

Verena de Fátima das Chagas

DISSERTAÇÃO APRESENTADA  
AO  
PROGRAMA DE PÓS-GRADUAÇÃO EM METEOROLOGIA  
DO  
INSTITUTO DE GEOCIÊNCIAS  
DO  
CENTRO DE CIÊNCIAS MATEMÁTICAS E DA NATUREZA  
DA  
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
PARA  
OBTENÇÃO DO TÍTULO  
DE  
MESTRE EM CIÊNCIAS (ÁREA: METEOROLOGIA)

Programa: Programa de Pós-graduação em Meteorologia

Orientador: Hugo Abi Karam

Durante o desenvolvimento deste trabalho a autora recebeu auxílio financeiro da CAPES

Rio de Janeiro, 26 de agosto de 2016

**Aplicação de Rede Neural Artificial para previsão do  
consumo de água residencial na Região Metropolitana  
do Rio de Janeiro em função das variáveis  
meteorológicas**

Esta dissertação de mestrado corresponde a versão original  
preparada por Verena de Fátima das Chagas.

# Aplicação de Rede Neural Artificial para previsão do consumo de água residencial na Região Metropolitana do Rio de Janeiro em função das variáveis meteorológicas

Esta dissertação de mestrado contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa realizada por Verena de Fátima das Chagas em 26/08/2016.

O original encontra-se disponível na Biblioteca do Centro de Ciências Matemáticas e da Natureza (CCMN) da Universidade Federal do Rio de Janeiro (UFRJ).

Comissão Julgadora:

- Prof. Dr. Hugo Abi Karam (orientador) - IGEO-UFRJ
- Prof. Dr. José Ricardo de Almeida França - IGEO-UFRJ
- Prof. Dr. Augusto José Pereira Filho - IAG-USP

# Resumo

CHAGAS, V. F. **Aplicação de Rede Neural Artificial para previsão do consumo de água residencial na Região Metropolitana do Rio de Janeiro em função das variáveis meteorológicas.**

2016. 100 f. Dissertação (Mestrado) - Meteorologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

Este trabalho investiga as condições meteorológicas associadas ao consumo de água residencial da população na Região Metropolitana do Rio de Janeiro (RMRJ). Além de uma análise baseada em Transformada de Ondeleita (TO), aplica-se uma Rede Neural Artificial (RNA) para modelagem do consumo em função de variáveis meteorológicas de superfície. As variáveis meteorológicas usadas como entrada da RNA foram estabelecidas a partir dos arquivos de dados horários da estação do Aeroporto Internacional Tom Jobim (Galeão), após reformatação e pareamento. Os dados de consumo de água potável foram obtidos de duas fontes independentes (ao longo de vários anos). A primeira, a partir de registros mensais de uma residência com 4 habitantes, localizada na região da Baixada Fluminense da RMRJ. A segunda, a partir de registros diários do consumo de um prédio alto residencial, localizado no Bairro das Laranjeiras na cidade do Rio de Janeiro-RJ, Brasil. Esta pesquisa apresentou os seguintes resultados: i) o treinamento otimizado de uma RNA com uma camada interna com 40 nós, para modelamento do consumo de água na RMRJ; ii) a caracterização da variação temporal do consumo de água potável a partir da evolução temporal de suas escalas dominantes obtidas por TO; iii) a estimativa da matriz de correlação entre o consumo de água e as variáveis meteorológicas da RMRJ. Em relação ao terceiro resultado, os níveis de consumo mostram-se tão elevados que predomina a autocorrelação do consumo, em comparação aos valores explicativos da correlação pelas variáveis meteorológicas. Apesar disto, a inclusão no treinamento da rede com variáveis meteorológicas, tem o papel de contribuir para a redução do erro sistemático, sendo uma vantagem. O trabalho se insere nas investigações da hidrometeorologia urbana das grandes cidades tropicais, como uma contribuição ao entendimento dos possíveis impactos de modificações da superfície e das mudanças climáticas sobre a disponibilidade de água potável.

**Palavras-chave:** Consumo de água potável; Modelo matemático do consumo de água;

Transformada de Ondeleta; Rede Neural Artificial; Região Metropolitana do Rio de Janeiro, Brasil.

# Abstract

CHAGAS, V. F. **Application of Network of Artificial Neural for forecast residential water consumption in the metropolitan area of Rio de Janeiro depending the meteorological variables.**

2016. 100 f. Dissertação (Mestrado) - Meteorologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

This work investigates weather conditions associated with residential water consumption of the population in the Metropolitan Area of Rio de Janeiro (MARJ), in Brazil. In addition to an analysis based on wavelet transform (TO), applies an Artificial Neural Network (ANN) for modeling the consumption due to surface weather variables. The meteorological variables used as input RNA were established from hourly data records of the Tom Jobim International Airport weather station, after pairing for deletion of missing data. Data from drinking water consumption were obtained from two independent sources over a period of several years. The first, from monthly records of a residence with 4 people, located in the region of Baixada Fluminense of RMRJ. The second, from daily records of consumption of a high residential building, located in the neighborhood of Laranjeiras in the city of Rio de Janeiro, RJ, Brazil. This research showed the following results: i) the optimal training of RNA with 40 knots in one inner layer allows the modeling of water consumption in RMRJ in first approximation; ii) the characterization of the temporal variation of the drinking water consumption presented in the form of temporal evolution of its dominant scales, obtained by TO; iii) an estimate of the correlation matrix between the consumption of water and meteorological variables observed in RMRJ. For this third result, consumption levels are so high that their variations are practically dominated by autocorrelation (persistence), i.e. compared with explanatory values associated with correlations with meteorological variables. Nevertheless, the inclusion in the training of the network of meteorological variables actually have the role of contributing to a small reduction of systematic error, which is an advantage. This work is part of the investigations of urban hydrometeorology of the larger tropical cities, considering the likely impact of surface modifications and climate change on the availability of drinking water.

**Keywords:** Potable Water Consumption; Mathematical Modelling of Water Consump-

tion; Wavelet Transform; Artificial Neural Network; Metropolitan Area do Rio de Janeiro, Brazil.

# Dedicatória

Dedico a minha mãe Cristina e meu avô Luiz Chagas.



# Agradecimentos

A Deus pelo dom da vida e capacidade de terminar essa caminhada com saúde.

A minha mãe Cristina e meu avô Luiz Chagas que sempre me apoiaram nas minhas decisões e acreditaram que esse sonho seria possível apesar das dificuldades.

Ao meu orientador Hugo Abi Karam por acreditar no meu potencial para o desenvolvimento desse trabalho, pela paciência, pelos anos de convivência e pela contribuição no meu desenvolvimento pessoal e profissional.

Ao meu querido Diogo, por sempre incentivar meu crescimento acadêmico apesar da distância que iria nos separar, pelo amor incondicional e apoio durante essa caminhada.

Aos amigos de laboratório, Suellen, Haroldo, Wagner e Diemerson que se tornaram minha família na UFRJ. Obrigada pelo companheirismo inclusive nos fins de semana no Lhydex.

As amigas, Bruna, Zâmara, Jéssica e Fernanda que sempre mandavam mensagens de apoio mesmo tão distantes.

A amiga Vivian Mariano, pelo apoio, incentivo e amizade nos bons e maus momentos.

Ao colega Wallace por compartilhar o script no programa R.

Ao Sr. Genival pela disponibilidade dos dados de consumo do prédio residencial.

Aos professores do PPGM-UFRJ pela contribuição no meu desenvolvimento acadêmico.

A CAPES pelo auxílio financeiro durante o desenvolvimento deste trabalho.

Enfim, quero agradecer a todos que ajudaram direta ou indiretamente na realização

desse trabalho.

# Sumário

Lista de Abreviaturas

Lista de Figuras

Lista de Tabelas

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Considerações Preliminares . . . . .	10
1.2	Objetivo . . . . .	10
<b>2</b>	<b>Materiais, conceitos e métodos</b>	<b>13</b>
2.1	Área de estudo . . . . .	13
2.2	Dados utilizados . . . . .	13
2.2.1	Dados meteorológicos (formato METAR) . . . . .	13
2.2.2	Dados de consumo de água potável . . . . .	14
2.2.3	Tratamento dos dados faltantes . . . . .	15
2.3	Ondeletas . . . . .	17
2.4	Redes Neurais Artificiais . . . . .	20
2.4.1	Índices estatísticos da RNA . . . . .	26
<b>3</b>	<b>Resultados</b>	<b>35</b>
3.1	Dados . . . . .	35
3.2	Consumo de água de uma casa da RMRJ . . . . .	47
3.3	Consumo de água em uma edificação-condomínio da RMRJ . . . . .	50
3.4	Transformada de Ondeleta do consumo de água (prédio residencial) . . . . .	52
3.5	Rede Neural Artificial . . . . .	56
3.6	Desempenho estatístico dos modelos . . . . .	71
<b>4</b>	<b>Conclusões</b>	<b>85</b>
<b>5</b>	<b>Anexo</b>	<b>87</b>
5.1	Programação . . . . .	87



# Lista de Abreviaturas

AGEVAP	Associação Pró-Gestão das Águas da Bacia Hidrográfica do Rio Paraíba do Sul ( <i>Pro-Management Association of Basin Waters of Paraíba do Sul River</i> )
ANA	Agência Nacional de Águas ( <i>National Water Agency</i> )
AQUASTAT	Sistema Global de Informação em Água e Agricultura da ONU ( <i>Information System on Water in Agriculture</i> )
BRIICS	Brasil, Rússia, Índia, Indonésia, China e Rússia
BRPS	Bacia do Rio Paraíba do Sul ( <i>Paraíba do Sul River Basin</i> )
CEDAE	Companhia Estadual de Águas e Esgotos ( <i>State Company for Water and Sewage</i> )
COMPERJ	Complexo Petroquímico do Rio de Janeiro ( <i>Petrochemical Complex of Rio de Janeiro</i> )
DECEA	Departamento de Controle do Espaço Aéreo ( <i>Department of Airspace Control</i> )
ENOS	El Niño Oscilação Sul ( <i>El Niño South Oscillation</i> )
ETA	Estação de Tratamento de Água ( <i>Water Treatment Station</i> )
FAO	Organização das Nações Unidas para a Alimentação e a Agricultura ( <i>Food and Agriculture Organization of the United Nations</i> )
OECD	Organização para a Cooperação e Desenvolvimento Econômico ( <i>Organization for Economic Co-operation and Development</i> )
ONU	Organização das Nações Unidas ( <i>United Nations Organization</i> )
PCJ	Piracicaba, Capivari e Jundiá
PNUD	Programa das Nações Unidas para Desenvolvimento ( <i>United Nations Program for Development</i> )
ROW	Resto do Mundo ( <i>Rest Of the World</i> )
SABESP	Companhia de Saneamento Básico do Estado de São Paulo ( <i>Basic Sanitation Company of the State of São Paulo</i> )
SNIS	Sistema Nacional de Informações sobre Saneamento ( <i>National System of Sanitation Information</i> )

## LISTA DE ABREVIATURAS

# Lista de Figuras

1.1	Demanda de água global, em 2000 e 2050. Os BRIICS (Brasil, Rússia, Índia, Indonésia, China, África do Sul); OECD (Organização para a Cooperação Econômica e Desenvolvimento); ROW (resto do mundo). Este gráfico apenas mede a demanda de “água azul” e não considera agricultura de sequeiro. Fonte: OECD Environmental Outlook Baseline; reprodução de IMAGE. . . . .	2
1.2	Bacia Hidrográfica do Rio Paraíba do Sul. Fonte: ANA . . . . .	7
1.3	Mapa da bacia hidrográfica do Guandu, abrangendo 15 municípios do RJ. Fonte: Tubbs Filho et al. (2012)[36] . . . . .	8
2.1	Ondeleta complexa de Morlet, a) ondeleta de Morlet com largura e amplitude arbitrária e com o tempo ao longo do eixo horizontal e b) construção da ondeleta de Morlet (tracejado) a partir de uma onda seno modulada por uma curva gaussiana. Fonte: Torrence e Compo (1998) . . . . .	19
2.2	Neurônios do sistema nervoso central dos vertebrados: a) Neurônio motor da medula espinhal; b) Célula de Ourkinje encontrada no cerebelo. Esta célula é notável pela extensa ramificação da sua árvore dentrital da qual apenas uma pequena parte é mostrada na figura, podendo ultrapassar várias dezenas de vezes as dimensões do corpo celular. Fonte: Kovács (2006) . . . . .	21
2.3	Modelo de RNA e suas classificações. Fonte: Cardon et al. (1994) . . . . .	22
2.4	Modelo de uma rede neural artificial . . . . .	24
2.5	Primeiras três amostras de $K = 10$ em que cada interação, 90% dos dados são utilizados para treinamento e os 10 % restantes são usados para o teste, sem repetição da amostra. Fonte: Bontempi (2003) . . . . .	34
3.1	Variação temporal da temperatura horária do ar na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	36
3.2	Histograma da temperatura do ar na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. A linha contínua representa a curva da Normal. . . . .	36
3.3	Quantil-Quantil da temperatura do ar na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	37

LISTA DE FIGURAS

3.4	Boxplot anual da temperatura horária do ar na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	37
3.5	Variação temporal da umidade relativa horária na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	38
3.6	Histograma da umidade relativa na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. A linha contínua representa a curva da Normal. . . . .	39
3.7	Quantil-Quantil da umidade relativa na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	40
3.8	Boxplot anual da umidade relativa na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	40
3.9	Histograma da pressão atmosférica na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. A linha contínua representa a curva da Normal. . . . .	41
3.10	Quantil-Quantil da pressão atmosférica na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	42
3.11	Boxplot anual da pressão atmosférica na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	42
3.12	Variação temporal da velocidade horária do vento na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	43
3.13	Histograma da velocidade do vento na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. A linha contínua representa a curva da Normal. . . . .	43
3.14	Quantil-Quantil da velocidade do vento na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	44
3.15	Boxplot anual da velocidade do vento na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. . . . .	45
3.16	Variação temporal do consumo diário de água (união de dados de uma residência e de um prédio) para o período de 2011 a 2015. . . . .	45
3.17	Histograma do consumo diário de água (união de dados de uma residência e de um prédio) para o período de 2011 a 2015. A linha contínua representa a curva da Normal. . . . .	46
3.18	Quantil-Quantil do consumo diário de água (união de dados de uma residência e de um prédio) para o período de 2011 a 2015. . . . .	47
3.19	Boxplot anual do consumo diário de água (união de dados de uma residência e de um prédio) para o período de 2011 a 2015. . . . .	48
3.20	Variação temporal do consumo acumulado por pessoa de uma família residente no município de São João de Meriti-RMJ, no período de novembro/2009 a maio/2015. . . . .	48

3.21	Evolução temporal por diferença simples e centrada do consumo diário de água individual na residência familiar em São João de Meriti-RMRJ, no período entre novembro de 2009 e maio de 2015. . . . .	49
3.22	Frequência relativa do consumo diário de água de cada membro da família residente em São João de Meriti-RMRJ, no período compreendido entre novembro de 2009 a maio de 2015. . . . .	50
3.23	Registro do consumo acumulado de água ao longo do tempo do prédio residencial em Laranjeiras, Rio de Janeiro, entre fevereiro de 2011 a dezembro de 2015. . . . .	51
3.24	Frequência relativa do consumo de água por pessoa residente em um prédio residencial em Laranjeiras, Rio de Janeiro. A curva contínua representa a função gaussiana ajustada. . . . .	51
3.25	Gráfico com ajuste de polinômio de 6 <sup>a</sup> ordem do consumo diário de água por pessoa residente em um prédio residencial em Laranjeiras, Rio de Janeiro. . . . .	52
3.26	Gráfico das funções de distribuição probabilidade e acumulada (f.d.p. e f.d.a.) por pessoa residente em um prédio residencial em Laranjeiras, Rio de Janeiro. . . . .	52
3.27	Taxa de consumo de água interpolada e original para a primeira semana do ano 2013. . . . .	53
3.28	Evolução temporal da parte real da transformada de ondeleta da taxa de consumo de água do prédio residencial. . . . .	53
3.29	Evolução temporal da potência da transformada de ondeleta da taxa de consumo de água do prédio residencial. . . . .	54
3.30	Evolução temporal da parte real da transformada de ondeleta da taxa de consumo ACUMULADO de água do prédio residencial. . . . .	55
3.31	Evolução temporal da potência da transformada de ondeleta da taxa de consumo ACUMULADO de água do prédio residencial. . . . .	55
3.32	Etapas 1, 2 e 3 de treinamento do modelo M1 (t, vel, p, ur e w-1). . . . .	58
3.33	Histogramas das etapas 1, 2 e 3 de treinamento do modelo M1 (t, vel, p, ur e w-1). . . . .	59
3.34	Gráfico Q-Q das etapas 1, 2 e 3 de treinamento do modelo M1 (t, vel, p, ur e w-1). . . . .	60
3.35	Etapas 1, 2 e 3 do teste do modelo M1 (t, vel, p, ur e w-1). . . . .	61
3.36	Histogramas das etapas 1, 2 e 3 do teste do modelo M1 (t, vel, p, ur e w-1). . . . .	62
3.37	Gráficos Q-Q das etapas 1, 2 e 3 do teste do modelo M1 (t, vel, p, ur e w-1). . . . .	63
3.38	Etapas de validação cruzada do modelo M1 (t, vel, p, ur e w-1). . . . .	64
3.39	Histograma da validação cruzada do modelo M1 (t, vel, p, ur e w-1). . . . .	65
3.40	Gráfico Q-Q da validação cruzada do modelo M1 (t, vel, p, ur e w-1). . . . .	65
3.41	Etapas de validação cruzada do modelo M2 (t, p, ur, vel, w-7). . . . .	66

LISTA DE FIGURAS

3.42	Histograma da validação cruzada do modelo M2 (t, p, ur, vel, w-7).	67
3.43	Etapas de validação cruzada do modelo M3 (ano, mês, nascer, pôr, t, p, ur, vel, u, v, w-1).	68
3.44	Histograma da validação cruzada do modelo M3 (ano, mês, nascer, pôr, t, p, ur, vel, u, v, w-1).	69
3.45	Etapas de validação cruzada do modelo M4 (w-1, w-2).	70
3.46	Histograma da validação cruzada do modelo M4 (w-1, w-2).	71
3.47	Etapas de validação cruzada do modelo M5 (w-1, w-2, w-3, w-4, w-5, w-6, w-7).	72
3.48	Histograma da validação cruzada do modelo M5 (w-1, w-2, w-3, w-4, w-5, w-6, w-7).	73
3.49	Etapas de validação cruzada do modelo M6 (t, vel, p, ur).	74
3.50	Histograma da validação cruzada do modelo M6 (t, vel, p, ur).	75
3.51	Etapas de validação cruzada do modelo M7 (vel, u, v, p, ur).	76
3.52	Histograma da validação cruzada do modelo M7 (vel, u, v, p, ur).	77
3.53	Etapas de validação cruzada do modelo M8 (t, ur).	78
3.54	Histograma da validação cruzada do modelo M8 (t, ur).	79
3.55	Tabela de contingência estilo 6x6 do modelo M1 (t, vel, p, ur e w-1).	80
3.56	Tabela de contingência estilo 6x6 do modelo M4 (w-1, w-2).	80
3.57	Tabela de contingência estilo 2x2 do modelo M1 (t, vel, p, ur e w-1).	81
3.58	Tabela de contingência estilo 2x2 do modelo M4 (w-1, w-2).	81
3.59	Distribuição dos dados observados e previstos do modelo M1 (t, vel, p, ur e w-1).	82
3.60	Distribuição dos dados observados e previstos do modelo M4 (w-1, w-2).	82
3.61	Matriz de correlação entre as variáveis de entrada da RNA.	83
3.62	Matriz de dispersão entre as variáveis da RNA, sendo a coluna 1 e linha 1: temperatura, coluna 2 e linha 2: velocidade do vento, coluna 3 e linha 3: pressão atmosférica, coluna 4 e linha 4: umidade relativa, coluna 5 e linha 5: consumo de dois dias anteriores e coluna 6 e linha 6: consumo de um dia anterior.	83

# Lista de Tabelas

1.1	Reservas de água da Terra. (Gleick, 1993) . . . . .	3
1.2	Reserva globais em porcentagens. (Fonte: Gleick, 1993) . . . . .	3
1.3	Reservas de água doce renováveis em $\text{km}^3 \text{ ano}^{-1}$ para os países mais ricos em água. . . . .	4
1.4	Reservas de água doce renováveis para a maioria dos países pobres em água. (Fonte: Gleick, 1993) . . . . .	4
1.5	Total de recursos hídricos renováveis (internos e externos) nos países da América do Sul. (Fonte: FAO, 2014) . . . . .	5
1.6	Demandas médias para abastecimento urbano nas regiões geográficas do Brasil, para os anos de 2005, 2015 e 2025. (Fonte: ANA, 2010b) . . . . .	6
2.1	Tabela de contingência. (Fonte:EUMETCAL, 2016) . . . . .	31
3.1	Estruturas dos modelos para previsão de consumo de água residencial na Região Metropolitana do Rio de Janeiro usando consumos de água anteriores e variáveis meteorológicas. . . . .	56
3.2	Desempenho estatístico dos modelos M1, M2, M3, M4, M5, M6, M7 e M8 para os índices estatísticos POD, FAR, DD, TS e ETS. . . . .	75
3.3	Desempenho estatístico dos modelos M1, M2, M3, M4, M5, M6, M7 e M8 para os índices estatísticos $\sigma_{obs}$ , $\sigma_{mod}$ , MAE, BIAS, $R^2$ e MSE. . . . .	77
3.4	Desempenho estatístico dos modelos M1, M2, M3, M4, M5, M6, M7 e M8 para os índices estatísticos $MSE_{pers}$ , SSA, $RMSE_{mod}$ , $RMSE_{ref}$ e $POD_{chance}$ . . . . .	79

## LISTA DE TABELAS

# Capítulo 1

## Introdução

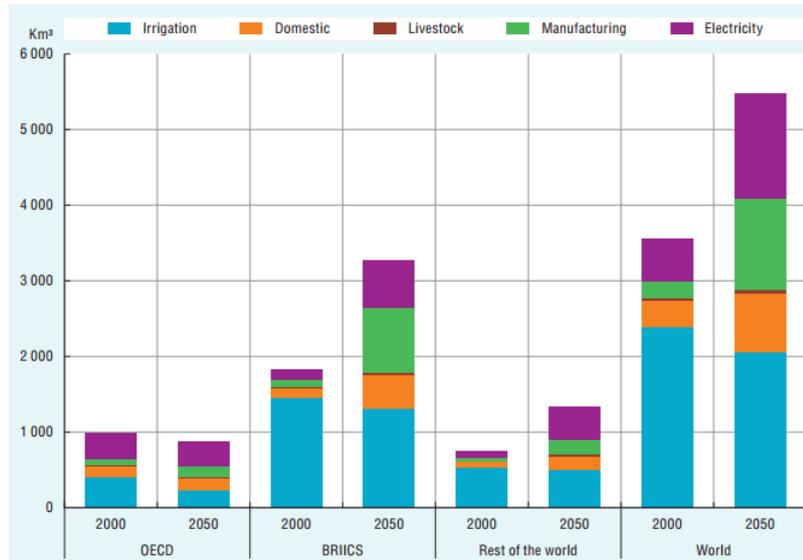
O acesso a água é uma necessidade humana elementar e um direito humano fundamental. Assim, em um mundo cada vez mais desenvolvido, há várias milhões de pessoas sem acesso a água potável. Essa inacessibilidade é ocasionada por diversos problemas como, orçamento limitado para distribuição de água em alguns países, o alto custo para obter água potável nos países mais pobres do mundo, dentre outros. Muitos países progrediram bastante no fornecimento de água aos seus cidadãos e um desenvolvimento crescente é observado no que diz respeito a inovações para aproveitamento da água. [59].

A água para sobrevivência engloba um conjunto diferente de desafios. Cerca de 1,4 bilhões de pessoas vivem em cursos d'água em que a utilização da água excede as taxas de recarga. O resultado dessa utilização são evidentes: os rios estão a secar, os lençóis freáticos a diminuir e os ecossistemas dependentes de água estão cada vez mais degradados. Sobretudo, o mundo está a perder um dos seus recursos naturais mais preciosos. [59]

Sabe-se que a água é o recurso natural mais importante da Terra, então, a Organização das Nações Unidas (ONU) trata com grande importância os problemas relacionados a esse recurso renovável. Desde 1972 até 2003, a ONU realiza conferências, fóruns e cúpulas mundiais para discutir e elaborar declarações, com relação ao direito de acesso a água e a condições sanitárias decentes. Em 2003 realizou-se o terceiro fórum mundial de água em Kyoto e nesta ocasião, assinou-se uma declaração final que consistiu nos objetivos de desenvolvimento do milênio, um deles era a redução pela metade até 2015 da proporção de pessoas sem acesso à água potável e/ou saneamento. [57]

O acesso a água é um direito essencial de cada indivíduo. A falta desse bem prejudica o desenvolvimento humano além de comprometer o desenvolvimento global. A aceleração da urbanização, o aumento da industrialização e a melhoria nos padrões de vida em geral, resultam em um aumento do consumo global de água nas cidades. A Figura 1.1 mostra a demanda mundial de água projetada para aumentar em 55% até 2050, devido principalmente ao aumento na demanda de produção (+400%), geração de energia elétrica (+140%) e uso doméstico (+130%), todos resultados da crescente urbanização de países em desenvolvimento. [56]

O volume da água em nosso planeta é praticamente constante. A água doce, utilizada com maior facilidade para consumo humano, é um bem natural renovável, apresentando variações de volume com as mudanças climáticas. [8]



**Figura 1.1:** Demanda de água global, em 2000 e 2050. Os BRICS (Brasil, Rússia, Índia, Indonésia, China, África do Sul); OECD (Organização para a Cooperação Econômica e Desenvolvimento); ROW (resto do mundo). Este gráfico apenas mede a demanda de “água azul” e não considera agricultura de sequeiro.

Fonte: OECD Environmental Outlook Baseline; reprodução de IMAGE.

As estimativas dos recursos hídricos que estão armazenados em diversos corpos d’água e em diferentes estados físicos são essenciais para o entendimento do ciclo natural da água e o efeito das atividades humanas. [39]

Os dados sobre o consumo de água global reunidos por cientistas soviéticos são apresentados na tabela 1.1. O volume total de estoque de água pura corresponde a 35 milhões de km<sup>3</sup>. Uma grande parte da água doce (24 milhões de km<sup>3</sup>) está na forma de gelo e neve. Lagos e rios de água doce são as principais fontes de consumo humano, contém, em média, cerca de 90 mil km<sup>3</sup> de água. [39]

A Tabela 1.2 evidencia as reservas de água globais em porcentagem, cerca de 96,5% da água existente no planeta Terra está presente nos oceanos. A quantidade total de água doce corresponde a 2,53%, mais da metade desse total é encontrado nas Geleiras e na Antártica e apenas 0,26% são encontradas em rios e lagos. A água presente no subsolo do planeta Terra (1,7% do total global) integra uma parte do ciclo hidrológico e é composta basicamente por água doce. [39]

A Tabela 1.3 mostra que o Brasil contém uma reserva 5 vezes maior que o Peru, nono colocado, enfatizando a abundância aquífera. Além disso, identifica-se que nove países compartilham cerca de 60% dos recursos hídricos renováveis. Os países que ocupam as três primeiras posições são Brasil, Rússia e Estados Unidos, respectivamente. [51]

Cerca de 12 países dividem 65% das reservas de água doce do planeta (Tabela 1.3). Brasil, Rússia e Canadá ocupam as três primeiras posições do ranking, ao passo que, Zaire ocupa a última posição. [40]

**Tabela 1.1:** *Reservas de água da Terra. (Gleick, 1993)*

	Área de distribuição (10 <sup>3</sup> km <sup>2</sup> )	Volume (10 <sup>3</sup> km <sup>3</sup> )	Camada (m)
Oceano mundial	361 300	1 338 000	3703
Água subterrânea	134 800	23 400	174
Água doce		10 530	
Umidade do solo		16,5	0,2
Geleiras e cobertura de neve permanente	16 227	24064,1	1483
Antártica	13 980	21 600	1545
Groenlândia	1802	2340	1299
Ilhas do Ártico	226	83,5	369
Regiões montanhosas	224	40,6	181
Permafrost (solo congelado)	21 000	300	14,3
Reserva de água em lagos	2058,7	176,4	85,7
Doce	1236,4	91	73,6
Salgada	822,3	85,4	104
Água do pântano	2682,6	11,47	4,276
Rios	148 800	2,12	0,014
Água biológica	510 000	1,12	0,002
Água atmosférica	510 000	12,9	0,025
Reservas totais	510 000	1 385 984,61	2718
Reservas de água doce	148 800	35 029,21	235

**Tabela 1.2:** *Reserva globais em porcentagens. (Fonte: Gleick, 1993)*

	Água total	Água doce
Oceano mundial	96,5	-
Água subterrânea	1,7	-
Água doce	0,76	30,1
Umidade do solo	0,001	0,05
Geleiras e cobertura de neve permanente	1,74	68,7
Antártica	1,56	61,7
Groenlândia	0,17	6,68
Ilhas do Ártico	0,006	0,24
Regiões montanhosas	0,003	0,12
Permafrost (solo congelado)	0,022	0,86
Reserva de água em lagos	0,013	-
Doce	0,007	0,26
Salgada	0,006	-
Água do pântano	0,0008	0,03
Rios	0,0002	0,006
Água biológica	0,0001	0,003
Água atmosférica	0,001	0,04
Reservas totais	100	-
Reservas de água doce	2,53	100

**Tabela 1.3:** *Reservas de água doce renováveis em  $\text{km}^3 \text{ ano}^{-1}$  para os países mais ricos em água.*

País	Margat, 1998 [ $\text{km}^3 \text{ ano}^{-1}$ ]	Gleick, 2000 [ $\text{km}^3 \text{ ano}^{-1}$ ]
Brasil	6220	6950
Rússia	4059	4498
EUA	3760	2478
Canadá	3220	2901
China	2800	2830
Indonésia	2530	2838
Índia	1850	1908
Venezuela	-	1317
Bangladesh	-	1211
Colômbia	1200	1070
Peru	1100	-
Myanmar	-	1046
Zaire	-	1019

**Tabela 1.4:** *Reservas de água doce renováveis para a maioria dos países pobres em água. (Fonte: Gleick, 1993)*

País	Reservas de água doce renováveis [ $\text{km}^3 \text{ ano}^{-1}$ ]
Jordânia	0,880
Singapura	0,600
Líbia	0,600
Djibouti	0,300
Cabo Verde	0,300
Emirados Árabes Unidos	0,150
Bahrein	0,116
Catar	0,053
Kuwait	0,020
Malta	0,016

Há diferenças entre as análises dos autores devido os dados hídricos de cada país possuírem variadas fontes, e ainda, esses dados são estimados por vários métodos e durante anos por diversos autores. [55]

A falta de acesso à água de qualidade não se limita aos países pobres, visto que esse bem também se torna raro para países ricos, isso pode ocasionar um problema mundial [62]. Os 54 países mais pobres em água representam apenas 1% do total das reservas [39]. Com isso, os dez países mais pobres em água são exibidos na Tabela 1.4. As últimas três posições da tabela são composta por Catar, Kuwait e Malta, respectivamente. Esses três países detêm juntos, apenas  $0,089 \text{ km}^3 \text{ ano}^{-1}$  das reservas de água doce renováveis.

A América do Sul é uma região detentora de grandes recursos hídricos. Nessa região estão algumas das bacias hidrográficas mais importantes do mundo, tal como dos rios Amazonas, Orinoco e da Prata. Entretanto, a distribuição da água é altamente desigual

**Tabela 1.5:** *Total de recursos hídricos renováveis (internos e externos) nos países da América do Sul. (Fonte: FAO, 2014)*

Países da América do Sul	Recursos hídricos renováveis ( $10^9 \text{ m}^3 \text{ ano}^{-1}$ )	Recursos hídricos renováveis per capita ( $\text{m}^3 \text{ habitante}^{-1} \text{ ano}^{-1}$ )
Argentina	876,2	20960
Bolívia	574	52913
Brasil	8647	42800
Chile	923,1	51938
Colômbia	2360	48232
Equador	442,4	27679
Guiana	271	337065
Paraguai	387,8	56057
Peru	1880	61100
Suriname	99	181985
Uruguai	172,2	50366
Venezuela	1325	42948

em termos geográficos. [21]

Grande parte da América do Sul possui regiões áridas ou semi-áridas como o Nordeste brasileiro, o Chaco Paraguaio, a Patagônia no Centro-Sul da Argentina, parte da Bolívia e do Peru (o Altiplano), destaca-se também o Atacama no litoral do Chile e o litoral sul do Peru como as regiões que estão entre as mais secas do mundo, tendo disponibilidade hídrica inferior a  $1 \text{ m}^3 \text{ habitante}^{-1} \text{ ano}^{-1}$ . [19]

A América do Sul possuía cerca de  $48.800 \text{ m}^3 \text{ pessoa}^{-1} \text{ ano}^{-1}$  na década de 80. Para o ano 2000 houve uma queda na disponibilidade de água devido o aumento da demanda, estimada em  $28.300 \text{ m}^3 \text{ pessoa}^{-1} \text{ ano}^{-1}$  [9]. Com isso, o acesso aos serviços de água em muitas cidades da América do Sul possui problemas crônicos, como a perda de água durante a distribuição e falta de racionalização no uso a nível doméstico e industrial [74].

Para entender a distribuição dos recursos hídricos na América do Sul, utiliza-se o AQUASTAT para as estatísticas totais de fluxo de água (internas ou externas, águas superficiais ou subterrâneas, medido ou calculado), i.e., não significa que esses fluxos são totalmente exploráveis e disponíveis (Tabela 1.5). [34]

Os recursos hídricos não estão distribuídos de forma uniforme no Brasil. Há regiões de escassez e outras com grande abundância. O Brasil detém 13% da água doce disponível do planeta, cerca de 80% está na bacia Amazônica, área com baixa demanda e com pouco mais de 5% da população brasileira, enquanto a região Nordeste possui quase 8% da população, incluindo as capitais Aracaju e Salvador que dispõem de menos de 0,4% das águas dos rios. [6]

Uma comparação da demanda de consumo de água foi realizada entre os anos de 2006 e 2010, em que o consumo aumentou 18%, passou de  $986 \text{ m}^3 \text{ s}^{-1}$  para  $1161 \text{ m}^3 \text{ s}^{-1}$ . Em

**Tabela 1.6:** Demandas médias para abastecimento urbano nas regiões geográficas do Brasil, para os anos de 2005, 2015 e 2025. (Fonte: ANA, 2010b)

Ano	Demanda por Região Geográfica ( $\text{m}^3 \text{s}^{-1}$ )					Total Brasil ( $\text{m}^3 \text{s}^{-1}$ )
	Norte	Nordeste	Centro-Oeste	Sudeste	Sul	
2005	34	115	33	247	65	494
2015	45	136	39	275	75	570
2025	54	151	44	298	83	630

2010, a irrigação foi responsável por 72% do consumo; para uso dos animais foi 11%; o uso urbano por 9%, o industrial por 7% e o rural (consumo das pessoas) por 1% do total consumido. [6]

Cerca de 16,4 mil km dos 105 mil km de rios federais no Brasil têm bastante importância para o governo, devido a sua qualidade e quantidade equivalente a 16% e a maior parte do país possui condição aceitável quanto à quantidade e à qualidade de água. As regiões hidrográficas Amazônica, Tocantins-Araguaia e Paraguai têm uma demanda de uso da água inferior em relação a outras regiões. [7]

Entretanto, os rios que estão em regiões metropolitanas, e.g., as bacias PCJ (Piracicaba, Capivari e Jundiaí), Paraíba do Sul e Alto Tietê têm nível crítico no que diz respeito a qualidade e quantidade, por causa da alta demanda e grande quantidade de carga orgânica jogada nos rios. Na região Nordeste há muitos rios com nível crítico em virtude da baixa disponibilidade hídrica, isso resulta na busca por água em aquíferos subterrâneos, visto que essa região possui bastante água, com capacidade de amenizar as secas que assolam a região. [7]

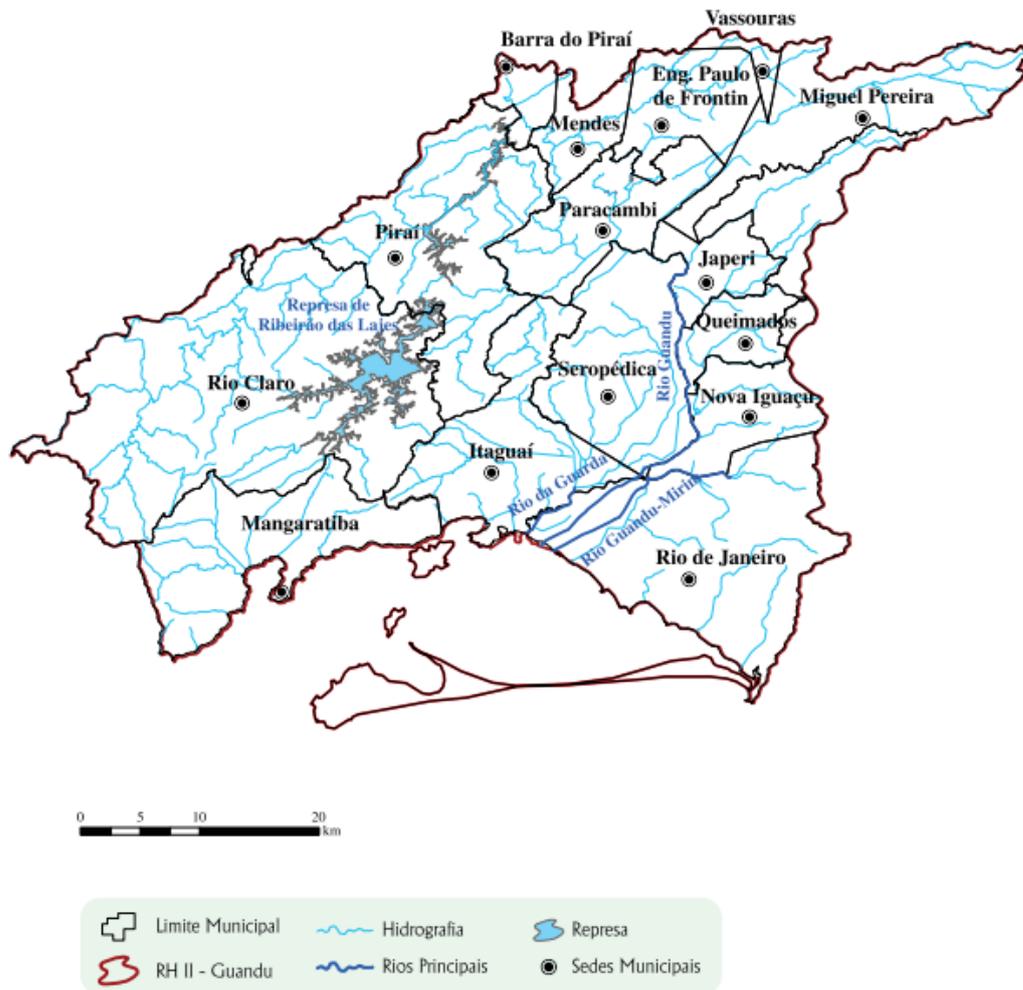
As projeções de abastecimento para a população urbana indicaram que as demandas hídricas totais do Brasil, no ano de 2025, correspondem a  $630 \text{ m}^3 \text{ s}^{-1}$  e  $695 \text{ m}^3 \text{ s}^{-1}$ , média e máxima respectivamente. Destaca-se as regiões Sudeste e Nordeste, seguidas pelas regiões Sul, Norte e Centro-Oeste (Tabela 1.6). As regiões Sudeste e Nordeste são detentoras de 71% da demanda de água projetada para o ano de 2025 e concentram 62% de todos os municípios do País. Estima-se que, entre os anos de 2005 e 2025, haverá um crescimento de 28% nas demandas médias para abastecimento da população urbana brasileira. [5]

A Bacia do Rio Paraíba do Sul (BRPS) fornece água a três estados da região sudeste do Brasil: Rio de Janeiro, São Paulo e Minas Gerais. Esta bacia hidrográfica é resultado da junção entre os rios Paraíba e Paraitinga e está localizada entre os paralelos  $20^{\circ}26'$  e  $23^{\circ}00'S$  e os meridianos  $41^{\circ}00'$  e  $46^{\circ}30'W$  (Figura 1.2). Os trechos/sub-bacias desta bacia hidrográfica conseguem suprir as necessidades atuais e futuras. Porém, algumas dessas cidades apresentam déficits de produção desproporcional a demanda, consequência da precariedade nas estruturas dos sistemas de abastecimento. [3].

Os sistemas de abastecimento da região Sudeste por muitas vezes são ineficientes devido ao grande número populacional de pessoas em uma área relativamente pequena do



O gerenciamento dos sistemas de abastecimento de água precisam de ajustes regulares por causa da alteração na demanda hídrica, com o propósito de diminuir os custos da distribuição [81]. Os sistemas de abastecimento e distribuição de água são alicerces para o consumo de água. A Estação de Tratamento de Água (ETA) do Guandu é responsável pelo abastecimento de 85% do município do Rio de Janeiro (Figura 1.3) e cerca de 70% dos municípios da Baixada Fluminense. Essa estação trata 43.000 litros de água por segundo, abastecendo 9 milhões de habitantes[23].



**Figura 1.3:** Mapa da bacia hidrográfica do Guandu, abrangendo 15 municípios do RJ. Fonte: Tubbs Filho et al. (2012)[36]

O estado do Rio de Janeiro apresenta um consumo de água com valor de 24,1% acima da média da região Sudeste e 52,2% acima da média do país. Esses valores elevados podem estar associados aos baixos índices de medição verificados no RJ, em que uma grande parte dos volumes consumidos são estimados[70].

O fornecimento e abastecimento de água sobre a região metropolitana do Rio de Janeiro deve atender também a qualidade da água recomendada. Uma variedade de subprodutos vem sendo depositada nos reservatórios naturais, resultado de atividades antrópicas [80]. Dentre esses elementos estão os metais pesados que permanecem na água

por mais tempo do que os poluentes orgânicos e se infiltram da superfície para as águas subterrâneas[58].

Um estudo realizado em duas regiões do RJ afirmou que as doenças causadas por água potável com altas concentrações de nitratos e certos metais como o alumínio tem aumentando a preocupação com relação os efeitos desses compostos na saúde[37].

Um levantamento sobre o consumo de água em  $\text{litros}^{-1}$  habitante $^{-1}$  dia $^{-1}$  foi elaborado pelo Sistema Nacional de Informações sobre Saneamento (SNIS) para os municípios do país. A pesquisa permitiu visualizar os dados de consumo de água da cidade do Rio de Janeiro e identificar o aumento progressivo desse consumo desde o ano de 2010 até 2013, com cerca de 329,8  $\text{litros}^{-1}$  habitante $^{-1}$  dia $^{-1}$  no ano de 2013[70].

O valor por pessoa de consumo de água potável para Rio de Janeiro é alto, devido ao grande número de ligações clandestinas de água, aos problemas de vazamentos nas linhas de abastecimento e ao desperdício da própria população no consumo diário, pois o metro cúbico de água de concessionária no Rio de Janeiro é mais barato, quando comparado ao estado de São Paulo[69].

A cidade do Rio de Janeiro possui cerca de 5,8 milhões de habitantes e consome cerca de  $704,077 \times 10^6$  ( $\text{m}^3$  ano $^{-1}$ ) de água[71]. Diante da crise hídrica ocorrida em 2014, a CEDAE promoveu uma campanha de conscientização do uso responsável da água que foi realizada em fevereiro de 2015 e obteve diminuição de 5 a 7% do consumo. Porém, após alguns meses, o consumo voltou ao mesmo número anterior a campanha.

A cidade do Rio de Janeiro tem uma alta demanda de água, então, é necessário uma solução para mitigar esse problema. Uma das soluções é o reúso e o aproveitamento de águas pluviais que podem ser utilizados para fins não potáveis nas cidades, tais como, limpeza e lavagem de ruas, ou mesmo em descargas de bacias sanitárias de banheiros públicos[11].

Com a intenção de identificar os hábitos do carioca com relação ao consumo de água, o laboratório de pesquisa do Centro Universitário Unicarioca, em 04 março de 2015, realizou uma enquete tendo como resultado que o carioca assume sua responsabilidade quanto ao desperdício de água e acredita que pode evitar o gasto desnecessário de água, quando adota-se algumas soluções como: reduzir o tempo de banho, fechar torneiras e registros [28]. O controle do consumo de água em residências tem uma influência significativa sobre a taxa de crescimento da demanda no futuro[53].

Para quantificação do consumo de água individual em uma residência é necessário ter conhecimento dos hábitos de uso da água pelos moradores. Existem fatores que influenciam o uso da água, como hábitos comportamentais, a exposição a restrições ou a escassez de água, a conscientização de economia da água, percepção de abundância de abastecimento ou de escassez de água e muitos outros fatores que podem não ser considerados[45].

Uma pesquisa sobre o levantamento do perfil do consumo residencial e uso final da água foi realizada em uma amostra de residências localizadas na zona oeste de São Paulo. Os resultados obtidos indicaram que a utilização do chuveiro tem maior relevância sobre

o consumo de água total nas residências. O uso da água destinado a higiene corporal (lavatório, caixa acoplada e chuveiro) ocorre todos os dias, assim como o consumo de água na pia da cozinha, enquanto que o consumo na máquina de lavar e no tanque ocorrem principalmente nos fins de semana. O maior dia de consumo é na sexta-feira. O consumo médio de água correspondeu a  $262,7 \text{ litros}^{-1} \text{ habitante}^{-1} \text{ dia}^{-1}$ . [11]

## 1.1 Considerações Preliminares

O desenvolvimento e crescimento da urbanização em grandes centros urbanos traz consigo uma grande demanda hídrica, e em alguns casos a crise hídrica acompanha esse crescimento. Em áreas urbanas, o fornecimento de água é alterado conforme a disponibilidade de água nas bacias hidrográficas, e este abastecimento depende muitas vezes das condições do ambiente. Em virtude disso, é necessário desenvolver um modelo de consumo de água para prever a quantidade de água fornecida pelos sistemas de abastecimento quando houver alteração nas condições atmosféricas.

As cidades com grande crescimento populacional já possuem uma disponibilidade hídrica insuficiente nas condições atuais [43]. Em situações de racionamento de água, a única solução não é o aumento da sua oferta, mas o desenvolvimento de campanhas permanentes de informação à população de uso consciente e apropriado da água disponível. [61]

No que se refere a disponibilidade de água para centros urbanos com grande atividade industrial, as indústrias presentes naquela região tendem a utilizar uma alta porção de água, para seu funcionamento e produção. A implantação do Complexo Petroquímico do Rio de Janeiro (COMPERJ) em Itaboraí e a construção do Arco Metropolitano ocasionarão no aumento da demanda de água. [4]

## 1.2 Objetivo

O objetivo principal da presente investigação é obter um melhor entendimento do consumo de água residencial em relação às condições meteorológicas de superfície observadas na RMRJ.

Para o entendimento do consumo de água potável residencial foram obtidos dados de uma residência e de um prédio residencial.

Além disso, as variáveis atmosféricas que serão correlacionadas ao consumo de água são: temperatura do ar, pressão atmosférica, umidade do ar e velocidade do vento.

Até onde se conhece da literatura sobre o consumo de água no RJ, não existem ainda trabalhos publicados considerando análises de Transformada de Ondeleta (TO) e de RNA aplicadas ao consumo de água potável residencial da população na RMRJ.

Para obtenção da análise do comportamento do consumo de água potável será feito a aplicação de TO.

Assim, propõe-se o desenvolvimento de um modelo de previsão da demanda de água, em função de variáveis meteorológicas com base no treinamento otimizado de uma RNA. Para isso faz-se necessário entender a correlação entre as variáveis de entrada e saída da rede. A otimização permite que a rede guarde a informação relacional entre as variáveis, de forma que a partir das variáveis de entrada possa modelar as variáveis de saída. No caso deste trabalho, a demanda de água potável da população é a única variável de saída.

A rede é treinada com a utilização de dados observados do consumo de água na RMRJ.

Este trabalho se insere no esforço feito para investigação das condições presentes da hidrometeorologia urbana das metrópoles tropicais, como o Rio de Janeiro no Brasil, aqui aborda-se o tema da disponibilidade de água potável para a população, seus níveis de consumo atuais e sua relação com as condições meteorológicas.



# Capítulo 2

## Materiais, conceitos e métodos

O uso da água de forma eficiente e sem desperdício é uma das tarefas mais importantes que a sociedade deve alcançar [61]. O aumento elevado da demanda está ultrapassando as previsões e a capacidade de oferta de água. Desta forma, é necessário desenvolver e aplicar modelos matemáticos para o planejamento de políticas otimizadas para uma gerenciamento adequado deste recurso [16]. Neste capítulo, apresenta-se a área de estudo, dados utilizados, programação, ondeletas e uma descrição de redes neurais.

### 2.1 Área de estudo

A Região Metropolitana do Rio de Janeiro (RMRJ) é composta por 18 municípios e concentra 75% da população do Estado do Rio de Janeiro. Quase a totalidade da demanda atual de abastecimento de água potável da RMRJ tem sido atendida por meio de três (3) sistemas produtores de água integrados: Acari, Guandu e Ribeirão das Lajes e por um sistema produtor independente Imunana/Laranjal. [4].

### 2.2 Dados utilizados

#### 2.2.1 Dados meteorológicos (formato METAR)

A fonte principal de dados meteorológicos empregada neste trabalho são dados medidos pela estação automática do Aeroporto Internacional Tom Jobim (Galeão), distribuídos em formato METAR pelo COMAER, Ministério da Aeronáutica do Brasil, com amostragem horária média, ao longo das 24 horas do dia.

Dados agrupados diários foram acessados na Internet via site [www.wunderground.com](http://www.wunderground.com) para a seguir serem baixados por um job (shell script) e empregar-se multiplamente o comando shell *wget*.

A extensão total destas observações (até o presente, 2016) totaliza quase 21 anos de dados completos, a partir de junho de 1996. O arquivo bruto concatenado contém cerca de 2/3 da informação necessária para o estabelecimento de uma climatologia (i.e, séries de 30 anos). [44]

### 2.2.2 Dados de consumo de água potável

Fatores como topografia local, clima e nível sócio econômico da população estão diretamente relacionados ao consumo médio per capita de água potável [69]. Em virtude disso, é necessário o entendimento do consumo de água potável residencial em uma metrópole como o Rio de Janeiro.

Para um determinado grupo de pessoas que habita uma cidade ocidental, como por exemplo, aquelas que moram em um prédio, em regime jurídico de condomínio, o consumo pode apresentar variações em função das horas do dia, i.e., apresentar uma variação horária. Em geral, a noite o consumo diminui a níveis muito baixos em comparação ao consumo durante o período diurno, isto devido a diminuição de atividades [11]. O consumo também pode variar de acordo com a época do ano, o que denomina-se variação sazonal, principalmente em cidades turísticas. [33]

Os dados usados neste trabalho provêm de duas fontes distintas: a primeira, de uma residência na Baixada Fluminense (Zona Oeste), e a segunda, obtida com um funcionário de um prédio de apartamentos em condomínio do consumo de água em Laranjeiras (Zona Sul) do Rio de Janeiro.

#### Dados de consumo de uma residência em São João de Meriti (Baixada Fluminense)

Os registros obtidos da residência são mensais e compreendem o período entre novembro de 2009 a maio de 2015. Um hidrômetro instalado na residência faz o registro do consumo de água e conforme a quantidade usada é feito um cálculo pela CEDAE para o valor que será cobrado na conta de água mensal [22]. A residência é composta por 4 pessoas e está localizada no bairro Parque São Nicolau, município de São João de Meriti, baixada fluminense. Os dados de localização estão descritos a seguir.

Longitude =  $43^{\circ} 23' 08''$  W =  $-(43+(23+08/60)/60) = -43,385555556$   
 graus

Latitude =  $22^{\circ} 48' 38''$  S =  $-(22+(48+38/60)/60) = -22,810555556$  graus

Altitude = 15 metros

Número de pessoas = 4

#### Dados de consumo de um edifício elevado residencial na Zona Sul do Rio de Janeiro

Para caracterização do consumo de água da zona sul da cidade do RJ foram utilizados dados fornecidos de um prédio localizado no bairro Laranjeiras, composto por 56 apartamentos. A coleta de dados de consumo diário é realizada todos os dias pelo funcionário que está trabalhando no horário da manhã.

Como em todo prédio residencial, a água é o segundo item de maior peso orçamentário nas cotas condominiais mensais [69]. O registro de dados do consumo diário possibilitou ao prédio, identificar quaisquer problemas de vazamento e trata-los imediatamente devido o controle diário de consumo de água do prédio por meio do hidrômetro. O período de

dados tem início em 3 de fevereiro de 2011 e término em 12 de dezembro de 2015. Abaixo temos os dados de localização desse prédio.

Longitude =  $43^\circ 10' 56''$  W =  $-(43+(10+56/60)/60) = -43,182222222$   
 graus  
 Latitude =  $22^\circ 56' 10''$  S =  $-(22+(56+10/60)/60) = -22,936111111$  graus  
 Altitude = 30 metros  
 Número de apartamentos = 56  
 Número de pessoas estimado (2 por apartamento) =  $2 * 56 = 112$

### 2.2.3 Tratamento dos dados faltantes

Para análises estatísticas consistentes, as falhas da série temporal devem ser consideradas [78].

Em geral, os momentos estatísticos de primeira e segunda ordem da distribuição devem ser conservados em qualquer método para tratamento de falhas, o que é particularmente importante para completar falhas. Neste trabalho, as falhas de variáveis individuais não são preenchidas, mas eliminadas, aonde ocorrem, de forma que somente as linhas de dados completas para todas as variáveis foram preservadas. Deste procedimento, obtendo-se assim os denominados *dados ajustados*.

Após parear os dados (em um programa fortran f90, chamado *reading\_dataset*), os valores foram verificados quanto à sua ocorrência dentro de um intervalo característico apropriado para cada variável meteorológica, por exemplo,  $\min(f_i) \leq f_i \leq \max(f_i)$ , sendo  $f_i$  a variável e  $\min(f_i)$  e  $\max(f_i)$  seus limites aceitáveis, mínimo e máximo, respectivamente. Em geral, estes limites de observação são escolhidos para representar mais ou menos 3 desvios-padrão das variáveis observadas. [25]

Ao final, uma interpolação de Barnes 2-D é aplicada ao longo da série para todas as variáveis meteorológicas [10]. O consumo de água potável também foi interpolado hora a hora, de forma a gerar um pareamento de todos os dados hora a hora.

O objetivo da interpolação foi também fazer o tratamento de dados duplicados ou quase duplicados presentes nas mensagens em codificação METAR (i.e., com horários de observação quase coincidentes ou coincidentes, correspondentes às mudança de tempo que ocorrem em escala de tempo sub-horária mas que devem ser registradas em mensagens de tempo de aeroportos por uma questão de segurança aeroportuária).

Nas interpolações 2-D, foi escolhida uma janela de observação de 3 horas, então aplicada na obtenção dos valores horários interpolados ao longo de toda a série. Desta forma, a série de variáveis filtradas e ajustadas foi gerada de forma consistente com a informação original.

A interpolação 2-D pode ser escrita como [10],

$$g_i(x, y) = \frac{\sum_j^M \eta(r_{ij}) f_j}{\sum_j^M \eta(r_{ij})} \quad (2.1)$$

onde  $\eta$  é o fator de peso,

$$\eta_{ij} = \exp\left(\frac{-r_{ij}^2}{4\kappa}\right) \quad (2.2)$$

sendo  $r$  a distância no sistema de coordenadas polar.

O erro de truncamento ( $\epsilon$ ) pode ser expresso

$$\epsilon = \exp\left(\frac{-R^2}{4\kappa}\right) \quad (2.3)$$

em função do raio de influência ( $R$ ). Somente observações dentro da vizinhança  $r < R$  são consideradas na somatória, o que implica no erro de truncamento  $\epsilon$ . Quando  $R \rightarrow \infty$  resulta que  $\eta$  será um função de distribuição, portanto com integral 1, e o erro da interpolação nulo, i.e.,  $\epsilon = 0$ .

$\kappa$  é o parâmetro de forma do filtro, associado à densidade das observações e ao comprimento-de-onda recuperado escrito em relação ao espaçamento de grade.

Uma generalização da interpolação de Barnes foi desenvolvida para a concepção de propriedades de filtragem específicas de um passa-banda. [29]

Filtragens temporal e espacial foram realizadas simultaneamente e obteve-se excelente continuidade temporal (i.e., uma meta em análise de cartas sinóticas), assim como obteve-se padrões espaciais suavizados ao máximo sem que informações em determinada banda espectral fossem filtradas. Portanto, uma solução de compromisso ótima do ponto de vista estatístico. [29]

Um esquema de análise objetiva foi desenvolvido e fundamentado na interpolação de Barnes. Propuseram um aplicativo auxiliar a análise e previsão de alta resolução (i.e., *nowcasting*), projetado para uso em computador interativo (*workstation*). Destacaram as vantagens do interpolador de Barnes como a possibilidade de implementação modular, de criação de interatividade com o usuário (via monitor do sistema), o que resultou na *análise objetiva* ou *estatística*. [47]

Entre as características da análise objetiva tem-se: possibilidade de sua aplicação a dados distribuídos irregularmente no espaço e tempo, interpolação de uma grade irregular para uma grade regular sobre o domínio de análise (previsão), controle interativo via possibilidade de rodar novamente com diferentes parâmetros de interpolação, consideradas as possibilidades funcionais de um monitor de processo amigável (aplicativo com botões iterativos), assim obtendo-se controle da taxa de convergência e controle dos parâmetros de filtragem do passa-banda. Este tipo de controle do passa-banda pode ser especialmente útil ao previsor de *nowcasting* para separação dos escoamentos de mesoescala e de escala sinótica.

## 2.3 Ondeletas

A ondeleta é uma função matemática usada para dividir uma determinada função ou sinal de tempo contínuo em componentes de escala diferentes. Normalmente, pode-se atribuir uma faixa de frequência de cada escala componente. Cada escala componente pode, então, ser estudada com uma resolução que corresponda à sua escala. [60]

A ondeleta tem como função caracterizar o sinal das variáveis. Ela consiste em uma oscilação de onda que tem início em zero, aumenta e em seguida diminui de volta para zero, definindo-a como uma breve ou pequena oscilação. Todas as ondeletas podem ser consideradas formas de representação em tempo e frequência e são ferramentas poderosas de processamento de sinais, devido a capacidade de decomposição das funções. [60]

Por meio da ondeleta pode-se decompor uma função, este processo é conhecido como Transformada de Ondeleta (TO) e possui variantes que se dividem em contínuos e discretos. Estas funções são bastante aplicadas na compressão de dados, eliminação de ruído, separação de componentes no sinal, identificação de singularidades, detecção de auto-similaridade, entre outras. [77]

Inicialmente, a TO foi utilizada para análise de sinais sísmicos. Esta foi desenvolvida na década de 1980 por pesquisadores como Morlet, Grossman, Meyer e Daubechies [35]. Em 1994, Kumar e Foufoula-Georgiou, escreveram que as TOs já estavam sendo exploradas para análise de diversos processos geofísicos, como a turbulência atmosférica, precipitação em tempo-espaço, ondas de ventos do oceano, batimetria do fundo do mar, estruturas de camadas geológicas, mudanças climáticas, dentre outros. [49]

A TO foi elaborada para o estudo de sinal não-estacionário. Em oposição, a transformada de Fourier foi aplicada para a análise de sinais estacionários, devido a natureza de preenchimento de espaço das funções trigonométricas. Enquanto, a transformada de Fourier concede a análise global do sinal e a sua reconstrução total, a TO permite estudar o sinal tanto localmente quanto globalmente, assim como a reconstrução de partes deste sinal. [35]

Diversos estudos foram desenvolvidos e estes empregaram a TO por ser um método com muitas vantagens. Uma implementação do “*algorithm à trous*” foi usado para calcular a TO, empregando sons musicais com duração de poucos segundos. [31]

A TO foi aplicada para o estudo da turbulência [35]. A TO foi utilizada para análise das frentes frias na camada limite e concluiu-se que a ondeleta fornece uma representação superior de fenômenos frontais quando comparados com métodos de transformada global. [38]

Um estudo sobre a organização da convecção sobre o Pacífico tropical foi realizado e indicou a TO como uma ferramenta poderosa para a análise de fenômenos que envolvem interações multiescalares que exibem localização em frequência e tempo. [76]

Uma análise de ondeleta é realizada, com exemplos tirados de séries temporais do fenômeno El Niño- Oscilação Sul (ENOS). Mostra-se que o nivelamento no tempo ou es-

cala pode ser usada para aumentar a confiança do espectro de ondeleta e que fórmulas empíricas são indicados para o efeito de suavização em níveis de significância e intervalos de confiança. [73]

A variabilidade climática na África foi estudada por meio de uma análise da TO aplicada ao registro de chuva e concluíram que fenômenos climáticos extremos associados remotamente ocorrem de forma generalizada sobre o continente. [46]

Uma análise das variabilidades intrasazonais sobre a América do Sul e oceanos adjacentes foi desenvolvida e aplicou-se o método da TO unidimensional para detectar as oscilações intrasazonais e verificar as escalas de tempo associadas ao sinal atmosférico mais energético. [75]

Verificou-se que o ciclo anual é o responsável pela maior parte da variabilidade climática no extremo sul do Brasil juntamente com as massas de ar e sistemas frontais. Ressaltou-se que a associação dos dados na TO com os eventos atmosféricos ocorreu devido a grande eficiência desta metodologia em decompor os sinais em domínios de tempo e escala, i.e., indicando a ferramenta como um discriminador eficiente dos sinais meteorológicos no tempo. [61]

O termo ondeleta corresponde a um conjunto de ondas formadas por dilatação [ $\psi(t) \rightarrow \psi(2t)$ ] e translação [ $\psi(t) \rightarrow \psi(t + 1)$ ] de uma única função  $\psi(t)$  que é quadraticamente integrável sobre o campo dos reais ou espaço [ $L^2(R)$ ] e possui energia finita. A função  $\psi(t)$  pode ser chamada de “ondeleta mãe”, “ondeleta básica” ou “ondeleta analisadora” enquanto que as funções dilatadas e transladadas derivadas da ondeleta mãe são chamadas simplesmente “ondeletas”. [76]

## Ondeletas contínuas e discretas

As ondas que formam uma ondeleta contínua transformada estão sujeitas ao princípio da incerteza da teoria respectiva de análise da amostragem de Fourier, atesta-se que dado um sinal com algum evento em que, não se pode atribuir um tempo exato e escala de resposta de frequência simultaneamente para esse evento.

As ondeletas discretas são utilizadas para a decomposição e filtragem de qualquer série-temporal. A sua aplicabilidade neste aspecto advém do fato de que estas ondeletas não provocam retundâncias de coeficientes entre escalas (frequências). [14]

A transformada de ondeleta contínua de um sinal  $f(t)$  é definida como [27] [76]:

$$W_{\ell,t'}(f) = \langle f, \psi_{\ell,t'}^* \rangle = \int f(t) \psi_{\ell,t'}^*(t) dt \quad (2.4)$$

onde  $\ell$  é o parâmetro de dilatação,  $t'$  é o parâmetro de translação e  $\psi^*$  é o complexo conjugado das ondeletas  $\psi_{\ell,t'}$ ,

$$\psi_{\ell,t'}(t) = \frac{1}{\sqrt{\ell}} \psi \left( \frac{t-t'}{\ell} \right) \quad (2.5)$$

onde  $\psi(t)$  é a ondeleta mãe,  $\ell$  é o parâmetro de dilatação (escala) e  $t'$  é o parâmetro de translação.

O cálculo da transformada inversa é determinado como [52][50]:

$$f(t) = \frac{1}{C_\psi} \int \int W_{\ell,t'}(f) \psi_{\ell,t'}(t) dt' \frac{d\ell}{\ell^2} \quad (2.6)$$

onde  $C_\psi$  aparece como um fator de normalização e é descrito como [24]:

$$C_\psi = 2\pi \int_0^\infty \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty \quad (2.7)$$

onde  $\hat{\psi}$  é a transformada de Fourier de  $\psi(t)$ .

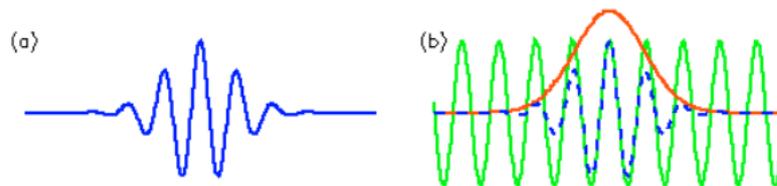
Para uma função  $\psi(t)$  se torna uma ondeleta mãe ela deve ser submetida a algumas condições [35]:

Admissibilidade relata que para uma função integrável, a sua média é nula.

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (2.8)$$

Para escolher a ondeleta mãe adequada devemos levar em consideração características semelhantes ao sinal que será estudado, como a assimetria e variação brusca ou suave no tempo [24].

A ondeleta mãe de Morlet, representada na Figura 2.1, é muito empregada em meteorologia pois é mais apropriada para sinais em forma de onda. Essa ondeleta consegue capturar variações nas frequências dos sinais geofísicos. [66]



**Figura 2.1:** Ondeleta complexa de Morlet, a) ondeleta de Morlet com largura e amplitude arbitrária e com o tempo ao longo do eixo horizontal e b) construção da ondeleta de Morlet (tracejado) a partir de uma onda seno modulada por uma curva gaussiana. Fonte: Torrence e Compo (1998)

A ondeleta de Morlet é descrita por

$$\psi(t) = \pi^{-1/4} e^{-\frac{t^2}{2}} e^{i\omega_0 t} \quad (2.9)$$

Veja a equação 2.9, sendo  $\omega_0$ , o parâmetro de Morlet. Este deve ser escolhido para satisfazer a condição de admissibilidade. O  $\omega_0$  corresponde a 5,4. [76]

## 2.4 Redes Neurais Artificiais

As chamadas Redes Neurais Artificiais ou RNAs são estruturas topológicas computacionais que foram relacionadas inicialmente ao funcionamento celular de neurônios acoplados. Estas redes são capazes armazenar um conjunto grande de informações por meio de um método de aprendizado (e.g., *backward propagation of information under constraints*), condicionado por minimização de métricas medias: MSE ou o que deve resultar no mesmo, a maximização de integrais: verossemelhança, *likelihood*.

A dinâmica e estrutura de cada rede de aprendizado está associada a uma função. A origem de sua teoria remonta aos modelos matemáticos aplicados como também a teoria de funcionamento de neurônios biológicos como redes topológicas. [48]

Em geral, uma interpretação estocástica é considerada aos sistemas não-determinísticos ou não-lineares. Neste caso, os coeficientes aplicados nas equações das séries temporais de entrada (i.e., das variáveis significativas) precisam ser determinados por otimização.

As RNAs foram introduzidas e apresentaram os fundamentos conceituais, princípios operacionais e métodos de modelagem. Com a finalidade de obter um tratamento do problema de aprendizado de máquina, em busca de solução de problemas complexos, deve-se aplicar métodos estatísticos e computacionais, basicamente. No caso de aplicações, o processo requer interpretações físicas dos resultados. [48]

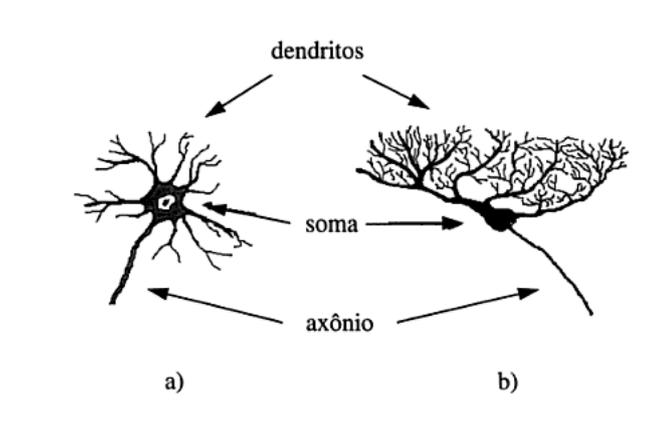
### O neurônio biológico

O neurônio, dita célula nervosa, incluído aqui por questões históricas, foi identificado anatomicamente e descrito com notável particularidade, pelo neurologista espanhol Cajal (no século XIX). [18]

O neurônio como qualquer célula biológica é composto por uma fina membrana celular ue além de desempenhar a função biológica, possui propriedades que são primordiais para o funcionamento da célula nervosa. [48]

O corpo celular, ou soma, representa o centro dos processos metabólicos da célula nervosa e a partir desse corpo desenvolvem-se os dendritos e o axônio que são extensões filamentosas. Os dendritos cobrem um volume muitas vezes maior do que o próprio corpo celular e formam uma árvore dendrital (Figura 2.2). [48]

A semelhança topológica de vários neurônios interconectados e de algumas redes relacionais de variáveis levou ao nome RNA. Entretanto, a semelhança deve é aparente, desde que a rede se refere a otimização de parâmetros associada a uma distribuição de probabilidade. Deste ponto de vista a rede é a forma topológica da solução otimizada de um problema de *distribuição posterior*, obtida a partir de uma *distribuição prior*, disponível a priori, que em geral é mal- informativa, portanto, de um problema *bayesiano*.



**Figura 2.2:** Neurônios do sistema nervoso central dos vertebrados: a) Neurônio motor da medula espinhal; b) Célula de Ourkinje encontrada no cerebelo. Esta célula é notável pela extensa ramificação da sua árvore dendritica da qual apenas uma pequena parte é mostrada na figura, podendo ultrapassar várias dezenas de vezes as dimensões do corpo celular. Fonte: Kovács (2006)

## Modelos de RNA

É possível realizar uma distinção entre modelos de RNA e suas características básicas, e.g., o tipo de entrada, a forma de conexão e o tipo de aprendizado. [20]

O tipo de entrada pode ser binária ou intervalar. Os modelos chamados binários aceitam entradas discretas, i.e., somente na forma de 0 e 1, inclusive nesse tipo de modelo temos os modelos de Hopfield e ART. Os modelos intervalares aceitam qualquer valor numérico como entrada (forma contínua) e temos como exemplo, o *backpropagation* e *Kohonen*.

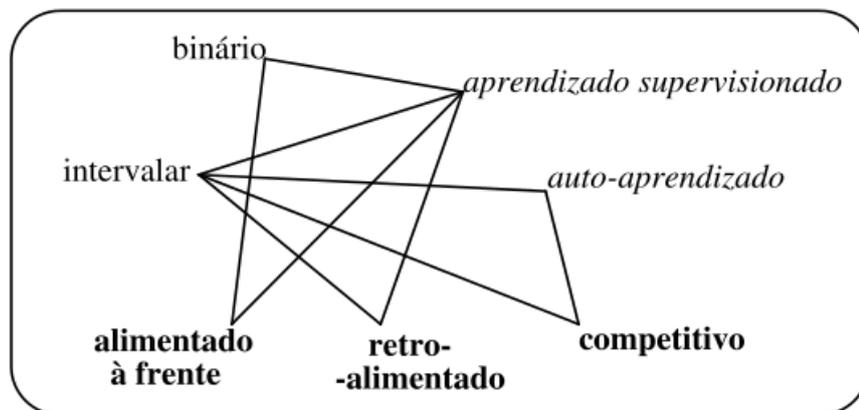
Por meio da forma de conexão temos como definir a maneira que os elementos da rede estão conectados e como o modelo matemático representa a transmissão dos sinais da rede. Com isso, temos três formas de conexão: alimentação à frente, retro-alimentação e competitiva. [20]

A forma de conexão por alimentação à frente consiste em que os sinais de entrada são simplesmente transformados em sinais de saída. A retro-alimentação indica que os sinais são alterados em diversas transições de estado, em que a saída também é alimentadora da entrada. Além das formas de conexão citadas anteriormente, temos a conexão competitiva que realiza a interação lateral dos sinais recebidos na entrada entre os elementos dentro de uma zona de vizinhança.

O tipo de aprendizado (treinamento) está ligado à existência ou não de um sinal de saída pré- definido para a rede. Quando há uma definição sobre qual saída se deseja para

a rede, chamamos de aprendizado supervisionado, i.e., o ajuste de pesos é forçado de modo a representar um sinal desejado. Em contrapartida, há o auto-aprendizado (não-supervisionado) que limita-se a se fazer uma representação da distribuição de probabilidade dos padrões de entrada na rede e este tipo de treinamento está ligado com a conexão competitiva.

Por meio da Figura 2.3, ilustram-se estes modelos e as classificações que foram citadas anteriormente. Essa figura mostra um esquema de combinação das classificações para compreensão das características dos modelos mais comumente usados atualmente.



**Figura 2.3:** Modelo de RNA e suas classificações. Fonte: Cardon et al. (1994)

As particularidades dos respectivos modelos de RNA [20] serão vistos a seguir:

- Perceptron/Adaline: Possui entrada intervalar, aprendizado supervisionado e conexão de alimentação à frente. Sua grande vantagem é a simplicidade na implementação desse algoritmo. Além disso, são necessários poucos parâmetros para ajustá-lo e o padrão de entrada não precisa de um processamento muito elaborado, dependendo da aplicação. Entretanto, esse algoritmo tem sua aplicação restrita a padrões não muito complexos que sejam linearmente separáveis.

As aplicações do perceptron/adaline está restrita a quantidade de padrões envolvidos e sua complexidade.

- *Backpropagation*: Esse modelo é derivado do perceptron/adaline obtendo assim as mesmas características, i.e., entrada intervalar, aprendizado supervisionado e conexão de alimentação à frente. A capacidade de abranger as mais variadas classificações de padrões é a grande vantagem deste modelo.

Algumas desvantagens também se aplicam a esse modelo como o tempo de treinamento que é bastante longo para a maioria das aplicações de uso prático (e.g. reconhecimento de caracteres). Além disso, há certos tipos de padrões envolvidos em séries temporais que precisam de uma grande necessidade de pré-processamento,

devido a complexidade da análise e classificação dos padrões em questão.

O *Backpropagation* é utilizado para reconhecimento de caracteres, composição musical, séries temporais, reconhecimento de sinais em geral, treinamento de robôs, etc.

- *Hopfield*: É caracterizado por entrada binária, aprendizado supervisionado e conexão de retro-alimentação. Essa rede tem como vantagem adequar-se quando deseja-se utilizar uma memória associativa, i.e., uma memória que armazena diversos padrões ao mesmo tempo. Com isso, cada um desses padrões pode ser referenciado dependendo de qual se assemelha ao padrão de entrada.

Em contrapartida, há duas desvantagens (restrições) nesse modelo. A primeira propõe que o número de padrões a ser ensinado deverá corresponder a no máximo 15% do total de nodos da rede e a segunda restrição é não permitir que um modelo exemplar compartilhe muitos bits com outro padrão, pois isto pode confundir a rede.

O conversor analógico-digital é uma das muitas aplicações desse modelo.

- *Kohonen*: Suas características são entrada intervalar, auto-aprendizado e conexão competitiva. Esse modelo possui vantagens nas aplicações em que os padrões tem alguma relação entre si, podendo desta forma, serem classificados. Entretanto, essa rede pode ser um pouco complexa em relação aos outros modelos, pois as variáveis como raio da vizinhança e coeficiente de aprendizado devem ser ajustados adequadamente, para que o aprendizado seja bem sucedido.

Como exemplos de aplicações pode ser citado a máquina de escrever fonética ou *Phonetic Typewriter*.

- ART (*Adaptative Resonance Theory*): A teoria da ressonância adaptativa resultou num classificador projetado por Carpenter e Grossberg e suas características são entrada binária, aprendizado supervisionado e conexão retro-alimentadora.

O modelo ART tem como vantagem a facilidade de não precisar de um aprendizado prévio, adaptando-se de acordo com a necessidade. Por outro lado, este modelo é muito sensível a padrões distorcidos não podendo fazer uma boa diferenciação entre padrões semelhantes, limitando assim suas aplicações.

Utiliza-se o ART para reconhecimento de imagens e mais comum para reconhecimento de caracteres.

## Normalização

### Escalonamento

As RNAs aceitam normalmente valores entre 0/(-1) e 1. Se os dados estão uniformemente distribuídos é considerada linear e quando seu comportamento é logarítmico denomina-se não linear.

### Normalização de vetores

Para normalização de uma série de escalares (na forma de elementos de vetores), subtrai-se o valor médio ( $\bar{x}$ ) e divide-se pelo desvio padrão  $\sigma_x$

$$Z = \frac{x - \bar{x}}{\sigma_x} \quad (2.10)$$

### Equações

A Figura 2.4 mostra o esquema de uma RNA contendo sinais de entrada ( $x_i+1$ ), uma camada intermediária (camada interna) representada por três neurônios e na saída há um neurônio atuando como função de ativação.

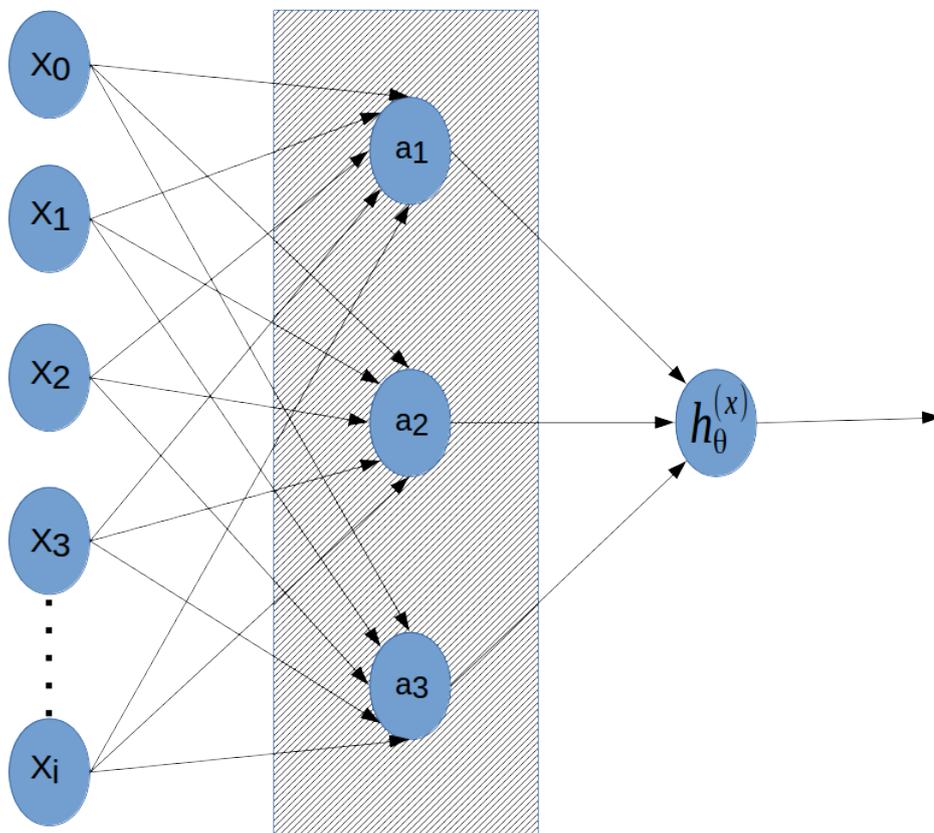


Figura 2.4: Modelo de uma rede neural artificial

na qual

$$\begin{cases} a_1 = \theta_0^{(1)}x_0 + \theta_1^{(1)}x_1 + \theta_2^{(1)}x_2 + \theta_3^{(1)}x_3 + \dots + \theta_i^{(1)}x_i \\ a_2 = \theta_0^{(2)}x_0 + \theta_1^{(2)}x_1 + \theta_2^{(2)}x_2 + \theta_3^{(2)}x_3 + \dots + \theta_i^{(2)}x_i \\ a_3 = \theta_0^{(3)}x_0 + \theta_1^{(3)}x_1 + \theta_2^{(3)}x_2 + \theta_3^{(3)}x_3 + \dots + \theta_i^{(1)}x_i \end{cases} \quad (2.11)$$

Temos o modelo logístico representado pela equação:

$$g(Z) = \frac{1}{1 + e^{-Z}} \quad (2.12)$$

Trabalhando este modelo em nossa RNA teremos

$$a_1 = g(x^T\theta^{(1)}); a_2 = g(x^T\theta^{(2)}); a_3 = g(x^T\theta^{(3)}) \quad (2.13)$$

Temos hipóteses de saída, conhecida como “output” descrito como:

$$h_\theta(x) = g(\vec{a}^T\theta^{(j)}) \text{ sendo, } \vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

sendo  $\theta$  usado para denotar os parâmetros do modelo,  $\theta^{(1)}, \dots, \theta^{(j)}$ .

Temos na 2.4, os parâmetros do modelo representados por  $(\mathbf{n}+1)\mathbf{m}$ , onde  $\mathbf{n}$  são os números de entrada,  $\mathbf{1}$  representa o número de saídas e  $\mathbf{m}$  são o número de neurônios internos.

A função custo é representada por

$$J(\theta) = \frac{1}{2} \sum_{i=1} (g^{(i)} - h_\theta(x^{(i)}))^2$$

## Aplicações de RNA

Um sistema de previsão a curtíssimo prazo foi desenvolvido usando técnicas de RNA para o Aeroporto Internacional Tom Jobim-RJ, dados meteorológicos (de vários anos) das estações de superfície e altitude do DECEA-Galeão, junto com dados acumulados horários de precipitação e registro de raios da Ilha do Governador e proximidades. A partir do grupo de dados pré-processados escolheram duas séries, uma para treinamento e outra para validar o processamento feito pela da RNA. Foram usados os registros de raios para melhor desempenho da rede e obteve-se médias de acertos e alarmes falsos do sistema de previsão para as três horas, tendo as respectivas respostas: a) 88% e 15% para primeira hora, b) 87% e 14% para segunda hora, c) 85% e 19% para terceira hora. Conclui-se que a RNA é capaz de identificar o perfil termodinâmico da atmosfera e a aplicação da mesma no sistema de previsão de curtíssimo prazo é muito importante no ramo da aviação. [65]

Um sistema de RNA foi usado para realizar um estudo de previsão da demanda de água na Região Metropolitana de São Paulo (RMSP) por meio do consumo de água (estação Cantareira) e variáveis meteorológicas e socioambiental, para o ano de 2005. Foram desenvolvidos oito configurações de modelos, sendo que o melhor desempenho foi obtido para a média de 12 horas nas variáveis de entrada. Concluíram que uma rede neural de alimentação típica de três camadas foi mais qualitativa para este estudo e o método proposto é uma ferramenta útil para a previsão e gerenciamento de sistemas de água. [68]

Um modelo de previsão de demandas de água permitiu a utilização de modelos de otimização da operação relacionado com modelos de simulação do fluxo do sistema. Esse modelo foi usado em alguns reservatórios do Sistema Adutor Metropolitano de São Paulo e obteve como resultados a precisão e colaboração no auxílio ao modelo de otimização, a melhoria no índice de atendimento à demanda e a redução do número de manobras fundamentais. [81]

Uma comparação de regressão multivariada e RNA foi realizada para prever a demanda de água a curto-prazo (i.e., semanal), principalmente os picos da demanda. Foram analisados cerca de 6 anos de dados de picos de demanda de água semanal e 2 variáveis meteorológicas em duas regiões diferentes da cidade de Nicosia, Chipre. Os modelos trabalhados foram de regressão linear múltipla e RNA usando três tipos de algoritmo. O método de RNA que empregou o algoritmo Levenberg-Marquardt forneceu uma previsão mais exata do pico da demanda de água semanal para as duas regiões e essa demanda foi melhor correlacionada com a ocorrência de precipitação. [1]

Uma modelagem e previsão da demanda do pico diário de água durante o verão aplicando redes neurais artificiais (RNA) foi realizada para a cidade de Ottawa, Canadá. Comparações entre a regressão linear, análise de séries temporais e RNA foram realizadas para modelar a previsão dessa demanda de água. Cerca de 10 anos de dados da demanda de água foram empregados juntamente com 2 variáveis meteorológicas. Os resultados obtidos apontaram que a RNA proporciona uma previsão melhor do pico de demanda diário no verão e os melhores resultados ocorrem quando a demanda atinge o pico do dia anterior. A temperatura máxima do dia atual e anterior e a ocorrência/não ocorrência de precipitação de cinco dias antes, foram usados como dados de entrada. O pico da demanda de água diária é melhor correlacionada com a ocorrência de chuva. [2]

### 2.4.1 Índices estatísticos da RNA

Há diversas técnicas de validação para RNA [15]. As mais comuns são:

- **Teste:** Uma sequência de testes independentes de  $D_N$  e distribuídas de acordo com a mesma distribuição de probabilidade é usada para avaliar a qualidade. Na prática, infelizmente, um conjunto adicional de observações de entrada/saída é raramente disponível.
- **Validação:** O método de validação, as vezes chamado de estimativa da amostra de teste, particiona os dados  $D_N$  em dois conjuntos mutuamente exclusivos, o conjunto de treinamento  $D_{N_{tr}}$  e a validação ou o conjunto de teste  $D_{N_{ts}}$  onde  $N = N_{tr} + N_{ts}$ .

- **Validação-cruzada de partição-K:** O conjunto  $D_N$  é dividido aleatoriamente em  $K$  partições de teste mutuamente exclusivas de tamanho aproximadamente igual. Os casos não encontrados, em cada partição de teste, são usados de forma independente para selecionar a hipótese que será testada na própria partição. O erro médio sobre todas as partições  $k$  é a taxa de erro da validação-cruzada.

Para verificação do desempenho da RNA do presente trabalho aplicaremos a validação-cruzada de partição-K [15]. O algoritmo é detalhado a seguir:

1. Dividir o conjunto de dados  $D_N$  em  $K$  partes de tamanhos iguais (i.e.,  $K$  subamostras).
2. Para a  $k$ -ésima parte, sendo  $k = 1, \dots, K$ , ajustar o modelo para outras partes  $K - 1$  partes do dado. A seguir, estimar a incerteza da previsão do modelo ajustado para cada uma das  $K$  subamostras.
3. Repetir os passos acima para as  $K - 1$  subamostras, e então, fazer a estimativa da incerteza média usando as  $K - 1$  estimativas da incerteza.

Desta forma obtém-se uma estimativa de validação-cruzada do erro quadrático médio (MSE) para aplicação na etapa de previsão (i.e., de validação).

$$\overline{MSE}_{CV} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i^{-k(i)})^2 \quad (2.15)$$

sendo  $\hat{y}_i^{-k(i)}$  o valor ajustado para outra  $i$ -ésima observação devolvida pelo modelo estimado com uma  $k$ -ésima parte dos dados removida.

O método acima é ilustrado na Figura 2.5. Nela temos a validação cruzada com 10 partições, i.e.,  $K = 10$ . Em cada interação 90% dos dados são usados para treinamento e o remanescente, 10%, para o teste.

O método da validação cruzada estabelece cada subamostra seja aplicada ao procedimento *leave-one-out* (i.e., deixa uma parte da amostra fora do treinamento para teste). Neste caso, o algoritmo durante o  $n$ -ésimo teste da série de  $K = 10$  subgrupos, separa a amostra na porção de treinamento (i.e., 9/10 da amostra) e na porção de teste (i.e., 1/10), escolhendo diferentes partes da amostra sem repetição para teste (Figura 2.5) para compreensão desta separação, exemplificada para os três primeiros testes da validação cruzada.

1. Realizou-se a identificação paramétrica e deixou-se aquela observação para fora do conjunto de treinamento,
2. Calculou-se o valor previsto para a  $i$ ésima observação, denotado por  $\hat{y}_i^{-i}$ .

A estimativa correspondente do erro de previsão MSE é

$$\overline{MSE}_{loo} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i^{-i})^2 \quad (2.16)$$

Em geral, a estimativa do ( $MSE$ ) associado ao subconjunto de amostras de validação das  $K$ -etapas é muito pequeno, estando próximo do limite de precisão em ponto-flutuante da máquina. Isto ocorre porque a máquina de aprendizado (rede) constitui-se como um sistema de variáveis interligadas com coeficientes ajustados pelo *método dos mínimos-quadrados*, como no caso deste trabalho, implicando portanto em uma estimativa do erro sistemático global muito próximo de zero.

$$\overline{MSE}_{CV} = \delta \quad (2.17)$$

sendo  $\delta$  da ordem do erro de truncamento da otimização dos coeficientes da rede. O que está por trás disto é a ideia de que quando se faz a aplicação sobre uma mesma amostra para treinar e para testar, o erro sistemático obtido na validação não será um bom estimador do erro real associado ao desempenho da rede em uma amostra independente.

Por isto o interesse em métodos como da validação-cruzada, por sua possibilidade de avaliação do  $MSE$  real durante cada um dos testes independentemente. Naturalmente, outras possibilidades existem, como o método de reamostragens *bootstrap* e outros métodos, como modelagem estatística pelo método de Monte-Carlo.

O valor pequeno,  $\delta$ , da incerteza do treinamento, entretanto, continua a ser uma estimativa razoável do ajuste possível da rede, mas que não é apropriada para avaliação da incerteza associada à aplicação da rede (modelo) a uma amostra diferente. Para avaliar a incerteza associada às amostras independentes do treinamento utiliza-se as estimativas denominadas *leave-one-out* (i.e., *deixa-uma-parte-fora*). [15]

O erro quadrático médio obtido de cada um dos  $K$ -testes da validação-cruzada, são mediados aritmeticamente para obter o valor  $MSE_{loo}$ . Em geral,

$$\overline{MSE}_{loo} < \overline{MSE}_{CV} \quad (2.18)$$

Uma aplicação importante da validação-cruzada, além da estimativa do  $MSE$  real, é a possibilidade de avaliar o número ótimo de camadas de nós da camada intermediária da rede, que deve ser superior a um mínimo de nós para que a rede tenha capacidade de armazenar a quantidade de informação associativa entre as variáveis. Mais adiante na etapa de aplicação da rede neste trabalho, após testes optou-se por 40 nós na camada intermediária da rede.

Após comparar o modelo de previsão com a observação é necessário avaliar o grau de confiança do modelo. Para que seja possível a avaliação do intervalo de confiança é preciso conhecer-se a distribuição dos erros do modelo em relação à observação. Isto não é conhecido a priori. Para fechar o problema em geral o intervalo de confiança é estimado.

A estimativa do intervalo de confiança é feita comumente a partir da suposição de que o erro do modelo em relação à observação pode ser aproximado por uma distribuição Chi-quadrado. A distribuição Chi-quadrado é aparentemente similar a uma distribuição Normal, mas apresenta caudas mais pesadas, ou seja, valores extremos são mais frequentes que na distribuição normal. Por isto, Chi-quadrado é apropriada para avaliar o intervalo de confiança, pois vai gerar valores de frequência não tão baixos como obtida por uma distribuição exponencial como a Normal.

Para tal, supõe-se que o erro obtido das inovações (modelo-observação) se distribua como a distribuição Chi-quadrado [64]. Para uma variável com distribuição quase normal como o consumo da água, há em geral 2 graus de liberdade, o primeiro associado à média, não pode ser considerado, e o segundo associada à variância. Supondo desconhecida a variância da inovação, o modelo produz resultados médios localmente. Assim tem-se apenas 1 grau de liberdade disponível para a distribuição Chi-quadrado. Supondo um intervalo de confiança de 90% de distribuição Chi-quadrado obtém-se o valor 2, 2. Logo, o intervalo de confiança de 90% ( $\sigma_{90\%}$ ) pode ser expresso por:

$$\sigma_{90\%} = \pm 2, 2\sqrt{MSE_{loo}} \quad (2.19)$$

Outra forma de estimar o intervalo de confiança é considerar diretamente as distribuições do modelo e da observação, sem recurso a um modelo paramétrico como Normal ou Chi-quadrado. Isto pode ser feito a partir da matriz de covariância (*cov*) entre modelo e observação. Neste caso, obtida a matriz de covariância, há dois caminhos: o primeiro é obter-se a decomposição de Cholesky de *COV*, resultando em duas matrizes, a primeira triangular inferior ou superior e a segunda, sua transposta. A norma da matriz de Cholesky é uma estimativa da variância dos erros entre modelo e observação. Por outro lado, pode-se calcular os autovalores e autovetores de *COV*. Isto é feito, para determinar os vetores de base do espaço vetorial, que permitem de forma compacta descrever a estrutura covariante com apenas a média e a escala dada por mais ou menos os máximos autovalores das distribuições fdp empíricas da observação e do modelo (i.e., neste caso, considerado também uma variável estocástica), vezes um fator Chi-quadrado.

## Índices estatísticos

Veremos os índices estatísticos a seguir [1]:

- O **coeficiente de determinação** ( $R^2$ ) calcula o grau de correlação entre os valores observados e previstos. É uma medida da capacidade do modelo na relação entre as variáveis de entrada e saída. Entende-se que quanto maior o valor de  $R^2$  (sendo 1, o valor máximo), melhor o desempenho do modelo.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (2.20)$$

com

$$\bar{y}_i = \frac{1}{n} \sum_{i=1}^n y_i \quad (2.21)$$

sendo  $n$  o número de pontos de dados usados,  $y_i$  a demanda de água semanal observada,  $\bar{y}_i$  o valor médio tomado sobre  $n$  e  $\hat{y}_i$  a demanda de água semanal prevista do modelo.

- A **raiz do erro quadrático médio** (*RMSE*, *root mean square error*) usado para avaliação da variância dos erros, dividindo-se por  $n$ , independentemente do tamanho da amostra.

$$RMSE = \sqrt{\frac{SEE}{n}} \quad (2.22)$$

com a **soma dos quadrados dos erros** (*SEE*, *sum of squared errors*)

$$SEE = \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (2.23)$$

quanto menor for o *RMSE*, melhor o desempenho do modelo.

- O **erro quadrático médio** (*MSE*, *mean square error*) é obtido pela divisão da soma dos quadrados dos erros, *SEE*, pelo número de dados na amostra,  $n$ , ou seja,

$$MSE = \frac{SEE}{n} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (2.24)$$

- O valor médio do erro absoluto relativo (*AARE*, *average absolute relative error*) também é uma estimativa útil da incerteza absoluta, particularmente aplicável na aferição da precisão de um modelo de previsões de uma variável flutuante em torno da média.

$$AARE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100\% \quad (2.25)$$

quanto menor for o valor obtido para *AARE*, melhor o desempenho do modelo.

O valor máximo do erro absoluto relativo (*maxAARE*, *maximum absolute relative error*) é também uma medida pontual da confiabilidade do modelo. Quanto menor, melhor o desempenho do modelo.

- O **índice de persistência** (*PI*, *persistence index*) permite uma comparação do desempenho do modelo em relação ao desempenho de um modelo de referência, e.g., um modelo de persistência ou de estrutura homogênea (i.e., com métrica de diferença invariante por translação). Isto pode ser expresso por

$$PI = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - y_{i-L})^2} \quad (2.26)$$

sendo  $y_{i-L}$  a demanda de água observada (entrada) no tempo  $(i-L)$ , para o qual  $L$  é o índice do vetor (tempo) associado ao tempo  $t$ , tal que  $(i-L)$  se refere ao tempo de observação anterior  $t-24$  ( $h$ ).

A probabilidade de explicação da persistência, i.e., do previsão do modelo homogêneo, é de  $(1 - \frac{1}{e}) \approx 63,212\%$ . Em geral, modelos operacionais devem ter uma performance associada a um  $PI > 0,63212$ , i.e., apresentar uma desempenho melhor do que o obtido para o modelo de persistência, considerado como modelo de referência.

O valor  $PI = 1$  reflete ao ajuste perfeito entre os valores previstos e observados (caso ideal). Quando  $PI = 0$  entende-se que o modelo não é melhor que um modelo homogêneo, sendo equipotente. Já  $PI < 0$  o modelo apresenta acurácia inferior em relação ao modelo homogêneo.

- A **tabela de contingência** ( $TC$ ) faz parte dos instrumentos usados em conexão à previsões não-probabilísticas, de forma que não contém diretamente expressões de incerteza [79]. Note-se que a verificação de previsões não-probabilísticas via preditores discretos está sendo praticada desde o século XIX. [54]

A  $TC$  tem como objetivo registrar e analisar a relação entre os pares de previsão e observação, sendo que o critério de acerto ou de erro da previsão baseia-se na ocorrência ou não do evento (binomial ou categorizado). [26]

A comparação das frequências registrada nas células previsão/observação da  $TC$  é efetuada por consolidação e avaliação de proporções, assim estabelecendo um quadro de resumo dos dados (Tabela 2.1). Em geral, os limiares de classe empregados para definir categorias de valores, i.e., usados na contagem das frequências absolutas, são obtidos pelos quantis. Um exemplo de  $TC$  binomial, associada ao quantil 50% é mostrada na Tabela 3.57.

**Tabela 2.1:** Tabela de contingência. (Fonte: EUMETCAL, 2016)

Eventos previstos	Eventos observados		
	Sim	Não	Total
Sim	a	b	a+b
Não	c	d	c+d
Total	a+c	b+d	a+b+c+d=n

sendo (em relação ao quantil alto, sim, ou de ocorrência binomial):

- $a$  frequência absoluta de acertos (i.e., o modelo previu e a observação confirmou);
- $b$  frequência absoluta de falsos alarmes (i.e., o modelo previu chuva e não ocorreu);

- $c$  frequência absoluta de falhas (i.e., o modelo não previu a ocorrência observada);
- $d$  frequência absoluta de acertos opostos (i.e., o modelo não previu a ocorrência não observada).

Com base nas categorias apresentadas acima é possível calcular alguns índices de avaliação, medidos como frequência relativa (razões entre contagens de amostras) ou probabilidade, como indicado a seguir.

**Razão de Falso Alarme (RAF)** O falso alarme é a proporção de previsões de ocorrência de chuva as quais não ocorreram. Quanto mais próximo for o valor de 1, pior é a previsão de chuva.

$$RAF = \frac{b}{a + b} \quad (2.27)$$

**Probabilidade de Detecção (POD)** O POD é a relação entre o evento de chuva previsto e os eventos de chuvas previstos confirmados ou não. Isto, indica o desempenho de detecção da chuva. Tem-se assim a melhor previsão quando a POD for igual a um.

$$POD = \frac{a}{a + c} \quad (2.28)$$

**Índice de Sucesso Crítico (CSI) ou ingl., Threat Score (TS)** A probabilidade *CSI* é aplicada quando a frequência de ocorrência é bem superior à frequência de não-ocorrência. O índice aponta para a melhoria da previsão a medida que  $CSI \rightarrow 1$ . [63]

$$TS = CSI = \frac{a}{a + b + c} \quad (2.29)$$

em que

$a$  é a frequência absoluta de acertos da previsão acima do limiar (i.e., acertos, *scores*);

$b$  é a frequência absoluta de falsos alarmes (i.e, falsos alarmes);

$c$  é a frequência absoluta de falhas (i.e., o modelo não previu o que foi observado).

Porém, devido a ocorrência de acertos aleatórios, *TS* tende a aumentar artificialmente o número de previsões corretas. Tendo em vista a necessidade de correção desta tendência, uma proporção derivada foi estabelecida, denominada *Equitable Threat Score ETS*.

**Índice de sucesso equitável em ingl., Equitable Threat Score (ETS)** corresponde a estimativa corrigida do (*CSI*), podendo ser obtido da seguinte razão,

$$ETS = \frac{a - a_r}{a + b + c - a_r} \quad (2.30)$$

sendo

- $a$  a frequência absoluta de acertos da previsão acima do limiar (i.e., acertos, *scores*);

- $a_r$  a frequência absoluta dos acertos por processos aleatórios (i.e., acertos por chance);
- $b$  a frequência absoluta de falsos alarmes (i.e, falsos alarmes);
- $c$  a frequência absoluta de falhas (i.e., o modelo não previu o que foi observado).

A frequência absoluta dos acertos por chance pode ser estimada por

$$a_r = \frac{(a+b)(a+c)}{n} \quad (2.31)$$

sendo  $n = a + b + c + d$  o número total de eventos observados (previstos).

O  $ETS$ , assim como também ( $CSI$ ), também ocorre no intervalo de 0 a 1. No caso de  $ETS = 1$  tem-se o caso ideal, dito *previsão perfeita*.<sup>1</sup>

**Grau de Desproporção (DD) em ingl., *Disproportion Degree*** permite inferir informação sobre o balanço das frequências (contagens de acertos) ao longo das categorias contingenciadas. Portanto, obtem-se informação sobre a distribuição da previsibilidade do modelo, sendo o valor ideal igual a 1.<sup>2</sup>

$$DD = \left| \frac{a}{a+b} - \frac{c}{c+d} \right| \quad (2.32)$$

$BIAS_{ratio}$  não apresenta uma medida da previsibilidade (i.e., probabilidade de acerto), mas sim, a razão entre as frequências previstas e observadas.

$$BIAS_{ratio} = \frac{a+b}{a+c} \quad (2.33)$$

A previsão pode subestimar (i.e., no caso de  $BIAS_{ratio} < 1$ ) ou superestimar a ocorrência do evento observado (i.e., no caso de  $BIAS_{ratio} > 1$ ).

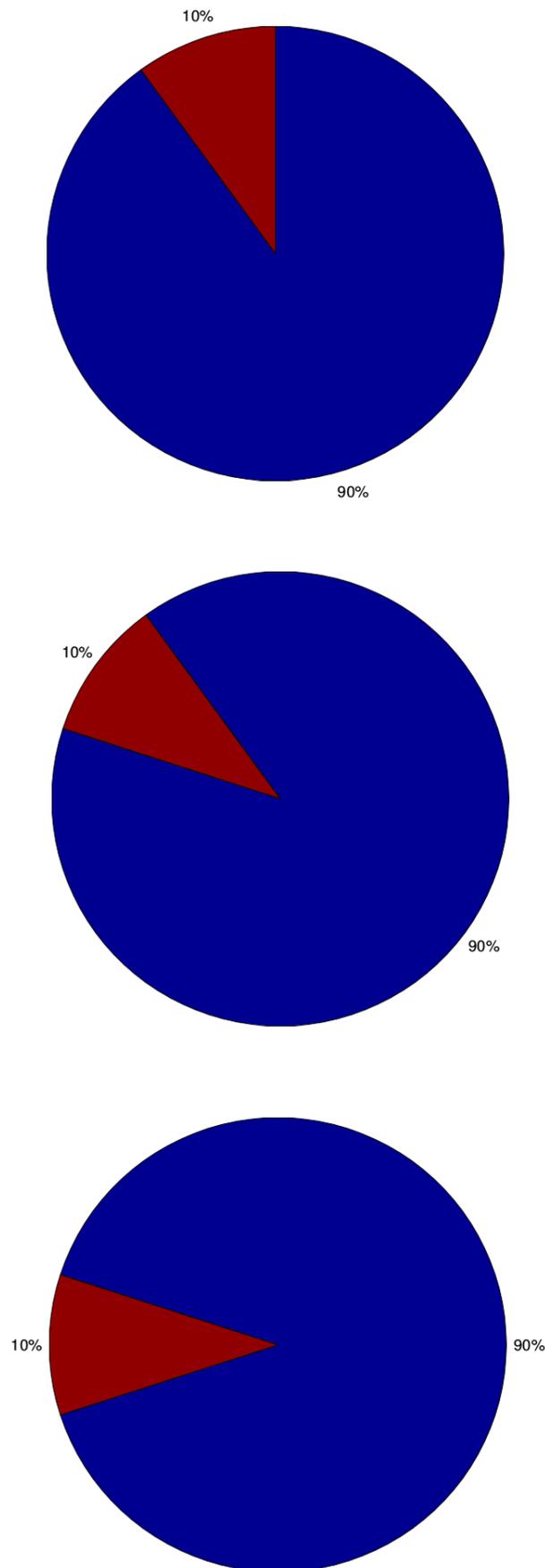
**Índice de habilidade associado (SSA) (em ingl., *Skill Score, Associated*)**

$$SSA = 1 - \frac{MSE}{MSE_{ref}} \quad (2.34)$$

sendo  $MSE_{ref}$  o índice  $MSE$  da previsão obtida com um modelo de referência. Uma comparação com a descrição do **índice de persistência (PI)** feita acima, indica que é conveniente neste trabalho a escolha do modelo de persistência de 1 dia para o modelo de referência. Outra possibilidade é um modelo estocástico de distribuição uniforme no intervalo das observações ou um modelo de distribuição Normal (i.e., f.d.p. e f.d.a. Normal), se as observações *aderirem bem* à distribuição Normal.

<sup>1</sup>Para uma referência eletrônica do índice *Grau de Desproporção* acesse o website sobre o Critical Success Index ( $CSI$ ) and Equitable Threat Score ( $ETS$ ) no endereço [www.eumetcal.org/resources/ukmeteocal/verificationSAV/www/english/msg/ver\\_catg\\_forec/uos2\\_Ko4.htm](http://www.eumetcal.org/resources/ukmeteocal/verificationSAV/www/english/msg/ver_catg_forec/uos2_Ko4.htm).

<sup>2</sup>Para uma referência eletrônica do índice *Grau de Desproporção* acesse o *Website for Statistical computation* no endereço [Vassarstats.net/tab2x2.html](http://Vassarstats.net/tab2x2.html).



**Figura 2.5:** Primeiras três amostras de  $K = 10$  em que cada interação, 90% dos dados são utilizados para treinamento e os 10 % restantes são usados para o teste, sem repetição da amostra. Fonte: Bontempi (2003)

# Capítulo 3

## Resultados

### 3.1 Dados

#### Temperatura

Para uma representatividade dos dados de entrada da RNA serão descritos nessa seção as seguintes variáveis: temperatura, umidade relativa, velocidade do vento e consumo de água. Apresentaremos gráficos do comportamento temporal das variáveis, histogramas com função de distribuição de probabilidade (f.d.p.), função de distribuição acumulada (f.d.a.), boxplot anual e quantil-quantil (Q-Q).

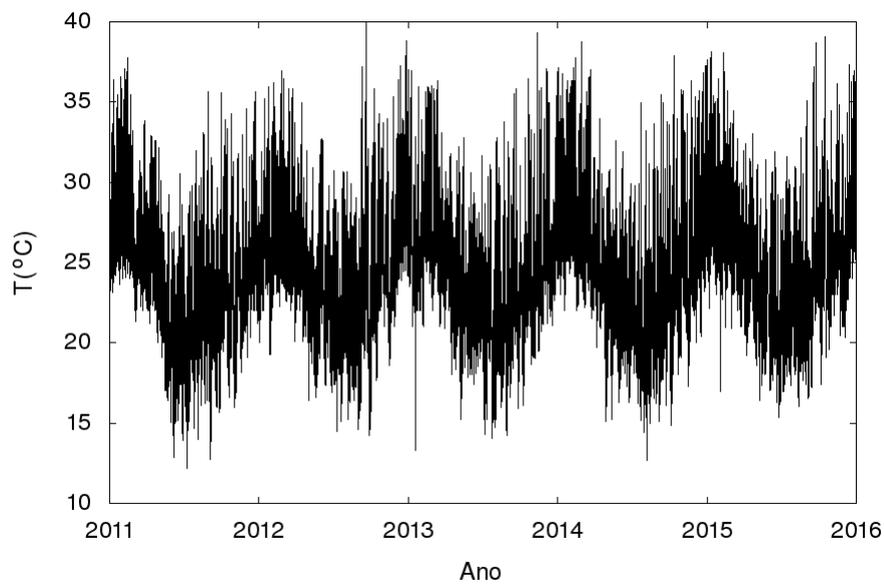
O comportamento da temperatura do ar está expresso em uma variação temporal, descrito na Figura 3.1. O gráfico presente na figura indica um ciclo anual bem definido, com um grande conjunto de temperaturas que ficam entre 20 e 30°C e uma média em torno de 25°C, que indica características de uma região tropical.

A partir do comportamento da temperatura gerou-se o histograma da f.d.p. (Figura 3.2) e juntamente foi traçada a curva da densidade da normal que indica a distribuição dos dados de forma quase normal, com dados iniciais abaixo da média e dados finais acima da média (Figura 3.3) que indica um aumento da temperatura do ar nos últimos anos.

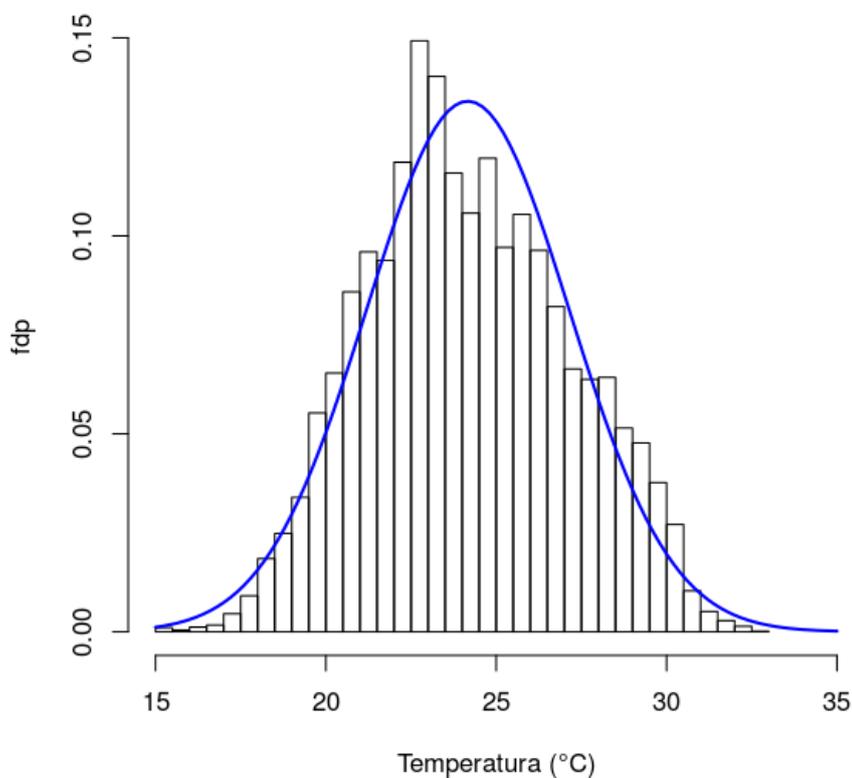
Para distribuição empírica dos dados de temperatura utilizou-se o boxplot anual, conforme a Figura 3.4. Nesse gráfico tem-se o limite inferior igual a 15°C, o limite superior igual a 33°C, a mediana igual a 24°C, primeiro quartil igual a 22°C e terceiro quartil com valor de 26°C. Por meio das posições dos quartis observamos que grande parte dos dados possuem valores acima da média e temos *outliers* com valores abaixo de 15°C.

#### Umidade relativa

A umidade relativa é uma variável muito importante no que diz respeito ao desconforto térmico que é avaliado a partir de medições em conjunto da umidade relativa e

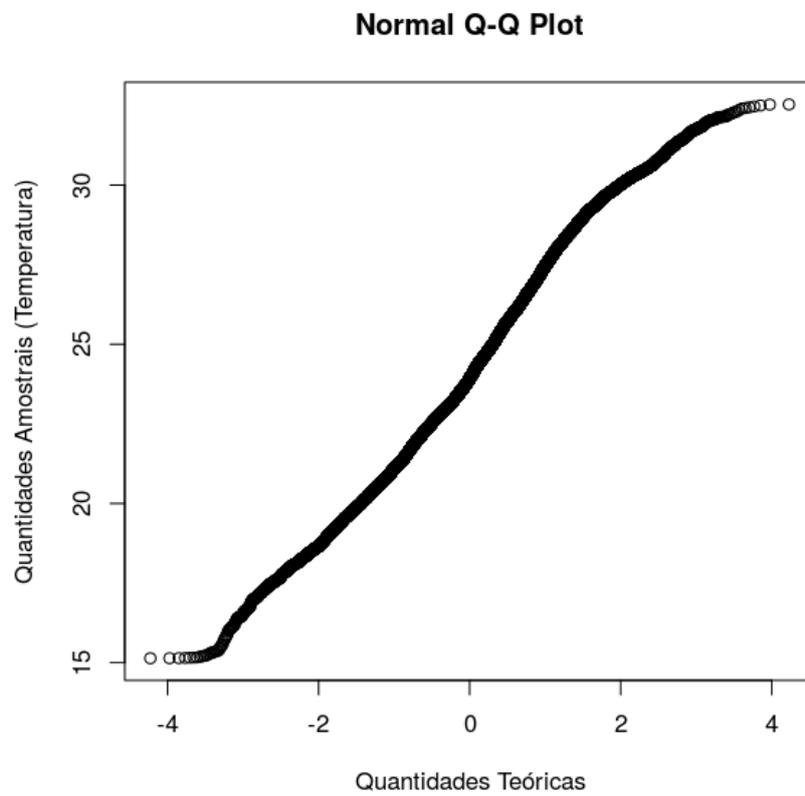


**Figura 3.1:** *Varição temporal da temperatura horária do ar na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*

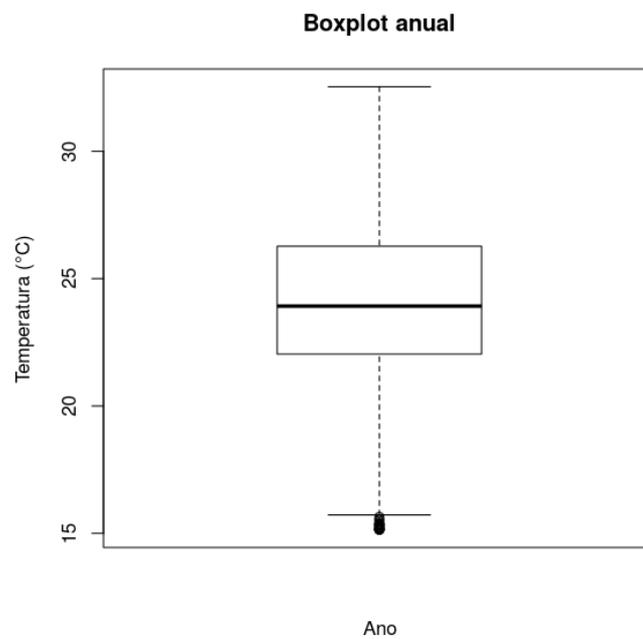


**Figura 3.2:** *Histograma da temperatura do ar na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. A linha contínua representa a curva da Normal.*

da temperatura do ar [12]. O comportamento temporal dessa variável é apresentado

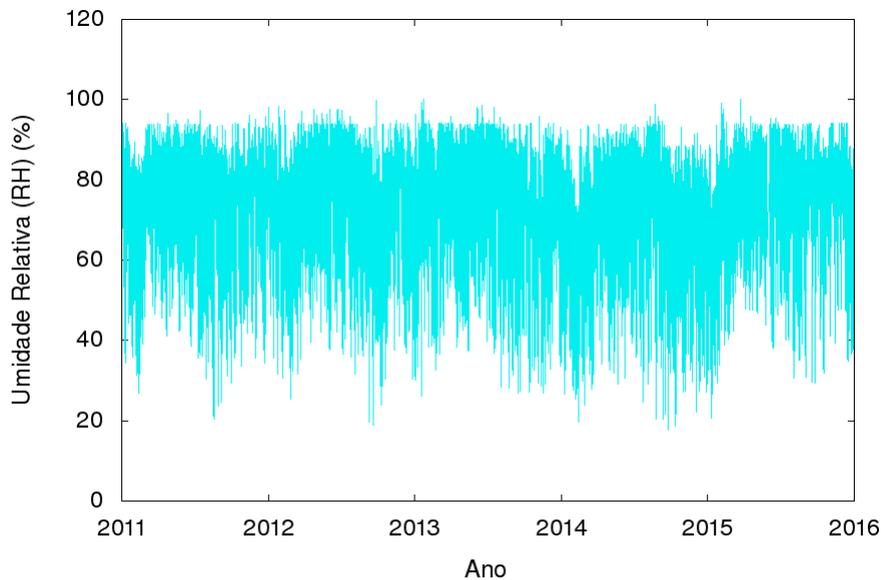


**Figura 3.3:** *Quantil-Quantil da temperatura do ar na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*



**Figura 3.4:** *Boxplot anual da temperatura horária do ar na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*

na Figura 3.5. É possível identificar uma variabilidade ao longo de toda a série que indica períodos com baixas e altas umidades e média em torno de 75%.



**Figura 3.5:** *Variação temporal da umidade relativa horária na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*

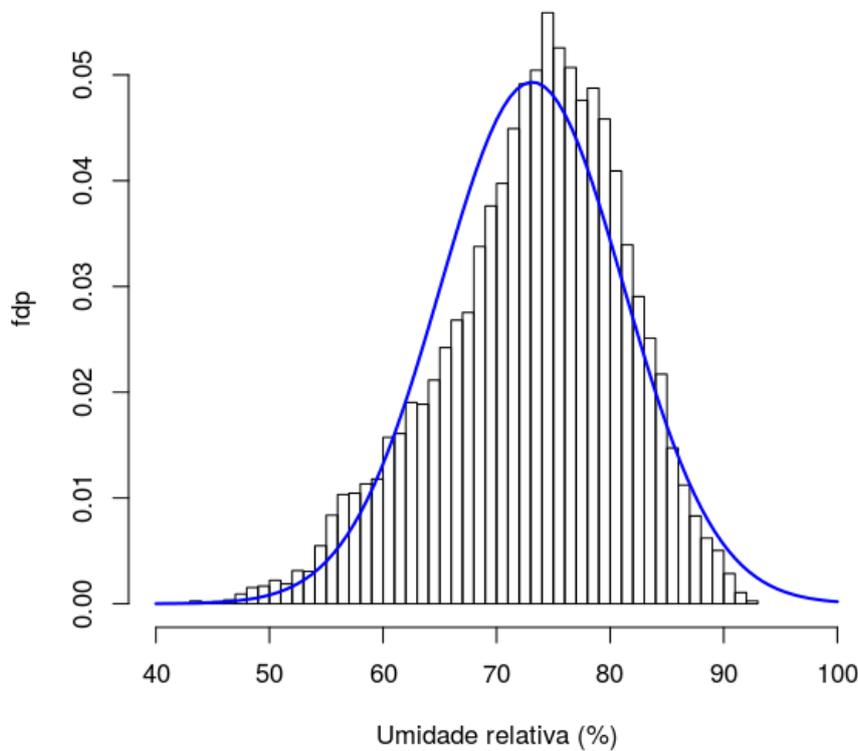
A f.d.p. da umidade relativa é descrita na Figura 3.6. O histograma e uma curva normal foram plotados juntos, com isso identificou-se que a curva não acompanha o histograma, i.e., os dados possuem assimetria devido a distribuição irregular dos dados dessa amostra, tem-se a maior frequência de dados no intervalo de 75%.

A comparação dos dados empíricos da umidade relativa com os valores ajustados dessa variável é descrita na Figura 3.7. A ligação entre as observações e a distribuição ajustada mostra uma superestimativa dos valores devido a maior concentração de dados ter valores acima de 68%.

O boxplot anual da umidade relativa é visto na Figura 3.8. Esse gráfico mostra que cerca de 50% dos dados centrais estão entre 68 e 79%. A mediana tem valor de 74% e a maioria dos dados entre o primeiro quartil e o terceiro quartil estão localizados abaixo da mediana. O limite superior e inferior correspondem a 92 e 52% respectivamente, além disso notamos valores abaixo do limite inferior com valor mínimo de 43%. Isso indica uma distribuição irregular dos dados.

## Pressão

A Figura 3.9 mostra o histograma da pressão atmosférica. No gráfico temos um estreitamento das barras justificado pela amostragem de dados ser grande. A curva plotada não acompanha a forma do histograma pelo fato dos dados serem multimodais.



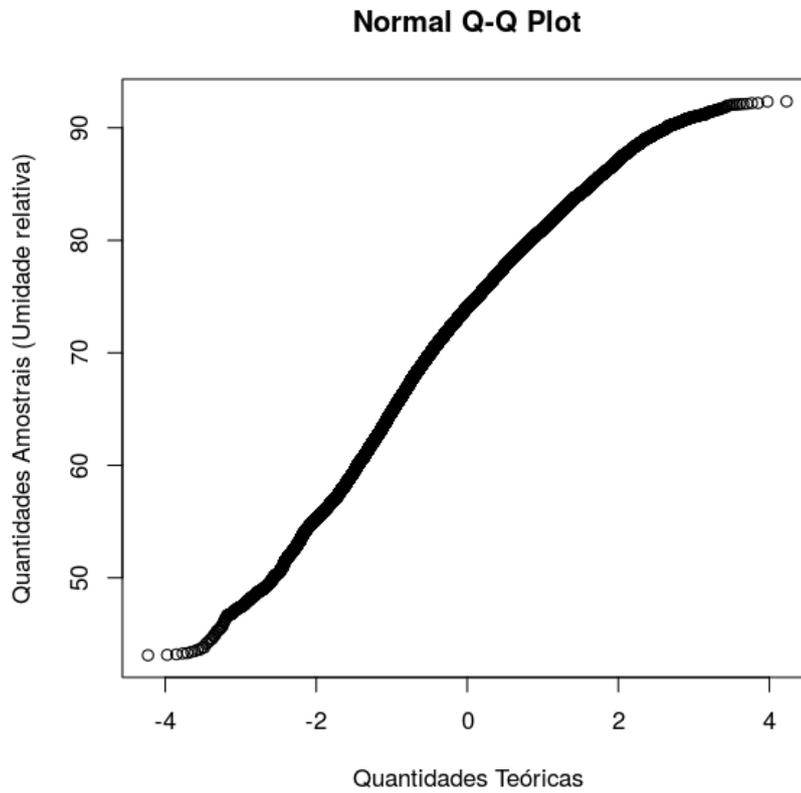
**Figura 3.6:** Histograma da umidade relativa na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. A linha contínua representa a curva da Normal.

O gráfico Q-Q é uma ferramenta muito útil para checar a adequação de distribuição de frequência dos dados. Com base nisso, aplicamos os dados de pressão para geração do gráfico, como visto na Figura 3.10 que mostra uma ligeira subestimativa sobre grande parte dos valores da série ordenada, i.e., essas discrepâncias podem estar atribuídas a uma ligeira variação em relação a Normal.

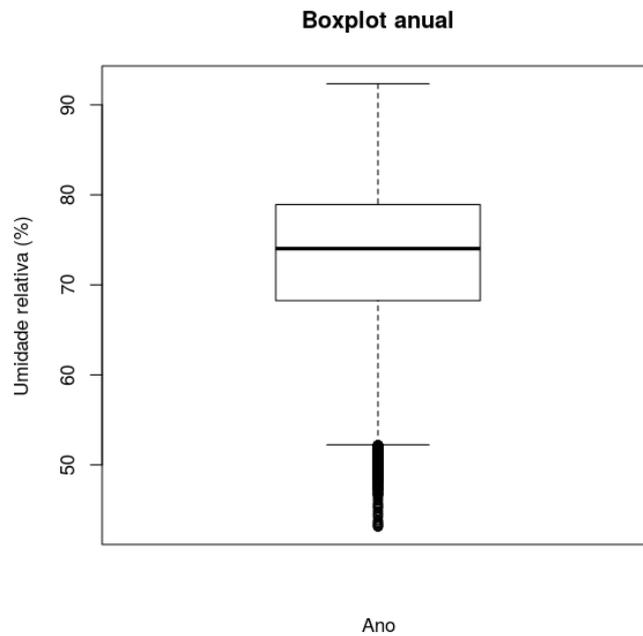
Para representação dos dados subjacentes utiliza-se o gráfico boxplot. Por meio da Figura 3.11 é descrito o gráfico boxplot anual de pressão atmosférica indicando que 50% dos dados estão entre 1012 e 1018 hPa. O limite inferior e superior correspondem a 1004 e 1027 hPa respectivamente. Como a mediana não está no centro da caixa e há valores extremos (e.g., 1031 hPa) acima do limite superior não existe simetria nos dados.

### Velocidade do vento

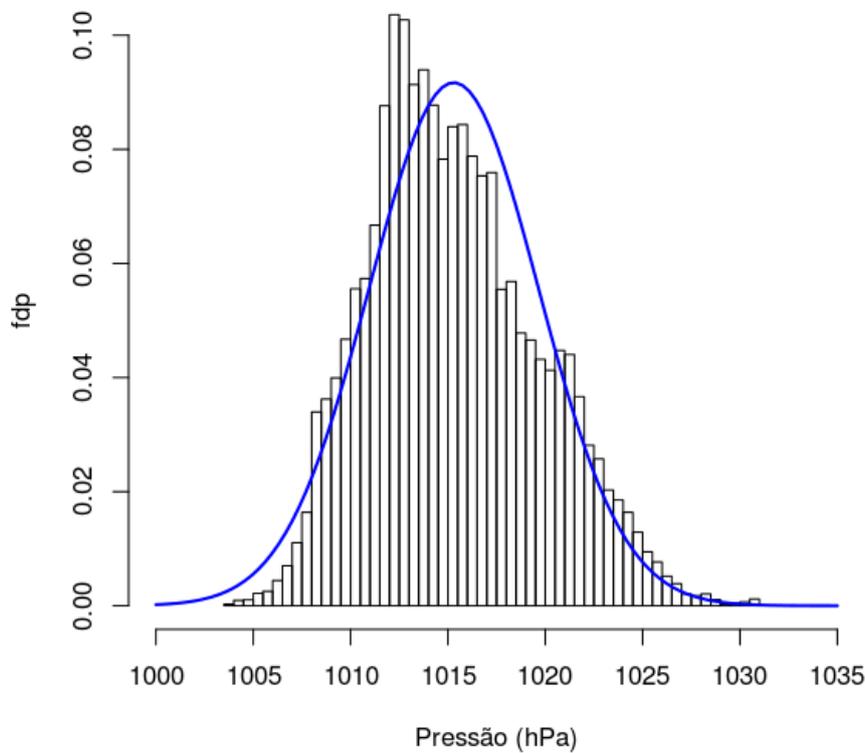
A velocidade do vento é uma variável meteorológica que tem influência sobre a sensação térmica. A Figura 3.12 mostra a exibição temporal da velocidade do vento durante o período estudado. A velocidade média do vento corresponde a  $3 \text{ m s}^{-1}$  e valores extremos mostrados no gráfico acima de  $20 \text{ m s}^{-1}$  foram considerados erros



**Figura 3.7:** *Quantil-Quantil da umidade relativa na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*



**Figura 3.8:** *Boxplot anual da umidade relativa na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*



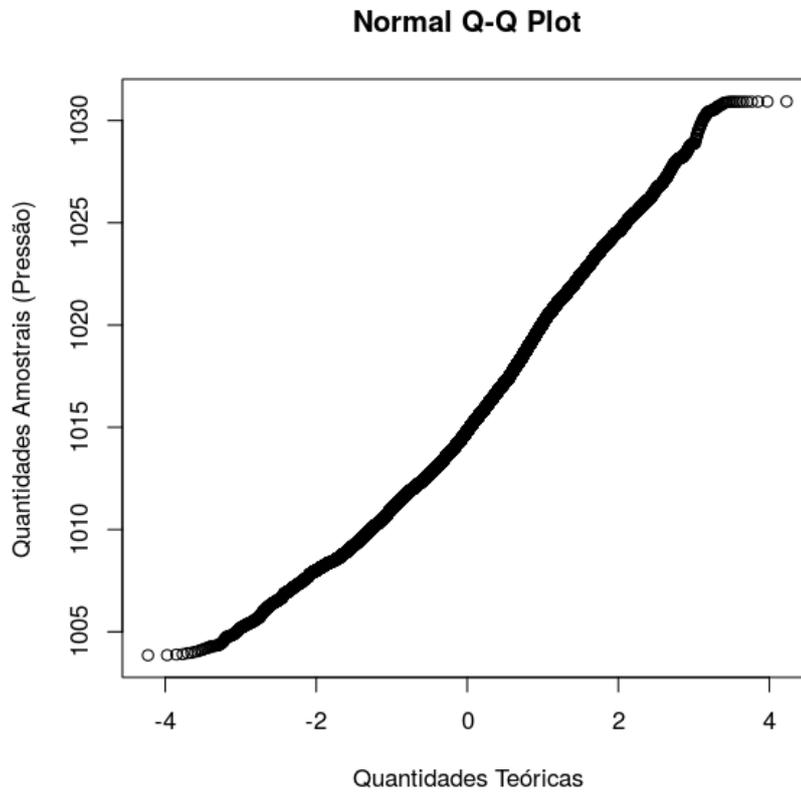
**Figura 3.9:** Histograma da pressão atmosférica na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. A linha contínua representa a curva da Normal.

instrumentais que sobreviveram ao processo de filtragem. Isso indica que a rotina de verificação dos dados deve ser melhorada implementando técnicas mais rigorosas nas próximas versões. Por exemplo, o padrão usado para definir o intervalo de filtragem poderia ser substituído por uma métrica entre os quantis de 10 e 90%, o que efetivamente eliminaria valores extremos (*spikes*).

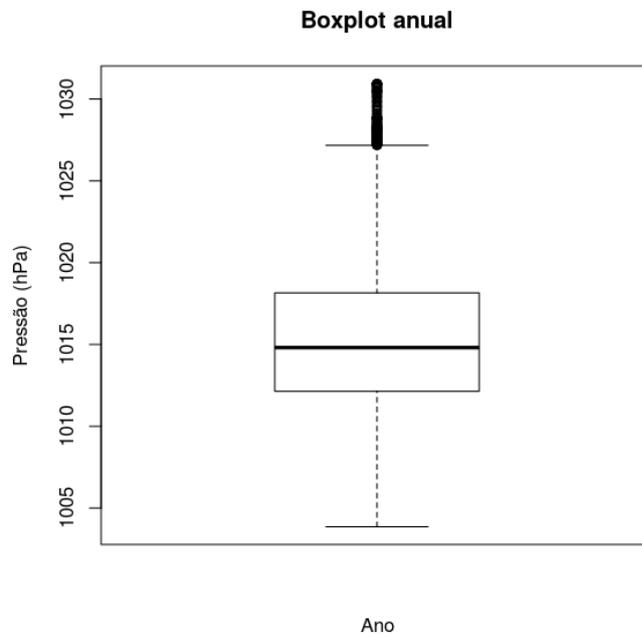
A distribuição dos dados de velocidade do vento é vista na Figura 3.13. O gráfico representado em forma de histograma da f.d.p. possui uma distribuição normal indicada pela curva em forma de sino que o acompanha. É possível identificar que 60% dos dados tem valor de  $3 \text{ m s}^{-1}$ .

A Figura 3.14 mostra a plotagem Q-Q da velocidade do vento. O gráfico Q-Q indica que a distribuição ajustada da velocidade do vento corresponde bem aos dados observados, durante boa parte do gráfico. Com isso, é possível observar que a função Q-Q avaliada nesse gráfico contém valores próximos dos dados observados, como visto na trajetória da curva expressa no gráfico.

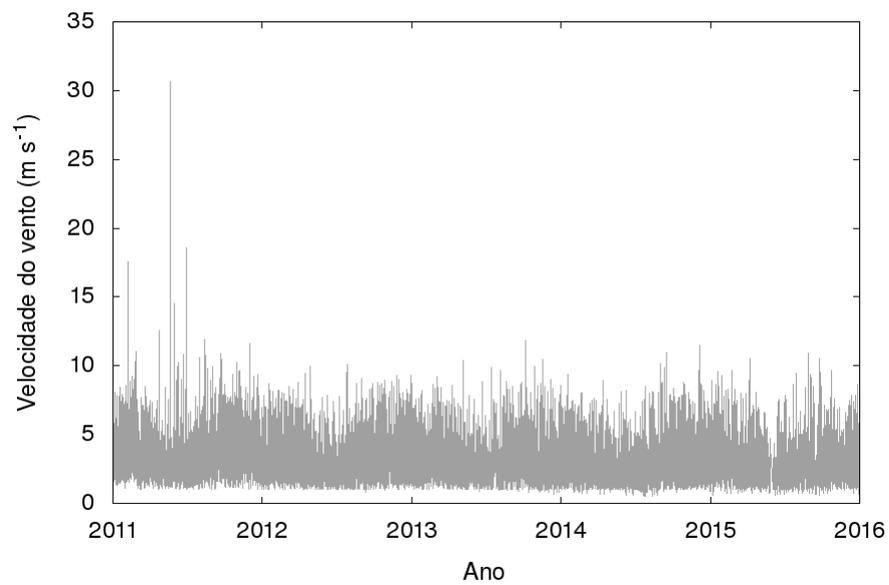
Para análise da distribuição dos dados subjacentes de velocidade do vento foi gerado um gráfico em boxplot anual, conforme visto na Figura 3.15. O primeiro quartil tem



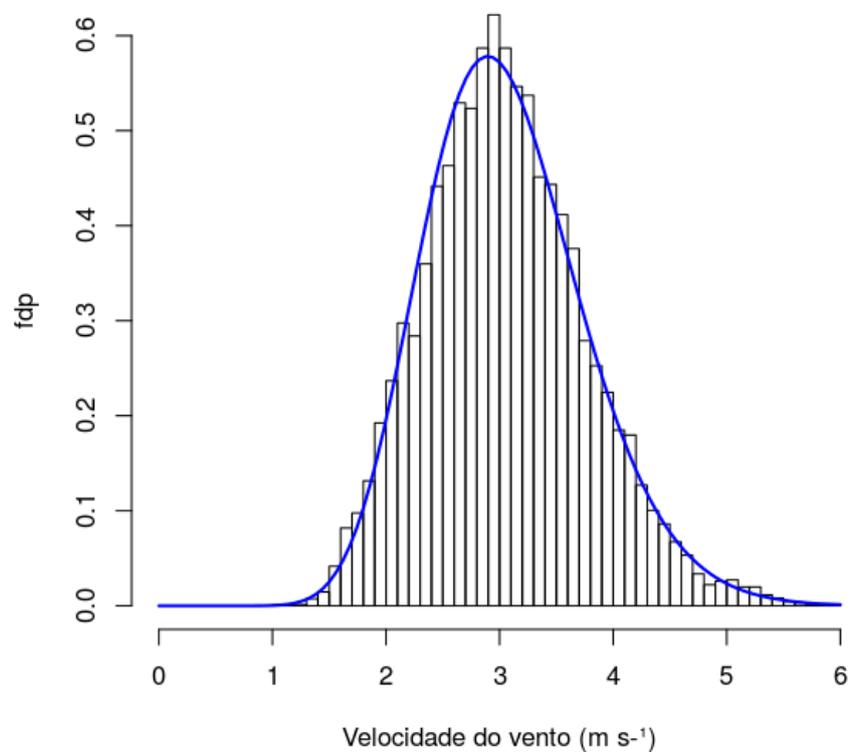
**Figura 3.10:** *Quantil-Quantil da pressão atmosférica na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*



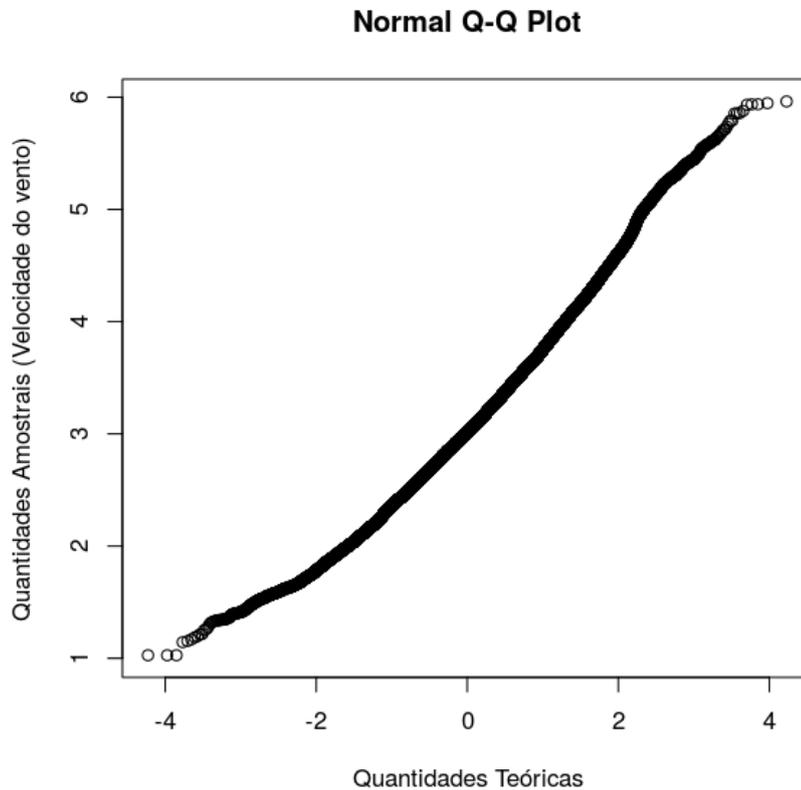
**Figura 3.11:** *Boxplot anual da pressão atmosférica na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*



**Figura 3.12:** *Varição temporal da velocidade horária do vento na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*



**Figura 3.13:** *Histograma da velocidade do vento na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015. A linha contínua representa a curva da Normal.*



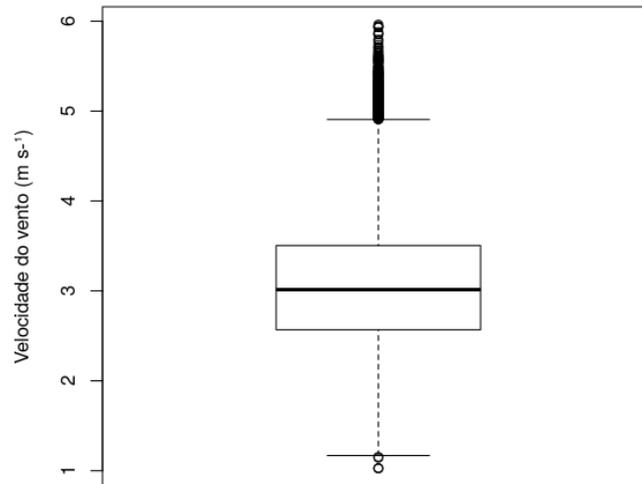
**Figura 3.14:** *Quantil-Quantil da velocidade do vento na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*

valor de  $2,6 \text{ m s}^{-1}$  e o terceiro quartil de  $3,6 \text{ m s}^{-1}$ . Além disso observa-se que a mediana apresenta valor de  $3 \text{ m s}^{-1}$ , i.e., a mediana está localizada no centro da caixa com 50% dos dados estão bem distribuídos. O limite inferior e superior correspondem a  $1,03$  e  $4,9 \text{ m s}^{-1}$  respectivamente e notamos a presença de outliers acima do intervalo de limite superior com valor máximo de  $6 \text{ m s}^{-1}$ .

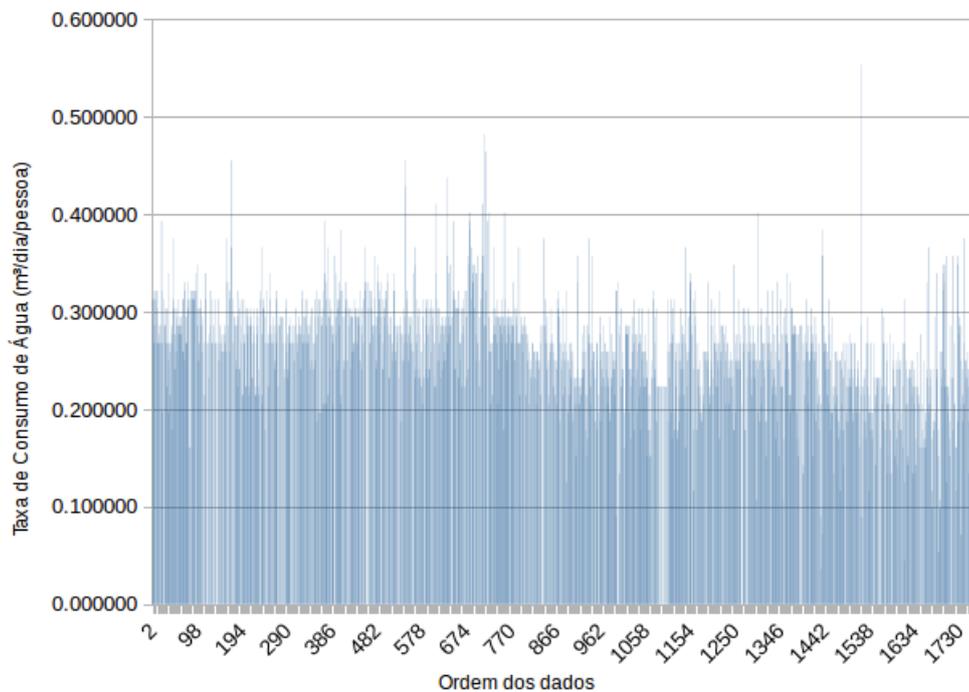
## Consumo de água

A variação temporal do consumo de água ( $\text{m}^3 \text{ dia}^{-1} \text{ pessoa}^{-1}$ ) (união de dados de uma residência e de um prédio) para o período estudado é mostrada na Figura 3.16. Os valores de consumo de água diários foram agregados em um arquivo único, tanto para residência quanto para o prédio, após digitação.

O comportamento da variação temporal mostra uma variabilidade anual bem caracterizada. É possível identificar uma diminuição do consumo ao longo dos últimos anos. Entretanto, é possível notar uma tendência de diminuição nos anos de 2013, 2014 e 2015. Isso indica que as pessoas utilizam menos água no edifício, ou fatores externos podem ser a causa disso como, maior custo, racionamento voluntário, campanhas de redução, menor fornecimento de água por redução da pressão, por causa de crises hídricas na região, entre outros.



**Figura 3.15:** *Boxplot anual da velocidade do vento na estação do Aeroporto Internacional Tom Jobim (Galeão-RJ) para o período de 2011 a 2015.*

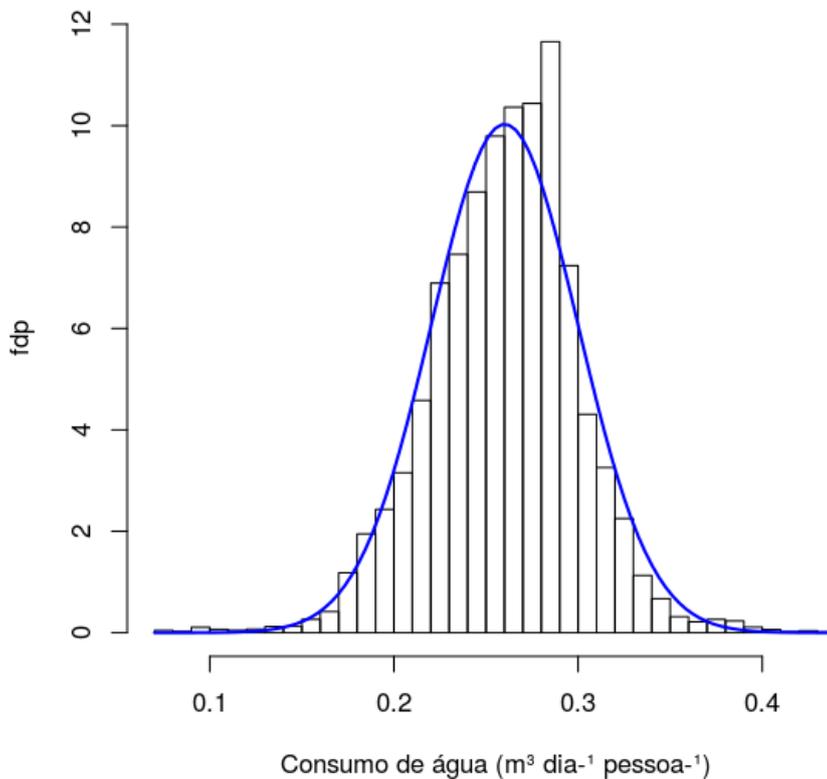


**Figura 3.16:** *Variação temporal do consumo diário de água (união de dados de uma residência e de um prédio) para o período de 2011 a 2015.*

Do ponto de vista estatístico, o histograma relativo ou f.d.p. do consumo de água (Figura 3.17) indica uma distribuição quase Normal, sendo que este modelo é mostrado pela curva que acompanha o histograma. O máximo de frequência é encontrado no

valor aproximado de  $0,28 \text{ (m}^3 \text{ dia}^{-1} \text{ pessoa}^{-1}\text{)}$ . Este valor corresponde a 25,5 vezes maior que o valor recomendado pela FAO. [57]

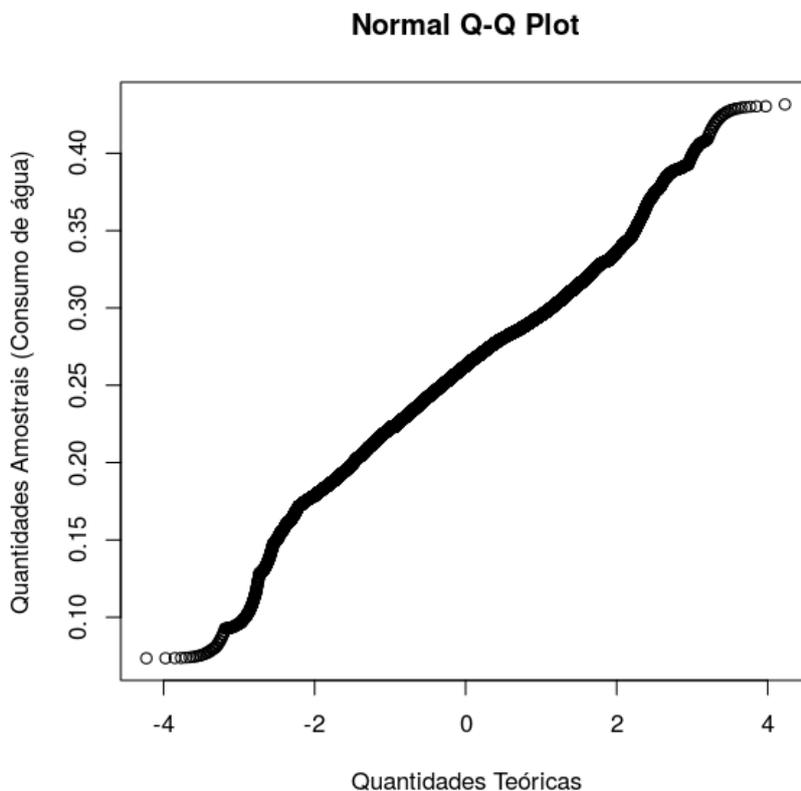
Segundo a CEDAE, a população da cidade do Rio de Janeiro, i.e., 6 milhões de pessoas, consome cerca de  $25 \text{ (m}^3 \text{ s}^{-1}\text{)}$  de água, isso equivale a  $2,16 \times 10^6 \text{ (m}^3 \text{ dia}^{-1}\text{)}$ . Dividindo o consumo de água pela população temos o valor de  $0,36 \text{ (m}^3 \text{ dia}^{-1} \text{ pessoa}^{-1}\text{)}$ . Então concluímos que os valores obtidos neste trabalho estão compatíveis com os dados da CEDAE.



**Figura 3.17:** Histograma do consumo diário de água (união de dados de uma residência e de um prédio) para o período de 2011 a 2015. A linha contínua representa a curva da Normal.

O gráfico Q-Q do consumo de água é mostrado na Figura 3.18. Este gráfico explica a relação entre a distribuição dos quantis do consumo de água observada e a distribuição correspondente Normal. Mais uma vez se confirma que a distribuição observada é quase Normal, mas não exatamente Normal pois apresenta duas peculiaridades, a primeira uma ligeira subestimativa dos primeiros quantis e segundo uma ligeira superestimativa dos maiores quantis.

O boxplot anual do consumo de água pode ser analisado a partir da Figura 3.19. Por meio da caixa é possível identificar os quartis inferior e superior, com valores de  $0,24$  e  $0,29 \text{ (m}^3 \text{ dia}^{-1} \text{ pessoa}^{-1}\text{)}$ , respectivamente. A mediana tem valor de



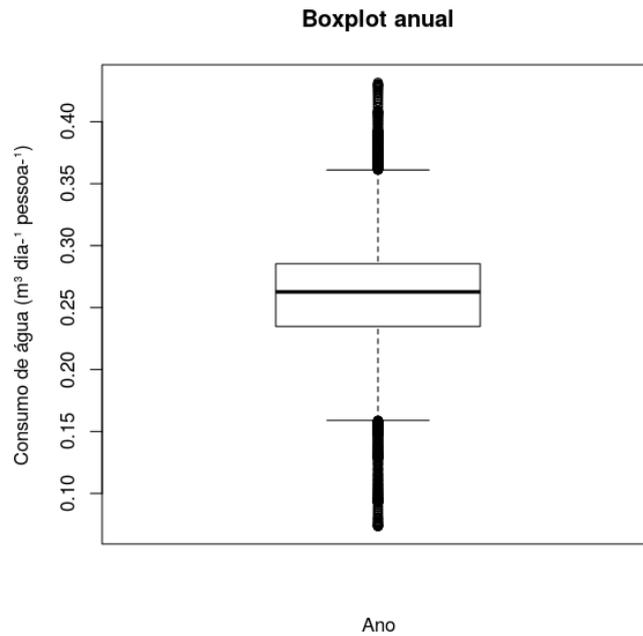
**Figura 3.18:** *Quantil-Quantil do consumo diário de água (união de dados de uma residência e de um prédio) para o período de 2011 a 2015.*

0,26 ( $m^3 \text{ dia}^{-1} \text{ pessoa}^{-1}$ ) que separa 50% dos dados abaixo dos 50% dos dados acima. Os limites inferior e superior são 0,16 e 0,35 ( $m^3 \text{ dia}^{-1} \text{ pessoa}^{-1}$ ) respectivamente. Além desses limites há pontos isolados (*outliers*) com valor mínimo de 0,07 e valor máximo de 0,43 ( $m^3 \text{ dia}^{-1} \text{ pessoa}^{-1}$ ), esses valores podem ser associados à vazamentos no sistema.

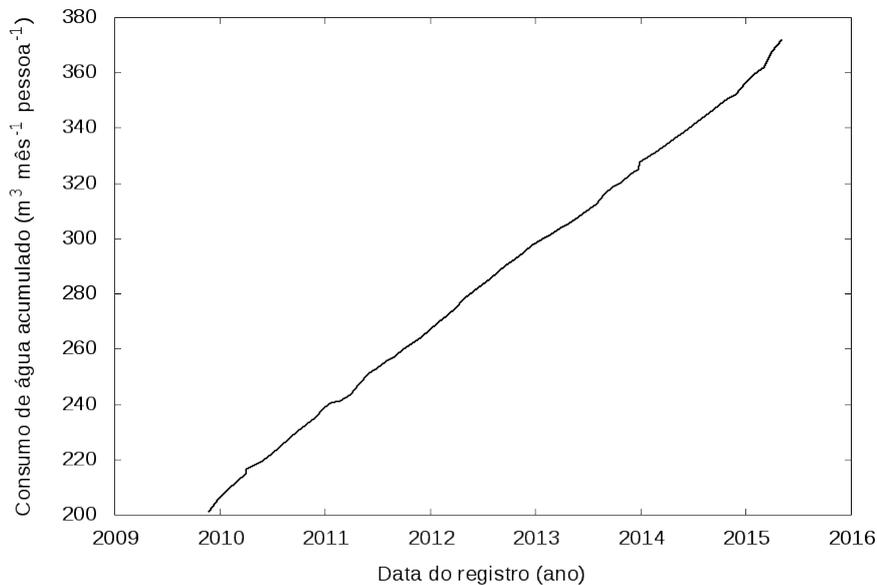
## 3.2 Consumo de água de uma casa da RMRJ

Os dados mensais de consumo de água de uma casa foram obtidos pelo hidrômetro no período de 26/11/2009 a 05/05/2015. A Figura 3.20 mostra a evolução temporal (em meses) do consumo acumulado de água por pessoa de uma família, residente no bairro São Nicolau, São João de Meriti (baixada fluminense). A variação mensal fica entre 201 e 371 ( $m^3 \text{ pessoa}^{-1} \text{ mes}^{-1}$ ).

A Figura 3.21 destaca o consumo diário por pessoa de uma família residente no bairro São Nicolau, São João de Meriti (baixada fluminense). O consumo de água é expresso por diferença finita simples (linha fina) e por diferença finita centrada (linha espessa). A diferença simples apresenta um erro de truncamento de primeira ordem enquanto a diferença centrada apresenta um erro de segunda, sendo portanto



**Figura 3.19:** *Boxplot anual do consumo diário de água (união de dados de uma residência e de um prédio) para o período de 2011 a 2015.*



**Figura 3.20:** *Variação temporal do consumo acumulado por pessoa de uma família residente no município de São João de Meriti-RMRJ, no período de novembro/2009 a maio/2015.*

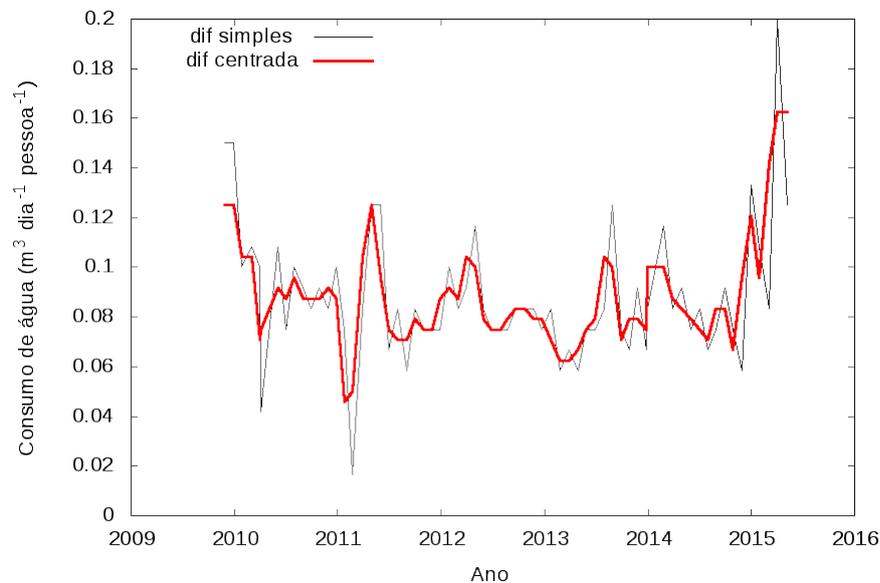
mais precisa (linha espessa).

Os máximos de taxa de consumo de água ocorreram no início dos anos de 2010 e 2011. Em janeiro de 2011 ocorreram chuvas intensas na região serrana do estado do Rio de Janeiro que ocasionaram no maior deslizamento da história [30]. Isso pode ter afetado o sistema com perdas ao longo da rede de distribuição e consequentemente a companhia de distribuição reduziu a oferta de água. Em contrapartida, pode ter

ocorrido um acúmulo de água pela família devido a baixa disponibilidade de água e assim um consumo de água acima do normal.

Outro ponto que pode ser ressaltado é a razão dos custos de serviços relacionados à água e sua política de distribuição [62]. No verão, espera-se um aumento de consumo, associado as necessidades fisiológicas. A quantidade de água que supre as necessidades fisiológicas é muito pequena em relação ao consumo individual observado, o que dá margem a uma possível redução do consumo. A oferta de água é controlada pela disponibilidade de água no ambiente, por períodos de manutenção da rede, por vazamentos, por estiagem, por secas, pela política de distribuição etc.

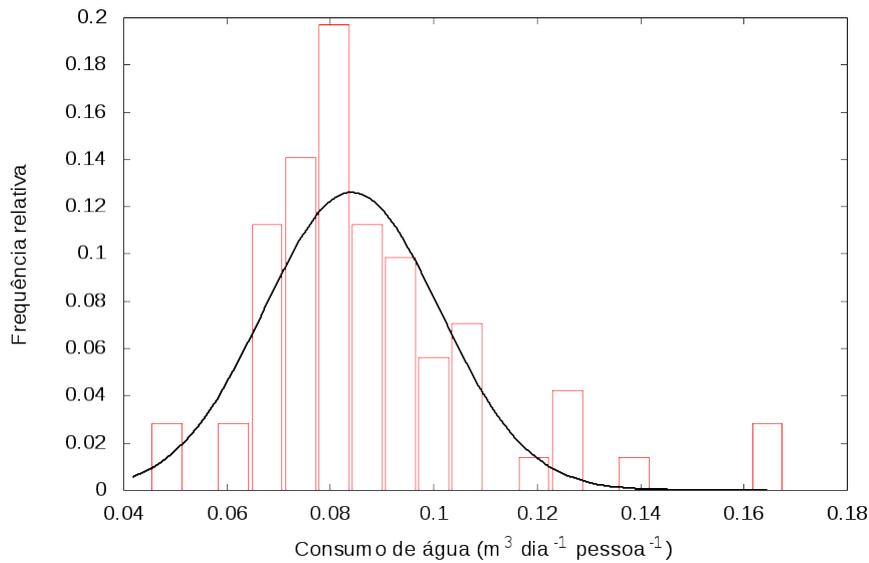
Os máximos de consumo de água ocorreram em novembro de 2009 e abril de 2015, essa maior demanda pode ter ligação com as altas temperaturas ou até mesmo com as perdas de água do sistema.



**Figura 3.21:** *Evolução temporal por diferença simples e centrada do consumo diário de água individual na residência familiar em São João de Meriti-RMRJ, no período entre novembro de 2009 e maio de 2015.*

A partir dos dados foi gerado o histograma de frequência relativa do consumo diário individual, ilustrado na Figura 3.22.

Uma curva gaussiana associada a fdp Normal foi ajustada ao histograma empírico. Cerca de 54% dos dados possuem valor igual ou maior que 0,08 ( $m^3 \text{ dia}^{-1} \text{ pessoa}^{-1}$ ), i.e., sendo que cada indivíduo da família consome em média aproximadamente 80 litros de água por dia, um gasto 7 vezes maior que o recomendado pela FAO[57] de 11 litros por dia, ou seja, 0,011 ( $m^3 \text{ dia}^{-1} \text{ pessoa}^{-1}$ ).



**Figura 3.22:** *Frequência relativa do consumo diário de água de cada membro da família residente em São João de Meriti-RMRJ, no período compreendido entre novembro de 2009 a maio de 2015.*

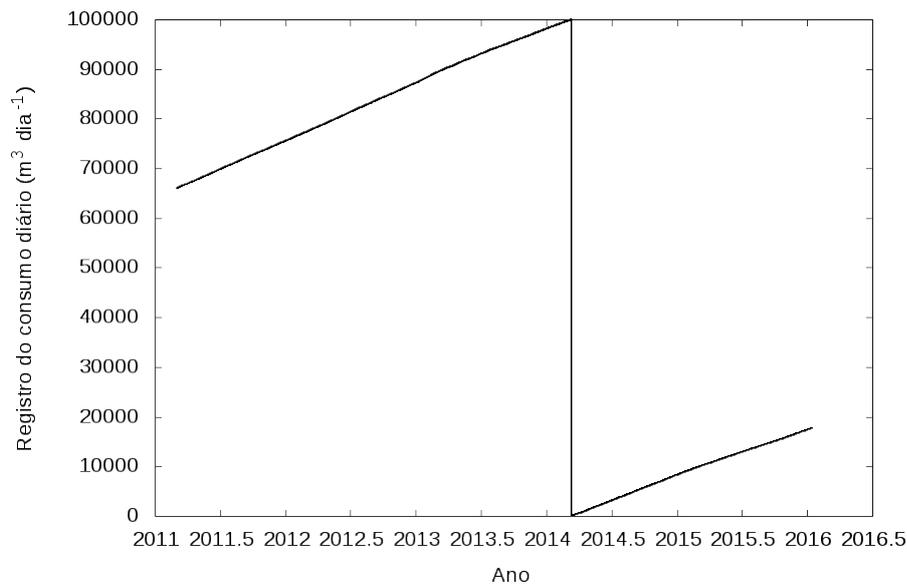
### 3.3 Consumo de água em uma edificação-condomínio da RMRJ

A Figura 3.23 mostra a evolução temporal do registro do consumo acumulado de água de um prédio residencial localizado no bairro de Laranjeiras, no Rio de Janeiro, RJ. Estima-se que nesse prédio com 56 apartamentos residam 112 pessoas (i.e., 2 por apartamento). Os dados foram medidos por um hidrômetro e anotados por um funcionário do prédio no período de 03/02/2011 a 12/12/2015. A descontinuidade de primeira ordem que aparece no gráfico indica a data do reinício da contagem do hidrômetro, a qual ocorre quando o acumulo totaliza o valor 100000 ( $m^3$ ).

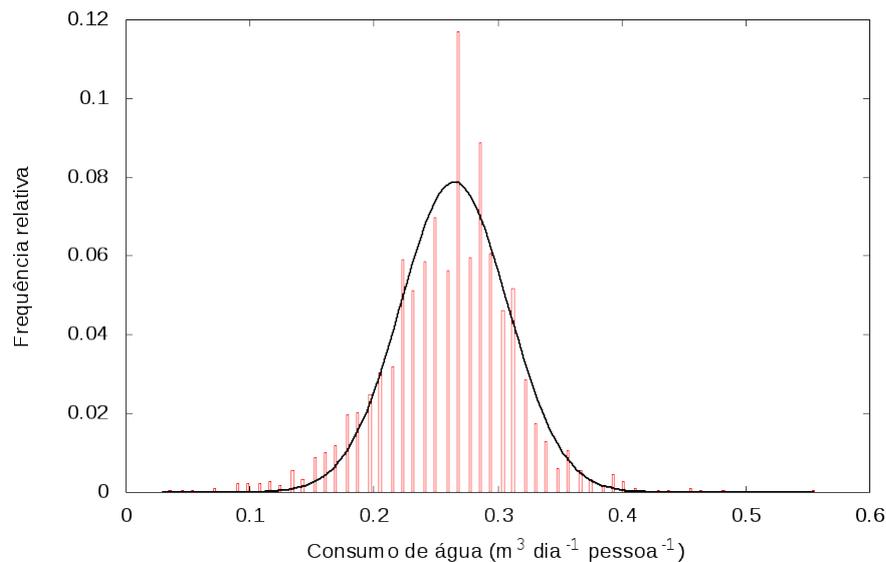
A Figura 3.24 apresenta a frequência relativa do consumo diário de água individual. O modelo Normal foi ajustado aos dados empíricos evidenciando que a distribuição é praticamente Normal e o valor da moda corresponde a 2.7 ( $m^3 \text{ dia}^{-1} \text{ pessoa}^{-1}$ ).

A Figura 3.25 apresenta um gráfico com ajuste de polinômio de 6<sup>a</sup> ordem do consumo diário individual. O ajuste polinomial mostra variações interanuais da taxa de consumo além de uma tendência de redução do próprio consumo ao longo do tempo.

A Figura 3.26 mostra a função *distribuição de probabilidade acumulada* (fda) e da *função de densidade de probabilidade* (fdp), ambas em função da taxa de consumo de água potável. A figura enfatiza a Normalidade da distribuição empírica, evidenciada pela forma sigmoideal da fda.



**Figura 3.23:** Registro do consumo acumulado de água ao longo do tempo do prédio residencial em Laranjeiras, Rio de Janeiro, entre fevereiro de 2011 a dezembro de 2015.

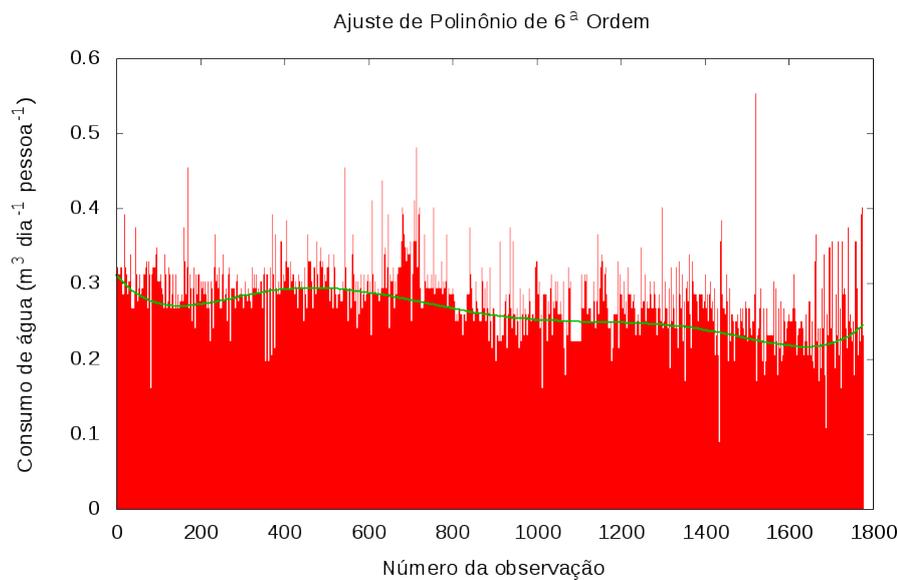


**Figura 3.24:** Frequência relativa do consumo de água por pessoa residente em um prédio residencial em Laranjeiras, Rio de Janeiro. A curva contínua representa a função gaussiana ajustada.

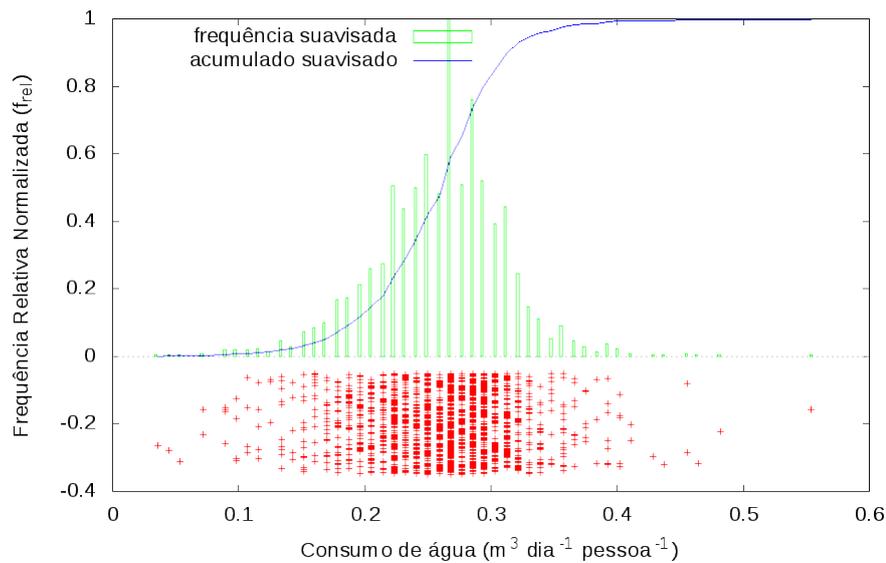
### Interpolação horária dos dados de consumo

Em geral, uma análise estatística, em particular uma análise de rede neural necessita que os dados estejam ajustados. Desde que os dados meteorológicos apresentam amostragem horária e os dados de consumo apresentam frequência diária foi realizada a harmonização aplicando-se o interpolador de Barnes a todos os dados, obtendo uma série horária de dados ajustados.

A Figura 3.27 mostra um pequeno trecho dos dados interpolados de consumo de água correspondente à primeira semana de 2013. Nota-se que a linha contínua de



**Figura 3.25:** Gráfico com ajuste de polinômio de 6ª ordem do consumo diário de água por pessoa residente em um prédio residencial em Laranjeiras, Rio de Janeiro.

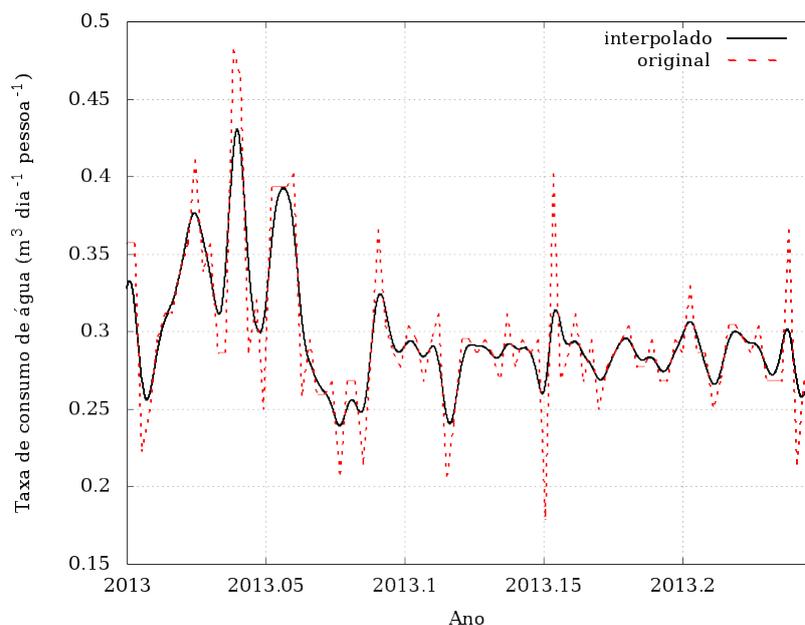


**Figura 3.26:** Gráfico das funções de distribuição probabilidade e acumulada (f.d.p. e f.d.a.) por pessoa residente em um prédio residencial em Laranjeiras, Rio de Janeiro.

interpolação acompanha de forma suavizada às variações temporais do dado original.

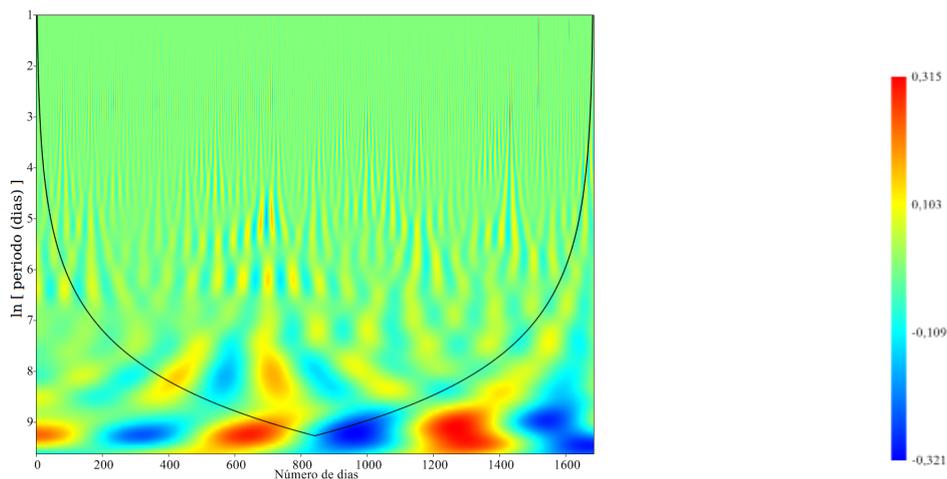
### 3.4 Transformada de Ondeleta do consumo de água (prédio residencial)

A Figura 3.28 mostra a evolução temporal da parte Real da transformada de ondeleta da taxa de consumo de água do prédio residencial.



**Figura 3.27:** Taxa de consumo de água interpolada e original para a primeira semana do ano 2013.

A TO distribui o sinal de acordo com escalas. No caso da variação temporal, as escalas são períodos. Na Figura 3.28, o eixo x é o tempo e o eixo  $y = \ln(\text{período})$ , com unidade do período em (dias), pois as amostras são diárias. O valor da TO é indicada pela escala de cores.

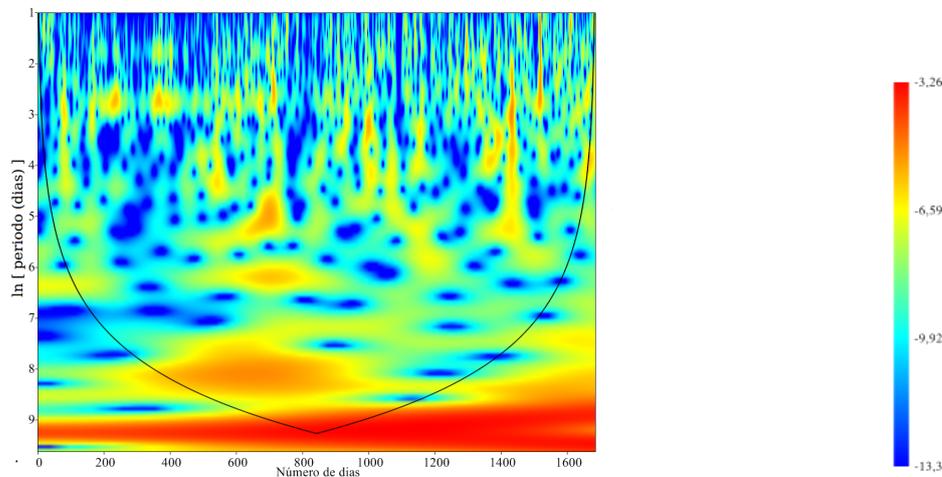


**Figura 3.28:** Evolução temporal da parte real da transformada de ondeleta da taxa de consumo de água do prédio residencial.

O que se nota na TO é que a variável apresenta uma distribuição de múltiplos períodos apresentando-se como um ruído vermelho, i.e., com sinal mais intenso nas baixas frequências ou períodos altos. O sinal é observado em todas as frequências mas o sinal é constantemente observado nas baixas frequências. A Figura 3.29 mos-

tra a evolução temporal da potência (i.e., quadrado do módulo) da transformada de ondeleta da taxa de consumo de água do prédio residencial.

O módulo da TO (i.e., a potência) indica que existem variações sazonais importantes, sendo que alguns meses ocorrem grandes aumentos do consumo, mas sem que observe-se uma regularidade sazonal (dentro do cone de significância). Talvez ocorra alguma regularidade para períodos de expoente 3,5 e 6,5, correspondentes à valores com período de 1 mês ( $\approx e^{3.5}$ ) e 2 anos ( $\approx e^{6.5}$ ). Isso pode ter associação com ciclos mensais e interanuais do consumo de água.



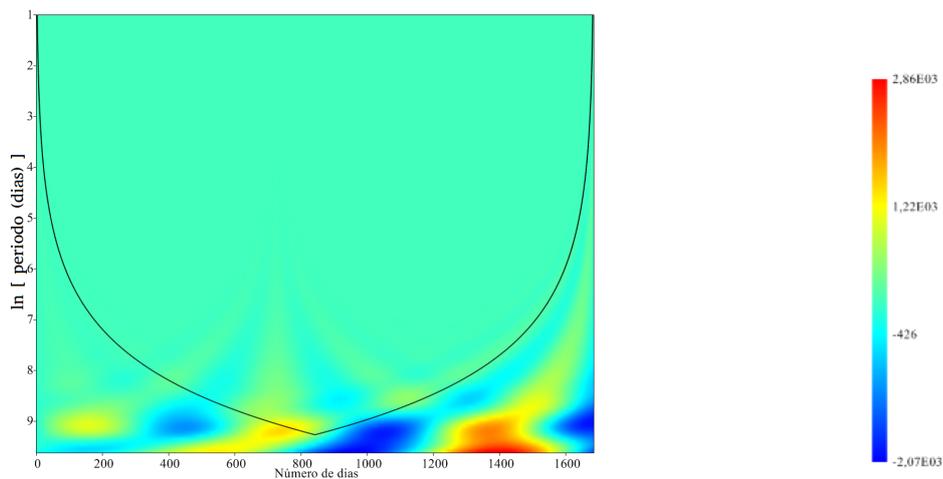
**Figura 3.29:** Evolução temporal da potência da transformada de ondeleta da taxa de consumo de água do prédio residencial.

A Figura 3.30 mostra a evolução temporal da parte Real da transformada de ondeleta da taxa de consumo ACUMULADO de água do prédio residencial. A TO do consumo acumulado mostra-se concentrada em baixas frequências, ou períodos muito longos. Isto indica que o sinal do consumo acumulado apresenta características de *ruído vermelho*, ou seja, de um sinal que dominam variações interanual.

A Figura 3.31 mostra a evolução temporal da potência (i.e., quadrado do módulo) da transformada de ondeleta da taxa de consumo ACUMULADO de água do prédio residencial.

Pode-se entender o domínio de baixas frequências como efeito das variações interanuais associadas a flutuação do clima que afetam a oferta e distribuição de água. Por exemplo, na tubulação do sistema de água pode alterar o consumo de água mesmo se o usuário mantiver a torneira aberta pelo mesmo período. Por hipótese, nos anos em que a precipitação é menor, a distribuição é utilizada com uma pressão de distribuição menor, que resulta em uma redução do consumo pela população.

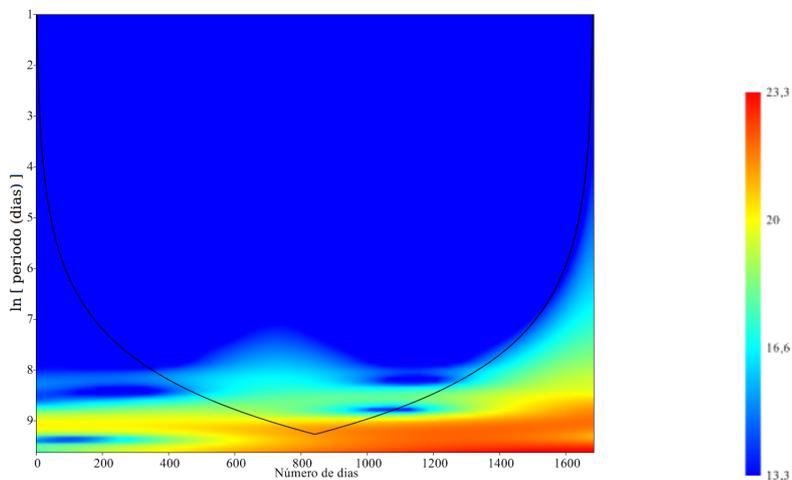
O caráter quase aleatório da distribuição de frequência do sinal da TO também está a indicar que os efeitos de variação diária (alta frequência) das variáveis meteoroló-



**Figura 3.30:** *Evolução temporal da parte real da transformada de ondeleta da taxa de consumo ACUMULADO de água do prédio residencial.*

gicas devem ser pequenos.

As características dos sinais cônicos com base na baixa frequência e vértices na alta frequência (fusos) indicam autocorrelação do sinal.



**Figura 3.31:** *Evolução temporal da potência da transformada de ondeleta da taxa de consumo ACUMULADO de água do prédio residencial.*

### 3.5 Rede Neural Artificial

As variáveis meteorológicas temperatura, umidade relativa, pressão atmosférica e vento foram incluídas como entradas da RNA. A variável de saída da RNA corresponde ao dado de consumo de água, que nesse caso com  $\delta t = 24 \text{ horas}$ . A partir do consumo diário de água em uma residência e de um prédio foi possível um treinamento que corresponde ao ajuste dos parâmetros das equações da rede. Além disso, foi possível verificar a qualidade do modelo em relação as observações.

O script Octave que implementa a rede de aprendizado utilizou palavras de máquina de 16 bits para representar as variáveis como números inteiros. Isto é, por imposição da biblioteca NNT utilizada. Uma implicação do software de 16 bits é que o número de dados de cada etapa de treinamento é limitada a 32 mil amostras.

A RNA apresentada nesse trabalho tem como estrutura: de duas a onze entradas, uma camada intermediária com 40 nós internos e uma saída. As diferentes redes empregadas nesse trabalho são listadas na Tabela 3.1. A variação temporal foi dividida em duas partes, a primeira dedicada aos treinamentos e verificação dos erros, a segunda para avaliação final. Dez treinamentos com testes cruzados foram realizados. Os resultados obtidos serão apresentados a seguir.

**Tabela 3.1:** Estruturas dos modelos para previsão de consumo de água residencial na Região Metropolitana do Rio de Janeiro usando consumos de água anteriores e variáveis meteorológicas.

Modelos	Entradas
M1	t, vel, p, ur, w-1
M2	t, vel, p, ur, w-7
M3	ano, mes, nascer, por, t, vel, u, v, p, ur, w-1
M4	w-2, w-1
M5	w-7, w-6, w-5, w-4, w-3, w-2, w-1
M6	t, vel, p, ur
M7	vel, u, v, p, ur
M8	t, ur

#### Modelo M1 (t, vel, p, ur e w-1)

O modelo M1 é composto por cinco entradas que tem como variáveis: temperatura do ar (t), velocidade do vento (vel), pressão atmosférica (p), umidade relativa (ur) e consumo de água do dia anterior (w-1). Como foi dito anteriormente, 10 testes de validação cruzada foram realizados com essa rede. Somente os três primeiros testes da validação serão discutidos a seguir.

A etapa de treinamento do modelo M1 é mostrada na Figura 3.32. Nela temos as etapas de treinamento 1, 2 e 3. Pode-se observar que o modelo segue a série observada com BIAS zero associado a otimização, ora superestimando na etapa 1, ora

subestimando na etapa 3. O mesmo comportamento volta a aparecer nas demais etapas.

O histograma das etapas 1, 2 e 3 de treinamento do modelo M1 (t, vel, p, ur e w-1) é visto na Figura 3.33. Os histogramas confirmam a tendência de um erro sistemático pequeno, decorrente do processo de otimização por mínimos quadrados da rede de aprendizado.

Gráficos Q-Q das etapas 1, 2 e 3 de treinamento do modelo M1 (t, vel, p, ur e w-1) são apresentados na Figura 3.34. A característica que se destaca é a tendência de um erro sistemático nulo considerando a distribuição dos quantis.

Os três primeiros testes de validação cruzada do modelo M1 são mostrados na Figura 3.35. A etapa de teste não mostra uma precisão tão grande quanto a etapa de validação como pode ser observado nos detalhes da variação da série modelada em relação a observada, como esperado em um teste de validação cruzada. Em particular, a simulação gera oscilações mais intensas e frequentes do que presentes na observação. Provavelmente, isso deve ser uma resposta para determinar a dinâmica das variáveis meteorológicas de superfície.

O histograma modelado mostra-se mais compacto do que os histogramas observados, não definindo os valores de moda observados (Figura 3.36).

Gráficos Q-Q do teste 1, 2 e 3 do modelo M1 são vistos na Figura 3.37. Os gráficos evidenciam a dificuldade do modelo em simular os extremos da distribuição.

A etapa de validação do modelo M1 (t, vel, p, ur e w-1) é apresentada na Figura 3.38 com as etapas de treinamento da validação, resultado da validação e validação com o intervalo de confiança. Nessa figura, é possível observar as séries temporais no treinamento da validação aplicada à metade dos dados. O resultado da validação é o conjunto de coeficientes do modelo M1. Os MSE médios obtidos durante as K etapas da validação cruzada, é empregado para estimar o desvio padrão do modelo, i.e.,  $\pm\sigma$ . O desvio padrão fornece uma ideia da escala de variabilidade local do modelo em torno da resposta média.

Para avaliação do intervalo de confiança partimos da suposição de que o erro do modelo em relação à observação pode ser aproximado por uma distribuição Chi-quadrado com intervalo de confiança de 90% ( $\sigma_{90\%}$ ) (área cinzenta na figura).

Para entender melhor a distribuição do modelo e observação apresenta-se a seguir o gráfico Quantil-Quantil ( $Q-Q$ ) (Figura 3.40) e os histograma de frequência obtidos na etapa de validação (Figura 3.39).

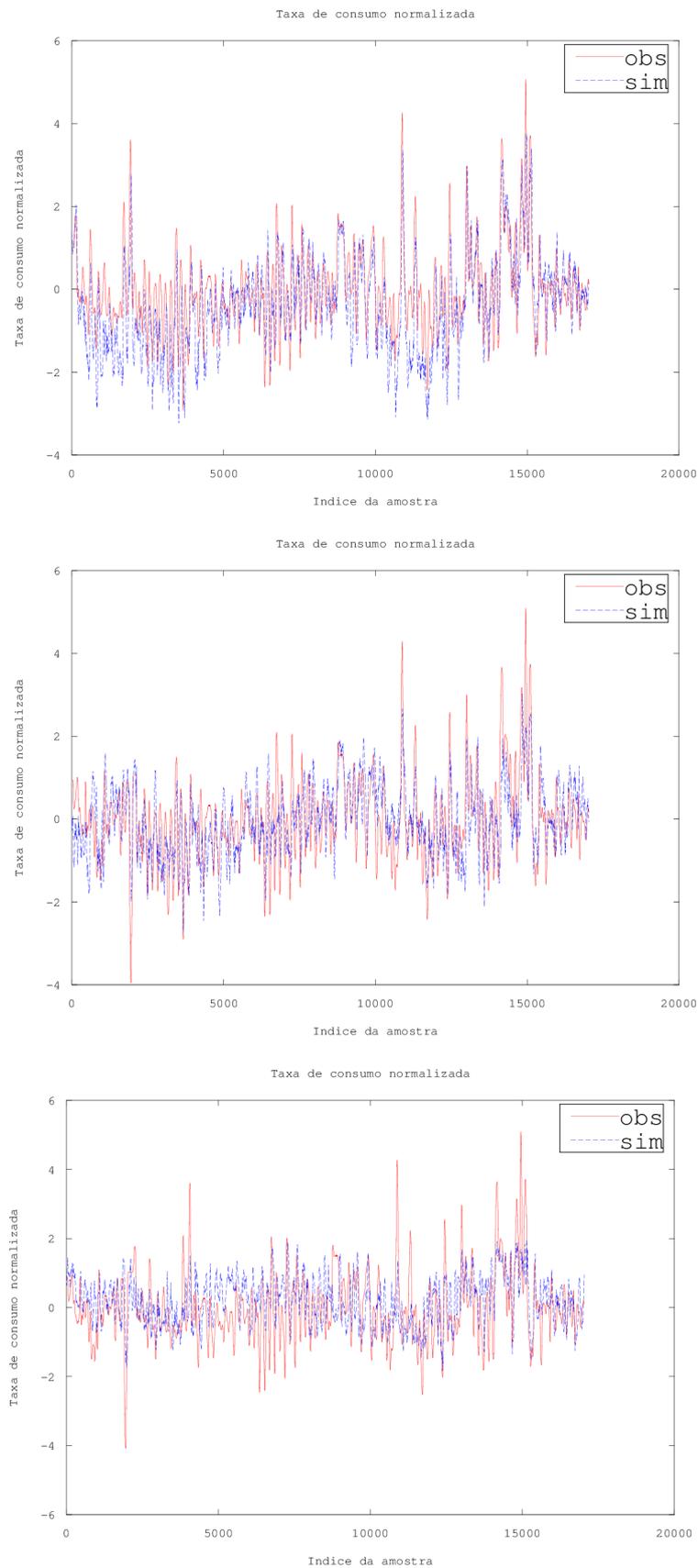


Figura 3.32: Etapas 1, 2 e 3 de treinamento do modelo  $M1$  ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).

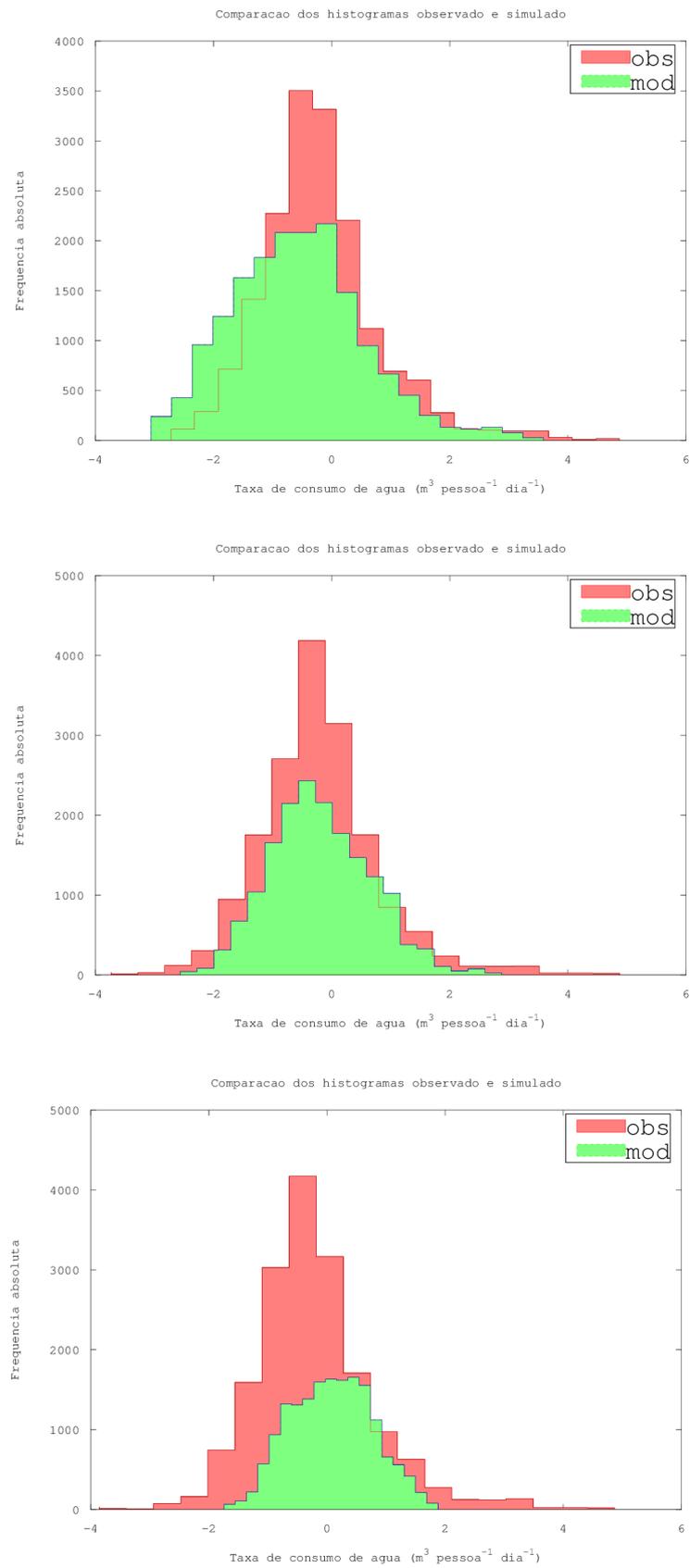
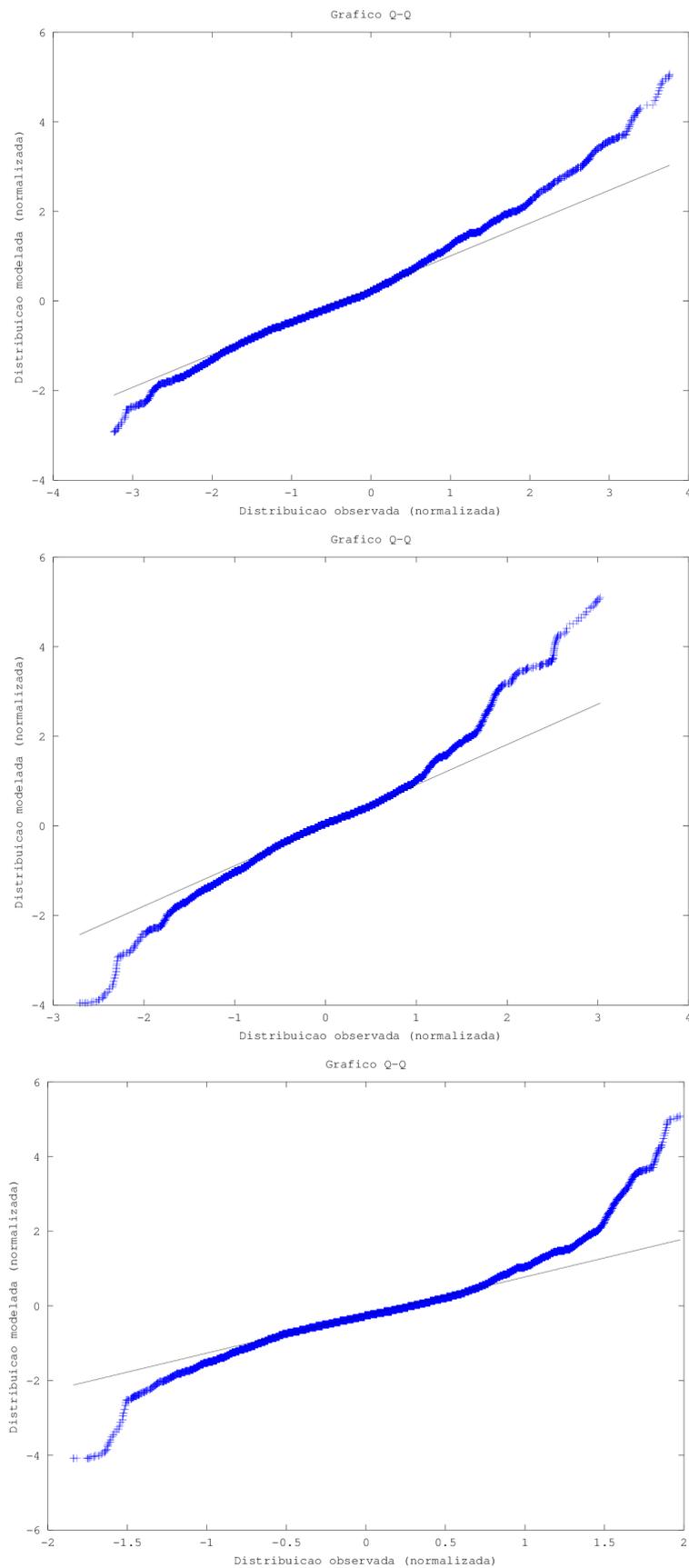
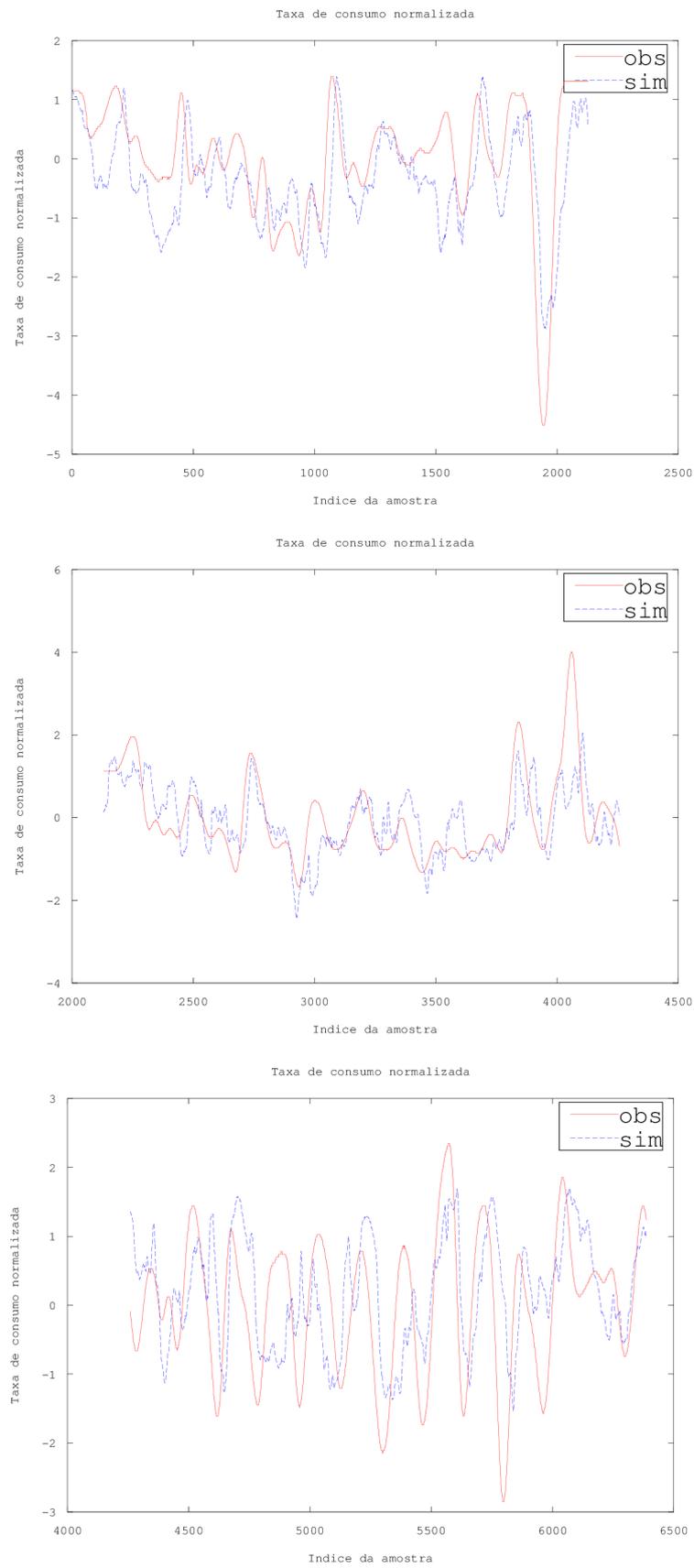


Figura 3.33: Histogramas das etapas 1, 2 e 3 de treinamento do modelo  $M1$  ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).



**Figura 3.34:** Gráfico Q-Q das etapas 1, 2 e 3 de treinamento do modelo M1 ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).



**Figura 3.35:** *Etapas 1, 2 e 3 do teste do modelo  $M1$  ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).*

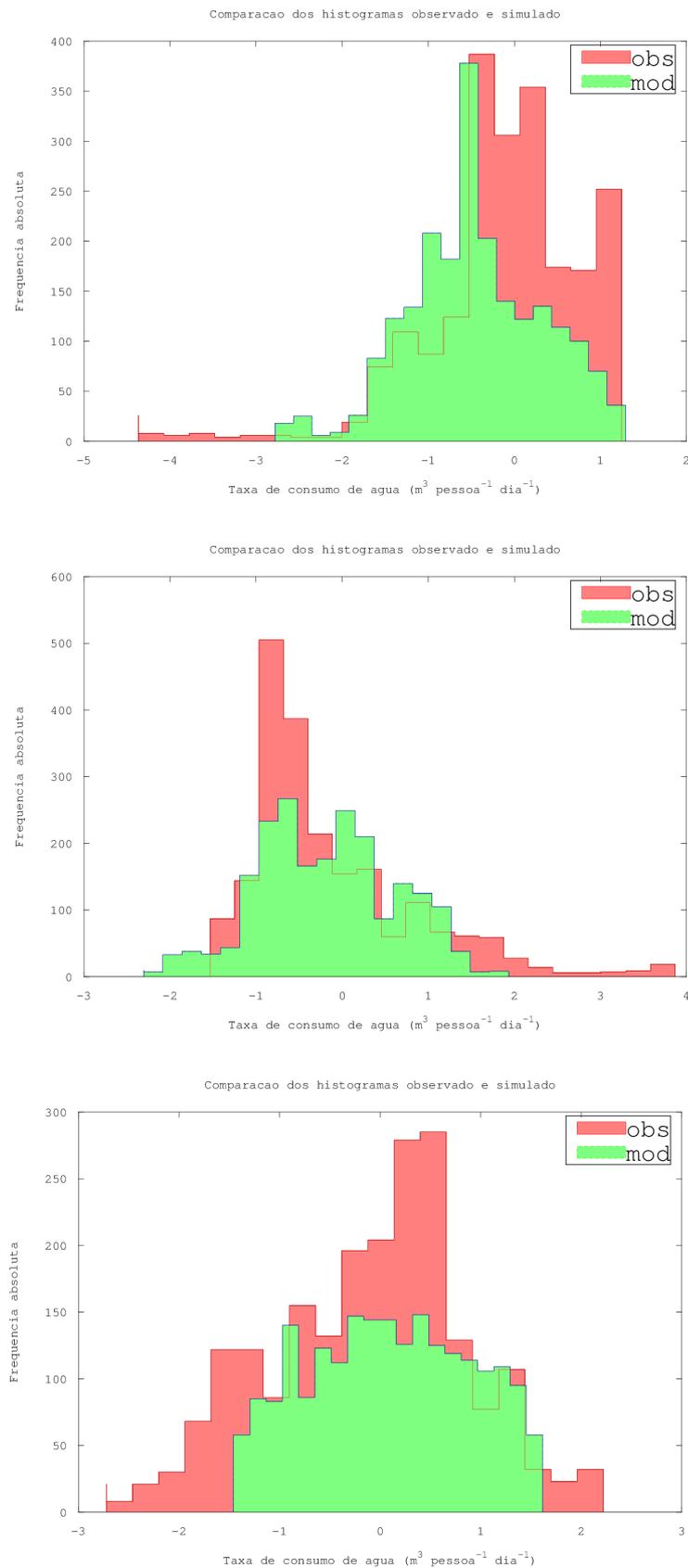


Figura 3.36: Histogramas das etapas 1, 2 e 3 do teste do modelo M1 ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).

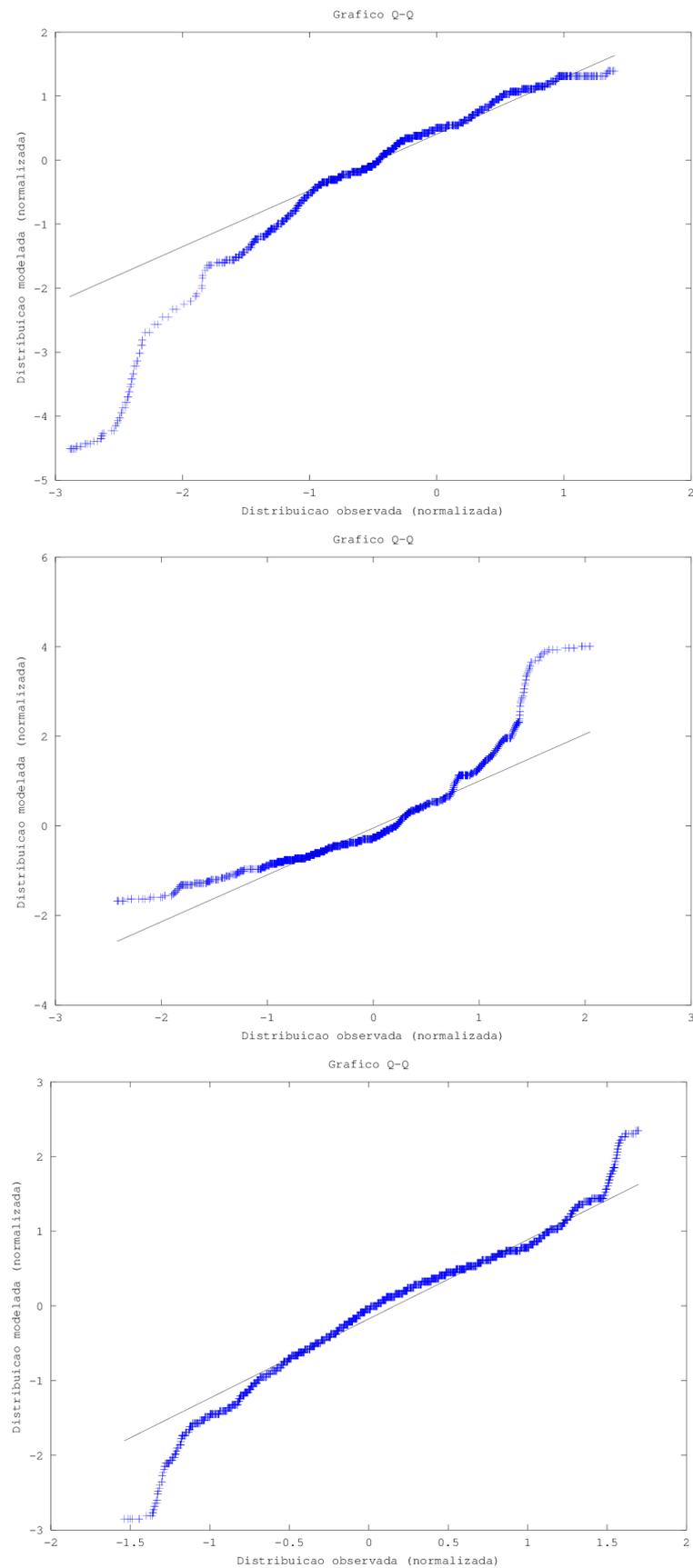


Figura 3.37: Gráficos Q-Q das etapas 1, 2 e 3 do teste do modelo  $M1$  ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).

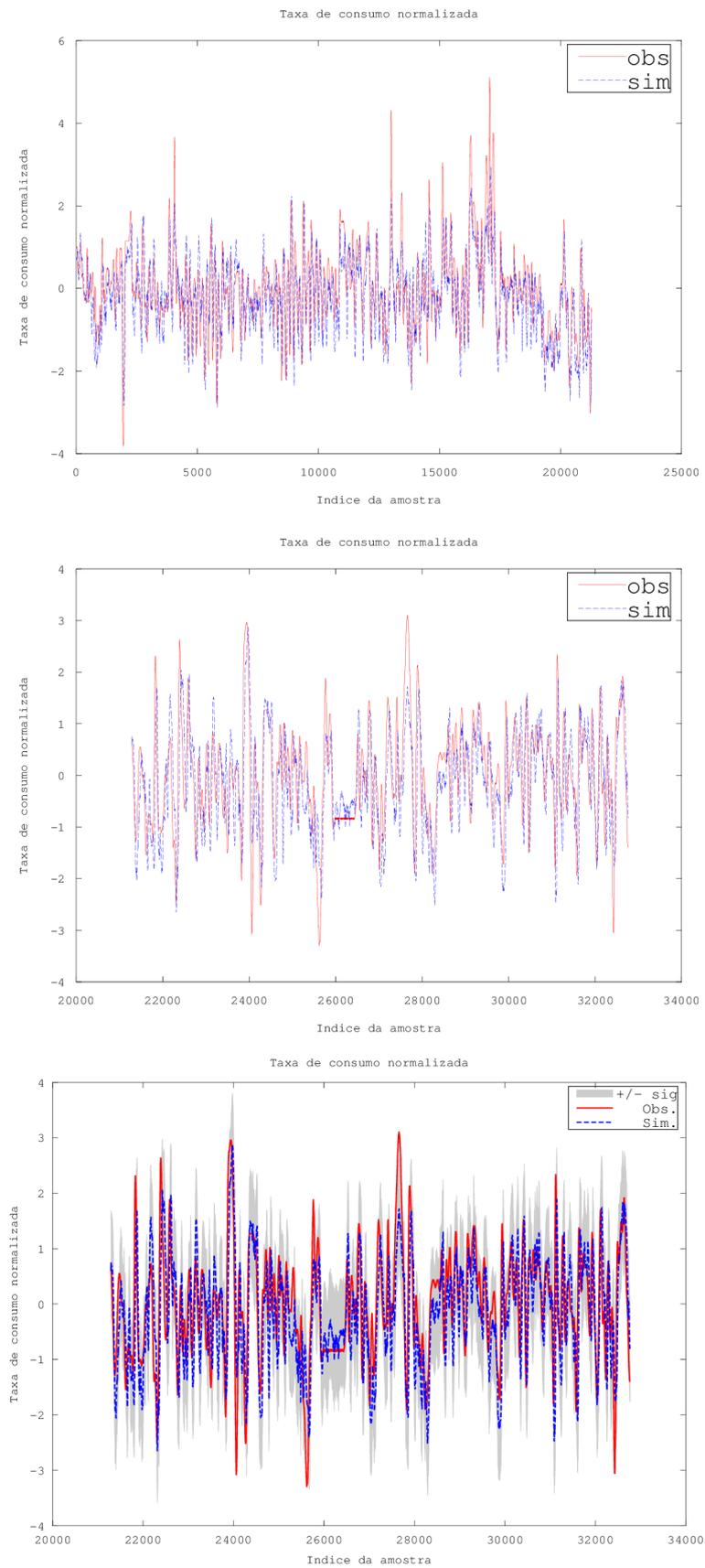
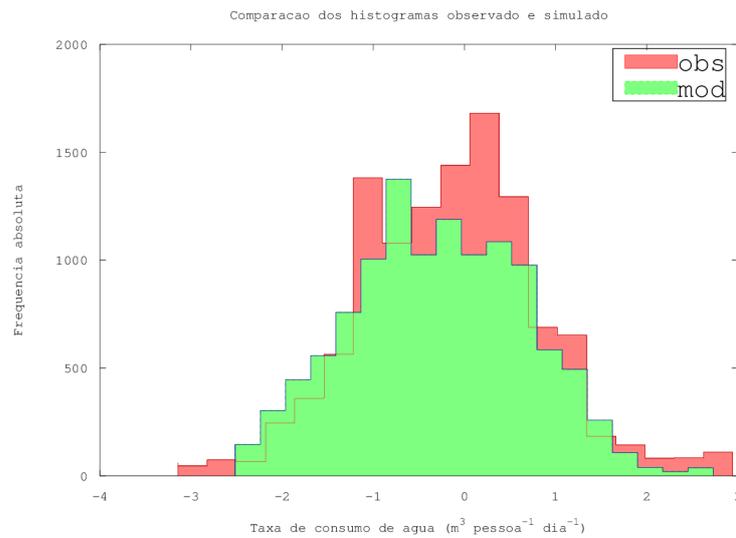
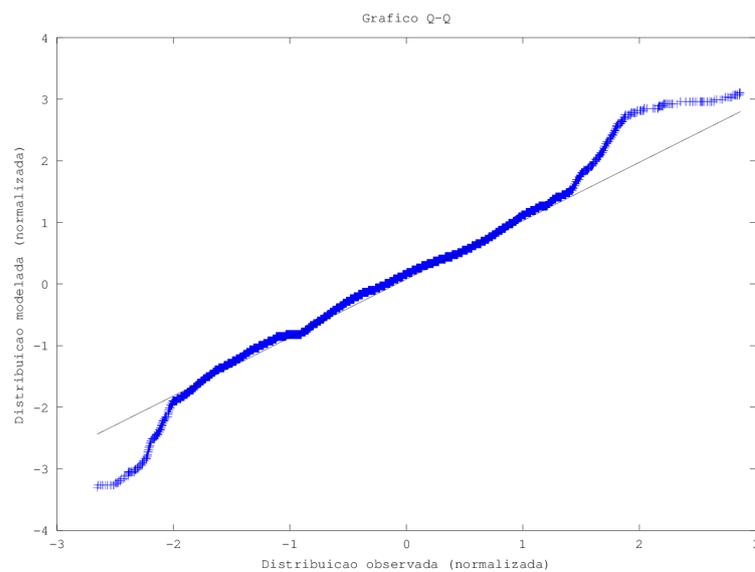


Figura 3.38: Etapas de validação cruzada do modelo  $M1$  ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).



**Figura 3.39:** Histograma da validação cruzada do modelo  $M1$  ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).



**Figura 3.40:** Gráfico  $Q-Q$  da validação cruzada do modelo  $M1$  ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).

### Modelo $M2$ ( $t$ , $p$ , $ur$ , $vel$ , $w-7$ )

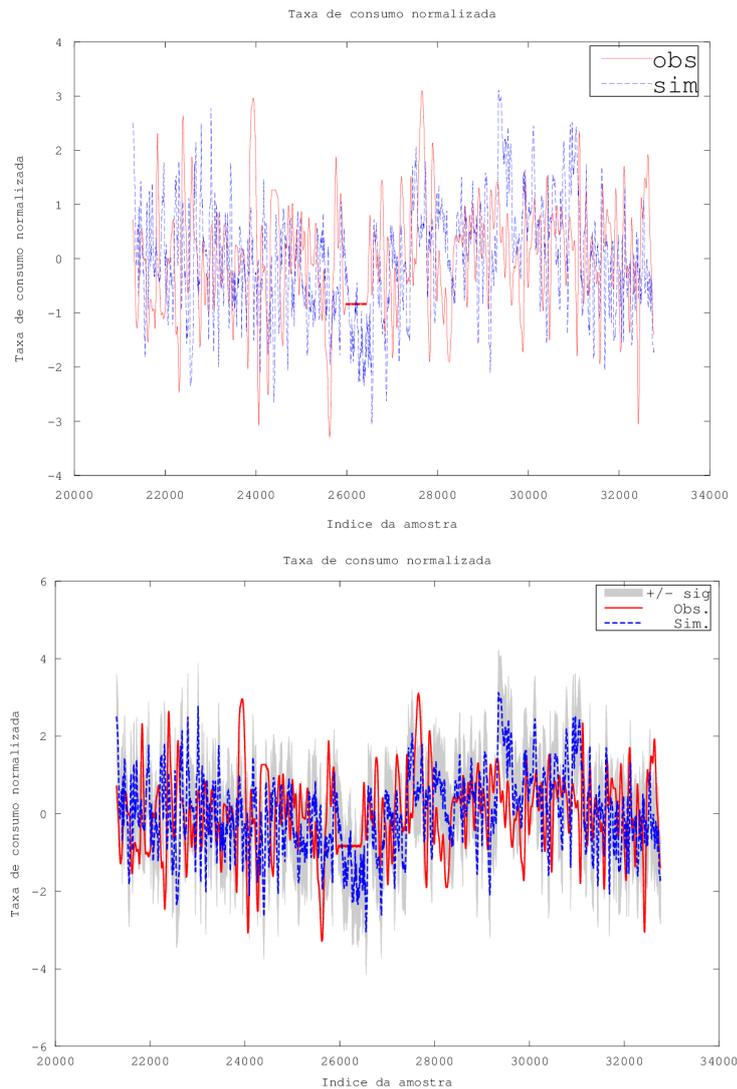
O modelo  $M2$  ( $t$ ,  $p$ ,  $ur$ ,  $vel$ ,  $w-7$ ) é similar ao modelo  $M1$ , sendo que a única diferença é a substituição do consumo anterior (denotado  $w-1$ ) pelo consumo de 7 dias antes (i.e.,  $w-7$ ). Portanto, pode ser caracterizado como um modelo de fraca dependência da persistência.

Serão mostrados apenas os resultados da validação do modelo  $M2$  (Figura 3.41), além de seu histograma na etapa de validação (Figura 3.42).

O modelo  $M2$  acompanha o movimento da média do consumo, o que é interessante para a finalidade da previsão. Entretanto, as oscilações do modelo são mais intensas

que as observadas, gerando superestimativas e subestimativas dos máximos e mínimos observados devido a fraca dependência da persistência.

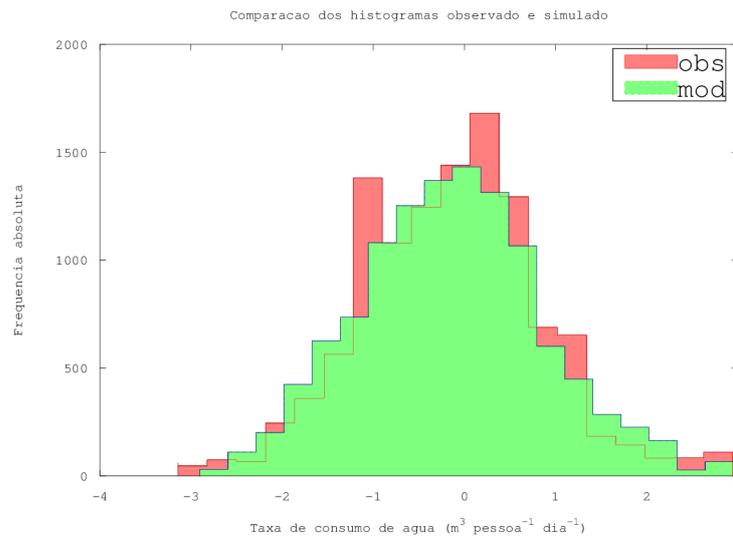
A semelhança dos histogramas modelado e observado reforça a característica do modelo M2 de acompanhar razoavelmente a média, apesar do pequeno papel da persistência usada na entrada. A correlação do consumo em relação ao observado na semana anterior é pequena, mas suficiente para ancorar bem o valor médio, mas não a amplitude da oscilação.



**Figura 3.41:** Etapas de validação cruzada do modelo M2 ( $t$ ,  $p$ ,  $ur$ ,  $vel$ ,  $w-7$ ).

### Modelo M3 (ano, mês, nascer, pôr, $t$ , $p$ , $ur$ , $vel$ , $u$ , $v$ , $w-1$ )

No modelo M3 aplicou-se 11 variáveis de entradas que são: ano, mês, hora do nascer, hora do pôr do Sol, temperatura, velocidade do vento, componente zonal do vento, componente meridional do vento, pressão atmosférica, umidade relativa e consumo



**Figura 3.42:** Histograma da validação cruzada do modelo M2 ( $t$ ,  $p$ ,  $ur$ ,  $vel$ ,  $w-7$ ).

de água do dia anterior (i.e., incluindo persistência).

Resultados obtidos para a etapa de validação são apresentados nas Figuras 3.43 e 3.44.

O modelo M3 tem um desempenho global médio, i.e., fornece o valor médio de longo período corretamente. Entretanto, apresentou superestimativa em aproximadamente metade do período de validação (período inicial) e subestimativa na segunda metade do período de validação.

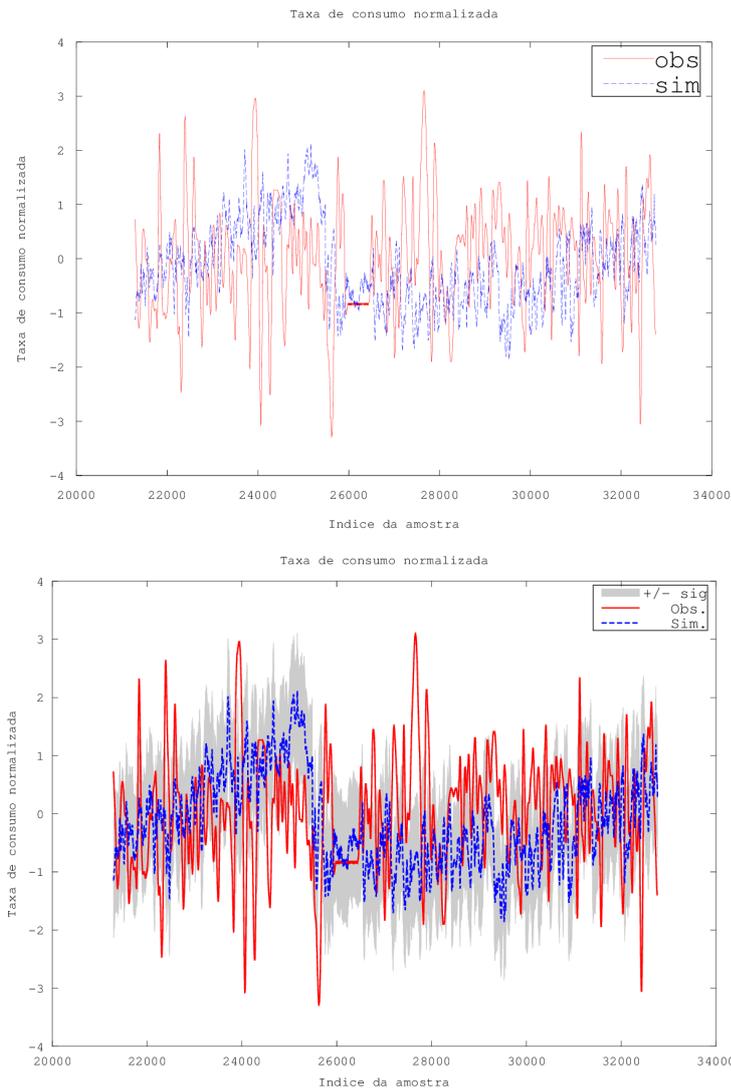
O histograma modelado não mostra o caráter bimodal do histograma observado, sendo monomodal. O característica bimodal observada se deve a diferença de consumo registrada nos últimos anos da série (2014 a 2015), i.e., um registro de redução do consumo, provavelmente associado à percepção da estiagem ocorrida no período 2013-2014 no SE do Brasil, particularmente em SP, onde se encontra as nascentes do rio Paraíba do Sul.

### Modelo M4 (w-1, w-2)

No modelo M4 (w-1, w-2) é um modelo de persistência de dois dias. As duas variáveis de entradas são, primeira o consumo de água do dia anterior e, a segunda, o consumo de água observado dois dias antes da data de previsão.

Resultados obtidos para a etapa de validação são apresentados nas Figuras 3.45 e 3.46.

Esse modelo apresenta uma boa resposta, verificadas a modelagem da média e a amplitude das oscilações do consumo ao longo do período de validação.



**Figura 3.43:** Etapas de validação cruzada do modelo  $M3$  (ano, mês, nascer, pôr,  $t$ ,  $p$ ,  $ur$ ,  $vel$ ,  $u$ ,  $v$ ,  $w-1$ ).

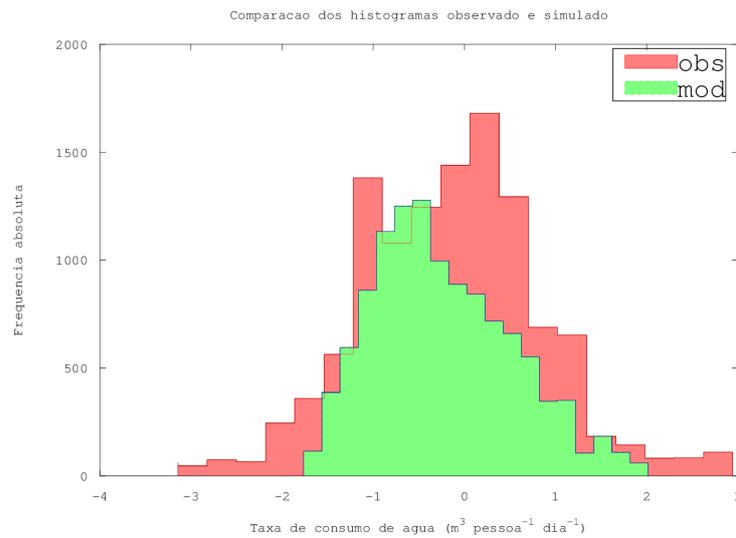
Embora, seja um dos modelos de melhor resposta obtidos neste trabalho, não reproduz o caráter bimodal do histograma de frequências.

### Modelo $M5$ (w-1, w-2, w-3, w-4, w-5, w-6, w-7)

No modelo  $M5$  (w-1, w-2, w-3, w-4, w-5, w-6, w-7) é um modelo de persistência de 7 dias. As 7 variáveis de entradas são os consumos de água dos 7 dias anteriores.

Resultados obtidos para a etapa de validação são apresentados nas Figuras 3.47 e 3.48.

O modelo  $M5$  (com 7 dias de persistência) apresenta as mesmas características des-



**Figura 3.44:** Histograma da validação cruzada do modelo M3 (ano, mês, nascer, pôr, t, p, ur, vel, u, v, w-1).

tacadas para o modelo M4 (com 2 dias de persistência), simulando de forma mais precisa a moda principal observada no histograma de frequências.

### Modelo M6 (t, vel, p, ur)

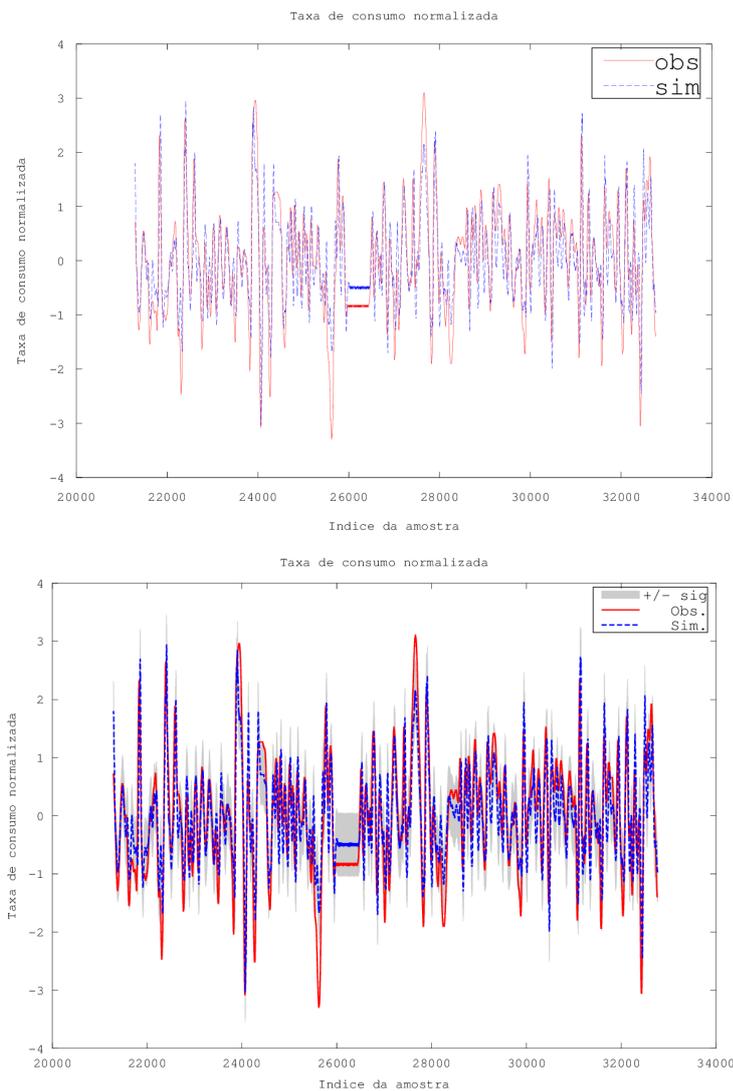
No modelo M6 (t, vel, p, ur) é um modelo de 4 variáveis de entrada, todas meteorológicas. Portanto, sem persistência. As 4 variáveis de entrada são temperatura do ar a superfície, velocidade do vento, pressão atmosférica e umidade relativa do ar.

Resultados obtidos para a etapa de validação são apresentados nas Figuras 3.49 e 3.50. O modelo M6 em função apenas de 4 variáveis meteorológicas não é capaz de prever adequadamente a amplitude das oscilações do consumo diário, permanecendo em torno da mediana global como visto no histograma.

As variações intra-anuais (i.e., oscilações de longo período) também não são simuladas apropriadamente pelo modelo M6.

### Modelo M7 (vel, u, v, p, ur)

No modelo M7 é um modelo de 5 variáveis de entrada que correspondem a velocidade do vento (vel), componente zonal do vento (u), componente meridional do vento (v), pressão atmosférica (p) e umidade relativa (ur). Todas essas variáveis são meteorológicas, no qual destaca-se a ausência da temperatura entre as variáveis de entrada. Outra característica escolhida é a ausência de persistência.

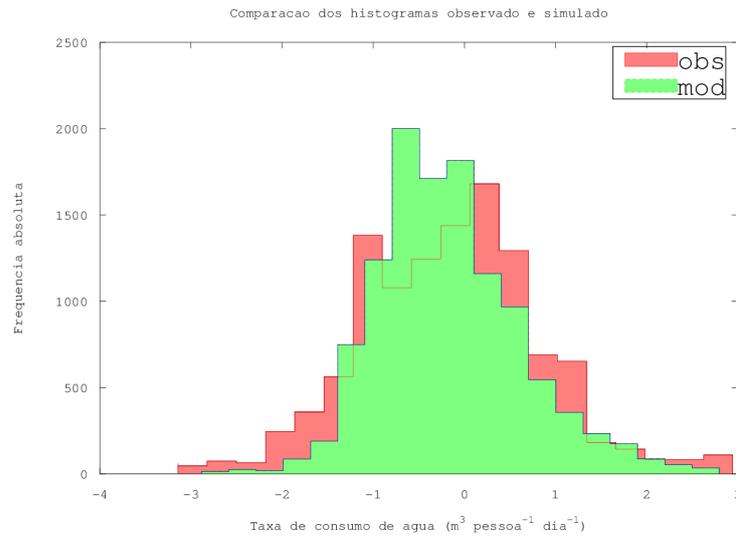


**Figura 3.45:** *Etapas de validação cruzada do modelo  $M_4$  ( $w-1$ ,  $w-2$ ).*

Resultados obtidos para a etapa de validação são apresentados nas Figuras 3.51 e 3.52.

O modelo  $M_7$  apresenta um desempenho global melhor que o modelo  $M_6$  em relação à amplitude das oscilações do consumo diário. Este modelo também foi escrito somente em função de variáveis meteorológicas, para as quais foi feito o acréscimo das componentes zonal e meridional do vento.

Considerou-se no design do modelo  $M_7$ , o papel relevante das componentes zonal e meridional do vento para o entendimento da meteorologia urbana do Rio de Janeiro. Em particular, a oscilação diária do vento, evidenciada pela oscilação Sul-Norte da componente meridional do vento, ocorre devido à circulação de brisa marítima-terrestre ao longo da linha de retorte da costa litorânea do RJ. Argumenta-se que a oscilação do vento em conjunto com a dinâmica da camada limite urbana possa explicar o fenômeno característico da Ilha de Calor Urbana da RMRJ, juntamente com os efeitos mecânico e térmico da topografia local e regional.



**Figura 3.46:** Histograma da validação cruzada do modelo  $M_4$  ( $w-1$ ,  $w-2$ ).

### Modelo M8 (t, ur)

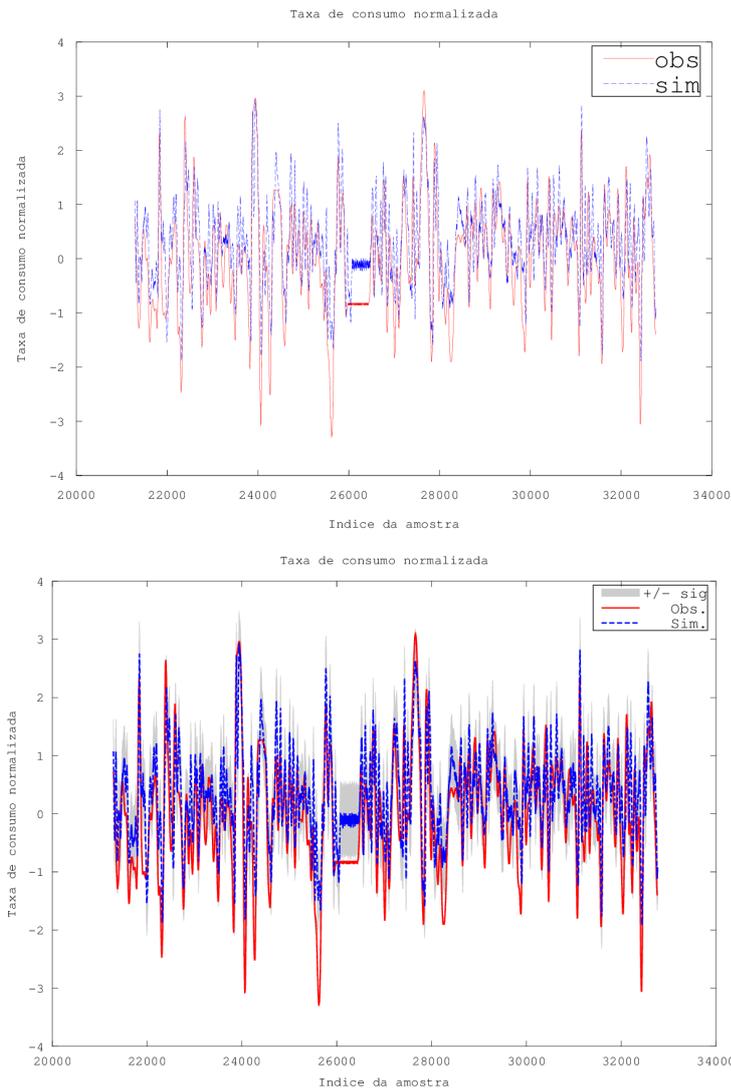
No modelo M8 (t, ur) é um modelo de apenas 2 variáveis de entrada meteorológicas que são temperatura e umidade relativa do ar. Com apenas duas variáveis meteorológicas como entrada este modelo é um dos mais compactos entre os modelos propostos.

Resultados obtidos para a etapa de validação são apresentados nas Figuras 3.53 e 3.54. Como em outros modelos apresentados anteriormente, esse modelo acompanha a média local e subestima a amplitude das oscilações. Sua principal vantagem, além da simplicidade, é a capacidade de simular o caráter bimodal observado no histograma de frequências.

## 3.6 Desempenho estatístico dos modelos

Os desempenhos dos modelos descritos anteriormente podem ser avaliados usando testes estatísticos que descrevem os erros associados com o modelo. Os índices estatísticos aplicados a todos os modelos estão descritos nas tabelas 3.2, 3.3 e 3.4.

A análise conjunta dos índices de avaliação do desempenho dos modelos testados na rede de aprendizado, indica que o melhor desempenho foi do modelo M5 ( $w-1$ ,  $w-2$ ,  $w-3$ ,  $w-4$ ,  $w-5$ ,  $w-6$ ,  $w-7$ ), seguido pelos modelos M4 ( $w-1$ ,  $w-2$ ) e M1 (t, vel, p, ur,  $w-1$ ).



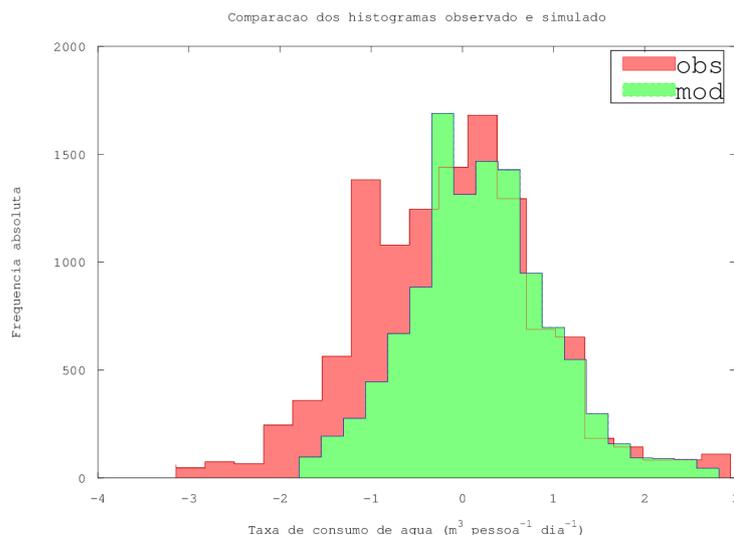
**Figura 3.47:** Etapas de validação cruzada do modelo M5 ( $w-1$ ,  $w-2$ ,  $w-3$ ,  $w-4$ ,  $w-5$ ,  $w-6$ ,  $w-7$ ).

O índice de persistência ( $SSA$ ) é maior para os modelos M5 ( $w-1$ ,  $w-2$ ,  $w-3$ ,  $w-4$ ,  $w-5$ ,  $w-6$ ,  $w-7$ ) e M4 ( $w-1$ ,  $w-2$ ), como esperado. Entretanto, a consideração de variáveis de entrada meteorológicas (como no modelo M1) é importante pois implica em redução do erro sistemático ( $BIAS$ ), importante do ponto de vista operacional.

O modelo M4 (i.e., modelo de persistência de 2 dias) alcançou o maior valor da probabilidade de detecção ( $POD$ ) dentre todos os modelos, com desvio padrão igual a 0,88 e também alcançou o menor índice de falso alarme ( $FAR$ ).

O erro quadrático do modelo de referência (i.e., considerando persistência de 24 ( $h$ ), ( $MSE_{ref}$ ), apresenta valor aproximado de 0,478. Este valor é menor que o valor correspondente aos  $MSE$  obtidos nos K-testes.

Os índices  $BIAS$  e  $POD_{chance}$  apresentam um menor erro para o modelo M1 mostrando a importância da inclusão de variáveis meteorológicas na entrada da rede de aprendizado para a previsão do consumo de água.



**Figura 3.48:** Histograma da validação cruzada do modelo M5 ( $w-1$ ,  $w-2$ ,  $w-3$ ,  $w-4$ ,  $w-5$ ,  $w-6$ ,  $w-7$ ).

A comparação das métricas estatísticas dos modelos, particularmente do índice  $POD_{chance}$ , em relação ao modelo M1, indica que a inclusão da meteorologia aumenta o caráter determinístico do modelo de simulação. Por outro lado, o caráter estocástico é destacado pela persistência e homogeneidade estatística (i.e., pela invariância por translação).

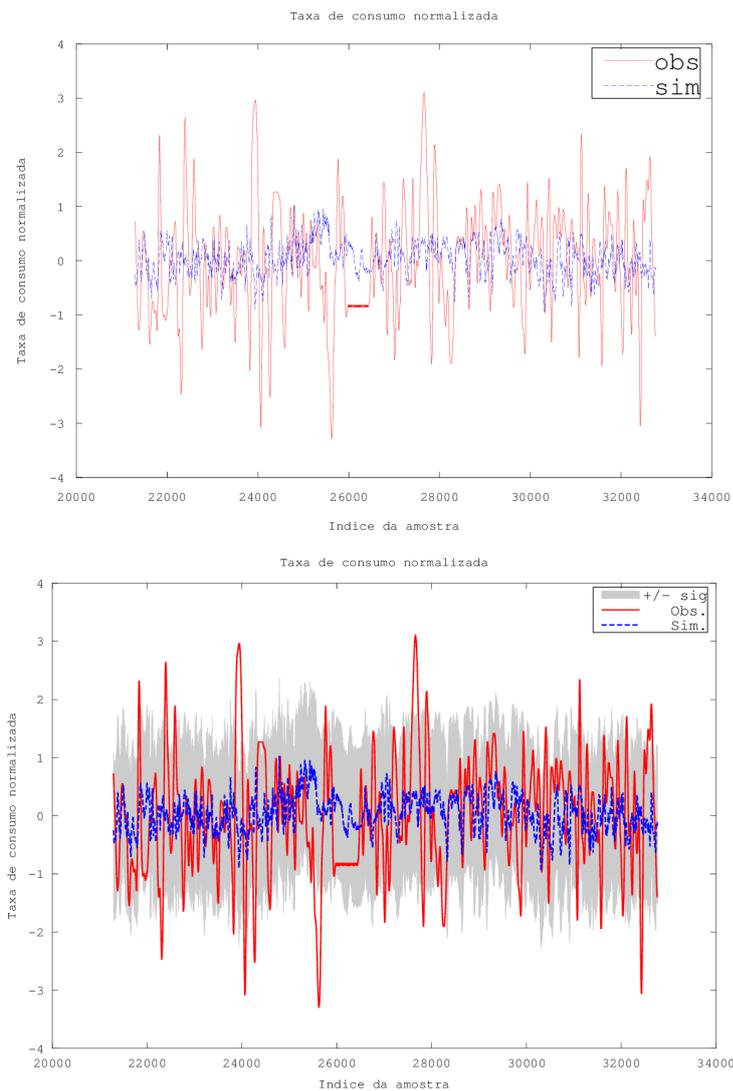
## Tabelas de contingência

A tabela de contingência ( $TC$ ) foi aplicada para todos os modelos. Entretanto, apenas as tabelas do modelo M1 e M4 serão mostradas pois são as que tiveram um melhor resultado nos índices estatísticos, como visto anteriormente.

Para uma descrição mais detalhada, foi gerada uma tabela de contingência categórica com 6 categorias (i.e., por 6 classes definidas pela distribuição de quantis observados) (Figura 3.55 para modelo M1 e Figura 3.56 para modelo 4).

Os resultados de contagem de frequência absoluta das 6 classes categóricas foram reagrupados em tabelas de 2 classes (usando a mediana ou quantil de 50% como critério de reagrupamento), apresentadas nas Figuras 3.57) e 3.58).

O que se infere das  $TC$ s é a menor variância do modelo em relação às observações. Isto gera alguma confusão estatística (i.e., um efeito se confunde com o outro) junto à diagonal de aferição (i.e., a segunda diagonal no caso das  $TC$  apresentadas), em relação às categorias associadas aos quantis menores e maiores. No caso ideal, todos os elementos ao longo da diagonal secundária das figuras da  $TC$ , deveriam mostrar



**Figura 3.49:** *Etapas de validação cruzada do modelo M6 ( $t$ ,  $vel$ ,  $p$ ,  $ur$ ).*

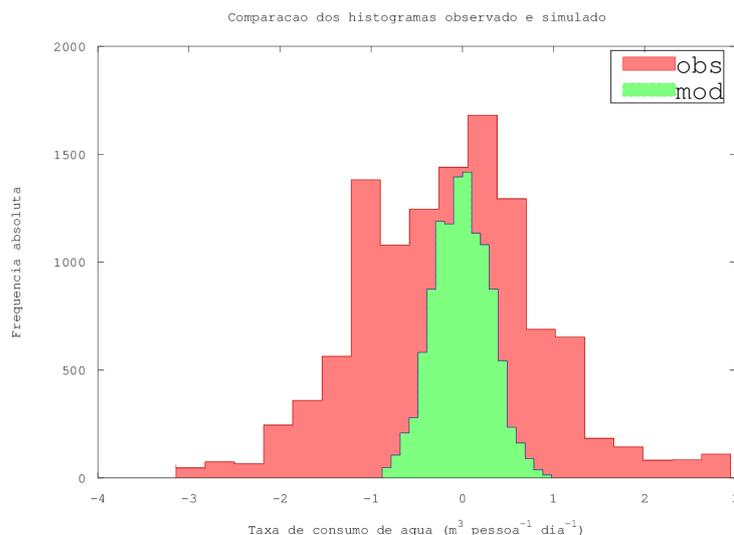
a cor cinza escura, usada para indicar alta frequência de contagem categórica.

## Matriz de correlação

A dispersão de pontos associada aos modelos M1 e M4 é mostrada nas Figuras 3.59 e 3.60, respectivamente. Observa-se a forma elíptica da distribuição da dispersão de pontos.

A elipse encontrada no centro do gráfico mostra a plotagem de raios menores e maior associação ao grau de confiabilidade estatística de 90%. Os autovetores da matriz de covariância ( $COV$ ) entre observação e simulação também foram representados.

Estas figura destacam o papel da autocorrelação na explicação do consumo, assim como o papel das variáveis meteorológicas para a dispersão dos pontos, ou seja, para



**Figura 3.50:** Histograma da validação cruzada do modelo M6 ( $t$ ,  $vel$ ,  $p$ ,  $ur$ ).

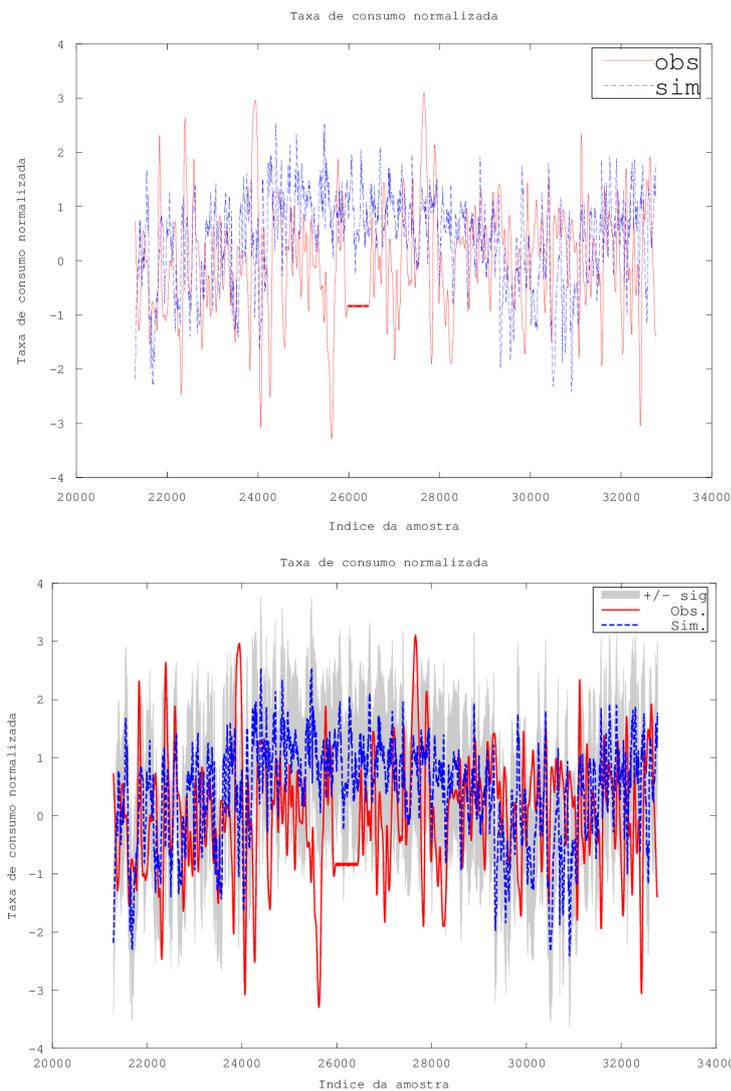
**Tabela 3.2:** Desempenho estatístico dos modelos M1, M2, M3, M4, M5, M6, M7 e M8 para os índices estatísticos POD, FAR, DD, TS e ETS.

Modelos	Índices Descrição/Intervalo estrutura da rede	POD [0;1] ideal=1	FAR [0;1] ideal=0	DD [0;1] ideal=1	TS ideal=1	ETS ideal=1
M1	( $t, vel, p, ur, w-1$ ); 5-40-1	0,74	0,19	0,56	0,63	0,39
M2	( $t, vel, p, ur, w-7$ ); 5-40-1	0,44	0,48	0,021	0,32	0,010
M3	( $ano, mes, nasch, porh,$ $t, vel, u, v, p, ur, w-1$ ); 11-40-1	0,57	0,21	0,45	0,50	0,26
M4	persistence( $w-2, w-1$ ); 2-40-1	0,82	0,13	0,69	0,73	0,53
M5	persistence( $w-7, w-6,$ $w-5, w-4, w-3, w-2, w-1$ ); 7-40-1	0,91	0,18	0,71	0,75	0,53
M6	( $t, vel, p, ur$ ); 4-40-1	0,49	0,52	0,058	0,32	- 0,028
M7	( $vel, u, v, p, ur$ ); 5-40-1	0,27	0,59	0,15	0,19	- 0,060
M8	( $t, ur$ ); 2-40-1	0,50	0,48	0,028	0,34	0,014

Nota: POD= probabilidade de detecção; FAR= razão de falso alarme; DD= grau de desproporção; TS= índice de sucesso; ETS= índice de sucesso corrigido.

a explicação da amplitude da dispersão observada.

A matriz de correlação ( $CORR$ ) das variáveis de entrada é mostrada na Figura 3.61.



**Figura 3.51:** Etapas de validação cruzada do modelo  $M7$  ( $vel$ ,  $u$ ,  $v$ ,  $p$ ,  $ur$ ).

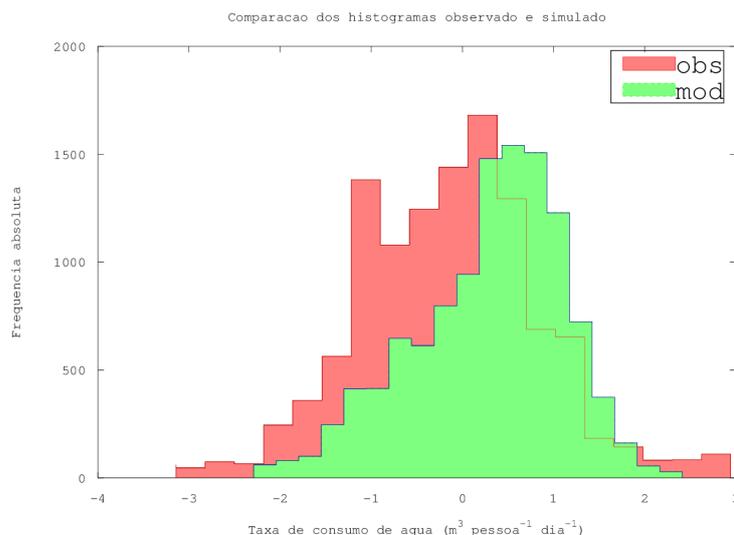
Em primeiro plano, destaca-se o papel da autocorrelação do consumo  $w - 1$  e  $w - 2$  na explicação do consumo observado. A correlação entre o consumo e as variáveis meteorológicas é menor que o efeito da autocorrelação.

A autocorrelação do consumo é maior para intervalo de 24 horas, reduzindo-se muito para intervalos maiores.

### Função adjunta de densidade de probabilidade

A Figura 3.62, temos a matriz de dispersão de seis variáveis meteorológicas: temperatura, velocidade do vento, pressão atmosférica, umidade relativa e duas variáveis de consumo de água, para o dia presente e o anterior.

A função adjunta de densidade de probabilidade entre observação e simulação pode



**Figura 3.52:** Histograma da validação cruzada do modelo M7 (*vel, u, v, p, ur*).

**Tabela 3.3:** Desempenho estatístico dos modelos M1, M2, M3, M4, M5, M6, M7 e M8 para os índices estatísticos  $\sigma_{obs}$ ,  $\sigma_{mod}$ , MAE, BIAS,  $R^2$  e MSE.

Modelos	Índices Descrição; estrutura da rede	$\sigma_{obs}$	$\sigma_{mod}$ ideal= $\sigma_{obs}$	MAE ideal=0	BIAS	$R^2$ ideal=0	MSE
M1	(t,vel,p,ur,w-1); 5-40-1	1,00	0,73	0,53	- 0,032*	0,40*	0,47
M2	(t,vel,p,ur,w-7); 5-40-1	1,00	0,40	0,84	0,058*	- 0,20*	1,2
M3	(ano,mes,nasch,porh, t,vel,u,v,p,ur,w-1); 11-40-1	1,00	0,72	0,62	0,20*	0,39*	0,65
M4	persistence(w-2,w-1); 2-40-1	1,00	0,88	0,42	0,040*	0,70*	0,30
M5	persistence(w-7,w-6, w-5,w-4,w-3,w-2,w-1); 7-40-1	1,00	0,81	0,38	0,26*	0,51*	0,24
M6	(t,vel,p,ur); 4-40-1	1,00	0,29	0,83	- 0,31*	-1,1*	2,1
M7	(vel,u,v,p,ur); 5-40-1	1,00	0,45	0,92	- 0,38*	- 0,47*	1,3
M8	(t,ur); 2-40-1	1,00	0,55	0,90	0,17*	- 0,15*	1,3

Nota:  $\sigma_{obs}$ = desvio padrão observado;  $\sigma_{mod}$ = desvio padrão modelado; MAE= erro médio absoluto; BIAS= erro sistemático;  $R^2$ = grau de correlação entre observação e previsão; MSE= erro quadrático médio.

ser estimada a partir do diagrama de dispersão dos pontos modelados e observados.

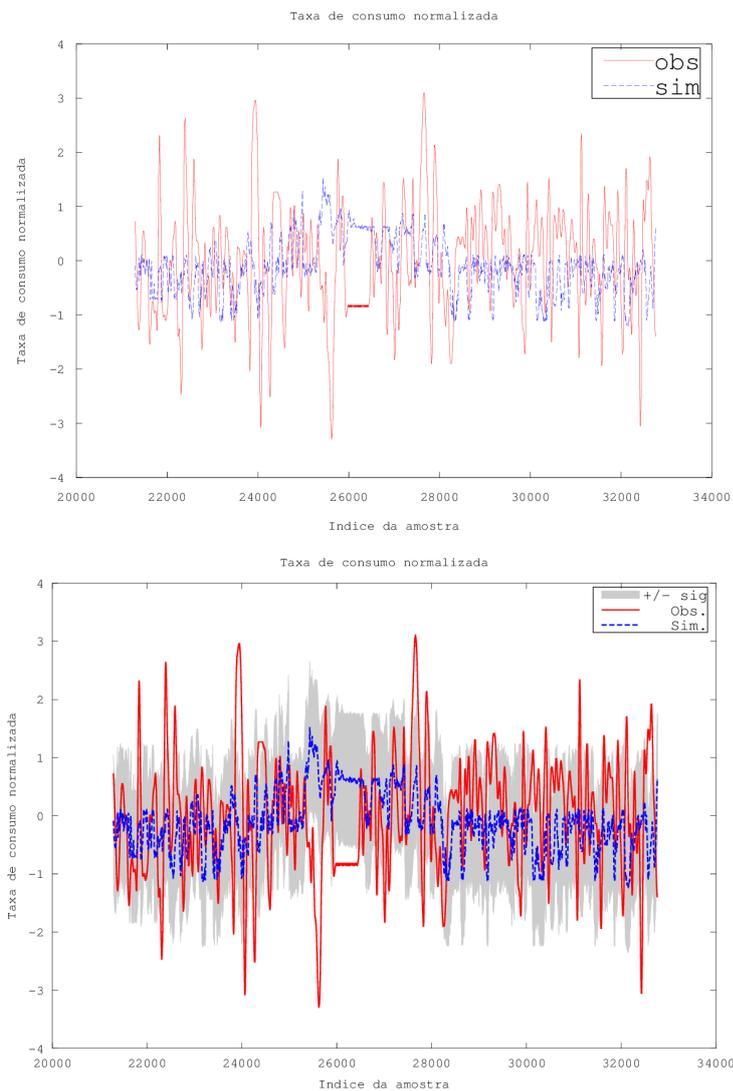
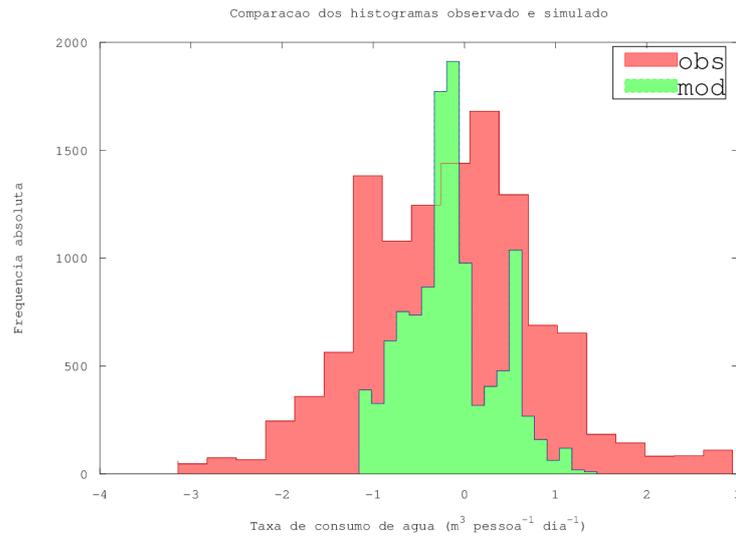


Figura 3.53: Etapas de validação cruzada do modelo  $M8(t, ur)$ .

## Discussões suplementares

Dado que diferentes microclimas urbanos são resultantes das alterações da superfície, i.e., nascem da urbanização que altera o balanço de água e energia originais, questiona-se, por exemplo, se a intensificação da Ilha de Calor Urbana, associada aos processo de urbanização, terá ou não como resultado uma modificação dos níveis atuais de consumo de água potável em cidades tropicais?

Se por um lado, a demanda urbana é por água potável, por outro, a cidade é fonte de poluição das águas utilizadas. Como a água é um solvente, o aumento da vazão de água potável pode levar a uma maior dissolução dos poluentes no meio líquido, caso fosse, de se aplicar um tratamento aos efluentes líquidos (i.e., fechar o ciclo de saneamento da água). No momento, não se pode entender isto sem um planejamento de saneamento das bacias urbanas, incluindo os rios. Diga-se que a legislação de proteção de áreas vegetadas pouco ou nada se aplica aos rios urbanos, i.e., em relação à presença de vegetação ciliar nas margens, o que causa perplexidade diante



**Figura 3.54:** Histograma da validação cruzada do modelo M8 ( $t, ur$ ).

**Tabela 3.4:** Desempenho estatístico dos modelos M1, M2, M3, M4, M5, M6, M7 e M8 para os índices estatísticos  $MSE_{pers}$ , SSA,  $RMSE_{mod}$ ,  $RMSE_{ref}$  e  $POD_{chance}$ .

Modelos	Índices Descrição; estrutura da rede	$MSE_{pers}$	SSA	$RMSE_{mod}$ ideal=0	$RMSE_{ref}$	$POD_{chance}$ ideal=0
M1	(t,vel,p,ur,w-1); 5-40-1	0,48	0,019	0,69	0,69	$\frac{2723}{11479} = 0,24$
M2	(t,vel,p,ur,w-7); 5-40-1	0,48	- 1,4	1,08	0,69	$\frac{2532}{11479} = 0,22$
M3	(ano,mes,nasch,porh, t,vel,u,v,p,ur,w-1); 11-40-1	0,48	- 0,36	0,81	0,69	$\frac{2157}{11479} = 0,19$
M4	persistence(w-2,w-1); 2-40-1	0,48	0,38	0,55	0,69	$\frac{2793}{11479} = 0,24$
M5	persistence(w-7,w-6, w-5,w-4,w-3,w-2,w-1); 7-40-1	0,48	0,49	0,50	0,69	$\frac{3319}{11479} = 0,29$
M6	(t,vel,p,ur); 4-40-1	0,48	- 1,4	1,06	0,69	$\frac{3049}{11479} = 0,27$
M7	(vel,u,v,p,ur); 5-40-1	0,48	- 1,8	1,2	0,69	$\frac{1923}{11479} = 0,17$
M8	(t,ur); 2-40-1	0,48	- 1,7	1,1	0,69	$\frac{2847}{11479} = 0,25$

Nota:  $MSE_{pers}$ = erro quadrático médio do modelo de persistência; SSA= índice de habilidade de previsão comparado com o modelo de referência;  $RMSE_{mod}$ = raiz do erro quadrático médio do modelo;  $RMSE_{ref}$ = raiz do erro quadrático médio do modelo de referência;  $POD_{chance}$ = probabilidade de detecção por chance.

	q1 (obs)	q2 (obs)	q3 (obs)	q4 (obs)	q5 (obs)	q6 (obs)
q6 (mod)	0.0	1.0	0.0	0.0	66.0	146.0
q5 (mod)	0.0	47.0	176.0	522.0	1393.0	156.0
q4 (mod)	5.0	204.0	574.0	1024.0	984.0	36.0
q3 (mod)	31.0	1591.0	1060.0	742.0	481.0	12.0
q2 (mod)	179.0	1327.0	365.0	177.0	120.0	1.0
q1 (mod)	24.0	32.0	3.0	0.0	0.0	0.0

**Figura 3.55:** Tabela de contingência estilo 6x6 do modelo  $M_1$  ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).

	q1 (obs)	q2 (obs)	q3 (obs)	q4 (obs)	q5 (obs)	q6 (obs)
q6 (mod)	0.0	1.0	0.0	5.0	92.0	168.0
q5 (mod)	0.0	5.0	111.0	499.0	1906.0	178.0
q4 (mod)	0.0	129.0	453.0	1162.0	758.0	5.0
q3 (mod)	5.0	736.0	1015.0	614.0	239.0	0.0
q2 (mod)	161.0	2305.0	599.0	185.0	49.0	0.0
q1 (mod)	73.0	26.0	0.0	0.0	0.0	0.0

**Figura 3.56:** Tabela de contingência estilo 6x6 do modelo  $M_4$  ( $w-1$ ,  $w-2$ ).

da necessidade de áreas verdes para lazer da população, de acordo com a ideia de utilização do espaço urbano de forma inclusiva.

No caso do Rio de Janeiro, na qual extensas áreas periféricas apresentam-se densamente habitadas levanta-se a questão das causas dos níveis elevados de consumo de água potável, em toda a cidade mas possivelmente incrementado nos bairros com melhores serviços públicos. Pergunta-se se a causa do consumo incrementado seria a existência de vazamentos ao longo da rede de distribuição? Mais provável entretanto é que o consumo da RMRJ possa ser entendido como resultado de um padrão de consumo diferente daquele que foi empregado pela ONU na elaboração de sua recomendação. Do ponto de vista econômico, níveis de consumo elevados podem induzir um aumento de preços pelo aumento da demanda. Do ponto de vista do

mod $\geq$ q(50%)	1007.0	4327.0
mod $<$ q(50%)	4612.0	1533.0
	obs $<$ q(50%)	obs $\geq$ q(50%)

**Figura 3.57:** Tabela de contingência estilo  $2 \times 2$  do modelo  $M_1$  ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).

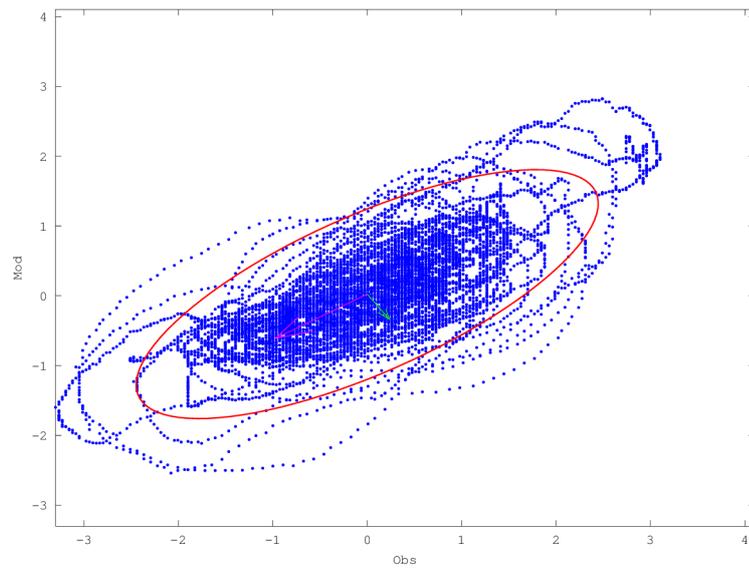
mod $\geq$ q(50%)	699.0	4773.0
mod $<$ q(50%)	4920.0	1087.0
	obs $<$ q(50%)	obs $\geq$ q(50%)

**Figura 3.58:** Tabela de contingência estilo  $2 \times 2$  do modelo  $M_4$  ( $w-1$ ,  $w-2$ ).

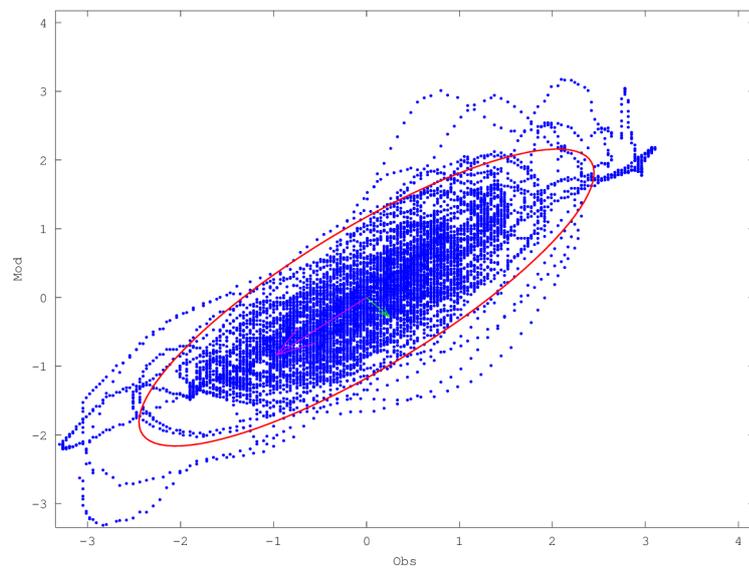
clima, a ocorrência de secas e de longos períodos de estiagem também pode implicar em aumentos do preço da água potável. É importante apontar ainda a inter-relação existente entre as produções de energia elétrica e de água potável.

## Considerações finais

Seria interessante continuar o trabalho utilizando dados de demanda de água global, como feito para a RMSP, para obter um entendimento do consumo pelos vários agentes, não somente para o consumo doméstico. A metodologia utilizada aqui pode ser aplicada a outras metrópoles e cidades, desde que é baseada principalmente em software livre, com saída gráfica automatizada em octave.



**Figura 3.59:** Distribuição dos dados observados e previstos do modelo  $M_1$  ( $t$ ,  $vel$ ,  $p$ ,  $ur$  e  $w-1$ ).



**Figura 3.60:** Distribuição dos dados observados e previstos do modelo  $M_4$  ( $w-1$ ,  $w-2$ ).

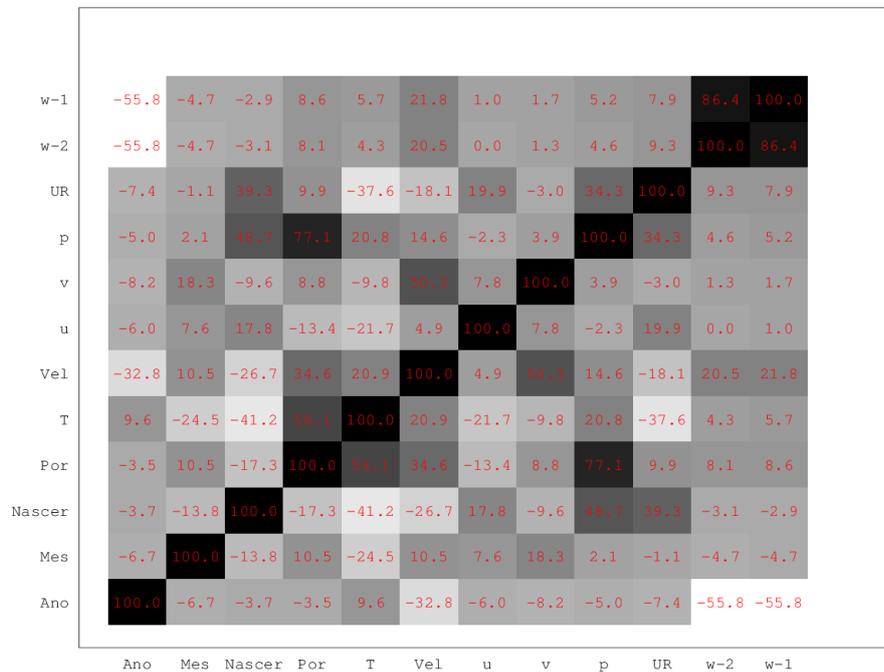


Figura 3.61: Matriz de correlação entre as variáveis de entrada da RNA.

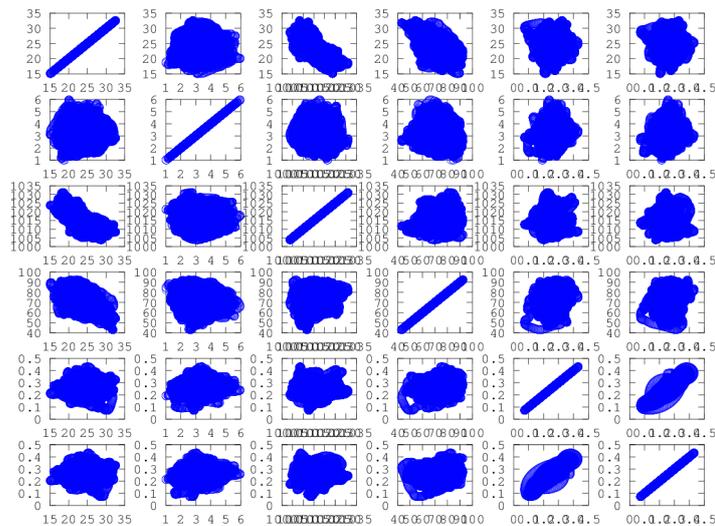


Figura 3.62: Matriz de dispersão entre as variáveis da RNA, sendo a coluna 1 e linha 1: temperatura, coluna 2 e linha 2: velocidade do vento, coluna 3 e linha 3: pressão atmosférica, coluna 4 e linha 4: umidade relativa, coluna 5 e linha 5: consumo de dois dias anteriores e coluna 6 e linha 6: consumo de um dia anterior.

O trabalho foi desenvolvido para ampliar o entendimento sobre o consumo de água potável na RMRJ na segunda década do século XIX, para que possa ser alcançada uma utilização mais racional da água.



# Capítulo 4

## Conclusões

1. Os dados meteorológicos foram organizados e analisados. Após isso, executou-se uma estatística básica, encontrando média, variância, desvio padrão, histogramas. Além disso, caracterizou-se os dados por meio da transformada de Fourier (TF) e da TO, fizemos a análise espectral para identificar quais períodos das variáveis meteorológicas são mais importantes ao longo da série temporal.
2. As TO que foram empregadas na análise do consumo do prédio residencial indicaram ciclos de 1 mês e 2 anos que caracteriza a mudança no hábito de consumo de água mensal e interanual por conta de mudanças climáticas.
3. No esquema da RNA foram empregadas entre 2 e 7 variáveis de entradas (i.e., tempo, temperatura, pressão, umidade do ar, vento, componentes zonal e meridional, consumo dos dias anteriores, de 1 a 7 dias), uma camada interna com 40 neurônios e 1 saída que corresponde ao consumo de água. Posteriormente, apresentou-se os resultados da rede neural e foram efetuados as avaliações dos estimadores estatísticos da incerteza como verificação, para todas as etapas de utilizada da rede topológica. Assim, modelou-se o consumo de água potável.
4. A rede que usou variáveis meteorológicas na entrada não apresenta uma vantagem aparente, para a métrica MSE, em relação ao modelo de persistência. Entretanto, apresenta uma pequena vantagem em reduzir o erro sistemático (*BIAS*).
5. Este trabalho permitiu estimar o valor do consumo de água residencial na Região Metropolitana do Rio de Janeiro, em aproximadamente  $0,28 (m^3 \text{ dia}^{-1} \text{ pessoa}^{-1})$ .
6. Os principais resultados obtidos neste trabalho foram submetidos na forma de uma artigo a uma revista científica da área.



# Capítulo 5

## Anexo

### 5.1 Programação

Os seguintes programas foram utilizados:

**wget\_metar** A concentração dos dados meteorológicos é realizada pelos comandos `wget` (<https://www.gnu.org/software/wget/>) e `cat` ([/unix\\_commands/cat.htm](/unix_commands/cat.htm)) do shell/linux. O arquivo concatenado na saída do `wget_metar` é a entrada de um programa Fortran chamado `reading_dataset`. Código-fonte: Laboratório de Hidrometeorologia Experimental.

**reading\_dataset** O programa chamado `reading_dataset.f90` organiza as variáveis meteorológicas em colunas de formato `XYZ.dat`. Também é aplicado para reformatação das variáveis, estabelecimento de uma escala de tempo comum e contínua (neste caso, o ano decimal em dupla precisão), localização geográfica (inclusão da longitude, latitude e altitude ANMS), validação e pareamento das variáveis (seleção de linhas consistentes), conferência se os valores escalares encontram-se dentro do intervalo esperado de desvio, e conversões das unidades físicas para o Sistema Internacional (S.I.) de unidades físicas. Código-fonte: Laboratório de Hidrometeorologia Experimental.

**script R** Script codificado do R para calcular as funções densidade de probabilidade (f.d.p.), funções de densidade acumulada (f.d.a.), plotagens quantil-quantil (q-q) e boxplot anual de cada variável. Código-fonte: Laboratório de Hidrometeorologia Experimental.

**PAST** A Transformada de Ondeleita (TO) é obtida com aplicativo PAleontological STatistics (PAST). De acordo com o site do aplicativo PAST trata-se de software livre para windows com capacidade de análise de dados científicos, funções de manipulação, plotagem de estatística uni- e multivariada, análise ecológica, análise de séries temporais e espaciais, análise morfométrica e estratigráfica. Site e download: <http://folk.uio.no/ohammer/past/>.

**NNET** A RNA foi implementada em Octave com base na biblioteca/library Neural Networks (`nnet`), versão: 0.1.13 de 02-Dez-2010, escrita pelo Dr. Michael Schmid. Esta biblioteca apresenta com capacidade de alimentação *forward* de uma rede neural de multicamadas com otimização usando mínimos quadrados. O desenvolvimento da biblioteca baseou-se nos trabalhos de Eaton (1997)[32],

The MathWorks[72], Bishop (2002)[13] e Hagon et al. (1996)[41]. Código-fonte é acessível via: <http://octave.sf.net> ou <http://octnnettb.sourceforge.net>.

# Referências Bibliográficas

- [1] J. Adamowski and C. Karapataki. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ann learning algorithms. *Journal of Hydrologic Engineering*, 15(10):729–743, 2010. 26, 29
- [2] J. F. Adamowski. Peak daily water demand forecast modeling using artificial neural networks. *Journal of Water Resources Planning and Management*, 134(2):119–128, 2008. 26
- [3] Agevap. Plano de recursos hídricos da bacia do rio paraíba do sul - resumo diagnóstico dos recursos hídricos relatório final. Technical report, Associação Pró-Gestão das Águas da Bacia Hidrográfica do Rio Paraíba do Sul - AGEVAP, 2006. 201 p. 6
- [4] ANA. Atlas brasil. abastecimento urbano de água: Região metropolitana do rio de janeiro. Agência Nacional de Águas, 2010a. 10, 13
- [5] ANA. Atlas brasil. abastecimento urbano de água: Panorama nacional. Agência Nacional de Águas, 2010b. 6, 7
- [6] ANA. Conjuntura dos recursos hídricos no brasil. Technical report, Agência Nacional de Águas, 2013. 5, 6
- [7] ANA. Conjuntura dos recursos hídricos no brasil. Agência Nacional de Águas, 2014. 6
- [8] L. G. S. Augusto, I. G. D. Gurgel, H. F. C. Neto, C. H. De Melo, and A. M. Costa. O contexto global e nacional frente aos desafios do acesso adequado à água para consumo humano. *Ciência e saúde coletiva*, .(.):., 2012. 1
- [9] M. B. Ayibotele. The world’s water: assessing the resource. In *International Conference on Water and the Environment: Development Issues for the 21 st Century*. Dublin, Ireland, 1992. 5
- [10] S. L. Barnes. A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, 3(4):396–409, 1964. 15
- [11] D. Barreto. Perfil do consumo residencial e usos finais da água. *Ambiente Construído*, 8(2):23–40, 2008. 9, 10, 14
- [12] R. G. Barry and R. J. Chorley. *Atmosfera, tempo e clima*. Bookman Editora, 2009. 36
- [13] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 2002. 88
- [14] M. J. A. Bolzan. Analysis of wavelet transform applied in geophysical. *Revista Brasileira de Ensino de Física*, 26(1):37–41, 2004. 18

- [15] G. Bontempi. Resampling techniques for statistical modeling. Technical report, Département d'Informatique, Computer Sciences Dept, 2003. 26, 27, 28
- [16] V. M. N. A. Borges. Acoplamento de um modelo de previsão de demanda de água a um modelo simulador em tempo real-estudo de caso: sistema adutor metropolitano de são paulo. Master's thesis, Universidade de São Paulo, 2003. 13
- [17] A. L. Britto. Gestão de serviços de saneamento em áreas metropolitanas: as alternativas existentes diante da necessidade de universalização dos serviços e preservação da qualidade ambiental. Technical report, BNDES, 2007. 7
- [18] S. R. Cajal. *Die Retina der Wirbeltiere.* -, 1894. 20
- [19] A. Campos and A. S. Soares. O acesso à água na américa do sul e em áfrica, diferenças e semelhanças. *Africanologia-Revista Lusófona de Estudos Africanos*, 1(2):., . 2009. . 5
- [20] A. Cardon, D. N. Müller, and P. Navaux. Introdução às redes neurais artificiais. Technical report, Instituto de Informática. Universidade Federal do Rio Grande do Sul. Porto Alegre, 1994. 21, 22
- [21] J. E. Castro. A gestão da água na américa latina. *Desafio do desenvolvimento. Instituto de Pesquisa Econômica Aplicada (IPEA)*, 74(-):-, - 2012. -. 5
- [22] CEDAE. Guia do usuário. Technical report, Companhia Estadual de Águas e Esgotos, 2012. 14
- [23] CEDAE. Informativo anual sobre a qualidade da água distribuída para a população do estado do rio de janeiro. Companhia Estadual de Águas e Esgotos, jan./dez. 2014. 8
- [24] S. Collineau and Y. Brunet. Detection of turbulent coherent motions in a forest canopy part i: wavelet analysis. *Boundary-Layer Meteorology*, 65(4):357–379, 1993. 19
- [25] S. M. B. B. Correa. *Probabilidade e Estatística.* PUC Minas Virtual, 2 edição edition, 2003. 15
- [26] CPTEC. Avaliação de modelos: Tabela de contingência. Centro de Previsão de Tempo e Estudos Climáticos, Junho 2016. 31
- [27] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7):909–996, 1988. 18
- [28] Laboratório de pesquisas (Unicarioca). Hábitos do carioca em relação ao consumo de água. Centro Universitário Unicarioca, 2015. 9
- [29] C. A. Doswell III. Obtaining meteorologically significant surface divergence fields through the filtering property of objective analysis. *Monthly Weather Review*, 105(7):885–892, 1977. 16
- [30] F. Dourado, T. C. Arraes, and M. Silva. O megadesastre da região serrana do rio de janeiro: as causas do evento, os mecanismos dos movimentos de massa e a distribuição espacial dos investimentos de reconstrução no pós-desastre. *Anuário do Instituto de Geociências*, 35(2):43–54, 2012. CCMN-IGEO-UFRJ. 48
- [31] P. Dutilleux. An implementation of the  $\hat{a}$  algorithm  $\hat{a}$  trous  $\hat{a}$  to compute the wavelet transform. In *Wavelets*, pages 298–304. Springer, 1989. 17

- [32] J. W. Eaton. *GNU Octave Manual*. PDF-Version, third edition, February 1997. 87
- [33] A. V. Falkenberg. Previsão de consumo urbano de água em curto prazo. Master's thesis, Universidade Federal do Paraná, 2005. 14
- [34] FAO. Water resources. Food and agriculture organization of the United Nations, 2014. 5
- [35] M. Farge. Wavelet transforms and their applications to turbulence. *Annual review of fluid mechanics*, 24(1):395–458, 1992. 17, 19
- [36] D. Tubbs Filho, J. C. O. Antunes, and J. Silva Vettorazzi. Bacia hidrográfica dos rios guandu, da guarda e guandu-mirim: experiências para a gestão dos recursos hídricos. In *Bacia hidrográfica dos Rios Guandu, da Guarda e Guandu-Mirim: experiências para a gestão dos recursos hídricos*. Instituto Estadual do Ambiente, 2012. , 8
- [37] M. B. Freitas, O. M. Brillhante, and L. M. Almeida. Importância da análise de água para a saúde pública em duas regiões do estado do rio de janeiro: enfoque para coliformes fecais, nitrato e alumínio. *Cadernos de Saúde Pública*, 17(3):651–660, 2001. 9
- [38] N. Gamage and W. Blumen. Comparative analysis of low-level cold fronts: wavelet, fourier, and empirical orthogonal function decompositions. *Monthly weather review*, 121(10):2867–2878, 1993. 17
- [39] P. H. Gleick. *Water in crisis: a guide to the world's fresh water resources*. Oxford University Press, Inc., 1993. 2, 4
- [40] P. H. Gleick. The worlds water 2000-2001. Technical report, 2000. 2
- [41] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesús. *Neural network design*, volume 20. PWS publishing company Boston, 1996. 88
- [42] L. Heller and V. L. de Pádua. *Abastecimento de água para consumo humano*. Editora UFMG, 2006. 7
- [43] I. Hespanhol. Um novo paradigma para a gestão de recursos hídricos. *Estudos avançados*, 22(63):131–158, 2008. 10
- [44] INMET. Climatologia. Instituto Nacional de Meteorologia, 2016. 13
- [45] B. Jorgensen, M. Graymore, and K. O'Toole. Household water use behavior: An integrated model. *Journal of environmental management*, 91(1):227–236, 2009. 9
- [46] M. R. Jury and J. L. Melice. Analysis of durban rainfall and nile river flow 1871-1999. *Theoretical and applied climatology*, 67(3-4):161–169, 2000. 18
- [47] S. E. Koch, M. DesJardins, and P. J. Kocin. An interactive barnes objective map analysis scheme for use with satellite and conventional data. *Journal of Climate and Applied Meteorology*, 22(9):1487–1503, 1983. 16
- [48] Z. L. Kovács. *Redes Neurais Artificiais - Fundamentos e Aplicações*. Editora Livraria da Física, 2006. 174 p. 20
- [49] P. Kumar and E. Foufoula-Georgiou. *Wavelet analysis in geophysics: An introduction*, volume 4. Academic Press New York, 1994. 17

- [50] K. M. Lau and H. Weng. Climate signal detection using wavelet transform: How to make a time series sing. *Bulletin of the American Meteorological Society*, 76(12):2391–2402, 1995. 19
- [51] J. Margat and J.R. Tiercelin. *L'eau en questions*. Enjeu du 21<sup>e</sup> siècle, 1998. 2
- [52] S. D Meyers, B. G. Kelly, and J. J. O'Brien. An introduction to wavelet analysis in oceanography and meteorology: With application to the dispersion of yanai waves. *Monthly weather review*, 121(10):2858–2866, 1993. 19
- [53] A. Mukhopadhyay, A. Akber, and E. Al-Awadi. Analysis of freshwater consumption patterns in the private residences of kuwait. *Urban water*, 3(1):53–62, 2001. 9
- [54] Allan H Murphy. The finley affair: A signal event in the history of forecast verification. *Weather and Forecasting*, 11(1):3–20, 1996. 31
- [55] A. Musy and C. Higy. *Hydrologie: Une science de la nature*, volume 21. PPUR presses polytechniques, 2004. 4
- [56] OECD. Oecd environmental outlook to 2050: The consequences of inaction. Technical report, Organization for Economic Co-operation and Development, 2012. 1
- [57] ONU. Terceiro fórum mundial da água. Technical report, Organização das Nações Unidas, Kyoto, Japão, 2003. 1, 46, 49
- [58] S. P. Pathak, S. Kumar, P. W. Ramteke, R. C. Murthy, J. W. Bhattacharjee, and K. Gopal. Potability of water sources in relation to metal and bacterial contamination in some northern and north-eastern districts of india. *Environmental monitoring and assessment*, 33(2):151–160, 1994. 9
- [59] PNUD. Relatório do desenvolvimento humano. a água para lá da escassez: poder, pobreza e a crise mundial da água, 2006. Technical report, Programa das Nações Unidas para o Desenvolvimento, 2006. 1
- [60] M. S. Reboita. Elementos da variabilidade climática no extremo sul do brasil, no período de 1990 a 2001. Master's thesis, Universidade Federal do Rio Grande do Sul, 2004. 17
- [61] A. C. Rebouças. *Uso inteligente da água*. Editora Escrituras, 2004. 10, 13, 18
- [62] W. C. Ribeiro. *Geografia política da água*. Annablume Editora, 2008. 4, 49
- [63] E. Rogers, T. L. Black, D. G. Deaven, G. J. DiMego, Q. Zhao, M. Baldwin, N. W. Junker, and Y. Lin. Changes to the operational 'early' eta analysis/-forecast system at the national centers for environmental prediction. *Weather and Forecasting*, 11(3):391–413, 1996. 32
- [64] J. M. C. Rosa. Intervalos de confiança, 2016. 29
- [65] A. C. Rosette. Previsão a curtíssimo prazo de eventos meteorológicos extremos para a área de influência do aeroporto internacional tom jobim utilizando técnicas de redes neurais artificiais. Master's thesis, Universidade Federal do Rio de Janeiro, 2011. 25
- [66] L. D. A. Sá, S. B. M. Sambatti, and G. P. Galvão. Ondeleta de morlet aplicada ao estudo da variabilidade do nível do rio paraguai em ladário, ms. *Pesquisa Agropecuária Brasileira*, 33(13):1775–1785, 1998. 19

- [67] C. C. Santos. *Previsão De Demanda De Água Na Região Metropolitana De São Paulo Com Redes Neurais Artificiais E Condições Socioambientais E Meteorológicas*. PhD thesis, Universidade de São Paulo, 2011. 7
- [68] C. C. Santos and A. J. Pereira Filho. Water demand forecasting model for the metropolitan area of são paulo, brazil. *Water Resources Management*, 28(13):4401–4414, 2014. 26
- [69] F. S. Santos, D. M. Silva, and D. M. Veneu. Reuso predial de águas de chuva aplicados em um edifício residencial no município do rio de janeiro. *Revista Episteme Transversalis*, 8(1), 2015. 9, 14
- [70] SNIS. Diagnóstico dos serviços de água e esgotos. Technical report, Sistema Nacional de Informações sobre Saneamento, 2013a. 8, 9
- [71] SNIS. Série histórica. Sistema Nacional de Informações sobre Saneamento, 2013b. 9
- [72] Inc The MathWorks. *MATLAB Help*. MATLAB Version 7.1 (R14SP3), Neural Network Toolbox Version 4.0.6 (R14SP3). 88
- [73] C. Torrence and G. P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78, 1998. 18
- [74] C. E. M. Tucci, J. C. Bertoni, et al. *Inundações urbanas na América do Sul*. Ed. dos Autores, 2003. 5
- [75] M. I. Vitorino. *Análise das oscilações intrasazonais sobre a América do Sul e oceanos adjacentes utilizando a análise de ondeletas*. PhD thesis, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2002. 18
- [76] H. Weng and K. M. Lau. Wavelets, period doubling, and time-frequency localization with application to organization of convection over the tropical western pacific. *Journal of the atmospheric sciences*, 51(17):2523–2541, 1994. 17, 18, 20
- [77] Wikipedia. Wavelet, 2015. 17
- [78] D. S. Wilks. *Statistical methods in the atmospheric sciences*, volume 100. Academic press, second edition, 2006. 15
- [79] D. S. Wilks. *Statistical methods in the atmospheric sciences*. Academic press, 2011. 31
- [80] M. J. S. Yabe and E. Oliveira. Metais pesados em águas superficiais como estratégia de caracterização de bacias hidrográficas. *Química Nova*, 21(5):551–556, 1998. 8
- [81] K. Zahed Filho. *Previsão de demanda de consumo em tempo real no desenvolvimento operacional de sistemas de distribuição de água*. PhD thesis, Universidade de São Paulo, 1990. 8, 26