

Guilherme Guilhermino Neto

**Métodos Univariados e Multivariados para Previsão da Demanda de Energia
Elétrica em Curto Prazo: Um Estudo Comparativo**

Dissertação apresentada ao Programa de Pós-graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do grau de mestre em Modelagem Computacional.

Orientador: Prof. Henrique Steinherz Hippert, D.Sc

Juiz de Fora

2014

Ficha catalográfica elaborada através do Programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Silva, Jonathan Esteban Arroyo.

Uma técnica explícita de marcha no tempo para ondas elásticas baseada em funções de Green calculadas localmente pelo MEF / Jonathan Esteban Arroyo Silva. -- 2014.

101 p. : il.

Orientador: Felipe dos Santos Loureiro

Coorientador: Luis Paulo da Silva Barra

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Engenharia. Programa de Pós-Graduação em Modelagem Computacional, 2014.

1. Funções de Green locais. 2. Princípio da causalidade. 3. Marcha no tempo. 4. MEF. 5. Ondas elásticas. I. Loureiro, Felipe dos Santos , orient. II. Barra, Luis Paulo da Silva , coorient. III. Título.

Guilherme Guilhermino Neto

**Métodos Univariados e Multivariados para Previsão da Demanda de Energia
Elétrica em Curto Prazo: Um Estudo Comparativo**

Dissertação apresentada ao Programa de Pós-graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do grau de mestre em Modelagem Computacional.

Aprovada em 20 de Agosto de 2014.

BANCA EXAMINADORA

Prof. Henrique Steinherz Hippert, D.Sc - Orientador
Universidade Federal de Juiz de Fora

Prof. Marcel de Toledo Vieira, Ph.D
Universidade Federal de Juiz de Fora

Prof^a Gláucia de Paula Falco, D.Sc
Instituto Vianna Júnior / Faculdade Machado Sobrinho

AGRADECIMENTOS

Aos meus pais, irmãos e avós, pelo fundamental apoio.

À Micheli, pela motivação e pela compreensão ao dividir algumas horas de mim com este texto.

Ao meu orientador, Henrique, pelo conhecimento transmitido.

Ao Samuel, por colaborar com os resultados e análises desta dissertação.

Ao Fabrizzio, por ter me introduzido às redes neurais.

Aos colegas Anna, Janaína e João, pelo companheirismo ao longo do curso.

À CAPES pelo auxílio financeiro.

*'Predictions are only parlor
games - a form of
entertainment.'*

Harry Browne

RESUMO

Previsões de demanda em curto prazo são fundamentais para o planejamento e o controle da produção em sistemas de energia elétrica. Como não é viável manter estoques de segurança para compensar demandas inesperadas, a programação da geração é baseada em previsões feitas com antecedência de algumas horas.

Ao longo dos anos, muitos métodos foram testados para a resolução do problema. Dentre os mais populares estão os univariados, em que a demanda é escrita como uma função linear de seu comportamento histórico e prevista por técnicas estatísticas. Também é frequente o uso de métodos multivariados, que levam em conta o efeito não-linear de variáveis climáticas, como a temperatura do ar, sobre o comportamento do consumidor. Para este caso, a literatura recente sugere o uso de previsores de inteligência computacional, como as redes neurais artificiais.

Embora alguns autores afirmem que deve-se considerar métodos multivariados, outros defendem que, para previsões de curto prazo (horizonte de poucas horas), a inclusão de variáveis climáticas traz poucos benefícios, posto que seus efeitos levam mais tempo para serem percebidos.

Neste trabalho, experimentamos diversos métodos univariados e multivariados a fim de comparar seu desempenho sobre uma base de dados da cidade do Rio de Janeiro. Para estes dados, mostramos que é possível obter, por meio de um simples previsor linear univariado (um modelo de curva de carga padrão cuja componente-base é prevista pelo amortecimento de Holt-Winters-Taylor), resultados próximos aos de técnicas mais complexas, porém, com as vantagens de maior robustez, parcimônia e economia de recursos computacionais.

Palavras-chave: Previsão de demanda. Energia elétrica. Amortecimento exponencial. Redes neurais.

ABSTRACT

Short-term demand forecasts are a vital part of the production plan and control on electrical power systems. As it is not possible to keep large inventories to meet sudden demand increases, the generation scheduling is based on forecasts made for some hours ahead.

Throughout the years, many methods have been proposed in order to solve the problem. Among the most popular are the univariate ones, on which the demand is written as a linear function of its historical behavior and forecast by statistical techniques. It is also common to use multivariate methods, which take into account also the nonlinear effects produced on the demand by weather-related variables, such as the air temperature. For this case, recent papers suggest the use of computational intelligence devices, such as artificial neural networks.

Although some authors claim that multivariate methods must be considered, some others state that, on a short-run (lead-times up to a few hours), adding weather-related variables brings little benefits, because its effects might take a longer time to affect the demand.

On this work, we experiment a large amount of univariate and multivariate methods aiming to compare its performance over a dataset from the city of Rio de Janeiro. For these data, we show that is possible to obtain, via a simple linear univariate method (a standard load curve model where the base load is forecast by the Holt-Winters-Taylor smoothing), results that are close enough to those achieved by more complex techniques, but bringing the advantages of more robustness, parsimony and computational economy.

Keywords: Load forecasting. Electrical energy. Exponential smoothing. Neural networks.

SUMÁRIO

1	INTRODUÇÃO.....	10
1.1	Motivação: a previsão de demanda em mercados competitivos	11
1.2	Objetivos	12
1.3	Organização do trabalho	13
2	MODELAGEM E PREVISÃO DA DEMANDA DE ENERGIA ELÉTRICA.....	14
2.1	Características das séries de demanda de energia elétrica	14
2.2	Modelos de curva de carga padrão	16
2.2.1	<i>Modelagem múltipla.....</i>	19
2.2.2	<i>Métodos de inteligência computacional.....</i>	20
2.2.3	<i>Combinação de previsores.....</i>	22
3	MATERIAL E MÉTODOS.....	24
3.1	Base de dados do Rio de Janeiro	24
3.2	Implementação dos métodos	27
3.3	Métodos univariados	27
3.3.1	<i>Tipo I: Previsores ingênuos.....</i>	27
3.3.2	<i>Tipo II: 168 previsores paralelos.....</i>	28
3.3.2.1	<i>Médias móveis simples.....</i>	29
3.3.2.2	<i>Amortecimento exponencial simples.....</i>	30
3.3.2.3	<i>Método linear de Holt.....</i>	31
3.3.2.4	<i>Modelos ARIMA.....</i>	31
3.3.3	<i>Tipo III: 24 previsores paralelos.....</i>	34
3.3.3.1	<i>Método de Holt-Winters.....</i>	35
3.3.3.2	<i>Modelos SARIMA.....</i>	35
3.3.4	<i>Tipo IV: Um método de sazonalidade múltipla.....</i>	36
3.4	Métodos multivariados de previsão	40
3.4.1	<i>Regressão linear múltipla.....</i>	40
3.4.2	<i>Redes neurais artificiais (RNAs).....</i>	41

3.5	Comparação com resultado externo	44
3.6	Métrica de erro	45
4	RESULTADOS E DISCUSSÃO	46
4.1	Métodos univariados	46
4.2	Métodos multivariados	49
4.2.1	<i>Comparação entre todos os métodos</i>	51
5	CONCLUSÕES	53
	APÊNDICES	54
A.1	Conceitos introdutórios	55
A.2	Terminologia e decomposição de séries temporais	56
	REFERÊNCIAS	58

1 INTRODUÇÃO

Através dos anos, a indústria de energia elétrica tem contribuído significativamente para o desenvolvimento tecnológico e o aumento da qualidade de vida do ser humano. Em consequência, a demanda por eletricidade experimenta notável crescimento ao redor do mundo. Particularmente no Brasil, apenas no ano de 2012, aproximadamente 1,8 milhão de novas unidades foram incorporadas à base de consumidores residenciais [1]. Além disso, estudos indicam que, entre 2011 e 2021, o consumo deve crescer em média 4,5% ao ano [2]. Frente à necessidade de se atender a estas demandas, empresas do setor elétrico são cada vez mais pressionadas a seguir padrões de qualidade, sob pena de multas da parte de órgãos regulatórios. Isto leva à necessidade de se adotar boas práticas de gestão a fim de melhorar a eficiência e reduzir custos.

A eletricidade recebida por consumidores industriais, comerciais e residenciais provém de um sistema de energia elétrica (SEE; retratado na Fig. 1.1). A produção em um SEE tem início na geração, etapa em que algum tipo de energia (hidráulica, calorífica, nuclear etc.) é produzida em uma usina e transformada em energia elétrica. Esta energia é transmitida em linhas de alta tensão aos centros de consumo, passando por subestações que reduzem sua tensão e, assim, a tornam própria para distribuição aos consumidores (para uma introdução mais abrangente, consultar [3]).

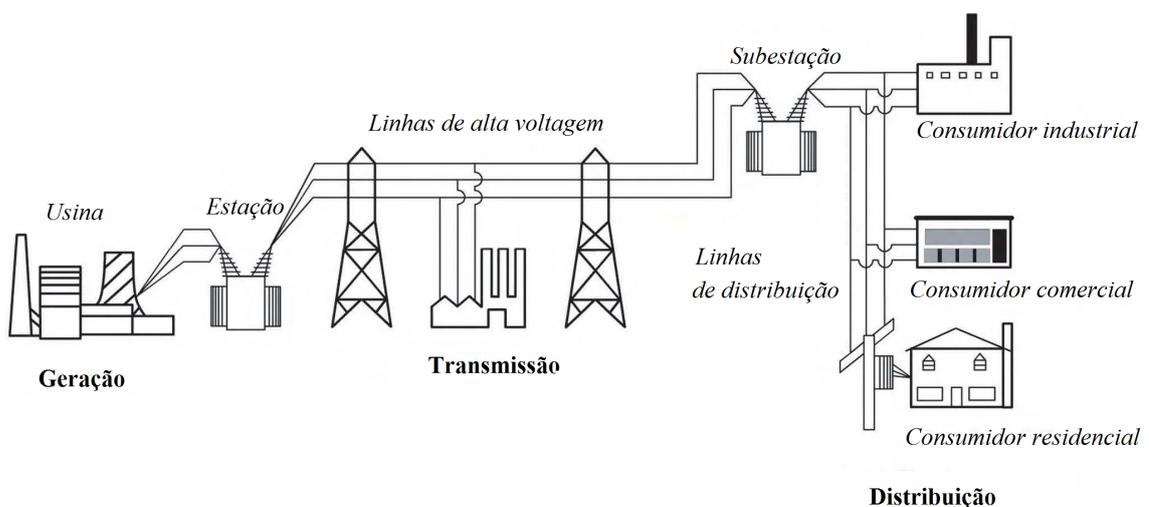


Figura 1.1: Esquema geral de um SEE ([3]; adaptada)

Os SEE são sistemas que devem entregar seu produto em tempo real; a energia

deve estar disponível no momento em que um consumidor acende uma lâmpada, por exemplo. Acontece que, diferentemente de recursos como água e gás, a eletricidade não pode ser armazenada em grandes quantidades para ser distribuída posteriormente. É necessário, portanto, calcular a quantidade de energia que deve estar disponível no SEE com antecedência.

A quantidade de energia total nos SEE é variável ao longo de um dia e também de um dia para outro e, por isso, a programação de quais geradores devem ser ligados e desligados (e em que sequência isto deve ser feito) deve ser montada diariamente. Este problema é chamado *alocação de unidades* (do inglês *unit commitment*) e sua solução ótima é pautada pela economia nas operações. [4]. Para fazer os cronogramas de alocação de unidades, são utilizadas previsões da demanda de energia elétrica dos clientes, espaçadas de hora em hora, para horizontes de uma hora até uma semana à frente [5]. Estas previsões são chamadas *previsões de curto prazo*, e são o foco desta dissertação de mestrado.

1.1 Motivação: a previsão de demanda em mercados competitivos

Estoques fazem parte da maioria dos sistemas produtivos. Armazenar material e produtos é uma estratégia logística comum para contornar a dificuldade em se prever a demanda com total certeza. Na produção de energia elétrica, no entanto, ocorre algo especial: como não é viável manter grandes estoques para atender demandas inesperadas, o excesso de oferta imediata converte-se em desperdício de energia. Ao mesmo tempo, entregar quantidades abaixo do necessário pode causar paradas nas atividades, acidentes e outros prejuízos ao consumidor.

Como explica Bunn [6], no início da década de 90, a previsão da demanda de energia elétrica em curto prazo havia atingido um estado confortável; e os métodos disponíveis permitiam operar os sistemas com erro relativamente baixo. Em países desenvolvidos, incertezas na demanda eram compensadas pela grande capacidade produtiva; enquanto que, em países em desenvolvimento, a solução era funcionar utilizando sempre o máximo dos recursos disponíveis.

Segue [6], com a privatização de concessionárias de energia e a desregulamentação do setor energético, surgiram mercados competitivos, onde operações de compra e venda

de eletricidade tornaram-se comuns. Neste cenário, pode-se, em situação de urgência, compensar a baixa produção comprando energia a qualquer hora, mas, como os mercados são tipo *spot* (imediatos), os preços são muito mais altos quando o pedido é feito em cima da hora; exigindo que as empresas tenham boa capacidade de antecipação. Outrossim, como a participação de consumidores aumentou significativamente, políticas de incentivo à redução de consumo e cumprimento com padrões de qualidade e desempenho ambiental tornaram-se essenciais à gestão da demanda, principalmente considerando que se espera o crescimento da autoprodução (geração feita pelo próprio consumidor).

Estes motivos, direta ou indiretamente, evidenciam a importância das previsões de curto prazo no contexto dos mercados competitivos, o que justifica a pesquisa por métodos cujos erros cometidos sejam mínimos.

1.2 Objetivos

A literatura em previsão da demanda de energia elétrica é ampla, com registros de experimentos com inúmeros métodos baseados em técnicas estatísticas e computacionais (e mistos destas). Estes métodos podem ser *univariados*, baseados no comportamento histórico da demanda, ou *multivariados*, considerando também as mudanças causadas por variáveis exógenas, principalmente as climáticas¹

Embora sejam muitos os casos de sucesso mediante o uso de ambos os métodos, parece não haver um consenso sobre qual seria mais adequado à previsão da demanda de energia em curto prazo. Gupta e Yamada [8] dizem que as abordagens univariadas são limitadas porque o clima é grande responsável por mudanças no consumo médio; e Rahman e Hazim [9] afirmam que a relação entre demanda e temperatura é o fator mais importante para que uma previsão seja confiável. Por outro lado, para Farmer e Potton [10], considerar a dependência entre a demanda e fatores climáticos dobra as dificuldades da previsão: segundo os autores, o único parâmetro que afeta o comportamento do consumidor, em

¹Os termos *univariável* e *univariado* e *multivariável* e *multivariado* normalmente causam alguma confusão. Em geral, o sufixo *variável* diz respeito à quantidade de variáveis de entrada, enquanto *variado* se refere às variáveis de saída de um modelo [7]. Por exemplo, um modelo multivariável possui múltiplas entradas, enquanto um multivariado possui múltiplas saídas. Como, neste trabalho, lidamos com métodos de todos os tipos possíveis (e, ainda, mistos destes, como os conjuntos de previsores paralelos, em que várias equações multivariáveis compõem um modelo de saída multivariada), adotamos, com alguma liberdade, uma terminologia simplificada, chamando *univariados* os métodos cuja entrada são apenas os valores passados de demanda de energia, e de *multivariados* aqueles que também levam em conta variáveis climáticas para explicar o comportamento da série de demandas.

uma escala horária, é a nebulosidade - quanto mais nuvens no céu, mais escuro o dia e mais a luz elétrica é utilizada. No entanto, é muito caro obter medições confiáveis sobre esta variável. Em outro trabalho, Taylor [11] também defende que, no curto prazo, variáveis como a temperatura tendem a mudar pouco, e, logo, influenciar pouco o consumo e, por isso, toda esta tendência deve ser capturada pelos próprios dados de demanda.

Neste trabalho, nos propomos a examinar o desempenho de métodos univariados e multivariados para previsões com horizonte de 1 até 24 horas à frente (ou previsões de *perfil*). Conduzimos experimentos sobre uma base de dados real, comparando inicialmente diversos previsores univariados de séries temporais. O previsor de melhor desempenho é utilizado como base para a construção de um modelo multivariado em que a mudança na temperatura explica os desvios do consumo esperado. É feita ainda uma comparação com os resultados obtidos por uma rede neural artificial multivariada do artigo [12].

1.3 Organização do trabalho

Esta dissertação está organizada em cinco capítulos.

Neste capítulo, fizemos uma introdução aos aspectos gerais dos SEE e destacamos a importância da previsão de demanda em curto prazo para a indústria de energia, e também explicamos os objetivos desta dissertação e a organização do trabalho.

No Capítulo 2, apresentamos as características das séries temporais de demanda de energia elétrica, e fazemos uma revisão sobre as estratégias mais comuns de modelagem e previsão para o problema.

No Capítulo 3, descrevemos os dados usados nos experimentos e os métodos univariados e multivariados analisados.

No Capítulo 4, mostramos os resultados obtidos por meio de gráficos e tabelas, os discutimos e comparamos.

As conclusões e comentários, baseados nos resultados obtidos, são apresentados no capítulo seguinte, juntamente com sugestões para trabalhos futuros.

Um breve Apêndice ao final do texto introduz conceitos-chave em análise e previsão de séries temporais e a terminologia empregada ao longo dos capítulos.

2 MODELAGEM E PREVISÃO DA DEMANDA DE ENERGIA ELÉTRICA

O processo decisório na indústria de energia exige previsões de demanda para diferentes horizontes. Decisões de longo prazo, como a instalação de novas usinas, baseiam-se em previsões feitas para até 20 anos à frente. Já decisões de médio prazo, como a compra de combustíveis e máquinas, necessitam de previsões cujo horizonte varia desde algumas semanas até um ano. [13].

Em mercados competitivos, é preciso determinar, diariamente, preços para compra e venda de energia. Por conta disso, a maior urgência é por previsões com horizonte de algumas horas à frente (curto prazo), especialmente a de *perfis diários*. Prever um perfil diário é reproduzir estimativas para as 24 horas de um dia [6], no dia anterior. Neste trabalho, consideramos dados espaçados de hora em hora e utilizamos os dados disponíveis até a última hora de um dia para fazer a previsão do perfil do dia seguinte.

Para escolher um método de previsão que extraia o máximo da informação contida nos dados, Chatfield diz que é imprescindível conhecer bem o fenômeno em análise [14]. Por isso, iniciamos o capítulo apresentando as características das séries temporais de demanda de energia elétrica. Depois, revisamos os principais métodos e estratégias para previsão de perfis difundidos na literatura.

2.1 Características das séries de demanda de energia elétrica

A principal característica das séries de demanda de energia elétrica é a existência de vários ciclos sazonais sobrepostos.

Na Fig. 2.1, temos um exemplo de uma série com forte sazonalidade anual. Para estes dados, medidos em um SEE americano [15], repete-se um padrão com demandas bem

mais baixas nos meses mais quentes.

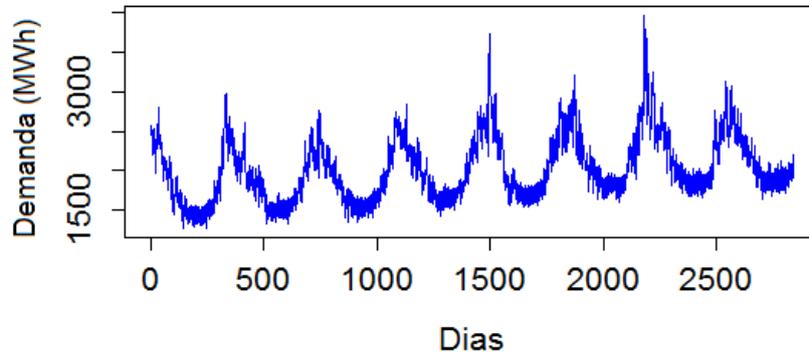


Figura 2.1: Média da demanda diária de energia nos anos de 1985 a 1992 (dados: Electric Power Research Institute)

Na mesma figura, há uma clara tendência de crescimento do consumo ao longo dos anos.

Na Fig. 2.2, observamos uma quinzena de verão da série americana. Notamos um padrão que se repete todos os dias, com demandas mais baixas nas madrugadas. Além disso, o consumo é similar entre os dias de semana, mas cai aos sábados e domingos. Temos, portanto, uma sazonalidade de extensão diária e outra semanal.

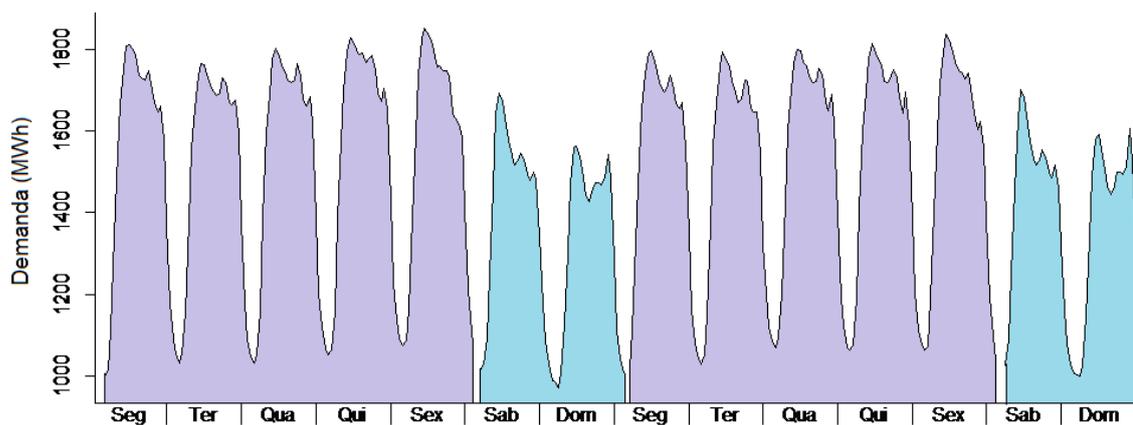


Figura 2.2: Demandas em uma quinzena de verão (dados: Electric Power Research Institute)

Os métodos discutidos nas seções seguintes são alternativas para tratar a complexa estrutura multi-sazonal das séries de carga e suas oscilações, modelando-a, cada um, à sua maneira.

2.2 Modelos de curva de carga padrão

Uma das formas mais utilizadas na indústria para prever a demanda de energia é a *curva de carga padrão* [16]. A ideia é tentar estimar a demanda média por meio de um modelo estatístico definido *a priori*.

A formulação aditiva geral é dada pela Eq. 2.1, onde S_t é chamada *componente-padrão* e R_t é a parcela de erro aleatório (ruído).

$$L_t = S_t + R_t \quad (2.1)$$

A componente-padrão S_t pode ser tratada como univariada e prevista apenas com base nos valores passados de carga. Neste caso, conforme mostra a coleção de artigos publicada por Bunn e Farmer [13], muitos autores modelam esta componente por funções trigonométricas, modelos de espaço de estado (filtros de Kalman) e, principalmente modelos ARIMA.

Outra forma de descrever S_t univariada é fazer sua decomposição em pelo menos três componentes [16]:

$$S_t = l_t + b_t + d_t \quad (2.2)$$

em que l_t é o nível médio da série, b_t é a tendência, e d_t é um fator sazonal semanal (*day-of-the-week effect*). Estas componentes podem ser estimadas utilizando amortecimentos exponenciais. Embora estes métodos tenham sido, historicamente, menos lembrados que outros –com menção para os trabalhos de Christiaanse [17] e Park [18]– a pesquisa recente de Taylor tem apresentado resultados interessantes. O autor introduz uma abordagem inovadora, ao adaptar amortecimentos tradicionais para acomodar mais de um ciclo sazonal em um só conjunto de equações [19]. Dedicamos atenção especial ao amortecimento exponencial de Holt-Winters-Taylor nos próximos capítulos.

Além dos modelos univariados, podemos supor que a forma funcional da componente-padrão seja descrita por flutuações em torno de uma *componente-base* B_t , e que estes desvios sejam explicados por uma componente W_t , afetada por variáveis climáticas. Então, vale a relação causal $S_t - B_t = W_t$, que resulta na extensão multivariada

$$S_t = B_t + W_t \quad (2.3)$$

Muitos fatores climáticos podem compor W_t e a escolha de quais devem ser incluídos no modelo depende da disponibilidade dos dados e claro, de quais são mais relevantes. Por exemplo, Davies [20], na primeira publicação sobre a correlação entre clima e demanda de energia, explica que, no Reino Unido, os efeitos mais importantes são os da temperatura ambiente e da nebulosidade, porque influenciam o uso de aparelhos de ar condicionado, aquecedores e da luz elétrica.

A maioria dos modelos multivariados trata a temperatura do ar como a principal variável exógena. Como a mudança de temperatura costuma ocorrer lentamente, a resposta observada no comportamento do consumidor não deve ser imediata [20], por isso normalmente se inclui no modelo não a temperatura exata no momento a ser previsto, mas valores defasados ou diferenças.

Uma das maiores dificuldades no uso da temperatura é o fato de sua correlação com a demanda de energia ser extremamente não-linear (Figs. 2.3, 2.4, 2.5), o que torna muito complicado definir um modelo para associar estas variáveis.

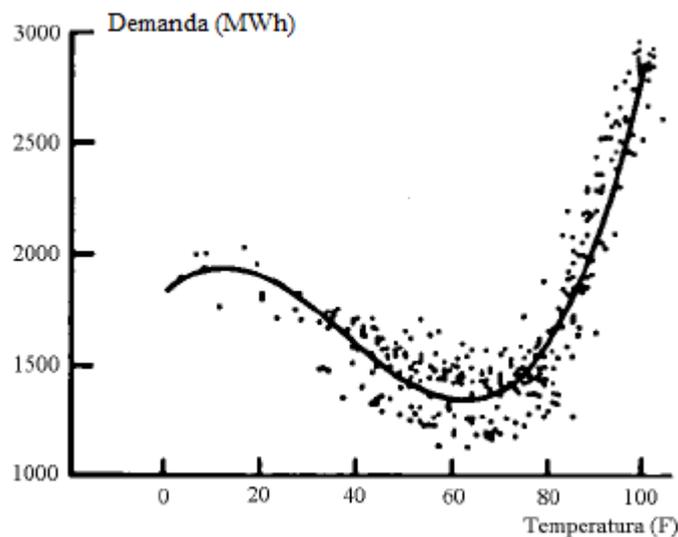


Figura 2.3: Demandas vs. temperaturas (dados: sudoeste dos EUA, 1983 [21])

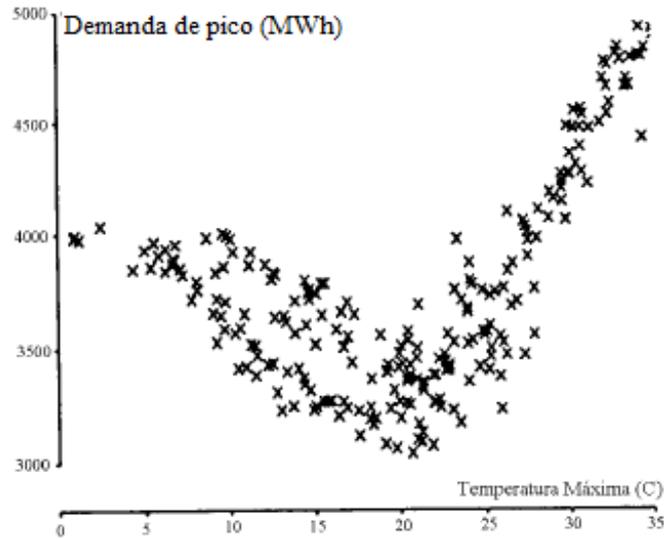


Figura 2.4: Demandas de pico vs. temperaturas (dados: Tóquio, 1990 [22])

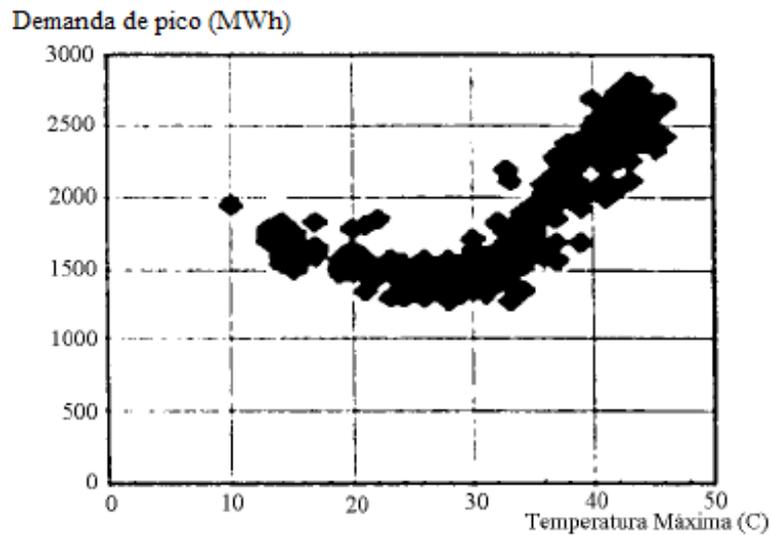


Figura 2.5: Demandas de pico vs. temperaturas (dados: Riyadh, 1990 [23])

Até o fim da década de 80 (pouco antes da popularização dos métodos de inteligência computacional), algumas soluções haviam sido propostas para modelar a não-linearidade. Por exemplo, Gupta e Yamada [8] propõem um modelo linear por partes, composto por duas equações de primeiro grau e uma reta constante:

$$W_t = \begin{cases} a(T_t - 70) & , \text{ se } T_t > 70^\circ F \\ 0 & , \text{ se } 60 < T_t < 70^\circ F \\ b(60 - T_t) & , \text{ se } T_t < 60^\circ F \end{cases} \quad (2.4)$$

T_t é a temperatura (em °F) na hora t e 60 e 70, estimativas do método de Stanton e Gupta [24], são temperaturas limítrofes de uma zona neutra, dentro da qual o consumidor

não sente a necessidade de usar aparelhos de aquecimento ou ar condicionado. Quanto mais fora da zona neutra, maior o efeito sobre da temperatura sobre demanda de energia (e, quanto mais próximo, mais este tende a zero).

Abou-Hussien *et al.* [25] aplicam um modelo similar, porém, mantendo alguma não-linearidade:

$$W_t = \begin{cases} a_0 + a_1(T_{t-1} - 70) + a_2(T_{t-1} - 70)^2 & \text{se } T_t > 70^\circ F \\ 0 & \text{se } 60 < T_t < 70^\circ F \\ b_0 + b_1(60 - T_{t-1}) + b_2(60 - T_{t-1})^2 & \text{se } T_t < 60^\circ F \end{cases} \quad (2.5)$$

Neste modelo, aparecem termos quadráticos e as temperaturas atuais são substituídas pelas defasadas em uma hora. Também adicionam-se vieses (a_0 e a_1). A Fig. 2.6 mostra uma comparação gráfica entre os modelos 2.4 e 2.5.

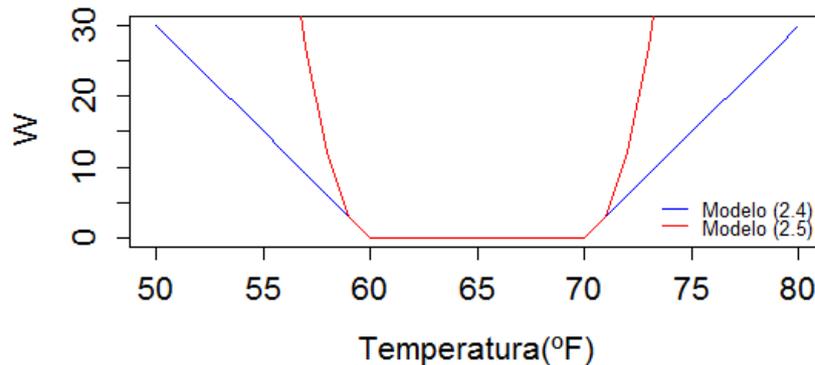


Figura 2.6: Comparação entre modelos para variável meteorológica baseados na temperatura

Além destes, modelos de alta complexidade foram testados por Engel *et al.*, como regressões semi-paramétricas baseadas na interpolação de *splines* cúbicas [26].

2.2.1 Modelagem múltipla

Modelos de curva de carga são métodos de saída univariada e possibilitam explorar bem a correlação entre as demandas de horas e dias distintos porque fazem previsões ao longo de toda a série de dados. Na prática, porém, demandas de horas distintas podem sofrer diferentes influências, o que leva muitos autores a adotarem um modelo para cada hora

do dia e aplicar previsores paralelamente a cada um.

Papalexopoulos e Hesterberg, quem provavelmente introduziram esta abordagem, constroem 168 modelos de regressão, um para cada hora de cada dia da semana [27]. Para evitar a superparametrização imposta pelo uso de tantos modelos, outros autores preferem fazer a divisão em 24 modelos, um para cada hora do dia: Park *et al.* [18] aplicam composições de filtros de Kalman e amortecimento exponencial a 24 séries; Cancelo *et al.* [28], 24 modelos ARIMA e Fan e Chen [29], 24 máquinas de vetor-suporte, por exemplo.

A consolidação deste tipo de estratégia deu-se com o trabalho de Ramanathan *et al.* [30], que adotam 24 modelos de regressão para as horas dos dias de semana e mais 24 para os fins de semana. Os resultados foram os melhores em uma competição promovida pelo *Electrical Power Research Institute* [15], superando diversos outros métodos.

Como a modelagem múltipla possibilita que o pesquisador explore melhor as características da demanda em uma determinada hora, começaram a surgir na literatura segmentações cada vez mais complexas para modelos de curva de carga padrão, para incluir o máximo de informação. Os 48 modelos feitos por Ramanathan *et al.*, por exemplo, são, cada um, uma decomposição da demanda em mais de 10 fatores, que vão desde variáveis indicadoras para feriados até médias de temperatura. Neste caso, o alto custo computacional para estimar os parâmetros foi bastante criticado [30].

2.2.2 Métodos de inteligência computacional

Ao invés de assumir um modelo *a priori*, podemos lançar mão de métodos capazes de reconhecer dependências entre os dados históricos para combiná-los e gerar previsões. Este é o caso das técnicas de *inteligência computacional* (IC).

A IC é uma área da inteligência de máquina que provê técnicas inspiradas no comportamento dos seres humanos e da natureza para descobrir (*aprender*) padrões de comportamento em um conjunto de dados e, posteriormente, aplicar o conhecimento adquirido em novas amostras [31]. A aprendizagem (ou o *treinamento*) se dá de forma adaptativa, com o sistema se corrigindo a partir dos próprios erros. A flexibilidade permite inúmeras aplicações práticas, que vão desde aproximação de funções e classificação (com destaque para o reconhecimento de caracteres e imagens) até controle de sistemas e otimização [32].

Com a evolução da capacidade de processamento dos computadores pessoais, a IC

tornou-se um grande atrativo para a previsão da demanda de energia, em especial devido à possibilidade de mapear as relações não-lineares entre demanda e temperatura sem a necessidade de um modelo previamente estabelecido. Foram testadas máquinas de vetor-suporte [33], [34]; sistemas especialistas [35], [9]; regressões *fuzzy* [36], entre outros, mas, decerto, nenhuma técnica tornou-se tão popular quanto as redes neurais artificiais (RNAs), como mostra uma revisão bibliográfica [37].

Uma RNA é um mecanismo que produz uma saída univariada ou multivariada a partir de complexas combinações não-lineares de uma ou mais variáveis de entrada. A transferência entre entrada e saída é feita pela interseção entre uma ou mais funções, normalmente sigmóides, dependendo do número de unidades de processamento (ou *neurônios*). Os pesos das combinações são encontrados por algum algoritmo de busca iterativa, que minimiza uma função de perda pré-definida.

Além da não-necessidade de se estipular um modelo previamente e da natureza não-linear, a flexibilidade na escolha das variáveis de entrada e saída, unidades de processamento, função de transferência e função de perda tornaram as RNAs métodos frequentemente utilizados para previsão.

Há muitas possibilidades de se montar as redes. Fidalgo e Lopes [38], por exemplo, trabalham com 74 variáveis de entrada (Fig. 2.7). Estas são funções trigonométricas (onde d é o dia da semana) que indicam as sazonalidades semanais e diárias, e também dados tipo $P(d, t)$, que são a demanda na hora t de um dia d . Existe uma camada intermediária com 6 neurônios a saída, multivariada, são as 24 horas de um perfil.

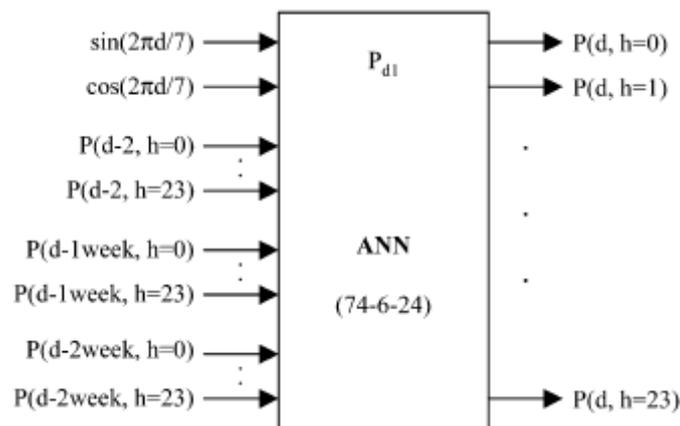


Figura 2.7: RNA para prever o perfil de um dia (Fonte: Fidalgo e Lopes [38])

Um dos maiores desafios ao se construir uma RNA é a escolha correta do número de neurônios na camada intermediária, o que varia de acordo com a complexidade do

problema. Não existe um conjunto fechado de regras para determinar este número, então é desejável escolher aquele que seja suficiente para atender à complexidade do problema, mas não tão grande a ponto de causar o sobreajuste dos dados e consequente perda da capacidade de generalização. [39]

Embora seja difícil superar as RNAs em termos dos erros de previsão, elas apresentam a desvantagem de consumirem muitos recursos computacionais e, principalmente, apresentarem baixa robustez em relação a outros métodos. Como os algoritmos de busca são bastante sensíveis às condições iniciais, é comum existir alguma variância entre os resultados obtidos por uma mesma rede executada várias vezes (não-raro com a obtenção de valores discrepantes). Além disso, as combinações não-lineares podem ser extremamente complexas, dificultando o julgamento dos resultados e sua discussão.

2.2.3 *Combinação de previsores*

O procedimento mais comum no cálculo de previsões é testar diversos previsores separadamente e compará-los, escolhendo o melhor entre eles. Newbold e Granger [40], no entanto, mostraram ser possível melhorar as previsões de métodos distintos realizando algum tipo de combinação entre seus resultados. Isto é indicado para quando:

- (i) Previsores levam em conta variáveis diferentes;
- (ii) Previsores abordam as relações entre variáveis de maneira diferente.

No artigo seminal, os autores combinam linearmente os resultados $F_{1,t}$ e $F_{2,t}$ de dois previsores com pesos k e $1 - k$, como na Eq. 2.6, e provam que a variância das previsões geradas pela combinação não pode ser maior que menor das variâncias separadas, ou seja, o resultado final não pode ser pior em termos de variância.

$$C_t = kF_{1,t} + (1 - k)F_{2,t} \quad (2.6)$$

A revisão bibliográfica feita por Clemen [41] mostra que o número de trabalhos com combinação de previsores aumentou substancialmente após a introdução do método. Apesar disso, a prática ainda parece ser pouco difundida na linha de previsão de demanda de energia elétrica. Destacamos o trabalho de Khotanzad *et al.* [42], em que são combinados resultados de RNAs: o programa desenvolvido pelos autores foi adotado por dezenas de SEEs americanos.

Vale mencionar ainda o trabalho de Taylor e Majhitia [43], sobre combinações de previsores univariados de *splines* cúbicas.

3 MATERIAL E MÉTODOS

3.1 Base de dados do Rio de Janeiro

A base de dados utilizada neste trabalho consiste de uma série de demandas horárias da cidade do Rio de Janeiro, para os anos de 1996 e 1997 (104 semanas, no total). As observações, fornecidas por uma concessionária local, foram medidas em MWh. Para os métodos multivariados, também utilizamos uma série de temperaturas do ar da mesma cidade, em °C, disponíveis a partir de 01/04/1996 até o final do ano de 1997.

Observações de demanda feitas em *dias especiais*, como feriados, são muito diferentes das demais, e, por isso, podem deturpar os resultados das previsões. Na prática, a demanda destes dias é descrita por modelos à parte [38], [44]. Como a modelagem dos dias especiais foge do escopo deste trabalho, optamos por substituir as observações de feriados pela média das observações da hora correspondente nas semanas anterior e seguinte.

Na Fig. 3.1, os gráficos de demanda e temperatura mostram uma forte sazonalidade anual: as demandas crescem nos meses mais quentes e diminuem nos mais frios. Este fenômeno ocorre devido ao aumento do uso de aparelhos de ar condicionado quando as temperaturas são mais altas. Existe também uma tendência de crescimento, embora muito suave.

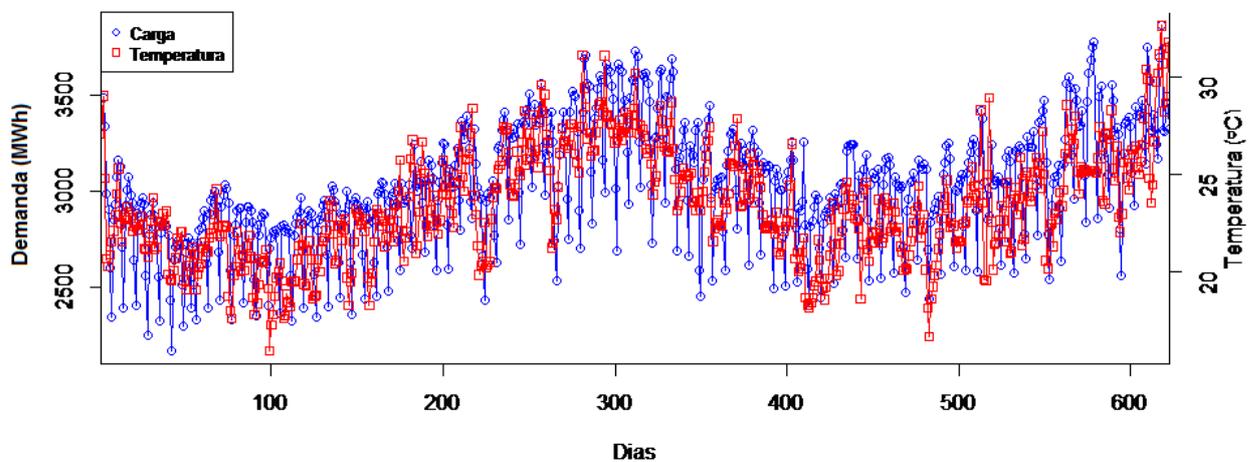


Figura 3.1: Médias diárias de demanda e energia para os anos de 1996 e 1997 (Dados: Rio de Janeiro)

Não incluímos a sazonalidade anual nos métodos de previsão, pois precisaríamos de mais dados do que o disponível para testar os métodos; seria necessário pelo menos um ano somente para a inicializá-los.

Na Fig. 3.2, observamos as demandas para uma semana de verão e outra de inverno. Há um padrão semanal, onde os dias de semana apresentam demandas similares, enquanto no sábado e no domingo os níveis são mais baixos. A distinção entre verão e inverno implica a necessidade de atualizar os parâmetros ao longo do ano, para capturar as mudanças de estação.

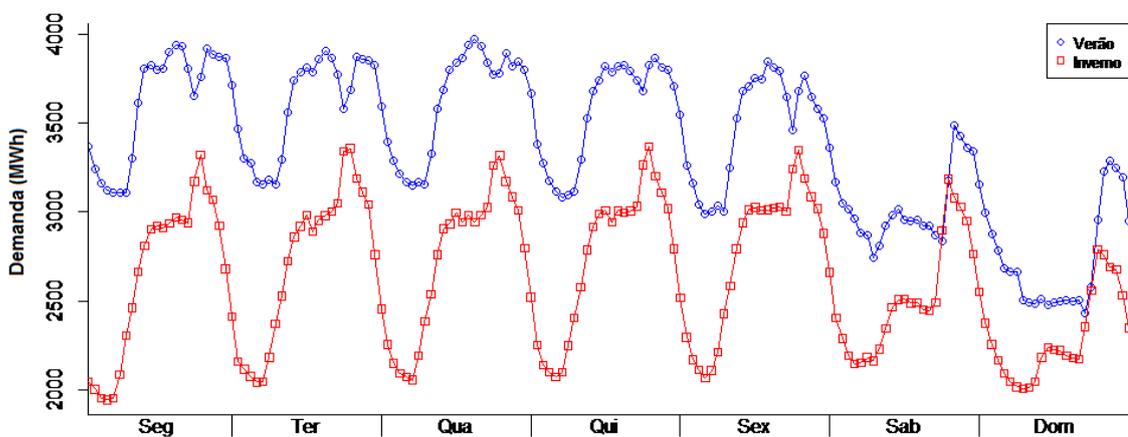


Figura 3.2: Demandas em uma semana de verão e outra de inverno (Dados: Rio de Janeiro)

Verificamos, ainda, padrões repetitivos nos perfis diários. Observando um perfil mais de perto, na Fig. 3.3, fica claro que há dois picos no verão, um às 15h e outro às 19h, enquanto no inverno há apenas um, às 18h.

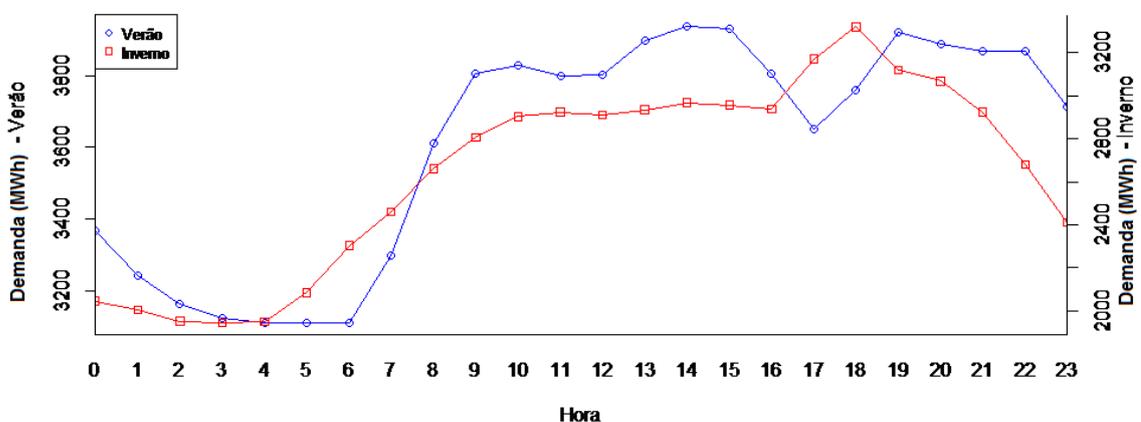


Figura 3.3: Demandas em um dia de verão e outro de inverno (Cidade do Rio de Janeiro)

A associação entre demanda e temperatura, exibida na Fig. 3.4, poderia ser modelada por uma função não-linear, com demandas mais altas relacionadas a temperaturas mais altas.

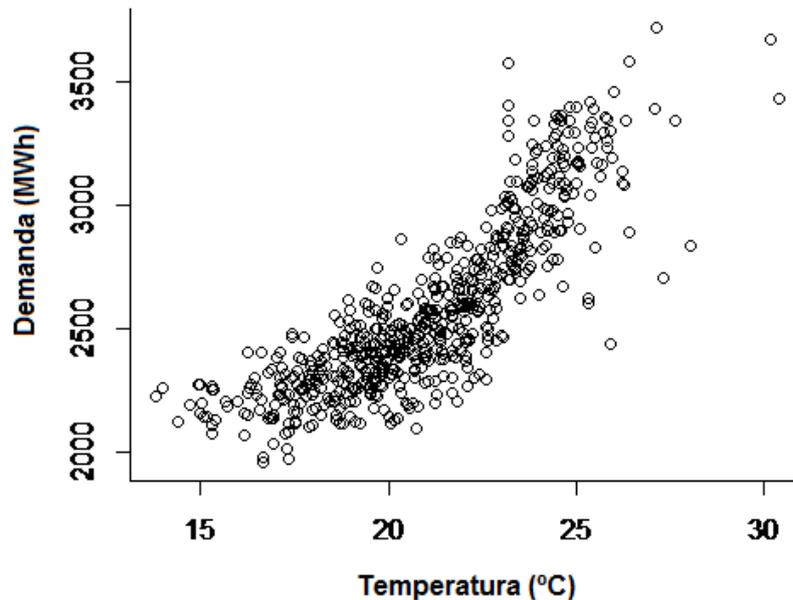


Figura 3.4: Demandas e temperaturas às 3h a.m. (Cidade do Rio de Janeiro)

Levando em conta as características sazonais da série de demandas do Rio de Janeiro e sua correlação com a temperatura, seguimos o roteiro abaixo para realizar nossos experimentos:

- (i) Experimentamos os métodos univariados listados na seção 3.3, que se baseiam na extrapolação de valores futuros a partir das demandas passadas. As primeiras 50 semanas da série de demandas são usadas para inicializar os métodos e ajustar as constantes e parâmetros, e as semanas 51 a 70, para testar a acurácia para previsão de perfis. Os perfis são previstos com os dados disponíveis até a última hora do dia anterior; isto é, a partir dos dados coletados até as 24h do dia d , prevemos as demandas das 24 horas do dia $d + 1$.
- (ii) Testamos um modelo multivariado de curva de carga padrão. O melhor método univariado entre os experimentados no item (i) é aplicado para prever a componente-base e, então, os desvios desta componente são modelados a partir da componente climática. Para isso, são aplicadas ou uma regressão linear múltipla ou uma rede

neural artificial (RNA). Usamos as semanas 1 a 84 para estimar os coeficientes do modelo de regressão e os pesos das redes e 85 a 104 para calcular os erros.

- (iii) Comparamos os resultados do melhor método univariado e dos métodos multivariados entre si e com os de uma RNA encontrada em Hippert *et al.* [12], para os mesmos dados.

3.2 Implementação dos métodos

Os métodos de previsão que testamos foram implementados em um computador pessoal, em linguagem e ambiente R, versão 3.1.1 [45]. O R é um projeto da Bell Laboratories (Lucent Technologies) com funções estatísticas, matemáticas, computacionais e gráficas. O aplicativo está disponível em formato de código aberto, para que os usuários possam criar, modificar e compartilhar rotinas.

Executamos tanto funções programadas quanto criadas por outros autores. Quando é este o caso, mencionamos quais foram estas funções e os pacotes que as contém.

3.3 Métodos univariados

Partimos de um modelo aditivo de curva de carga univariado (Eq. 3.1), onde S_t é uma componente-padrão, função das demandas passadas, e R_t é um resíduo. Experimentamos calcular S_t por meio de alguns métodos univariados.

$$L_t = S_t + R_t \quad (3.1)$$

Agrupamos os métodos em quatro tipos. Para mais detalhes sobre os previsores dos tipos I a III, ver Chatfield [14], Makridakis *et al.* [46], Morettin e Tolói [47] e Box *et al.* [48]. Para o previsor IV, as referências são as publicações de Taylor [11], [19].

3.3.1 Tipo I: Previsores ingênuos

Um previsor ingênuo (*naive forecaster*) é um método que considera a previsão para o próximo instante como sendo o último valor observado disponível. Embora estas previsões não forneçam informações muito precisas sobre o padrão de comportamento da série,

os resultados podem ser utilizados para a validação de outros métodos: caso um certo previsor mostre desempenho similar ou inferior ao de um *naive*, deve-se reestimar seus coeficientes ou parâmetros ou mesmo considerar descartá-lo.

Implementamos dois previsores ingênuos, os quais chamamos de *naive-24* e *naive-168*. No *naive-24*, Eq. 3.2, a demanda em uma hora de um dia é igual à da mesma hora do dia anterior:

$$\hat{L}_t = L_{t-24} \quad (3.2)$$

É fácil perceber que este método leva a grandes erros antes e depois dos fins de semana. Como os perfis diários são muito diferentes, prever as demandas de segunda-feira como sendo iguais às de domingo, por exemplo, produz resultados distantes da realidade. O previsor *naive-168* tenta contornar este problema ao escrever a demanda em uma hora de um dia como o valor observado na mesma hora do dia correspondente na semana anterior:

$$\hat{L}_t = L_{t-168} \quad (3.3)$$

Esta proposta remove a sazonalidade. O problema é que, agora, o dado utilizado na previsão é velho de sete dias e, assim, toda a informação que se passa durante a semana é totalmente descartada.

3.3.2 Tipo II: 168 previsores paralelos

A exemplo de Papalexoupoulos e Hesterberg [27], dividimos os dados em 168 séries, criamos um modelo para cada hora da semana, e aplicamos previsores separadamente. As séries de horas, por exemplo a da Fig. 3.5, não possuem nenhuma sazonalidade além da anual (recordamos, desconsiderada neste trabalho), o que nos permite utilizar métodos de previsão sem fatores sazonais. Escolhemos, então, médias móveis simples, amortecimentos exponenciais e modelos ARIMA.

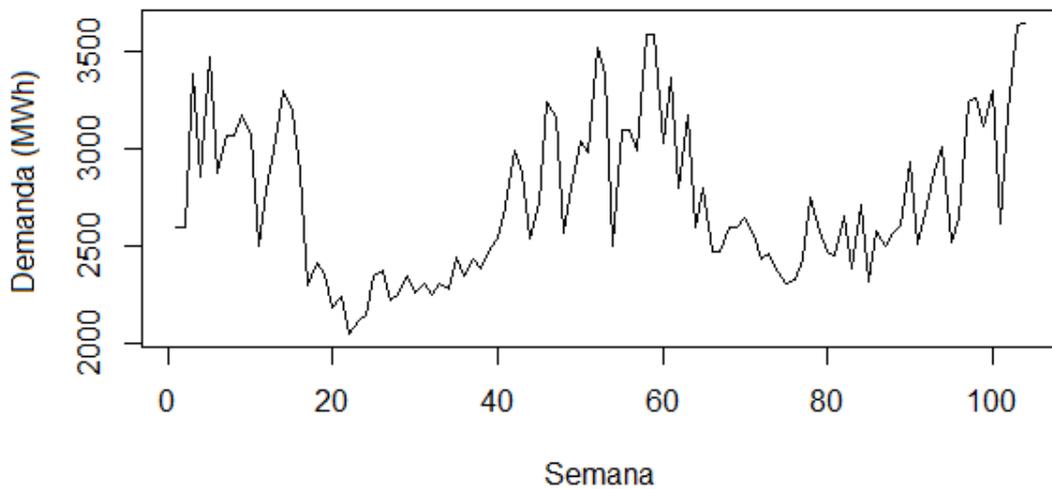


Figura 3.5: Demandas semanais de 0h, aos domingos, para 1996 e 1997 (Dados: Rio de Janeiro)

3.3.2.1 Médias móveis simples

A previsão por médias móveis simples para um período à frente, feita no instante $t - 1$, é igual à média aritmética simples das últimas N observações:

$$\hat{L}_t = \frac{L_{t-1} + L_{t-2} + \dots + L_{t-N}}{N} \quad (3.4)$$

O valor N é chamado de *tamanho da janela* da média móvel. À medida que os dados avançam no tempo, a janela também se desloca, daí o nome do método.

Uma média móvel simples é centrada no período $t - (N + 1)/2$, o que implica que a estimativa do nível local tende a estar atrás de seu verdadeiro valor em cerca de $(N + 1)/2$ períodos. Dizemos que esta é a quantidade de tempo que o método demora para reconhecer mudanças repentinas nos dados. Por exemplo, se calcularmos a média das últimas 7 horas, as previsões estarão atrasadas em 4 horas com relação a pontos de mudança brusca de nível.

O ponto-chave na implementação é decidir o tamanho da janela. Valores próximos a 1 fazem com que os resultados tendam ao predictor ingênuo $\hat{L}_t = L_{t-1}$, enquanto valores muito grandes podem criar uma suavização excessiva. Neste trabalho, seguimos a sugestão de Barros [49] e experimentamos inteiros de 1 a 10, escolhendo o tamanho de janela que

minimiza o erro médio quadrático das previsões (*mean-squared error*, MSE; Eq. 3.5, onde n é o número de observações).

$$MSE = \frac{1}{n} \sum_{i=1}^n (L_t - \hat{L}_t)^2 \quad (3.5)$$

3.3.2.2 Amortecimento exponencial simples

Uma extensão do método das médias móveis simples foi proposta por Brown em 1956 [50] e chamada de *amortecimento exponencial* (posteriormente, *amortecimento exponencial simples*; ou SES, *simple exponential smoothing*). A ideia é escrever uma observação como uma média ponderada das observações passadas cujos pesos são decrescentes: quanto mais antigas as observações, menos peso elas devem receber. Esta estratégia ajuda a rastrear melhor o nível local, dado que a informação mais recente é priorizada.

A ponderação é feita por uma constante de amortecimento α segundo a Eq. 3.6. A demanda durante os n períodos anteriores é sempre ponderada por $(1 - \alpha)^n$, daí o nome *exponencial*.

$$\hat{L}_{t+1} = \alpha \sum_{n=0}^t (1 - \alpha)^n L_{t-n} \quad (3.6)$$

A escolha de α , restrito a valores reais entre 0 e 1, governa a influência que as observações passadas exercem sobre as demandas futuras. Se o valor é próximo de 0, o decaimento dos pesos é lento, o que significa que o período-base é longo e, portanto, observações mais recentes não recebem pesos muito maiores que as mais antigas. Quanto maior o valor da constante, mais curto é o período-base, e maior a importância da última observação com relação às demais.

A Eq. 3.6 pode ser reescrita de forma recursiva. Desenvolvendo-a, ficamos com

$$\hat{L}_{t+1} = \alpha L_t + \alpha(1 - \alpha)L_{t-1} + \alpha(1 - \alpha)^2 L_{t-2} + \dots + \alpha(1 - \alpha)^t L_0 \quad (3.7)$$

Pode ser demonstrado que isto nos leva à formulação simplificada

$$\hat{L}_{t+1} = \alpha L_t + (1 - \alpha)\hat{L}_t \quad (3.8)$$

A Eq. 3.8 mostra outra vantagem do método: somente os últimos dados de observação e previsão precisam ser armazenados, reduzindo o custo computacional com relação ao

método de médias móveis.

Utilizamos a função *HoltWinters* do R [51], excluindo as componentes de tendência e sazonalidade, para o ajuste da constante α , e a função *forecast* para efetuar as previsões.

3.3.2.3 Método linear de Holt

Em seu primeiro estudo, Brown sugere como estimar uma parcela de tendência linear para séries com mudança de nível usando o método dos mínimos quadrados [50]. Paralelamente, Holt [52] desenvolveu outro amortecimento exponencial para séries com tendência, o qual chamamos *método linear de Holt*. Este método decompõe a série em uma soma do nível médio local l_t com uma tendência linear b_t , estimada pelas diferenças sucessivas entre níveis ($l_t - l_{t-1}$). O conjunto de equações do método é:

$$\begin{aligned} l_t &= \lambda L_t + (1 - \lambda)(l_{t-1} + b_{t-1}) \\ b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \\ \hat{L}_{t+1} &= l_t + b_t \end{aligned} \tag{3.9}$$

As constantes de amortecimento, λ e β , são números reais entre 0 e 1. Mais uma vez, efetuamos a estimação e a previsão pelas funções *HoltWinters* (excluindo a componente sazonal) e *forecast*, do R.

3.3.2.4 Modelos ARIMA

Modelos ARIMA (ou modelos de Box & Jenkins) são uma forma de descrever uma série temporal procurando reproduzir a autocorrelação entre os valores L_t e os passados L_{t-k} . As possibilidades são polinômios auto-regressivos (AR; *auto-regressive*), de médias móveis (MA; *moving average*) ou mistos (ARMA).

Um polinômio AR de ordem p , denotado AR(p), é a forma mais simples de se modelar a autocorrelação. Relaciona-se um valor L_t com os p passados por meio da combinação linear

$$L_t = c + \phi_1 L_{t-1} + \phi_2 L_{t-2} + \dots + \phi_p L_{t-p} + \varepsilon_t \tag{3.10}$$

onde c é o nível médio, ϕ_1, \dots, ϕ_k são os parâmetros do modelo e ε_t é parte de uma série de *ruídos brancos*, ou seja, variáveis aleatórias independentes e identicamente distribuídas (iid) com distribuição normal de média zero e variância constante.

Também podemos escrever a Eq. 3.13 como um polinômio de *operador-retardo*¹. Centralizando os valores Z_t , temos

$$\bar{L}_t = \phi_1 \bar{L}_{t-1} + \phi_2 \bar{L}_{t-2} + \dots + \phi_p \bar{L}_{t-p} + \varepsilon_t \quad (3.11)$$

Reescrevendo a Eq. 3.11, ficamos com

$$\begin{aligned} \varepsilon_t &= \bar{L}_t - \phi_1 \bar{L}_{t-1} + \phi_2 \bar{L}_{t-2} + \dots + \phi_p \bar{L}_{t-p} \\ &= \bar{L}_t - \phi_1 B \bar{L}_t + \phi_2 B^2 \bar{L}_t + \dots + \phi_p B^p \bar{L}_t \\ &= (1 - \phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p) \bar{L}_t \\ &= \Phi_p(B) \bar{L}_t, \end{aligned} \quad (3.12)$$

A autocorrelação também pode ser explicada pela combinação linear da série ε_t de ruídos passados, o que chamamos de polinômio MA de ordem q , ou MA(q):

$$L_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (3.13)$$

onde μ é o nível médio e $\theta_1, \dots, \theta_k$ são os parâmetros do modelo.

Da mesma forma que fizemos para o AR(p), podemos escrever o MA(q) como polinômio de operador retardo:

$$\begin{aligned} \bar{L}_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \\ &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \\ &= \Theta_q(B) \varepsilon_t, \end{aligned} \quad (3.14)$$

As Eqs. 3.12 e 3.14 também podem ser unidas para formar um polinômio misto. Como $\Phi(B) \bar{L}_t = \varepsilon_t$ (AR) e $\bar{L}_t = \Theta(B) \varepsilon_t$ (MA), tem-se

$$\Phi_p(B) \bar{L}_t = \Theta_q(B) \varepsilon_t \quad (3.15)$$

que é chamado ARMA(p, q).

Os polinômios AR(p), MA(q) e ARMA(p, q) ajustam-se melhor a séries cujo nível médio e a variância são constantes (séries *estacionárias*). Quando estes pressupostos não são atendidos, é preciso lançar mão do *operador de diferenciação* ∇^d , para remover as mudanças no nível e tornar a série estacionária.

¹Em análise de séries temporais, o operador-retardo B serve para defasar uma observação. Por exemplo, $BL_t = L_{t-1}$. Em geral, vale $B^k L_t = L_{t-k}$

O operador de diferenciação ∇^d subtrai algum valor defasado de uma observação de série temporal. Por exemplo, $\nabla L_t = L_t - L_{t-1}$; ou, analogamente, $\nabla L_t = (1 - B)L_t$. De modo geral, um operador ∇^d , de ordem d , calcula a diferença $\nabla L_d = \nabla^d L_t = \nabla \nabla \dots \nabla L_t = (1 - B)^d L_t$.

Aplicando o operador de diferenciação à Eq. 3.16, ficamos com o modelo geral ARIMA(p,d,q):

$$\Phi_p(B)\nabla^d \bar{L}_t = \Theta_q(B)\varepsilon_t \quad (3.16)$$

Os modelos ARIMA trazem como desafio a necessidade de determinação das ordens dos polinômios e estimação dos parâmetros. A análise das funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) pode auxiliar na escolha da ordem; contudo, como neste trabalho lidamos com 168 modelos, seria muito dispendioso verificar cada um individualmente. Adotamos, pois, o procedimento alternativo de testar, primeiro, modelos mais simples e depois aumentar sua complexidade aumentando termos e comparando os resultados. Este processo de busca está implementado na função *auto.arima*, parte do pacote *forecast*, do R [53], que determina a ordem por varredura em grade e estima os parâmetros pelo método da máxima verossimilhança.

Por padrão, o critério para seleção de modelos no R é a comparação do valor do critério de informação de Akaike (*Akaike Information Criterion*; AIC [54]). Este critério é uma função da variância dos erros (σ_a^2) e do número de parâmetros. A intenção é penalizar modelos com excesso de parâmetros: dado um conjunto de modelos candidatos, o modelo escolhido será aquele que minimizar o AIC. A penalização controla o sobreajuste e mantém a parcimônia.

A forma original do AIC é dada pela Eq. 3.17, onde L é a função de verossimilhança (*likelihood function*) do modelo e m é o número de parâmetros ($p + q + 1$; incluindo uma constante).

$$AIC = -2 \ln L + 2m \quad (3.17)$$

Como pode ser custoso calcular L , alguns programas utilizam a estimativa da Eq.

3.18, com N sendo o número total de observações na série.

$$-2 \ln L \cong N[1 + \ln 2\pi] + N \ln(\hat{\sigma}_a^2) \quad (3.18)$$

O AIC, portanto, fica

$$AIC \cong N[1 + \ln 2\pi] + N \ln(\hat{\sigma}_a^2) + 2m \quad (3.19)$$

Comparamos modelos sobre a mesma série e, logo, o valor N é igual. Podemos, então, suprimir o termo $N[1 + \ln 2\pi]$, que será o mesmo para todos os modelos. Assim, a Eq. 3.20 nos dá uma estimativa final do AIC.

$$AIC \cong N \ln(\hat{\sigma}_a^2) + 2m \quad (3.20)$$

É importante ressaltar que o AIC é uma medida para comparação de modelos candidatos, e não um teste (de hipóteses) de aderência e, portanto, não diz nada sobre a qualidade do ajuste.

3.3.3 Tipo III: 24 previsores paralelos

Neste caso, adotamos 24 modelos, um para cada hora do dia, aos quais os previsores são aplicados paralelamente.

As 24 séries apresentam sazonalidade semanal, com demandas mais baixas nos fins de semana (Fig. 3.6), e, por isso, selecionamos métodos capazes de estimar esta sazonalidade; o método de Holt-Winters e os modelos SARIMA.

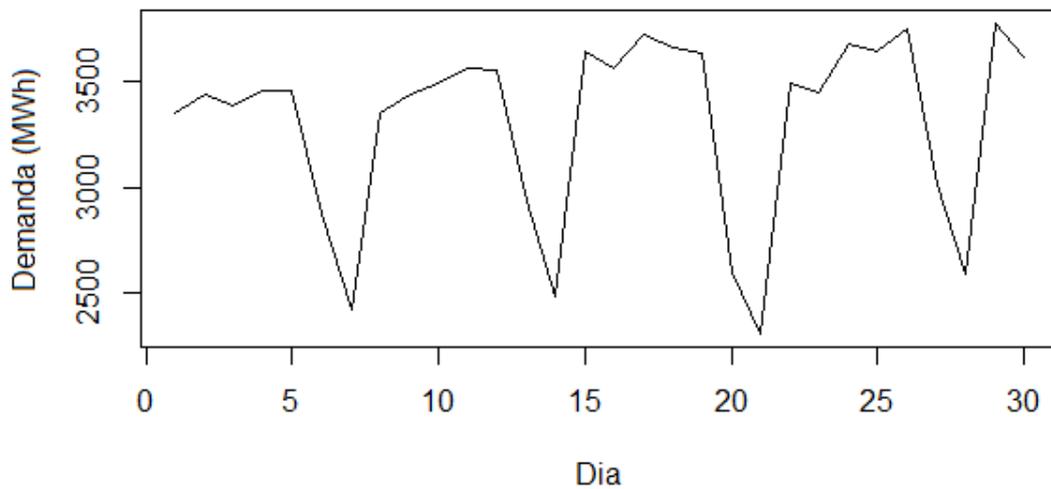


Figura 3.6: Cargas diárias de 11h a.m. para o mês de janeiro (Dados: Rio de Janeiro)

3.3.3.1 Método de Holt-Winters

O método de Holt-Winters, proposto por Winters [55], é adequado para séries com um ciclo sazonal. Similar ao método de Holt, a versão de sazonalidade multiplicativa do método é representada pelas Eqs. 3.21. É adicionada uma equação para estimar o fator sazonal local d_t , cujo período tem tamanho s (em nosso caso, $s = 24$ horas), pelo cálculo da razão entre o valor observado L_t e o nível l_t .

$$\begin{aligned}
 l_t &= \lambda(L_t/d_{t-s}) + (1 - \lambda)(l_{t-1} + b_{t-1}) \\
 b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \\
 d_t &= \delta(L_t/l_t) + (1 - \delta)d_{t-s} \\
 \hat{L}_{t+k} &= (l_t + kb_t)d_{t-s+k}
 \end{aligned} \tag{3.21}$$

A exemplo do amortecimento exponencial simples e do método linear de Holt, também usamos a função *HoltWinters*, do R, para estimar as constantes, reais entre 0 e 1, e a função *forecast* para prever os perfis.

3.3.3.2 Modelos SARIMA

Os modelos ARIMA podem ser adaptados para descrever a autocorrelação em séries sazonais. Por exemplo, uma série de dados horários cuja repetição do padrão sazonal

acontece diariamente (ou seja, de período sazonal s igual a 24), pode ser modelada pelo polinômio AR abaixo:

$$L_t = c_t + \phi_1 L_{t-24} + \phi_2 L_{t-48} + \phi_3 L_{t-72} \quad (3.22)$$

Em termos de operador-retardo, temos:

$$\begin{aligned} (1 - \phi_1 B^{24} - \phi_2 B^{48} - \phi_3 B^{72}) L_t &= c_t \\ \therefore \Phi(B^{24}) L_t &= c_t \end{aligned} \quad (3.23)$$

O modelo da Eq. 3.23 é um AR sazonal de período 24 com três defasagens, o qual chamamos de SAR(3).

A generalização é feita estendendo a inclusão da sazonalidade a um ARIMA(p,d,q). Seja $\nabla_s^D = (1 - B^s)^D$, onde D é o número de diferenciações e s é o tamanho do ciclo sazonal, o operador de diferença sazonal. Então,

$$\Phi_P(B^s) \nabla_s^D L_t = \Theta_Q(B^s) \varepsilon_t \quad (3.24)$$

é o polinômio ARIMA sazonal SARIMA(P, D, Q) _{s} . P é a ordem do polinômio autorregressivo sazonal SAR(P) dado por $\Phi_P(B^s)$; D é o número de diferenciações; Q é a ordem do polinômio de médias móveis sazonal SMA(Q) dado por $\Theta_Q(B^s)$; e s é o tamanho do ciclo sazonal.

Dadas as inúmeras possibilidades de ajuste para os 24 modelos, lançamos mão da função *auto.arima*, do pacote *forecast* do R, para escolher os mais parcimoniosos, a julgar pelo AIC.

3.3.4 Tipo IV: Um método de sazonalidade múltipla

Chamamos de métodos de sazonalidade múltipla aqueles capazes de acomodar, simultaneamente, mais de uma sazonalidade em um único modelo, que é o mesmo para todas as horas. Estes previsores tem sido objeto de estudo do professor James W. Taylor, que se dedica a explorar possibilidades de adaptações de métodos clássicos, principalmente os de amortecimento exponencial. Cabe destacar que, como diz a pesquisa de Hyndman *et al.* [56], amortecimentos exponenciais de múltipla sazonalidade não haviam sido considerados anteriormente.

Dos métodos propostos por Taylor, optamos por implementar o *amortecimento exponencial de Holt-Winters-Taylor* (HWT) para múltiplas sazonalidades, por conta de seu excelente desempenho comparado a métodos de formulações mais complexas, como mostram várias publicações do autor, e também por ainda ter sido pouco explorado segundo nossas pesquisas.

Inicialmente chamado de *amortecimento exponencial duplamente sazonal*, o HWT foi apresentado pela primeira vez no artigo [19], como uma forma de adaptar o método de Holt-Winters para acomodar dois ciclos sazonais ao mesmo tempo. O método é proposto na forma 3.25 e usado para modelar dados de demanda de energia elétrica da Inglaterra e do País de Gales.

$$\begin{aligned}
 l_t &= \alpha \left(\frac{L_t}{d_{t-s_1} w_{t-s_2}} \right) + (1 - \alpha)(l_{t-1} + b_{t-1}) \\
 b_t &= \gamma(l_t - l_{t-1}) + (1 - \gamma)b_{t-1} \\
 d_t &= \delta \left(\frac{L_t}{l_t w_{t-s_2}} \right) + (1 - \delta)d_{t-s_1} \\
 w_t &= \omega \left(\frac{L_t}{l_t d_{t-s_1}} \right) + (1 - \omega)w_{t-s_2} \\
 \hat{L}_t &= (l_t + kb_t)d_{t-s_1+k}w_{t-s_2+k}
 \end{aligned} \tag{3.25}$$

Esta primeira formulação, de tendência linear b_t aditiva e sazonalidade multiplicativa é similar à do método Holt-Winters das Eqs. 3.21. A novidade é que são considerados dois fatores sazonais separados; um relativo ao ciclo diário (d_t , de período s_1) e outro ao ciclo semanal (w_t , de período s_2). α , β , δ e ω são as constantes de amortecimento.

Na execução do método, nível, tendência e sazonalidades locais são reestimados de hora em hora. Para a previsão da demanda de um instante $t + k$, são somadas a atualização mais recente do nível, l_t , e da tendência, b_t , às atualizações dos fatores sazonais no período correspondente mais recente. Por exemplo, para prever as 15h de uma terça-feira, usamos a estimativa do nível local às 24h de segunda, do índice diário às 15h de segunda, e do índice semanal às 15h da terça passada.

Esta proposta, além de possibilitar que se explore a correlação da demanda de uma hora com as demandas na mesma hora no dia anterior e na semana anterior, contorna o problema da superparametrização imposta pela divisão dos dados em 24 e 168 modelos. Isto tudo é feito com um único previsor de apenas três coeficientes, reduzindo muito o custo computacional e fazendo do HWT um método convidativo.

No artigo seminal [19], o método é comparado com duas formulações do Holt-Winters tradicional, uma para sazonalidade diária e outra para sazonalidade semanal. Por não levar em conta a diferença entre dias de semana e fins de semana, os resultados do primeiro são tão ruins que são imediatamente descartados. O Holt-Winters com fator sazonal semanal produz melhores previsões, mas o HWT o desbanca e mostra que levar em conta as duas sazonalidades pode melhorar consideravelmente os resultados.

No mesmo artigo, Taylor propõe um refinamento dos resultados (procedimento que seria feito em todas as suas publicações posteriores sobre o HWT), acrescentando um termo de ajuste. Como os erros de previsão um passo à frente, $\varepsilon_t = L_t - \hat{L}_t$, são altamente correlacionados, existe a suspeita de que o método não capture toda a informação trazida pelos dados. Para melhorar a qualidade das previsões, Taylor baseia-se em Chatfield [57], e efetua um ajuste que envolve modificar as previsões k passos à frente da origem t adicionando um termo de ajuste do último erro disponível, $\phi^k \varepsilon_t$, como na Eq. 3.26.

$$\hat{L}_t = (l_t + kb_t)d_{t-s_1+k}w_{t-s_2+k} + \phi^k \varepsilon_t \quad (3.26)$$

Os resultados mostram que o acréscimo deste termo, além de diminuir os erros, causa mudanças significativas nos valores das outras constantes. A equação do nível médio, inclusive, torna-se redundante, com $\lambda = 0$, o que quer dizer que o nível passa a ser pressuposto constante a todo momento.

Com o sucesso dos primeiros experimentos, o autor decidiu investigar melhor o método, e passou a fazer diversos testes e adaptações. Taylor *et al.* [58] o aplicam a dados da cidade do Rio de Janeiro (os mesmos desta dissertação) e País de Gales e Inglaterra, comparando com os resultados de um modelo ARIMA e de uma rede neural artificial. Surpreendentemente, o HWT, com suas equações mais simples, supera os resultados de ambos.

Um ano depois, Taylor *et al.*, em [59], mostram uma aplicação mais abrangente na qual o HWT é confrontado com um modelo ARIMA, um AR periódico, uma nova adaptação de amortecimento exponencial (análoga do HWT, com fatores sazonais diferentes para cada dia da semana) e um método baseado na análise de componentes principais (*principal component analysis*, PCA). Desta vez, as previsões são feitas para diversas bases de dados de SEEs europeus (Bélgica, Finlândia, França, Grã-Bretanha, Irlanda, Portugal, Itália, Noruega, Espanha e Suécia). Para todas as séries, o HWT funciona melhor que os

previsores concorrentes.

Para verificar como o HWT atuaria na resolução de outro tipo de problema, Taylor também o aplicou à previsão de chamadas em *call-centers*, com dados de 5 empresas britânicas [60]. Neste artigo somos apresentados a uma versão do HWT para sazonalidade aditiva, dada pelas Eqs. 3.27.

$$\begin{aligned}
\hat{L}_{t+k} &= l_t + kb_t + d_{t-s_1+k} + w_{t-s_2+k} + \phi^k(L_t - (l_{t-1} + b_{t-1} + d_{t-s_1} + w_{t-s_2})) \\
l_t &= \lambda(L_t - d_{t-s_1} - w_{t-s_2}) + (1 - \lambda)(l_{t-1} + b_{t-1}) \\
b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \\
d_t &= \delta(L_t - l_{t-1} - w_{t-s_2}) + (1 - \delta)d_{t-s_1} \\
w_t &= \omega(L_t - l_{t-1} - d_{t-s_1}) + (1 - \omega)w_{t-s_2}
\end{aligned} \tag{3.27}$$

Outra situação diferente testada foi a previsão da demanda de energia em curtíssimo prazo (para horizontes de 1 a 30 minutos) [61]. Aqui o HWT funciona, mais uma vez, muito bem, e ainda mostra-se que sua combinação com o resultado de outros métodos pode melhorar ainda mais as previsões.

Trabalhos mais recentes mostram comparações com outros métodos de múltipla sazonalidade [62] (artigo em que o método, pela primeira vez, é chamado de amortecimento exponencial de Holt-Winters-Taylor); uma extensão para acomodar um terceiro ciclo sazonal, anual, adequado para quando se tem uma série histórica de muitos anos; [11]; e modelos de espaço de estados, para variáveis aleatórias contínuas [63], [64].

Nesta dissertação, a formulação do HWT é aditiva e sem a equação de tendência (3.28), já que a análise gráfica sugere que a mudança de nível nos dados do Rio de Janeiro é bastante discreta.

$$\begin{aligned}
\hat{L}_{t+k} &= l_t + d_{t-s_1+k} + w_{t-s_2+k} + \phi^k(L_t - (l_{t-1} + d_{t-s_1} + w_{t-s_2})) \\
l_t &= \lambda(L_t - d_{t-s_1} - w_{t-s_2}) + (1 - \lambda)l_{t-1} \\
d_t &= \delta(L_t - l_{t-1} - w_{t-s_2}) + (1 - \delta)d_{t-s_1} \\
w_t &= \omega(L_t - l_{t-1} - d_{t-s_1}) + (1 - \omega)w_{t-s_2}
\end{aligned} \tag{3.28}$$

Como lidamos com dados espaçados de hora em hora, o período do ciclo diário é $s_1 = 24$ e o do semanal é $s_2 = 168$. Fazemos também a correção pela ponderação do último erro com peso ϕ^k .

Para inicialização do método, estimamos o nível \hat{l}_{168} como a média dos valores observados em todas as horas da primeira semana, e os fatores sazonais da primeira semana (diários e semanais) como a diferença entre os valores observados e o nível estimado, $L_t - \hat{l}_{168}$. Para estimar os valores das constantes, restritos a reais entre 0 e 1, usamos a função *optim*, do R, um mecanismo de busca iterativa dirigido pelo algoritmo BFGS (BroydenFletcherGoldfarbShanno) de memória limitada [65]. O BFGS método numérico *quasi*-Newton para otimização de funções não lineares que, neste caso, escolhe os valores das constantes que minimizam o MSE das previsões.

3.4 Métodos multivariados de previsão

O modelo multivariado considerado foi o aditivo de curva de carga da Eq. 3.29, com a componente base, B_t , relacionada às demandas passadas, e W_t relacionada à temperatura do ar.

$$L_t = B_t + W_t + R_t \quad (3.29)$$

Primeiramente, a componente-base é estimada pelo melhor predictor univariado entre os testados. Depois, os desvios $L_t - \hat{B}_t$ são previstos como função da componente climática W_t ; relação modelada por regressão linear múltipla e por uma rede neural artificial.

3.4.1 Regressão linear múltipla

Apesar de a correlação entre demanda e temperatura, na base de dados do Rio de Janeiro, ser não-linear, seguimos a sugestão de Davies [20] e começamos os testes por um modelo menos complexo.

Como discutimos no Cap. 2, os efeitos da temperatura do ar sobre a demanda de energia não são facilmente perceptíveis em uma base horária, motivo pelo qual escolhemos escrever um modelo em que W_t não somente como função da temperatura mais atual, como também das diferenças entre a temperatura atual e a de um dia e uma semana atrás (Eq. 3.30).

$$W_t = \beta_0 + \beta_1 T_t + \beta_2 (T_t - T_{t-24}) + \beta_3 (T_t - T_{t-168}) \quad (3.30)$$

Foi ajustado um único modelo para todas as horas. Para encontrar o viés β_0 e os coeficientes β_i , fizemos o uso da função lm , do R, que ajusta um modelo de regressão linear pelo método dos mínimos quadrados [66]

Pormenores sobre a teoria de regressão linear e estimação por mínimos quadrados podem ser encontrados em Draper e Smith [67].

3.4.2 Redes neurais artificiais (RNAs)

As RNAs são um mecanismo de processamento paralelo de grande porte cuja unidade básica de processamento é o *neurônio artificial*, estrutura representada na Fig. 3.7. Os principais elementos de um neurônio são seus *nós de entrada*, ponderados por pesos w_i ; um *combinador linear*, que soma as entradas ponderadas; e uma *função de ativação* f , função contínua e diferenciável que usa a combinação linear como argumento para produzir a saída do neurônio.

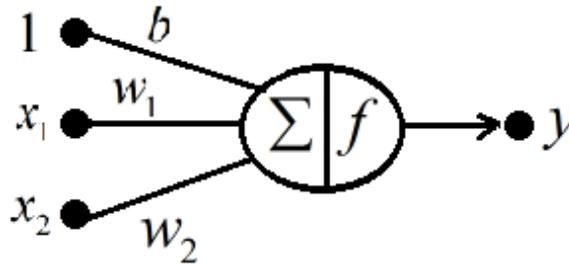


Figura 3.7: Exemplo de neurônio artificial

Como função de ativação, são normalmente escolhidas a função degrau, para problemas de classificação, e as funções sigmoidais, para problemas de aproximação de funções. Exemplos de sigmoidais (Fig. 3.8) são a função logística (Eq. 3.31), para valores de saída entre 0 e 1, e a tangente hiperbólica (Eq. 3.32), para valores de saída entre -1 e 1.

$$y(x_i) = \frac{1}{(1 + e^{-x_i})} \quad (3.31)$$

$$y(x_i) = \tanh(x_i) = \frac{1 - e^{-x}}{1 + e^{-x}} \quad (3.32)$$

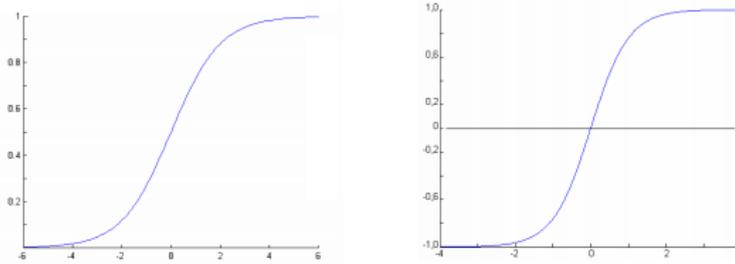


Figura 3.8: Funções logística (à esquerda) e tangente hiperbólica (à direita)

O neurônio da Fig. 3.7 possui três entradas: um viés, representado pela unidade ponderada por um peso b , e duas variáveis, x_1 e x_2 , ponderadas por pesos w_1 e w_2 . O combinador linear Σ produz $b + w_1x_1 + w_2x_2$, que é usada como argumento de f . Se esta função for a logística, por exemplo, teremos, como saída do neurônio, $y(x_i) = 1/(1 + e^{-(b+w_1x_1+w_2x_2)})$.

Uma rede neural é um agrupamento de neurônios segundo alguma arquitetura. A arquitetura mais popular para previsão da demanda de energia elétrica é a do tipo *perceptron multicamadas* (MLP; *multilayer perceptron*). Em um MLP, os neurônios são organizados em duas ou mais camadas e a informação se propaga de maneira *feedforward*, isto é, uma camada tem como entrada as saídas de sua anterior.

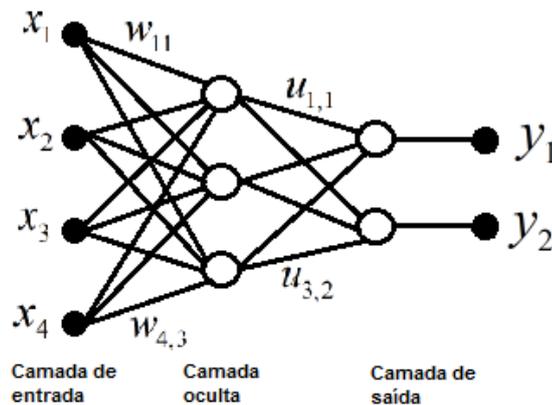


Figura 3.9: Exemplo de MLP com uma camada oculta

No exemplo da Fig. 3.9, o MLP possui quatro entradas e duas saídas, com uma camada interna (ou *oculta*) de três neurônios. A informação flui da camada de nós de entrada para a camada de saída, com os neurônios anteriores conectados aos seguintes (mas nunca lateralmente). No exemplo da figura, em cada neurônio da camada interna é feita uma combinação linear dos nós de entrada, com pesos $w_{i,j}$, que é passada como argumento de uma função de ativação, por exemplo, uma tangente hiperbólica. Os neurônios da

camada de saída fazem, então, uma combinação linear, com pesos $u_{j,k}$, das tangentes hiperbólicas da camada anterior. Estas combinações servem de argumento para outra função de ativação, digamos, uma função identidade ($y(x_i) = x_i$), e são enviadas para a camada de saída. Assim, as saídas y_k tem a forma

$$y_k = \sum_{i=1}^4 (u_{j,k} \tanh \sum_{i=1}^3 w_{i,j} x_i) \quad (3.33)$$

Uma das formas de estimar os pesos (i.e. *treinar* a RNA), é o procedimento de *aprendizagem supervisionada*. Neste procedimento, a rede recebe uma série de vetores de entrada, acompanhados da respectiva saída desejada. A diferença entre a saída desejada e aquela gerada pela rede é utilizada para alterar os pesos a fim de que as saídas se aproximem dos valores desejados. O ajuste é repetido até que algum critério de parada seja atingido. As atualizações são feitas por algoritmos de busca iterativa para minimização de uma função de perda (normalmente uma função quadrática dos erros) relativamente ao valor dos pesos. A maioria destes algoritmos buscam pelo mínimo caminhando na direção contrária do gradiente da função de perda.

As RNAs testadas neste trabalho são MLPs de uma camada oculta (funções de ativação logísticas) cujos nós de entrada são as mesmas variáveis da regressão linear (T_t , $T_t - T_{t-24}$, $T_t - T_{t-168}$), e a saída, univariada, é uma estimativa do desvio $L_t - B_t$.

Na implementação computacional, usamos função *mlp*, do pacote RSNNS (*R Stuttgart Neural Network Simulator*), do R [68]. O algoritmo de treinamento é o gradiente conjugado escalonado (SCG; *scaled conjugate gradient*). Um algoritmo de gradiente conjugado tenta, iterativamente, se aproximar do mínimo, caminhando a cada passo na direção conjugada dos passos anteriores. O SCG difere de outros métodos de gradiente conjugado pela forma com que estima a matriz Hessiana (matriz de derivadas de segunda ordem), o que o torna consideravelmente mais rápido que outros métodos de busca [69].

Como critérios de parada do algoritmo de busca, escolhemos o número máximo de épocas e o momento em que o erro no processo de validação cruzada começa a aumentar. Esperamos, após treinada uma rede, que ela seja capaz de produzir saídas razoáveis não somente para os dados apresentados no processo de treinamento, mas também para entradas desconhecidas. Isto significa que a rede deve ter boa capacidade de generalizar seu conhecimento para novas situações. Uma forma de aprimorar a capacidade de generalização é o uso do método de *validação cruzada* (*cross-validation*). Dividimos os

dados de treinamento (amostra de ajuste II) em dois grupos distintos, sendo 80% das amostras separadas para treinamento e 20% para validação. A cada época (iteração em que são apresentados todos os padrões disponíveis na amostra de treino), calculamos o erro na amostra de validação. Interrompemos o treinamento se o erro na amostra de validação começar a aumentar ou 200 épocas forem executadas, o que ocorrer primeiro.

A ideia por trás da condição relacionada à validação cruzada é evitar o sobreajuste aos dados de treinamento (*overfitting*). Na Fig. 3.10, vemos que o erro na amostra de validação inicialmente diminui com o número de épocas, mas existe um ponto em que ele passa a aumentar, consequência do *overfitting*. Tomando por referência os erros cometidos em uma amostra não vista previamente, nos asseguramos que a rede apresente boa capacidade de generalização.

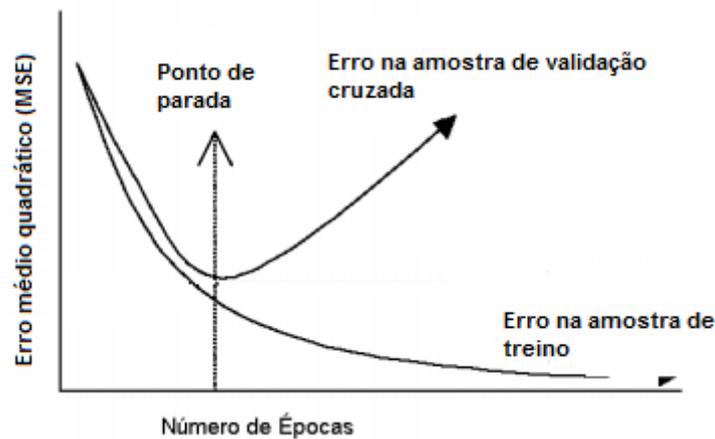


Figura 3.10: Erros nas amostras de treino e validação cruzada ([70], adaptada)

3.5 Comparação com resultado externo

Os resultados obtidos neste trabalho foram comparados com os obtidos por Hippert *et al.* [12]. No trabalho destes autores, é concebido, para as mesmas amostras da mesma base de dados, um MLP de uma camada oculta cujas entradas são as 48 demandas dos últimos dois dias e mais duas *dummies* para indicar se é fim de semana ou dia de semana. As saídas são as 24 horas do perfil do dia seguinte.

3.6 Métrica de erro

Para a avaliação dos métodos testados é usado o erro médio percentual absoluto (MAPE), comumente utilizada nas publicações em previsão da demanda de energia. A vantagem desta medida está na interpretação direta, posto que os resultados, calculados pela Eq. 3.35, são dados em pontos percentuais.

Calculamos o erro percentual absoluto (APE) como a razão do módulo da diferença entre um valor previsto e um observado e o valor observado, multiplicada por 100, como na Eq. 3.34.

$$APE = \frac{|L_t - \hat{L}_t|}{L_t} \times 100 \quad (3.34)$$

O MAPE de uma amostra de n previsões é a média de seus APEs (Eq. 3.35).

$$MAPE = \sum_{i=0}^n \frac{|L_t - \hat{L}_t|}{L_t} \times \frac{100}{n} \quad (3.35)$$

4 RESULTADOS E DISCUSSÃO

Para a implementação dos métodos, os dados foram divididos em quatro amostras:

- Semanas 1 a 50 - amostra de ajuste I
- Semanas 51 a 70 - amostra de teste I
- Semanas 1 a 84 - amostra de ajuste II
- Semanas 85 a 104 - amostra de teste II

4.1 Métodos univariados

Primeiramente, testamos os previsores univariados discutidos na Seção 3.3 para o modelo de curva de carga da Eq. 4.1.

$$L_t = S_t + R_t \quad (4.1)$$

Utilizamos a amostra de ajuste I para inicializar os métodos e calcular os valores das constantes e parâmetros. Com os valores encontrados, computamos o MAPE para previsões de perfis na amostra de teste I. Os valores dos erros estão dispostos na Tabela 4.1.

Tabela 4.1: MAPEs (%) dos métodos univariados na amostra de teste I

		MAPE
Tipo I - Previsores ingênuos	<i>naive-24</i>	5,23
	<i>naive-168</i>	7,87
Tipo II - 168 previsores paralelos	Médias móveis simples	4,82
	Am. exp. simples	5,37
	Método linear de Holt	5,08
	ARIMA	4,84
Tipo III - 24 previsores paralelos	Método de Holt-Winters	3,83
	SARIMA	3,36
Tipo IV - Um previsor longitudinal	Método de Holt-Winters-Taylor	2,76

A leitura da Tabela 4.1 mostra que a divisão em 24 previsores paralelos supera a divisão em 168 previsores. No entanto, é o uso de um único previsor com duas sazonalidades, o

HWT, produz melhores previsões que as segmentações em modelos individuais, indicando que pouco benefício é obtido ao se aplicar um previsor por hora (alternativa de maior custo computacional).

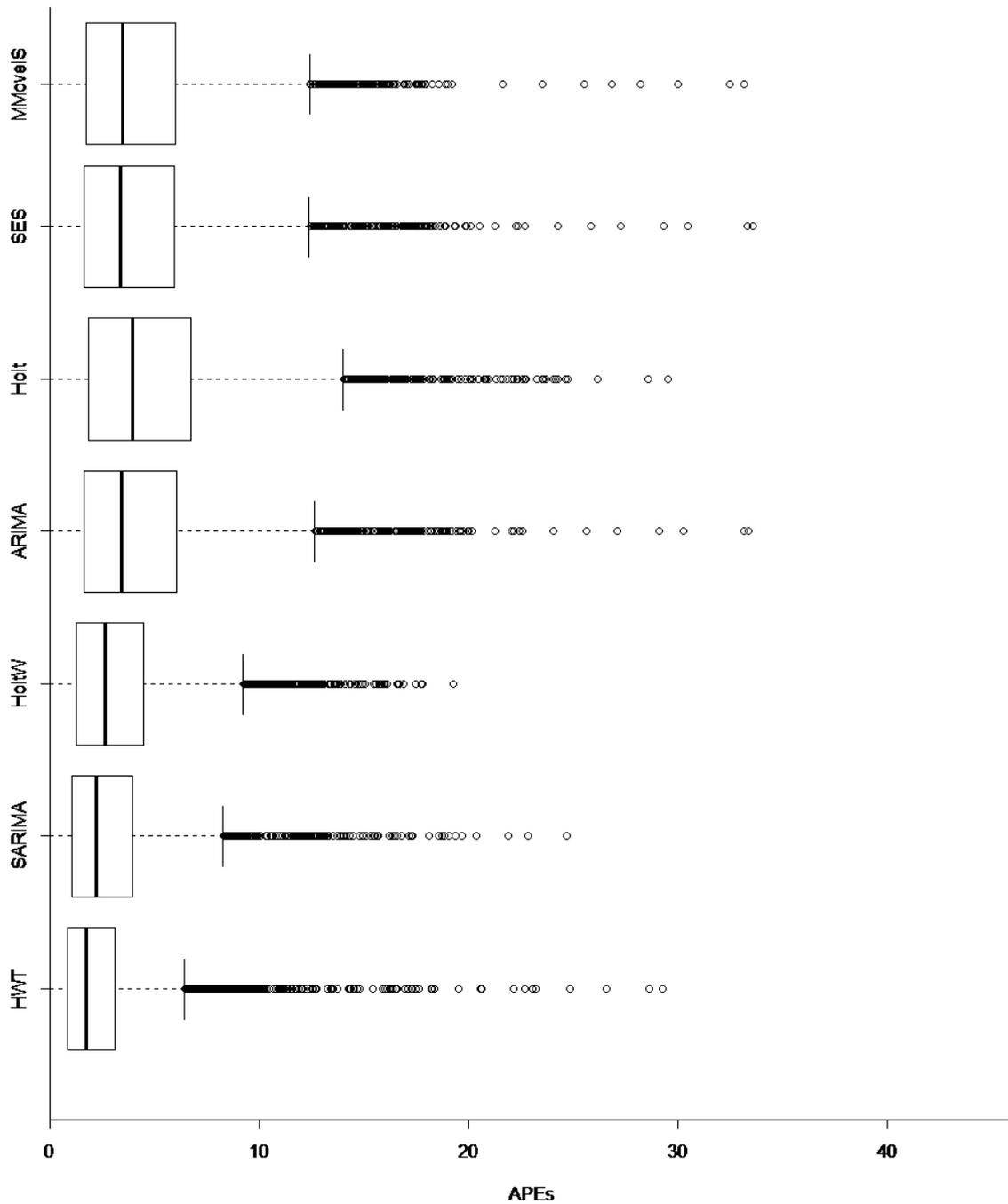


Figura 4.1: APEs dos métodos univariados

Deduzimos da Fig. 4.1 que, além de menor MAPE, o HWT também apresenta menor dispersão dos APEs, conferindo robustez ao método. Isto é muito útil para a aplicação prática em sistemas de previsão automatizados, já que o tempo de reação do operador do

sistema a alguma anomalia é baixo, principalmente no contexto do curto prazo.

Os valores das constantes das equações do HWT, calculados na amostra de ajuste I e dispostos na Tabela 4.2, remontam aos obtidos por Taylor em [19]; ao considerar a correção para a autocorrelação dos erros de primeira ordem (constante ϕ), a equação do nível torna-se redundante ($\lambda = 0$). É razoável pressupor o nível constante, já que, em dois anos de dados, a tendência da demanda no Rio de Janeiro parece evoluir muito discretamente, como vimos na análise gráfica da Seção 3.1.

Tabela 4.2: Constantes de amortecimento do HWT (estimados na amostra de ajuste I)

Constante	Valor
ϕ	0,94
λ	0,00
δ	0,21
ω	0,22

Traçando um gráfico de linha dos MAPEs das previsões do HWT por hora do dia (Fig. 4.2), percebemos que as horas pertencentes ao intervalo entre 7h a 17h parecem mais fáceis de se prever. Provavelmente isto se deve ao fato de a rotina dos consumidores, no horário comercial, variar pouco.

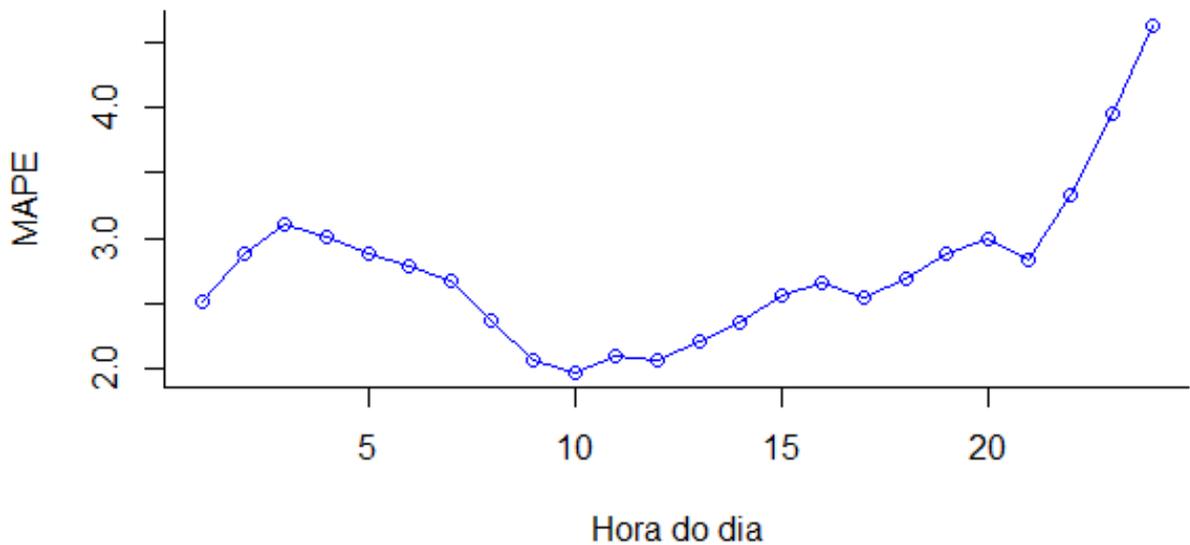


Figura 4.2: MAPEs por hora do dia (HWT, amostra de teste I)

Já a observação dos MAPEs das previsões do HWT por dia da semana (Fig. 4.3) deixa claro que os fins de semana (incluindo sexta-feira) devem ser mais difíceis de se prever. Podemos atribuir esta dificuldade à maior variação nas atividades cotidianas das pessoas

durante os fins de semana.

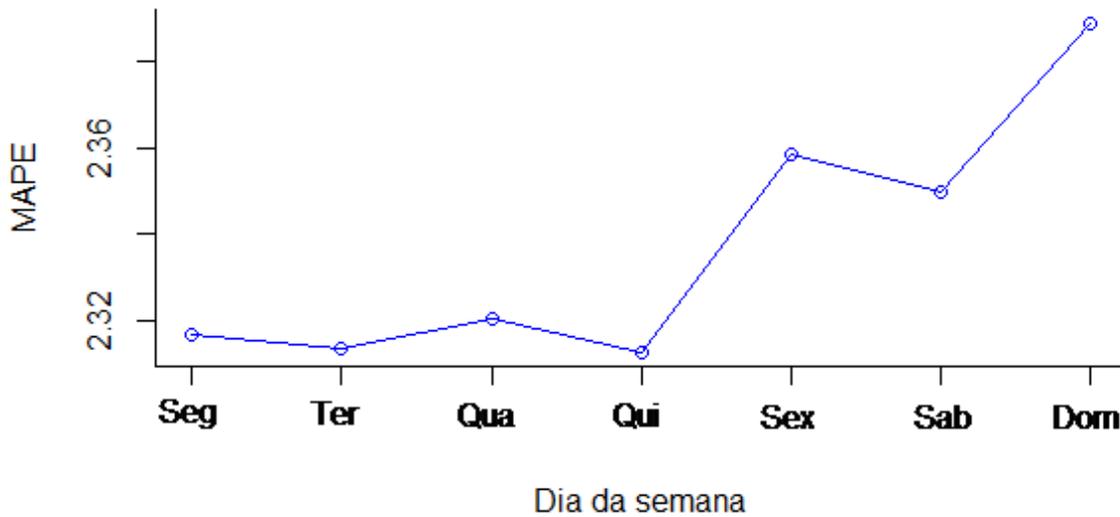


Figura 4.3: MAPEs por dia da semana (HWT, amostra de teste I)

4.2 Métodos multivariados

Selecionado o HWT como melhor previsor univariado, ajustamos um modelo de curva de carga padrão no formato 4.2.

$$L_t = B_t + W_t + R_t \quad (4.2)$$

Primeiro, prevemos B_t utilizando o HWT. Em seguida, modelamos os erros de previsão como função da componente climática W_t , isto é, $W_t = L_t - \hat{B}_t$. Para isto, escrevemos W_t como uma combinação linear de variáveis relacionadas à temperatura: a temperatura T no instante t e as diferenças entre esta e as observadas no dia e na semana anteriores (Eq. 4.3). Os coeficientes são encontrados calculando os erros na amostra de ajuste II.

$$W_t = \beta_0 + \beta_1 T_t + \beta_2 (T_t - T_{t-24}) + \beta_3 (T_t - T_{t-168}) \quad (4.3)$$

Embora os valores-p (Tabela 4.3) sugiram que todas as variáveis sejam significativas, o coeficiente de correlação linear de Pearson (R^2) é de apenas 0,05, indicando correlação linear muito baixa entre as variáveis de entrada e saída. De fato, a escala dos coeficientes da Tabela 4.3 mostra que a magnitude da componente climática é ínfima perto dos valores de B_t , que são da ordem de milhares de kWh (exemplos na Tabela 4.4). Logo, W_t pouco

acrescenta à previsão da demanda.

Tabela 4.3: Coeficientes do modelo de regressão usando as temperaturas reais

Coeficiente	Valor	Valor-p
β_0	-20,44	0,00
β_1	0,86	0,00
β_2	-6,08	0,00
β_3	-2,82	0,00

Tabela 4.4: Valores das componentes base e climática para as primeiras 12 horas em MWh (dia 1 da semana 85)

Hora do dia	B_t	W_t
01	2704,31	7,52
02	2302,17	4,36
03	2187,51	8,44
04	2141,65	9,82
05	2122,26	6,62
06	2158,89	3,54
07	2258,40	1,32
08	2456,45	-0,87
09	2615,96	1,34
10	2891,39	6,92
11	3116,45	11,85
12	3116,45	14,01

Depois dos testes com regressão linear, fazemos uma nova tentativa de prever os erros $L_t - B_t$ com base em W_t , desta vez utilizando uma rede neural artificial (RNA) do tipo MLP com uma camada oculta. Testamos 4 MLPs, cujas entradas são as mesmas variáveis explicativas da regressão linear ($T_t, T_t - T_{t-24}, T_t - T_{t-168}$) e a variável de saída também é o erro $L_t - B_t$. As quatro redes diferem entre si quanto ao número de neurônios na camada oculta, 1, 3, 5 e 10. Como as RNAs são métodos iterativos sensíveis a condições iniciais, os resultados provenientes de treinamentos diferentes de uma rede com mesmas especificações (variáveis, números de neurônios etc.) e esperado que os MAPEs apresentem alguma variância. Portanto, simulamos 100 treinamentos para cada rede e dispusemos os MAPEs, calculados na amostra de teste II, para cada treinamento de cada rede, nos *boxplots* da Fig. 4.4. Destacamos que, apesar de os resultados com três ou mais neurônios (similares entre si) desbancarem os da rede com um neurônio apenas, a variância é tão baixa que não se justifica investir mais recursos computacionais em redes com mais neurônios, que requerem o cálculo de mais combinações de funções sigmoidais.

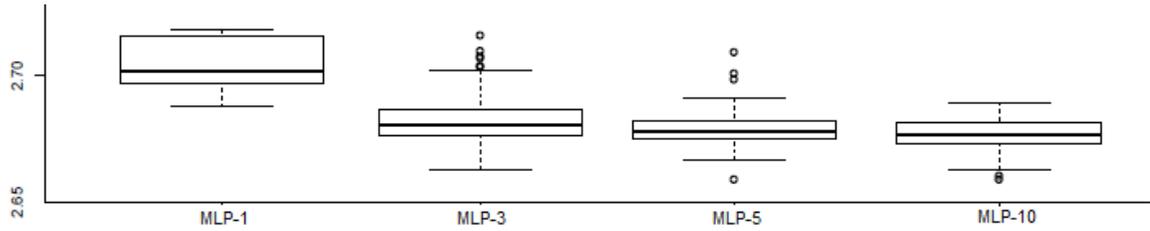


Figura 4.4: MAPEs (%) para 100 treinamentos de cada RNA

4.2.1 Comparação entre todos os métodos

Por fim, comparamos os resultados do modelo univariado previsto pelo HWT, do modelo multivariado previsto por HWT e regressão linear, do modelo de curva de carga previsto por HWT e RNAs e pela RNA publicada em [12].

Recapitulamos que a RNA da publicação é um MLP de uma camada oculta cujas entradas são as 48 demandas dos últimos dois dias e mais duas *dummies* para indicar se é fim de semana ou dia de semana e cujas saídas são as 24 horas do perfil do dia seguinte.

Para esta comparação, reestimamos as constantes do HWT (4.5) usando a amostra de ajuste II, a fim de calcular os erros na amostra de teste II e comparar os resultados com os dos métodos multivariados.

Tabela 4.5: Constantes de amortecimento do HWT (amostra de ajuste II)

Constante	Valor
ϕ	0,94
λ	0,00
δ	0,20
ω	0,17

Na Tabela 4.6 vemos que, embora tenhamos conseguido reduzir o MAPE com o uso de um modelo multivariado de curva de carga, esta redução é bastante discreta, restringindo-se a, no máximo, 0,16% pontos percentuais (caso da regressão linear). Este resultado, em uma situação prática, nos leva a optar pelo uso do HWT de dupla sazonalidade (previsões na Fig. 4.5), visto que, com o uso de muito menos recursos computacionais e uma formulação relativamente simples, este método pode proporcionar resultados de acurácia similar à de previsores mais complexos, e levando em conta apenas uma variável de entrada; o que garante maior robustez, já que o sistema não é alimentado por fontes externas (com previsões de temperatura, por exemplo).

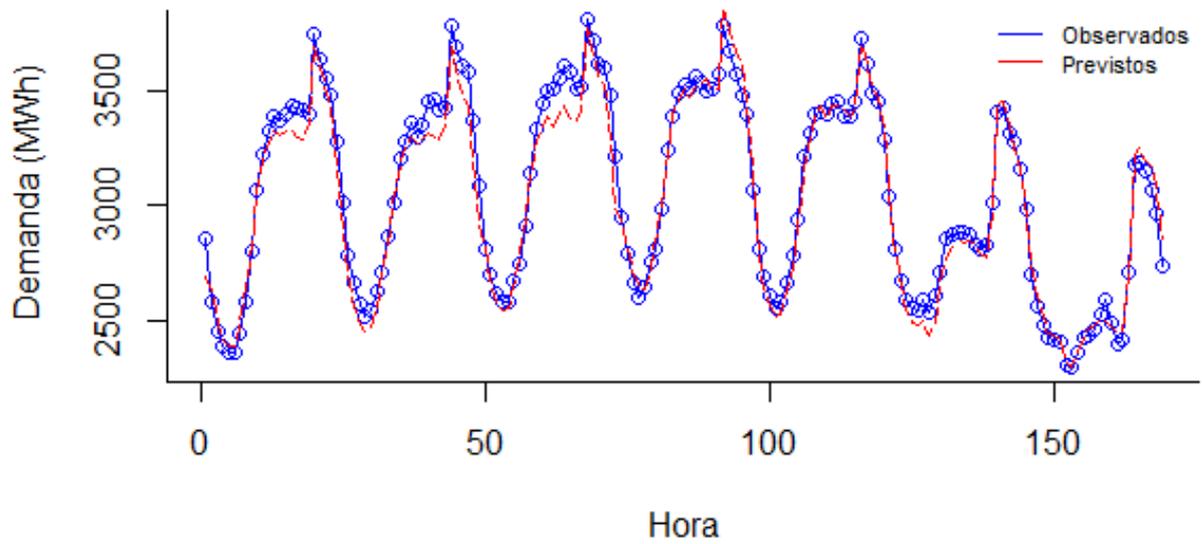


Figura 4.5: Valores observados e previstos pelo HWT (semana 85, amostra de teste II)

Tabela 4.6: MAPEs calculados na amostra de teste II

		MAPE (%)
HWT		2,47
Regressão linear+HWT		2,29
HWT+MLP-1	Mínimo	2,41
	Médio	2,44
	Máximo	2,47
HWT+MLP-3	Mínimo	2,36
	Médio	2,39
	Máximo	2,47
HWT+MLP-5	Mínimo	2,36
	Médio	2,38
	Máximo	2,44
HWT+MLP-10	Mínimo	2,35
	Médio	2,37
	Máximo	2,40

5 CONCLUSÕES

Métodos estatísticos baseados na autocorrelação dos valores em séries temporais dominaram a literatura em previsão da demanda de energia elétrica por muitos anos, até que, com a popularização das ferramentas de inteligência computacional, maior ênfase passou a ser dada às redes neurais artificiais (RNAs). O principal motivo, sem dúvida, é o fato de a grande flexibilidade na especificação das redes torná-las capazes de aproximar os tipos mais complexos de função entre dois conjuntos de variáveis. Com isso, toda relação da demanda com seus valores passados e, principalmente, com variáveis exógenas (como as climáticas), passou a poder ser mapeada sem a necessidade de modelos pré-determinados. No entanto, como mostramos nesta dissertação, um simples método univariado de amortecimento exponencial pode produzir resultados de acurácia similar à de uma RNA, com uma formulação bem menos complexa e menos consumo de recursos computacionais.

Como a inclusão de variáveis climáticas pouco acrescentou ao modelo de curva de carga padrão, os resultados deste trabalho corroboram a opinião dos autores (como discutido no Cap. 1) que afirmam que, no curto prazo, variáveis como a temperatura exercem pouca influência sobre o consumo de energia elétrica, e, portanto, as mudanças no nível já devem estar implícita na própria série de demandas.

O previsor HWT, que se destacou entre os demais, provou-se um método valioso para aplicação prática, já que é relativamente simples de se compreender e implementar, e se mostra acurado para previsão de demanda de energia em curto prazo. Outra vantagem é que é preciso calcular o valor de apenas quatro constantes, processo de baixo custo se comparado com a estimação dos pesos de uma RNA, por exemplo. Por fim, vale mencionar que o uso de um modelo univariado linear garante:

- (i) *Robustez de especificação*; porque o modelo possui boa capacidade de generalização, operando bem em partes da série temporal para a qual é sub-ótimo.
- (ii) *Robustez a falha externa*; porque o modelo é baseado apenas nas demandas passadas e, portanto, o sistema precisa ser alimentado com variáveis exógenas de fontes externas, não estando sujeito a erros nestas variáveis ou falhas de comunicação.

Extensões deste trabalho foram publicadas como resumo no *PanAm (VIII Pan-American Workshop Applied and Computational Mathematics)* [71], apresentadas em pôster no XI SIMMEC / II EMMCOMP 2014 (XI Simpósio de Mecânica Computacional / II Encontro Mineiro de Modelagem Computacional) [72] e submetidas como trabalhos completos para publicações nos anais do XXI SIMPEP (Simpósio de Engenharia de Produção).

Para trabalhos futuros, sugerimos experimentar os mesmos métodos em outras bases de dados e também experimentar modificações no HWT, por exemplo, para modelar separadamente os fatores sazonais dos dias de semana e fins de semana, de forma análoga à tentativa de Taylor com seu *amortecimento exponencial intracíclico (IC exponential smoothing)* [62].

APÊNDICE A - CONCEITOS BÁSICOS EM SÉRIES TEMPORAIS

Neste apêndice, fazemos uma breve introdução ao estudo de séries temporais e, ao mesmo tempo, apresentamos a terminologia empregada nesta dissertação. Para maiores referências, ver Chatfield [14], Makridakis et al. [46], Morettin e Tolói [47] e Box et al. [48].

A.1 Conceitos introdutórios

Uma série temporal é um conjunto de observações feitas sequencialmente ao longo do tempo. Normalmente estas observações são tomadas em intervalos equidistantes (por exemplo, de hora em hora). A principal característica deste tipo de dados é a presença de *autocorrelação* entre as observações subsequentes, no sentido de que os valores futuros dependem dos passados. Por conta disso, as técnicas de análise e modelagem são totalmente distintas das convencionais, que normalmente trabalham com o pressuposto de independência dos dados.

O estudo de séries temporais pode servir a vários objetivos, dentre os quais nos interessam, para este trabalho, a *análise* e, principalmente, a *previsão*. A análise de séries temporais é uma metodologia baseada em ferramentas diversas, principalmente gráficas, que ajudam a identificar componentes das séries, como mudanças bruscas e padrões repetitivos. Já as previsões são extrapolações de valores futuros a partir de informações extraídas de valores passados. Previsões são subsídio para a tomada de decisões em diversas áreas do conhecimento, desempenhando um papel crucial para as empresas, que necessitam dispor de uma estimativa de sua demanda para programar as atividades com a antecedência exigida.

A.2 Terminologia e decomposição de séries temporais

O valor de uma variável Z observado no instante t (que pode ser uma hora, um dia, um ano etc.) é dado por Z_t . Uma previsão para Z_t é dada por \hat{Z}_t . Podemos expressar a previsão em termo de seu *horizonte* k : uma previsão para o valor de Z no instante $t + k$, feita no instante t , é dada por $\hat{Z}_t(k)$. Esta previsão também pode ser chamada de *previsão para Z k passos à frente de t* .

Uma forma de analisar uma série temporal é decompô-la na forma da Eq. A.1, onde S_t é um padrão e R_t é um erro.

$$Z_t = S_t + R_t \quad (\text{A.1})$$

O padrão S_t pode ser decomposto em função de uma componente de nível (l_t) e outra de sazonalidade (d_t). O termo que resta é um erro (ou *resíduo*) aleatório, R_t , decorrente de fatores de natureza imprevisível.

O *nível* da série é o valor médio ao redor do qual oscilam as observações da série. A componente de *sazonalidade*, por sua vez, corresponde a flutuações que se repetem periodicamente.

A relação entre as componentes de uma série temporal pode ser aditiva:

$$Z_t = l_t + d_t + R_t \quad (\text{A.2})$$

Alternativamente, a decomposição multiplicativa é dada por

$$Z_t = l_t \times d_t \times R_t \quad (\text{A.3})$$

Na Fig. A.1, ilustramos a decomposição graficamente; nos painéis, temos, em ordem, a série original, a componente de nível, a componente sazonal e o resíduo.

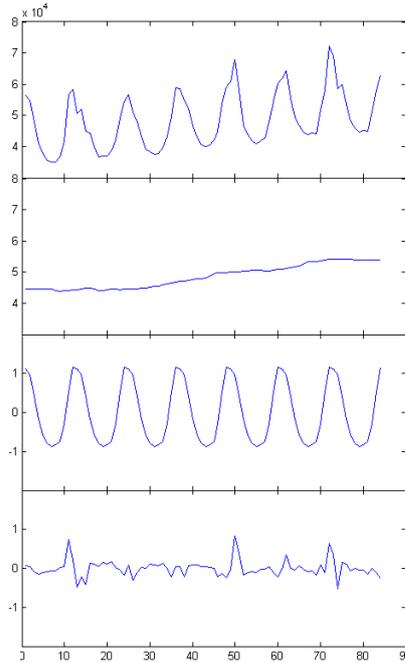


Figura A.1: Gráfico de decomposição (série de demandas de energia elétrica dos EUA [15])

Também pode-se incluir uma componente para a *tendência* da série. A tendência em uma série temporal, denotada por b_t , é a componente que reflete o comportamento de longo prazo da série, indicando evoluções no valor médio (também chamado *valor esperado* ou {nível da série}). Agregando a tendência à formulação aditiva, ficamos com

$$Z_t = l_t + b_t + s_t + R_t \quad (\text{A.4})$$

Neste caso, os valores da série flutuam (periodicamente, devido à sazonalidade, ou erráticamente, devido ao resíduo) em torno de um valor médio previsível l_t , o nível, o qual, com o tempo, desloca-se para cima ou para baixo devido à tendência.

Métodos clássicos para a estimação das componentes de uma série temporal, baseados em médias móveis, tem origem na década de 20, ainda são muito usados. Recentemente, com o avanço tecnológico, tem sido propostos métodos mais complexos, embora muitos deles ainda se fundamentem na decomposição clássica.

No Capítulo 3, mostramos vários métodos para decompor uma série temporal e estimar suas componentes.

REFERÊNCIAS

- [1] EMPRESA DE PESQUISA ENERGÉTICA (EPE), “Resenha Mensal do Mercado de Energia Elétrica”, <http://www.epe.gov.br/>, 2013.
- [2] EMPRESA DE PESQUISA ENERGÉTICA (EPE), “Demanda de energia elétrica - 10 Anos (informe à imprensa)”, <http://www.epe.gov.br/>, 2012.
- [3] BLUME, S., *Electric Power System Basics for the Nonelectrical Professional*. Wiley: Hoboken, 2007.
- [4] GRAINGER, J.J.; STEVENSON JR., W., *Power System Analysis*. McGraw Hill: New York, 1994.
- [5] GROSS, G.; GALIANA, F., “Short-term Load Forecasting”, *Proceedings of the IEEE*, v. 75, n. 12, pp. 1558–1573, 1987.
- [6] BUNN, D., “Forecasting Loads and Prices in Competitive Power Markets”, *Proceedings of the IEEE*, v. 88, n. 2, pp. 163–169, 2000.
- [7] KATZ, M., *Multivariable Analysis: A Practical Guide for Clinicians*. Cambridge University Press: Cambridge, 1999.
- [8] GUPTA, P.C.; YAMADA, K., “Adaptive Short-term Forecasting of Hourly Loads Using Weather Information”, *IEEE Transactions on Power Apparatus and Systems*, v. PAS-91, n. 5, pp. 2085–2094, 1972.
- [9] RAHMAN, S.; HAZIM, O., “A Generalized Knowledge-based Short-term Load Forecasting Technique”, *IEEE Transactions on Power Systems*, v. 8, n. 2, pp. 508–514, 1993.
- [10] FARMER, E.D.; POTTON, M., “Development of Online Load Prediction Techniques with Results from Trials in the South-west Region of the CEGB”, *Proceedings of the Institution of Electrical Engineers*, v. 115, n. 10, pp. 1549–1558, 1968.
- [11] TAYLOR, J., “Triple Seasonal Methods for Short-term Electricity Demand Forecasting”, *European Journal of Operational Research*, v. 204, pp. 139–152, 2010.

- [12] HIPPERT, H.S.; BUNN, D. S. R., “Large Neural Networks for Electricity Load Forecasting: Are they overfitted?” *International Journal of Forecasting*, v. 21, n. 3, pp. 425–434, 2005.
- [13] BUNN, D.W.; FARMER, E., *Comparative Models for Electrical Load Forecast*. Wiley: New York, 1985.
- [14] CHATFIELD, C., *The Analysis of Time Series: An Introduction*. 6th ed. CRC Press: Boca Raton, 2002.
- [15] ELECTRIC POWER RESEARCH INSTITUTE (EPRI), “Probabilistic methods in forecasting hourly loads”, 1993.
- [16] BUNN, D., “Short-term Forecasting: A Review of Procedures in the Electricity Supply Industry”, *Journal of the Operational Research Society*, v. 33, n. 6, pp. 533–545, 1982.
- [17] CHRISTIAANSE, W., “Short-term Load Forecasting Using General Exponential Smoothing”, *IEEE Transactions on Power Apparatus and Systems*, v. PAS-90, n. 2, pp. 900–911, 1971.
- [18] PARK, J.H.; PARK, Y. L. K., “Composite Modeling for Adaptive Short-term Load Forecasting”, *IEEE Transactions on Power Systems*, v. 6, n. 2, pp. 450–457, 1991.
- [19] TAYLOR, J., “Short-term Electricity Demand Forecasting Using Double-seasonal Exponential Smoothing”, *Journal of Operational Research Society*, v. 54, n. 8, pp. 799–805, 2003.
- [20] DAVIES, M., “The Relationship Between Weather and Electricity Demand”, *Proceedings of the IEE - Part C: Monographs*, v. 106, n. 9, pp. 27–37, 1959.
- [21] HAGAN, M.T.; BEHR, S., “The Time Series Approach to Short Term Load Forecasting”, *IEEE Transactions on Power Systems*, v. 2, n. 3, pp. 785–791, 1987.
- [22] HAIDA, T.; MUTO, S., “Regression Based Peak Load Forecasting Using a Transformation Technique”, *IEEE Transactions on Power Systems*, v. 9, n. 2, pp. 1788–1794, 1994.

- [23] BARAKAT, E.H.; AL-QASEM, J., “Methodology for Weekly Load Forecasting”, *IEEE Transactions on Power Systems*, v. 13, n. 4, pp. 1548–1555, 1998.
- [24] STANTON, K.N.; GUPTA, P., “Forecasting Annual or Seasonal Peak Demand on Electric Utility Systems”, *IEEE Transactions on Power Apparatus and Systems*, v. PAS-89, n. 5, pp. 951–959, 1970.
- [25] ABOU-HUSSSIEN, M.S.; KANDIL, M. T. M. F. S., “An accurate model for short-term load forecasting”, *IEEE Transactions on Power Apparatus and Systems*, v. PAS-100, n. 9, pp. 4158–4165, 1981.
- [26] ENGLE, R.F.; GRANGER, C. R. J. W. A., “Semiparametric Estimates of the Relation Between Weather and Electricity Sales”, *Journal of the American Statistical Association*, v. 81, n. 394, pp. 310–320, 1986.
- [27] PAPALEXOPOULOS, A.D.; HESTERBERG, T., “A Regression-based Approach for Short-term System Load Forecasting”, *IEEE Transactions on Power Systems*, v. 5, n. 4, pp. 1535–1547, 1990.
- [28] CANCELO, J.R.; ESPASA, A. G. R., “Forecasting the Electricity Load from One Day to One Week Ahead for the Spanish System Operator”, *International Journal of Forecasting*, v. 24, n. 4, pp. 588–602, 2008.
- [29] FAN, S.; CHEN, L., “Short-term Load Forecasting Based on an Adaptive Hybrid Method”, *IEEE Transactions on Power Systems*, v. 21, n. 1, pp. 392–401, 2006.
- [30] RAMANATHAN, R., ENGLE, R., GRANGER, C., VAHID-ARAGHI, F., BRACE, C., “Short-run Forecasts of Electricity Loads and Peaks”, *International Journal of Forecasting*, v. 13, n. 2, pp. 161–174, 1997.
- [31] CRAENEN, B.G.W.; EIBEN, A., “Computational Intelligence”, In: *Encyclopedia of Life Support Systems*, EOLSS Co. Ltd., 2003.
- [32] JAIN, ANIL K.; MAO, J. M. K., “Artificial Neural Networks: A Tutorial”, *Computer*, v. 29, n. 3, pp. 31–44, 1996.
- [33] CHEN, B.; CHANGE, M. L. C., “Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001”, *IEEE Transactions on Power Systems*, v. 19, n. 4, pp. 1821–1830, 2004.

- [34] ELATTAR, E.E.; GOULERMAS, J. W. Q., “Electric Load Forecasting Based on Locally Weighted Support Vector Regression”, *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, v. 40, n. 4, pp. 438–447, 2010.
- [35] RAHMAN, S.; BHATNAGAR, R., “An Expert System Based Algorithm for Short Term Load Forecast”, *IEEE Transactions on Power Systems*, v. 3, n. 2, pp. 392–399, 1988.
- [36] NAZARKO, J.; ZALEWSKI, W., “The Fuzzy Regression Approach to Peak Load Estimation in Power Distribution Systems”, *IEEE Transactions on Power Systems*, v. 14, n. 3, pp. 809–814, 1999.
- [37] HIPPERT, H.S.; PEDREIRA, C. S. R., “Neural Networks for Short-term Load Forecasting: A Review and Evaluation”, *IEEE Transactions on Systems*, v. 16, n. 1, pp. 44–55, 2001.
- [38] FIDALGO, J.N.; PEÇAS LOPES, J., “Load Forecast Performance Enhancement when Facing Anomalous Events”, *IEEE Transactions on Power Systems*, v. 20, n. 1, pp. 408–415, 2005.
- [39] OSMAN, Z.H.; AWAD, M. M. T., “Neural Network Based Approach for Short-term Load Forecasting”. In: *Power Systems Conference and Exposition*, pp. 1–8, 2009.
- [40] NEWBOLD, P.; GRANGER, C., “Experience with Forecasting Univariate Time Series and the Combination of Forecasts”, *Journal of the Royal Statistical Society*, v. 137, n. 2, pp. 131–165, 1974.
- [41] CLEMEN, R., “Combining Forecasts: A Review and Annotated Bibliography”, *International Journal of Forecasting*, v. 5, n. 4, pp. 559–583, 1990.
- [42] KHOTANZAD, A.; AFKHAMI-ROHANI, R. M. D., “ANNSTLF - Artificial Neural Network Short-term Load Forecaster - Generation Three”, *IEEE Transactions on Power Systems*, v. 13, n. 4, pp. 1413–1422, 1998.

- [43] TAYLOR, J.W.; MAJHITIA, S., “Using Combined Forecasts with Changing Weights for Electricity Demand Profiling”, *Journal of the Operational Research Society*, v. 51, n. 1, pp. 72–82, 2000.
- [44] SONG, K.; BAEK, Y. H. D. J. G., “Short-term Load Forecasting for the Holidays Using Fuzzy Linear Regression Method”, *IEEE Transactions on Power Systems*, v. 20, n. 1, pp. 96–101, 2005.
- [45] CRAN (THE COMPREHENSIVE R ARCHIVE NEWTORK), “What is R?” <http://www.r-project.org/>, 2014.
- [46] MAKRIDAKIS, S.; WHEELWRIGHT, S. H. R., *Forecasting: Methods and Applications*. 3rd ed. Wiley: New York, 1998.
- [47] MORETTIN, P.A.; TOLOI, C., *Análise de Séries Temporais*. 2nd ed. Edgard Blucher: São Paulo, 2006.
- [48] BOX, G.E.P.; JENKINS, G. R. G., *Time Series Analysis: Forecasting and Control*. 4th ed. Wiley: New York, 2008.
- [49] BARROS, M., “Capítulo 5: Introdução às Séries Temporais e aos Modelos ARIMA”, <http://www.mbarros.com/documentos/upload/Livro Processos Cap 5 parte.pdf>, 2014.
- [50] BROWN, R., “Exponential Smoothing for Predicting Demand”. In: *Tenth national meeting of the Operation Research Society of America*, 1956.
- [51] CRAN (THE COMPREHENSIVE R ARCHIVE NEWTORK), “Holt-Winters Filtering”, <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/HoltWinters.html>, 2014.
- [52] HOLT, C., *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*. N. 52, Carnegie Institute of Technologie, 1957.
- [53] HYNDMAN, R., “Forecasting package for R”, <http://robjhyndman.com/software/forecast/>, 2014.
- [54] AKAIKE, H., “A New Look at the Statistical Model Identification”, *IEEE Transactions on Automatic Control*, v. 19, n. 16, pp. 716–723, 1974.

- [55] WINTERS, ., “Forecasting Sales by Exponentially Weighted Moving Averages”, *Management Science*, v. 6, pp. 324–342, 1960.
- [56] HYNDMAN, R.J.; KOEHLER, A. G. S., “A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods”, *International Journal of Forecasting*, v. 18, n. 3, pp. 439–454, 2002.
- [57] CHATFIELD, C., “The Holt-Winters Forecasting Procedure”, *Applied Statistics*, v. 27, pp. 264–279, 1978.
- [58] TAYLOR, J.W.; DE MENEZES, L. M. P., “A Comparison of Univariate Methods for Forecasting Electricity Demand Up to a Day Ahead”, *International Journal of Forecasting*, v. 22, n. 1, pp. 1–16, 2006.
- [59] TAYLOR, J.W.; MCSHARRY, P., “Short-term Load Forecasting Methods: An Evaluation Based on European Data”, *IEEE Transactions on Power Systems*, v. 22, n. 4, pp. 2213–2219, 2007.
- [60] TAYLOR, J., “A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center”, *Management Science*, v. 54, n. 2, pp. 253–265, 2008.
- [61] TAYLOR, J., “An Evaluation of Methods for Very Short-Term Load Forecasting Using Minute-by-Minute British Data”, *International Journal of Forecasting*, v. 24, n. 4, pp. 645–658, 2008.
- [62] TAYLOR, J., “Exponentially Weighted Methods for Forecasting Intraday Time Series with Multiple Seasonal Cycles”, *International Journal of Forecasting*, v. 26, n. 4, pp. 627–646, 2010.
- [63] TAYLOR, J., “Density Forecasting of Intraday Call Center Arrivals using Models Based on Exponential Smoothing”, *Management Science*, v. 58, n. 3, pp. 534–549, 2012.
- [64] TAYLOR, J.W.; SNYDER, R., “Forecasting Intraday Time Series with Multiple Seasonal Cycles Using Parsimonious Seasonal Exponential Smoothing”, *Omega*, v. 40, n. 6, pp. 748–757, 2012.

- [65] SHANNO, D., “Conditioning of Quasi-Newton Methods for Function Minimization”, *Math. Comput.*, v. 24, pp. 647–656, 1970.
- [66] CRAN (THE COMPREHENSIVE R ARCHIVE NEWTORK), “Fitting linear models”, <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/lm.html>, 2014.
- [67] DRAPER, N.; SMITH, H., *Applied Regression Analysis*. 3rd ed. Wiley: New York, 1998.
- [68] CRAN (THE COMPREHENSIVE R ARCHIVE NEWTORK), “RSNNS: Neural Networks in R using the Stuttgart Neural Network Simulator (RSNNS)”, <http://cran.r-project.org/web/packages/RSNNS/>, 2014.
- [69] MOLLER, M., “A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning”, *Neural Networks*, v. AC-19, n. 4, pp. 525–533, 1993.
- [70] SILVA, R., *Redes Neurais Artificiais Aplicadas à Detecção de Intrusão em Redes TCP/IP*, Dissertação de mestrado, Pontifícia Universidade Católica (PUC), Rio de Janeiro, 2005.
- [71] GUILHERMINO NETO, G., “Short-Term Load Forecasting Using an Updated Load Curve Formulation”. In: *PanAm: VIII Pan-American Workshop Applied and Computational Mathematics*, 2014.
- [72] GUILHERMINO NETO, G., DEFILIPPO, S., CARNEIRO, A., HIPPERT, H., “Previsão da Demanda de Energia Elétrica em Curto Prazo por Modelos de Curva de Carga: Um Estudo Comparativo”. In: *XI SIMMEC / II EMMCOMP: XI Simpósio de Mecânica Computacional (SIMMEC) / II Encontro Mineiro de Modelagem Computacional*, 2014.