

Douglas de Oliveira Matos Braga

**Previsão de Cargas Elétricas a Curto Prazo por Combinação de Previsões
via Regressão Simbólica**

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

2017

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

de Oliveira Matos Braga, Douglas.

Previsão de Cargas Elétricas a Curto Prazo por Combinação de Previsões via Regressão Simbólica / Douglas de Oliveira Matos Braga. -- 2017.

72 p.

Orientador: Henrique Steinherz Hippert

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, ICE/Engenharia. Programa de Pós-Graduação em Modelagem Computacional, 2017.

1. Séries Temporais. 2. Previsão de Cargas Elétricas. 3. Combinação de Previsões. 4. Regressão Simbólica. I. Steinherz Hippert, Henrique, orient. II. Título.

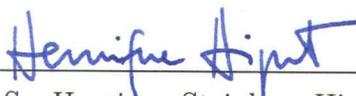
Douglas de Oliveira Matos Braga

Previsão de Cargas Elétricas a Curto Prazo por Combinação de Previsões
via Regressão Simbólica

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 31 de Agosto de 2017.

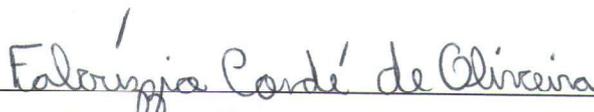
BANCA EXAMINADORA



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



Prof. PhD. Ronaldo Rocha Bastos
Universidade Federal de Juiz de Fora



Prof. D.Sc. Fabrizzio Condé de Oliveira
Universidade Salgado de Oliveira

AGRADECIMENTOS

À minha mãe Marina e à minha família pela presença e apoio constantes aos meus objetivos.

À Marcela pelo incentivo e companheirismo.

Ao Henrique por aceitar me orientar mesmo à distância e pela paciência, disponibilidade e os conselhos que tornaram este trabalho possível.

Aos professores do PPGMC pelos ensinamentos proporcionados.

Aos funcionários do PPGMC pelo excelente suporte dado aos alunos, sempre buscando resolver os problemas da melhor forma.

Aos colegas pelas boas discussões que puderam enriquecer este trabalho.

A todos que contribuíram de alguma forma para este trabalho, meu muito obrigado.

RESUMO

O planejamento energético é base para as tomadas de decisões nas companhias de energia elétrica e, para isto, depende fortemente da disponibilidade de previsões acuradas para as cargas. Devido à inviabilidade de armazenamentos em larga-escala e o custo elevado de compras de energia a curto prazo, além da possibilidade de multas e sanções de órgãos governamentais, previsões em curto prazo são importantes para a otimização da alocação de recursos e da geração de energia.

Neste trabalho utilizamos nove métodos univariados de séries temporais para a previsão de cargas a curto prazo, com horizontes de 1 a 24 horas a frente. Buscando melhorar a acurácia das previsões, propomos um método de combinação de previsões através de *Regressão Simbólica*, que combina de forma não-linear as previsões obtidas pelos nove métodos de séries temporais utilizados. Diferente de outros métodos não-lineares de regressão, a Regressão Simbólica não precisa de uma especificação prévia da forma funcional.

O método proposto é aplicado em uma série real da cidade do Rio de Janeiro (RJ), que contém cargas horárias de 104 semanas dos anos de 1996 e 1997. Comparamos, através de critérios indicados na literatura, os resultados obtidos pelo método proposto com os resultados obtidos por métodos tradicionais de combinação de previsões e ao resultado de simulações de redes neurais artificiais aplicados ao mesmo conjunto de dados. O método proposto obteve melhores resultados, que indicam que a não-linearidade pode ser aspecto importante para combinação de previsões no problema de previsão de carga a curto prazo.

Palavras-chave: Séries Temporais. Previsão de Cargas Elétricas. Combinação de Previsões. Regressão Simbólica.

ABSTRACT

Decision-making in energy companies relies heavily on the availability of accurate load forecasts. Because storing electricity on a large scale is not viable, the cost of short-term energy purchasing is high, and there are government fines and sanctions for failing to supply energy on demand, short-term load forecasts are important for the optimization of resource allocation and energy production.

In this work we used nine univariate time series methods for short-term load forecasts, with forecast horizons ranging from 1 to 24 hours ahead. In order to improve the accuracy of forecasts, we propose a method of combining forecasts through *Symbolic Regression*, which combines in a non-linear way the forecasts obtained by the nine methods of the time series used. Unlike other non-linear regression methods, Symbolic Regression does not need a previous specification of the function structure.

We applied the proposed method to a real time series of the city of Rio de Janeiro (RJ), which contains data on hourly loads of 104 weeks in the years 1996 and 1997. We compare, through the criteria indicated in the literature, the results obtained by the proposed method with the results obtained by traditional methods of forecasts combination and the result obtained by artificial neural networks applied to the same dataset. The method has yielded better results, indicating that non-linearity may be important in combining predictions in short term load forecasts.

Keywords: Time Series. Eletrical Load Forecasting. Forecast Combination. Symbolic Regression.

SUMÁRIO

1	INTRODUÇÃO.....	11
2	MATERIAL E MÉTODOS.....	15
2.1	Séries de Cargas de Energia Elétrica	15
2.2	Métodos de Previsão	18
2.2.1	<i>Previsores Naïve</i>	19
2.2.2	<i>Previsores por Médias Móveis</i>	20
2.2.3	<i>Previsores por Amortecimento Exponencial</i>	20
2.2.3.1	<i>Amortecimento Exponencial Simples (AES)</i>	21
2.2.3.2	<i>Amortecimento Exponencial de Holt</i>	21
2.2.3.3	<i>Holt-Winters</i>	22
2.2.3.4	<i>Holt-Winters com Dupla Sazonalidade</i>	23
2.2.4	<i>Previsores do tipo ARIMA</i>	24
2.3	Combinação de Previsões	27
2.3.1	<i>Métodos de Combinação</i>	29
2.4	Regressão Simbólica	32
2.4.1	<i>Representação dos Indivíduos</i>	33
2.4.2	<i>Inicialização da População</i>	35
2.4.3	<i>Seleção</i>	36
2.4.4	<i>Combinação</i>	38
2.4.5	<i>Mutação</i>	39
2.4.6	<i>Critérios de Parada</i>	41
2.5	Avaliação das previsões	41
2.5.1	<i>Medidas de Erro</i>	41
2.5.2	<i>Avaliação de Combinação de Previsões</i>	42
3	RESULTADOS E DISCUSSÕES	44
3.1	Previsões Individuais	44
3.2	Combinação de Previsões por Regressão Simbólica	45
3.3	Comparação com outros métodos de combinação de previsões	56

<i>3.3.1 Assimetria dos Resíduos de Previsão</i>	59
<i>3.3.2 Autocorrelação dos Resíduos de Previsão</i>	60
4 CONCLUSÕES	66
REFERÊNCIAS	68

LISTA DE ILUSTRAÇÕES

2.1	Série de Cargas com a partição em Amostras de <i>Treino</i> e de <i>Teste</i>	16
2.2	Cargas horárias em duas semanas do verão e duas do inverno.	17
2.3	Série de cargas diárias dos anos de 1996 e 1997 na cidade do Rio de Janeiro. . .	18
2.4	Gráfico da Função de Autocorrelação da série horária de cargas.	19
2.5	Fluxograma básico de um algoritmo de regressão simbólica	34
2.6	Representação de uma função como Árvore Sintática.	35
2.7	Exemplo de construção de um indivíduo pelo método <i>grow</i>	37
2.8	Exemplo de Combinação entre 2 indivíduos	39
2.9	Exemplo de Mutação de um indivíduo	40
3.1	Gráficos das Funções de Autocorrelação dos Resíduos das Previsões Individuais.	45
3.2	Histograma dos MAPEs das soluções para Amostra de Teste	47
3.3	Boxplot dos MAPEs das soluções para Amostra de Teste	48
3.4	Série de Cargas e Previsão via Regressão Simbólica para semanas selecionadas da Amostra de Teste	49
3.5	Resíduos da Previsão da carga na Amostra de Teste pela combinação usando Regressão Simbólica	50
3.6	Histograma dos Resíduos de Previsão na Amostra de Teste	51
3.7	Função de Autocorrelação dos Resíduos de Previsão na Amostra de Teste . . .	52
3.8	Boxplot dos APEs por Dia da Semana	53
3.9	Boxplot dos APEs por Hora do Dia	53
3.10	Boxplot dos APEs por Dia da Semana desconsiderando os dias do Natal de 1997.	54
3.11	Boxplot dos APEs por Horário do Dia desconsiderando os dias do Natal de 1997.	55
3.12	Boxplot dos APEs das Combinações de Previsões	58
3.13	Boxplot dos APEs das Combinações de Previsões desconsiderando os dias do Natal de 1997.	59
3.14	Histogramas dos Resíduos das diferentes Combinações de Previsões	61
3.15	Gráficos das Funções de Autocorrelação dos Resíduos das Combinações de Previsões.	63

3.16 Gráficos das Funções de Autocorrelação dos Resíduos das Combinações de Previsões desconsiderando os dias do Natal de 1997.	64
--	----

LISTA DE TABELAS

2.1	Algoritmo para geração de indivíduos pelo método <i>grow</i> . Adaptado de Poli <i>et al.</i> (2008)	36
3.1	Resultados das previsões individuais na Amostras de Treino II e Teste	44
3.2	Parâmetros utilizados na aplicação da Regressão Simbólica	46
3.3	Parâmetros estimados da forma funcional para as previsões individuais utilizando a Amostra de Treino II	56
3.4	Resultados para Combinações de Previsões na Amostra de Teste considerando ou não a semana do Natal de 1997	57
3.5	Coefficientes Momento de Assimetria dos Resíduos das Combinações de Previsões na Amostra de Teste considerando ou não a semana do Natal de 1997	60
3.6	Estatísticas de Ljung-Box para Autocorrelação dos Resíduos das Combinações de Previsões na Amostra de Teste considerando ou não a semana do Natal de 1997.	65

1 INTRODUÇÃO

A energia elétrica é um insumo essencial na sociedade atual devido à sua aplicabilidade nos mais diversos setores de produção, e também na vida cotidiana. O desenvolvimento tecnológico ocorrido nas últimas décadas foi possibilitado principalmente pela disponibilidade da energia elétrica. No Brasil, segundo a Empresa de Pesquisa Energética (2015), o consumo anual de energia elétrica cresceu de 243074 GWh, em 1995, para 473393 GWh, em 2014, o que representa um aumento de 94%; o setor comercial é o responsável pelo maior aumento proporcional de 32276 GWh para 89819 GWh, ou seja, 178%. Assim, pesquisas sobre a indústria de energia elétrica são relevantes para a sociedade. Esta relevância somada à crescente demanda de energia elétrica, ao seu alto custo de produção e à impossibilidade de estocagem em larga escala, levam à necessidade de um bom planejamento do setor elétrico. Para isto, é essencial que sejam feitas previsões das cargas de energia elétrica em diferentes prazos.

Previsões a longo prazo das cargas de energia elétrica, com horizontes de anos à frente, são importantes para o planejamento do sistema de geração e distribuição de energia elétrica; por exemplo, para a tomada de decisões sobre a construção de usinas, linhas de distribuição ou desativação das mesmas. A médio prazo (semanas ou meses à frente) são necessárias previsões para planejar a compra de insumos para produção, bem como a utilização da capacidade de produção disponível. A curto prazo (algumas horas ou poucos dias à frente), previsões são necessárias para o planejamento da utilização e otimização dos recursos disponíveis para geração, como por exemplo para programar a ativação de turbinas, ou o despacho de carga. Além disso, estas previsões são a base para a tomada de decisões financeiras, pois desde as reestruturações do setor elétrico ocorridas em diversos países através de privatizações e abertura de mercados, a energia elétrica passou a ser negociada como uma *commodity*, aumentando a competitividade dos mercados (Almeida, 2007). A comercialização pode ocorrer através de contratos de longo prazo, ou em operações de mercado *spot*, que possuem um preço mais elevado, definido com base principalmente na carga prevista.

Neste trabalho, trataremos das previsões a curto prazo dos perfis diários de carga, isto é, da previsão da carga a cada hora do dia, com base em informações dos dias

anteriores, o que requer previsões com horizontes de 1 a 24 horas a frente. No contexto do setor energético, tais previsões são importantes para evitar a necessidade de compras de energia no mercado *spot*, minimizar o risco de falhas de abastecimento - que são passíveis de multas e sanções pelo órgão regulador - e ainda evitar o desperdício na geração de energia, devido ao seu alto custo.

Para a previsão de carga de energia elétrica, são geralmente utilizados métodos de séries temporais, uma vez que é conhecido que os valores de carga possuem uma dependência temporal. Além disto, a carga de energia elétrica também é dependente de fatores externos, como temperatura, luminosidade e outros fatores, relacionados principalmente ao clima (Manna *et al.* , 2017). Os métodos que consideram apenas o próprio histórico da carga, são chamados *univariáveis*, enquanto os métodos que consideram as variáveis exógenas são chamados de *multivariáveis* (Neto *et al.* , 2015).

Entre os métodos univariáveis utilizados para a previsão de carga, destacam-se aplicações de métodos de base estatística e de inteligência computacional. Os métodos estatísticos de séries temporais utilizados incluem modelos ARIMA (Pappas *et al.* , 2010) e métodos de amortecimento exponencial, com destaque para o método de Holt-Winters com dupla sazonalidade (Taylor & McSharry, 2007). Dentre os métodos de inteligência computacional, é comum o uso de Redes Neurais Artificiais (RNA), como em Zhang *et al.* (2017a), que utilizam Redes Neurais Recorrentes e uma estratégia de múltiplas séries temporais com diferentes resoluções para modelar as complexas relações a curto e longo prazo das cargas, além de suas interações. Outra estratégia de previsão para séries univariadas, proposta por Alvarez *et al.* (2011), é baseada na identificação de sequências de padrões das séries através dos algoritmos de *Pattern Sequence-based Forecast*.

Para previsões com variáveis exógenas é comum o uso de modelos multivariáveis de séries temporais como ARIMAX e SARIMAX (Tarsitano & Amerise, 2017), e métodos de inteligência computacional, com destaque para as Redes Neurais Artificiais, que possuem extensa aplicação para previsão de cargas com uso de covariáveis (Hippert *et al.* , 2001). Exemplos de aplicações de RNA são em previsões de cargas de sistemas utilizando covariáveis climáticas (Hippert *et al.* , 2005) e em previsões de cargas de domicílios, utilizando informações de área construída e número de ocupantes (Rodrigues *et al.* , 2017).

Para ambos os métodos de previsão, estão sendo amplamente utilizadas estratégias

híbridas de previsão, que consistem em criar novos métodos de previsão através da junção de diferentes métodos de previsão e/ou heurísticas. Combinações entre heurísticas e métodos de inteligência computacional são encontradas para previsões univariáveis, como López *et al.* (2017), que utilizam redes neurais de ondaletas com otimização por enxame de partículas (PSO) para a decomposição da série em funções intrínsecas e identificação das relevantes para a previsão; e para previsões multivariadas, como o trabalho de Zhang *et al.* (2017b) que utiliza *Extreme Learning Machines* - um tipo específico de rede neural - com a estimação dos parâmetros sendo feita através de heurísticas de busca gravitacional básica com otimização por enxame de partículas. Mayrink & Hippert (2016) propõem um híbrido entre métodos de séries temporais e inteligência computacional que consiste em corrigir os erros de uma previsão por Holt-Winters com dupla sazonalidade através do algoritmo de *Gradient Boosting*, um *ensemble* de árvores de decisão. Há, ainda, estratégias híbridas que combinam previsões por sequências de padrões com outros métodos, como Dudek (2016) que utiliza regressões lineares baseadas em sequências de padrões para as previsões; e Alkhatib *et al.* (2017) propõem um método híbrido para previsões de *grids* através de técnicas de agrupamentos de edificações com previsões por sequências de padrões.

Para a previsão de cargas, encontramos usos da Regressão Simbólica tanto como método multivariável de previsão (Ghareeb & El Saadany, 2013; Castelli *et al.*, 2015), como parte de método híbrido de previsão (Lee & Tong, 2011a,b), onde é utilizada para modelagem dos resíduos de previsão obtidos por métodos univariáveis de previsão.

De acordo com Taylor *et al.* (2006) para previsões a curto prazo, métodos univariáveis são suficientemente acurados e devem ser utilizados, especialmente quando o acesso às variáveis climáticas é caro e difícil, como ocorre no Brasil. Por esta razão, neste trabalho utilizaremos apenas previsões através de métodos univariáveis.

Diferentes métodos de previsão univariáveis possuem diferentes limitações em captar as estruturas temporais presentes no dados. Neste trabalho utilizaremos alguns métodos de previsão univariáveis para a previsão de uma série de cargas horárias da cidade do Rio de Janeiro e posteriormente faremos combinações, buscando mitigar suas limitações. Estudaremos diferentes métodos de combinação de previsões apresentados na literatura, avaliando-os através de medidas de erro de previsão e da análise da distribuição dos erros.

Particularmente, iremos propor uma nova aplicação para a Regressão Simbólica no

problema de previsão de cargas. Ela será utilizada como um método não linear de combinação de previsões devido à sua propriedade de ser uma meta-heurística que não depende de uma forma funcional pré-estabelecida. Avaliaremos sua capacidade de gerar boas previsões, e de descobrir boas formas funcionais para combinação de previsões.

Os resultados obtidos serão comparados com os de Hippert (2001) e Hippert *et al.* (2005), que utilizam métodos multivariáveis (Redes Neurais Artificiais) para o mesmo conjunto de dados, considerando a temperatura como variável exógena. Também comparamos nossos resultados com os de Carneiro (2014), que aplica um método *bayesiano* para combinar os mesmos métodos de previsão univariáveis utilizados neste trabalho. Uma importante diferença é que, naquele trabalho as previsões foram combinadas em pares, enquanto que neste são todas combinadas simultaneamente, o que nos permite avaliar a capacidade do método de combinação em identificar previsões não relevantes.

2 MATERIAL E MÉTODOS

2.1 Séries de Cargas de Energia Elétrica

Para a aplicação e avaliação dos métodos de previsão discutidos neste trabalho, utilizamos uma série temporal de cargas horárias de energia elétrica na cidade do Rio de Janeiro dos anos de 1996 e 1997. Esta série contém 17472 observações, correspondentes a um período de 104 semanas. Antes de utilizarmos esta série nas aplicações, ela passou por uma etapa de pré-processamento, na qual fizemos a crítica dos dados e a imputação de cargas de feriados e datas especiais. As cargas em tais datas possuem um perfil diferenciado das demais datas e necessitam de modelos específicos para sua previsão (Arora & Taylor, 2013). Para imputar estas cargas, usamos a média entre os valores da série para a mesma hora e mesmo dia da semana anterior e seguinte.

Diferentemente do que acontece em outras classes de problemas de previsão, na previsão de séries temporais a ordenação dos dados é relevante para a aplicação, pois deseja-se que o método seja capaz de prever o futuro utilizando dados apenas do passado. Devido a isto, a partição da série em amostras de *treino* e de *teste* para validação dos métodos deve ser sequencial. Para permitir que os resultados obtidos sejam comparados com os encontrados nos trabalhos de Hippert (2001), Hippert *et al.* (2005) e Carneiro (2014), adotamos a seguinte partição para a série de cargas

- Semanas de 1 a 54: Amostra de Treino I;
- Semanas de 55 a 84: Amostra de Treino II;
- Semanas de 85 a 104: Amostra de Teste,

divisão esta, ilustrada na figura 2.1.

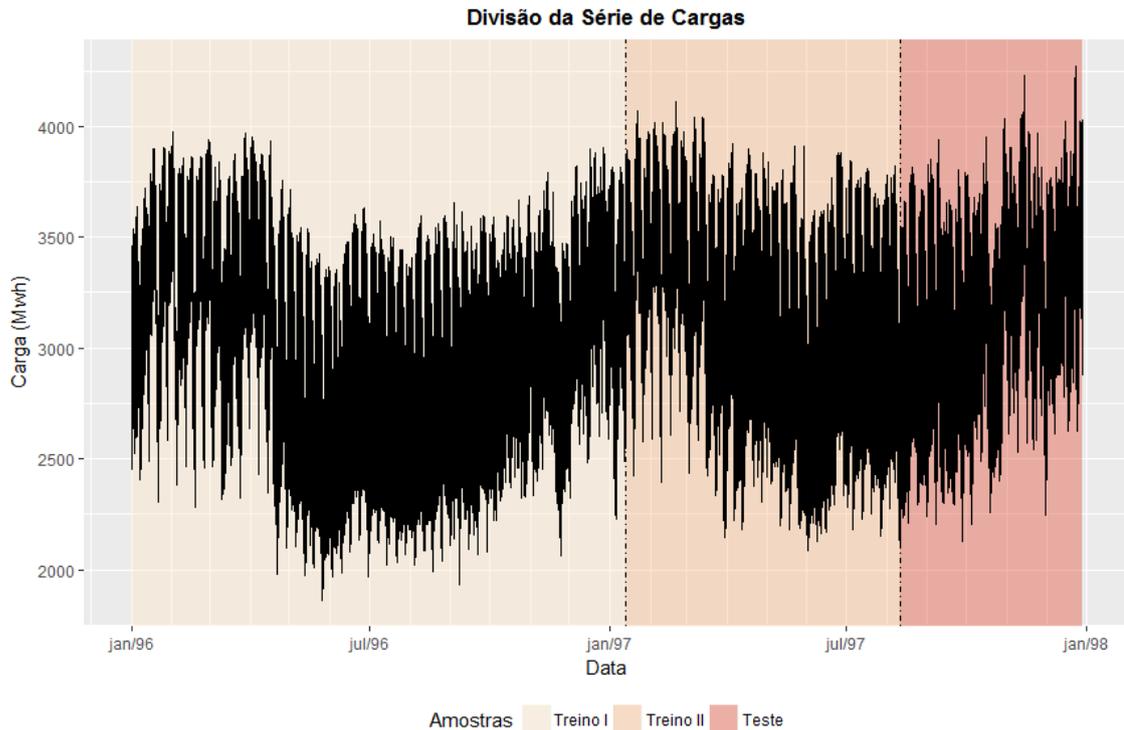


Figura 2.1: Série de Cargas com a partição em Amostras de *Treino* e de *Teste*.

Em geral, séries de cargas elétricas possuem uma parte do seu comportamento que se repete periodicamente, este comportamento é chamado de sazonalidade. Nesta série, é possível observar três padrões de sazonalidade: Diária, Semanal e Anual.

Dois destes padrões sazonais podem ser observados na Figura 2.2, que apresenta as cargas horárias de duas semanas do verão e de duas semanas do inverno. Em cada um destes gráficos, há um comportamento que se repete todos os dias, a cada 24 horas, com diferenças nos sábados e domingos. Apesar disto acontecer em ambas as estações, os perfis diários se diferem - no verão há dois picos diários, enquanto no inverno há apenas um.

A sazonalidade semanal é caracterizada por média de cargas mais altas de segunda a sexta com quedas no sábado e no domingo; este comportamento está presente em ambas as estações.

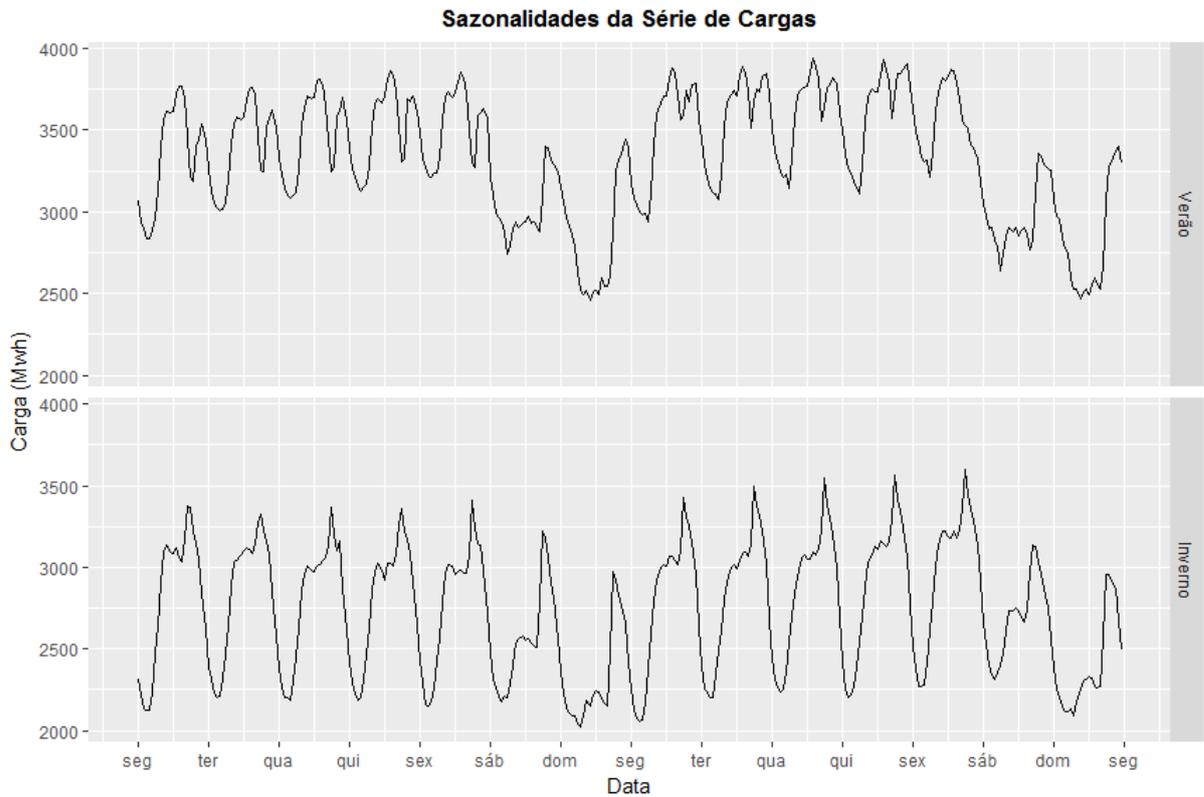


Figura 2.2: Cargas horárias em duas semanas do verão e duas do inverno.

O componente sazonal anual pode ser visto na Figura 2.3, que mostra a carga diária total nos anos de 1996 e 1997. As cargas são mais altas nos meses de verão do que nos de inverno, devido às altas temperaturas que exigem uso de aparelhos de ar-condicionado. Esta sazonalidade não será modelada neste trabalho, pois seria necessário uma série mais longa para a estimação de modelos para os perfis anuais, e sua validação.

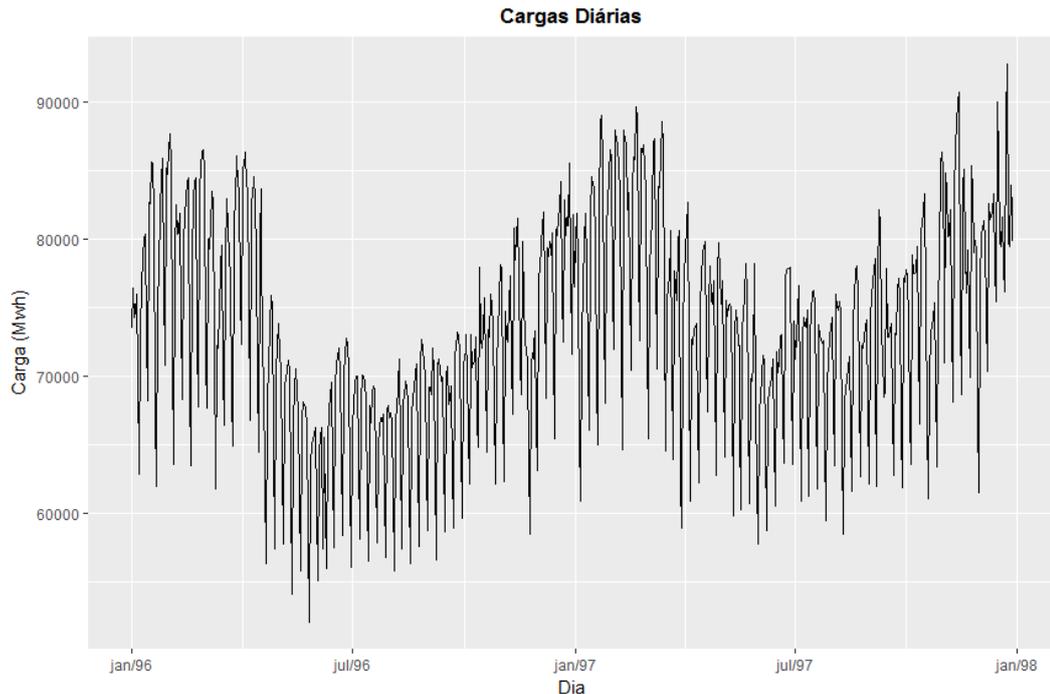


Figura 2.3: Série de cargas diárias dos anos de 1996 e 1997 na cidade do Rio de Janeiro.

2.2 Métodos de Previsão

As duas sazonalidades apresentadas podem ser modeladas a partir de duas estratégias distintas. Uma estratégia, que chamaremos de “previsão sequencial”, é a utilização de modelos que descrevam simultaneamente as duas componentes sazonais. Em geral, modelos deste tipo possuem uma maior complexidade em seus métodos de estimação.

Outra estratégia é a utilização de modelos em paralelo, que descrevam apenas uma componente sazonal de cada vez. Podem ser utilizados 24 modelos em paralelo, um para cada hora do dia, que contemplem apenas a sazonalidade semanal; ou 168 modelos em paralelo, um para cada hora da semana, que não considerem componentes sazonais. Tais estratégias podem ser justificadas pelo gráfico da Função de Autocorrelação da Série (Figura 2.4), que apresenta picos para defasagens de 168 e 24 horas.

Neste trabalho, para a previsão da carga de energia elétrica, utilizamos métodos univariáveis, nos quais a previsão da carga \hat{L}_t para um instante t depende apenas das informações da própria carga em instantes anteriores (L_{t-1}, \dots, L_{t-k}). Além disto, foram utilizadas previsões de *perfis diários*, isto é, previsões das 24 horas de um dia realizadas com informações dos dias anteriores, o que exigiu previsões com horizonte variando de uma a 24 horas à frente. A seguir são apresentados os métodos de previsão individuais

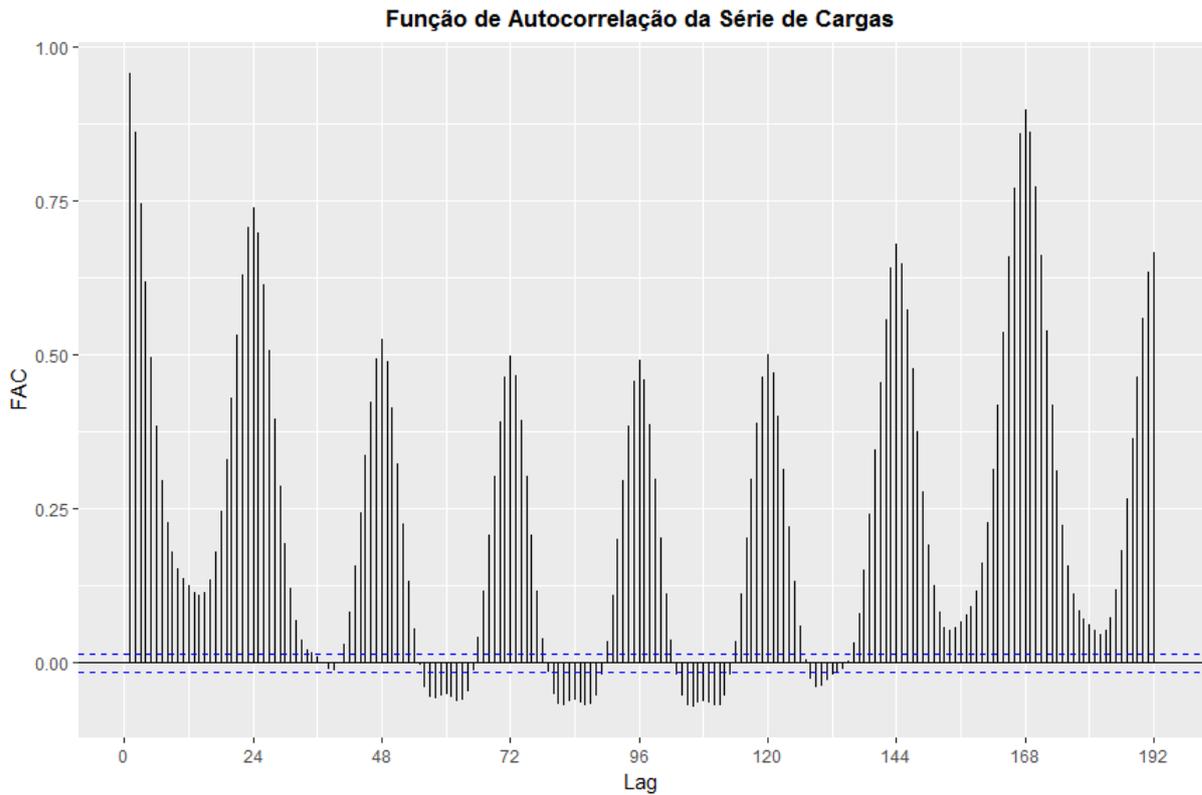


Figura 2.4: Gráfico da Função de Autocorrelação da série horária de cargas.

utilizados neste trabalho para ambas as estratégias de modelagem.

2.2.1 *Previsores Naïve*

O método *Naïve* ("ingênuo") é um método de previsão simples, utilizado para estimar o nível de modelos sem tendências ou sazonalidade, representados por

$$Z_t = \mu + \epsilon_t. \quad (2.1)$$

Neste tipo de modelo, a previsão para um dado instante é a estimativa do nível da série. No método *Naïve*, esta estimativa é dada pela observação feita no momento imediatamente anterior, ou seja, $\hat{\mu}_t = Z_{t-1}$. Neste trabalho, usamos o método *Naïve* para 24 e 168 modelos em paralelo. As equações de previsão são, para cada hora do dia nos 24 modelos em paralelo:

$$\hat{L}_t = L_{t-24} \quad (2.2)$$

e, para cada hora da semana, utilizando 168 modelos em paralelo:

$$\hat{L}_t = L_{t-168}. \quad (2.3)$$

Por não considerar a sazonalidade semanal, o *Naïve* com 24 previsores em paralelo apresentará grandes erros nas previsões de segunda-feira, por exemplo, pois seus valores serão calculados com base nos de domingo, que têm um perfil muito diferente.

Devido a sua simplicidade, estes métodos são usualmente utilizados como *benchmarks* para avaliar a qualidade das previsões feitas por outros métodos.

2.2.2 *Previsores por Médias Móveis*

O método de Médias Móveis (MM) é utilizado para a estimar o nível do modelo representado em (2.1). A previsão, assim como no método *Naïve*, é dada pela estimativa do nível da série. A diferença entre dois métodos está no número de observações utilizadas para a estimativa deste nível - enquanto o *Naïve* utiliza apenas a última observação, o método de Médias Móveis utiliza a média das K últimas observações. Sendo assim, o predictor *Naïve* pode ser considerado um predictor de Médias Móveis de $K = 1$.

Neste trabalho, utilizamos o método de Médias Móveis apenas para a previsão com 168 modelos em paralelo, devido a não possibilidade de inclusão de sazonalidade. A definição do valor K (chamado de "janela") utilizado foi feita através de busca em grade, entre os valores 2 e 10, do K que minimiza a função de custo MAPE na amostra de estimação.

A previsão feita a partir deste método para os 168 modelos em paralelo segue a equação

$$\hat{L}_t = \frac{1}{K} \sum_{i=1}^K L_{t-168i}. \quad (2.4)$$

2.2.3 *Previsores por Amortecimento Exponencial*

O método de Médias Móveis utiliza as últimas K observações da série para a previsão, todas com o mesmo peso. Os métodos baseados em Amortecimento Exponencial possuem em sua fundamentação a ideia de que previsões mais recentes devem ser mais importantes para as previsões do que as observações mais antigas, sendo portanto necessário empregarmos uma ponderação decrescente. Esta ponderação é feita através de potências sucessivas de uma ou mais constantes de valores entre 0 e 1, que decaem

para as observações mais antigas.

São apresentados a seguir quatro previsores baseados na ideia de Amortecimento Exponencial, que se diferenciam pelo modelo utilizado. A estimação das constantes de amortecimento dos métodos apresentados foram feitas através de função disponível no pacote *forecast* (Hyndman, 2016; Hyndman & Khandakar, 2008).

2.2.3.1 Amortecimento Exponencial Simples (AES)

O Amortecimento Exponencial Simples é um método utilizado para previsões seguindo o modelo de nível constante apresentado em (2.1). A estimação do nível da série é feita utilizando uma média ponderada com pesos que decaem de forma exponencial de acordo com uma constante $0 < \alpha \leq 1$ de tal forma que

$$\hat{Z}_{t+1} = M_t = \alpha Z_t + \alpha(1-\alpha)Z_{t-1} + \dots + \alpha(1-\alpha)^i Z_{t-i} + \alpha(1-\alpha)^t Z_0 = \sum_{i=0}^t \alpha(1-\alpha)^i Z_{t-i}. \quad (2.5)$$

É possível também construir uma equação recursiva de previsão que é mais facilmente computável

$$\hat{Z}_{t+1} = M_t = \alpha Z_t + (1 - \alpha)M_{t-1} = \alpha Z_t(1 - \alpha) + \hat{Z}_t \quad (2.6)$$

Como este método não considera a sazonalidade dos dados, foi utilizado apenas para os 168 modelos em paralelo. A equação de previsão neste caso é dada por

$$\hat{L}_t = \alpha_j L_{t-168} + (1 - \alpha_j)\hat{L}_{t-168} \quad \forall j \in 1, \dots, 168. \quad (2.7)$$

Estimamos uma constante α_j diferente para cada um dos 168 modelos.

2.2.3.2 Amortecimento Exponencial de Holt

O Método de Amortecimento Exponencial de Holt é utilizado para modelos que possuem tendência linear. Tais modelos podem ser descritos por

$$Z_t = a + bt + \epsilon_t. \quad (2.8)$$

O Método de Holt utiliza duas equações de para estimar o nível a e a declividade b da

série:

$$\begin{aligned}\hat{a}_t &= \alpha Z_t + (1 - \alpha)(\hat{a}_{t-1} - \hat{b}_{t-1}), \\ \hat{b}_t &= \beta(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta)\hat{b}_{t-1},\end{aligned}\tag{2.9}$$

com $0 < \alpha, \beta < 1$. Assim como o AES, este método de previsão foi utilizado para os 168 modelos em paralelo por não considerar o comportamento sazonal em seu modelo. A equação de previsão da carga utilizada para o Método de Holt foi

$$\hat{L}_t = \hat{a}_{t-168} + \hat{b}_{t-168},\tag{2.10}$$

com

$$\begin{aligned}\hat{a}_t &= \alpha_j L_t + (1 - \alpha_j)(\hat{a}_{t-168} - \hat{b}_{t-168}), \\ \hat{b}_t &= \beta_j(\hat{a}_t - \hat{a}_{t-168}) + (1 - \beta_j)\hat{b}_{t-168} \quad \forall j \in 1, \dots, 168.\end{aligned}\tag{2.11}$$

As constantes $\{\alpha_j, \beta_j\}$ foram estimadas separadamente para cada um dos 168 modelos utilizados.

2.2.3.3 Holt-Winters

Os métodos de Amortecimento Exponencial apresentados anteriormente são válidos para modelos que não possuam estrutura sazonal. Como visto, os dados de carga horária de energia elétrica possuem comportamento sazonal. O Método de Holt-Winters é utilizado para modelos lineares com sazonalidade multiplicativa da forma

$$Z_t = (a + bt)\rho^{d(t)} + \epsilon_t,\tag{2.12}$$

onde $\rho^{d(t)}$ é um fator sazonal associado ao instante t . As estimativas dos parâmetros do modelo são feitas por três equações:

$$\begin{aligned}\hat{a}_t &= \alpha \frac{Z_t}{\hat{\rho}_{t-1}^{d(t)}} + (1 - \alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}), \\ \hat{b}_t &= \beta(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta)\hat{b}_{t-1}, \\ \hat{\rho}_t^{d(t)} &= \gamma \frac{Z_t}{\hat{a}_t} + (1 - \gamma)\hat{\rho}_{t-1}^{d(t)}\end{aligned}\tag{2.13}$$

Como este método é utilizado para modelos que possuem apenas uma periodicidade sazonal, ele foi utilizado para a estratégia de 24 modelos paralelos, um para cada hora do dia. Serão então utilizados sete fatores sazonais (um por dia da semana) em cada modelo. A equação de previsão utilizada a partir deste método é

$$\hat{L}_t = (\hat{a}_{t-24} + \hat{b}_{t-24})\hat{\rho}_{t-24}^{d(t)}, \quad (2.14)$$

onde

$$\begin{aligned} \hat{a}_t &= \alpha_j \frac{L_t}{\hat{\rho}_{t-24}^{d(t)}} + (1 - \alpha_j)(\hat{a}_{t-24} + \hat{b}_{t-24}), \\ \hat{b}_t &= \beta_j(\hat{a}_t - \hat{a}_{t-24}) + (1 - \beta_j)\hat{b}_{t-24}, \\ \hat{\rho}_t^{d(t)} &= \gamma_j \frac{L_t}{\hat{a}_t} + (1 - \gamma_j)\hat{\rho}_{t-24}^{d(t)} \end{aligned} \quad (2.15)$$

sendo $d(t)$ o dia da semana referente ao instante t . Os fatores $\hat{\rho}_t^{d(t)}$ precisam de ser normalizados a cada semana, de forma que $\sum_{i=1}^7 \hat{\rho}_t^{d(i)} = 7$. As constantes $\{\alpha_j, \beta_j, \gamma_j\}$ foram estimadas separadamente para cada um dos 24 modelos utilizando o método de *Limited Memory BGGs* de Byrd *et al.* (1995) implementado na função *optim* do software R.

2.2.3.4 Holt-Winters com Dupla Sazonalidade

Para modelar simultaneamente as duas sazonalidades, Taylor (2003) propôs uma expansão do método de Holt-Winters, utilizando dupla sazonalidade multiplicativa através da inclusão de mais um fator sazonal em (2.12). Posteriormente, Taylor (2010) propõe um método com dupla sazonalidade aditiva para a previsão de cargas a curto prazo, denominado método de Holt-Winters com dupla sazonalidade (DSHW), que utilizamos neste trabalho.

Como este método permite a modelagem de duas sazonalidades ao mesmo tempo, foi utilizado para as previsões sequenciais. A equação de previsão deste método é dada por

$$\hat{L}_{t+K} = \hat{a}_t + \hat{\rho}_t^{d(t+K)} + \hat{\psi}_t^{h(t+K)} + \phi^K(L_t - (\hat{a}_{t-1} + \hat{\rho}_{t-1}^{d(t)} + \hat{\psi}_{t-1}^{h(t)})) \quad K = 1, \dots, 24, \quad (2.16)$$

onde $\rho^{d(t)}$ é o fator sazonal associado ao dia da semana, e $\psi^{h(t)}$ é o fator sazonal associado

à hora do dia, ambos correspondentes ao instante t . A constante ϕ^K é um refinamento do método, utilizado para correção da autocorrelação de primeira ordem dos erros. A variação de K ocorre de 1 a 24, pois as previsões são feitas até 24 horas a frente.

As equações utilizadas para a estimativa do nível a_t e dos fatores sazonais são

$$\begin{aligned}\hat{a}_t &= \alpha(L_t - \hat{\rho}_{t-1}^{d(t)} - \hat{\psi}_{t-1}^{h(t)} + (1 - \alpha)\hat{a}_{t-1}; \\ \hat{\rho}_t^{d(t)} &= \gamma(L_t - \hat{a}_{t-1} - \hat{\psi}_{t-1}^{h(t)}) + (1 - \gamma)\hat{\rho}_{t-1}^{d(t)}; \\ \hat{\psi}_t^{h(t)} &= \omega(L_t - \hat{a}_{t-1} - \hat{\rho}_{t-1}^{d(t)}) + (1 - \omega)\hat{\psi}_{t-1}^{h(t)},\end{aligned}\tag{2.17}$$

e as constantes $\{\alpha, \gamma, \omega, \phi\}$ foram estimadas através de otimização da soma dos quadrados dos resíduos na Amostra de Treino I.

2.2.4 Previsores do tipo ARIMA

Os modelos ARIMA (Box & Jenkins, 1994) são uma classe de modelos estatísticos para modelagem e previsão de séries temporais - o nome é um acrônimo de *autoregressive integrated moving average* -. Esta classe de modelos se baseia na teoria de filtros lineares e define uma série temporal como a realização de um processo estocástico.

Para um conjunto T de índices, um processo estocástico é uma família de variáveis aleatórias $Z = Z_t, t \in T$, onde Z_t são os estados do processo; o conjunto destes estados é o espaço de estados. Se T é enumerável, o processo estocástico é considerado de tempo discreto. Caso contrário, o processo é dito ser de tempo contínuo.

Ao utilizar modelos ARIMA para modelagem de séries temporais, considera-se que cada observação da série em um instante t possui um componente aleatório de erro ϵ_t , e que estes erros ϵ_t da série são independentes e identicamente distribuídos com distribuição Normal, de média 0 e variância σ^2 constantes.

Antes de utilizarmos os modelos ARIMA, é preciso verificar a *estacionariedade* do processo. Um processo é dito estacionário caso $E[Z_t] = \mu_Z \forall t$, $E[Z_t^2] = C < \infty \forall t$ e $Cov[Z_t, Z_{t+k}]$ dependa apenas de k para todo $t \in T$ (Morettin & Toloi, 2006). Essas três condições indicam que os momentos de primeira e segunda ordem do processo são invariantes no tempo. Processos ditos estacionários são modelados por uma sub-classe de modelos ARIMA, os modelos ARMA; processos não estacionários são modelados por modelos ARIMA, que incluem uma etapa de diferenciação que busca tornar estes processos

estacionários.

Os modelos ARMA modelam uma série temporal Z_t que possua dependência temporal como função de uma série de choques a_t , independentes e identicamente distribuídos, com distribuição $Normal(0, \sigma^2)$ (esta série de choques também é chamada de *ruído branco*). As funções usadas são filtros lineares, que transformam a série de ruídos brancos a_t na série Z_t :

$$a_t \rightarrow [FT] \rightarrow Z_t. \quad (2.18)$$

Um filtro é invertido se é capaz de retirar a informação da série, transformando-a em um ruído branco:

$$a_t \leftarrow [FT'] \leftarrow Z_t, \quad (2.19)$$

Seja B um *operador de retardo* tal que $B^k Z_t = Z_{t-k}$, Podemos então escrever um filtro linear utilizando B para o valor centrado \bar{Z}_t , onde $\bar{Z}_t = Z_t - E[Z]$, na forma:

$$\bar{Z}_t = (1 + \psi_1 B + \psi_2 B^2 + \dots) a_t = \mathbf{\Psi}(B) a_t, \quad (2.20)$$

e o filtro linear invertido

$$\mathbf{\Psi}^{-1}(B) a_t = \mathbf{\Pi}(B) a_t. \quad (2.21)$$

O processo estocástico será invertível se existe $\mathbf{\Pi}(B)$ que torne a equação 2.21 verdadeira. As condições de estacionariedade e invertibilidade do filtro estão relacionadas a seus parâmetros ψ e π . Um filtro é estacionário se $\sum_i |\psi_i| < \infty$ e invertível se $\sum_i |\pi_i| < \infty$.

Um processo escrito da forma

$$(1 - \phi B - \dots - \phi_p B^p) \bar{Z}_t = \mathbf{\Phi}(B) \bar{Z}_t = a_t \quad (2.22)$$

é dito um processo auto-regressivo de ordem p , podendo ser modelado por um filtro linear AR(p). Pela definição apresentada, este filtro já é um filtro invertido.

Um processo de médias móveis de ordem q pode ser escrito como

$$\bar{Z}_t = (1 - \theta B - \dots - \theta_q B^q) a_t = \mathbf{\Theta}(B) a_t, \quad (2.23)$$

sendo modelado por um filtro MA(q), que será sempre estacionário.

A combinação destes dois filtros permite a escrita de um processo auto-regressivo de

médias móveis ARMA(p, q) tal que

$$\Phi(B)\bar{Z}_t = \Theta(B)a_t. \quad (2.24)$$

Os modelos ARMA dependem do pressuposto de estacionariedade da série. Para séries não estacionárias, é necessária a etapa de diferenciação a fim de se obter uma série estacionária. O operador de diferenciação ∇ é o análogo, em domínios discretos, à operação de derivação nos domínios contínuos. Ele pode ser escrito em função do operador de retardo B , tal que $\nabla = 1 - B$. O operador de ordem d é definido recursivamente como

$$\nabla^d Z_t = \nabla(\nabla^{d-1} Z_t). \quad (2.25)$$

Com a utilização do operador de diferenciação, um modelo ARIMA(p, d, q) tem como forma

$$\Phi(B)\nabla^d \bar{Z}_t = \Theta(B)a_t. \quad (2.26)$$

Modelos ARIMA como os representados pela equação 2.26 não são capazes de modelar componentes sazonais dos dados, por isto foram utilizados para a estratégia de previsão de 168 modelos em paralelo.

Com uma modificação no operador de diferenciação, é possível modelar um componente sazonal da série temporal através de modelos do tipo ARIMA. Para isto, é necessário o operador de diferenciação com defasamento s , tal que $\nabla_s = 1 - B^s$. Modelos que utilizam este operador para modelar a sazonalidade são chamados de SARIMA.

Com o operador de diferenciação com defasamento, é possível diferenciar uma série de sazonalidade com período S e obter um modelo SARIMA(P, Q, D) $_S$, dado por

$$\Phi(B^S)\nabla_S^D \bar{Z}_t = \Theta(B^S)a_t. \quad (2.27)$$

Este modelo, porém, modela apenas a relação do instante t com os instantes $t - kS$, não permitindo modelar a relação de t com os instantes imediatamente anteriores. Isto acaba gerando uma autocorrelação nos erros. Para corrigir isto, uma solução é ajustar um modelo ARIMA(p, d, q) aos erros, obtemos então um modelo SARIMA(p, d, q)(P, D, Q) $_S$ da forma

$$\phi(B)\Phi(B^S)\nabla_S^D \nabla^d \bar{Z}_t = \theta(B)\Theta(B^S)a_t. \quad (2.28)$$

Devido à capacidade do modelo SARIMA de modelar um componente sazonal, ele foi utilizado para 24 modelos em paralelo, em que cada modelo SARIMA(p, d, q)(P, D, Q)₇ é capaz de modelar a sazonalidade semanal para uma determinada hora do dia.

Para a previsão através dos modelos ARIMA e SARIMA, foi utilizado o pacote *forecast* (Hyndman, 2016; Hyndman & Khandakar, 2008) para a determinação da ordem dos modelos, estimação dos parâmetros e previsão. A determinação da ordem dos modelos é feita através do Critério de Informação de Akaike (AIC), que considera a qualidade do ajuste do modelo e o penaliza por sua complexidade (medida pelo número de parâmetros). O AIC de um modelo é calculado através de

$$AIC = 2k - 2\ln(L), \quad (2.29)$$

onde k é o número de parâmetros estimados no modelo e L é o valor máximo da função de verossimilhança do modelo. Segundo este critério, o melhor modelo é o que apresenta menor AIC, ou seja, menor complexidade e maior verossimilhança.

2.3 Combinação de Previsões

Devido à grande diversidade de modelos disponíveis para previsão de séries temporais, com distintas estruturas e características de relação entre a predição e os preditores, e à complementaridade e similaridade entre eles, a escolha de qual método utilizar se torna crucial para a qualidade da previsão. Para a previsão de carga em curto prazo, essa situação é agravada pela possibilidade de utilização de estratégias tanto de modelos sequenciais quanto dos paralelos, aumentando assim a gama de previsores possíveis.

Comumente utilizados para escolhas de modelos em diversas áreas da Estatística, os teste de hipóteses enfrentam muitas dificuldades quando utilizados em séries temporais (Zou & Yang, 2004). A realização de múltiplos testes sequenciais dificulta o cálculo das probabilidades dos erros associados ao processo e muitos métodos de previsão não atendem a pressupostos geralmente requeridos para os testes. Além disto, a escolha de modelos através de testes de hipóteses não garante uma melhor performance quanto ao risco de previsão.

Como uma alternativa aos testes de hipóteses, são geralmente utilizadas para seleção de modelos algumas medidas de erro, como Erro Quadrático Médio (*Mean Square*

Error, MSE) e Erro Absoluto Percentual Médio (*Mean Absolute Percent Error*, MAPE), apresentados na Seção 2.5.1, além de critérios de seleção de modelos como o *Akaike Information Criterion* (AIC) e o *Bayesian Information Criterion* (BIC), que levam em consideração não apenas a qualidade do ajuste do modelo, mas também a sua complexidade. Estas métricas são utilizadas para comparações entre modelos e, em geral, é escolhido o modelo que apresenta o menor valor. Um problema deste método de escolha é sua instabilidade, já que a diferença entre os valores das métricas para diferentes modelos costuma ser pequena, especialmente quando o número de observações é de moderado a pequeno. Desta forma, uma pequena mudança nos dados pode levar a escolha de um modelo diferente. Este fato é chamado "Incerteza de Modelo" (Chatfield, 1996).

Além disto, na tentativa de se obter a melhor previsão possível, usualmente a previsão feita pelo modelo considerado o melhor (independente dos critérios), é aceita e as outras previsões são descartadas. Segundo Bates & Granger (1969), este procedimento pode não ser o mais sábio, pois os modelos descartados costumam conter informações úteis para a previsão que são independentes das contidas no modelo aceito. Esta perda de informação pode ocorrer de duas formas:

- A previsão descartada utiliza variáveis e/ou informações que não são consideradas na previsão aceita;
- Os modelos fazem suposições diferentes sobre a relação entre as variáveis.

Uma método proposto por Reid (1968, 1969) e Bates & Granger (1969) para se evitar os problemas da instabilidade da escolha dos modelos e da perda de informação é a combinação de previsões, que gera uma nova previsão com base nos valores das previsões obtidas por diferentes métodos ou modelos. Devido à sua estratégia de combinar informações advindas de diferentes modelos, a combinação de previsões também pode ser vista como um método de *ensemble* (Dietterich, 2000), técnica amplamente difundida no aprendizado de máquinas.

Clemen (1989) faz uma ampla revisão da bibliografia até a data e afirma que os resultados encontrados corroboram a ideia de que a combinação de previsões leva a uma maior acurácia da previsão e que, além disto, em muitos casos uma simples média aritmética já produz melhoras na performance das previsões.

2.3.1 Métodos de Combinação

A seguir são apresentados alguns métodos de combinação das previsões, discutidos por Bates & Granger (1969) e De Menezes *et al.* (2000).

Seja $\mathbf{f} = f_1, f_2, \dots, f_P$ um conjunto de P previsões distintas para os tempos $t = 1, \dots, T$ e $\Psi(\mathbf{f})$ a função que represente a combinação dessas previsões.

- **Média Aritmética Simples:** Um dos métodos mais simples de combinação, consiste em utilizar a média aritmética simples das previsões em cada instante de tempo. É amplamente utilizado como método de combinação devido à sua simplicidade e, segundo Armstrong (1989), em geral já possui performance melhor que as previsões individuais. A previsão combinada no instante t é dada por

$$\Psi_{MAS}(\mathbf{f}_t) = \frac{1}{P} \sum_{i=1}^P f_{i,t}; \quad (2.30)$$

- **Mediana:** A mediana também é um método de simplicidade conceitual, onde para cada instante do tempo, a combinação das previsões é dada pela mediana das previsões para aquele instante. É um método mais robusto que o da média, por não ser influenciado por valores extremos das previsões, mas há a necessidade que se ordene as previsões em cada instante de tempo. A previsão em um instante de tempo t é dada por

$$\Psi_{MD}(\mathbf{f}_t) = \begin{cases} f_{(\frac{P+1}{2}),t}, & \text{se } P \text{ é ímpar} \\ (f_{(\frac{P}{2}),t} + f_{(\frac{P+1}{2}),t}) & \text{se } P \text{ é par,} \end{cases} \quad (2.31)$$

Onde $f_{(i),t}$ representa o i -ésimo maior valor das previsões no instante t ;

- **Média Podada:** A combinação por média podada é uma versão da combinação por média aritmética simples, mais robusta a valores de previsões discrepantes. Este método de combinação consiste em calcular a média aritmética desconsiderando as J previsões mais baixas e as J previsões mais altas para um instante t . Também há a necessidade de ordenação das previsões, como no método da mediana. A

combinação é dada por

$$\Psi_{MP}(\mathbf{f}_t) = \frac{1}{P - 2J} \sum_{i=J+1}^{P-J} f_{(i),t}. \quad (2.32)$$

Neste trabalho foi utilizado $J = 2$ para as combinações de 10 previsões, resultando na média das 6 previsões centrais em cada instante de tempo;

- **Média dos Extremos:** Ao contrário do método da média podada, esse método é altamente influenciado por previsões discrepantes, e pode ser útil para prever comportamentos voláteis. Consiste na média entre a maior e a menor previsão em cada instante de tempo

$$\Psi_{MEX}(\mathbf{f}_t) = \frac{f_{(1),t} + f_{(P),t}}{2} \quad (2.33)$$

- **Média Ponderada Ótima:** é uma média aritmética ponderada, onde os pesos das previsões são definidos de forma a minimizar a soma dos quadrados dos erros de previsão. A combinação das previsões é dada por

$$\Psi_{MPO}(\mathbf{f}) = \sum_{i=1}^P \omega_i f_i; \quad (2.34)$$

Este método utiliza a matriz de covariância S dos resíduos das previsões individuais para estimar o vetor de pesos ω por

$$\omega = \frac{S^{-1}\mathbf{1}}{\mathbf{1}'S^{-1}\mathbf{1}}, \quad (2.35)$$

onde $\mathbf{1}$ é um vetor $1_{1 \times P}$ (Bates & Granger, 1969). Os pesos estimados são utilizados para todos os instantes de tempo e este método garante que a variância do resíduo da combinação de previsões seja menor ou igual à variância dos resíduos de todas as previsões utilizadas na combinação, no intervalo utilizado para estimar S .

Neste estudo, utilizamos os resíduos das previsões individuais na Amostra de Treino II para a estimação da matriz de covariância S .

- **Média Geométrica:** Esse é um método não-linear de combinação de previsões com uma grande facilidade conceitual. As previsões são combinadas através da média

geométrica delas em cada instante t , de tal forma que

$$\Psi_{MG}(\mathbf{f}_t) = \sqrt{\prod_{i=1}^P f_{i,t}}. \quad (2.36)$$

Este método só pode ser utilizado para previsões que possuam apenas valores positivos, caso P seja par.

- **Regressão Linear:** A regressão linear é utilizada para modelar relações lineares entre uma variável dependente e uma ou mais variáveis explicativas. No paradigma de combinação de previsões, é utilizada a carga como variável dependente e as previsões como variáveis explicativas, resultando em um modelo

$$L_t = \beta_0 + \beta_1 f_{1,t} + \beta_2 f_{2,t} + \dots + \beta_P f_{P,t} + \epsilon_t. \quad (2.37)$$

Os parâmetros β_i são estimados, utilizando os dados na Amostra de Treino II, através do método de mínimos quadrados ordinários, que possui solução analítica.

Neste estudo não são feitos testes de hipóteses sobre os parâmetros β devido à esperada alta multicolinearidade entre as previsões individuais, o que impossibilita a análise de previsores individuais, mas não afeta a capacidade de predição do método como um todo (Gujarati & Porter, 2003).

Após a estimativa dos parâmetros $\hat{\beta}$, a combinação das previsões é feita através da equação

$$\Psi_{RL}(\mathbf{f}) = \hat{\beta}_0 + \sum_{i=1}^P \hat{\beta}_i f_i. \quad (2.38)$$

Com exceção da Média Geométrica, todos os métodos de combinação apresentados acima combinam linearmente as previsões individuais. Contrapondo este paradigma de combinação linear, Shi & Liu (1993) apresentam cada previsão individual f_i como um subsistema de processamento de informação, e a combinação dos modelos $\Psi(\mathbf{f})$ como um sistema complexo de processamento de informação, e argumentam que a integração de um sistema complexo deve ser mais que a soma de seus subsistemas, o que requer que Ψ seja uma função não-linear.

A abordagem tradicional em regressão não-linear consiste em assumir que $\Psi(\mathbf{f})$ possui uma forma padrão (e.g: $\beta_0 + \beta_1 f_1 + \beta_2 f_2^2 f_3^{\beta_3}$), sendo necessário apenas encontrar os

coeficientes β 's que ajustem melhor a função aos dados observados. Tal abordagem é dependente da habilidade de quem escolhe a forma da função a ser utilizada e, por não conhecermos a natureza da relação entre as informações das previsões individuais, a escolha desta função se torna um problema não trivial. Uma forma de abordar esse problema é apresentada por Shi & Liu (1993); Shi *et al.* (1999), que utilizam Redes Neurais Artificiais (RNA) para a combinação de previsões, devido à capacidade deste método de aproximar funções não-lineares.

Neste trabalho, propomos um método não-linear de combinação através de *Regressão Simbólica* para a determinação da função de combinação Ψ no problema de previsão de cargas a curto prazo. Diferentemente das outras abordagens, a Regressão Simbólica não objetiva apenas estimar os coeficientes de uma função de regressão ou aproximá-la, mas encontrar a própria função, seja esta linear ou não, além de selecionar quais previsões devem ser incluídas na combinação.

2.4 Regressão Simbólica

A Regressão Simbólica (RS) é uma meta-heurística de modelagem baseada em Computação Evolucionária. é uma aplicação da *Programação Genética* (PG), paradigma primeiramente desenvolvido por Koza (1989, 1992), baseado nos trabalho de Holland (1975) em Algoritmos Genéticos (AG). Tais métodos, inspirados na evolução natural das espécies e na genética, são utilizados para solução de problemas de otimização.

Poli *et al.* (2008) descrevem as seguintes características dos problemas onde a Regressão Simbólica tem se mostrado produtiva:

- A relação entre as variáveis regressoras é desconhecida ou pouco compreendida;
- Encontrar a forma funcional da solução é uma parte significativa do problema;
- Existe uma quantidade razoável de dados para teste;
- O problema não possui solução analítica;
- Uma solução aproximada é aceitável (ou a única possível);
- Pequenas melhoras na qualidade da previsão trazem grandes resultados.

A partir disto e das características do problema de combinação de previsões para séries temporais de carga a curto prazo, a Regressão Simbólica se mostra uma alternativa interessante à ser utilizada para determinar a função Ψ a ser utilizada para a combinação das previsões.

Os métodos de computação evolucionária possuem como fundamento a geração de um conjunto de soluções candidatas, representados através de uma estrutura computacional, denominado *população de indivíduos*, e a seguir a evolução destes indivíduos é feita com base em métodos de seleção, combinação e mutação. Esta evolução gera novos indivíduos na população, buscando otimizar uma ou mais funções denominadas *funções de aptidão*. A cada vez que é executada uma iteração destes operadores para a criação de novos indivíduos, dizemos que houve uma nova Geração. A forma de representar os indivíduos, assim como os operadores de seleção, combinação e mutação e os critérios de parada, variam de acordo com a técnica e o algoritmo escolhidos.

A Regressão Simbólica tem como objetivo, dado um conjunto de dados com variáveis regressoras X_1, X_2, \dots, X_P , e uma variável independente Y , ajustar uma função $\Psi(X_1, X_2, \dots, X_P)$ que minimize uma função de aptidão, geralmente uma função de custo relacionada ao resíduo do ajuste dada por $C[Y - \Psi(X_1, X_2, \dots, X_P)]$. O fluxograma básico de um algoritmo de Regressão Simbólica é apresentado na Figura 2.5.

Para a utilização da Regressão Simbólica é portanto necessário definir qual representação dos indivíduos será utilizada, como será feita a inicialização da população, quais os operadores de seleção, combinação e mutação, além da função de aptidão e os critérios de parada para o algoritmo. A seguir são apresentados os que foram utilizados neste trabalho.

2.4.1 Representação dos Indivíduos

Na Programação Genética, o indivíduo é um candidato a solução do problema de otimização. Na Regressão Simbólica, esse indivíduo é uma função das variáveis regressoras e é necessário que seja representado computacionalmente, para a aplicação dos operadores da Programação Genética. Na literatura, existem propostas de diversas estruturas para representação das funções, tais como listas e grafos; contudo, neste trabalho, utilizaremos a representação clássica de Koza (1992) utilizando árvores sintáticas. A representação de uma função como uma árvore sintática pode ser vista na Figura 2.6, que representa a

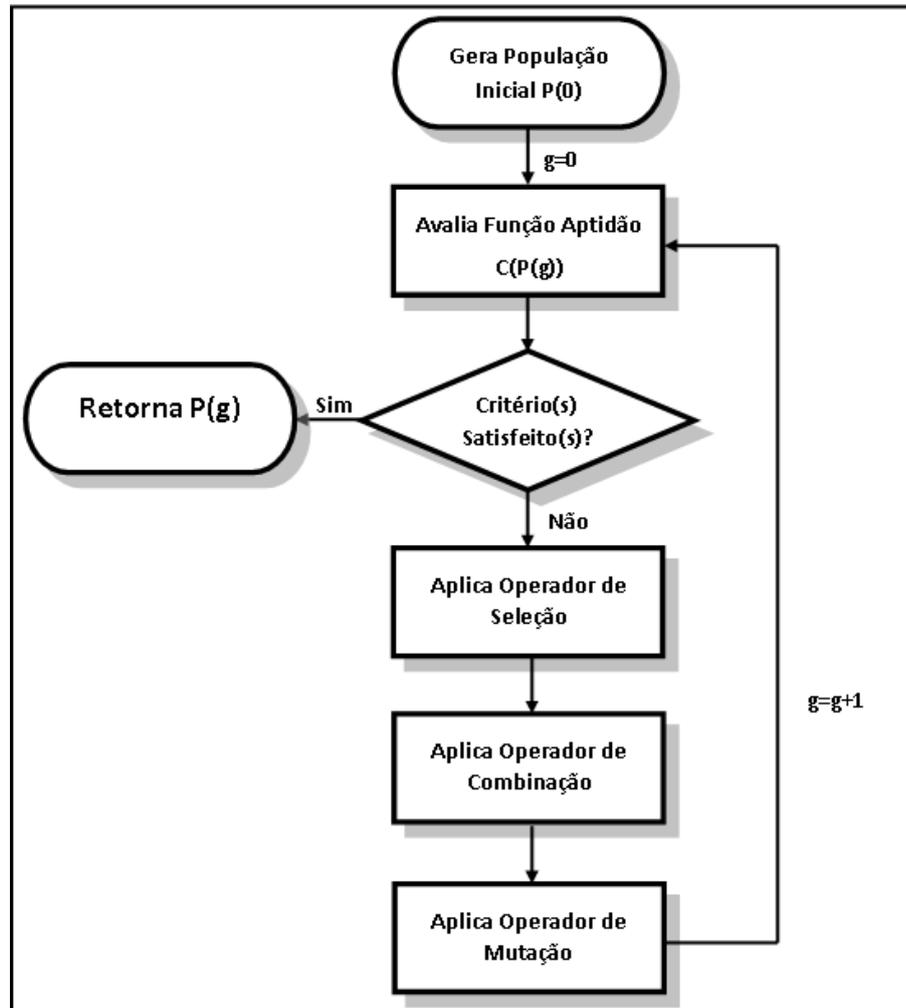


Figura 2.5: Fluxograma básico de um algoritmo de regressão simbólica

função $2y + \frac{\text{sen}(x)}{5}$. A árvore sintática, neste paradigma, é chamada de *Genótipo*, e a função representada, chamada *Fenótipo*. Vale notar que genótipos diferentes podem representar um mesmo fenótipo.

Os nós da árvore sintática podem ser de dois tipos, nós *funções* ou nós *terminais*. Em Regressão Simbólica, os nós funções usualmente são operadores aritméticos e os nós terminais são as variáveis regressoras e constantes numéricas. Há ainda um conceito importante relativo aos nós da árvore sintática, que é o conceito de *aridade*. A aridade de um nó define o seu número de nós filhos, e depende da operação presente no nó - nós com operadores binários (ex: +; -; *; /) possuem aridade 2, enquanto operadores unários (ex: *sen*; *cos*; *exp*; *log*) possuem aridade 1, e nós terminais possuem aridade 0.

Além da estrutura utilizada para a representação, é necessário definir as funções que poderão compor os nós e como serão geradas as constantes; esta escolha define o espaço de busca dos fenótipos. Neste estudo, utilizamos as funções binárias de adição, subtração,

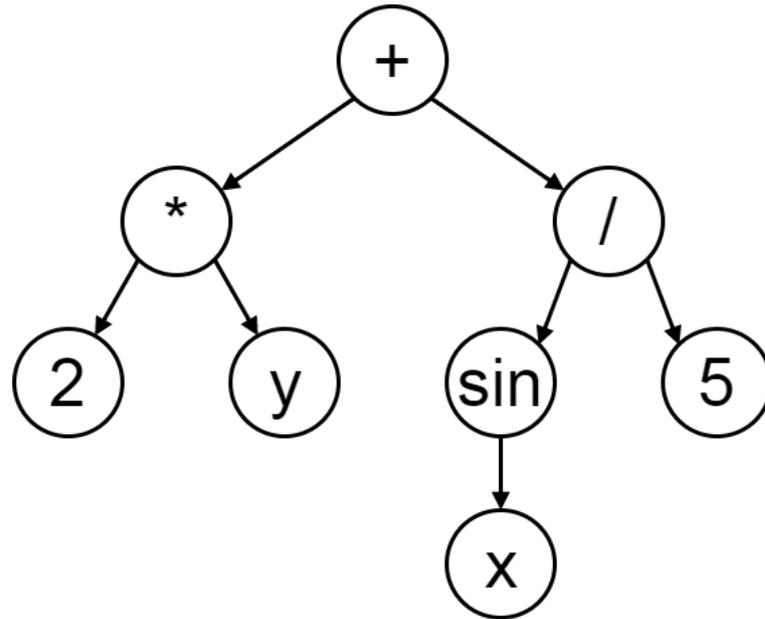


Figura 2.6: Representação de uma função como Árvore Sintática.

multiplicação e divisão (protegida¹) e as funções unárias seno, cosseno, exponencial, logaritmo natural (protegido¹) e raiz-quadrada (protegida¹). Para os nós terminais, utilizamos as previsões individuais como variáveis regressoras, e números aleatórios sorteados de uma distribuição normal padrão como constantes.

2.4.2 Inicialização da População

Após definida a estrutura utilizada para representar os indivíduos da população, é gerada uma população inicial. Para isto, é necessário definir dois parâmetros: o primeiro é o tamanho da população a ser utilizada na evolução (N_p), ou seja, a quantidade de indivíduos a serem gerados; o segundo é a altura máxima das árvores a serem geradas (h), que influencia na complexidade das funções obtidas.

Definidos os parâmetros, existem dois métodos principais para geração dos indivíduos, o método *full*, que gera sempre árvores cheias (com todos os nós possíveis) com altura igual à altura máxima estabelecida, e o método *grow*, utilizado neste trabalho, que permite a criação de árvores mais diversificadas em relação à forma e ao tamanho. A Tabela 2.1 mostra um algoritmo recursivo de geração de indivíduos pelo método *grow*, e a Figura 2.7 mostra um exemplo das etapas de geração de um indivíduo por este método.

O tamanho da população é um parâmetro de grande importância para a Regressão

¹As versões protegidas de uma função retornam 0 caso um argumento não esteja contido no domínio

Função: `gera_ind_grow(conj_func,conj_term,h_max)`
1: **se** `h_max =0` **ou** `aleatorio()` $\frac{|conj_term|}{|conj_term|+|conj_func|}$
2: `expr = escolhe_elemento_aleatorio(conj_term)`
3: **senão**
4: `func=escolhe_elemento_aleatorio(conj_func)`
5: **para** `i=1 até aridade(func)` **faça**
6: `arg_i=gera_ind_grow(conj_func,conj_term,h_max-1)`
7: **fim para**
8: `expr=(func,arg_1,arg_2,...)`
9: **fim se**
10: **retorna** `expr`

Notas: **conj_func** é o conjunto das funções que podem ser utilizadas, **conj_term** é o conjunto de terminais (variáveis regressoras e constantes), **h_max** é a altura máxima definida para a árvore, **aleatorio()** gera um número aleatório a partir de uma distribuição Uniforme(0, 1) e **expr** é a função resultante em notação pré-fixada.

Tabela 2.1: Algoritmo para geração de indivíduos pelo método *grow*. Adaptado de Poli *et al.* (2008)

Simbólica, e muitas vezes o sucesso do método depende de uma população de muitos indivíduos. Poli *et al.* (2008) indicam que a população deve ser de 500 ou mais indivíduos; Augusto & Barbosa (2000), Barmpalexis *et al.* (2011) utilizam populações que variam entre 500 e 1500 indivíduos em seus experimentos. Para a geração de uma população de tamanho N_p , são feitas N_p execuções do algoritmo da Tabela 2.1.

A escolha do tipo de representação dos indivíduos, dos nós funções e terminais a serem utilizados e da altura máxima da árvore a ser utilizada definem o **espaço de busca** do método. Este espaço de busca é composto por todas as funções Ψ que podem ser representadas através da árvore sintática de altura máxima h com os nós funções e terminais definidos. As etapas de seleção, combinação e mutação são realizadas para a busca dentro deste espaço da função Ψ que minimize a função de aptidão.

2.4.3 Seleção

Após inicializada a população, a próxima etapa é a seleção dos indivíduos que serão utilizados para produzir a próxima geração. Inspirados no processo evolutivo natural, os métodos de computação evolucionária utilizam uma seleção dos indivíduos com base em sua adaptação.

Na Regressão Simbólica, esta adaptação é medida através de funções dos erros de

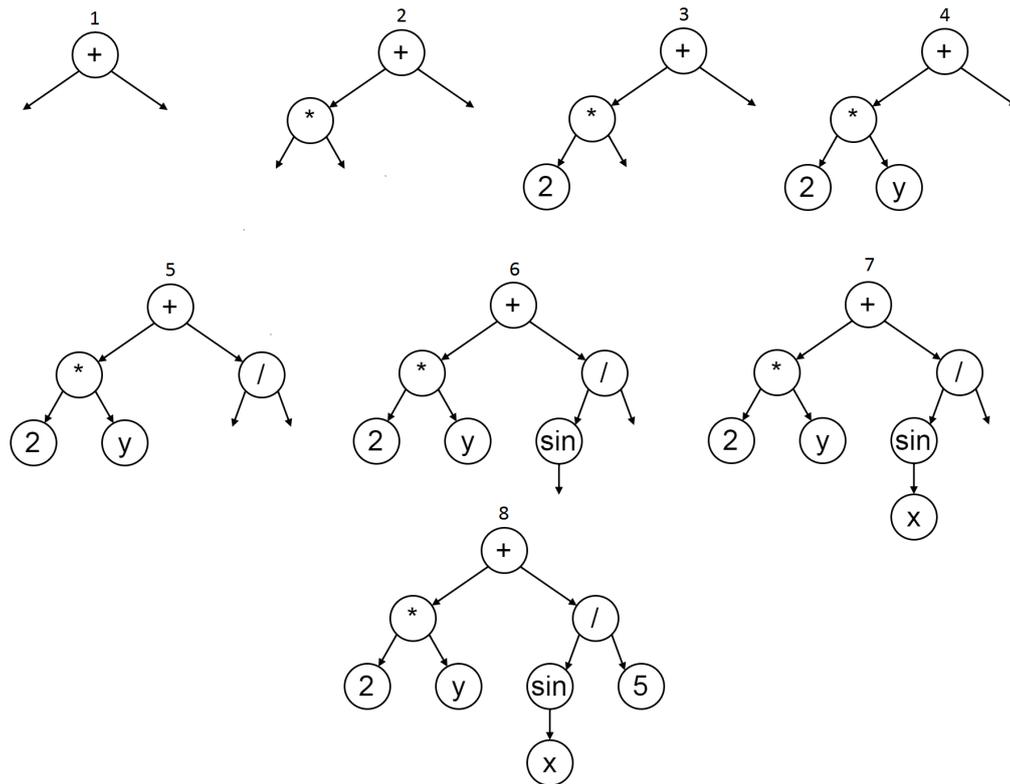


Figura 2.7: Exemplo de construção de um indivíduo pelo método *grow*

previsão do indivíduo na amostra de treino utilizada, como o MAPE e o MSE (ver na Seção 2.5.1); o objetivo da Regressão Simbólica é encontrar indivíduos que minimizem estas funções. Para a etapa de seleção, é necessário computar essa função, na amostra de treino utilizada, para cada indivíduo na população.

Um primeiro mecanismo de seleção que pode ser utilizado é o *elitismo*, que garante que uma fração f_e da população contendo os indivíduos que possuem a melhor adaptação seja incluída na próxima geração, sem passar por operadores de combinação e mutação; ou seja, estes indivíduos são mantidos exatamente iguais na geração seguinte. Enquanto esta estratégia garante que não haja uma perda dos melhores indivíduos, ela também diminui a diversidade da população, podendo levar à convergência para ótimos locais. Para evitar isto, é recomendável que, caso o mecanismo seja utilizado, seja escolhido um valor pequeno para f_e .

Outro passo da seleção é a escolha de indivíduos para o *pool* de reprodução, que contém os indivíduos que gerarão os indivíduos da próxima geração (também chamados de "filhos") através dos operadores de combinação e de mutação. O número de indivíduos neste *pool* (N_r) depende da utilização de elitismo e do operador de combinação utilizado.

Pelos métodos escolhidos neste trabalho, utilizamos $N_r = (1 - f)N_p$.

Para a escolha dos indivíduos que compõe o *pool*, é recomendável que haja, além de critérios de aptidão, um componente estocástico, para preservar a diversidade na população (Eiben *et al.*, 2003). Métodos de seleção comumente utilizados incluem Roleta Proporcional à Aptidão, Torneios de 2 ou 4 indivíduos e Seleção Baseada em Ranking. O Torneio de 2 Indivíduos Determinístico foi utilizado neste trabalho e é apresentado a seguir.

O Torneio de 2 Indivíduos consiste na seleção aleatória com reposição de 2 indivíduos da população. Estes indivíduos são comparados utilizando seus valores de função aptidão. No caso de Torneio Determinístico, o indivíduo que possuir melhor aptidão é considerado o "vencedor" do Torneio e incluído no *pool* de reprodução. Este procedimento é repetido N_r vezes, para se obter todos os indivíduos que serão utilizados para a reprodução.

2.4.4 Combinação

Inspirada no *crossover* genético, a combinação é a forma de reprodução utilizada na Programação Genética que combina o genótipo de dois indivíduos "pais" previamente selecionados para reprodução, criando assim dois novos indivíduos "filhos". Este operador permite que indivíduos herdem características das gerações anteriores, mas possibilita um aumento da diversidade, aumentando a cobertura do espaço de busca.

No operador de combinação são selecionados, sem reposição, dois indivíduos do *pool* de reprodução e com uma probabilidade p_c são combinados. Caso não haja a combinação, os filhos serão idênticos aos pais. Caso haja, são selecionados aleatoriamente um nó em cada um dos pais e as subárvores desses nós são trocadas entre si. Como este é o principal operador de geração de indivíduos da Regressão Simbólica, é recomendável que p_c seja alta.

Um exemplo de combinação pode ser visto na Figura 2.8, onde os genótipos de dois indivíduos pais são combinados, dando origem a dois indivíduos filhos. Os fenótipos combinados são $2y + \frac{\text{sen}(x)}{5}$ e $e^{x-y} \frac{3}{x}$ dando origem à $2y + x - y$ e $e^{\frac{\text{sen}(x)}{5}} \frac{3}{x}$. é possível perceber que os filhos podem ter alturas maiores ou menores que os pais, dependendo do nó sorteado para que seja feito o corte.

Além disso, o primeiro fenótipo é equivalente à $x - y$, que pode ser representado com apenas três nós. Apesar de possuírem o mesmo fenótipo, a diferença no genótipo faz com

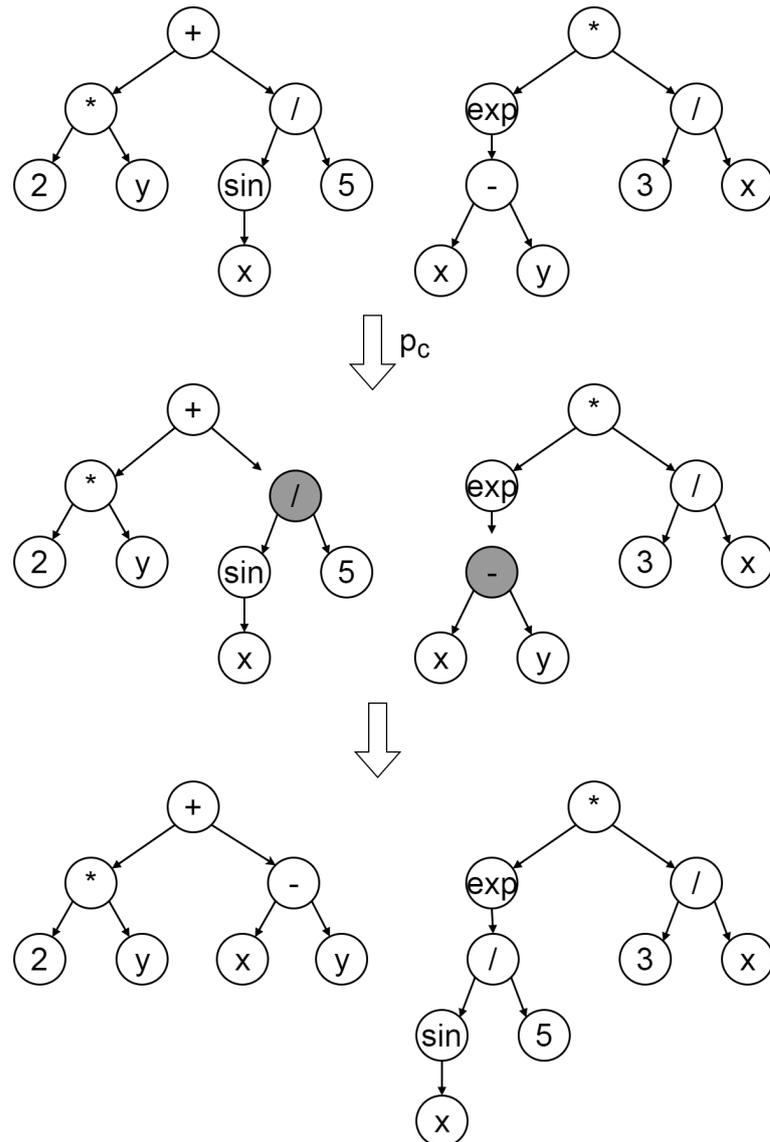


Figura 2.8: Exemplo de Combinação entre 2 indivíduos

que possam gerar filhos distintos, permitindo uma maior cobertura do espaço de busca.

2.4.5 Mutação

A *mutação* é um operador genético utilizado para aumentar a diversidade da população de indivíduos. é útil para evitar que a população convirja para um ótimo local, reinserindo nós que podem ter sido extintos no processo evolutivo, e ampliando a cobertura do espaço de busca.

Para cada indivíduo filho gerado na etapa anterior, é sorteado, com probabilidade p_m se o indivíduo passará ou não pelo operador de mutação. Caso passe, um nó do indivíduo é sorteado, e a partir desse nó é gerada aleatoriamente uma sub-árvore. O novo indivíduo

filho substitui o original na população de filhos. Banzhaf *et al.* (1998) recomendam que p_m seja um número pequeno, no máximo 0,05.

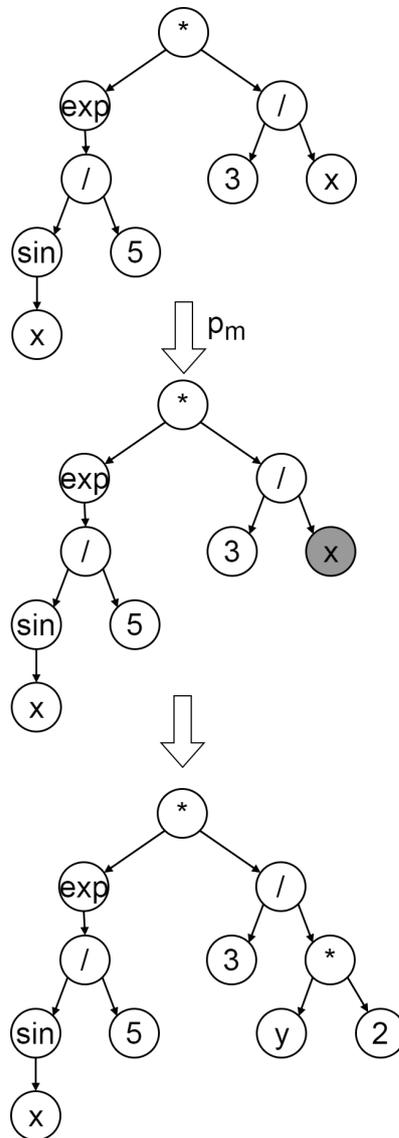


Figura 2.9: Exemplo de Mutaç o de um indiv duo

A Figura 2.9 apresenta um exemplo de muta o para um dos filhos gerados na Figura 2.8. Ap s a muta o, o fen tipo $e^{\frac{\sin(x)}{5}} \frac{3}{x}$ se torna $e^{\frac{\sin(x)}{5}} \frac{3}{2y}$, reintroduzindo a vari vel y que havia sido perdida na opera o de combina o.

Ap s a etapa de muta o, os filhos gerados, junto com os indiv duos selecionados por elitismo, substituem a popula o original, dando in cio   uma nova gera o. Esta nova gera o ser  submetida por sua vez aos operadores de sele o, combina o e muta o, e tal procedimento se repetir  at  que seja atendido um crit rio de parada pr -estabelecido.

2.4.6 Critérios de Parada

É necessário definir os critérios que serão utilizados para a parada do algoritmo de Regressão Simbólica. Estratégias comuns são a) fixar o tempo de execução; b) fixar o número de gerações; c) estabelecer um valor para função de aptidão que, caso alcançado, causa o término do algoritmo; d) fixar o número de avaliações da função de aptidão. Esta última estratégia é especialmente útil quando se deseja comparar mais de um algoritmo evolucionista com parâmetros distintos, dado que muitas vezes a operação com maior custo computacional é a avaliação da função de aptidão. Estas estratégias também podem ser combinadas através de disjunção ou conjunção.

Após o critério ser atendido, o algoritmo é interrompido, retornando a população da última geração como solução. Usualmente, o fenótipo do indivíduo com maior aptidão é selecionado como melhor modelo de regressão e utilizado para análise e predição. Contudo, em alguns problemas pode ser interessante examinar os demais indivíduos com boa função de aptidão, para observar a forma funcional das soluções, e obter alguma interpretação prática do problema.

Para o problema de combinação de previsões deste trabalho, escolhemos o indivíduo com melhor função de aptidão para a realização da combinação das previsões individuais, gerando uma nova previsão.

2.5 Avaliação das previsões

Como discutido na Seção 2.3, existem diversas métricas para a avaliação e comparação de previsões em séries temporais. Neste trabalho, utilizamos duas medidas relacionadas ao erro de previsão, de conceituação simples e fácil implementação. Além destas, para a avaliação da combinação de previsões, utilizamos os critérios propostos por De Menezes *et al.* (2000).

2.5.1 Medidas de Erro

O Erro Quadrático Médio (*Mean Square Error*, MSE) é utilizado neste trabalho como função de aptidão na Regressão Simbólica, além de ser utilizado como função a ser minimizada na estimação das constantes e dos parâmetros dos métodos de previsão individuais e da combinação por Regressão Linear. Para uma série temporal L_t e uma

previsão f_t , $t = 1, \dots, T$, calcula-se o MSE por

$$MSE = \frac{1}{T} \sum_{t=1}^T (L_t - f_t)^2. \quad (2.39)$$

Devido à sua forma quadrática, o MSE leva em conta não apenas a magnitude dos erros de previsão, como também sua variância. Apesar desta vantagem, esta forma quadrática faz que o MSE não tenha grande interpretabilidade prática; é mais comumente utilizada para a comparação de previsões de carga uma outra medida, o Erro Absoluto Percentual Médio (*Mean Absolute Percent Error*, MAPE), que é adimensional.

O MAPE é uma média dos Erros Absolutos Percentuais (APEs) dados por

$$APE_t = 100 \left| \frac{L_t - f_t}{L_t} \right|. \quad (2.40)$$

Além da Média dos APEs, outra medida útil para a comparação é a sua variância, que proporciona informação sobre a dispersão dos erros, ou seja, sobre a estabilidade das previsões.

Apesar da interpretabilidade do MAPE, ele não é recomendado para séries que contenham observações iguais ou próximas de 0, pois estas observações teriam pesos altíssimos no cálculo, causando uma distorção no valor do MAPE.

2.5.2 Avaliação de Combinação de Previsões

Como discutido, a Combinação de Previsões, além da redução do erro de previsão, pode levar à outros benefícios para a previsão de uma série temporal. A fim de avaliar a qualidade de uma combinação de previsões, De Menezes *et al.* (2000) propuseram três critérios a serem analisados além do erro de previsões:

- **Variância dos Erros de Previsão:** Em uma boa combinação de previsões, a variabilidade dos resíduos deve ser a menor que a de todas as previsões individuais, proporcionando um aumento na estabilidade das previsões. Uma forma de avaliar este critério é através da Variância dos APEs da combinação de previsões.
- **Assimetria dos Erros de Previsão:** A forma da distribuição dos erros de previsão influencia diretamente na avaliação de riscos e incertezas das previsões. A assimetria da distribuição dos erros é uma das medidas que podem ser utilizadas para avaliar

a forma de uma distribuição e pode ser medida pelo coeficiente da assimetria ou avaliada graficamente através de um histograma. Uma distribuição simétrica dos erros implica em um menor risco de previsão.

- **Correlação Serial dos Erros de Previsão:** Apesar de séries temporais possuírem correlação serial, o mesmo não deve acontecer com os resíduos. A presença de correlação serial nos resíduos implica na presença de estruturas temporais nos dados que o método de previsão utilizado não foi capaz de modelar, além de afetar a estimação da variabilidade dos erros de previsão. Esta correlação serial pode ser avaliada através dos gráficos da Função de Autocorrelação (FAC) dos resíduos.

3 RESULTADOS E DISCUSSÕES

Os métodos apresentados anteriormente foram utilizados para a previsão dos dados de carga horária de energia elétrica da cidade do Rio de Janeiro (RJ) entre os anos de 1996 e 1997, considerando a partição apresentada na Seção 2.1. Foram realizadas previsões individuais e combinadas através de diferentes métodos, buscando obter melhores previsões da carga de energia elétrica. Os resultados obtidos em cada etapa deste trabalho são apresentados e discutidos a seguir.

3.1 Previsões Individuais

A partir dos nove métodos de previsão apresentados na Seção 2.2, realizamos previsões 24 horas-à-frente para a série de cargas. Para os métodos em que há a necessidade de se estimar parâmetros, foram utilizados os dados das 54 semanas iniciais da série (Amostra de Treino) para esta estimação. A Tabela 3.1 apresenta o Erro Absoluto Percentual Médio, encontrado para cada um destes métodos de previsão.

Tipo	Método		MAPE (%)	
			Treino II	Teste
Sequencial	DSHW	f_1	2,39	2,52
24 métodos em paralelo	<i>Naïve</i>	f_2	7,34	7,37
	HW	f_3	2,96	3,74
	SARIMA	f_4	4,01	4,68
168 métodos em paralelo	<i>Naïve</i>	f_5	4,65	6,12
	MM	f_6	4,27	5,82
	AES	f_7	4,20	5,10
	Holt	f_8	4,22	5,04
	ARIMA	f_9	4,37	5,61

Tabela 3.1: Resultados das previsões individuais na Amostras de Treino II e Teste

Assim como no trabalho de Carneiro (2014), o método de Holt-Winters com Dupla Sazonalidade (DSHW), utilizando uma estratégia sequencial para previsão, apresentou resultado bem superior em ambas amostras, quando comparado aos demais que utilizam um modelo para cada hora do dia ou da semana. Entre os métodos que foram utilizados

para a estratégia de modelos em paralelo, o que apresentou melhor resultado foi o método de Holt-Winters utilizando 24 modelos. Na mesma tabela, é possível observar que os erros foram menores para a Amostra de Treino II em comparação com a Amostra de Teste.

Na Figura 3.1 são apresentados gráficos da Função de Autocorrelação dos resíduos das previsões univariadas na Amostra de Teste. Estes gráficos indicam a existência de autocorrelação nos resíduos para todas as previsões, um indício de que os métodos de previsão não conseguiram, individualmente, captar toda a estrutura temporal presente nos dados.

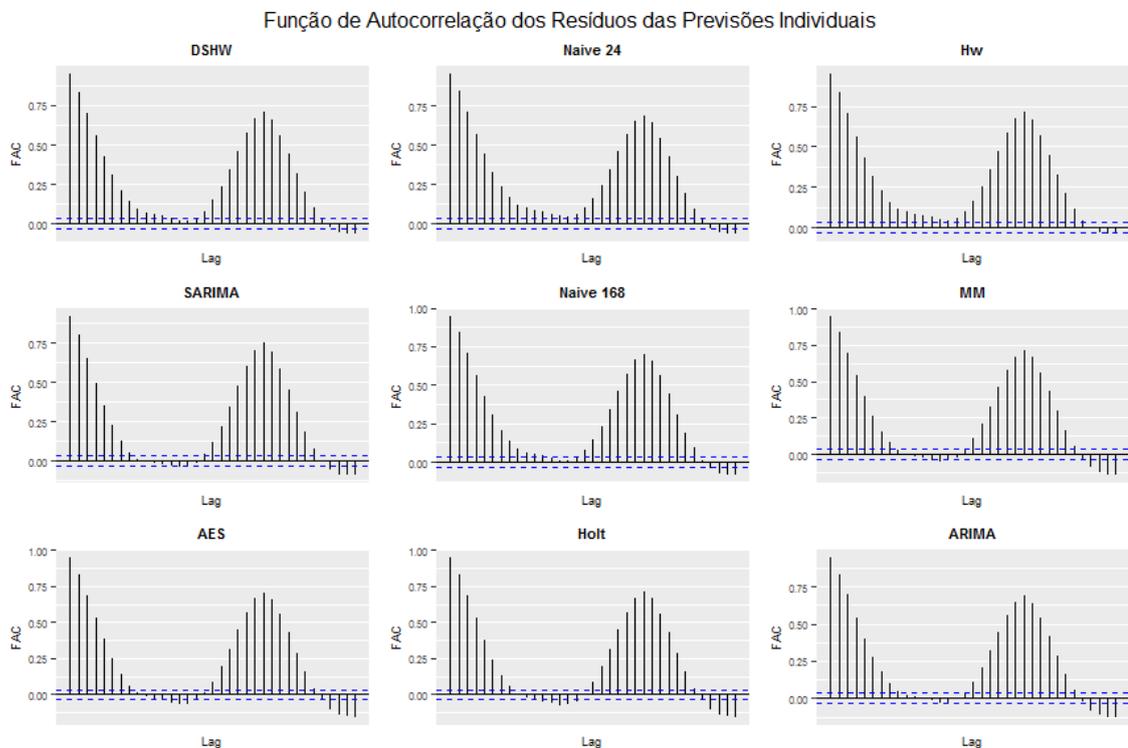


Figura 3.1: Gráficos das Funções de Autocorrelação dos Resíduos das Previsões Individuais.

Através da combinação das previsões, esperamos conseguir captar melhor esta estrutura temporal.

3.2 Combinação de Previsões por Regressão Simbólica

Para avaliar o desempenho da Regressão Simbólica como um método de combinação de previsões, fizemos uma implementação do método no *software* R (R Core Team, 2016)

adaptando funções incluídas no pacote *rgp* (Flasch *et al.*, 2014; Flasch, 2015). Com esta implementação, fizemos execuções do programa a fim de encontrar uma função $\Psi(\mathbf{f})$ que combinasse as previsões individuais f_i , minimizando uma medida de erro.

Os parâmetros escolhidos para implementação seguem as recomendações de Poli *et al.* (2008). O Tamanho da População escolhido para a aplicação foi de 1000 indivíduos. A altura máxima de 10 nós para as árvores sintáticas, geradas inicialmente pelo método *grow*, com as constantes sorteadas a partir de uma distribuição normal padrão. A função objetivo utilizada para avaliar a aptidão dos indivíduos foi o Erro Quadrático Médio (MSE). Esta função foi escolhida para facilitar a comparação com os métodos de combinação por Média Ponderada Ótima, e com o de Regressão Linear que também buscam minimizar o MSE na Amostra de Treino II, e para evitar o vício de avaliação, uma vez que a medida utilizada para compará-las é o MAPE.

Utilizamos uma fração de elitismo f_e igual à 1%; isso implica que os 10 indivíduos com melhor função objetivo em uma geração são automaticamente incluídos na próxima. A probabilidade de ocorrer uma Combinação (p_c) foi definida como 0,8; a probabilidade de ocorrer uma Mutação (p_m) como 0,05. Como critério de parada, estabelecemos um valor máximo de 200000 avaliações da função objetivo. Um resumo dos parâmetros utilizados na aplicação da Regressão Simbólica é encontrado na Tabela 3.2.

	Parâmetro	Valor
Funções	Aridade = 1	<i>sen, cos, exp, log*</i> , e $\sqrt{}$ *
	Aridade = 2	+ , - , * , / *
Variáveis Regressoras		Previsões Individuais
Método de Inicialização		<i>grow</i>
Gerador de Constantes		Normal(0, 1)
Tamanho da População N_P		1000
Altura Máxima		10 nós
Fração de Elitismo f_e		1%
Probabilidade de Combinação p_c		0,8
Probabilidade de Mutação p_m		0,05
Função Objetivo		MSE
Avaliações da Função Objetivo		200000

*Operações Protegidas.

Tabela 3.2: Parâmetros utilizados na aplicação da Regressão Simbólica

Devido à característica estocástica do método de Regressão Simbólica, os resultados encontrados podem diferir entre execuções. Para evitar que o resultado seja afetado por esta aleatoriedade, e também para verificar a estabilidade das soluções encontradas pela combinação de previsões através da Regressão Simbólica, executamos 1000 vezes o programa implementado. Para cada execução, o indivíduo que apresentava o menor valor da função objetivo na última geração foi tomado como a solução, e seu fenótipo Ψ armazenado para as avaliações.

Para cada uma destas soluções, foi calculado o MAPE e a Variância dos APEs para a Amostra de Teste. O Menor MAPE encontrado foi de 2,18%, O MAPE mediano foi de 2,19%, o MAPE médio 2,22% e o maior encontrado foi 2,34%. Quanto à variância dos APEs a menor foi de 4,31 e a maior de 5,29, tendo média de 4,59 e mediana de 4,61.

No Histograma (Figura 3.2), é possível perceber que há uma concentração de soluções com MAPEs menores que 2,20% e uma outra com MAPEs em torno de 2,30%. No Boxplot na Figura 3.3, observa-se que mais de 75% das soluções obtiveram MAPE abaixo de 2,20% na Amostra de Teste, indicando que a combinação de previsões conseguiu melhorar o resultado encontrado pela melhor previsão individual, o DSHW, que obteve MAPE de 2,52% nesta mesma Amostra.

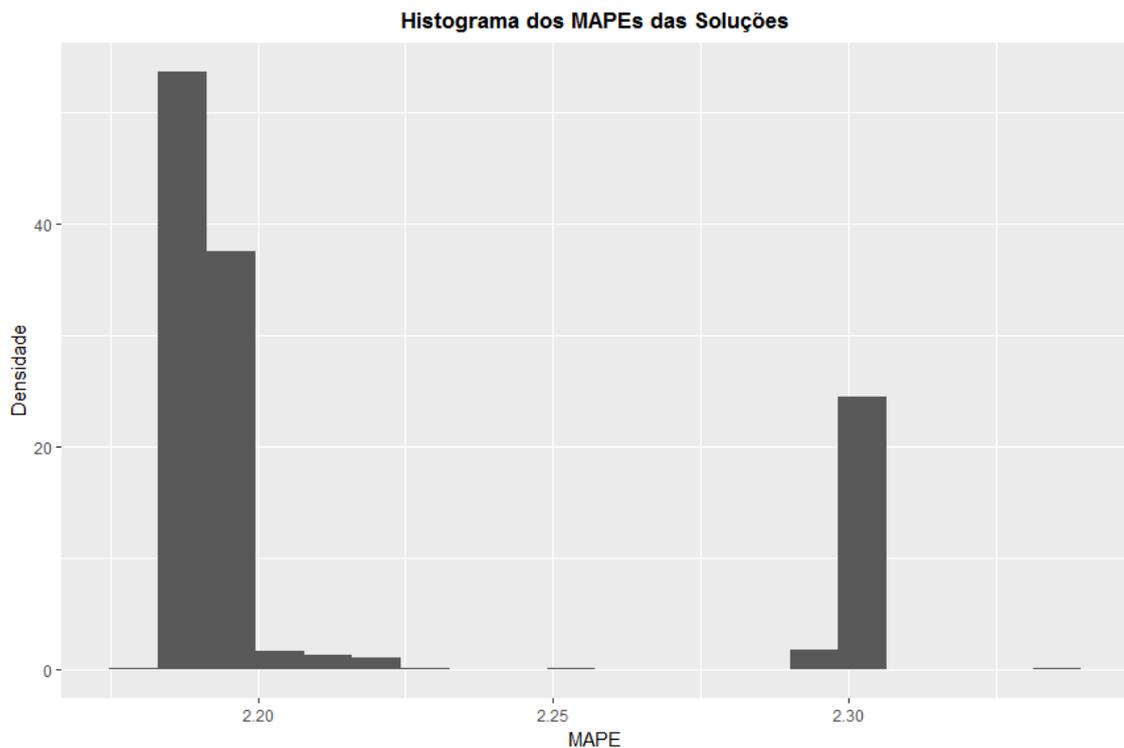


Figura 3.2: Histograma dos MAPEs das soluções para Amostra de Teste

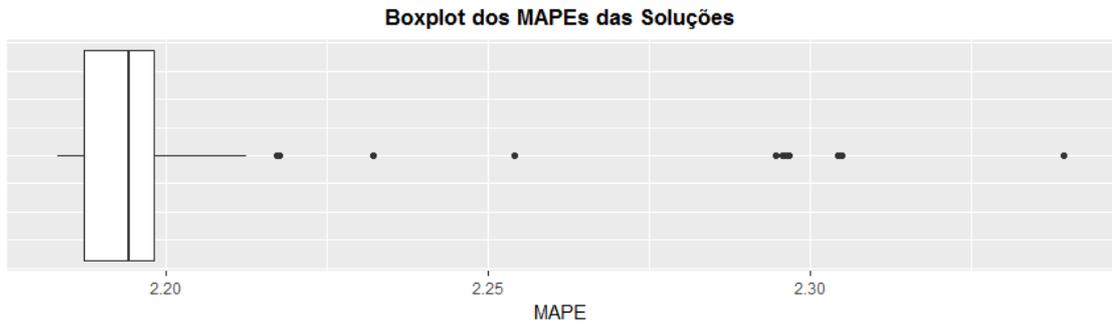


Figura 3.3: Boxplot dos MAPEs das soluções para Amostra de Teste

A seguir, avaliamos mais detalhadamente a solução que apresentou o menor MAPE. A função representada pela árvore sintática deste indivíduo é

$$\Psi_{RS}(\mathbf{f}) = \sqrt{\sqrt{f_3} \sqrt{\sqrt{\sqrt{f_3} \sqrt{\sqrt{f_5} \sqrt{f_1 f_3} \sqrt{f_1 f_1}} \sqrt{\sqrt{f_1 f_3} \sqrt{\sqrt{\sqrt{f_1 f_3 f_3} \sqrt{\sqrt{\sqrt{f_3 f_3 f_3 f_3} \sqrt{f_1 f_1}}}}}}}}}} \quad (3.1)$$

Na função acima, vemos que as operações que foram mais importantes para a solução encontrada foram a raiz quadrada e a multiplicação. Também percebe-se a presença apenas das previsões por DSHW, HW e a previsão Naïve (f_1 , f_3 e f_5), para 168 modelos em paralelo.

A equação 3.1 não permite um conhecimento mais profundo da relação encontrada entre as previsões, mas isto pode ser facilitado se fizermos uma simplificação, onde obtém-se

$$\Psi_{RS}(\mathbf{f}) = f_1^{\frac{173}{256}} f_3^{\frac{79}{256}} f_5^{\frac{4}{256}} \quad (3.2)$$

Em 3.2, é possível perceber que a solução encontrada é não-linear, e que as duas previsões mais representativas na combinação são as do DSHW e as do HW, justamente as que possuíam menor MAPE individualmente, o que era esperado. Em contraponto, aparece na solução, mesmo que com um impacto menor por estar elevada a um expoente pequeno, a previsão Naïve para 168 modelos em paralelo, previsão que apresentou um pior MAPE que as demais deste tipo.

Na Figura 3.4 é apresentado o gráfico da série temporal das cargas horárias, juntamente com as previsões obtidas pela combinação de previsões individuais por Regressão Simbólica para algumas semanas da Amostra de Teste. Este gráfico mostra que a previsão conseguiu capturar bem o comportamento dos dados, principalmente nas semanas iniciais da Amostra de Teste. Outra ferramenta que pode ser utilizada para análise visual das previsões é o gráfico dos resíduos (Figura 3.5). Este gráfico apresenta valores em torno do 0, e parece aumentar sua volatilidade na parte final da série, apresentando picos positivos e negativos. Tal comportamento será discutido em uma análise seguinte.

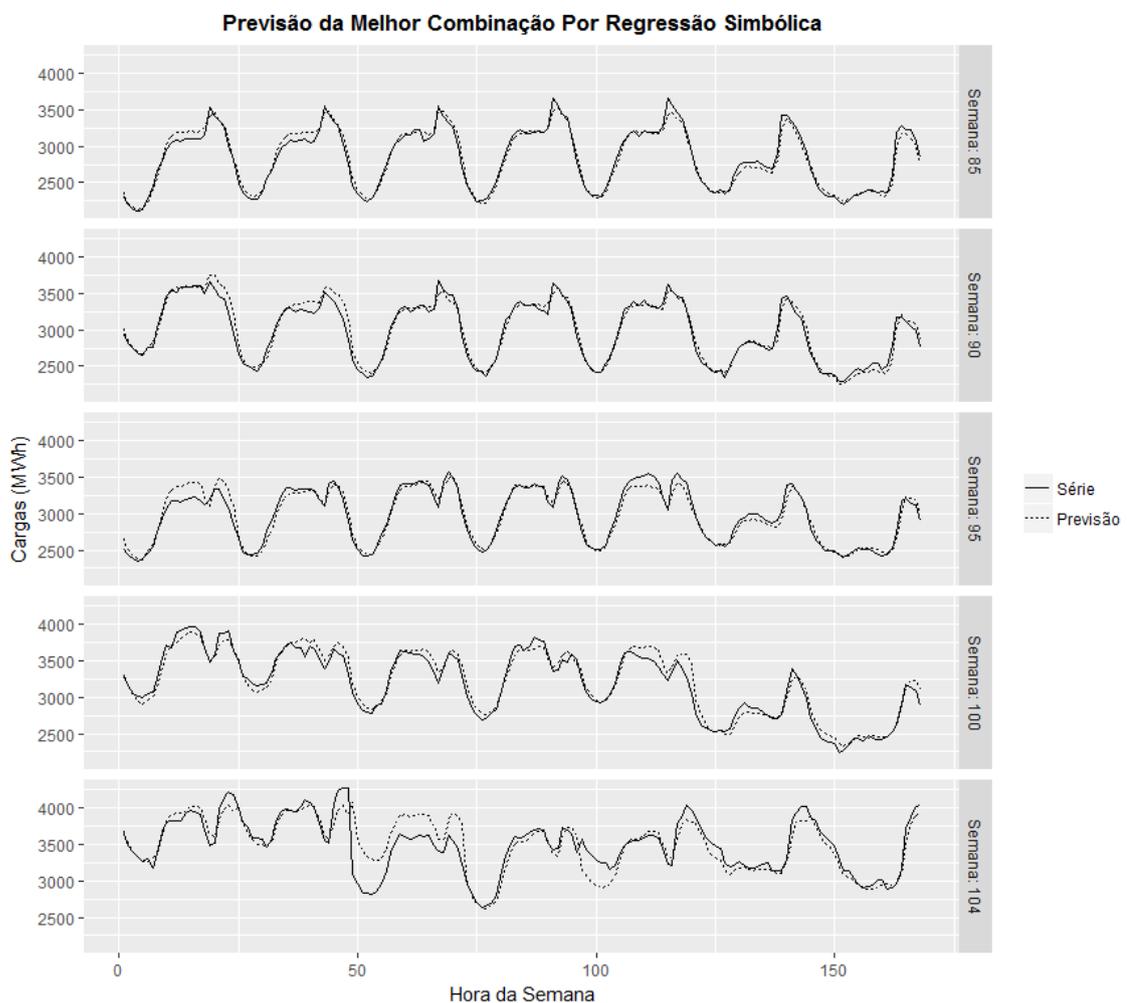


Figura 3.4: Série de Cargas e Previsão via Regressão Simbólica para semanas selecionadas da Amostra de Teste

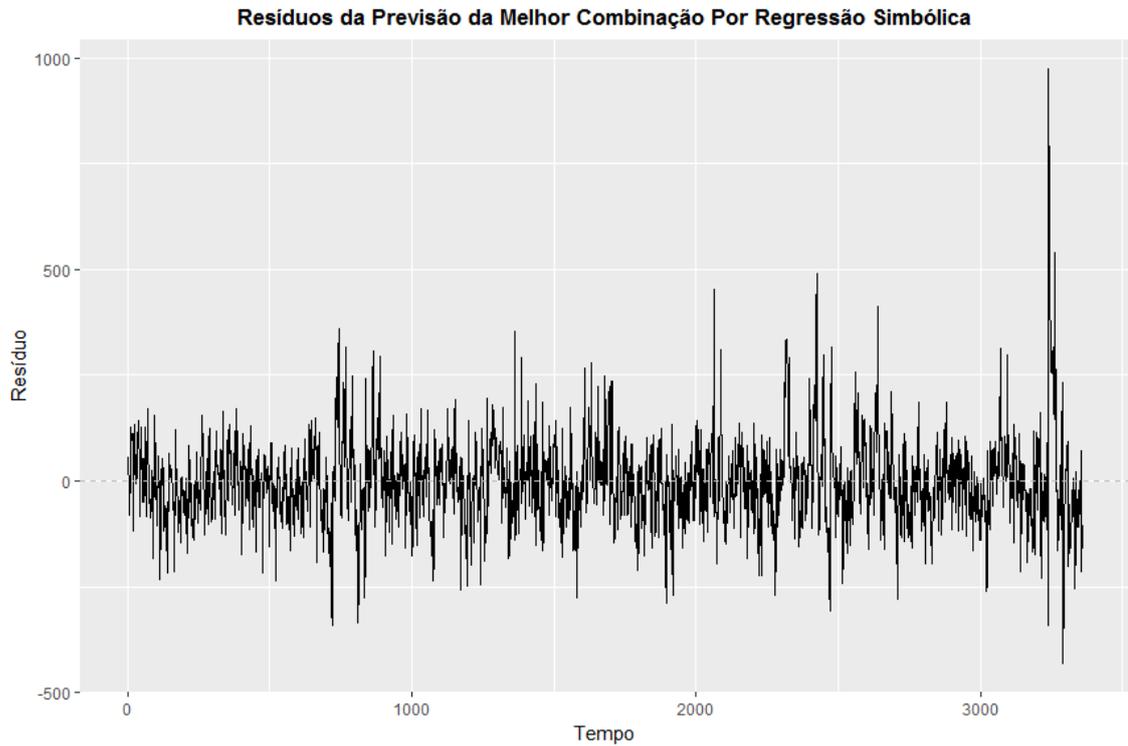


Figura 3.5: Resíduos da Previsão da carga na Amostra de Teste pela combinação usando Regressão Simbólica

Segundo as recomendações de De Menezes *et al.* (2000), apresentadas na Seção 2.5.2, avaliamos a distribuição dos resíduos. No histograma da Figura 3.6 é possível perceber uma distribuição aproximadamente Normal, com uma leve assimetria à direita. Esta assimetria é confirmada pelo coeficiente momento de assimetria, de valor 0,88. Tal assimetria pode ser devido aos picos apresentados nos resíduos ao final da Amostra de Teste.

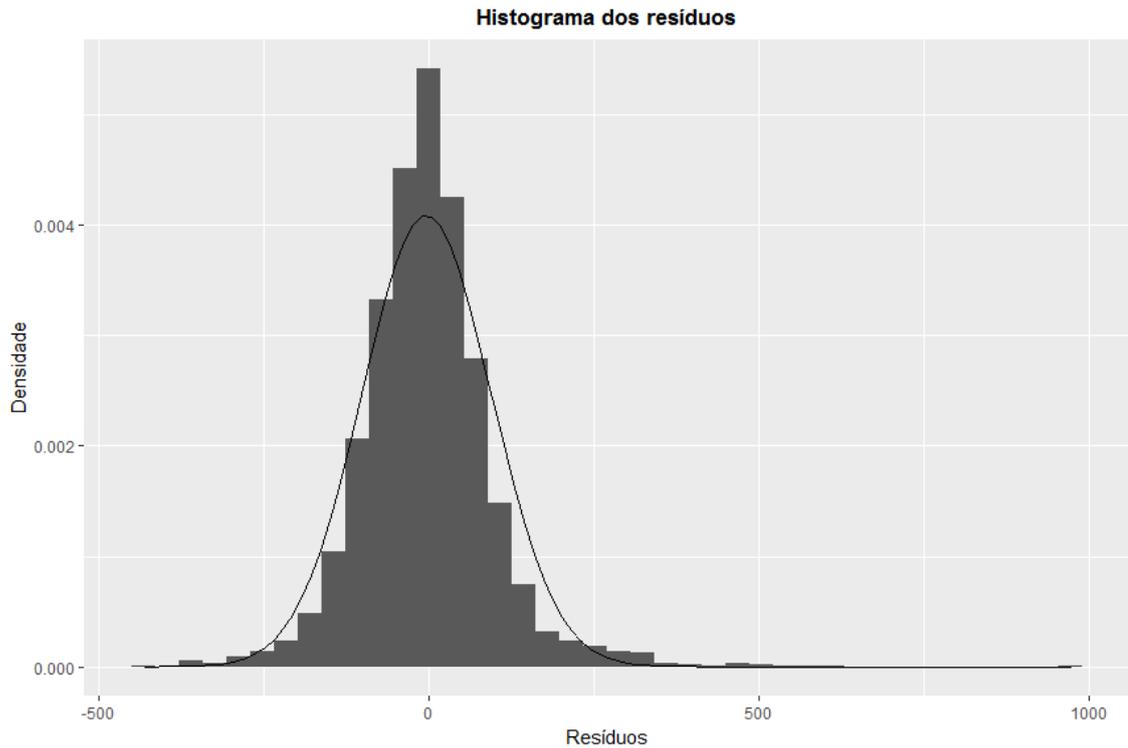


Figura 3.6: Histograma dos Resíduos de Previsão na Amostra de Teste

Outro elemento a ser avaliado em uma previsão é a autocorrelação dos resíduos. A Figura 3.7 mostra que os resíduos ainda possuem alguma autocorrelação. Contudo, quando comparamos este gráfico com os das previsões individuais (Figura 3.1), vemos que houve uma redução nos valores da autocorrelação, o que indica que a combinação de previsões pode capturar mais da estrutura temporal que qualquer uma das previsões individuais utilizadas para compô-la.

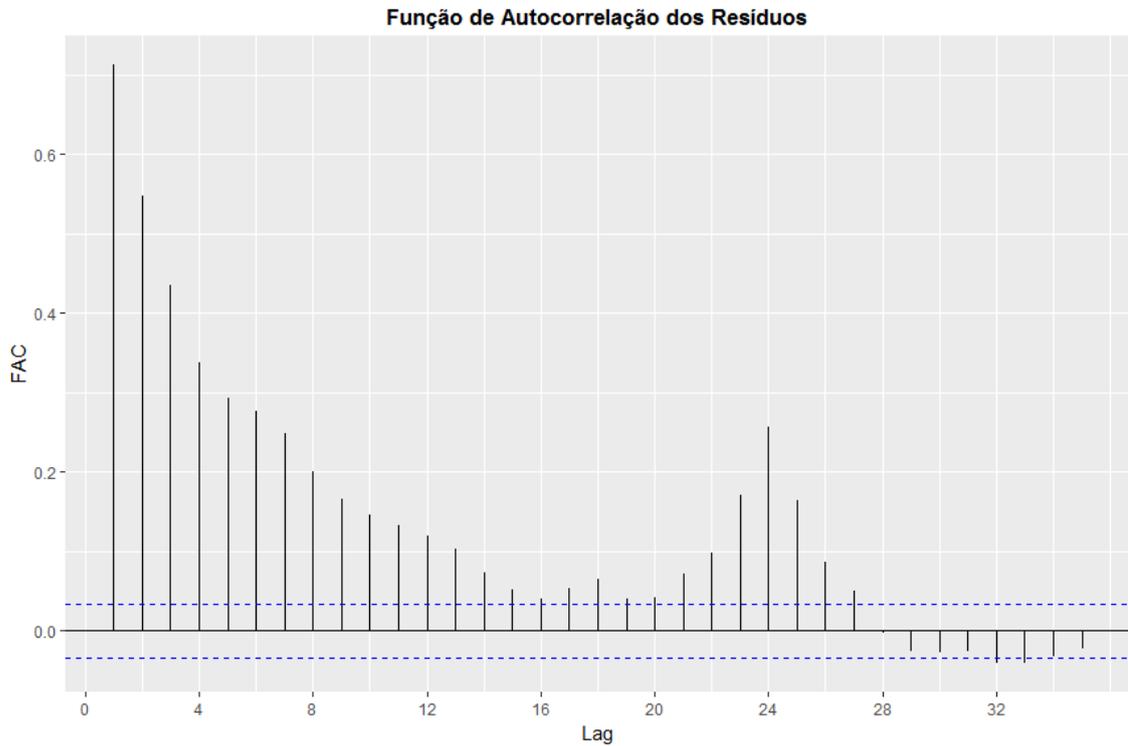


Figura 3.7: Função de Autocorrelação dos Resíduos de Previsão na Amostra de Teste

Outra análise interessante que também pode ser feita, é observar o comportamento dos APEs por dia da semana e também por hora do dia, tentando identificar padrões que podem ser úteis para explicar o erro de previsão. Estas distribuições podem ser vistas nos Boxplots nas Figuras 3.8 e 3.9.

Analisando o gráfico da Figura 3.8, é possível observar um grande *outlier* numa Quarta-feira, referente ao grande pico que aparecia no final do gráfico temporal dos resíduos. Após investigações na base de dados, descobrimos que este dia era a véspera de Natal de 1997, cuja carga não havia sofrido imputação (e, como comentado na Seção 2.1, feriados deveriam ser tratados separadamente). Decidimos, então, fazer a análise dos resultados considerando esta semana do Natal, para ser possível a comparação com os trabalho de Hippert (2001); Hippert *et al.* (2005); Carneiro (2014), mas também repetir as mesmas análises desconsiderando esta semana, para poder analisar os resultados sem a contaminação causada por estes dados.

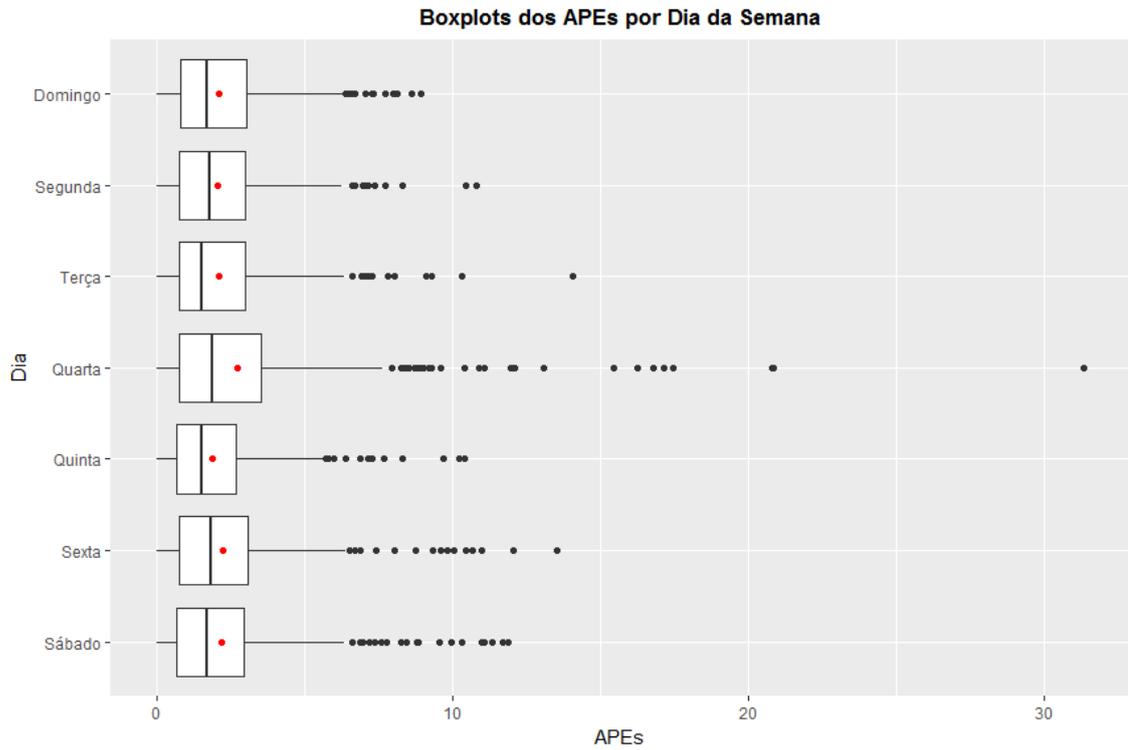


Figura 3.8: Boxplot dos APEs por Dia da Semana

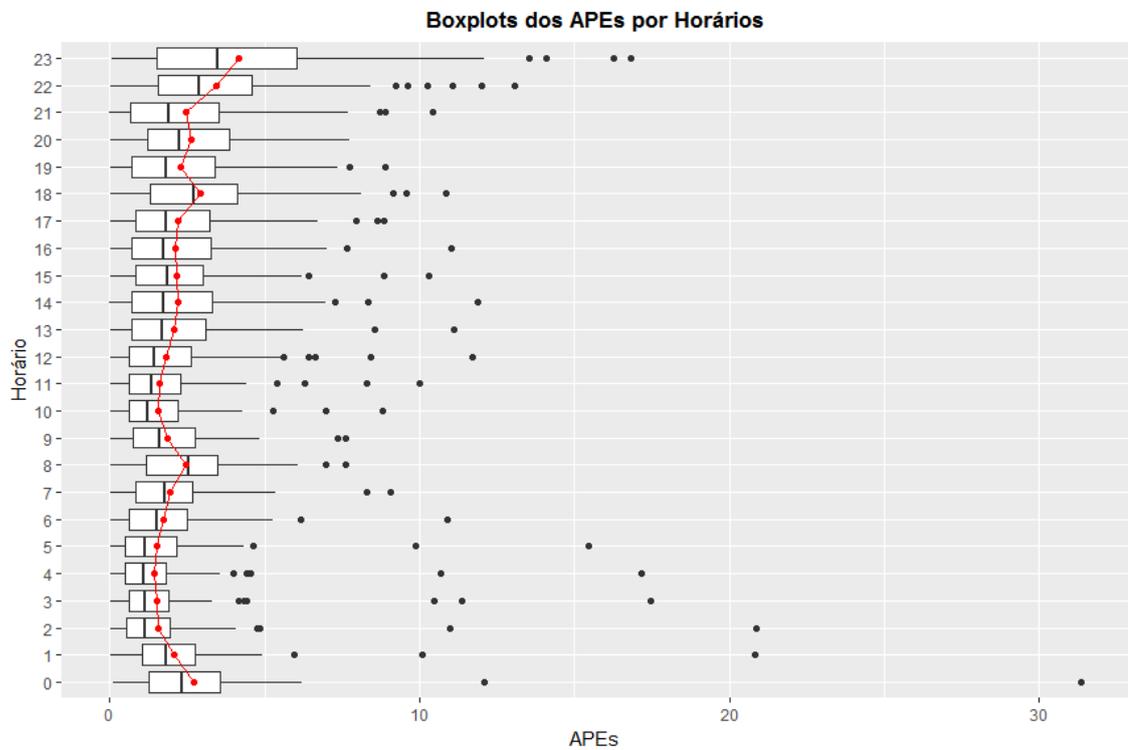


Figura 3.9: Boxplot dos APEs por Hora do Dia

Considerando apenas os dados anteriores ao Natal de 1997, o melhor MAPE encontrado através da combinação de previsões por Regressão Simbólica passa de 2,18%

para 2,09% e a variância dos APEs de 4,31 para 3,30. Além destes, o coeficiente de assimetria é substancialmente reduzido de 0,88 para 0,38. Tais diferenças também podem ser observadas nas novas distribuições dos APEs apresentadas nas Figuras 3.10 e 3.11.

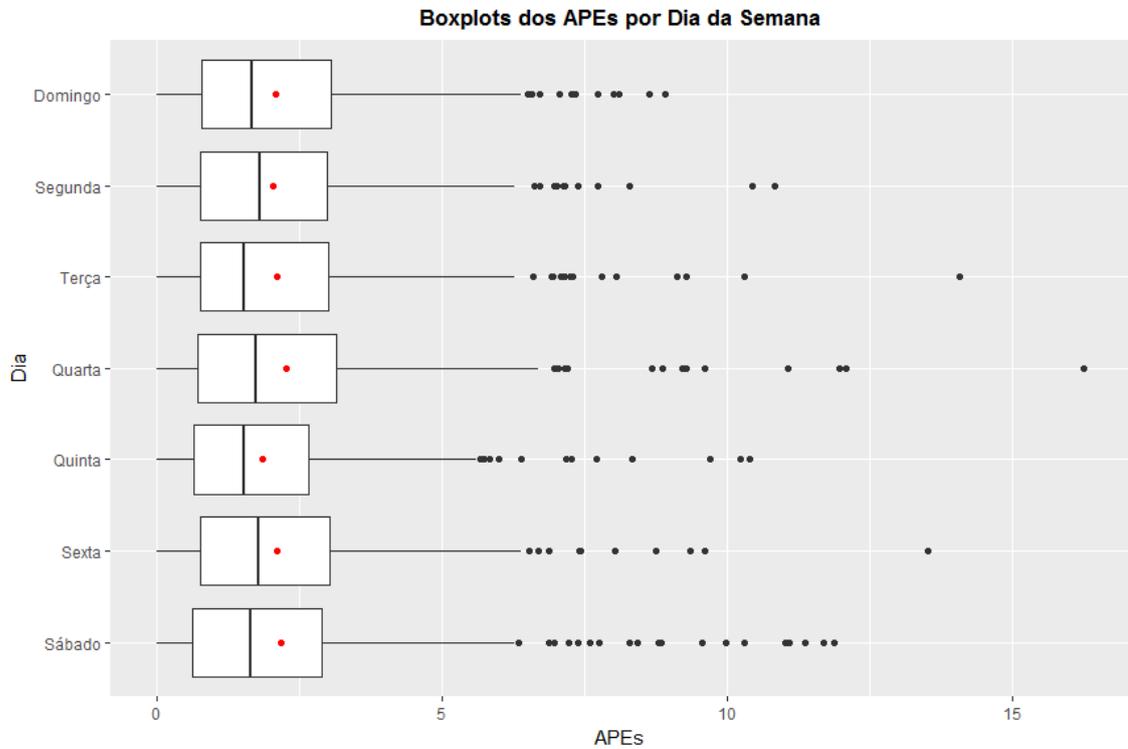


Figura 3.10: Boxplot dos APEs por Dia da Semana desconsiderando os dias do Natal de 1997.

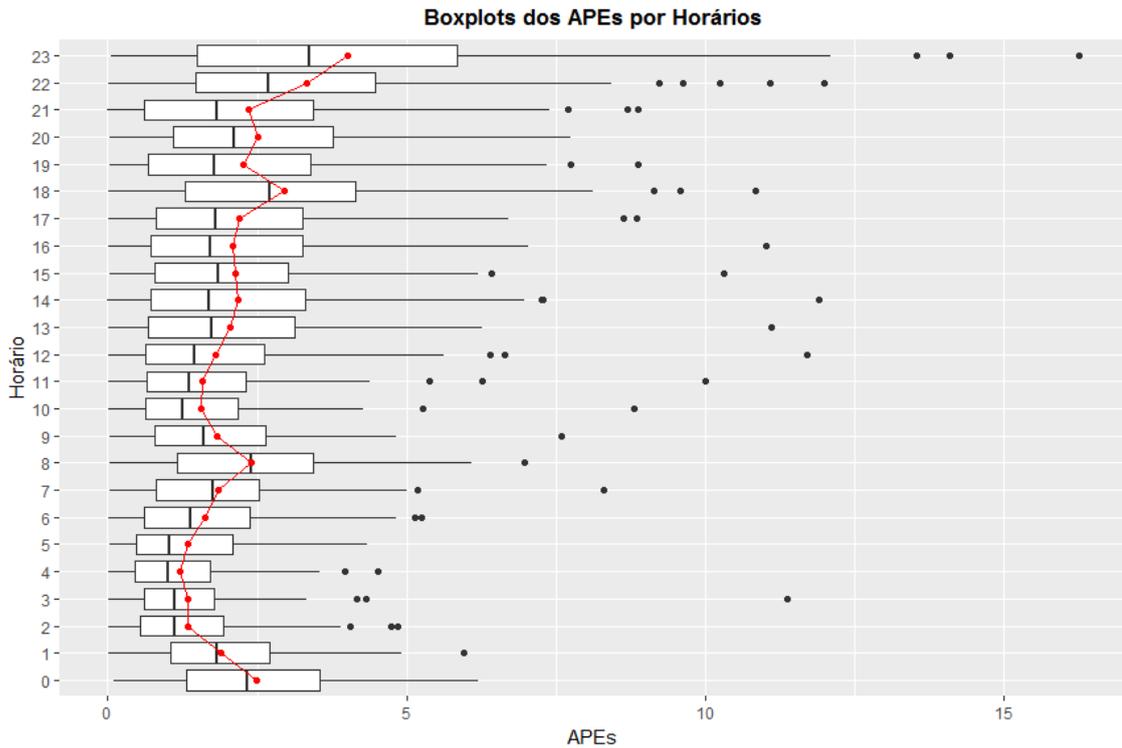


Figura 3.11: Boxplot dos APEs por Horário do Dia desconsiderando os dias do Natal de 1997.

Além de ser utilizada para a combinação das variáveis, a Regressão Simbólica pode ser utilizada também para descobrir relações entre as variáveis regressoras, ou no caso da combinação de previsões, para encontrar uma forma funcional para combinar as previsões. A função 3.2 sugere uma forma funcional do tipo

$$\Psi_F(\mathbf{f}) = \prod_{i=1}^9 f_i^{\beta_i}. \quad (3.3)$$

Para avaliar a Regressão Simbólica como um método de descobrir formas funcionais para a combinação de previsões, encontramos os β_i que minimizam o MSE na Amostra de Treino II, apresentados na Tabela 3.3. A estimativa desses parâmetros foi menor que o obtido pela Regressão Simbólica para a DSHW e maior para *Naïve* com 168 modelos em paralelo e Holt-Winters, além de terem sido obtidos $\beta_i > 0$ para *Naïve* com 24 modelos em paralelo e método de Holt, que não estavam presentes na melhor solução obtida por Regressão Simbólica. A partir destes β_i , fizemos a previsão pela Equação 3.3 para a Amostra de Teste e encontramos MAPE de 2,24% considerando a semana do Natal de 1997, e de 2,14% desconsiderando-a. Tais resultados não foram melhores que os encontrados pela Regressão Simbólica. Uma comparação mais detalhada entre as combinações é feita

a seguir.

Previsões	Sequencial	24 em paralelo				168 em paralelo			
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9
β_i	0,45	0,03	0,29	0,00	0,16	0,00	0,00	0,07	0,00

Tabela 3.3: Parâmetros estimados da forma funcional para as previsões individuais utilizando a Amostra de Treino II

3.3 Comparação com outros métodos de combinação de previsões

Os métodos de combinação de previsões apresentados na Seção 2.3.1 foram utilizados para gerar previsões para a Amostra de Teste. Para os métodos que precisam de estimação de parâmetros (Média Ponderada Ótima e Regressão Linear), foi utilizada a Amostra de Treino II, assim como no caso da Regressão Simbólica e da Forma Funcional descoberta através da Regressão Simbólica. Os resultados obtidos para os MAPEs e Variâncias dos APEs por estes métodos são apresentados na Tabela 3.4, considerando e não considerando a semana do Natal de 1997, podendo assim também haver comparabilidade com os trabalhos de Hippert (2001) e Hippert *et al.* (2005), que utilizaram Redes Neurais Artificiais para a previsão neste mesmo conjunto de dados e Carneiro (2014) que utilizou o método de combinação *outperformance* para combinar 2 a 2 os mesmos métodos de previsões individuais para esta série. Além destes, é apresentado o melhor resultado das previsões individuais, o DSHW, para fins de comparação.

Através das tabelas, é possível perceber que, independente de incluirmos a semana do Natal de 1997, a Regressão Simbólica obteve melhor MAPE que os demais métodos de combinação tanto em seu valor mínimo (2,18%), quanto na média (2,22%) e na mediana (2,19%). Considerando a semana do Natal, a pior Execução da Regressão Simbólica (2,34%) só apresentou MAPE maior que as combinações por Média Ponderada Ótima (2,31%) e por Regressão Linear (2,30%), sendo estes três os únicos métodos de combinação que conseguiram MAPE inferior ao obtido pelo DSHW (2,52%). Desconsiderando a semana do Natal de 1997, o pior MAPE obtido pela Regressão Simbólica (2,22%) não foi maior que o obtido por nenhum dos outros métodos de combinação.

Método	Com semana do Natal		Sem semana do Natal		
	MAPE (%)	Var(APE)	MAPE (%)	Var(APE)	
Média	3,82	8,92	3,78	8,69	
Mediana	3,96	10,20	3,93	9,97	
Média Podada	3,82	9,14	3,77	8,87	
Média dos Extremos	4,61	14,20	4,62	14,28	
Média Ponderada Ótima	2,31	4,27	2,22	3,43	
Média Geométrica	3,84	8,94	3,80	8,71	
Regressão Linear	2,30	4,27	2,22	3,43	
Regressão Simbólica	Mínimo	2,18	4,31	2,09	3,30
	Mediana	2,19	4,61	2,10	3,46
	Média	2,22	4,59	2,12	3,47
	Máximo	2,34	5,29	2,22	3,97
Forma Funcional		2,24	4,52	2,14	3,42
DSHW		2,52	7,30	2,38	5,30
DSHW (Carneiro, 2014)		2,47	7,15	-	-
<i>outperformance</i> (Carneiro, 2014)		2,48	7,15	-	-
Redes Neurais (Hippert, 2001)	Mínimo	2,26	-	-	-
	Média	2,44	-	-	-
	Máximo	2,75	-	-	-
Redes Neurais (Hippert <i>et al.</i> , 2005)	Mínimo	2,35	-	-	-
	Máximo	2,65	-	-	-

Tabela 3.4: Resultados para Combinações de Previsões na Amostra de Teste considerando ou não a semana do Natal de 1997

Além da média dos APEs, é importante analisar sua variância. Quando consideradas a semana do Natal de 1997, a variância dos APEs obtidos pelos métodos de Média Ponderada Ótima e de Regressão Linear (4,27) foi menor do que a Variância obtida por qualquer execução da Regressão Simbólica (4,31 a 5,29). No entanto, quando considerados apenas os dados anteriores à semana do Natal, a melhor execução da Regressão Simbólica obteve variância dos APEs (3,30) inferior aos demais métodos de combinação. Analisando a distribuição dos APEs através dos boxplots (Figuras 3.12 e 3.13), é possível perceber que os limites da caixa do boxplot têm seus menores valores para a previsão obtida pela melhor execução da Regressão Simbólica em ambos os casos. Além disto, é importante ressaltar que a combinação de previsões, através tanto da Regressão Simbólica quanto dos métodos lineares de Média Ponderada Ótima e de Regressão Linear, conseguiu reduzir substancialmente a variância dos APES obtida na previsão individual do DSHW (5,30).

Comparando com os resultados do trabalho de Carneiro (2014), todas as execuções da

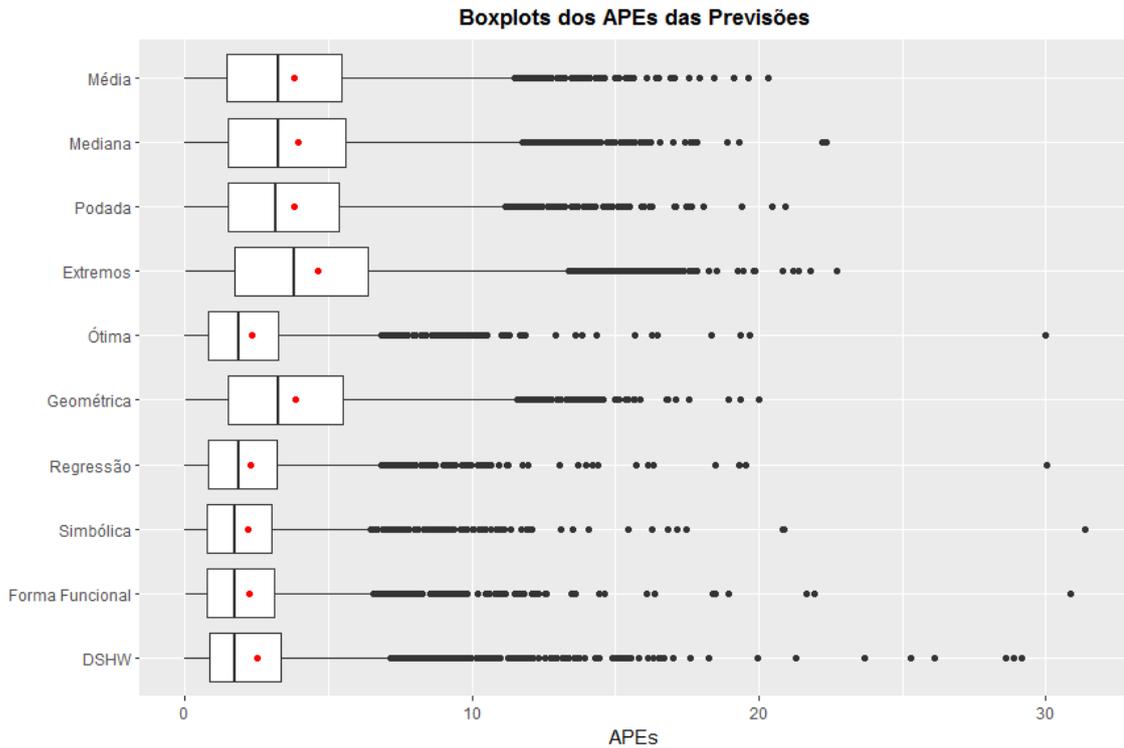


Figura 3.12: Boxplot dos APEs das Combinações de Previsões

Regressão Simbólica obtiveram resultados melhores do que os obtidos com o método de *outperformance*, quando comparados pelos MAPEs e pelas variâncias dos APES. Comparando com os métodos de previsões multivariáveis através de Redes Neurais de Hippert (2001) e Hippert (2001), a combinação através de Regressão Simbólica obteve valores Mínimos, Médios e Máximos inferiores aos obtidos no primeiro trabalho, tendo o maior valor de MAPE inferior ao menor valor obtido no segundo trabalho. Vale destacar que estes trabalhos utilizaram uma variável exógena, a temperatura, como um preditor da carga, enquanto os métodos utilizados neste trabalho utilizam apenas as observações anteriores da própria carga.

Além das previsões obtidas pela própria Regressão Simbólica, foram avaliadas as medidas de erro de previsão obtidas pela combinação através da Forma Funcional descoberta na melhor execução da Regressão Simbólica (3.3). Independente de considerarmos a semana do Natal de 1997, o MAPE obtido através desta previsão foi 0,02% superior à média dos MAPEs da Regressão Simbólica, e sempre inferior ao obtido pelos demais métodos de combinação. Além disto, a variância dos APEs obtida foi inferior a média das variâncias obtidas nas execuções da Regressão Simbólica. Comparada aos demais métodos, obteve uma variância maior quando consideramos a semana do Natal de

1997 e menor quando não consideramos.

Apesar de não ter obtido nenhuma medida de erro melhor que a melhor execução da Regressão Simbólica, tais resultados apontam para uma viabilidade da utilização desta forma funcional como método de combinação, devido aos seus resultados superiores quando comparada com os demais métodos e à sua natureza não estocástica, quando comparado com a Regressão Simbólica, o que pode trazer mais estabilidade às estimações e previsões obtidas.

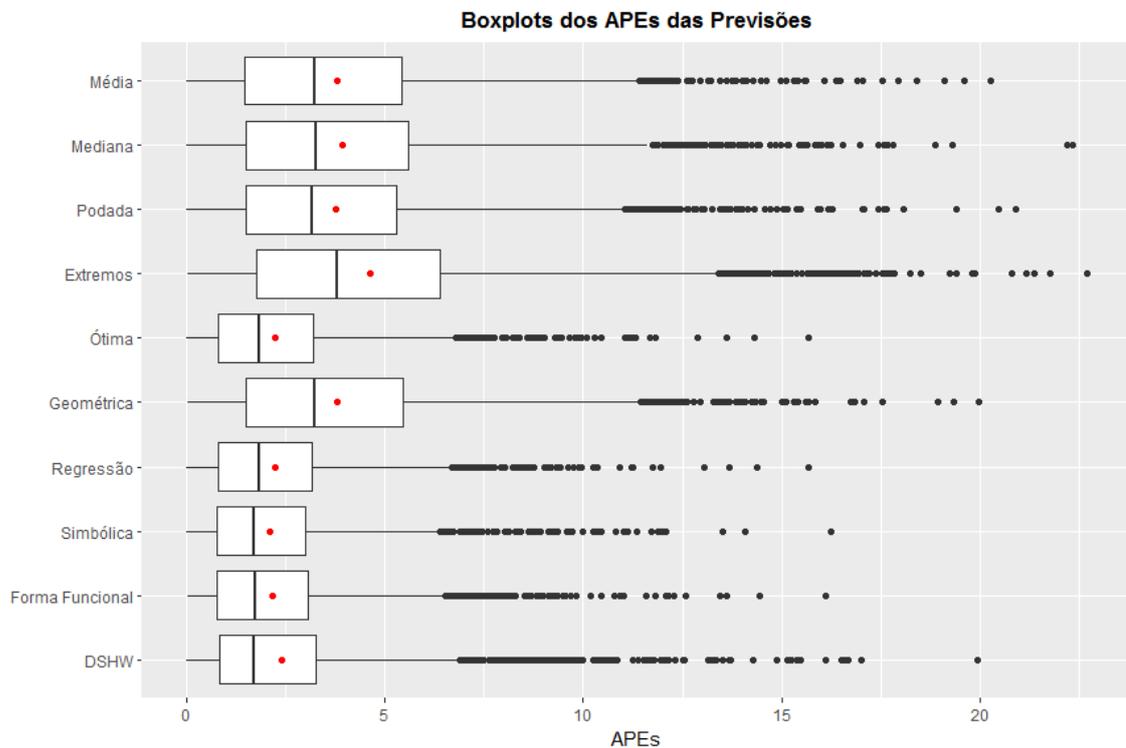


Figura 3.13: Boxplot dos APEs das Combinações de Previsões desconsiderando os dias do Natal de 1997.

3.3.1 Assimetria dos Resíduos de Previsão

Seguindo as recomendações em De Menezes *et al.* (2000), consideramos a assimetria ao avaliar e comparar as distribuições dos resíduos obtidos por cada método de combinação. Para isto, utilizamos o Coeficiente Momento de Assimetria (Tabela 3.5) e os histogramas dos resíduos (Figura 3.14).

Método	Coeficiente de Assimetria	
	Com semana do Natal	Sem semana do Natal
Média	0,04	0,08
Mediana	0,04	0,11
Média Podada	0,04	0,09
Média dos Extremos	-0,43	-0,47
Média Ponderada Ótima	0,80	0,41
Média Geométrica	-0,01	0,02
Regressão Linear	0,83	0,44
Regressão Simbólica	0,88	0,38
Forma Funcional	0,92	0,42
DSHW	0,93	0,43

Tabela 3.5: Coeficientes Momento de Assimetria dos Resíduos das Combinações de Previsões na Amostra de Teste considerando ou não a semana do Natal de 1997

Pelo coeficiente de assimetria, as combinações que se apresentaram mais simétricas foram a Média Geométrica (0,02), a Média (0,08), a Média Podada (0,09), e a Mediana (0,11); porém todas estas tiveram uma grande dispersão dos resíduos, como apresentado nos histogramas. Entre as combinações que apresentaram baixa dispersão, a Regressão Simbólica obteve o menor coeficiente de assimetria (0,38). Tal situação se inverte quando considerada a semana do Natal de 1997, resultando em um coeficiente de assimetria de 0,88 para a Regressão Simbólica, superior ao da Média Ponderada Ótima (0,80) e ao da Regressão Linear (0,83). Estes resultados mostram também o impacto de se considerar a semana do Natal para a simetria, pois com a exclusão destes valores atípicos, o coeficiente de assimetria é bastante reduzido para as combinações de previsão que obtiveram uma menor dispersão.

3.3.2 Autocorrelação dos Resíduos de Previsão

Outro critério a ser avaliado é a autocorrelação dos resíduos de previsão, visualizada através dos gráficos da FAC na Figura 3.15, onde se considera a semana do Natal de 1997, e na Figura 3.16, onde esta semana é desconsiderada.

Através destes gráficos, é possível perceber que em todas as combinações ainda há uma estrutura temporal que não é modelada, pois todos os métodos possuem uma autocorrelação significativa para diversas defasagens. Tal fato pode ocorrer caso as previsões individuais utilizadas na combinação não sejam capazes de, conjuntamente,



Figura 3.14: Histogramas dos Resíduos das diferentes Combinações de Previsões

modelar toda a estrutura temporal dos dados.

Nos gráficos observa-se que a combinação por Regressão Simbólica apresentou uma queda mais rápida da FAC que as demais, apesar de apresentar uma maior autocorrelação para *lags* de 24 horas que a previsão individual por DSHW. Quando comparada com a combinação por Regressão Linear, a Regressão Simbólica obteve menores valores de autocorrelação para todos os *lags*, indicando que a não-linearidade pode ser importante para se obter uma melhor combinação das estruturas temporais específicas de cada método de previsão individual, e modelar melhor a estrutura temporal presente nos dados.

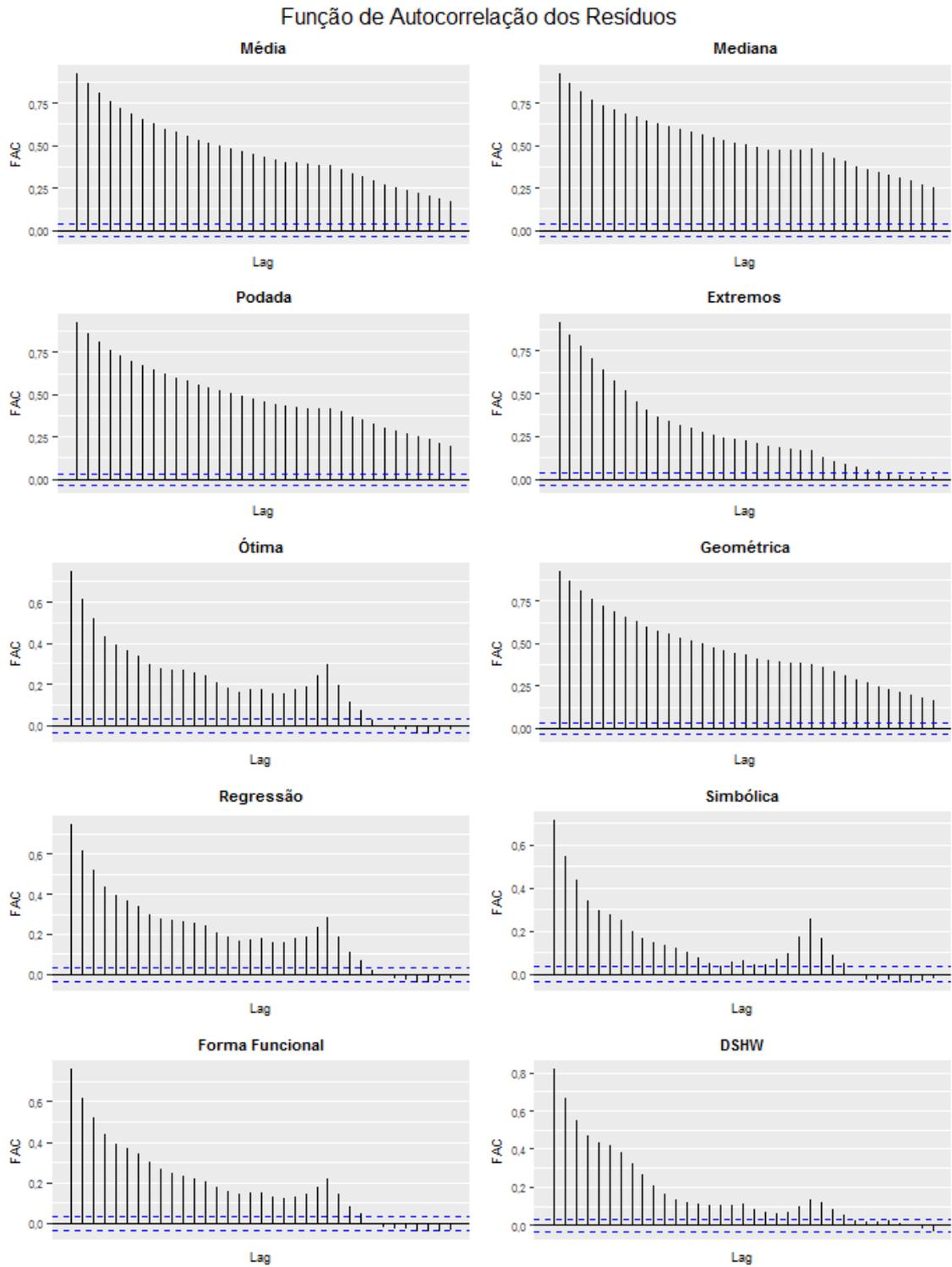


Figura 3.15: Gráficos das Funções de Autocorrelação dos Resíduos das Combinações de Previsões.

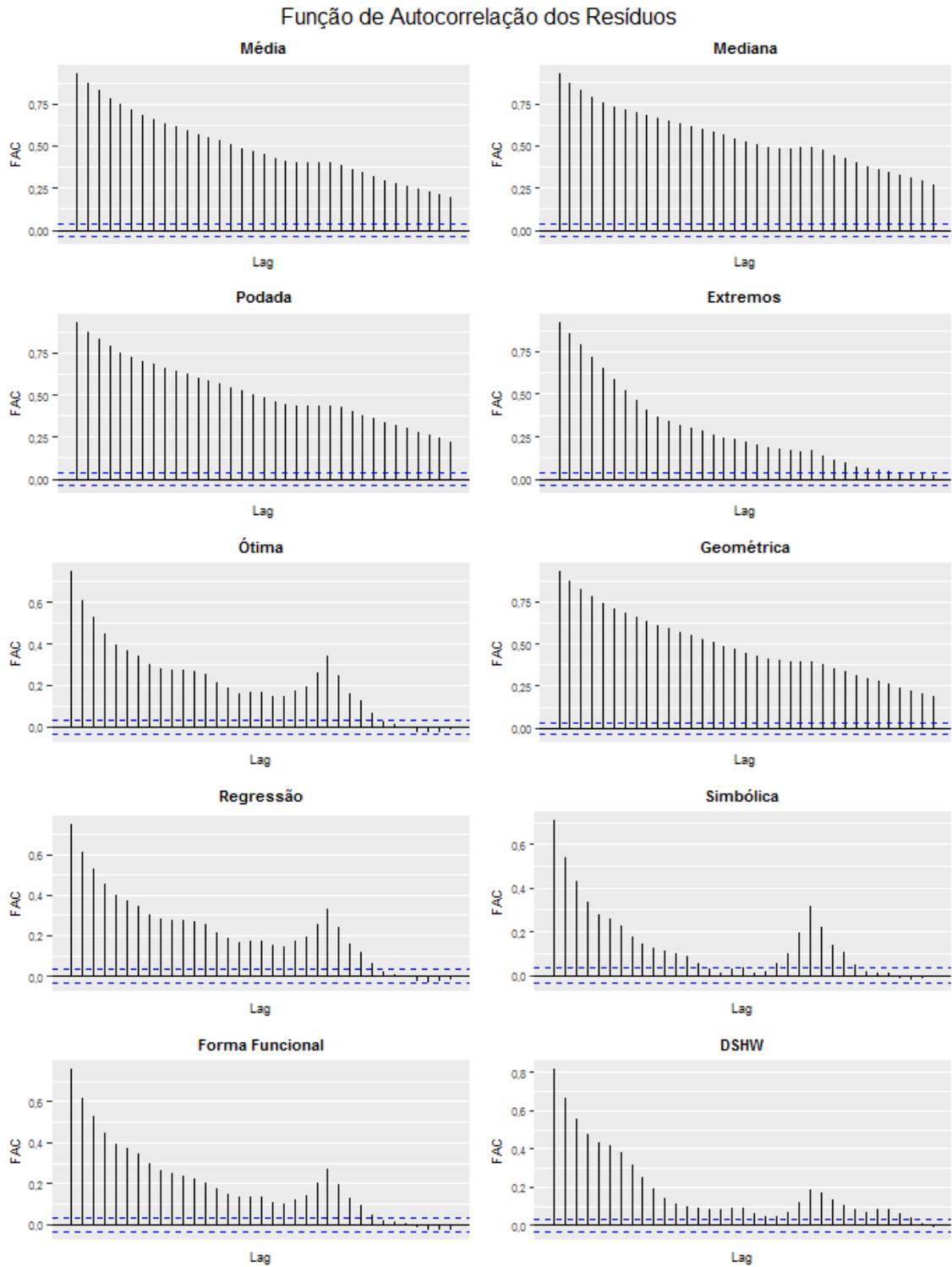


Figura 3.16: Gráficos das Funções de Autocorrelação dos Resíduos das Combinações de Previsões desconsiderando os dias do Natal de 1997.

Aplicar um teste de significância para a autocorrelação dos resíduos é desnecessário, pois vemos claramente nos gráficos das FACs que os resíduos de todas as previsões possuem autocorrelação serial. Ainda assim, podemos utilizar a estatística de Ljung-Box (Morettin & Toloi, 2006) para comparar quantitativamente a autocorrelação presente entre nas combinações de previsões. Na Tabela 3.6, vemos que a combinação por Regressão Simbólica apresentou menores valores da estatística de Ljung-Box considerando ou não a semana do Natal de 1997, o que indica que seus resíduos são menos autocorrelacionados.

Método	Estatística de Ljung-Box	
	Com semana do Natal	Sem semana do Natal
Média	2866,2	2777,2
Mediana	2860,6	2769,1
Média Podada	2874,2	2781,7
Média dos Extremos	2808,9	2722,8
Média Ponderada Ótima	1882,5	1804,6
Média Geométrica	2869,8	2779,6
Regressão Linear	1883,6	1806,9
Regressão Simbólica	1706,5	1617,2
Forma Funcional	1931,9	1839,9
DSHW	2248,5	2136,0

Tabela 3.6: Estatísticas de Ljung-Box para Autocorrelação dos Resíduos das Combinações de Previsões na Amostra de Teste considerando ou não a semana do Natal de 1997.

4 CONCLUSÕES

A previsão de cargas a curto prazo é um problema de relevância para o setor elétrico. Dos nove métodos de previsão individual aplicados para previsões em uma série de cargas do Rio de Janeiro, o método DSHW teve melhor desempenho que os demais, obtendo menores médias de erros e de variabilidade. Este foi o único método utilizado capaz de modelar simultaneamente duas componentes sazonais, uma diária e uma semanal.

Dos métodos usuais utilizados em combinações de previsões, apenas as combinações por Média Ponderada Ótima e por Regressão Linear conseguiram melhorar os resultados obtidos pela melhor previsão individual, sendo os restantes influenciados por previsões individuais ruins, obtendo erros superiores ao DSHW.

O método proposto de combinação não-linear através de Regressão Simbólica foi capaz de melhorar os resultados obtidos pelas previsões individuais e ainda que os resultados obtidos pelas combinações lineares, indicando assim que a não-linearidade pode ser aspecto importante a considerar em uma combinação de previsões. Na ausência de dados de dias especiais, a Regressão Simbólica apresentou o menor MAPE e a menor variabilidade dos APEs, um maior decaimento na função de autocorrelação e, comparada somente às combinações que conseguiram melhora nas medidas de erro, obteve o menor coeficiente de assimetria dos resíduos.

A função de combinação encontrada pela Regressão Simbólica foi capaz de identificar previsões ruins e excluí-las, além de identificar a relevância de cada previsão para a combinação, podendo ser utilizada como um método de seleção de atributos. Por conter previsões obtidas através das três estratégias de previsão utilizadas - sequencial e 24 e 168 modelos em paralelo - a função indica que as informações modeladas por cada uma das estratégias podem ser complementares.

A Regressão Simbólica também foi capaz de descobrir uma forma funcional não linear para combinação de previsões que obteve melhor resultado que os métodos de combinação lineares. Apesar de ter obtido piores resultados que a maioria das execuções da Regressão Simbólica, esta combinação tem a vantagem de não ser estocástica, proporcionando maior estabilidade nas previsões. A combinação através dessa forma funcional pode ser interessante de ser aplicada em problemas que há necessidade de previsões recorrentes,

devido ao seu baixo custo computacional quando comparada com a Regressão Simbólica.

A combinação por Regressão Simbólica obteve em todas as suas execuções resultados superiores à combinação de previsões individuais por *outperformance* de Carneiro (2014) e por Hippert *et al.* (2005), que utilizaram um método multivariado através de Redes Neurais. Comparado os resultados de Hippert (2001), o método proposto obteve menores valores mínimos, médios e máximos de MAPE.

Conseguimos, neste trabalho, propor uma inovação na abordagem do problema de combinação de previsões, introduzindo uma nova forma aplicação da Regressão Simbólica para o problema de previsão de carga a curto prazo. Esta aplicação mostrou resultados promissores e pode ser mais explorada em trabalhos posteriores.

Apesar destas vantagens, a combinação através de Regressão Simbólica utilizando estes previsores individuais ainda apresenta assimetria e autocorrelação nos resíduos. Este último problema é um indício que há estrutura temporal nos dados que não foi captada pela modelagem. Sugestões de trabalhos futuros incluem a inclusão de mais previsões individuais para a combinação, para tentar capturar a estrutura temporal não alcançada nesta aplicação, além da aplicação de combinação de previsões através de outros métodos não-lineares (ex: Redes Neurais Artificiais, Florestas Aleatórias e Algoritmos de *Boosting*).

Além destas, há a possibilidade de melhorias no algoritmo de Regressão Simbólica implementado. Das funções base utilizada, apenas duas apareceram no resultado final. Pode ser aumentado o número de funções base, ou podem ser utilizados métodos para preservar a diversidade de funções na população. Outro ponto passível de melhoria é a geração de constantes; no resultado final não houve nenhum nó constante, o que pode ser devido a um mecanismo ruim de geração de constantes. As alternativas incluem a estimação das constantes através de procedimentos de otimização em cada novo indivíduo, que pode aumentar muito o custo computacional, ou a utilização de outras distribuições de probabilidade para a geração de constantes.

REFERÊNCIAS

- Alkhatib, Wael, Alhamoud, Alaa, Böhnstedt, Doreen, & Steinmetz, Ralf. 2017. *Hybrid Models for Short-Term Load Forecasting Using Clustering and Time Series*. Cham: Springer International Publishing. Pages 104–115.
- Almeida, Álvaro Augusto de. 2007. *O Livre Mercado de Energia Elétrica Brasileiro Parte I*. Disponível em <http://raceadm3.nuca.ie.ufrj.br/buscarace/Docs/alaalmeida1.pdf>. Acessado em 28/04/2017.
- Alvarez, Francisco Martinez, Troncoso, Alicia, Riquelme, Jose C, & Ruiz, Jesus S Aguilar. 2011. Energy time series forecasting based on pattern sequence similarity. *IEEE Transactions on Knowledge and Data Engineering*, **23**(8), 1230–1243.
- Armstrong, J.Scott. 1989. Combining forecasts: The end of the beginning or the beginning of the end? *International Journal of Forecasting*, **5**(4), 585 – 588.
- Arora, Siddharth, & Taylor, James W. 2013. Short-term forecasting of anomalous load using rule-based triple seasonal methods. *IEEE transactions on Power Systems*, **28**(3), 3235–3242.
- Augusto, Douglas Adriano, & Barbosa, Helio JC. 2000. Symbolic regression via genetic programming. *Pages 173–178 of: Neural Networks, 2000. Proceedings. Sixth Brazilian Symposium on*. IEEE.
- Banzhaf, Wolfgang, Nordin, Peter, Keller, Robert E, & Francone, Frank D. 1998. *Genetic programming: an introduction*. Vol. 1. Morgan Kaufmann San Francisco.
- Barmpalexis, P, Kachrimanis, K, Tsakonas, A, & Georgarakis, E. 2011. Symbolic regression via genetic programming in the optimization of a controlled release

- pharmaceutical formulation. *Chemometrics and Intelligent Laboratory Systems*, **107**(1), 75–82.
- Bates, J. M., & Granger, C. W. J. 1969. The Combination of Forecasts. *OR*, **20**(4), 451–468.
- Box, George Edward Pelham, & Jenkins, Gwilym M. 1994. *Time Series Analysis: Forecasting and Control*.
- Byrd, Richard H, Lu, Peihuang, Nocedal, Jorge, & Zhu, Ciyou. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, **16**(5), 1190–1208.
- Carneiro, Anna Cláudia Mancini da Silva. 2014. *Previsão do consumo de energia elétrica a curto prazo, usando combinações de métodos univariados*. Dissertação de Mestrado, Universidade Federal de Juiz de Fora (UFJF).
- Castelli, Mauro, Trujillo, Leonardo, & Vanneschi, Leonardo. 2015. Energy consumption forecasting using semantic-based genetic programming with local search optimizer. *Computational intelligence and neuroscience*, **2015**, 57.
- Chatfield, Chris. 1996. Model uncertainty and forecast accuracy. *Journal of Forecasting*, **15**(7), 495–508.
- Clemen, Robert T. 1989. Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, **5**(4), 559–583.
- De Menezes, Lilian M, Bunn, Derek W, & Taylor, James W. 2000. Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, **120**(1), 190–204.
- Dietterich, Thomas G. 2000. Ensemble methods in machine learning. *Pages 1–15 of: International workshop on multiple classifier systems*. Springer.
- Dudek, Grzegorz. 2016. Pattern-based local linear regression models for short-term load forecasting. *Electric Power Systems Research*, **130**, 139 – 147.
- Eiben, Agoston E, Smith, James E, *et al.* . 2003. *Introduction to evolutionary computing*. Vol. 53. Springer.

- Empresa de Pesquisa Energética. 2015. *Consumo anual de energia elétrica por classe (nacional) 1995-2014*. <http://www.epe.gov.br/mercado/Paginas/Consumonacionaldeenergiael%C3%A9tricaporclasse%E2%80%932009.aspx>. Acessado em 28/04/2017.
- Flasch, Oliver. 2015. *A Modular Genetic Programming System*. Ph.D. thesis, Technischen Universität Dortmund.
- Flasch, Oliver, Mersmann, Olaf, & Bartz-beielstein, Thomas. 2014. *RGP: An Open Source Genetic Programming System for the R Environment*.
- Ghareeb, WT, & El Saadany, EF. 2013. Multi-Gene Genetic Programming for Short Term Load Forecasting. *Pages 1–5 of: Electric Power and Energy Conversion Systems (EPECS), 2013 3rd International Conference on*. IEEE.
- Gujarati, Damodar, & Porter, D. 2003. Multicollinearity: What happens if the regressors are correlated. *Basic econometrics*, **363**.
- Hippert, Henrique Steinherz, Pedreira, Carlos Eduardo, & Souza, Reinaldo Castro. 2001. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems*, **16**(1), 44–55.
- Hippert, HS. 2001. *Previsão de cargas a curto prazo: uma avaliação da viabilidade do uso de redes neurais artificiais*. Ph.D. thesis, Tese de Doutorado, PUC-Rio, Departamento de Engenharia Elétrica.
- Hippert, HS, Bunn, DW, & Souza, RC. 2005. Large neural networks for electricity load forecasting: Are they overfitted? *International Journal of forecasting*, **21**(3), 425–434.
- Holland, John H. 1975. *Adaptation in natural and artificial systems. An introductory analysis with application to biology, control, and artificial intelligence*. Ann Arbor, MI: University of Michigan Press.
- Hyndman, Rob J. 2016. *forecast: Forecasting functions for time series and linear models*. R package version 7.3.
- Hyndman, Rob J, & Khandakar, Yeasmin. 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, **26**(3), 1–22.

- Koza, John R. 1989. Hierarchical Genetic Algorithms Operating on Populations of Computer Programs. *Pages 768–774 of: IJCAI*.
- Koza, John R. 1992. *Genetic programming: on the programming of computers by means of natural selection*. Vol. 1. MIT press.
- Lee, Yi-Shian, & Tong, Lee-Ing. 2011a. Forecasting energy consumption using a grey model improved by incorporating genetic programming. *Energy Conversion and Management*, **52**(1), 147–152.
- Lee, Yi-Shian, & Tong, Lee-Ing. 2011b. Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. *Knowledge-Based Systems*, **24**(1), 66–72.
- López, Cristian, Zhong, Wei, & Zheng, MengLian. 2017. Short-term Electric Load Forecasting Based on Wavelet Neural Network, Particle Swarm Optimization and Ensemble Empirical Mode Decomposition. *Energy Procedia*, **105**, 3677 – 3682. 8th International Conference on Applied Energy, ICAE2016, 8-11 October 2016, Beijing, China.
- Manna, Nimai Sundar, Choudhury, Binoy Krishna, & Das, Rajib. 2017. Relationship between Electricity Consumption and Ambient Temperature: A Non-Linear Time Varying Approach. *EVIDENCE BASED MANAGEMENT*, 265.
- Mayrink, V., & Hippert, H. S. 2016 (Nov). A hybrid method using Exponential Smoothing and Gradient Boosting for electrical short-term load forecasting. *Pages 1–6 of: 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*.
- Morettin, Pedro A, & Toloí, Clélia. 2006. *Análise de séries temporais*. Blucher.
- Neto, Guilherme G, Defilippo, Samuel B, & Hippert, Henrique S. 2015. Univariate versus Multivariate Models for Short-term Electricity Load Forecasting. *In: XIII Simposio Argentino de Investigación Operativa (SIO)-JAIIO 44 (Rosario, 2015)*.
- Pappas, S Sp, Ekonomou, L, Karampelas, P, Karamousantas, DC, Katsikas, SK, Chatzarakis, GE, & Skafidas, PD. 2010. Electricity demand load forecasting of the Hellenic power system using an ARMA model. *Electric Power Systems Research*, **80**(3), 256–264.

- Poli, Riccardo, Langdon, William B, McPhee, Nicholas F, & Koza, John R. 2008. *A field guide to genetic programming*. Lulu. com.
- R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reid, David J. 1968. Combining three estimates of gross domestic product. *Economica*, **35**(140), 431–444.
- Reid, David J. 1969. *A comparative study of time series prediction techniques on economic data*. University of Nottingham, Library Photographic Unit.
- Rodrigues, Filipe, Cardeira, Carlos, & Calado, J. M. F. 2017. *Neural Networks Applied to Short Term Load Forecasting: A Case Study*. Cham: Springer International Publishing. Pages 173–197.
- Shi, Shan Ming, Da Xu, Li, & Liu, Bao. 1999. Improving the accuracy of nonlinear combined forecasting using neural networks. *Expert Systems with Applications*, **16**(1), 49–54.
- Shi, Shanming, & Liu, Bao. 1993. Nonlinear combination of forecasts with neural networks. *Pages 959–962 of: Neural Networks, 1993. IJCNN'93-Nagoya. Proceedings of 1993 International Joint Conference on*, vol. 1. IEEE.
- Tarsitano, Agostino, & Amerise, Ilaria L. 2017. Short-term load forecasting using a two-stage sarimax model. *Energy*, **133**, 108 – 114.
- Taylor, J. W., & McSharry, P. E. 2007. Short-Term Load Forecasting Methods: An Evaluation Based on European Data. *IEEE Transactions on Power Systems*, **22**(4), 2213–2219.
- Taylor, James W. 2003. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, **54**(8), 799–805.
- Taylor, James W. 2010. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, **204**(1), 139–152.

- Taylor, James W., de Menezes, Lilian M., & McSharry, Patrick E. 2006. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, **22**(1), 1 – 16.
- Zhang, Bing, Wu, Jhen-Long, & Chang, Pei-Chann. 2017a. A multiple time series-based recurrent neural network for short-term load forecasting. *Soft Computing*, 1–14.
- Zhang, W., Hua, H., & Cao, J. 2017b (April). Short Term Load Forecasting Based on IGSA-ELM Algorithm. *Pages 296–301 of: 2017 IEEE International Conference on Energy Internet (ICEI)*.
- Zou, Hui, & Yang, Yuhong. 2004. Combining time series models for forecasting. *International Journal of Forecasting*, **20**(1), 69–84.