

Universidade Federal de Juiz de Fora  
Instituto de Ciências Exatas / Faculdade de Engenharia  
Programa de Pós-Graduação em Modelagem Computacional

**Iago Mansur de Azevedo**

**Previsão de carga elétrica a curto prazo combinando métodos lineares  
univariados e variáveis exógenas através de algoritmos genéticos**

Juiz de Fora

2019

Iago Mansur de Azevedo

**Previsão de carga elétrica a curto prazo combinando métodos lineares univariados e variáveis exógenas através de algoritmos genéticos**

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, na área de concentração Interdisciplinar, como requisito parcial para obtenção do título de Mestre em Modelagem Computacional.

Orientador: Henrique Steinherz Hippert

Juiz de Fora

2019

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

Azevedo, Iago Mansur de.

Previsão de carga elétrica a curto prazo combinando métodos lineares univariados e variáveis exógenas através de algoritmos genéticos / Iago Mansur de Azevedo. – 2019.

74 f. : il.

Orientador: Henrique Steinherz Hippert

Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas / Faculdade de Engenharia. Programa de Pós-Graduação em Modelagem Computacional, 2019.

1. Séries temporais. 2. Amortecimento exponencial. 3. Algoritmos genéticos. 4. Combinação de previsões. 5. Previsão de cargas. I. Hippert, Henrique Steinherz, orient. II. Título.

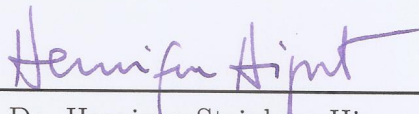
Iago Mansur de Azevedo

Previsão de carga elétrica a curto prazo combinando métodos lineares univariados e variáveis exógenas através de algoritmos genéticos

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, na área de concentração Interdisciplinar, como requisito parcial para obtenção do título de Mestre em Modelagem Computacional.

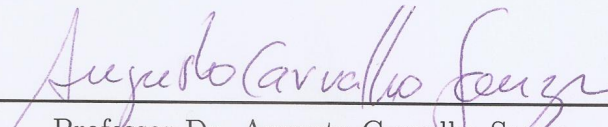
Aprovada em: 13 de março de 2019

BANCA EXAMINADORA



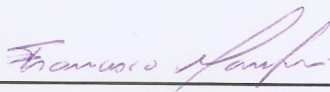
---

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



---

Professor Dr. Augusto Carvalho Souza  
Universidade Federal de Juiz de Fora



---

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

## AGRADECIMENTOS

À minha família, em especial minha mãe (Jacqueline) e minha namorada (Camila).

Ao orientador Henrique.

Aos professores, funcionários, amigos e colegas do Programa de Pós-Graduação em Modelagem Computacional.

À CAPES e à UFJF por apoiar neste estudo.

A todos que contribuíram, direta ou indiretamente, para essa trajetória.

*“There’s more to be seen than can ever be seen  
More to do than can ever be done”*

JOHN, E.; RICE, T. Circle of Life. Santa Monica, CA: BOP Recording Studios, 1994.  
(4:51).

## RESUMO

As previsões da carga elétrica a curto prazo são cruciais para a operação dos sistemas energéticos, pois permitem otimizar os custos de operação e evitar desperdícios. Além disso, as previsões são necessárias para estabelecer o preço da energia no mercado e para definir os contratos de fornecimento. Neste trabalho, desenvolvemos um sistema baseado na combinação de previsões. Primeiramente, são feitas previsões univariadas por meio de quatro métodos de amortecimento exponencial (extensões do método de Holt-Winters); em seguida, estas previsões são combinadas linearmente entre si, e também com algumas funções não-lineares de variáveis exógenas (especialmente a temperatura), por meio de algoritmos genéticos. Fizemos experimentos com duas implementações diferentes destes algoritmos e testamos o sistema sobre uma base de dados reais, obtidos na cidade do Rio de Janeiro - RJ. Os resultados obtidos indicaram que a introdução de variáveis exógenas não melhorou a acurácia das previsões; além disso, que a combinação usando algoritmos genéticos não conseguiu previsões melhores do que as feitas usando um método de otimização já bem conhecido (no caso, o método de Nelder-Mead). O que indica que os métodos baseados em inteligência computacional podem não apresentar resultados melhores, em todas as situações, que os dos métodos lineares usuais e que a sua aplicação deve ser bem justificada, em cada problema particular.

Palavras-chave: Séries temporais. Amortecimento exponencial. Algoritmos genéticos. Combinação de previsões. Previsão de cargas.

## ABSTRACT

Short-term load forecasts are crucial to the operation of energy systems, since they allow optimizing operating costs and avoiding waste. In addition, the forecasts are needed to set the price of energy in the market and to define supply contracts. In this dissertation, we developed a system based on the combination of forecasts. First, univariate forecasts are made by four exponential smoothing methods (extensions of the Holt-Winters method); then, these forecasts are combined linearly with each other, and also with some non-linear functions of exogenous variables (especially temperature), through genetic algorithms. We performed experiments with two different implementations of these algorithms and tested the system on a real database, obtained in the city of Rio de Janeiro. The results indicated that the introduction of exogenous variables did not improve the accuracy of the forecasts; in addition, that the combination using genetic algorithms did not get better forecasts than that made using an already well-known optimization method (in this case, the Nelder-Mead method). This shows that methods based on computational intelligence may not achieve better results, in all circumstances, than the usual linear methods and that their application must be well justified in each particular problem.

Key-words: Time series. Exponential smoothing. Genetic algorithms. Combination of forecasts. Short-term load forecasting.



## LISTA DE ILUSTRAÇÕES

Figura 1 – Esquema de uma SEE . . . . .	14
Figura 2 – Características das séries temporais . . . . .	18
Figura 3 – Quadro comparativo entre a seleção natural e computação evolutiva . . . . .	25
Figura 4 – Ilustração da configuração genética de um indivíduo . . . . .	26
Figura 5 – Fluxograma de um algoritmo evolutivo . . . . .	28
Figura 6 – Cruzamento de um e “ $n$ ” pontos . . . . .	34
Figura 7 – Cruzamento uniforme . . . . .	34
Figura 8 – Ilustração da recombinação por representação com valores reais . . . . .	35
Figura 9 – Mutação para a representação binária . . . . .	36
Figura 10 – Distribuição da população em um algoritmo evolucionista . . . . .	39
Figura 11 – Progresso do maior valor de aptidão ao passar das gerações . . . . .	39
Figura 12 – Séries de cargas e temperaturas do Rio de Janeiro, de 1996 a 1997 . . . . .	40
Figura 13 – Série de cargas para duas semanas (de 01/07/1996 a 14/07/1996), mostrando a sazonalidade semanal e sazonalidade diária . . . . .	41
Figura 14 – Diagrama de dispersão da cargas em função da temperatura, para os dados da Carolina de Norte (EUA), durante o período de 01/01/2004 a 07/07/2008 . . . . .	42
Figura 15 – Esquema da partição 1 e 2, para a combinação 1 . . . . .	44
Figura 16 – Esquema da partição 1, para a combinação 2 . . . . .	45
Figura 17 – Previsor <i>naïve</i> , no período de 10/11/1997 a 16/11/1997 . . . . .	46
Figura 18 – Método de Holt-Winters-Taylor, sazonalidade aditiva, entre o período de 10/11/1997 a 16/11/1997 . . . . .	47
Figura 19 – Fator sazonal diário e fator sazonal para uma semana . . . . .	49
Figura 20 – Fluxograma do AG . . . . .	55
Figura 21 – Componentes de um indivíduo no AG . . . . .	56
Figura 22 – Iterações do AG . . . . .	65
Figura 23 – Gráfico de iterações sem restrição . . . . .	67
Figura 24 – Boxplot do AG com dez séries e combinação 2 . . . . .	69

## LISTA DE TABELAS

Tabela 1 – Partição das séries de cargas e de temperaturas . . . . .	43
Tabela 2 – Representação do AG . . . . .	56
Tabela 3 – Tamanho do gene . . . . .	57
Tabela 4 – Configuração dos experimentos . . . . .	62
Tabela 5 – Análise do passo . . . . .	63
Tabela 6 – Análise dos Experimentos . . . . .	63
Tabela 7 – Resultados (MAPE [%]) . . . . .	64
Tabela 8 – Resultados (MAPE [%]) da combinação 1 com as funções do R . . . . .	66
Tabela 9 – Comparação (MAPE [%]) dos resultados em relação a restrição, combinação 1 . . . . .	67
Tabela 10 – Resultados (MAPE [%]) para as combinações 1 e 2, com quatro e dez séries . . . . .	68
Tabela 11 – Resultados (MAPE [%]), com dez séries para a combinação 2 . . . . .	69

## LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agência Nacional de Energia Elétrica
SEE	Sistema de Energia Elétrica
AG	Algoritmo Genético
AES	Amortecimento Exponencial Simples
AEH	Amortecimento Exponencial de Holt
HW	Amortecimento Exponencial de Holt-Winters
HWT	Amortecimento Exponencial de Holt-Winters-Taylor
APE	Erro Percentual Absoluto ( <i>Absolute Percentage Error</i> )
MAPE	Erro Percentual Absoluto Médio ( <i>Mean Absolute Percentage Error</i> )
HWT1	Holt-Winters-Taylor Original com Sazonalidade Aditiva
HWT2	Holt-Winters-Taylor com Duas Sazonalidades Diárias e Uma Sazonalidade Semanal
HWT3	Holt-Winters-Taylor com Três Sazonalidades Diárias e Uma Sazonalidade Semanal
HWT4	Holt-Winters-Taylor com Três Sazonalidades Diárias, sem Sazonalidade Semanal
RNA	Redes Neurais Artificiais
ARIMA	Auto-Regressivo Integrado de Médias Móveis

## LISTA DE SÍMBOLOS

$Y$	Valor observado
$l$	Nível da série
$e$	Erro aleatório
$t$	Tempo
$N$	Distribuição normal
$\sigma$	Desvio-padrão
$X$	Valor previsto
$\alpha, \lambda, \beta, \rho, \delta, \omega, \phi$	Constantes de amortecimento
$b$	Tendência
$k$	Passo a frente do instante de previsão
$r$	Fator sazonal
$d$	Fator sazonal diário
$w$	Fator sazonal semanal
$s, s_1, s_2$	Período: sazonal, diário, semanal
$n$	Número genérico / número de elementos da série
$p$	Probabilidade
$c$	Comprimento
$z, u, v, \nu$	Variáveis genéricas
$i$	Índice / Variável <i>dummy</i>
$\nu$	Coefficiente genérico
$\theta$	Coefficiente da combinação
$T$	Temperatura
$T'$	Diferença de temperatura
$q$	Parâmetro
$\Gamma$	Intervalo do domínio

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>13</b>
1.1	NATUREZA DO PROBLEMA . . . . .	14
1.2	OBJETIVOS . . . . .	15
1.3	RELEVÂNCIA DO ESTUDO . . . . .	16
1.4	ORGANIZAÇÃO DO TRABALHO . . . . .	16
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA . . . . .</b>	<b>17</b>
2.1	MÉTODOS DE PREVISÃO . . . . .	17
2.1.1	<b>Modelos de amortecimento exponencial . . . . .</b>	<b>17</b>
2.1.1.1	<i>Amortecimento exponencial simples (AES) . . . . .</i>	<i>18</i>
2.1.1.2	<i>Amortecimento exponencial de Holt (AEH) . . . . .</i>	<i>19</i>
2.1.1.3	<i>Amortecimento exponencial de Holt-Winters (HW) . . . . .</i>	<i>20</i>
2.1.1.3.1	Modelo sazonal multiplicativo . . . . .	21
2.1.1.3.2	Modelo sazonal aditivo . . . . .	22
2.1.1.4	<i>Amortecimento exponencial de Holt-Winters-Taylor (HWT) . . . . .</i>	<i>23</i>
2.2	ALGORITMOS EVOLUTIVOS . . . . .	25
2.3	ALGORITMO GENÉTICO . . . . .	29
2.3.1	<b>Representação . . . . .</b>	<b>29</b>
2.3.2	<b>Função de aptidão . . . . .</b>	<b>31</b>
2.3.3	<b>População . . . . .</b>	<b>31</b>
2.3.4	<b>Mecanismo de Seleção dos pais . . . . .</b>	<b>32</b>
2.3.5	<b>Operadores de Variação . . . . .</b>	<b>33</b>
2.3.5.1	<i>Recombinação . . . . .</i>	<i>33</i>
2.3.5.2	<i>Mutação . . . . .</i>	<i>36</i>
2.3.6	<b>Mecanismo de sobrevivência . . . . .</b>	<b>37</b>
2.3.7	<b>Inicialização e condição de parada . . . . .</b>	<b>38</b>
2.3.8	<b>Comportamento típico de um AG . . . . .</b>	<b>39</b>
<b>3</b>	<b>METODOLOGIA . . . . .</b>	<b>40</b>
3.1	SÉRIE . . . . .	40
3.2	PARTIÇÃO DAS SÉRIES . . . . .	43
3.2.1	<b>Combinação 1 . . . . .</b>	<b>43</b>
3.2.2	<b>Combinação 2 . . . . .</b>	<b>44</b>
3.3	MÉTODOS DE PREVISÃO . . . . .	45
3.3.1	<b>Previsor <i>Naïve</i> . . . . .</b>	<b>45</b>
3.3.2	<b>Previsor Holt-Winters-Taylor . . . . .</b>	<b>47</b>

3.3.2.1	<i>Holt-Winters-Taylor original</i> . . . . .	47
3.3.2.2	<i>Holt-Winters-Taylor com duas sazonalidades diárias e uma sazonalidade semanal</i> . . . . .	49
3.3.2.3	<i>Holt-Winters-Taylor com três sazonalidades diárias e uma sazonalidade semanal</i> . . . . .	51
3.3.2.4	<i>Holt-Winters-Taylor com três sazonalidades diárias, sem a sazonalidade semanal</i> . . . . .	53
3.4	<b>ALGORITMO GENÉTICO</b> . . . . .	55
3.4.1	<b>Representação</b> . . . . .	55
3.4.2	<b>População</b> . . . . .	57
3.4.3	<b>Função de aptidão</b> . . . . .	57
3.4.4	<b>Mecanismos de seleção dos pais</b> . . . . .	59
3.4.5	<b>Operadores de variação</b> . . . . .	59
3.4.5.1	<i>Recombinação</i> . . . . .	60
3.4.5.2	<i>Mutação</i> . . . . .	60
3.4.6	<b>Mecanismo de sobrevivência</b> . . . . .	60
3.4.7	<b>Inicialização e Condição de parada</b> . . . . .	61
3.4.8	<b>Definição dos parâmetros do AG</b> . . . . .	61
4	<b>RESULTADOS E DISCUSSÕES</b> . . . . .	64
4.1	<b>VALIDAÇÃO DOS MODELOS</b> . . . . .	64
4.2	<b>EXPERIMENTOS DA COMBINAÇÃO COM DEZ SÉRIES</b> . . . . .	66
4.3	<b>COMPARAÇÃO ENTRE MÉTODOS LINEARES E RNAS</b> . . . . .	70
5	<b>CONCLUSÃO</b> . . . . .	72
	<b>REFERÊNCIAS</b> . . . . .	73

## 1 INTRODUÇÃO

A oferta de energia viabilizou o crescimento e as transformações da economia mundial, tanto que atualmente o consumo de energia pode ser considerado um dos principais indicadores de desenvolvimento social. Porém, devido aos impactos ambientais causados pelo crescimento populacional, foi necessária uma nova visão mundial: a necessidade do desenvolvimento sustentável. A indústria de energia elétrica tem contribuído significativamente para o desenvolvimento tecnológico e o aumento da qualidade de vida do ser humano, por isso, seria inviável reduzir a oferta de energia, “O equilíbrio entre oferta e demanda não é alcançado apenas aumentando a oferta. É possível e desejável atuar também pelo lado da demanda” (ANEEL, 2008).

O mercado de energia brasileiro passou por duas grandes mudanças recentes, a primeira foi em 1996 com as privatizações das operadoras, que instituiu a Agência Nacional de Energia Elétrica, (ANEEL), e a segunda em 2004, com a introdução do “Novo Modelo do Setor Elétrico”, tendo como um dos seus principais objetivos a garantia da segurança no suprimento de energia. O governo brasileiro tem se mostrado interessado em estudos de previsão e projeções para o mercado de energia, tanto para construção e desativação de usinas, bem como numa visão de curto prazo para definir os preços a serem regulamentados. Como os leilões de concessão são baseados no menor preço, métodos de previsão podem ser um grande diferencial em propostas mais arrojadas (ANEEL, 2008).

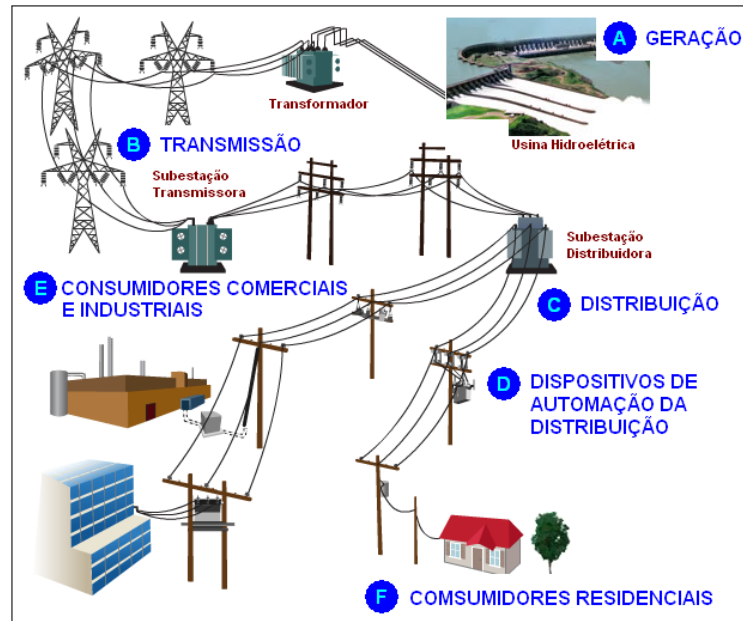
Do ponto de vista das concessionárias de energia, com o aumento de padrões de qualidades cobrados sobre as empresas do setor, sob pena de multas de órgãos regulatórios, é necessário adotar uma gestão energética mais eficiente, no intuito de reduzir desperdícios, bem como custos (NETO, 2014).

Para grandes consumidores, são firmados contratos entre a concessionária e clientes, diferentemente dos pequenos consumidores, que têm suas tarifas regulamentadas pelo governo (TEIXEIRA, 2017). Assim, a previsão com base no consumo de energia elétrica de cada cliente, estabelece as cláusulas dos contratos. Esses se baseiam, nos tipos de tarifa e perfil de consumo, essa decisão será mantida durante toda a vigência do contrato, o que exige uma confiabilidade na previsão. Outra questão que os métodos de previsão auxiliam no mercado de energia, é com relação a programação de produção da concessionária, visto, que são grandes instalações e não permitem uma resposta rápida de acordo com as flutuações de consumo de mercado, desta forma, a concessionária deve ter uma programação para permitir a operação sem que falte energia no mercado e principalmente evitar desperdícios de produção no intuito de aumentar seu lucro.

Para um melhor entendimento do contexto desse estudo, a Figura 1 apresenta um esquema geral de um Sistema de Energia Elétrica (SEE). Esse sistema se inicia através das usinas (“A”, Figura 1) onde a energia elétrica é gerada e posteriormente é distribuída

nas redes de transmissão (“B”, Figura 1). Nessa parte, a energia elétrica trafega, em alta voltagem, até chegar às subestações das distribuidoras (“C”, Figura 1), onde a tensão é rebaixada por meio de transformadores. Posteriormente como indicado na Figura 1 (D), a energia é distribuída para os consumidores, que se dividem em: consumidores comerciais e industriais (“E”, Figura 1) e consumidores residenciais (“F”, Figura 1).

Figura 1 – Esquema de uma SEE



Fonte: Bauer (2017)

## 1.1 NATUREZA DO PROBLEMA

Os métodos de previsão são divididos, em geral, pelo o horizonte de previsão. Embora não haja um consenso quanto a classificação de horizontes nas publicações, neste estudo, segue-se a classificação de Hong e Fan (2016), que distingue em quatro categorias:

- Muito curto prazo: previsões de frações de horas até um dia a frente;
- Curto prazo: previsões após um dia, até duas semana a frente;
- Médio prazo: previsão para mais de duas semana, até três anos;
- Longo prazo: previsão para mais de três anos.

Embora nas publicações recentes já exista a diferenciação entre curto prazo e muito curto prazo, para manter a coerência com os outros estudos feitos com essa mesma série de carga, nesse estudo “curto prazo” irá abranger “muito curto prazo”.

A previsão de carga elétrica a curto prazo é crucial para a programação do sistema de energia, para otimizar custos de operação e melhorar a confiança do sistema de distribuição



(ARORA; TAYLOR, 2013). Tudo isso está ligado às questões financeiras, o que aumenta a cobrança por modelos mais acurados. Na maioria dos sistemas produtivos, podem-se armazenar os produtos para mitigar os efeitos da sazonalidade no consumo de mercado; contudo, na produção de energia, essa estratégia não é possível, devido às limitações técnicas para o armazenamento de energia elétrica. Assim o excesso de energia se converte em perdas do processo e a produção abaixo do necessário se converte em prejuízos para o consumidor, como “apagões” (NETO, 2014).

Com o planejamento, é possível identificar quando a demanda de carga está acima da capacidade de geração de uma usina e buscar outras fontes de energia, além das hidrelétricas, para evitar a falta de energia aos consumidores. Contudo, essas outras fontes podem ser mais caras, e com os modelos de previsão é possível mitigar esse ônus, visto que a distribuidora terá tempo para se programar e atender a demanda sem a necessidade de comprar energia emergencial de outras fontes de energia, como de termoeletricas, que são mais caras. As fontes de energia estão classificadas em duas categorias e podem ser, segundo ANEEL (2008):

#### 1. Fontes renováveis:

- Hidráulica/Hidrelétrica;
- Biomassa;
- Eólica;
- Solar;
- Biogás;
- Geotérmica;
- Mar.

#### 2. Fontes não-renováveis:

- Gás natural;
- Derivados de Petróleo;
- Nuclear;
- Carvão Mineral.

## 1.2 OBJETIVOS

O objetivo geral deste trabalho é propor um método de previsão para uma série de carga. O objetivo específico é combinar um conjunto de previsores e a série de temperatura através de um algoritmo genético (AG), no intuito de reduzir o erro entre a série original e a previsão.

### 1.3 RELEVÂNCIA DO ESTUDO

Do ponto de vista operacional e comercial, a previsão de carga a curto prazo é essencial em um gerenciamento efetivo do comércio de carga elétrica (SONG et al., 2006). Do ponto de vista acadêmico, pode se justificar que trabalhos dessa linha são oportunidades para testar diferentes métodos e avaliar se apresentam melhorias. Embora já existam muitos métodos de previsão, as séries de carga, em geral, apresentam um conjunto de características que esses métodos não são capazes de modelar (concomitantemente) como será explicado na seção 3.1, exemplo: não linearidade, uma parcela de comportamento aleatório da carga, condições climáticas, etc. Assim torna-se necessário avaliar novos métodos e possivelmente estabelecer novas técnicas na literatura.

Como cada metodologia de previsão possui vantagens e desvantagens, não existe uma metodologia perfeita de previsão; cada série vai depender de um estudo específico para aprimorar sua previsão. Contudo, para aproveitar as vantagens de vários previsores e mitigar a desvantagem de cada método, pode-se fazer combinação destes na busca de melhorar a previsão; segundo Hong e Fan (2016), essa técnica é a melhor prática para previsão de carga.

### 1.4 ORGANIZAÇÃO DO TRABALHO

Esse estudo está estruturado em cinco capítulos, são eles:

1. Introdução: neste capítulo são apresentadas as ideias iniciais, além do contexto do estudo, bem como os objetivos do mesmo;
2. Revisão bibliográfica: onde são apresentados os conceitos introdutórios relacionados com a metodologia utilizada no estudo;
3. Metodologia: são apresentados os métodos de estudo, materiais e relação destes para encontrar os resultados do estudo;
4. Resultados e discussões: neste capítulo são apresentados os resultados encontrados e são realizadas discussões a partir destes;
5. Conclusão: neste capítulo final, são feitas conclusões de todo o estudo e são apresentadas sugestões de trabalhos futuros.

## 2 REVISÃO BIBLIOGRÁFICA

Neste capítulo, dividido em três tópicos, são apresentados os conceitos teóricos para o entendimento da metodologia a ser aplicada no Capítulo 3. O primeiro apresenta os conceitos de previsão, em particular os previsores por amortecimento exponencial. O segundo tópico apresenta os conceitos gerais dos algoritmos evolutivos de forma a esclarecer e introduzir o terceiro tópico que é o AG, usado para otimizar a combinação dos previsores.

### 2.1 MÉTODOS DE PREVISÃO

Segundo Morettin e Toloí (2006): “uma série temporal é qualquer conjunto de observações ordenadas no tempo”, entretanto, Ehlers (2009) explica que a variável tempo pode ser substituída por outras variáveis como: espaço, profundidade, etc; que também destaca que ordem das observações são cruciais, o que não acontece para modelos de regressão linear por exemplo. Estas séries podem ser discretas (exemplo: consumo mensal de energia elétrica) ou contínuas (exemplo: série de temperaturas), vale notar que séries contínuas podem ser armazenadas como séries discretas, se forem definidos intervalos iguais para a tomada dos dados, como exemplo a série de temperaturas registrada a cada hora do dia.

Para Morettin e Toloí (2006) “[...] um modelo é uma descrição probabilística de uma série temporal[...]”, e após selecionado o modelo da série, é possível obter as previsões dos valores futuros das séries com base nos métodos de previsão adequados ao modelo selecionado. Neste trabalho serão apresentados previsores paramétricos, no domínio temporal, como o amortecimento exponencial. De um modo geral, as previsões das séries podem ser de curto prazo, como para séries de vendas, produção ou estoque, ou a longo prazo, como para séries populacionais, de produtividade, etc. Neste estudo, as previsões trabalhadas são de curto prazo.

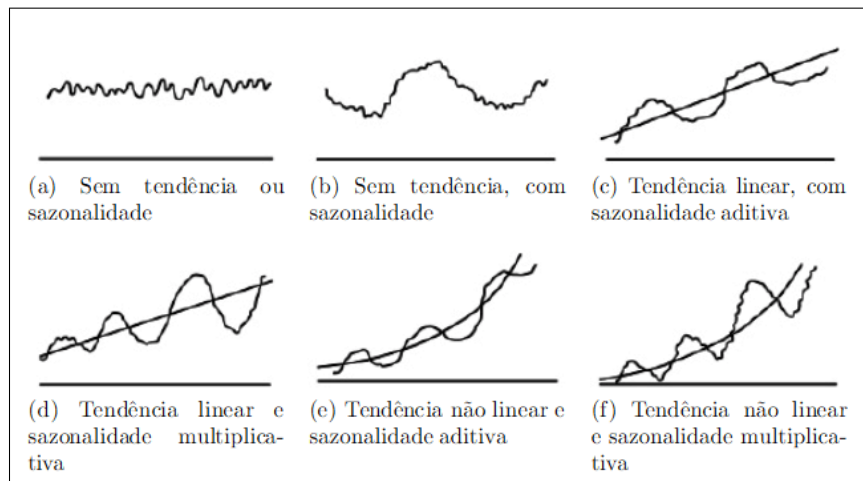
#### 2.1.1 Modelos de amortecimento exponencial

Os métodos de previsão se baseiam no pressuposto, que as observações feitas fornecem informações sobre um padrão de comportamento da série temporal (MORETTIN; TOLOÍ, 2006). Os amortecimentos são técnicas que interpretam que os valores de uma série representam aleatoriedade, e amortecendo esses valores pode-se encontrar um padrão para a série. Ou seja, caso o valor da série apresente uma grande variação de um instante para o outro, com a utilização do amortecimento exponencial, a previsão para o valor seguinte não será condenada pela variação entre os dois instantes anteriores. Então para saber qual método de previsão utilizar, é necessário notar que as séries podem apresentar três características, são elas:

- nível (constante ou variável);
- Tendência;
- Sazonalidade.

Cada uma dessas características podem ser observadas na Figura 2 que ilustra seis séries teóricas. Os diferentes métodos de previsão que podem ser aplicados nessas séries, serão explorados nas seções seguintes. Apesar de sua simplicidade na aplicação, os modelos de amortecimento exponencial apresentam uma dificuldade para encontrar suas constantes. Suas estimativas podem ser achadas por otimização, de forma a minimizar o erro.

Figura 2 – Características das séries temporais



Fonte: Pichiliani (2015)

#### 2.1.1.1 Amortecimento exponencial simples (AES)

Considere o modelo de nível constante como descrito na Equação 2.1.

$$Y_t = l_t + e_t \quad (2.1)$$

Onde,

$Y_t$ : valor observado;

$l_t$ : nível da série;

$e_t$ : ruído observado,  $e_t \sim N(0, \sigma^2)$ ;

$t$ : tempo.

Pode-se dizer que a AES é uma média ponderada que dá pesos maiores às observações mais recentes, em relação às observações mais antigas. A previsão dos valores futuros é feita a partir do último valor amortecido, calculado conforme a Equação 2.2.

$$\hat{X}_{t+1} = \alpha \cdot Y_t + (1 - \alpha) \cdot \hat{X}_t \quad (2.2)$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$\alpha$ : constante de amortecimento, com seu valor variando no intervalo  $[0, 1]$ ;

$t$ : tempo;

Desta forma, não é necessário armazenar toda a série para fazer a previsão de um instante  $t$ , bastando para o cálculo a última observação ( $Y_{t-1}$ ), a previsão imediatamente anterior ( $\hat{X}_{t-1}$ ) e o coeficiente  $\alpha$ . Contudo, pode-se demonstrar que a previsão está sempre alerta a mudanças no nível de série (MORETTIN; TOLOI, 2006).

Sobre o valor de  $\alpha$ , quanto menor seu valor, mais estável será a previsão final. Uma vez que a utilização de baixo valor do coeficiente implica que pesos maiores serão dados às previsões passadas e, conseqüentemente, qualquer flutuação aleatória no presente exercerá um peso menor no cálculo de previsão. Em geral, quanto mais aleatória for a série estudada, menores serão os valores das constantes de amortecimento (MORETTIN; TOLOI, 2006). Seu valor ideal é encontrando quando o valor da previsão apresentar o menor erro em relação aos dados já obtidos.

#### 2.1.1.2 Amortecimento exponencial de Holt (AEH)

Considere uma série temporal não sazonal que é composta localmente da soma do nível, tendência, resíduo aleatório e seu valor observado em um instante  $t$ , calculado conforme Equação 2.3.

$$Y_t = l_t + b_t + e_t \quad (2.3)$$

Onde,

$Y_t$ : valor observado;

$l_t$ : nível da série;

$b_t$ : tendência da série;

$e_t$ : ruído observado,  $e_t \sim N(0, \sigma^2)$ ;

$t$ : tempo.

Para as séries deste tipo, a AES não é uma boa alternativa, pois, irá fornecer previsões que subestimam (ou superestimam) continuamente os valores observados. Então, é necessário aplicar o AEH. Este método, além de amortecer o nível como o AES, inclui também uma constante para modelar a tendência da série. A Equação 2.4 dá a previsão  $k$  passos a frente.

$$\begin{aligned}\hat{l}_t &= \lambda \cdot Y_t + (1 - \lambda) \cdot (\hat{l}_{t-1} + \hat{b}_{t-1}) \\ \hat{b}_t &= \beta \cdot (\hat{l}_t - \hat{l}_{t-1}) + (1 - \beta) \cdot \hat{b}_{t-1}\end{aligned}\tag{2.4}$$

$$\hat{X}_{t+k} = \hat{l}_t + \hat{b}_t \cdot k$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

$\hat{l}_t$ : estimativa do nível para o instante  $t$ ;

$\hat{b}_t$ : estimativa da tendência para o instante  $t$ ;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$\beta, \lambda$ : constantes de amortecimento, com valores variando no intervalo  $[0, 1]$ ;

$t$ : tempo;

$k$ : passo a frente do instante  $t$ .

Assim, para a implementação do método, é necessário atribuir os valores iniciais, que podem ser:  $b_2 = Y_2 - Y_1$  e  $l_2 = Y_2$ .

As séries com tendência também podem ser previstas pela amortecimento exponencial linear de Brown (tendência linear) e a amortecimento quadrática de Brown (tendência quadrática). Estes não serão apresentados neste trabalho, pois não foram realizados testes com eles.

### 2.1.1.3 Amortecimento exponencial de Holt-Winters (HW)

Para séries com sazonalidade, pode-se usar a metodologia de HW. Esse método se baseia em três equações com constantes de amortecimento que variam de acordo com as características da série. Observa-se que a série é composta de três componentes: nível, tendência e sazonalidade. Há dois tipos de métodos HW, dependendo da forma como o fator sazonal será usado.

### 2.1.1.3.1 Modelo sazonal multiplicativo

Neste modelo o fator sazonal multiplica o nível da série. A Equação 2.5 representa o modelo.

$$Y_t = (l_t + b_t) \cdot r_t + e_t \quad (2.5)$$

Onde,

$Y_t$ : valor observado;

$l_t$ : nível da série;

$b_t$ : tendência da série;

$r_t$ : fator sazonal;

$e_t$ : ruído observado,  $e_t \sim N(0, \sigma^2)$ ;

$t$ : tempo.

Neste caso, o conjunto de Equações 2.6 apresenta o cálculo das estimativas das componentes da série, considerando  $s$  como o período sazonal.

$$\begin{aligned} \hat{l}_t &= \lambda \cdot \left( \frac{Y_t}{\hat{r}_{t-s}} \right) + (1 - \lambda) \cdot (\hat{l}_{t-1} + \hat{b}_{t-1}) \\ \hat{b}_t &= \beta \cdot (\hat{l}_t - \hat{l}_{t-1}) + (1 - \beta) \cdot \hat{b}_{t-1} \\ \hat{r}_t &= \rho \cdot \left( \frac{Y_t}{\hat{l}_t} \right) + (1 - \rho) \cdot \hat{r}_{t-s} \end{aligned} \quad (2.6)$$

$$\hat{X}_{t+k} = (\hat{l}_t + \hat{b}_t \cdot k) \cdot \hat{r}_{t-s+k}$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

$\hat{l}_t$ : estimativa do nível para o instante  $t$ ;

$\hat{b}_t$ : estimativa da tendência para o instante  $t$ ;

$\hat{r}_t$ : estimativa do fator sazonal para o instante  $t$ ;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$\beta, \lambda, \rho$ : constantes de amortecimento, com valores variando no intervalo  $[0, 1]$ ;

$t$ : tempo;

$s$ : período sazonal;

$k$ : passos a frente do instante  $t$ .

As expressões em 2.7, apresentam maneiras de se inicializarem os valores para realizar a previsão de  $\hat{X}_{t+k}$  (MORETTIN; TOLOI, 2006).

$$r_j = \frac{Y_j}{\left(\frac{1}{s}\right) \cdot \sum_{k=1}^s Y_k}, \quad j = 1, 2, \dots, s \quad l_s = \frac{1}{s} \cdot \sum_{k=1}^s Y_k; \quad l_s = 0. \quad (2.7)$$

### 2.1.1.3.2 Modelo sazonal aditivo

Diferentemente do processo anterior, para a sazonalidade aditiva, o fator sazonal é somado ao nível e tendência. O modelo da série pode ser representado segundo a Equação 2.8.

$$Y_t = l_t + b_t + r_t + e_t \quad (2.8)$$

Onde,

$Y_t$ : valor observado;

$l_t$ : nível da série;

$b_t$ : tendência da série;

$r_t$ : fator sazonal;

$e_t$ : ruído observado,  $e_t \sim N(0, \sigma^2)$ ;

$t$ : tempo.

Como no caso anterior, a série contém três componentes: fator sazonal, nível e tendência que são calculados pelo conjunto de Equações 2.9, além da previsão “ $k$ ” passos a frente  $\hat{X}_{t+k}$ .

$$\hat{l}_t = \lambda \cdot (Y_t - \hat{l}_{t-s}) + (1 - \lambda) \cdot (\hat{l}_{t-1} + \hat{b}_{t-1})$$

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

$$\hat{r}_t = \rho \cdot (Y_t - \hat{l}_t) + (1 - \rho) \cdot \hat{r}_{t-s}$$

$$\hat{X}_{t+k} = \hat{l}_t + \hat{b}_t \cdot k + \hat{r}_{t-s+k}$$

(2.9)



Onde,

$Y_t$ : valor observado no instante  $t$ ;

$\hat{l}_t$ : estimativa do nível para o instante  $t$ ;

$\hat{b}_t$ : estimativa da tendência para o instante  $t$ ;

$\hat{r}_t$ : estimativa do fator sazonal para o instante  $t$ ;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$\beta, \lambda, \rho$ : constantes de amortecimento, com valores variando no intervalo  $[0, 1]$ ;

$t$ : tempo;

$s$ : período sazonal;

$k$ : passos a frente do instante  $t$ .

#### 2.1.1.4 Amortecimento exponencial de Holt-Winters-Taylor (HWT)

O HW se apresenta como um bom candidato a método de previsão, mas considera apenas uma sazonalidade. Se a série apresentar duas sazonalidades (diária e semanal, por exemplo), não é a melhor opção. Taylor (2003) modificou o HW para atender mais uma sazonalidade, conforme o modelo 2.10

$$Y_t = (l_t + b_t) \cdot d_t \cdot w_t + e_t \quad (2.10)$$

Onde,

$Y_t$ : valor observado;

$l_t$ : nível da série;

$b_t$ : tendência da série;

$d_t$ : sazonalidade diária da série;

$w_t$ : sazonalidade semanal da série;

$e_t$ : ruído observado,  $e_t \sim N(0, \sigma^2)$ ;

$t$ : tempo.

Segundo Taylor (2003), a versão multiplicativa é mais comum de ser usada. A formulação feita por ele, para a atualização dos componentes, pode ser vista nas Equações 2.11, onde deve-se considerar,  $s_1$  e  $s_2$ , respectivamente: período diário e período semanal.

$$\begin{aligned}
\hat{l}_t &= \lambda \cdot \left( \frac{Y_t}{\hat{d}_{t-s_1} \cdot \hat{w}_{t-s_2}} \right) + (1 - \lambda) \cdot (\hat{l}_{t-1} + \hat{b}_{t-1}) \\
\hat{b}_t &= \beta \cdot (\hat{l}_t - \hat{l}_{t-1}) + (1 - \beta) \cdot \hat{b}_{t-1} \\
\hat{d}_t &= \delta \cdot \left( \frac{Y_t}{\hat{l}_t \cdot \hat{w}_{t-s_2}} \right) + (1 - \delta) \cdot \hat{d}_{t-s_1} \\
\hat{w}_t &= \omega \cdot \left( \frac{Y_t}{\hat{l}_t \cdot \hat{d}_{t-s_1}} \right) + (1 - \omega) \cdot \hat{w}_{t-s_2}
\end{aligned} \tag{2.11}$$

$$\hat{X}_{t+k} = (\hat{l}_t + \hat{b}_t \cdot k) \cdot \hat{d}_{t-s_1+k} \cdot \hat{w}_{t-s_2+k}$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

$\hat{l}_t$ : estimativa do nível para o instante  $t$ ;

$\hat{b}_t$ : estimativa da tendência para o instante  $t$ ;

$\hat{d}_t$ : estimativa do fator sazonal diário para o instante  $t$ ;

$\hat{w}_t$ : estimativa do fator sazonal semanal para o instante  $t$ ;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$\beta, \delta, \lambda, \omega$ : constantes de amortecimento, com valores variando no intervalo  $[0, 1]$ ;

$t$ : tempo;

$s_1$ : período diário;

$s_2$ : período semanal;

$k$ : passos a frente do instante  $t$ .

Para Taylor (2003), o método para sazonalidade aditiva pode ser implementado de forma similar, também com dupla sazonalidade. Assim, como o método pode ser estendido de forma parecida, em ambos os casos (sazonalidade aditiva e multiplicativa), para três ou mais sazonalidades.

## 2.2 ALGORITMOS EVOLUTIVOS

Nesta seção, será apresentado o contexto teórico dos algoritmos evolutivos com ênfase no algoritmo genético (AG), baseado em Eiben e Smith (2015). A Computação Evolutiva é um ramo da ciência da computação inspirada por processos naturais vistos na teoria de Darwin, com aplicações em diversos problemas. Para os problemas de otimização, sua motivação se estabelece naturalmente como: “*the power of evolution in nature is evident in the diverse species that make up our world, each tailored to survive well in its own niche.*”(EIBEN; SMITH, 2015).

Pelo princípio da “evolução natural”, em um determinado ambiente, região geográfica, existe uma população de indivíduos que ocupam esse espaço e se adaptam a ele, para sobrevivência e reprodução. O grau de adaptação pode ser medido pela *aptidão*, que mede a relação entre o indivíduo e o ambiente em que vive; assim é possível avaliar as chances de sobrevivência de cada indivíduo e o desenvolvimento de sua prole, atingindo suas necessidades de forma a perpetuar sua espécie naquele ambiente. Este conceito pode ser interpretada para o ambiente computacional conforme Figura 3.

Figura 3 – Quadro comparativo entre a seleção natural e computação evolutiva

Evolução	Solução do Problema
Ambiente	Problema
Indivíduo	Candidato a solução
Aptidão	Qualidade

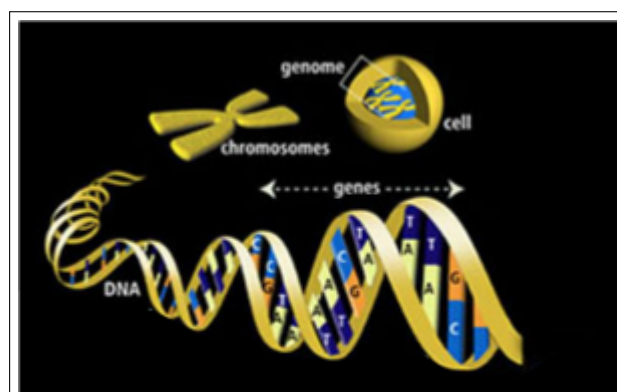
**Fonte:** adaptado de Eiben e Smith (2015)

Pelo conceito de seleção natural de Darwin (1859 *apud* EIBEN; SMITH, 2015), dado um ambiente limitado a um certo número de indivíduos, se estiverem adaptados, a reprodução irá ocorrer naturalmente, assim a seleção se torna inevitável para definir quais indivíduos perpetuaram, caso a população não tenha um crescimento exponencial. A seleção natural favorece os indivíduos que competem efetivamente pelos recursos existentes, ou seja, aqueles que estão adaptados ou ajustados às condições do ambiente de melhor maneira. Assim, competição é uma dos dois principais fatores para o progresso evolutivo, além da variação das características dos indivíduos entre os membros da população.

Os conceitos de fenótipo e genótipo são necessários no entendimento das técnicas. Para entender o fenótipo, é preciso de uma visão microscópica, no interior da célula, onde está o genoma (Figura 4) que compreende toda a informação genética do indivíduo. No genoma estão armazenados os *cromossomos*, que são formados por conjuntos de genes. Os genes podem ser encontrados em diferentes formas, cada forma é chamado de *alelo*. O conjunto dessa configuração definirá todas as características do indivíduo e essa informação é transmitida dos pais para os filhos através dos cromossomos. Então, pode-se dizer que os indivíduos são duas entidades: por fora fenótipo e por dentro *genótipo*, que se trata da

composição genética de um indivíduo como descrito.

Figura 4 – Ilustração da configuração genética de um indivíduo



Fonte: Traxco (2010)

Os fenótipos são comportamentos e características do indivíduo que afetam diretamente sua resposta ao ambiente (inclusive seu convívio com os outros indivíduos), assim determinando sua adaptação. Cada indivíduo representa uma combinação única do fenótipo, que é avaliada pelo ambiente. Esta avaliação sendo favorável, o indivíduo tem grandes chances de procriar e o mesmo fenótipo irá se perpetuar nos indivíduos da geração seguinte. Durante o processo de reprodução podem ocorrer pequenas variações de fenótipo, chamadas *mutações*.

Resumindo a ideia básica do modelo, a população consiste em um número de indivíduos. Eles formam unidades de seleção, que irão reproduzir entre si. Seu sucesso em reproduzir vai depender do quão bem um indivíduo está adaptado ao ambiente em relação aos outros indivíduos. Ocasionalmente ocorrerão mutações que gerarão novos indivíduos para serem testados. Com o tempo, a população irá sofrer alterações, assim a população se torna uma unidade de evolução.

Na computação evolutiva, a combinação das características dos pais em indivíduos da prole, é chamada de “cruzamento”. Neste mecanismo, onde serão trocadas as partes dos cromossomos dos pais, são escolhidos aleatoriamente os pontos onde os cromossomos sofrerão quebras e serão recombinados. A consequência dessa abordagem é que o material genético da população só poderá surgir das variações aleatórias e seleção natural, e não do aprendizado do indivíduo. Se faz necessário lembrar que as variações ocorrem a nível do *genótipo* e a seleção ocorre da adaptação do indivíduo ao sistema, ou seja a nível do *fenótipo*. Dessa forma os indivíduos da prole serão distintos dos pais, pois são formados pela combinação dos pais e de variações na reprodução. Assim essa variação irá aparecer no fenótipo dos novos indivíduos, que serão avaliados pela seleção.

Cada geração de indivíduos será uma nova amostra de todas as possibilidades de indivíduos daquele espaço, esses serão formados pela variação de suas características. Esta técnica pode ser pensada como um algoritmo para gerar e testar possibilidades de

respostas a problemas. E esta técnica pode ser repetida várias vezes até encontrar uma melhor resposta para o problema em estudo.

A ideia básica de todas as técnicas é a mesma: dada uma população de indivíduos em mesmo ambiente com recursos limitados, estes estarão em constante competição que será a causa da seleção natural (sobrevivência do mais apto), assim, após muitas gerações a população tente a permanecer com as mesmas características, com pouca variação no fenótipo dos indivíduos. Assim, é fácil entender que os mais adaptados da população irão se perpetuar e desenvolver a população de indivíduos. Dada uma função de qualidade para ser maximizada, pode-se criar aleatoriamente um conjunto de candidatos a solução, ou seja, elementos do domínio da função.

Aplica-se a função de qualidade para estes, como uma medida geral de aptidão; quanto maior seu valor, mais apto é o indivíduo. Assim, os melhores candidatos são escolhidos para gerar a próxima geração de indivíduos, através de recombinação e mutação. A *recombinação* é um operador que se aplicada a dois indivíduos da população selecionados (pais), que produzem um ou mais candidatos (filhos). Já a *mutação* é outro operador, que se aplica a um indivíduo da nova geração (filhos). Assim, aplicando estes operadores, será criado um grupo de novos candidatos (prole). Estes são analisados pela função de aptidão e competem com os indivíduos da geração anterior, para formar a próxima geração (*steady-state*). Esse processo pode ser repetido sucessivas vezes até o encontro da qualidade suficiente (solução), ou até que responda as limitações impostas através da programação, ou recursos computacionais.

Existem duas frentes para formular a ideia de evolução de sistemas:

- Operadores de variação (recombinação e mutação): criam a diversidade necessária dentro da população, o que viabiliza o surgimento de novas características;
- Seleção: age aumentando a qualidade média das soluções na população.

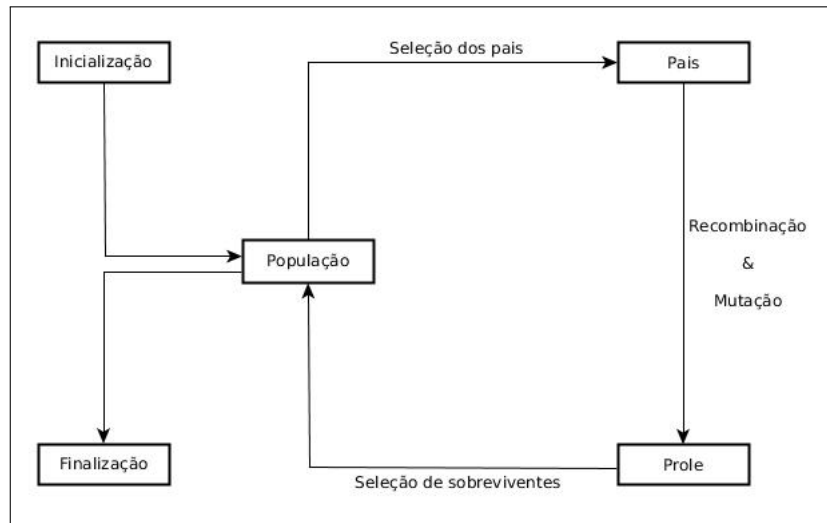
A combinação de variação e seleção geralmente melhora o valor de aptidão das gerações seguintes. Nota-se que o processo de evolução é otimizado (ou aproximações de valores ótimos) para a função de aptidão, pois aproxima os valores a cada nova geração.

Um visão alternativa é que a evolução pode ser vista como um processo de adaptação. Assim, a função de aptidão não é vista como uma função objetivo, mas como requisito do ambiente. Seguindo essa ideia, notamos que após gerações, o processo resulta em populações mais bem adaptadas com o ambiente.

Trata-se de um processo aleatório, pois no processo de seleção ainda podem ser encontrados indivíduos menos adaptados; a função objetivo não cresce monotonicamente. O mesmo pode ser dito da mutação, pois o intervalo do gene que será alterado e o valor que substituirá o alelo, são escolhidos aleatoriamente.

A Figura 5 mostra a inspiração básica para a criação dos algoritmos evolutivos.

Figura 5 – Fluxograma de um algoritmo evolutivo



Fonte: adaptado de Eiben e Smith (2015)

Para a implementação desses algoritmos é necessário especificar cada um dos componentes citados abaixo e definir os processos de inicialização e finalização. Caso seja necessário que o algoritmo termine em determinado momento, também é necessário especificar essa condição. Na próxima seção, serão explicados com mais detalhes cada um dos componentes, focando na implementação do AG.

- representação;
- função de aptidão;
- população;
- mecanismo de seleção dos pais;
- operadores de variação;
- mecanismo de sobrevivência.

## 2.3 ALGORITMO GENÉTICO

A família dos “algoritmos evolutivos” compreende: algoritmo genético, estratégias evolutivas, programação evolutiva, programação genética, otimização por enxame de partículas, etc. Segue uma apresentação mais detalhada, baseando em Eiben e Smith (2015), do algoritmo evolutivo em especial o AG, pois esse estudo apresenta apenas sua implementação.

Os princípios de Darwin para automatizar a solução de problemas foram usados antes da década de 1940 (EIBEN; SMITH, 2015), mas foi na década de 1960 que Holland e seus alunos desenvolveram o *Algoritmo Genético* (PEREIRA, 2013), seguindo os princípios da computação evolutiva. Para Eiben e Smith (2015), o AG em sua forma canônica ou “AG simple”, é o algoritmo evolutivo mais conhecido. Este compreende uma representação binária, com seleção por proporção do valor de aptidão, com baixa taxa de mutação e ênfase na inspiração de recombinação genética como principal forma de encontrar candidatos a solução. Para esse caso particular, os operadores do algoritmo evolutivo são:

- Representação: *bit-string*;
- Recombinação: cruzamento de um ponto;
- Mutação: troca de *bit*;
- Seleção dos pais: roleta;
- Sobrevivência: geracional.

Tradicionalmente, dada uma população inicial, é gerada uma população intermediária de mesmo tamanho. Essa população é criada aleatoriamente por pares de pais, aplicando o cruzamento que ocorre com uma certa probabilidade; caso contrário os filhos representam simples cópias dos pais. A nova população intermediária passa para a mutação, indivíduo por indivíduo, onde cada *bit* do indivíduo pode ser alterado com probabilidades iguais. Essa geração intermediária vai formar a próxima geração, substituindo toda a população anterior (geracional). Contudo existem outras formas de implementar um AG, assim, nas próximas seções, serão apresentados mecanismos para cada um dos principais componentes do algoritmo.

### 2.3.1 Representação

Para implementar o algoritmo evolutivo, é necessária criar uma representação do problema real que atenda as funcionalidades do algoritmo. Isso implica em simplificações, que representem bem o problema real, tornando viável a armazenagem e processamento pelo computador.

Essa representação do problema real é chamado de *fenótipo* e sua codificação de *genótipo*, este deve ser capaz de representar todos os possíveis fenótipos do problema real. A julgar pela representação, o genótipo pode ser consideravelmente diferente do fenótipo. Pressupõe-se que qualquer solução do problema poder ser representada pelo fenótipo e genótipo; assim, será viável encontrar uma solução ótima. O caminho inverso, isto é, partindo do genótipo para se obter o fenótipo, é chamado decodificação. Além da codificação, a representação vai interferir no uso de operadores, como a mutação, visto que a cada codificação existirá um mecanismo diferente de implementação.

A escolha de qual representação deve ser usada está associado às características do problema. “*Often this only comes with practice and a good knowledge of the application domain*”(EIBEN; SMITH, 2015); essa decisão vai interferir também nos operadores de variação a serem usados. Então, para decidir qual representação a usar, é necessário conhecer as principais representações, assim como a integração destas com os operadores de variação que serão detalhados de forma mais clara a seguir.

- Representação binária:

Neste caso, o cromossomo é representado por uma simples *string* de dígitos binários. Para uma aplicação, devemos decidir qual o tamanho dessa *string* e como interpretar o fenótipo a partir dela. Essa decisão deve garantir que a decodificação de genótipo para fenótipo seja possível para qualquer candidato a solução, assim como a codificação fenótipo para genótipo. As possibilidades da *string* devem resultar somente em possíveis soluções do problema;

- Representação inteira:

Para certos casos a representação binária não possibilita a representação de todos os candidatos a solução do problema, assim são necessários outros tipos de representações. É desejado que a codificação e operadores de variação tenha alguma relação natural com o problema, entre o valores possíveis e os atributos assumidos. Para os atributos que possam ser representados por números ordinários a representação por números inteiros é vantajosa, se comparado com representação de números cardiais. Diferentemente da representação binária, para números inteiros a escala de crescimento é similar, por exemplo, o atributo dois e três estão próximos, ao passo que o mil está muito distante;

- Representação real:

“*Often the most sensible way to represent a candidate solution to a problem is to have a string of real values*” (EIBEN; SMITH, 2015). Este caso ocorre quando os valores, que representam o gene, são contínuos como: altura, largura, etc. Assim, o genótipo consiste em um vetor  $z = \{(z_i, \dots, z_n) \mid z_i \in \mathbb{R}\}$ ;



- Permutação:

Diferentemente das outras representações, onde os valores do campo de busca podem se repetir, para os casos que os valores não podem repetir, usa-se a permutação. Neste casos fixa-se os valores de possíveis soluções em uma ordem numérica de valores inteiros.

### 2.3.2 Função de aptidão

A noção de *função de aptidão* se baseia no fato que a população deve se adaptar e quão apta a população está no sistema, o que forma a base para seleção e viabiliza o desenvolvimento (progresso). Essa função mede a qualidade de um determinado genótipo, de forma a avaliá-lo na forma de fenótipo no contexto do problema real.

Em um problema de otimização, a função objetivo será usada para avaliar o resultado das soluções encontradas e se o valor da função calculado, for o maior valor possível, então estamos falando do valor ótimo para um problema de maximização. No caso de um algoritmo evolutivo, essa função pode ser idêntica à função de aptidão ou adaptada para implementação computacional.

### 2.3.3 População

Um conjunto de vários genótipos possíveis forma uma população, esses devem ter obrigatoriamente sua representação do fenótipo. A população é a unidade de evolução, visto que um indivíduo não se altera para adaptar ao ambiente, mas sim à população que torna essa adaptação possível. Com essa concepção, é notável que os indivíduos mais adaptados terão maior probabilidade de serem escolhidos como os progenitores da próxima geração, ou que os piores indivíduos serão escolhidos para serem trocados por novos indivíduos. Essa ideia é oposta a ideia do operadores de variação que atuam diretamente nos indivíduos, e não na população como um conjunto.

É comum nas aplicações dos algoritmos evolutivos, que a população permaneça com o mesmo tamanho, criando um ambiente propenso a competição. A diversidade da população pode ser medida pela quantidade de soluções possíveis. É comum que essa medida esteja relacionada com os valores de aptidão presentes e com os diferentes tipos de fenótipos ou de genótipos. Ressalta-se que, assim como um mesmo fenótipo pode estar associado a diferentes genótipos, um determinado valor de aptidão pode estar associado a diferentes fenótipos. Pode-se concluir que dado um genótipo, a ele estará associado apenas um fenótipo e um valor de aptidão, assim como a definição de domínio e contra domínio em uma função.

Para a implementação do algoritmo, existem duas abordagens para a população. Então, o algoritmo começa com uma população inicial e posteriormente são selecionados

indivíduos chamados de pais para criar novos indivíduos. Desta forma, alguns indivíduos são evoluídos devido aos operadores de variação e outros remanescem como cópias do pais. Na primeira abordagem, chamada geracional, toda a população é trocada por indivíduos selecionados dos filhos, então chamada de próxima geração. Para Eiben e Smith (2015), no modelo típico de AG, a população, tanto de pais selecionados quanto filhos, tem o mesmo tamanho, então toda a geração é trocada pelos filhos. Na segunda abordagem, chamada de *steady-state*, a princípio não se altera toda a população. Uma parcela da população original (pais), será trocado por indivíduos selecionados da prole, ou se for gerada uma prole menor que o tamanho da população original, a prole irá substituir uma parte da população original.

### 2.3.4 Mecanismo de Seleção dos pais

A seleção dos pais é feita com base em suas respectivas aptidões. Assim, os melhores indivíduos serão os progenitores da geração seguinte, contudo não se exclui a possibilidade que os piores indivíduos sejam progenitores também. Desta forma, temos que o processo de seleção dos pais é visto como um processo probabilístico, sendo que os indivíduos mais aptos têm maior probabilidade de serem selecionados como pais e os indivíduos com menores aptidões têm menor, embora suas chances sejam diferentes de zero. Caso essa condição não fosse respeitada, a população poderia encontrar uma região de ótimo local e permanecer nessa região, sem encontrar o ótimo global, como será discutido na seção 2.3.8. Assim, esse mecanismo é responsável por melhorar a qualidade dos indivíduos (melhor aptidão) das gerações seguintes. Segue abaixo os três principais mecanismos para esse componente.

- Seleção por roleta:

Essa seleção depende do valor absoluto de aptidão de cada indivíduo em relação a soma da aptidão de todos indivíduos, nota-se que o somatório da probabilidade de todos indivíduos deve ser um. Assim, o indivíduo mais apto tem maior probabilidade de ser selecionado. Contudo esse mecanismo apresenta alguns contrapontos, como: indivíduos mais aptos tendem a ser escolhido com maior frequência, restringindo o espaço de busca, o que pode ocasionar em uma convergência prematura; quando os indivíduos apresentam valores de aptidão próximos, reduzem quase a zero a pressão de seleção. Assim a seleção passa a ser aleatória, então o valor de aptidão do melhor indivíduo deixa de ser importante;

- Seleção por ranking:

Essa seleção preserva a pressão de seleção, por ordenar a população de indivíduos de acordo com seu valor de aptidão. Dessa forma os melhores receberão o número de indivíduos da população  $n$ ,  $n - 1$ , ao passo que os piores receberão número menores.

A partir daí os mecanismos são similares ao método da roleta. Esta seleção atribuirá maiores chances de escolha dos melhores indivíduos, quanto maior for a pressão de seleção necessária, maior ênfase será dada aos indivíduos que apresentarem valor de aptidão acima da média;

- Seleção por torneio:

Caso a população seja muito grande ou bem distribuída, para as seleções anteriores, gasta-se muito tempo ou até se tornam impossíveis de serem processadas. Sem contar que assumem que a aptidão é um valor quantitativo baseado em uma função objetivo, o que em alguns casos pode não ser. A seleção por torneio tem um formato parecido com os torneios em geral, onde são selecionados aleatoriamente dois ou mais competidores para duelarem entre si e o mais apto ganha. Segue a ideia do *ranking*, mas comparando apenas dois indivíduos por instância. Essas instâncias se repetem até chegar em um número ideal de pais selecionados para formarem os novos indivíduos. Assim, quanto maior o tamanho do torneio, maior a probabilidade de selecionar indivíduos com a aptidão acima da média e menor a probabilidade de indivíduos abaixo da média, tornando-se um parâmetro para a pressão de seleção.

### 2.3.5 Operadores de Variação

Os operadores de variação são responsáveis por criar novos indivíduos a partir dos indivíduos já existentes, ou seja, criar novos candidatos à solução por meio da mutação e recombinação. A aplicação dos operadores de variação, sob um aspecto mais abrangente, consiste em estabelecer um parâmetro no algoritmo, então é gerada uma probabilidade  $p$ . Se o valor de  $p$  for menor ou igual ao parâmetro configurado, executa-se os operadores de variação.

#### 2.3.5.1 Recombinação

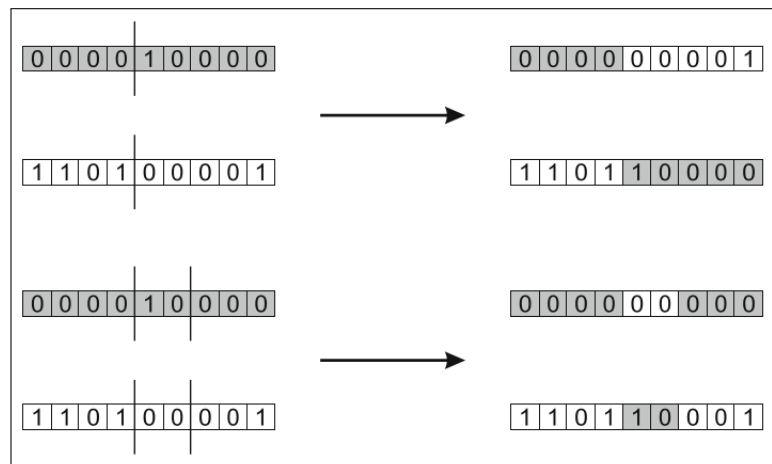
A recombinação é o operador binário de variação, trata-se de um operador que combina informações dos dois pais, ou seja, escolhe parte aleatoriamente do genoma de cada indivíduo e combina em um novo indivíduo. Para Eiben e Smith (2015), a recombinação é o principal operador de busca para o algoritmo genético e como a mutação, também é um processo probabilístico. Através da recombinação é possível trazer para um novo indivíduo características desejadas dos pais, criando um novo indivíduo que harmoniza essas duas características. Ressalta-se que a forma de recombinação vai depender da representação usada para o desenvolvimento do estudo.

- Recombinação para a representação binária:

Em geral a recombinação para a representação binária é feita de três formas, todas são feitas a partir da seleção de dois pais e geram dois filhos. A primeira forma é

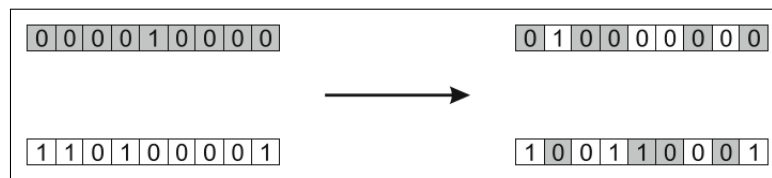
chamado de cruzamento de um ponto, assim, dado o comprimento da *string* ( $c$ ), seleciona-se um pedaço do genótipo, que varia de 1 a  $(c - 1)$ . Onde o genótipo irá se dividir em dois, para posteriormente a primeira parte do “pai um” se combinar com a segunda parte do “pai dois”, para gerar o “filho um”. Já o “filho dois” é criado combinando as partes restantes dos pais (Figura 6, parte superior). A segunda forma é chamado de cruzamento de “ $n$ ” pontos, este, trata-se de uma generalização do caso anterior, visto que o genótipo será dividido em mais de duas partes, e os filhos serão gerados alterando entre os genes dos pais (Figura 6, parte inferior). Por fim, a terceira forma é o cruzamento uniforme: diferentemente dos casos anteriores, nesse caso cada alelo dos filhos é selecionado com base em uma distribuição uniforme em  $[0, 1]$ , assim cada pai tem 50% de chances de fornecer seu alelo para cada *bit* do filho um. O filho dois é criado do genótipo inverso do filho um (Figura 7);

Figura 6 – Cruzamento de um e “ $n$ ” pontos



Fonte: Eiben e Smith (2015)

Figura 7 – Cruzamento uniforme



Fonte: Eiben e Smith (2015)

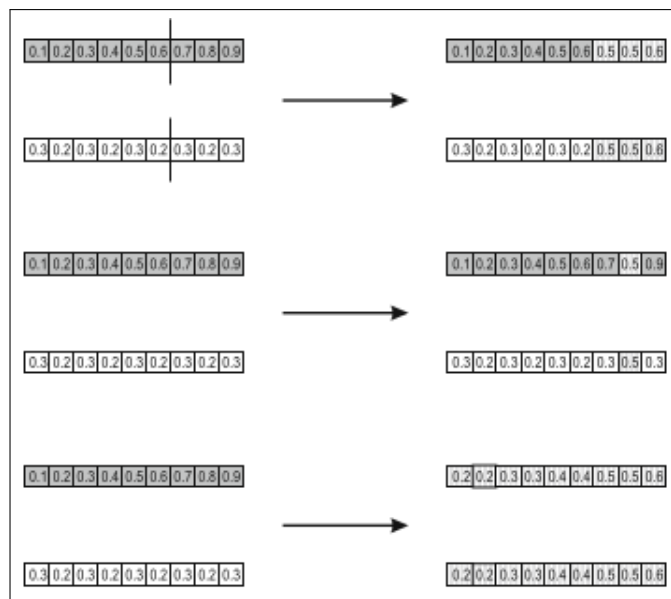
- Recombinação para representação inteira:

São usados processos similares ao processo de recombinação binária. Contudo, ao invés dos valores dos alelos se restringirem em 0 e 1, neste caso, os valores dos alelos poderão ser qualquer valor do conjunto  $\mathbb{N}$ ;

- Recombinação para representação real:

Para a recombinação, em geral são usados três métodos. O primeiro se trata de copiar o alelo dos pais nos filhos, decidindo em cada alelo o valor a ser copiado do pai um ou pai dois. O segundo método é através de combinação linear dos alelos dos pais para formar os alelos dos filhos (Figura 8), da forma :  $z_i = \nu \cdot u_i + (1 - \nu) \cdot v_i$ , onde  $u$  representa o alelo do pai um e  $v$  do pai dois,  $z$  o alelo do filho e  $i$  a posição do alelo, com  $\nu$  variando dentro do intervalo  $[0,1]$ , em geral usa-se  $\nu = 0,5$ . Por último, pode-se criar um novo valor de alelo, para cada posição, maior que os valores de alelo daquela posição dos pais, ou menor. Conforme vimos na representação binária, a recombinação real usa mecanismos parecidos. Assim o cruzamento pode ser a partir de um certo ponto do genótipo (Figura 8, região superior), em um alelo aleatório (Figura 8, região central) e como no último cruzamento da Figura 8, onde é feita a combinação linear de todo o genótipo dos pais criando dois novos indivíduos iguais.

Figura 8 – Ilustração da recombinação por representação com valores reais



Fonte: Eiben e Smith (2015)

- Recombinação por permutação:

Essa recombinação não será apresentada neste texto, devido sua complexidade. Pois, a recombinação para essa representação exige maior atenção, visto que é necessário manter as propriedades da permutação, mesmo combinando informações dos pais. Assim, uma explicação mais detalhada pode ser consultada em Eiben e Smith (2015), na seção 4.5.2 do capítulo quatro.

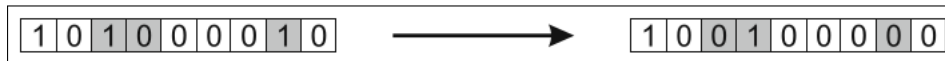
### 2.3.5.2 Mutação

A mutação é um operador de variação unário, pois altera apenas um genótipo e cria uma modificação na população. Como esse operador irá atuar de forma aleatória, podem ser criados filhos que representem novos candidatos à solução, com características diferentes das apresentadas no pais. Assim, mesmo que o algoritmo se aproxime de uma solução, ótimo local, será possível, através da mutação, aumentar o espaço e a diversidade de indivíduos.

- Mutação para a representação binária:

Entre as possibilidades de mutação para a representação binária, segundo Eiben e Smith (2015), a mais comum é considerar cada gene separadamente e permitir que cada *bit* mude de um para zero ou de zero para um, com uma pequena probabilidade. A quantidade de valores a ser alterado não é fixa, essa decisão é feita de forma aleatória e está relacionado com o tamanho da *string*. Como podemos ver na Figura 9, uma *string* de nove *bits*, cada *bit* representa 1/9 do tamanho do cromossomo, onde encontra-se um alelo (valor zero ou um). Foram escolhidos, o terceiro, quarto e oitavo *bit* para a ocorrência da mutação, como destacado em cinza no lado esquerdo da Figura 9. Após ocorrida a mutação, nota-se no cromossomo a direita da Figura 9, que cada *bit* selecionado teve o seu valor de alelo alterado;

Figura 9 – Mutação para a representação binária



Fonte: Eiben e Smith (2015)

- Mutação para a representação inteira:

Neste caso, existem duas principais formas de mutação usadas. A primeira é a redefinição aleatória: dado um valor de probabilidade  $p$  e de acordo com a configuração dos parâmetro de mutação, caso seja maior ou igual a  $p$ , ocorrerá a mutação, onde será escolhido aleatoriamente um alelo do cromossomo para ser trocado de forma aleatória por outro. A outra forma é a mutação *creep*, para esta mutação é usada uma distribuição estatística simétrica com média zero. Assim, dado um  $p$ , caso seja menor ou igual ao parâmetro de mutação, então serão gerados valores aleatórios, orientados pela distribuição escolhida, que serão somados ao valor já existente do alelo, criando pequenas alterações no gene;

- Mutação para a representação real:

Devido às limitações de *hardware*, é comum trocar o valor dos alelos por outros valores aleatórios, que pertencem ao domínio do problema. Para esse mecanismo é dado um limite inferior e superior que irá restringir o domínio, assim será sorteado

um valor para substituir os já existentes. Existem duas distribuições usadas para isto: a uniforme e não uniforme, para o segundo caso usa-se a distribuição gaussiana com média zero e desvio-padrão como parâmetro do algoritmo (tamanho do passo da mutação);

- Mutação para a permutação:

Existem quatro formas. A primeira é através da troca; nesse caso são escolhidos, aleatoriamente, duas posições do cromossomo para que os alelos troquem de posição. A segunda é a inserção: dois alelos são escolhidos aleatoriamente, sendo que o segundo alelo é movido para o lado do primeira alelo. Os outros alelos permanecem na ordem. A terceira forma, é a mutação *scramble*: todo o cromossomo, ou parte dele, é escolhido para que sejam executadas trocas aleatórias dos valores. E por último a inversão, a qual são escolhidos dois alelos aleatórios do cromossomo e, dentro desse intervalo (de um alelo a outro), inverte-se toda a ordem dos alelos desse intervalo. Esse tipo de mutação pode ocorrer em todo cromossomo também.

### 2.3.6 Mecanismo de sobrevivência

A seleção de sobreviventes ou substituição, é um mecanismo parecido com seleção dos pais. Contudo é usado após a criação dos novos indivíduos (prole), onde somados os indivíduos da prole e população (pais e filhos) irá ultrapassar o tamanho da população inicial, Contudo mantendo seu tamanho, irá promover a competição e seleção dos indivíduos. Esse mecanismo se torna uma estratégia comparável à seleção natural. O critério de seleção, em geral, é o cálculo de aptidão favorecendo os que apresentam maior valor. Diferente dos mecanismos apresentados até então, este é determinístico e pode ser aplicado pela simples troca da população antiga (pais) pela nova população (filhos), assim chamado de substituição. Independente de como será a implementação do mecanismo de sobrevivência, ao final desse processo a prole substituirá a população para a próxima geração (AG geracional).

O mecanismo de sobrevivência é responsável em reduzir a memória do trabalho de um algoritmo evolutivo. A princípio qualquer mecanismo de escolha dos pais pode ser usado nessa etapa também, contudo existem estratégias próprias sugeridas e usadas de forma inteligente.

- Reposição por idade:

Para esse mecanismo não considera-se o valor de aptidão do indivíduo, ao contrário, é feito para que cada indivíduo permaneça na população por tempo constante (número de interações ou gerações). Isso não significa que cópias de indivíduos com alto valor de aptidão não existam na população. Contudo, como não se considera o valor de aptidão na escolha, o crescimento médio e até do melhor indivíduo de aptidão pode

ser lento. Costuma ser usada em algoritmo genético quando o tamanho da população não muda, ou seja, trocar toda a população pela nova geração. Outra possibilidade é escolher aleatoriamente pais para serem substituídos por indivíduos da nova geração;

- Mecanismos baseados no valor de aptidão:

Estes mecanismos consideram o valor de aptidão dos indivíduos da população. A primeira alternativa é a substituição do pior indivíduo. Neste, é escolhido o indivíduo com o pior valor de aptidão para ser substituído por um indivíduo da geração anterior. Contudo, tem-se um rápido crescimento da aptidão média da população, focando no indivíduo de melhor aptidão atual. O elitismo é a segunda alternativa, neste caso, é comum que este mecanismo seja usado junto à seleção baseado em idade, para evitar a perda do atual maior valor de aptidão. Nesse mecanismo é escolhido o indivíduo com maior aptidão na população a ser substituída (pais). Este é comparado com os indivíduos da prole, caso o indivíduo (da população dos pais) não apresente o valor de aptidão superior a algum indivíduo da prole, o mesmo é descartado e a prole permanece inalterada, caso contrário esse substitui o indivíduo avaliado da prole.

### 2.3.7 Inicialização e condição de parada

Em geral a inicialização é feita pela população inicial. Esta é gerada aleatoriamente criando um determinado número de indivíduos factíveis (indivíduos que atendam aos requisitos para existir em seu ambiente). O número de indivíduos vai depender do tamanho da população a ser trabalhada.

A condição de parada vai adequar-se ao formato do problema. No caso de um problema de otimização, que se tem uma função objetivo, essa condição vai depender do valor que a função objetivo vai fornecer. Já em casos que o modelo passou por simplificações e ruído nos dados, será importante considerar uma tolerância de erro na resposta final. Contudo, como os algoritmos evolutivos são processos empíricos e existe a possibilidade de nunca encontrar o ótimo global, torna-se difícil implementar essa noção de tolerância de erro em relação ao valor ótimo. Para evitar que o algoritmo nunca pare, em geral são feitas as seguintes condições:

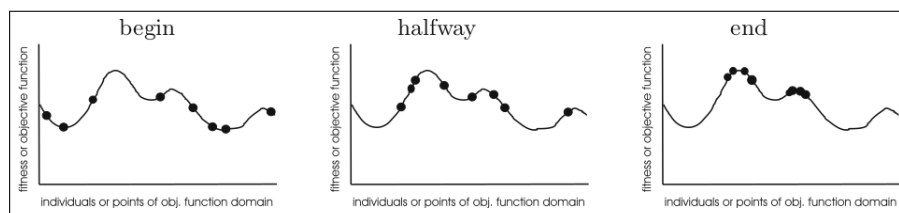
- O máximo de processamento que a CPU permitir;
- Existir um limite de gerações;
- O valor de aptidão não apresentar um crescimento;
- A diversidade da população chegar a uma determinada tolerância.



### 2.3.8 Comportamento típico de um AG

Como podemos ver na Figura 10, que apresenta uma determinada função objetivo a ser maximizada, inicialmente a população de candidatos a solução está aleatoriamente distribuída ao longo da imagem dessa função como vemos em “*begin*”. Passado um determinado número de gerações (iterações do algoritmo), visto que a população possui operadores de seleção e recombinação, é possível notar que os indivíduos começam se aproximar de ótimos locais (*halfway*), ou seja, elevando o valor de aptidão dos indivíduos da população. Já próximo ao critério de parada, em “*end*” podemos observar que os indivíduos estão concentrados em ótimos locais da função. Como podemos notar, é possível obter como resposta final de um ótimo local, ao invés de um ótimo global como desejado. Dessa forma é necessário analisar os parâmetros do algoritmo para assegurar o melhor valor de aptidão daquela população, conforme explica Eiben e Smith (2015): “*Premature convergence is the well-known effect of losing population diversity too quickly, and getting trapped in a local optimum*”. Assim, ao permanecer com muitos indivíduos da geração antiga, pode significar uma busca ineficiente, enquanto que a alta renovação de população tende a agilizar a busca.

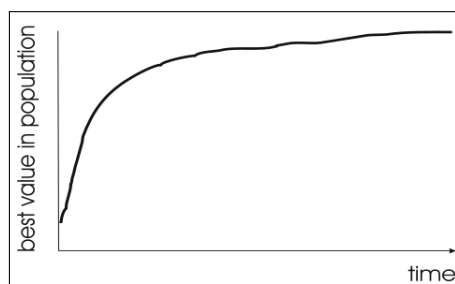
Figura 10 – Distribuição da população em um algoritmo evolucionista



Fonte: Eiben e Smith (2015)

Sobre o crescimento da aptidão da população, Figura 11, que apresenta o maior valor de aptidão ao longo das gerações da população, nota-se um rápido crescimento nas primeiras gerações e um amortecimento da curva após um determinado número de gerações. Assim, do ponto de vista da inicialização e terminação do algoritmo, podemos notar que um número grande de interações podem resultar em baixo crescimento a partir de um certo número de gerações (interações), o que dependendo do problema a ser resolvido, tem-se como uma estratégia pouco interessante.

Figura 11 – Progresso do maior valor de aptidão ao passar das gerações



Fonte: Eiben e Smith (2015)

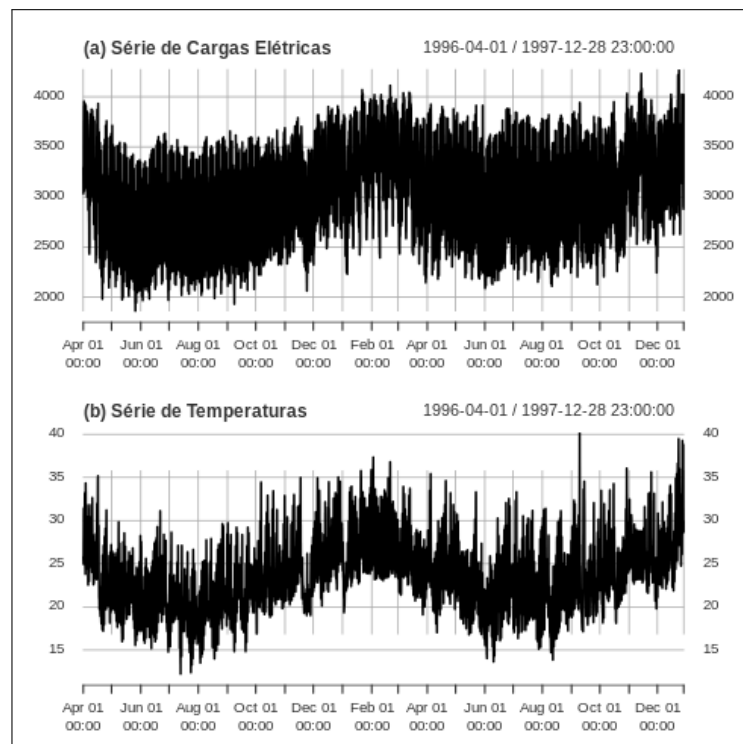
### 3 METODOLOGIA

Este capítulo apresenta os dados utilizados para o estudo, suas divisões, métodos utilizados para fazer as previsões, bem como a implementação do algoritmo genético. O capítulo está dividido em quatro seções.

#### 3.1 SÉRIE

As séries de dados utilizadas para este estudo são: a carga elétrica e a temperatura, a cada hora do dia na cidade Rio de Janeiro - RJ. Ambas referem-se ao período de 1996 a 1997, constituído por 104 semanas, ou seja, 17472 observações. Em ambas as séries, foram excluídos alguns dias do final de 1997 para que a série tivesse exatamente 104 semanas. As unidades de medidas das séries são, respectivamente, Megawatt-hora [*MWh*] e graus Celsius [ $^{\circ}C$ ]. Diferentemente da série de carga, a série de temperatura não contém todos os dados dos dois anos; apenas no período de 01/04/1996 até o fim de 1997. Para facilitar a organização do trabalho, considera-se que a semana se inicia em 01:00:00 de segunda-feira e termina às 0:00:00 de domingo, contabilizando 24 observações por dia e 168 por semana. As séries de dados iniciam-se na segunda-feira (01/01/1996) e terminam no domingo (28/12/1997), e são mostradas na Figura 12.

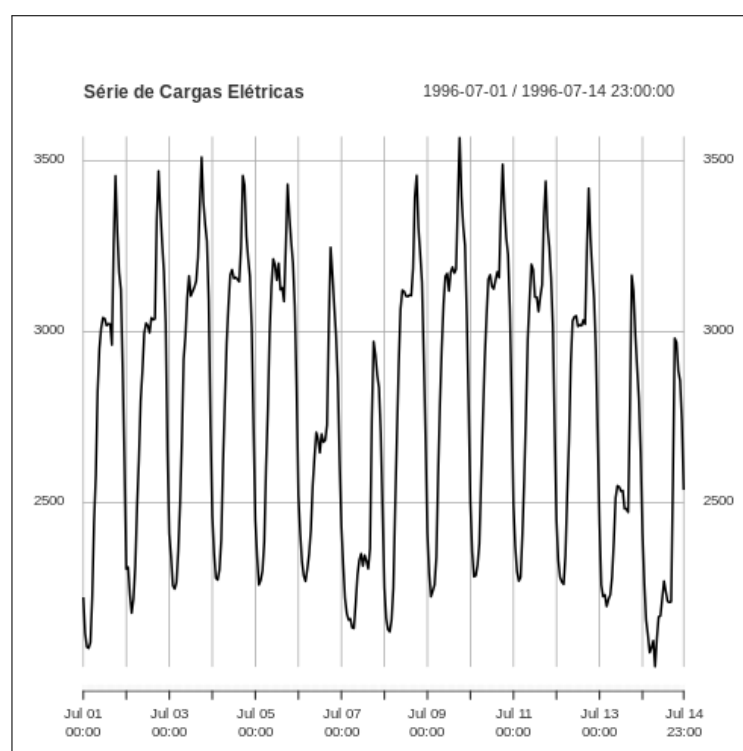
Figura 12 – Séries de cargas e temperaturas do Rio de Janeiro, de 1996 a 1997



Fonte: autor

Analisando a Figura 12, percebe-se um comportamento parecido de ambas as séries; isto indica que “*the weather is a major factor driving the electricity demand*” (HONG; FAN, 2016), o que justifica a ideia de usar a série de temperaturas combinada com a de cargas elétricas. Na Figura 12, nota-se que ambas as séries têm uma sazonalidade anual, com picos nas estações mais quentes, o que pode ser explicado pelo uso de ar condicionado, e com vales nas estações amenas (SONG et al., 2006). Também observa-se uma leve tendência, embora esta não seja considerada no estudo, pois as séries de dados são de apenas dois anos, o que significa pouco tempo para avaliar a mesma, e uma sazonalidade anual. Restringindo os dados da série de cargas para as duas primeiras semanas (Figura 13), nota-se que a série contém uma sazonalidade semanal. Isto ocorre também quando se analisa um dia, ou seja, esta série apresenta três sazonalidades: anual, semanal e diária; além da tendência. Este conjunto de propriedades, em uma mesma série, motiva o estudo de previsão, por ser desafiador encontrar um previsor capaz de considerar todas essas propriedades.

Figura 13 – Série de cargas para duas semanas (de 01/07/1996 a 14/07/1996), mostrando a sazonalidade semanal e sazonalidade diária

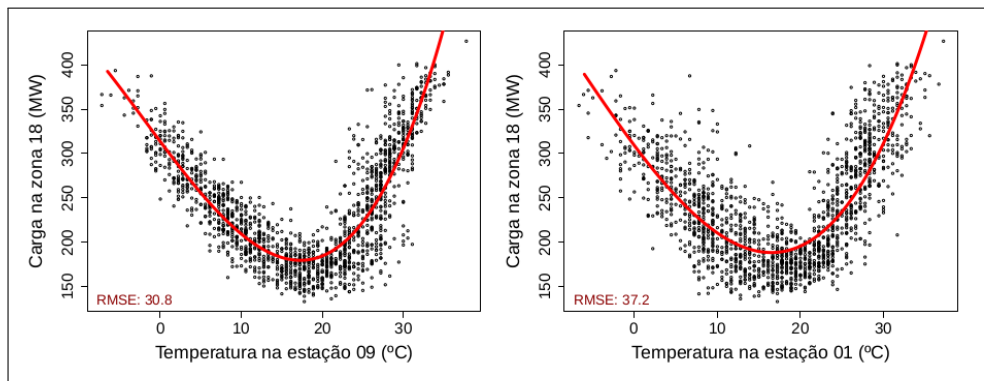


Fonte: autor

Zhu, Chow e Tse (2018), discutem que existe uma pequena diferença de erro da previsão de carga com o uso de dados das temperaturas registradas no instante da previsão e do uso de previsão da temperatura para o mesmo ponto. Assim esse trabalho se atém a utilizar os dados registrados na hora em que se pretende prever a carga. Complementando a análise da relação da série de carga e de temperatura, tem-se a Figura 14, a qual apresenta o diagrama de dispersão da carga elétrica em função da temperatura, para parte

dos dados observados em duas regiões distintas da Carolina de Norte (EUA), durante o período de 01/01/2004 a 07/07/2008. Optou-se por usar os dados de outra região, diferente da região do estudo por uma questão didática, pois na Figura 14, fica evidente que a relação entre carga e temperatura é não linear. Para modelar a relação entre carga e temperatura, deve-se selecionar dentre as diversas possibilidades, como por exemplo os polinômios de segundo ou terceiro grau, além de funções lineares definidas por partes (HONG; FAN, 2016). Mayrink (2016) desenvolveu um polinômio de terceiro grau para ajustar os valores de carga em função das temperaturas observadas em seu estudo. Seu resultado pode ser visto em vermelho na Figura 14. Seguindo o raciocínio similar, nesse estudo, experimenta-se o uso de uma função quadrática,  $f(x) = T^2$ , para avaliar seu ganho em simplicidade, levando em consideração a perda de resultado em relação ao erro de previsão.

Figura 14 – Diagrama de dispersão da cargas em função da temperatura, para os dados da Carolina de Norte (EUA), durante o período de 01/01/2004 a 07/07/2008



Fonte: Mayrink (2016)

Analisando a série de temperatura diariamente, em especial a cada hora do dia é possível observar que a temperatura varia pouco de um dia para o outro; por isso, existem autores que defendem que a informação da temperatura já está agregada no histórico da série de carga elétrica, sendo relevante somente quando ocorre algum fenômeno natural, que leve a uma forte alteração de temperatura (DEFILIPPO, 2017). Contudo, se a região apresentar condições climáticas voláteis e existir uma forte relação entre a temperatura e carga, pode ser uma desvantagem para o método não considerar a variação de temperatura (HONG; FAN, 2016). Para esta série é mais interessante utilizar a diferença de temperatura entre um dia e outro, conforme resultados obtidos por Defilippo (2017) em seus experimentos. Neste estudo, leva-se em consideração a diferença de temperatura entre o valor observado no instante da previsão e o valor da mesma hora no dia anterior, conforme Equação 3.1. Também será experimentado o uso de uma função quadrática para as diferenças de temperatura,  $f(x) = (T')^2$ .

$$T'_{t+k} = T_{t+k} - T_{t+k-24} \quad (3.1)$$

Em geral os métodos de previsão são desenvolvidos para dias normais; quando se tenta fazer previsões com o dias especiais, como feriados, o predictor acaba tendo um erro maior. Uma vez que os perfis de carga para dias especiais são diferentes daqueles dos dias normais. Fica evidente que é necessário um tratamento diferente para esses dias. Para não comprometer o trabalho com essas exceções, os mesmo foram substituídos pela média das observações dos dados no mesmo horário da semana seguinte e anterior.

### 3.2 PARTIÇÃO DAS SÉRIES

A implementação dos previsores é desenvolvida como Pedreira (2018), que particionou a série conforme a Tabela 1. Assim, as previsões são feitas iniciando nas amostras de treino com vinte semanas, onde estimam-se os valores das constantes de previsão, e posteriormente, usa-se a amostra de teste, com um período de dez semanas, para avaliar os resultados da previsão. Para as combinações, neste estudo são experimentadas duas formas, e ambos os casos são explicados tomando a partição 1 como exemplo.

Tabela 1 – Partição das séries de cargas e de temperaturas

Partição	Amostra de Treino [semanas]	Amostra de Teste [semanas]	Índice da Observação da Amostra de Teste
1	[35,54]	[55,64]	[9073,10752]
2	[45,64]	[65,74]	[10753,12432]
3	[55,74]	[75,84]	[12433,14112]
4	[65,84]	[85,94]	[14113,15792]
5	[75,94]	[95,104]	[15793,17472]

Fonte: autor

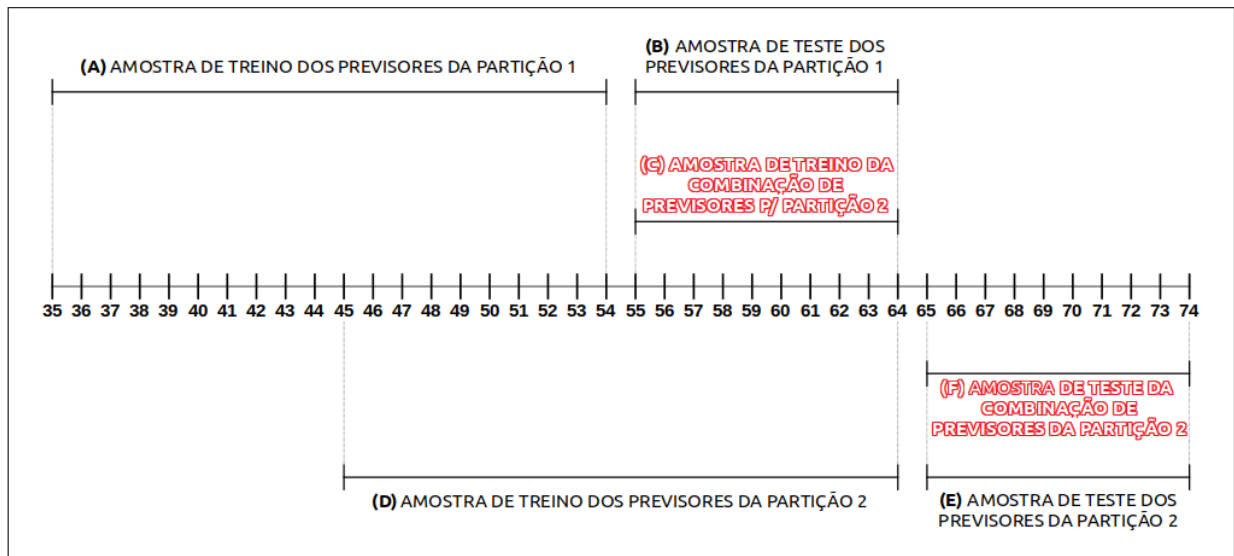
#### 3.2.1 Combinação 1

Para explicar a combinação 1, considera-se partição 1 como exemplo, conforme ilustrado na Figura 15. Para ajustar as constantes dos previsores, usa-se a amostra de treino com vinte semanas. Esta amostra inicia-se na semana 35 e termina na semana 54, Figura 15 (A). Definidos os modelos de previsão, os mesmos são testados na amostra de teste, com dez semanas (semanas 55 até 64), referenciado como (B) na Figura 15. Nessa mesma amostra as previsões da partição 1 são usadas para otimização dos coeficientes da combinação dos previsores, Figura 15 (C).

Note que a previsão de cada partição se limita às dez semanas seguintes a amostra de treino. Assim para obter as previsões da semana 65 a 74, é necessário estimar novamente as contantes dos previsores com a amostra de treino da partição 2, Figura 15 (D). Posteriormente suas previsões são testadas na amostra de teste, inciando na semana 65 e se encerrando na semana 74 conforme a Figura 15 (E).

Então, a combinação dos previsores é testada na amostra de teste da partição 2 (semanas 65 e 74), indicado por **F** na Figura 15, com as previsões da amostra de teste da partição 2 (Figura 15 na parte **E**), Porém usando os coeficientes da combinação encontrada na amostra de teste da partição 1, região **C** da Figura 15.

Figura 15 – Esquema da partição 1 e 2, para a combinação 1



Fonte: autor

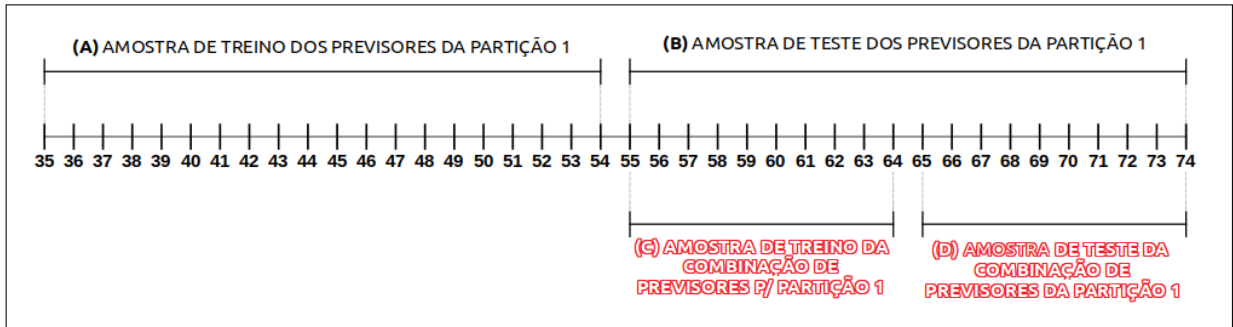
Portanto, os métodos ajustados na partição 1 servem para o ajuste dos coeficientes de combinação, e esses mesmos coeficientes são aplicados aos previsores da partição 2. Assim, essa combinação é avaliada na amostra de teste da partição 2, o que inviabiliza a avaliação da combinação de previsores para a partição 1.

### 3.2.2 Combinação 2

Como no caso anterior, para facilitar o entendimento da combinação 2, considera-se novamente a partição 1 como exemplo, a qual é ilustrada na Figura 16. Para ajustar as constantes dos previsores, usa-se a amostra de treino com vinte semanas, esta amostra se inicia na semana 35 e termina na semana 54, Figura 16 (A). Definidos os modelos de previsão, os mesmos são testados na amostra de teste; neste caso, porém, ao invés de limitar as previsões às dez semanas à frente, as constantes de previsão permaneceram inalteradas para as vinte semanas à frente. Assim a amostra de teste da partição 1 fica entre a semana 55 a semana 74, referenciada como **B** na Figura 16.

Contudo, para aplicar a combinação 2, é necessário dividir a amostra de teste da partição 1 em duas amostras. A primeira amostra corresponde a amostra para ajuste (semanas 55 a 64), Figura 16 (C), onde são estimados os coeficientes da combinação 2 dos previsores. Com esses coeficientes calcula-se a combinação dos previsores na segunda amostra, que é a de teste (semanas 65 a 74), Figura 15 (D), onde testa-se a eficácia do método.

Figura 16 – Esquema da partição 1, para a combinação 2



Fonte: autor

Para facilitar a comparação dos métodos, levando em consideração que o intervalo de teste da combinação 2 corresponde ao intervalo de teste da partição 2 na Tabela 1, então será chamado de resultado da partição 2 para a combinação 2. O que novamente inviabiliza a avaliação da combinação de previsores para a partição 1.

### 3.3 MÉTODOS DE PREVISÃO

Nesta seção, explica-se quais e como são implementados os métodos de previsão. Contudo, é necessário esclarecer que os métodos de previsão não foram implementados nesse estudo, e sim aproveitados do trabalho de Pedreira (2018), que foram enviados através de arquivos salvos no formato “txt”, contendo as séries de previsão desenvolvidas.

#### 3.3.1 Previsor *Naïve*

Segundo Morettin e Tolo (2006), este é método de previsão mais simples, também chamado de previsor ingênuo. Neste caso, o valor mais recente é usado como previsão para valores futuros. Portanto, é uma boa forma de avaliar os métodos de previsão, visto sua simplicidade na implementação, pois se este previsor apresentar melhor estimativa, é melhor manter sua utilização.

Para esta aplicação (previsão da carga elétrica) é comum o uso do previsor *naïve* com um dia de atraso; contudo após os dias especiais, como fins de semanas e feriados, este modelo apresenta grandes erros, devido as diferenças entres o perfis diários. Para contornar essa situação; usa-se a previsão da mesma hora do dia de uma semana atrás, neste caso, os perfis de carga são semelhantes, embora é preciso ter atenção, pois perde-se parte da informação da série, ao decorrer de sete dias. Neste estudo, utilizam-se ambos os métodos. O previsor *naïve* para a mesma hora de uma semana atrás, chamado de “*naïve168*” e calculado de acordo com a Equação 3.2.

$$\hat{X}_{t+k} = Y_{t+k-168} \quad (3.2)$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$t$ : tempo;

$k$ : passo a frente do instante  $t$ .

Para tentar aproveitar os benefícios de cada método, também é usado o previsor *naïve* da mesma hora do dia anterior, chamado de “*naïve24*” e calculado conforme Equação 3.3.

$$\hat{X}_{t+k} = Y_{t+k-24} \quad (3.3)$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

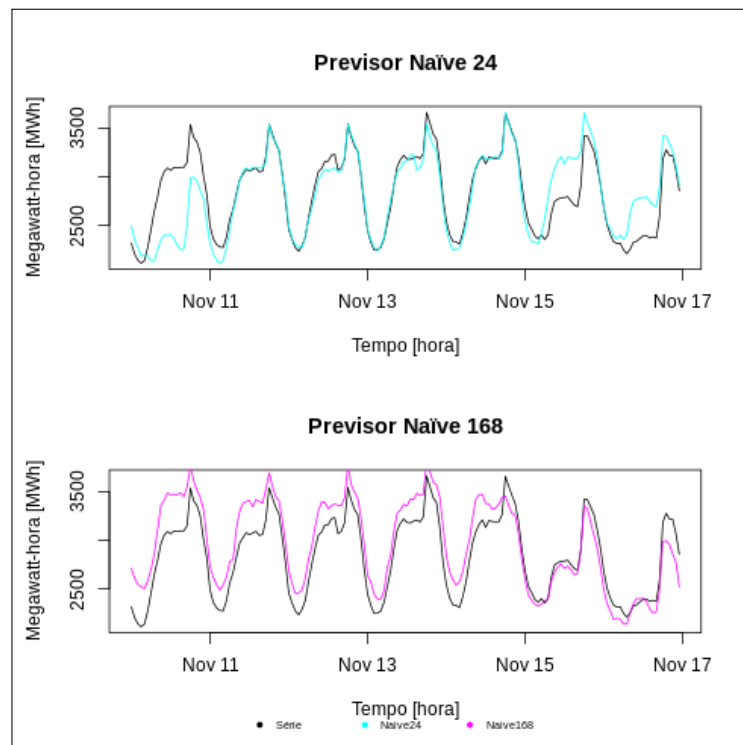
$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$t$ : tempo;

$k$ : passos a frente do instante  $t$ .

A Figura 17 mostra a série original e a previsão com cada previsor *naïve*, para uma semana do mês.

Figura 17 – Previsor *naïve*, no período de 10/11/1997 a 16/11/1997



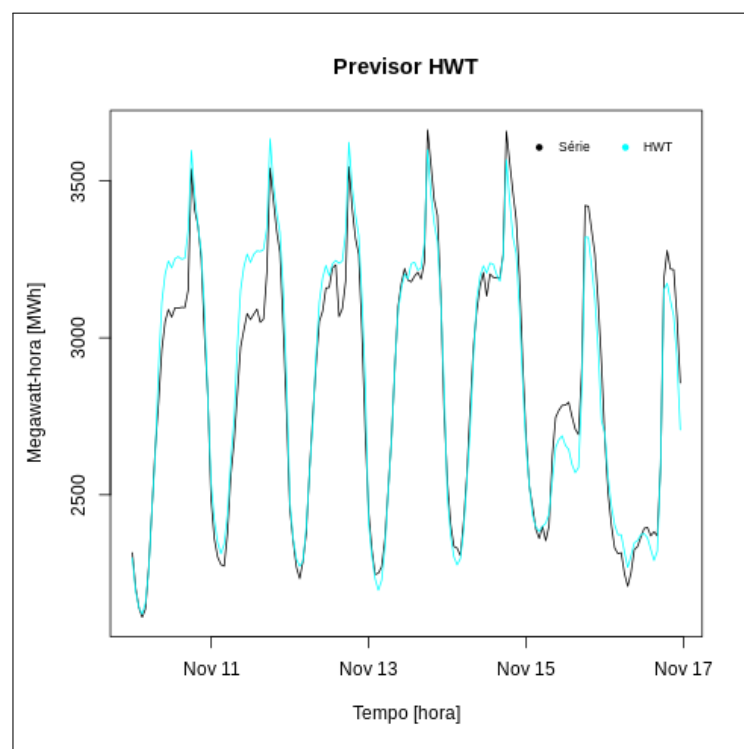
Fonte: autor



### 3.3.2 Previsor Holt-Winters-Taylor

Segundo Neto (2014), o amortecimento exponencial de Holt-Winters-Taylor (HWT) para múltiplas sazonalidades apresenta desempenho considerável, se comparado a métodos de formulações mais complexas. Assim como Pedreira (2018), nesse trabalho apresentam-se as mesmas versões do previsor de Holt-Winters-Taylor, que serão detalhados a seguir. A Figura 18 ilustra a previsão pelo método de Holt-Winters-Taylor com sazonalidade aditiva, além da série original de carga.

Figura 18 – Método de Holt-Winters-Taylor, sazonalidade aditiva, entre o período de 10/11/1997 a 16/11/1997



Fonte: autor

#### 3.3.2.1 Holt-Winters-Taylor original

Como descrito na seção 2.1.1.4, Taylor (2003) afirma que o modelo HWT original (versão multiplicativa) pode ser adaptado para a versão aditiva, modelo 3.4.

$$Y_t = l_t + d_t + w_t + e_t \quad (3.4)$$

Onde,

$Y_t$ : valor observado;

$l_t$ : nível da série;

$d_t$ : sazonalidade diária da série;

$w_t$ : sazonalidade semanal da série;

$e_t$ : ruído observado,  $e_t \sim N(0, \sigma^2)$ ;

$t$ : tempo.

Pedreira (2018) implementou esta versão (sazonalidade aditiva), neste trabalho chamada de *hwt1*. Sua previsão é feita conforme o conjunto de Equações 3.5.

$$\begin{aligned}\hat{l}_t &= \lambda \cdot (Y_t - \hat{d}_{t-s_1} - \hat{w}_{t-s_2}) + (1 - \lambda) \cdot \hat{l}_{t-1} \\ \hat{d}_t &= \delta \cdot (Y_t - \hat{l}_t - \hat{w}_{t-s_2}) + (1 - \delta) \cdot \hat{d}_{t-s_1} \\ \hat{w}_t &= \omega \cdot (Y_t - \hat{l}_t - \hat{d}_t) + (1 - \omega) \cdot \hat{w}_{t-s_2}\end{aligned}\tag{3.5}$$

$$\hat{X}_{t+k} = \hat{l}_t + \hat{d}_{t-s_1+k} + \hat{w}_{t-s_2+k} + \phi^k \cdot (Y_t - \hat{l}_{t-1} - \hat{d}_{t-s_1} - \hat{w}_{t-s_2})$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

$\hat{l}_t$ : estimativa do nível para o instante  $t$ ;

$\hat{d}_t$ : estimativa do fator sazonal diário para o instante  $t$ ;

$\hat{w}_t$ : estimativa do fator sazonal semanal para o instante  $t$ ;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$\delta, \lambda, \phi, \omega$ : constantes de amortecimento, com seu valor variando no intervalo  $[0, 1]$ ;

$t$ : tempo;

$s_1$ : período diário;

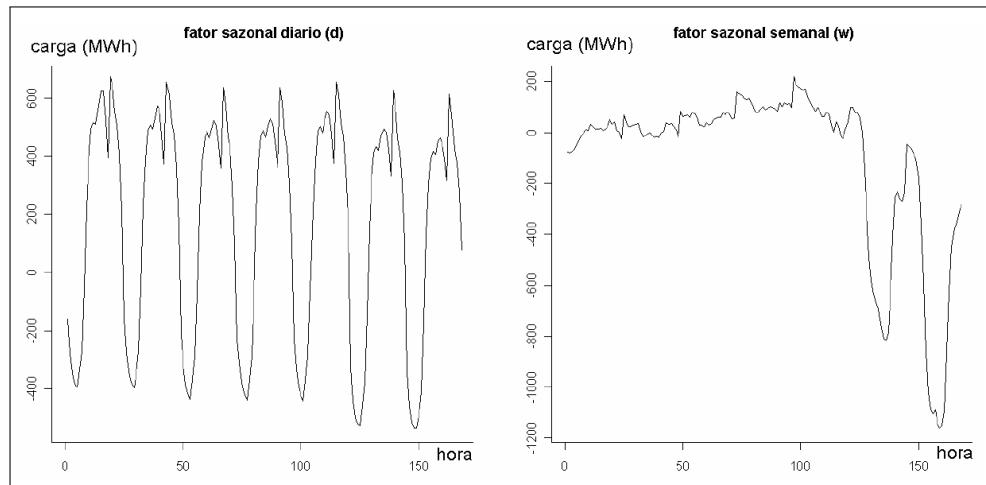
$s_2$ : período semanal;

$k$ : passos a frente do instante  $t$ .

De acordo com Pedreira (2018), os fatores sazonais diários formam um vetor de 24 valores, para modelar as observações registradas em um dia. O que ocorre de fato para os dias úteis e mantém esse comportamento aos fins de semana, como pode ser visto no gráfico do lado esquerdo da Figura 19. No lado direito da Figura 19, nota-se que o

fator sazonal semanal tem um comportamento aproximadamente constante nos dias úteis, diferentemente dos fins de semana. Pela análise dos gráficos, é possível identificar que ao invés de modelar a sazonalidade da semana, o vetor de sazonalidade semanal está corrigindo o vetor de sazonalidade diária para os fins de semana, sendo aproximadamente nulo aos dias úteis.

Figura 19 – Fator sazonal diário e fator sazonal para uma semana



Fonte: Pedreira (2018)

### 3.3.2.2 Holt-Winters-Taylor com duas sazonalidades diárias e uma sazonalidade semanal

Partindo dos fatos observados na seção anterior, uma nova proposta de Pedreira (2018) é dividir o fator sazonal diário em dois fatores: um para dias úteis e outro para fins de semana, além da sazonalidade semanal como o modelo anterior. Conforme Equação 3.6

$$Y_t = l_t + d_t^{[1]} + d_t^{[2]} + w_t + e_t \quad (3.6)$$

Onde,

$Y_t$ : valor observado;

$l_t$ : nível da série;

$d_t^{[1]}$ : sazonalidade da série para dias úteis;

$d_t^{[2]}$ : sazonalidade da série para fins de semana;

$w_t$ : sazonalidade semanal da série;

$e_t$ : ruído observado,  $e_t \sim N(0, \sigma^2)$ ;

$t$ : tempo.

Segundo Pedreira (2018) o conjunto de Equações 3.7 apresenta as componentes do método de previsão, chamado nesse estudo de *hwt2*.

$$\begin{aligned}
\hat{l}_t &= \lambda \cdot (Y_t - \hat{d}_{t-s_1}^{[1]} - \hat{d}_{t-s_2}^{[2]} - \hat{w}_{t-s_2}) + (1 - \lambda) \cdot \hat{l}_{t-1} \\
\hat{d}_t^{[1]} &= i_1 \cdot [\delta_1 \cdot (Y_t - \hat{l}_t - \hat{w}_{t-s_2}) + (1 - \delta_1) \cdot \hat{d}_{t-s_1}^{[1]}] + \\
&\quad + i_2 \cdot [\delta_1 \cdot (Y_t - \hat{l}_t - \hat{d}_{t-s_2}^{[2]} - \hat{w}_{t-s_2}) + (1 - \delta_1) \cdot \hat{d}_{t-s_1}^{[1]}] \\
\hat{d}_t^{[2]} &= i_2 \cdot [\delta_2 \cdot (Y_t - \hat{l}_t - \hat{d}_t^{[1]} - \hat{w}_{t-s_2}) + (1 - \delta_2) \cdot \hat{d}_{t-s_2}^{[2]}] \\
\hat{w}_t &= \omega \cdot (Y_t - \hat{l}_t - \hat{d}_t^{[1]} - \hat{d}_t^{[2]}) + (1 - \omega) \cdot \hat{w}_{t-s_2} \\
\hat{X}_{t+k} &= \hat{l}_t + \hat{d}_{t-s_1+k}^{[1]} + i_2 \cdot \hat{d}_{t-s_2+k}^{[2]} + \hat{w}_{t-s_2+k} + \\
&\quad + \phi^k \cdot (Y_t - \hat{l}_{t-1} - \hat{d}_{t-s_1}^{[1]} - i_2 \cdot \hat{d}_{t-s_2}^{[2]} - \hat{w}_{t-s_2})
\end{aligned} \tag{3.7}$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

$\hat{l}_t$ : estimativa do nível para o instante  $t$ ;

$\hat{d}_t^{[1]}$ : estimativa do fator sazonal para dias úteis no instante  $t$ ;

$\hat{d}_t^{[2]}$ : estimativa do fator sazonal para fins de semana no instante  $t$ ;

$\hat{w}_t$ : estimativa do fator sazonal semanal para o instante  $t$ ;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$\delta, \lambda, \phi, \omega$ : constantes de amortecimento, com seu valor variando no intervalo  $[0, 1]$ ;

$t$ : tempo;

$s_1$ : período diário;

$s_2$ : período semanal;

$k$ : passos a frente do instante  $t$ ;

$i$ : variáveis *dummy*.

No conjunto de Equações 3.7 aparecem quatro variáveis novas. A variável  $d$  passa a ser dividida em duas novas variáveis: o  $d^{[1]}$  que é a sazonalidade para os dias de semana e  $d^{[2]}$  que passa a ser a sazonalidade para os fins de semana. Assim são necessárias as variáveis *dummy* “ $i$ ”, que são usadas para indicar como o fator sazonal é calculado a depender do dia da semana.

Este método parece mais lógico do que o método HWT\_1, uma vez que separa os fatores sazonais que modelam os dias de semana daqueles que modelam os fins-de-semana. Contudo, ainda se pode objetar que é incoerente usar um mesmo vetor de fatores sazonais para modelar tanto o sábado quanto o domingo, que têm perfis claramente diferentes. (PEDREIRA, 2018)

### 3.3.2.3 *Holt-Winters-Taylor com três sazonalidades diárias e uma sazonalidade semanal*

A partir da explicação de Pedreira (2018), na seção anterior, desenvolve-se o modelo 3.8 para testar a viabilidade da separação dos fatores diários em dias úteis e subdividir os fatores sazonais para fins de semanas em dois: um fator sazonal para sábados e outro para domingos.

$$Y_t = l_t + d_t^{[1]} + d_t^{[2]} + d_t^{[3]} + w_t + e_t \quad (3.8)$$

Onde,

$Y_t$ : valor observado;

$l_t$ : nível da série;

$d_t^{[1]}$ : sazonalidade da série para dias úteis;

$d_t^{[2]}$ : sazonalidade da série para sábados;

$d_t^{[3]}$ : sazonalidade da série para domingos;

$w_t$ : sazonalidade semanal da série;

$e_t$ : ruído observado,  $e_t \sim N(0, \sigma^2)$ ;

$t$ : tempo.

A previsão para esse método é realizada conforme as Equações 3.9, referenciado nesse estudo como *hwt3*.

$$\begin{aligned}
\hat{l}_t &= \lambda \cdot (Y_t - \hat{d}_{t-s_1}^{[1]} - \hat{d}_{t-s_2}^{[2]} - \hat{d}_{t-s_2}^{[3]} - \hat{w}_{t-s_2}) + (1 - \lambda) \cdot \hat{l}_{t-1} \\
\hat{d}_t^{[1]} &= i_1 \cdot [\delta_1 \cdot (Y_t - \hat{l}_t - \hat{w}_{t-s_2}) + (1 - \delta_1) \cdot \hat{d}_{t-s_1}^{[1]}] + \\
&\quad + i_2 \cdot [\delta_1 \cdot (Y_t - \hat{l}_t - \hat{d}_{t-s_2}^{[2]} - \hat{w}_{t-s_2}) + (1 - \delta_1) \cdot \hat{d}_{t-s_1}^{[1]}] + \\
&\quad + i_3 \cdot [\delta_1 \cdot (Y_t - \hat{l}_t - \hat{d}_{t-s_2}^{[3]} - \hat{w}_{t-s_2}) + (1 - \delta_1) \cdot \hat{d}_{t-s_1}^{[1]}] \\
\hat{d}_t^{[2]} &= i_2 \cdot [\delta_2 \cdot (Y_t - \hat{l}_t - \hat{d}_t^{[1]} - \hat{w}_{t-s_2}) + (1 - \delta_2) \cdot \hat{d}_{t-s_2}^{[2]}] \\
\hat{d}_t^{[3]} &= i_3 \cdot [\delta_3 \cdot (Y_t - \hat{l}_t - \hat{d}_t^{[1]} - \hat{w}_{t-s_2}) + (1 - \delta_3) \cdot \hat{d}_{t-s_2}^{[3]}] \\
\hat{w}_t &= \omega \cdot (Y_t - \hat{l}_t - \hat{d}_t^{[1]} - \hat{d}_t^{[2]} - \hat{d}_t^{[3]}) + (1 - \omega) \cdot \hat{w}_{t-s_2} \\
\hat{X}_{t+k} &= \hat{l}_t + \hat{d}_{t-s_1+k}^{[1]} + i_2 \cdot \hat{d}_{t-s_2+k}^{[2]} + i_3 \cdot \hat{d}_{t-s_2+k}^{[3]} + \hat{w}_{t-s_2+k} + \\
&\quad + \phi^k \cdot (Y_t - \hat{l}_{t-1} - \hat{d}_{t-s_1}^{[1]} - i_2 \cdot \hat{d}_{t-s_2}^{[2]} - i_3 \cdot \hat{d}_{t-s_2}^{[3]} - \hat{w}_{t-s_2})
\end{aligned} \tag{3.9}$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

$\hat{l}_t$ : estimativa do nível para o instante  $t$ ;

$\hat{d}_t^{[1]}$ : estimativa do fator sazonal para dias úteis no instante  $t$ ;

$\hat{d}_t^{[2]}$ : estimativa do fator sazonal para sábados no instante  $t$ ;

$\hat{d}_t^{[3]}$ : estimativa do fator sazonal para domingos no instante  $t$ ;

$\hat{w}_t$ : estimativa do fator sazonal semanal para o instante  $t$ ;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$\delta, \lambda, \phi, \omega$ : constantes de amortecimento, com seu valor variando no intervalo  $[0, 1]$ ;

$t$ : tempo;

$s_1$ : período diário;

$s_2$ : período semanal;

$k$ : passos a frente do instante  $t$ ;

$i$ : variáveis *dummy*.

No conjunto de Equações 3.9, temos  $d^{[2]}$  e  $d^{[3]}$  respectivamente como fatores sazonais para sábados e domingos, além de  $d^{[1]}$  como fator sazonal para dias úteis. Considera-se  $i_1$ ,  $i_2$  e  $i_3$  como variáveis *dummy*, com o objetivo de adequar o modelo de acordo com o dia da semana.

#### 3.3.2.4 Holt-Winters-Taylor com três sazonalidades diárias, sem a sazonalidade semanal

Uma vez que no método HWT\_3 são usados três vetores de fatores sazonais diários diferentes, é possível que o fator sazonal semanal  $w_t$  se torne desnecessário, já que sua função nos métodos anteriores, era principalmente a de compensar as deficiências dos fatores sazonais diários. No método HWT\_4, fazemos uma simplificação do HWT\_3, eliminando o fator sazonal  $w_t$ . (PEDREIRA, 2018)

O modelo 3.10 representa o valor observado em qualquer instante  $t$ , para a formulação citada acima.

$$Y_t = l_t + d_t^{[1]} + d_t^{[2]} + d_t^{[3]} + e_t \quad (3.10)$$

Onde,

$Y_t$ : valor observado;

$l_t$ : nível da série;

$d_t^{[1]}$ : sazonalidade da série para dias úteis;

$d_t^{[2]}$ : sazonalidade da série para sábados;

$d_t^{[3]}$ : sazonalidade da série para domingos;

$e_t$ : ruído observado,  $e_t \sim N(0, \sigma^2)$ ;

$t$ : tempo.

A previsão, conforme o modelo 3.10, pode ser calculada segundo o conjunto de Equações 3.11, chamado de *hwt4* nesse estudo.

$$\begin{aligned}
\hat{l}_t &= \lambda \cdot (Y_t - \hat{d}_{t-s_1}^{[1]} - \hat{d}_{t-s_2}^{[2]} - \hat{d}_{t-s_2}^{[3]}) + (1 - \lambda) \cdot \hat{l}_{t-1} \\
\hat{d}_t^{[1]} &= i_1 \cdot [\delta_1 \cdot (Y_t - \hat{l}_t) + (1 - \delta_1) \cdot \hat{d}_{t-s_1}^{[1]}] + \\
&\quad + i_2 \cdot [\delta_1 \cdot (Y_t - \hat{l}_t - \hat{d}_{t-s_2}^{[2]}) + (1 - \delta_1) \cdot \hat{d}_{t-s_1}^{[1]}] + \\
&\quad + i_3 \cdot [\delta_1 \cdot (Y_t - \hat{l}_t - \hat{d}_{t-s_2}^{[3]}) + (1 - \delta_1) \cdot \hat{d}_{t-s_1}^{[1]}] \\
\hat{d}_t^{[2]} &= i_2 \cdot [\delta_2 \cdot (Y_t - \hat{l}_t - \hat{d}_t^{[1]}) + (1 - \delta_2) \cdot \hat{d}_{t-s_2}^{[2]}] \\
\hat{d}_t^{[3]} &= i_3 \cdot [\delta_3 \cdot (Y_t - \hat{l}_t - \hat{d}_t^{[1]}) + (1 - \delta_3) \cdot \hat{d}_{t-s_2}^{[3]}] \\
\hat{X}_{t+k} &= \hat{l}_t + \hat{d}_{t-s_1+k}^{[1]} + i_2 \cdot \hat{d}_{t-s_2+k}^{[2]} + i_3 \cdot \hat{d}_{t-s_2+k}^{[3]} + \\
&\quad + \phi^k \cdot (Y_t - \hat{l}_{t-1} - \hat{d}_{t-s_1}^{[1]} - i_2 \cdot \hat{d}_{t-s_2}^{[2]} - i_3 \cdot \hat{d}_{t-s_2}^{[3]})
\end{aligned} \tag{3.11}$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

$\hat{l}_t$ : estimativa do nível para o instante  $t$ ;

$\hat{d}_t^{[1]}$ : estimativa do fator sazonal para dias úteis no instante  $t$ ;

$\hat{d}_t^{[2]}$ : estimativa do fator sazonal para sábados no instante  $t$ ;

$\hat{d}_t^{[3]}$ : estimativa do fator sazonal para domingos no instante  $t$ ;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$\delta, \lambda, \phi$ : constantes de amortecimento, com seu valor variando no intervalo  $[0, 1]$ ;

$t$ : tempo;

$s_1$ : período diário;

$s_2$ : período semanal;

$k$ : passos a frente do instante  $t$ ;

$i$ : variáveis *dummy*.

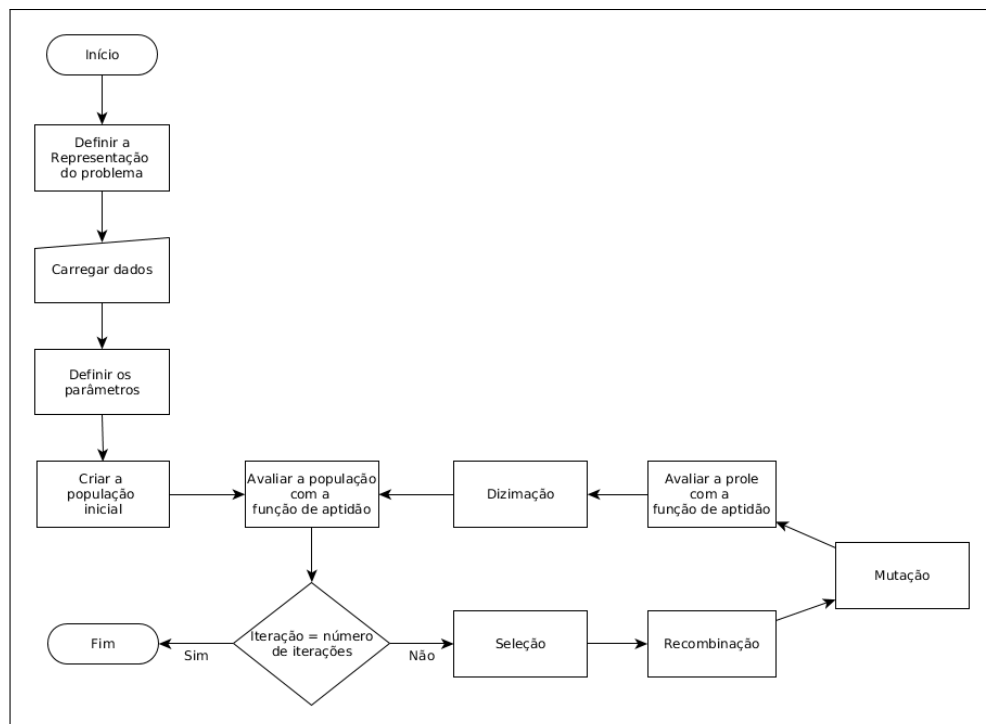
Após a explicação dos métodos de previsão utilizados e as partições da série, na próxima seção, explica-se a implementação do algoritmo para a estimação das constantes de combinação que viabiliza a combinação de todos os previsores apresentadas nesse estudo.



### 3.4 ALGORITMO GENÉTICO

A implementação do algoritmo seguiu os princípios descritos no fluxograma da Figura 20, baseado em Colherinhas (2016). Iniciando pela definição da representação do problema, carregam-se os dados e definem-se os parâmetros do algoritmo. Assim é criada a população inicial, que é avaliada através da função de aptidão. Nesse momento começa a primeira iteração ou geração, quando são selecionados indivíduos da população para gerar novos indivíduos, com auxílio dos *operadores de variação* (*recombinação* e *mutação*). Posteriormente, a prole é novamente avaliada pela função de aptidão. Contudo, como se mantém a ideia de restrição do tamanho da população, é necessário criar um critério para selecionar quais os indivíduos vão fazer o papel da população na geração seguinte, chamado *dizimação*. As interações ocorrem sucessivamente, como descrito, até encontrar os critérios de parada. Os precedimentos do AG serão detalhados nas próximas seções.

Figura 20 – Fluxograma do AG

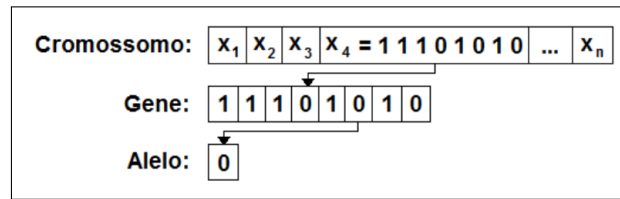


Fonte: autor

#### 3.4.1 Representação

A representação foi mantida como no caso canônico, pois a codificação binária possui a característica que uma mudança simples; a alteração, em um *bit*, pode causar uma significativa alteração no fenótipo, assim a mutação nessa representação causa uma grande variabilidade (EIBEN; SMITH, 2015). A Figura 21 esclarece cada componente de um indivíduo para a codificação binária no AG.

Figura 21 – Componentes de um indivíduo no AG



Fonte: Colherinhas (2016)

Como os coeficientes são reais, é preciso adaptar a forma de encontrar as possíveis soluções, o que leva a implementação da representação conforme Brum (2016). Na Tabela 2, os valores reais são codificados na representação binária.

Tabela 2 – Representação do AG

Índice	Valor Real	Representação Binária
0	0	0000
1	0,1	0001
2	0,2	0010
3	0,3	0011
4	0,4	0100
5	0,5	0101
6	0,6	0110
7	0,7	0111
8	0,8	1000
9	0,9	1001
10	1	1010
11	<i>Random</i>	1011
12	<i>Random</i>	1100
13	<i>Random</i>	1101
14	<i>Random</i>	1110
15	<i>Random</i>	1111

Fonte: autor

Interessante notar que essa representação possibilita limitar o espaço de busca; como podemos ver na Tabela 2, o espaço de busca é  $\Gamma \in [0, 1]$  e passo de  $10^{-1}$ . A definição do passo pode impactar na resposta final do algoritmo. Como não se trata de uma decisão trivial, essa informação foi considerada como parâmetro nesse estudo, variando  $10^n$ ,  $\{n \in \mathbb{N} \mid 1 \leq n \leq 4\}$ .

Dependendo da quantidade de *bits* utilizados para representar o gene, tem-se o número de representações binárias quantificadas por  $2^n$ ,  $\{n \in \mathbb{N}\}$ . Assim, podem acontecer casos nos quais a quantidade de representações divergem da quantidade de possíveis valores do problema. A Tabela 2 mostra que, para esse caso, existem somente 11 valores possíveis. Contudo se forem escolhidos três *bits*, tem-se  $2^3 = 8$ , que é um número insuficiente de representações para o problema. Com quatro *bits*, tem-se  $2^4 = 16$ , que viabiliza a

representação de todas as possibilidades, restando 5 opções vagas, chamados de “*random*” na tabela. A estratégia adotada nesse algoritmo, baseado em Brum (2016), caso seja escolhido um “*random*”, deve-se escolher de forma aleatória qualquer uma das 11 possíveis soluções. Conseqüentemente, o tamanho do gene varia de acordo com o tamanho do passo, Tabela 3, que resulta na variação do tamanho do cromossomo também. Assim, é possível notar que o espaço de busca é grande, visto que para cada constante necessária para a combinação, tem-se a variedade de possibilidades, como visto na Tabela 3.

Tabela 3 – Tamanho do gene

Passo	Quantidade de Valores Possíveis	Quantidade de <i>Bits</i>
$10^{-1}$	11	4
$10^{-2}$	101	7
$10^{-3}$	1001	10
$10^{-4}$	10001	14

Fonte: autor

### 3.4.2 População

O tamanho da população foi definido como parâmetro do AG. Supondo uma população de cem indivíduos, foram gerados cem vetores aleatórios com valores dos coeficientes. “A dimensão da população a escolher para um caso concreto depende de vários fatores, mas os valores padrões utilizados na maioria das situações pertencem ao intervalo [100, 500]” (PEREIRA, 2013). Foram escolhidos os valores de cem, trezentos e quinhentos para serem testados nos experimentos.

### 3.4.3 Função de aptidão

Mantendo as partições conforme descrito em 3.1, implementa-se a proposta deste trabalho, que é combinar os métodos de previsão e as série de temperaturas, conforme Equação 3.12.

$$\begin{aligned}
 \hat{X}_{t+k} = & \theta_1 \cdot \hat{X}_{t+k}^{[1]} + \theta_2 \cdot \hat{X}_{t+k}^{[2]} + \theta_3 \cdot \hat{X}_{t+k}^{[3]} + \theta_4 \cdot \hat{X}_{t+k}^{[4]} + \\
 & + \theta_5 \cdot \hat{X}_{t+k}^{[5]} + \theta_6 \cdot \hat{X}_{t+k}^{[6]} + \\
 & + \theta_7 \cdot T_{t+k} + \theta_8 \cdot T_{t+k}^2 + \\
 & + \theta_9 \cdot T'_{t+k} + \theta_{10} \cdot (T'_{t+k})^2
 \end{aligned} \tag{3.12}$$

Onde:

$\hat{X}^{[1]}$ : previsão com *hwt1*;

$\hat{X}^{[2]}$ : previsão com *hwt2*;

$\hat{X}^{[3]}$ : previsão com *hwt3*;

$\hat{X}^{[4]}$ : previsão com *hwt4*;

$\hat{X}^{[5]}$ : previsão com *naïve24*;

$\hat{X}^{[6]}$ : previsão com *naïve168*;

$T$ : temperatura;

$T'$ : diferença de temperatura;

$\theta$ : coeficiente da combinação;

$t$ : tempo;

$k$ : passos a frente do instante  $t$ .

Os métodos de previsão são avaliados através de medidas de erro. Na literatura podem ser encontrados diversas formas de medir o erro entre o valor previsto e o valor observado. No entanto, na área de previsão de carga elétrica é comum o uso do erro percentual absoluto médio, *Mean Absolute Percentage Error* (MAPE). Então para cálculo de aptidão dos indivíduos, usa-se o MAPE, assim como Pedreira (2018) que utiliza a mesma medida para calcular o erro de previsão. As equações 3.13 apresentam o cálculo do MAPE.

$$\begin{aligned}
 e_t &= Y_t - \hat{X}_t \\
 APE &= 100 \cdot \left| \frac{e_t}{Y_t} \right| \\
 MAPE &= \frac{1}{n} \cdot \sum_{t=1}^n APE_t
 \end{aligned} \tag{3.13}$$

Onde,

$Y_t$ : valor observado no instante  $t$ ;

$e$ : erro de previsão;

$t$ : tempo;

$n$ : número de elementos da série;

$\hat{X}_t$ : estimativa do valor no instante  $t$ ;

$APE$ : erro percentual absoluto (*Absolute Percentage Error*);

$MAPE$ : erro percentual absoluto médio (*Mean Absolute Percentage Error*).

Como o objetivo do AG é encontrar o valor dos coeficientes  $(\theta_1, \dots, \theta_{10})$  que combinam as previsão e variáveis exógenas para fazer uma nova previsão, essa combinação é executada através da multiplicação de um vetor e uma matriz. Essa matriz contém em cada linha uma série de previsões ou variáveis exógenas e o vetor contém os valores dos coeficientes da combinação (indivíduo do AG), cada individuo é representado por um vetor

com os valores de coeficientes factíveis. Multiplicando esse vetor pela matriz forma-se um novo vetor onde se observa a previsão combinada. Seus valores foram comparados com os valores reais de carga através do MAPE, assim calculando a aptidão dos indivíduos. Ressalta-se que este estudo é um caso de minimização, assim quanto melhor o indivíduo, menor será o seu MAPE e por consequência, menor o valor da função de aptidão.

Para a previsão, existem dois tipos de erro, chamadas de sobrestimação e subestimação. O primeiro ocorre quando o erro de previsão é maior que o valor real observado; o segundo, quando o valor da previsão é inferior ao valor observado. Como nesse estudo a combinação é uma média ponderada, os indivíduos factíveis são aqueles cuja a soma dos coeficientes é igual a um. Caso essa soma seja maior que um, é provável que ocorra uma sobrestimação do valor; de forma oposta, se o somatório for menor que um, é possível que ocorra uma subestimação. Para os dois casos os indivíduos são penalizados pelo cálculo do MAPE, um vez que os valores dos erros observados para esses casos são maiores, em valor absoluto, tornando os indivíduos não factíveis. Desta forma, não foi necessário impor penalizações arbitrárias aos indivíduos não factíveis, pois devem ser eliminados naturalmente pela seleção.

#### 3.4.4 Mecanismos de seleção dos pais

A partir da população, em cada iteração são escolhidos dois indivíduos (pais) para que sejam criados outros dois novos indivíduos (filhos). Essa seleção ocorre até que a quantidade de filhos seja igual ao tamanho da população.

A seleção dos pais é realizado pelo *ranking*, uma variação do método da roleta usado no AG canônico, pois a probabilidade de seleção de cada indivíduo tende a zero na gerações finais e inviabiliza o uso do método da roleta neste caso. No *ranking*, os indivíduos são ordenados de forma decrescente em relação aos seus valores da função de aptidão (MAPE); numerando essa ordenação, os melhores indivíduos recebem números maiores. A probabilidade de seleção do indivíduo é calculado dividindo sua posição no *ranking* pelo o somatório de todas as posições. Desta forma a probabilidade do indivíduo ser escolhido é proporcional ao valor de aptidão. Definida a probabilidade de cada indivíduo, é gerado um valor aleatório, e o intervalo que conter este valor corresponde ao indivíduo selecionado como pai.

#### 3.4.5 Operadores de variação

Para cada par de indivíduos da população, escolhidos na seleção dos pais, estes passam pela função de recombinação e posteriormente todos os indivíduos gerados (filhos), passam pela função de mutação. Ambos os operadores de variação ocorrem da seguinte forma: primeiro é sorteado um valor “ $p$ ”, que é comparado com a respectiva taxa do

operador de variação ( $q$ ), se  $p \leq q$ , então aplica-se o operador conforme descrito abaixo, caso contrário os indivíduos serão uma simples cópia.

#### 3.4.5.1 *Recombinação*

A recombinação é executada pelo cruzamento em “ $n$ ” pontos, uma generalização do método de cruzamento usado no AG canônico (cruzamento de um ponto). Os pais são formados por um cromossomo composto por “ $n$ ” genes. De forma aleatória, é definido o número de divisões a ser executada nestes cromossomos. Seja “ $i$ ” o  $i$ -ésimo gene, escolhido aleatoriamente; desta maneira o  $i$ -ésimo gene do pai *um* será trocado pelo  $i$ -ésimo gene do pai *dois*. Essa operação é feita sucessivamente até completar o número de divisões, formando o primeiro indivíduo. A Figura 6, da seção 2.3.5.1, ilustra esse mecanismo. O indivíduo *dois* é criado de forma análoga ao primeiro. Segundo Pereira (2013), em geral os valores dos parâmetros de recombinação são altos variando de 0,7 a 1,0, pois esta variação é o principal responsável por construir boas soluções. Para esse estudo, foram feitos experimentos com taxas de cruzamento variando entre os seguintes valores 0,7, 0,85 e 1,0.

#### 3.4.5.2 *Mutação*

A função de mutação é aplicada a todos os indivíduos gerados da recombinação (filhos). Este operador (implementado assim como no AG canônico), escolhe um *bit* aleatório do cromossomo para inverter o valor do alelo, dando origem a um novo indivíduo. Segundo Pereira (2013), o valor de taxa de mutação em geral é baixo; podem-se encontrar valores entre 0,001 e 0,01, pois sua função é aumentar a diversidade de populações e evitar que a solução fique presa em ótimos locais. Seguindo essa ideia, nesse estudo são experimentados os seguintes valores: 0,001, 0,0055 e 0,01.

#### 3.4.6 **Mecanismo de sobrevivência**

O mecanismo de sobrevivência escolhido foi o *elitismo*, diferentemente do que é usado no AG canônico que é de forma *geracional*. Esse mecanismo é executado após a geração da prole (filhos) pela aplicação dos operadores de variação. A aptidão da prole é avaliada, assim é selecionado o indivíduo menos apto da prole, para ser comparado com o indivíduo mais apto da população. Se o valor de aptidão do indivíduo da prole for pior que o indivíduo da população, ou seja, se o valor da aptidão do filho for maior que a do pai, esse indivíduo é substituído na prole pelo pai, pois se trata de um problema de minimização. Após essa operação, a prole assume o posto da população, eliminando os outros indivíduos.

### 3.4.7 Inicialização e Condição de parada

Com base nos conceitos apresentados em 3.4.2 gera-se a população inicial através de uma função que sorteia valores factíveis para os coeficientes e cria um indivíduo, satisfazendo a condição estabelecida em 3.4.3, que exige que o somatório dos coeficientes deve ser igual a um. Esta operação é repetida até gerar um número suficiente de indivíduos, para atender o parâmetro do tamanho da população, permitindo o início do AG. Por outro lado, a condição de parada foi estabelecida como cem iterações.

### 3.4.8 Definição dos parâmetros do AG

Inicialmente a implementação do AG é desenvolvida com base em Pedreira (2018), onde os quatro previsores de HWT são combinados conforme Equação 3.14, uma simplificação da Equação 3.12. Os experimentos são implementados em linguagem R e executados em uma máquina com oito *gigabyte* (GB) de memória RAM, um *terabyte* (TB) de HD, processador Intel® Core™ i7 CPU 950 e oito núcleos de 3,07 *gigahertz* (GHz), placa de vídeo NVIDIA Corporation GF100GL e sistema operacional Linux Fedora 27.

$$\hat{X}_{t+k} = \theta_1 \cdot \hat{X}_{t+k}^{[1]} + \theta_2 \cdot \hat{X}_{t+k}^{[2]} + \theta_3 \cdot \hat{X}_{t+k}^{[3]} + \theta_4 \cdot \hat{X}_{t+k}^{[4]} \quad (3.14)$$

Onde:

$\hat{X}^{[1]}$ : previsão com *hwt1*;

$\hat{X}^{[2]}$ : previsão com *hwt2*;

$\hat{X}^{[3]}$ : previsão com *hwt3*;

$\hat{X}^{[4]}$ : previsão com *hwt4*;

$\theta$ : coeficiente da combinação.

Para definir os parâmetros apresentados, foram realizados 27 experimentos. A configuração dos experimentos pode ser vista na Tabela 4. Cada experimento foi replicado trinta vezes, pois “Embora existam exceções, 30 repetições é um valor padrão normalmente aceito pela comunidade da computação evolutiva” (PEREIRA, 2013).

Tabela 4 – Configuração dos experimentos

Tamanho da População	Taxa de Cruzamento	Taxa de Mutação	Experimento
100	0,7	0,001	1
		0,0055	2
		0,01	3
	0,85	0,001	4
		0,0055	5
		0,01	6
	1,0	0,001	7
		0,0055	8
		0,01	9
300	0,7	0,001	10
		0,0055	11
		0,01	12
	0,85	0,001	13
		0,0055	14
		0,01	15
	1,0	0,001	16
		0,0055	17
		0,01	18
500	0,7	0,001	19
		0,0055	20
		0,01	21
	0,85	0,001	22
		0,0055	23
		0,01	24
	1,0	0,001	25
		0,0055	26
		0,01	27

Fonte: autor

Como dito na seção 3.4.1, o tamanho do passo dos valores no espaço de busca foram variados, conforme Tabela 3. Assim, esses mesmos 27 experimentos foram replicados para cada tamanho de passo. Não foram incluídas neste texto todas as tabelas geradas durante os experimentos, mas apenas uma tabela resumindo os resultados obtidos. Na Tabela 5, pode-se notar: o tempo de processamento, melhor valor de aptidão encontrado com cada valor de passo e os indivíduos que as representam (valor dos coeficientes). Note que, por uma pequena diferença de MAPE, o passo de  $10^{-3}$  apresenta melhor resultado, principalmente se levarmos em consideração o tempo de execução de cada experimento que aumentou em cerca de 900% do passo de  $10^{-3}$  para  $10^{-4}$ . Desta forma, o passo do espaço de busca foi definido como  $10^{-3}$ .



Tabela 5 – Análise do passo

Passo	Início	Fim	Tempo de Execução	Coefficientes do Melhor Indivíduo	Melhor Aptidão (MAPE)
$10^{-1}$	2018-07-27 17:02:54	2018-07-28 02:18:39	09:15:45	0,3 0,5 0,2 0	2,3163%
$10^{-2}$	2018-07-27 17:18:01	2018-07-28 02:35:22	09:17:21	0,28 0,5 0,22 0	2,3159%
$10^{-3}$	2018-07-28 10:10:13	2018-07-28 23:00:22	12:50:09	0,365 0,393 0,226 0,013	2,3037%
$10^{-4}$	2018-07-28 10:14:47	2018-08-02 01:45:21	111:30:34	0,2668 0,5087 0,2226 0,0019	2,3159%

Fonte: autor

Para Pereira (2013): “O AG poder ser avaliado pela média das melhores soluções e/ou pela melhor solução encontrada”. Assim, ao final de cada replicação foi armazenado o valor do indivíduo mais apto, formando um grupo 30 indivíduos. Formaram-se então 27 grupos composto de 30 elementos; em cada grupo foi calculada a média, moda, mediana, desvio padrão, e encontrado o indivíduo mais apto. Os resultados, podem ser vistos na Tabela 6. Pode-se observar que os coeficientes dos melhores indivíduos de cada experimento apresentam valores similares, assim como os valores de aptidão dos melhores indivíduos a cada replicação. As medidas de dispersão estão na Tabela 6 (média, moda, mediana e desvio-padrão). A priori, de um ponto de vista estatístico, o processo parece ter baixa variação. Contudo, do ponto de vista do AG, esses resultados podem indicar uma baixa variabilidade da população, o que alerta para a possibilidade de convergência prematura, que será analisado posteriormente. Concluída essa etapa, manteve-se o critério de escolha do melhor indivíduo, que aponta o experimento 25 como a melhor configuração de parâmetros: o tamanho da população com quinhentos indivíduos, a taxa de cruzamento em 1,0 e taxa de mutação no valor de 0,001, utilizando o passo de tamanho igual a  $10^{-3}$ .

Tabela 6 – Análise dos Experimentos

Experimento	Coefficientes do Melhor Indivíduo	Melhor Valor de Aptidão	Média dos Valores de Aptidão	Moda dos Valores de Aptidão	Mediana dos Valores de Aptidão	Desvio Padrão dos Valores de Aptidão
1	0,347 0,416 0,016 0,218	2,3041%	2,3089%	2,3041%	2,3079%	0,0039%
2	0,363 0,395 0,208 0,031	2,3038%	2,3061%	2,3038%	2,3053%	0,0023%
3	0,366 0,403 0,159 0,069	2,3038%	2,3065%	2,3038%	2,3057%	0,0023%
4	0,416 0,353 0,181 0,047	2,3039%	2,3072%	2,3039%	2,3063%	0,0027%
5	0,411 0,366 0,033 0,187	2,3042%	2,3067%	2,3042%	2,3060%	0,0022%
6	0,313 0,44 0,243 0,001	2,3040%	2,3071%	2,3040%	2,3063%	0,0028%
7	0,4 0,365 0,127 0,105	2,3039%	2,3075%	2,3039%	2,3059%	0,0037%
8	0,308 0,44 0,219 0,03	2,3041%	2,3072%	2,3041%	2,3065%	0,0026%
9	0,362 0,39 0,041 0,204	2,3041%	2,3060%	2,3041%	2,3055%	0,0018%
10	0,388 0,38 0,2 0,029	2,3038%	2,3053%	2,3038%	2,3050%	0,0016%
11	0,377 0,391 0,038 0,191	2,3040%	2,3052%	2,3040%	2,3048%	0,0010%
12	0,349 0,412 0,219 0,017	2,3038%	2,3055%	2,3038%	2,3050%	0,0016%
13	0,355 0,407 0,219 0,016	2,3038%	2,3051%	2,3038%	2,3048%	0,0012%
14	0,329 0,425 0,225 0,018	2,3039%	2,3051%	2,3039%	2,3049%	0,0011%
15	0,351 0,412 0,183 0,051	2,3038%	2,3050%	2,3038%	2,3047%	0,0011%
16	0,374 0,395 0,123 0,105	2,3039%	2,3046%	2,3039%	2,3045%	0,0006%
17	0,364 0,395 0,211 0,027	2,3037%	2,3047%	2,3037%	2,3047%	0,0007%
18	0,341 0,418 0,221 0,017	2,3038%	2,3050%	2,3038%	2,3047%	0,0010%
19	0,355 0,402 0,232 0,008	2,3038%	2,3046%	2,3038%	2,3044%	0,0007%
20	0,368 0,39 0,106 0,133	2,3039%	2,3048%	2,3039%	2,3046%	0,0009%
21	0,356 0,403 0,183 0,055	2,3038%	2,3045%	2,3038%	2,3043%	0,0006%
22	0,393 0,375 0,155 0,074	2,3038%	2,3044%	2,3038%	2,3043%	0,0005%
23	0,376 0,39 0,194 0,037	2,3038%	2,3044%	2,3038%	2,3041%	0,0007%
24	0,353 0,4 0,237 0,007	2,3038%	2,3048%	2,3038%	2,3045%	0,0008%
<b>25</b>	<b>0,365 0,393 0,226 0,013</b>	<b>2,3037%</b>	<b>2,3041%</b>	<b>2,3037%</b>	<b>2,3040%</b>	<b>0,0004%</b>
26	0,351 0,411 0,233 0,002	2,3038%	2,3044%	2,3038%	2,3041%	0,0007%
27	0,385 0,38 0,163 0,069	2,30379%	2,30429%	2,30379%	2,30419%	4,9730E-06%

Fonte: autor

## 4 RESULTADOS E DISCUSSÕES

Esse capítulo trata dos resultados obtidos e discussões pertinentes aos valores encontrados, bem como cometários oriundos da aplicação dos métodos. Primeiramente, valida-se a metodologia em relação ao estudo desenvolvido por Pedreira (2018) e a partir desse ponto são feitos novos experimentos com base na metodologia apresentada.

### 4.1 VALIDAÇÃO DOS MODELOS

Definidos os parâmetros do AG, é necessário replicar os experimentos feitos por Pedreira (2018) usando as quatro versões do HWT e combinando-as usando a função “optim” do R (combinação 0). Seus resultados são comparados, na Tabela 7, com os obtidos usando o AG, no intuito de validar o modelo e viabilizar novos experimentos, como proposto na Equação 3.12.

Tabela 7 – Resultados (MAPE [%])

Método	Partição 2	Partição 3	Partição 4	Partição 5	Média
naïve24*	7,68	6,54	7,11	7,63	7,34
naïve168*	4,66	3,84	5,63	6,60	5,03
hwt1*	2,89	1,71	2,15	2,98	2,43
hwt2*	2,95	1,83	2,25	2,98	2,50
hwt3*	3,11	2,11	2,29	3,04	2,64
hwt4*	3,18	2,08	2,33	3,08	2,67
combinação 0*	2,80	1,66	2,11	2,75	2,33
combinação 1**	2,87	1,67	2,12	2,96	2,41
combinação 2**	2,87	1,68	2,12	2,70	2,34

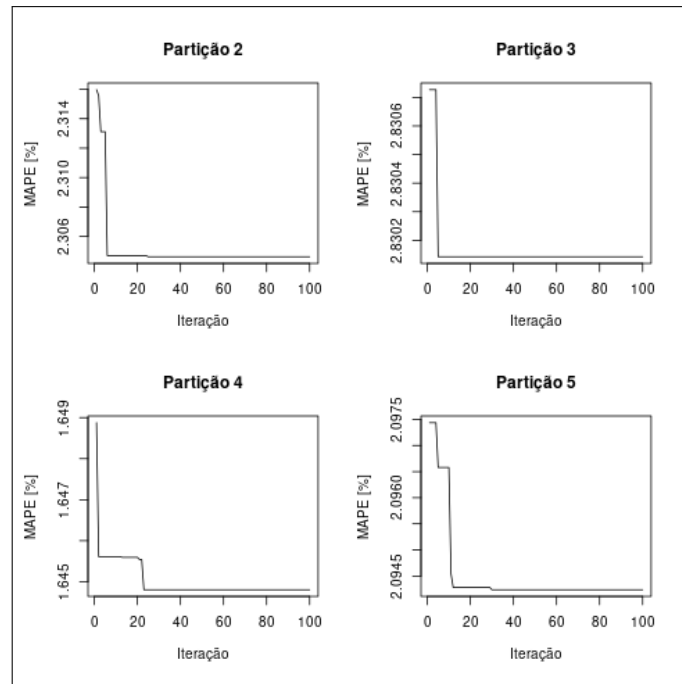
Fonte: Pedreira (2018)\*, autor\*\*

Analisando os resultados, é possível notar que as duas formas de combinação apresentam resultados similares (tanto em cada partição, quanto na média geral). Contudo, a combinação 2 apresenta uma certa vantagem. Comparando, nota-se que esta combinação apresenta valores mais próximos da combinação 0 apresentada por Pedreira (2018). Levando em consideração a média de cada previsão, nota-se um erro relativo, em relação à combinação 0, de 3,43% ( $|2,33 - 2,41|/2,33$ ) para a combinação 1 e 0,43% ( $|2,33 - 2,34|/2,33$ ) para a combinação 2. Contudo, mesmo que a combinação 1 apresente resultados inferiores, esta ainda não será descartada, pois a diferença entre os previsores é muito pequena. Como no próximo experimento serão avaliados dez parâmetros, como proposto na Equação 3.12, é possível que a combinação 1 apresente ainda resultado satisfatório, embora não seja provável.

Nesse momento cabe a análise sobre a possibilidade de convergência prematura do AG, discutida antes. Tomando a combinação 1 como exemplo (pois ambas as combinações apresentam análise de convergência parecidas), observa-se na Figura 22 o desenvolvimento

do indivíduo mais apto a cada iteração, mostrando uma queda rápida de aptidão. Isto não quer dizer que a convergência seja prematura; as vezes, a solução ótima pode ser encontrada na primeira interação, como explica Brum (2016): “É importante observar que, após uma primeira avaliação da população inicial, uma solução para o problema pode ser obtida e então finaliza-se o algoritmo”. Seguindo essa ideia, a restrição feita na criação da população inicial, a qual restringiu consideravelmente o espaço de busca, pode ter levado ao encontro da resposta final nas primeiras interações, o que não implica necessariamente em convergência prematura.

Figura 22 – Iterações do AG



Fonte: autor

Complementando a informação do gráfico, tem-se que foram necessárias seis interações para partição 2, cinco iterações para a partição 3, 21 iterações para a partição 4 e por fim 12 iterações para a partição 5. Constatando que o número de iterações varia, mas, não foram necessárias mais de 30 iterações, o que sugere ser este um número suficiente. Embora este número não seja teoricamente limitado, visto que os experimentos com dez séries (previsores e temperatura) podem apresentar um comportamento diferente, assim são mantidas as cem iterações para os próximos experimentos.

Para fins de comparação, também foram feitos os mesmos experimentos com as funções “ga” e “optim” do R. Os resultados obtidos para a combinação 1 podem ser vistos na Tabela 8. A função “optim” do pacote “stats” foi implementada utilizando 0,5 como valor inicial para todas as variáveis, e definindo a função objetivo como sendo a função de aptidão do AG, em que se busca minimizar o valor do MAPE (3.4.3), e o método de otimização "L-BFGS-B", assim como desenvolvido por Pedreira (2018). Foi estabelecido o

valor de -1 como o limite inferior do domínio de busca para as soluções e 1 para o limite superior. A função “ga” do pacote “GA”, foi implementada com a codificação real, com os mesmo parâmetros já apresentados. São eles: tamanho da população quinhentos indivíduos, taxa de cruzamento 1,0, taxa de mutação 0,001 e elitismo um, com cem interações; e o mesmo domínio de busca da função “optim”, assim como a função de aptidão. Em ambos os casos, não foram feitas replicações dos experimentos, pois o intuito desse experimento não é validar a função implementada ao longo desse estudo, mas somente observar sua coerência, se comparada com as funções disponíveis do *software*, além de avaliar se existe diferença entre o uso de representação real e binária para este estudo.

Tabela 8 – Resultados (MAPE [%]) da combinação 1 com as funções do R

Função	Partição 2	Partição 3	Partição 4	Partição 5	Média
GA	2,88	1,64	2,10	2,96	2,40
OPTIM	2,88	1,64	2,10	2,96	2,40

Fonte: autor

Durante os testes realizados com as funções, foi possível notar que a função “optim” respondeu mais rápido, embora a função “ga” possa ter uma resposta mais ágil dependendo das configurações de seus parâmetros, principalmente pelo número de iterações; contudo, ambas responderam em questão de segundos. Diante dos resultados apresentados uma primeira observação importante é a semelhança entre os resultados apresentados, pelas duas funções, o que mostra que em termos de performance, ambas são competentes. Assim, comparando os resultados da Tabela 8 com os resultados apresentados pela combinação 1 na Tabela 7, entende-se que o modelo está coerente, embora seja notório que existam variações nos resultados de cada partição e na média geral; mas estes podem ser considerados não significantes.

Após a análise dos resultados apresentados, principalmente ao comparar os resultados das tabelas apresentadas nessa seção, têm-se evidências suficientes para validar o modelo, viabilizando os testes da proposta do trabalho, conforme a Equação 3.12.

## 4.2 EXPERIMENTOS DA COMBINAÇÃO COM DEZ SÉRIES

Após o entendimento dos mecanismos dos experimentos, realizaram-se os primeiros testes, utilizando a máquina descrita na seção 3.4.8, conforme a Equação 3.12. É importante revelar que esse algoritmo levou mais de uma semana de processamento, sem encerrar. Assim, foi necessário simplificar o código pois constatou-se que a restrição de que o somatório dos coeficientes de cada indivíduo deve ser igual a um não é viável para combinações de uma quantidade maior de séries. Considerando o tempo de execução do algoritmo, observa-se que para os experimentos de validação, com o uso de quatro previsores HWT, tem-se uma média de uma hora e dez minutos para a combinação 1,

enquanto que para a combinação 2, tem-se uma média de 53 minutos. Assim, nessa etapa de experimentos, conclui-se que não é viável o uso dessa restrição na população inicial para a combinação com dez séries de dados (previsores e séries de temperaturas). Uma nova proposta de experimento foi sortear os valores de coeficientes, de forma aleatória dentro do intervalo  $\Gamma \in [0, 1]$  sem impor nenhuma restrição. Contudo, essa alteração no algoritmo pode comprometer os resultados. Para verificar isto, os experimentos foram realizados novamente, conforme a etapa de validação (com quatro previsores), mas agora sem nenhuma restrição para a geração da população inicial, (Tabela 9). Nesse cenário, temos um redução para 51 minutos o tempo de execução para encontrar a combinação 1, única combinação avaliada nesse momento.

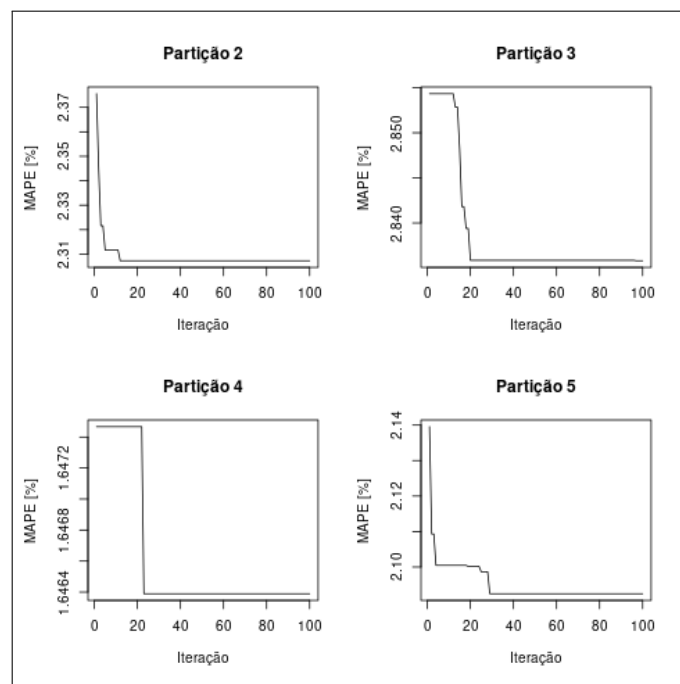
Tabela 9 – Comparação (MAPE [%]) dos resultados em relação a restrição, combinação 1

Modelo	Partição 2	Partição 3	Partição 4	Partição 5	Média
com restrição	2,87	1,67	2,12	2,96	2,41
sem restrição	2,88	1,67	2,12	2,96	2,41

Fonte: autor

Nota-se que os resultados são similares aos apresentados antes, com a exceção da partição 2, onde houve uma diferença: de 2,87% para 2,88%, esta diferença é pequena e não invalida o trabalho. Aparentemente, o maior destaque dessa alteração está no gráfico de convergência, como pode ser visto na Figura 23. Neste cenário foram necessários 11 interações para a partição 2, vinte para a partição 3, 23 para a partição 4 e 25 para a partição 5. Embora o número de interações aumente, pode-se dizer que o limite estabelecido antes de 30 interações atenderia neste caso também.

Figura 23 – Gráfico de iterações sem restrição



Fonte: autor

Assim, finalmente pode-se experimentar a proposta da Equação 3.12 (utilizando a máquina descrita na seção 3.4.8), cujos resultados podem ser vistos na Tabela 10. Esse experimento levou em média uma hora e sete minutos para combinação 1 e 54 minutos para a combinação 2. Comparando os resultados desses experimentos com os resultados de Pedreira (2018), Tabela 10, pode-se notar uma pequena diminuição do erro de previsão. Comparando os resultados do AG para quatro e dez séries, na combinação 1 a média geral diminuiu, considerando o erro relativo em 0,42% ( $|2,41 - 2,40|/2,41$ ) e, para a combinação 2 diminuiu 0,86% ( $|2,34 - 2,32|/2,34$ ). Em ambas as combinações, observa-se que a partição 4 apresenta o melhor resultado; contudo as outras partições apresentam variações dos valores apresentados antes. Podemos notar ainda que, a única partição que apresenta vantagem, nessa metodologia, é a partição 4 e que a combinação 2 permanece apresentando vantagem em relação a combinação 1.

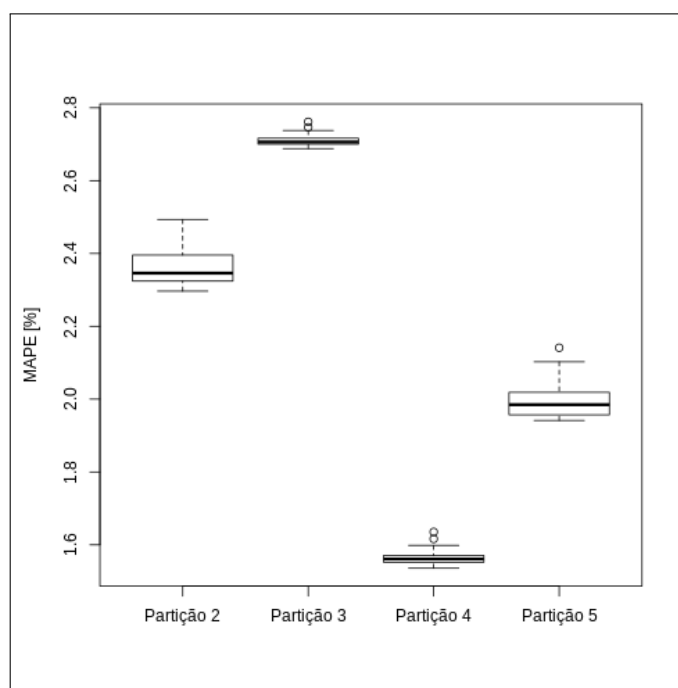
Tabela 10 – Resultados (MAPE [%]) para as combinações 1 e 2, com quatro e dez séries

Combinação	Séries	Partição 2	Partição 3	Partição 4	Partição 5	Média
combinação 0*	4	2,80	1,66	2,11	2,75	2,33
combinação 1**	4	2,87	1,67	2,12	2,96	2,41
combinação 2**	4	2,87	1,68	2,12	2,70	2,34
combinação 1**	10	2,90	1,70	2,02	2,96	2,40
combinação 2**	10	2,86	1,72	2,00	2,70	2,32

Fonte: Pedreira (2018)\*, autor\*\*

Complementando o estudo, a Figura 24 apresenta os gráficos *boxplot* comparando os resultados das trinta replicações dos experimentos para a combinação 2, onde foram armazenados somente os melhores valores de cada replicação do experimento. A figura mostra que não houve uma grande variação de resposta, e dispensa o número elevado de replicações do experimento. Ressalta-se que os *boxplots* e gráficos de iteração são referentes às amostras de treino da combinação, por isso apresentam um resultado melhor que os resultados finais obtidos nas amostras de teste. Porém, não é coerente usá-los como resposta final, visto que na implementação de métodos de previsão, não faz sentido prever para um intervalo de tempo em que já exista o valor observado. Por isso os resultados são simulados como se não tivesse a resposta futura, para que depois possam ser comparados simulando uma previsão real, na amostra de teste.

Figura 24 – Boxplot do AG com dez séries e combinação 2



Fonte: autor

Assim como na etapa de validação, os experimentos foram replicados com as funções do R: “ga” e “optim”. Para a combinação 2, seguem os resultados na Tabela 11. Nota-se que a função “optim” apresenta resultados melhores, mas não foram feitas replicações dos experimentos; dessa forma não é possível comprometer a função “ga”. Se compararmos esses resultados com os apresentados pela Tabela 10, pode-se dizer que os valores são coerentes, mas apresentam as maiores variações observadas nos resultados apresentados desse estudo.

Tabela 11 – Resultados (MAPE [%]), com dez séries para a combinação 2

Função	Partição 2	Partição 3	Partição 4	Partição 5	Média
GA	2,76	1,75	2,31	2,83	2,41
OPTIM	2,98	1,69	2,02	2,73	2,35

Fonte: autor

Diante dos resultados apresentados sobre a escolha do método de otimização, entre o “optim” e “ga” do R, cabe concluir que, no fim, ambos apresentam resultados similares. Contudo, o estudo necessário para aplicar os conceitos inerentes na função “optim” são mais demorados, se comparado ao aprendizado de um AG, que é bem simples intuitivo. Entretanto, o AG apresenta maior variação nos seus resultados, o que se apresenta como uma desvantagem, pois é necessário um aumento de replicações dos experimentos; portanto, os resultados do “optim” são encontrados em menor tempo.

Analisando os coeficientes das combinações encontradas em todos os experimentos, entende-se que não é necessário que seu somatório seja igual a um, derrubando, com base nos experimentos, a premissa feita na metodologia, principalmente no caso de dez séries.

### 4.3 COMPARAÇÃO ENTRE MÉTODOS LINEARES E RNAS

Esse estudo mostra uma pequena vantagem em combinar previsores e séries de temperatura, Neto (2014) comenta a existência de autores defendendo que para previsões de curto prazo, as variáveis climáticas podem apresentar pouca vantagem, como foi constatado nesse estudo. Então, se pensar na dificuldade em obter tanto a série de consumo quando a série de temperatura com os mesmos instantes de observação, pode inviabilizar sua aplicação na rotina de uma distribuidora. Assim, sua aplicação estará atrelada ao ganho financeiro que essa previsão pode oferecer, justificando investimentos para a manutenção dos dados para que seja viável a aplicação dessa metodologia rotineiramente. Desta forma, essa diferença no erro de previsão pode não justificar o uso dessa combinação.

Segundo Carneiro (2014), a relação entre carga elétrica e temperatura é não linear. Em estudos já apresentados com esta mesma série de dados, é possível notar uma vantagem do uso de redes neurais artificiais (RNA), pois “Uma RNA é um mecanismo que produz uma saída univariada ou multivariada a partir de complexas combinações não-lineares de uma ou mais variáveis de entrada” (NETO, 2014). Contudo, sua implementação apresenta consideráveis desvantagens:

“Embora seja difícil superar as RNAs em termos dos erros de previsão, elas apresentam a desvantagem de consumirem muitos recursos computacionais e, principalmente, apresentarem baixa robustez em relação a outros métodos. Como os algoritmos de busca são bastante sensíveis às condições iniciais, é comum existir alguma variância entre os resultados obtidos por uma mesma rede executada várias vezes (não-raro com a obtenção de valores discrepantes). Além disso, as combinações não-lineares podem ser extremamente complexas, dificultando o julgamento dos resultados e sua discussão.” (NETO, 2014)

Levando em consideração que uma RNA de fato apresente melhores resultados, como pode-se evidenciar no trabalho desenvolvido por Defilippo (2017), é preciso considerar que devido ao volume de parâmetros exigidos em uma RNA, tem-se uma grande sensibilidade em relação a temperatura. Fato também observado por Makridakis, Spiliotis e Assimakopoulos (2018), que explicam que uma RNA apresenta dificuldade em distinguir entre oscilações nos dados da série e o ruído. Segundo estes autores, os métodos estatísticos tradicionais foram desenvolvidos de acordo com as características que uma série pode apresentar, ou seja, já possuem mecanismos (pré configurado pelo usuário) para corrigir essas oscilações, como a sazonalidade, tendência, etc; enquanto que a RNA, se trata de um *blackbox*. Desta forma



a RNA aprende com a série sem intervenção do usuário, tendendo para um *over-fitting* com base nas características da série e ruídos, o que conseqüentemente irá prejudicar a previsão. Em geral, as publicações nessa área usam a temperatura do instante observado como feito nesse estudo. Contudo, os métodos lineares, como os apresentados nesse estudo, aparentemente são menos sensíveis a ruídos nos valores de temperatura, enquanto que na RNA pode levar a um erro maior, conforme conclui Defilippo (2017).

Curiosamente, como explica Defilippo (2017), a RNA vem apresentando grandes avanços nas publicações recentes, contudo essas vantagens não são observadas em competições de previsão *online*, que mostram que as técnicas lineares (como a regressão) apresentam melhores resultados. Para Makridakis, Spiliotis e Assimakopoulos (2018), que comparam métodos estatísticos com métodos de inteligência computacional (IC), em especial a RNA, nota-se que os métodos estatísticos apresentaram melhores performances. Propõem estudos para investigar os motivos das RNA apresentarem erros maiores, levando em consideração que o ARIMA (auto-regressivo integrado de médias móveis), um exemplo de método estatístico, apresenta um mecanismo parecido com a rede neural, em relação a otimizar parâmetros com o objetivo de minimizar o erro de previsão. Ainda comparando uma RNA com o ARIMA, outro ponto destacado por Makridakis, Spiliotis e Assimakopoulos (2018), é que a RNA pode apresentar um erro menor na amostra de treino, se comparado ao ARIMA; contudo, na amostra de teste, o ARIMA apresenta um erro menor. Isto mostra que devem existir estudos para evitar o *over-fitting* e que os atuais não são eficazes. Nesse caso em particular, Neto e Hippert (2018), testaram o ARIMA de forma similar aos testes desse estudo, não apresentando vantagens nos resultados, em relação a metodologia apresentada nesse estudo.

Apesar dos avanços computacionais, deve-se levar em conta a complexidade dos algoritmos, uma RNA apresenta uma complexidade muito maior se comparado com os métodos estatísticos. Assim, pode-se dizer que embora as RNA apresentem melhores valores de previsão, é possível dizer que o HWT é um previsor mais confiável em relação a resposta, muito pela menor dispersão dos resultados encontrados com HWT que na RNA.

## 5 CONCLUSÃO

Não existe método perfeito de previsão; os resultados de cada método sempre vão depender da análise da série em estudo. Assim, comparando os resultados obtidos nesse estudo com o resultados apresentados por Pedreira (2018), fica evidente que apenas a combinação linear das variantes do método de HWT é necessária para se fazer uma boa previsão. Compartilhando a mesma conclusão de Carneiro (2014) e (NETO, 2014) (trabalharam com combinações de previsores nesta mesma série de dados), que muitos autores comentam sobre o fato de acrescentar a série de temperaturas pode não apresentar grandes vantagens. O que parece ser uma vantagem de fato, é o uso da metodologia de *rolling window*, isto é, redividir a série para que a amostra de treino seja menor e consequentemente a amostra de teste também seja menor, o que representa uma vantagem para avaliar quais o melhores métodos para a série, visto que nenhum método é soberano em todas as partições. Pelos resultados apresentados nesse estudo, é possível notar que métodos diferentes podem apresentar melhores resultados em cada partição; o que parece ser promissor o estudo desse padrão ao longo dos anos da série de dados, com potencial para aprimorar o resultados de previsão.

Como o método apresentado nesse estudo não apresentou ganho em relação aos melhores resultados da literatura, possivelmente devido a existência de muitos trabalhos com essa mesma série, o que pode indicar o encontro de um ótimo local e até mesmo um ótimo global. Assim, é importante aprender com esse cenário, para evitar trabalhos futuros nessa linha. Pode-se elencar que as justificativas para os resultados são: como o método de HWT é univariado, neste caso, parte-se do pressuposto que a série já carrega as informações dos dados temperatura. Assim, para encontrar resultados melhores seria mais interessantes usar métodos multivariados. Outro ponto, é que nesse estudo foi desenvolvida uma combinação linear dos previsores e dados de temperatura; como a relação entre carga e temperatura é não linear, mesmo se usarmos funções quadrática para os dados de temperatura, a combinação continua sendo linear, pois a previsão combinada é encontrada pela soma dos termos, cada termo representa uma multiplicação de uma constante por uma variável independente.

Nesse trabalho ficam quatro propostas para trabalhos futuros. A primeira trata-se de estimar as contantes de amortecimento dos previsores para otimizar a combinação, ou seja, usar o AG para encontrar as contantes de amortecimento dos previsores e coeficientes de combinação concomitantemente. Uma outra ideia, é substituir as constantes por funções para tentar extrair o melhor de cada previsor em cada instante da série. A terceira proposta, é separar as séries em grupos de dados, avaliando a previsão em dias úteis e dias especiais separadamente, tomando os devidos cuidados para fazer uma melhor previsão em dias especiais. A última ideia, é dividir a série de acordo com as estações do ano e desenvolver métodos de previsão para cada estação, avaliando a existência de um padrão.

## REFERÊNCIAS

- ANEEL. *Atlas de energia elétrica do Brasil*. 3. ed. Brasília: Agência Nacional de Energia Elétrica, 2008. ISBN 9788587491107.
- ARORA, S.; TAYLOR, J. W. Short-term forecasting of anomalous load using rule-based triple seasonal methods. *IEEE transactions on Power Systems*, IEEE, v. 28, n. 3, p. 3235–3242, 2013.
- BAUER, R. S. M. *Cobrança Indevida na energia elétrica*. 2017. <https://blogdobauer.com.br/2016/12/02/cobranca-indevida-na-energia-eletrica/>. Acesso em 16 de julho de 2018.
- BRUM, T. C. *Sistema de engenharia Kansei para seleção de alternativas de embalagens com critérios sustentáveis de materiais*. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, Juiz de Fora, 4 2016.
- CARNEIRO, A. C. M. da S. *Previsão do consumo de energia elétrica a curto prazo, usando combinações de métodos univariados*. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, Juiz de Fora, 9 2014.
- COLHERINHAS, G. B. *Ferramenta de otimização via algoritmos genéticos com aplicações em engenharia*. Dissertação (Mestrado) — Universidade de Brasília, Brasília, 8 2016.
- DEFILIPPO, S. B. *Previsão da demanda de energia elétrica por combinações de modelos lineares e de inteligência computacional*. Tese (Doutorado) — Universidade Federal de Juiz de Fora, Juiz de Fora, 9 2017.
- EHLERS, R. S. *Análise de séries temporais*. 2009. <http://www.icmc.usp.br/ehlers/stemp/stemp.pdf>. Acesso em 06 de agosto de 2018.
- EIBEN, A. E.; SMITH, J. E. *Introduction to evolutionary computing*. 2. ed. Heidelberg: Springer, 2015. ISBN 9783662448731.
- HONG, T.; FAN, S. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, Elsevier, v. 32, n. 3, p. 914–938, 2016.
- MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, Public Library of Science, v. 13, n. 3, p. e0194889, 2018.
- MAYRINK, V. T. de M. *Avaliação do algoritmo gradient boosting em aplicações de previsão de carga elétrica a curto prazo*. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, Juiz de Fora, 8 2016.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. 2. ed. São Paulo: Egard Blucher, 2006. ISBN 978-85-212-0389-6.
- NETO, G. G. *Métodos univariados e multivariados para previsão da demanda de energia elétrica em curto prazo: um estudo comparativo*. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, Juiz de Fora, 8 2014.

NETO, G. G.; HIPPERT, H. S. Previsão de carga elétrica em curto prazo utilizando modelos de box-jenkins especificados via otimização por enxame de partículas. In: XXV SIMPEP. *Inovação e Sustentabilidade na Gestão ne Processos ne Negócios*. Bauru, São Paulo, Brasil, 2018.

PEDREIRA, T. de M. *Um estudo sobre os métodos de amortecimento exponencial para a previsão de carga a curto prazo*. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, Juiz de Fora, 9 2018.

PEREIRA, F. J. B. Algoritmos genéticos. In: GASPAR-CUNHA, A.; TAKAHASHI, R.; ANTUNES, C. H. (Ed.). *Manual de computação evolutiva e metaheurística*. Belo Horizonte: Imprensa da Universidade de Coimbra, 2013. cap. 2, p. 25–47.

PICHILIANI, M. *Análise de dados do e-commerce: indo além do analytics*. 2015. <https://www.ecommercebrasil.com.br/artigos/analise-de-dados-e-commerce-indo-alem-analytics/>. Acesso em 08 de agosto de 2018.

SONG, K.-B. et al. Hybrid load forecasting method with analysis of temperature sensitivities. *IEEE Transactions on Power Systems*, IEEE, v. 21, n. 2, p. 869–876, 2006.

TAYLOR, J. W. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, v. 54, n. 8, p. 799–805, 3 2003.

TEIXEIRA, R. F. F. *Estudo de técnicas de apoio a definições em contratos de energia elétrica*. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, Juiz de Fora, 8 2017.

TRAXCO. *El genoma de la soja*. 2010. <https://www.traxco.es/blog/noticias-agricolas/genoma-de-la-soja-alimentos-y-biocombustibles>. Acesso em 07 de junho de 2018.

ZHU, G.; CHOW, T.-T.; TSE, N. Short-term load forecasting coupled with weather profile generation methodology. *Building Services Engineering Research and Technology*, SAGE Publications Sage UK: London, England, v. 39, n. 3, p. 310–327, 2018.