

ANO  
2017

LAIS LEDRA | MÉTODOS DE PREENCHIMENTO DE FALHAS DE DADOS PLUVIOMÉTRICOS:  
COMPARAÇÕES PARA DADOS DE 24 HORAS NA CIDADE DE JOINVILLE, SC



**UDESC**

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC  
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT  
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA CIVIL

DISSERTAÇÃO DE MESTRADO

**MÉTODOS DE PREENCHIMENTO DE  
FALHAS DE DADOS PLUVIOMÉTRICOS:  
COMPARAÇÕES PARA DADOS DE 24  
HORAS NA CIDADE DE JOINVILLE, SC**

LAIS LEDRA

JOINVILLE, 2017

Séries históricas são fundamentais no desenvolvimento de estudos hidrológicos. Entretanto, a maior dificuldade encontrada muitas vezes é a indisponibilidade dos dados e o número de falhas encontrado nessas séries disponibilizadas pela rede de estações pluviométricas. Em função dessas lacunas, é necessário realizar o preenchimento de falhas a partir de métodos aplicáveis, e verificar a consistência dos dados após o preenchimento. Neste trabalho o objetivo é realizar o preenchimento de falhas de dados de precipitação de 24 horas da rede de monitoramento da Defesa Civil do Município de Joinville (DCMJ), SC, aplicando os métodos de Interpolação Linear, Regressão Linear Múltipla, Método das Correlações, Redes Neurais Artificiais e Imputação Múltipla, todos eles utilizando o *software R*.

Orientadora: Prof<sup>a</sup> Dr<sup>a</sup>. Virgínia Grace Barros

Joinville, 2017

**LAIS LEDRA**

**MÉTODOS DE PREENCHIMENTO DE FALHAS DE DADOS PLUVIOMÉTRICOS:  
COMPARAÇÕES PARA DADOS DE 24 HORAS NA CIDADE DE JOINVILLE, SC.**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Civil, da Universidade do Estado de Santa Catarina, como requisito parcial para obtenção do grau de Mestre em Engenharia Civil.

Orientador: Prof<sup>a</sup> Dr<sup>a</sup>. Virgínia Grace Barros

**JOINVILLE**

**2017**

Ficha catalográfica elaborada pelo(a) autor(a), com  
auxílio do programa de geração automática da  
Biblioteca Setorial do CCT/UDESC

Ledra, Lais

MÉTODOS DE PREENCHIMENTO DE FALHAS DE DADOS  
PLUVIOMÉTRICOS: COMPARAÇÕES PARA DADOS DE 24 HORAS NA  
CIDADE DE JOINVILLE, SC. / Lais Ledra. -  
Joinville , 2017.  
168 p.

Orientadora: Virgínia Grace Barros  
Dissertação (Mestrado) - Universidade do Estado de  
Santa Catarina, Centro de Ciências Tecnológicas,  
Programa de Pós-Graduação em Engenharia Civil,  
Joinville, 2017.

1. Precipitação. 2. Preenchimento de falhas de  
dados. 3. Dados de 24 horas. I. Grace Barros,  
Virgínia . II. Universidade do Estado de Santa  
Catarina. Programa de Pós-Graduação. III. Título.

**Métodos de Preenchimento de Falhas de Dados Pluviométricos: Comparações  
para Dados de 24 Horas na Cidade de Joinville, SC**

por

**Lais Ledra**

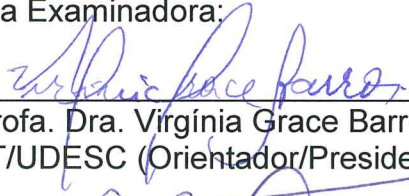
Esta dissertação foi julgada adequada para obtenção do título de


**MESTRE EM ENGENHARIA CIVIL**

Área de concentração em “Engenharia Urbana e da Construção Civil”  
e aprovada em sua forma final pelo

**CURSO DE MESTRADO EM ENGENHARIA CIVIL  
CENTRO DE CIÊNCIAS TECNOLÓGICAS DA  
UNIVERSIDADE DO ESTADO DE SANTA CATARINA.**

Banca Examinadora:

  
\_\_\_\_\_  
Profa. Dra. Virgínia Grace Barros  
CCT/UDESC (Orientador/Presidente)

  
\_\_\_\_\_  
Prof. Dr. Doalcey Antunes Ramos  
CCT/UDESC

  
\_\_\_\_\_  
Profa. Dra. Therezinha Maria Novais  
de Oliveira - Univille

**Joinville, SC, 29 de agosto de 2017.**



## **AGRADECIMENTOS**

Primeiramente quero agradecer a Deus por ter colocado esta oportunidade em meu caminho, além de fazer tudo acontecer da forma que tinha que ser. Me dando saúde e me guiando sempre para o caminho correto principalmente quando encontrei dificuldades. Agradecer aos meus pais, Ana Maria e Evaristo, pois sem eles não teria conseguido esta conquista do grau de Mestre em Engenharia Civil. Eles e toda minha família, incluindo minha irmã Mariana, meu cunhado Hueyton e minha sobrinha Amanda, que sempre apoiaram minhas decisões, me dando forças pra continuar e nunca desistir, sendo indispensáveis nas idas e vindas para Rio do Sul. Aos meus colegas de mestrado e também da sala H14, Dalila, José, Gabriela, Rodrigo, Tânia, Natássia e Adriane, pelas conversas e cafés. Aos meus amigos externos ao universo acadêmico que se fizeram presentes em todos os momentos de descontração, Alline, Bruna, Carol, Dalvana e Fernanda. Às minhas colegas de “casa” Geovana e Suzy, pela companhia e incentivo. Aos demais colegas e professores do Programa de Pós-Graduação, em especial a Professora Elisa Henning, por participar constantemente deste trabalho como co-orientadora extra oficial. Aos profissionais da Defesa Civil Municipal de Joinville, Maiko e Dieter, pela disponibilização de dados importantes que contribuíram para o trabalho. À minha orientadora, Profa. Dra. Virgínia Grace Barros por ter aceitado esse desafio e acreditar no meu potencial, me fazendo crer que era capaz e por enriquecer esse trabalho compartilhando seus conhecimentos. A todos que de certa forma participaram dessa conquista, o meu muito obrigado.

## RESUMO

Séries históricas são fundamentais no desenvolvimento de estudos hidrológicos. Entretanto, a maior dificuldade encontrada muitas vezes é a indisponibilidade dos dados e o número de falhas encontrado nessas séries disponibilizadas pela rede de estações pluviométricas. Em função dessas lacunas, é necessário realizar o preenchimento de falhas a partir de métodos aplicáveis, e verificar a consistência dos dados após o preenchimento. Portanto, o objetivo deste trabalho é realizar o preenchimento de falhas de dados de precipitação de 24 horas da rede de monitoramento da Defesa Civil do Município de Joinville (DCMJ), SC, aplicando os métodos de Interpolação Linear, Regressão Linear Múltipla, Método das Correlações, Redes Neurais Artificiais e Imputação Múltipla, todos eles utilizando o *software* R. A metodologia consiste em aplicar dois preenchimentos para cada método, o primeiro para preencher as falhas originais oriundas de problemas nos equipamentos, e o segundo para preencher somente os dados reais que foram retirados. As falhas da rede de monitoramento da DCMJ, que possui dez estações pluviométricas, todas localizadas na planície, foram preenchidas e analisadas através das seguintes medidas de erro: Raiz do erro quadrático médio (RMSE), Erro relativo médio (MRE), e Erro médio absoluto (MAE), além do Coeficiente de determinação ( $R^2$ ). A consistência dos dados após o preenchimento foi verificada pelo Método da Dupla Massa. Os melhores resultados tanto para as medidas de erro, quanto para o  $R^2$ , foram alcançados com o Método das Correlações e com o Método da Regressão Linear Múltipla, com valores de RMSE = 0,617 a 14,333 mm, MRE = 0,009 a 4,005, MAE = 0,035 a 7,204 mm e  $R^2$  = 0,003 a 1. Embora, todas as estações tenham demonstrado consistência nos dados após a aplicação dos métodos de preenchimento de falhas, atingindo valores de  $R^2$  no Método da Dupla Massa entre 0,834 e 0,999.

**Palavras-chave:** Precipitação. Preenchimento de falhas de dados. Dados de 24 horas.

## ABSTRACT

Times series, it is essential at the development of hydrological researches to perform statistical analysis and probability. However, the greater difficulties are the unavailability of the data and the amount of gaps found in these time series provided by the network of rainfall stations. Due to these concerns, it is necessary to fill gap using statistical methods, and check the consistency of the data soon after. The major aim in this paper is to fill gaps in 24 hour precipitation data of Joinville Civil Defense's (JCD), SC, applying Linear Interpolation, Multiple Linear Regression, Method of Correlations, Artificial Neural Networks and Multiple Imputation methods, all of them using R software. The methodology consists of applying two fills to each method, the first to fill the original gaps caused by equipment problems, and the second one to fill the gaps that were simulated. The gaps of the JCD monitoring network, which has ten rainfall stations, was completed and analyzed using the following error measures: Root Mean Square Error (RMSE), Mean Relative Error (MRE), and Mean Absolute Error (MAE) besides the coefficient of determination ( $R^2$ ). The Double Mass Method verified the consistency of the data after filling. The best results for both error measurements and  $R^2$  were obtained with the Correlation and Multiple Linear Regression methods, however, all the stations showed consistency in the data after the application of the fault filling methods

**Keywords** – Precipitation; Filling gap data; 24 hour data.

## LISTA DE FIGURAS

Figura 1 - Número de ocorrência de desastres por continente de 2000 a 2015	25
Figura 2 – Esquema geral de elementos presentes em uma rede neural artificial padrão.....	39
Figura 3 - Mapa de pluviosidade de Santa Catarina.....	47
Figura 4 - Delimitação das Bacias Hidrográficas de Joinville e Localização das Estações da Rede de Monitoramento de Joinville – SC.....	48
Figura 5 - Histogramas de frequência de cada ano hidrológico.....	66
Figura 6 – Histograma de frequência para as estações estudadas.....	67
Figura 7 - RMSE para o primeiro ano hidrológico.....	84
Figura 8 - RMSE para o segundo ano hidrológico.....	85
Figura 9 - RMSE para o terceiro ano hidrológico.....	87
Figura 10 - RMSE para o quarto ano hidrológico.....	88
Figura 11 - MRE para o primeiro ano hidrológico.....	89
Figura 12- MRE para o segundo ano hidrológico.....	91
Figura 13- MRE para o terceiro ano hidrológico.....	92
Figura 14- MRE para o quarto ano hidrológico.....	93
Figura 15 - MAE para o primeiro ano hidrológico.....	94
Figura 16 - MAE para o segundo ano hidrológico.....	95
Figura 17 - MAE para o terceiro ano hidrológico.....	96
Figura 18 - MAE para o quarto ano hidrológico.....	97
Figura 19 - R <sup>2</sup> para o primeiro ano hidrológico.....	99
Figura 20 - R <sup>2</sup> para o segundo ano hidrológico.....	100
Figura 21 - R <sup>2</sup> para o terceiro ano hidrológico.....	101
Figura 22 - R <sup>2</sup> para o quarto ano hidrológico.....	102
Figura 23 - Dupla Massa para Estação Bucarein sem preenchimento de Outubro de 2012 até Setembro de 2016.....	104
Figura 24 - Dupla Massa para Estação Bucarein preenchido por Imputação Múltipla de Outubro de 2012 até Setembro de 2016.....	104
Figura 25 - Dupla Massa para Estação Bucarein preenchido por Interpolação Linear de Outubro de 2012 até Setembro de 2016.....	105
Figura 26 - Dupla Massa para Estação Bucarein preenchido pelo Método das Correlações de Outubro de 2012 até Setembro de 2016.....	105
Figura 27 - Dupla Massa para Estação Bucarein preenchido por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016.....	106
Figura 28 - Dupla Massa para Estação Bucarein preenchido por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016.....	106
Figura 29 - Dupla Massa para Estação JIC sem preenchimento de Outubro de 2012 até Setembro de 2016.....	107
Figura 30 - Dupla Massa para Estação JIC preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016.....	108
Figura 31 - Dupla Massa para Estação JIC preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016.....	108
Figura 32 - Dupla Massa para Estação JIC preenchido pelo Método das Correlações de Outubro de 2012 até Setembro de 2016.....	109
Figura 33 - Dupla Massa para Estação JIC preenchido por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016.....	109
Figura 34 - Dupla Massa para Estação JIC preenchido por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016.....	110

Figura 35 - Dupla Massa para Estação Jativoca sem preenchimento de Outubro de 2012 até Setembro de 2014 .....	111
Figura 36 - Dupla Massa para Estação Jativoca preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2014 .....	111
Figura 37 - Dupla Massa para Estação Jativoca preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2014 .....	112
Figura 38 - Dupla Massa para Estação Jativoca preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2014 .....	112
Figura 39 - Dupla Massa para Estação Jativoca preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2014 .....	113
Figura 40 - Dupla Massa para Estação Jativoca preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2014 .....	113
Figura 41 - Dupla Massa para Estação FlotFlux sem preenchimento de Outubro de 2012 até Setembro de 2016 .....	114
Figura 42 - Dupla Massa para Estação FlotFlux preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016 .....	115
Figura 43 - Dupla Massa para Estação FlotFlux preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016 .....	115
Figura 44 - Dupla Massa para Estação FlotFlux preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016 .....	116
Figura 45 - Dupla Massa para Estação FlotFlux preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016 .....	116
Figura 46 - Dupla Massa para Estação FlotFlux preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016 .....	117
Figura 47 - Dupla Massa para Estação Cubatão sem preenchimento de Outubro de 2012 até Setembro de 2016 .....	118
Figura 48 - Dupla Massa para Estação Cubatão preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016 .....	118
Figura 49 - Dupla Massa para Estação Cubatão preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016 .....	119
Figura 50 - Dupla Massa para Estação Cubatão preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016 .....	119
Figura 51 - Dupla Massa para Estação Cubatão preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016 .....	120
Figura 52 - Dupla Massa para Estação Cubatão preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016 .....	120
Figura 53 - Dupla Massa para Estação Guanabara sem preenchimento de Outubro de 2012 até Setembro de 2016 .....	121
Figura 54 - Dupla Massa para Estação Guanabara preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016 .....	122
Figura 55 - Dupla Massa para Estação Guanabara preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016 .....	122
Figura 56 - Dupla Massa para Estação Guanabara preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016 .....	123
Figura 57 - Dupla Massa para Estação Guanabara preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016 .....	123
Figura 58 - Dupla Massa para Estação Guanabara preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016 .....	124
Figura 59 - Dupla Massa para Estação Itaum sem preenchimento de Outubro de 2012 até Setembro de 2016 .....	125

Figura 60 - Dupla Massa para Estação Itaum preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016.....	125
Figura 61 - Dupla Massa para Estação Itaum preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016.....	126
Figura 62 - Dupla Massa para Estação Itaum preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016.....	126
Figura 63 - Dupla Massa para Estação Itaum preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016.....	127
Figura 64 - Dupla Massa para Estação Itaum preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016 .....	127
Figura 65 - Dupla Massa para Estação Paraíso sem preenchimento de Outubro de 2012 até Setembro de 2016 .....	128
Figura 66 - Dupla Massa para Estação Paraíso preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016.....	129
Figura 67 - Dupla Massa para Estação Paraíso preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016.....	129
Figura 68 - Dupla Massa para Estação Paraíso preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016 .....	130
Figura 69 - Dupla Massa para Estação Paraíso preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016 .....	130
Figura 70 -Dupla Massa para Estação Paraíso preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016 .....	131
Figura 71 - Dupla Massa para Estação Rodovia do Arroz sem preenchimento de Outubro de 2012 até Setembro de 2014.....	132
Figura 72 - Dupla Massa para Estação Rodovia do Arroz preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2014 .....	132
Figura 73 - Dupla Massa para Estação Rodovia do Arroz preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2014.....	133
Figura 74 - Dupla Massa para Estação Rodovia do Arroz preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2014 .....	133
Figura 75 - Dupla Massa para Estação Rodovia do Arroz preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2014 .....	134
Figura 76 - Dupla Massa para Estação Rodovia do Arroz preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2014 .....	134
Figura 77 - Dupla Massa para Estação Unidade de Obras sem preenchimento de Outubro de 2012 até Setembro de 2016.....	135
Figura 78 - Dupla Massa para Estação Unidade de Obras preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016 .....	136
Figura 79 - Dupla Massa para Estação Unidade de Obras preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016.....	136
Figura 80 - Dupla Massa para Estação Unidade de Obras preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016 .....	137
Figura 81 - Dupla Massa para Estação Unidade de Obras preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016 .....	137
Figura 82 - Dupla Massa para Estação Unidade de Obras preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016 .....	138

## LISTA DE TABELAS

Tabela 1 - Localização das estações pluviométricas da rede de monitoramento da Defesa Civil do município de Joinville - SC.....	49
Tabela 2 – Distância, em metros, entre as estações da rede de monitoramento da Defesa Civil.....	50
Tabela 3 – Falhas e simulações do primeiro ano hidrológico .....	53
Tabela 4 - Falhas e simulações do segundo ano hidrológico .....	54
Tabela 5 - Falhas e simulações do terceiro ano hidrológico .....	54
Tabela 6 - Falhas e simulações do quarto ano hidrológico .....	55
Tabela 7- Valores discrepantes identificados na estação Unidade de Obras..	62
Tabela 8 - Valores discrepantes identificados na estação Paraíso.....	63
Tabela 9 - Valor discrepante identificado na estação JIC .....	63
Tabela 10 - Valor discrepante identificado na estação Rodovia do Arroz.....	63
Tabela 11 - Número de falhas para cada estação por ano hidrológico (Continua)	64
Tabela 13 - Medidas de erros para cada estação em todos os 4 anos hidrológicos para o método de regressão linear múltipla (Continua) .....	72
Tabela 14 - Medidas de erros para cada estação em todos os 4 anos hidrológicos para o método das correlações.....	75
Tabela 15 - Medidas de erros para cada estação em todos os 4 anos hidrológicos para o método das redes neurais artificiais (Continua) .....	78
Tabela 16 - Medidas de erros para cada estação em todos os 4 anos hidrológicos para o método de imputação múltipla .....	81
Tabela 17 - Erro RMSE para o primeiro ano hidrológico .....	83
Tabela 18- Erro RMSE para o segundo ano hidrológico .....	85
Tabela 19 - Erro RMSE para o terceiro ano hidrológico .....	86
Tabela 20 - Erro RMSE para o quarto ano hidrológico, unidade em mm .....	87
Tabela 21 - Erro MRE para o primeiro ano hidrológico.....	89
Tabela 22 - Erro MRE para o segundo ano hidrológico .....	90
Tabela 23- Erro MRE para o terceiro ano hidrológico.....	91
Tabela 24 - Erro MRE para o quarto ano hidrológico .....	92
Tabela 25- Erro MAE para o primeiro ano hidrológico, unidade em mm .....	94
Tabela 26 - Erro MAE para o segundo ano hidrológico .....	95
Tabela 27 - Erro MAE para o terceiro ano hidrológico .....	96
Tabela 28 - Erro MAE para o quarto ano hidrológico.....	97
Tabela 29 - Coeficiente de determinação para o primeiro ano hidrológico .....	98
Tabela 30 - Coeficiente de determinação para o segundo ano hidrológico .....	99
Tabela 31 - Coeficiente de determinação para o terceiro ano hidrológico .....	101
Tabela 32 - Coeficiente de determinação para o quarto ano hidrológico.....	102

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>23</b>
1.1	OBJETIVOS .....	26
1.1.1	Geral.....	26
1.1.2	Específicos .....	26
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA.....</b>	<b>28</b>
2.1	REDES DE MONITORAMENTO DE PRECIPITAÇÃO.....	28
2.2	SÉRIES DE DADOS .....	30
2.3	SOFTWARE R.....	32
2.4	MÉTODOS DE PREENCHIMENTO DE FALHAS DE DADOS DE PRECIPITAÇÃO.....	33
2.4.1	Interpolação Linear .....	34
2.4.2	Regressão Linear Múltipla - RLM .....	35
2.4.3	Método das correlações .....	36
2.4.4	Redes Neurais Artificiais - RNAs .....	38
2.4.5	Imputação múltipla .....	41
2.5	CONSISTÊNCIA DOS DADOS .....	42
2.6	MEDIDAS DE ERRO .....	43
2.7	COEFICIENTE DE DETERMINAÇÃO – $R^2$ .....	45
<b>3</b>	<b>METODOLOGIA .....</b>	<b>46</b>
3.1	ÁREA DE ESTUDO .....	46
3.2	REDE DE MONITORAMENTO.....	46
3.3	SOFTWARES UTILIZADOS .....	50
3.4	CRITÉRIOS UTILIZADOS .....	51
3.5	SÉRIE DE DADOS .....	51
3.6	APLICAÇÃO DOS MÉTODOS DE PREENCHIMENTO DE FALHAS .....	52
3.7	MÉTODOS ESTATÍSTICOS UTILIZADOS .....	55
3.7.1	Interpolação linear .....	55
3.7.2	Regressão Linear Múltipla .....	56
3.7.3	Método das correlações .....	56
3.7.4	Redes neurais artificiais.....	57
3.7.5	Método de Imputação múltipla por equações vinculadas.....	59
3.8	CONSISTÊNCIA E HOMOGENEIDADE DOS DADOS .....	60



<b>4</b>	<b>RESULTADOS E DISCUSSÕES .....</b>	<b>62</b>
4.1	ANÁLISE EXPLORATÓRIA DOS DADOS .....	62
4.2	MÉTODOS DE PREENCHIMENTO DE FALHAS .....	69
4.2.1	Interpolação linear.....	69
4.2.2	Regressão linear múltipla.....	71
4.2.3	Método das Correlações.....	74
4.2.4	Redes Neurais Artificiais .....	77
4.2.5	Imputação Múltipla – Pacote mice .....	80
4.3	MEDIDAS DE ERRO.....	83
4.3.1	Raiz do Erro Quadrático Médio - RMSE.....	83
4.3.2	Erro Médio Relativo - MRE.....	88
4.3.3	Erro Médio Absoluto - MAE .....	93
4.4	COEFICIENTE DE DETERMINAÇÃO - $R^2$ .....	98
4.5	ANÁLISE DE DUPLA MASSA.....	103
<b>5</b>	<b>CONCLUSÃO.....</b>	<b>139</b>
<b>6</b>	<b>LIMITAÇÕES E SUGESTÕES DE TRABALHOS FUTUROS.....</b>	<b>141</b>
APÊNDICE A – Rotina no R para o método de Imputação Múltipla .....		142
APÊNDICE B – Rotina no R para o método de Interpolação Linear .....		144
APÊNDICE C – Rotina no R para o Método das Correlações .....		145
APÊNDICE D – Rotina no R para o método de Redes Neurais Artificiais .....		147
APÊNDICE E – Rotina no R para o método de Regressão Linear Múltipla.....		152
REFERÊNCIAS.....		153

## 1 INTRODUÇÃO

A precipitação é um processo natural que ocorre por meio de diferentes mecanismos e podem ser classificadas em: convectivas, orográficas e frontais. As chuvas convectivas são provocadas pelo aquecimento desigual da superfície do solo, que faz com que a massa de ar ascenda, se resfrie e precipite. Costumam ser violentas, de curta duração e alta intensidade. As chuvas orográficas ocorrem devido aos ventos quentes e úmidos vindos do mar para a terra, sendo elevados quando encontram acidentes geográficos naturais como montanhas, e resfriados dando origem a precipitação que geralmente é de baixa intensidade e média duração, muito comum no litoral catarinense e Serra do Mar. Já as chuvas frontais (frias e quentes), originadas nos polos e Equador respectivamente, são provocadas pelo choque entre as massas de ar que se deslocam, resfriam e por fim precipitam. São em geral de grande duração e média intensidade (BARBOSA, 2006 e CARVALHO; SILVA, 2006).

Segundo Reboita et al. (2009) a América do Sul, por sua grande extensão latitudinal e diferentes formas de relevo, permite a existência de 8 diferentes regimes de precipitação, dos quais 4 deles se encontram no Brasil, devido à sua extensão territorial. Na região Sul do país, por exemplo, devido à sua localização latitudinal, a maior influência é dos sistemas de latitudes médias, onde os sistemas frontais são os principais causadores de chuvas durante o ano. Ao longo de quase todo o território do Sul do Brasil, a média anual da precipitação varia de 1250 a 2000 mm (CLIMANÁLISE, 2009). Em Santa Catarina, Monteiro (2001) *apud* UFSC (2011) relatam que a Massa Tropical Atlântica e a Massa Polar Atlântica são frequentes nos meses de inverno formando as chuvas frontais e no verão predominam as chuvas convectivas. Os autores destacam ainda que o relevo do Estado é um contribuinte fundamental para a distribuição da precipitação diversificada e os maiores índices de precipitação se encontram em municípios próximos à encosta da Serra Geral.

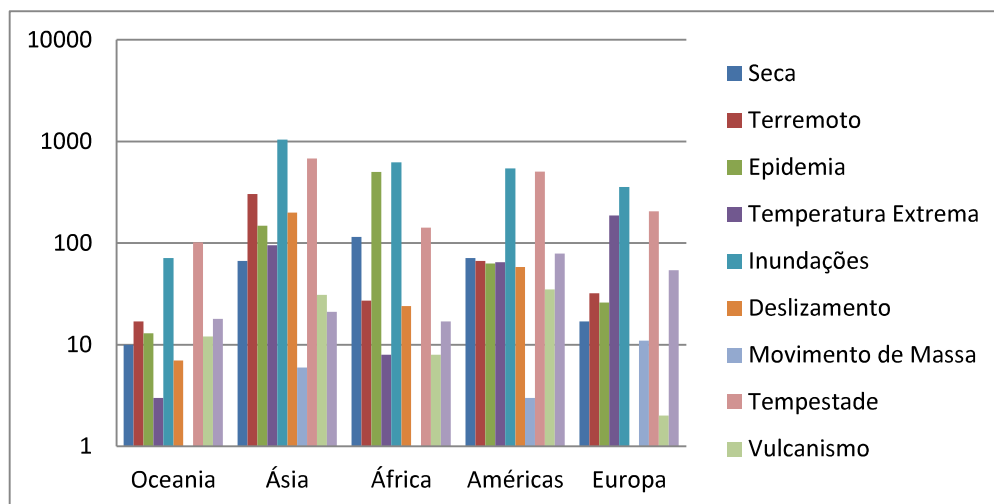
Conhecer o regime de precipitação de uma determinada região, bem como o regime hídrico de uma bacia hidrográfica, é bastante significativo tanto no que diz respeito ao manejo de bacias hidrográficas para estimar, por exemplo, a necessidade de abastecimento de água e irrigação (MELLO, 2012) quanto à importância dessa variável meteorológica para estudos sobre a prevenção dos efeitos relacionados ao seu excesso (MONTEIRO, 2000 *apud* SOUZA, 2008).

Os efeitos causados pelo excesso de precipitação se encaixam em uma das categorias de desastres. Os desastres que ocorrem no mundo podem ser classificados em diferentes tipologias e atualmente estão sendo designados também de eventos extremos, pois alguns autores questionam a nomenclatura desastre natural (Dias, 2011), já que muitos dos desastres que ocorrem no mundo são induzidos por ações antrópicas associadas aos fenômenos naturais. No glossário da Estratégia Internacional para Redução de Desastres, elaborado pela Organização das Nações Unidas (ONU-EIRD, 2004), a palavra desastre está definida como:

“séria interrupção do funcionamento de uma comunidade ou sociedade que causa perdas humanas e/ou importantes perdas materiais, econômicas ou ambientais; que excedem a capacidade da comunidade ou sociedade afetada de lidar com a situação utilizando seus processos de risco. Resulta da combinação de ameaças, condições de vulnerabilidade e insuficiente capacidade ou medidas para reduzir as consequências negativas e potenciais do risco.”

Ao realizar uma análise dos tipos de desastres que ocorreram no mundo entre 2000 e 2015 (EM-DAT, 2016), como mostra a Figura 1, dividindo o globo em continentes, pode-se perceber uma vulnerabilidade bastante alta de todos os continentes às inundações. De modo geral, na Europa, Ásia e África, as inundações se sobressaem perante aos demais tipos de eventos extremos, já na Oceania e América, as tempestades são superiores e igualadas às inundações, respectivamente. Quanto aos deslizamentos, que também são frequentes, porém numa escala menor, o número de ocorrência também está disposto na Figura 1.

Figura 1 - Número de ocorrência de desastres por continente de 2000 a 2015



Fonte: EM-DAT: OFDA/CRED International Disaster Database - Université Catholique de Louvain, Belgium (2016, tradução nossa).

Dias (2011) afirma que apesar do aumento da complexidade dos modelos e cenários consistentes ao clima, a simulação dos eventos extremos ainda é precária. Esses eventos englobam o estudo de processos hidrológicos estocásticos, como a precipitação, que são governados por leis de probabilidade, porém contém componentes aleatórias que estão acima das regularidades eventualmente explicáveis (NAGHETTINI; PINTO, 2007).

Os estudos que tratam de dados hidrológicos, em especial a precipitação, trabalham em conjunto com a probabilidade e a estatística, e por isso, devem contar com séries históricas consistentes. Porém, frequentemente não se dispõe de séries suficientemente grandes para estudar os processos hidrológicos (LANNA, p. 80, 1997). Além da não disponibilidade de séries históricas, outro contratempo que dificulta a análise dos pesquisadores é o número de falhas encontrado nas séries. Mesmo que o objetivo dos postos de medição de chuvas seja justamente obter uma série com o menor número de falhas possível de precipitação ao longo dos anos (BERTONI, 1997), é inevitável a existência de períodos sem informações ou com falhas nas observações. Em função disso, é necessário realizar o preenchimento de falhas e verificar a consistência dos dados após o preenchimento (TEEGAVARAPU; CHANDRAMOULI, 2005; COULIBALY;

EVORA, 2007; TEEGAVARAPU; TUFAIL; ORMSBEE, 2009; MWALE; ADELOYE; RUSTUM, 2012; BÁRDOSSY; PEGRAM, 2014).

A Agência Nacional de Águas - ANA (2012) apresenta as diretrizes para consistência de dados pluviométricos baseada principalmente nos órgãos Departamento Nacional de Águas e Energia Elétrica – DNAEE e *World Meteorological Organization – WMO*. No relatório de diretrizes, a ANA afirma que a análise de consistência dos dados pluviométricos tem como objetivo a identificação e correção de erros, bem como o preenchimento de falhas das séries pluviométricas. Ainda no mesmo relatório, sobre a condição espacial da precipitação, o órgão expressa a necessidade de se analisar os dados de estações pluviométricas adjacentes para permitir o preenchimento de falhas nos registros ou a substituição de dados observados e considerados errôneos.

## 1.1 OBJETIVOS

### 1.1.1 Geral

Verificar qual o melhor método de preenchimento de falhas de dados de 24 horas para a região da cidade de Joinville, em Santa Catarina, que sofre inundações desde o início do seu desenvolvimento (SILVEIRA, p.2, 2008), visando uma análise comparativa das diferentes metodologias aplicáveis ao preenchimento de falhas de dados de precipitação das estações da rede de monitoramento da Defesa Civil Municipal.

### 1.1.2 Específicos

Para alcançar o propósito final, alguns objetivos específicos são importantes:

- a) localizar geograficamente as estações pluviométricas da rede de monitoramento da Defesa Civil em Joinville;
- b) obter dados de precipitação das estações;
- c) identificar dados faltantes ou com erros nas séries das estações;
- d) preparar dados obtidos;
- e) preencher falhas dos dados com diferentes métodos utilizando *software* livre;
- f) validar o método com menor erro para os dados de precipitação para região de Joinville;

g) reconstruir as séries para uso dos dados;

Com estes objetivos traçados espera-se alcançar o escopo do trabalho ao final do desenvolvimento desta pesquisa.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 REDES DE MONITORAMENTO DE PRECIPITAÇÃO

A precipitação pode exercer forte influência sobre as condições ambientais, como por exemplo, eventos extremos de chuva que causam inundações intensas e movimentos de massa resultando em prejuízos socioeconômicos (AYOADE, 1998; TUCCI, 2001; CAVALCANTI *et al.*, 2009; MELLO, 2015), o que justifica a importância de uma rede de monitoramento para essa variável hidrológica.

A hidrometria é a área da hidrologia que permite fazer medições sistemáticas das variáveis hidrológicas, tornando possível acompanhar em que condições se encontram as bacias hidrográficas, e por sua vez, efetuar pesquisas relacionadas aos recursos hídricos (BLAINSKI, GARBOSSA e ANTUNES, p.7, 2012); e a hidrometeorologia é definida como um estudo científico da interação entre fenômenos meteorológicos e hidrológicos, incluindo basicamente a transferência de água e energia da superfície para a atmosfera (HOUGHTON MIFFLIN, 2002).

Outro fator relevante é o monitoramento hidrometeorológico realizado por redes de estações hidrométricas. É um recurso importante de gerenciamento de dados, composto por ações de coleta, tratamento, armazenamento, recuperação e disponibilização de informações históricas referentes às condições atmosféricas e vazão dos rios, além de constituir uma etapa fundamental no gerenciamento de recursos hídricos (FLORIANÓPOLIS, p. 2, 2012). O monitoramento hidrometeorológico apesar de ser complexo devido à variabilidade dos fenômenos de interesse no tempo e no espaço, é uma prática comum dentro da hidrologia, e tem como principal objeto estudar a água na Terra, sua ocorrência, circulação e distribuição, suas propriedades físicas e químicas e sua relação com o meio ambiente, permitindo coletar informações para criar uma base de dados auxiliando na tomada de decisões, e também, determinar o regime hidrológico da região (BLAINSKI, GARBOSSA e ANTUNES, 2012).

As estações das redes de monitoramento são projetadas para proporcionar a melhor representação possível de um estado de campo meteorológico que varia continuamente no tempo e no espaço (KAMEL, H.F., SLIMANI, M., CUDENNEC, C.

2010). E dependendo dos tipos de dispositivos que são associados a essas estações é possível classificá-las em convencionais ou automáticas.

Para alguns autores e órgãos responsáveis pelo monitoramento hidrometeorológico nacional, as redes de estações hidrométricas podem ser classificadas em dois grandes grupos: automáticas e convencionais (ANA, 2012; BLAINSKI, GARBOSSA e ANTUNES, p. 8, 2012). Portanto, o Instituto Nacional de Meteorologia – INMET (2011) ainda considera uma terceira classe, a rede de observação de altitude ou radiossonda, descrita como um conjunto de equipamentos e sensores utilizados para medir a temperatura, umidade relativa e pressão atmosférica. Apesar da divisão, os autores concordam quanto à definição de cada uma delas, sendo as estações automáticas, aquelas dotadas de um *datalogger* para armazenar os dados que se ligam aos sensores eletrônicos específicos para cada parâmetro meteorológico a ser monitorado, seja ele pressão atmosférica, temperatura, precipitação, radiação solar, direção e velocidade do vento entre outros. O INMET (2011) considera estações convencionais como as que registram os mesmos parâmetros meteorológicos das estações automáticas, porém de forma isolada. As estações convencionais têm os dados registrados por um observador, responsável por enviar as informações a um centro coletor. Em Santa Catarina, as estações mais usuais instaladas pela Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (Epagri) por intermédio do Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina (Ciram), são divididas em quatro grupos: pluviométricas automáticas, agrometeorológicas, fluviométricas automáticas e meteorológicas (BLAINSKI, GARBOSSA e ANTUNES, p. 8, 2012).

As estações requerem alguns dispositivos para obter, armazenar e transmitir os dados. Braga (2005) afirma que os dados são produzidos mediante sensores eletrônicos específicos para cada parâmetro meteorológico a ser monitorado, enquanto os *dataloggers* possuem função restrita ao armazenamento de dados em campo, onde o registro é feito de maneira digital, diminuindo o intervalo de tempo de visita ao posto de medição. Para a transmissão dos dados, seja ela via rádio, telefone ou satélite, existem configurações que podem ser adicionadas às estações permitindo um monitoramento em tempo real. E ainda, para eventos extremos, dispositivos de rádio alerta podem ser



programados caso seja necessário. A tecnologia de transmissão dos dados, segundo ANA (2012) é uma escolha livre da entidade responsável por operar as estações, porém, alguns critérios como: frequência de aquisição e transmissão de uma hora, e dados enviados em tempo real são exigidos. Essa escolha da tecnologia de transmissão pode variar de acordo com a localização, disponibilidade de serviços além de custo de implantação e manutenção.

Quanto à manutenção e cuidado com os dispositivos, os sensores eletrônicos fazem parte da ligação mais frágil do processo eletrônico de medição, pois podem sofrer interferências externas atrapalhando ou até inviabilizando a operação dos mesmos, exigindo constante calibração devido à sua sensibilidade térmica e fadiga dos amplificadores eletrônicos (BRAGA, 2005). Em geral as estações podem apresentar problemas gerados no equipamento de transmissão de dados e falta de manutenção dos dispositivos acoplados, ou seja, o desempenho das estações automáticas depende da integração adequada de seus componentes, que por muitas vezes provém de fabricantes diferentes, e apesar da facilidade de aquisição desses equipamentos, o desafio passa a ser a padronização, manutenção e operação de grandes redes automatizadas (ANA, 2012).

## 2.2 SÉRIES DE DADOS

As variáveis hidrológicas descrevem as variações temporais e espaciais dos fenômenos do ciclo hidrológico (NAGHETTINI; PINTO, p. 6-8, 2007), e as séries temporais registram essa variabilidade (NAGHETTINI; PINTO, p. 6-8, 2007; RENNO e BORMA, 2016). Tucci (1997) relata que as sequências hidrológicas não seguem um padrão de comportamento devido à variabilidade natural da precipitação.

Para estudos hidrológicos é comum trabalhar com o período denominado ano hidrológico, que difere do ano civil. O ano hidrológico é descrito como: "período contínuo de doze meses durante o qual ocorre um ciclo anual climático completo e que é escolhido por permitir uma comparação mais significativa dos dados meteorológicos" (DNAEE, 1976). Alguns autores ainda definem o ano hidrológico como sendo o ano que inicia no período chuvoso e vai até o fim do período seco, contabilizando 12 meses fixos

(NAGHETTINI; PINTO, 2007; TUCCI, 2012; SIMON; PICKBRENNER; MARCUZZO, 2013).

Sobre a classificação das séries hidrológicas, Naghettini e Pinto (2007) e Rennó e Borma (2016) afirmam que quando uma série hidrológica apresenta uma tendência ou periodicidade ao longo do tempo ela é dita como não estacionária, caso contrário ela é estacionária. Ainda sobre a classificação das séries hidrológicas, os autores afirmam que elas podem ser homogêneas, quando é identificado um padrão de variabilidade ou heterogêneas na ausência desse padrão. Para Naghettini e Pinto (2007) essas séries temporais podem ser consideradas completas ou reduzidas, no primeiro caso as observações são coletadas em intervalos de tempo regular durante vários anos e nenhum registro é desprezado, já nas séries reduzidas são considerados apenas alguns dados de valores característicos, geralmente os máximos ou médias mensais de precipitação.

As séries hidrológicas históricas são objetos de estudos de muitos pesquisadores, seja para identificar tendências de precipitação (PINHEIRO *et al.*, 2013; PENEREIRO; ORLANDO, 2013, PORTELA *et al.*, 2011), analisar a homogeneidade dos dados (RIBEIRO *et al.*, 2016), ou para realizar previsões de precipitação (WANG; FENG; LIU, 2013) entre outros objetivos. No entanto, a presença de erros nessas séries é inevitável.

Alguns dos erros mais comuns para Rennó e Borma (2016) são denominados como pontuais ou isolados, quando há falha na leitura ou no armazenamento dos dados, e os erros sistemáticos ou grosseiros, causados principalmente devido à falta de calibração do aparelho. Na primeira categoria, os erros ocorrem de forma aleatória, e flutuam em torno do valor verdadeiro da variável de interesse (NAGHETTINI; PINTO, 2007). Os erros ditos como sistemáticos são os que apresentam um viés nos resultados e podem ser originados devido à mudança de técnica empregada na medição, por exemplo. E os erros grosseiros são usualmente provocados por falhas humanas, tanto na leitura incorreta dos instrumentos, quanto na falta de cuidado durante a execução de uma medição. De acordo com Chow *et al.* (p. 177, 1988), que sequenciaram a medição de fenômenos hidrológicos, os erros mais comuns para a variável precipitação são verificados na sincronização dos dados automáticos das medições e também nas informações gravadas que foram perdidas durante a transmissão dos dados.

Além dos erros, é possível também, identificar algumas falhas nas observações das séries temporais, como por exemplo, a falta de informações em determinados períodos, segundo Tucci (p. 182, 1997). Este mesmo autor afirma ainda que antes de utilizar essas séries é necessário submetê-las a uma análise para verificar os erros grosseiros que podem estar atribuídos a vários fatores como: interrupção de medidas por danificação do aparelho, obstrução próxima ao posto de medição, danificação do aparelho, problemas mecânicos, falha humana conforme relatam também os autores Elshorbagy, Panu e Simonovic (2000), entre outros. A análise e consequente identificação desses erros são na verdade uma preparação dos dados para o tratamento estatístico que antecede o processo de preenchimento de falhas.

Segundo Kar *et al.* (2015), autores que utilizaram técnicas de multicritério para tomada de decisão a fim de realizar um projeto de rede pluviométrica para previsão de enchentes na Índia, são várias as principais dificuldades encontradas para conseguir reunir informações de todos os pluviômetros instalados naquela região, entre elas: falhas, não gravação dos dados e a própria dificuldade de transmissão de informação.

Silveira (2012) relata a complexidade da dinâmica da chuva que difere do modelo da própria climatologia, confirmando assim a necessidade de determinar a homogeneidade das séries históricas de precipitação devido à importância do uso de dados confiáveis, registrados nas estações pluviométricas.

## 2.3 SOFTWARE R

O emprego de *softwares* e pacotes estatísticos para a análise de dados é cada vez mais utilizado, no entanto, devido ao alto custo de aquisição da licença desses programas, os pesquisadores estão sendo encorajados a optar pelos chamados *softwares* livres (MELLO; PETERNELLI, 2013).

Um dos *softwares* que vem se destacando na categoria de domínio público é o *R* (R CORE TEAM, 2017). Verzani (2011) alega que além do *R* ser gratuito, é um projeto de código-fonte aberto que possui ambiente de desenvolvimento integrado, o que significa dizer que pode ser elaborado e alterado a qualquer instante e por qualquer usuário, ou seja, é uma ferramenta respeitável na manipulação de dados por contar com colaboradores renomados.

O programa foi criado primeiramente como um projeto de pesquisa pelos autores Robert Gentleman e Ross Ihaka que tiveram influência de duas linguagens antigas, a linguagem S e a linguagem Scheme (SUSSMAN; STEELE JR., 1998; IHAKA; GENTLEMAN, 1996). O R abrange mais que sua linguagem núcleo, ele tem um sistema de repositório, conhecido como CRAN - *Comprehensive R Archive Network*, para usuários desenvolvedores de pacotes que complementam a distribuição base. Hoje o R tem funcionalidade para lidar com inúmeros problemas, inclusive análise de séries temporais, pois apesar de ser projetado em torno do seu núcleo de linguagem, ele permite a integração dos códigos compilados em C, C++, Fortran, Java entre outras (VERZANI, 2011).

Para Mello e Peternelli (2013) o *RStudio* é um *software* complementar ao R, que dispõe de um ambiente de desenvolvimento integrado, *Integrated Development Environment - IDE* da sigla em inglês, e é projetado para facilitar o trabalho do programa. Com a interface do *RStudio* por meio do *IDE*, o usuário tem, além da facilidade de utilizar os comandos, a grande vantagem de interagir com o R por intermédio do editor de textos fornecido (MELLO; PETERNELLI, 2013; VERZANI, 2011). Ou seja, de forma simplificada, o R tem seu funcionamento baseado nos comandos digitados pelo usuário, que são processados por meio dos pacotes armazenados na memória e os resultados são demonstrados na interface *RStudio* ou em um objeto na memória do programa.

## 2.4 MÉTODOS DE PREENCHIMENTO DE FALHAS DE DADOS DE PRECIPITAÇÃO

A continuidade das informações em séries históricas é comprometida pelas falhas nos registros de dados (OLIVEIRA *et al*, 2010). Essas falhas são comumente encontradas em séries de dados históricos de precipitação e para seu preenchimento existem diversos métodos. O objetivo do estudo de métodos para o preenchimento de falhas em séries temporais é contribuir para a geração de dados mais confiáveis que promovam o avanço de pesquisas climatológicas (MELLO, p. 22, 2015), pois, através de dados anteriores é possível prever futuras ocorrências de fenômenos que venham a se repetir por influência do clima (FERNANDEZ, 2007).

Os métodos de preenchimento de falhas visam completar as lacunas das séries históricas e posteriormente verificar a consistência dos dados. Em recursos hídricos,

conforme Elshorbagy, Panu e Simonovic (2000), as análises baseadas em regressão, séries temporais, redes neurais artificiais e interpolação tem sido as técnicas mais utilizadas para preenchimento de falhas de dados. Para Oliveira *et al* (2010), além desses métodos usuais devem ser incluídos ainda os de ponderação regional (PR), ponderação regional com base em regressões lineares (RS) e vetor regional (VR).

O pesquisador deve atentar quanto aos períodos de falhas, pois alguns métodos são aplicáveis para séries mensais e anuais, enquanto outros apenas para uma delas. Bertoni e Tucci (2007) *apud* Depiné *et al* (2014) afirmam que nenhum dos métodos usualmente empregados é satisfatório quando se pretende preencher dados de intervalos diários e horários, pois nessas escalas os padrões espaço-temporais do processo de chuva são complexos, não lineares e altamente variáveis, o que dificulta a aplicação de métodos clássicos.

Os métodos aplicados nesta pesquisa são descritos nos itens a seguir.

#### 2.4.1 Interpolação Linear

Uma interpolação linear é utilizada quando se tem a necessidade de obter um valor intermediário que não está tabelado, ou seja, encontrar um valor correspondente de  $y$  para um dado  $x$  não pertencente à tabela. Isto é resolvido obtendo-se uma função que relaciona as variáveis  $x$  e  $y$ , e essa função geralmente são as funções polinomiais (CAMPOS, F. F., 2007).

Para Campos (2007), a função polinomial mais simples é a de grau 1, que envolve a solução de um sistema de equações lineares. São conhecidos dois pontos-base  $(x_0, y_0)$  e  $(x_1, y_1)$ , com  $x_0 \neq x_1$ , pertencentes a uma função  $y = f(x)$ . Para estimar  $f(z)$ ,  $z \in (x_0, x_1)$  faz-se:  $f(x) \approx P_1(x) = a_0 + a_1x$ . O polinômio interpolador de grau 1,  $P_1(x)$ , passa pelos dois pontos-base gerando um sistema de equações lineares de ordem 2, e a solução desse sistema é:

$$a_1 = \frac{y_1 - y_0}{x_1 - x_0} \quad e \quad a_0 = y_0 - a_1x \quad (1)$$

sendo:

$a_1$ , o coeficiente angular.

$a_0$ , o coeficiente linear.

$x_0, y_0$ , as coordenadas do ponto-base 0.

$x_1, y_1$ , as coordenadas do ponto-base 1.

O polinômio interpolador pode ser sumarizado pela Equação (2):

$$P_1(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) \quad (2)$$

Entretanto, quando utilizada para preenchimento de dados faltantes, a interpolação linear não leva em consideração as estações vizinhas, ela apenas preenche os dados faltantes de uma estação a partir dos dados já registrados naquela estação. Isto é, têm-se 2 pontos-base conhecidos, um imediatamente antes da falha e um depois, e o método preenche o dado faltante com base na função polinomial de grau 1, utilizando esses 2 pontos.

#### 2.4.2 Regressão Linear Múltipla - RLM

Outra metodologia, também bastante desenvolvida, é a regressão linear múltipla. Essa metodologia pode ser adotada por compensar a falta de interação entre as estações na interpolação linear, pois na regressão múltipla, as informações pluviométricas do posto independente são correlacionadas com as correspondentes observações das estações vizinhas.

Montgomery e Runger (2009) explicam que o modelo de regressão linear múltipla utiliza mais de um regressor, ou seja, mínimo de duas variáveis independentes, porém, a variável resposta ou dependente pode estar relacionada a  $n$  variáveis independentes ou regressoras, podendo ser descrito pela equação:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (3)$$

sendo:

$Y_i$ , o valor estimado para a variável dependente no  $i$ -ésimo nível da variável independente.

$\beta_0$ , a constante de regressão (interseção do plano).

$\beta_1, \beta_2, \dots, \beta_n$ , os coeficientes parciais de regressão.

$X_1, X_2, \dots, X_n$ , o  $i$ -ésimo nível da variável independente  $X$ .

$\epsilon$ , o erro aleatório.

Logo, neste caso, a estação contendo falhas é preenchida com o valor predito pelo modelo de regressão que utiliza as estações vizinhas como variáveis independentes. Os autores explicam ainda que o cálculo dos parâmetros  $\beta_0, \beta_1, \beta_2 \dots \beta_n$  é estimado por Mínimos Quadrados utilizando uma abordagem matricial para facilitar as operações, e para medir a adequação do modelo alguns testes de hipótese são úteis. Esses testes verificam se existe uma relação linear entre a variável resposta  $y$  e um subconjunto de regressores  $x_n$ , muito semelhante ao procedimento usado na regressão linear simples. Um indicador que é bastante usado para medir a qualidade do ajustamento de modelos de regressão é o coeficiente de determinação, simbolizado por  $R^2$  detalhado no item 2.6.

A verificação da adequação do modelo exige uma análise residual. Os resíduos correspondem à diferença entre o valor observado e o valor predito pelo modelo, estes resíduos do modelo de regressão linear múltipla devem ser independentes e apresentar distribuição normal.

### 2.4.3 Método das correlações

Este método faz uso de correlações entre as estações que poderiam ser aproveitada em outros métodos, e do pacote *hyfo* e sua dinâmica de trabalho disponibilizada na interface *RStudio* em conjunto com o cálculo do coeficiente de correlação entre as variáveis de entrada, bem como gráficos produzidos para análise dos resultados. Esse método das correlações disponibilizado por meio do pacote *hyfo* é baseado em correlação e regressão linear, ou seja, ele utiliza duas metodologias mais simples para preencher os dados faltantes de uma estação.

O método da regressão linear é semelhante ao que está descrito no item 2.3.3, porém, com apenas a constante de regressão  $\beta_0$ , e um coeficiente angular da reta  $\beta$  conforme equação (4):

$$Y_i = \beta_0 + \beta x_i + \epsilon \quad (4)$$

Conforme Montgomery e Runger (2009) a correlação é uma das medidas de relação entre duas variáveis aleatórias e pode ser denotada por:

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (5)$$

onde:

$cov(X, Y)$  é a covariância entre as variáveis aleatórias  $X$  e  $Y$ .

$\sigma_X$  e  $\sigma_Y$  correspondem aos desvios padrão da variável  $X$  e  $Y$  respectivamente.

Os autores explicam que devido aos valores dos desvios padrão serem maiores que zero, a magnitude da correlação é dependente direta do valor da covariância, ou seja, se a covariância entre as variáveis for positiva a correlação entre elas será também positiva. Sendo assim, a correlação é uma grandeza usada para comparar as relações lineares entre pares de variáveis, com o intervalo:

$$-1 \leq \rho_{XY} \leq +1 \quad (6)$$

onde:

$\rho_{XY}$ , é a correlação entre a variável  $X$  e  $Y$ .

Nesse pacote *hyfo*, o *software* calcula o coeficiente de correlação ( $r$ ) para cada par de estações, e esse coeficiente pode ser escolhido para ser calculado no intervalo de tempo diário, mensal ou anual. Posteriormente é obtida a matriz de correlação e com base nessa matriz são identificadas as ordens de correlação para cada estação. Portanto, se for identificado um dado faltante em uma determinada estação num determinado dia, esse valor é preenchido com a estação de maior correlação. No caso do dado ser faltante também nessa estação, a estação com a segunda maior correlação é utilizada para o preenchimento e assim sucessivamente. Logo após é estimado o coeficiente linear a cada par de estações a partir da regressão linear simples. E então, é feito o preenchimento dos dados faltantes fundamentado nas ordens obtidas das correlações entre as estações.

A função final obtida para o dado faltante é o coeficiente linear multiplicado pelo valor lido na estação com maior correlação. Isto é, o método utiliza a correlação para identificar qual a melhor estação vizinha que pode ser utilizada para realizar o preenchimento de cada estação contendo dados faltantes. Diferentemente da interpolação linear, a função *fillGap* do pacote *hyfo* faz o preenchimento de forma instantânea para todas as estações (HIRSCH; MAIDMENT, 1992).



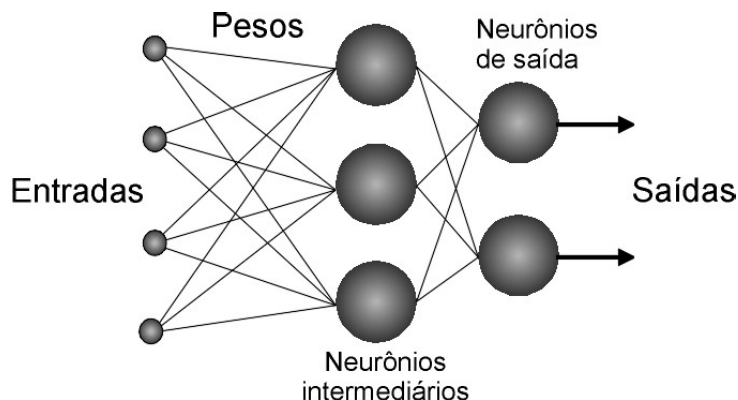
#### 2.4.4 Redes Neurais Artificiais - RNAs

As RNAs são amplamente aplicadas, tanto para a previsão de chuva quanto para o preenchimento de falhas em diferentes lugares do mundo, Nastos *et al* (2013), Chattopadhyay e Chattopadhyay (2008), Belayneh e Adamowski (2012), Nkuna e Odiyo (2011) entre outros. Além de não ter restrição quanto ao intervalo temporal, os resultados dos estudos de Teegavarapu e Chandramouli (2005), bem como os dos autores Coulibaly e Evora (2007) na comparação de diferentes RNA's para preenchimento de registros diários de temperatura e precipitação se mostraram bastante promissores em relação a outros métodos de preenchimento de falhas.

Mc Culloch e Pits (1943) foram os pioneiros nos estudos sobre neurocomputação ao realizarem uma simulação matemática sobre células nervosas vivas para reproduzir o comportamento do neurônio natural, definindo essa metodologia, hoje conhecida como rede neural artificial, como um paradigma computacional que está apto a imitar a estrutura biológica do cérebro.

De acordo com Nkuna e Odiyo (2011) as redes neurais *feed-forward* empregadas para modelagem de recursos hídricos incluem os tipos de *Multilayer Perceptrons (MLP)* e *Radial Basis Function Neural Networks (RBF)*. De forma simplificada, conforme ilustrado na Figura 2, as redes neurais artificiais de função MLP são estruturadas em 3 camadas. Na camada de entrada são inseridas as variáveis de entrada independentes, na camada oculta (intermediária), onde acontecem os processamentos está representado o número de neurônios que segundo a teoria de HELCHT-NIELSEN *apud* De Deus (2012) pode ser  $2n + 1$ , onde  $n$  é o número de variáveis de entrada. E a camada de saída apresenta o resultado final que foi processado na rede (FERNANDES, 2003; De DEUS, 2012).

Figura 2 – Esquema geral de elementos presentes em uma rede neural artificial padrão



Fonte: De DEUS (2012).

Com a capacidade de gerenciar conjuntos de dados complexos e extensos (Şahin et al., 2013) a aplicabilidade da inteligência artificial como método de preenchimento de falhas vem apresentando resultados promissores (TEEGAVARAPU *et al.*, 2009; WANDERLEY *et al.*, 2014).

A RNA não necessita informações detalhadas sobre as entradas, ela aprende, por meio de processo iterativo das variáveis, as relações entre os dados e verifica as tendências anteriores com a aplicação de um algoritmo que minimiza o erro quadrático médio a cada iteração, produzindo assim valores de saída associados a uma previsão (DEO; ŞAHIN, 2015). De Deus (2012) relata que a função de ativação de uma rede neural é responsável por gerar a resposta dos neurônios a partir dos valores de vetores de peso e de entrada. E o peso utilizado na rede é basicamente a influência que ele tem sobre o sinal de saída, onde valores negativos significam fatores de inibição e valores positivos são fatores de reforço do sinal de entrada.

Os resultados produzidos pelas RNAs podem ser avaliados por meio de diferentes estatísticas que mensuram a diferença entre os valores simulados pelo modelo e os valores observados. Essas estatísticas são: o coeficiente *Nash-Sutcliffe-NS* (DEPINÉ *et al.*, 2013; NKUNA; ODIYO, 2011), a Raiz do Erro Quadrático Médio – REQM (ABBOT; MAROHASY, 2014), ou apenas o Erro Quadrático Médio – EQM (MISLAN *et al.*, 2015; MEKANIK *et al.*, 2013).

Depiné *et al* (2013) testaram uma metodologia com RNAs para preencher falhas em séries horárias de dados de precipitação, utilizando 9 estações pluviométricas em uma bacia com 78 km<sup>2</sup> situada no Rio Grande do Sul. Os autores experimentaram três alternativas quanto à extensão e disposição das séries horárias e verificaram que a série longa, que continha aproximadamente 365 dias, apresentou dificuldades na compatibilidade dos dados gerados, pois apesar de reproduzir resultados coerentes, em alguns momentos os valores calculados eram bastante divergentes dos observados nos postos vizinhos. No estudo com a série longa o coeficiente NS foi inferior a 0,85, representando um resultado não suficiente, já que quanto mais próximo o NS é da unidade, melhores os resultados obtidos com as simulações. Os melhores resultados, cujos coeficientes NS superaram a marca de 0,9, foram obtidos com as séries curtas de 12 a 30 dias, concluindo que mesmo em bacias com precipitações consideradas homogêneas, as relações entre os postos pluviométricos não são constantes.

Assim como Depiné *et al* (2013), Nkuna e Odiyo (2011) também utilizaram a metodologia RNA para preencher dados faltantes, e verificaram por meio do coeficiente NS que apesar das RNAs se mostrarem adequadas para estimar dados faltantes de precipitação na bacia hidrográfica do Rio Luvuvhu, África do Sul, com o conjunto de dados utilizados, os autores recomendam o uso de mais estações que mostrem a variação espacial da precipitação e também as estimativas de precipitação por satélite devido à distribuição não uniforme dos pluviômetros.

Na Austrália, Abbot e Marohasy (2014) e Deo e Şahin (2015), destacaram que os resultados produzidos pela aplicação de RNAs para fazer previsão de precipitação mensal foram melhores se comparados com as previsões do modelo *POAMA – Predictive Ocean Atmosphere Model for Australia*, modelo de previsão oficial de chuvas sazonais para a região, baseados nas estatísticas de REQM, erro médio absoluto e coeficientes de correlação.

Mekanik *et al* (2013) quando compararam estatisticamente as metodologias de Regressão Múltipla e RNA para modelos de previsão de precipitação de larga escala, enfatizaram o potencial das Redes Neurais Artificiais quanto ao coeficiente de correlação e aos erros do conjunto do teste.

### 2.4.5 Imputação múltipla

Essa metodologia é uma ferramenta utilizada na área de reanálise meteorológica de dados e foi escolhida pela sua disponibilidade desenvolvida no *R* a partir do pacote *mice*. A reanálise é realizada pelo *National Centers for Environmental Prediction* – NCEP em parceria com o *National Center for Atmospheric Research* – NCAR que disponibilizam séries históricas sintéticas desde 1948 até a atualidade em malhas de até 0,5° x 0,5° de resolução espacial, o equivalente a cerca de 210 km resolução horizontal (KISTLER *et al* 2001; SAHA *et al* 2010). Para que isso seja possível, são necessárias observações assimiladas que necessitam de equipamentos como radiossonda, navio, aviões e satélites para controlar os dados. Esses dados são gerados a partir de campos globais atmosféricos e fluxos superficiais dos sistemas de previsão numérica do próprio NCEP, denominado como interpolação estatística espectral (PARRISH; DERBER, 1992). Com a reanálise é possível verificar anomalias na climatologia em geral e também, através das malhas disponibilizadas para cada região, obter uma previsão de dados faltantes para estações pertencentes a essa malha.

Segundo Lima e Nunes (2012), a qualidade dos dados meteorológicos, principalmente da precipitação, é uma das grandes dificuldades encontradas nos estudos climáticos, e essa dificuldade se acentua ainda mais quando é considerada uma escala temporal e espacial devido às características peculiares de cada cidade. As autoras fizeram uma comparação básica pontual dos dados climatológicos e os dados numéricos do *Climate Prediction Centre* - CPC para preenchimento de falha de dados mensais de precipitação, e afirmam que a precipitação orográfica é um dos fatores com mais implicação no preenchimento de falhas de dados pontuais.

Dias e Albieri (1992), assim como Nunes *et al.* (2010), afirmam que a imputação é uma técnica utilizada para preencher os dados faltantes com valores plausíveis ou por estimativas dos mesmos.

Dias e Albieri (1992) relatam que as primeiras técnicas de imputação envolviam métodos que substituíam dados faltantes pela média ou pela mediana da variável, por interpolação ou até por regressão linear, denominadas de “imputação simples”. E o que torna esta técnica atrativa, além da fácil implementação, é que após a imputação dos

dados com este método, o pesquisador pode utilizar técnicas tradicionais para realizar análises estatísticas de dados completos (NUNES *et al.*, 2010).

A técnica de imputação múltipla foi proposta por Rubin em 1978 para tratar de problemas complexos de conjunto de dados incompletos (VAN BUUREN; OUDSHOORN, 2011) com a tentativa de construir um método capaz de refletir a incerteza sobre as previsões dos dados faltantes, no qual cada valor ausente é substituído por um conjunto de valores aceitáveis que representam a incerteza sobre o valor a ser imputado (da SILVA, 2012).

Essa metodologia está disponível nos principais *softwares* estatísticos gratuitos (CAMARGOS *et al.*, 2011), inclusive no *R* por meio do pacote *mice*, no qual é baseada em imputação múltipla por equações vinculadas, que conseguem detectar quais variáveis do conjunto de dados têm falhas e realiza as imputações dos dados faltantes de acordo com o método *Predictive Mean Matching – PMM*, uma abordagem de imputação semi-paramétrica semelhante à regressão, porém este método garante que os valores imputados são realistas (UCLA, 2016). A série de precipitação original com falhas é colocada lado a lado dos quatro pontos de grade mais próximos da localidade estudada. Essa grade é adquirida por meio do CPC/NOAA – *National Oceanic and Atmospheric Administration* que são baseadas no método de interpolação ótima.

Essa técnica é utilizada para preencher dados faltantes em séries históricas, Costa *et al* (2012) utilizaram o pacote *mice* para imputar dados diários faltantes de precipitação em seis séries no Estado da Paraíba e obtiveram resultados promissores de correlações entre os dados diários observados e imputados, bem como quando acumularam estes dados mensalmente.

## 2.5 CONSISTÊNCIA DOS DADOS

Após a imputação dos dados e análise dos resíduos do método para obter o melhor modelo ajustado, pode haver ainda algumas incoerências nos dados preenchidos, e para que não ocorram erros grosseiros, é verificada a consistência dos dados. Essa etapa é fundamentada na avaliação da consistência dos resultados obtidos com o preenchimento e permite identificar o grau de homogeneidade dos dados de cada estação selecionada em relação aos valores observados nas estações vizinhas (COELHO, p. 79, 2014).

O método escolhido para verificar a consistência dos dados foi o Método da Dupla Massa, desenvolvido pelo *United States Geological Survey – USGS* (1960) para ser utilizado na verificação da consistência de diferentes tipos de dados hidrológicos por meio da comparação entre diferentes estações situadas numa região climatológica semelhante. Este método assume que a relação entre a quantidade de uma variável X e Y pode ser expressa pela equação (7) (SEARCY; HARDISON, 1960):

$$Y = b \times X \quad (7)$$

sendo:

*b*, a declividade da curva da dupla massa.

Entretanto, Searcy e Hardison (p. 34, 1960) afirmam que essa hipótese é substancialmente correta para relações envolvendo dados de precipitação apenas.

## 2.6 MEDIDAS DE ERRO

Na estatística existem algumas medidas de erro que podem ser utilizados para comparar o desempenho dos métodos de preenchimento de falhas. As mais reconhecidas e comumente utilizadas com base nos valores observados e simulados são: Raiz do erro quadrático médio (RMSE), Erro médio absoluto (MAE) e Erro relativo médio (MRE) (TEEGAVARAPU; CHANDRAMOULI, 2005; TEEGAVARAPU; TUFAL; ORMSBEE, 2009). Além destas medidas de erro, a estatística denominada Coeficiente de determinação também faz parte da análise de resultados, visto que é uma das formas de avaliar a qualidade do ajuste do modelo (ABBOT; MAROHASY, 2014; BÁRDOSSY; PEGRAM, 2014). As medidas de erro têm suas unidades expressas nas mesmas medidas dos parâmetros que são utilizados (neste caso milímetros (mm)), com exceção do MRE que é adimensional.

Pela desvantagem do erro quadrático médio em penalizar previsões que contém erros elevados e valorizar os erros de menor magnitude, é comum utilizar a raiz do erro quadrático médio para adequar essa desproporcionalidade (SOUZA *et al.*, 2008).

A raiz do erro quadrático médio, ou *Root Mean Squared Error – RMSE* da sigla em inglês, é utilizada como estatística métrica padrão na climatologia para previsão e análise de regressão ao medir o desempenho do modelo. É uma medida de verificação de

variáveis padronizadas definida como a raiz quadrada da média das diferenças quadradas entre elementos correspondentes das previsões e observações (BARNSTON, 1992; CHAI; DRAXLER, 2014), equação 8:

$$RMSE = \sqrt{\frac{\sum (sim-obs)^2}{n}} \quad (8)$$

sendo:

sim, valores dos dados simulados pelo método de preenchimento de falhas.

obs, valores observados do conjunto de dados.

n, número de simulações realizadas pelo método de preenchimento de falhas.

Ou seja, essa medida é o desvio padrão dos resíduos, e esses resíduos dão uma ideia do grau de dispersão em torno da linha de regressão. Quando utilizada a RMSE, as unidades tornam-se mais similares, pois os erros maiores ficam com o mesmo peso que os erros menores no modelo (OLIVEIRA; HENNING, 2015).

Para Willmott e Matsuura (2005) o erro médio absoluto, do inglês *Mean Absolute Error* – MAE é relativamente simples de ser calculado em relação ao RMSE, pois o MAE envolve o cálculo da soma dos valores absolutos dos erros para obter o erro total e em seguida a divisão pelo número de observações. Oliveira (2015) afirmam que:

“Na fórmula do MAE o erro de previsão é considerado em módulo, sendo assim tratado com uma distância, caracterizando um bom critério para avaliar a precisão dos métodos de previsão com intuito de verificar o quanto o modelo de previsão desenvolvido está desviando da série estudada. Quanto menor o MAE, mais adequado será o modelo.”

conforme equação 9:

$$MAE = \frac{1}{n} \sum_{i=1}^n |sim - obs| \quad (9)$$

Chai e Draxler (2014) afirmam que o MAE é mais adequado para descrever erros uniformemente distribuídos e que as melhores medidas métricas de estatísticas devem fornecer além de uma medida de desempenho, uma representação da distribuição de erros. Porém, em situações em que os erros do modelo provavelmente terão uma distribuição normal em vez de uma distribuição uniforme, o RMSE deve ser a melhor medida métrica para apresentar do que o MAE.

Com o cálculo semelhante ao MAE, o erro relativo médio ou MRE do inglês *Mean Relative Error* também é um critério usado para escolha de melhor ajuste do modelo.

Entretanto, o *MRE* representa a magnitude da diferença entre os valores observados e simulados em relação ao próprio valor observado:

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{sim-obs}{obs} \right| \quad (10)$$

## 2.7 COEFICIENTE DE DETERMINAÇÃO – $R^2$

Esse coeficiente é uma medida largamente utilizada para julgar a adequação de um modelo de regressão, e quando se tratam de duas variáveis aleatórias distribuídas conjuntamente em uma regressão linear simples, o  $R^2$  é matematicamente igual ao quadrado do coeficiente de correlação entre estas duas variáveis (MONTGOMERY; RUNGER, p. 251, 2009).

Esse coeficiente indica quanto da variação total é comum aos elementos que constituem os pares (valores simulados – valores observados), ou seja, é uma medida de ajustamento de um modelo estatístico linear generalizado em relação aos valores observados (CONTI, 2009).

O coeficiente de determinação foi escolhido para ser utilizado neste trabalho como indicador da proporção que os valores de precipitações preditos pelos métodos se aproximam dos valores observados nos postos de medição, porém, é preciso cuidado ao analisar os resultados, já que o  $R^2$  sempre aumentará quando adicionada uma variável ao modelo (MONTGOMERY; RUNGER, 2009). Este coeficiente é adimensional com valor variando entre 0 a 1 e pode ser obtido conforme descrito na equação (11):

$$R^2 = r^2 = \frac{\sum(obs - média_{obs})^2 - \sum(obs - sim)^2}{\sum(obs - média_{obs})^2} \quad (11)$$

sendo:

*sim* são os valores simulados pelos métodos;

*obs* são os valores observados;

*média<sub>obs</sub>* é a média dos valores observados;

*r* é o Coeficiente de Correlação de Pearson.



### 3 METODOLOGIA

#### 3.1 ÁREA DE ESTUDO

O município de Joinville está localizado na latitude 26° 18' 05" S e na longitude 48° 50' 38" W, na região Nordeste de Santa Catarina, possui uma área total de 1.135,05 km<sup>2</sup>, sendo 212,6 km<sup>2</sup> de área urbana e 922,45 km<sup>2</sup> de área rural. (MAIA *et al*, 2014).

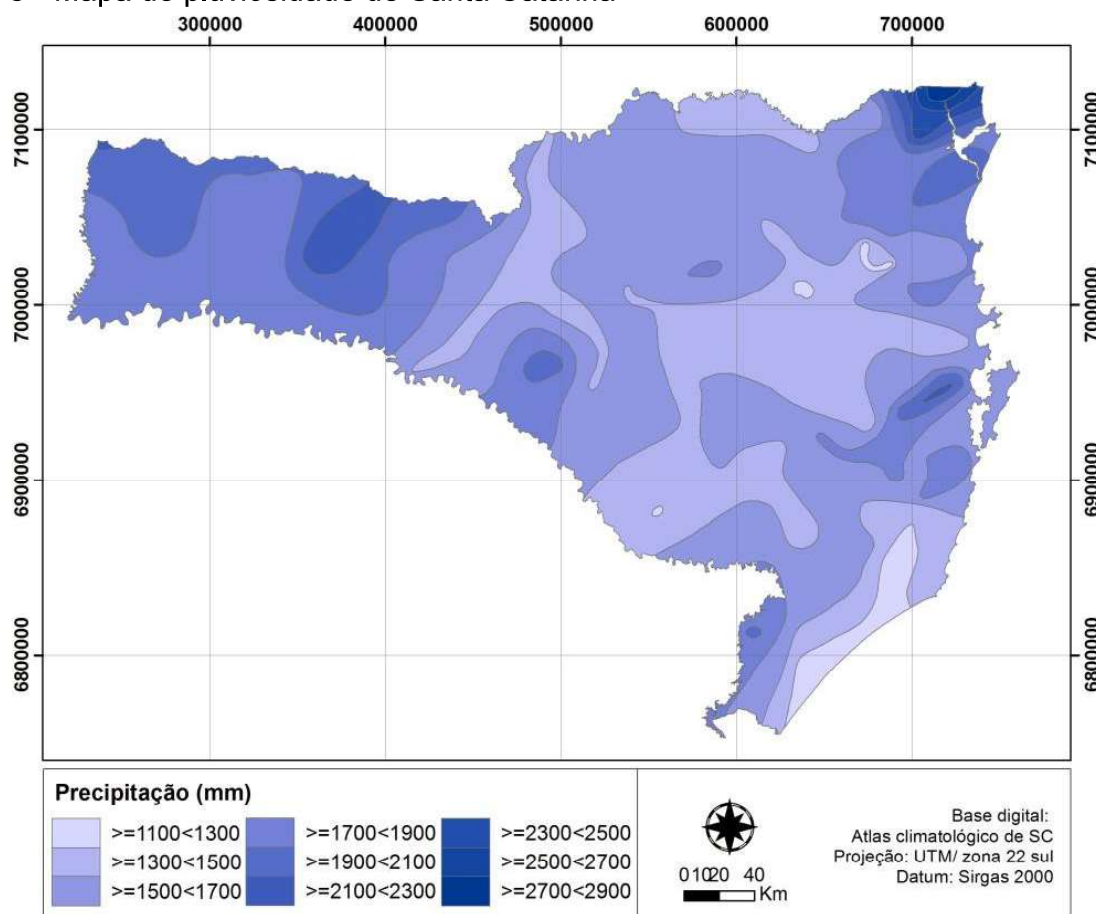
A cidade está situada na Região Hidrográfica do Atlântico Sul, região que predominam os rios de pequeno porte, sendo composta por sete bacias hidrográficas conforme Maia *et al* (2014):

- a) Bacia Hidrográfica do Rio Palmital;
- b) Bacia Hidrográfica do Rio Cubatão (Norte);
- c) Bacia Hidrográfica do Rio Piraí;
- d) Bacia Hidrográfica do Rio Itapocuzinho;
- e) Bacia Hidrográfica do Rio Cachoeira;
- f) Bacias Hidrográficas Independentes da Vertente Leste;
- g) Bacias Hidrográficas Independentes da Vertente Sul.

#### 3.2 REDE DE MONITORAMENTO

Santa Catarina é considerado um dos Estados brasileiros com melhor distribuição pluviométrica de precipitação durante o ano, Figura 3 (MONTEIRO, 2001). Segundo Mello *et al*. (2012), a região nordeste de Santa Catarina, na qual se insere o município de Joinville, apresenta um regime pluviométrico anual que varia entre 1.500 mm e 2.000 mm, sendo que nos meses de primavera e verão são registrados os maiores índices. Entretanto, existem valores pontuais de precipitação que excedem os 2.900 mm registrados no Atlas Climatológico de Santa Catarina (MELLO, 2015), e como pode ser observado, a região de Joinville é uma dessas áreas.

Figura 3 - Mapa de pluviosidade de Santa Catarina



Fonte: Mello (2015).

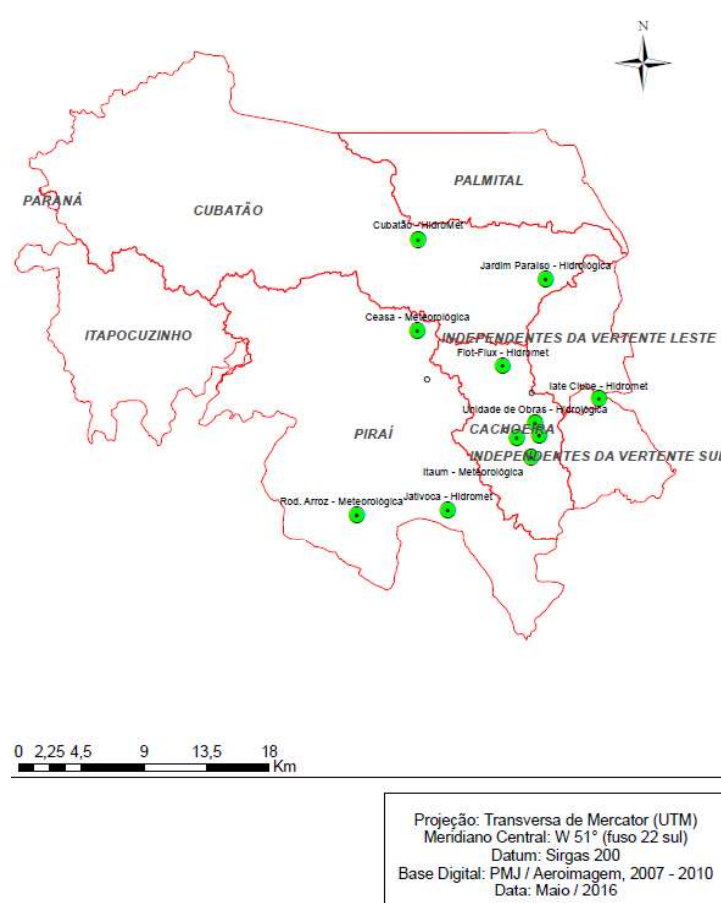
Para a execução do trabalho, foi utilizada a rede de monitoramento da Defesa Civil de Joinville, que foi instalada seguindo as recomendações dos padrões mínimos exigidos pela Organização Mundial de Meteorologia – OMM.

A rede tem suas 11 estações classificadas como automáticas, sendo 3 meteorológicas, 3 hidrológicas e 5 hidrometeorológicas, que por sua vez, estão distribuídas em quatro bacias hidrográficas: Rio Cubatão do Norte, Rio Pirai, Rio Cachoeira e Vertente Sul, conforme Figura 4.

As estações meteorológicas estão distribuídas da seguinte maneira, 1 na bacia hidrográfica do Rio Cachoeira e 2 no Rio Pirai, todas as 3 dotadas de sensores eletrônicos que medem: velocidade e direção dos ventos, umidade relativa e temperatura, intensidade e quantidade de chuva, e radiação solar. Nas estações hidrológicas são aferidas as variáveis precipitação e nível do rio, sendo duas estações na bacia do Rio

Cachoeira e uma no Rio Cubatão do Norte. As estações hidrometeorológicas são capazes de mensurar as variáveis de ambas as estações, tanto das meteorológicas quanto das hidrológicas. Estas estão situadas, uma em cada, nas bacias do Rio Cubatão do Norte, Rio Pirai, Vertente Sul e 2 no Rio Cachoeira. Para a pesquisa apenas a estação meteorológica Ceasa, localizada na Bacia do Rio Pirai, foi excluída, pelo fato de conter inúmeras leituras questionáveis, que poderiam impactar nas análises.

Figura 4 - Delimitação das Bacias Hidrográficas de Joinville e Localização das Estações da Rede de Monitoramento de Joinville – SC



Fonte: Defesa Civil (2016).

As coordenadas de cada estação podem ser observadas na Tabela 1.

Tabela 1 - Localização das estações pluviométricas da rede de monitoramento da Defesa Civil do município de Joinville - SC

Estações pluviométricas	Bacia	Latitude	Longitude	Altitude (m)
Cubatão	Rio Cubatão do Norte	26°11'40.00"S	48°54'39.83"O	24
Cachoeira Área Central (FlotFlux)	Rio Cachoeira	26°16'30.34"S	48°50'57.31"O	6
Paraíso	Rio Cubatão do Norte	26°13'11.31"S	48°49'9.21"O	6
Ceasa	Rio Pirai	26°15'12.44"S	48°54'54.74"O	39
Cia. Águas (Bucarein)	Rio Cachoeira	26°19'18.38"S	48°50'16.90"O	4
Itaum	Rio Cachoeira	26°20'2.12"S	48°49'38.68"O	53
Guanabara	Rio Cachoeira	26°19'10.89"S	48°49'19.86"O	1
Joinville Iate Clube (JIC)	Vertente Sul	26°17'34.18"S	48°46'48.70"O	2
Unidade de Obras	Rio Cachoeira	26°18'42.24"S	48°49'30.53"O	1
Jativoca	Rio Pirai	26°22'8.88"S	48°53'12.40"O	7
Rodovia do Arroz	Rio Pirai	26°22'23.17"S	48°57'7.84"O	11

Fonte: Defesa Civil (2016).

A partir da Tabela 1, pode-se verificar que as estações utilizadas nesta pesquisa pertencem à região climatológica semelhante, todas estão localizadas na planície.

O relatório de diretrizes da ANA (2012) indica que as estações a serem analisadas sejam separadas em estações principais e estações secundárias. As estações principais são consideradas a partir de algumas características como: localização na bacia, extensão do período de observação, poucos períodos contendo interrupção dos dados, ou seja, as estações que possuem menor número de falhas nos dados e possam ser tomadas como base. As estações secundárias verificam a existência de erros de observação ou transcrição nos dados, servindo como apoio às estações principais. Essas últimas são selecionadas conforme alguns critérios: maior coeficiente de correlação, menor distância entre as estações (máximo 200 km), períodos de dados coincidentes e altitudes semelhantes.

As distâncias entre as estações escolhidas para este trabalho estão apresentadas na Tabela 2.

Tabela 2 – Distância, em metros, entre as estações da rede de monitoramento da Defesa Civil

Estações	1	2	3	4	5	6	7	8	9	10
JIC (1)		4965	7160	9046	17012	19346	13610	6567	6611	5166
Unidade de Obras (2)	4965		4680	10270	15550	14390	8860	2480	1693	950
FlotFlux (3)	7160	4680		6910	10880	14931	11053	6837	2555	5608
Paraíso (4)	9046	10270	6910		9567	21609	17942	12740	11520	11155
Cubatão (5)	17012	15550	10880	9567		20195	19495	17553	15867	16470
Rodovia do Arroz (6)	19346	14390	14931	21609	20195		6538	13188	12740	14250
Jativoca (7)	13610	8860	11053	17942	19495	6538		7100	7170	8454
Itaum (8)	6567	2480	6837	12740	17553	13188	7100		1712	1646
Bucarein (9)	6611	1693	2555	11520	15867	12740	7170	1712		1582
Guanabara (10)	5166	950	5608	11155	16470	14250	8454	1646	1582	

Fonte: Defesa Civil (2016).

### 3.3 SOFTWARES UTILIZADOS

Para o processo de aplicação dos diferentes métodos de preenchimento de falhas de dados foi utilizado o *software R* versão 3.4.0 (21/04/2017). Na interface *RStudio* os pacotes *forecast*, *hyfo*, *mice*, *neuralnet* (HYNDMAN, 2017; YUANCHAO, 2017; BUUREN; OUDSHOORN, 2011; FRITSCH; GUENTHER, 2016) foram aplicados para o desenvolvimento dos métodos de interpolação linear, regressão linear múltipla, método das correlações, imputação múltipla e redes neurais artificiais respectivamente. As falhas foram contabilizadas primeiramente por meio do pacote *RMAWGEN* (CORDANO ;ECCEL, 2012) e depois por intermédio do pacote *mice* os valores discrepantes foram identificados pelo pacote *forecast*. Para os gráficos de resultados optou-se por trabalhar com o pacote *ggplot2* (WICKHAM, 2009), já os histogramas foram reproduzidos a partir da própria biblioteca de funções *default* do programa.

As planilhas de dados, resultados e análise foram todas organizadas utilizando o *Microsoft Excel* 2013 ®.

### 3.4 CRITÉRIOS UTILIZADOS

Os critérios estabelecidos para selecionar as estações a serem utilizadas no preenchimento de estações com falhas, em cada método, visaram agrupar as estações vizinhas mais próximas e sem falhas no mesmo período. Entretanto, se essas estações vizinhas não atingissem níveis de significância de até 0,05 no teste de hipóteses, valor este tradicional de corte para rejeitar a hipótese nula (FERREIRA; PATINO, 2015), da regressão linear múltipla, as mesmas eram retiradas da equação do modelo de regressão. Em casos nos quais essa retirada de estações não atingisse o número mínimo desejável pelo modelo, que devem ser 3 estações, foi utilizado um segundo critério.

O segundo critério buscou priorizar as estações com menor número de falhas ao longo do ano hidrológico em questão. Caso nenhum destes dois critérios pudessem ser atendidos, foram colocadas todas as estações restantes na equação e estas foram excluídas uma a uma do modelo até que duas ou mais estações atingissem nível de significância de 0,10 como variáveis independentes, visto que o nível de significância de 5% era algo difícil de alcançar. Essas condições foram respeitadas no preenchimento dos dados faltantes originais, bem como, no preenchimento de falhas simuladas em todos os anos da série hidrológica.

### 3.5 SÉRIE DE DADOS

A rede de monitoramento escolhida para este trabalho possui série reduzida, iniciando sua coleta de dados a partir de 2012 em intervalos de 5 minutos, o que permitiu trabalhar com dados de 24 h de precipitação, medidos das 00:00 horas do dia atual até as 00:00 horas do dia seguinte.

Seguindo os objetivos específicos que foram citados, o primeiro passo foi a coleta de dados realizada por meio eletrônico no site da prefeitura da cidade: <https://intranet.joinville.sc.gov.br/defesanet/estacao/lista>. Os dados são de acesso público e para acessá-los é necessário realizar um cadastro e logo após é possível fazer o *download* dos dados que estão disponíveis para cada estação e exportá-los para o *software Excel*. Concluído o *download* dos dados diários para as 10 estações em estudo, na etapa seguinte colocou-se em prática a preparação dos dados. Iniciando a análise

pelas datas dos arquivos, foi possível identificar, para cada estação, em quais dias se necessitava de preenchimento. Em determinados períodos algumas datas não constavam no arquivo, essa ausência de datas foi considerada como falha e assumiu-se que nessas datas também era necessário o preenchimento.

A série foi dividida em períodos de ano hidrológico, que segundo Bolzan (2014) baseado nos registros do Comitê Cubatão Cachoeira Joinville – CCJ, concluiu que o ano hidrológico para a cidade deveria começar na primavera. Sendo assim, o primeiro ano hidrológico teve início no mês de Outubro de 2012 seguindo-se até o mês de Setembro do ano seguinte e o quarto ano hidrológico terminou em Setembro de 2016. Para cada ano hidrológico foi gerado um arquivo no *Excel* com as correspondentes datas.

A identificação de falhas aconteceu para todas as 10 estações pluviométricas, que estão assim denominadas: Bucarein, Cubatão, FlotFlux Guanabara, Itaum, Jativoca, JIC, Paraíso, Rodovia do Arroz e Unidade de Obras.

Antes de passar para o próximo estágio, foi necessário realizar a análise exploratória dos dados, que permitiu apontar os possíveis valores discrepantes e identificar a forma de distribuição da variável. Sendo assim, novamente com o auxílio do *R*, utilizando o pacote *forecast* e o comando *tsoutliers*, foram observadas as posições em que se encontravam os valores considerados discrepantes pelo programa. Esses valores foram analisados em conjunto com os valores das outras estações para a mesma data no *Excel* e logo após a retirada dos valores discrepantes, foi realizada a construção do histograma de frequência para verificar a distribuição da variável hidrológica.

### 3.6 APLICAÇÃO DOS MÉTODOS DE PREENCHIMENTO DE FALHAS

Posterior ao processo de tratamento dos dados foram aplicados os métodos de preenchimento de falhas. Considerando que a presente pesquisa foi desenvolvida com a aplicação de diferentes metodologias para o preenchimento de falha de dados de 24 horas, e que o objetivo é identificar quais destes métodos preenche melhor as lacunas das estações que contemplam a rede de monitoramento da Defesa Civil de Joinville, optou-se por trabalhar com: Interpolação linear, Regressão linear múltipla, método das correlações, Redes neurais artificiais e o método de Imputação múltipla por equações

vinculadas. Estes métodos foram escolhidos por serem bastante utilizados por pesquisadores, além da disponibilidade dentro do *software R*.

Cada método foi aplicado duas vezes. A primeira delas para preencher os dados faltantes originais, que são os dados com falhas geradas pelo mau funcionamento dos equipamentos da rede de monitoramento da Defesa Civil Municipal ou interrupção na transmissão de dados, ou seja, as falhas encontradas ao coletar os dados. Na segunda vez, o método deveria preencher os períodos de registros reais de precipitação que foram retirados das planilhas em cada estação. Nesse segundo preenchimento, cada resposta obtidas com os métodos de preenchimento foi denominada de dado simulado, ou simulação.

As medidas de erro citadas no item 2.6 foram aplicadas para apurar o desempenho dos métodos comparando os valores obtidos pelo método de preenchimento (dados simulados) com os valores observados. Para realizar o preenchimento das falhas simuladas foi considerado o percentual de dados faltantes contido em cada estação para cada ano hidrológico. No entanto, para estações que permitiam o número de dados simulados iguais ao número de falhas originais, assim foi feito.

No primeiro ano hidrológico, por exemplo, as duas estações com falhas, Paraíso e Unidade de Obras, tiveram a quantidade de dados simulados igual ao número de falhas originais, totalizando 8 para a estação Paraíso e 73 para a estação Unidade de Obras, vide Tabela 3.

Tabela 3 – Falhas e simulações do primeiro ano hidrológico

Estação	nº de falhas originais	nº de simulações	nº de observações	% de falhas
Paraíso	8	8	357	2,192
Unidade de Obras	73	73	292	20

Fonte: A autora.

O segundo ano teve as estações Jativoca, Guanabara e Rodovia do Arroz com a mesma situação do primeiro ano hidrológico, e para as estações Cubatão e Unidade de Obras houve uma semelhança, com o número de simulações pouco inferior ao número



de falhas originais. A estação Bucarein teve seu número de simulações bem inferior devido à porcentagem de falhas, conforme mostra a Tabela 4.

Tabela 4 - Falhas e simulações do segundo ano hidrológico

Estação	nº de falhas originais	nº de simulações	nº de observações	% de falhas
Bucarein	270	35	95	73,973
Jativoca	111	111	254	30,411
Cubatão	14	11	351	3,835
Guanabara	58	58	307	15,890
Rodovia do Arroz	99	99	266	27,123
Unidade de Obras	40	41	324	10,959

Fonte: A autora.

Para o terceiro ano hidrológico o procedimento foi bastante parecido com segundo ano, tendo algumas estações com número de falhas e simulações idênticos, bem como algumas estações contendo uma pequena diferença entre esses números. Com exceção das estações Bucarein e Itaum que tiveram o número de simulações bastante inferior ao número de falhas e a estação Paraíso que tinha apenas uma única falha, contabilizando 100 simulações, ver Tabela 5.

Tabela 5 - Falhas e simulações do terceiro ano hidrológico

Estação	nº de falhas originais	nº de simulações	nº de observações	% de falhas
Bucarein	297	21	68	81,370
JIC	76	76	289	20,822
Cubatão	23	23	342	6,301
Guanabara	75	73	290	20,548
FlotFlux	127	119	238	34,795
Unidade de Obras	12	7	353	3,288
Itaum	233	53	132	63,836
Paraíso	1	100	364	0,274

Fonte: A autora.

No quarto ano hidrológico a maioria das estações não manteve o número de simulações igual ou próximo ao número de falhas, com exceção das estações Bucarein, FlotFlux e Itaum que respeitaram exatamente o mesmo valor para ambos os índices, apresentados na Tabela 6.

Tabela 6 - Falhas e simulações do quarto ano hidrológico

Estação	nº de falhas originais	nº de simulações	nº de observações	% de falhas
Bucarein	58	58	308	15,890
Cubatão	100	65	265	27,397
Guanabara	102	52	263	27,945
FlotFlux	3	3	362	0,822
Unidade de Obras	264	37	102	72,329
Itaum	13	13	352	3,562
Paraíso	1	47	364	0,274

Fonte: A autora.

### 3.7 MÉTODOS ESTATÍSTICOS UTILIZADOS

#### 3.7.1 Interpolação linear

A interpolação linear é um método simples se comparado aos outros escolhidos, porém, sua presença na pesquisa é justificada pela posterior comparação com os métodos mais sofisticados, como por exemplo, o método de imputação múltipla (pacote *mice*). Essa função, de interpolação linear, está disponível no pacote *forecast* com o comando *na.interp*, este comando é aplicado para cada estação individualmente, ou seja, para cada estação existe uma função. Uma desvantagem desse método se deve ao fato de não levar em conta os valores de precipitação das outras estações para o mesmo dia do dado faltante, fazendo com que o valor interpolado seja baseado apenas em valores de precipitação já ocorridos naquela estação. Isso ocorre porque o pacote *forecast* foi criado com o intuito de atender séries temporais univariadas, para fazer previsões e não preenchimento de falhas, logo, considera o conjunto de dados de apenas uma estação.

O processo é o mesmo para o segundo ano hidrológico, entretanto, o que difere de um ano para o outro são as estações a serem preenchidas.

### 3.7.2 Regressão Linear Múltipla

A regressão linear múltipla foi escolhida por ser um método de fácil aplicação e também pelo bom desempenho em outros trabalhos com o mesmo propósito de preenchimento de falhas pluviométricas. Por exemplo, Mello (2015) confirmou que o método da regressão linear múltipla foi o que melhor se adequou para a região do município de Joinville, utilizando dados mensais.

Antes da aplicação do método, foi necessário selecionar quais estações poderiam ser associadas, visando a melhor correlação entre elas, baseado nos critérios descritos no item 3.4. Para este procedimento de escolha das estações cada estação a ser preenchida se tornou a variável dependente das outras estações que foram agrupadas, trabalhando sempre com o mínimo de três estações, uma dependente e duas independentes.

Na regressão linear múltipla foram obtidas funções diferentes a cada estação a ser preenchida. Por exemplo, no primeiro ano hidrológico, as estações com falhas e o número delas são: Unidade de Obras (73) , Paraíso (8), Cubatão (1) e Guanabara (1), o que significa dizer que para preencher cada uma dessas 4 estações, poderiam ser utilizadas apenas as estações que não continham falhas no mesmo período. Essa condição foi respeitada em todos os anos hidrológicos da série.

No preenchimento das falhas originais o método também foi colocado em prática a cada ano hidrológico, porém, para cada um deles existiam diferentes estações contendo lacunas, o que implicou em diferentes equações de modelo para cada estação e também para cada período. Ao preencher as falhas simuladas algumas estações mantiveram as mesmas variáveis (estações) melhores correlacionadas, igualmente obtidas no primeiro preenchimento. No entanto, a maior dificuldade encontrada tanto para as falhas originais, quanto para as falhas simuladas, foi no terceiro e quarto ano hidrológico, pois em ambos os períodos, todas as estações continham falhas.

### 3.7.3 Método das correlações

Outra metodologia utilizada para este trabalho é o método das correlações que está contido no pacote *hyfo*. Este método é baseado na correlação e regressão linear

para preencher falhas em séries temporais de precipitação (HIRSCH; MAIDMENT, 1992). Como os dados para este trabalho são de 24 horas, este foi o período utilizado para obter o coeficiente de correlação a cada par de estações.

Para efetuar o método das correlações (pacote *hyfo*) exportaram-se os arquivos separados por ano hidrológico para o *R* e executou-se o comando necessário para armazenar os resultados em uma planilha compatível com o *Excel*. Após repetir esse procedimento para cada ano hidrológico, foi possível verificar, no ambiente integrado do *RStudio*, a matriz de correlação e as ordens de correlação para cada estação, assim como os coeficientes lineares para cada par de estações, que foram obtidos a partir da regressão linear simples. Os dados faltantes das estações nos determinados dias ou períodos foram preenchidos baseados em todos os registros de precipitação referente ao ano hidrológico em aplicação, simultaneamente para todas as estações. Este mecanismo foi utilizado tanto no preenchimento de falhas originais quanto no preenchimento das falhas simuladas.

### 3.7.4 Redes neurais artificiais

Para esta pesquisa a aplicação de RNA foi escolhida por se tratar de um método mais sofisticado e também por ser aplicado nesta área (TEEGAVARAPU; CHANDRAMOULI, 2005; COULIBALY; EVORA, 2007), enfatizando o poder do método de trabalhar com dados de 24 horas de precipitação.

Nesta pesquisa optou-se por trabalhar com a MLP visto que no *R* o algoritmo de aprendizado da rede utilizado é o *Backpropagation* ou retropropagação de erro (RPROP), disposto no pacote *neuralnet*.

Ao aplicar este método foram necessárias três tentativas para convergir a rede. A primeira delas foi aplicada ano a ano, igual ao procedimento utilizado nos outros métodos desta pesquisa. Na segunda tentativa optou-se por agrupar os dados utilizando o critério de sazonalidade iniciando na Primavera de 2012. Porém, ao utilizar estas alternativas a rede além de não convergir, continuava apontando erros grosseiros. Então, foi realizada a terceira tentativa separando os dados mês a mês e, assim, as redes convergiram mais rapidamente e alcançaram erros aceitáveis, tal como Depiné *et al* (2014) que na tentativa

de realizar o preenchimento de dados de precipitação, utilizaram séries curtas com aproximadamente 12 a 30 dias.

O procedimento consiste em primeiramente exportar o arquivo de cada mês para o *R*, e depois, utilizando o comando *neuralnet* fez-se com que a fórmula da regressão linear tivesse as variáveis independentes e dependentes corretamente escolhidas. As variáveis de entrada independentes neste estudo são as estações usadas para preencher a estação com falha, geralmente as que possuem maior correlação com a estação contendo dados faltantes, seguindo os critérios estabelecidos no item 3.4. E a resposta são os valores preditos para os dados faltantes da estação.

Algumas das características inseridas pelo usuário são: a magnitude da precisão (*threshold*) desejada para obter os valores preditos pelo modelo; o número máximo de passos (*stepmax*) que a rede deve dar até convergir; o tipo da função de ativação (*act.fct*) dos neurônios da rede; tipo da função erro (*err.fct*) que calcula o erro total da rede; tipo de saída das respostas (*linear.output*) se saída linear ou não; e o número de neurônios (*hidden*) na camada oculta onde ocorrem as iterações. Neste trabalho os respectivos parâmetros utilizados foram:

- a) Threshold = 0,01
- b) Stepmax =  $1 \times 10^7$
- c) Act.fct = *logistic*
- d) Err.fct = *sse*
- e) Linear.output = *true*
- f) Hidden = 15

Foram feitas tentativas com 5, 10 e 15 neurônios, porém, esta última foi escolhida para ser utilizada por atingir a convergência da rede mais rapidamente.

Neste método deve-se atentar tanto para as correlações entre as estações, pois a função da RNA é baseada na regressão linear, quanto para a ocorrência de falha nas estações, já que o método não permite fazer o preenchimento de uma estação com falhas utilizando estações que também contém dados faltantes.

O procedimento é o mesmo para todos os meses que contemplam os 4 anos hidrológicos da série escolhida para o estudo. O que difere de um mês para o outro são

as estações a serem preenchidas e as estações que devem ser utilizadas como variáveis independentes, já que a cada mês, o período contendo falhas é diferente. O que significa dizer, que para cada mês foi obtida uma rede neural artificial diferente, com pesos atribuídos diferentes, porém, todas com o mesmo número de neurônios.

Para cada neurônio é atribuído o parâmetro peso e calculada a média e o desvio padrão. Ao final da RNA é mostrada a matriz resposta, que contém os valores da soma quadrática dos erros, alcance limite para as derivadas da função erro, número de passos e os pesos associados a cada estação que está sendo usada. Com essas estatísticas foi possível calcular o coeficiente de determinação ( $R^2$ ), Raiz do erro quadrático médio (RMSE), Erro relativo médio (MRE), e Erro médio absoluto (MAE), para avaliar o desempenho da rede neural artificial empregada.

Tanto no preenchimento das falhas originais quanto no preenchimento das falhas simuladas, ocorreram situações em que a estação a ser preenchida não tinha registros de precipitação e conseqüentemente não poderia ser preenchida pelo método. Nestes casos foram atribuídas precipitação igual a zero para um único dia desta estação, que deveria corresponder à precipitação igual a zero nas demais estações também. Isto foi respeitado para todas as estações em todos os anos hidrológicos, com exceção das estações Rodovia do Arroz e Jativoca, que não puderam ser preenchidas em função do número de falhas encontradas.

### **3.7.5 Método de Imputação múltipla por equações vinculadas**

Este método foi escolhido inicialmente por ser útil em estudos de prática de reanálise meteorológica (COSTA *et al.*, 2012), além da sua aplicabilidade quanto ao intervalo dos dados utilizados.

A aplicação do método iniciou com o armazenamento em um vetor do comando *mice* com o conjunto de dados, usando o *default* do programa que são 5 imputações múltiplas, 5 iterações e método de imputação *pmm*. Neste método de imputação *pmm* são calculados os valores preditos para os valores observados e faltantes, e para cada valor faltante que foi predito, é procurada uma unidade observada com valor predito mais próximo, utilizando o valor observado como o valor a ser imputado (NUNES *et al.*, 2009).

Na primeira etapa é realizada uma imputação simples para cada valor faltante do conjunto de dados. Posteriormente, esses valores imputados são ajustados novamente como faltantes e são realizadas novas imputações, o processo se repete por quantas vezes o usuário indicar, conforme explicação anterior. Terminado o procedimento de imputação em cinco iterações, o método preencheu os valores faltantes no conjunto de dados por previsões feitas a partir do modelo e exportou para uma planilha em Excel.

### 3.8 CONSISTÊNCIA E HOMOGENEIDADE DOS DADOS

Para verificação da consistência e homogeneidade dos dados se faz uso do Método da Dupla Massa, cujo procedimento consiste primeiramente em tabelar os valores de precipitação anual ou mensal da série e plotar um gráfico cartesiano, no qual o eixo das ordenadas é composto pelos valores acumulados de precipitação da estação que se deseja verificar a consistência dos dados e o eixo das abcissas representa a média acumulada dos valores de precipitação do conjunto de estações vizinhas (USGS, 1960; TUCCI, 2007; COELHO, 2014). A consistência é obtida quando houver uma tendência linear entre os dados, verificada pelo ajuste da equação da reta e seu coeficiente de determinação ( $R^2$ ) (MELLO, 2015).

Neste trabalho o conjunto das estações vizinhas era formado por três estações, as quais deveriam obedecer a distância máxima de 10000 m, porém, quando não era possível obedecer tal critério, eram utilizadas as demais estações. Os dados das estações foram acumulados mensalmente no período entre Outubro de 2012 até Setembro de 2016, com exceção das estações Jativoca e Rodovia do Arroz que tiveram apenas dados de Outubro de 2012 até Setembro de 2014.

Para verificar a consistência da estação desejada, criou-se primeiramente uma planilha no Excel com as falhas originais da estação em análise e as falhas originais das três estações vizinhas à esta. Posteriormente, uma segunda planilha foi construída com os dados destas mesmas estações preenchidas com o método de interpolação linear, este procedimento se repetiu para o método da RLM, MDC, RNA bem como IM. Após este procedimento, foram elaborados gráficos de cada método de preenchimento para analisar a consistência dos dados simulados.

As estações Jativoca e Rodovia do Arroz, tiveram apenas os dados do primeiro e segundo ano analisados, isto porque não puderam ser preenchidas com nenhum método para o terceiro e quarto ano hidrológico.



## 4 RESULTADOS E DISCUSSÕES

### 4.1 ANÁLISE EXPLORATÓRIA DOS DADOS

A série hidrológica em análise no presente estudo apontou várias falhas distribuídas ao longo das dez estações pertencentes ao conjunto de dados. Além destas falhas, alguns valores discrepantes de precipitação de 24 horas também foram identificados. Esses valores, considerados discrepantes, foram retirados por serem classificados como erros grosseiros e posteriormente declarados como falhas. Por exemplo, a Estação Unidade de Obras no dia 08/12/2013 registrou uma precipitação de 698,2 mm. Esta mesma estação obteve o maior número de dados considerados como discrepantes, 24 no total, devido à diferença dos registros de precipitação em relação às outras estações (ver Tabela 7).

Tabela 7- Valores discrepantes identificados na estação Unidade de Obras

Data	Unidade de Obras (mm)	Média das estações (mm)
16/11/2012	235,7	13,641
07/02/2013	466,6	7,855
25/05/2013	242,3	11,456
22/06/2013	279,6	51,473
22/07/2013	274,7	32,053
27/08/2013	267,5	41,054
23/09/2013	287,9	59,574
26/11/2013	237,8	6,602
08/12/2013	698,2	14,189
16/12/2013	228,6	0
05/01/2014	246,9	18,836
06/03/2014	235,2	6,002
13/03/2014	253,8	25,750
14/03/2014	250,1	20,567
20/04/2014	242,9	16,251
11/05/2014	236,6	8,144
29/06/2014	244,2	10,662
18/07/2014	228,8	0,686
28/09/2014	238,9	9,413
06/02/2015	67,89	12,970
14/02/2015	243,5	10,740
15/02/2015	281,2	30,400
07/11/2015	465,7	5,144
09/11/2015	251,9	19,660

Fonte: A autora.

Na estação Paraíso, foram identificados 2 valores discrepantes (Tabela 8), as estações JIC e Rodovia do Arroz apenas um cada, conforme consta na Tabela 9 e Tabela 10 respectivamente.

Tabela 8 - Valores discrepantes identificados na estação Paraíso

Data	Estação Paraíso (mm)	Média das estações (mm)
09/07/2015	274,8	13,717
22/10/2015	194	12,308

Fonte: A autora.

Tabela 9 - Valor discrepante identificado na estação JIC

Data	Estação JIC (mm)	Média das estações (mm)
04/02/2015	69,16	10,370

Fonte: A autora.

Tabela 10 - Valor discrepante identificado na estação Rodovia do Arroz

Data	Estação Rodovia do Arroz (mm)	Média das estações (mm)
14/02/2014	107,1	21,874

Fonte: A autora.

Cada estação do conjunto de dados utilizados nessa pesquisa deveria conter 1461 registros de precipitação, totalizando 14610 dados considerando todas as dez estações utilizadas. Entretanto, foram identificadas 3495 falhas distribuídas ao longo destas dez estações e 28 valores discrepantes durante os quatro anos hidrológicos da série histórica. Logo, 3523 dados foram computados como falhas originais, sendo a estação Paraíso a que menos apresentou falhas, apenas 11, e a estação Jativoca a que mais apresentou falhas 842 no total.

Após a retirada dos valores discrepantes foi possível apontar a quantidade de dados faltantes para cada estação em cada ano hidrológico, e também, identificar as

estações com menor e maior quantidade de falhas. A Tabela 11 apresenta, por ano hidrológico, os números de falhas para cada estação e a porcentagem de dados faltantes.

Tabela 11 - Número de falhas para cada estação por ano hidrológico (Continua)

Ano hidrológico	Estação	nº de falhas originais	nº de observações	% de falhas
1º	Bucarein	0	365	0
	Jativoca	0	365	0
	JIC	0	365	0
	Cubatão	1	364	0,274
	Guanabara	1	364	0,274
	Rodoviado Arroz	0	365	0
	FlotFlux	0	365	0
	Unidade de Obras	73	292	20
	Itaum	0	365	0
	Paraíso	8	357	2,192
2º	Bucarein	270	95	73,973
	Jativoca	111	254	30,411
	JIC	0	365	0,000
	Cubatão	14	351	3,836
	Guanabara	58	307	15,890
	Rodovia do Arroz	99	266	27,123
	FlotFlux	0	365	0,000
	Unidade de Obras	41	324	11,233
	Itaum	0	365	0,000
	Paraíso	0	365	0,000
3º	Bucarein	297	68	81,370
	Jativoca	365	0	100,000
	JIC	76	289	20,822
	Cubatão	23	342	6,301
	Guanabara	75	290	20,548
	Rodovia do Arroz	365	0	100,000
	FlotFlux	127	238	34,795
	Unidade de Obras	12	353	3,288
	Itaum	233	132	63,836

	Paraíso	1	364	0,274
4º	Bucarein	58	308	15,847
	Jativoca	366	0	100,000
	JIC	1	365	0,273
	Cubatão	101	265	27,596
	Guanabara	103	263	28,142
	Rodovia do Arroz	366	0	100,000
	FlotFlux	4	362	1,093
	Unidade de Obras	264	102	72,131
	Itaum	14	352	3,825
	Paraíso	2	364	0,546

Fonte: A autora.

As estações Jativoca e Rodovia do Arroz não puderam ser preenchidas no terceiro e quarto ano hidrológico com nenhum método, tanto para o preenchimento de falhas originais, quanto no preenchimento das falhas simuladas. Ambas as estações registraram falhas durante todo este período, contabilizando 365 falhas no terceiro ano hidrológico e 366 falhas no quarto ano para cada estação.

Nas estações utilizadas por Coulibaly e Evora (2007), ao aplicar seis diferentes métodos de redes neurais artificiais para preencher dados diários de precipitação e temperatura, foram identificados percentuais de dados faltantes desde 2,1% até 65,2% em séries hidrológicas que variavam de 1 até 23 anos. Ao contrário disso, os autores Mwale *et al* (2012) quando preencheram dados faltantes de precipitação na bacia do Rio Shire em Malawi utilizaram séries hidrológicas que variavam de no mínimo 29 anos e no máximo 71 anos, nas quais foram encontradas porcentagem de falhas que atingiam no máximo 17,1% dos dados.

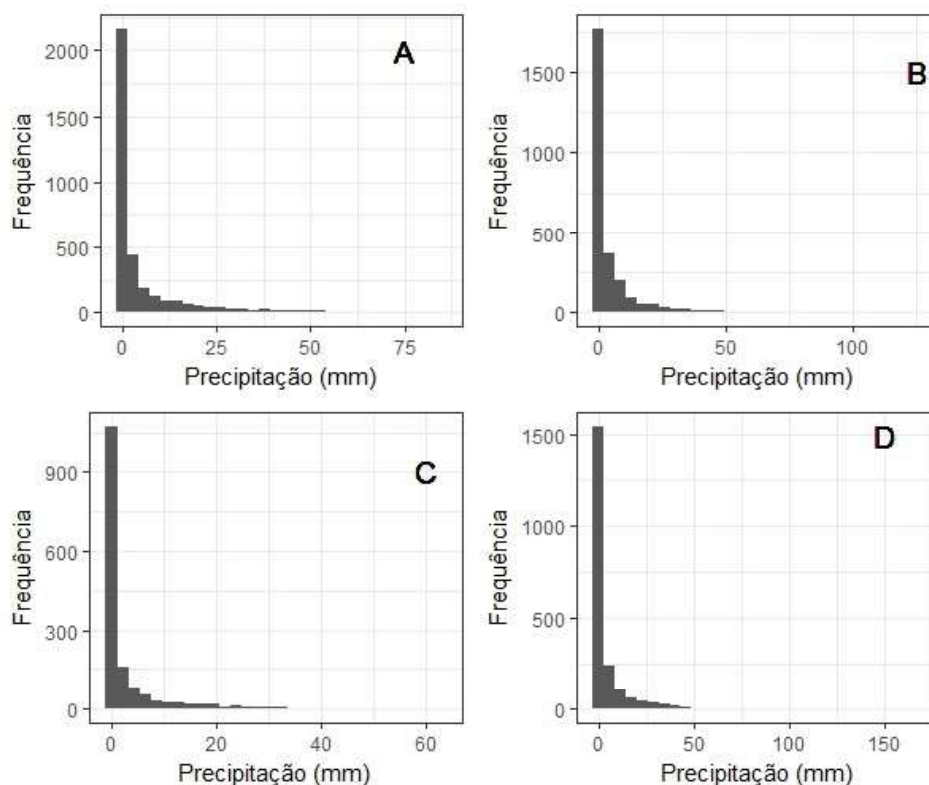
Em função dos números mostrados na Tabela 11, fica claro que no início da série, no primeiro ano hidrológico, as estações contêm poucas falhas, são seis estações com 100% dos dados observados. Entretanto à medida que o tempo vai passando, esse número de falhas tende a aumentar, ou então, diminui o número de estações sem falhas, passando de seis para quatro estações no segundo ano hidrológico. Ao chegar no terceiro e quarto ano hidrológico nenhuma estação possui 100% dos dados observados. Esse aumento de falhas pode acontecer devido a interrupções na transmissão dos dados,

problemas no equipamento ou ainda, falta de manutenção, já que o início da série coincide com o ano de instalação dessas estações e todas são automatizadas com leitura de dados transmitida via rádio.

Após a exclusão dos valores discrepantes e construção dos histogramas de frequência, considerando os dias sem chuva, Figura 5, foi verificado que a variável precipitação, para todos os anos hidrológicos considerados, possui distribuição assimétrica à direita ou positiva, o que indica que os valores mais baixos são mais frequentes, ou seja, o menor valor admissível é zero e o limite superior é desconhecido. Tucci (1997) afirma que esse tipo de assimetria costuma ser típico das distribuições de variáveis hidrológicas.

Na Figura 5, estão representados os histogramas de frequência para o primeiro (a), segundo (b), terceiro (c) e quarto (d) ano hidrológico, considerando toda a série hidrológica utilizada neste trabalho, confirmando a distribuição assimétrica à direita.

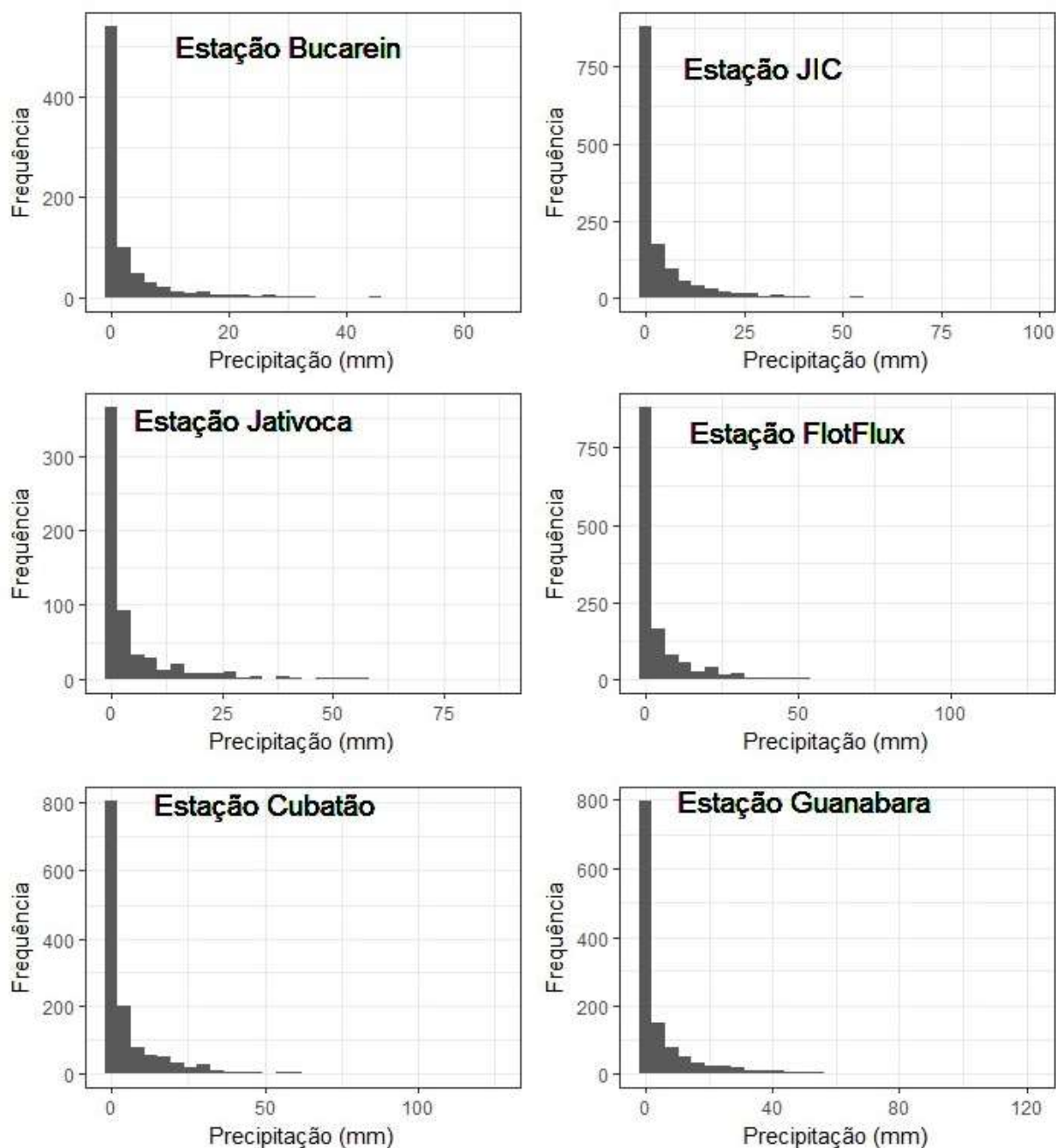
Figura 5 - Histogramas de frequência de cada ano hidrológico

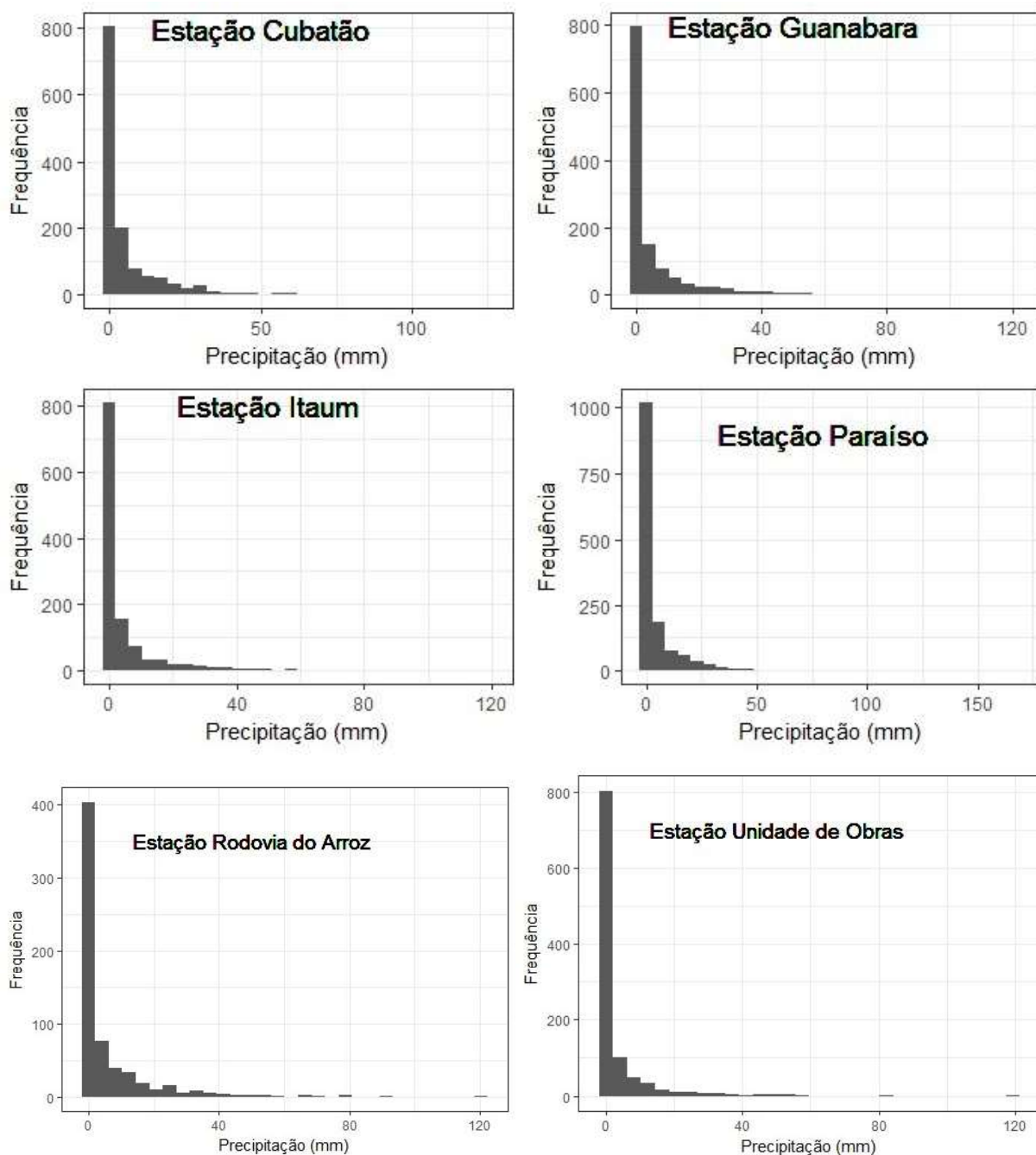


Fonte: A autora.

Já a Figura 6 representa os histogramas de frequência para cada estação do conjunto de dados.

Figura 6 – Histograma de frequência para as estações estudadas





Fonte: A autora.

Os histogramas da Figura 5 e Figura 6 são semelhantes aos histogramas obtidos no trabalho de Costa *et al* (2012), que utilizaram a técnica imputação múltipla para preencher dados diários de precipitação em seis estações com falhas no Estado da Paraíba, contemplando uma série de dados de 31 anos entre o período de 1979 e 2010.

Essas estações apresentavam porcentagem de dados diários faltantes de 8,60% até 51,84% considerando todo o conjunto de dados.

## 4.2 MÉTODOS DE PREENCHIMENTO DE FALHAS

Todos os métodos de preenchimento de falhas foram aplicados utilizando o *software R* e os resultados estão apresentados em um *link* (<http://imputacaometeorolo.wixsite.com/ppgec>) contendo um arquivo, compatível com o *Excel*, para cada método de preenchimento utilizado por ano hidrológico.

Com a intenção de evidenciar os valores obtidos pelos métodos ao preencher as falhas originais, estes valores foram destacados em vermelho nos arquivos que constam no *site* disponibilizado anteriormente.

### 4.2.1 Interpolação linear

Os dados faltantes preenchidos pelo método de interpolação linear estão destacados em vermelho nas tabelas e podem ser acessados no *link*: <http://imputacaometeorolo.wixsite.com/ppgec>. Os arquivos estão nomeados como IL-AH1, IL-AH2, IL-AH3 e IL-AH4 para o primeiro, segundo, terceiro e quarto ano hidrológico respectivamente.

Em geral, no preenchimento das falhas originais, não houve dificuldade ao aplicar o método da interpolação linear para o primeiro ano hidrológico.

Mas em algumas estações, como: estação Jativoca no segundo ano hidrológico; FlotFlux e Unidade de Obras no terceiro ano hidrológico; e Paraíso, Cubatão e FlotFlux no quarto ano hidrológico, todos os dados simulados foram iguais a zero. Isso ocorreu porque o período de dados retirados para as simulações estava situado entre dois dias sem chuva, o que afetou diretamente a correlação, que não pôde ser calculada quando todos os dados simulados são iguais a zero, logo, o coeficiente de determinação ( $R^2$ ) foi tomado como zero para estas situações.

As medidas de erro citadas no item 2.5 foram calculadas após o preenchimento das falhas simuladas, comparando os valores observados e os resultados gerados pelo método. Essas medidas de erros obtidas no segundo preenchimento com o método de interpolação linear podem ser observadas na Tabela 12.



Tabela 12 - Medidas de erros para cada estação em todos os 4 anos hidrológicos para o método de interpolação linear

Ano hidrológico	Estação	nº falhas originais	nº simulações	RMSE (mm)	MRE	MAE (mm)	R <sup>2</sup>	(%) dados faltantes
1º	Paraíso	8	8	23,511	1,790	17,832	0,001	2,192
	Unidade de Obras	73	73	12,227	3,863	1,983	0,002	20,000
2º	Bucarein	270	35	5,597	2,086	3,939	0,029	73,973
	Jativoca	111	111	10,974	0,640	5,263	0,000	30,411
	Cubatão	14	11	8,979	3,785	3,678	0,570	3,836
	Guanabara	58	58	5,459	2,123	1,860	0,000	15,890
	Rodovia do Arroz	99	99	9,286	0,462	4,497	0,027	27,123
	Unidade de Obras	40	41	23,176	0,374	9,020	0,033	10,959
3º	Bucarein	297	21	0,688	0,165	0,118	0,163	81,370
	JIC	76	76	11,389	4,117	1,712	0,001	20,822
	Cubatão	23	23	18,530	0,065	11,119	0,231	6,301
	Guanabara	75	73	11,356	1,913	2,638	0,104	20,548
	FlotFlux	127	119	9,220	0,479	3,875	0,000	34,795
	Unidade de Obras	12	7	5,067	1,000	3,624	0,000	3,288
	Itaum	233	53	9,244	1,313	2,135	0,044	63,836
	Paraíso	1	100	11,960	0,273	6,250	0,009	0,274
4º	Bucarein	58	58	16,570	9,994	8,837	0,001	15,890
	Cubatão	100	65	17,680	0,877	10,035	0,000	27,397
	Guanabara	102	52	19,541	6,079	10,104	0,000	27,945
	FlotFlux	3	3	5,165	1,000	4,191	0,000	0,822
	Unidade de Obras	264	37	16,069	0,653	8,829	0,000	72,329
	Itaum	13	13	13,190	0,806	7,945	0,220	3,562
	Paraíso	1	47	8,951	0,894	5,798	0,000	0,274

Fonte: A autora.

Na Tabela 12 é possível perceber que os valores de maior magnitude são referentes à medida de erro *RMSE*, os resultados obtidos para o erro médio absoluto (*MAE*) estão entre 0 e 18 mm aproximadamente, e para o erro relativo médio (*MRE*) os valores se concentram entre 0 e 4.

A estação com menores *RMSE* e *MAE* é a estação Bucarein no terceiro ano hidrológico com valores de 0,688 e 0,118 mm respectivamente, mesmo tendo o maior índice de dados faltantes (81,37%). A estação Cubatão obteve o menor erro relativo médio (*MRE*) no terceiro ano hidrológico, 0,065. Uma possível razão para esses dois casos pode ser o período de dados reais retirados, que além de estar situado entre duas observações próximas a zero, os valores registrados em sua maioria também eram próximos a zero. Isso acarretou em resultados de interpolações similares às leituras originais de precipitação e consequentemente baixos valores para as medidas de erro.

Para o método de interpolação linear não foram encontradas bibliografias que utilizavam este método para o preenchimento de falhas de dados de precipitação.

#### 4.2.2 Regressão linear múltipla

Os resultados dos valores simulados pelo método da regressão linear múltipla, destacados em vermelho, podem ser acessados no link: <http://imputacaometeorolo.wixsite.com/ppgec>. Os arquivos estão nomeados como RL-AH1, RL-AH2, RL-AH3 e RL-AH4 sendo o primeiro, segundo, terceiro e quarto ano hidrológico respectivamente.

No preenchimento das falhas originais, os parâmetros estatísticos obtidos com a regressão linear múltipla para cada estação com falha no primeiro ano hidrológico mostraram coeficientes de determinação ( $R^2$ ) bastante aceitáveis. Em geral esses coeficientes ficaram próximos de 80%, ou seja, na maioria das vezes o modelo mostrou um bom comportamento para se ajustar aos dados.

Para o segundo ano hidrológico nem todos os modelos de regressão linear múltipla alcançaram bons resultados. A estação Bucarein, por exemplo, obteve  $R^2=0,477$ , indicando o pior modelo do segundo ano hidrológico, porém as outras estações atingiram coeficiente de determinação próximos a 70%, ressaltando a estação Unidade de Obras que com  $R^2=0,972$  foi o melhor modelo ajustado.

No terceiro ano da série hidrológica, todas as dez estações do conjunto de dados apresentavam falhas. Portanto, para algumas estações foram elaboradas mais de uma equação do modelo, por exemplo, para a estação Bucarein o preenchimento teve que ser dividido em dois períodos, pois do dia 01/10/204 até o dia 06/01/205 apenas as estações

Paraíso, JIC e FlotFlux puderam ser utilizadas como variáveis independentes no preenchimento. E do dia 04/03/2015 até 17/09/2015 as estações Unidade de Obras e Cubatão estavam disponíveis. Isso foi feito para obedecer os critérios descritos no item 3.4.

Não conseguiu-se ajustar os dados das estações Guanabara e Unidade de Obras, deste período da série, aos critérios estabelecidos, pois na maioria das vezes os dias com falhas dessas duas estações coincidiam com o período de falhas das estações que poderiam ser utilizadas como variáveis independentes, e por esse motivo preferiu-se não preencher as estações Guanabara e Unidade de Obras com este método no terceiro ano hidrológico. As demais estações puderam ser preenchidas sem problemas.

Neste terceiro ano hidrológico a estação Itaum obteve o melhor ajuste do modelo, com  $R^2 = 0,934$  e a estação Bucarein novamente obteve o menor valor para o coeficiente de determinação,  $R^2 = 0,024$ , que pode ser explicado pelo preenchimento realizado em dois períodos diferentes.

Ainda no preenchimento das falhas originais, no quarto ano hidrológico não houve bons resultados quanto nos anos anteriores, porém, a estação Unidade de Obras alcançou o coeficiente de determinação de 0,934, o melhor para este ano. A estação Paraíso com  $R^2 = 0,424$  obteve o menor coeficiente, seguida pela estação FlotFlux, Bucarein, Itaum, com  $R^2$  de 0,464, 0,519 e 0,585 respectivamente.

O preenchimento dos dados reais que foram retirados, apontaram as medidas de erros apresentadas na Tabela 12.

Tabela 12 - Medidas de erros para cada estação em todos os 4 anos hidrológicos para o método de regressão linear múltipla (Continua)

Ano hidrológico	Estação	nº falhas originais	nº simulações	RMSE (mm)	MRE	MAE (mm)	$R^2$	(%) dados faltantes
1º	Paraíso	8	8	10,637	0,467	2,644	0,189	2,192
	Unidade de Obras	73	73	3,047	0,247	0,724	0,948	20,000
2º	Bucarein	270	35	6,322	3,338	3,842	0,406	73,973
	Jativoca	111	111	5,590	0,459	2,639	0,689	30,411
	Cubatão	14	11	4,184	0,515	2,857	1,000	3,836
	Guanabara	58	58	1,941	0,016	0,060	0,872	15,890

	Rodovia do Arroz	99	99	10,718	0,767	2,760	0,449	27,123
	Unidade de Obras	40	41	0,617	0,010	0,035	0,999	10,959
3º	Bucarein	297	21	2,205	3,693	3,689	0,003	81,370
	JIC	76	76	11,989	4,005	1,099	0,369	20,822
	Cubatão	23	23	11,927	0,647	5,619	0,743	6,301
	Guanabara	75	73	4,638	0,018	0,367	0,755	20,548
	FlotFlux	127	119	2,488	0,097	0,100	0,912	34,795
	Unidade de Obras	12	7	5,600	0,855	3,689	0,693	3,288
	Itaum	233	53	5,664	0,196	0,158	0,712	63,836
4º	Paraíso	1	100	8,369	0,334	0,182	0,480	0,274
	Bucarein	58	58	10,209	0,254	5,742	0,825	15,890
	Cubatão	100	65	6,830	2,234	0,398	0,834	27,397
	Guanabara	102	52	10,089	0,403	3,563	0,748	27,945
	FlotFlux	3	3	3,709	0,099	1,334	0,156	0,822
	Unidade de Obras	264	37	6,060	0,051	1,247	0,805	72,329
	Itaum	13	13	11,819	0,118	5,779	0,046	3,562
	Paraíso	1	47	14,333	2,234	6,919	0,567	0,274

Fonte: A autora.

No preenchimento das falhas simuladas, a estação Unidade de Obras obteve os melhores resultados de medidas de erro no primeiro ano hidrológico, se comparada com a estação Paraíso. Percebe-se na Tabela 12 que mesmo com o maior percentual de dados faltantes entre elas, a estação Unidade de Obras alcançou o melhor valor para  $R^2$  (0,948). No segundo ano hidrológico a performance dessa estação, para o método de regressão linear múltipla, é de novo superior às demais, com  $R^2$  ainda maior que no primeiro ano hidrológico (0,999).

Divergindo do preenchimento das falhas originais, a estação Bucarein obteve o menor valor para o *RMSE* do terceiro ano hidrológico, mas o pior para o coeficiente de determinação. A estação FlotFlux pode ser considerada a estação com melhor desempenho neste método no terceiro ano da série hidrológica.

O quarto ano hidrológico é marcado pelos maiores valores de *RMSE* e *MAE*, no qual a estação Paraíso atingiu 14,333, e 6,919 mm respectivamente, ainda que seja a estação com menor percentual de dados faltantes para toda a série hidrológica.

Fernandez (2007) utilizou regressão linear múltipla com 13 estações do Rio Grande do Sul e Mello (2015) com 42 estações de Santa Catarina para o preenchimento de falhas de dados de precipitação mensal. Os autores compararam diferentes técnicas estatísticas de preenchimento de falhas, além da regressão múltipla e constataram que este método obteve os melhores resultados. Oliveira *et al* (2010) destacaram o desempenho da metodologia RLM em relação aos métodos de ponderação regional (PR) e ponderação regional com base em regressões lineares (PRBRL) para preenchimento de falhas de dados mensais e anuais de precipitação e verificaram que a regressão linear múltipla apresentou os menores valores de desvios relativos ao ser comparada com os métodos citados, contudo, constataram uma melhoria nas estimativas do preenchimento se combinada à metodologia do vetor regional (VR) com os demais métodos. Zeilhofer *et al.* (2003) estabelecem que para séries temporais das vazões diárias e da precipitação, o preenchimento deve ser feito escolhendo em cada execução pares de pontos com correlação máxima, definindo uma regressão linear entre os conjuntos de dados.

Os dados da Defesa Civil de Joinville quando preenchidos com o método da regressão linear múltipla, também apresentaram bons resultados, conforme mostra a Tabela 12, alcançando valores de MRE e MAE próximos a zero, além de RMSE de no máximo 12 mm.

#### **4.2.3 Método das Correlações**

Os valores estimados por meio do método das correlações estão destacados em vermelho no *link*: <http://imputacaometeorolo.wixsite.com/ppgec>, nomeados como MDC-AH1, MDC-AH2, MDC-AH3 e MDC-AH4, para o primeiro, segundo, terceiro e quarto ano hidrológico respectivamente.

Diferentemente dos métodos anteriores, nos quais foi necessário agrupar as estações conforme os critérios estabelecidos, este método das correlações trabalhou simultaneamente com todas as estações do conjunto de dados, facilitando sua aplicação, pois não era necessário escolher qual estação utilizar, o próprio método trabalhou com todas as dez estações ao mesmo tempo e identificou quais as mais apropriadas para preencher as falhas de cada uma delas. Com restrição das estações Jativoca e Rodovia do Arroz que não puderam ser preenchidas, as estações remanescentes tiveram suas

falhas originais devidamente completadas. O preenchimento das simulações resultaram nos valores das medidas de erro e coeficiente de determinação apresentados na Tabela 13, na qual constam também a porcentagem de dados faltantes para todas as estações, bem como o número de falhas e simulações em cada ano hidrológico.

Tabela 13 - Medidas de erros para cada estação em todos os 4 anos hidrológicos para o método das correlações

Ano hidrológico	Estação	nº falhas originais	nº simulações	RMSE (mm)	MRE	MAE (mm)	R <sup>2</sup>	(%) dados faltantes
1º	Paraíso	8	8	2,950	0,224	0,894	0,864	2,192
	Unidade de Obras	73	73	2,193	0,009	0,296	0,963	20,000
2º	Bucarein	270	35	3,989	1,904	2,390	0,433	73,973
	Jativoca	111	111	8,500	0,500	3,422	0,533	30,411
	Cubatão	14	11	4,461	0,375	2,964	0,999	3,836
	Guanabara	58	58	1,525	0,085	0,313	0,918	15,890
	Rodovia do Arroz	99	99	12,599	0,589	2,780	0,345	27,123
	Unidade de Obras	40	41	3,027	0,048	0,656	0,982	10,959
3º	Bucarein	297	21	1,242	1,057	0,154	0,006	81,370
	JIC	76	76	12,242	1,038	3,060	0,369	20,822
	Cubatão	23	23	12,050	0,370	7,204	0,650	6,301
	Guanabara	75	73	3,448	0,085	1,115	0,948	20,548
	FlotFlux	127	119	2,488	0,097	0,100	0,912	34,795
	Unidade de Obras	12	7	3,801	0,889	2,901	0,872	3,288
	Itaum	233	53	9,254	0,537	0,278	0,239	63,836
	Paraíso	1	100	8,804	0,468	0,063	0,472	0,274
4º	Bucarein	58	58	11,951	0,621	7,098	0,329	15,890
	Cubatão	100	65	13,027	0,479	5,406	0,342	27,397
	Guanabara	102	52	8,310	0,129	1,673	0,782	27,945
	FlotFlux	3	3	4,813	0,128	2,147	0,111	0,822
	Unidade de Obras	264	37	13,491	0,035	3,604	0,201	72,329
	Itaum	13	13	11,904	0,110	5,034	0,031	3,562
	Paraíso	1	47	6,540	1,282	1,179	0,456	0,274

Fonte: A autora.

A partir da Tabela 13, é possível perceber que para o primeiro ano hidrológico, apesar das estações Paraíso e Unidade de Obras alcançarem valores similares para as medidas de erro, a estação Unidade de Obras obteve o menor erro relativo médio (*MRE*), inclusive o melhor do método, atingindo o valor de 0,009. Os valores dos coeficientes de determinação  $R^2$  também foram parecidos, sendo 0,864 para a estação Paraíso, e 0,963 para a Unidade de Obras.

A estação Guanabara, no segundo ano hidrológico, se destacou pelos baixos valores das medidas de erro, *RMSE*=1,525 mm, *MRE*=0,085, *MAE*=0,313 mm e também pelo bom  $R^2$ =0,918. Entretanto, a estação com melhor coeficiente de determinação foi a estação Cubatão com  $R^2$ =0,999. Neste mesmo ano, a estação Rodovia do Arroz obteve o pior índice para o *RMSE*, o terceiro maior do método, 12,599 mm.

No terceiro ano a estação Bucarein, com mais de 80% de dados faltantes, apresenta o menor valor de  $R^2$ , 0,006. Entretanto, os valores obtidos para as medidas de erro *RMSE*, *MRE* e *MAE* foram respectivamente 1,242 mm, 1,057 e 0,154 mm, considerados ótimos resultados. A estação Cubatão além do maior erro médio absoluto (*MAE*=7,204 mm), também apresenta o segundo maior *RMSE* no terceiro ano, 12,050 mm, tendo valor melhor apenas que a estação JIC (*RMSE*=12,242 mm) para este erro. Ainda no terceiro ano hidrológico, o melhor erro médio relativo (*MRE*) e coeficiente de determinação foram obtidos pela estação Guanabara, são eles 0,085 e 0,948 respectivamente. A estação Paraíso, com 100 simulações no terceiro ano hidrológico, apresentou o menor erro médio absoluto do ano e também para o método, *MAE*=0,063 mm.

O quarto ano hidrológico apresenta os valores máximos para a medida *RMSE* no método das correlações, o maior valor 13,491mm foi obtido com a estação Unidade de Obras. Curiosamente, essa mesma estação obteve o menor erro relativo médio para o quarto ano hidrológico, *MRE*=0,035. A única estação com  $R^2$  acima de 50% foi a estação Guanabara com aproximadamente 78%, as demais ficaram entre 10% e 45%.

Para o método das correlações não foram encontradas bibliografias específicas que utilizavam o pacote *hyfo* no preenchimento de falhas de dados de precipitação. Porém, como este método é baseado em correlação e regressão linear, o seu desempenho pode ser analisado em conjunto com os trabalhos referenciados no método

de Regressão Linear Múltipla. Os métodos de regressão linear múltipla e o método das correlações apresentaram, além de bons resultados, desempenho semelhantes por tratarem de metodologias similares.

#### **4.2.4 Redes Neurais Artificiais**

Os resultados para este método também se encontram no link: <http://imputacaometeorolo.wixsite.com/ppgec>, nomeados como: RNA-AH1, RNA-AH2, RNA-AH3 e RNA-AH4 correspondentes ao primeiro, segundo, terceiro e quarto ano hidrológico respectivamente.

Ao aplicar este método a cada ano hidrológico, conforme foi feito com os outros métodos, houve uma peculiaridade. Além da rede apresentar erros grosseiros, não conseguia alcançar sua convergência. Então, foram feitas duas tentativas para que o método tivesse bons resultados, a primeira delas foi agrupar os dados utilizando o critério de sazonalidade iniciando na Primavera de 2012. Porém, ainda assim a rede apontava erros ou não convergia. Já na segunda tentativa, os dados foram separados mês a mês e então as redes convergiram mais rapidamente e alcançaram erros aceitáveis. Em todas as três tentativas foram utilizados 15 neurônios.

Para o primeiro ano hidrológico as 83 falhas originais foram preenchidas com redes neurais diferentes, no mínimo uma para cada mês. Se no mesmo mês houvesse falhas em mais de uma estação, eram criadas uma rede neural para cada estação referente ao mesmo mês. O primeiro ano hidrológico resultou em 16 redes neurais artificiais, sendo 12 para a estação Unidade de Obras (1 rede por mês), duas para a estação Paraíso e uma para a Cubatão, bem como para a Guanabara. Com o ajuste dos parâmetros da rede neural, a estação Paraíso para o mês de novembro de 2012 obteve o maior erro de precisão, aproximadamente 0,11, enquanto o valor deste erro para as redes neurais das demais estações ficou abaixo de 0,016.

No segundo ano hidrológico foram contabilizadas 592 falhas originais ao todo, e criadas, em torno de 20 redes neurais artificiais. A estação Cubatão obteve o maior erro para este segundo ano, no mês de outubro de 2013, com o valor de 0,035 aproximadamente, as demais estações permaneceram com o valor do erro abaixo de 0,015.



O terceiro ano da série hidrológica teve o maior número de redes neurais criadas, foram cerca de 50. A maioria destas redes obtiveram erros próximos a zero, com exceção da estação JIC no mês de fevereiro de 2015 que alcançou o valor de 2,2086 para o erro total da rede, e a estação Guanabara neste mesmo mês, com erro de 21,48.

Para o quarto ano hidrológico foram necessárias 34 redes neurais artificiais para preencher as falhas originais e o maior erro obtido foi com a estação Bucarein no mês de maio de 2016, atingindo o valor de 2,84.

As falhas simuladas foram preenchidas da mesma maneira, no mínimo uma rede neural artificial para cada mês. Os resultados das medidas de erros calculadas após o preenchimento das falhas simuladas estão na Tabela 14.

Tabela 14 - Medidas de erros para cada estação em todos os 4 anos hidrológicos para o método das redes neurais artificiais (Continua)

Ano hidrológico	Estação	nº falhas originais	nº simulações	RMSE (mm)	MRE	MAE (mm)	R <sup>2</sup>	(%) dados faltantes
1º	Paraíso	8	8	15,279	1,254	9,770	0,384	2,192
	Unidade de Obras	73	73	12,858	0,848	6,900	0,069	20,000
2º	Bucarein	270	35	5,684	1,969	3,603	0,001	73,973
	Jativoca	111	111	9,775	3,107	5,307	0,088	30,411
	Cubatão	14	11	5,631	0,592	3,915	0,874	3,836
	Guanabara	58	58	4,923	1,364	0,123	0,090	15,890
	Rodovia do Arroz	99	99	8,607	3,331	1,460	0,126	27,123
	Unidade de Obras	40	41	21,648	0,243	7,674	0,372	10,959
3º	Bucarein	297	21	0,627	0,509	0,248	0,329	81,370
	JIC	76	76	10,317	1,444	2,912	0,369	20,822
	Cubatão	23	23	13,031	1,476	3,662	0,360	6,301
	Guanabara	75	73	10,397	0,467	3,957	0,016	20,548
	FlotFlux	127	119	8,643	0,422	3,497	0,244	34,795
	Unidade de Obras	12	7	4,982	1,106	3,656	0,556	3,288
	Itaum	233	53	8,502	0,937	1,900	0,142	63,836
	Paraíso	1	100	10,525	1,165	2,850	0,071	0,274
4º	Bucarein	58	58	8,623	1,297	2,291	0,368	15,890
	Cubatão	100	65	20,168	0,036	5,873	0,001	27,397
	Guanabara	102	52	17,175	0,302	4,127	0,018	27,945

FlotFlux	3	3	3,377	0,454	2,300	0,956	0,822
Unidade de Obras	264	37	15,589	0,096	4,004	0,077	72,329
Itaum	13	13	9,041	0,521	4,771	0,768	3,562
Paraíso	1	47	9,819	0,846	4,288	0,035	0,274

Fonte: A autora.

Com base na Tabela 14 é possível perceber que as estações do primeiro ano hidrológico não atingiram bons resultados para as medidas de erro *RMSE*, *MAE* e  $R^2$ , apenas o erro médio relativo (*MRE*) dessas duas estações atingiu bons valores.

No segundo ano hidrológico, a estação Guanabara obteve o menor *MAE* (0,123 mm) com 58 dados simulados. A estação Unidade de Obras obteve o maior *MAE* e o pior *RMSE* do segundo ano hidrológico, em compensação, atingiu o melhor valor para o erro relativo médio do ano. Os valores do coeficiente de determinação para este mesmo ano foram bem abaixo da unidade, entre 00,1 e 0,037, com exceção da estação Cubatão que obteve  $R^2=0,874$ .

No terceiro ano hidrológico a estação Bucarein se destacou não só pelo menor *RMSE* (0,627 mm) do método, como também pelos baixos valores das medidas de erro *MAE* e *MRE*. Novamente a maioria das estações não alcançaram coeficientes de determinação acima de 40%, demonstrando também uma baixa correlação entre os valores observados e os valores simulados.

A estação Cubatão no quarto ano hidrológico obteve o segundo maior *RMSE* e o menor *MRE* do método RNA. O destaque do  $R^2$  ficou com a estação FlotFlux que atingiu 0,956, o melhor do método.

Com o objetivo de utilizar redes neurais artificiais na interpolação espacial de dados pluviométricos médios mensais no Estado de Alagoas, Wanderley *et al* (2014) com dados disponíveis compreendidos entre 1965 e 1980 de 245 estações pluviométricas, encontraram valores de *RMSE* que variavam entre 29,89 a 41,63 mm, bem como, valores de *MAE* entre 21,15 a 30,77 mm para o mês de maior variabilidade espacial. Para o mês de menor variabilidade espacial os valores de *RMSE* ficaram entre 0,67 a 4,38 mm.

Ao empregar diferentes métodos de redes neurais artificiais para preencher dados diários de precipitação e temperatura, Coulibaly e Evora (2007) constataram que a rede neural artificial do tipo MLP é mais adequada para estimar dados de precipitação ao

comparar o erro médio absoluto (MAE) e coeficiente de correlação com as demais arquiteturas de redes neurais utilizadas. Estes autores encontraram valores de MAE entre 1,21 e 1,96 mm para a arquitetura do tipo MLP, considerado os melhores, e valores entre 1,60 e 2,33 mm para a arquitetura *RBF - Radial Basis Function*.

Para os dados de Joinville utilizados nesta pesquisa também foi utilizada a rede neural com arquitetura MLP, entretanto, os resultados obtidos com este método não foram os melhores, sendo superados por métodos menos sofisticados como regressão linear múltipla e método das correlações.

Mello (2015) também apontou a regressão linear múltipla como o melhor método de preenchimento de falhas para a região de Joinville, porém, não utilizou redes neurais artificiais nas suas comparações. Já os autores Depiné *et al* (2013) relataram que as séries curtas, situação semelhante a desta pesquisa, obtiveram os melhores resultados para o coeficiente NS.

#### **4.2.5 Imputação Múltipla – Pacote mice**

Assim como os outros métodos aplicados nesta pesquisa, os resultados deste método de Imputação Múltipla para o preenchimento das falhas originais, também se encontram disponíveis no *link*: <http://imputacaometeorolo.wixsite.com/ppgec>, com os seguintes nomes: IM-AH1, IM-AH2, IM-AH3 e IM-AH4, representando o primeiro, segundo, terceiro e quarto ano hidrológico respectivamente.

No preenchimento das falhas originais com este método não houve dificuldades para os dois primeiros anos da série hidrológica, porém, para o terceiro e quarto ano as estações Jativoca e Rodovia do Arroz não puderam ser preenchidas. Isso ocorreu porque o método requer ao menos uma observação para conseguir realizar a imputação dos valores faltantes, e essas duas estações não tinham registro de precipitação nos dois últimos anos da série.

Como no primeiro preenchimento constam os resultados obtidos pelo método para as falhas originais, do segundo preenchimento nos interessa como o método se comportou ao preencher as falhas simuladas. Na Tabela 16 estão apresentadas as estatísticas para cada estação pluviométrica em cada ano hidrológico e seus respectivos

erros, coeficiente de determinação e também a porcentagem de dados faltantes e número de simulações que foram realizadas em cada período.

Tabela 15 - Medidas de erros para cada estação em todos os 4 anos hidrológicos para o método de imputação múltipla

Ano hidrológico	Estação	nº falhas originais	nº simulações	RMSE (mm)	MRE	MAE (mm)	R <sup>2</sup>	(%) dados faltantes
1º	Paraíso	8	8	4,867	0,442	3,607	0,811	2,192
	Unidade de Obras	73	73	6,868	0,051	3,091	0,867	20,000
2º	Bucarein	270	35	4,960	1,551	2,758	0,173	73,973
	Jativoca	111	111	12,563	3,751	7,893	0,132	30,411
	Cubatão	14	11	4,441	0,539	3,091	0,874	3,836
	Guanabara	58	58	8,401	1,028	4,320	0,852	15,890
	Rodovia do Arroz	99	99	14,056	1,981	4,525	0,436	27,123
	Unidade de Obras	40	41	10,445	0,090	2,705	0,913	10,959
3º	Bucarein	297	21	0,782	0,524	0,370	0,002	81,370
	JIC	76	76	9,145	1,526	2,121	0,369	20,822
	Cubatão	23	23	17,434	1,666	7,590	0,059	6,301
	Guanabara	75	73	9,897	0,298	3,367	0,003	20,548
	FlotFlux	127	119	15,825	5,296	6,815	0,056	34,795
	Unidade de Obras	12	7	5,067	1,000	3,624	0,000	3,288
	Itaum	233	53	17,919	7,014	11,807	0,220	63,836
	Paraíso	1	100	12,290	0,596	3,935	0,055	0,274
4º	Bucarein	58	58	13,335	0,881	8,138	0,057	15,890
	Cubatão	100	65	15,496	0,280	1,664	0,278	27,397
	Guanabara	102	52	23,604	3,184	14,832	0,257	27,945
	FlotFlux	3	3	4,699	0,057	1,905	0,077	0,822
	Unidade de Obras	264	37	6,654	0,436	1,013	0,761	72,329
	Itaum	13	13	11,858	0,756	6,987	0,555	3,562
	Paraíso	1	47	10,888	1,280	2,056	0,241	0,274

Fonte: A autora.

O preenchimento das falhas simuladas do primeiro ano hidrológico resultou em valores de *RMSE* e *MAE* similares para as estações Paraíso e Unidade de Obras, com exceção do erro relativo médio, que atingiu o valor de 0,051 para a estação Unidade de

Obras, o melhor do método inclusive, e 0,442 para a estação Paraíso. A estação Unidade de Obras também superou a estação Paraíso no coeficiente de determinação, obtendo  $R^2=0,867$ .

Para o segundo ano da série os valores de *MRE* e *MAE* das estações também foram próximos, com destaque novamente para a estação Unidade de Obras que com 0,090 obteve o menor erro relativo médio do terceiro ano hidrológico e o melhor coeficiente de determinação,  $R^2=0,913$ . Já os valores de *RMSE*, apesar de não serem os maiores, foram bastante altos.

No terceiro ano todas as estações apresentaram valores baixos de  $R^2$ , a maioria delas com menos de 10%, o que representa uma correlação ainda menor entre os dados observados e simulados pelo método. Entretanto, neste mesmo ano foram obtidos os menores valores para as medidas de erro *RMSE* e *MRE* do método, ambos pertencentes à estação Bucarein, 0,782 mm e 0,370 respectivamente.

O quarto ano envolve as maiores medidas de *RMSE*, sendo a estação Guanabara com maior valor, 23,604 mm. Esta mesma estação também obteve o maior *MAE*, 14,832 mm, enquanto os coeficientes de determinação das estações neste último ano da série permaneceram abaixo de 80%, os valores apresentados para o erro médio relativo alcançaram resultados aceitáveis.

Costa *et al* (2012) também utilizaram a técnica *Multivariate Imputation by Chained Equations - MICE* de imputação múltipla disponível no *software* livre R para imputar dados diários e mensais de precipitação em seis estações com falhas do Estado da Paraíba entre 1979 e 2010. E obtiveram valores de correlações entre os dados observados e imputados de 0,7 e 1,0 para os dados acumulados mensais, e valores entre 0,4 e 0,8 para os dados diários.

Os dados das estações da Defesa Civil de Joinville não apresentaram coeficientes de determinação satisfatórios para todos os anos da série hidrológica, em geral esses valores ficaram abaixo de 0,4. Apenas o primeiro ano obteve valor de  $R^2$  acima de 0,8 para todas as suas estações, o segundo ano atingiu três bons valores, o terceiro ano não obteve nenhum acima de 0,4 e no quarto ano apenas a estação Unidade de Obras, com  $R^2 = 0,761$ , obteve um bom resultado.

### 4.3 MEDIDAS DE ERRO

As medidas de erro descritas no item 2.5 foram calculadas para cada estação em todos os anos da série hidrológica, após a retirada de dados originais, e apresentadas na forma de tabelas em seus respectivos métodos de preenchimento conforme demonstrado a partir do item 4.2.

Essas medidas de erros estão expostas em forma de gráficos, nos quais constam todas as estações do conjunto de dados e os respectivos valores das medidas de erro obtidas por elas em cada método de preenchimento aplicado.

#### 4.3.1 Raiz do Erro Quadrático Médio - RMSE

A raiz do erro quadrático médio (*RMSE*) foi utilizada como uma medida de erro de previsão e calculada para cada método de preenchimento em todos os anos hidrológicos.

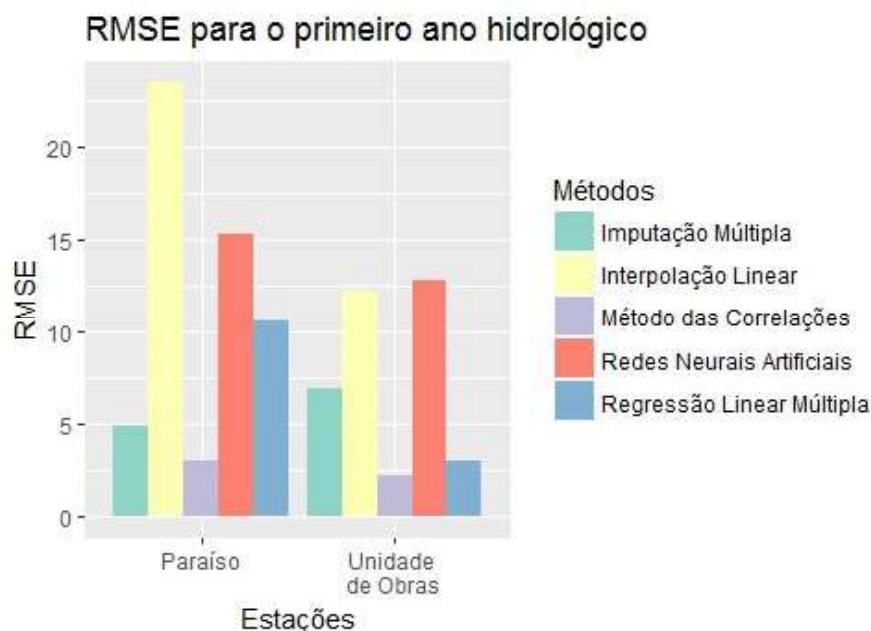
Para o primeiro ano hidrológico, a Tabela 16 e Figura 7 apresentam os valores obtidos de *RMSE* para o primeiro ano hidrológico, referente aos métodos e estações utilizadas.

Tabela 16 - Erro RMSE para o primeiro ano hidrológico

Estação	Métodos de preenchimento				
	IL (mm)	RLM (mm)	MDC (mm)	RNA (mm)	IM (mm)
Paraíso	23,511	10,637	2,950	15,279	4,867
Unidade de Obras	12,227	3,047	2,193	12,859	6,868

Fonte: A autora.

Figura 7 - RMSE para o primeiro ano hidrológico



Fonte: A autora.

A partir da Figura 7 e Tabela 16 podemos perceber que a estação Paraíso obteve os maiores valores de RMSE para os métodos de interpolação linear, regressão linear múltipla, método das correlações e redes neurais artificiais. Isso significa dizer que apenas no método de imputação múltipla essa estação obteve menor erro RMSE que a estação Unidade de Obras.

O método das correlações se mostrou melhor, por obter os menores valores de RMSE, em ambas as estações, logo, pode-se dizer que para este erro (RMSE), o método com melhor desempenho no primeiro ano hidrológico é o método das correlações, enquanto que a interpolação linear é o pior método para este mesmo período.

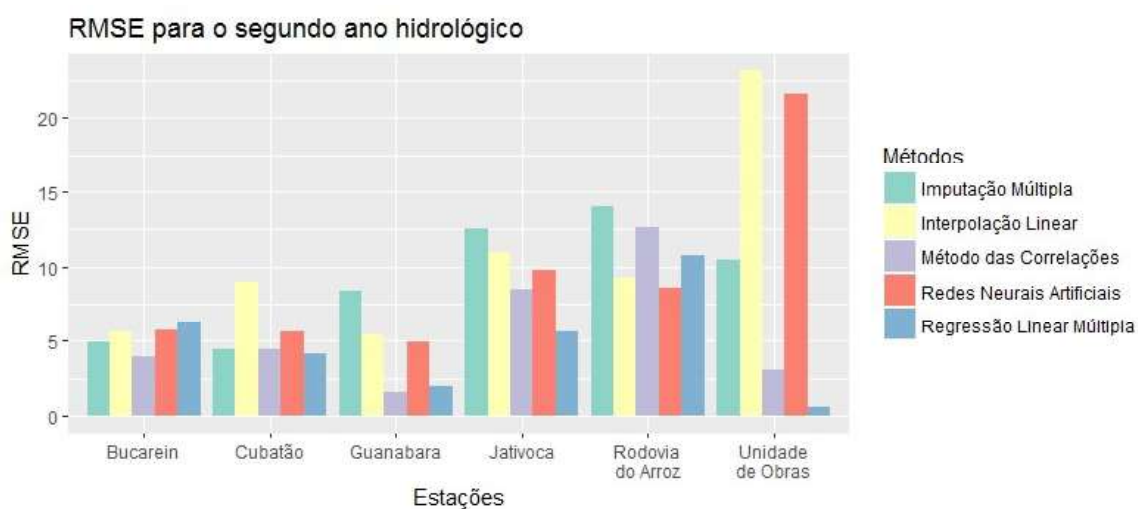
No segundo ano da série hidrológica seis das dez estações tiveram suas falhas simuladas preenchidas, os valores estão apresentados na Tabela 17 e Figura 8.

Tabela 17- Erro RMSE para o segundo ano hidrológico

Estação	Métodos de preenchimento				
	IL (mm)	RLM (mm)	MDC (mm)	RNA (mm)	IM (mm)
Bucarein	5,597	6,322	3,989	5,684	4,960
Jativoca	10,974	5,590	8,500	9,775	12,563
Cubatão	8,979	4,184	4,461	5,631	4,441
Guanabara	5,459	1,941	1,525	4,923	8,401
Rodovia do Arroz	9,286	10,718	12,599	8,607	14,056
Unidade de Obras	23,176	0,617	3,027	21,648	10,445

Fonte: A autora.

Figura 8 - RMSE para o segundo ano hidrológico



Fonte: A autora.

Neste segundo ano hidrológico os valores de RMSE foram em geral menores que os valores obtidos no primeiro ano, porém, o desempenho dos métodos de interpolação linear e redes neurais artificiais para a estação Unidade de Obras não foi bom, apresentando valores de 23,176 e 21,648 mm respectivamente.

Com base na Tabela 17 e Figura 8 podemos afirmar que o método com melhor desempenho nesse segundo ano da série hidrológica é a regressão linear múltipla, por



obter os menores valores de RMSE, mas o método das correlações também se mostrou bastante eficaz. Já a interpolação linear, com a maioria dos valores de RMSE maiores que os outros métodos, não se mostrou adequada para o preenchimento de falhas.

Os valores de RMSE no terceiro ano hidrológico ficaram abaixo de 20 mm e as estações obtiveram valores bastante próximos entre os métodos, como mostra a Tabela 18 e Figura 9.

Tabela 18 - Erro RMSE para o terceiro ano hidrológico

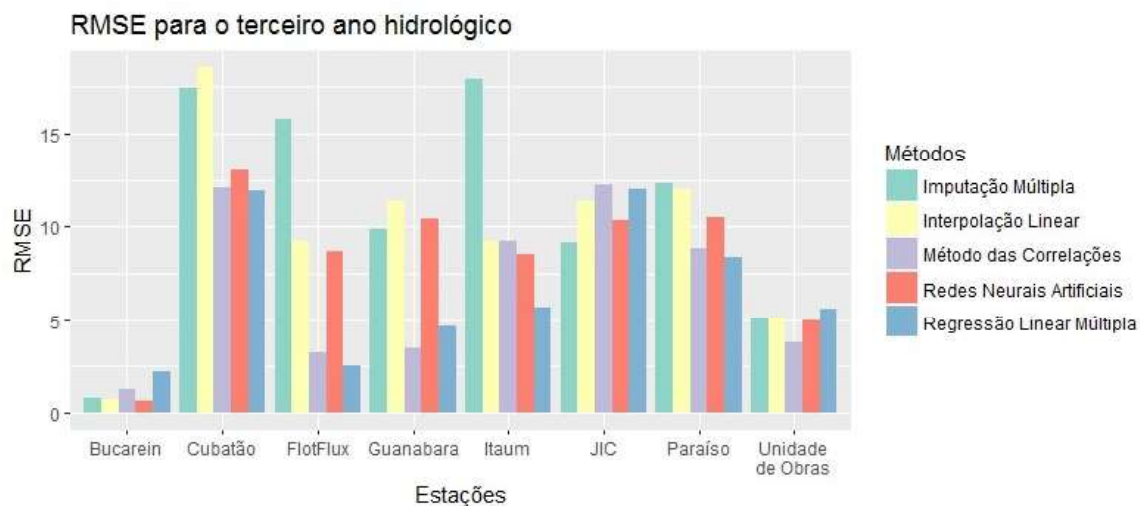
Estação	Métodos de preenchimento				
	IL (mm)	RLM (mm)	MDC (mm)	RNA (mm)	IM (mm)
Bucarein	0,688	2,205	1,242	0,627	0,782
JIC	11,389	11,989	12,242	10,317	9,145
Cubatão	18,529	11,927	12,049	13,031	17,434
Guanabara	11,356	4,638	3,448	10,400	9,897
FlotFlux	9,220	2,488	2,488	8,643	15,825
Unidade de Obras	5,067	5,600	3,801	4,982	5,067
Itaum	9,243	5,664	9,254	8,502	17,919
Paraíso	11,960	8,369	8,804	10,525	12,29

Fonte: A autora.

A estação Cubatão obteve o pior desempenho nos métodos de interpolação linear e redes neurais artificiais além de ficar bem próximo dos maiores valores de RMSE para os métodos de regressão linear múltipla, método das correlações e imputação múltipla.

Neste terceiro ano da série os métodos de regressão linear múltipla de método das correlações obtiveram os melhores resultados para o erro RMSE com valores bem próximos um do outro.

Figura 9 - RMSE para o terceiro ano hidrológico



Fonte: A autora.

O quarto ano hidrológico envolve registros de precipitação de sete estações, Tabela 19 e Figura 10.

Tabela 19 - Erro RMSE para o quarto ano hidrológico, unidade em mm

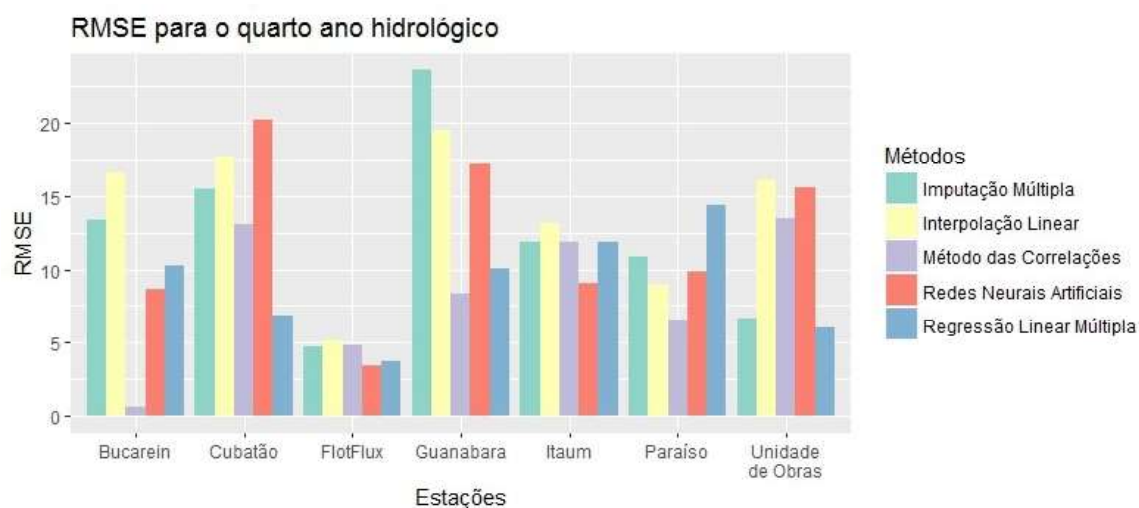
Estação	Métodos de preenchimento				
	IL (mm)	RLM (mm)	MDC (mm)	RNA (mm)	IM (mm)
Bucarein	16,569	10,209	0,621	8,623	13,335
Cubatão	17,680	6,830	13,027	20,168	15,496
Guanabara	19,541	10,089	8,310	17,175	23,604
FlotFlux	5,1651	3,709	4,813	3,377	4,699
Unidade de Obras	16,069	6,060	13,491	15,589	6,654
Itaum	13,189	11,819	11,904	9,041	11,858
Paraíso	8,951	14,333	6,539	9,819	10,888

Fonte: A autora.

Para o quarto ano da série o maior valor de RMSE obtido foi de 23,604 mm para a estação Guanabara no método de imputação múltipla. Porém, o método de interpolação linear obteve os maiores valores do erro RMSE para as estações Bucarein, FlotFlux,

Unidade de Obras e Itaum, caracterizando o pior desempenho dentre os métodos de preenchimento aplicados na pesquisa no quarto ano hidrológico.

Figura 10 - RMSE para o quarto ano hidrológico



Fonte: A autora.

A partir da Figura 10 pode-se destacar o desempenho de três métodos, são eles: regressão linear múltipla, método das correlações e redes neurais artificiais. Cada um desses métodos obteve dois melhores resultados para o valor de RMSE, sendo estação Cubatão e Unidade de Obras para o método da regressão linear múltipla, Bucarein e Paraíso para o método das correlações e FlotFlux e Itaum para as redes neurais artificiais.

#### 4.3.2 Erro Médio Relativo - MRE

O erro médio relativo (*MRE*) também foi utilizado como uma medida de erro e calculado para cada método de preenchimento em todos os anos hidrológicos.

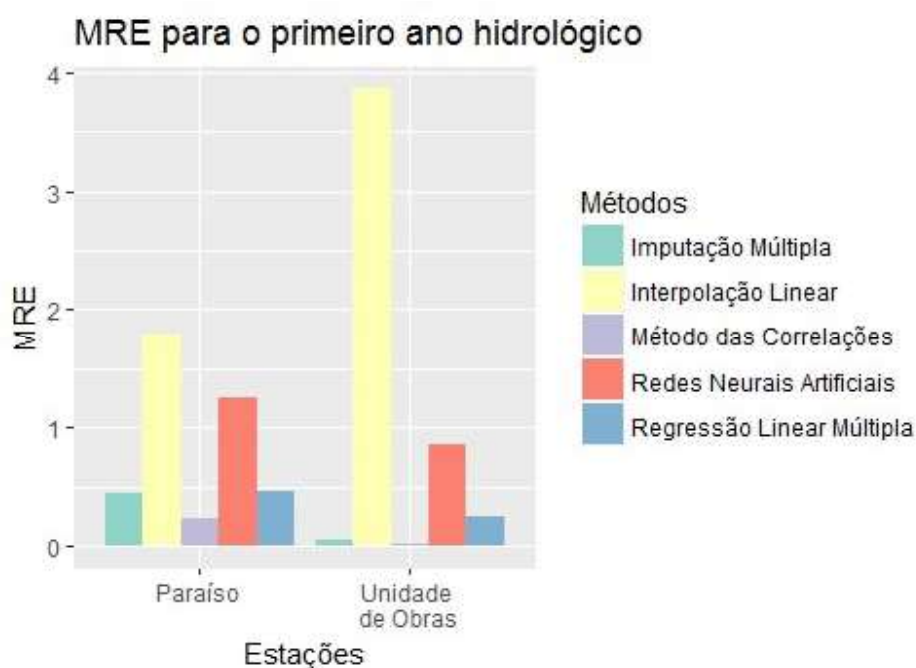
Para o primeiro ano hidrológico, a Tabela 20 e Figura 11 apresentam os valores obtidos de *MRE* para o primeiro ano hidrológico, referente aos métodos e estações utilizadas.

Tabela 20 - Erro MRE para o primeiro ano hidrológico

Estação	Métodos de preenchimento				
	IL	RLM	MDC	RNA	IM
Paraíso	1,789	0,467	0,224	1,254	0,442
Unidade de Obras	3,863	0,247	0,009	0,848	0,051

Fonte: A autora.

Figura 11 - MRE para o primeiro ano hidrológico



Fonte: A autora.

Diferentemente do RMSE, o erro relativo médio obteve valores bem baixos para todos os métodos e estações utilizadas no preenchimento de falhas simuladas. Apesar de bem próximos os valores de MRE para os métodos de regressão linear múltipla, imputação múltipla e método das correlações, este último demonstrou melhor desempenho, com valores de erro abaixo de 0,25 para as estações Paraíso e Unidade de Obras. Já o método da interpolação linear não obteve o mesmo sucesso, ficando com o pior desempenho dentre os métodos para este primeiro ano da série hidrológica.

O segundo ano hidrológico com sete estações também mostrou bons resultados para o erro MRE, conforme a Tabela 21 e Figura 12.

Tabela 21 - Erro MRE para o segundo ano hidrológico

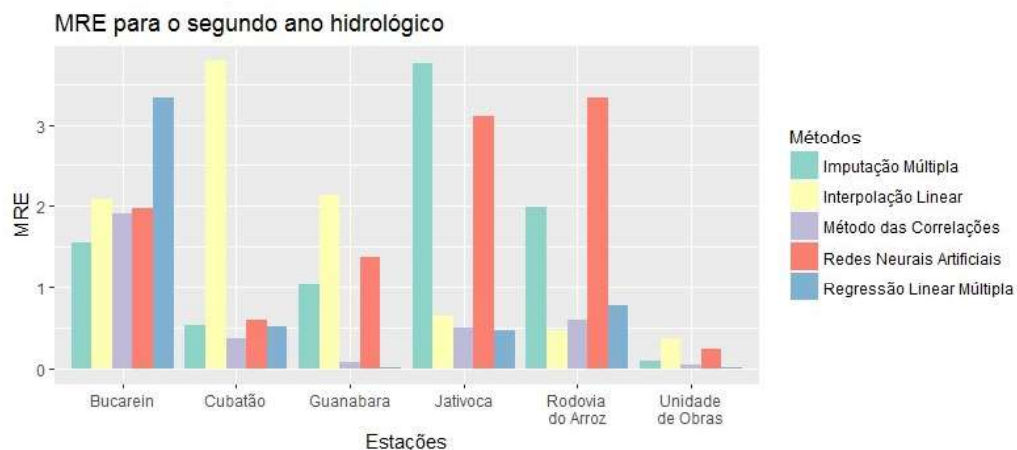
Estação	Métodos de preenchimento				
	IL	RLM	MDC	RNA	IM
Bucarein	2,086	3,338	1,904	1,969	1,551
Jativoca	0,640	0,459	0,500	3,107	3,751
Cubatão	3,785	0,515	0,375	0,592	0,539
Guanabara	2,123	0,016	0,085	1,364	1,028
Rodovia do Arroz	0,462	0,767	0,589	3,331	1,981
Unidade de Obras	0,374	0,010	0,048	0,243	0,090

Fonte: A autora.

A partir da Tabela 21 podemos aferir que a estação Unidade de Obras obteve os valores de MRE mais baixos para todos os métodos, caracterizando o melhor desempenho entre as estações. E com base na Figura 12 podemos perceber que todos os métodos e estações obtiveram valores de MRE abaixo de 4,00, que também indica bons resultados.

O método das correlações e regressão linear múltipla foram os que apresentaram valores mais próximos de zero e consequentemente, melhor desempenho quanto ao MRE referente ao segundo ano hidrológico da série.

Figura 12- MRE para o segundo ano hidrológico



Fonte: A autora.

O terceiro ano da série hidrológica contou com o maior número de estações utilizadas no preenchimento de falhas simuladas. E os valores de MRE obtidos estão apresentados na Tabela 22 e Figura 13.

Tabela 22- Erro MRE para o terceiro ano hidrológico

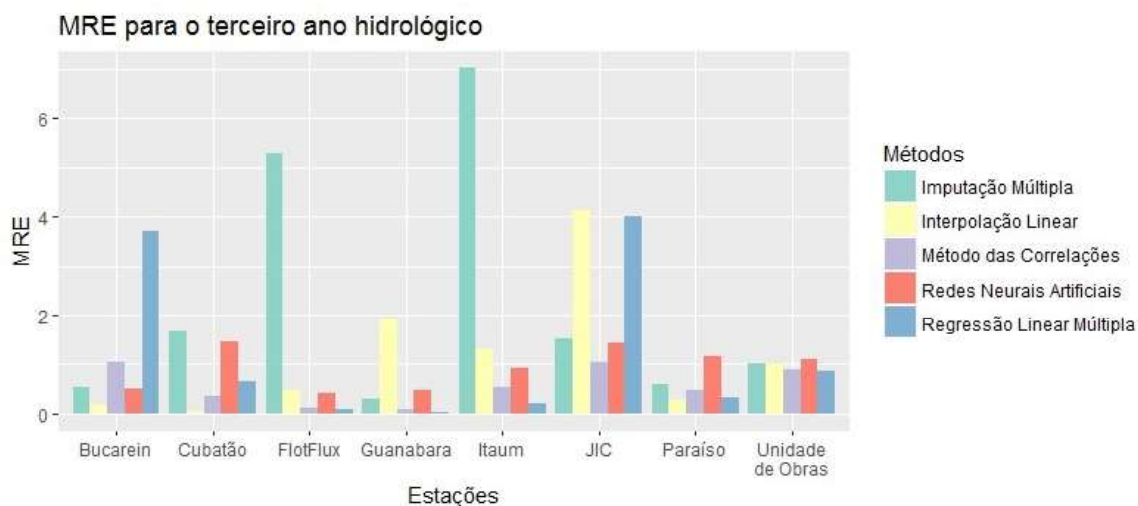
Estação	Métodos de preenchimento				
	IL	RLM	MDC	RNA	IM
Bucarein	0,165	3,693	1,057	0,509	0,524
JIC	4,117	4,005	1,038	1,444	1,526
Cubatão	0,065	0,647	0,370	1,476	1,666
Guanabara	1,913	0,018	0,085	0,467	0,298
FlotFlux	0,479	0,097	0,097	0,422	5,296
Unidade de Obras	1,000	0,855	0,889	1,106	1,000
Itaum	1,313	0,196	0,537	0,937	7,014
Paraíso	0,273	0,334	0,468	1,165	0,596

Fonte: A autora.

A partir da Figura 13, pode-se verificar que o método das correlações manteve seus valores de MRE próximos à unidade em todas as estações, caracterizando um ótimo resultado. As redes neurais artificiais também demonstraram um bom desempenho com

os valores de MRE abaixo de 1,5. Na imputação múltipla, com exceção dos valores altos nas estações FlotFlux e Itaum, os resultados foram bons.

Figura 13- MRE para o terceiro ano hidrológico



Fonte: A autora.

Já o último ano da série hidrológica tem seus resultados apresentados na Tabela 23 e Figura 14.

Tabela 23 - Erro MRE para o quarto ano hidrológico

Estação	Métodos de preenchimento				
	IL	RLM	MDC	RNA	IM
Bucarein	9,994	0,254	0,621	1,297	0,881
Cubatão	0,877	2,234	0,479	0,036	0,280
Guanabara	6,079	0,403	0,129	0,302	3,184
FlotFlux	1,000	0,099	0,128	0,454	0,057
Unidade de Obras	0,653	0,051	0,035	0,096	0,436
Itaum	0,806	0,118	0,110	0,521	0,756
Paraíso	0,894	2,234	1,282	0,846	1,280

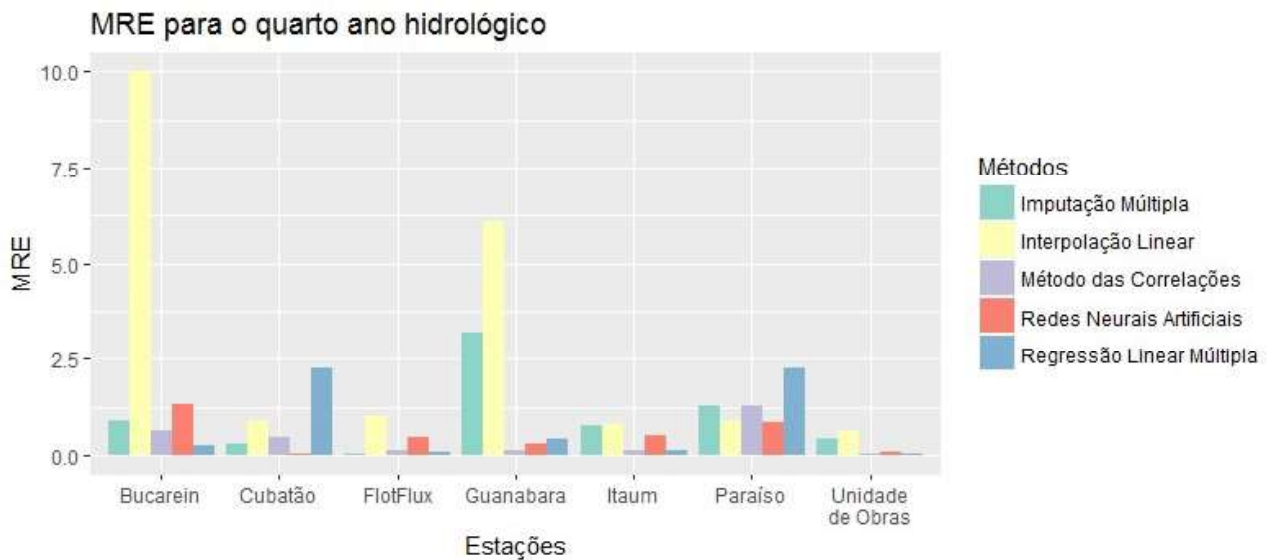
Fonte: A autora.

Da Tabela 23 podemos perceber dois valores de MRE bem acima do intervalo de valores que se concentra entre 0 e 3,0 para os demais métodos. Estes valores foram

obtidos com o método de interpolação linear nas estações Bucarein e Guanabara, indicando o pior desempenho dentre os métodos.

Na Figura 14 evidencia-se que o método de interpolação linear não é o mais adequado para preenchimento de falhas se comparado o erro MRE com os outros métodos aplicados neste quarto ano da série. Ainda na Figura 14 podemos verificar que o método das correlações obteve resultados de MRE relativamente melhores que as redes neurais artificiais, indicando o melhor desempenho para este tipo de erro no último ano da série hidrológica.

Figura 14- MRE para o quarto ano hidrológico



Fonte: A autora.

#### 4.3.3 Erro Médio Absoluto - MAE

O erro médio absoluto (MAE) também foi calculado conforme descrito no item 2.5 e os resultados para o primeiro ano hidrológico estão expostos na Tabela 14 e Figura 15.



Tabela 24- Erro MAE para o primeiro ano hidrológico, unidade em mm

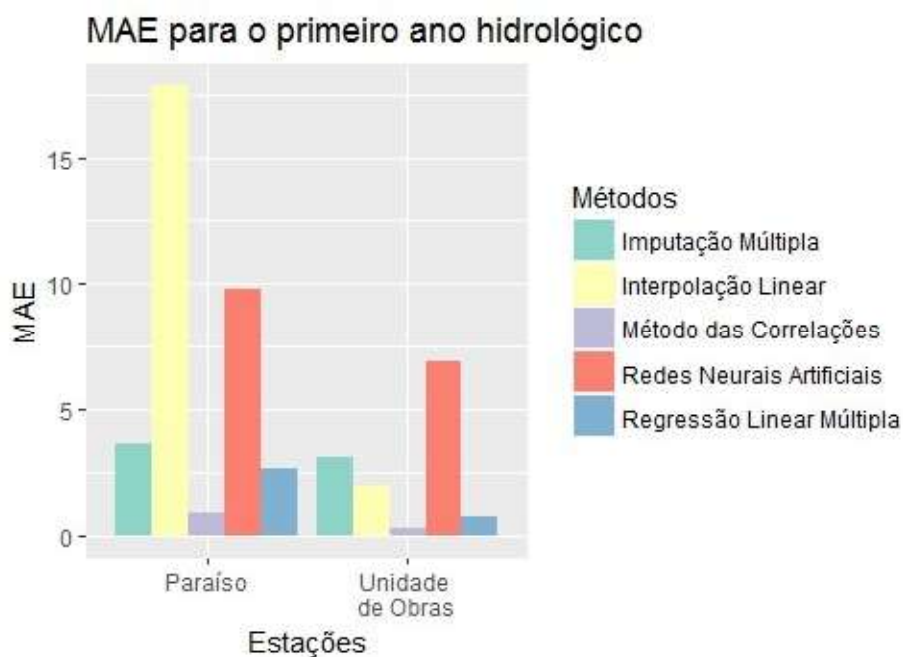
Estação	Métodos de preenchimento				
	IL	RLM	MDC	RNA	IM
Paraíso	17,832	2,644	0,894	9,770	3,607
Unidade de Obras	1,983	0,724	0,296	6,900	3,091

Fonte: A autora.

Da Tabela 24 verifica-se que o método de interpolação linear na estação Paraíso obteve o pior resultado, porém, na estação Unidade de Obras o valor de MAE foi aceitável. Apesar de ser um dos métodos mais sofisticados utilizados nesta pesquisa, as redes neurais artificiais não atingiram o desempenho esperado, podendo ser comparados ao método de interpolação linear.

A Figura 15 reforça o mau desempenho dos métodos de interpolação linear e redes neurais, e o bom desempenho do método das correlações neste primeiro ano da série hidrológica.

Figura 15 - MAE para o primeiro ano hidrológico



Fonte: A autora.

Para o segundo ano hidrológico, os valores de MAE ficaram concentrados entre 0 e 10 mm, de acordo com a Tabela 25.

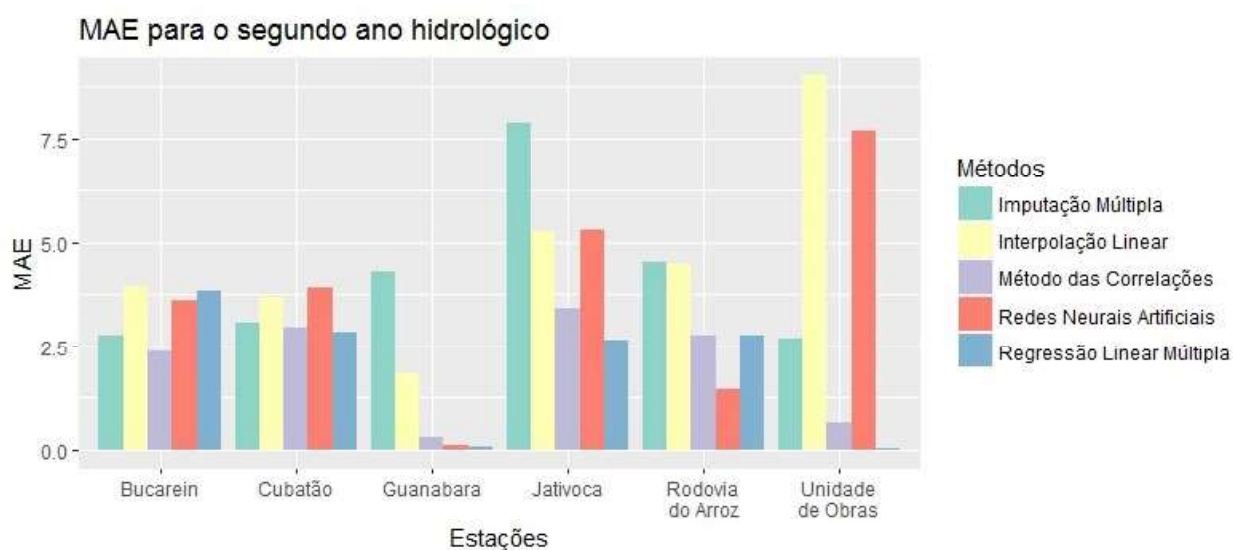
Tabela 25 - Erro MAE para o segundo ano hidrológico

Estação	Métodos de preenchimento				
	IL (mm)	RLM (mm)	MDC (mm)	RNA (mm)	IM (mm)
Bucarein	3,939	3,842	2,390	3,603	2,758
Jativoca	5,263	2,639	3,422	5,307	7,893
Cubatão	3,678	2,857	2,964	3,915	3,091
Guanabara	1,860	0,060	0,313	0,123	4,320
Rodovia do Arroz	4,497	2,760	2,780	1,460	4,525
Unidade de Obras	9,020	0,035	0,656	7,674	2,705

Fonte: A autora.

A Figura 16 representa graficamente os valores da Tabela 25 e a partir dela podemos comprovar mais uma vez a boa performance do método das correlações e regressão linear múltipla.

Figura 16 - MAE para o segundo ano hidrológico



Fonte: A autora.

O terceiro ano da série hidrológica também obteve alguns resultados acima da faixa de valores concentrada entre 0 e 7,5 mm para o erro MAE conforme mostra a Tabela 26.

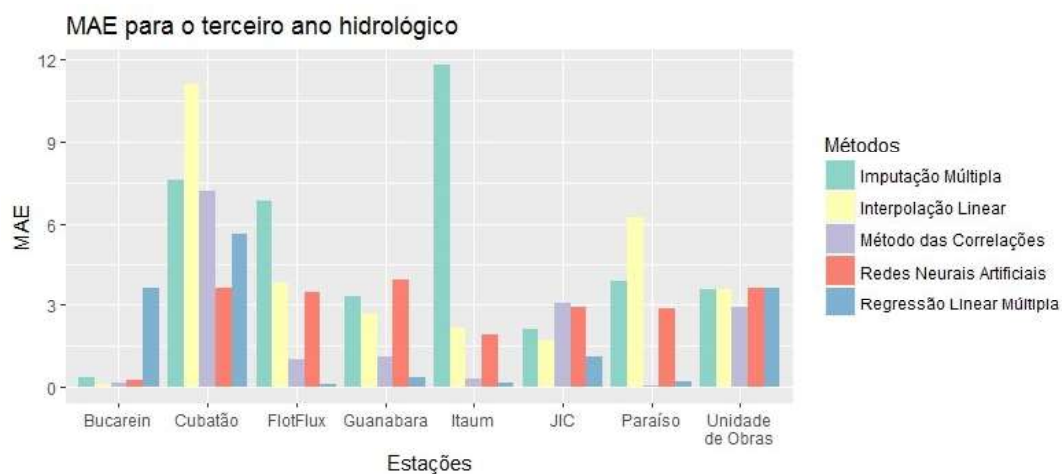
Tabela 26 - Erro MAE para o terceiro ano hidrológico

Estação	Métodos de preenchimento				
	IL (mm)	RLM (mm)	MDC (mm)	RNA (mm)	IM (mm)
Bucarein	0,118	3,689	0,154	0,248	0,370
JIC	1,712	1,099	3,060	2,912	2,121
Cubatão	11,119	5,619	7,204	3,662	7,590
Guanabara	2,638	0,367	1,115	3,957	3,367
FlotFlux	3,875	0,100	0,100	3,497	6,815
Unidade de Obras	3,624	3,689	2,901	3,656	3,624
Itaum	2,135	0,158	0,278	1,900	11,807
Paraíso	6,250	0,182	0,063	2,850	3,935

Fonte: A autora.

Na Figura 17 os dados indicam uma proximidade entre os valores de MAE obtidos com o método da regressão linear múltipla e método das correlações, que de novo são classificados como os métodos mais adequados para preenchimento de falhas deste terceiro ano hidrológico se considerado o erro médio absoluto.

Figura 17 - MAE para o terceiro ano hidrológico



Fonte: A autora.

A Tabela 27 exibe os valores obtidos para o erro MAE no último ano da série hidrológica utilizada nesta pesquisa.

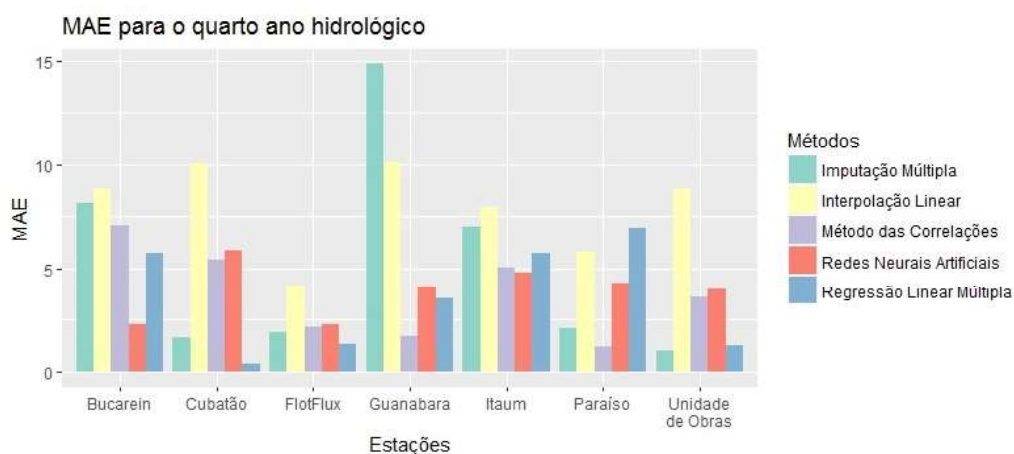
Tabela 27 - Erro MAE para o quarto ano hidrológico

Estação	Métodos de preenchimento				
	IL (mm)	RLM (mm)	MDC (mm)	RNA (mm)	IM (mm)
Bucarein	8,837	5,742	7,098	2,291	8,138
Cubatão	10,035	0,398	5,406	5,873	1,664
Guanabara	10,104	3,563	1,673	4,127	14,832
FlotFlux	4,191	1,334	2,147	2,300	1,905
Unidade de Obras	8,829	1,247	3,604	4,004	1,013
Itaum	7,945	5,779	5,034	4,771	6,987
Paraíso	5,798	6,919	1,179	4,288	2,056

Fonte: A autora.

Com base na Figura 18 é possível verificar que o método de imputação múltipla apresenta o maior valor de MAE obtido no quarto ano hidrológico, que por sua vez se encontra na estação Guanabara. E enquanto o método de interpolação linear encontra-se acima da faixa concentrada de valores de MAE, os métodos de regressão linear múltipla, redes neurais artificiais e método das correlações apresentam os melhores resultados deste quarto ano hidrológico da série.

Figura 18 - MAE para o quarto ano hidrológico



Fonte: A autora.

Bárdossy e Pegram (2014) compararam um novo método baseado em cópula espacial com outras técnicas para preencher dados de precipitação diária observados por 13 estações durante o período de 32 anos em Cabo do Sul na África do Sul. Os métodos foram avaliados com base no MAE, RMSE e coeficiente de correlação de Pearson. O RMSE apresentou valores abaixo de 5 mm quando utilizada toda a série hidrológica, mas atingiu valores próximos a 30 mm quando foram utilizados apenas 8 anos da série. O mesmo aconteceu com os outros erros citados, que aumentaram conforme a retirada de anos da série hidrológica. O trabalho destes autores se aproxima desta pesquisa justamente por simular uma série hidrológica pequena, de 8 anos, o que acontece de fato neste trabalho contendo dados apenas de 4 anos hidrológicos, e por isso justificam-se os valores altos de RMSE.

#### 4.4 COEFICIENTE DE DETERMINAÇÃO - $R^2$

O coeficiente de determinação ( $R^2$ ) é mais uma das medidas utilizadas para avaliar o desempenho dos métodos no preenchimento de falhas de dados de 24 horas de precipitação. Os valores desse coeficiente para o primeiro ano da série hidrológica podem ser observados na Tabela 28.

Tabela 28 - Coeficiente de determinação para o primeiro ano hidrológico

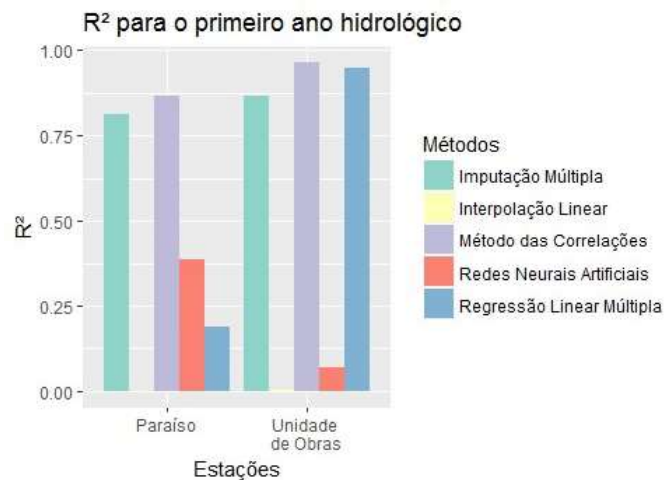
Estação	Métodos de preenchimento				
	IL	RLM	MDC	RNA	IM
Paraíso	0,001	0,189	0,864	0,384	0,811
Unidade de Obras	0,002	0,948	0,963	0,069	0,867

Fonte: A autora.

Ao analisar os dados da Tabela 28, podemos perceber que os métodos de interpolação linear e redes neurais artificiais não alcançaram valores significativos de  $R^2$ , ficando abaixo de 40%. Entretanto, o método da imputação múltipla e método das correlações, juntamente com a regressão linear múltipla para somente a estação Unidade de Obras, atingiram valores de  $R^2$  acima de 80%, o que indica um ótimo desempenho destes métodos e também correlação entre os valores observados e simulados.

A interpolação linear, representada pela coluna amarela na Figura 19, obteve valores de  $R^2$  bem próximos a zero. Já o método das correlações, seguido pela imputação múltipla, alcançaram os melhores resultados para o coeficiente de determinação neste primeiro ano da série hidrológica.

Figura 19 -  $R^2$  para o primeiro ano hidrológico



Fonte: A autora.

O segundo ano hidrológico tem seus coeficientes de determinação expostos na Tabela 29.

Tabela 29 - Coeficiente de determinação para o segundo ano hidrológico

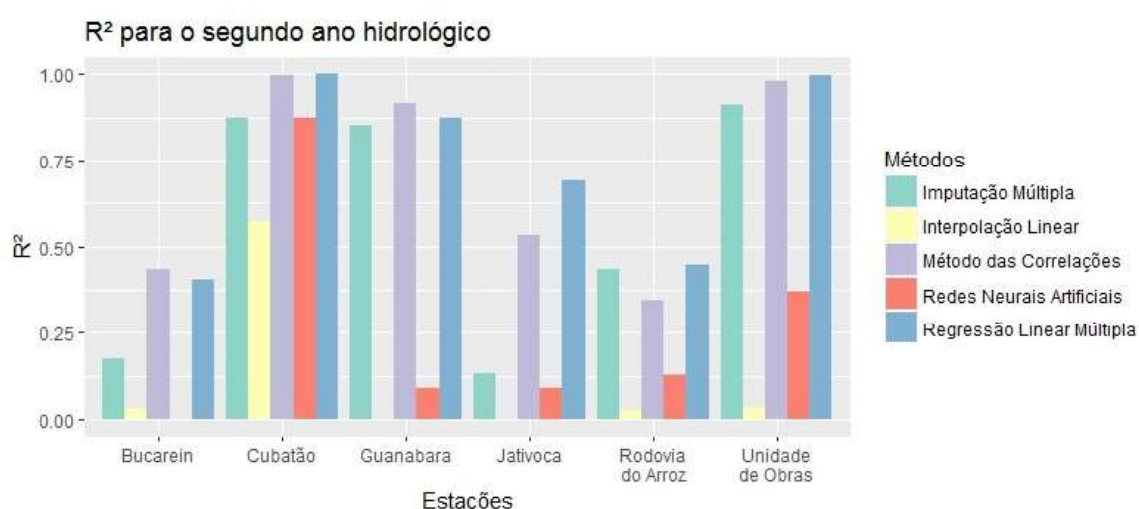
Estação	Métodos de preenchimento				
	IL	RLM	MDC	RNA	IM
Bucarein	0,029	0,406	0,433	0,001	0,173
Jativoca	0,000	0,689	0,533	0,088	0,132
Cubatão	0,570	1,000	0,999	0,874	0,874
Guanabara	0,000	0,872	0,918	0,090	0,852
Rodovia do Arroz	0,027	0,449	0,345	0,126	0,436
Unidade de Obras	0,033	0,999	0,982	0,372	0,913

Fonte: A autora.

Com base na Tabela 29 é possível perceber que todos os métodos de preenchimento englobam valores de  $R^2$  bastante diferentes, valores estes que estão entre 0% e 100%. Por exemplo, o método de interpolação linear tem a estação Guanabara com  $R^2=0$  e a estação Cubatão com  $R^2=0,57$ , como o menor e maior valor respectivamente. Bem como o método de redes neurais artificiais com  $R^2=0,001$  até  $R^2=0,874$ . Se analisarmos por estação, a que obteve os melhores valores de coeficiente de determinação foi a estação Cubatão, mantendo valores acima de 50% em todos os métodos aplicados.

Na Figura 20 é reafirmada a boa performance da estação Cubatão, e também fica claro que o método das correlações, bem como a regressão linear múltipla, alcançaram o melhor desempenho quanto ao coeficiente de determinação neste segundo ano da série hidrológica.

Figura 20 -  $R^2$  para o segundo ano hidrológico



Fonte: A autora.

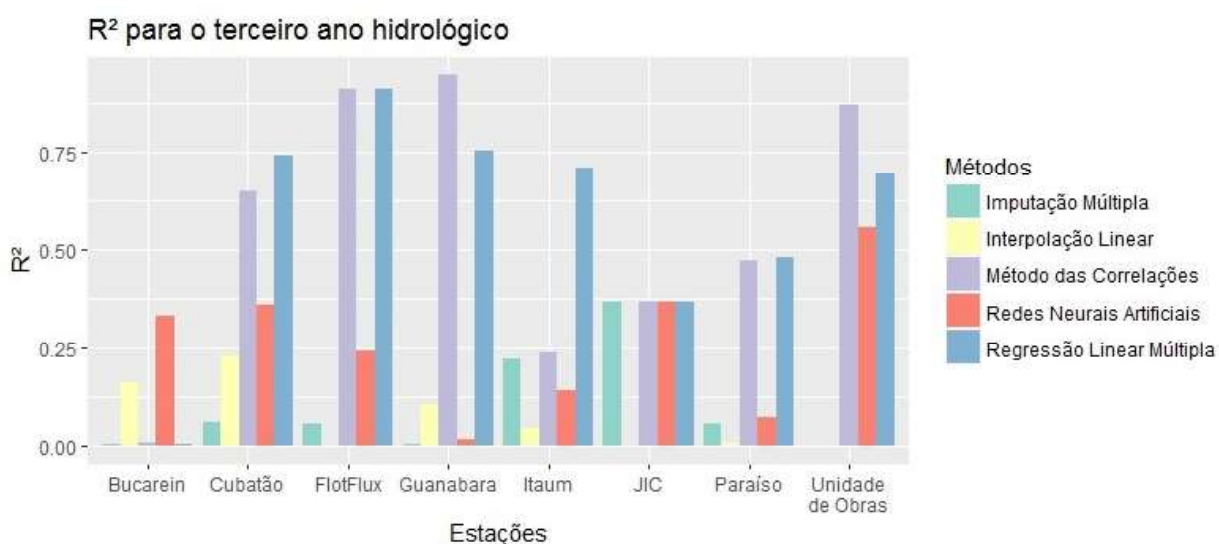
Na Tabela 30 constam os valores do coeficiente de determinação para o terceiro ano da série hidrológica, na qual podemos notar que para cada método de preenchimento aplicado, existe no mínimo um valor de  $R^2$  bem próximo a zero.

Tabela 30 - Coeficiente de determinação para o terceiro ano hidrológico

Estação	Métodos de preenchimento				
	IL	RLM	MDC	RNA	IM
Bucarein	0,163	0,003	0,006	0,329	0,002
JIC	0,001	0,369	0,369	0,369	0,369
Cubatão	0,231	0,743	0,650	0,360	0,059
Guanabara	0,104	0,755	0,948	0,016	0,003
FlotFlux	0,000	0,912	0,912	0,244	0,056
Unidade de Obras	0,000	0,693	0,872	0,556	0,000
Itaum	0,044	0,712	0,239	0,142	0,220
Paraíso	0,009	0,480	0,472	0,071	0,055

Fonte: A autora.

A representação gráfica da Figura 21 facilita o entendimento da magnitude do coeficiente de determinação, com base nela podemos afirmar que o método das correlações e regressão linear múltipla alcançaram valores próximos ou acima de 70% para as estações Cubatão, Guanabara, FlotFlux e Unidade de Obras, o que indica um bom desempenho referente à estatística  $R^2$ .

Figura 21 -  $R^2$  para o terceiro ano hidrológico

Fonte: A autora.



A Tabela 31 além de demonstrar os resultados do coeficiente de determinação para todos os métodos de preenchimento aplicados neste quarto e último ano da série hidrológica, revela também valores de  $R^2$  muito próximos a zero para o método de interpolação linear em sua totalidade.

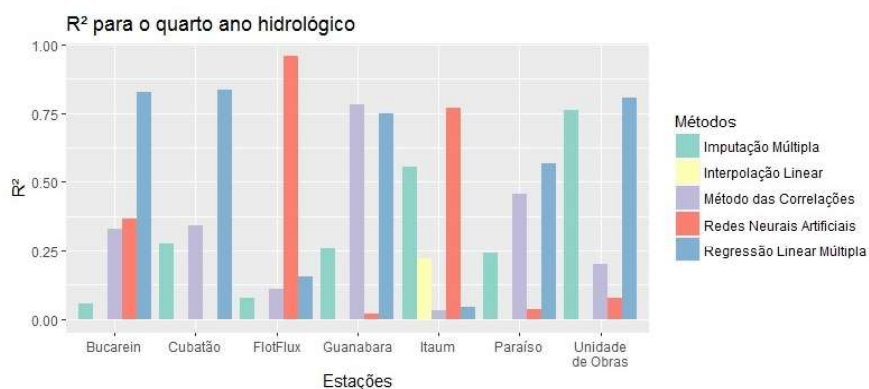
Tabela 31 - Coeficiente de determinação para o quarto ano hidrológico

Estação	Métodos de preenchimento				
	IL	RLM	MDC	RNA	IM
Bucarein	0,001	0,825	0,329	0,368	0,057
Cubatão	0,000	0,834	0,342	0,001	0,278
Guanabara	0,000	0,748	0,782	0,018	0,257
FlotFlux	0,000	0,156	0,111	0,956	0,077
Unidade de Obras	0,000	0,805	0,201	0,077	0,761
Itaum	0,220	0,046	0,031	0,768	0,555
Paraíso	0,000	0,567	0,456	0,035	0,241

Fonte: A autora.

Da Figura 22 podemos salientar o desempenho discreto do método de interpolação linear, representado pela barra amarela, no quarto ano hidrológico. No entanto, o método da regressão linear múltipla obteve valores de  $R^2$  acima de 70%, nas estações Bucarein, Cubatão, Guanabara e Unidade de Obras, o que indica um bom desempenho para estas quatro estações, sendo menor nas demais.

Figura 22 -  $R^2$  para o quarto ano hidrológico



Fonte: A autora.

No trabalho de Teegavarapu e Chandramouli (2005) foi aplicado o método tradicional de ponderação do inverso da distância (IDWM) e alguns métodos que surgiram das revisões conceituais deste mesmo método em conjunto com o método de estimativa das redes neurais artificiais, todos eles com o intuito de preencher dados diários de precipitação de 20 estações pluviométricas do estado de Kentucky nos Estados Unidos. Os autores analisaram os resultados com base no RMSE, MAE, MRE,  $R^2$  e encontraram valores entre 0,225 e 1,906 mm, 0,010 e 0,612 mm, 1,912 e 4,608 mm, 0,378 e 0,759 respectivamente.

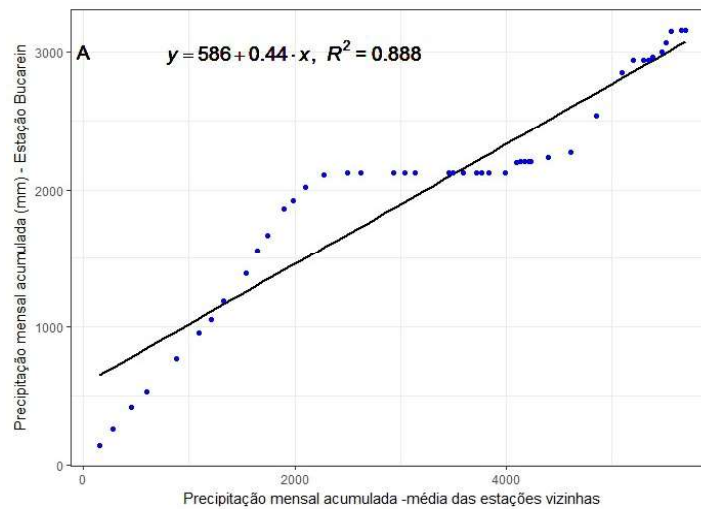
Neste mesmo estado de Kentucky (EUA) os autores Teegavarapu *et al* (2009), agora utilizando 15 estações e os métodos de IDWM, CCWM e fórmulas funcionais ótimas para estimar dados diários de precipitação, alcançaram valores de RMSE e MAE muito bons, próximos a zero, bem como coeficientes de correlação próximos a unidade. Porém, para os dados apresentados nesta pesquisa, os bons valores dos coeficientes de determinação nem sempre vinham acompanhados por valores bons de RMSE e MAE.

Seguindo o viés da interpolação para o preenchimento de falhas de dados de precipitação, os autores Eischeid *et al* (1995 e 2000) utilizaram 22 estações, uma série hidrológica de 40 anos e seis técnicas de interpolação com estações meteorológicas próximas, dentre elas: relação normal, inverso da distância, interpolação ideal, regressão múltipla, melhor estimador único e mediana. Estes autores relataram que o método da regressão múltipla atingiu a maior eficiência no preenchimento de falhas e posteriormente, constataram também, que este o mesmo método obteve os maiores coeficientes de correlação não só para as precipitações bem como para as temperaturas.

#### 4.5 ANÁLISE DE DUPLA MASSA

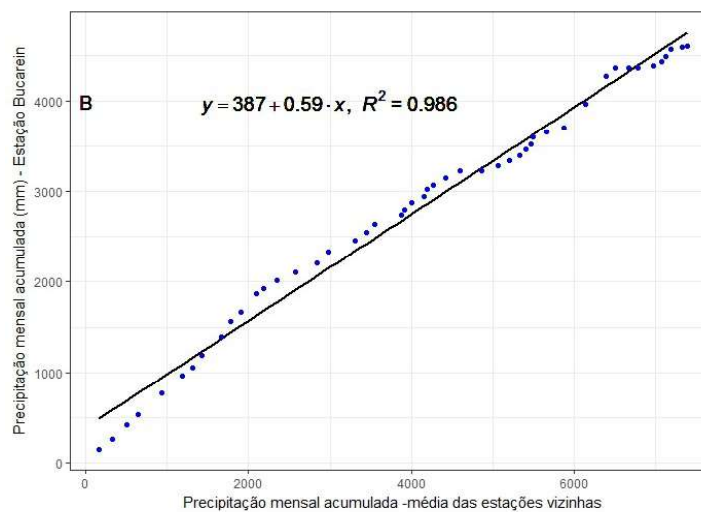
O conjunto de estações vizinhas utilizado para análise da consistência dos dados da estação Bucarein foi composto pelas estações Itaum, Unidade de Obras e FlotFlux. Os resultados obtidos na análise desta estação (a) sem preenchimento, (b) imputação múltipla, (c) interpolação linear, (d) método das correlações, (e) redes neurais artificiais, e (f) regressão linear múltipla, estão representados nas Figura 23, Figura 24, Figura 25, Figura 26, Figura 27 e Figura 28.

Figura 23 - Dupla Massa para Estação Bucarein sem preenchimento de Outubro de 2012 até Setembro de 2016



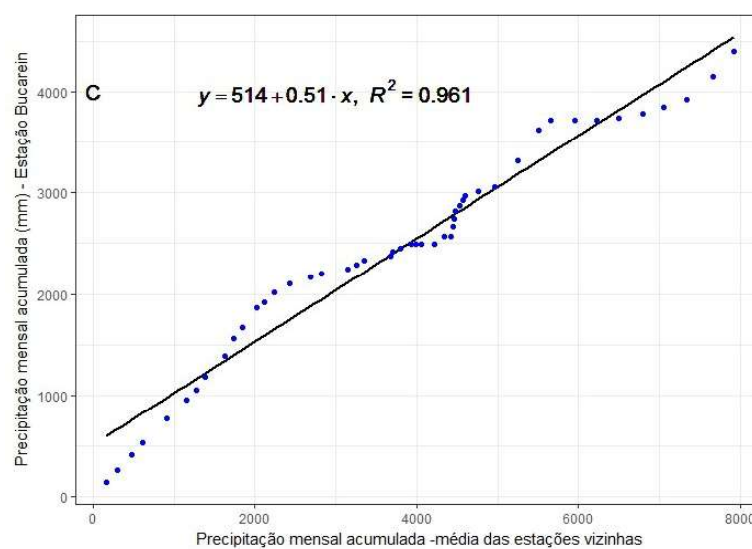
Fonte: A autora.

Figura 24 - Dupla Massa para Estação Bucarein preenchido por Imputação Múltipla de Outubro de 2012 até Setembro de 2016



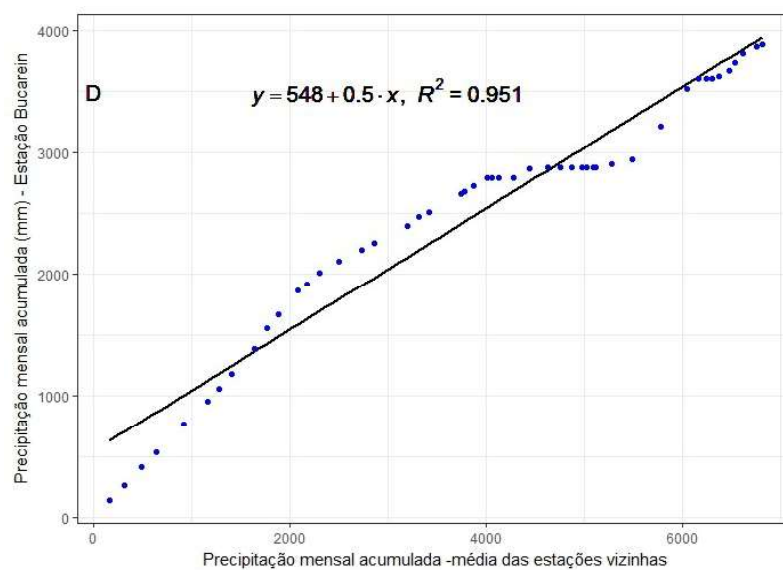
Fonte: A autora.

Figura 25 - Dupla Massa para Estação Bucarein preenchido por Interpolação Linear de Outubro de 2012 até Setembro de 2016



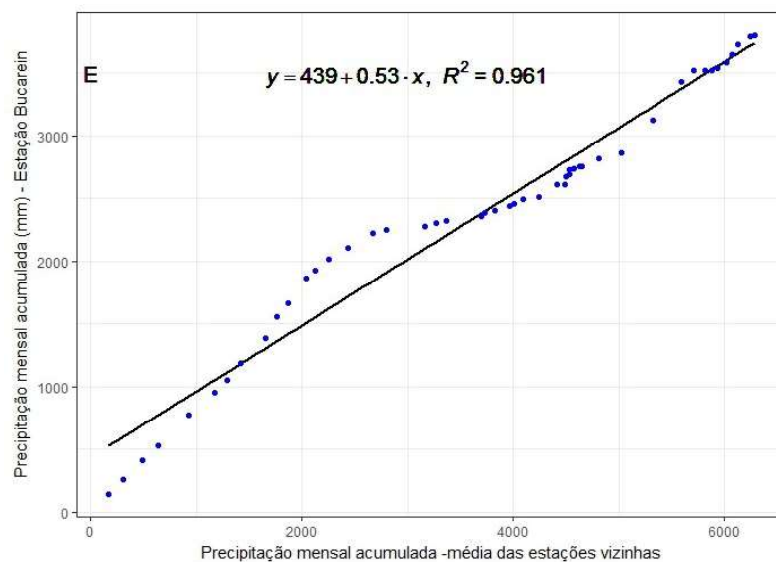
Fonte: A autora.

Figura 26 - Dupla Massa para Estação Bucarein preenchido pelo Método das Correlações de Outubro de 2012 até Setembro de 2016



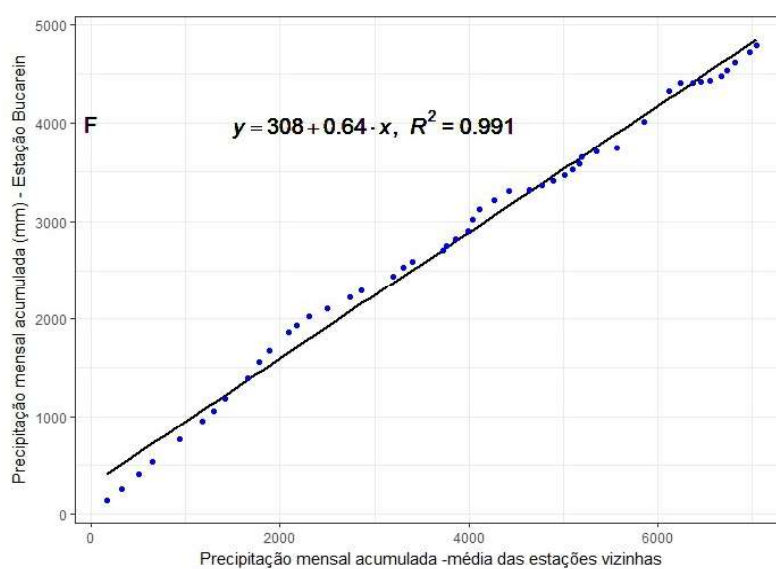
Fonte: A autora.

Figura 27 - Dupla Massa para Estação Bucarein preenchido por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

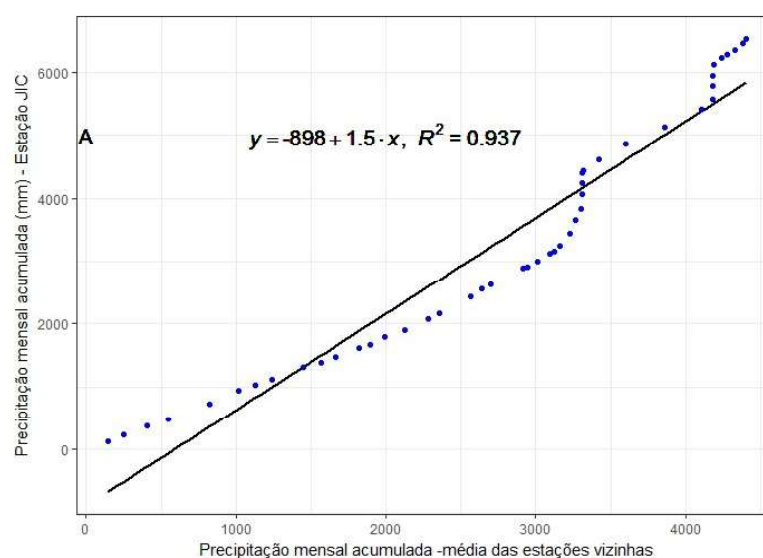
Figura 28 - Dupla Massa para Estação Bucarein preenchido por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

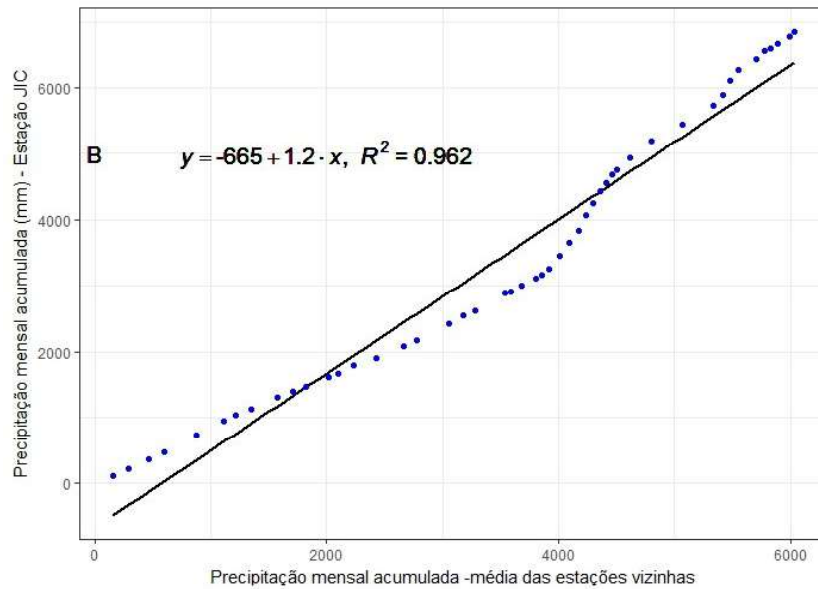
Para a estação JIC foram utilizadas as estações Unidade de Obras, Itaum e Bucarein como estações vizinhas, e os resultados da análise de consistência dos dados (a) sem preenchimento, (b) imputação múltipla, (c) interpolação linear, (d) método das correlações, (e) redes neurais artificiais, e (f) regressão linear múltipla, podem ser vistos nas Figura 29, Figura 30, Figura 31, Figura 32, Figura 33 e Figura 34.

Figura 29 - Dupla Massa para Estação JIC sem preenchimento de Outubro de 2012 até Setembro de 2016



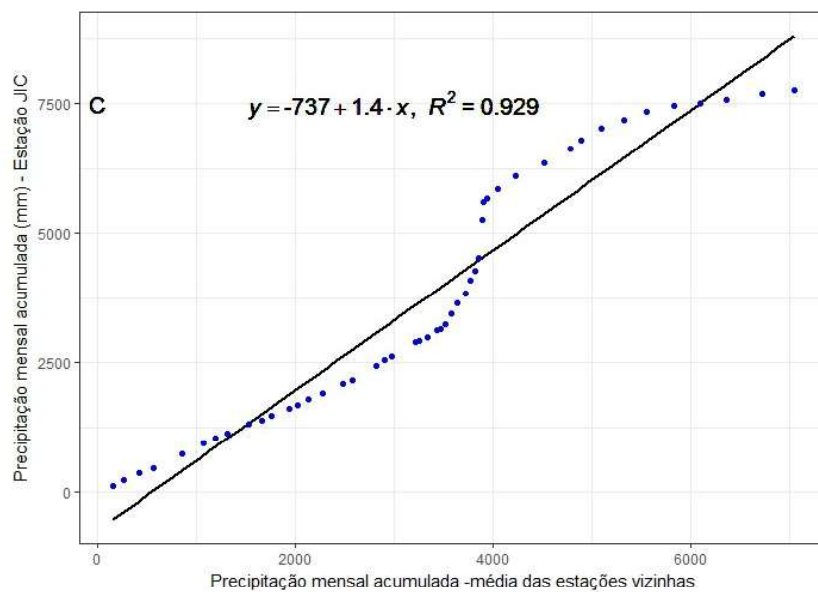
Fonte: A autora.

Figura 30 - Dupla Massa para Estação JIC preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016



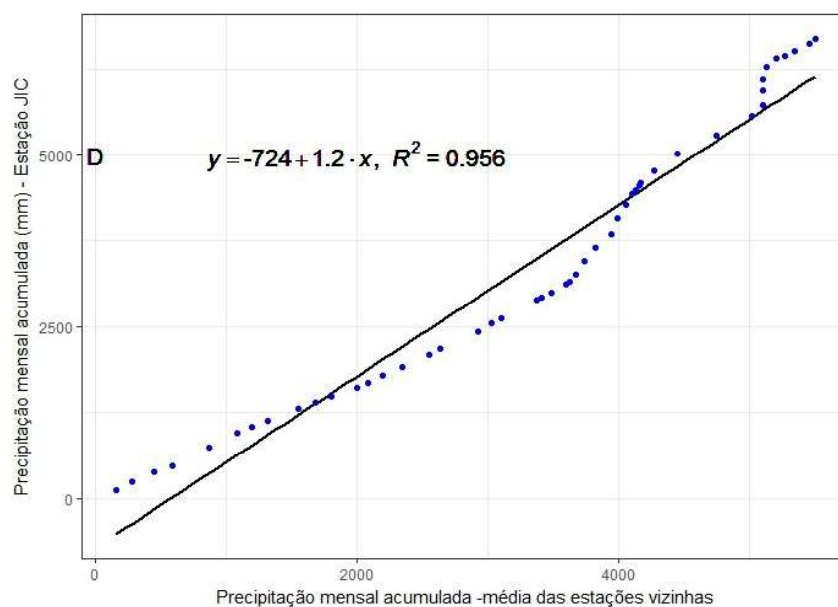
Fonte: A autora.

Figura 31 - Dupla Massa para Estação JIC preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016



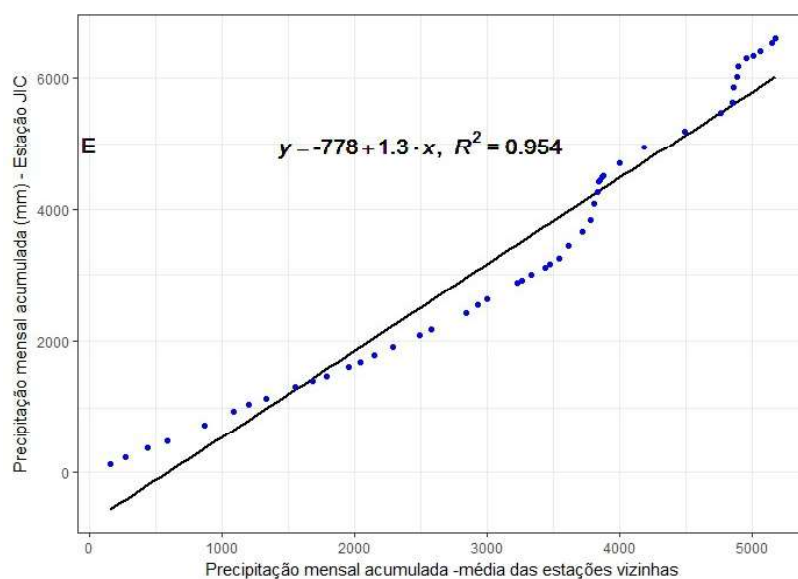
Fonte: A autora.

Figura 32 - Dupla Massa para Estação JIC preenchido pelo Método das Correlações de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

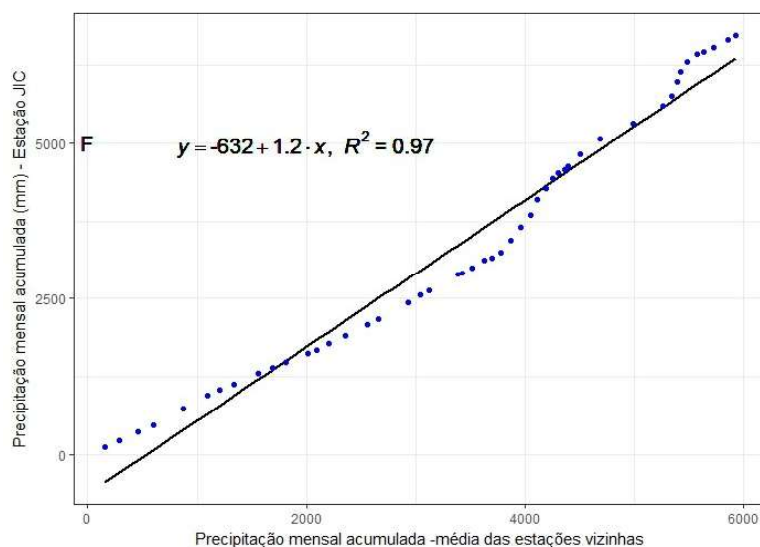
Figura 33 - Dupla Massa para Estação JIC preenchido por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016



Fonte: A autora.



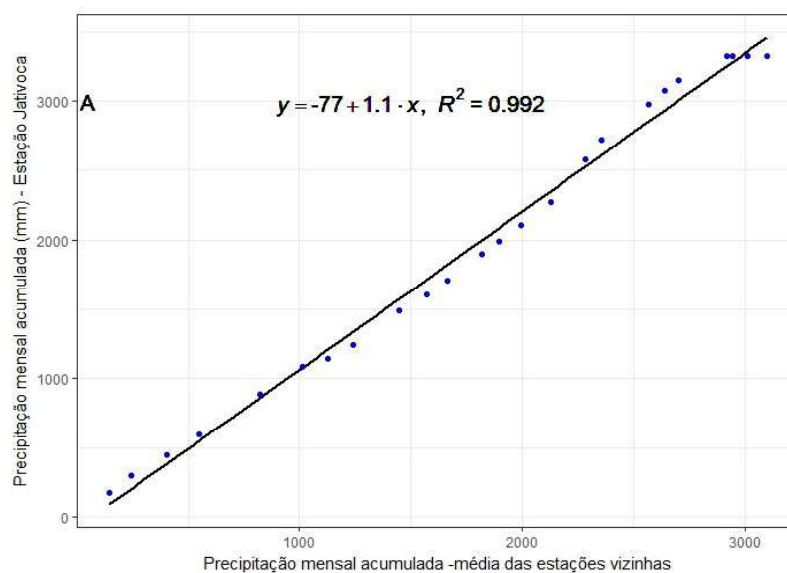
Figura 34 - Dupla Massa para Estação JIC preenchido por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

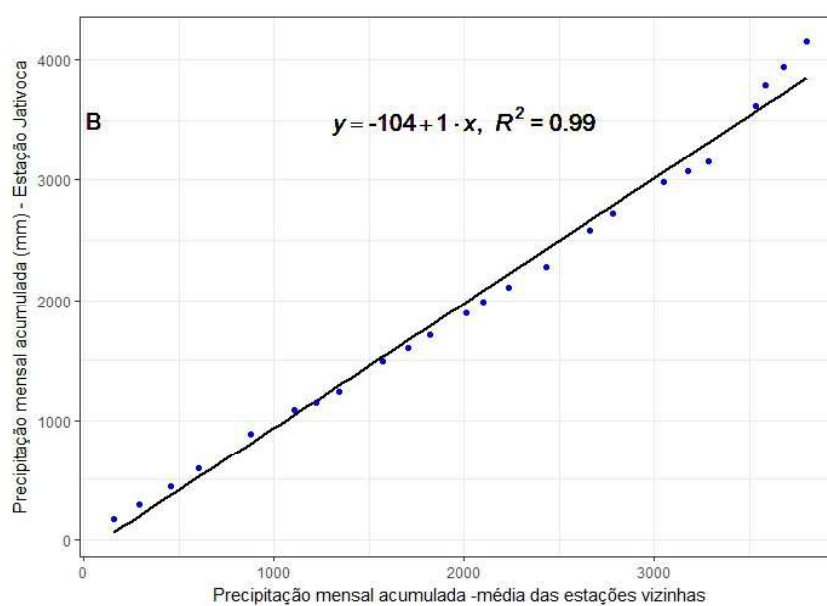
Na estação Jativoca também foram utilizadas as estações Unidade de Obras, Itaum e Bucarein como estações vizinhas, e os resultados da análise de consistência dos dados (a) sem preenchimento, (b) imputação múltipla, (c) interpolação linear, (d) método das correlações, (e) redes neurais artificiais, e (f) regressão linear múltipla, podem ser vistos nas Figura 35, Figura 36, Figura 37, Figura 38, Figura 39 e Figura 40.

Figura 35 - Dupla Massa para Estação Jativoca sem preenchimento de Outubro de 2012 até Setembro de 2014



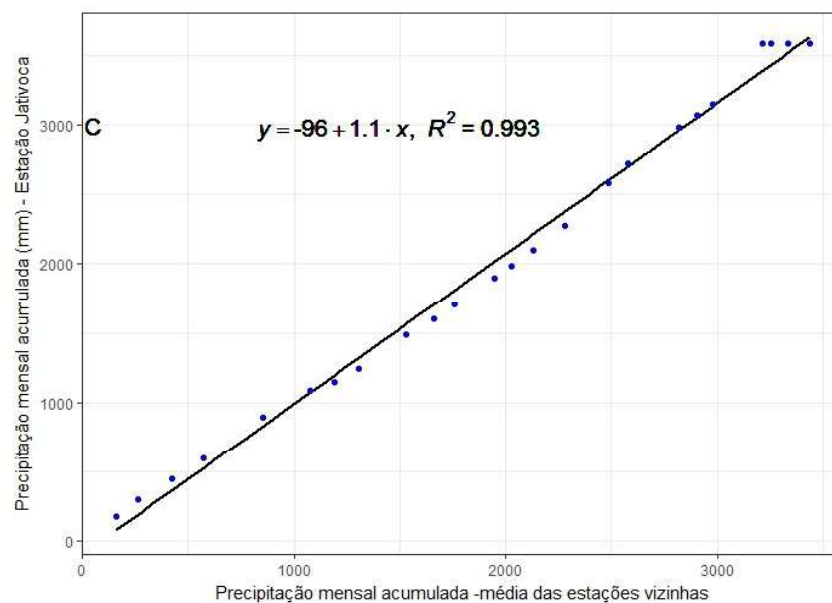
Fonte: A autora.

Figura 36 - Dupla Massa para Estação Jativoca preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2014



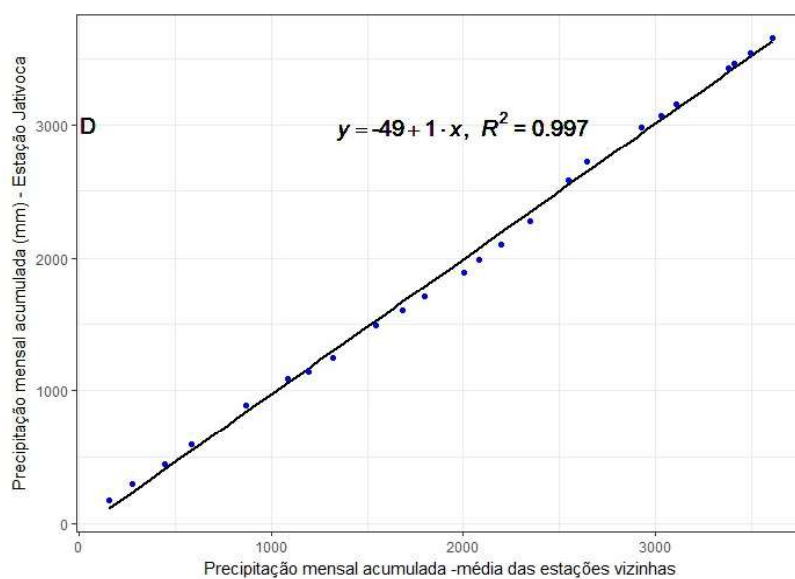
Fonte: A autora.

Figura 37 - Dupla Massa para Estação Jativoca preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2014



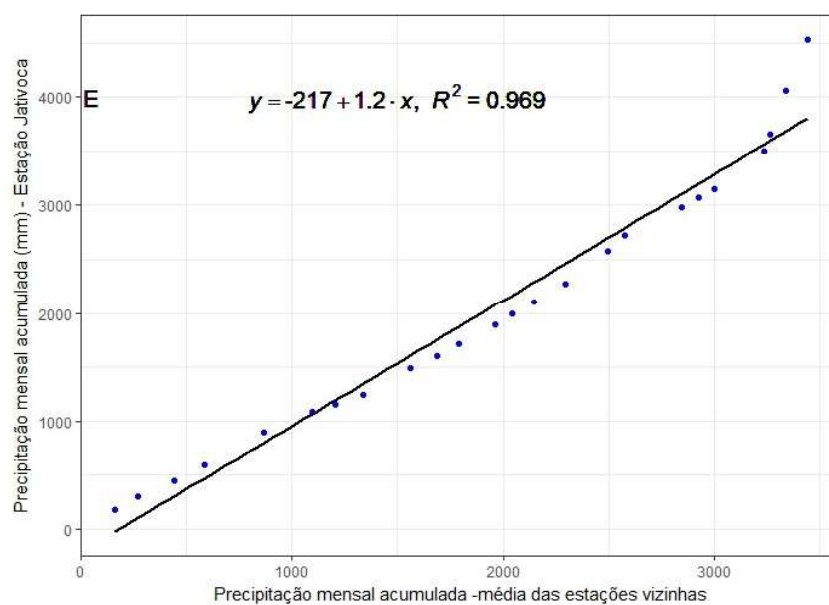
Fonte: A autora.

Figura 38 - Dupla Massa para Estação Jativoca preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2014



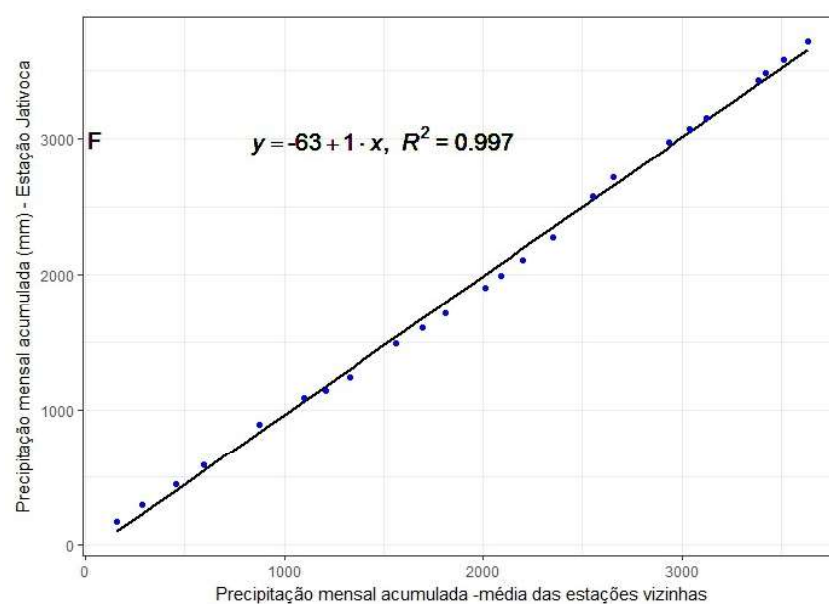
Fonte: A autora.

Figura 39 - Dupla Massa para Estação Jativoca preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2014



Fonte: A autora.

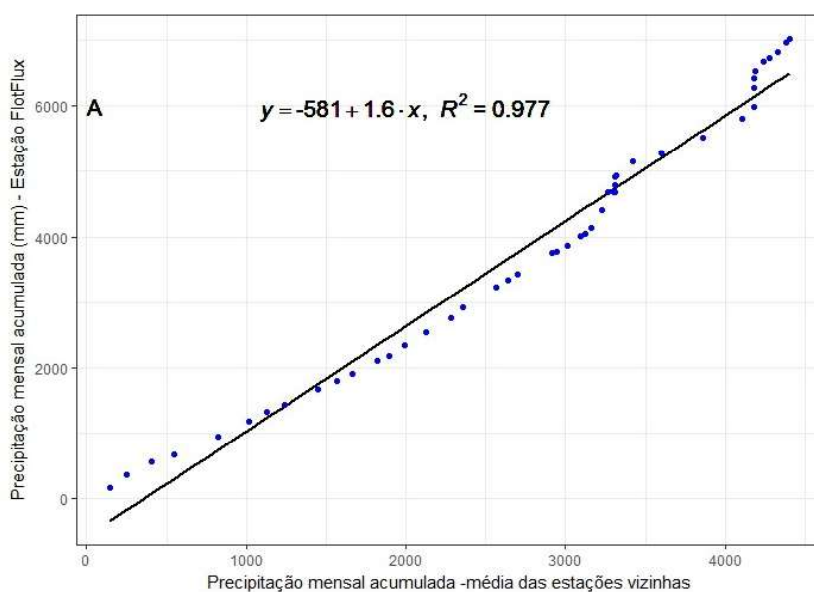
Figura 40 - Dupla Massa para Estação Jativoca preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2014



Fonte: A autora.

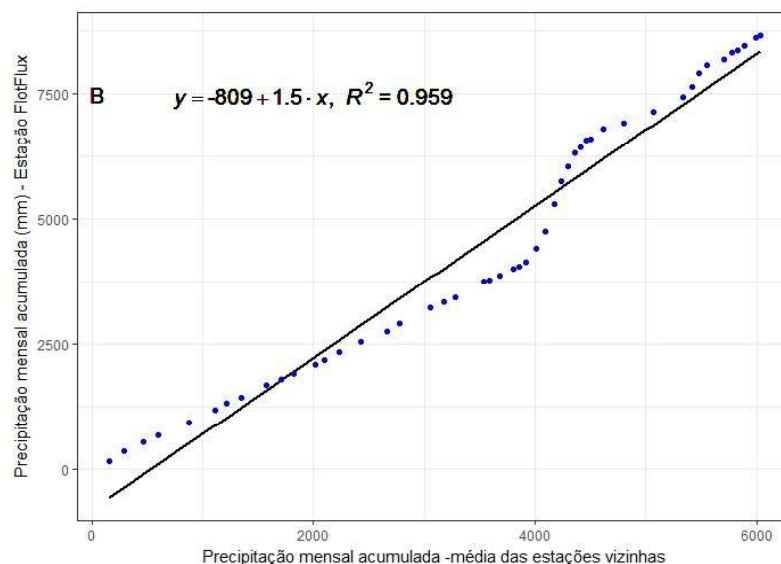
Na estação FlotFlux novamente foram utilizadas as estações Unidade de Obras, Itaum e Bucarein como estações vizinhas, e os resultados da análise de consistência dos dados (a) sem preenchimento, (b) imputação múltipla, (c) interpolação linear, (d) método das correlações, (e) redes neurais artificiais, e (f) regressão linear múltipla, podem ser vistos nas Figura 41, Figura 42, Figura 43, Figura 44, Figura 45 e Figura 46.

Figura 41 - Dupla Massa para Estação FlotFlux sem preenchimento de Outubro de 2012 até Setembro de 2016



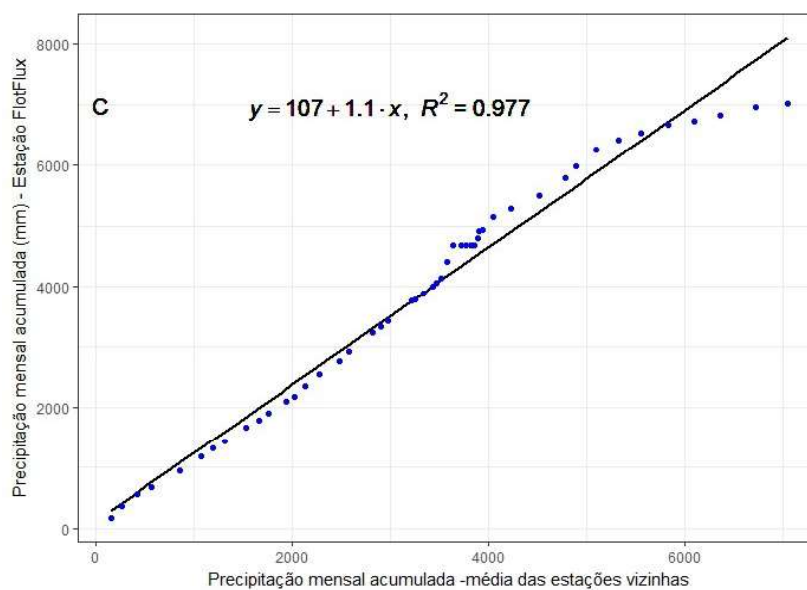
Fonte: A autora.

Figura 42 - Dupla Massa para Estação FlotFlux preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016



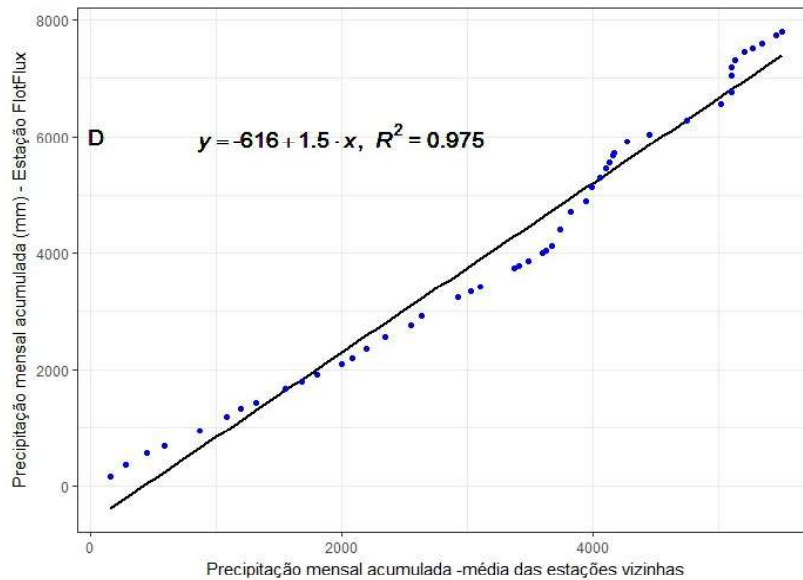
Fonte: A autora.

Figura 43 - Dupla Massa para Estação FlotFlux preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016



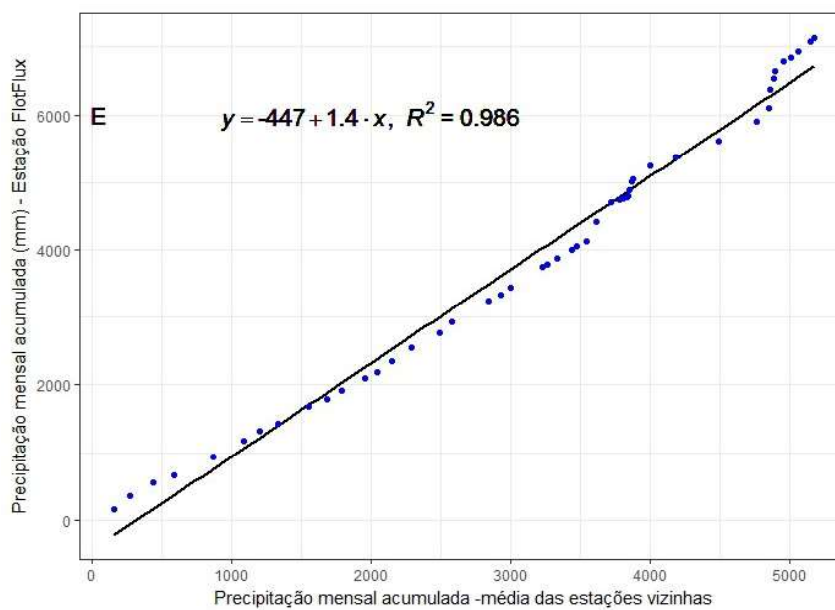
Fonte: A autora.

Figura 44 - Dupla Massa para Estação FlotFlux preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016



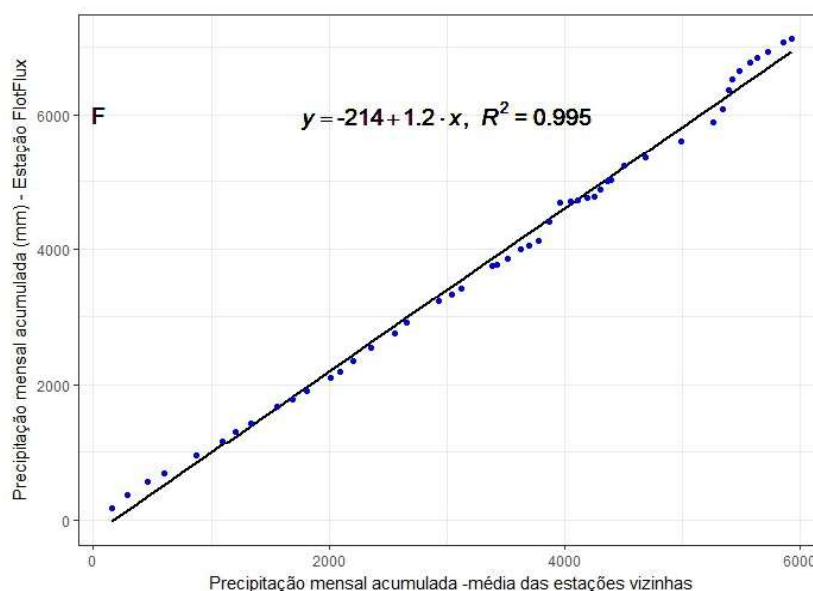
Fonte: A autora.

Figura 45 - Dupla Massa para Estação FlotFlux preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

Figura 46 - Dupla Massa para Estação FlotFlux preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016

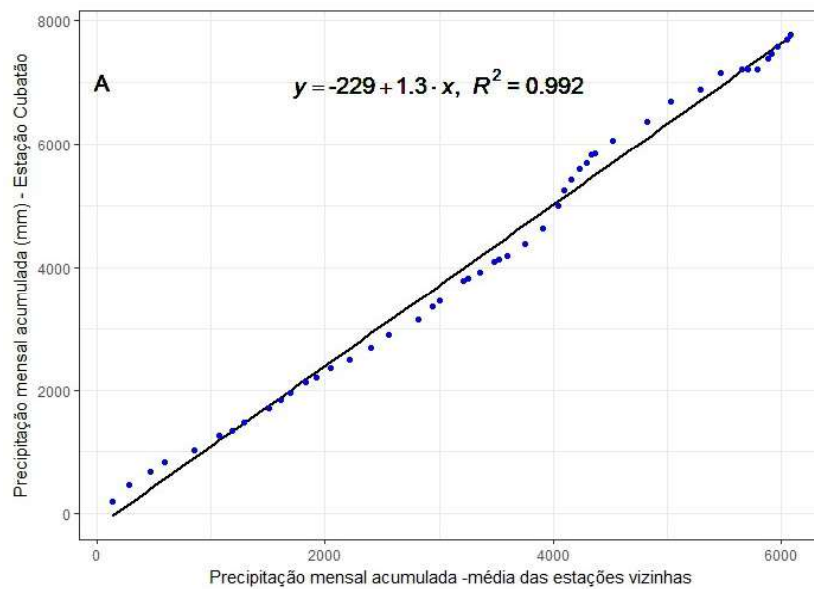


Fonte: A autora.

Na estação Cubatão, foram utilizadas as estações vizinhas Paraíso, FlotFlux e Unidade de Obras, e os resultados da análise de consistência dos dados (a) sem preenchimento, (b) imputação múltipla, (c) interpolação linear, (d) método das correlações, (e) redes neurais artificiais, e (f) regressão linear múltipla, podem ser vistos nas Figura 47, Figura 48, Figura 49, Figura 50, Figura 51 e Figura 52.

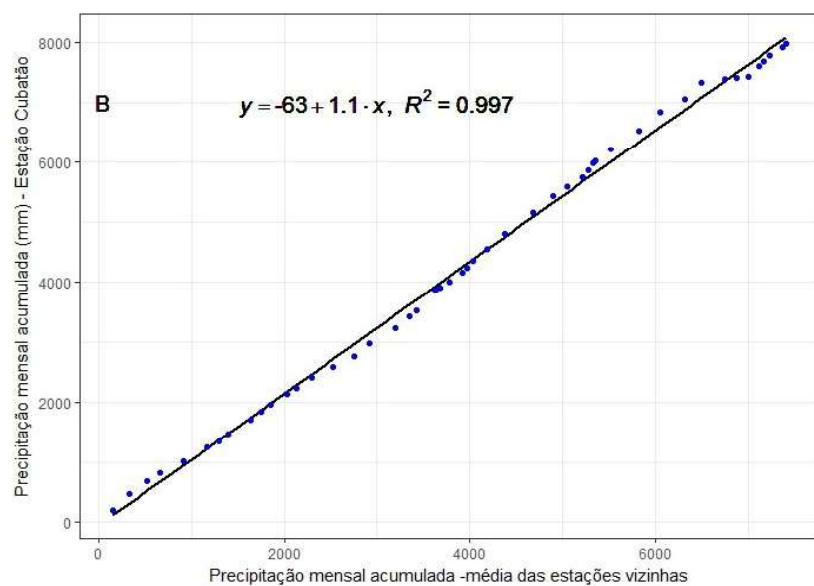


Figura 47 - Dupla Massa para Estação Cubatão sem preenchimento de Outubro de 2012 até Setembro de 2016



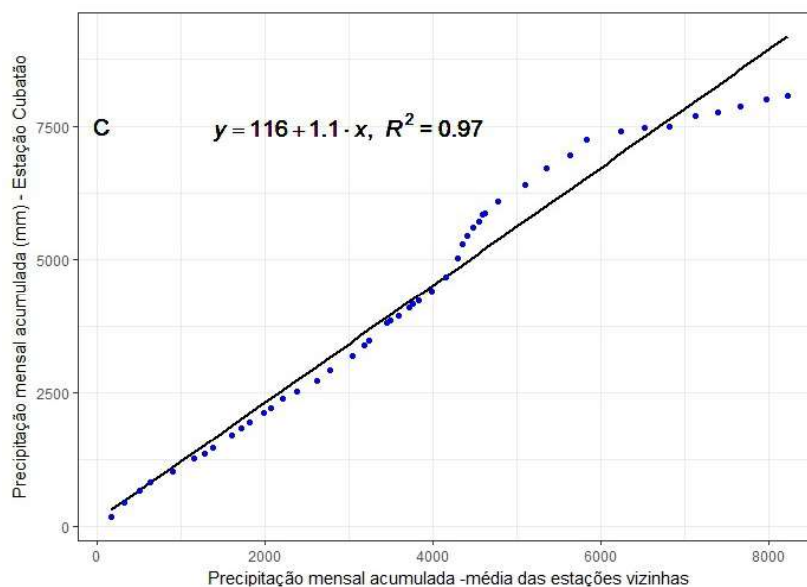
Fonte: A autora.

Figura 48 - Dupla Massa para Estação Cubatão preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016



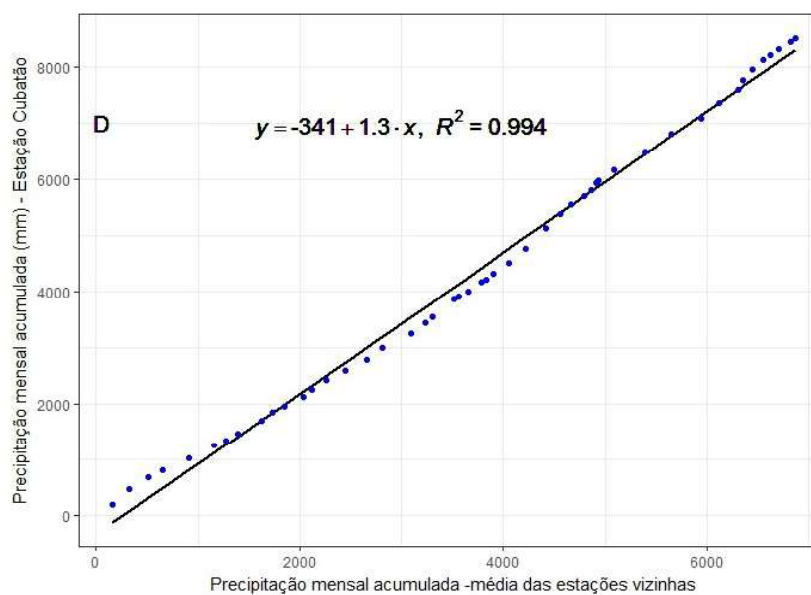
Fonte: A autora.

Figura 49 - Dupla Massa para Estação Cubatão preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016



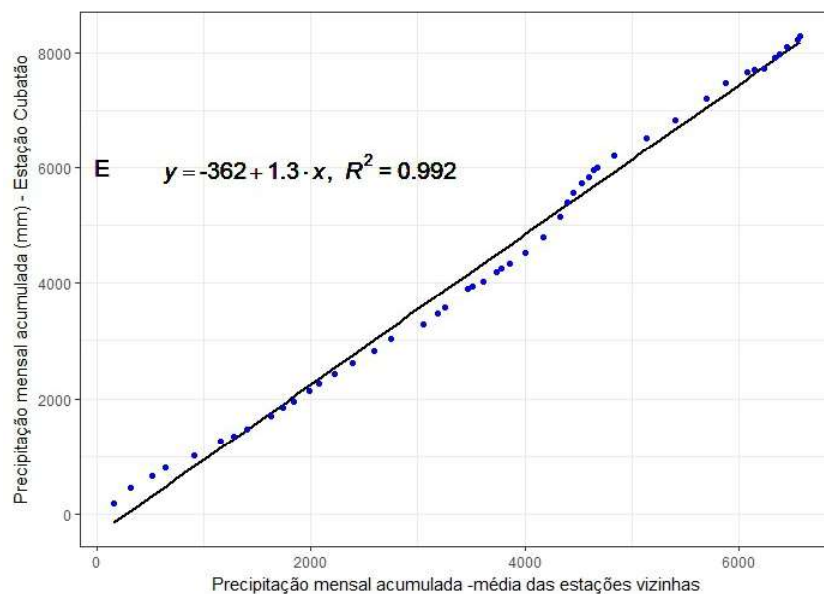
Fonte: A autora.

Figura 50 - Dupla Massa para Estação Cubatão preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016



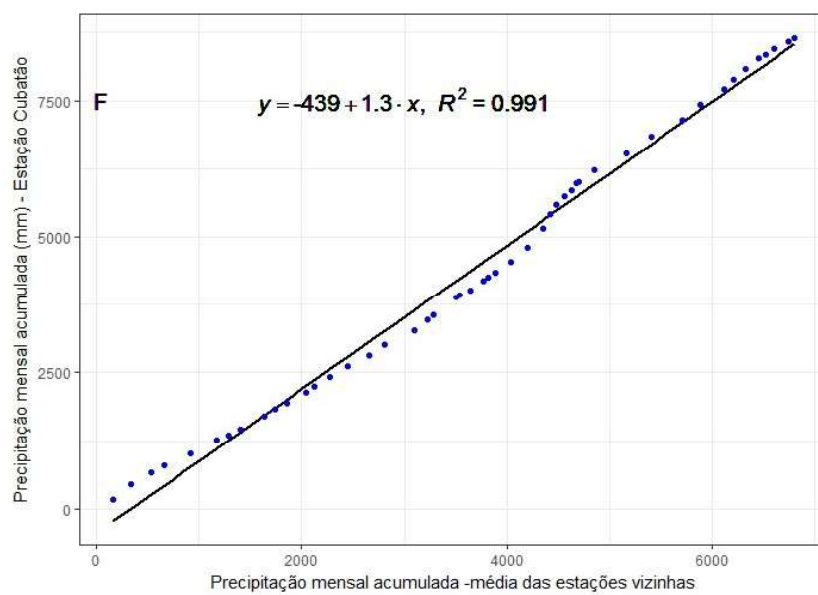
Fonte: A autora.

Figura 51 - Dupla Massa para Estação Cubatão preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

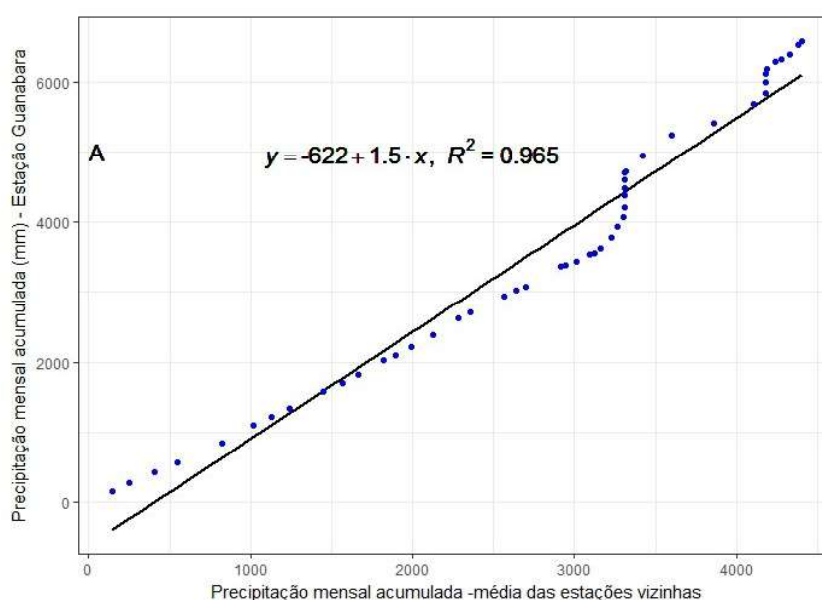
Figura 52 - Dupla Massa para Estação Cubatão preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

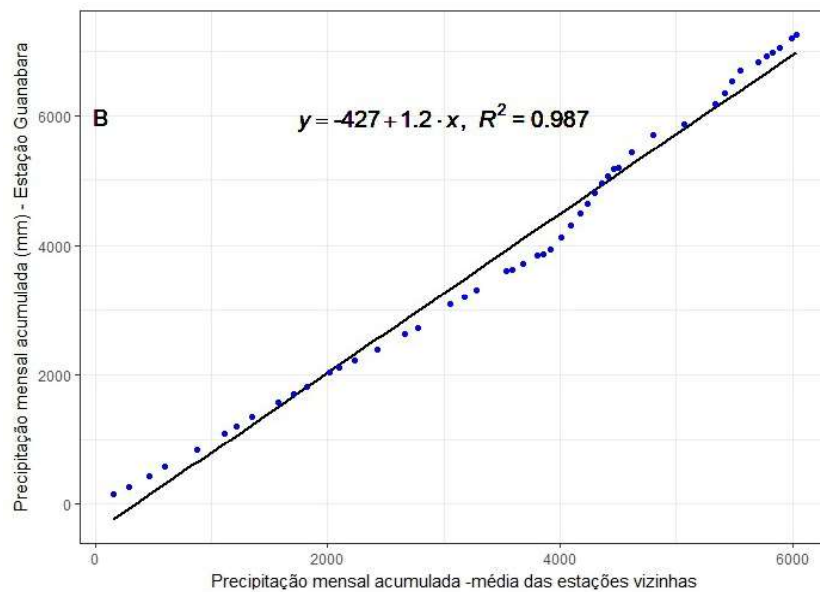
Na estação Guanabara as estações vizinhas foram: Unidade de Obras, Itaum e Bucarein, e os resultados da análise de consistência dos dados (a) sem preenchimento, (b) imputação múltipla, (c) interpolação linear, (d) método das correlações, (e) redes neurais artificiais, e (f) regressão linear múltipla, podem ser vistos nas Figura 53, Figura 54, Figura 55, Figura 56, Figura 57 e Figura 58.

Figura 53 - Dupla Massa para Estação Guanabara sem preenchimento de Outubro de 2012 até Setembro de 2016



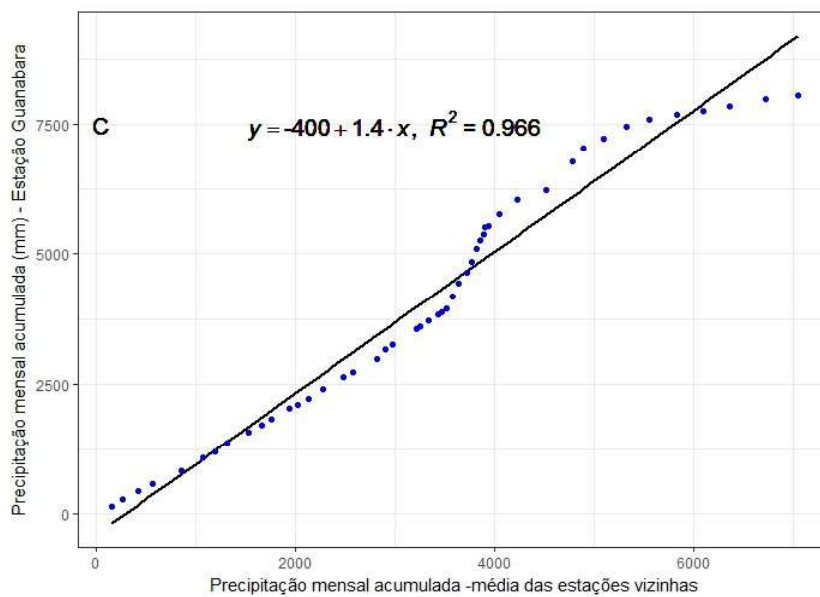
Fonte: A autora.

Figura 54 - Dupla Massa para Estação Guanabara preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016



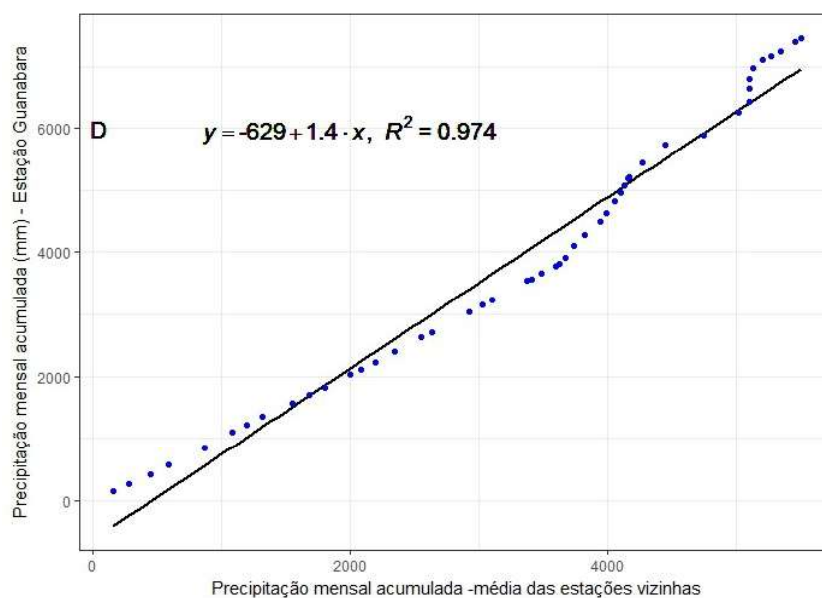
Fonte: A autora.

Figura 55 - Dupla Massa para Estação Guanabara preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016



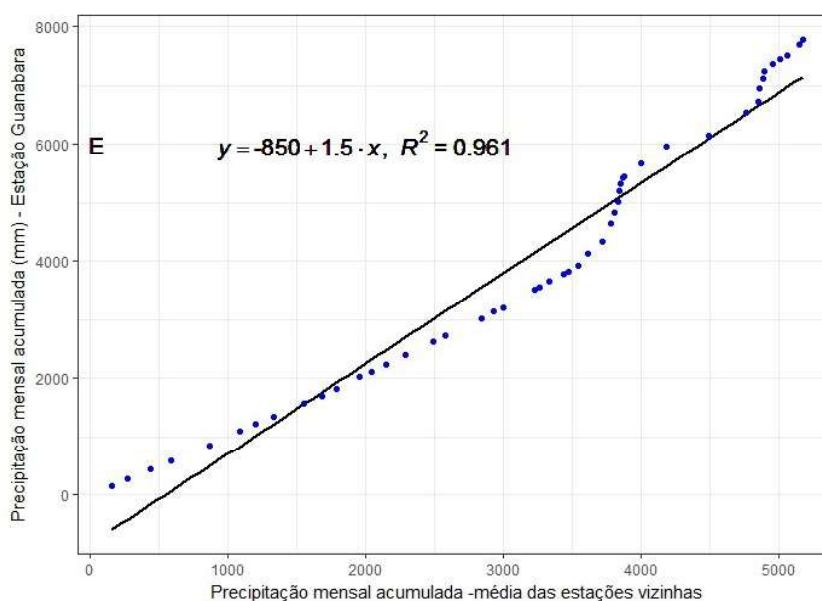
Fonte: A autora.

Figura 56 - Dupla Massa para Estação Guanabara preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016



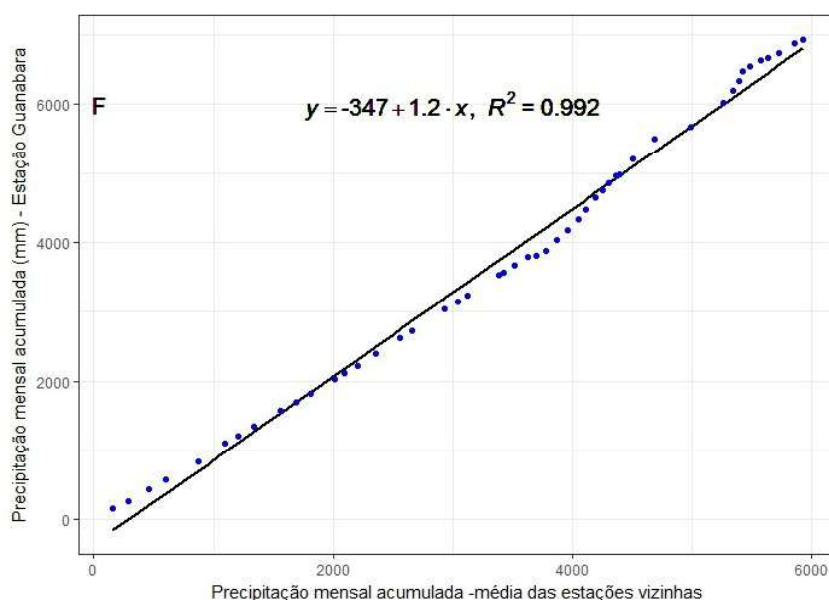
Fonte: A autora.

Figura 57 - Dupla Massa para Estação Guanabara preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

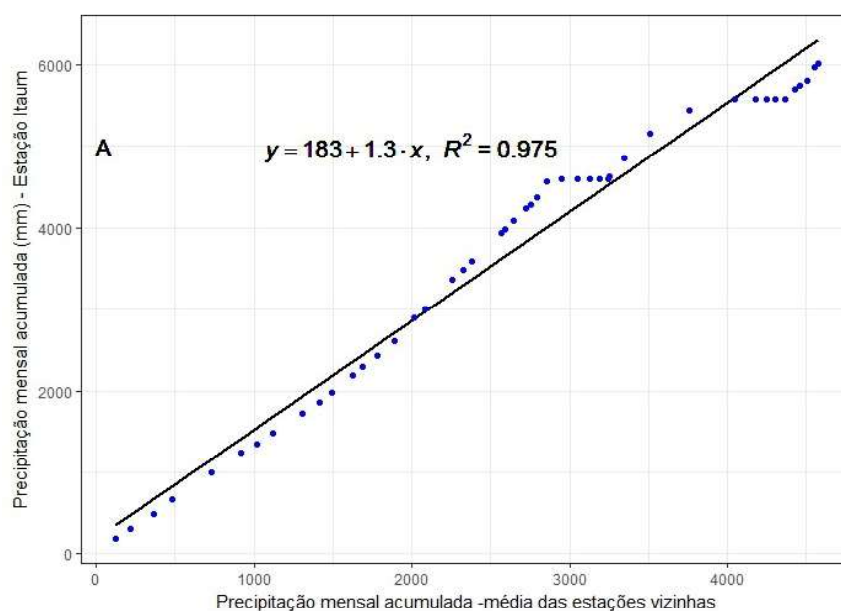
Figura 58 - Dupla Massa para Estação Guanabara preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

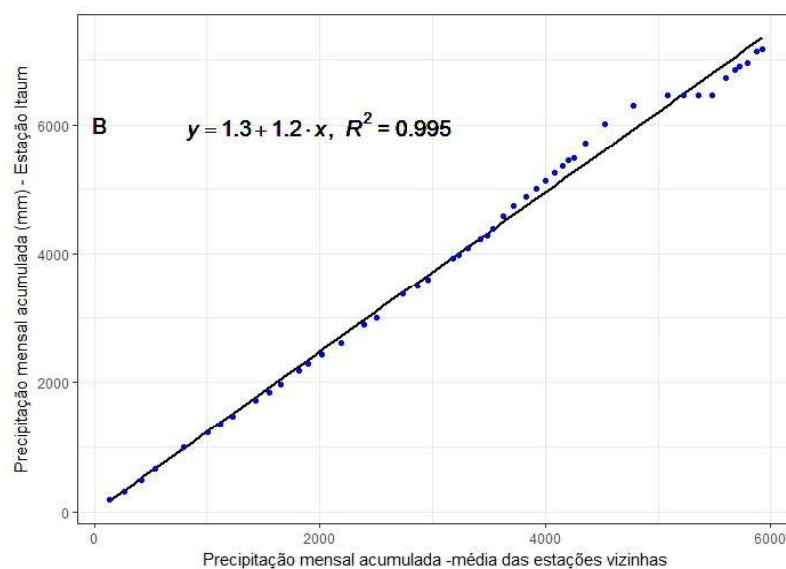
Na estação Itaum, as estações vizinhas utilizadas foram: Bucarien, Unidade de Obras e JIC. Os resultados da análise de consistência dos dados (a) sem preenchimento, (b) imputação múltipla, (c) interpolação linear, (d) método das correlações, (e) redes neurais artificiais, e (f) regressão linear múltipla, podem ser vistos nas Figura 59, Figura 60, Figura 61, Figura 62, Figura 63 e Figura 64.

Figura 59 - Dupla Massa para Estação Itaum sem preenchimento de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

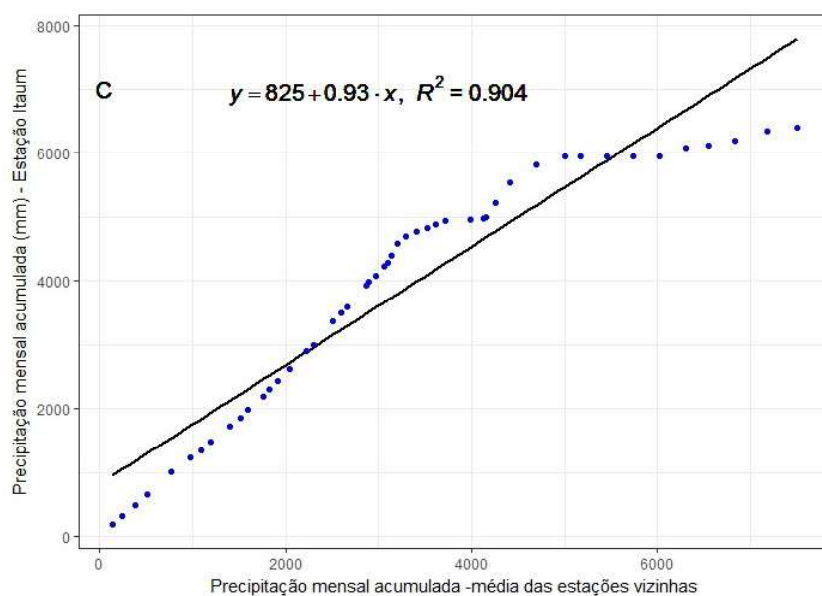
Figura 60 - Dupla Massa para Estação Itaum preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

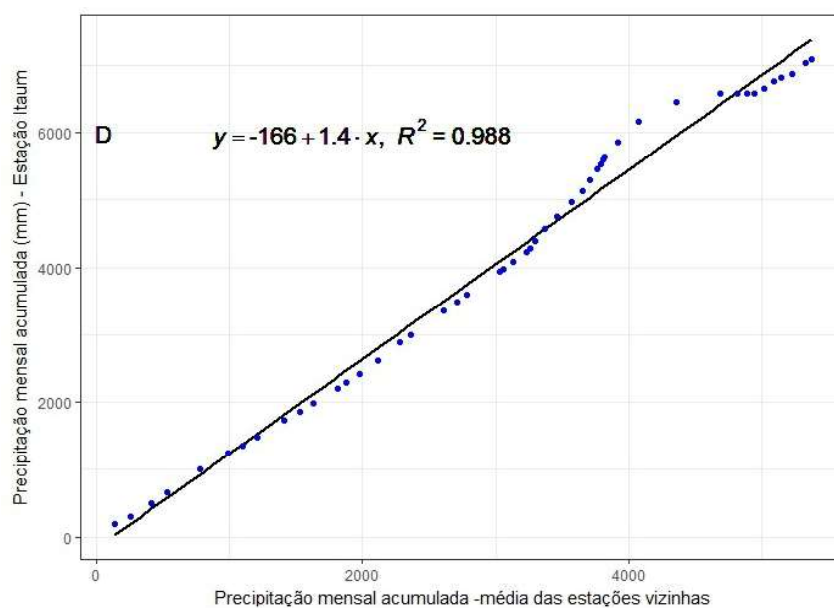


Figura 61 - Dupla Massa para Estação Itaum preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016



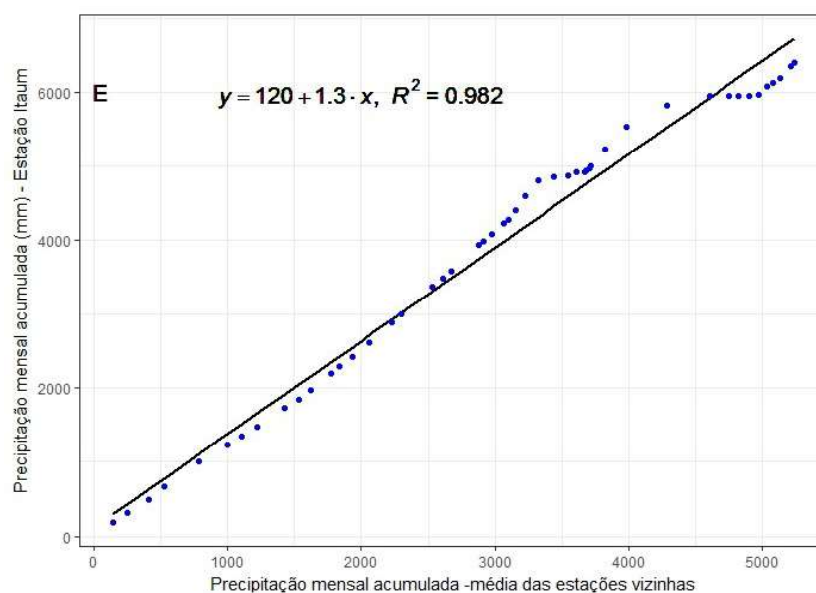
Fonte: A autora.

Figura 62 - Dupla Massa para Estação Itaum preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016



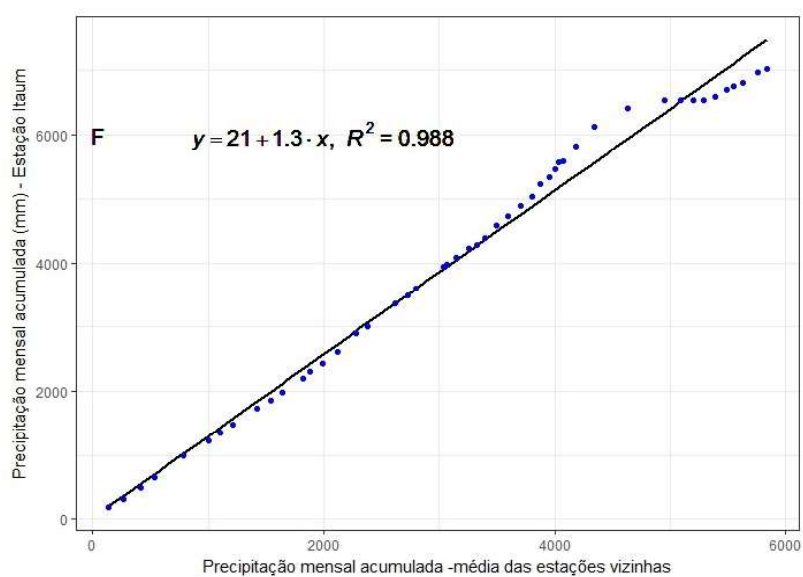
Fonte: A autora.

Figura 63 - Dupla Massa para Estação Itaum preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

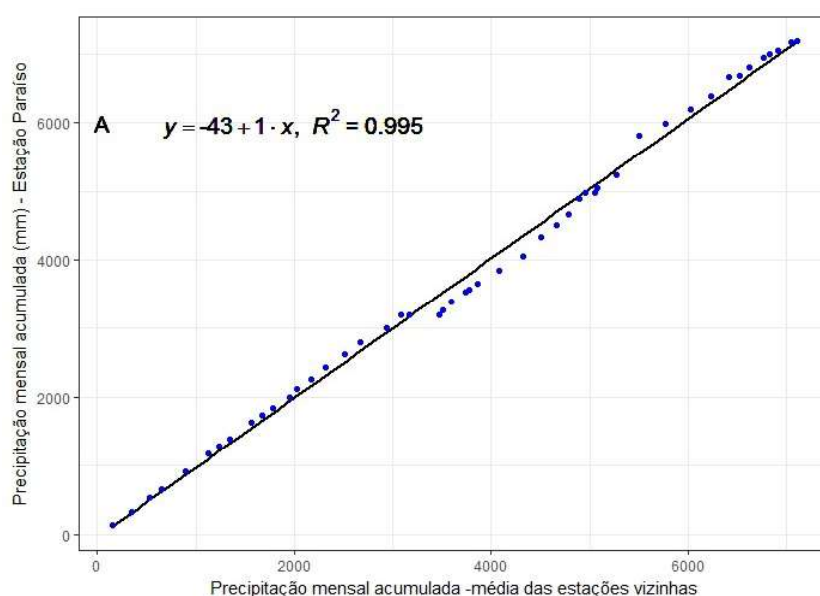
Figura 64 - Dupla Massa para Estação Itaum preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

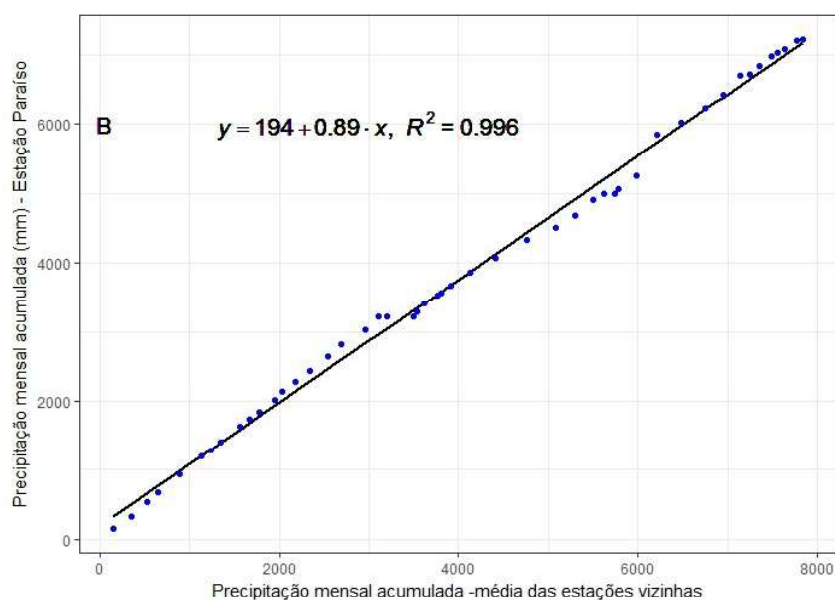
Na estação Paraíso as estações utilizadas como estações vizinhas foram: FlotFlux, JIC e Cubatão. Os resultados da análise de consistência dos dados (a) sem preenchimento, (b) imputação múltipla, (c) interpolação linear, (d) método das correlações, (e) redes neurais artificiais, e (f) regressão linear múltipla, podem ser vistos nas Figura 65, Figura 66, Figura 67, Figura 68, Figura 69 e Figura 70.

Figura 65 - Dupla Massa para Estação Paraíso sem preenchimento de Outubro de 2012 até Setembro de 2016



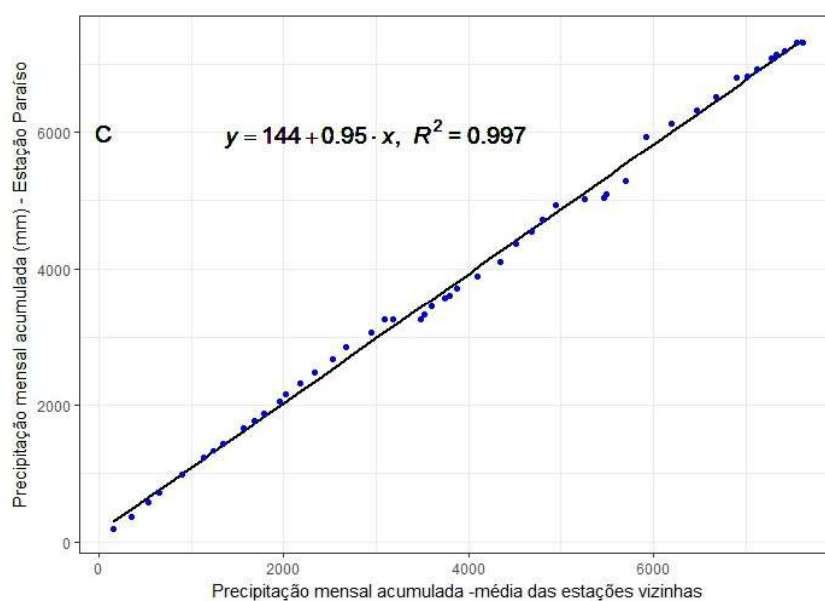
Fonte: A autora.

Figura 66 - Dupla Massa para Estação Paraíso preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016



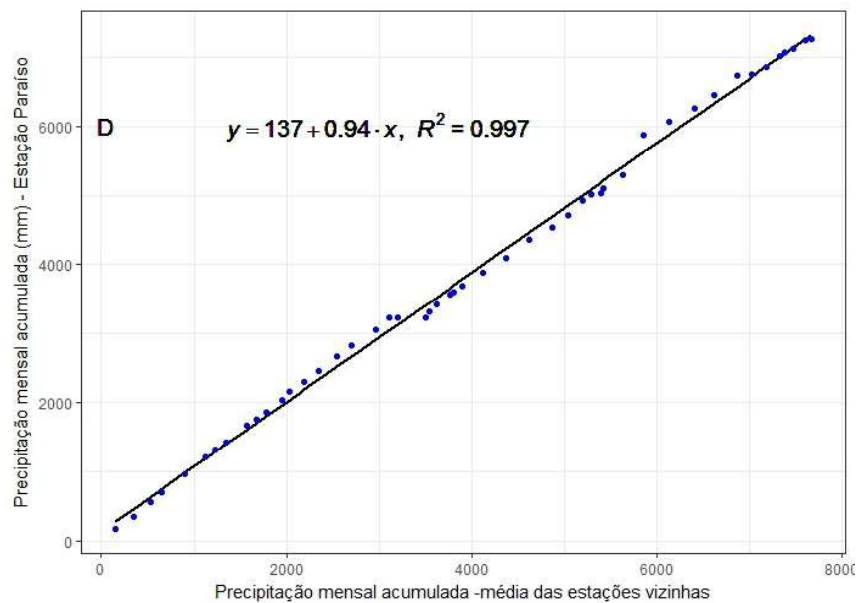
Fonte: A autora.

Figura 67 - Dupla Massa para Estação Paraíso preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016



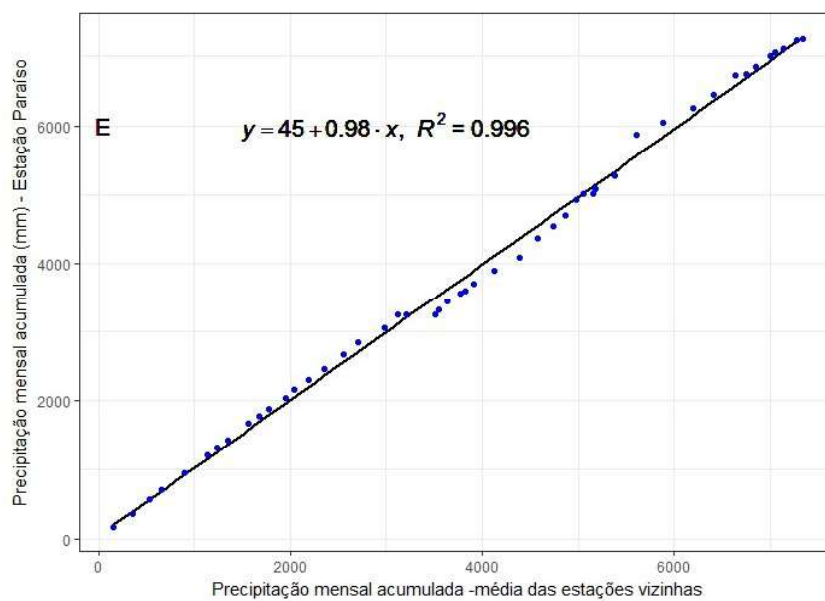
Fonte: A autora.

Figura 68 - Dupla Massa para Estação Paraíso preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016



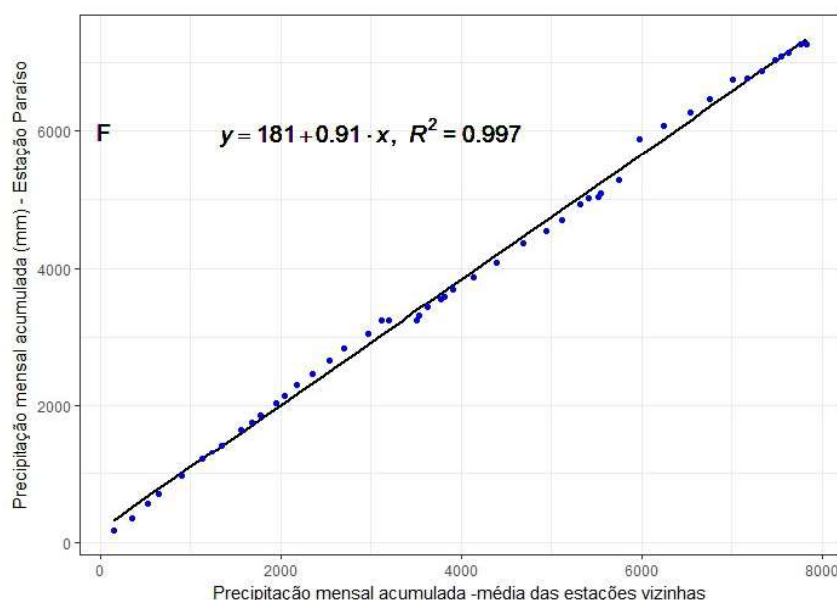
Fonte: A autora.

Figura 69 - Dupla Massa para Estação Paraíso preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

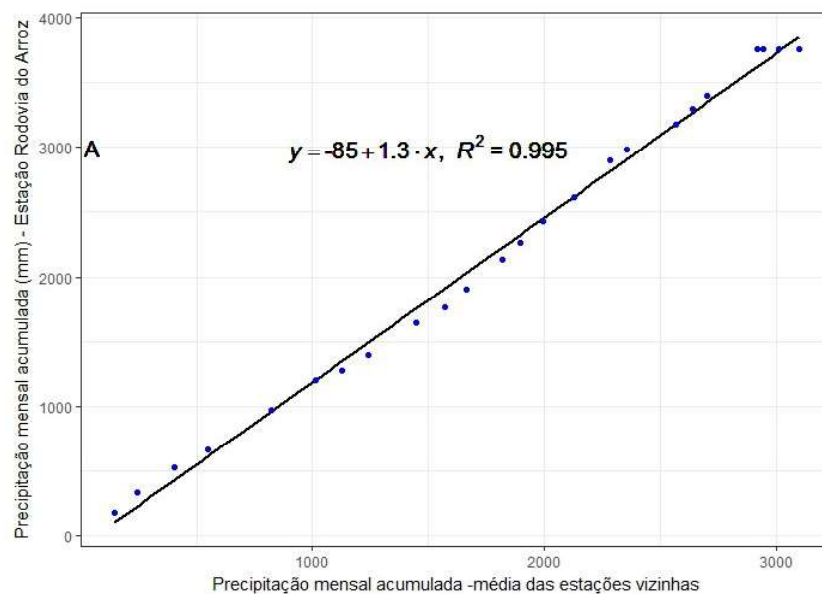
Figura 70 -Dupla Massa para Estação Paraíso preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

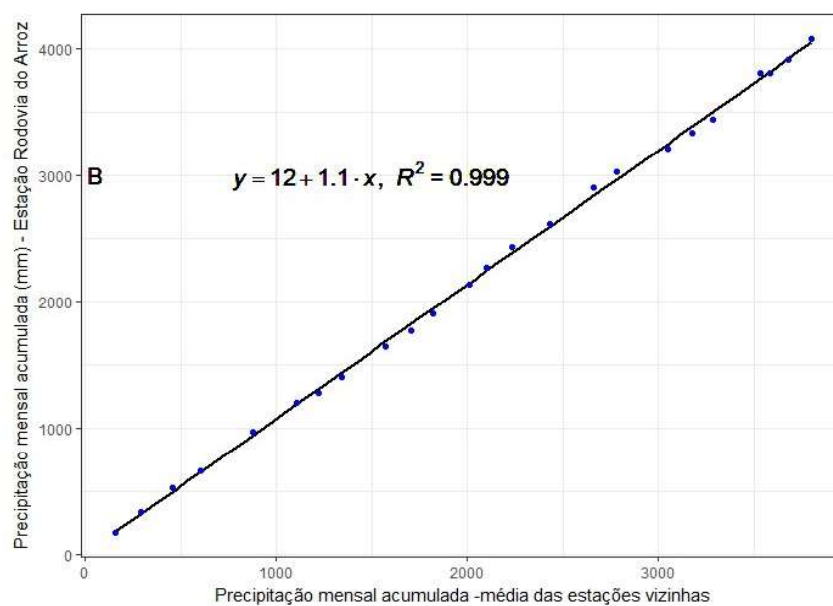
Para a estação Rodovia do Arroz as estações utilizadas como estações vizinhas foram: Bucarein, Itaum e Unidade de Obras. Os resultados da análise de consistência dos dados (a) sem preenchimento, (b) imputação múltipla, (c) interpolação linear, (d) método das correlações, (e) redes neurais artificiais, e (f) regressão linear múltipla, podem ser vistos nas Figura 71, Figura 72, Figura 73, Figura 74, Figura 75 e Figura 76.

Figura 71 - Dupla Massa para Estação Rodovia do Arroz sem preenchimento de Outubro de 2012 até Setembro de 2014



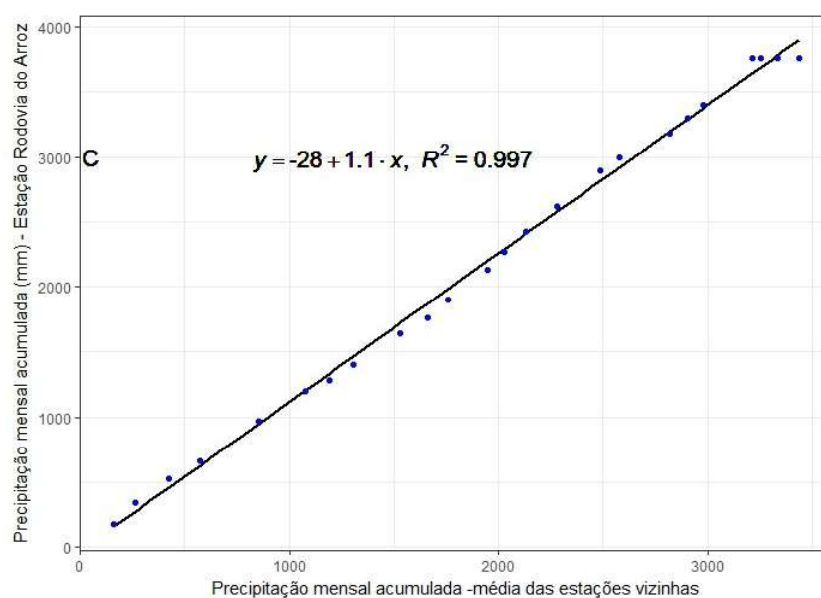
Fonte: A autora.

Figura 72 - Dupla Massa para Estação Rodovia do Arroz preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2014



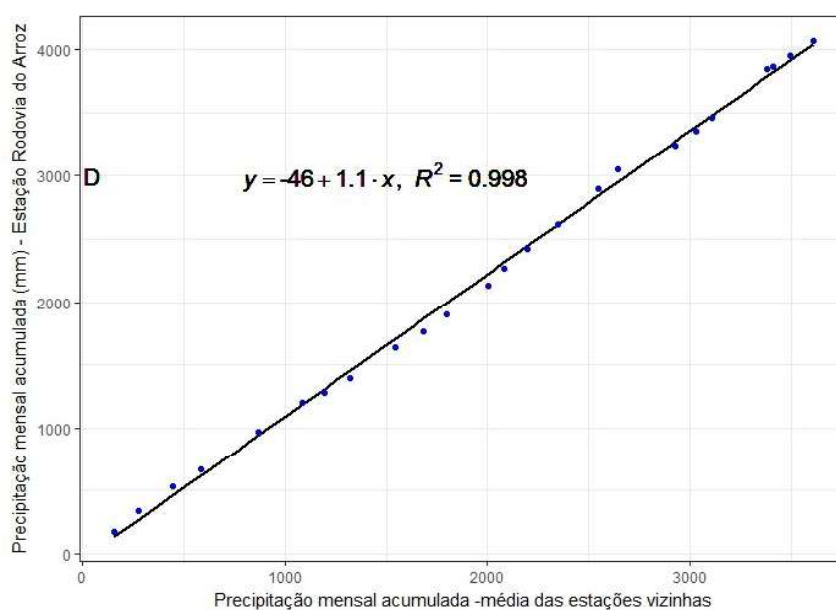
Fonte: A autora.

Figura 73 - Dupla Massa para Estação Rodovia do Arroz preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2014



Fonte: A autora.

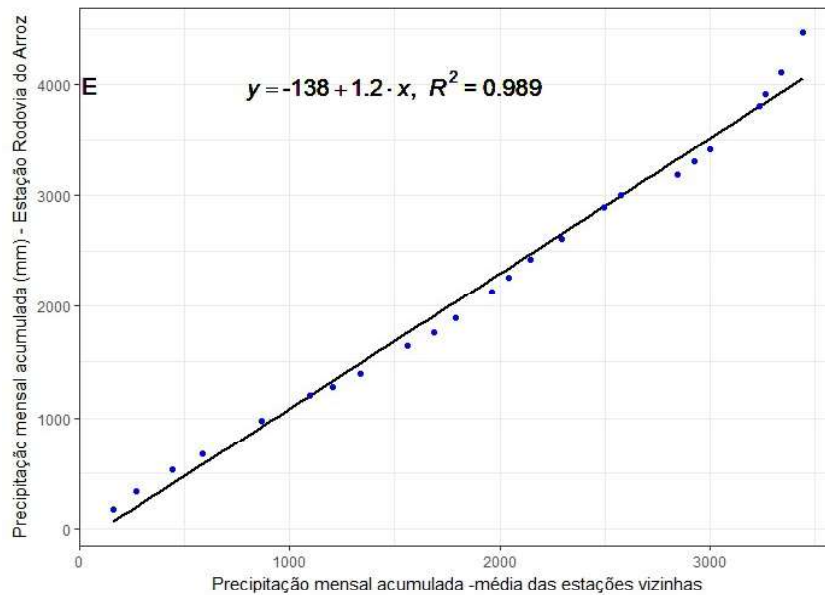
Figura 74 - Dupla Massa para Estação Rodovia do Arroz preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2014



Fonte: A autora.

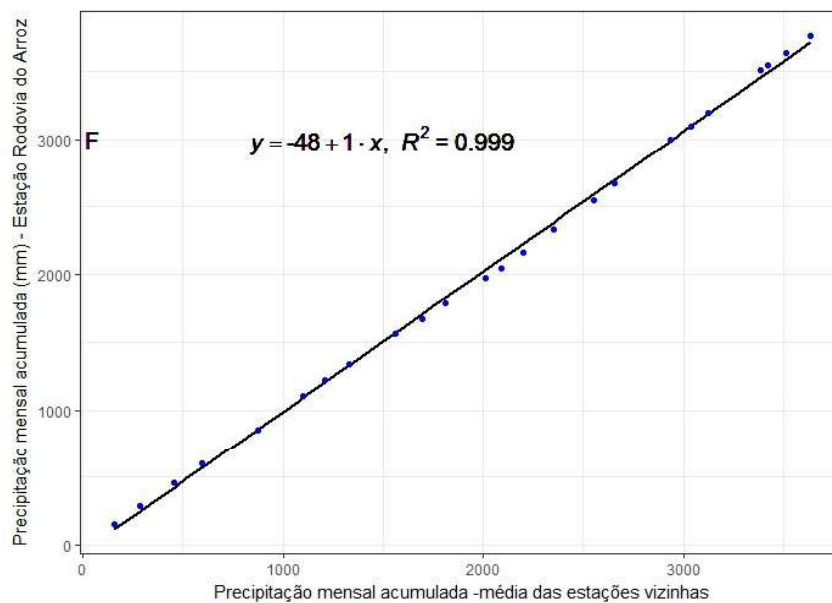


Figura 75 - Dupla Massa para Estação Rodovia do Arroz preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2014



Fonte: A autora.

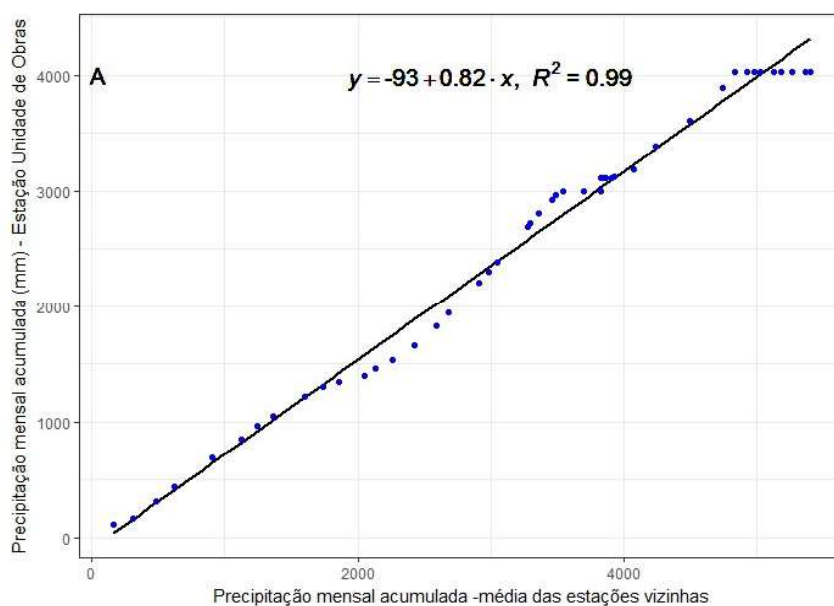
Figura 76 - Dupla Massa para Estação Rodovia do Arroz preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2014



Fonte: A autora.

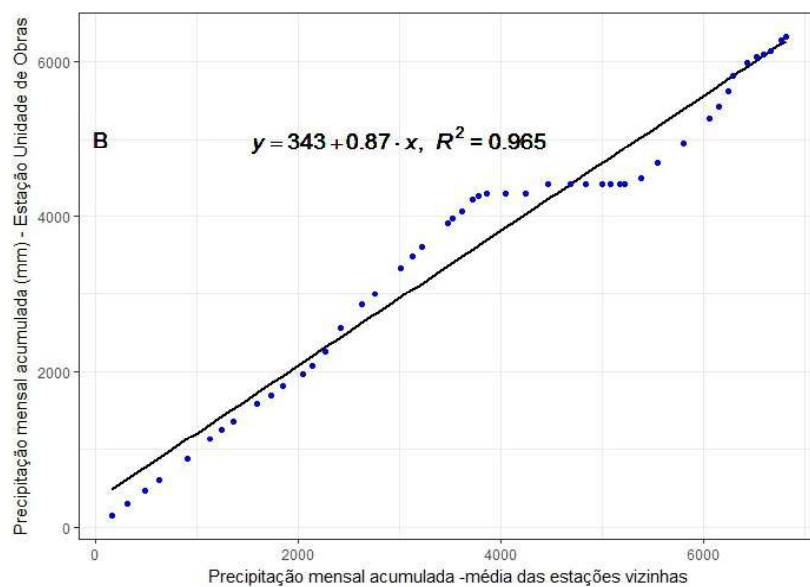
Enfim, para a estação Unidade de Obras as estações vizinhas foram: FlotFlux, Itaum e Bucarein. Os resultados da análise de consistência dos dados (a) sem preenchimento, (b) imputação múltipla, (c) interpolação linear, (d) método das correlações, (e) redes neurais artificiais, e (f) regressão linear múltipla, podem ser vistos nas Figura 77, Figura 78, Figura 79, Figura 80, Figura 81 e Figura 82.

Figura 77 - Dupla Massa para Estação Unidade de Obras sem preenchimento de Outubro de 2012 até Setembro de 2016



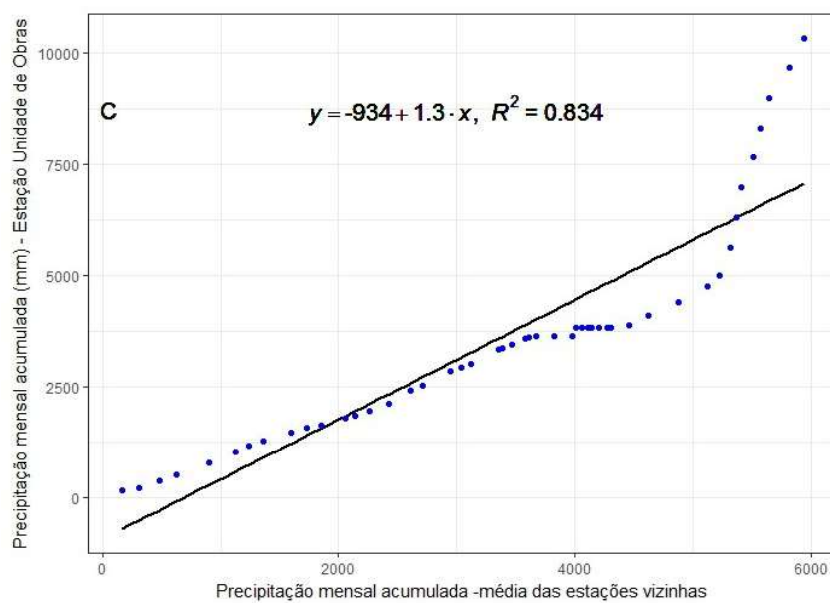
Fonte: A autora.

Figura 78 - Dupla Massa para Estação Unidade de Obras preenchida por Imputação Múltipla de Outubro de 2012 até Setembro de 2016



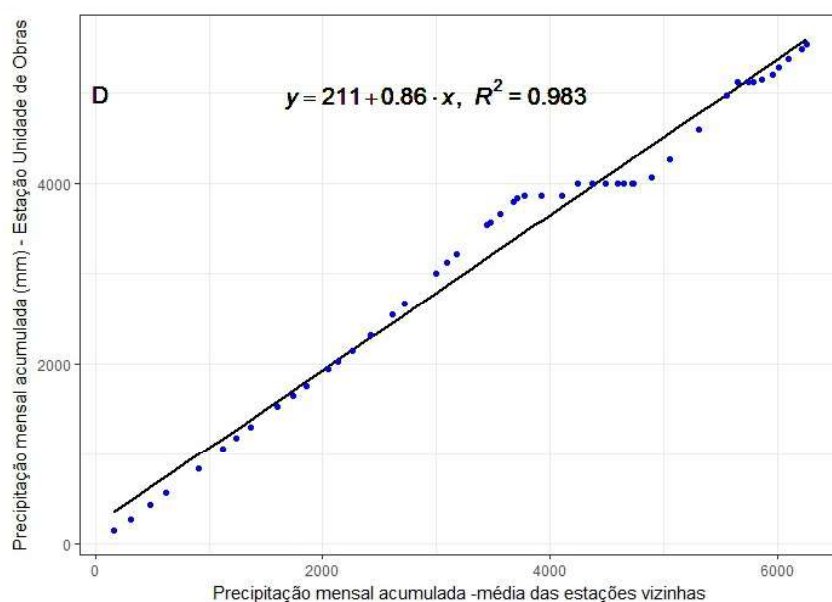
Fonte: A autora.

Figura 79 - Dupla Massa para Estação Unidade de Obras preenchida por Interpolação Linear de Outubro de 2012 até Setembro de 2016



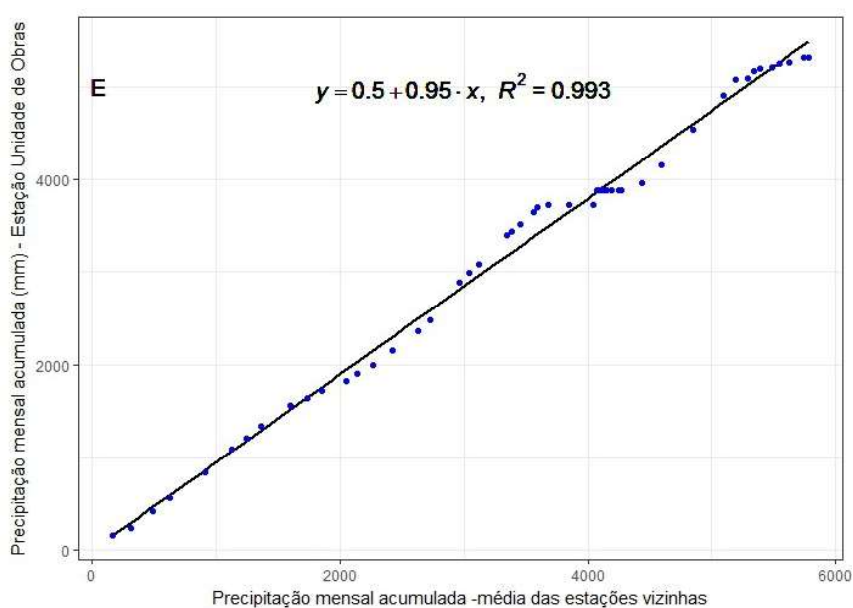
Fonte: A autora.

Figura 80 - Dupla Massa para Estação Unidade de Obras preenchida pelo Método das Correlações de Outubro de 2012 até Setembro de 2016



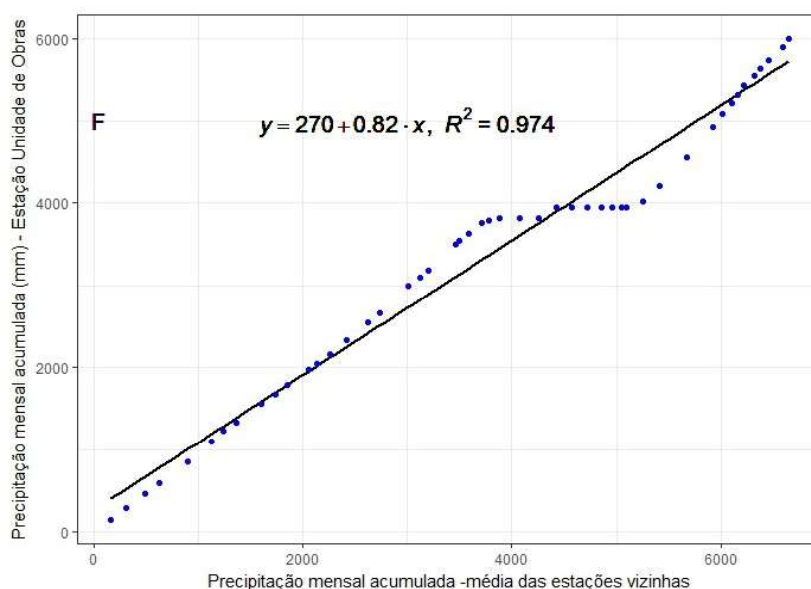
Fonte: A autora.

Figura 81 - Dupla Massa para Estação Unidade de Obras preenchida por Redes Neurais Artificiais de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

Figura 82 - Dupla Massa para Estação Unidade de Obras preenchida por Regressão Linear Múltipla de Outubro de 2012 até Setembro de 2016



Fonte: A autora.

Os autores Oliveira *et al* (2010), Coelho (2014) e Mello (2015) também utilizaram o Método da Dupla Massa para verificar a consistência dos dados ao comparar os acumulados de precipitação anual entre os postos pluviométricos por eles utilizados.

Estes autores encontraram valores de  $R^2$  bastante próximos a 1, semelhantes aos valores encontrados para as estações da Defesa Civil de Joinville neste trabalho. Em geral as estações utilizadas nesta pesquisa mostraram consistência nos dados antes e após todos os métodos de preenchimento aplicados, alcançando valores de  $R^2$  no Método da Dupla Massa entre 0,834, no método de interpolação linear para a estação Unidade de Obras, e 0,999 para a estação Rodovia do Arroz no método de imputação múltipla. Isto demonstra o bom desempenho dos métodos de preenchimento e consequentemente, disponibiliza uma série hidrológica consistente para ser utilizadas nos próximos estudos.

## 5 CONCLUSÃO

A presente pesquisa teve como objetivo principal preencher dados faltantes e com erros de precipitação de 24 horas nas estações pluviométricas da rede de monitoramento hidrometeorológico da Defesa Civil de Joinville (SC). O município apresenta altos índices pluviométricos, mas não dispõe de uma longa série de dados pluviográficos para ser analisada e explorada a fim de auxiliar em previsões de precipitação e planejamento urbano, por exemplo.

Os métodos de preenchimento de falhas aplicados neste trabalho tinham o intuito de preencher as falhas originais e valores discrepantes primeiramente. E posteriormente preencher as falhas simuladas que correspondiam ao período de dados reais retirados da série hidrológica. Os melhores resultados foram obtidos com o método de regressão linear múltipla e método das correlações, ambos apresentando valores bastante satisfatórios, com valores de *Root Mean Square Error - RMSE* = 0,617 a 14,333 mm, *Mean Relative Error - MRE* = 0,009 a 4,005, *Mean Absolute Error - MAE* = 0,035 a 7,204 mm e coeficiente de determinação  $R^2$  = 0,003 a 1. Estes dois métodos apresentaram medidas de erro próximas a zero em sua maioria, e melhores que os demais métodos utilizados neste trabalho, embora isto não tenha sido observado em 100% dos casos.

Quando comparadas as medidas de erro, os maiores valores foram obtidos no erro *Root Mean Square Error - RMSE* (24,333 mm) e os menores para o *Mean Relative Error - MRE* (0,009), o que indica que o erro relativo médio apresentou bons resultados, próximos à zero, mas o *RMSE* teve variações em sua magnitude, alcançando valores bastante altos se comparados com as outras medidas de erro. Contudo, esses valores altos de *RMSE* são aceitáveis, pois tratam-se de precipitações de 24 horas.

É importante ressaltar que o pesquisador deve conhecer o local que está trabalhando quando realizar o preenchimento de falhas, pois valores de precipitação que parecem discrepantes para determinada região, podem não ser assim decretados para outras. Os métodos de preenchimento aplicados à esta pesquisa e suas rotinas em *R* podem ser reaplicados para outras regiões com o intuito de realizar o preenchimento de falhas de dados pluviométricos, entretanto, os resultados aqui obtidos foram fortemente influenciados pela semelhança climatológica das estações, pois todas estão localizadas

na planície. Isto é, para outras regiões, podem ser encontrados outros métodos mais adequados ao realizar o preenchimento de falhas de dados de precipitação.

Algumas estações tiveram seu preenchimento inviabilizados em decorrência do alto número de falhas registrados, muitas vezes devido à falta de manutenção dos equipamentos, como por exemplo no caso da estação Bucarein que não consegue transmitir os dados devido à danificação no rádio. A estação Rodovia do Arroz também apresenta defeitos no rádio, além dos sensores de vento e temperatura. Na estação Jativoca o problema está no *datalogger*, no rádio de transmissão e ainda, houve furto da bateria e conversor da placa solar.

Em geral os dados das estações demonstraram consistência após os métodos de preenchimento aplicados, atingindo valores de  $R^2$  no Método da Dupla Massa entre 0,834 e 0,999, embora a série hidrológica utilizada nesta pesquisa seja bastante curta, e no caso das estações Jativoca e Rodovia do Arroz a série analisada ainda foi reduzida em 2 anos.

Os resultados obtidos no presente trabalho levam a considerar que o método menos apropriado para o preenchimento de falhas de dados de precipitação nas estações e condições analisadas é o método da interpolação linear devido aos maiores valores das medidas de erros obtidas ( $RMSE = 23,510$  mm;  $MRE = 9,990$ ;  $MAE = 17,830$  mm) e também, aos menores valores de  $R^2$  (0 a 0,57), apesar da forte tendência linear entre os dados dos postos pluviométricos utilizados na análise de consistência. Contudo, em relação às medidas de erros, coeficientes de determinação e análise de consistência de dados, indica-se o método das correlações como o mais adequado para o preenchimento de dados de precipitação de 24 horas para a cidade de Joinville. Destaca-se também o método da regressão linear múltipla que obteve desempenho semelhante ao método das correlações quanto às medidas de erro e coeficiente de determinação, podendo também ser utilizado no preenchimento de falhas de dados de precipitação para a região estudada.

## 6 LIMITAÇÕES E SUGESTÕES DE TRABALHOS FUTUROS

O trabalho desenvolvido contou com algumas limitações, como por exemplo, a série hidrológica de apenas 4 anos de observação. Outro fator relevante foi a quantidade de falhas encontradas nas séries de dados, inviabilizando o preenchimento de algumas estações, como nas estações Jativoca e Rodovia do Arroz durante o terceiro e quarto ano hidrológico, sem nenhuma observação registrada devido à falta de manutenção e falhas no equipamento de transmissão de dados.

Uma sugestão para trabalhos futuros, seria analisar a consistência dos dados utilizando as estações do Centro Nacional de Monitoramento e Alertas de Desastres Naturais – Cemaden como referência a título de comparação. Identificando a estação do Cemaden que está localizada mais próxima à estação a ser preenchida da Defesa Civil e comparando os dados em ambas as estações o que poderia auxiliar na verificação de dados classificados como duvidosos, por exemplo.

Outra sugestão seria repetir este estudo para a mesma região dispondo de uma série hidrológica de longa duração a fim de aplicar os métodos indicados para o preenchimento de falhas e verificar a consistência novamente.



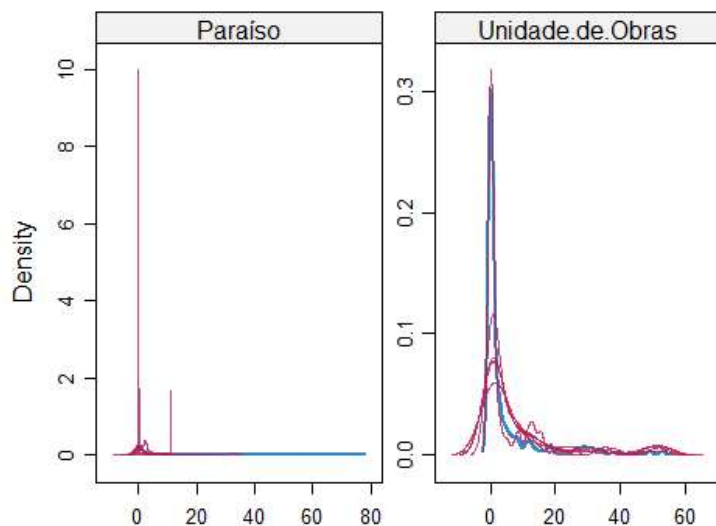
## APÊNDICE A – Rotina no R para o método de Imputação Múltipla

```
##primeiro ano hidrológico
anohidroum.semnan.semout <- read.csv2("F:/anohidroum-semnan-semout.csv")
attach(anohidroum.semnan.semout)
library(mice)
require(lattice)

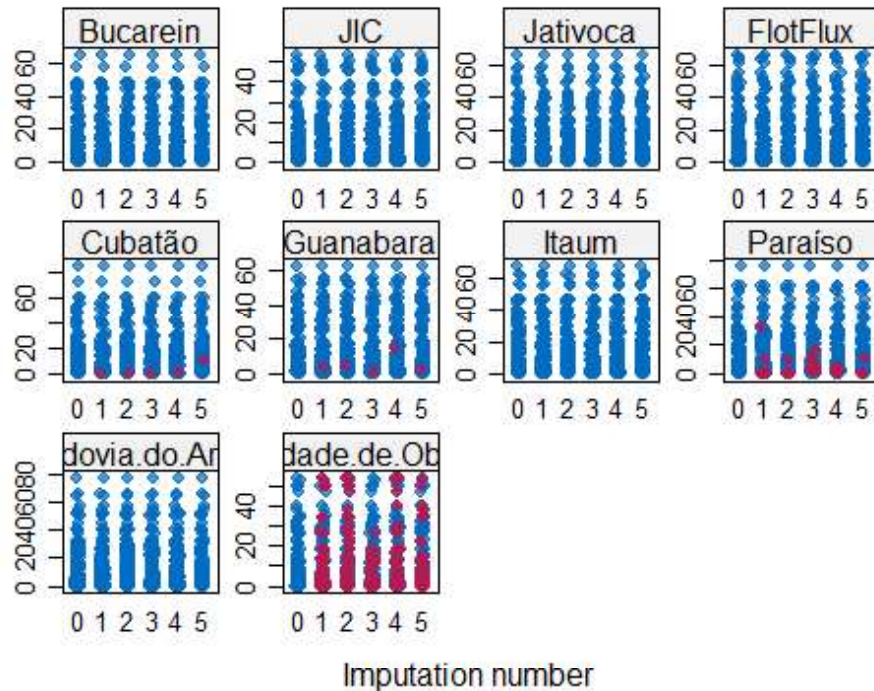
imp <-mice(anohidroum.semnan.semout)
## iter imp variable
## 1 1 Cubatão Guanabara Paraíso Unidade.de.Obras
## 1 2 Cubatão Guanabara Paraíso Unidade.de.Obras
## 1 3 Cubatão Guanabara Paraíso Unidade.de.Obras
## 1 4 Cubatão Guanabara Paraíso Unidade.de.Obras
## 1 5 Cubatão Guanabara Paraíso Unidade.de.Obras
## 2 1 Cubatão Guanabara Paraíso Unidade.de.Obras
## 2 2 Cubatão Guanabara Paraíso Unidade.de.Obras
## 2 3 Cubatão Guanabara Paraíso Unidade.de.Obras
## 2 4 Cubatão Guanabara Paraíso Unidade.de.Obras
## 2 5 Cubatão Guanabara Paraíso Unidade.de.Obras
## 3 1 Cubatão Guanabara Paraíso Unidade.de.Obras
## 3 2 Cubatão Guanabara Paraíso Unidade.de.Obras
## 3 3 Cubatão Guanabara Paraíso Unidade.de.Obras
## 3 4 Cubatão Guanabara Paraíso Unidade.de.Obras
## 3 5 Cubatão Guanabara Paraíso Unidade.de.Obras
## 4 1 Cubatão Guanabara Paraíso Unidade.de.Obras
## 4 2 Cubatão Guanabara Paraíso Unidade.de.Obras
## 4 3 Cubatão Guanabara Paraíso Unidade.de.Obras
## 4 4 Cubatão Guanabara Paraíso Unidade.de.Obras
## 4 5 Cubatão Guanabara Paraíso Unidade.de.Obras
## 5 1 Cubatão Guanabara Paraíso Unidade.de.Obras
## 5 2 Cubatão Guanabara Paraíso Unidade.de.Obras
## 5 3 Cubatão Guanabara Paraíso Unidade.de.Obras
## 5 4 Cubatão Guanabara Paraíso Unidade.de.Obras
## 5 5 Cubatão Guanabara Paraíso Unidade.de.Obras

miceum=complete(imp)
write.table(miceum, file = "F:/miceum.csv", sep = ";", dec = ",", col.names = NA, qme
thod = "double")
complete(imp)
```

```
densityplot(imp)
```



```
stripplot(imp, pch = 20, cex = 1.2)
```



## APÊNDICE B – Rotina no R para o método de Interpolação Linear

```
## Rotina - Interpolação linear
diariosemoutlier <- read.csv2("F:/diariosemoutlier.csv")
attach(diariosemoutlier)
library(forecast)
library(utils)

#Primeiro ano hidrológico - Outubro/12 até Setembro/13
anohidroum.semnan.semout <- read.csv2("F:/anohidroum-semnan-semout.csv")
attach(anohidroum.semnan.semout)
summary(anohidroum.semnan.semout)
```

##	data	Bucarein	JIC	Jativoca
##	01/01/2013: 1	Min. : 0.000	Min. : 0.000	Min. : 0.000
##	01/02/2013: 1	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000
##	01/03/2013: 1	Median : 0.229	Median : 0.229	Median : 0.457
##	01/04/2013: 1	Mean : 5.108	Mean : 4.428	Mean : 5.193
##	01/05/2013: 1	3rd Qu.: 4.572	3rd Qu.: 3.886	3rd Qu.: 4.343
##	01/06/2013: 1	Max. : 65.070	Max. : 53.640	Max. : 66.500
##	(Other) :359			
##	FlotFlux	Cubatão	Guanabara	Itaum
##	Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. : 0.000
##	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000
##	Median : 0.229	Median : 0.457	Median : 0.457	Median : 0.457
##	Mean : 5.756	Mean : 5.859	Mean : 5.564	Mean : 6.009
##	3rd Qu.: 4.572	3rd Qu.: 5.543	3rd Qu.: 5.029	3rd Qu.: 5.029
##	Max. : 66.290	Max. : 84.600	Max. : 62.150	Max. : 67.540
##		NA's :1	NA's :1	
##	Paraíso	Rodovia.do.Arroz	Unidade.de.Obras	
##	Min. : 0.000	Min. : 0.000	Min. : 0.000	
##	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	
##	Median : 0.457	Median : 0.457	Median : 0.229	
##	Mean : 5.609	Mean : 5.837	Mean : 4.782	
##	3rd Qu.: 4.809	3rd Qu.: 5.029	3rd Qu.: 3.715	
##	Max. : 74.960	Max. : 77.670	Max. : 53.620	
##	NA's :8		NA's :73	

```
C= na.interp(Cubatão)
write.table(C, file = "F:/interpcubatao.csv", sep = ";", dec = ",", col.names = NA, q
method = "double")
G= na.interp(Guanabara)
write.table(G, file = "F:/interpguana.csv", sep = ";", dec = ",", col.names = NA, qme
thod = "double")
P= na.interp(Paraíso)
write.table(P, file = "F:/interparaiso.csv", sep = ";", dec = ",", col.names = NA, qm
ethod = "double")
U= na.interp(Unidade.de.Obras)
write.table(U, file = "F:/interpunidade.csv", sep = ";", dec = ",", col.names = NA, q
method = "double")
```

## APÊNDICE C – Rotina no R para o Método das Correlações

```
##primeiro ano
anohidroum.semnan.semout <- read.csv2("F:/anohidroum-semnan-semout.csv")

attach(anohidroum.semnan.semout)
library(hyfo)

um= fillGap(anohidroum.semnan.semout)
## Correlation Coefficient
##
##      Bucarein      JIC  Jativoca  FlotFlux  Cubatão
## Bucarein      1.000000 0.8364152 0.8030002 0.8488447 0.7334335
## JIC            0.8364152 1.0000000 0.8504732 0.9543339 0.7798861
## Jativoca       0.8030002 0.8504732 1.0000000 0.8479037 0.7924917
## FlotFlux       0.8488447 0.9543339 0.8479037 1.0000000 0.8299451
## Cubatão        0.7334335 0.7798861 0.7924917 0.8299451 1.0000000
## Guanabara      0.9077778 0.9259846 0.8750172 0.9224978 0.7744182
## Itaum          0.8913385 0.9102155 0.8927729 0.8956825 0.7811353
## Paraíso        0.8010634 0.8929185 0.8321064 0.9287824 0.9007485
## Rodovia.do.Arroz 0.8299785 0.8448643 0.9251474 0.8871979 0.8144245
## Unidade.de.Obras 0.8971945 0.9552749 0.8814326 0.9559656 0.8037821
##
##      Guanabara      Itaum      Paraíso Rodovia.do.Arroz
## Bucarein      0.9077778 0.8913385 0.8010634      0.8299785
## JIC            0.9259846 0.9102155 0.8929185      0.8448643
## Jativoca       0.8750172 0.8927729 0.8321064      0.9251474
## FlotFlux       0.9224978 0.8956825 0.9287824      0.8871979
## Cubatão        0.7744182 0.7811353 0.9007485      0.8144245
## Guanabara      1.0000000 0.9659091 0.8642078      0.8908332
## Itaum          0.9659091 1.0000000 0.8584729      0.8804421
## Paraíso        0.8642078 0.8584729 1.0000000      0.8364972
## Rodovia.do.Arroz 0.8908332 0.8804421 0.8364972      1.0000000
## Unidade.de.Obras 0.9813780 0.9592123 0.8988128      0.9047652
##
##      Unidade.de.Obras
## Bucarein      0.8971945
## JIC            0.9552749
## Jativoca       0.8814326
## FlotFlux       0.9559656
## Cubatão        0.8037821
## Guanabara      0.9813780
## Itaum          0.9592123
## Paraíso        0.8988128
## Rodovia.do.Arroz 0.9047652
## Unidade.de.Obras 1.0000000
##

## Correlation Order
##
##      1      2      3
## Bucarein "Guanabara" "Unidade.de.Obras" "Itaum"
## JIC      "Unidade.de.Obras" "FlotFlux" "Guanabara"
## Jativoca "Rodovia.do.Arroz" "Itaum" "Unidade.de.Obras"
## FlotFlux "Unidade.de.Obras" "JIC" "Paraíso"
## Cubatão  "Paraíso" "FlotFlux" "Rodovia.do.Arroz"
## Guanabara "Unidade.de.Obras" "Itaum" "JIC"
```

```

## Itaum "Guanabara" "Unidade.de.Obras" "JIC"
## Paraíso "FlotFlux" "Cubatão" "Unidade.de.Obras"
## Rodovia.do.Arroz "Jativoca" "Unidade.de.Obras" "Guanabara"
## Unidade.de.Obras "Guanabara" "Itaum" "FlotFlux"
## 4 5 6
## Bucarein "FlotFlux" "JIC" "Rodovia.do.Arroz"
## JIC "Itaum" "Paraíso" "Jativoca"
## Jativoca "Guanabara" "JIC" "FlotFlux"
## FlotFlux "Guanabara" "Itaum" "Rodovia.do.Arroz"
## Cubatão "Unidade.de.Obras" "Jativoca" "Itaum"
## Guanabara "FlotFlux" "Bucarein" "Rodovia.do.Arroz"
## Itaum "FlotFlux" "Jativoca" "Bucarein"
## Paraíso "JIC" "Guanabara" "Itaum"
## Rodovia.do.Arroz "FlotFlux" "Itaum" "JIC"
## Unidade.de.Obras "JIC" "Rodovia.do.Arroz" "Paraíso"
## 7 8 9
## Bucarein "Jativoca" "Paraíso" "Cubatão"
## JIC "Rodovia.do.Arroz" "Bucarein" "Cubatão"
## Jativoca "Paraíso" "Bucarein" "Cubatão"
## FlotFlux "Bucarein" "Jativoca" "Cubatão"
## Cubatão "JIC" "Guanabara" "Bucarein"
## Guanabara "Jativoca" "Paraíso" "Cubatão"
## Itaum "Rodovia.do.Arroz" "Paraíso" "Cubatão"
## Paraíso "Rodovia.do.Arroz" "Jativoca" "Bucarein"
## Rodovia.do.Arroz "Paraíso" "Bucarein" "Cubatão"
## Unidade.de.Obras "Bucarein" "Jativoca" "Cubatão"
##
## Linear Coefficients
## 1 2 3 4 5
## Bucarein 0.8804455 0.8781123 0.8136784 0.7847244 1.0153755
## JIC 0.8263883 0.7249868 0.7652683 0.7125675 0.7466685
## Jativoca 0.8399717 0.8231750 0.8830751 0.8695655 1.0336716
## FlotFlux 1.0481385 1.2435858 0.9921715 0.9966666 0.9170236
## Cubatão 0.9816624 0.8870705 0.8693311 0.9077804 0.9579464
## Guanabara 1.0153327 0.9163924 1.1587334 0.8798385 1.0183736
## Itaum 1.0411014 1.0882104 1.2264254 0.9201317 1.0490559
## Paraíso 0.8935319 0.8802846 0.9671582 1.1468588 0.9124968
## Rodovia.do.Arroz 1.0667741 0.9515544 0.9775324 0.8988758 0.9022578
## Unidade.de.Obras 0.9553626 0.8588707 0.8855205 1.1223232 0.8947807
## 6 7 8 9
## Bucarein 0.7796083 0.8685994 0.8014985 0.7286931
## JIC 0.7653743 0.6655116 0.7753183 0.6519916
## Jativoca 0.7890447 0.8270739 0.8957393 0.7561333
## FlotFlux 0.8989361 1.0278180 1.0021634 0.8892831
## Cubatão 0.8403391 1.1155606 0.9017742 0.9520278
## Guanabara 0.8630414 0.9749229 0.8967959 0.7981723
## Itaum 1.0693536 0.9053767 0.9492037 0.8453151
## Paraíso 0.8475731 0.8299103 0.9435927 0.9432292
## Rodovia.do.Arroz 1.1414902 0.9244003 1.0210485 0.8714656
## Unidade.de.Obras 0.8706195 0.9565415 0.9248555 0.7778360
summary(anohidroum.semnan.semout)

write.table(um, file = "F:/um.csv", sep = ";", dec = ",", col.names = NA, qmethod = "
double")

```

## APÊNDICE D – Rotina no R para o método de Redes Neurais Artificiais

```
## aplicando RNA para primeiro ano hidrológico outubro 2012 até setembro 2013
library(neuralnet)
##Preenchendo as falhas originais

#outubro 12, Paraíso e Unidade
outubro12 <- read.csv2("F:/outubro12.csv")
#Paraíso
nn1 <- neuralnet(Paraíso~JIC+Cubatão+FlotFlux+Guanabara+Bucarein+Rodovia.do.Arroz+Jat
ivoca+Itaum, data = outubro12, hidden = 15, threshold = 0.01, stepmax = 1e+07, err.fct
= "sse", act.fct = "logistic", linear.output = TRUE)
head(nn1$result.matrix)

##
## error 1
## reached.threshold 0.007612072666
## steps 610.0000000000000
## Intercept.to.1layhid1 0.882115257507
## JIC.to.1layhid1 -2.571773984628
## Cubatão.to.1layhid1 -1.724952370138

## $result.matrix

##
1
## error 0.00
7612072666
## reached.threshold 0.00
7593084777
## steps 610.00
0000000000
## Intercept.to.1layhid1 0.88
2115257507
## JIC.to.1layhid1 -2.57
1773984628
## Cubatão.to.1layhid1 -1.72
4952370138
## FlotFlux.to.1layhid1 0.03
2713075845
## Guanabara.to.1layhid1 -1.24
3013835028
## Bucarein.to.1layhid1 0.68
5081634749
## Rodovia.do.Arroz.to.1layhid1 -0.44
1032915331
## Jativoca.to.1layhid1 2.65
4959036829
## Itaum.to.1layhid1 0.90
9442301983
## Intercept.to.1layhid2 5.49
2169849854
## JIC.to.1layhid2 -0.87
9424455967
## Cubatão.to.1layhid2 0.75
8805159306
## FlotFlux.to.1layhid2 -1.65
5676560701
## Guanabara.to.1layhid2 -0.34
9484085829
## Bucarein.to.1layhid2 -0.61
5983787343
## Rodovia.do.Arroz.to.1layhid2 1.13
5068742065
## Jativoca.to.1layhid2 2.14
8935365762
## Itaum.to.1layhid2 -1.01
9129161948
## Intercept.to.1layhid3 10.27
3617126728
## JIC.to.1layhid3 0.77
4308456264
## Cubatão.to.1layhid3 -0.23
1071440229
## FlotFlux.to.1layhid3 -0.24
4006501314
## Guanabara.to.1layhid3 -0.63
7120464797
## Bucarein.to.1layhid3 -0.78
8943399113
## Rodovia.do.Arroz.to.1layhid3 0.55
8364129820
```

## Jativoca.to.1layhid3	-0.58	7132869848	
0231381060		## Itaum.to.1layhid6	-1.41
## Itaum.to.1layhid3	0.78	6035194626	
1139723765		## Intercept.to.1layhid7	-3.26
## Intercept.to.1layhid4	-0.12	6233928157	
8387251112		## JIC.to.1layhid7	-0.83
## JIC.to.1layhid4	-1.28	1541405509	
7430115947		## Cubatão.to.1layhid7	-0.36
## Cubatão.to.1layhid4	-0.54	6002487017	
5673887921		## FlotFlux.to.1layhid7	0.77
## FlotFlux.to.1layhid4	0.82	8410255252	
8112253177		## Guanabara.to.1layhid7	1.51
## Guanabara.to.1layhid4	1.72	3660798739	
7206204164		## Bucarein.to.1layhid7	-0.78
## Bucarein.to.1layhid4	-0.79	1994380121	
9426109534		## Rodovia.do.Arroz.to.1layhid7	0.41
## Rodovia.do.Arroz.to.1layhid4	-1.41	8968650380	
2763242754		## Jativoca.to.1layhid7	-1.42
## Jativoca.to.1layhid4	0.30	0988404105	
8261438314		## Itaum.to.1layhid7	0.67
## Itaum.to.1layhid4	0.15	7896039340	
6960068013		## Intercept.to.1layhid8	1.99
## Intercept.to.1layhid5	-6.35	9873756055	
7143012477		## JIC.to.1layhid8	0.20
## JIC.to.1layhid5	0.81	1755078050	
2683003779		## Cubatão.to.1layhid8	1.09
## Cubatão.to.1layhid5	0.73	3429771924	
9866813047		## FlotFlux.to.1layhid8	0.52
## FlotFlux.to.1layhid5	1.30	1725535841	
4382776738		## Guanabara.to.1layhid8	0.91
## Guanabara.to.1layhid5	-1.99	4813429888	
6951402464		## Bucarein.to.1layhid8	-1.03
## Bucarein.to.1layhid5	0.83	2144996875	
5244817827		## Rodovia.do.Arroz.to.1layhid8	-0.80
## Rodovia.do.Arroz.to.1layhid5	-1.85	1833479828	
4904063072		## Jativoca.to.1layhid8	0.13
## Jativoca.to.1layhid5	0.81	2903730728	
4147170377		## Itaum.to.1layhid8	-1.21
## Itaum.to.1layhid5	-0.23	2214660232	
6662829938		## Intercept.to.1layhid9	-5.06
## Intercept.to.1layhid6	2.90	4945537591	
6873717679		## JIC.to.1layhid9	0.88
## JIC.to.1layhid6	-0.31	5114335004	
0168548335		## Cubatão.to.1layhid9	0.40
## Cubatão.to.1layhid6	1.23	7838454089	
7985104438		## FlotFlux.to.1layhid9	0.27
## FlotFlux.to.1layhid6	1.89	5438854016	
6061406649		## Guanabara.to.1layhid9	-1.60
## Guanabara.to.1layhid6	-0.85	5740752520	
9559718913		## Bucarein.to.1layhid9	-0.04
## Bucarein.to.1layhid6	0.02	1879627469	
6086047983		## Rodovia.do.Arroz.to.1layhid9	-1.91
## Rodovia.do.Arroz.to.1layhid6	-0.42	1342788480	
0497007770		## Jativoca.to.1layhid9	0.91
## Jativoca.to.1layhid6	-0.26	9038751960	

## Itaum.to.1layhid9	0.61	3026466988	
9628433614		## Intercept.to.1layhid13	-5.07
## Intercept.to.1layhid10	-1.03	3153509607	
6113891974		## JIC.to.1layhid13	-0.01
## JIC.to.1layhid10	-6.04	2271961518	
3660709384		## Cubatão.to.1layhid13	1.24
## Cubatão.to.1layhid10	3.22	4439377177	
1373691952		## FlotFlux.to.1layhid13	-0.02
## FlotFlux.to.1layhid10	-0.04	5341482192	
8300813543		## Guanabara.to.1layhid13	-0.66
## Guanabara.to.1layhid10	-8.48	0140556701	
4946560228		## Bucarein.to.1layhid13	-0.91
## Bucarein.to.1layhid10	3.63	1339197700	
6483025663		## Rodovia.do.Arroz.to.1layhid13	-1.63
## Rodovia.do.Arroz.to.1layhid10	-0.09	2444955212	
6053954201		## Jativoca.to.1layhid13	1.66
## Jativoca.to.1layhid10	4.60	2579633027	
2217557074		## Itaum.to.1layhid13	0.13
## Itaum.to.1layhid10	-1.32	4562643423	
2667861188		## Intercept.to.1layhid14	5.13
## Intercept.to.1layhid11	3.70	4959946565	
3237669566		## JIC.to.1layhid14	-1.07
## JIC.to.1layhid11	-0.04	0145275947	
4185878489		## Cubatão.to.1layhid14	-0.79
## Cubatão.to.1layhid11	-0.66	2987170293	
0282434312		## FlotFlux.to.1layhid14	0.02
## FlotFlux.to.1layhid11	1.05	9045041404	
8221862062		## Guanabara.to.1layhid14	1.02
## Guanabara.to.1layhid11	0.02	9965297577	
0314574721		## Bucarein.to.1layhid14	-0.76
## Bucarein.to.1layhid11	-0.84	1091184999	
1448612731		## Rodovia.do.Arroz.to.1layhid14	0.40
## Rodovia.do.Arroz.to.1layhid11	-0.10	0854208812	
3345224844		## Jativoca.to.1layhid14	-1.69
## Jativoca.to.1layhid11	-1.30	1602955386	
7869249050		## Itaum.to.1layhid14	0.89
## Itaum.to.1layhid11	0.85	7717865657	
1162791679		## Intercept.to.1layhid15	2.16
## Intercept.to.1layhid12	-2.21	1368744712	
9876374173		## JIC.to.1layhid15	-0.02
## JIC.to.1layhid12	-1.09	4758351288	
7011438613		## Cubatão.to.1layhid15	2.69
## Cubatão.to.1layhid12	-0.16	1289999405	
4798989707		## FlotFlux.to.1layhid15	-1.87
## FlotFlux.to.1layhid12	-0.29	7771835899	
1452674877		## Guanabara.to.1layhid15	0.21
## Guanabara.to.1layhid12	-0.58	7292182572	
8630021494		## Bucarein.to.1layhid15	2.58
## Bucarein.to.1layhid12	0.24	6841628221	
9118058910		## Rodovia.do.Arroz.to.1layhid15	0.68
## Rodovia.do.Arroz.to.1layhid12	0.08	8493252940	
6737052173		## Jativoca.to.1layhid15	-1.99
## Jativoca.to.1layhid12	1.76	2955103313	
6480307327		## Itaum.to.1layhid15	-1.39
## Itaum.to.1layhid12	0.22	7837769893	



## Intercept.to.Paraíso	7.55	## 1layhid.8.to.Paraíso	8.54
1851811186		6773691360	
## 1layhid.1.to.Paraíso	2.81	## 1layhid.9.to.Paraíso	9.79
0800009917		0232301003	
## 1layhid.2.to.Paraíso	-8.75	## 1layhid.10.to.Paraíso	-1.57
8542557577		5187837436	
## 1layhid.3.to.Paraíso	-8.52	## 1layhid.11.to.Paraíso	-6.84
8065161718		7808692675	
## 1layhid.4.to.Paraíso	-3.40	## 1layhid.12.to.Paraíso	5.78
6742558745		9198790423	
## 1layhid.5.to.Paraíso	8.23	## 1layhid.13.to.Paraíso	12.90
1272560621		9733229823	
## 1layhid.6.to.Paraíso	7.44	## 1layhid.14.to.Paraíso	-3.68
6370164901		3501910450	
## 1layhid.7.to.Paraíso	10.42	## 1layhid.15.to.Paraíso	4.87
4476142556		6619888934	

```
## attr(,"class")
## [1] "nn"
```

```
dados.Valida<-outubro12[, ]
previsao<-compute(nn1,dados.Valida[, c(2,5,4,6,1,9,3,7)])
```

## \$net.result		## [15,] -0.0008812447836
## [1,]	0.2284281825145	## [16,] 0.1301169883692
## [2,]	4.5712515956162	## [17,] 0.0053799384221
## [3,]	0.0053799384221	## [18,] 0.0053799384221
## [4,]	-0.2956043926625	## [19,] 0.0053799384221
## [5,]	1.5989169762699	## [20,] 0.0697062650310
## [6,]	0.4606822979733	## [21,] 0.0053799384221
## [7,]	-0.0011483570888	## [22,] 5.0292118192667
## [8,]	0.4670191907660	## [23,] 74.9598079350102
## [9,]	0.0053799384221	## [24,] 1.2745376674047
## [10,]	0.0053799384221	## [25,] 0.2230825970224
## [11,]	7.6914009749784	## [26,] 11.1057496137321
## [12,]	25.8298538365980	## [27,] 22.6099794662829
## [13,]	40.9715632415012	## [28,] 0.0053799384221
## [14,]	6.5855821889409	## [29,] 0.2316480429978

```
## [30,] -0.0012316840805
## [31,] -0.0142263189710
```

```
write.table(previsao, file = "F:/previsaooutubro12-paraíso.csv", sep = ";", dec = ",",
col.names = NA, qmethod = "double")
```

## APÊNDICE E – Rotina no R para o método de Regressão Linear Múltipla

```
## Regressão Linear Múltipla para o Primeiro Ano Hidrológico

anohidroum.semnan.semout <- read.csv2("F:/anohidroum-semnan-semout.csv")
summary(anohidroum.semnan.semout)

attach(anohidroum.semnan.semout)
#estações que tem falhas: Paraíso(8), Guanabara(1), Cubatão(1) e Unidade (73)
##preenchendo a estação Cubatão
options(na.action = na.exclude)
estacaocubatao <- lm(Cubatão~FlotFlux+Rodovia.do.Arroz)
summary(lm(Cubatão~FlotFlux+Rodovia.do.Arroz))

##
## Call:
## lm(formula = Cubatão ~ FlotFlux + Rodovia.do.Arroz)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -32.435  -0.617  -0.617   0.479  31.922
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.61667    0.33439   1.844   0.066 .
## FlotFlux       0.54112    0.05148  10.510 < 2e-16 ***
## Rodovia.do.Arroz 0.36255    0.05166   7.017 1.12e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.704 on 361 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.7736, Adjusted R-squared:  0.7724
## F-statistic: 616.9 on 2 and 361 DF, p-value: < 2.2e-16

valores= fitted(estacaocubatao)
valores[1]<-predict.lm(estacaocubatao,anohidroum.semnan.semout[365,2:10])
valores[1]

##      1
## 1.44211
```

## REFERÊNCIAS

ABBOT, J.; MAROHASY, J. Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. **Atmospheric Research**, 2014, v. 138, p. 166-178.

AGÊNCIA NACIONAL DE ÁGUAS (ANA). Diretrizes e análises recomendadas para a consistência de dados pluviométricos. Brasília: Superintendência de Gestão da Rede Hidrometeorológica – Sgh, 2012. 18 p. Disponível em: <<http://arquivos.ana.gov.br/infohidrologicas/cadastro/DiretrizesEAnalisesRecomendaDasParaConsistenciaDeDadosPluviometricos-VersaoJan12.pdf>>. Acesso em: 16 set. 2016.

AYOADE, J. O. **Introdução à climatologia para os trópicos**. 5<sup>a</sup> ed. – Rio de Janeiro: Bertrand Brasil, 1998. 332p.

BARBOSA J. D. R.P, Notas de Aulas – Disciplina de Hidrologia Aplicada, Escola Politécnica, Departamento de Recursos Hídricos e Meio Ambiente, 2006, UFRJ, Rio de Janeiro.

BÁRDOSSY, A.; PEGRAM, G. Infilling missing precipitation records – A comparison of a new copula-based method with other techniques. **Journal of Hydrology**, 2014, v. 519, p. 1162-1170.

BARNSTON, A. G. 1992. Correspondence Among the Correlation, RMSE, and Heidke Forecast Verification Measures; Refinement of the Heidke Score. **Weather and Forecasting** 7: 699-709.

BELAYNEH, A.; ADAMOWSKI, J. Standard precipitation index drought forecasting using neural networks, wavelet neural networks and support vector regression, **Journal of applied computational intelligence and soft computing**, 2012, p. 1-13.

BERTONI, J. C.; TUCCI, Carlos E.M. Precipitação. In: TUCCI, Carlos E.M. (Org.). **Hidrologia - Ciência e Aplicação**. Porto Alegre: Editora da UFRGS, 1997, p. 177-231.

BLAINSKI, E.; GARBOSSA, L.H.P.; ANTUNES, E.N. Estações hidrometeorológicas automáticas: recomendações técnicas para instalação. Florianópolis: Epagri, 2012, 43p.<[http://ciram.epagri.sc.gov.br/recomendacoes\\_tecnicas\\_para\\_instalacao\\_de\\_estacoes.pdf](http://ciram.epagri.sc.gov.br/recomendacoes_tecnicas_para_instalacao_de_estacoes.pdf)>. Acesso em: 25 de abr. de 2016.

BOLZAN, Lucas Tedesco. **Desenvolvimento de software para preenchimento de falhas de séries pluviométricas diárias**. 2014. 86 f. TCC (Graduação) - Curso de Engenharia Civil, Universidade do Estado de Santa Catarina, Joinville, 2014.

Disponível em:

<<http://sistemabu.udesc.br/pergamumweb/vinculos/000003/00000351.pdf>>. Acesso em: 20 set. 2016.

BRAGA, S. M. **Análise do potencial de utilização de sensores automáticos hidroambientais**: estudo de caso da bacia do Rio Barigui. Curitiba, PR, 2005. Originalmente apresentada como dissertação de mestrado, Universidade Federal do Paraná.

CAMARGOS, V. P.; CÉSAR, C. C.; CAIAFFA, W. T.; XAVIER, C. C.; PROIETTI, F. A. Imputação múltipla e análise de casos completos em modelos de regressão logística: uma avaliação prática do impacto das perdas em covariáveis. **Cardeneta de Saúde Pública**, v. 26, 2299-2313, dez, 2011. Disponível em:

<<http://www.scielo.br/pdf/csp/v27n12/03.pdf>>. Acesso em: 07 jun. 2017.

CAMPOS, F. F. 2007. **Algoritmos numéricos. 2 ed.** Rio de Janeiro: LTC.

CANHOLI, Aluísio Pardo. Drenagem urbana e controle de enchentes. São Paulo: Oficina de Textos, 2005. 302 p.

CARVALHO, D. F. de; SILVA, L. D. B. da. Apostila de Hidrologia: Precipitação. Rio de Janeiro: Universidade Federal Rural do Rio de Janeiro, 2006. cap 2. p.11-14.

Disponível em:

<<http://www.ufrj.br/institutos/it/deng/leonardo/downloads/APOSTILA/HIDRO-Cap4-PPT.pdf>>. Acesso em: 08 set. 2016.

CAVALCANTI, I. F. A.; FERREIRA, N. J.; DIAS, M. A. F. da S.; SILVA, M. G. A. J. da (Org.). **Tempo e clima no Brasil**. São Paulo: Oficina de Textos, 2009.

CHATTOPADHYAY, S.; CHATTOPADHYAY, G. Comparative study among different neuralnet learning algorithms applied to rainfall time series. **Meteorological Applications**. 2008, v. 15, p. 273–280. Disponível em:

<<http://onlinelibrary.wiley.com/doi/10.1002/met.71/pdf>>. Acesso em: 21 set. 2016.

CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. **Geoscientific Model Development**, v.7, 1247-1250 p, 2014. Disponível em:<<https://doi.org/10.5194/gmd-7-1247-2014>>. Acesso em: 31 maio 2017.

CHOW, V.T.; MAIDMENT, D.R.; MAYS, L.W. Applied Hydrology. Santafé de Bogotá: McGraw-Hill Interamericana, 1988.

CLIMANÁLISE. **Boletim de Climatologia de Precipitação e Temperatura.** Cachoeira Paulista –SP: INPE/CPTEC, v. 19, n. 1, 2009. Disponível em: <<http://climanalise.cptec.inpe.br/~rclimanl/boletim/cliesp10a/chuesp.html>>. Acesso em: 10 jun. 2016.

COELHO, Gustavo de Almeida. **Análise de não estacionariedade da precipitação em São Paulo / SP e implicações sobre os sistemas de drenagem urbana.** São Paulo, 2014. 202 p. Dissertação de Mestrado (Mestre em Engenharia) – Escola Politécnica.

CONTI, Fátima. Regressão e Correlação. 10 p. Disponível em: <<http://www.cultura.ufpa.br/dicas/>>. Acesso em: 26 maio 2017.

CORDANO, E.; ECCEL, E. RMAWGEN: RMAWGEN (R Multi-site Autoregressive Weather GENerator): a package to generate daily time series from monthly mean values. R package version 1.2.5.3. Disponível em: <<https://cran.r-project.org/package=RMAWGEN>>. Acesso em: 10 ago. 2016.

COSTA, R. L.; SILVA, F. D. dos S.; SARMANHO, G. F.; LUCIO, P. S. Imputação multivariada de dados diários de precipitação e análise de índices de extremos climáticos. **Revista Brasileira de Geografia Física**, 2012, v. 3, p. 661-675.

COULIBALY; P.; EVORA; N. D. Comparison of neural network methods for infilling missing daily weather records. **Journal of Hydrology**, v. 341, p. 27-41. 2007.

DEUS, Evelyne Braga de. **Estudo analítico entre redes neurais com função base radial e o perceptron multicamadas em um caso específico:** diagnóstico de falhas incipientes em transformadores de potência. Ouro Preto, 2012. Trabalho de graduação, Universidade Federal de Ouro Preto.

DEFESA CIVIL, Delimitação das Bacias Hidrográficas de Joinville e Localização das Estações da Rede de Monitoramento de Joinville – SC. Joinville, 2016.

DEFESA CIVIL, Localização das estações pluviométricas da rede de monitoramento da Defesa Civil. **Joinville, 2016.**

DEO, R. C.; ŞAHIN, M. Application of the Artificial Neural Network model for prediction of monthly standardized precipitation and evapotranspiration index using

hydrometeorological parameters and climate indices in eastern Australia. **Atmospheric Research**, 2015, v. 161-162, p. 65-81. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0169809515001003>>. Acesso em: 08 ago. 2016.

DEPINÉ, H.; CASTRO, N. M. dos. R.; PINHEIRO, A.; PEDROLLO, O. Preenchimento de falhas de dados horários de precipitação utilizando redes neurais artificiais. **Revista Brasileira de Recursos Hídricos**, v. 19, n. 1, p.51-63, 2013. Disponível em: <<https://www.abrh.org.br/SGCv3/index.php?PUB=1&ID=161&SUMARIO=4347>>. Acesso em: 08 ago. 2016

DIAS, A. J. R., ALBIERI, S. Uso de imputação em pesquisas domiciliares. VIII Encontro Nacional de Estudos Populacionais. **Anais**, v. 1: Informação Demográfica, Fecundidade, Demografia Histórica, p. 11:26, São Paulo: ABEP, 1992.

DIAS, M. A. F. S. Eventos extremos: Fenômenos Naturais ou Consequências das Ações Humanas?. Universidade de São Paulo: Departamento de Ciências Atmosféricas, 2011. Disponível em: <<http://www.abas.org/eventosextrmos/iag.pdf>>. Acesso em: 08 set. 2016.

DNAEE - DEPARTAMENTO NACIONAL DE ÁGUAS E ENERGIA ELÉTRICA. Glossário de termos hidrológicos. Brasília, Ministério de Minas e Energia, 1976.

ELSHORBAGY, A. ;PANU , U. S.; SIMONOVIC, S. P. Group-based estimation of missing hydrological data: I. approach and general methodology. **Hydrological Sciences Journal**, 2000, v. 45, n. 6, p. 849-866.

EM-DAT - The International Disaster data Base. **Centre for Research on the Epidemiology of Disasters – CRED**. Université Catholique de Louvain, Belgium. Disponível em: <<http://www.emdat.be/database>>. Acesso em: 22 maio. 2016.

EISCHEID, J. K.; BAKER, C. B.; KARL, T. R.; DIAZ, H. F. The quality control of long-term climatological data using objective data analysis. **Journal of Applied Meteorology**, v. 34, 2787–2795 p. 1995. Disponível em: <[http://journals.ametsoc.org/doi/abs/10.1175/1520-0450\(1995\)034%3C2787%3ATQCOLT%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0450(1995)034%3C2787%3ATQCOLT%3E2.0.CO%3B2)>. Acesso em: 15 jul. 2016.

——, PASTERIS, P. A.; DIAZ, H. F.; PLANTICO, M. S.; LOTT, N. J. Creating a serially complete, national daily time series of temperature and precipitation for the western United States. **Journal of Applied Meteorology**, v. 39, 1580–1591p. 2000. Disponível em: < <http://journals.ametsoc.org/doi/full/10.1175/1520->

0450%282000%29039%3C1580%3ACASCND%3E2.0.CO%3B2 >. Acesso em: 15 jul. 2016.

FERNANDES, Anita M. da Rocha., “Inteligência Artificial Noções Gerais”, Florianópolis: Bookstore, 2003.

FERNANDEZ, Marfiza. Negrine.; **Preenchimento de falhas em series temporais**. Rio Grande, 2007. 106 p. Dissertação de Mestrado (Mestre em Engenharia Oceânica) – FURG. Disponível em: <<http://www.engenhariaoceanica.furg.br/arquivos/fernandez,m.n.pdf>>. Acesso em: 30 jun. 2017.

FERREIRA, C., J.; PATINO, M., C. O que realmente significa o valor-p?. **Sociedade Brasileira de Pneumologia e Tisiologia**. 2015. Disponível em: <[http://www.scielo.br/pdf/jbpneu/v41n5/pt\\_1806-3713-jbpneu-41-05-00485.pdf](http://www.scielo.br/pdf/jbpneu/v41n5/pt_1806-3713-jbpneu-41-05-00485.pdf)>. Acesso em: 04 ago. 2017.

FLORIANÓPOLIS. CIRAM. Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina. Rede de monitoramento hidrometeorológico para a Bacia Hidrográfica do Rio Camboriú – SC. 2012. Disponível em: <<http://www.comitecamboriu.com.br/wp-content/uploads/2014/11/CIRAM-2011.pdf>>. Acesso em: 28 abr. 2016.

FRITSCH, S.; GUENTHER, F. **neuralnet: Training of Neural Networks**. R package version 1.33, 2016. Disponível em: <<https://CRAN.R-project.org/package=neuralnet>>. Acesso em: 27 abr. 2017.

HIRSCH, R. M.; MAIDMENT, D. R. Statistical analysis of hydrologic data. In: MAIDMENT, D. R. (Org.). **Handbook of Hydrology**. Nova Iorque: Editora McGRAW-HILL, 1992. Cap 17.

HOUGHTON MIFFLIN (Estados Unidos da América). Houghton Mifflin (Ed.). The American Heritage: Science Dictionary. 2002. Disponível em: <<http://www.dictionary.com/browse/hydrometeorology>>. Acesso em: 20 maio 2016.

HYNDMAN, R. J. forecast: Forecasting functions for time **series and linear models**. R package version 8.1, 2017. Disponível em: <<http://github.com/robjhyndman/forecast>>. Acesso em: 01 set. 2016.

IHAKA, R.; GENTLEMAN, R. R: a language for data analysis and graphics. **Journal of Computational and Graphical Statistics**, v. 5, n. 3, p. 299-314, 1996. Disponível

em: <<https://www.stat.auckland.ac.nz/~ihaka/downloads/R-paper.pdf>>. Acesso em: 19 set. 2016.

INMET (Instituto Nacional de Meteorologia), 2011, Rede de Estações Meteorológicas Automáticas do INMET. In: Nota Técnica n. 001/2011/SEGER/ LAIME/CSC/INMET, Brasília, MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. JOINVILLE. Defesa Civil. Defesa Civil de Joinville. **Rede de monitoramento das estações meteorológicas de Joinville/SC**. 2011. Disponível em: <<https://defesacivil.joinville.sc.gov.br/monitoramento>>. Acesso em: 20 abr. 2016.

KAMEL, H.F., SLIMANI, M., CUDENNEC, C. A comparison of three geostatistical procedures for rainfall network optimization. In **International Renewable Energy Congress**, nov. 2010, Tunísia. <<http://2010.irec-conference.com/papers/REA/Paper%20ID51.pdf>>. Acesso em: 28 abr. 2016.

KAR, A. K.; LOHANI, A.K.; GOEL, N.K.; ROY<sup>a</sup>, G.P. Rain gauge network design for flood forecasting using multi-criteria decision analysis and clustering techniques in lower Mahanadi river basin, India. **Journal of Hydrology: Regional Studies**, 2015, v. 4, parte B, p. 313-332. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2214581815000877>>. Acesso em: 02 maio 2016.

KISTLER, Robert et al. The NCEP- NCAR 50-year reanalysis: monthly means CD-ROM and documentation. **Bulletin of the American Meteorological Society**, 2001, v. 82, n. 2, p. 247-268.

LANNA, A. E.; TUCCI, Carlos E. M. Elementos de Estatística e Probabilidades: Introdução. In: TUCCI, Carlos E. M. (Org.). **Hidrologia - Ciência e Aplicação**. 2. ed. Porto Alegre: Editora da Universidade/UFRGS: ABRH, 1997. Cap. 4. p. 80-81.

LIMA, M. V.; NUNES, A. B. Preenchimento de falhas de dados mensais de precipitação: comparação básica pontual para Pelotas- RS. **XVII Congresso Brasileiro de Meteorologia**, 2012, Gramado, RS.

MAIA, B. G. O.; KLOSTERMANN, D.; RIBEIRO, J. M. G., SIMM, M.; OLIVEIRA, T.M.N., BARROS, V. G. **Bacias Hidrográficas da Região de Joinville**. Blumenau: Editora 3 de Maio Ltda, 2014.

MCCULLOCH, W.S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 1943, v. 5, n. 4, p. 115-133. Disponível em: <<http://link.springer.com/article/10.1007%2fbf02478259>>. Acesso em: 20 set. 2016.



MEKANIK, F.; IMTEAZ, M. A.; GATO-TRINIDAD, S.; ELMAHDI, A. Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. **Journal of Hydrology**, 2013, v. 503, p. 11-21.

MELLO, M. P.; PETERNELLI, L. A. Algo sobre o R: Introdução. In: MELLO, M. P.; PETERNELLI, L. A. Conhecendo o R: Uma visão mais que Estatística. Viçosa: UFV, 2013. Cap. 1. p. 15-24.

MELLO, Yara de. **Distribuição de precipitação no município de Joinville (SC) e sua relação com a incidência de leptospirose**. Joinville, 2015. 92 p. Dissertação de Mestrado (Mestre em Saúde e Meio Ambiente) – UNIVILLE

MELLO, Y. R.; OLIVEIRA, T. M. N. de; VAZ, C.; NASS, D. N. Distribuição de precipitação anual na bacia hidrográfica do rio Cubatão Norte. **Revista GEONORTE**, Manaus, v.1, n. 5, Edição Especial 2, nov. 2012. Disponível em: <http://www.periodicos.ufam.edu.br/revista-geonorte/article/view/2392>>. Acesso em: 20 set. 2016

MISLAN; HAVILUDDIN; HARDWINARTO, S.; SUMARYONO; AIPASSA, M. Rainfall monthly prediction based on artificial neural network: a case study in Tenggara Station, East Kalimantan – Indonesia. **Procedia Computer Science**, 2015, v. 59, p. 142-151.

MONTEIRO, M. A. Caracterização climática do Estado de Santa Catarina: uma abordagem dos principais sistemas atmosféricos que atuam sobre o ano. **Geosul**, Florianópolis, v.16, n.31, p.69-78, jan./jun., 2001.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 2009, Rio de Janeiro: LTC, ed. 4.

MWALE, F.D.; ADELOYE, A. J.; RUSTUM, R. Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi – A self organizing map approach. **Physics and Chemistry of the Earth**, v. 50-52, p. 34-43, 2012.

NAGHETTINI, M.; PINTO, E. J. de A. **Hidrologia Estatística**. Belo Horizonte: CPRM, 2007. 552 p. Disponível em: <http://www.cprm.gov.br/publique/Hidrologia/Mapas-e-Publicacoes/Livro-%22Hidrologia-Estatistica%22-981.html>>. Acesso em: 15 set. 2016.

NASTOS, P.; PALIATSOS, A.; KOUKOULETSOS, K.; LARISSI, I.; MOUSTRIS, K. Artificial neural networks modeling for forecasting the maximum daily total

precipitation at Athens, Greece. **Atmospheric Research**. 2014, v. 144, p. 141–150. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0169809513003360>>. Acesso em: 08 ago. 2016.

NUNES, L. N.; KLUCK, M. M.; FACHEL, J. M. G. Uso da imputação múltipla de dados faltantes: uma simulação utilizando dados epidemiológicos. **Caderneta de Saúde Pública**, Rio de Janeiro, v. 25, n. 2, p. 268-278, Fev. 2009. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-311X2009000200005&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2009000200005&lng=en&nrm=iso)>. Acesso em 11 Abr. 2017.

—KLUCK, M. M.; FACHEL, J. M. G. Comparação de métodos de imputação única e múltipla usando como exemplo um modelo de risco para mortalidade cirúrgica. **Revista Brasileira Epidemiologia**. 2010, vol.13, n.4, pp.596-606. ISSN 1415-790X. Disponível em: <<http://dx.doi.org/10.1590/S1415-790X2010000400005>>. Acesso em 15 Ago. 2016

NKUNA, T. R.; ODIYO, J. O. Filling of missing rainfall data in Luvuvhu River Catchment using artificial neural networks. **Physics and chemistry of the Earth**, 2011, v. 36, p. 830-835

OLIVEIRA, L. F.C.; FIOREZA, A. P.; MEDEIROS, A. M. M.; SILVA, M. A. S. Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 2010, v. 14, n. 11, p. 1186-1192.

OLIVEIRA, Tatiana Cristina de. **Proposta para monitorar o erro de previsão de demanda de um produto aplicado à refrigeração**. TCC (Graduação) – Curso de Engenharia de Produção, Universidade do Estado de Santa Catarina, Joinville, 2015.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Estratégia Internacional para Redução de Desastres – ONU/EIRD (2009), Terminologia: termos básicos relativos à redução do risco de desastres ONU/EIRD, 2004. Disponível em: <<http://www.smad.rs.gov.br/downloads/documentos/ONU-UNISDR%20-%20Glossario%20EIRD%20Estrategia%20Internac%20Reducao%20Desastres.pdf>>. Acesso em: 10 jun. 2016.

PARRISH, D. F.; DERBER, J. C. The National Meteorological Center's spectral statistical-interpolation analysis system. **American Meteorological Society**, 1992, p. 1747-1763.

PENEREIRO, J.C.; ORLANDO, D.V. Análise de tendências em séries temporais anuais de dados climáticos e hidrológicos na bacia do Rio Parnaíba entre os Estados do Maranhão e Piauí/ Brasil. **Revista Geográfica Acadêmica**. 2013, v. 7, n. 2, p. 5-21.

PINHEIRO, A.; GRACIANO, R. L. G. G.; SEVERO, D. L. Tendência das séries temporais de precipitação da região sul do Brasil. **Revista Brasileira de Meteorologia**, v. 28, n. 3, 281-290, 2013. Disponível em: <<http://dx.doi.org/10.1590/S0102-77862013000300005>>. Acesso em: 20 ago. 2016.

PORTELA, M. M.; QUINTELA, A.C.; SANTOS, J.F.; VAZ, C.; MARTINS, C. Tendências em séries temporais de variáveis hidrológicas. **Recursos Hídricos**. 2011, v. 32, p. 43-60.

R DEVELOPMENTCORE TEAM (2017). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://www.R-project.org/>>.

REBOITA, M. S.; GAN, M. A.; Da ROCHA, R. P.; AMBRIZZI, T. REGIMES DE PRECIPITAÇÃO NA AMÉRICA DO SUL: UMA REVISÃO BIBLIOGRÁFICA. **Revista Brasileira de Meteorologia**, São José dos Campos, v. 25, n. 2, p.185-204, out. 2009. Disponível em: <<http://www.scielo.br/pdf/rbmet/v25n2/a04v25n2.pdf>>. Acesso em: 16 ago. 2016.

RENNÓ, C. D.; BORMA, L. de S. **Métodos estatísticos aplicados à hidrologia**. 2016. Disponível em: <<http://www.dpi.inpe.br/~camilo/prochidr/>>. Acesso em: 14 set. 2016.

RIBEIRO, S.; CAINETA, J.; COSTA, A. C.; HENRIQUES, R.; SOARES, A.; 2016. Detection of inhomogeneities in precipitation time series in Portugal using direct sequential simulation. **Atmospheric Research**, v. 171, p. 147-158.

R.UCLA: **Statistical Consulting Group**. Disponível em: <[http://www.ats.ucla.edu/stat/r/faq/R\\_pmm\\_mi.htm](http://www.ats.ucla.edu/stat/r/faq/R_pmm_mi.htm)>. Acesso em: 15 out. 2016.

SAHA, Suranjana et al. The NCEP climate forecast system reanalysis. **American Meteorological Society**, 2010, p. 1016- 1057.

ŞAHİN, M., Kaya, Y., Uyar, M. Comparison of ANN and MLR models for estimating

solar radiation in Turkey using NOAA/AVHRR data. **Advances in Space Research**, 2013, v. 51, p. 891–904.

SEARCY, J. K.; HARDINSON, C. H. General Surface-Water Techniques. In: NOLAN, T. B. (Diretor). **Double-Mass Curves**. Washington: United States Government Printing Office, 1960, p. 31-65.

SILVA, Maria Joseane Cruz da. **Imputação Múltipla: comparação e eficiência em experimentos multiambientais**. Piracicaba, 2012. 122 p. Dissertação de mestrado (Mestre em Ciências) – USP.

SILVEIRA, Paola da Costa. **Precipitação pluviométrica na Bacia Hidrográfica do Rio Vacacaí, RS**. Santa Maria, 2012. 54 p. Dissertação (Mestrado em Geografia e Geociências). Programa de Pós-Graduação em Geografia e Geociências, Universidade Federal de Santa Maria. 2012.

SILVEIRA, Wivian Nereida. **Análise histórica de inundação no município de Joinville – SC, com enfoque na bacia hidrográfica do rio Cubatão do Norte**. 2008. 165 f. Dissertação (Mestrado) - Curso de Programa de Pós-graduação em Engenharia Ambiental, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2008. Disponível em: <<https://repositorio.ufsc.br/xmlui/handle/123456789/91644>>. Acesso em: 30 maio 2017.

SIMON, F. W. ; PICKBRENNER, K. ; MARCUZZO, F. F. N. Estudo do regime hídrico pluvial e fluvial em bacia hidrográfica com precipitação homogênea. In: **XX Simpósio Brasileiro de Recursos Hídricos**, 2013, Bento Gonçalves. **Anais**. Bento Gonçalves: ABRH, 2013. Artigos, p. 1-8. CD-ROM. Acesso em: 20 set. 2016.

SOUZA, G. P., MIRANDA, R., SAMOBYL, R. W. Métodos Simplificados em Previsão Empresarial. Rio de Janeiro: Editora Ciência Moderna, 2008.

SOUZA, R. R. Análise Preliminar da Preferência das Chuvas na Amazônia Mato-Grossense no Período de 2004 a 2007 (Janeiro, Fevereiro e Março). **Revista Geográfica Acadêmica**. Brasil, v. 2. n. 1, p. 56- 72, 2008. Disponível em: <<http://132.248.9.34/hevila/Revistageograficaacademica/2008/vol2/no1/6.pdf>>. Acesso em: 20 set. 2016.

SUSSMAN, G. J.; STEELE JR., G. L. Scheme: A interpreter for extended lambda calculus. Estados Unidos. **Higher-Order and Symbolic Computation**, 1998, v. 11, n. 4, p. 405-439. Disponível em: < <https://cs.au.dk/~hosc/local/HOSC-11-4-pp405-439.pdf>>. Acesso em: 20 set. 2016.

TEEGAVARAPU, R. S. V.; CHANDRAMOULI V. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. **Journal of Hydrology**, v. 312, p. 191-206. 2005.

TEEGAVARAPU, R. S.V.; TUFAIL, M.; ORMSBEE, L. Optimal functional forms for estimation of missing precipitation data. **Journal of Hydrology**, v.374, p. 106-115. 2009.

TUCCI, C. E. M. (Org.). Hidrologia – Ciência e Aplicação. ABRH/EDUSP/Editora da UFRGS, Porto Alegre, 1997.

TUCCI, C. E. M. **Recursos hídricos e meio ambiente**. 2012. Disponível em: <<http://rhama.net/wordpress/?cat=7>>. Acesso em: 04 set. 2016.

UFSC - Universidade Federal de Santa Catarina. Centro Universitário de Estudos e Pesquisas sobre Desastres. **Atlas brasileiro de desastres naturais 1991 a 2010: volume Santa Catarina**/ Centro Universitário de Estudos e Pesquisas sobre Desastres. Florianópolis: CEPED UFSC, 2011. 20 p. Disponível em: <<http://150.162.127.14:8080/e-soll.ceped.atlas.aspx>>. Acesso em: 10 jun. 2016.

USGS – United States Geological Survey, 1960: Double-Mass Curves. Manual of Hydrology: Part I. General Surface-Water Techniques, by James K. Searcy and Clayton H. Hardison, United States Department of the Interior.

VAN BUUREN, S.; OUDSHOORN, G. K. 2011. MICE: Multivariate Imputation by Chained Equations in R. **Journal of Statistical Software**, v. 45, p. 1-67. Disponível em: <<http://www.jstatsoft.org/v45/i03/>>. Acesso em 04 out. 2016.

VERZANI, John. Overview, Installation. In: VERZANI, John. Getting Started with RStudio. Estados Unidos: O'reilly Production Service, 2011. Cap. 1. p. 1-11. Disponível em: <<http://www.cs.utexas.edu/~cannata/dataVis/Class Notes/Getting Started with RStudio.pdf>>. Acesso em: 18 set. 2016.

WANG, S.; FENG, J.; LIU, G. 2013. Application of seasonal time series model in the precipitation forecast. **Mathematical and Computer Modelling**, v. 58, p. 677-683.

WANDERLEY, H. S.; AMORIM, R. F. C.; CARVALHO, F. O. Interpolação espacial de dados médios mensais pluviométricos com redes neurais artificiais. **Revista Brasileira de Meteorologia**, v. 29, n. 3. 389-396 p. 2014. Disponível em: <<http://dx.doi.org/10.1590/0102-778620130639>>. Acesso em : 07 jun. 2017.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. R package. Springer-Verlag New York, 2009.

WILMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance.

**Climate Research**, v. 30, 79-82 p., 2005. Disponível em: <<http://www.int-res.com/abstracts/cr/v30/n1/p79-82/>>. Acesso em: 31 maio 2017.

YUANCHAO, X. **hyfo: Hydrology and Climate Forecasting**. R package version 1.3.9, 2017. Disponível em: <<https://CRAN.R-project.org/package=hyfo>>. Acesso em: 29 ago. 2016.

ZEILHOFER, P.; LIMA, E. B. R.; SANTOS, F. M.; RIGO Júnior, L. O. Um ambiente SIG para modelagem integrada da qualidade da água utilizando Qual2e. **Caminhos de Geografia**, v.8, n.10, p.107-125, 2003.