

Anna Cláudia Mancini da Silva Carneiro

**Previsão do consumo de energia elétrica a curto prazo, usando combinações
de métodos univariados**

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. D.Sc. Henrique Steinherz Hippert

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)

Mancini da Silva Carneiro, Anna Claudia .

Previsão do consumo de energia elétrica a curto prazo, usando combinações de métodos univariados / Anna Claudia Mancini da Silva Carneiro. -- 2014.
60 f.

Orientador: Henrique Steinherz Hippert

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

1. Previsão de cargas elétricas. 2. Perfis de carga. 3. Modelos univariados de previsão. 4. Combinação de previsões. 5. Séries temporais. I. Steinherz Hippert, Henrique, orient. II. Título.

Anna Cláudia Mancini da Silva Carneiro

**Previsão do consumo de energia elétrica a curto prazo, usando combinações
de métodos univariados**

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 26 de Setembro de 2014.

BANCA EXAMINADORA

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

Prof. Ph.D Ronaldo Rocha Bastos
Universidade Federal de Juiz de Fora

D.Sc. José Márcio Faier
ICSystems-Sistemas de Inteligência Computacional LTDA

À minha mãe, Kátia.

AGRADECIMENTOS

À minha mãe, Kátia, meus avós Vera e Élcio, ao Eduardo, ao Gabriel e à tia Karla, por estarem sempre comigo, nos momentos de alegria e tristeza, sempre me apoiando e nunca permitindo que eu desistisse; por todo amor, carinho e compreensão. Em especial, à minha mãe e aos meus avós, pois devo a eles tudo o que eu sou e conquistei até hoje.

Ao Pedro, por todos os anos ao meu lado, tornando os meus dias mais felizes, por todo amor, companheirismo e dedicação.

Ao Henrique, pela paciência, atenção e confiança. Sem ele, não teria sido possível finalizar este trabalho.

Aos colegas do mestrado, principalmente à Janaína e ao João, por toda ajuda nesses dois anos.

Às minhas amigas Anna Paula, Érica e Ana Luíza, por mais de uma década de amizade, e que, mesmo longe, torcem por mim e fazem parte da minha felicidade.

*‘No meio da dificuldade
encontra-se a oportunidade.’*

Albert Einstein

RESUMO

A previsão de cargas elétricas é fundamental para o planejamento das empresas de energia. O foco deste estudo são as previsões a curto prazo; assim, aplicamos métodos univariados de previsão de séries temporais a uma série real de cargas elétricas de 104 semanas no Rio de Janeiro, nos anos de 1996 e 1997, e experimentamos várias combinações dos métodos de melhor desempenho.

As combinações foram feitas pelo método *outperformance*, uma combinação linear simples, com pesos fixos. Os resultados das combinações foram comparados ao de simulações de redes neurais artificiais que solucionam o mesmo problema, e ao resultado de um método de amortecimento de dupla sazonalidade aditiva. No geral, este método de amortecimento obteve os melhores resultados, e talvez seja o mais adequado e confiável para aplicações práticas, embora necessite de melhorias para garantir a extração completa da informação contida nos dados.

Palavras-chave: Previsão de cargas elétricas. Perfis de cargas. Modelos univariados de previsão. Combinação de previsões. Séries temporais.

ABSTRACT

Forecasting the demand for electric power is crucial for the production planning in energy utilities. The focus of this study are the short-term forecasts. We apply univariate time series methods to the forecasting of a series containing observations of the energy consumption of 104 weeks in Rio de Janeiro, in 1996 and 1997, and experiment with several combinations of the methods which have the best performance.

These combinations are done by the *outperformance* method, a simple linear combination with fixed weights. The results were compared to those obtained by neural networks on the same problem, and with the results of a exponential smoothing method for dual additive seasonality. Overall, the exponential smoothing method achieved the best results, and was shown to be perhaps the most reliable and suitable for practical applications, even though it needs improvements to ensure complete extraction of the information contained in the data.

Keywords: Electrical load forecasting. Load profile. Univariate forecasting models. Combined forecasts. Time series.

SUMÁRIO

1	INTRODUÇÃO.....	12
2	CARACTERIZAÇÃO DO PROBLEMA	14
2.1	A previsão de cargas elétricas	14
2.2	Modelos usuais de previsão de cargas elétricas	16
3	MATERIAIS E MÉTODOS.....	19
3.1	Materiais	19
3.2	Métodos	23
3.2.1	<i>Introdução</i>	23
3.2.2	<i>Método Naive</i>	25
3.2.3	<i>Método das Médias Móveis</i>	26
3.2.4	<i>Amortecimento Exponencial Simples</i>	27
3.2.5	<i>Amortecimento Exponencial de Holt</i>	29
3.2.6	<i>Método de Holt-Winters</i>	29
3.2.7	<i>Método de Holt-Winters-Taylor</i>	30
3.2.8	<i>Modelos ARIMA</i>	31
3.2.9	<i>Combinação de previsões</i>	35
3.2.10	<i>Medidas de erro de previsão</i>	37
3.2.11	<i>Função $optim()$</i>	39
3.2.12	<i>Modelos propostos</i>	40
4	RESULTADOS E DISCUSSÕES	43
5	CONCLUSÕES	52
	REFERÊNCIAS	55

LISTA DE ILUSTRAÇÕES

2.1	Previsões de perfis de cargas elétricas (adaptado de [1]).	18
3.1	Efeito do feriado no comportamento padrão da série de cargas (adaptado de [2]).	19
3.2	Utilização das amostras antes das combinações.	20
3.3	Utilização das amostras após combinações.	20
3.4	Cargas médias semanais nos anos de 1996 e 1997.	21
3.5	Perfis de cargas de quinzenas de verão e inverno.	22
3.6	Perfis de cargas de dias de verão e inverno.	22
4.1	<i>Boxplots</i> dos APEs dos métodos simples para a amostra de teste I.	45
4.2	<i>Boxplots</i> dos APEs das combinações e dos métodos comparativos para a amostra de teste II.	47
4.3	Histograma dos desvios do previsor de HWT para a amostra de teste II. . . .	48
4.4	Gráfico de dispersão dos desvios do previsor de HWT para a amostra de teste II.	49
4.5	Função de autocorrelação dos desvios do previsor de HWT para a amostra de teste II.	49
4.6	Função de autocorrelação parcial dos desvios do previsor de HWT para a amostra de teste II.	50

LISTA DE TABELAS

4.1	MAPEs dos métodos simples para a amostra de teste I.	44
4.2	Pesos das combinações.	45
4.3	MAPEs das combinações e dos métodos comparativos para a amostra de teste II.	46

1 INTRODUÇÃO

A energia elétrica é uma das formas de energia mais utilizada no mundo, como fonte de calor, luz e força, estando diretamente ligada ao desenvolvimento tecnológico e econômico. Sob um ponto de vista técnico, uma concessionária de energia elétrica pode ser vista como um sistema produtivo, em que usinas, turbinas e geradores combinados com combustível ou água (*inputs*) produzem a energia (*output*), que será distribuída ao cliente.

Para o funcionamento adequado e eficiente de um sistema produtivo, é necessário o planejamento e controle da produção em termos de prazo. No setor energético, decisões a longo prazo são a respeito da redução ou ampliação da capacidade de produção, transmissão e distribuição, envolvendo as unidades geradoras de energia — instalações de novas plantas ou linhas de produção, por exemplo, que necessitam de anos para serem concretizadas. A médio prazo, em intervalos de semanas ou meses, podem ser resolvidas questões sobre a utilização da capacidade produtiva, como a compra de combustível e máquinas. A curto prazo, em intervalos de horas ou dias, são tomadas decisões a respeito da otimização do uso de máquinas e instalações.

A previsão de cargas elétricas, então, é fundamental para o planejamento das empresas de energia [3]. A produção de energia deve atender à demanda, sem que haja desperdícios, uma vez que a energia produzida em excesso não pode ser estocada e pode gerar prejuízos para a empresa, com o gasto desnecessário de combustível, água e equipamentos, além de multas aplicadas pela Agência Nacional de Energia Elétrica (ANEEL), no caso do Brasil. Por outro lado, a produção de energia abaixo da esperada pode causar falhas no abastecimento.

As privatizações das concessionárias e a desregulamentação dos sistemas de energia levaram à criação de mercados de energia altamente competitivos [4], com foco na produção de energia com altos padrões de qualidade, a custos baixos. As empresas passaram a depender de previsões de cargas mais acuradas devido ao alto custo dos erros de previsão. Particularmente as previsões com horizonte de previsão de um dia a frente [4] se tornaram cada vez mais importantes para a operação das empresas de energia.

Os dados sobre a demanda de energia se apresentam na forma de séries temporais de cargas elétricas. A previsão dos valores futuros dessas séries pode ser feita através de

diversos métodos, baseados em modelos *univariados* — em que a carga de um instante futuro é função das cargas de instantes anteriores — ou *multivariados* — em que a carga esperada no futuro é função das cargas passadas e de outras variáveis exógenas, principalmente as relacionadas ao clima.¹

Os modelos multivariados mais comuns são os modelos de regressão linear e aqueles que decompõem a carga em um componente básico linear e um componente não-linear dependente de variáveis meteorológicas. Para a previsão desse componente não-linear, grande parte da literatura aborda o uso de técnicas de Inteligência Artificial [5], principalmente as redes neurais artificiais, devido à sua flexibilidade e à sua capacidade de reproduzir relações não-lineares.

Este trabalho tem como objetivo estudar combinações de modelos univariados para a previsão de cargas elétricas a curto prazo, com base nas considerações de Taylor *et al* [6] de que, embora a carga seja afetada a curto prazo por condições meteorológicas, efeitos sazonais e eventos especiais, essas variações são críticas apenas nas previsões com horizontes de mais de um dia a frente e, portanto, um modelo univariado é suficiente para modelar o comportamento da série de cargas a curto prazo.

No Capítulo 2, o problema da previsão de cargas elétricas é brevemente abordado, para uma melhor compreensão sobre o assunto. O Capítulo 3 é dedicado aos materiais e métodos utilizados no trabalho, e os resultados obtidos são apresentados no Capítulo 4. Por fim, as conclusões e discussões acerca dos resultados e a sugestão para possíveis estudos futuros estão no Capítulo 5.

¹Pesquisadores defendem o uso do termo *multivariado* para modelos com múltiplas saídas. Os modelos com múltiplas entradas seriam então modelos *multivariáveis*. Entretanto, quando usamos esse termo, neste trabalho, estamos nos referindo a modelos com múltiplas entradas, mas também com múltiplas saídas, e assim optamos, por simplicidade, por adotar o termo mais comum. O mesmo serve para o termo *univariado*.

2 CARACTERIZAÇÃO DO PROBLEMA

2.1 A previsão de cargas elétricas

Os sistemas de energia de diversos países do mundo vêm sendo privatizados e desregulamentados desde o início da década de 1990. Com isso, criou-se um mercado de energia altamente competitivo, com preços de energia elétrica fixados de acordo com o interesse das empresas de energia. O padrão de desempenho, então, aumentou, e os consumidores se tornaram mais exigentes [4].

Antes das privatizações, as previsões de cargas já eram necessárias para a tomada de decisões relacionadas à programação de partida e parada dos geradores e ao despacho de cargas [7]. Após as privatizações, contudo, as mudanças no cenário do abastecimento de energia levaram à necessidade de previsões de cargas mais acuradas, uma vez que os erros de previsão passaram a resultar em perdas financeiras para as empresas. Com a presença de um mercado de energia, o consumidor, se estiver insatisfeito, tem a opção de mudar de fornecedor, reduzir seu consumo, ou produzir sua própria energia.

As previsões de carga elétrica também são necessárias como base para a atribuição ao preço da energia. O nível de acurácia desejado nessa previsão é obtido quando os custos de aperfeiçoamento do sistema são maiores que os custos dos erros de previsão, ou quando é atingido o ponto de equilíbrio entre os custos dos erros e os custos da própria previsão [7].

Artigos já publicados indicam que é justificável o investimento na pesquisa em previsão de cargas elétricas, uma vez que os custos do erro de previsão, em geral, são altos (para detalhes ver [4],[8], [9], [10]). Por outro lado, estes custos são variáveis, dependendo das características da concessionária. Hobbs *et al* [9], por exemplo, estima que uma redução de 1% nos erros de previsão poderia representar, para concessionárias de tamanho médio, uma economia anual de 0,6 a 1,6 milhões de dólares. Este artigo e outros são discutidos em uma revisão sobre o tema em [2].

De acordo com Bunn e Farmer [7], a qualidade do controle do sistema de energia e a

economia nos custos operacionais são altamente sensíveis aos erros de previsão. Algumas consequências dos erros de previsão são [9]:

1. Se a carga observada for menor que a prevista (*overforecasts*):
 - (a) unidades podem ter sido iniciadas sem necessidade, aumentando custos de produção;
 - (b) energia pode ter sido comprada desnecessariamente, a preços altos;
 - (c) água pode ter sido desperdiçada na produção de energia hidroelétrica não utilizada.

2. Se a carga observada for maior que a prevista (*underforecasts*):
 - (a) podem faltar recursos para atender às necessidades de segurança;
 - (b) pode ser necessário comprar energia no mercado, ou apelar para unidades geradoras antieconômicas;
 - (c) energia produzida pode ter tido seu preço fixado em um nível muito baixo.

A privatização dos sistemas de energia levou à ênfase nas previsões de perfis de carga, para fixação do preço da energia a cada momento do dia. A previsão de perfil nada mais é que a previsão das cargas das 24 horas do dia seguinte, usando um mesmo modelo para todas as horas, ou modelos diferentes para cada hora do dia. As dificuldades encontradas nesse tipo de previsão são a autocorrelação das séries de cargas – previsão da carga da hora h do dia seguinte é baseada na carga da hora $h-1$ também desconhecida – e a alta dimensionalidade dos problemas, já que os perfis são vetores de 24 dimensões.

As pesquisas na área de previsão de cargas, então, visam determinar modelos de previsão que resultem em erros os menores possíveis. É intuitivo crer que os melhores modelos para previsão de cargas são aqueles que consideram os fatores que afetam a demanda de energia, geralmente fatores meteorológicos, embora isso nem sempre seja verdade. Antes de avaliarmos a importância desses fatores, devemos considerar o horizonte de previsão necessário para atender o propósito do estudo.

Diversos fatores meteorológicos podem gerar mudança no comportamento das cargas, como temperatura, umidade e velocidade do ar, presença de nuvens, luminosidade do dia, precipitação. Davies [11] foi um dos primeiros a constatar que os dois primeiros

fatores controlam a demanda de carga elétrica no aquecimento, e os demais são utilizados para estimar o nível de iluminação da luz do dia. Em outra situação, Heinemann *et al* [12] estudaram a relação entre a demanda de carga elétrica e o clima de Newark, Nova Jérsei, EUA, durante os meses de verão. Eles mostram que o aumento da temperatura resultou no aumento do uso de aparelhos de ar-condicionado e outros equipamentos. Esse comportamento, além de gerar o crescimento da carga média demandada no verão, gerou também a variabilidade de cargas de acordo com a resposta dos usuários às mudanças climáticas. Essa resposta, de acordo com [11], demora cerca de um dia para ser notada. Asbury [13] afirma que o efeito dos fatores meteorológicos em sistemas de carga depende das características da demanda de energia e do clima, e que um modelo ajustado para um determinado sistema não deve ser usado para prever a demanda em sistemas com características diferentes.

Atualmente, existem diversos artigos já publicados sobre previsão de cargas, apresentando diferentes métodos e abordagens. Na seção seguinte, tratamos de alguns modelos de previsão mais utilizados, para melhor compreensão sobre o tema.

2.2 Modelos usuais de previsão de cargas elétricas

Os modelos usados para previsão de cargas podem ser univariados ou multivariados. No primeiro caso, a carga é modelada como uma função de seus valores observados no passado e no segundo, como função de variáveis exógenas, principalmente fatores meteorológicos [5].

Como exemplos de modelos multivariados usados na previsão de cargas podemos citar funções de transferência de Box e Jenkins [14], [15], modelos ARMAX [16], [17] e ARIMAX [18], modelos de regressão linear [19], [20] e, principalmente, modelos que decompõem a carga em um componente básico linear (B_t), que deve refletir o comportamento normal da série, e um componente não-linear relacionado ao clima (W_t), que modela a influência das variáveis meteorológicas nas mudanças ocorridas na carga [7], [21]. Esse modelo tem como forma geral:

$$L_t = B_t + W_t + \varepsilon_t$$

Como W_t é uma função não linear, diversas técnicas de Inteligência Artificial (IA) têm sido aplicadas para a estimação desse componente. Artigos publicados apresentam

modelos baseados em lógica *fuzzy* [22], [23], sistemas especialistas e, em sua maioria, redes neurais artificiais (RNA) para resolução desse problema [5], [24], [25]. A vantagem das RNAs é que são modelos não-paramétricos, ideais para reproduzir relações não-lineares, que permitem também saída multivariada. Por outro lado, existem duas dificuldades que podem ocorrer durante a implementação: o *super-ajuste* — modelo treinado em excesso — e a *superparametrização* do modelo — modelo muito complexo em relação ao problema.

Devido ao foco deste estudo, modelos univariados, consideramos W_t nulo, sendo então a carga modelada apenas em função de seus valores passados e um erro aleatório. Encontramos como modelos univariados já aplicados modelos auto-regressivos (AR) multiplicativos [26], modelos ARMA [27], modelos baseados em Filtro de Kalman [28], [1], amortecimento exponencial baseado em decomposição por valores singulares (SVD) [29], modelos de regressão de vetor suporte (SVR) [30], [31], e amortecimentos exponenciais com dupla e tripla sazonalidade[32], [33].

Os perfis de carga elétrica podem ser previstos por conjuntos de modelos de saída univariada, ou por um modelo único de saída multivariada – modelo não-linear com 24 saídas, que utiliza o perfil de cargas do dia anterior, geralmente realizado por RNAs. Existem duas maneiras de fazer as previsões de perfis, ilustradas na Figura 2.1:

1. Previsão iterativa ou sequencial: um valor de carga, horário, previsto de cada vez, que é agregado à série de valores de entrada do modelo, de forma que as previsões futuras sejam baseadas nas anteriores.
2. Previsão por modelos em paralelo: perfil modelado por um conjunto de modelos similares, um para a carga de cada hora.

Os modelos em paralelo recebem uma atenção especial neste trabalho. Sua desvantagem é que consideram cada hora do dia separadamente e, portanto, não exploram a autocorrelação entre as cargas de diferentes horas. O uso dos 24 modelos em paralelo gera previsões para as cargas de segunda-feira, por exemplo, baseadas nas cargas de domingo. Devido a isso, alguns autores são a favor do uso de 168 modelos para a previsão de perfis de cargas que, por outro lado, tem como desvantagem o grande número de parâmetros a serem estimados e previsões baseadas em dados antigos, de uma semana antes. A facilidade e o baixo custo na implementação devem sempre ser considerados na escolha do modelo a ser usado.

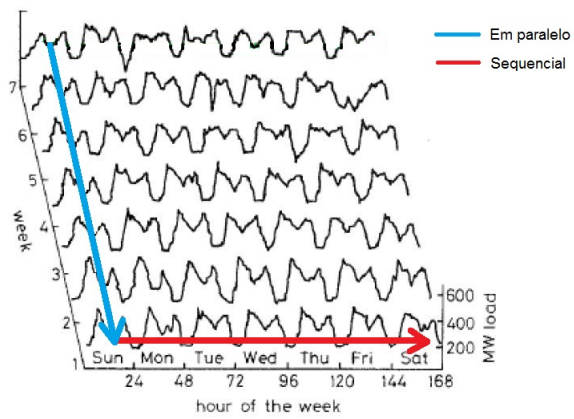


Figura 2.1: Previsões de perfis de cargas elétricas (adaptado de [1]).

Neste estudo, fazemos previsões de uma série de cargas real por modelos univariados, sequencial e em paralelo, e por combinações desses modelos. No Capítulo 3, a seguir, detalhamos a metodologia utilizada.

3 MATERIAIS E MÉTODOS

3.1 Materiais

Para estudarmos a eficiência de modelos univariados para a previsão de perfis de cargas elétricas, utilizamos uma série temporal (ST) de cargas diárias, contendo 17472 observações referentes às cargas diárias consumidas no Rio de Janeiro nos anos de 1996 e 1997, disponibilizada por uma antiga concessionária de energia da cidade.

A série de cargas passou por uma etapa de pré-tratamento, para ajustes de feriados e dados faltantes. Ela não apresentou valores faltantes, nem valores discrepantes, e os dados de feriados foram substituídos pela média das cargas do dia equivalente nas semanas anterior e posterior. Os ajustes para os dados de feriado são importantes uma vez que a presença de dados não usuais observados em feriados afeta o comportamento padrão da série de cargas, como mostrado na Figura 3.1. As cargas de feriados e outros dias especiais (feriados, férias e greves, por exemplo) são previstas por modelos especializados, como os apresentados em [34] e [35]. As técnicas utilizadas para preencher valores faltantes, ou para substituir valores discrepantes por outros mais razoáveis, são chamadas de técnicas de *imputação* de valores.

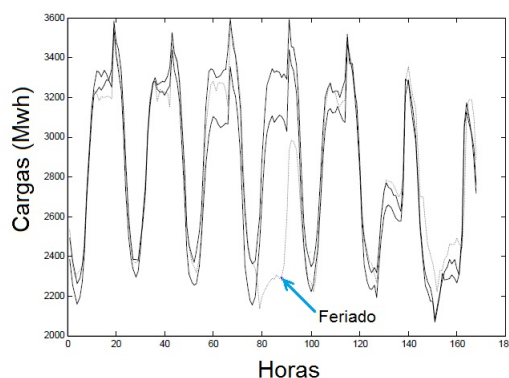


Figura 3.1: Efeito do feriado no comportamento padrão da série de cargas (adaptado de [2]).

Separamos essa série em três partes, de modo que as observações das primeiras 54 semanas (amostra de ajuste), foram usadas para o ajuste dos parâmetros dos modelos de previsão. Ajustados os modelos, calculamos as medidas de erro no intervalo incluindo as

semanas 55 a 84 (amostra de teste I), para escolhermos os melhores. Por fim, combinamos estes modelos escolhidos, calculamos os erros de previsão finais para o último intervalo de dados, semanas 85 a 104 (amostra de teste II), para compararmos a acurácia das combinações com resultados já publicados para o mesmo problema. As Figuras 3.2 e 3.3 mostram como as amostras foram utilizadas neste estudo.

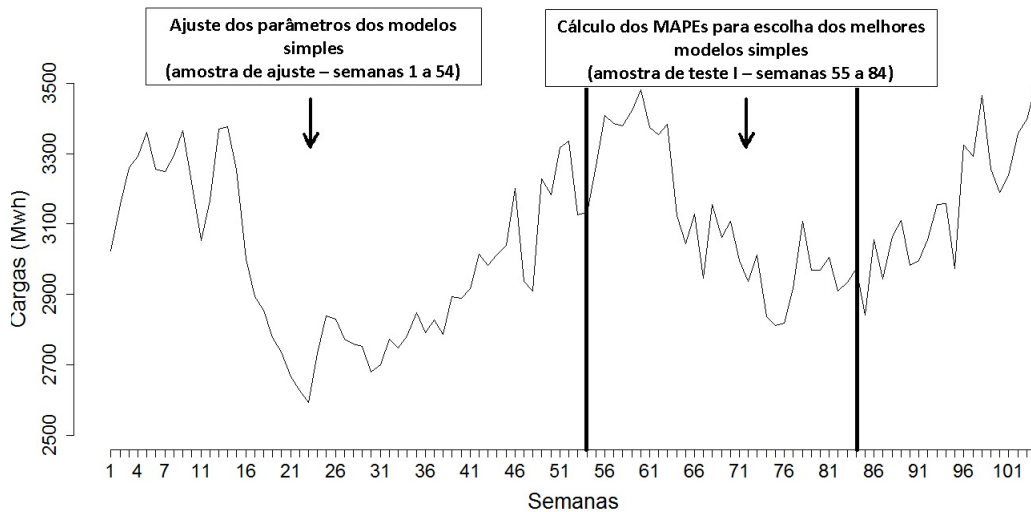


Figura 3.2: Utilização das amostras antes das combinações.

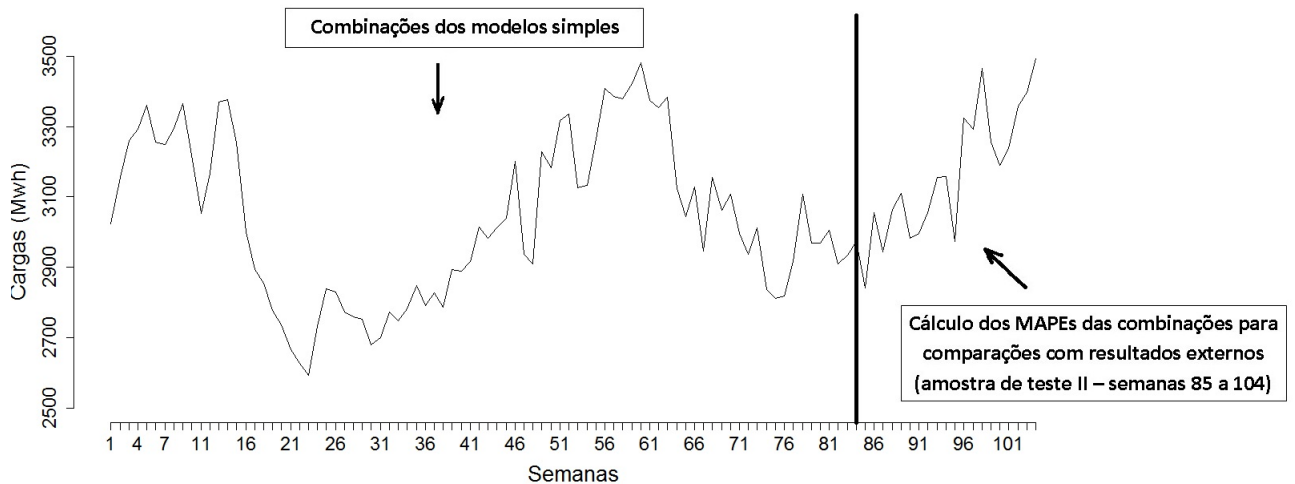


Figura 3.3: Utilização das amostras após combinações.

Séries de cargas elétricas têm padrões de comportamento complexo, com três sazonalidades: uma diária, uma semanal e uma anual. Na Figura 3.4, que mostra as cargas médias semanais nos dois anos, é possível verificar a presença de sazonalidade anual na série: as cargas são mais altas nos primeiros meses do ano, devido ao uso de

aparelhos de ar-condicionado, e mais baixas no meio do ano. Para a estimação do valor inicial do fator sazonal anual seriam necessários dados dos primeiros anos da série, que seriam descartados para previsões futuras, restando, assim, poucas observações para as etapas de ajuste, escolha e validação do modelo. Dessa forma, optamos por desconsiderar a sazonalidade anual da série.

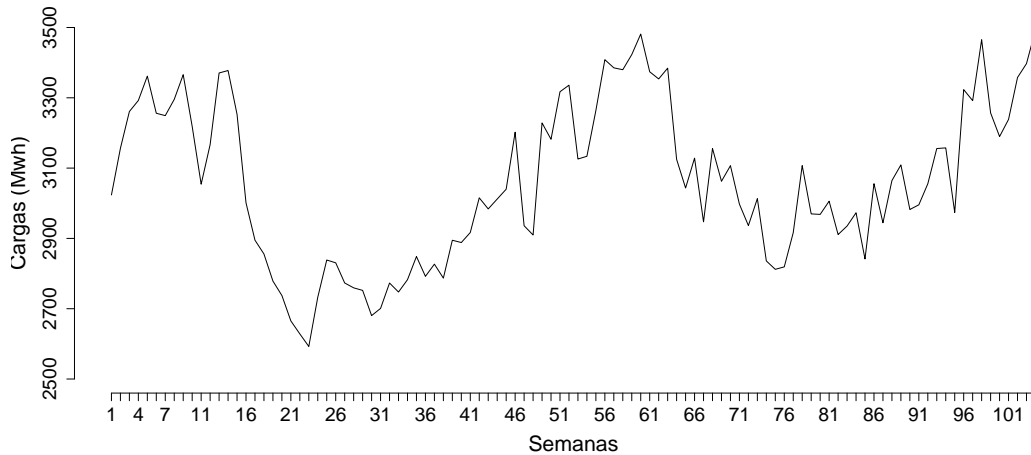


Figura 3.4: Cargas médias semanais nos anos de 1996 e 1997.

A Figura 3.5 apresenta dados de uma quinzena de verão e uma de inverno. Em ambas quinzenas identificamos um certo padrão no comportamento das cargas: mais elevado e aproximado entre os dias de semana e mais baixo em fins de semana, caracterizando a sazonalidade semanal da série. Como as cargas nos meses frios são menos sensíveis a alterações no clima, o consumo de energia no inverno tende a ser mais estável do que no verão.

No verão, existem dois picos de cargas durante os dias de semana, um às 15 horas e outro às 23 horas, geralmente relacionados ao uso de aparelhos de condicionamento de ar e chuveiros elétricos. Por outro lado, no inverno existe apenas um pico, às 19 horas, que pode ser explicado pelo fim do expediente de escolas, comércios e repartições públicas, por exemplo. Esse comportamento indica o ciclo de sazonalidade diária da série de cargas, mostrado na Figura 3.6, que apresenta as cargas de uma quarta-feira de verão e uma de inverno.

A Seção 3.2 a seguir, traz conceitos importantes sobre séries temporais, incluindo o de sazonalidade, além dos modelos e métodos para previsão de perfis de cargas usados neste estudo.

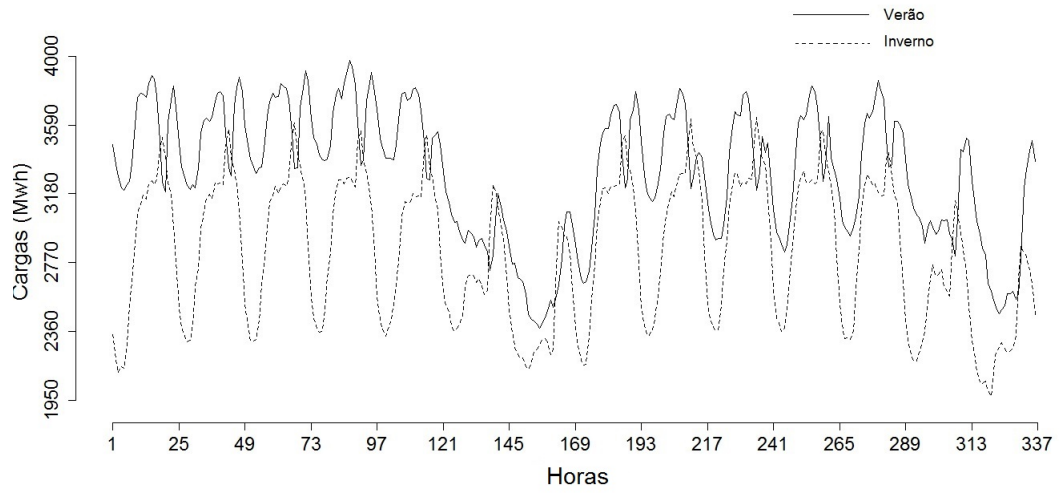


Figura 3.5: Perfis de cargas de quinzenas de verão e inverno.

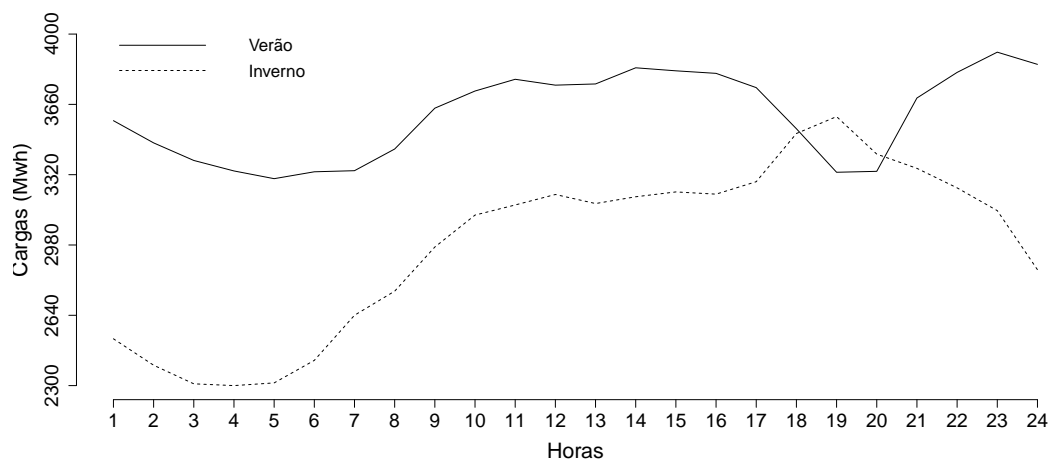


Figura 3.6: Perfis de cargas de dias de verão e inverno.

Implementamos esses modelos através do software estatístico livre R [36], versão 3.0.3. Para tal, foi necessária a instalação do pacote *forecast*, que contém métodos e ferramentas para realização e análise de previsão de séries temporais univariadas [37].

As funções presentes nesse pacote foram utilizadas para a previsão da série de cargas por modelos ARIMA (Subseção 3.2.8). Todos os demais métodos foram implementados livremente, sem o uso de funções prontas. Durante as simulações, usamos a função *optim()* para determinar os parâmetros dos métodos de previsão e os pesos dos previsores nas combinações. A Subseção 3.2.11 é dedicada a uma breve explicação sobre o funcionamento dessa função.

3.2 Métodos

No decorrer deste Capítulo, apresentamos uma base teórica sobre os métodos de previsão de STs que usamos neste estudo, para melhor compreensão do assunto.

3.2.1 Introdução

Séries temporais são coleções de observações feitas sequencialmente ao longo do tempo [38], e podem ser contínuas ou discretas. As primeiras são caracterizadas pela medição, em tempo contínuo, da(s) variável(is) de interesse. STs discretas são aquelas cuja medição é feita em tempo discreto, geralmente em instantes equidistantes. Existem ainda as STs discretizadas, em que o agrupamento de valores observados em intervalos de tempo de séries contínuas resulta em séries discretas. As séries podem ser, também, univariadas, em que apenas uma variável é observada, ou multivariadas, em que k variáveis são medidas no tempo, simultaneamente.

É necessário identificar as características da ST para o uso de técnicas adequadas para sua análise e previsão. A maior parte das teorias presentes na Estatística são aplicadas a amostras aleatórias de observações independentes; entretanto, as observações de uma ST não são independentes, e isso deve ser levado em consideração na hora de analisá-la.

Como as observações de uma ST são dependentes entre si, dizemos que elas são *autocorrelacionadas*. Os conceitos de *autocovariância* e *autocorrelação* têm a mesma interpretação dos conceitos de covariância e correlação. A diferença está no fato de que, quando utilizamos os coeficientes de covariância e correlação, estamos interessados

em verificar se há relação entre duas variáveis distintas, e o grau dessa relação; através das funções de autocovariância e autocorrelação pretendemos estudar o relacionamento entre duas observações de uma mesma variável feitas em instantes distintos. Logo, a formulação dessas funções é análoga à dos coeficientes usuais de covariância e correlação. Para detalhes, ver [38], [39].

O *nível médio* de uma série é um valor μ_t em torno do qual ela oscila. Quando existe uma mudança a longo prazo no nível da série, dizemos que essa série apresenta uma *tendência*.

A tendência pode ocorrer de forma linear, com inclinação positiva ou negativa ou polinomial — geralmente quadrática —, exponencial ou logística. Quando o método de diferenciação é utilizado, a tendência da série é eliminada, tornando o seu nível constante. A tendência não necessariamente deve ser retirada dos dados; ela pode ser estimada de algumas maneiras:

1. através do ajuste de uma função polinomial ao longo tempo,
2. a partir de um componente de tendência incluído no modelo, estimado por regressão, método das Médias Móveis, amortecimento de Holt ou Filtro de Kalman, por exemplo.

Embora seja intuitivo o conceito de *sazonalidade*, formalmente é difícil defini-lo. A sazonalidade é entendida como o padrão, regular e periódico, de um determinado fenômeno — no nosso caso, a realização de uma ST. Esse padrão geralmente é devido a fatores físicos externos à série, e pode ser medido de forma explícita, ou ser removido dos dados.

O componente sazonal de um modelo pode ser *aditivo* ou *multiplicativo*, e pode ser modelado através de *fatores sazonais* ou de *funções trigonométricas*. Os fatores sazonais são os mais usados na prática, e consistem em séries de S valores observados, um para cada período do ciclo sazonal. Um modelo sazonal, então, tem três componentes: um nível médio μ_t , que pode apresentar tendência, um fator sazonal e um erro aleatório ε_t . Supondo um fator sazonal mensal, denotado por $\rho^{m(t)}$, temos que os componentes mencionados podem ser combinados de duas formas, dado que Z_t é o valor observado da variável Z no instante t :

$$Z_t = \mu_t + \rho^{m(t)} + \varepsilon_t \quad (\text{Modelo aditivo}) \quad (3.1)$$

$$Z_t = \mu_t \rho^{m(t)} + \varepsilon_t \quad (\text{Modelo multiplicativo}) \quad (3.2)$$

Os componentes de tendência e sazonalidade de um modelo são muito relacionados e a influência da tendência sobre o componente sazonal pode ser muito forte por duas razões [39]: os métodos de estimação da sazonalidade podem ser afetados se não considerarmos a tendência, e a especificação do componente sazonal depende da especificação do componente de tendência.

No caso de modelos sazonais aditivos, as flutuações sazonais desconsideram a média global da série, sendo aproximadamente de variância constante. Essas flutuações sazonais, no caso dos modelos multiplicativos, levam em consideração a média global, e por isso possuem amplitudes diferentes ao longo da série.

Dada uma ST, e determinado um modelo que se ajuste bem aos seus dados, podemos estar interessados em fazer previsões dos valores em instantes futuros da ST, para fins de planejamento e controle. A correlação entre as observações permite que valores futuros, não observados, da série sejam previstos a partir de valores passados. Assim, a previsão de um valor futuro é dada pelo valor esperado condicional da série no instante desejado, dados os valores passados da série. Para simplificar, dada uma série Z , um instante t , a previsão k passos a frente é:

$$\hat{Z}_{t+k} = E(Z_{t+k} | Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_1)$$

Na prática, os métodos de previsão variam bastante, podendo ser simples e intuitivos ou mais complexos. Uma análise descritiva dos dados da série é muito importante para a escolha do modelo de previsão mais adequado. A definição da adequação desse modelo, contudo, será feita a partir da avaliação dos erros de previsão resultantes. As medidas de erro mais utilizadas são apresentadas na Subseção 3.2.10.

As subseções seguintes são dedicadas aos métodos de previsão de séries temporais que utilizamos neste estudo. Um conteúdo completo sobre este tema pode ser encontrado em [38], [39], [40].

3.2.2 Método Naive

O método conhecido como “*Naive*” (do inglês “ingênuo”) é usado para estimar o nível de modelos não-sazonais sem tendência, que podem ser escritos como:

$$Z_t = \mu + \varepsilon_t$$

Para esses modelos, a previsão dos valores futuros da série é a estimativa do nível, que chamamos de a :

$$\hat{Z}_{t+k} = \hat{\mu}_t = \hat{a}_t$$

Há várias maneiras de obter essa estimativa de a_t . No método *Naive*, método mais simples, que agrega o mínimo possível de informação à previsão, a estimativa para o instante $t + 1$ é dada, no instante t , pelo último valor observado, Z_t :

$$\hat{Z}_{t+1|t} = Z_t$$

Pela simplicidade, esse método geralmente apresenta altos erros de previsão. Neste estudo, as previsões pelo método *Naive* são usadas como referência para a qualidade das previsões. Um previsor que apresente um erro maior que o erro resultante do *Naive* não deve ser considerado como uma boa opção.

3.2.3 Método das Médias Móveis

O método das *Médias Móveis* (MM) é usado para estimar a tendência de modelos não-sazonais constantes, a partir do cálculo da média aritmética dos n valores mais recentes da série. Essa estimativa não considera observações mais antigas, o que é plausível, levando em conta o fato de que o nível varia suavemente ao longo da série.

A previsão para o instante $t + 1$ é a média dos n valores observados até o instante t , para o instante $t + 2$ calculamos a média dos n valores observados até o instante $t + 1$, e assim por diante:

$$\hat{Z}_{t+1} = \frac{Z_t + Z_{t-1} + \dots + Z_{t-n+1}}{n} = M_t$$

Para determinarmos o parâmetro N do método — o tamanho da janela da média móvel —, é necessário testar diversos valores diferentes, em uma amostra selecionada para ajuste do modelo; o valor do parâmetro que resultar em menor erro de previsão é o escolhido.

O método das Médias Móveis pode ser usado também para a estimação da tendência de séries não-sazonais lineares ou quadráticas, da forma:

$$Z_t = a + b_t + \varepsilon_t \tag{3.3}$$

$$Z_t = a + bt + ct^2 + \varepsilon_t \quad (3.4)$$

As previsões para 3.3 e 3.4 são dadas, respectivamente por:

$$\hat{Z}_{t+k} = \hat{\mu}_t = \hat{a} + \hat{b}t$$

$$\hat{Z}_{t+k} = \hat{\mu}_t = \hat{a} + \hat{b}t + \hat{c}t^2$$

Para esses modelos, é necessário estimar mais de um parâmetro no método. Dessa forma, para a série com tendência linear, o método das Médias Móveis torna-se, então, “Médias Móveis Duplas”(MMD), em que uma média móvel de tamanho n é aplicada à série de médias móveis de tamanho n aplicadas à série original, e os parâmetros a e b são atualizados iterativamente, de forma [39], [38]:

$$M_t^{[2]} = \frac{M_t + M_{t-1} + \dots + M_{t-n+1}}{n}$$

$$\hat{a}_t \approx 2M_t - M_t^{[2]}$$

$$\hat{b}_t \approx \frac{2}{n-1}[M_t - M_t^{[2]}]$$

Para séries com tendência quadrática, o raciocínio é análogo, dando origem às “Médias Móveis Triplas”[39], [38]. Neste estudo, usamos apenas o método das Médias Móveis Simples e o método de Médias Móveis Duplas.

3.2.4 *Amortecimento Exponencial Simples*

Para modelos não-sazonais constantes, a ideia de um método de amortecimento é de que a previsão seja dada pela estimativa mais recente do nível, calculada através de uma forma de média ponderada, em que as observações mais antigas recebem pesos menores que as observações mais atuais.

Existem diversos métodos de amortecimento exponencial, usados de acordo com o modelo da série de interesse. Alguns exemplos são os amortecimentos de Brown e Holt (maiores detalhes ver [38], [39]).

Tomando essa proposta de médias ponderadas, a previsão do nível médio poderia ser dada por várias constantes α_i , multiplicando as observações da série:

$$\hat{a}_t = M_t = \alpha_1 Z_t + \alpha_2 Z_{t-1} + \alpha_3 Z_{t-2} + \dots, \quad \text{com } \alpha_1 > \alpha_2 > \alpha_3 > \dots$$

Esse método de previsão seria muito custoso, uma vez que teríamos que determinar todas as constantes α_i . O *Amortecimento Exponencial Simples*, ou AES, utiliza apenas uma constante α , e os pesos são definidos como potências sucessivas dessa constante:

$$\hat{Z}_{t+1} = M_t = \alpha Z_t + \alpha(1 - \alpha)Z_{t-1} + \cdots + \alpha(1 - \alpha)^k Z_{t-k} + \cdots + \alpha(1 - \alpha)^t Z_0$$

Para que os pesos decaiam à medida em que as observações se afastam do instante da previsão, o valor de α deve estar entre 0 e 1. Assim, os pesos $(1 - \alpha)^k$ tendem a 0 quando k tende ao infinito. Quanto mais próxima de 1 for a constante, mais rápido é o decaimento da sequência, fazendo com que as observações mais novas recebam muito mais peso que as mais antigas. Se o valor de α for pequeno, a diferença entre os pesos atribuídos às observações mais atuais e mais antigas é pequena, fazendo com que o AES seja mais estável. Quando α é igual a 1, temos o método *Naive*. A determinação do valor ideal da constante é feita experimentalmente, em uma amostra da série separada para teste. Neste trabalho, a escolha de α foi feita através da função *optim()* do software R, que será detalhada na Subseção 3.2.11.

Mesmo com uma constante só, ainda é trabalhoso calcular a estimativa da média usando todos os valores da série. Existe uma maneira mais simples de calculá-la, através de recursividade. Consideramos que a média é uma combinação linear entre a média do instante anterior e a observação mais atual, na forma:

$$M_t = \alpha Z_t + (1 - \alpha)M_{t-1} \quad (3.5)$$

Mas, dado que $\hat{Z}_{t+1} = M_t$ e $\hat{Z}_t = M_{t-1}$, podemos reescrever 3.5 como:

$$\hat{Z}_{t+1} = \alpha Z_t + (1 - \alpha)\hat{Z}_t \quad (3.6)$$

Podemos demonstrar que, de 3.6, a expressão final para a previsão da série no instante t é:

$$\hat{Z}_{t+1} = \hat{Z}_t + \alpha(Z_t - \hat{Z}_t) \quad (3.7)$$

A equação dada por 3.7 pode ser entendida como uma correção do erro de previsão, muito útil quando o nível médio oscila ao longo da série. Essa correção é feita automaticamente: se o valor previsto para o instante t for grande, a parcela do erro $(Z_t - \hat{Z}_t)$ vai ser positiva, e a previsão para o instante $t + 1$ será aumentada, e vice-versa.

3.2.5 Amortecimento Exponencial de Holt

O Amortecimento Exponencial de Holt é usado para estimar a tendência de modelos não-sazonais lineares, da forma em 3.3. Essa estimativa é baseada em duas equações recursivas de atualização: uma para o parâmetro a , nível da série, e uma para a declividade b :

$$\hat{a}_t = \alpha Z_t + (1 - \alpha)(\hat{a}_{t-1} + \hat{b}_{t-1})$$

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta)\hat{b}_{t-1}$$

A previsão é da forma:

$$\hat{Z}_{t+1} = \hat{a}_t + \hat{b}_t$$

As duas equações para atualização dos parâmetros aplicam a ideia básica dos métodos de amortecimento exponencial, que obtêm estimativas de forma recursiva:

$$\text{estimativa atual} = \xi(\text{estimativa mais recente}) + (1 - \xi)(\text{estimativa antiga})$$

A estimativa inicial do nível a_1 pode ser a primeira observação, Z_1 . Para a declividade, existem vários tipos de inicialização. Um deles é a diferença entre as duas primeiras observações Z_1 e Z_2 . As constantes de amortecimento α e β podem ser determinadas, por exemplo, por uma busca em grade, em que criamos um vetor de possíveis valores de α , simulamos as previsões com cada um deles, comparamos os erros e o valor escolhido para α é aquele que resulte em menor erro de previsão [41]. Neste estudo, usamos novamente a função *optim()* do R.

3.2.6 Método de Holt-Winters

O método de Holt-Winters (HW) é usado para modelar séries com tendência linear e sazonalidade multiplicativa. O modelo básico, considerando por exemplo uma sazonalidade mensal, é dado por:

$$Z_t = (a + bt)\rho^{m(t)} + \varepsilon_t$$

A previsão para esse modelo depende de três equações recursivas de atualização: uma para o nível a da série, uma para a tendência b , e uma para os fatores sazonais $\rho^{m(t)}$. Ela é da forma:

$$\hat{Z}_{t+1|t} = (\hat{a}_t + \hat{b}_t)\rho_t^{m(t+1)}$$

Com:

$$\hat{a}_t = \alpha \left[\frac{Z_t}{\hat{\rho}_{t-1}^{m(t)}} \right] + (1 - \alpha)(\hat{a}_{t-1} + \hat{b}_{t-1})$$

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta)\hat{b}_{t-1}$$

$$\hat{\rho}_t^{m(t)} = \gamma \left[\frac{Z_t}{\hat{a}_t} \right] + (1 - \gamma)\hat{\rho}_{t-1}^{m(t)}$$

Para os fatores sazonais:

$$\hat{\rho}_t^{m(j)} = \hat{\rho}_{t-1}^{m(j)} \quad \forall j \neq t, \quad j = 1, 2, \dots, S$$

Isso indica que os fatores sazonais são atualizados uma vez por ano, no mês correspondente, e no resto do ano permanecem os mesmos. É necessário reajustar todos os fatores a cada instante, através de uma normalização da série de fatores sazonais, obedecendo à seguinte restrição:

$$\sum_{i=1}^S \rho_t^{m(i)} = S$$

Temos que S é o número de fatores sazonais do modelo e, denotando os fatores sazonais antes da normalização por $\hat{\rho}_t^{m(j)*}$, a normalização é:

$$\hat{\rho}_t^{m(j)} = \hat{\rho}_t^{m(j)*} * \frac{S}{\sum_{i=1}^S \rho_t^{m(i)*}}, \quad j = 1, 2, \dots, S$$

Essa é apenas uma introdução sobre o método de Holt-Winters. Para maiores detalhes, ver [38], [39].

3.2.7 Método de Holt-Winters-Taylor

Uma série de cargas elétricas apresenta três sazonalidades: uma diária, uma semanal e uma anual. Para modelar esse comportamento, o amortecimento de Holt-Winters requer uma pequena extensão de sua formulação, incluindo os componentes sazonais capazes de abranger todo o padrão da série. No caso deste estudo, devido à não disponibilidade de dados, optamos por desconsiderar a sazonalidade anual da série de cargas.

Taylor [33] propõe um modelo de amortecimento exponencial para dupla sazonalidade aditiva, com:

$$l_t = \lambda(Z_t - d_{t-s_1} - w_{t-s_2}) + (1 - \lambda)l_{t-1} \quad (3.8)$$

$$d_t = \delta(Z_t - l_{t-1} - w_{t-s_2}) + (1 - \delta)d_{t-s_1} \quad (3.9)$$

$$w_t = \omega(Z_t - l_{t-1} - d_{t-s_1}) + (1 - \omega)w_{t-s_2} \quad (3.10)$$

Em que s_1 e s_2 são os períodos sazonais diário e semanal, respectivamente, l_t é o nível da série amortecido, d_t e w_t são os fatores sazonais diários e semanais, e λ , δ e ω são as constantes de amortecimento do método. Combinando 3.8, 3.9 e 3.10, temos que a previsão k passos a frente pelo *método de Holt-Winters-Taylor* é dada por:

$$\hat{Z}_{t+k} = l_t + d_{t-s_1+k} + w_{t-s_2+k} + \phi^k(z_t - (l_{t-1} + d_{t-s_1} + w_{t-s_2})) \quad (3.11)$$

A equação 3.11 só é válida se $k \leq s_1$. O termo envolvendo o parâmetro ϕ em 3.11 é um fator de correção para a autocorrelação de primeira ordem da série de erros, e tanto ele quanto os demais parâmetros são estimados por otimização. Os valores iniciais do nível e dos fatores sazonais são estimados pela média das observações iniciais da série.

3.2.8 Modelos ARIMA

Os *modelos ARIMA*, ou *modelos de Box e Jenkins*, são métodos de previsão de base estatística, e consideram que uma série temporal Z é uma realização de um processo estocástico, sendo atribuída à série uma determinada distribuição de probabilidades. Por ser uma única realização do processo, uma série é uma amostra com apenas um elemento, que é o conjunto de observações que a compõe.

Formalmente, um processo estocástico é uma família de variáveis aleatórias $Z = Z_t, t \in T$. Os valores Z_t são chamados *estados* do processo, e o conjunto dessas observações é conhecido como *espaço de estados*. Se T for finito ou enumerável, o processo estocástico é dito ser *a tempo discreto*; se $T \in \mathbb{R}$, é dito ser *a tempo contínuo*.

Dado que uma ST é uma realização de um processo estocástico, cada observação possui uma parcela que representa um erro aleatório ε_t , com as seguintes propriedades:

1. ε_t são independentes e identicamente distribuídos, com distribuição normal,
2. $E(\varepsilon_t) = 0$,

$$3. \text{Var}(\varepsilon_t) = \sigma^2.$$

Informalmente, podemos dizer que um processo estocástico é *estacionário* se ele varia em torno de seu nível médio, sem alteração de suas propriedades estatísticas. Se o processo for estacionário e gaussiano, então ele é dito ser estritamente estacionário. Modelos de previsão deste tipo são conhecidos como *modelos ARMA*, baseados em filtros lineares. Se o processo é não estacionário, os modelos ARIMA, também baseados em filtros lineares, podem ser utilizados. Neste caso, o processo precisa passar pela etapa de diferenciação antes do ajuste do modelo.

Uma série que não apresenta autocorrelação entre observações sucessivas, sob o ponto de vista estatístico, não contém informação útil, já que a informação de valores passados não implicará em possível conhecimento de valores futuros. Neste caso, dizemos que a série é um *ruído branco*. Os modelos ARIMA traduzem qualquer série autocorrelacionada como função de uma série de choques independentes a_t , identicamente distribuídos, com distribuição normal com média 0 e variância σ^2 , sendo a_t um ruído branco.

Um *filtro linear* é uma função de transferência definida por uma combinação linear infinita dos valores da ST. Em outras palavras, o filtro linear recebe uma série de choques aleatórios a_t , transformando-os na série Z_t :

$$a_t \longrightarrow [FT] \longrightarrow Z_t$$

Esse filtro pode ser apresentado na forma invertida, em que retira a autocorrelação de uma série Z_t , retornando uma série sem informação estatística a_t :

$$a_t \longleftarrow [FT'] \longleftarrow Z_t$$

Vamos considerar um *operador de retardo* B que, quando aplicado à série, produz o valor da série no instante anterior ($B^k Z_t = Z_{t-k}$). Um filtro linear pode ser definido, em termos de B , como, dado que o valor centrado de Z é $\bar{Z}_t = Z_t - \mu$:

$$\bar{Z}_t = (1 + \psi_1 B + \psi_2 B^2 + \dots) a_t = \Psi(B) a_t \quad (3.12)$$

O filtro linear invertido, em termos de B , é da forma:

$$\Psi^{-1}(B) \bar{Z}_t = \Pi(B) \bar{Z}_t = a_t \quad (3.13)$$

Um processo estacionário é *invertível* se a forma em 3.12 puder ser escrita como em 3.13. As condições de estacionariedade e invertibilidade dos filtros lineares são dadas por:

1. o filtro é *estacionário* se: $\sum_j |\Psi_j| < \infty$;
2. o filtro é *invertível* se: $\sum_j |\Pi_j| < \infty$

Existem três casos particulares de filtros lineares: os *modelos de médias móveis* (MA, do inglês “*moving average*”), os *modelos autorregressivos* (AR), e os *modelos autorregressivos/médias móveis* (ARMA).

Um modelo de médias móveis de ordem q , conhecido como $MA(q)$, é gerado por uma combinação linear finita dos q choques aleatórios passados. Supondo um operador de retardo de média móvel, um processo $MA(q)$ tem como forma geral:

$$\bar{Z}_t = (1 - \theta B - \dots - \theta_q B^q) a_t = \Theta(B) a_t \quad (3.14)$$

Um processo autorregressivo de ordem p ($AR(p)$) é formado pela regressão do valor atual da série nos p valores anteriores e em um choque a_t . Usando um operador B de ordem p , temos:

$$(1 - \phi B - \dots - \phi_p B^p) \bar{Z}_t = \Phi(B) \bar{Z}_t = a_t \quad (3.15)$$

Combinando 3.14 e 3.15, temos um processo autorregressivo/médias móveis, ou $ARMA(p, q)$, da forma:

$$\Phi(B) \bar{Z}_t = \Theta(B) a_t$$

Dadas as condições de estacionariedade e os modelos apresentados, podemos concluir que o modelo $AR(p)$ será sempre invertível e o $MA(q)$ será sempre estacionário.

Quando a série não é estacionária, é interessante relaxar a condição de estacionariedade do processo $AR(p)$, de forma que as raízes G_p da equação a seguir estejam fora do círculo unitário ($|B| > 1$):

$$1 - \phi_1 B - \dots - \phi_p B^p = (1 - G_1^{-1} B)(1 - G_2^{-1} B) \dots (1 - G_p^{-1} B)$$

Quando acontece de uma dessas raízes G_p ser igual a 1, usamos os modelos ARIMA para estudar a ST de interesse, e nesse caso geralmente usamos a *diferenciação*. O método de diferenciação consiste em tomar diferenças sucessivas da série original, até se obter

uma série estacionária [39]. Esse processo pode ser comparado à derivação do cálculo, mas aplicado a funções discretas. Um operador de diferenciação pode ser aplicado sobre outro, pode sofrer defasamento e permite uma operação inversa, equivalente à integração. Definimos o operador de diferenciação de n -ésima ordem como:

$$\nabla^n Z_t = \nabla(\nabla^{n-1} Z_t)$$

Temos que o operador de diferenciação, em termos de B é dado por:

$$\nabla = 1 - B \tag{3.16}$$

O operador de diferenciação representado por 3.16 tem as seguintes propriedades:
Com:

1. Operador de diferenciação de ordem d : $\nabla^d = (1 - B)^d$
2. Operador de diferenciação com defasamento k : $\nabla_k = 1 - B^k$
3. Operador de somatório infinito: $\nabla^{-1} = (1 - B)^{-1} = S$

Um modelo ARIMA(p, d, q), então, tem como forma geral básica:

$$\Phi_p(B)\nabla^d Z_t = \Theta_q(B)a_t$$

Quando o processo é não-estacionário e sazonal, ele pode ser modelado pela teoria dos modelos ARIMA de duas formas: por processos puramente sazonais (SAR, SARI, MA sazonal) ou por modelos mistos SARIMA. Considerando uma série sazonal não-estacionária, de sazonalidade S , definimos o operador de diferenciação de ordem D para modelos sazonais como:

$$\nabla_S^D = (1 - B^S)^D$$

Generalizando, temos um modelo SARIMA(P, D, Q) $_S$, com expressão geral:

$$\Phi(B^S)\nabla_S^D Z_t = \Theta(B^S)b_t \tag{3.17}$$

Nesses modelos SARIMA, não há autocorrelação entre os valores de Z_t e Z_{t-1} . Podemos imaginar, então, que a série original é um conjunto de S processos independentes. Os erros b_t desses modelos, por isso, não seriam independentes, já que o modelo descreve,

por exemplo, a relação entre um valor de um determinado mês com o valor do mesmo mês do ano anterior, mas não dá informações sobre a relação desse valor com o do mês anterior. Para corrigir isso, é necessário ajustar um modelo $ARIMA(p, d, q)$ aos erros b_t :

$$\phi(B)\nabla^d b_t = \theta(B)a_t \quad (3.18)$$

Unindo 3.17 e 3.18, obtemos o modelo misto sazonal multiplicativo $SARIMA(p,d,q)(P,D,Q)_S$:

$$\phi(B)\Phi(B^S)\nabla_S^D\nabla^d Z_t = \theta(B)\Theta(B^S)a_t$$

Esta Subseção apresenta uma pequena introdução sobre os modelos ARIMA. Para maiores detalhes, consultar [38], [40], [39].

O ajuste desses modelos foi feito através da função `auto.arima()` do pacote `forecast` do software R, que retorna o melhor modelo ARIMA, sazonal ou não, de acordo com o critério de informação de Akaike (AIC). O AIC é uma medida de qualidade de um modelo estatístico, baseado na teoria da informação: considerando a complexidade do modelo proposto, ele produz uma estimativa da informação perdida quando esse modelo é usado para representar o processo gerador dos dados em questão [37], pela fórmula:

$$AIC = 2k - 2\ln(L) \quad (3.19)$$

Em 3.19, k é o número de parâmetros a serem estimados e L é o valor maximizado da função de verossimilhança do modelo.

O modelo escolhido será aquele de menor AIC, ou seja, o de maior verossimilhança e o de menor complexidade.

Para as previsões da série de cargas, usamos a função `forecast()`, para previsões de modelos de séries temporais. Em outras palavras, a função `forecast()` retornou previsões baseadas nos resultados da função `auto.arima()`.

3.2.9 Combinação de previsões

Desde o final da década de 1960, estudos sobre combinação de previsões vêm sendo realizados por pesquisadores da área. Em 1969, Bates e Granger [42] publicaram o primeiro artigo sobre o tema, apresentando métodos de combinação linear para dois

previsores, com base no fato de que, se há duas previsões e uma delas for descartada, esta poderia ter apresentado informações úteis, de duas maneiras:

1. uma previsão talvez seja baseada em variáveis ou informações que a outra não considerou,
2. a previsão talvez faça suposição diferente sobre a relação entre as variáveis.

Bates e Granger apresentam cinco métodos para determinar os pesos das combinações, de forma a minimizar a variância dos erros, e mostram que a combinação entre duas previsões individuais diferentes reduz a variância total do erro de previsão, em comparação às variâncias individuais. Embora os resultados sejam para combinações de duas previsões, eles podem ser estendidos para combinações de mais de duas previsões diferentes.

Esses pesos podem ser fixos ou variar ao longo do tempo. Diversos artigos sugerem que a utilização de pesos variantes resulta em melhores resultados para combinações lineares [42], [43], [44], [45].

Em 1985, Bunn [46] afirma que o problema da combinação é por si só um problema de previsão, cuja dificuldade é determinar como essa combinação pode ser feita buscando atingir uma menor variância dos erros. Para uma combinação eficiente, existem vários fatores que devem ser considerados, como a dependência entre os modelos de previsão individuais, razões das variâncias dos erros, tamanho da amostra e valores discrepantes. Por outro lado, ele afirma também que é importante observar a diferença entre um modelo de previsão simples e uma combinação: o primeiro combina entradas para obter modelos de previsão eficientes, e o segundo, combina as saídas desses modelos [47].

A revisão bibliográfica feita por Clemen [48] é citada em diversos estudos da área, desde sua publicação, e serviu como base para novos estudos na área. Muitos anos depois, Menezes *et al* [49] publicaram uma nova revisão bibliográfica, propondo um conjunto de critérios para a escolha do melhor método de combinação, envolvendo a variância, a assimetria da distribuição e a autocorrelação dos erros de previsão.

Grande parte dos estudos de combinações de previsões consideram um único objetivo, geralmente representado por uma medida de acurácia da previsão final. Reeves e Lawrence [50] defenderam a combinação de previsões como um processo de decisão multi-objetivos, uma vez que em muitas situações é interessante prever, além da magnitude, a direção da mudança da variável de interesse. Menezes *et al* reafirmaram essa proposta, devido à

dificuldade da escolha dos critérios e restrições a serem adotados na escolha do método de combinação.

Os experimentos realizados por Maines [51] para investigar como as combinações de previsões são feitas subjetivamente a partir de previsões realizadas por especialistas mostram que geralmente a média entre as previsões é usada quando não há informações que permitem distinguir entre os previsores, e que as combinações que incorporam apenas a acurácia dos previsores são pelo menos tão boas quanto aquelas que incorporam acurácia e dependência entre as previsões.

Neste estudo, usaremos o método *outperformance* de combinação, proposto por Bunn em dois artigos publicados na década de 1970 [52],[53], em que, dadas duas previsões p_1 e p_2 , a previsão final é:

$$P = \alpha_1 p_1 + \alpha_2 p_2 \quad (3.20)$$

onde, $\alpha_1 + \alpha_2 = 1$ e $0 \leq \alpha_1 \leq 1$.

Nesse método, a probabilidade de um modelo de previsão é interpretada como uma avaliação sobre a eficiência desse modelo em relação ao outro utilizado na combinação. Em geral, esse método apresenta bons resultados para amostras pequenas. Embora seja apresentado para combinações de dois previsores, pode ser usado para combinações maiores. Por sua simplicidade e facilidade de interpretação, esse foi o método escolhido para combinar as previsões neste estudo. Os valores das constantes α_1 e α_2 foram obtidos usando a função *optim()*.

3.2.10 Medidas de erro de previsão

O *erro de previsão*, ou *desvio*, é uma medida utilizada para comparar o valor previsto da série com o valor observado, possibilitando, assim, avaliar a qualidade do modelo de previsão. A ideia é que o modelo escolhido minimize o máximo possível os erros. O erro da previsão k passos a frente, feita no instante t , é dado por:

$$e_{t+k|t} = Z_{t+k|t} - \hat{Z}_{t+k|t}$$

Para a previsão 1 passo a frente, podemos simplificar a notação:

$$e_{t|t-1} = Z_t - \hat{Z}_{t|t-1}$$

O *erro médio* (ME) é uma média simples de todos os valores e_t , e é o ponto de partida para as medidas de erro mais adotadas na análise da adequação dos modelos de previsão. Assim, temos:

$$ME = \frac{1}{N} \sum_{t=1}^N e_t$$

A partir dele, obtemos todas as demais medidas de erro utilizadas, sendo elas:

1. Erro absoluto médio (MAE):

$$MAE = \frac{1}{N} \sum_{t=1}^N |e_t|$$

2. Erro percentual médio (MPE):

$$MPE = \frac{100}{N} \sum_{t=1}^N \frac{e_t}{Z_t}$$

3. Erro absoluto percentual médio (MAPE):

$$MAPE = \frac{100}{N} \sum_{t=1}^N \frac{|e_t|}{|Z_t|}$$

4. Erro quadrático médio (MSE):

$$MSE = \frac{1}{N} \sum_{t=1}^N e_t^2$$

Uma das expressões de erro mais utilizadas em estudos na área da Estatística, é o MAPE, devido à sua facilidade de compreensão: ele expressa a porcentagem média de desvios cometidos na previsão da ST, então quanto menor ele for, melhor é a previsão. O MSE também é muito utilizado, em diversas áreas, embora sua interpretação seja mais complicada por sua unidade de medida ser elevada ao quadrado. Como o MSE é uma média dos quadrados dos desvios, quanto maiores forem os desvios, maior ele será; por outro lado, se os desvios forem pequenos, ele será pequeno também.

É importante observar que diferentes medidas de erro podem ser minimizadas por ajustes diferentes de um mesmo modelo, ou seja, nem sempre o ajuste que minimiza o MSE minimiza também o MAPE, por exemplo.

Usamos neste trabalho, para verificar a adequação dos modelos, o erro MAPE, por sua vantagem na interpretação, e por ser a medida mais usada nas publicações sobre previsão de cargas, como mencionado em [5]. Usamos, também, o erro percentual absoluto (APE), dado por:

$$APE = 100 \frac{|e_t|}{|Z_t|}$$

3.2.11 Função *optim()*

A função *optim()* do software R foi utilizada durante os experimentos para determinação dos parâmetros dos métodos de previsão e dos pesos das combinações de previsores. Essa rotina permite obter os parâmetros que minimizam uma determinada função, com base nos métodos Nelder-Mead, quase-Newton ou Gradiente Conjugado.

Os argumentos obrigatórios da função são os valores iniciais dos parâmetros e a função a ser otimizada. Dentre os argumentos opcionais, está o método de otimização. Em todos os casos em que a *optim()* foi utilizada, optamos pelo método *L-BFGS-B*, que consiste em um método de otimização quase-Newton com memória limitada para funções não-lineares sujeitas a restrições simples, e com um grande número de parâmetros ou grande complexidade.

O método de Newton nos permite obter o mínimo de uma função $f(x)$, a partir da expansão em série de Taylor de sua derivada:

$$f'(x) = f'(x_0) + (x - x_0)f''(x_0) \quad (3.21)$$

Mas no ponto de mínimo, $f'(x) = 0$, então, de 3.21, temos:

$$x = x_0 - \frac{f'(x_0)}{f''(x_0)} \quad (3.22)$$

No caso multivariado, 3.22 passa a ser:

$$\mathbf{x} = \mathbf{x}_0 - H^{-1} \nabla f \quad (3.23)$$

Em 3.23, H é a Hessiana, matriz de derivadas segundas de $f(x)$, e ∇f é o vetor gradiente, contendo as derivadas primeiras da função:

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_N} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial x_N x_1} & \frac{\partial^2 f}{\partial x_N x_2} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}$$

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_N} \right)$$

A Hessiana, então, é uma matriz quadrada de dimensão N . Se N é grande, é necessário muito tempo e recursos computacionais para o cálculo de H e sua inversão. Os métodos quase-Newton usam uma aproximação da Hessiana para a otimização.

Um dos métodos quase-Newton tradicionais é o método BFGS, que armazena uma aproximação $N \times N$ da inversa da matriz Hessiana a cada iteração. O L-BFGS é uma adaptação do BFGS, usando memória limitada, e armazena, ao invés de uma matriz quadrada $N \times N$, pares de vetores da forma (dado que k é iteração atual do algoritmo):

$$\{x_{k+1} - x_k; \nabla f_{k+1} - \nabla f_k\}$$

A aproximação da Hessiana é calculada a partir de operações envolvendo esses pares de vetores. A cada iteração, um novo par de vetores é armazenado, e o mais antigo, descartado. O método L-BFGS-B é uma extensão do método L-BFGS, para funções não-lineares sujeitas a restrições simples.

Maiores informações sobre a função *optim()* e seus possíveis métodos de otimização podem ser encontradas em [36].

3.2.12 Modelos propostos

Propomos neste trabalho uma previsão de perfis horários de cargas elétricas usando modelos univariados e combinações de previsões desses modelos.

Consideramos um modelo de carga padrão, com forma geral:

$$L_{t,h} = B_{t,h} + \varepsilon_{t,h} \quad (3.24)$$

A previsão de 3.24 é dada por:

$$\hat{L}_{t,h} = \hat{B}_{t,h}$$

Sabemos que B_t é uma função linear dos valores passados da série de cargas, e sua estimativa é feita por modelos de STs univariados.

Para as previsões, utilizamos diferentes técnicas:

1. Prever perfis de cargas por conjuntos de modelos em paralelo:
 - (a) usando 168 modelos em paralelo, um para cada hora da semana, pelos métodos MM e MMD, por métodos de amortecimento (AES e Holt) e por modelos ARIMA (Método I);
 - (b) usando 24 modelos em paralelo, um para cada hora do dia, por métodos sazonais: HW e modelos SARIMA (Método II).
2. Prever perfis de cargas usando o método de Holt-Winters-Taylor (HWT) sequencialmente (Método III);
3. Prever perfis de cargas por uma combinação de modelos sequenciais com conjuntos de modelos em paralelo:
 - (a) (Método IV) Considerando

$$B_{t,h} = m_t + d_{t,h} \quad (3.25)$$

em que:

- i. cargas médias diárias (m_t) previstas por HW sequencialmente;
- ii. desvios de carga horários em relação às médias diárias ($d_{t,h}$) previstos por MM, AES, Holt (168 modelos em paralelo) e HW (24 modelos em paralelo).

- (b) (Método V) Considerando

$$B_{t,h} = l_t + e_{t,h} \quad (3.26)$$

em que:

- i. cargas (l_t) previstas por HTW sequencialmente;
- ii. desvios do previsor de HWT em relação às cargas observadas ($e_{t,h}$) previstos por MM, AES, Holt (168 modelos em paralelo) e HW (24 modelos em paralelo).

4. Prever perfis de cargas combinando linearmente os melhores previsores simples, obtidos pelos métodos I, II, III, IV e V.

Como os desvios horários em relação à média são uma série sem tendência, o método IV permite que a tendência da série de cargas elétricas seja retirada. O método V, por sua vez, tenta ajustar um modelo aos desvios do previsor de HWT.

Os previsores com menores percentuais de erro na amostra de teste I são combinados linearmente dois a dois. Os resultados finais das combinações das previsões por esses métodos mencionados são comparados com os das redes neurais usadas em [2] e [24] para fazer as previsões da série de cargas do Rio de Janeiro.

Em [2], foram usados dados de temperaturas diárias previstas e de cargas horárias passadas como entradas de uma RN do tipo *perceptron* de múltiplas camadas. Essa RN foi treinada 15 vezes, e a cada uma delas o resultado foi diferente. Usamos para fazer a comparação entre os estudos os erros MAPEs do pior, do médio e do melhor caso da RN.

A RN proposta em [24] apresenta os mesmos dados de entrada, e foi treinada, também, 15 vezes. Os resultados, porém, são um pouco diferentes dos encontrados pela outra rede. Apenas os MAPEs do pior e do melhor caso foram publicados, e estão presentes na tabela com os resultados das combinações, no Capítulo 4.

4 RESULTADOS E DISCUSSÕES

Utilizamos cinco métodos diferentes para previsão de perfis de cargas sobre a série de demanda de energia do Rio de Janeiro. Ajustamos os modelos de previsão na amostra de ajuste (semanas 1 a 54) e calculamos os erros absolutos médios percentuais no intervalo da amostra de teste I (semanas 55 a 84), como explicado na Seção 3.1. Os modelos usados para as combinações são escolhidos de acordo com esses erros, e os resultados finais são comparados, na amostra de teste II (semanas 85 a 104), com as redes neurais em [2] e [24], modelos multivariados que consideram a influência da temperatura na série de cargas elétricas.

A Tabela 4.1 a seguir mostra os erros calculados na amostra de teste I para os modelos dos métodos I, II, III, IV e V, mencionados no Capítulo 3, que chamamos neste Capítulo de métodos simples.

Os métodos I e II consistem em aplicações de métodos tradicionais univariados para previsão de séries temporais por meio de conjuntos de modelos em paralelo. Ajustamos 168 modelos, um para cada hora da semana, quando usamos métodos não-sazonais (MM, MMD, AES, Holt e ARIMA) na previsão, e 24 modelos, um para cada hora do dia, no caso de métodos sazonais (HW e SARIMA). Dentre todos eles, os métodos com menor percentual de erro na amostra de teste I são o HW e o SARIMA, o que já era esperado, dada a sazonalidade da série de cargas.

O amortecimento exponencial com sazonalidade dupla (HWT) utilizado para a previsão sequencialmente, método III, é o que obtém o melhor desempenho na amostra de teste I, em comparação aos demais métodos testados.

O método IV tem um resultado um pouco melhor quando a previsão final (como em 3.25) é dada pela série de cargas médias diárias (m_t) previstas por HW combinada com os desvios de carga horários ($d_{t,h}$) em relação às médias diárias, previstos por AES.

Com relação ao método V, o MAPE na amostra de teste I é menor quando o previsor (l_t) pelo método de HWT é combinado com a previsão de seus desvios horários ($e_{t,h}$) por AES, como em 3.26.

No geral, o método proposto por Taylor [33], HWT, gera o previsor mais eficiente. Com excessão de sua combinação com os desvios previstos por HW, apresentou, em todos

		MAPE (amostra de teste I)
Método I (168 modelos em paralelo)	<i>Naive</i>	4,80
	MM	4,32
	MMD	5,39
	AES	4,26
	Holt	4,59
	ARIMA	4,45
Método II (24 modelos em paralelo)	<i>Naive</i>	7,34
	HW	3,28
	SARIMA	2,85
Método III	HWT	2,39
Método IV (modelo $L_{t,h} = m_t + d_{t,h} + \varepsilon_{t,h}$)	HW+MM	3,35
	HW+AES	3,27
	HW+Holt	3,34
	HW+HW	13,32
Método V (modelo $L_{t,h} = l_t + e_{t,h} + \varepsilon_{t,h}$)	HWT+MM	2,83
	HWT+AES	2,63
	HWT+Holt	3,16
	HWT+HW	9,71

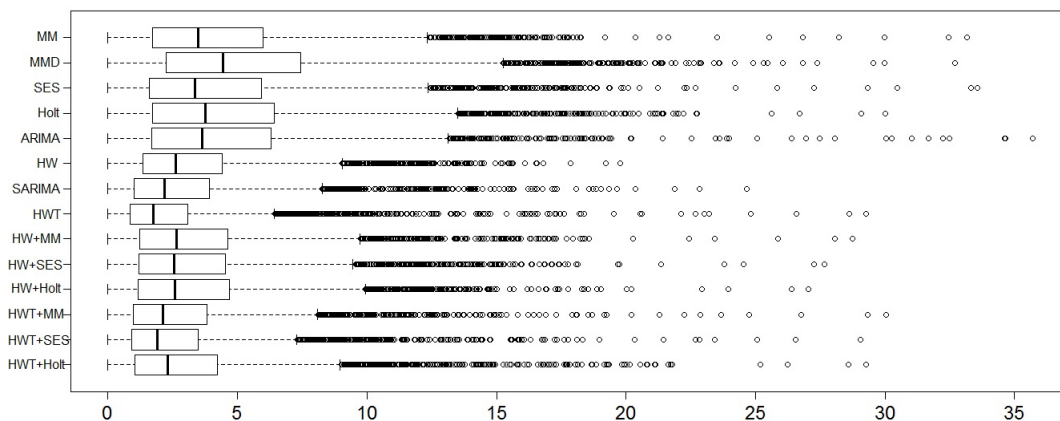
Tabela 4.1: MAPEs dos métodos simples para a amostra de teste I.

	P1	P2	α_1	α_2
Combinação I	HW	SARIMA	0,20	0,80
Combinação II	HW	HW+AES	0,99	0,01
Combinação III	SARIMA	HW+AES	0,99	0,01
Combinação IV	HW	HWT+AES	0,99	0,01
Combinação V	SARIMA	HWT+AES	0,99	0,01
Combinação VI	HWT	HW	0,90	0,10
Combinação VII	HWT	SARIMA	0,72	0,28
Combinação VIII	HWT	HW+AES	0,99	0,01
Combinação IX	HWT	HWT+AES	0,99	0,01
Combinação X	HW+AES	HWT+AES	0,65	0,35

Tabela 4.2: Pesos das combinações.

os demais casos, percentuais de erros baixos.

A Figura 4.1 mostra os *boxplots* dos erros percentuais absolutos dos métodos simples testados, calculados na amostra de teste I. Os modelos com MAPEs muito altos foram descartados, para evitar problemas na construção do gráfico final. O método III, HWT aplicado sequencialmente, é o método com menor dispersão dos erros, aparentemente.

Figura 4.1: *Boxplots* dos APEs dos métodos simples para a amostra de teste I.

Para as combinações, escolhemos os métodos que resultaram em menores MAPEs na amostra de teste I, a partir dos resultados da Tabela 4.1. As combinações foram feitas de acordo com o método *outperformance* [52], [53], com forma dada por 3.20, e seus pesos, obtidos usando a função *optim()*, são apresentados na Tabela 4.2.

		MAPE (amostra de teste II)	Variância APes (amostra de teste II)
Combinação	I	3,75	11,14
	II	4,10	13,19
	III	3,74	11,10
	IV	4,09	13,08
	V	3,72	11,00
	VI	2,52	7,13
	VII	2,61	6,80
	VIII	2,48	7,15
	IX	2,48	7,16
	X	3,23	8,77
Método III	HWT	2,47	7,15
Redes Neurais Resultados de Hippert [2]	Mínimo	2,26	-
	Média	2,44	-
	Máximo	2,75	-
Redes Neurais Resultados de Hippert <i>et al</i> [24]	Mínimo	2,35	-
	Máximo	2,65	-

Tabela 4.3: MAPEs das combinações e dos métodos comparativos para a amostra de teste II.

O previsor de HWT é o que contribui com mais informação para a previsão final em todas as combinações em que foi utilizado. Em três, das quatro combinações das quais faz parte, a previsão por HWT é responsável por noventa por cento ou mais do resultado da previsão final.

O objetivo das combinações de previsões é alcançar um melhor desempenho, embora isso nem sempre implique em menores erros — o que é garantido nessa técnica é a diminuição na variância total dos erros, conforme Subseção 3.2.9.

Os erros das combinações, e das redes neurais propostas por Hippert [2] e Hippert *et al* [24], calculados no intervalo da amostra de teste II, estão na Tabela 4.3. Devido aos resultados encontrados a partir das combinações, optamos por comparar os erros das combinações também com os erros, no mesmo intervalo, do método III.

As combinações mais eficientes, em termos de MAPE, são aquelas em que um dos dois previsores é o de HWT (combinações VI, VII, VIII, IX). Entretanto, nenhuma delas deve ser considerada uma boa opção. No caso das combinações VI e VII, em comparação aos resultados do previsor de HWT, a previsão final tem maior média de erro, e pouca

diminuição na variância dos APEs.

Em média, os erros das previsões finais das combinações VIII e IX quase não são diferentes do HWT, assim como as variâncias dos erros. Isso pode ser explicado pelos pesos atribuídos ao HWT nessas combinações, resultando em previsões finais praticamente idênticas à previsão do método III.

A Figura 4.2 permite concluir que, de fato, as previsões combinadas ao predictor de HWT contribuem com pouquíssima informação na previsão final. Os *boxplots* dos APEs dessas combinações mostram que esses erros se comportam quase que identicamente aos erros do método III, tanto em dispersão quanto em valor médio.

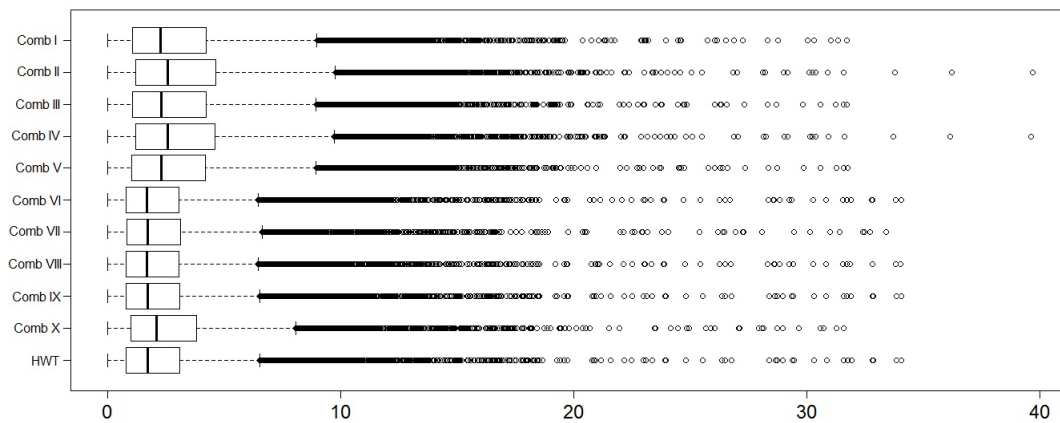


Figura 4.2: *Boxplots* dos APEs das combinações e dos métodos comparativos para a amostra de teste II.

Em relação à rede neural de [2], em média, o HWT tem um desempenho inferior. Porém, pelos resultados, podemos inferir que em cinquenta por cento das vezes em que a RN foi simulada, aproximadamente, a previsão da rede tem maior percentual de erro que o predictor de HWT.

A RN de [24] obtém um resultado levemente melhor que o HWT. Embora não dispomos de seu resultado médio, podemos inferir o mesmo que foi concluído sobre a RN de [2].

Para ambos os casos, além do que já foi mencionado, as RNs são modelos mais complexos do que o HWT simples, dependem fortemente dos valores iniciais atribuídos aos seus parâmetros, e seus MAPEs são obtidos em uma única amostra, não sendo possível prever se esses resultados se repetiriam.

Portanto, o método III tem o menor erro para a previsão da série de cargas,

dentre todos os métodos simples testados, combinações experimentadas e RNs propostas. Contudo, para afirmarmos que o HWT simples é um método ótimo, é preciso analisar os erros da previsão final. O ideal é que esses erros tenham uma distribuição normal, com média zero e variância constante, e sejam descorrelacionados. Em teoria, essa análise dos resíduos é necessária [40], entretanto, ela não é vista com frequência nos estudos da área de previsões de cargas a curto prazo, conforme apontado por [5].

A primeira suposição a ser feita é a de normalidade dos desvios da previsão. O histograma da Figura 4.3 mostra que a dispersão dos desvios é grande, e os testes de normalidade de Lilliefors e Shapiro-Wilk, com estatísticas de teste iguais a 0,1037 e 0,9084, respectivamente, rejeitam a hipótese nula de normalidade dos desvios, a um nível de significância de 0,05, com valor p próximo a zero.

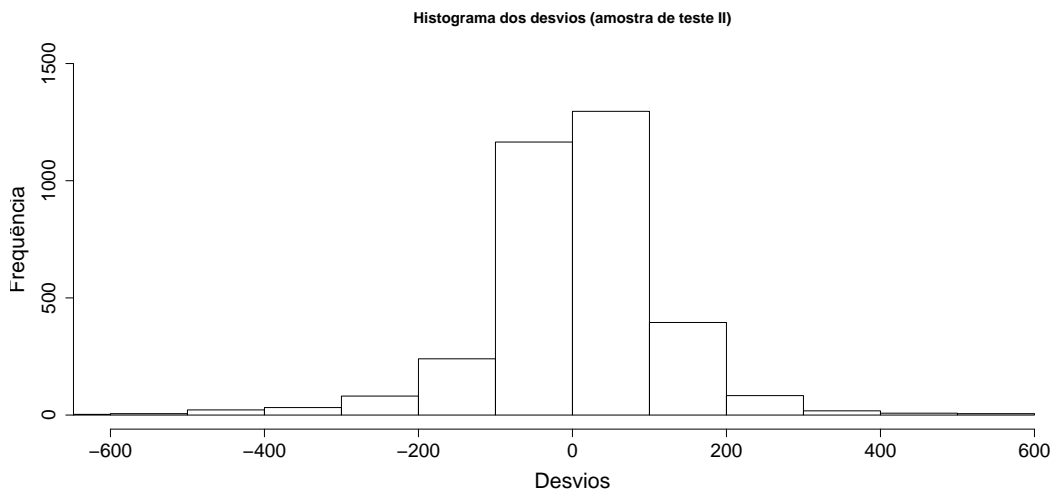


Figura 4.3: Histograma dos desvios do predictor de HWT para a amostra de teste II.

O gráfico de dispersão dos desvios sugere que eles estão, em sua maioria, próximos de zero, mas que não são homocedásticos (de variância constante).

A função de autocorrelação (FAC) mede o grau de autocorrelação entre as observações de uma ST. A Figura 4.5 indica que os desvios da previsão do HWT são autocorrelacionados.

A função de autocorrelação parcial (FACP) permite a identificação do grau de defasagem de um modelo autorregressivo. A partir da Figura 4.6, verificamos que a autocorrelação dos desvios é de defasagem 1; ou seja, essa autocorrelação existe entre uma determinada observação da série e a observação imediatamente anterior a ela.

Todavia, para afirmarmos a autocorrelação da série de desvios, realizamos dois testes.

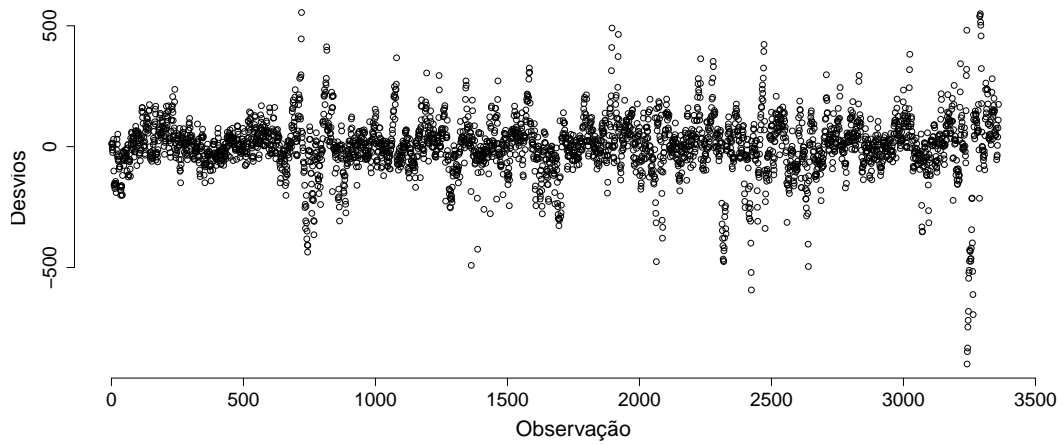


Figura 4.4: Gráfico de dispersão dos desvios do predictor de HWT para a amostra de teste II.

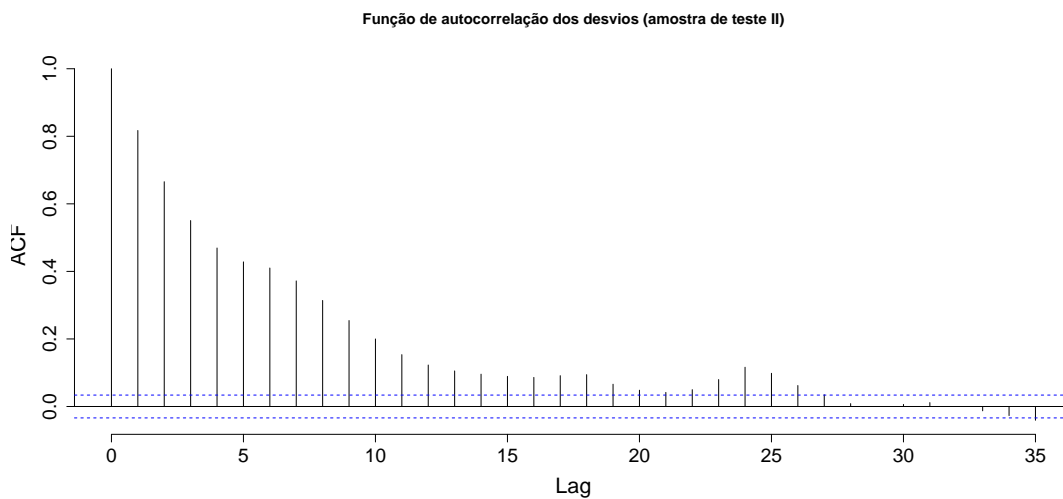


Figura 4.5: Função de autocorrelação dos desvios do predictor de HWT para a amostra de teste II.

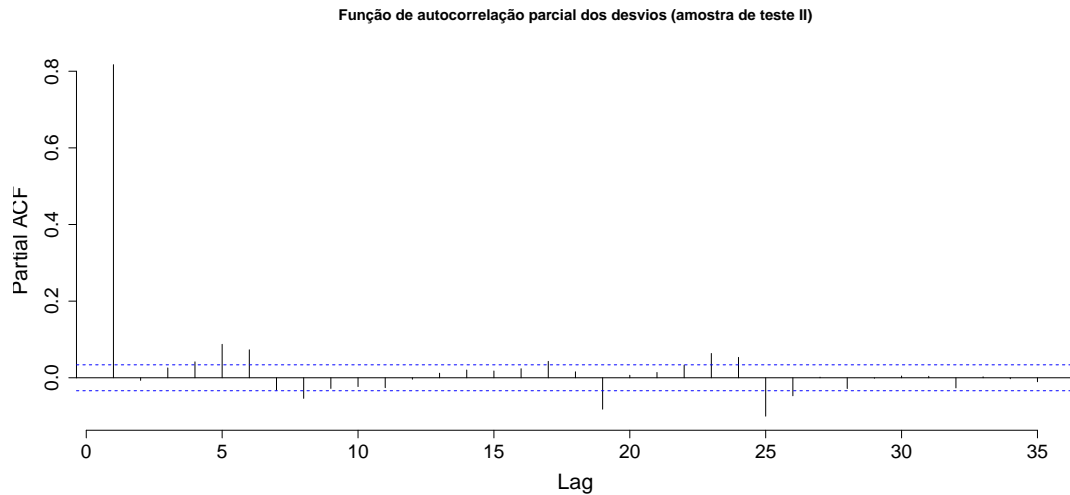


Figura 4.6: Função de autocorrelação parcial dos desvios do predictor de HWT para a amostra de teste II.

O primeiro, de Ljung-Box, tem hipótese nula de independência entre as observações, considerando qualquer grau de defasagem entre elas. A estatística encontrada é 2246,54 e o valor p é próximo de zero, então, a um nível de significância de 0,05, concluímos que os desvios do HWT são autocorrelacionados.

A estatística de Durbin-Watson serve para testar a independência dos desvios, com base na suposição de que esses desvios são gerados por um processo autorregressivo de primeira ordem, e assume valores entre 0 e 4. Quanto mais próxima de 2 ela for, menor é a autocorrelação da série. Encontramos um valor aproximadamente igual a 0,37 para essa estatística, sugerindo uma forte autocorrelação positiva entre os desvios.

Verificamos, então, que os desvios não são normalmente distribuídos, são heterocedásticos e fortemente autocorrelacionados. Assim, embora o método de HWT resulte em erros percentuais baixos, em média, e com variância pequena, ele necessita, talvez, de alguns ajustes para extrair toda a informação contida nos dados.

Com relação ao erro de previsão final do HWT, podemos concluir que o resultado está de acordo com o esperado. Estudos que utilizaram como dados a mesma série de cargas do Rio de Janeiro apresentaram erros aproximados aos que encontramos [6], [24],[54]. O nível de erro depende das características da série de cargas, que, por outro lado, depende da empresa de energia de onde os dados foram obtidos e do país de origem. Séries de cargas inglesas, por exemplo, têm previsões com erros menores, entre 1,5% e 2% [6].

Devido à falta dos dados simulados pelas redes neurais, não há como testar se há, de

fato, significância estatística entre a diferença dos resultados do HWT e das RNs. Além da análise dos resíduos não ser muito utilizada na prática, como mencionado anteriormente, testes de significância também não são comuns em estudos na área de previsões. Apenas alguns poucos artigos publicados atualmente parecem tratar do assunto [55], [30], [31].

Temos, como limitação deste estudo, o tamanho da série de cargas. A série utilizada é pequena, e por isso não permite a identificação da tendência e a modelagem da sazonalidade anual. Além disso, dispomos de apenas uma série de cargas, enquanto o ideal seria repetir este estudo em outras séries de cargas.

5 CONCLUSÕES

A previsão de cargas elétricas é fundamental para o planejamento das empresas de energia. O foco deste estudo são as previsões a curto prazo; assim, aplicamos métodos univariados de previsão de séries temporais a uma série real de demanda de energia do Rio de Janeiro, nos anos de 1996 e 1997.

Utilizamos cinco técnicas diferentes para a previsão dos perfis de cargas:

1. Método I: 168 modelos em paralelo, um para cada hora da semana, usando os métodos das MM, MMD, SES, Holt e modelos ARIMA;
2. Método II: 24 modelos em paralelo, um para cada hora do dia, usando HW e modelos SARIMA;
3. Método III: HWT aplicado sequencialmente;
4. Método IV: considerando um modelo $L_{t,h} = m_t + d_{t,h} + \varepsilon_{t,h}$, prevendo cargas médias diárias (m_t) por HW e desvios de carga horários (d_t) em relação às médias diárias por MM, SES, Holt e HW;
5. Método V: considerando um modelo $L_{t,h} = l_t + e_{t,h} + \varepsilon_{t,h}$, e usando HWT (l_t), combinado com a previsão por MM, SES, Holt e HW dos desvios (d_t) do previsor de HWT em relação às cargas observadas.

Os melhores previsores desses cinco métodos foram combinados linearmente dois a dois pelo método *outperformance* de combinação, dado por 3.20, e os resultados das combinações foram comparados aos das RNS em [2] e [24].

Proposto por [33], o método de HWT, para dupla sazonalidade aditiva, aparentemente produz previsões para perfis de cargas muito boas, e pudemos constatar sua eficiência sobre a série de cargas do Rio de Janeiro.

Em termos de MAPE, o HWT é o método de melhor desempenho em relação aos demais métodos simples testados. Em relação às combinações, as melhores previsões finais são obtidas quando um dos previsores individuais é o de HWT. Devido a esses resultados, optamos por compará-los, também, aos resultados do previsor de HWT simples.

Os erros das combinações, em média, são maiores que os do HWT ou quase não apresentam diferença; o mesmo acontece com as variâncias dos APEs das combinações em relação ao método simples. Isso pode ser explicado pelos pesos atribuídos a cada previsão individual no momento da combinação: na maioria das combinações das quais faz parte, a previsão pelo método III é responsável por noventa por cento ou mais do resultado da previsão final — resultando em previsões finais quase idênticas às do Holt-Winters-Taylor.

A rede neural em [2] tem um desempenho superior ao HWT, em média, enquanto a RN proposta em [24] é um pouco melhor que o método III. Entretanto, podemos inferir que, em aproximadamente cinquenta por cento das vezes em que as redes foram simuladas, suas previsões têm maior percentual de erro que o previsor de HWT.

Além disso, os resultados de simulações da rede são muito dependentes dos valores iniciais atribuídos aos seus parâmetros, e são obtidos em apenas uma amostra, não sendo possível prever se esses resultados se repetiriam em novas simulações. Por esses motivos, atrelados ao fato de que RNs são modelos complexos, o HWT se torna, talvez, um método mais adequado e confiável para aplicações práticas.

Tentativas de melhorar o HWT, como a proposta de somar às previsões do método III um componente dependente de variáveis meteorológicas, modelado por regressão linear [56], também não obtiveram resultados melhores do que os do HWT simples. Esse modelo de regressão foi aplicado a uma série de temperaturas observadas, mas geralmente usamos dados de previsões de temperatura, que possivelmente levariam a um pior desempenho do modelo de regressão proposto.

Apesar dos resultados satisfatórios, o HWT não pode ser considerado como um método ótimo de previsão. Uma análise dos desvios da previsão do HWT mostra que eles não são normalmente distribuídos, são heterocedásticos e fortemente autocorrelacionados. Assim, concluímos que embora resulte em erros percentuais baixos, em média, e com variância pequena, o método de Holt-Winters-Taylor necessita, talvez, de alguns ajustes para que consiga extrair toda a informação contida nos dados.

Os resultados obtidos neste estudo estão de acordo com a literatura, embora não seja possível testar se a diferença entre eles e os resultados externos das RNs é significativa. Assim como a análise dos desvios, testes de significância não são utilizados com frequência nos estudos da área.

Além disso, encontramos como limitação a inutilização dos modelos propostos para

previsões de dias especiais, como feriados, greves ou dias de jogos de futebol.

Como proposta para estudos futuros, podemos testar melhorias no método de HWT, combinações com pesos variantes ou outros modelos de combinação, ou até mesmo métodos de previsão diferentes dos que experimentamos neste trabalho.

REFERÊNCIAS

- [1] SARGUNARAJ, S., SEN GUPTA, D. P., DEVI, S., “Short-term load forecasting for demand side management”, *IEEE Proceedings - Generation, Transmission and Distribution*, v. 144, n. 1, pp. 68–74, 1997.
- [2] HIPPERT, H. S., *Previsão de cargas a curto prazo - Uma avaliação da viabilidade do uso de redes neurais artificiais*, Ph.D. Thesis, Departamento de Engenharia Elétrica da PUC, Rio de Janeiro, RJ, Brasil, 2001.
- [3] HAHN, H., MEYER-NIEBERG, S., PICKL, S., “Electric load forecasting methods: Tools for decision making”, *European Journal of Operational Research*, v. 199, pp. 902–907, 2009.
- [4] BUNN, D. W., “Forecasting loads and prices in competitive power markets”, *Proceedings of the IEEE*, v. 88, n. 2, pp. 163–169, 2000.
- [5] HIPPERT, H. S., PEDREIRA, C. E., SOUZA, R. C., “Neural networks for short-term load forecasting: a review and evaluation”, *IEEE Transactions on Power Systems*, v. 16, n. 1, pp. 44–55, 2001.
- [6] TAYLOR, J. W., MENEZES, L. M., MCSHARRY, P. E., “A comparison of univariate methods for forecasting electricity demand up to a day ahead”, *International Journal of Forecasting*, v. 22, pp. 1–16, 2006.
- [7] BUNN, D. W., FARMER, E. D., *Comparative models for electrical load forecasting*. John Wiley & Sons, 1985.
- [8] DOUGLAS, A. P., BREIPOHL, A. M., LEE, F. N., ADAPA, R., “Risk due to load forecast uncertainty in short term power system planning”, *IEEE Transactions Power Systems*, v. 13, n. 4, pp. 1493–1499, 1998.
- [9] HOBBS, B. F., JITPRAPAIKULSARN, S., KONDA, S., CHANKONG, V., LOPARO, K. A., J., M. D., “Analysis of the value for unit commitment of improved load forecasting”, *IEEE Transactions Power Systems*, v. 14, n. 4, pp. 141–149, 1999.

- [10] RANAWEERA, D. K., KARADY, G. G., FARMER, R. G., “Economic impact analysis of load forecasting”, *IEEE Transactions Power Systems*, v. 12, n. 3, pp. 1388–1392, 1997.
- [11] DAVIES, M., “The relationship between weather and electricity demand”, *The Institution of Electrical Engineers Monograph*, v. 314, pp. 27–37, 1958.
- [12] HEINEMANN, G. T., NORDMAN, D. A., PLANT, E. C., “The relationship between summer weather and summer loads - A regression analysis”, *IEEE Transactions on Power Apparatus and Systems*, v. PAS-85, n. 11, pp. 1144–1154, 1966.
- [13] ASBURY, C. E., “Weather load model for electric demand and energy forecasting”, *IEEE Transactions on Power Apparatus and Systems*, v. PAS-94, n. 4, pp. 1111–1116, 1975.
- [14] HAGAN, M. T., BEHR, S. M., “The time series approach to short term load forecasting”, *IEEE Transactions Power Systems*, v. PWRS-2, n. 3, pp. 785–791, 1987.
- [15] JENKINS, G. M., “Practical experiences with modeling and forecast”, *Time Series*, 1979.
- [16] YANG, H. T., HUANG, C. M., “A new short-term load forecasting approach using self-organizing fuzzy ARMAX models”, *IEEE Transactions Power Systems*, v. 13, n. 1, pp. 217–225, 1998.
- [17] WANG, B., TAI, N., ZHAI, H., YE, J., ZHU, J., QI, L., “A new ARMAX model based on evolutionary algorithm and particle swarm optimization for short-term load forecasting”, *Electric Power Systems Research*, v. 78, pp. 1679–1685, 2008.
- [18] BENNETT, C., STEWART, R. A., LU, J., “Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks”, *Energies*, v. 7, pp. 2938–2960, 2014.
- [19] RAMANATHAN, R., ENGLE, R., GRANGER, C. W. J., VAHID-ARAGHI, F., BRACE, C., “Short-run forecasts of electricity loads and peaks”, *International Journal of Forecasts*, v. 13, pp. 161–174, 1997.

- [20] SOLIMAN, S. A., PERSAUD, S., EL-NAGAR, K., EL-HAWARY, M. E., “Application of least absolute value parameter estimation based on linear programming to short-term load forecasting”, *Electrical Power & Energy Systems*, v. 19, n. 13, pp. 209–216, 1997.
- [21] HYDE, O., HODNETT, P. F., “An adaptable automated procedure for short-term electricity load forecasting”, *IEEE Transactions Power Systems*, v. 12, n. 1, pp. 84–93, 1997.
- [22] HINOJOSA, V. H., HOESE, A., “Short-term load forecasting using fuzzy inductive reasoning and evolutionary algorithms”, *IEEE Transactions on Power Systems*, v. 25, pp. 565–574, 2010.
- [23] HANMANDLU, M., CHAUHAN, B. K., “Load forecasting using hybrid models”, *IEEE Transactions on Power Systems*, v. 26, n. 1, pp. 20–29, 2011.
- [24] HIPPERT, H. S., BUN, D. W., SOUZA, R. C., “Large neural networks for electricity load forecasting: Are they overfitted?” *International Journal of Forecasting*, v. 21, pp. 425–434, 2005.
- [25] HIPPERT, H. S., TAYLOR, J. W., “An evaluation of bayesian techniques for controlling model complexity in a neural network for short-term load forecasting”, *Neural Networks*, v. 23, pp. 386–395, 2010.
- [26] MBAMALU, G. A. N., EL-HAWARY, M. E., “Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation”, *IEEE Transactions Power Systems*, v. 8, n. 1, pp. 343–348, 1993.
- [27] PAPPAS, S. S., EKONOMOU, L., KARAMPELAS, P., KARAMOUSANTAS, D. C., KATSIKAS, S. K., CHATZARAKIS, G. E., SKAFIDAS, P. D., “Electricity demand load forecasting of the Hellenic power system using an ARMA model”, *Electric Power Systems Research*, v. 80, pp. 256–264, 2010.
- [28] INFIELD, D. G., HILL, D. C., “Optimal smoothing for trend removal in short term electricity demand forecasting”, *IEEE Transactions Power Systems*, v. 13, n. 3, pp. 1115–1120, 1998.

- [29] TAYLOR, J. W., “Short-term load forecasting with exponentially weighted methods”, *IEEE Transactions on Power Systems*, v. 27, pp. 458–464, 2012.
- [30] HONG, W., “Application of chaotic ant swarm optimization in electric load forecasting”, *Energy Policy*, v. 38, pp. 5830–5839, 2010.
- [31] HONG, W., “Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm”, *Energy*, v. 36, pp. 5568–5578, 2011.
- [32] TAYLOR, J. W., “Short-term electricity demand forecasting using double seasonal exponential smoothing”, *Journal of the Operational Research Society*, v. 54, pp. 799–805, 2003.
- [33] TAYLOR, J. W., “Triple seasonal methods for short-term electricity demand forecasting”, *European Journal of Operational Research*, v. 204, pp. 139–152, 2010.
- [34] LAMEDICA, R., PRUDENZI, A., SFORNA, M., CACIOTTA, M., CENCELLI, V. O., “A neural network based technique for short-term forecasting of anomalous load periods”, *IEEE Transactions on Power Systems*, v. 11, n. 4, pp. 1749–1756, 1996.
- [35] ARORA, S., TAYLOR, J. W., “Short-term forecasting of anomalous load using rule-based triple seasonal methods”, *IEEE Transactions on Power Systems*, v. 28, n. 3, pp. 3235–3242, 2013.
- [36] “The R Project for Statistical Computing”, <http://www.r-project.org/>. Último acesso em 27 de julho de 2014.
- [37] “CRAN, The Comprehensive R Archive Network”, <http://cran.r-project.org/>. Último acesso em 27 de julho de 2014.
- [38] CHATFIELD, C., *The analysis of time series - An introduction*. Chapman and Hall/CRC, 6a ed., 2003.
- [39] MORETTIN, P. A., TOLOI, C. M. C., *Análise de séries temporais*. Edgar Blucher, 2a ed., 2006.

- [40] BOX, G. E. P., JENKINS, G. M., REINSEL, G. C., *Time series analysis: forecasting and control*. Prentice-Hall, 4a ed., 2008.
- [41] MAKRIDAKIS, S., WHEELWRIGHT, S. C., HYNDMAN, R. J., *Forecasting - Methods and Applications*. John Wiley & Sons, 3a ed., 1998.
- [42] BATES, J. M., GRANGER, C. W. J., “The combination of forecasts”, *Journal of the Operational Research Society*, v. 20, n. 4, pp. 451–468, 1969.
- [43] DEUTSCH, M., GRANGER, C. W. J., TERÄSVIRTA, T., “The combination of forecasts using changing weights”, *International Journal of Forecasting*, v. 10, pp. 47–57, 1994.
- [44] TAYLOR, J. W., MAJITHIA, 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.
- [45] TERUI, N., DIJK, H. K., “Combined forecasts from linear and nonlinear time series models”, *International Journal of Forecasting*, v. 18, pp. 421–438, 2002.
- [46] BUNN, D. W., “Statistical efficiency in the linear combination of forecasts”, *International Journal of Forecasting*, v. 1, pp. 151–163, 1985.
- [47] BUNN, D. W., “Combining forecasts”, *European Journal of Operational Research*, v. 33, pp. 223–229, 1988.
- [48] CLEMEN, R. T., “Combining forecasts: a review and annotated bibliography”, *International Journal of Forecasting*, v. 5, pp. 559–583, 1989.
- [49] MENEZES, L. M., BUNN, D. W., TAYLOR, J. W., “Review of guidelines for the use of combined forecasts”, *European Journal of Operational Research*, v. 120, pp. 190–204, 2000.
- [50] REEVES, G. R., LAWRENCE, K. D., “Combining forecasts given different types of objectives”, *European Journal of Operational Research*, v. 51, pp. 65–72, 1991.
- [51] MAINES, L. A., “An experimental examination of subjective forecast combination”, *International Journal of Forecasting*, v. 12, pp. 223–233, 1996.

- [52] BUNN, D. W., “A Bayesian approach to the linear combination of forecasts”, *Operational Research Quarterly*, v. 26, pp. 325–329, 1975.
- [53] BUNN, D. W., “The synthesis of forecasting models in decision analysis”, *Birkhauser. Basel*, 1978.
- [54] CARPINTEIRO, O. A. S., SILVA, A. P. A., “A hierarchical self-organizing map model in short-term load forecasting”, *Journal of Intelligent and Robotic Systems*, v. 31, pp. 105–113, 2001.
- [55] HONG, W., “Electric load forecasting by support vector model”, *Applied Mathematical Modelling*, v. 33, pp. 2444–2454, 2009.
- [56] GUILHERMINO NETO, G., *Métodos univariados e multivariados para previsão de demanda de energia elétrica em curto prazo: Um estudo comparativo*, Master’s Thesis, Programa de Pós-Graduação em Modelagem Computacional, Faculdade de Engenharia, UFJF, Juiz de Fora, MG, Brasil, 2014.