondo uma função densidade de probabilidade Gaussiana

Na Eq. (4.38), s é o tamanho do período sazonal. Se $s = 1$, não existe sazonalidade e a previsão para o instante t será igual ao valor da última observação disponível.

Sabemos que os dados de carga seguem, no curto prazo, dois padrões sazonais: diário, com cargas mais altas nas horas de meio do dia; e semanal, com padrões distintos para cargas de dia de semana e de final de semana. À vista disso, utilizaremos como *benchmarks* dois previsores *naive* sazonais, com períodos iguais $m = 24$ horas e $m = 168$ horas. No primeiro caso, a carga prevista será igual à observada na mesma hora no dia anterior e, no segundo, igual à observada na mesma hora na semana anterior.

Estes métodos são os mais simples que se pode conceber, e é isso que os torna *benchmarks* atrativos: caso alguma proposta mais complexa não produza previsões melhores, este é um sinal de que deve ela ser revista.

Apesar de simples, um previsor *naive* com $s = 1$ é ótimo quando os dados seguem um passeio aleatório, e os *naive* sazonais costumam funcionar razoavelmente bem para dados da Economia [69].

5 MÉTODOS DE PREVISÃO: REGRESSÃO

Dado um conjunto de **variáveis de entrada** (ou **preditoras**, ou **independentes**), X_1, X_2, \dots, X_p e uma **variável de saída** (ou **resposta**, ou **independente**) y , a **análise de regressão** é um conjunto de métodos para determinar uma equação relacionando a saída com as entradas [99].

No Capítulo 4, apresentamos os modelos de Box & Jenkins, que descrevem a carga elétrica como função das cargas passadas. Aqueles modelos são auto-regressivos, porque a regressão da carga é uma combinação de valores da própria variável e dos erros em outros instantes de tempo. Já neste capítulo, tratamos a carga como a combinação linear de **variáveis exógenas**, como as temperaturas.

5.1 INTRODUÇÃO À REGRESSÃO LINEAR

Um modelo de **regressão linear** tem a seguinte forma:

$$y = \beta_0 + \sum_{j=1}^p \beta_j X_j + \varepsilon \quad (5.1)$$

onde $X_j, j = 1, \dots, p$ são as p variáveis de entrada, e y é a variável de saída. ε é uma variável aleatória não-observável que denota o erro de previsão. Os valores β_0, \dots, β_p são os parâmetros do modelo. O modelo é linear nos parâmetros, por isso o nome regressão **linear** [101].

Para um exemplo, extraímos, aleatoriamente, de nossa base de dados, as cargas médias diárias de uma sequência de 30 dias da primavera, da zona de carga D - WEST, e as 30 temperaturas médias diárias observadas na estação BUF correspondente aos mesmos dias. Exibimos os pares de temperatura e carga (T_t, y_t) no diagrama de dispersão da Figura 5.1:

Vamos supor que tentemos modelar a relação entre as cargas e temperaturas usando regressão linear. Utilizando a função *lm*, do pacote *stats* do R [100], que ajusta modelos de regressão linear para conjuntos de dados, chegamos ao seguinte modelo:

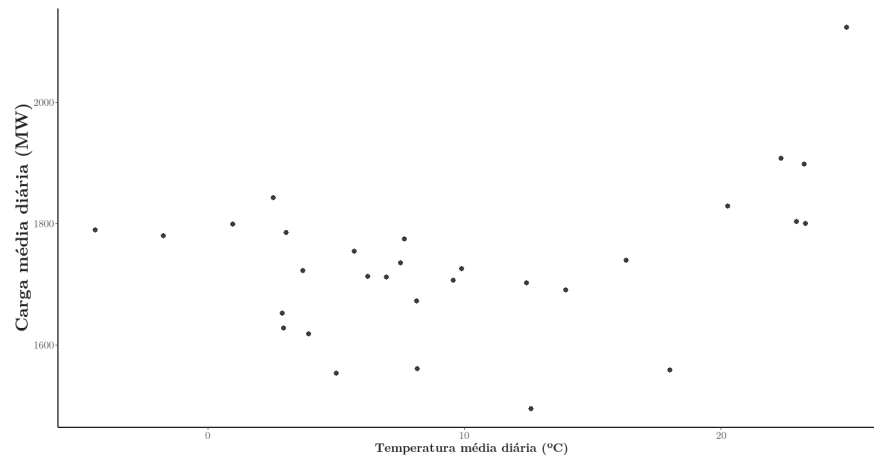
$$y = 1678,72T + 5,71 + \varepsilon \quad (5.2)$$

Para calcular uma previsão \hat{y}_t , supomos o erro igual a zero e lançamos o valor X_t correspondente na seguinte aproximação:

$$\hat{y}_t = 1678,72T_t + 5,71 \quad (5.3)$$

Então, por exemplo, para uma temperatura de 20,2 °C, a Eq. (5.3) forneceria uma carga prevista de 1792,92 MW.

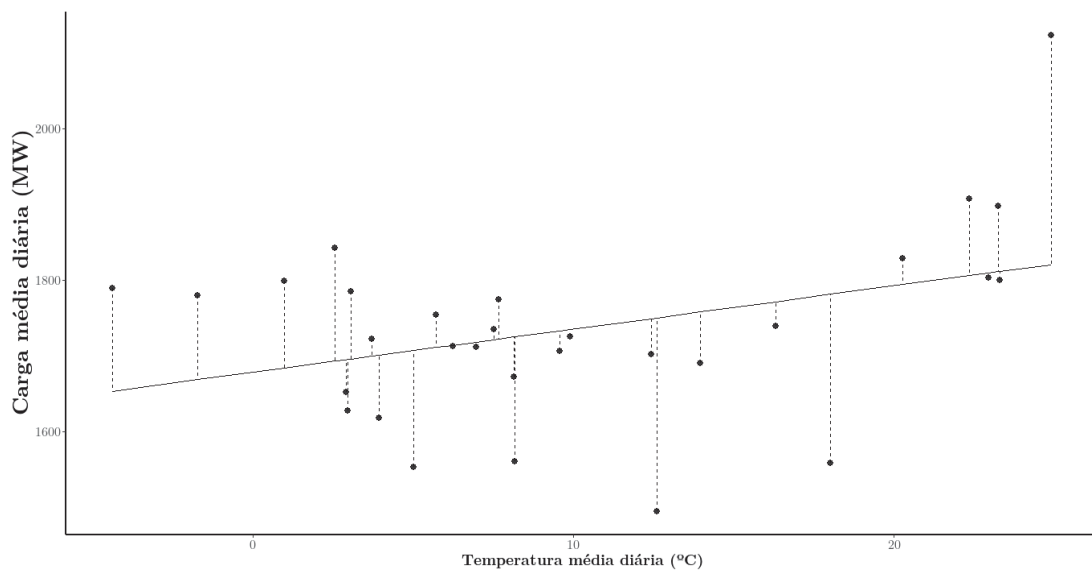
Figura 5.1 - - Cargas médias na zona D - WEST *versus* temperaturas médias na estação BUF, diárias, para 30 dias amostrados por conveniência.



Lançando os 30 valores da série $\{X_t\}$ que já temos na Eq. 5.3, chegamos a uma série de valores $\{\hat{y}_t\}$. Como já dispomos das cargas observadas para estes dados, podemos calcular os erros de previsão, $\varepsilon_t = \hat{y}_t - y_t$.

Na Figura 5.2, mostramos a reta de regressão, além dos erros de previsão, representados pelas linhas pontilhadas, ou seja, as distâncias entre os valores preditos e observados:

Figura 5.2 - - Reta de regressão e erros para cargas médias na zona D - WEST *versus* temperaturas médias na estação BUF, diárias, para 30 dias amostrados por conveniência.



Há duas questões importantes há se levantar sobre este exemplo.

Primeiramente, a reta que ajustamos não é a única possível. Teríamos que lançar

mão de algum método de escolha dos coeficientes β_0 e β_1 de modo a garantir que as previsões sejam as de menor erro e que o modelo generalize bem para novos dados ¹.

Segundo, a correlação entre os dados não parece linear. Vamos calcular o **coeficiente de correlação linear de Pearson**, para mensurar esta correlação. A estimativa para o coeficiente, a partir de uma amostra de n observações, para duas variáveis, x e y , com médias amostrais \bar{x} e \bar{y} , é dada por:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (5.4)$$

O coeficiente de correlação assume valores no intervalo $[-1,1]$. Quanto mais próximo de -1 , mais forte é a correlação linear negativa (o acréscimo no valor de uma variável leva ao decréscimo no valor de outra), ao passo que, quanto mais próximo de 1 , mais forte é a correlação linear positiva (o acréscimos ou decréscimos em uma variável implicam, respectivamente em acréscimos ou decréscimos em outra). Quando o coeficiente é zero, não há correlação linear.

O coeficiente de correlação linear de Pearson para estes dados é de $0,38$. Como não é tão próximo de -1 ou 1 , uma reta talvez não seja o modelo mais adequado.

O erro médio percentual absoluto para as previsões, com o modelo de regressão da Equação 5.2 é de $4,86\%$. Podemos tentar melhorar a acurácia testando um modelo com não-linearidade nas variáveis de entrada².

Vamos considerar, por exemplo, o seguinte modelo:

$$y = \sum_{j=1}^5 \beta_j X^j + \varepsilon \quad (5.5)$$

O modelo da Eq. (5.5) é de um tipo que pode ser chamado de *regressão polinomial*, já que a forma funcional é um polinômio (de quinto grau, no caso). Estimando os coeficientes utilizando a função *lm* do pacote *stats* do R, chegamos ao modelo seguinte:

$$y = -29,49\beta_1 T - 2,77\beta_2 T^2 + 0,76\beta_3 T^3 - 0,20\beta_4 T^4 + 0,02\beta_5 T^5 + 1767 + \varepsilon$$

Para utilizar a equação para previsão, reescrevemos como a seguir:

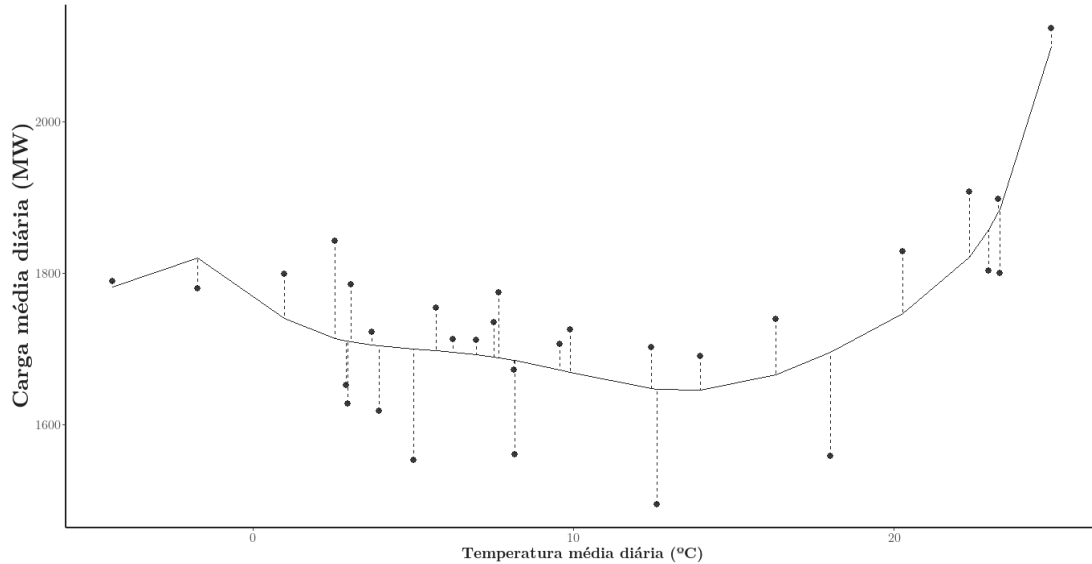
$$\hat{y}_t = 29,49T_t - 2,77T_t^2 + 0,76T_t^3 - 0,20T_t^4 + 0,02T_t^5 + 1767$$

Na Figura 5.3, mostramos o polinômio de regressão, além dos erros de previsão, representados pelas linhas pontilhadas:

¹Na verdade, como uma reta (ou um plano, ou hiperplano) não é uma forma funcional muito complexa, ela acaba não sendo tão específica para os dados e, por isso, não costuma levar ao sobreajuste.

²Reforçamos que não introduzimos não-linearidade nos coeficientes, o que faria com que o modelo deixasse de ser uma regressão linear [101]

Figura 5.3 - - Regressão com polinômio de quinto grau e erros para cargas médias na zona D - WEST *versus* temperaturas médias na estação BUF, diárias, para 30 dias amostrados por conveniência.



Comparando com a Figura 5.2, verificamos que o modelo da Figura 5.3 parece acompanhar melhor formato não linear. O erro médio percentual absoluto é de 3,87%, contra 4,86% do modelo de primeiro grau.

No entanto, é necessário ter cautela para que o modelo, por ser mais específico para os dados, não fique sobreajustado e generalize mal.

5.2 ESTIMAÇÃO DOS COEFICIENTES DE UM MODELO DE REGRESSÃO LINEAR

5.2.1 Equações normais

Para estimar os coeficientes β_0, \dots, β_p em um modelo de regressão linear, partimos de um conjunto de dados observados $(\mathbf{x}_1, y_1) \dots (\mathbf{x}_N, y_N)$. Cada $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{i,p-1})$ é um vetor de observações das variáveis X_1, X_2, \dots, X_{p-1} para o i -ésimo caso.

Vamos reescrever a Eq. 5.1 em formato matricial.

Considerando:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{N1} & \dots & x_{N,p-1} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix} \quad (5.6)$$

A forma matricial será dada por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} \quad (5.7)$$

Um método bastante utilizado para estimar $\boldsymbol{\beta}$ é o **método dos mínimos quadrados** [101], que consiste em minimizar o quadrado da norma euclidiana do vetor de resíduos

$r = \mathbf{y} - \beta\mathbf{X}$, ou seja,

$$\min_{\beta} \|\mathbf{r}\|_2^2 = \min_{\beta} \|\beta\mathbf{X} - \mathbf{y}\|_2^2 \quad (5.8)$$

Expandindo a norma, temos:

$$\begin{aligned} \|\mathbf{r}\|_2^2 &= \mathbf{r}^T \mathbf{r} = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \\ &= \mathbf{y}^T \mathbf{y} - 2\beta^T \mathbf{X}^T \mathbf{y} + \beta^T \mathbf{X}^T \mathbf{X} \mathbf{y} \end{aligned} \quad (5.9)$$

Para encontrar β que resolva o problema de minimização, tomamos a derivada de \mathbf{y} em relação a β e igualamos a zero:

$$2\mathbf{X}^T \mathbf{X} \beta - 2\mathbf{X}^T \mathbf{y} = \mathbf{0} \quad (5.10)$$

Ficamos com este sistema linear $n \times n$ de equações:

$$\mathbf{X}^T \mathbf{X} \beta = \mathbf{X}^T \mathbf{y} \quad (5.11)$$

O sistema da Eq. (5.5) é chamado de sistema de **equações normais**.

O uso das equações normais para resolver o problema dos mínimos quadrados é razoável em teoria, mas, na prática, pode produzir resultados inacurados [101]. Para o problema $\mathbf{y} = \mathbf{X}\beta$, o uso das equações normais implica em elevar o número de condicionamento de \mathbf{X} ao quadrado³, já que temos que formar $\mathbf{X}^T \mathbf{X}$. Isso implica que, se \mathbf{X} e \mathbf{y} são conhecidos com N dígitos significativos, e o número de condicionamento de \mathbf{X} para $\kappa(\mathbf{X}) \approx 10^k$, β será estimado com $N - 2k$ dígitos significativos. Perde-se, portanto, o dobro de dígitos significativos.

No contexto de um problema de regressão, as colunas de \mathbf{X} correspondem a variáveis explicativas; por exemplo, valores de temperatura que serão utilizadas para descrever a carga elétrica. É relativamente comum que exista **multicolinearidade** (*i.e.* que algumas variáveis explicativas sejam correlacionadas). No caso da previsão de carga, isto é comum quando utilizamos, por exemplo, observações de temperaturas em horários diferentes em magnitudes diferentes (ao quadrado, ao cubo, *etc.*). Quando há multicolinearidade, as colunas das variáveis correlacionadas na matriz \mathbf{X} são linearmente dependentes e isto faz com que a matriz seja mal condicionada (*e.g.* $\det(\mathbf{X}^T \mathbf{X}) \cong 0$).

Uma estratégia para evitar estas dificuldades numéricas trazidas pelas equações normais é reescrever o problema de mínimos quadrados de outra forma, que seja mais robusta às mudanças nas entradas, e que preserve a solução do problema original.

5.2.2 Otimização numérica para estimação dos coeficientes de regressão

Vamos decompor \mathbf{X} em sua fatoração QR [101]:

$$\mathbf{X} = \mathbf{Q}\mathbf{R} \quad (5.12)$$

³O **número de condicionamento** é uma função que mede o quanto mudanças nas variáveis de saída são sensíveis a mudanças nas variáveis de entrada [101]. Quando o número de condicionamento é alto, o problema é **mal-condicionado** e sua solução é difícil de encontrar (em outras palavras, $\det(\mathbf{X}^T \mathbf{X}) \cong 0$).

onde \mathbf{Q} é ortogonal e \mathbf{R} é triangular superior. Isto facilita a resolução do sistema $\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{y}$, porque reescreveríamos $\mathbf{X}^T \mathbf{X}$ como:

$$\mathbf{X}^T \mathbf{X} = \mathbf{R}^T \mathbf{Q}^T \mathbf{Q} \mathbf{R} = \mathbf{R}^T \mathbf{R} \quad (5.13)$$

O sistema ficaria:

$$\mathbf{R}^T \mathbf{R} \boldsymbol{\beta} = \mathbf{R}^T \mathbf{Q}^T \mathbf{y} \quad (5.14)$$

Supondo que \mathbf{R}^T seja uma matriz de posto cheio, ficamos com:

$$\mathbf{R} \boldsymbol{\beta} = \mathbf{Q}^T \mathbf{y} \quad (5.15)$$

Como \mathbf{R} é triangular superior, a solução torna-se trivial.

O método de fatoração QR que, computacionalmente, produz melhores resultados para resolver os problemas de mínimos quadrados são as **reflexões de Householder** [101].

Um reflexão de Householder é uma transformação linear que faz uma reflexão sobre um hiperplano. O mapeamento de um vetor \mathbf{x} usando a reflexão de Householder é dado por:

$$\mathbf{x} \Rightarrow \mathbf{x} - 2 \langle \mathbf{x}, \mathbf{v} \rangle \mathbf{v} \quad (5.16)$$

Em formato de matriz, a Eq. (5.16) pode ser reescrita como:

$$\mathbf{T} = \mathbf{I} - 2\mathbf{v}\mathbf{v}^T \quad (5.17)$$

Podemos refletir um vetor \mathbf{x} sobre um determinado hiperplano de modo que os elementos de \mathbf{x} , exceto o primeiro, sejam zerados. Seja \mathbf{Q}_x a reflexão que zere os elementos de \mathbf{x} , exceto o primeiro. Se denotamos a primeira coluna de \mathbf{X} como \mathbf{x}_1 , então $\mathbf{Q}_{\mathbf{x}_1} \mathbf{X}$ é a matriz que zera os elementos da primeira coluna, exceto o primeiro, resultando na seguinte forma:

$$\begin{pmatrix} \alpha_1 & * & \dots & * \\ 0 & & & \\ 0 & \mathbf{X}_2 & & \\ 0 & & & \end{pmatrix} \quad (5.18)$$

Podemos seguir aplicando este procedimento à matriz \mathbf{X}_2 , de modo que os elementos da primeira coluna, abaixo do primeiro, sejam zerados. Resulta uma matriz neste formato:

$$\mathbf{Q}_2 \mathbf{Q}_1 \mathbf{X} = \begin{pmatrix} \alpha_1 & \alpha_{12} & * & \dots & * \\ 0 & \alpha_{22} & * & \dots & * \\ 0 & & & & \\ 0 & \mathbf{X}_3 & & & \\ 0 & & & & \end{pmatrix} \quad (5.19)$$

Continuando o procedimento, criaremos uma seqüência de reflexões $\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_p$, que reduzem \mathbf{X} a uma matriz triangular superior:

$$\mathbf{Q}_p, \mathbf{Q}_{p-1} \dots \mathbf{Q}_1 \mathbf{X} = \mathbf{R} \quad (5.20)$$

Cada matriz de reflexão é ortogonal, então podemos inverter \mathbf{Q} de modo a ficar com:

$$\mathbf{X} = \mathbf{Q}_1^T \mathbf{Q}_2^T \dots \mathbf{Q}_p^T \mathbf{R} \quad (5.21)$$

Fazendo $\mathbf{Q} = \mathbf{Q}_1^T \mathbf{Q}_2^T \dots \mathbf{Q}_p^T$, temos a decomposição \mathbf{QR} de \mathbf{X} . Basta, então, resolver $\mathbf{R}\boldsymbol{\beta} = \mathbf{Q}^T \mathbf{y}$ para encontrar os coeficientes da regressão de uma forma satisfatória quanto aos aspectos numéricos.

5.3 UM PREVISOR *NAIVE* PARA CARGAS ELÉTRICAS BASEADO EM REGRESSÃO LINEAR

A literatura traz várias aplicações de regressão linear à previsão de carga elétrica além daquela que apresentamos no Capítulo 3.

Um modelo de interesse para nossa pesquisa é o *Tao's Vanilla Benchmark*, que chamaremos de TVB. Este modelo foi proposto como *benchmark* para a pesquisa na área [102]. A ideia dos proponentes foi estabelecer um modelo de *benchmark naive* de maior valor prático no contexto da previsão de carga elétrica, levando em conta variáveis que frequentemente aparecem nos modelos, como indicadores para os dias da semana, e temperaturas.

O TVB foi utilizado como *benchmark* na competição GEFCom 2012 [20] e é o único modelo testado até o momento para a testar métodos de seleção e combinação de estações meteorológicas.

O TVB descreve a carga y no instante t da seguinte forma:

$$\begin{aligned} y_t = & \beta_0 + \beta_1 \times Trend_t + \beta_2 \times Day_t \times Hour_t + \beta_3 \times Month_t \\ & + \beta_4 \times Month_t \times + \beta_5 \times Month_t \times TMP_t^2 \\ & + \beta_6 \times Month_t \times TMP_t^3 + \beta_7 \times Hour_t \times TMP_t \\ & + \beta_8 \times Hour_t \times TMP_t^2 + \beta_9 \times Hour_t \times TMP_t^3 \end{aligned} \quad (5.22)$$

As variáveis no modelo são:

- (i) *Trend*: variável quantitativa para capturar a tendência local de crescimento ou decrescimento. É atribuído um número natural a cada hora, em ordem crescente, começando do 1;
- (ii) *Hour*: hora do dia (qualitativa, 24 classes);
- (iii) *Day*: dia da semana (qualitativa, 7 classes);

- (iv) *Month*: mês do ano (qualitativa, 12 classes);
- (v) *TMP*: temperatura.

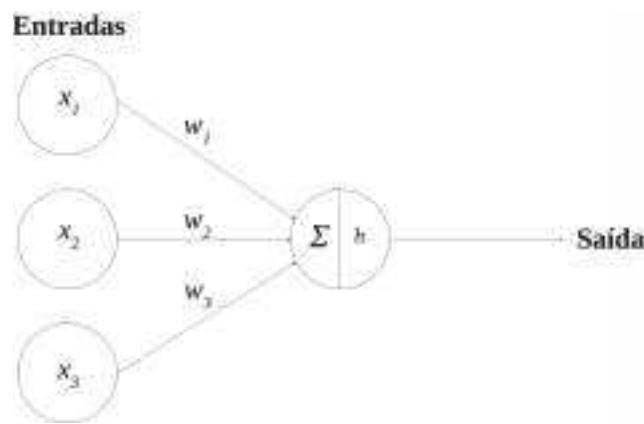
Os coeficientes do modelo podem ser estimados pelo método dos mínimos quadrados.

6 MÉTODOS DE PREVISÃO: REDES NEURAIIS

No contexto do aprendizado de máquina, um **neurônio** é um mapeamento funcional que recebe um conjunto de valores numéricos como entradas e produz um único valor como saída. O mapeamento é não-linear: os valores de entrada são combinados linearmente e o resultado passa por uma **função de ativação**, transformação que inclui a não-linearidade.

Na Figura 6.1, mostramos um esquema para um neurônio (no contexto de uma rede neural):

Figura 6.1 - - Exemplo de neurônio com três entradas.



No exemplo da Figura 6.1, o neurônio possui três entradas, as quais poderíamos representar por um **vetor de atributos**, $\mathbf{x} = [x_1, x_2, x_3]^T$. Os atributos são as variáveis do conjunto de dados utilizadas para descrever a variável de saída [75]. Por exemplo, poderíamos considerar a temperatura máxima prevista para amanhã, a temperatura mínima prevista para amanhã e a carga às 13:00h de hoje como atributos para prever da carga das 13:00h de amanhã.

Os valores $\mathbf{w} = [w_1, w_2, w_3]^T$, número reais, são os **pesos** correspondentes à conexão de cada entrada com o neurônio.

Inicialmente, o neurônio faz a combinação linear dos valores de entrada com os pesos, na Figura 6.1 representada pela letra grega Σ . Neste exemplo de três entradas, o resultado da combinação seria $w_1x_1 + w_2x_2 + w_3x_3$; idêntico, portanto, a uma regressão linear múltipla. Depois, o resultado é passado como argumento para uma função de ativação $h(\cdot)$, produzindo a saída $h(w_1x_1 + w_2x_2 + w_3x_3)$.

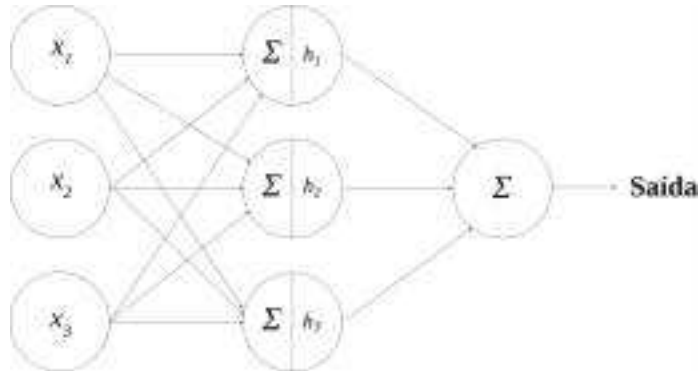
A *tangente hiperbólica*, a **função logística** e a **linear retificada (ReLU, rectified linear unit)** são algumas escolhas populares de função de ativação [75]. Estas funções introduzem não-linearidade ao mapeamento.

Uma **rede neural** é um conjunto de neurônios conectados, com o objetivo de fazer mapeamentos mais complexos por meio de funções compostas.

6.1 REDES NEURAIIS *FEEDFORWARD* (FFNN)

Mostramos um exemplo de rede neural na 6.2:

Figura 6.2 - - Exemplo de rede neural com três entradas e três neurônios na camada oculta.



Na Figura 6.2, a rede do exemplo recebe como entrada um vetor de atributos $w_1x_1 + w_2x_2 + w_3x_3$. Estas entradas são passadas para três neurônios. A cada neurônio j , $j = 1, 2, 3$, está associado um vetor $\mathbf{w}_j = [w_{j,1}, w_{j,2}, w_{j,3}]^T$, cujos elementos são os pesos de suas conexões com as entradas da rede. A cada neurônio também está associada uma função de ativação h_j .

Ao receber as entradas, os neurônios fazem a combinação linear e depois aplicam a função de ativação não-linear. A saída da rede é dada pela combinação linear das saídas dos neurônios.

A cada coluna de nós no grafo na Figura 6.2 dá-se o nome de **camada**. A primeira camada é a **camada de entrada** e a última é a **camada de saída**. A camada intermediária recebe o nome de **camada oculta**. A rede do exemplo é uma rede do tipo *feedforward* (FFNN; *feedforward neural network*), porque o fluxo não tem recursão. Esta rede também é do tipo **totalmente conectada**, porque as unidades em uma camada conectam-se com todas as ativações da camada seguinte.

O **Teorema da Aproximação Universal** demonstra que uma FFNN de uma camada oculta com um número finito de neurônios é capaz de aproximar qualquer função contínua, desde que as funções de ativação obedeçam a certas características, como a diferenciabilidade [103]. A capacidade de aproximação universal de funções das FFNNs as torna muito atrativas para a previsão de carga, já que as redes podem modelar as complexas relações não-lineares entre carga e temperatura sem que seja necessário que os desenvolvedores escolham uma forma funcional explicitamente.

Para que uma rede neural faça o mapeamento de entradas e saídas com menor erro, os pesos das conexões devem ser otimizados, em um procedimento chamado **treinamento** da rede. Para isso, não são utilizados todos os dados disponíveis, e por uma razão óbvia:

se todos os exemplos fossem usados, tanto para ajustar pesos e chegar ao modelo final, quanto para avaliar o desempenho, seria difícil saber se o modelo funciona bem para novos vetores de atributos que serão observados (em outras palavras, se a rede terá bom poder de generalização). Por isso, é habitual segmentar os dados em três amostras: **treino**, **validação** e **teste** [75].

Existem alguns esquemas possíveis para treino, validação e teste, mas, essencialmente, os dados da amostra de treino são os exemplos apresentados à rede para ajuste dos pesos e a amostra de validação é utilizada para verificar o desempenho *out-of-sample*, isto é, fora dos dados já “conhecidos”. A amostra de teste é um conjunto de dados separados para comparar o resultado da rede como os de outros modelos.

A escolha do tamanho de cada amostra é um procedimento empírico, cabendo aos desenvolvedores fazer a escolha por tentativa e erro ou guiar-se por orientações da literatura especializada. Uma regra empírica muito difundida é separar 80% dos dados para treino e 20% para teste, sendo, do conjunto de treino, extraídos 20% para validação [75].

A depender do tipo dos dados, podem-se compor as amostras de maneira aleatória. Isso não se aplica, no entanto, quando se tratam de dados de séries temporais devido à existência de correlação serial. A aleatorização pode implicar na composição de amostras sem sequência temporal, de modo que uma amostra de treino contenha dados futuros em relação a uma amostra de teste. Nesse caso, construiríamos um modelo que faz previsões já sabendo parte do futuro; uma espécie de trapaça.

6.1.1 Notação para modelagem matemática de FFNNs

Antes de explicar as principais ideias por trás do ajuste dos pesos de uma rede neural, faremos algumas definições, utilizando como exemplo uma rede de duas camadas (uma camada oculta e outra de saída), a fim de estabelecer uma notação para a modelagem matemática. Esta seção é baseada em [104, 107].

Podemos escrever uma rede neural como uma sequência de transformações funcionais. Primeiro, construímos M combinações lineares das variáveis de entrada x_1, x_2, \dots, x_D na forma

$$a_j = \sum_{i=0}^D w_{j,i}^{(1)} x_i \quad (6.1)$$

onde $j = 1, \dots, M$ são os índices para os neurônios, o índice sobrescrito (1) denota a primeira camada da rede e $w_{j,i}^{(1)}$ são os pesos¹. Os valores a_j são as *ativações* correspondentes a cada neurônio j , e cada ativação passa, depois, por uma função de ativação não-linear $h(\cdot)$, chegando a

$$z_j = h(a_j) \quad (6.2)$$

¹Exceto $w_{j,0}^{(1)}$, que são vieses.

As saídas no formato da Eq. (6.2) são os neurônio ocultos. Estes valores são novamente combinados, de modo que cheguemos aos neurônios de saída:

$$a_k = \sum_{i=0}^M w_{j,i}^{(2)} z_j \quad (6.3)$$

onde $k = 1, \dots, K$ é o total de saídas. Estas saídas são transformadas, utilizando mais uma função de ativação não-linear, de modo que se obtenham uma série de saídas \hat{y}_k .

A escolha das funções de ativação não-lineares depende, dentre outros aspectos, da natureza do problema [107].

A **tangente hiperbólica** recebe qualquer número real como entrada e produz uma saída no intervalo $[-1,1]$. Quanto maior a entrada, mais próxima a saída estará de 1; ao passo que, quanto menor a entrada, mais próxima a saída será de -1. Seu cálculo é dado pela razão entre o seno hiperbólico $\sinh a = \frac{1-e^{-2a}}{2e^{-a}}$ e o cosseno hiperbólico $\cosh a = \frac{1+e^{-2a}}{2e^{-a}}$:

$$h(a) = \frac{1 - e^{-2a}}{1 + e^{-2a}} \quad (6.4)$$

A **função logística** recebe qualquer número real como entrada e produz saídas no intervalo $[0,1]$. Quanto maior a entrada, mais próxima de 1 a saída; quanto menor a entrada, mais próxima de 0 a saída.

A função logística de um número real a é dada pela fórmula

$$h(a) = \frac{1}{e^a + 1} \quad (6.5)$$

A **função linear retificada** (ReLU), recebe um número real que, caso seja negativo, produz uma saída igual a 0; caso contrário, o próprio valor será retornado.

A ReLU para um número real a é dada pela seguinte fórmula:

$$h(a) = \begin{cases} 0 & \text{se } a < 0 \\ a & \text{c.c.} \end{cases} \quad (6.6)$$

A equação geral para a FFNN totalmente conectada de uma camada oculta fica, então,

$$y_k(\mathbf{x}, \mathbf{w}) = h \left(\sum_{j=0}^M w_{k,j}^{(2)} h \left(\sum_{i=0}^D w_{j,i}^{(1)} x_i \right) \right) \quad (6.7)$$

onde \mathbf{x} é o vetor de atributos e \mathbf{w} é o vetor de pesos.

A Eq. (6.7) pode ser facilmente estendida para redes com mais camadas, bastando ampliar a composição de funções.

6.1.2 Treinamento de uma FFNN

Com a topologia da rede definida, temos que determinar os pesos \mathbf{w} de forma a minimizar uma função de custo $J(\mathbf{w})$. Uma possibilidade é o erro médio quadrático:

$$J(\mathbf{w}) = \frac{1}{N} \sum_{n=1}^N \{y(\mathbf{x}_n, \mathbf{w}) - t_n\}^2 \quad (6.8)$$

onde $y(\mathbf{x}_n, \mathbf{w})$ é a saída da rede, dados o vetor de atributos \mathbf{x}_n , o vetor de pesos \mathbf{w} e os valores observados (também chamados de **alvos**, do inglês *targets*) t_n .

A minimização pode ser feita utilizando informações do gradiente da função de custo, por exemplo, através do *método de descida de gradiente*, aprofundado no Capítulo 4 desta tese. A equação de atualização dos pesos fica, então:

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla J(\mathbf{w}^{(\tau)}) \quad (6.9)$$

onde τ é o índice para as iterações (no contexto de redes neurais também chamadas de *épocas*) e $\eta \geq 0$ é a *taxa de aprendizado*.

Para executar o procedimento de minimização, temos que avaliar o gradiente da função de custo, $\nabla J(\mathbf{w})$. Vamos fazê-lo recorrendo ao algoritmo *backpropagation* (ou “retropropagação dos erros”). Faremos a dedução do algoritmo conforme [104, 107], para uma rede geral com topologia FFNN genérica, e funções de ativação não-lineares diferenciáveis genéricas e uma função de erro genérica.

Em uma FFNN genérica, um neurônio calcula a combinação linear de suas entradas da seguinte forma:

$$a_j = \sum_i w_{j,i} z_i \quad (6.10)$$

onde z_i é a ativação conectada a j com peso $w_{j,i}$. Notemos que uma ou mais variáveis z_i podem ser entradas e, nesse caso, denotaríamos por x_i .

A soma na Eq. (6.10) sofre uma transformação, por meio de uma função ativação não-linear $h(\cdot)$, levando à ativação z_j do neurônio j :

$$z_j = h(a_j) \quad (6.11)$$

z_j pode vir de um neurônio de saída e, nesse caso, denotaríamos por y_k .

Vamos supor que, para cada exemplo na amostra de treino, tenhamos feito o *forward propagation* (“propagação para frente”), ou seja, apresentamos os exemplos à rede e calculamos as saídas, utilizando as Equações (6.10) e (6.11).

Temos que determinar valores para os pesos de forma que uma função de custo seja minimizada. Por razões práticas, consideremos funções de custo que podem ser escritas como a soma dos erros individuais para cada exemplo na amostra de treino (o que é o caso da função na Eq. (6.8), por exemplo).

Vamos buscar por um procedimento para avaliar a derivada de J_n (ou seja, da função de custo para o n -ésimo exemplo dos dados de treino) em relação aos pesos e vieses da rede. Consideremos um peso arbitrário $w_{j,i}$. Aplicado a regra da cadeia, temos:

$$\frac{\partial J_n}{\partial w_{j,i}} = \frac{\partial J_n}{\partial a_j} \frac{\partial a_j}{\partial w_{j,i}} \quad (6.12)$$

Introduzimos a seguinte notação:

$$\delta_j = \frac{\partial J_n}{\partial a_j} \quad (6.13)$$

A partir da Eq. (6.10), podemos escrever:

$$\frac{\partial a_j}{\partial w_{j,i}} = z_i \quad (6.14)$$

Substituindo as equações (6.13) e (6.14) em (6.12), ficamos com:

$$\frac{\partial J_n}{\partial w_{j,i}} = \delta_j z_i \quad (6.15)$$

Logo, para avaliar as derivadas nos neurônios ocultos (o que é nosso objetivo para atualizar os pesos conforme a Eq. (6.9)), temos que calcular os erros δ_j .

Para os neurônios na camada de saída, o cálculo de δ_k fica:

$$\delta_k = \frac{\partial J_n}{\partial a_k} = h'(a_k) \frac{\partial J_n}{\partial y_k} \quad (6.16)$$

Já para os neurônios nas camadas ocultas, podemos utilizar a regra de cadeia e fazer o seguinte:

$$\delta_j = \frac{\partial J_n}{\partial a_j} = \sum_k \frac{\partial J_n}{\partial a_k} \frac{\partial a_k}{\partial a_j} \quad (6.17)$$

onde a soma é feita para os k neurônios com os quais o neurônio j se conecta.

Substituindo a Eq. (6.13) em (6.17), e utilizando (6.10) e (6.11), chegamos à fórmula de *backpropagation*:

$$\delta_j = h'(a_j) \sum_k w_{k,j} \delta_k \quad (6.18)$$

Esta fórmula mostra que, para calcular δ para um neurônio oculto, propagamos para trás os δ s que estão em neurônios à frente na rede.

Como os valores de δ para a camada de saída são conhecidos, podemos calcular os valores para os neurônios da camada oculta de qualquer FFNN utilizando a regra de cadeia de maneira recursiva.

Tendo deduzido o cálculo dos δ s pelo *backpropagation*, temos condições de utilizá-lo em um algoritmo de otimização (ou de “treinamento”, no contexto das redes neurais) baseado em informações do gradiente. No caso da descida de gradiente, por exemplo, combinando as equações (6.13) e (6.9), temos a atualização do peso $w_{j,i}$ dado pela equação a seguir:

$$w_{j,i}^{\tau+1} = w_{j,i}^{\tau} - \eta \delta_j x_i \quad (6.19)$$

A Eq. (6.19) corresponde ao chamado **treinamento on-line**, onde um exemplo da amostra de treino é apresentado por vez.

Outra possibilidade é fazer o treinamento em **batches** (ou “lotes”), em que um grupo de exemplos é apresentado por vez. Nesse caso, outros algoritmos, como os *quasi-Newton* do Capítulo 4 são mais adequados [107].

A opção por treinamento *on-line* ou em *batches* é uma escolha dos desenvolvedores. Há que se levar em conta que, normalmente, o treinamento em *batches* tem convergência mais rápida, mas é computacionalmente mais caro [106].

6.2 PRÉ-PROCESSAMENTO DE DADOS PARA REDES NEURAI

Antes de utilizar um conjunto de dados para treinar uma rede neural, é aconselhável realizar o pré-processamento dos dados para evitar que ocorram problemas numéricos na otimização.

Quando há variáveis de entrada cujas escalas dos valores típicos diferem significativamente, é útil fazer o **reescalonamento** [104]. É o caso da previsão de carga elétrica, onde as cargas, medidas em MW, costumam ter valores na casa dos milhares, enquanto as temperaturas, medidas em °C, não passam da casa das dezenas. Esta diferença nas magnitudes pode fazer com que a importância relativa das variáveis para o cálculo das saídas não seja estabelecida de forma adequada.

Uma maneira de fazer com que todas as variáveis passem a ter valores similares é a aplicação de uma transformação linear. Tratamos cada variável de maneira independente e, para cada variável x_i , calculamos a média \bar{x}_i e a variância σ_i^2 em relação à amostra de teste, utilizando:

$$\begin{aligned}\bar{x}_i &= \frac{1}{N} \sum_{n=1}^N x_i^n \\ \sigma_i^2 &= \frac{1}{N-1} \sum_{n=1}^N (x_i^n - \bar{x}_i)^2\end{aligned}\tag{6.20}$$

onde $n = 1, \dots, N$ enumeram os exemplos.

O reescalonamento de um valor x_i fica, então:

$$z_i = \frac{x_i - \bar{x}_i}{\sigma_i}\tag{6.21}$$

As variáveis z_i têm média zero e desvio padrão um. No caso de problemas de regressão, é recomendável aplicar a mesma transformação às variáveis de saída [104].

7 SELEÇÃO E COMBINAÇÃO DE ESTAÇÕES METEOROLÓGICAS

Neste capítulo, entramos no problema da combinação de estações meteorológicas.

Falamos das propostas encontradas na literatura, introduzimos uma heurística que tem sido utilizada para ranquear as estações de acordo com a acurácia, além de métodos para combinar estas estações ranqueadas.

Em seguida, apontamos possíveis caminhos para aprimorar os resultados.

7.1 UMA HEURÍSTICA PARA COMBINAÇÃO DE ESTAÇÕES METEOROLÓGICAS

No Capítulo 1, falamos sobre a importância da seleção de estações meteorológicas para a previsão de carga elétrica quando o problema envolve múltiplas zonas de carga em um território grande. Ao lidar com este problema, profissionais têm optado por fixar um número de estações meteorológicas, que serão as mesmas para todas as zonas [36]. Este procedimento é, para Hong, contraintuitivo, porque não leva em conta aspectos essenciais da previsão hierárquica de carga [36]. Considerando a diversidade geográfica, a diversidade demográfica e o perfil do consumidor final, o número de estações não deveria ser obrigatoriamente o mesmo para todas as zonas de carga e deveria haver uma preocupação maior na escolha de quais estações meteorológicas devem ser associadas a cada região.

Percebendo que o problema parecia negligenciado, autores propuseram uma reformulação do problema, que passa a ter como mote a busca pelo melhor subconjunto de estações para um território de interesse [36]. Com esta nova abordagem, pretende-se descobrir quantas estações meteorológicas devem ser usadas em um território e quais são estas estações.

Em linhas gerais, o roteiro de [36] seria o seguinte:

- (i) Avaliar separadamente o desempenho de cada uma das N estações meteorológicas disponíveis, segundo algum critério pré-estabelecido (erro médio das previsões, número de valores faltantes, *etc.*), e ranqueá-las de acordo com este critério;
- (ii) Combinar os dados de temperatura das estações, começando com as melhores ranqueadas;
- (iii) Avaliar as combinações das estações conforme o critério de desempenho estabelecido, e ranquear as combinações.

Para executar este plano, os autores propõem a heurística a seguir. É necessário separar dos dados duas amostras, que chamaremos de A e B , antes de começar. Supondo que se tenha acesso a dados de N estações meteorológicas:

(1) Denote-se a série de temperaturas de cada estação meteorológica como W_i , com $i = 1, \dots, N$. Ajustar um modelo de previsão M , utilizando os dados de temperatura W_i , na amostra A. Prever a carga na amostra B.

(2) Calcular uma medida de erro, como o $MAPE_i$ (*mean absolute percentage error*, erro médio percentual absoluto) na amostra A, para cada estação W_i .

(3) Listar as medidas de erro em ordem crescente. Usar o índice j para re-rotular cada estação meteorológica de acordo com sua colocação na lista. Denotar a j -ésima série de temperatura colocada como $W_j, j = 1, \dots, N$.

(4) Combinar as previsões de carga das $k, k = 1, \dots, N$ estações meteorológicas mais bem posicionadas para criar uma nova previsão. Uma forma simples de combinação poderia ser a média aritmética simples das previsões das k estações mais bem colocadas.

(5) Calcular a medida de erro das previsões combinadas, na amostra B.

(6) O índice k correspondente ao menor erro representa o número de estações meteorológicas a ser selecionado. As estações correspondentes podem ser determinadas observando a listagem do passo (3).

A maior vantagem de um sistema de ranqueamento como este é a redução do espaço de busca. Caso se utilizasse busca exaustiva, n estações levariam a 2^n subconjuntos possíveis. No caso dos dados com os quais iremos trabalhar, 11 estações levariam a 2048, subconjuntos, elevando consideravelmente o custo computacional em relação à heurística.

O algoritmo foi testado inicialmente para dados de um sistema elétrico americano e dados da competição GEFCom2012 [36]. O modelo de previsão M escolhido foi um modelo de regressão linear, o TVB¹. A combinação dos dados de temperatura foi feita usando uma média aritmética simples. Comparada à simples combinação de todas as estações, a heurística melhora a acurácia das previsões, o que os autores atribuem à remoção da restrição do uso das mesmas estações para todas as zonas de carga.

Apesar de os autores mencionarem que fatores como a distância da estação meteorológica até o território da zona de carga, o histórico da estação, a porcentagem de dados faltantes, restrições de orçamento e o perfil do consumidor são importantes e deveriam ser considerados para a ponderação na hora de combinar as estações meteorológicas, a única abordagem testada para combinar os dados de temperatura foi a média aritmética simples; portanto, todas as estações receberam o mesmo peso.

7.2 MAIS ABORDAGENS PARA COMBINAÇÃO DAS TEMPERATURAS

Utilizando a mesma heurística para seleção das estações meteorológicas e usando o mesmo modelo de regressão linear e os dados da GEFCom2012, [37] testaram outras abordagens para combinação dos dados de temperatura além da média aritmética simples,

¹O modelo de regressão *Tao's Vanilla Benchmark*, visto no Capítulo 4.

como uma combinação linear, uma combinação com pesos exponencialmente decrescentes e uma combinação com pesos definidos por algoritmo genético.

Apesar da quantidade de novas formas de combinação introduzidas, o trabalho pouco avança em relação em que se tinha antes. As combinações ainda não levam em conta aspectos práticos, como as características geográficas da região. Também não foram testados outros modelos de previsão além do TVB e tampouco outros dados.

A última publicação em meio científico sobre o tema até este momento foi [38].

Neste trabalho, os autores propõem o uso do ACI (*average coherence index*, ou **índice médio de coerência**), baseado no conceito de *wavelets*, originário da teoria de Processamento de Sinais. A ideia é comparar as frequências das séries temporais de temperaturas e combinar aquelas que sejam mais “próximas” (ou seja, que tenham maior ACI).

Este trabalho vai um pouco além dos demais: os autores experimentam com dados de carga, velocidade do vento, radiação solar e temperatura do ar, de 2004 a 2008, para a Cidade do Panamá. São utilizadas 18 zonas de cargas e 11 estações meteorológicas.

Mais uma vez, é utilizada a heurística introduzida em [36] e o modelo de previsão utilizado também é o TVB. Das estações melhores ranqueadas, obtém-se a média aritmética simples dos dados para chegar a uma estação virtual.

Os autores comparam os resultados do ranqueamento pelo ACI com uma combinação linear, uma combinação exponencial e a média de ambas. Na maioria dos casos, o MAPE obtido utilizando o método proposto é mais baixo que o dos *benchmarks*, o que indica que sua proposta tem melhor desempenho para os dados em questão.

7.3 POSSIBILIDADES DE APRIMORAMENTO PARA A SELEÇÃO E COMBINAÇÃO DE ESTAÇÕES METEOROLÓGICAS

Considerando o estado da arte do problema da seleção e combinação de estações meteorológica, enxergamos espaço para aprimoramento destes modelos, em alguns aspectos sobre os quais discorreremos a seguir.

(1) **Pode-se validar os métodos em novos conjuntos de dados.**

Os trabalhos [36, 37, 38], que abordam a seleção e combinação de estações meteorológicas trabalham com os dados da GEFCom, e o último deles apresenta resultados também para dados do Panamá.

Seria importante experimentar a heurística de seleção de estações e os métodos de combinação em mais bases de dados para validar sua eficiência.

(2) **Podem-se testar outros modelos de previsão.**

Os três trabalhos que citamos utilizam o TVB como modelo de previsão. Este

modelo foi proposto como *benchmark Naive* para problemas de previsão de carga elétrica.

A ideia de um *benchmark Naive* para previsão de carga é interessante, em especial no contexto das competições, como a GefCom. O problema é que o TVB tem pouca utilidade na prática, já que é um modelo muito específico e que já pressupõe qual a forma funcional da relação entre carga e temperatura. Utilizar o modelo na prática seria contraitutivo, porque partiríamos do modelo para os dados; e não o contrário.

Pode-se, então, testar outros modelos de previsão para ter uma ideia de como seria utilizar a heurística de ranqueamento na prática.

(3) Podem-se levar em conta aspectos práticos para fazer as combinações.

Dentre as formas de combinação de estações meteorológicas propostas até então, a maioria são operações aritméticas com pouca contextualização. Não são levados em conta aspectos de ordem prática, como a localização das estações meteorológicas.

Poderíamos experimentar médias ponderadas que atribuem pesos com base em aspectos práticos, como a proximidade da estação meteorológica da zona de carga.

Pretendemos trabalhar com as ideias (1) a (3).

Para a primeira, trabalharemos com os conjuntos de dados de carga e temperatura do NYISO e do NOAA, introduzidos no Capítulo 2.

Para a segunda, experimentaremos redes neurais como modelos de previsão. As redes trazem a vantagem do mapeamento das relações não-lineares entre cargas e temperaturas, razão pela qual acreditamos que trarão resultados interessantes.

Para a terceira, iremos propor, a partir da próxima seção, novas abordagens para combinação de estações meteorológicas.

7.4 NOVAS ABORDAGENS PARA COMBINAÇÃO DE ESTAÇÕES METEOROLÓGICAS

Vimos que os métodos de combinação de estações meteorológicas na literatura consistem, em sua maioria, em operações aritméticas, que não levam em conta fatores de ordem prática que podem influenciar na importância das estações.

Pensando nisso, propomos duas novas abordagens para combinação de dados de estações meteorológicas, em que consideramos fatores de natureza prática.

7.4.1 Método C1: Ponderação pela distância da estação meteorológica até o centro geográfico da zona de carga

Caso pudéssemos instalar uma estação meteorológica dentro do território da zona de carga, a fim de utilizar os dados coletados como entradas dos modelos de previsão,

onde posicionaríamos esta estação? Acreditamos que localizá-la no *centro geográfico* da região permitiria capturar o perfil meteorológico da área.

Definimos como centro geográfico o centróide da figura bidimensional de uma região na superfície da Terra, projetada radialmente ao nível do mar (ou em uma superfície geóide²) [109].

Como primeiro método de combinação, propomos uma combinação linear das temperaturas de cada estação meteorológica, às quais são atribuídos pesos iguais ao inverso da distância da respectiva estação até o centro geográfico da zona de carga. Assim, as estações mais próximas do centro geográfico recebem maiores pesos.

Devemos dar atenção especial à forma de se calcularem as distâncias das estações até o centro da zona de carga. Devido a distorções causadas por projeções em uma superfície plana, uma linha reta em um mapa não será necessariamente a distância mais curta entre dois pontos [110]. Por isso, preferimos trabalhar com a **distância do grande círculo** (também chamada **ortodromia**). Esta medida é tomada ao longo da superfície da esfera, e não traçando uma linha reta que atravessa o interior da esfera. É sabido que, devido ao fato de a Terra ser aproximadamente esférica, a ortodromia entre dois pontos consegue aproximar a distância entre pontos na superfície terrestre com erro de no máximo 0,5%, considerando 1' de latitude igual a uma milha náutica internacional (1852 m) [111].

Sejam P e Q dois pontos em uma esfera, e ϕ_1, ϕ_2 e λ_1, λ_2 suas latitudes e longitudes. Seja $d_{P,Q}$ a distância do grande círculo entre estes pontos, e R o raio da Terra. Então,

$$d_{P,Q}^2 = 2R^2 - 2R^2 \cos \phi_1 \cos \phi_2 \cos(\Delta\lambda) - 2R^2 \sin \phi_1 \sin \phi_2. \quad (7.1)$$

Para a derivação deste cálculo, ver [112].

Vamos à combinação de estações meteorológicas. Seja $d_{t,c}$ a distância (do grande círculo) entre as coordenadas da estação t até o centro geográfico da zona de carga c . O peso atribuído à estação z será:

$$w_t = \frac{1}{d_{t,c}} \quad (7.2)$$

Sendo T_t a série de temperaturas da estação meteorológica t , a combinação dos dados de n estações será dada por:

$$T_{comb} = \frac{\sum_{t=1}^n (w_t T_t)}{\sum_{t=1}^n (w_t)} \quad (7.3)$$

7.4.2 Método C2: Ponderação pela correlação não-linear entre cargas e temperaturas

A intensidade da correlação linear entre duas variáveis x e y é dada pelo coeficiente de correlação linear de Pearson, definido no Capítulo 5, na Eq. (5.4).

²Também podemos entender como o ponto de equilíbrio ou centro de gravidade desta figura.

Os valores de r_{xy} variam entre -1 e 1. Um valor de -1 significa que a correlação linear é fortemente negativa, ou seja, enquanto os valores de uma variável crescem, os da outra decrescem, linearmente. Já um valor de 1 significa que a correlação linear é fortemente positiva, isto é, enquanto os valores de uma variável crescem ou decrescem, o da outra caminham, linearmente, juntos.

Falamos nos Capítulos 1 e 2, e mostramos através de exemplos, que as cargas mais altas costumam ser observadas quando há registros de temperatura extremas, e a correlação entre cargas e temperaturas costuma ter um formato não-linear. O coeficiente de correlação de Pearson serve para correlações lineares e, logo, não seria adequado para mensurar a intensidade da correlação entre carga e temperatura.

Pensando na necessidade de uma medida para a correlação não-linear, [113] desenvolveram a medida **nlcor** (*nonlinear correlation coefficient*). O **nlcor** é uma medida entre 0 e 1 e, quanto mais próximo de 1 é seu valor, maior a intensidade da correlação não-linear.

Para chegar ao **nlcor**, os dados são divididos em segmentos, de modo a maximizar a média dos coeficientes de correlação linear de Pearson de cada segmento, em valor absoluto.

Além de maximizar a média citada anteriormente, a segmentação é também otimizada de modo excluir segmentos ruidosos. Isto é feito descartando o coeficiente de Pearson dos segmentos para os quais o valor- p do coeficiente é maior que 5%³.

O valor do **nlcor** será a média dos valores absolutos dos coeficientes de correlação linear de Pearson estatisticamente significativos (*i.e.* valor- p menor que 5%) de cada segmento em que foram divididos os dados.

Consideremos os dados simulados da Figura 7.1, onde existe correlação não-linear cíclica.

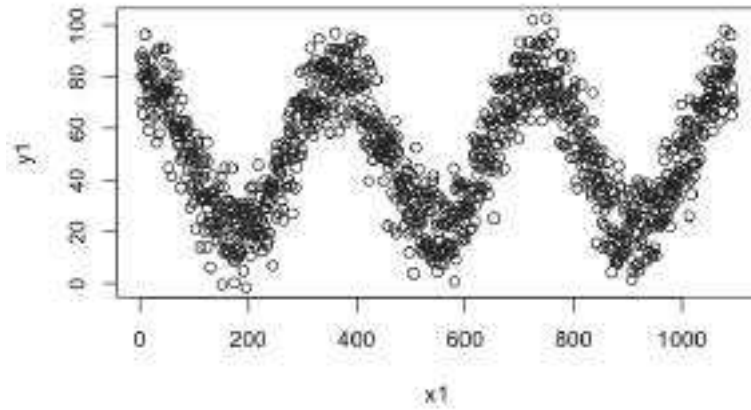
O coeficiente de correlação linear de Pearson para estes dados tem intensidade 0,008 (próximo de zero, como esperado).

O **nlcor** segmenta os dados em seis partes, como na Figura 7.2, calcula o coeficiente de correlação linear de Pearson em cada segmento e combina, através da média dos módulos, os valores estatisticamente significativos, chegando a uma intensidade de correlação (não-linear) de 0,8689, o que parece bastante razoável considerando que a medida pode assumir valores entre 0 e 1.

A implementação do **nlcor** está disponível como funções de um pacote do *software* livre R. O pacote foi submetido ao repositório oficial de pacotes do R, o CRAN (*The Comprehensive R Archive Network*). O código fonte, o qual analisamos para discorrer sobre o método está disponível em [113]. Propomos um método de combinação das estações

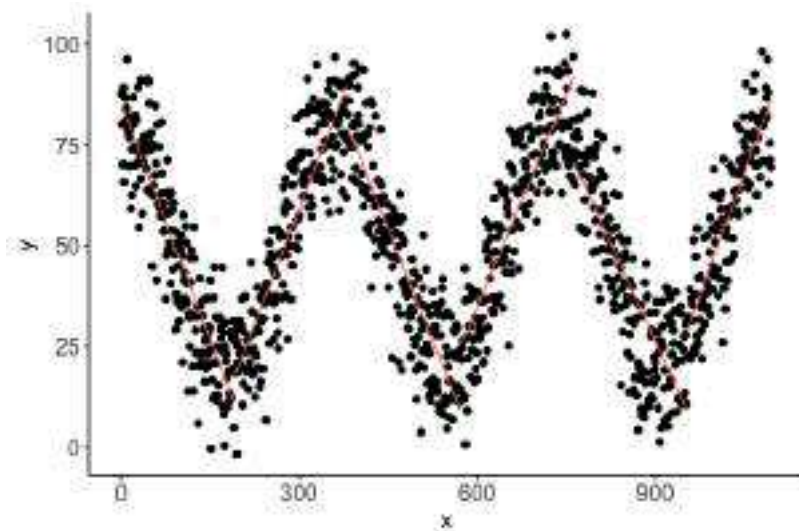
³O valor- p do coeficiente de Pearson é a probabilidade de que o coeficiente de correlação linear seja, na verdade, igual a zero.

Figura 7.1 – Dados simulados com correlação não-linear cíclica



Fonte: [114]

Figura 7.2 – Segmentação dos dados simulados da Figura 8.1 usando o nlcor



Fonte: [114]

meteorológicas que consiste em uma combinação linear dos dados de temperatura das estações meteorológicas, cujo peso w_t de cada estação t é o *nlcor* entre seus dados de temperatura e a série de cargas da zona z :

$$w_t = nlcor_{z,t} \quad (7.4)$$

Recebem maior peso as estações cujos dados de temperatura tenham maior intensidade de correlação não-linear com os dados de carga. A combinação será dada pela Eq. (7.4).

8 SIMULAÇÕES E RESULTADOS

Neste capítulo, descrevemos as simulações que realizamos utilizando os dados, os métodos e os modelos apresentados nos capítulos anteriores, e discutimos os resultados.

8.1 ROTEIRO DAS SIMULAÇÕES

Primeiramente, explicamos o roteiro das simulações, com ênfase nos métodos e, eventualmente, recuperando conceitos dos capítulos anteriores.

Denotamos as cargas observadas e previstas para o instante t , respectivamente, por y_t e \hat{y}_t . Nosso interesse está na previsão de **perfis de carga**: na vigésima-quarta hora de um dia, produzimos a série de previsões \hat{y}_{t+k} , para $k = 1, \dots, 24$, ou seja, o perfil do dia seguinte.

8.1.1 Heurística para seleção e combinação de estações meteorológicas

Para ranquear e selecionar as estações meteorológicas, um dos focos desta tese, utilizamos a heurística proposta em [36], apresentada no Capítulo 7.

Para combinar as estações meteorológicas ranqueadas, testamos os dois métodos definidos no Capítulo 7, C1 e C2.

Testamos também um método que chamaremos de C3, dado pela média aritmética simples das temperaturas das estações meteorológicas. Utilizado em alguns trabalhos na literatura, como visto no Capítulo 7, o método é dado pela Equação que segue:

$$W_{comb} = \frac{\sum_{i=1}^N W_i}{N} \quad (8.1)$$

C1 e C2 são inéditos e propostos por nós, enquanto o uso de C3 já foi amplamente documentado.

8.1.2 Métodos de séries temporais

Como *benchmarks*, utilizamos os métodos de séries temporais que apresentamos no Capítulo 4:

- Um previsor ingênuo (*naive*) sazonal com período $m = 24$ horas (chamaremos de **NF24**);
- Um previsor ingênuo (*naive*) sazonal com período $m = 168$ horas (chamaremos de **NF168**);
- O amortecimento exponencial de Holt-Winters-Taylor para duas sazonalidades, diária e semanal (chamaremos de **HWT**);

- Modelos de Box & Jenkins.

As formulações dos modelos de Box & Jenkins que apresentamos no Capítulo 4 não permitem prever de maneira direta para horizontes de previsão à frente de $t + 1$. Para fazer as previsões de perfil, ajustamos 24 modelos, um para cada hora do dia, utilizando-os para prever a hora seguinte (*i.e.*, a mesma hora do dia, no próximo dia).

8.1.3 Redes neurais *Feedforward* (FFNN)

O principal método para o qual gostaríamos de testar a heurística de seleção e combinação, e os métodos de combinação, são as redes neurais *feedforward* (FFNN).

A arquitetura que utilizamos foi uma rede de uma camada oculta ¹.

As saídas da rede são as 24 cargas que compõem o perfil do dia seguinte (dia $d + 1$).

As entradas são:

- (i) 24 cargas do dia d ;
- (ii) 24 cargas do dia $d - 1$;
- (iii) 7 variáveis *dummy*, para os dias da semana, para $d + 1$;
- (iv) 4 variáveis *dummy*, para as estações do ano, para $d + 1$;
- (v) 24 temperaturas do dia $d - 1$;
- (vi) 24 temperaturas do dia d ;
- (vii) 24 temperaturas do dia $d + 1$.

Esta arquitetura é inspirada em [30], e engloba os principais aspectos das séries de carga e temperatura, averiguados na análise exploratória do Capítulo 2.

Os grupos de entradas (i) a (iv) modelam a sazonalidade diária e a tendência mais recente da curva de carga. Os demais incluem os efeitos da temperatura sobre a carga.

É importante mencionar que, embora utilizemos temperaturas do dia $d + 1$ no modelo, na prática é impossível ter acesso a estes dados. Como as previsões são feitas no dia d , a série de temperaturas para $d + 1$ nem sequer foi observada! Para contornar o problema, costumam-se utilizar **previsões de temperatura**. Como estas previsões contêm erros (afinal, são previsões), fizemos uma análise de sensibilidade, que mostraremos mais adiante, injetando vários níveis de ruído nos dados de temperatura. A intenção é verificar como erros nos dados de previsão de temperatura (elaboradas *in company* ou adquiridas de terceiros) afetam o desempenho de nossa rede.

¹Testamos até quatro camadas, mas não percebemos benefícios em utilizar mais do que uma.

Experimentamos a **Floresta Aleatória** como método para reduzir o número de entradas. No entanto, como há multicolinearidade entre as variáveis devido à dependência temporal dos dados, o método não foi benéfico.

8.1.3.1 Ajuste dos hiperparâmetros das FFNNs

A primeira fase na implantação de uma FFNN é o **ajuste dos hiperparâmetros**, uma gama de configurações que tem efeito sobre o desempenho das redes e cujos valores devem ser definidos pelos usuários.

Adotamos uma tática sugerida por [106]. Começamos decidindo quais hiperparâmetros gostaríamos de ajustar², e mantivemos fixos os valores dos demais. Do hiperparâmetro mais para o menos importante (com base em uma lista recomendada pelo autor), testamos uma quantidade fixa de valores aleatórios para este hiperâmetro, sem modificar os demais. Escolhemos o valor que leva ao melhor desempenho³. Este procedimento tem se mostrado mais eficiente que a busca em grade para escolha de hiperparâmetros [106].

Pela ordem de importância, escolhemos ajustar os seguintes hiperparâmetros:

- η (**taxa de aprendizado**) - valor inicial: 0,1 (sugerido por [106]); outros: cinco aleatórios da distribuição uniforme entre 10^{-6} e 0,1 (sugerido por [106]);
- B (**tamanho do *batch***) - valor inicial: 32 (*default* do Keras, biblioteca que utilizamos para implementação computacional); outros: 1 (treinamento *on-line*), 64, 128, 256 (conforme sugerido por [106]);
- i (**número de épocas**) - valor inicial: 250; outros: 50 a 500, em intervalos de 50 (estabelecido por nós; verificamos que em mais de 500 épocas o tempo de treinamento chegava a dobrar, mas sem reduzir o erro);
- n_h (**número de neurônios na (s) camada (s) oculta (s)**) - valor inicial: 32 (valor máximo em [30]); outros: aleatórios inteiros de 10 (mínimo de [30]) a 32.;
- λ (**coeficiente de regularização L2**) - valor inicial: 0; outros: 10^{-6} , 10^{-3} e um número da distribuição uniforme entre ambos (baseado em processo de tentativa e erro).

Outros hiperparâmetros foram o **número de camadas ocultas** (uma), o **algoritmo de aprendizado** (Adam) e a **função de ativação** (ReLU). Percebemos que não houve melhora testando outros valores ou funções e fixamos os que estão sinalizados entre parênteses.

²O que pode ser limitado pelos usuários, a depender dos recursos computacionais e de tempo disponíveis.

³Utilizamos o MSE como critério.

8.1.4 Divisão dos dados

Para propósito de elaboração de modelos e comparação dos resultados, dividimos os dados em duas amostras, A e B, cujos usos detalhamos no Quadro 8.2:

Passo	Amostra A (anos de 2015 a 2017)	Amostra B (ano de 2018)
Classificação de estações	Ajuste de hiperparâmetros, treinamento e validação das redes neurais com dados de cada estação meteorológica em separado.	Previsões para elaboração do ranque das estações meteorológicas, pelo MAPE.
Combinação de estações	Ajuste de hiperparâmetros, treinamento e validação das redes neurais com estações combinadas.	Previsões para elaboração do ranque das combinações, pelo MAPE, com e sem ruído nos dados de temperatura. Comparação com os <i>benchmarks</i> .
<i>Benchmark</i>	Otimização das constantes do HWT e determinação da ordem e parâmetros dos modelos de Box & Jenkins.	Previsões para comparação com as redes neurais.

Quadro 8.1 – Divisão dos dados.

Como mostra o quadro, utilizamos os anos de 2015 a 2017 cheios dos conjuntos de dados (36 meses) para ajustar o HWT e os modelos de Box & Jenkins, para ajustar os hiperparâmetros e treinar as redes neurais. Utilizamos o ano de 2018 cheio (12 meses) para comparar o desempenho dos modelos.

Ao ajustar hiperparâmetros ou treinar uma FFNN em um conjunto de dados limitado, é comum utilizar alguma técnica de reamostragem, como a **validação cruzada**, a fim de garantir maior poder de generalização. A validação cruzada do tipo *k-fold* é um dos métodos mais utilizados para este propósito: os dados são embaralhados de maneira aleatória e, então, divididos em *k* grupos. Os dados de cada grupo são tomados como uma amostra *out-of-sample* (o que seria nossa amostra A) e o restante é tomado como amostra *in-sample* (o que seria nossa amostra B). O modelo é treinado nos dados *in-sample* e validado nos dados *out-of-sample*, e a avaliação final será um resumo (por exemplo, a mediana) do desempenho nos *k* grupos.

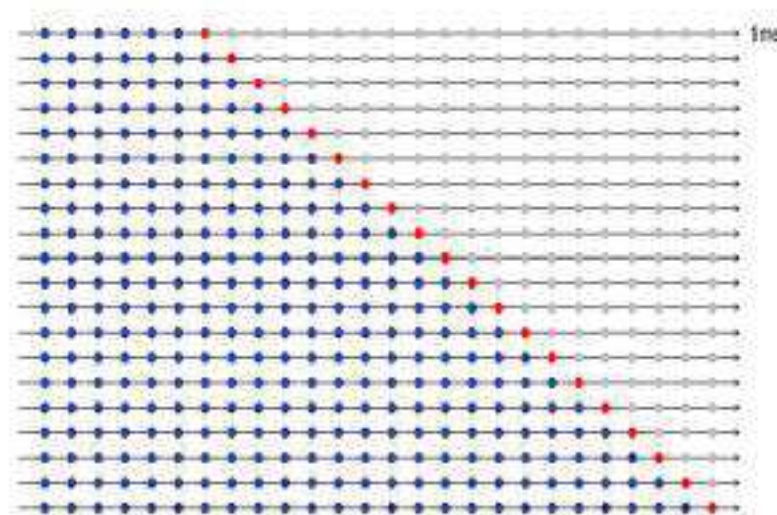
É necessário, no entanto, atentarmos para o fato de nossos dados apresentarem dependência temporal. Como as amostras são provenientes de séries temporais, o embaralhamento do *k-fold* pode gerar uma situação em que utilizamos dados do futuro para treinar o modelo e dados do passado para validar.

No caso de dados de séries temporais, uma tática útil é a **origem móvel de previsão** (do inglês *rolling forecasting origin*). Nesse caso, escolhe-se um ponto no tempo

que corresponde à amostra de validação, e os conjuntos de treino serão os pontos que ocorreram antes [118].

Na Figura 8.1, mostramos um esquema genérico para o procedimento. Pontos azuis representam os conjuntos de treino, e pontos vermelhos são os conjuntos de validação. Nenhuma observação futura será utilizada para treino.

Figura 8.1 – Esquema genérico para *rolling forecasting origin*.



Fonte: [118]

Como dispomos de alguns milhares de pontos de dados, definimos os conjuntos de teste como um mês inteiro de observações (no lugar de uma observação apenas, como é usual). Para cada mês de validação, consideramos os cinco meses anteriores para treino.

Na Tabela 8.1 trazemos a segmentação dos dados da amostra A para treino e validação das FFNN:

Conjunto	Meses de treino	Mês de validação
1	[1,5]	6
2	[1,11]	12
3	[1,17]	18
4	[1,23]	24
5	[1,29]	30
6	[1,35]	36

Tabela 8.1 – Segmentação da amostra A com *rolling forecasting origin*.

8.1.5 Avaliação do desempenho dos modelos

Para treinar as redes neurais e otimizar as constantes do HWT e os modelos de Box & Jenkins, optamos pela minimização do erro médio quadrático (MSE), especialmente com a intenção de minimizar a variância dos erros nas redes neurais.

Para comparar todos os modelos, escolhemos o erro médio percentual absoluto (*MAPE*) que, além de facilmente interpretável, costuma ser a medida padrão na literatura de previsão de carga.

Recapitulando, para uma amostra de N pontos, o *MAPE* é dado por:

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (8.2)$$

8.1.6 Recursos computacionais

Os experimentos foram executados utilizando GPU Nvidia GeForce 1080Ti com 11 GB de RAM para as redes neurais e CPU Intel Intel® Core™ i7-8700K com 3,70GHz para os demais métodos.

As implementações foram escritas em linguagem R [119], utilizando funções dos pacotes **keras** [120], para redes neurais; **forecast** para os modelos de Box & Jenkins [121]; e **geosphere** [122] para as distâncias na superfície terrestre.

Para reprodutibilidade, os códigos estão disponíveis em repositório na plataforma GitHub [123].

8.1.7 Regressão linear múltipla: considerações sobre o modelo TVB

Uma de nossas intenções era trabalhar com o modelo de regressão múltipla proposto por [102] como modelo *naive* para previsão de carga, o Tao's Vanilla Benchmark (TVB), que apresentamos no Capítulo 5.

Recordemos que a carga y no instante t é descrita pelo TVB como:

$$\begin{aligned} y_t = & \beta_0 + \beta_1 \times Trend_t + \beta_2 \times Day_t \times Hour_t + \beta_3 \times Month_t \\ & + \beta_4 \times Month_t \times TMP_t^2 + \beta_5 \times Month_t \times TMP_t^2 \\ & + \beta_6 \times Month_t \times TMP_t^3 + \beta_7 \times Hour_t \times TMP_t \\ & + \beta_8 \times Hour_t \times TMP_t^2 + \beta_9 \times Hour_t \times TMP_t^3 \end{aligned} \quad (8.3)$$

As variáveis mencionadas na Equação (8.3) são:

Trend: variável quantitativa para capturar a tendência local de crescimento ou decréscimo. É atribuído um número natural a cada hora, em ordem crescente, começando do 1;

Hour: hora do dia (qualitativa, 24 classes);

Day: dia da semana (qualitativa, 7 classes);

Month: mês do ano (qualitativa, 12 classes);

TMP: temperatura.

O desempenho deste modelo não foi satisfatório para nossos dados.

Na Tabela 8.2, exibimos os *MAPEs* das previsões, na amostra B (ano de 2018) para as zonas de carga A até G, utilizando dados de cada uma das estações meteorológicas em separado como entradas do TVB:

Zona de carga	Estação meteorológica									
	ALB	ART	BGM	BUF	HPN	JFK	MSS	RME	ROC	SYR
A - WEST	6,27	7,29	6,6	7,03	6,54	6,73	6,56	7,01	7,04	6,67
B - GENESE	7,15	8,73	7,66	8,36	5,97	6,18	5,91	8,57	8,45	8,2
C - CENTRL	5,18	5,68	5,10	5,37	4,63	5,97	5,68	5,68	5,28	5,12
D - NORTH	7,81	9,57	8,49	9,08	6,23	5,78	6,11	9,4	9,3	9,06
E - MHKVL	6,05	7,89	6,68	7,54	5,54	6,2	5,76	7,67	7,55	7,21
F - CAPITL	5,28	6,59	5,96	6,43	5,78	6,13	5,84	6,37	6,45	6,10
G - HUDVL	6,48	6,49	6,34	6,10	6,94	7,22	6,94	6,64	5,98	6,19

Tabela 8.2 – *MAPEs* (%; amostra B) para o TVB, para zonas de carga A a G, por estação meteorológica.

Para se ter uma ideia, o desempenho do previsor ingênuo NF24 para a mesma amostra difere pouco dos melhores desempenhos nesta tabela (destacados em negrito). Para as zonas de carga A a G, o NF24 chega, respectivamente aos *MAPEs* (%) 5,40; 6,64; 5,84; 4,40; 6,25; 5,82; 6,96. Outro *benchmark* linear, o HWT supera o TVB em todos os casos.

O desempenho pouco satisfatório do TVB provavelmente se deve ao fato de que, apesar de ser idealizado como um previsor ingênuo, o modelo é bastante específico e considera efeitos que talvez não sejam importantes para os dados utilizados.

Como não pensamos que seja interessante investir esforços para trabalhar em cima de um modelo aparentemente inadequado para nossos dados, e que apresenta desempenho similar ao um previsor ingênuo muito mais simples, optamos por não nos aprofundarmos no TVB⁴.

⁴Ou em outro modelo de regressão; o que fica como sugestão para futuros trabalhos.

8.2 RESULTADOS

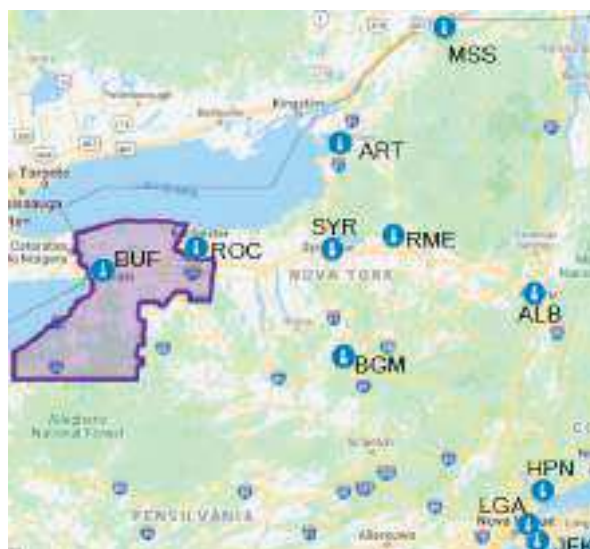
Devido a especificidades de cada zona de carga, faremos as análises em separado para cada uma delas.

8.2.1 Zona de carga A - WEST

Começamos pela zona de carga A - WEST, localizada no extremo oeste do mapa do estado de Nova Iorque, na região de Buffalo, nas planícies Eirie-Ontario.

Na Figura 8.2, vemos que existe uma estação meteorológica dentro do território da zona de carga (estação BUF):

Figura 8.2 – Zona de carga A - WEST em lilás, com estações meteorológicas indicadas em azul.



8.2.1.1 Benchmarks de séries temporais

Primeiramente, implementamos os métodos de séries temporais NF24, NF168, HWT e Box & Jenkins, estabelecidos na literatura.

Para o HWT, utilizamos o algoritmo l-BFGS, apresentado no Capítulo 4, para otimizar as constantes, minimizando o erro médio quadrático (MSE) para os dados *in-sample* (anos de 2015 a 2017).

Já para os modelos de Box & Jenkins, encontramos a ordem e os coeficientes dos modelos utilizando o algoritmo de Hyndman-Khandakar (também discutido no Capítulo 4), que minimiza o Critério de Informação de Akaike (AIC), para encontrar modelos mais parcimoniosos.

Os procedimentos para implementação do HWT e dos modelos de Box & Jenkins foram os mesmos para os demais conjuntos de dados, portanto deixaremos isto implícito nas próximas análises.

Na Tabela 8.3, mostramos os resultados dos *benchmarks* para zona de carga A-WEST, na amostra B (ano de 2018):

Método	MAPE (%)
HWT	3,39
Box & Jenkins	4,26
NF24	5,39
NF168	7,07

Tabela 8.3 – *MAPEs* (%) na amostra B para *benchmarks* univariados (zona de carga A-WEST)

Como esperado, os previsores ingênuos têm pior desempenho. Utilizando o *MAPE* como critério, o HWT se sai melhor que os modelos de Box & Jenkins para A - WEST.

Na Tabela 8.4, mostramos as constantes para o HWT. A constante para atualização do nível, α é igual a zero, sugerindo que o nível médio da carga seja constante para o período de tempo em questão.

A constante ϕ , para ajuste da autocorrelação dos erros de primeira ordem, por outro lado, é muito importante para os dados de carga, como visto em [27]. Esta constante corrige as previsões pelo erro mais recente: se o valor previsto está abaixo do observado, é feita uma correção para cima; se está acima do observado, a correção é feita para baixo. A correção decai exponencialmente conforme o horizonte de previsão.

As constantes δ e ω , para os dados de A - WEST, sugerem que os fatores sazonais diário e semanal influenciam a carga de maneira similar.

α	ϕ	δ	ω
0,00	0,99	0,16	0,13

Tabela 8.4 – Constantes de amortecimento do HWT para zona A - WEST.

8.2.1.2 FFNN e ranqueamento das estações meteorológicas

Como já mencionamos, embora tenhamos posse dos dados de temperatura para o dia $d + 1$, na prática dispomos apenas de previsões para estas temperaturas.

Gostaríamos de avaliar nossos métodos para uma ampla gama de possibilidades, razão pela qual testamos as redes também com ruído injetado nos dados de teste, a fim de simular o efeito dos erros de previsão de temperatura.

Fizemos a injeção de ruído aos dados de temperatura utilizando a Eq. seguinte:

$$W_{t_{\text{ruído}}} = W_t + X \sim N(0, (r \times \mu_W)^2) \quad (8.4)$$

Na Equação (8.4), W_t é a temperatura observada no instante t e X é uma variável aleatória da distribuição normal com média zero e variância igual a $(l \times \mu_W)^2$, onde r é uma constante e μ_W é a média da série de temperaturas.

Para cada estação meteorológica, seguimos estes passos:

- Treinamos 100 FFNNs. Como a inicialização das redes na fase de treinamento pode fazer com que os resultados variem, treinando mais redes conseguimos ter uma noção dessa variabilidade;
- Adicionamos ruído aos dados de teste (ano de 2018);
- Calculamos o MAPE mediano das previsões das 100 FFNNs nos dados de teste, sem ruído e com ruído.

Experimentamos este procedimento para diferentes níveis de ruído, com $l = 0, 1\%, 2\%, 3\%, 4\%, 5\%$.

Fizemos a avaliação final do desempenho *out-of-sample* pelo MAPE mediano dos desempenhos das FFNNs em todos os níveis de ruído, o que julgamos interessante porque cobre uma ampla gama de possibilidades advindas: (1) da natureza das redes neurais, em que a variabilidade surge como consequência de palpites iniciais, e (2) do ruído nos dados, por conta dos erros nas previsões meteorológicas.

A exemplo do que decidimos para os *benchmarks*, para que o texto fique mais conciso, omitimos esta explicação nas seções seguintes, já que o procedimento é o mesmo.

Os resultados para os 11 conjuntos de FFNNs, cada um utilizando dados de uma estação meteorológica estão dispostos na Tabela a seguir:

Estação	MAPE (%)						MAPE mediano (todos os l)
	l = 0	l = 1%	l = 2%	l = 3%	l = 4%	l = 5%	
ALB	3,02	3,10	3,32	3,59	3,92	4,36	3,51
ART	2,80	2,87	3,09	3,39	3,78	4,21	3,29
BGM	3,08	3,15	3,33	3,58	3,84	4,34	3,57
BUF	2,51	2,55	2,77	3,05	3,40	3,83	2,97
HPN	3,32	3,39	3,58	3,87	4,20	4,57	3,73
JFK	3,75	3,80	3,98	4,21	4,52	4,95	4,14
LGA	3,59	3,52	3,71	3,95	4,27	4,76	3,97
MSS	3,52	3,54	3,69	3,96	4,29	4,75	3,92
RME	3,39	3,36	3,54	3,77	4,07	4,50	3,73
ROC	2,55	2,60	2,79	3,06	3,40	3,81	2,97
SYR	3,16	3,11	3,28	3,53	3,85	4,25	3,52

Tabela 8.5 – MAPEs medianos na amostra B, para 100 execuções de FFNNs, por estação meteorológicas, por nível de ruído (dados da zona de carga A - WEST).

Agora, começando a heurística de seleção e combinação de estações, construímos o ranque das estações, partindo daquela com menor MAPE mediano (última coluna da Tabela 8.5). Mostramos o ranque na Tabela 8.6, onde também exibimos a distância de cada estação ao centróide da zona de carga, além do nlor entre os dados de carga e temperatura:

Estação	MAPE mediano (%)	Distância até o centróide de A - WEST	nlor com dados de A - WEST
BUF	2,97	0,32	0,63
ROC	2,97	0,90	0,62
ART	3,29	2,54	0,59
ALB	3,51	3,90	0,57
SYR	3,52	2,07	0,61
BGM	3,57	2,19	0,60
HPN	3,73	4,41	0,54
RME	3,73	2,66	0,59
MSS	3,92	3,92	0,59
LGA	3,97	4,44	0,52
JFK	4,14	4,59	0,51

Tabela 8.6 – Ranque das estações meteorológicas para A - WEST (MAPEs medianos na amostra B para FFNNs).

A Tabela 8.6 sugere que as estações mais próximas do centróide de A - WEST, e com maior nlor, têm melhor desempenho para as FFNNs, quando seus dados são utilizados.

Isto fica mais claro quando observamos os diagramas de dispersão das Figuras 8.3 e 8.4, que explicitam que as estações BUF e ROC, de menor *MAPE*, são também as mais próximas do centróide e com maior *n*lcor:

Figura 8.3 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e distância até o centróide de A - WEST.

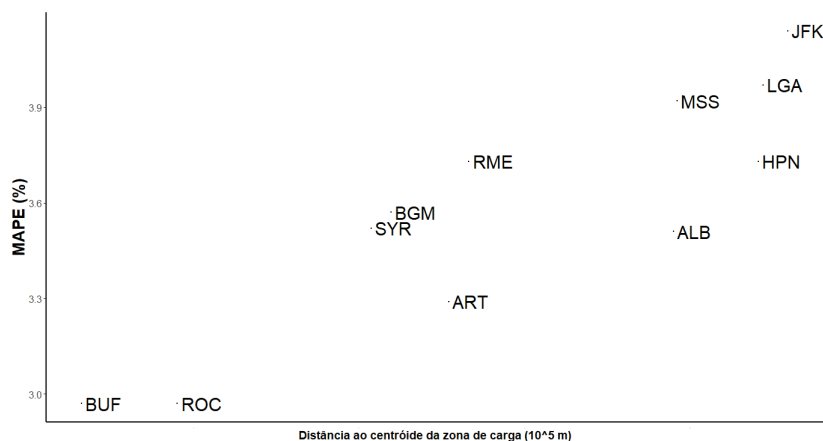
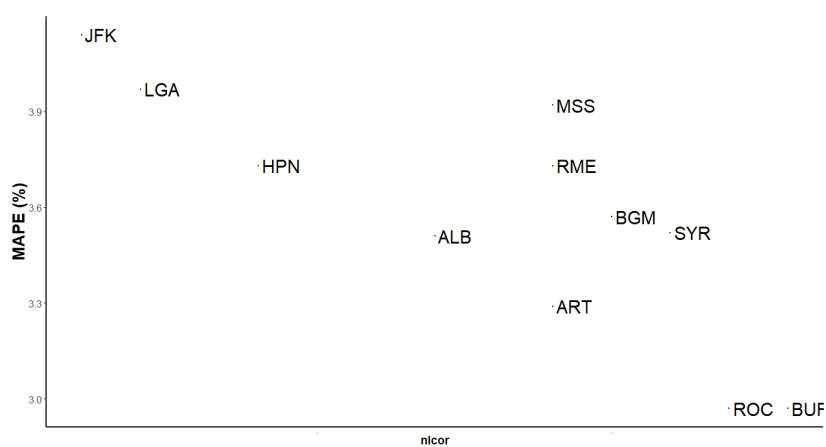


Figura 8.4 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e *n*lcor das cargas de A - WEST.



8.2.1.3 Combinação das estações meteorológicas

Utilizando os métodos C1, C2 e C3, combinamos os dados das estações meteorológicas para utilizar como entradas das redes neurais. Para cada combinação, treinamos, mais uma vez 100 FFNNs.

Na Tabela 8.7, listamos os *MAPEs* medianos destas redes, na amostra B, para uma até 11 estações meteorológicas, combinadas pela ordem do ranque da Tabela 8.6:

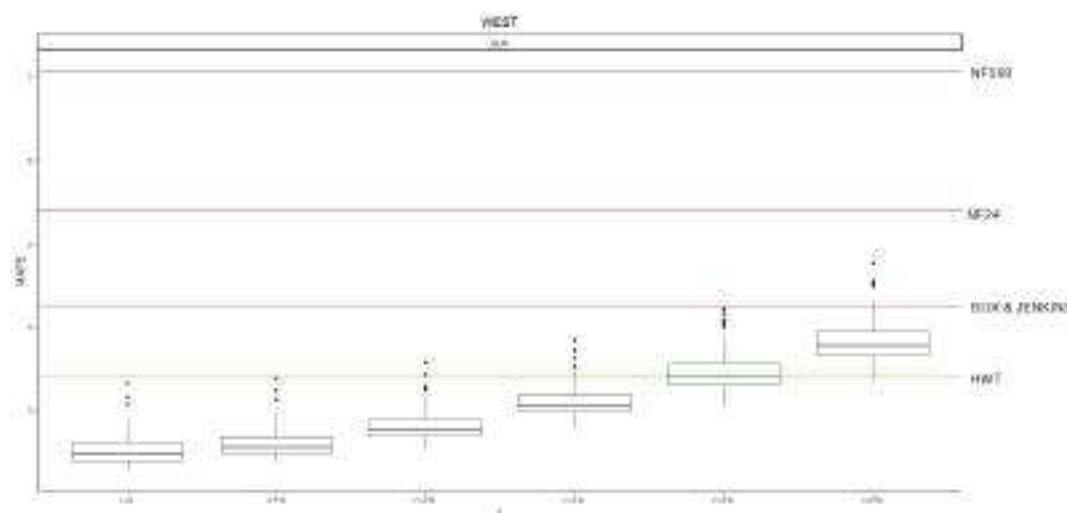
Número de estações	Estações	MAPE mediano (%)		
		C1	C2	C3
1	BUF	2,97	2,97	2,97
2	BUF + ROC	2,99	3,08	3,10
3	BUF + ROC + ART	3,02	3,27	3,24
4	BUF + ROC + ART + ALB	3,01	3,19	3,21
5	BUF + ROC + ART + ALB + SYR	3,04	3,19	3,13
6	BUF + ROC + ART + ALB + SYR + BGM	2,97	3,17	3,15
7	BUF + ROC + ART + ALB + SYR + BGM + HPN	3,00	3,13	3,14
8	BUF + ROC + ART + ALB + SYR + BGM + HPN + RME	3,02	3,19	3,18
9	BUF + ROC + ART + ALB + SYR + BGM + HPN + RME + MSS	3,03	3,24	3,23
10	BUF + ROC + ART + ALB + SYR + BGM + HPN + RME + MSS + LGA	3,02	3,30	3,26
11	BUF + ROC + ART + ALB + SYR + BGM + HPN + RME + MSS + LGA + JFK	3,08	3,31	3,32

Tabela 8.7 – *MAPEs* medianos (amostra B; ano de 2018) para 100 FFNNs, utilizando os métodos C1, C2 e C3 para gerar as temperaturas de entrada das redes (zona A - WEST).

Para A - WEST, o mais interessante parece ser escolher uma estação meteorológica, apenas (BUF). Isto parece razoável, já que havíamos notado que esta estação está dentro do território da zona de carga (e é a única). Também é interessante do ponto de vista prático, já que pode-se ter um bom desempenho sem depender de dados de múltiplas fontes.

Na Figura 8.5, mostramos os resultados das redes neurais para a estação BUF, em comparação com os *benchmarks*, para A - WEST.

Figura 8.5 – *Boxplots* para os *MAPEs* das redes neurais (estação BUF) e dos *benchmarks* de séries temporais.



Quando há pouco ou nenhum ruído, o desempenho das redes neurais é bastante superior. Alguma piora começa a ocorrer à medida que o nível de ruído aumenta. A partir de 4% de ruído, o HWT passa a se mostrar competitivo, o que definitivamente ocorre com 5%, quando o *benchmark* supera o resultado da maior parte das redes.

Os resultados para previsões de carga de A - WEST no ano de 2018 utilizando FFNN com a melhor estação meteorológica, e o dos *benchmarks*, podem ser vistos na Tabela 8.8:

		Ruído					
FFNN		0	1%	2%	3%	4%	5%
(estação BUF)							
	mínimo	2,27	2,36	2,54	2,77	3,05	3,35
	mediana	2,47	2,55	2,77	3,05	3,40	3,83
	máximo	3,33	3,39	3,57	3,84	4,21	4,77
HWT			3,39				
Box & Jenkins			4,26				
NF24			5,39				
NF168			7,07				

Tabela 8.8 – *MAPEs* (amostra B; ano de 2018) para FFFNs e *benchmarks*, dados de A - WEST.

Na Figura 8.6, destacamos a estação BUF no mapa:

Figura 8.6 – Zona de carga A - WEST em lilás, com estações meteorológicas indicadas em azul, e a estação BUF, de melhor desempenho, em verde.



8.2.2 Zona de carga B - GENESE

A zona de carga B - GENESE, na região do Condado de Genesee, é composta de duas regiões disjuntas. Como vimos no Capítulo 2, uma parte localiza-se majoritariamente nas Planícies Eirie-Ontario e outra no Planalto Apalache, regiões fisiográficas com características distintas. Além disso, na Figura 8.7, vemos que há uma estação meteorológica disponível dentro do território da zona de carga (ROC).

Figura 8.7 – Zona de carga B - GENESE em verde, com estações meteorológicas indicadas em azul.



8.2.2.1 Benchmarks de séries temporais

Mostramos na Tabela 8.9, os resultados dos *benchmarks* para zona de carga B-GENESE, na amostra B (ano de 2018):

Método	MAPE (%)
HWT	4,15
Box & Jenkins	5,31
NF24	6,64
NF168	8,75

Tabela 8.9 – *MAPEs* (%) na amostra B para *benchmarks* univariados (zona de carga B-GENESE)

De forma similar ao que ocorre com A - WEST, o HWT tem o melhor desempenho, seguido dos modelos de Box & Jenkins e dos previsores ingênuos.

Exibimos na Tabela 8.10, as constantes para o HWT. Os resultados também se assemelham ao que havíamos observado para A - WEST, com a constante para atualização

do nível, α próxima de zero (*i.e.* nível constante, sem tendência), constante ϕ , para ajuste da autocorrelação dos erros de primeira ordem com bastante relevância, e constantes δ e ω mostrando que os fatores sazonais diário e semanal parecem exercer influência com intensidade próxima.

α	ϕ	δ	ω
0,02	1,00	0,14	0,15

Tabela 8.10 – Constantes de amortecimento do HWT para zona B - GENESE.

8.2.2.2 FFNN e ranqueamento das estações meteorológicas

Para a zona de carga B - GENESE, dispomos os resultados para os 11 conjuntos de FFNNs, na seguinte Tabela:

Estação	MAPE (%)						MAPE mediano (todos os l)
	l = 0	l = 1%	l = 2%	l = 3%	l = 4%	l = 5%	
ALB	3,63	3,69	3,83	4,02	4,26	4,54	3,99
ART	3,32	3,36	3,47	3,64	3,86	4,10	3,61
BGM	3,45	3,49	3,57	3,70	3,89	4,13	3,68
BUF	3,18	3,21	3,30	3,44	3,69	3,77	3,41
HPN	4,61	4,68	4,77	4,94	5,18	5,44	4,94
JFK	4,16	4,22	4,36	4,54	4,76	5,01	4,53
LGA	4,50	4,57	4,71	4,90	5,12	5,39	4,90
MSS	3,49	3,57	3,72	3,93	4,18	4,50	3,87
RME	3,48	3,53	3,63	3,80	3,98	4,23	3,79
ROC	2,87	2,94	3,10	3,32	3,59	3,87	3,31
SYR	3,09	3,12	3,21	3,38	3,57	3,80	3,39

Tabela 8.11 – MAPEs medianos na amostra B para 100 execuções de FFNNs, por estação meteorológicas, por nível de ruído (dados da zona de carga B - GENESE).

A partir dos resultados da última coluna da Tabela 8.11, construímos o ranque das estações meteorológicas, que dispomos juntamente com a distância de cada estação ao centróide da zona de carga, além do *n*lcor entre os dados de carga e temperatura na Tabela 8.12. Como B - GENESE é composta de duas regiões disjuntas, utilizamos como valor para distância até o centróide a média das distâncias até os centróides das duas regiões.

Estação	MAPE mediano (%)	Distância até o centróide de B - GENESE	<i>n</i>lcor com dados de B - GENESE
ROC	3,31	0,66	0,63
SYR	3,39	1,33	0,61
BUF	3,41	1,06	0,61
ART	3,61	1,91	0,60
BGM	3,68	1,56	0,60
RME	3,79	1,91	0,60
MSS	3,87	3,31	0,60
ALB	3,99	3,15	0,58
JFK	4,53	4,00	0,52
LGA	4,90	3,84	0,53
HPN	4,94	3,78	0,55

Tabela 8.12 – Ranque das estações meteorológicas para B - GENESE (MAPEs medianos na amostra B para FFNNs).

Os resultados da Tabela 8.12 guardam alguma semelhança com o que ocorre para a estação A - WEST, com as estações mais próximas do centróide de A - WEST, e com maior *n*lcor, apresentando melhor desempenho para as FFNNs.

Os diagramas de dispersão das Figuras 8.8 e 8.9, destacam o desempenho das estações que as estações ROC, BUF e SYR, as mais próximas do centróide e com maior *n*lcor apresentando menor *MAPE*:

Figura 8.8 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e distância até o centróide de B - GENESE.

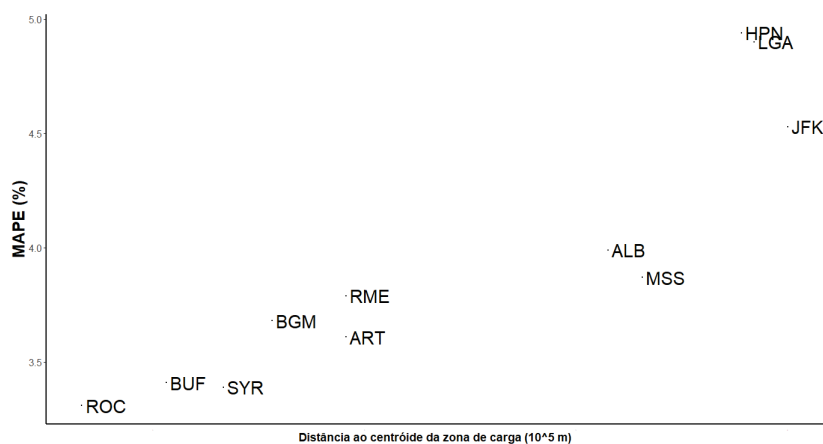
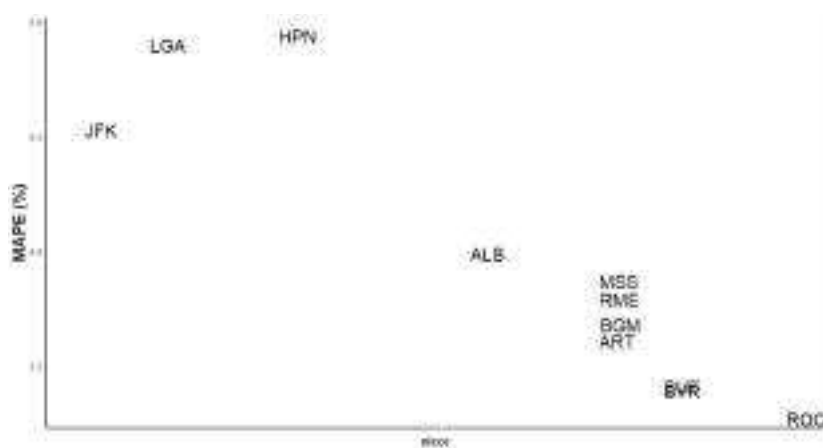


Figura 8.9 – Diagrama de dispersão entre MAPE mediano na amostra B das FFNN e nlor das cargas de B - GENESE.



8.2.2.3 Combinação das estações meteorológicas

Apresentamos na Tabela 8.13, os *MAPEs* medianos das redes cujos dados temperatura utilizados nas entradas são obtidos pelos métodos C1, C2 e C3, para uma a 11 estações meteorológicas, combinadas pela ordem do ranque da Tabela 8.12:

Número de estações	Estações	MAPE mediano (%)		
		C1	C2	C3
1	ROC	3,31	3,31	3,31
2	ROC + SYR	2,91	3,13	3,14
3	ROC + SYR + BUF	2,95	3,08	3,10
4	ROC + SYR + BUF + ART	2,96	3,13	3,12
5	ROC + SYR + BUF + ART + BGM	2,92	3,13	3,11
6	ROC + SYR + BUF + ART + BGM + RME	3,05	3,21	3,19
7	ROC + SYR + BUF + ART + BGM + RME + MSS	3,05	3,20	3,18
8	ROC + SYR + BUF + ART + BGM + RME + MSS + ALB	3,06	3,23	3,26
9	ROC + SYR + BUF + ART + BGM + RME + MSS + ALB + JFK	3,04	3,30	3,32
10	ROC + SYR + BUF + ART + BGM + RME + MSS + ALB + JFK + LGA	3,12	3,35	3,29
11	ROC + SYR + BUF + ART + BGM + RME + ALB + JFK + LGA + HPN	3,10	3,37	3,43

Tabela 8.13 – *MAPEs* medianos (amostra B; ano de 2018) para 100 FFNNs, utilizando os métodos C1, C2 e C3 para gerar as temperaturas de entrada das redes (estação B - GENESE).

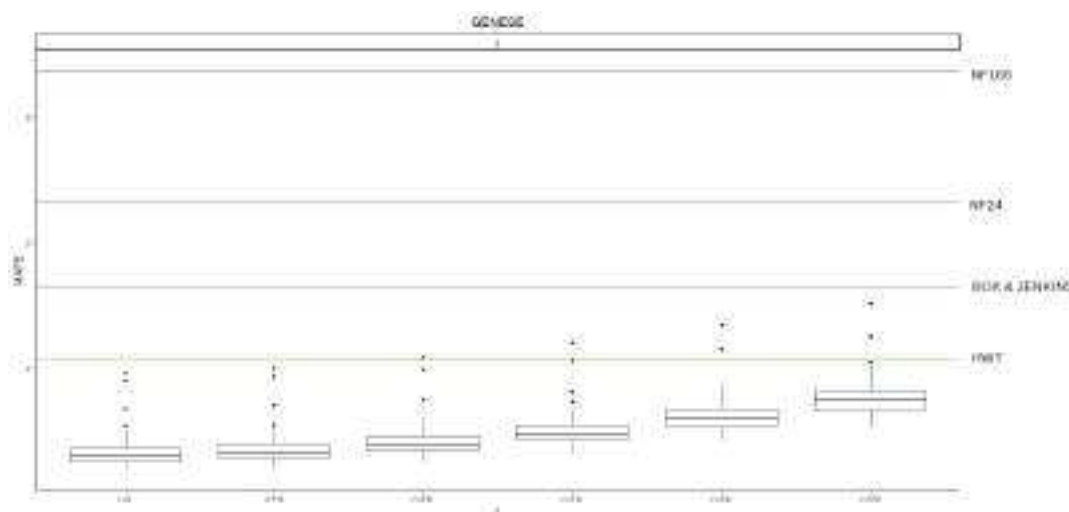
No caso de B - GENESE, combinar estações meteorológicas trouxe benefícios para a acurácia mediana das previsões. As estações ROC, que se localiza dentro do território da zona de carga, e SYR, próxima, que haviam apresentado melhores resultados individuais, combinadas produziram resultados ainda melhores, reduzindo o *MAPE* de 3,31% para 2,91% quando combinadas utilizando o método C1 (ponderação pelo inverso da distância até o centróide).

As estações escolhidas são a primeira e a terceira mais próximas do centróide de

B - GENESE, e combinar com BUF, que é a segunda mais próxima, não melhorou as previsões.

Na Figura 8.10, mostramos os resultados das redes neurais combinando via método C1 ROC e SYR, em comparação com os *benchmarks*, para B - GENESE.

Figura 8.10 – *Boxplots* para os *MAPEs* das redes neurais (combinação C1 de ROC e SYR) e dos *benchmarks* de séries temporais.



O desempenho das redes neurais piora conforme se adiciona ruído. Porém, para B - GENESE, é superior ao dos *benchmarks* lineares em praticamente todos os casos, exceto por seis (de 400) redes, que ocorrem em níveis mais altos de ruído. A combinação C1 de ROC e SYR produz, portanto, resultados consideravelmente robustos.

Os resultados para previsões de carga de B - GENESE no ano de 2018 utilizando FFNNs com as melhores estações meteorológicas e os melhores métodos de combinação, e o dos *benchmarks*, podem ser vistos na Tabela 8.14:

	Ruído					
FFNN	0	1%	2%	3%	4%	5%
(C1, estações ROC e SYR)						
mínimo	2,37	2,40	2,50	2,65	2,85	3,07
mediana	2,59	2,64	2,78	2,97	3,19	3,48
máximo	3,92	4,00	4,17	4,41	4,69	5,03
HWT		4,15				
Box & Jenkins		5,31				
NF24		6,64				
NF168		8,75				

Tabela 8.14 – *MAPEs* (amostra B; ano de 2018) para FFNNs e *benchmarks*, dados de B - GENESE.

Na Figura 8.11, destacamos as estações ROC e SYR no mapa:

Figura 8.11 – Zona de carga B - GENESE em verde, com estações meteorológicas indicadas em azul, e as estações ROC e SYR, de melhor desempenho, em lilás.



8.2.3 Zona de carga C - CENTRL

C - CENTRL é uma zona de carga que se localiza próximo ao centro do estado de Nova Iorque, na região dos Apalaches. Existem duas estações meteorológicas disponíveis em seu território (BGM e SYR), como vemos na Figura 8.12.

Figura 8.12 – Zona de carga C - CENTRL em azul claro, com estações meteorológicas indicadas em azul escuro.



8.2.3.1 *Benchmarks* de séries temporais

Mostramos na Tabela 8.15, os resultados dos *benchmarks* para zona de carga C - CENTRL, na amostra B (ano de 2018):

Método	MAPE (%)
HWT	3,75
Box & Jenkins	4,75
NF24	5,84
NF168	7,50

Tabela 8.15 – *MAPEs* (%) na amostra B para *benchmarks* univariados (zona de carga C - CENTRL)

Mais uma vez, o HWT tem o melhor desempenho e, logo depois, vêm os modelos de Box & Jenkins e os previsores ingênuos.

A Tabela 8.16 traz as constantes para o HWT, e aqui temos alguma diferença para as outras duas zonas de carga. Além de a constante ϕ contribuir um pouco menos (ainda assim sendo bastante importante), o fator sazonal semanal parece ser razoavelmente mais importante que o diário, como sugerem as constantes δ e ω .

α	ϕ	δ	ω
0,00	0,90	0,12	0,18

Tabela 8.16 – Constantes de amortecimento do HWT para zona C - CENTRL.

8.2.3.2 FFNN e ranqueamento das estações meteorológicas

Na Tabela que segue, apresentamos os resultados para os 11 conjuntos de FFNNs para a zona de carga C - CENTRL:

Estação	MAPE (%)						MAPE mediano (todos os l)
	l = 0	l = 1%	l = 2%	l = 3%	l = 4%	l = 5%	
ALB	3,17	3,21	3,32	3,48	3,68	3,94	3,49
ART	3,04	3,10	3,22	3,41	3,65	3,92	3,40
BGM	2,68	2,72	2,84	3,03	3,24	3,47	2,99
BUF	2,88	2,91	3,02	3,20	3,41	3,63	3,15
HPN	3,22	3,29	3,43	3,62	3,87	4,13	3,56
JFK	3,63	3,71	3,86	4,05	4,31	4,58	4,01
LGA	3,41	3,45	3,59	3,76	3,97	4,24	3,76
MSS	3,21	3,24	3,35	3,54	3,76	4,04	3,49
RME	2,67	2,71	2,83	3,00	3,20	3,46	2,93
ROC	2,74	2,77	2,86	3,02	3,21	3,44	3,00
SYR	2,71	2,74	2,87	3,08	3,36	3,64	3,05

Tabela 8.17 – MAPEs medianos na amostra B, para 100 execuções de FFNNs, por estação meteorológicas, por nível de ruído (dados da zona de carga C - CENTRL).

Com os resultados da última coluna da Tabela 8.17, construímos o ranque das estações meteorológicas. Apresentamos a classificação, a distância de cada estação ao centróide da zona de carga e o nlor entre os dados de carga e temperatura na Tabela 8.18.

Estação	MAPE mediano (%)	Distância até o centróide de C - CENTRL	nlor com dados de C - CENTRL
RME	2,93	1,24	0,67
BGM	2,99	0,72	0,67
ROC	3,00	1,00	0,66
SYR	3,05	0,73	0,68
BUF	3,15	1,72	0,65
ART	3,40	1,63	0,65
ALB	3,49	2,35	0,65
MSS	3,49	2,98	0,65
HPN	3,56	2,99	0,63
LGA	3,76	3,09	0,60
JFK	4,01	3,25	0,58

Tabela 8.18 – Ranque das estações meteorológicas para C - CENTRL (MAPEs medianos na amostra B para FFNNs).

A Tabela 8.18 mostra que as quatro primeiras estações são as mais próximas do centróide da zona de carga e com maior nlor. No entanto, chama atenção a ordem de classificação, já que nem BGM e nem SYR, que estão dentro da zona de carga, ficaram com a primeira posição. Poder-se-ia verificar a razão disto, que pode ir desde as diferenças climáticas entre as localizações das estações até a qualidade de seus dados.

Nas Figuras 8.13 e 8.14, vemos, no diagrama de dispersão, o aglomerado formado por RME, BGM, ROC e SYR, com desempenho mais interessante que o das demais estações:

Figura 8.13 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e distância até o centróide de C - CENTRL.

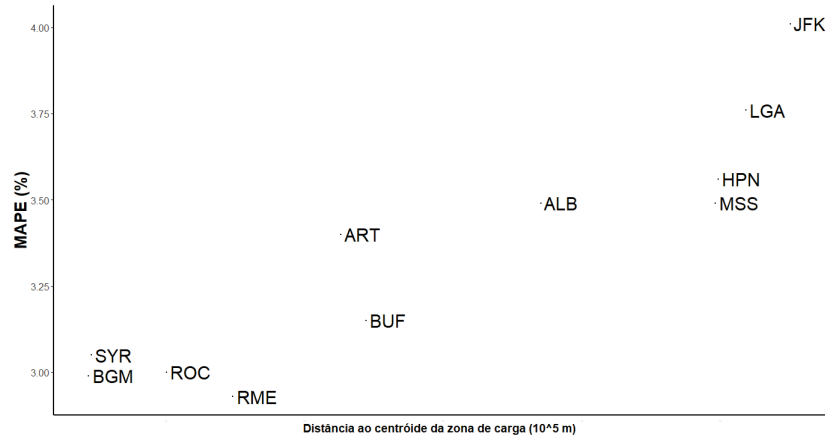
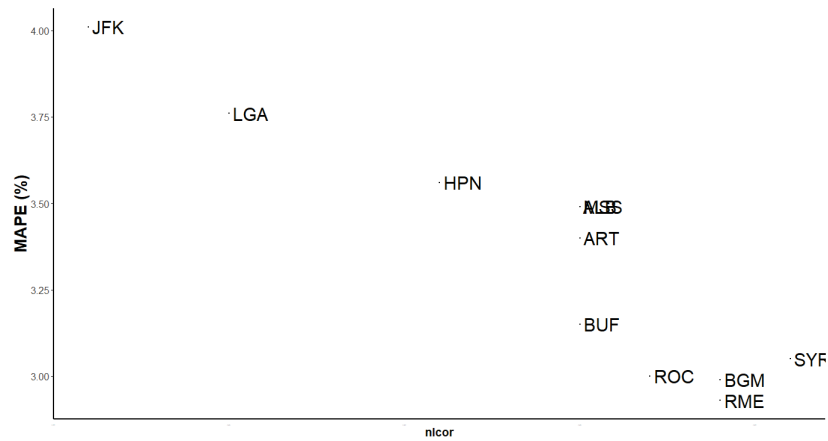


Figura 8.14 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e nlor das cargas de C - CENTRL.



8.2.3.3 Combinação das estações meteorológicas

Mostramos na Tabela 8.19, os *MAPEs* medianos das redes cujos dados temperatura utilizados nas entradas são obtidos pelos métodos C1, C2 e C3, para uma até 11 estações meteorológicas, combinadas pela ordem do ranque da Tabela 8.18:

Número de estações	Estações	MAPE mediano (%)		
		C1	C2	C3
1	RME	2,93	2,93	2,93
2	RME + BGM	2,94	2,88	2,87
3	RME + BGM + ROC	2,85	2,79	2,79
4	RME + BGM + ROC + SYR	2,77	2,80	2,80
5	RME + BGM + ROC + SYR + BUF	2,83	2,91	2,91
6	RME + BGM + ROC + SYR + BUF + ART	2,90	2,94	2,96
7	RME + BGM + ROC + SYR + BUF + ART + ALB	2,89	2,97	2,95
8	RME + BGM + ROC + SYR + BUF + ART + ALB + MSS	2,93	2,95	2,91
9	RME + BGM + ROC + SYR + BUF + ART + ALB + MSS + HPN	2,85	2,93	2,92
10	RME + BGM + ROC + SYR + BUF + ART + ALB + MSS + HPN + LGA	2,90	2,91	2,94
11	RME + BGM + ROC + SYR + BUF + ART + ALB + MSS + HPN + LGA + JFK	2,83	2,91	2,97

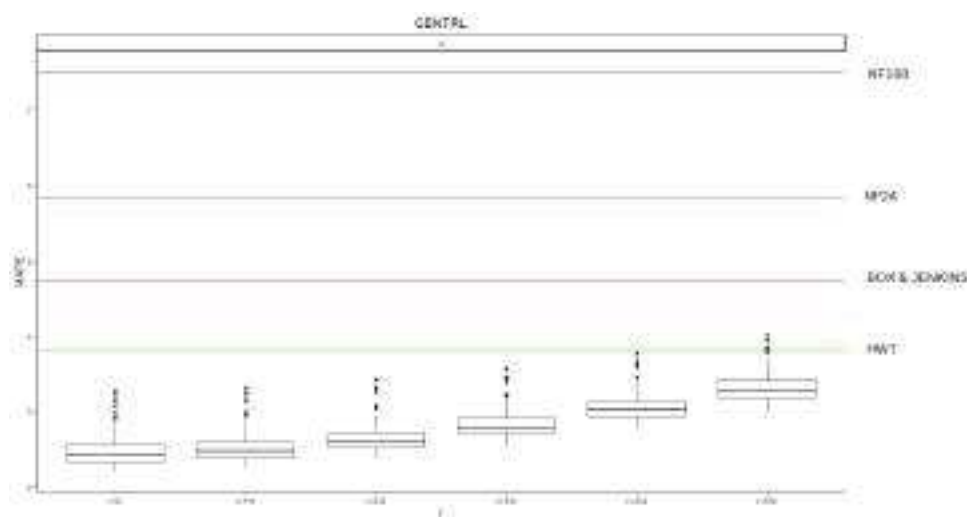
Tabela 8.19 – *MAPEs* medianos (amostra B; ano de 2018) para 100 FFNNs, utilizando os métodos C1, C2 e C3 para gerar as temperaturas de entrada das redes (estação C - CENTRL).

Assim como para a zona de carga B - GENESE, a combinação de estações meteorológicas melhora o desempenho. As estações combinadas foram as quatro mais próximas do centróide da zona de carga, RME, BGM, ROC e SYR (sendo BGM e SYR localizadas dentro do território da zona de carga). O método C1 (ponderação pelo inverso da distância até o centróide) foi mais uma vez aquele que chegou ao menor *MAPE* mediano.

Apresentamos nos *boxplots* da Figura 8.15, mostramos os resultados das redes neurais combinando via método C1 RME, BGM, ROC e SYR, em comparação com os

benchmarks, para C - CENTRL.

Figura 8.15 – *Boxplots* para os *MAPEs* das redes neurais (combinação C1 de RME, BGM, ROC e SYR) e dos *benchmarks* de séries temporais (zona de carga C - CENTRL).



O resultado das combinações é robusto, e mesmo com injeção de ruído o desempenho é superior ao dos *benchmarks* à exceção de apenas três redes (com ruído de 5%).

Na Tabela 8.20, vemos resultados das FFNNs com o melhor método, em comparação com os *benchmarks*, para os dados de teste (ano de 2018):

	Ruído					
FFNN	0	1%	2%	3%	4%	5%
(C1, estações RME, BGM, ROC e SYR)						
mínimo	2,20	2,25	2,37	2,55	2,76	3,00
mediana	2,43	2,48	2,70	2,79	3,07	3,29
máximo	3,23	3,32	3,24	3,57	3,77	4,01
HWT		3,79				
Box & Jenkins		4,75				
NF24		5,84				
NF168		7,50				

Tabela 8.20 – *MAPEs* (amostra B; ano de 2018) para FFFNs e *benchmarks*, dados de C - CENTRL.

Destacamos na Figura 8.16, as estações RME, BGM, ROC e SYR no mapa. Interessante notar como três destas estações formam um triângulo ao redor da região, provavelmente captando, em conjunto, uma variedade de suas características climáticas.

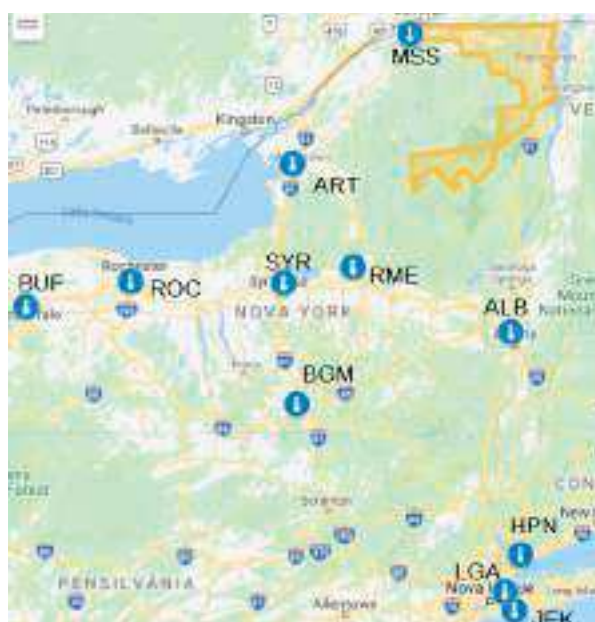
Figura 8.16 – Zona de carga C - CENTRL em azul claro, com estações meteorológicas indicadas em azul, e as estações RME, BGM, ROC e SYR, de melhor desempenho, em verde escuro.



8.2.4 Zona de carga D - NORTH

A zona de carga D - NORTH será provavelmente mais desafiadora para as previsões com dados de estações meteorológicas (nem sequer há uma estação dentro de seu território). Na Figura 8.17, vemos que esta é uma zona distante da maioria das estações meteorológicas. Além disso, também vimos no Capítulo 2 que é uma zona menos populosa que as demais e com características climáticas um pouco distintas. Ainda que a estação MSS esteja relativamente próxima da zona de carga, se comparada com as demais, ela está localizada em uma região fisiográfica diferente.

Figura 8.17 – Zona de carga D - NORTH em amarelo, com estações meteorológicas indicadas em azul escuro.



8.2.4.1 Benchmarks de séries temporais

Exibimos na Tabela 8.21, os resultados dos *benchmarks* para zona de carga D - NORTH, na amostra B (ano de 2018):

Método	MAPE (%)
HWT	3,57
Box & Jenkins	3,94
NF24	4,40
NF168	5,45

Tabela 8.21 – *MAPEs* (%) na amostra B para *benchmarks* univariados (zona de carga D - NORTH)

É curioso como, apesar de os *benchmarks* terem a mesma ordem de desempenho de acordo com o *MAPE*, em relação às zonas de carga A até C, nesse caso os previsores

ingênuos não são tão menos acurados quanto com as zonas analisadas anteriormente.

Trazemos na 8.22 as constantes para o HWT, cujos valores se assemelham ao que encontramos para as zonas A e B: nível constante, fator de correção para as autocorrelações dos erros de primeira ordem com importância significativa e fatores sazonais diário e semanal de intensidades próximas.

α	ϕ	δ	ω
0,00	0,96	0,14	0,15

Tabela 8.22 – Constantes de amortecimento do HWT para zona D - NORTH.

8.2.4.2 FFNN e ranqueamento das estações meteorológicas

Exibimos na próxima Tabela os resultados para os 11 conjuntos de FFNNs para a zona de carga D - NORTH:

Estação	MAPE (%)						MAPE mediano (todos os l)
	l = 0	l = 1%	l = 2%	l = 3%	l = 4%	l = 5%	
ALB	3,81	3,80	3,87	4,02	4,25	4,54	4,07
ART	3,52	3,52	3,62	3,80	4,05	4,35	3,78
BGM	3,71	3,71	3,80	4,00	4,26	4,59	3,98
BUF	3,78	3,77	3,84	3,99	4,21	4,52	4,04
HPN	4,11	4,11	4,20	4,36	4,56	4,84	4,38
JFK	3,84	3,87	4,01	4,26	4,61	5,10	4,18
LGA	3,76	3,76	3,89	4,11	4,43	4,82	4,13
MSS	3,50'	3,49	3,58	3,77	4,02	4,35	3,80
RME	3,67	3,65	3,73	3,87	4,11	4,40	3,94
ROC	3,65	3,64	3,76	3,94	4,21	4,52	3,89
SYR	3,89	3,88	3,96	4,14	4,36	4,68	4,17

Tabela 8.23 – MAPEs medianos na amostra B (ano de 2018), para 100 execuções de FFNNs, por estação meteorológicas, por nível de ruído (dados da zona de carga D - NORTH).

Elaboramos a partir dos resultados da última coluna da Tabela 8.23, o ranque das estações meteorológicas. A classificação, a distância de cada estação ao centróide da zona de carga e o *n*lcor entre os dados de carga e temperatura estão na Tabela 8.24.

Estação	MAPE mediano (%)	Distância até o centróide de D - NORTH	<i>n</i>lcor com dados de D - NORTH
ART	3,78	1,71	0,69
MSS	3,80	0,81	0,72
ROC	3,89	3,33	0,68
RME	3,94	1,82	0,69
BGM	3,98	3,03	0,67
LGA	4,13	4,17	0,65
SYR	4,17	2,30	0,69
JFK	4,18	4,32	0,65
HPN	4,38	3,84	0,67
BUF	4,04	4,19	0,67
ALB	4,07	1,98	0,69

Tabela 8.24 – Ranque das estações meteorológicas para D - NORTH (MAPEs medianos na amostra B para FFNNs).

Diferentemente do que se passa com as estações anteriores, para D - NORTH a Tabela 8.24 sugere que não há associação entre a acurácia mediana das FFNNs e a distância das estações meteorológicas até o centróide da zona de carga, ou o *n*lcor entre os dados de temperatura e de carga. Isto fica mais claro observando os diagramas de dispersão das Figuras 8.18 e 8.19. Impressiona também o fato de nenhuma das FFNNs ter desempenho mediano superior ao do HWT (*MAPE* de 3,57%).

Figura 8.18 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e distância até o centróide de D - NORTH.

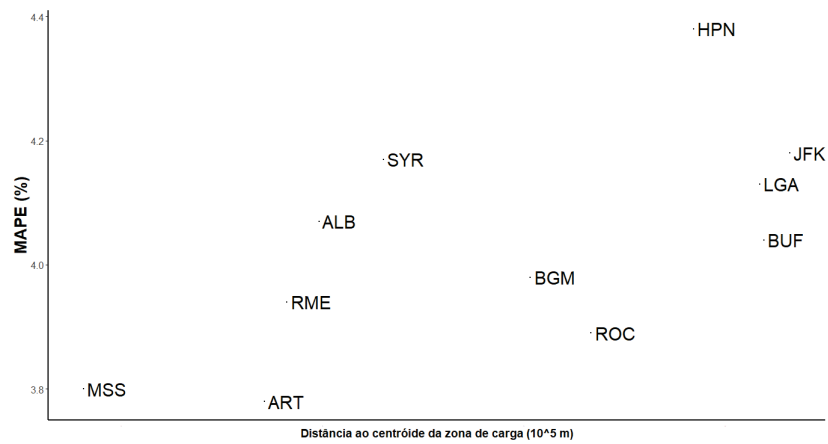
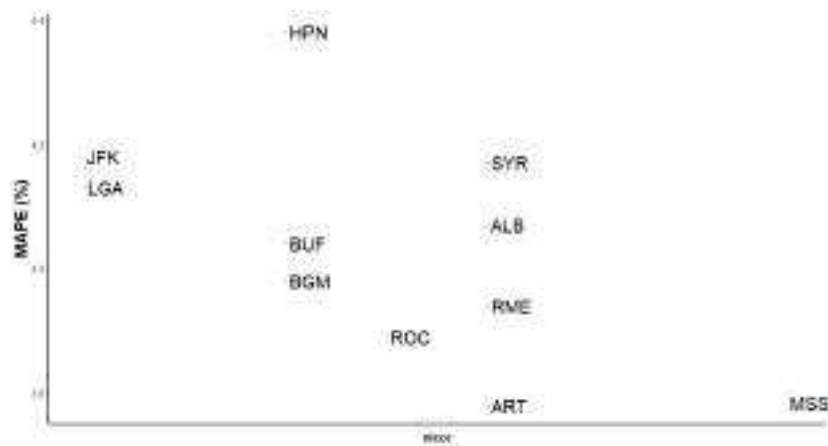


Figura 8.19 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e valor das cargas de D - NORTH.



8.2.4.3 Combinação das estações meteorológicas

Na Tabela 8.25, listamos os *MAPEs* medianos das redes com dados de temperatura de entrada obtidos pelos métodos C1, C2 e C3, para uma até 11 estações meteorológicas, combinadas pela ordem do ranque da Tabela 8.24:

Número de estações	Estações	MAPE mediano (%)		
		C1	C2	C3
1	ART	3,78	3,78	3,78
2	ART + MSS	3,75	3,80	3,92
3	ART + MSS + ROC	3,76	3,94	3,78
4	ART + MSS + ROC + RME	3,71	3,81	3,98
5	ART + MSS + ROC + RME + BGM	3,73	3,99	3,89
6	ART + MSS + ROC + RME + BGM + LGA	3,77	3,93	3,93
7	ART + MSS + ROC + RME + BGM + LGA + SYR	3,77	3,97	3,96
8	ART + MSS + ROC + RME + BGM + LGA + SYR + JFK	3,83	3,89	4,01
9	ART + MSS + ROC + RME + BGM + LGA + SYR + JFK + HPN	3,85	3,93	3,94
10	ART + MSS + ROC + RME + BGM + LGA + SYR + JFK + HPN + BUF	3,84	4,00	3,97
11	ART + MSS + ROC + RME + BGM + LGA + SYR + JFK + HPN + BUF + ALB	3,86	3,99	3,99

Tabela 8.25 – *MAPEs* medianos (amostra B; ano de 2018) para 100 FFNNs, utilizando os métodos C1, C2 e C3 para gerar as temperaturas de entrada das redes (estação D - NORTH).

Vemos na Tabela 8.25 que, para D - NORTH, algumas combinações de estações meteorológicas melhoram o desempenho das redes. Ainda assim, veremos que os resultados carecem de robustez, comparados aos *benchmarks*.

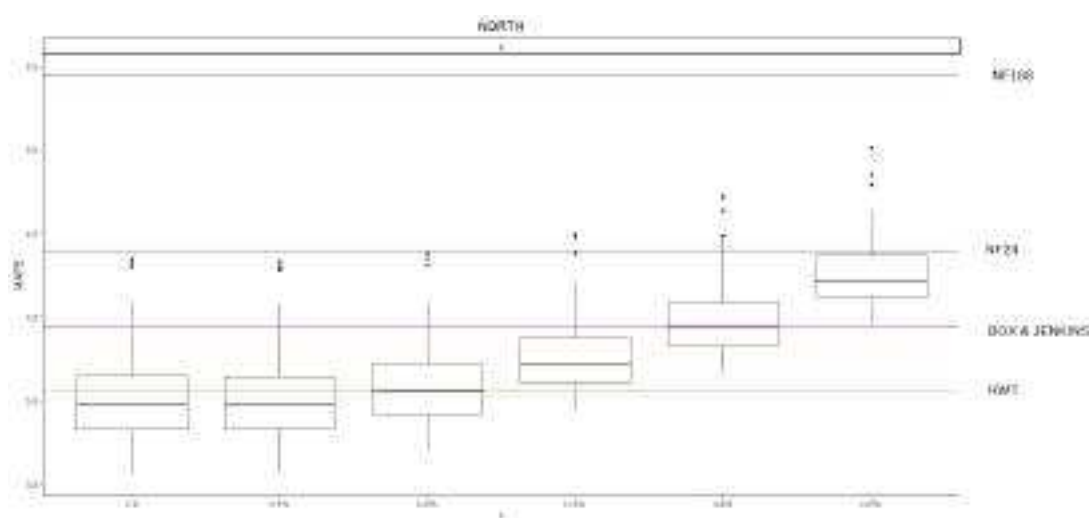
Na Tabela 8.26, exibimos os resultados das FFNNs com o melhor método de combinação, em comparação com os *benchmarks*, para os dados de teste (ano de 2018):

FFNN (C1, estações ART, MSS, ROC, RME)	Ruído					
	0	1%	2%	3%	4%	5%
mínimo	3,06	3,08	3,21	3,43	3,67	3,96
mediana	3,47	3,47	3,57	3,71	3,94	4,22
máximo	4,35	4,33	4,38	4,49	4,72	5,01
HWT	3,57					
Box & Jenkins	3,94					
NF24	4,40					
NF168	5,45					

Tabela 8.26 – *MAPEs* (amostra B; ano de 2018) para FFFNs e *benchmarks*, dados de D - NORTH.

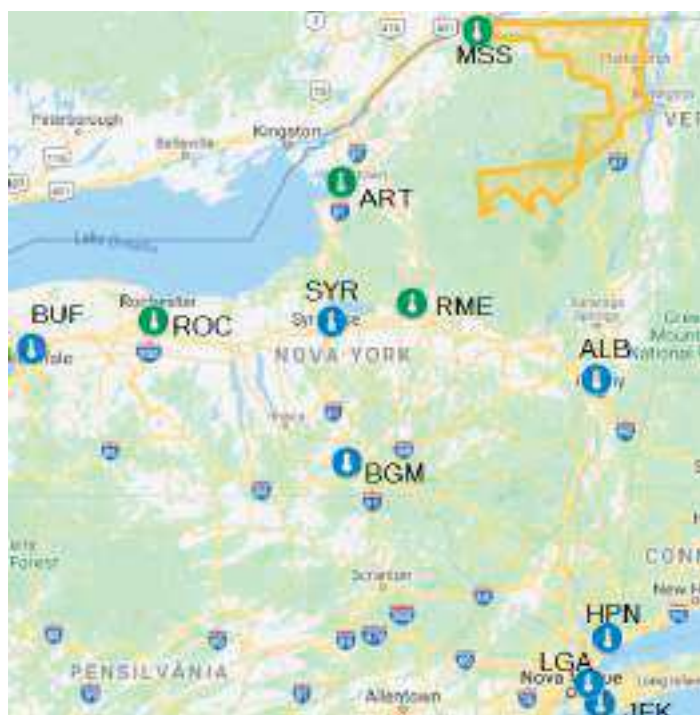
Na Figura 8.20 temos a apresentação gráfica da dispersão das redes em comparação com os *benchmarks*. Em conjunto com os dados da Tabela 8.26, estes resultados nos alertam para o fato de que, a partir de um nível de ruído de 2% nos dados de temperatura, 50% ou mais das redes apresentam desempenho inferior ao HWT - o que destaca, para estes dados, a robustez deste método simples e computacionalmente barato.

Figura 8.20 – *Boxplots* para os *MAPEs* das redes neurais (combinação C1 de ART, MSS, ROC e RME) e dos *benchmarks* de séries temporais (zona de carga D - NORTH).



Encerramos as análises para D - NORTH destacando na Figura 8.21, as estações ART, MSS, ROC e RME no mapa. Vale ressaltar o quão longe se encontra a estação ROC da zona de carga.

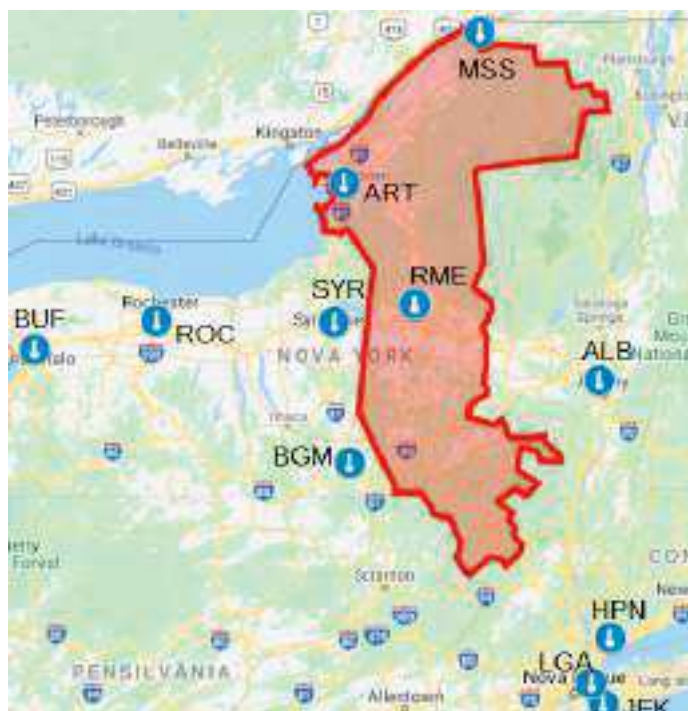
Figura 8.21 – Zona de carga D - NORTH em amarelo, com estações meteorológicas indicadas em azul, e as estações ART, MSS, ROC e RME, de melhor desempenho, em verde escuro.



8.2.5 Zona de carga E - MHKVL

Como mostra a Figura 8.22, existem três estações meteorológicas na zona de carga (de norte a sul MSS, ART e RME).

Figura 8.22 – Zona de carga E - MHKVL em vermelho, com estações meteorológicas indicadas em azul escuro.



8.2.5.1 Benchmarks de séries temporais

A 8.27 apresenta os resultados dos *benchmarks* para zona de carga E - MHKVL, na amostra B (ano de 2018):

Método	MAPE (%)
HWT	4,14
Box & Jenkins	5,35
NF24	6,25
NF168	8,30

Tabela 8.27 – MAPEs (%) na amostra B para *benchmarks* univariados (zona de carga E - MHKVL)

Ainda que os *benchmarks* sigam padrão similar ao que ocorre nas zonas de carga de A até D, a Tabela 8.28 nos mostra que para E - MHKVL a constante para sazonalidade semanal (ω) tem o dobro do peso da sazonalidade diária (δ):

α	ϕ	δ	ω
0,00	0,98	0,11	0,21

Tabela 8.28 – Constantes de amortecimento do HWT para zona D - NORTH.

8.2.5.2 FFNN e ranqueamento das estações meteorológicas

A Tabela a seguir traz os resultados para os 11 conjuntos de FFNNs para a zona de carga E - MHKVL:

Estação	MAPE (%)						MAPE mediano (todos os l)
	l = 0	l = 1%	l = 2%	l = 3%	l = 4%	l = 5%	
ALB	3,57	3,66	3,84	4,08	4,38	4,72	4,01
ART	3,67	3,72	3,87	4,10	4,40	4,74	4,01
BGM	3,64	3,73	3,89	4,12	4,42	4,72	4,11
BUF	4,04	4,08	4,23	4,45	4,73	5,10	4,43
HPN	3,74	3,82	4,01	4,25	4,54	4,87	4,20
JFK	3,89	3,96	4,15	4,39	4,68	5,01	4,33
LGA	3,68	3,76	3,94	4,20	4,49	4,81	4,10
MSS	3,91	3,94	4,08	4,27	4,54	4,87	4,26
RME	3,45	3,52	3,68	3,93	4,23	4,58	3,85
ROC	3,83	3,88	4,02	4,22	4,48	4,81	4,22
SYR	3,42	3,47	3,62	3,86	4,14	4,49	3,80

Tabela 8.29 – MAPEs medianos na amostra B, para 100 execuções de FFNNs, por estação meteorológicas, por nível de ruído (dados da zona de carga E - MHKVL).

Elaboramos o ranque das estações meteorológicas a partir da última coluna da Tabela 8.29 e o mostramos na Tabela 8.30, juntamete com a distância de cada estação ao centróide da zona de carga e o nlor entre os dados de carga e temperatura.

Estação	MAPE mediano (%)	Distância até o centróide de E - MHKVL	nlor com dados de E - MHKVL
SYR	3,80	0,88	0,65
RME	3,85	0,33	0,64
ALB	4,01	1,33	0,65
ART	4,01	0,94	0,62
HPN	4,20	2,90	0,65
ROC	4,22	2,09	0,62
MSS	4,26	1,67	0,63
JFK	4,33	3,32	0,63
BUF	4,43	2,98	0,61
LGA	4,10	3,15	0,63
BGM	4,11	1,55	0,64

Tabela 8.30 – Ranque das estações meteorológicas para E - MHKVL (MAPEs medianos na amostra B para FFNNs).

A Tabela 8.30 indica que não parece haver correlação entre o *MAPE* e a distância da estação meteorológica até o centróide da zona de carga. Também notamos que há pouca dispersão para o nlor. Os diagramas das Figuras 8.23 e 8.24 reforçam estes padrões.

Figura 8.23 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e distância até o centróide de E - MHKVL.

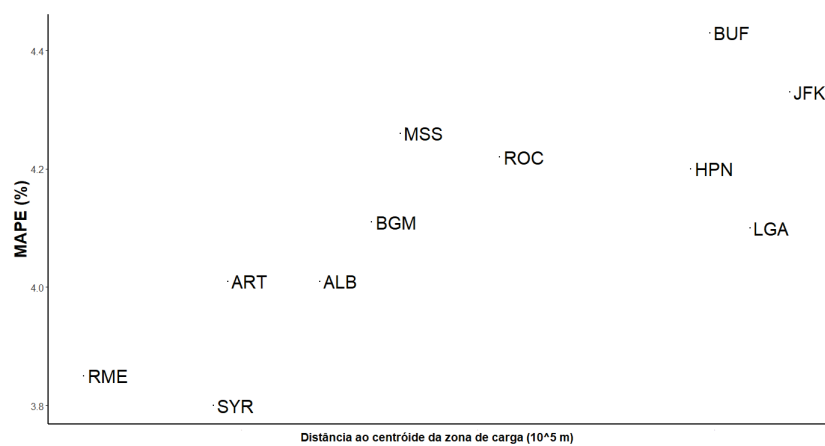
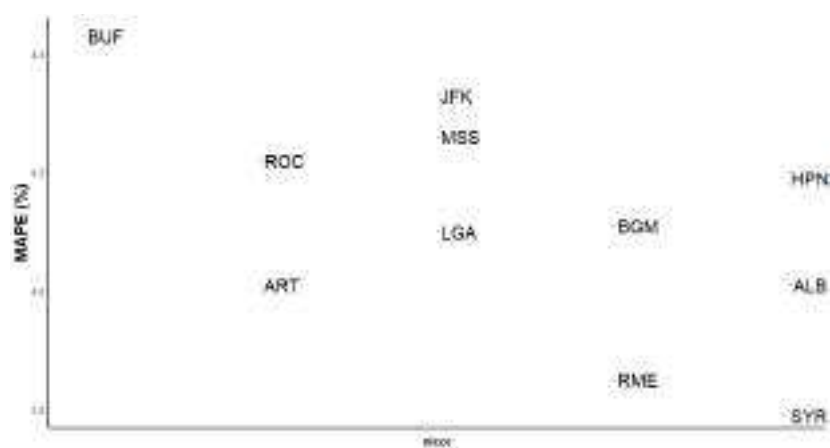


Figura 8.24 – Diagrama de dispersão entre MAPE mediano na amostra B das FFNN e nlor das cargas de E - MHKVL.



8.2.5.3 Combinação das estações meteorológicas

Mostramos na Tabela 8.31, *MAPEs* medianos das redes com dados de temperatura de entrada obtidos pelos métodos C1, C2 e C3, para uma até 11 estações meteorológicas, combinadas pela ordem do ranque da Tabela 8.30:

Número de estações	Estações	MAPE mediano (%)		
		C1	C2	C3
1	SYR	3,80	3,80	3,80
2	SYR + RME	3,75	3,76	3,77
3	SYR + RME + ALB	3,80	3,76	3,79
4	SYR + RME + ALB + ART	3,77	3,71	3,71
5	SYR + RME + ALB + ART + HPN	3,80	3,74	3,69
6	SYR + RME + ALB + ART + HPN + ROC	3,76	3,72	3,76
7	SYR + RME + ALB + ART + HPN + ROC + MSS	3,72	3,69	3,76
8	SYR + RME + ALB + ART + HPN + ROC + MSS + JFK	3,75	3,72	3,75
9	SYR + RME + ALB + ART + HPN + ROC + MSS + JFK + BUF	3,71	3,71	3,78
10	SYR + RME + ALB + ART + HPN + ROC + MSS + JFK + BUF + LGA	3,76	3,72	3,73
11	SYR + RME + ALB + ART + HPN + ROC + MSS + JFK + BUF + LGA + BGM	3,71	3,77	3,75

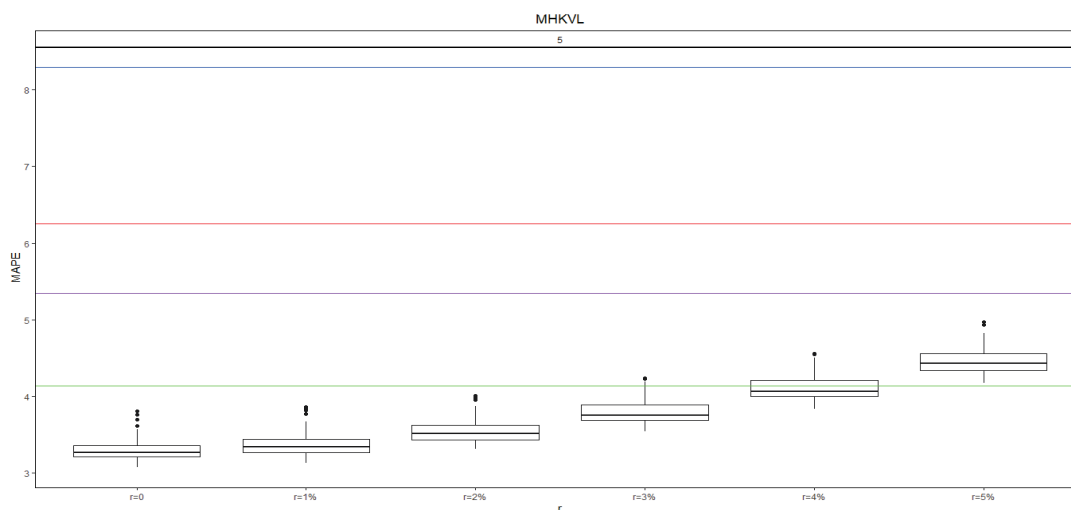
Tabela 8.31 – *MAPEs* medianos (amostra B; ano de 2018) para 100 FFNNs, utilizando os métodos C1, C2 e C3 para gerar as temperaturas de entrada das redes (estação E - MHKVL).

Na Tabela 8.31 vemos que tanto a combinação C3 (média aritmética simples) com cinco estações quanto a combinação C2 (ponderação pelo inverso da distância até o centróide) obtém um *MAPE* mediano de 3,69%. Preferimos C3 devido porque o mesmo resultado mediano é obtido dependendo de menos dados de fontes externas.

Na Figura 8.25 mostramos graficamente a dispersão das redes em comparação com os *benchmarks*. O desempenho das redes é melhor que o dos *benchmarks* para até 2% de nível de ruído, e consideravelmente pior na maioria das vezes quando o ruído é de no

mínimo 4%. Mais uma vez, fica destacada a robustez de um método univariado como o HWT quando a qualidade dos dados externos é comprometida.

Figura 8.25 – *Boxplots* para os *MAPEs* das redes neurais (combinação C3 de SYR, RME, ALB, ART e HPN) e dos *benchmarks* de séries temporais (zona de carga E - MHKVL).



Os resultados das FFNNs com o melhor método de combinação, em comparação com os *benchmarks*, para os dados de teste (ano de 2018), estão na Tabela 8.32

	Ruído					
FFNN	0	1%	2%	3%	4%	5%
(C3, estações SYR, RME, ALB, ART, HPN)						
mínimo	3,07	3,12	3,31	3,54	3,84	4,17
mediana	3,27	3,36	3,54	3,75	4,10	4,43
máximo	3,81	4,20	4,01	4,24	4,56	4,97
HWT		4,14				
Box & Jenkins		5,35				
NF24		6,25				
NF168		8,30				

Tabela 8.32 – *MAPEs* (amostra B; ano de 2018) para FFFNs e *benchmarks*, dados de E - MHKVL.

Na Figura 8.26, destacamos as estações SYR, RME, ALB, ART e HPN no mapa. Como a zona de carga tem grande extensão longitudinal, é razoável pensar que sejam necessárias mais estações meteorológicas combinadas para descrever o clima local. Além disso, também faz sentido supor que, como há necessidade de capturar o clima de regiões diversas, a distância até o centróide não faça tanta diferença (na verdade, acaba sendo preferível contar com estações dispersas ao longo do território). Chama atenção a inclusão da estação HPN, mais a sul, e não inclusão de MSS, mais a norte (e dentro do território da zona de carga). Isto possivelmente acontece porque o clima em HPN é mais parecido com o de alguma parte da zona de carga do que o de MSS, por conta das diferenças fisiográficas.

Figura 8.26 – Zona de carga E - MHKVL em vermelho, com estações meteorológicas indicadas em azul, e as estações SYR, RME, ALB, ART e HPN, de melhor desempenho, em verde escuro.



8.2.6 Zona de carga F - CAPITL

F - CAPITL é a zona de carga que atende à capital do estado de Nova Iorque, Albany, e seus arredores. Vemos na Figura 8.27 que uma estação meteorológica (ALB) se localiza no interior da zona de carga.

Figura 8.27 – Zona de carga F - CAPITL em verde, com estações meteorológicas indicadas em azul escuro.



8.2.6.1 Benchmarks de séries temporais

Os resultados dos *benchmarks* para zona de carga F - CAPITL, na amostra B (ano de 2018) estão exibidos na Tabela 8.33:

Método	MAPE (%)
HWT	4,10
Box & Jenkins	5,14
NF24	5,82
NF168	8,41

Tabela 8.33 – *MAPEs* (%) na amostra B para *benchmarks* univariados (zona de carga F - CAPITL)

O padrão dos resultados não difere do que vimos até agora, com o HWT sendo o *benchmark* de melhor desempenho. Quanto às constantes do HWT (Tabela 8.34), a constante para atualização do fator sazonal semanal é ligeiramente maior que a do fator diário.

α	ϕ	δ	ω
0,00	0,99	0,12	0,16

Tabela 8.34 – Constantes de amortecimento do HWT para zona F - CAPITL.

8.2.6.2 FFNN e ranqueamento das estações meteorológicas

Na próxima Tabela, apresentamos os resultados para os 11 conjuntos de FFNNs para a zona de carga F - CAPITL:

Estação	MAPE (%)						MAPE mediano (todos os l)
	l = 0	l = 1%	l = 2%	l = 3%	l = 4%	l = 5%	
ALB	2,94	3,01	3,19	3,45	3,77	4,17	3,40
ART	3,50	3,55	3,68	3,93	4,24	4,58	3,84
BGM	3,59	3,65	3,79	3,99	4,27	4,58	4,04
BUF	3,88	3,93	4,10	4,28	4,55	4,88	4,34
HPN	3,77	3,82	3,97	4,22	4,48	4,77	4,21
JFK	3,97	4,04	4,21	4,40	4,65	4,93	4,41
LGA	3,84	3,88	4,04	4,22	4,49	4,77	4,24
MSS	3,65	3,70	3,85	4,06	4,33	4,63	4,02
RME	3,32	3,36	3,50	3,73	3,99	4,30	3,69
ROC	3,52	3,58	3,72	3,97	4,27	4,60	3,90
SYR	3,32	3,36	3,52	3,76	4,07	4,45	3,72

Tabela 8.35 – MAPEs medianos na amostra B, para 100 execuções de FFNNs, por estação meteorológicas, por nível de ruído (dados da zona de carga F - CAPITL).

Baseados na última coluna da Tabela 8.35, construímos o ranque, apresentando-o na Tabela 8.36, com uma coluna para a distância de cada estação ao centróide da zona de carga e outra para o nlor entre os dados de carga e temperatura.

Estação	MAPE mediano (%)	Distância até o centróide de F - CAPITL	nlor com dados de F - CAPITL
ALB	3,40	0,52	0,65
RME	3,69	1,16	0,63
SYR	3,72	1,74	0,63
ART	3,84	1,87	0,60
ROC	3,90	3,00	0,61
MSS	4,02	2,06	0,62
BGM	4,04	1,98	0,63
HPN	4,21	2,38	0,63
LGA	4,24	2,70	0,61
BUF	4,34	3,88	0,60
JFK	4,41	2,85	0,60

Tabela 8.36 – Ranque das estações meteorológicas para F - CAPITL (MAPEs medianos na amostra B para FFNNs).

Observando os resultados da Tabela 8.36 e os diagramas de dispersão nas Figuras 8.28 e 8.29, fica clara a vantagem da estação ALB, que está no interior da zona de carga, em relação às demais. Vale recordar, do Capítulo 2, que ALB é uma das estações com menos valores faltantes (7 pontos, apenas).

Figura 8.28 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e distância até o centróide de F - CAPITL.

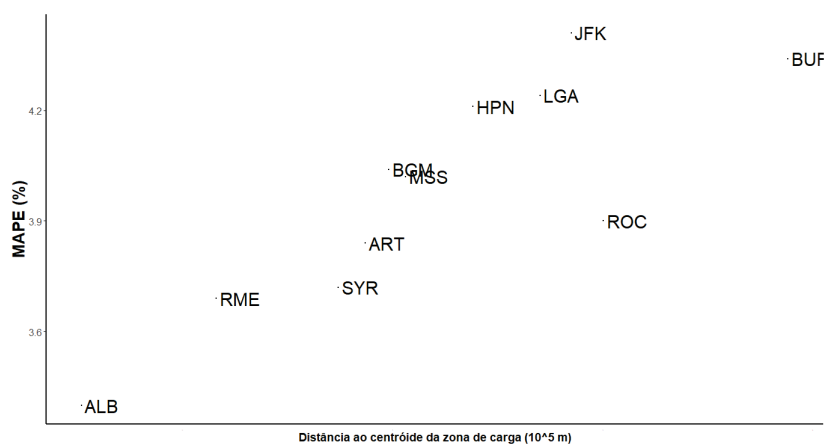
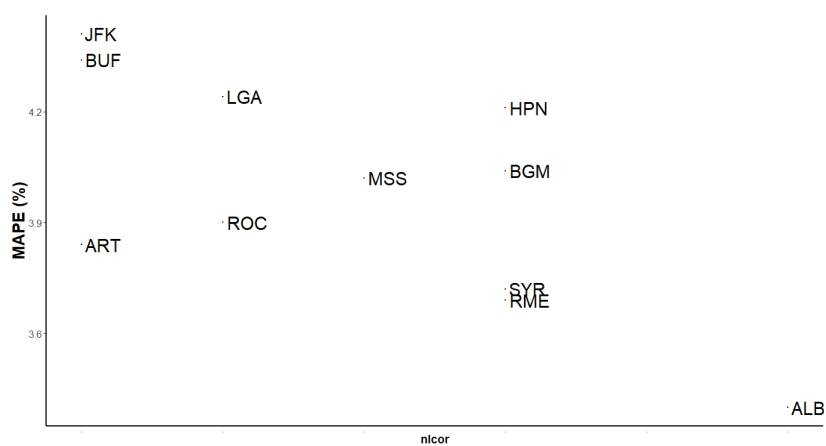


Figura 8.29 – Diagrama de dispersão entre MAPE mediano na amostra B das FFNN e nlor das cargas de F - CAPITL.



8.2.6.3 Combinação das estações meteorológicas

Na Tabela 8.37, exibimos *MAPEs* medianos das redes com dados de temperatura de entrada obtidos pelos métodos C1, C2 e C3, para uma até 11 estações meteorológicas, combinadas pela ordem do ranque da Tabela 8.36:

Número de estações	Estações	MAPE mediano (%)		
		C1	C2	C3
1	ALB	3,40	3,40	3,40
2	ALB + RME	3,56	3,43	3,49
3	ALB + RME + SYR	3,57	3,61	3,57
4	ALB + RME + SYR + ART	3,55	3,56	3,57
5	ALB + RME + SYR + ART + ROC	3,57	3,54	3,57
6	ALB + RME + SYR + ART + ROC + MSS	3,56	3,62	3,57
7	ALB + RME + SYR + ART + ROC + MSS + BGM	3,51	3,59	3,56
8	ALB + RME + SYR + ART + ROC + MSS + BGM + HPN	3,61	3,56	3,57
9	ALB + RME + SYR + ART + ROC + MSS + BGM + HPN + LGA	3,54	3,52	3,56
10	ALB + RME + SYR + ART + ROC + MSS + BGM + HPN + LGA + BUF	3,58	3,57	3,55
11	ALB + RME + SYR + ART + ROC + MSS + BGM + HPN + LGA + BUF + JFK	3,59	3,62	3,58

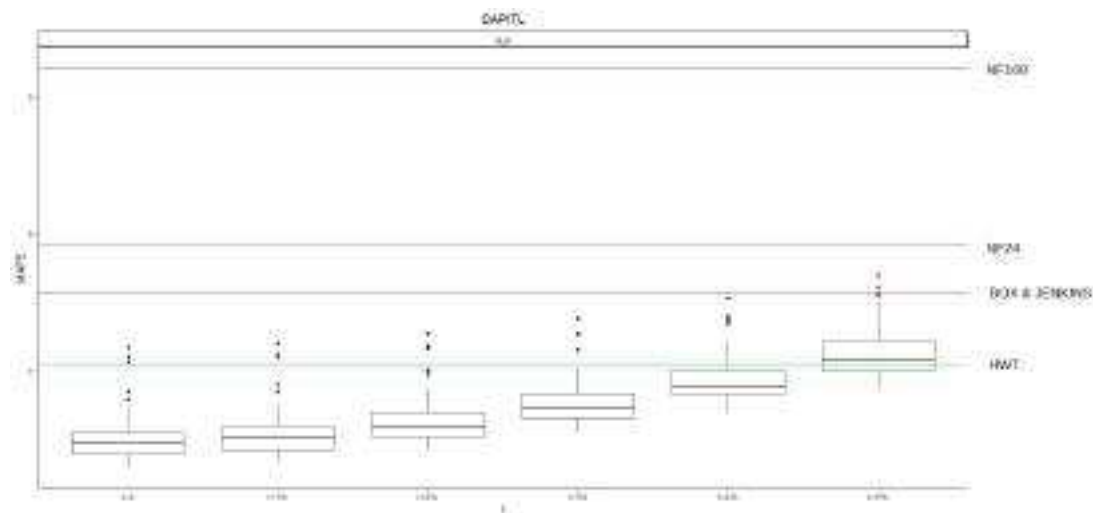
Tabela 8.37 – *MAPEs* medianos (amostra B; ano de 2018) para 100 FFNNs, utilizando os métodos C1, C2 e C3 para gerar as temperaturas de entrada das redes (estação F - CAPITL).

A Tabela 8.37 corrobora o que havíamos mencionado: os resultados utilizando os dados de ALB são tão melhores que os das demais estações que, a exemplo do que ocorre com a estação BUF na zona A - WEST, uma estação basta para se obterem as melhores previsões. Vale repetir que isto é uma vantagem, dada a menor dependência de dados de fontes externas.

Mostramos na Figura 8.30, através de *boxplots*, a dispersão dos resultados das redes em comparação com os *benchmarks*. Apesar do bom desempenho das FFNNs utilizando

os dados da estação ALB, há que se tomar cuidado para níveis de ruído a partir de 4%, quando o HWT passa a ser mais vantajoso.

Figura 8.30 – *Boxplots* para os *MAPEs* das redes neurais (dados de temperatura da estação ALB) e dos *benchmarks* de séries temporais (zona de carga F - CAPITL).



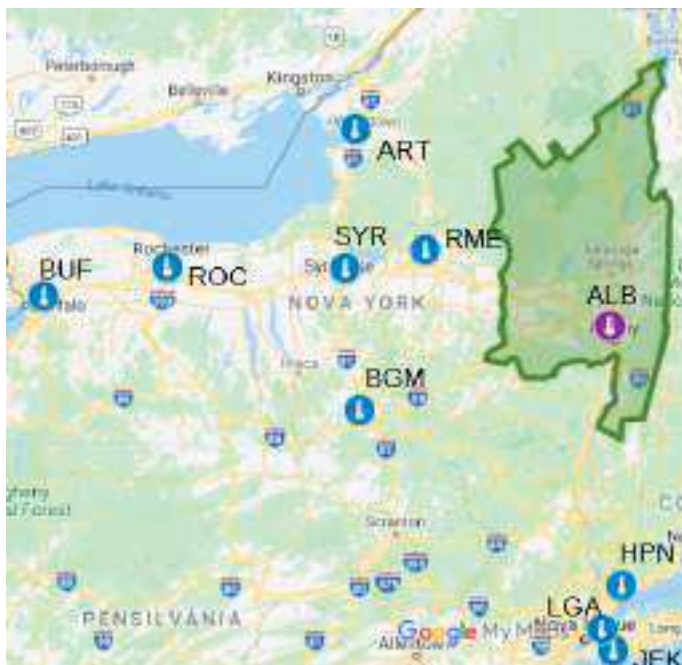
Disponemos os resultados das FFNNs com o melhor método de combinação, em comparação com os *benchmarks*, para os dados de teste (ano de 2018) na Tabela 8.38:

FFNN (estação ALB)	Ruído					
	0	1%	2%	3%	4%	5%
mínimo	2,58	2,65	2,85	3,10	3,38	3,70
mediana	2,94	3,01	3,19	3,45	3,77	4,16
máximo	4,33	4,40	4,53	4,76	5,05	5,39
HWT	4,10					
Box & Jenkins	5,14					
NF24	5,82					
NF168	8,41					

Tabela 8.38 – *MAPEs* (amostra B; ano de 2018) para FFFNs e *benchmarks*, dados de F - CAPITL.

Destacamos no mapa da Figura 8.31 a estação ALB:

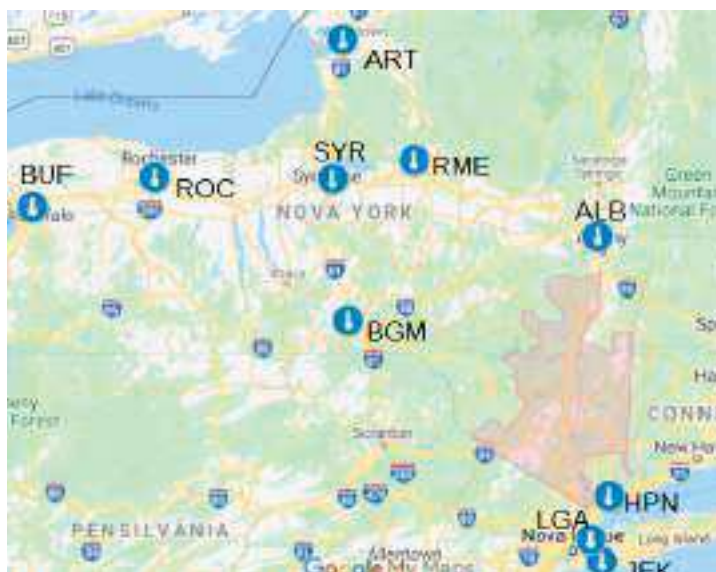
Figura 8.31 – Zona de carga F - CAPITL em verde, com estações meteorológicas indicadas em azul, e a estação ALB, de melhor desempenho, em roxo.



8.2.7 Zona de carga G - HUDVL

A última zona de carga para qual experimentamos nossos métodos é G - HUDVL. Na Figura 8.32, vemos que não há estação meteorológica no interior da zona de carga, que possui território próximo a região de Albany e também ao litoral. As três estações mais próximas, HPN, LGA e JFK estão próximas às planícies litorâneas.

Figura 8.32 – Zona de carga G - HUDVL em rosa, com estações meteorológicas indicadas em azul escuro.



8.2.7.1 Benchmarks de séries temporais

Para zona de carga G - HUDVL, exibimos os resultados dos *benchmarks* na amostra B (ano de 2018) estão exibidos na Tabela 8.39:

Método	MAPE (%)
HWT	5,56
Box & Jenkins	6,26
NF24	6,96
NF168	10,01

Tabela 8.39 – *MAPEs* (%) na amostra B para *benchmarks* univariados (zona de carga G - HUDVL)

Para todas as sete zonas de carga, confirma-se o HWT como *benchmark* de melhor desempenho.

Na Tabela 8.40), vemos algo que até agora não havia ocorrido, com a constante para o fator sazonal diário com intensidade razoavelmente maior que a do fator sazonal semanal:

α	ϕ	δ	ω
0,01	1,00	0,18	0,12

Tabela 8.40 – Constantes de amortecimento do HWT para zona G - HUDVL.

8.2.7.2 FFNN e ranqueamento das estações meteorológicas

Os resultados para os 11 conjuntos de FFNNs para a zona de carga G - HUDVL encontram-se na Tabela a seguir:

Estação	MAPE (%)						MAPE mediano (todos os l)
	l = 0	l = 1%	l = 2%	l = 3%	l = 4%	l = 5%	
ALB	4,31	4,35	4,45	4,58	4,80	4,99	4,62
ART	5,10	5,10	5,15	5,29	5,47	5,69	5,32
BGM	4,10	4,15	4,23	4,40	4,60	4,87	4,40
BUF	4,83	4,85	4,95	5,12	5,32	5,55	5,10
HPN	3,51	3,55	3,70	3,93	4,21	4,51	3,96
JFK	4,20	4,21	4,35	4,54	4,75	4,98	4,56
LGA	3,96	4,00	4,10	4,26	4,45	4,64	4,27
MSS	5,10	5,13	5,22	5,36	5,55	5,78	5,38
RME	4,73	4,76	4,85	4,99	5,16	5,41	4,99
ROC	4,77	4,79	4,88	5,00	5,21	5,43	5,05
SYR	4,76	4,78	4,88	5,02	5,20	5,40	4,99

Tabela 8.41 – MAPEs medianos na amostra B, para 100 execuções de FFNNs, por estação meteorológicas, por nível de ruído (dados da zona de carga G - HUDVL).

A partir da última coluna da Tabela 8.41, construímos o ranque da Tabela 8.42, acrescido de uma coluna para a distância de cada estação ao centróide da zona de carga e outra para o nlor entre os dados de carga e temperatura.

Estação	MAPE mediano (%)	Distância até o centróide de G - HUDVL	nlor com dados de G - HUDVL
HPN	3,96	0,80	0,66
LGA	4,27	1,08	0,64
BGM	4,40	1,67	0,63
JFK	4,56	1,24	0,62
ALB	4,62	1,15	0,65
RME	4,99	2,00	0,61
SYR	4,99	2,28	0,61
ROC	5,05	3,34	0,59
BUF	5,10	4,06	0,58
ART	5,32	2,98	0,57
MSS	5,38	3,62	0,59

Tabela 8.42 – Ranque das estações meteorológicas para G - HUDVL (MAPEs medianos na amostra B para FFNNs).

Na Tabela 8.42, verificamos que as estações HPN, LGA e JFK, que se encontram nas planícies litorâneas, figuram entre as primeiras. BGM também tem um bom desempenho, o que pode dever-se ao fato de captar algum fator climático mais ao norte da zona de carga, que possui territórios em altitudes diferentes (planalto Apalache a norte e planícies litorâneas a sul). Os resultados estão representados graficamente nas Figuras 8.33 e 8.34.

Figura 8.33 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e distância até o centróide de G - HUDVL.

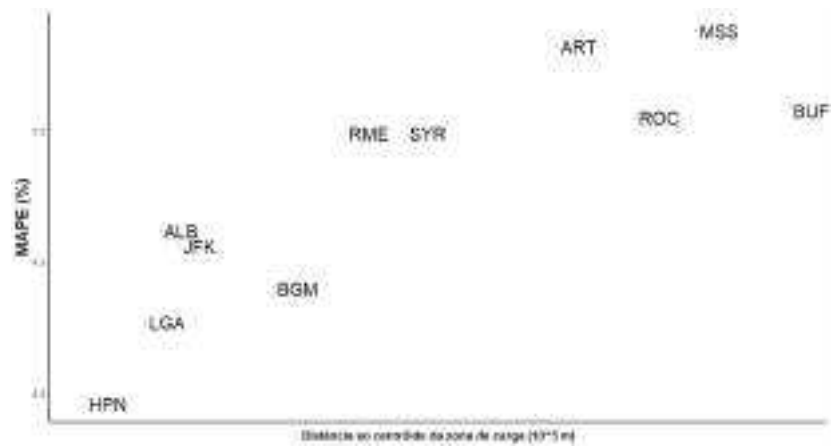
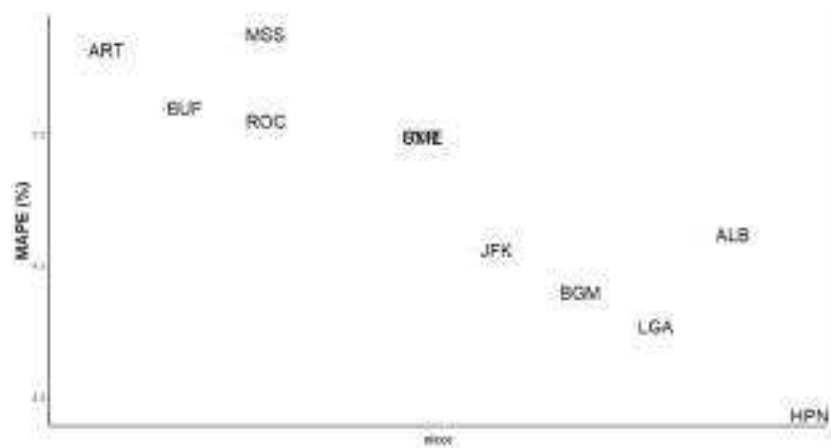


Figura 8.34 – Diagrama de dispersão entre *MAPE* mediano na amostra B das FFNN e nlor das cargas de G - HUDVL.



8.2.7.3 Combinação das estações meteorológicas

Na Tabela 8.44, estão dispostos *MAPEs* medianos das redes com dados de temperatura de entrada obtidos pelos métodos C1, C2 e C3, para uma até 11 estações meteorológicas, combinadas pela ordem do ranque da Tabela 8.42:

Número de estações	Estações	MAPE mediano (%)		
		C1	C2	C3
1	HPN	3,96	3,96	3,96
2	HPN + LGA	4,10	3,94	3,98
3	HPN + LGA + BGM	3,82	3,62	3,78
4	HPN + LGA + BGM + JFK	3,93	3,79	3,75
5	HPN + LGA + BGM + JFK + ALB	3,75	3,71	3,73
6	HPN + LGA + BGM + JFK + ALB + RME	3,74	3,74	3,76
7	HPN + LGA + BGM + JFK + ALB + RME + SYR	3,80	3,88	3,91
8	HPN + LGA + BGM + JFK + ALB + RME + SYR + ROC	3,83	3,97	4,11
9	HPN + LGA + BGM + JFK + ALB + RME + SYR + ROC + BUF	3,77	4,05	4,08
10	HPN + LGA + BGM + JFK + ALB + RME + SYR + ROC + BUF + ART	3,93	4,14	4,26
11	HPN + LGA + BGM + JFK + ALB + RME + SYR + ROC + BUF + ART + MSS	3,83	4,24	4,33

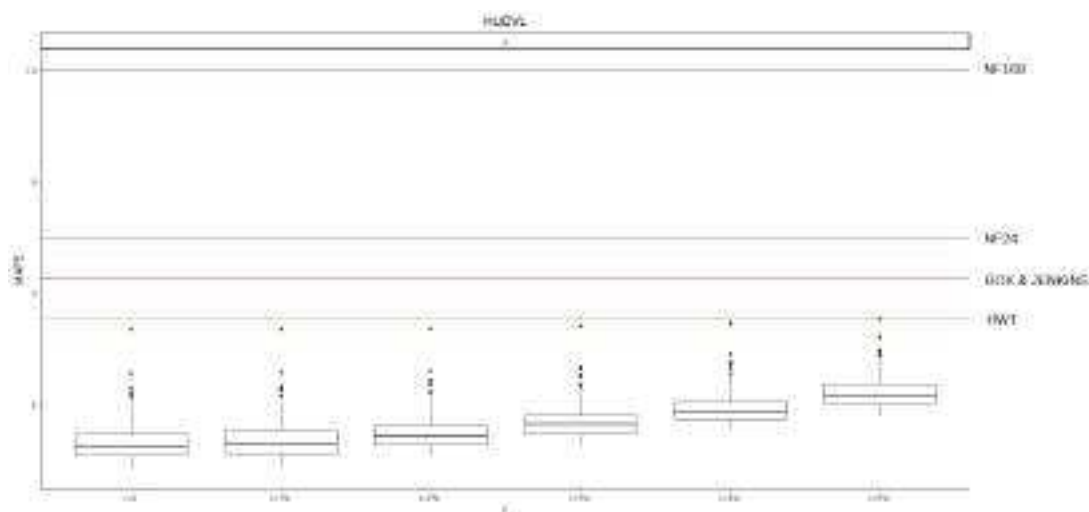
Tabela 8.43 – *MAPEs* medianos (amostra B; ano de 2018) para 100 FFNNs, utilizando os métodos C1, C2 e C3 para gerar as temperaturas de entrada das redes (estação G - HUDVL).

Na Tabela 8.44, vemos que os melhores resultados, em mediana dos *MAPEs*, é obtido pela combinação de três estações, HPN, LGA (estas nas planícies litorâneas) e BGM (no planalto Apalache). O método de combinação que produz estes resultados é o C2 (ponderação pelo nlcor).

Na Figura 8.35 exibimos os *boxplots* para os resultados das redes em comparação com os *benchmarks*. Impressiona a robustez do método a que chegamos, com as redes cujas entradas são combinações dos dados de HPN, LGA e BGM pelo método C2. Mesmo

com nível de ruído de 5%, não há nenhum resultado batido por algum *benchmark*.

Figura 8.35 – *Boxplots* para os *MAPEs* das redes neurais (dados de temperatura das estações HPN, LGA e BGM, combinados pelo método C2) e dos *benchmarks* de séries temporais (zona de carga G - HUDVL).



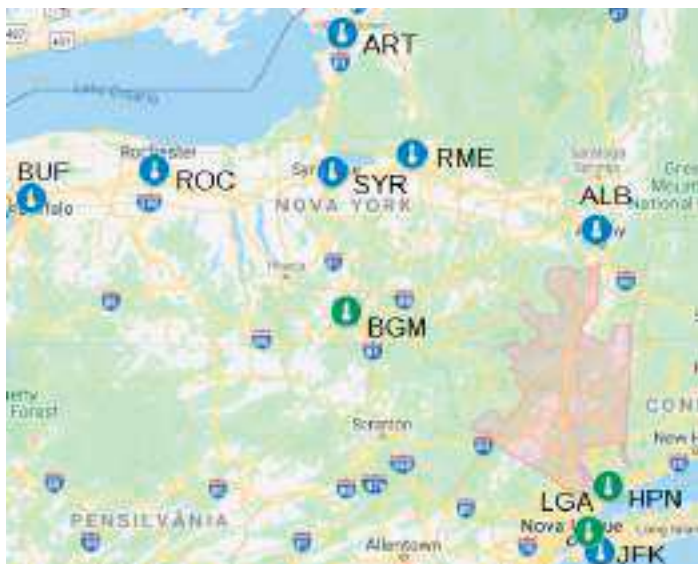
Os resultados das FFNNs com o melhor método de combinação, em comparação com os *benchmarks*, para os dados de teste (ano de 2018), estão na Tabela 8.44:

	Ruído					
FFNN						
(estações HPN, LGA e BGM, combinadas pelo método C2)	0	1%	2%	3%	4%	5%
mínimo	2,83	2,88	3,05	3,29	3,55	3,78
mediana	3,24	3,29	3,43	3,64	3,87	4,14
máximo	5,34	5,34	5,36	5,40	5,46	5,53
HWT			5,56			
Box & Jenkins			6,26			
NF24			6,96			
NF168			10,01			

Tabela 8.44 – *MAPEs* (amostra B; ano de 2018) para FFFNs e *benchmarks*, dados de G - HUDVL.

Destacamos no mapa da Figura 8.36 as estações HPN, LGA e BGM:

Figura 8.36 – Zona de carga G - HUDVL em rosa, com estações meteorológicas indicadas em azul, e as estações HPN, LGA e BGM, de melhor desempenho, em verde.



8.3 RESUMO DOS RESULTADOS

Um resumo dos resultados, com os melhores métodos para cada conjunto de dados, pode ser visto na Tabela 8.45. Os resultados evidenciam o bom desempenho dos métodos introduzidos nesta tese, em especial da combinação C1 com previsores de redes neurais.

Zona de carga	Nível de ruído					
	0	1%	2%	3%	4%	5%
A - WEST	FFNN	FFNN	FFNN	FFNN	HWT	HWT
B - GENESE	C1	C1	C1	C1	C1	C1
C - CENTRL	C1	C1	C1	C1	C1	C1
D - NORTH	C1	C1	HWT	HWT	HWT	HWT
E - MHKVL	C3	C3	C3	C3	HWT	HWT
F - CAPITL	FFNN	FFNN	FFNN	FFNN	FFNN	HWT
G - HUDVL	C2	C2	C2	C2	C2	C2

Tabela 8.45 – Métodos de melhor desempenho (menor MAPE mediano para preditor de redes neurais ou menor MAPE para preditor de séries temporais) para cada zona de carga, por nível de ruído.

9 CONCLUSÕES

A previsão de carga elétrica é um problema clássico para as empresas do setor, discutido há décadas na Academia e no Mercado.

Entretanto, mudanças no mercado de energia, na economia global e nos aspectos climáticos nos anos mais recentes trouxeram ainda mais desafios para a cadeia de suprimentos do setor elétrico, ratificando a importância da boa previsão para os custos de operação.

Abordamos, nesta tese, o tema da seleção e da combinação de estações meteorológicas para previsão de carga elétrica em curto prazo. Em um contexto em que cada vez mais dados são coletados e disponibilizados, e em que *smart grids* permitem monitorar por sensores a carga em diversos pontos de um território, torna-se crucial escolher e/ou combinar as fontes dos dados externos que serão utilizados como entrada para os modelos de previsão.

Apesar da importância, o problema havia, até então, sido pouco explorado, com a maioria das (poucas) publicações reportando pouca ou nenhuma atenção ao tema.

Em nosso trabalho, validamos uma heurística para seleção e combinação de estações meteorológicas para sete novos conjuntos de dados, dois novos métodos de combinação e um método de previsão mais complexo.

Os conjuntos de dados foram cargas horárias para os anos de 2015 a 2018 de sete zonas do NYISO, o Operador Independente que atende ao estado de Nova Iorque, nos EUA.

Como métodos de combinação, apresentamos duas propostas inéditas: C1, a ponderação pelo inverso da distância da estação meteorológica ao centróide da zona de carga, com as estações mais próximas recebendo maior peso; e C2, a ponderação pela intensidade da correlação não-linear entre os dados da estação meteorológica e os dados de carga (utilizando, para isso, uma medida desenvolvida recentemente para capturar a intensidade desta correlação). Comparamos os resultados a C3, a média aritmética simples das estações; método até hoje mais recorrente na literatura.

Para as previsões, utilizamos redes neurais do tipo *feedforward* (FFNN), com atenção especial para táticas de validação. Também demos atenção especial à sensibilidade dos resultados a ruídos nos dados de temperatura, posto que, na prática, não dispomos dos dados reais, mas sim de previsões (que, logo, contém erros).

Comparamos os resultados aos de quatro *benchmarks* univariados, cujas previsões dependem apenas do histórico da carga elétrica, e não de dados exógenos.

Não encontramos um método que fosse infalível para todas as situações, e a característica de cada conjunto de dados foi determinante para os resultados que encontramos.

Para as zonas de carga A - WEST e F - CAPITL, localizadas nos extremos oeste e leste do estado, respectivamente, percebemos que utilizar os dados da única estação meteorológica disponível dentro da zona de carga já foi suficiente para produzir os melhores resultados utilizando FFNNs. Não houve benefício em acrescentar dados de mais fontes, o que reduz os custos de operação. É necessário ter atenção, no entanto, à qualidade dos dados exógenos: para níveis de ruído a partir de 4%, tanto para A - WEST quanto para F - CAPITL, as FFNNs podem produzir resultados cada vez menos acurados, ficando aquém do desempenho de um dos *benchmarks*, o amortecimento exponencial HWT.

Já para as zonas B - GENESE e C - CENTRL, a combinação das estações meteorológicas mais próximas das zonas de carga utilizando o método C1, proposto por nós, chegou, com as FFNNs, aos melhores resultados. Destacamos os resultados para B - GENESE: mesmo para os mais altos níveis de ruído nos dados exógenos, nenhum resultado teve desempenho abaixo dos *benchmarks*.

A zona G - HUDVL, cuja localização abrange regiões fisiográficas com diferentes altitudes (planaltos e planícies litorâneas), beneficiou-se da combinação de estações em regiões distintas (no planalto Apalache e nas planícies litorâneas). Nesse caso, o método C2, introduzido por nós, trouxe, com FFNNs, os melhores resultados, inclusive com robustez à injeção de ruídos: mesmo com os mais altos níveis de ruído, também não houve resultado inferior ao dos *benchmarks*, a exemplo de B - GENESE.

E - MHKVL é uma zona de carga com ampla extensão longitudinal, atravessando regiões de climas distintos. As melhores previsões, em mediana, vieram de FFNNs utilizando como entrada a média aritmética simples de algumas estações localizadas de norte a sul da zona de carga. Isto é plausível, já que capturar cada aspecto climático, sem distinção de pesos, parece razoável no caso de um território mais amplo e diverso. Porém, aqui há mais uma vez que se ter cautela quando o nível de ruído chega aos 4%, o que piora o desempenho das redes e torna, em alguns casos, preferível utilizar o HWT.

Finalmente, os resultados para D - NORTH foram os mais distintos do restante. De certa forma, esperávamos alguma diferença, já que esta zona de carga é a mais distante das estações meteorológicas e se localiza em uma região de clima bastante específico para que estações distantes consigam capturá-lo. Por conta disso, na maioria dos casos o HWT tem um desempenho melhor que o das redes neurais, especialmente para níveis de ruído de 2% ou mais.

Os resultados nos trazem indícios que corroboram a tese de [20]: considerar fatores práticos, como aqueles de ordem geográfica, pode trazer benefícios na hora de escolher ou combinar dados de diferentes fontes para previsão de carga dentro de um território diverso. Os métodos C1 e C2, propostos por nós, mostraram-se competitivos ao considerar fatores práticos como a geografia local e a correlação não-linear entre os dados, para situações para as quais é necessário combinar estações para ilustrar a diversidade climática local.

Além disso, os resultados de B - GENESE e G - HUDVL permitem dizer que os métodos demonstraram robustez à injeção de ruído, para estes dados.

Falando na robustez, vale retomar [19], quando, há vinte anos, os autores apontavam problemas na forma como se fazia a pesquisa em redes neurais. Era¹ comum os trabalhos darem pouca atenção ao fato de que as redes neurais podem ter dificuldades para generalização e, por isso, pode não ser suficiente exibir resultados para uma quantidade limitada de bases de dados, tamanho limitado de amostras, pequeno número de execuções (dada a sensibilidade à inicialização dos dados), *etc.*.

A análise de sensibilidade, em que treinamos 100 redes para cada um dos 5 níveis de ruído (e também sem ruído) nos dados de temperatura nos permitiu enxergar situações em que, dependendo da qualidade dos dados, seria possivelmente melhor abrir mão de métodos complexos como as redes neurais e utilizar um *benchmark* de séries temporais, reduzindo significativamente o custo computacional (e operacional, já que o método não depende de dados de fontes externas).

Finalmente, os resultados destacam a importância do conhecimento das bases de dados com as quais trabalhamos para o desempenho dos métodos e modelos, e da interpretação dos resultados. No contexto atual, em que dados são produzidos e disponibilizados em volumes cada vez maiores, é sedutor alimentar os algoritmos sem nos atentarmos a um aspecto fundamental da Estatística: a análise exploratória dos dados. As análises do Capítulo 2 nos permitiram conhecer melhor os dados com os quais trabalhamos e isso possibilitou um entendimento bastante aprofundado dos resultados - o que é fundamental para a tomada de decisões.

Importante mencionar como limitação o fato de não termos feito previsões probabilísticas; importantes para a previsão hierárquica de carga, mas que fogem do escopo deste trabalho.

9.1 TRABALHOS FUTUROS

Como possibilidades para trabalhos futuros, listamos:

- Experimentar os mesmos métodos em novos conjuntos de dados;
- Experimentar a combinação das estações meteorológicas atribuindo maior peso às estações meteorológicas cuja diferença de altitude até o centróide da zona de carga seja menor (*insight* vindo dos resultados de G - HUDVL, que combinou estações de planícies e planaltos);
- Experimentar como métodos de previsão redes neurais que consideram explicitamente a dependência temporal, como as redes LSTM;

¹E ainda parece ser.

- Experimentar o nllcor como medida para seleção de atributos de entrada nas redes neurais;
- No lugar de uma rede neural de 24 saídas, experimentar 24 redes, uma para cada hora do dia;
- Experimentar modelos de regressão linear múltipla com as mesmas entradas que as redes neurais;
- Alterar as séries temporais para características, para utilizar florestas aleatórias ou outros algoritmos de seleção de atributos;
- Experimentar os métodos de combinação no contexto de previsões probabilísticas.

REFERÊNCIAS

- 1 CRETÍ, Anna; FONTINI, Fulvio. **Economics of Electricity: Markets, Competitions and Rules**. Cambridge: Cambridge University Press, 2019.
- 2 DE OLIVEIRA, Adilson. Liberalização interrompida. **Revista USP**, [S. l.], n. 104, p. 51-62, 2015. DOI: 10.11606/issn.2316-9036.v0i104p51-62. Disponível em: <https://www.revistas.usp.br/revusp/article/view/106753>. Acesso em: 19 jan. 2021.
- 3 BUNN, Derek W. Forecasting loads and prices in competitive power markets. **Proceedings of the IEEE**, v. 88, n. 2, p. 51-62, 2000.
- 4 WERON, Rafał. **Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach**. Chichester: John Wiley & Sons Ltd., 2006.
- 5 OPERADOR NACIONAL DO SISTEMA ELÉTRICO - ONS. **O Sistema Interligado Nacional**. Disponível em: <http://www.bdt.org.br/sma/entendendo/atual.htm>. Acesso em: 22 jan. 2020.
- 6 SUPERINTENDÊNCIA DE REGULAÇÃO DOS SERVIÇOS DE DISTRIBUIÇÃO - SRD. **Regulação dos Serviços de Distribuição**. Disponível em: <https://www.aneel.gov.br/regulacao-dos-servicos-de-distribuicao>. Acesso em: 28 dez. 2021.
- 7 NAÇÕES UNIDAS BRASIL. **Como as Nações Unidas apoiam os Objetivos de Desenvolvimento Sustentável no Brasil**. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 12 abr. 2022.
- 8 NATIONAL AERONAUTICS AND SPACE ADMINISTRATION - NASA. **Climate Change: How do We Know?**. Disponível em: <https://climate.nasa.gov/evidence/>. Acesso em: 11 mar. 2021.
- 9 INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE - IPCC. **The Intergovernmental Panel on Climate Change**. Disponível em: <https://www.ipcc.ch/>. Acesso em: 11 mar. 2021.
- 10 INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE - IPCC. **Summary for policymakers**. Disponível em: https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_summary-for-policymakers.pdf. Acesso em: 11 mar. 2021.
- 11 NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY - NIST. **Engineering Statistics Handbook**. Disponível em: <http://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm>. Acesso em 16 Jun. 2021.
- 12 NASA's GODDARD INSTITUTE FOR SPACE STUDIES - NASA/GISS. **Land-Ocean Temperature Index**. Disponível em: https://data.giss.nasa.gov/gistemp/graphs/graph_data/Global_Mean_Estimates_based_on_Land_and_Ocean_Data/graph.txt. Acesso em: 11 mar. 2021.

- 13 EUROPEAN COMMISSION. **Renewable energy directive**. Disponível em: <https://ec.europa.eu/energy/topics/renewable-energy/renewable-energy-directive>. Acesso em: 11 mar. 2021.
- 14 HONG, Tao; FAN, Shu. Probabilistic electric load forecasting: a tutorial review. **International Journal of Forecasting**, v. 32, n. 3, p. 914-938, 2016.
- 15 UNITED NATIONS. Department of Economic and Social Affairs. **World Urbanization Prospects 2018**. Disponível em: <https://population.un.org/wup/Download/>. Acesso em: 16 fev. 2021.
- 16 GASSMANN, Oliver; BÖHM, Jonas; PALMIÉ, Maximilian. **Smart Cities: Introducing Digital Innovation to Cities**. Bingley: Emerald Publishing Limited, 2019.
- 17 TAYLOR, James W. An evaluation of methods for very short term electricity demand forecasting using minute-by-minute British data. **International Journal of Forecasting**, v. 24, n. 4, p. 645-658, 2008.
- 18 BUNN, Derek W.; FARMER, E. Dillon. **Comparative Models for Electrical Load Forecasting**. New York: Wiley, 1985.
- 19 HIPPERT, Henrique S.; PEDREIRA, Carlos Eduardo; SOUZA, Reinaldo C. Neural networks for short-term load forecasting: a review and evaluation. **IEEE Transactions on Power Systems**, v. 16, n. 1, p. 44-55, 2001.
- 20 HONG, Tao; PINSON, Pierre; FAN, Shu. Global Energy Forecasting Competition 2012. **International Journal of Forecasting**, v. 30, n. 2, p. 357-363, 2014.
- 21 HONG, Tao.; PINSON, Pierre; FAN, Shu; ZAREIPOUR, Hamidreza; TROCCOLI, Alberto; HYNDMAN, Rob J. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. **International Journal of Forecasting**, v. 32, n. 3, p. 896-913, 2016.
- 22 THE EUROPEAN NETWORK OF EXCELLENCE ON INTELLIGENT TECHNOLOGIES FOR SMART ADAPTIVE SYSTEMS - EUNITE. **World-wide competition within the EUNITE network**. Disponível em: <http://neuron.tuke.sk/competition/index.php>. Acesso em: 05 Mar. 2017.
- 23 INTERNET ARCHIVE. **Wayback Machine**. Disponível em: <https://archive.org/web/>. Acesso em 18 mar. 2021.
- 24 EL-SHARKAWI, Mohamed A. EE 555: Fundamentals of Intelligent Systems; Data. Disponível em: https://class.ece.uw.edu/555/el-sharkawi/index_files/Page3404.htm. Acesso em 18 Mar. 2021.
- 25 ARORA, Siddharth; TAYLOR, James W. Short-term forecasting of anomalous load using rule-based triple seasonal methods. **IEEE Transactions on Power Systems**, v. 28, n. 3, p. 3235-3242, 2013.
- 26 SUPPES, Gallen J. e STORVICK, Truman S. **Sustainable Nuclear Power: A volume in sustainable world**. Cambridge: Academic Press, 2007.
- 27 TAYLOR, James .W. Short-term load forecasting with exponentially weighted methods. **IEEE Transactions on Power Systems**, v. 27, n. 1, p. 458-464, 2011.

- 28 HONG, Tao. **Temperature-based Models vs. Time Series Models**. Disponível em: <http://blog.drhongtao.com/2018/12/temperature-based-models-vs-time-series-models.html>. Acesso em: 22 jan. 2020.
- 29 GUILHERMINO NETO, Guilherme. **Métodos Univariados e Multivariados para Previsão de Demanda de Energia Elétrica em Curto Prazo: Um Estudo Comparativo**. Orientador: Henrique Steinherz Hippert, D.Sc. 2014. 69 f. Dissertação (mestrado) - Programa de Pós-Graduação em Modelagem Computacional - Universidade Federal de Juiz de Fora (UFJF). Defendida em 20 ago. 2014. Disponível em: <https://www2.ufjf.br/pgmc/teses-e-dissertacoes/dissertacoes/>. Acesso em 28 dez. 2021.
- 30 DEFILIPPO, Samuel Belini. **Previsão da Demanda de Energia Elétrica por Combinações de Modelos Lineares e de Inteligência Computacional**. Orientador: Henrique Steinherz Hippert, D.Sc. 2017. 109 f. Tese (doutorado). Programa de Pós-Graduação em Modelagem Computacional - Universidade Federal de Juiz de Fora (UFJF). Defendida em 20 set. 2017. Disponível em: <https://repositorio.ufjf.br/jspui/bitstream/ufjf/6036/1/samuelbelinidefilippo.pdf>. Acesso em 28 dez. 2021.
- 31 MAYRINK, Victor Teixeira de Melo. **Avaliação do algoritmo gradient boosting em aplicações de previsão de carga elétrica a curto prazo**. Orientador: Henrique Steinherz Hippert, D.Sc. 2016. 91 f. Dissertação (mestrado) - Programa de Pós-Graduação em Modelagem Computacional - Universidade Federal de Juiz de Fora (UFJF). Defendida em 31 ago. 2016. Disponível em: <https://www2.ufjf.br/pgmc/teses-e-dissertacoes/dissertacoes/>. Acesso em 12 abr. 2022.
- 32 TEIXEIRA, Raphael Francisco Firmiano. **Estudo de técnicas de apoio a definições em contratos de energia elétrica**. Orientador: Henrique Steinherz Hippert, D.Sc. 2017. 57 f. Dissertação (mestrado) - Programa de Pós-Graduação em Modelagem Computacional - Universidade Federal de Juiz de Fora (UFJF). Defendida em 30 ago. 2017. Disponível em: <https://www2.ufjf.br/pgmc/teses-e-dissertacoes/dissertacoes/>. Acesso em 12 abr. 2022.
- 33 PEDREIRA, Taís de Medeiros. **Um estudo sobre os métodos de amortecimento exponencial para a previsão de carga a curto prazo**. Orientador: Henrique Steinherz Hippert, D.Sc. 2018. 49 f. Dissertação (mestrado) - Programa de Pós-Graduação em Modelagem Computacional - Universidade Federal de Juiz de Fora (UFJF). Defendida em 05 set. 2018. Disponível em: <https://www2.ufjf.br/pgmc/teses-e-dissertacoes/dissertacoes/>. Acesso em 12 abr. 2022.
- 34 DE AZEVEDO, Iago Mansur. **Previsão de carga elétrica a curto prazo combinando métodos lineares univariados e variáveis exógenas através de algoritmos genéticos**. Orientador: Henrique Steinherz Hippert, D.Sc. 2019. 76 f. Dissertação (mestrado) - Programa de Pós-Graduação em Modelagem Computacional - Universidade Federal de Juiz de Fora (UFJF). Defendida em 13 mar. 2019. Disponível em: <https://www2.ufjf.br/pgmc/teses-e-dissertacoes/dissertacoes/>. Acesso em 12 abr. 2022.

- 35 ACURITE. **What is a weather station?** Disponível em:
<https://www.acurite.com/learn/weather-stations/what-is-a-weather-station>. Acesso em: 22 jan. 2020.
- 36 HONG, Tao; WANG, Pu. e WHITE, Laura. Weather Station Selection for Electric Load Forecasting. **IEEE Transactions on Power Systems**, v. 31, n. 2, p. 286-295, 2015.
- 37 SOBHANI, Masoud; CAMPBELL, Alisson., SANGAMWAR, Saurabh., LI, Changlin. e HONG, Tao. Combining Weather Stations for Electric Load Forecasting. **Energies**, v. 12, n. 8, 1510, 2019.
- 38 BOYA, Carlos; ARDILA-REY, Jorge. A Method for Weather Station Selection Based on Wavelet Squared Coherence for Electric Load Forecasting. **IEEE Access**, v. 8, n.8, pp.197431-197438, 2020.
- 39 NEW YORK STATE - DEPARTMENT OF PUBLIC SERVICE. **New York Independent System Operator (NYISO)**. Disponível em:
<http://www3.dps.ny.gov/>. Acesso em: 26 jan. 2020.
- 40 RTO INSIDER. **NYISO Management Committee Briefs: March 28,2018**. Disponível em:
<https://rtoinsider.com/nyiso-management-committee-deliverability-rights-89415/>. Acesso em 26 jan. 2020.
- 41 NEW YORK STATE. **New York State Civil Boundaries**. Disponível em:
<http://gis.ny.gov/civil-boundaries/>. Acesso em 10 mai. 2020.
- 42 EUROPEAN NETWORK OF TRANSMISSION SYSTEM OPERADORES FOR ELECTRICITY (ENTSO-E). **Transparency platform**. Disponível em:
<https://transparency.entsoe.eu/dashboard/show>. Acesso em: 31 mar. 2021.
- 43 NEW YORK INDEPENDENT SYSTEM OPERATOR - NYISO. **Real-time dashboard**. Disponível em: <https://www.nyiso.com/real-time-dashboard>. Acesso em: 28 jan. 2020.
- 44 NYISO ENERGY MARKET OPERATIONS. **Day-ahead scheduling manual**. Rensselaer: New York Independent System Operator, 2017.
- 45 BRASIL. Lei N^o 9.074, de 7 de julho de 1995. **Estabelece normas para outorga e prorrogações das concessões e permissões de serviços públicos e dá outras providências**. Disponível em:
http://www.planalto.gov.br/ccivil_03/leis/19074compilada.htm. Acesso em 10 mai. 2021.
- 46 IGAVA, Louise. **Zone A outline**. Disponível em:
<https://www.arcgis.com/home/item.html?id=57daa251b89e48d49838351441f31ac2>. Acesso em 10 mai. 2021.
- 47 IGAVA, Louise. **Zone B one outline**. Disponível em:
<https://www.arcgis.com/home/item.html?id=99fa9c4d22274fa7bc813e48d0ee28cd>. Acesso em 10 mai. 2021.

- 48 IGAVA, Louise. **Zone B two outline**. Disponível em:
<https://www.arcgis.com/home/item.html?id=a6054c840e9047d9a648ed89d1c741ea>.
Acesso em 10 mai. 2021.
- 49 IGAVA, Louise. **Zone C outline**. Disponível em:
<https://www.arcgis.com/home/item.html?id=203cf0c4c66c451bb8f220d685bffba8>.
Acesso em 10 mai. 2021.
- 50 IGAVA, Louise. **Zone D outline**. Disponível em:
<https://www.arcgis.com/home/item.html?id=e6094f4bc6af4f83b532e8439661cf4a>.
Acesso em 10 mai. 2021.
- 51 IGAVA, Louise. **Zone E outline**. Disponível em:
<https://www.arcgis.com/home/item.html?id=c5f1ca6b8a87454da717fd28872a3f6e>.
Acesso em 10 mai. 2021.
- 52 IGAVA, Louise. **Zone F outline**. Disponível em:
<https://www.arcgis.com/home/item.html?id=a30b7eeb350249f48de26276e1c3139d>.
Acesso em 10 mai. 2021.
- 53 IGAVA, Louise. **Zone G outline**. Disponível em:
<https://www.arcgis.com/home/item.html?id=0a4d949c53844654904efa0ee38953c3>.
Acesso em 10 mai. 2021.
- 54 NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION - NOAA. **About our agency**. Disponível em: <https://www.noaa.gov/about-our-agency>.
Acesso em 26. jan. 2020.
- 55 NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION. **Integrated Surface Database**. Disponível em
<https://www.ncei.noaa.gov/products/land-based-station/integrated-surface-database>.
Acesso em 29 ago. 2021.
- 56 VASQUEZ, Tim. **Weather Station Identifiers**. Disponível em:
<http://www.weathergraphics.com/identifiers/>. Acesso em 26. jan. 2020.
- 57 NOAA National Centers for Environmental Information. **Global Surface Hourly** (subset 722160, 726227, 725150, 725280, 725156, 725037, 744860, 725030, 726225, 726223, 725145, 725036, 725196, 725290, 725038, 725190 - 2015 to 2019). NOAA National Centers for Environmental Information. Acesso em 28 ago. 2021.
- 58 U.S. STATE DEPARTMENT. **American Holidays**. Disponível em
<https://usa.usembassy.de/holidays.htm>. Acesso em 11 mai. 2021.
- 59 CURRENT RESULTS. **Weather and science facts**. Disponível em
<https://www.currentresults.com/>. Acesso em 29 ago. 2021.
- 60 Campbell, ALAN K.; SCUDIÈRE, PAUL JOSEPH. **New York**. Encyclopedia Britannica. 22 out. 2020, Disponível em:
<https://www.britannica.com/place/New-York-state>. Acesso em: 3 abr. 2021.

- 61 MICHIGAN SEA GRANT. **Teaching Great Lakes Science - Lessons & Data Sets**: Microclimates. Disponível em <https://www.michiganseagrant.org/lessons/lessons/by-broad-concept/earth-science/microclimates/>. Acesso em 10 mai. 2021.
- 62 NOAA (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION). **What is a Lake Effect Snow?** Disponível em <https://www.weather.gov/safety/winter-lake-effect-snow>. Acesso em 10 mai. 2021.
- 63 ADIRODACK COUNCIL. **Climate Change**. Disponível em <https://www.adirondackcouncil.org/page/climate-change-88.html>. Acesso em 10 mai. 2021.
- 64 ST. LAWRENCE COUNTY GOVERNMENT. **St. Lawrence County Characteristics**. Disponível em <https://www.stlawco.org/About/AboutCharacteristics>. Acesso em 10 mai. 2021.
- 65 NATIONAL GEOGRAPHIC. **Season**. Disponível em: <https://www.nationalgeographic.org/encyclopedia/season/>. Acesso em 26. jan. 2020.
- 66 US ENERGY INFORMATION ADMINISTRATION - EIA. **Household Energy Use in New York**: A closer look at residential energy consumption, 2009.
- 67 CUBIT. **New York Cities by Population**. Disponível em: https://www.newyork-demographics.com/cities_by_population. Acesso em 13 ago. 2021.
- 68 CALS - COLLEGE OF AGRICULTURE AND LIFE SCIENCES. **Climate change in the Hudson Valley**. Disponível em <https://wri.cals.cornell.edu/hudson-river-estuary/climate-change-hudson-river-estuary/resources-resilience/climate-projections-nys/>. Acesso em 29 ago. 2021.
- 69 HYNDMAN, Robert J.; ATHANASOPOULOS, George. **Forecasting**: Principles and practice. 2 ed. Melbourne: Otexts, 2018. Disponível em [Otexts.com/fpp2](https://www.otexts.com/fpp2). Acesso em 26 jan 2020.
- 70 WORLD HEALTH ORGANIZATION. **WHO Coronavirus (COVID-19) Dashboard**. Disponível em <https://covid19.who.int/>. Acesso em 09 jul 2021.
- 71 WORLD HEALTH ORGANIZATION. **Global - Brazil**. Disponível em <https://covid19.who.int/region/amro/country/br>. Acesso em 09 jul 2021.
- 72 KRARTI, Moncef; ALDUBYAN, Mohammad. Review analysis of COVID-19 impact on electricity demand for residential buildings. **Renewable and Sustainable Energy Reviews**, v. 143, 2021. doi.org/10.1016/j.rser.2021.110888
- 73 RAMANATHAN, Ramu; ROBERT, Engle; GRANGER, Clive W.J.; VAHID-ARAGHI, Farshid; BRACE, Casey. Short-run forecasts of electricity loads and peaks. **International Journal of Forecasting**, vol. 13, n.2, pp. 161-174, 1997.

- 74 COCHRANE, D.; ORCUTT, G. H. Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms. **Journal of the American Statistical Association**, vol. 44, n. 245, pp. 32–61, 1949.
- 75 BURKOV, Andriy. **The Hundred-Page Machine Learning Book**. Monee: 2019.
- 76 WIDROW, Bernard; LEHR, Michael A. 30 years of adaptive neural networks: perceptron, Madaline, and backpropagation. **Proceedings of the IEEE**, v. 78, n. 9, pp. 1415-1442, 1990.
- 77 KHOTANZAD, Alireza; AFKHAMI-ROHANI, Reza; LU, Tsun-Liang; ABAYE, Alireza.; DAVIS, Malcolm.; MARATUKULAM, Dominic J. ANNSTLF - a neural-network-based electric load forecasting system. **IEEE Transactions on Neural Networks**, vol. 8, n.4, pp. 835–846, 1997.
- 78 SOUTHERN METHODIST UNIVERSITY. **A Service of Memory**. Disponível em <https://calendar.smu.edu/site/perkins/event/university-memorial-service/>. Acesso em 09 jul. 2021.
- 79 ENVERUS TRADING & RISK. About PRT. Disponível em <https://www.prtforecast.com/about/>. Acesso em 09 jul. 2021.
- 80 TAYLOR, James W. Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing. **Journal of Operational Research Society**, v. 54, n. 8, pp. 799-805, 2003.
- 81 PEDREIRA, Taís de Medeiros. **Um Estudo Sobre os Métodos de Amortecimento Exponencial para Previsão de Carga a Curto Prazo**. Orientador: Henrique Steinherz Hippert, D.Sc. 2018. 47 f. Dissertação (mestrado). Programa de Pós-Graduação em Modelagem Computacional - Universidade Federal de Juiz de Fora (UFJF). Defendida em 5 set. 2018. Disponível em: <https://repositorio.ufjf.br/jspui/bitstream/ufjf/8020/1/taisdemedeirospedreira.pdf>. Acesso em 28 dez. 2021.
- 82 VALLEY NEWS. **Milestones - Robert Goodell Brown**. Disponível em: <https://www.vnews.com/Archives/2013/10/Robert-Brown-obit-vn-100813> . Acesso em 21 jul. 2021.
- 83 ARTHUR D. LITTLE, INC.; BROWN, Robert Goodell. Exponential Smoothing for Predicting Demand. **Philip Morris Records; Master Settlement Agreement**. Unknown. 1956. Disponível em: <https://www.industrydocuments.ucsf.edu/docs/jzlc0130> . Acesso em 21 jul. 2021.
- 84 HOLT, Charles. C. Forecasting seasonals and trends by exponentially weighted moving averages. **International Journal of Forecasting**, v. 20, n. 1, pp. 5–10, 2004.
- 85 JADHAV, Pritish. **Proving Convexity of the MSE Loss Function**. Disponível em <https://medium.com/swlh/proving-convexity-of-mean-squared-error-loss-function-68b139fe6ea7>. Acesso em 14 Jul. 2021.
- 86 NOCEDAL, Jorge; WRIGHT, Stephen J. **Numerical Optimization**. New York: Springer, 2006.

- 87 SINGER, Yaron. **Am 221 - advanced optimization:** Lecture 9. 25 jan. 2019, 27 apr. 2019. Notas de Aula.
- 88 TIBSHIRANI, Ryan. **10-725 - Optimization:** Lecture 6. 27 ago. 2013, 13 dez. 2013. Notas de Aula.
- 89 POCZOS, Barnabas; TIBSHIRANI, Ryan. **10-725 - Optimization:** Lecture 9: Newton Method. 27 ago. 2013, 13 dez. 2013. Notas de Aula.
- 90 LAM, Adrian. **BFGS in a nutshell:** An Introduction to Quasi-Newton Methods. Disponível em <https://towardsdatascience.com/bfgs-in-a-nutshell-an-introduction-to-quasi-newton-methods-21b0e13ee504> . Acesso em 19 jul. 2021.
- 91 PENG, Roger D. **Advanced Statistical Computing.** Disponível em <https://leanpub.com/advstatcomp/> . Acesso em 17 abr. 2022.
- 92 DENNIS JR., John E. **A page of memories.** Disponível em <https://www.caam.rice.edu/~dennis/pa.html> . Acesso em 20 jul. 2021.
- 93 BOX, George E.P.; JENKINS, Gwilym M.; REINSEL, Gregory C. e Ljung, Greta M. **Time Series Anaysis:** Forecasting and Control. 5. ed. Hoboken: Wiley, 2015.
- 94 CARO, E; JÉSUS, J. e CARA, J. Periodically correlated models for short-term electricity load forecasting. **Applied Mathematics and Computation**, v. 364, p. 1-13, 2020.
- 95 HYNDMAN, Rob. J.; KHANDAKAR, Yeasmin. Automatic Time Series Forecasting: The forecast Package for R. **Journal of Statistical Software**, v. 27, n. 1, p. 1-22, 2008.
- 96 KWIATKOWSKI, Denis; PHILLIPS, Peter C.B.; SCHMIDT, Peter; SHIN, Yongcheol. Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. **Journal of Econometrics**, v. 54, n. 1-3, pp. 159–178, 1992.
- 97 CANOVA, Fabio; HANSEN, Bruce E. Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability. **Journal of Business and Economic Statistics**, v. 13, n. 3, pp. 237–252., 1995.
- 98 TRIACCA, Umberto. **Lesson 12:** Estimation of the parameters of an ARMA model. Disponível em <http://www.phdeconomics.sssup.it/documents/Lesson12.pdf> . Acesso em 28 jul. 2021.
- 99 KOTZ, Samuel (Founder and Editor-in-Chief); BALAKRISHNAN, N., READ, Campbell B., VIDA KOVIC, Brani (Editors-in-Chief); JOHNSON, Norman L. (Founder and Chairman of Editorial Board). **Encyclopedia of Statistical Sciences.** Hoboken: Wiley, 2006.
- 100 R CORE TEAM. **Stats.** R Package. Version 4.2.0. 2022.
- 101 HEATH, Michael T. **Scientific Computing:** An Introductory Survey. Nova Iorque: McGraw Hill, 1997.

- 102 HONG, Tao; WANG, Pu. e WILLIS, H. Lee. “A naive multiple linear regression benchmark for short term load forecasting”. **IEEE 2011 Power and Energy Society General Meeting**. Detroit, 2011.
- 103 CYBERENKO, G. Approximations by superpositions of sigmoidal functions. **Mathematics of Control, Signals, and Systems**, v. 2, n. 4, p. 303-314, 1989.
- 104 BISHOP, Christopher M. **Neural Networks for Pattern Recognition**. New York: Oxford University Press, 1995.
- 105 MUELLER, Vincent. Backpropagation in Neural Networks. Disponível em <<https://towardsdatascience.com/backpropagation-in-neural-networks-6561e1268da8>> . Acesso em 07 Out. 2021.
- 106 BENGIO, Yoshua. Practical recommendations for gradient-based training of deep architectures. *In: Neural networks: Tricks of the trade*, pp. 437–478. Springer, 2012,
- 107 BISHOP, Christopher M. **Pattern Recognition and Machine Learning**. Cambridge: Springer, 2006.
- 108 GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. MIT Press, 2016. Disponível em <<http://www.deeplearningbook.org>>
- 109 US DEPARTMENT OF THE INTERIOR. **Geographic Centers of the United States**. Survey: 4. 1964.
- 110 RODRIGUE, Jean Paul. **The Geography of Transport Systems**. Overview. Disponível em: <https://transportgeography.org/>. Acesso em 08 Dez. 2021.
- 111 HER MAJESTY’S STATIONERY OFFICE. **Admiralty Manual of Navigation**, Volume 1. p. 10. London, 1987.
- 112 CAMBRIDGE UNIVERSITY. **Underground Mathematics: The great circle distance**. Disponível em <https://undergroundmathematics.org/trigonometry-compound-angles/the-great-circle-distance> . Acesso em 18 abr. 2022.
- 113 RANJAN, Chitta. e SISI, Vahab Najari. **Processminer/nlcor**. Disponível em: <https://github.com/ProcessMiner/nlcor/blob/master/DESCRIPTION>. Acesso em 26 jan. 2020.
- 114 RANJAN, Chitta. **Estimating Non-linear Correlation in R**. Disponível em: <<https://towardsdatascience.com/estimating-non-linear-correlation-in-r-62c6571cb1db>>. Acesso em 09 Dez. 2021.
- 115 MAGALHÃES, Marcos Nascimento e DE LIMA, Antonio Carlos Pedroso. **Noções de Probabilidade e Estatística**. 7. ed. revista. São Paulo: Edusp, 2015.
- 116 CHATFIELD, Chris. **The Analysis of Time Series: An Introduction**. 5th ed. Boca Raton: Chapman & Hall/CRC, 1996.

- 117 HIPPERT, Henrique S. **Projeto Sigma**: Introdução às Séries temporais. Disponível em: <https://projetosigma.org/series-temporais/>. Acesso em 09 Jan. 2022.
- 118 HYNDMAN, Rob J. **Cross-validation for time series**. Disponível em: <https://robjhyndman.com/hyndsight/tscv/>. Acesso em: 30 Jan. 2020.
- 119 R CORE TEAM. **R**: A Language and Environment for Statistical Computing. Vienna, Austria. Disponível em <https://www.R-project.org>. Acesso em 12 Jan. 2022.
- 120 FALBEL, D.; ALLAIRE, J.J.; CHOLLET, F.; Rstudio; Google; TANGM Y.; VAN DER BIJL, W.; STUDER, M.; KEYDANA, S. **keras**: R Interface to Keras. R package version 2.3.0.0. 2020.
- 121 HYNDMAN, R.; ATHANASOPOULOS, R.; BERGMEIR, C.; CACERES, G.; CHHAY, L.; O'HARA-WILD, M.; PETROPOULOS, F.; RAZBASH, S. **forecast**: Forecasting functions for time series and linear models. R package version 8.13. 2020.
- 122 HIJMANS, R.J.; WILLIAMS, E.; VENNESE, C. **geosphere**: Spherical Trigonometry. R package version 1.5-10. 2019.
- 123 GUILHERMINO NETO, Guilherme. **ggn4488/tese**. Disponível em <https://github.com/ggn4488/tesePGMC>. Acesso em 12 Jan. 2022.

APÊNDICE A – Séries temporais: conceitos e notação

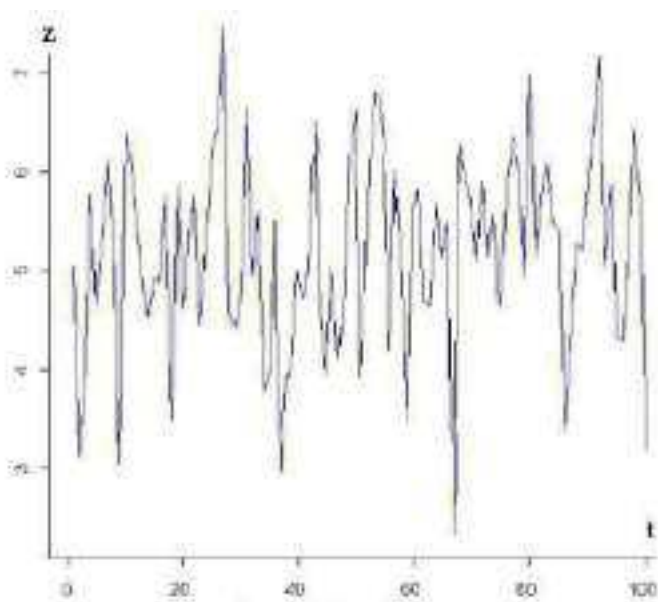
Uma **série temporal** é um conjunto de observações feitas ao longo do tempo [116]. Dados de séries temporais aparecem em várias áreas do conhecimento: na Engenharia, podemos acompanhar a temperatura de um material minuto a minuto; na Economia, podemos analisar o valor de um indicador para a inflação mensalmente; no Controle de Processos, podemos monitorar dado parâmetro a cada hora, e assim por diante. Os principais objetivos de se trabalhar com séries temporais no contexto de cargas elétricas é descrever **padrões** ao longo do tempo e utilizar estes padrões para fazer previsões para instantes de tempo futuros.

A.1 PADRÕES EM UMA SÉRIE TEMPORAL

Os principais padrões em uma série temporal são o **nível**, a **tendência** e a **sazonalidade**.

Chamamos de **nível** o valor médio de uma série temporal, se excluirmos os outros padrões. A série da Figura A.1 tem um nível médio constante $l = 5$, com flutuações aleatórias em torno deste nível ao longo do tempo:

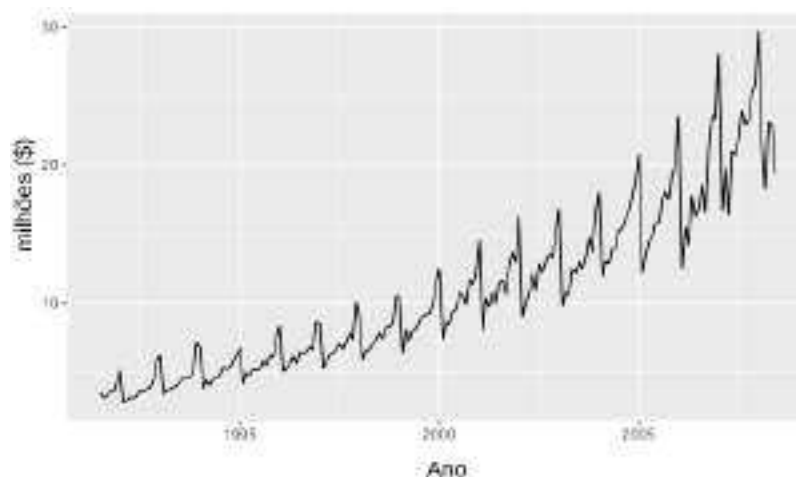
Figura A.1 – Série de dados simulados.



Fonte: [117]

Uma série possui **tendência** se há acréscimos ou decréscimos de longo prazo nos dados [69]. Não necessariamente a tendência é linear. A série da Figura A.2 possui tendência:

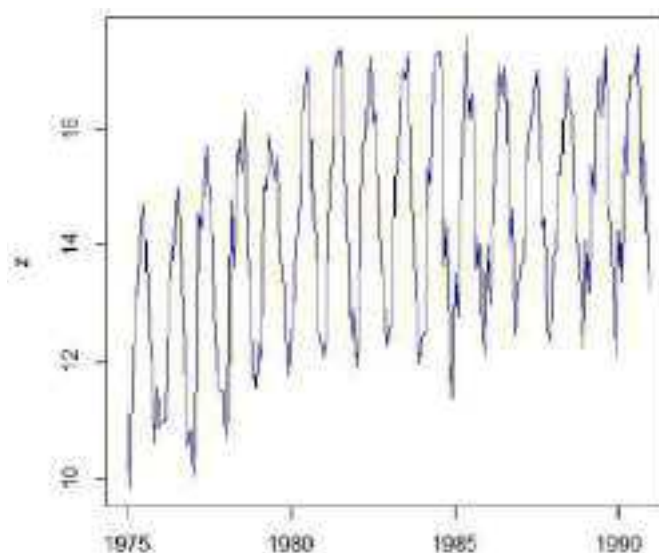
Figura A.2 – Vendas mensais de drogas antidiabéticas na Austrália.



Fonte: [69]

A **sazonalidade** é um padrão que acontece quando a série temporal é afetada por fatores sazonais, como a época do ano ou o dia da semana, com uma frequência fixa e conhecida [69]. Na Figura A.3, vemos uma série que, além da tendência de crescimento, possui sazonalidade, com valores mais altos acontecendo sempre no meio do ano:

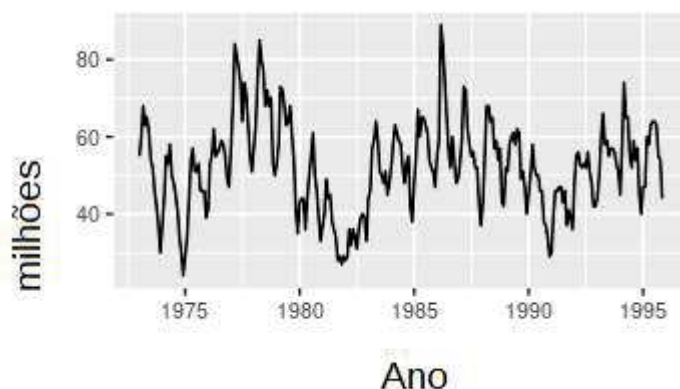
Figura A.3 – Venda de cerveja na Austrália, em milhões de barris.



Fonte: [117]

Existe outro padrão possível para flutuações periódicas, o chamado **ciclo**. A diferença entre ciclo e sazonalidade é que, quando há ciclo, os dados exibem sobes e descas que não têm uma frequência fixa [69]. Nos dados da Figura A.4, a venda mensal de casas apresenta sazonalidade dentro de cada ano, mas também comportamento cíclico, em períodos de seis a dez anos:

Figura A.4 – Venda de novas casas nos EUA.



Fonte: [69]

A.2 NOTAÇÃO E TERMINOLOGIA

Nesta tese, utilizamos a notação y_t para o valor de uma série temporal y_1, y_2, \dots, y_T , observado no instante y_t , sendo T o último instante.

O **horizonte de previsão**, que denotamos por k , será o número de passos à frente para o qual gostaríamos de fazer a previsão.

Para diferenciar um valor **previsto**, de um **observado** no instante t , acrescentamos o acento circunflexo e escrevemos \hat{y}_t quando estamos falando de uma previsão.

Expressamos uma previsão k passos à frente de T , feita no instante T (alternativamente, utilizando todos os dados disponíveis até T), como a seguir:

$$\hat{y}_{T+k|T} \tag{A.1}$$

Ocasionalmente omitimos o ponto de partida da previsão e escrevemos apenas \hat{Y}_{T+k} , porque costumamos utilizar todos os dados disponíveis até o último instante para fazer as previsões.

O **erro de previsão** no instante t é a diferença entre o valor observado e o previsto. O expressamos, portanto, da seguinte forma:

$$e_t = y_t - \hat{y}_t \tag{A.2}$$

APÊNDICE B – Mais gráficos e estatísticas para análise exploratória dos dados.

Limitamos a quantidade de figuras e tabelas no Capítulo 2 para que o texto ficasse mais conciso. Neste apêndice trazemos os gráficos e as estatísticas referentes a todas as zonas de carga e estações meteorológicas de nossa base de dados.

Na Figura B.1, mostramos as cargas de uma semana de inverno e outra de verão (amostradas por conveniência) para as estações B até G¹.

Nas Figuras B.2 a B.8, mostramos os diagramas de dispersão para cargas das zonas A até G *versus* temperaturas de todas as estações meteorológicas. As cargas e temperaturas nos diagramas são as médias diárias, e o nlor para os dados também está nas figuras.

Nas Tabelas B.1 a B.11, mostramos estatísticas descritivas para as temperaturas de cada estação meteorológica, separadas por estação do ano.

¹O gráfico para A - WEST está no Capítulo 2.

Figura B.1 – Cargas de uma semana de inverno e outra de verão, amostradas por conveniência; zonas de carga B a G.

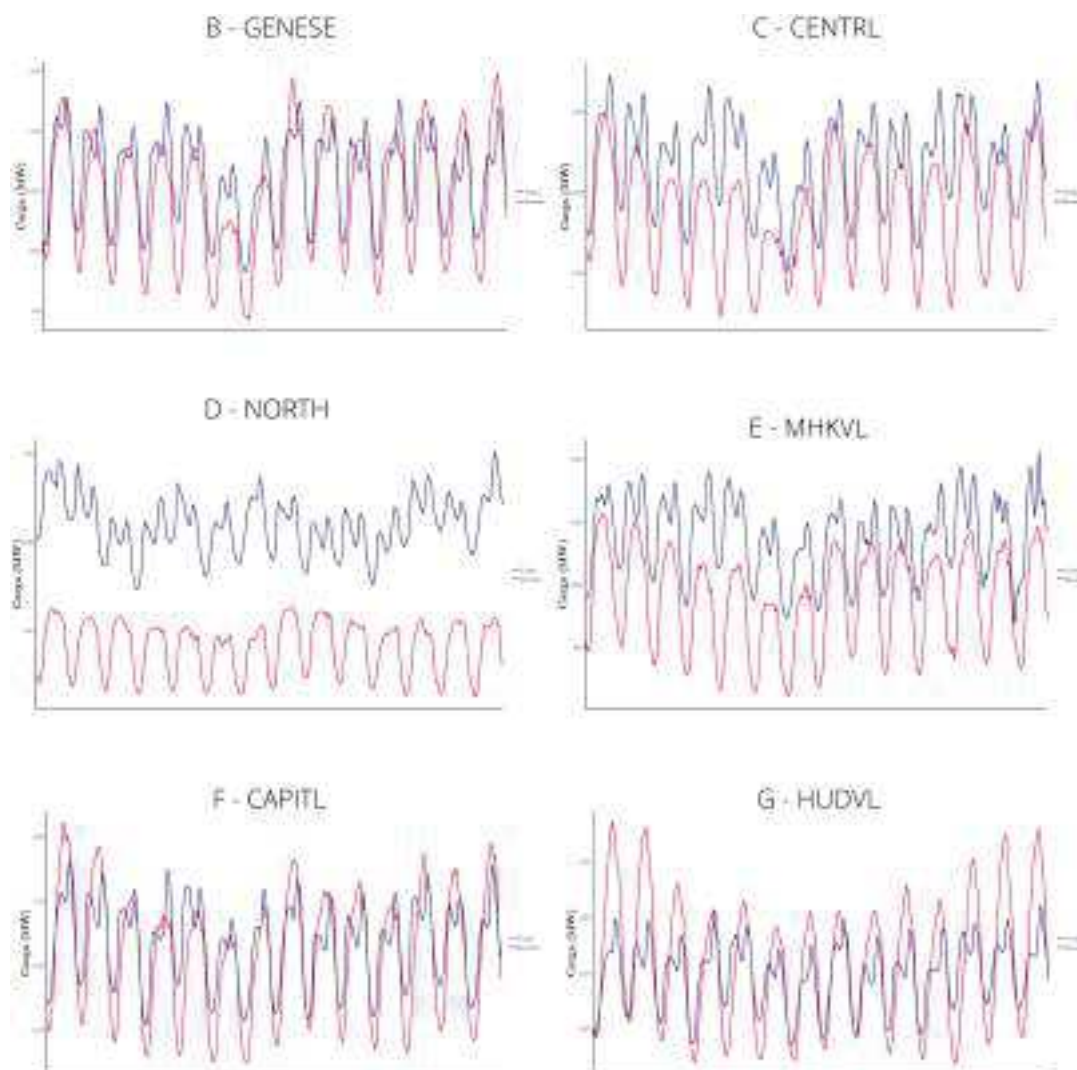


Figura B.2 – Cargas *versus* temperaturas (médias diárias) para a zona A-WEST e todas as estações meteorológicas.

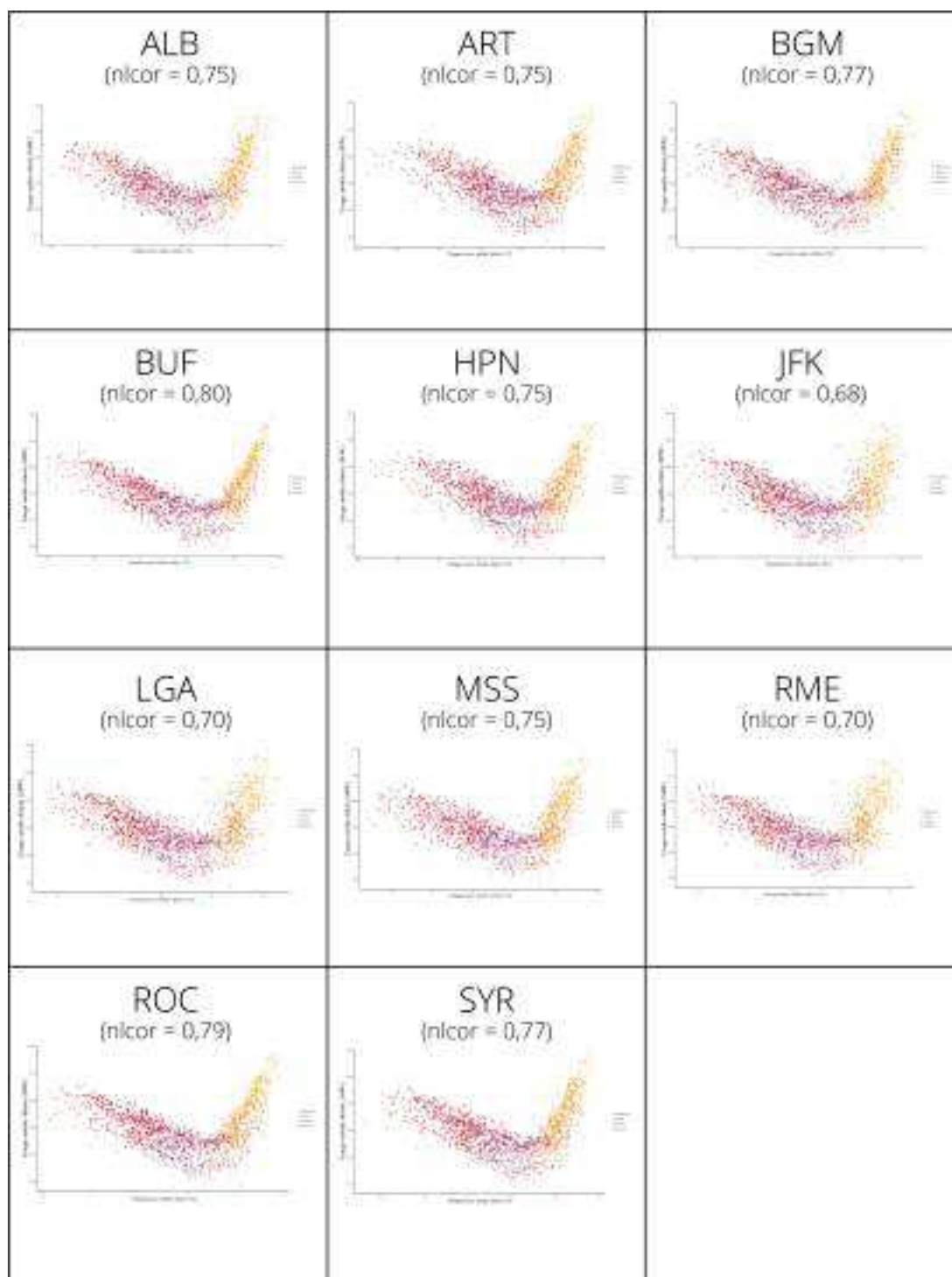


Figura B.3 – Cargas *versus* temperaturas (médias diárias) para a zona B-GENESE e todas as estações meteorológicas.

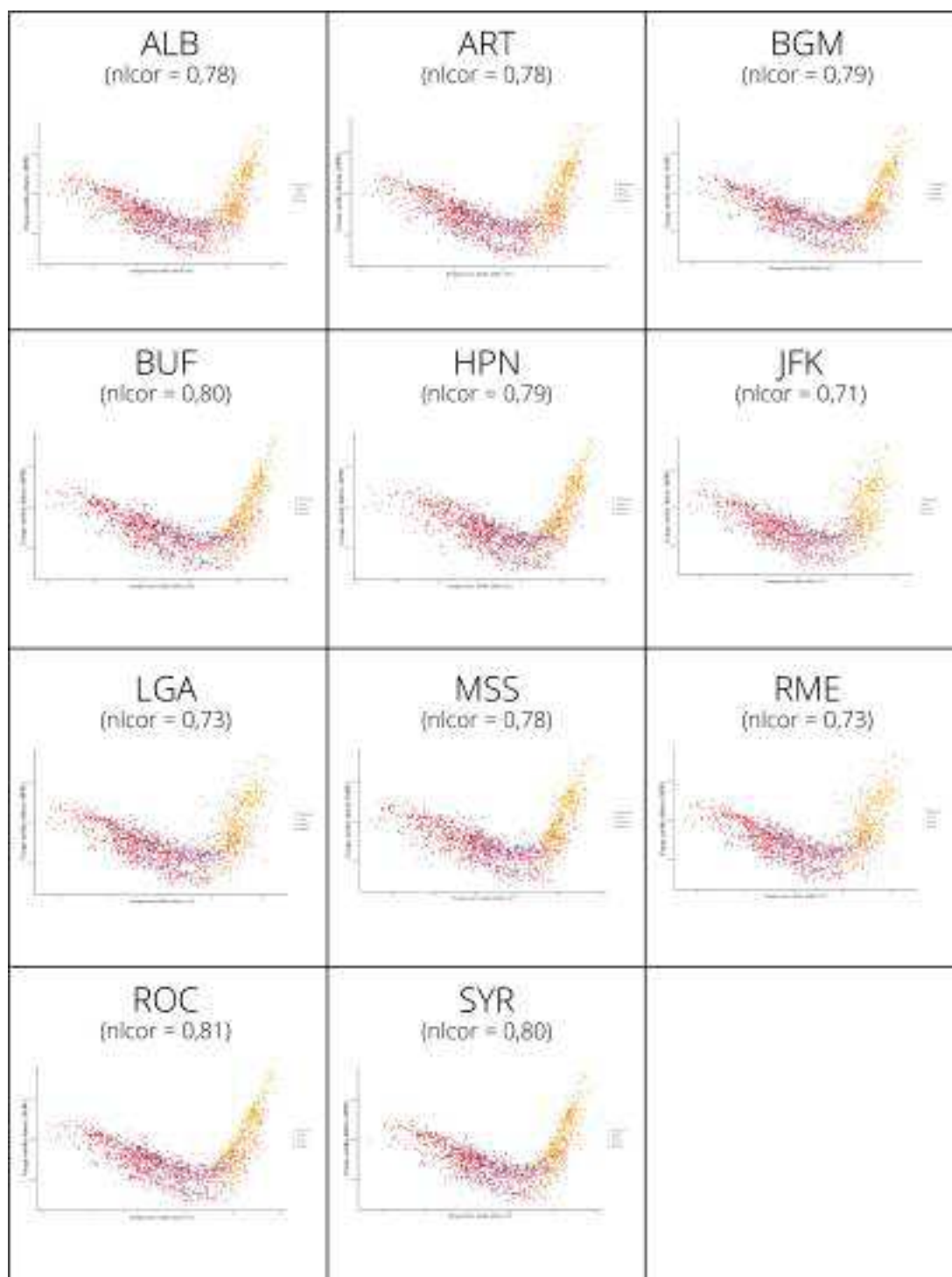


Figura B.4 – Cargas *versus* temperaturas (médias diárias) para a zona C-CENTRL e todas as estações meteorológicas.

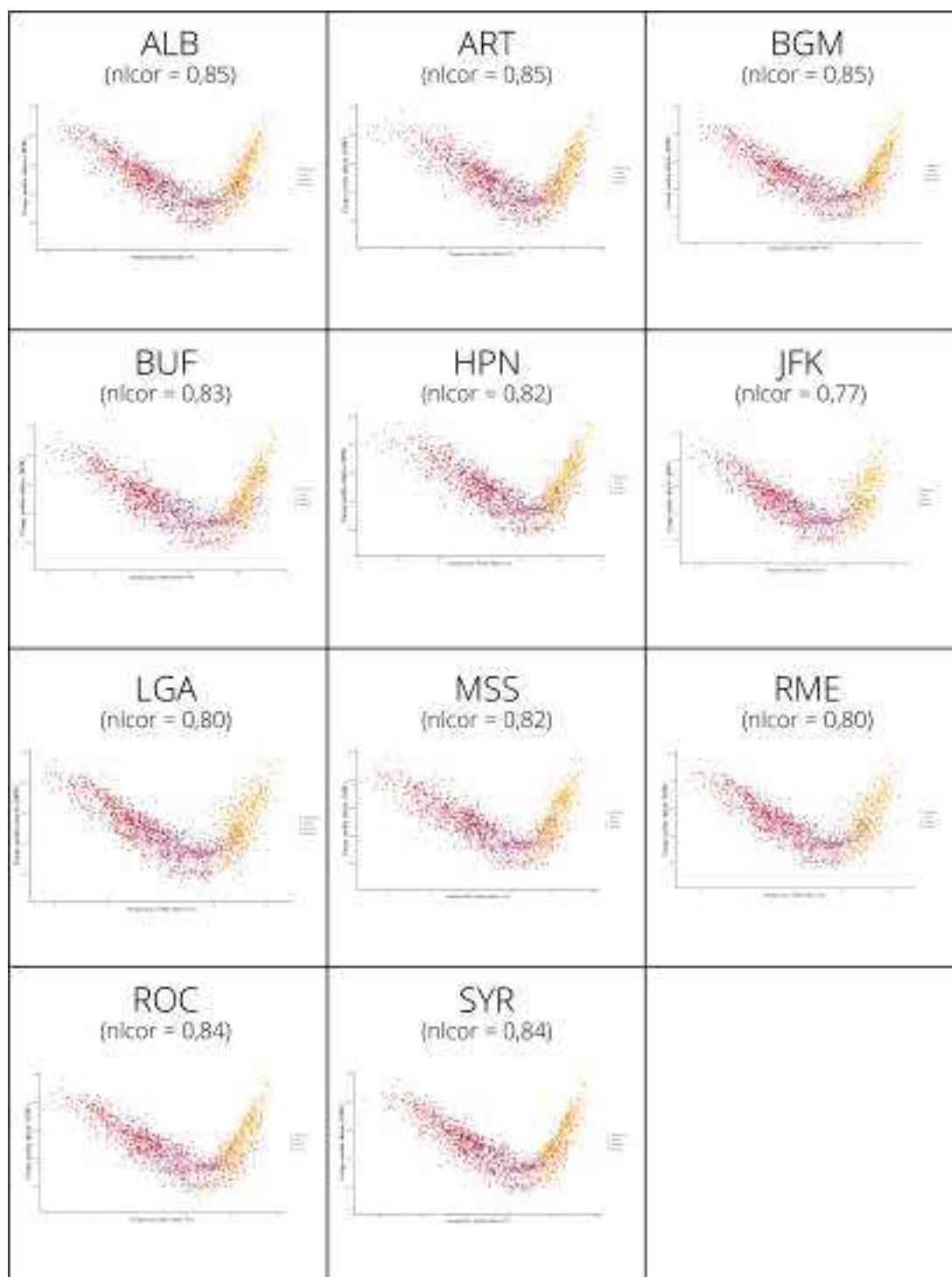


Figura B.5 – Cargas *versus* temperaturas (médias diárias) para a zona D-NORTH e todas as estações meteorológicas.

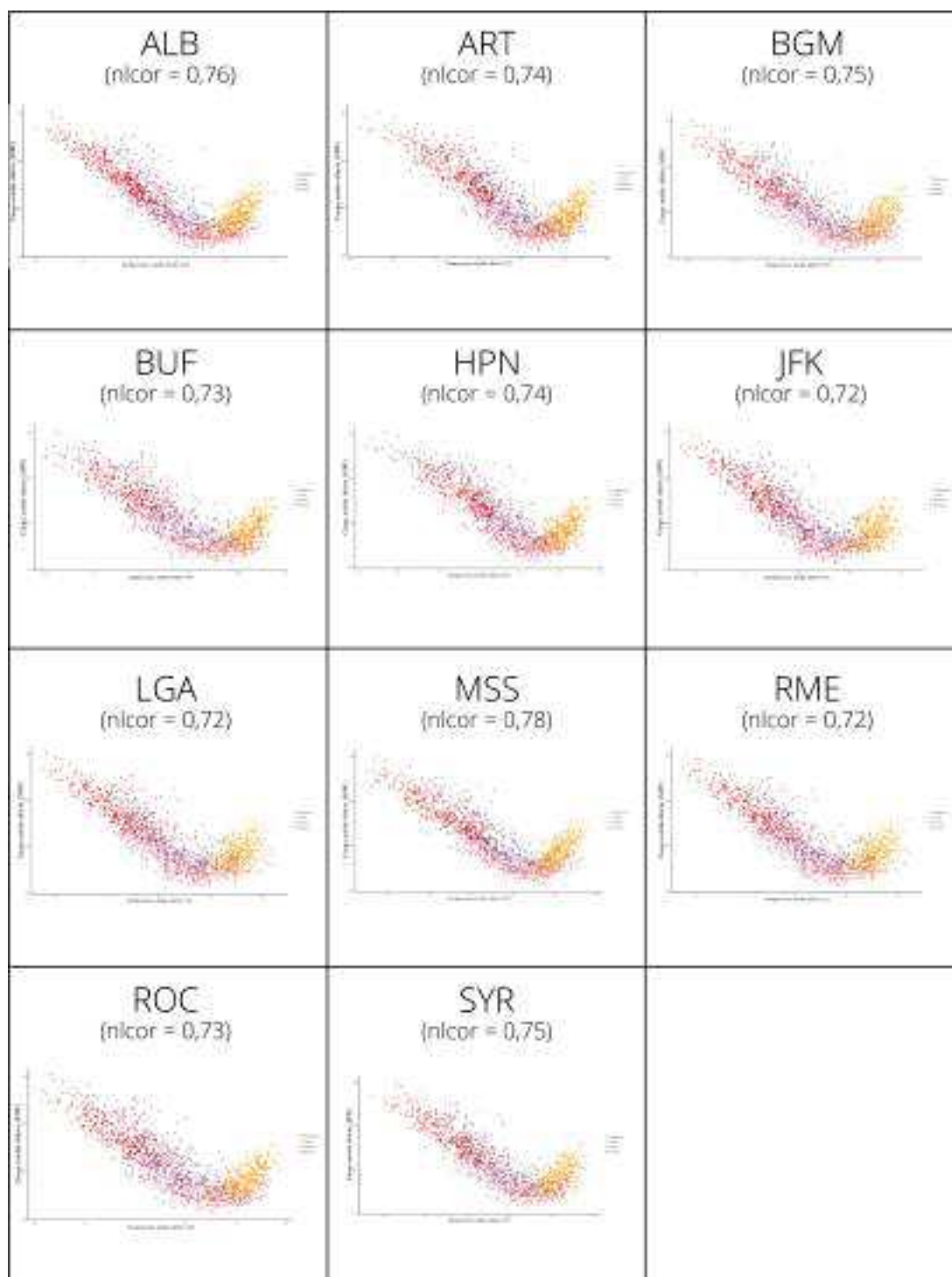


Figura B.6 – Cargas *versus* temperaturas (médias diárias) para a zona E-MHKVL e todas as estações meteorológicas.

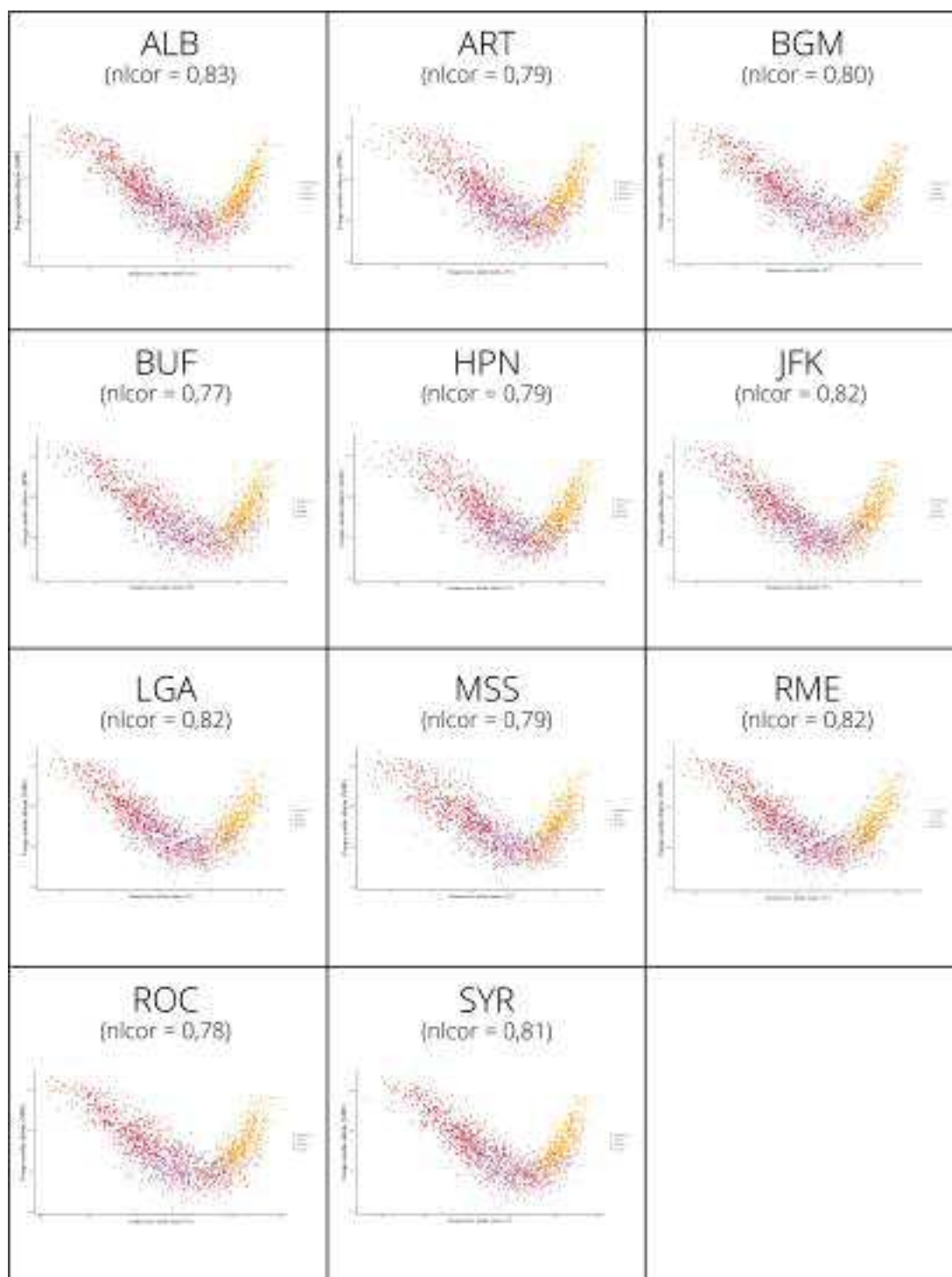


Figura B.7 – Cargas *versus* temperaturas (médias diárias) para a zona F-CAPITL e todas as estações meteorológicas.

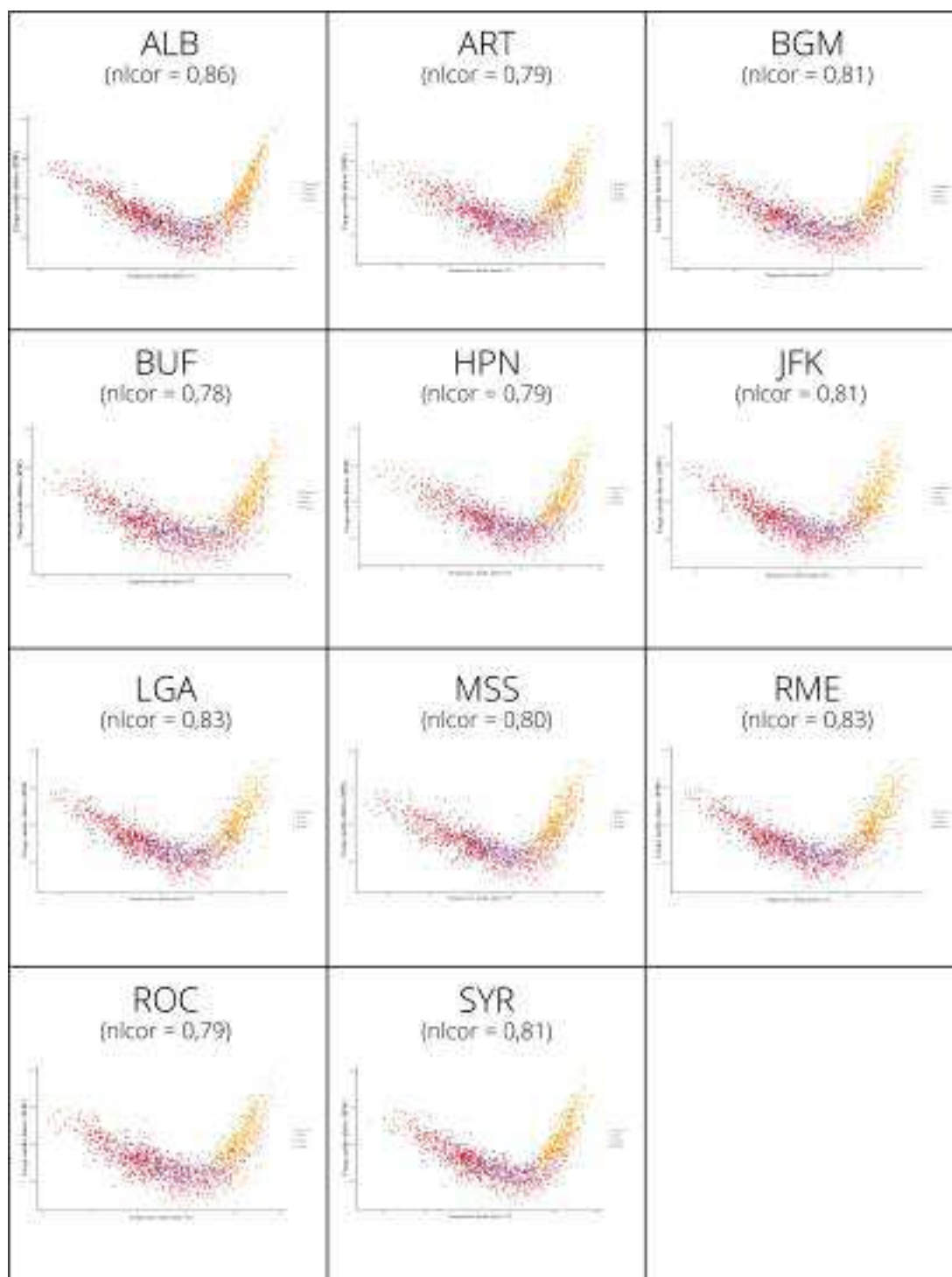


Figura B.8 – Cargas *versus* temperaturas (médias diárias) para a zona G-HUDVL e todas as estações meteorológicas.

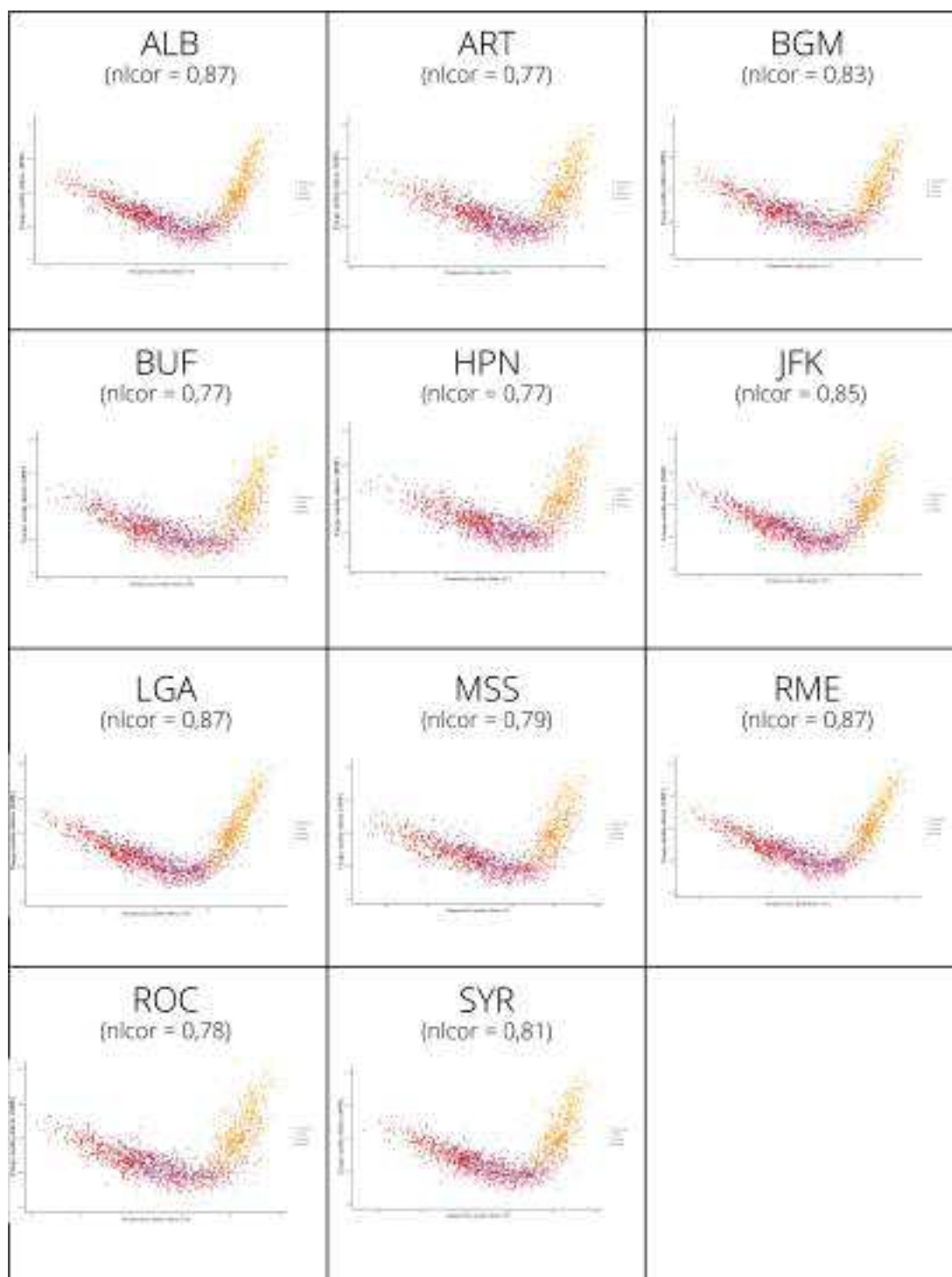


Tabela B .1 – Medidas descritivas, em °C, para temperaturas da estação meteorológica ALB

	Primavera	Verão	Outono	Inverno
Média	12,99	22,12	7,44	-1,80
Mediana	13,35	22,25	7,25	-1,15
Mínimo	-11,75	4,45	-17,25	-25,05
Máximo	35,05	36,15	32,85	26,75
Amplitude	46,80	31,0	50,10	51,80
Desvio padrão	8,27	4,79	7,72	7,43

Tabela B .2 – Medidas descritivas, em °C, para temperaturas da estação meteorológica ART

	Primavera	Verão	Outono	Inverno
Média	10,66	20,33	6,27	-4,10
Mediana	10,65	20,65	6,15	-3,35
Mínimo	-16,15	2,25	-26,75	-38,35
Máximo	35,85	32,85	28,95	20,05
Amplitude	49,00	30,60	55,70	58,40
Desvio padrão	8,50	5,16	8,06	8,54

Tabela B .3 – Medidas descritivas, em °C, para temperaturas da estação meteorológica BGM

	Primavera	Verão	Outono	Inverno
Média	10,94	19,61	5,86	-3,37
Mediana	11,15	19,45	5,65	-3,35
Mínimo	-11,75	3,95	-17,25	-27,85
Máximo	32,25	32,85	30,05	22,25
Amplitude	44,00	28,90	47,30	50,10
Desvio padrão	8,34	4,43	7,78	7,41

Tabela B .4 – Medidas descritivas, em °C, para temperaturas da estação meteorológica BUF

	Primavera	Verão	Outono	Inverno
Média	12,07	21,6	7,79	-2,22
Mediana	12,25	21,75	7,65	-1,75
Mínimo	-8,95	7,25	-12,85	-24,45
Máximo	32,25	33,35	32,25	21,16
Amplitude	41,20	26,10	45,10	45,61
Desvio padrão	8,35	4,42	7,63	7,23

Tabela B .5 – Medidas descritivas, em °C, para temperaturas da estação meteorológica HPN

	Primavera	Verão	Outono	Inverno
Média	10,66	20,33	6,27	-4,10
Mediana	10,65	20,65	6,15	-3,35
Mínimo	-16,15	2,25	-26,75	-38,35
Máximo	32,85	32,85	28,95	20,05
Amplitude	49,00	30,60	55,70	58,40
Desvio padrão	8,50	5,16	8,06	8,54

Tabela B .6 – Medidas descritivas, em °C, para temperaturas da estação meteorológica JFK

	Primavera	Verão	Outono	Inverno
Média	14,14	24,26	11,41	2,48
Mediana	13,95	24,00	11,75	2,85
Mínimo	-4,45	12,85	-9,45	-17,25
Máximo	33,95	36,15	32,85	20,65
Amplitude	38,40	23,30	42,30	37,90
Desvio padrão	6,58	3,47	6,71	6,03

Tabela B .7 – Medidas descritivas, em °C, para temperaturas da estação meteorológica LGA

	Primavera	Verão	Outono	Inverno
Média	15,34	25,26	12,07	2,88
Mediana	15,05	25,05	11,75	3,35
Mínimo	-4,45	13,35	-8,35	-17,25
Máximo	37,85	36,15	32,25	25,65
Amplitude	42,30	22,80	40,60	42,90
Desvio padrão	7,25	3,76	6,86	6,42

Tabela B .8 – Medidas descritivas, em °C, para temperaturas da estação meteorológica MSS

	Primavera	Verão	Outono	Inverno
Média	11,15	20,37	5,14	-6,03
Mediana	11,15	20,65	5,05	-5,65
Mínimo	-15,05	2,85	-27,25	-31,15
Máximo	32,25	34,45	31,75	19,45
Amplitude	47,30	31,60	59,00	50,60
Desvio padrão	8,85	5,34	8,38	8,36

Tabela B .9 – Medidas descritivas, em °C, para temperaturas da estação meteorológica RME

	Primavera	Verão	Outono	Inverno
Média	15,34	25,26	12,07	2,88
Mediana	15,05	25,05	11,75	3,35
Mínimo	-4,45	13,35	-8,35	-17,25
Máximo	37,85	36,15	32,25	25,65
Amplitude	42,30	22,80	40,60	42,90
Desvio padrão	7,25	3,76	6,86	6,42

Tabela B .10 – Medidas descritivas, em °C, para temperaturas da estação meteorológica ROC

	Primavera	Verão	Outono	Inverno
Média	12,51	21,73	7,91	-1,72
Mediana	12,85	21,75	7,25	-1,15
Mínimo	-10,05	6,75	-15,05	-23,95
Máximo	33,35	35,05	32,85	23,95
Amplitude	43,40	28,30	47,90	47,90
Desvio padrão	8,52	4,95	7,63	7,40

Tabela B .11 – Medidas descritivas, em °C, para temperaturas da estação meteorológica SYR

	Primavera	Verão	Outono	Inverno
Média	11,95	21,23	6,96	-2,94
Mediana	12,25	21,15	6,75	-2,25
Mínimo	-12,85	4,45	-18,95	-30,05
Máximo	33,35	35,05	32,25	25,05
Amplitude	46,20	30,60	51,20	55,10
Desvio padrão	8,63	4,90	7,83	7,66