



UNIVERSIDADE FEDERAL DE ALAGOAS  
CENTRO DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS HÍDRICOS  
E SANEAMENTO

RICARDO LUIS BARRETO DE SALES REIS

**ESTUDO DE TELECONEXÕES CLIMÁTICAS PARA A  
PREVISÃO HIDROLÓGICA DE LONGO PRAZO DA BACIA  
DO RIO PIANCÓ NO SEMIARIDO PARAIBANO**

Maceió  
2015

RICARDO LUIS BARRETO DE SALES REIS

**ESTUDO DE TELECONEXÕES CLIMÁTICAS PARA A  
PREVISÃO HIDROLÓGICA DE LONGO PRAZO DA BACIA  
DO RIO PIANCÓ NO SEMIARIDO PARAIBANO**

Dissertação apresentada ao Programa de Pós-Graduação em Recursos Hídricos e Saneamento, Centro de Tecnologia da Universidade Federal de Alagoas, como requisito para obtenção do título de Mestre em Recursos Hídricos e Saneamento.

Orientador: Prof. Dr. Adriano Rolim da Paz

Co-orientador: Prof. Dr. Carlos Ruberto Fragoso Jr.

Maceió  
2015

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

- R375e Reis, Ricardo Luis Barreto de Sales.  
Estudo de teleconexões climáticas para a previsão hidrológica de longo prazo da  
bacia do rio Piancó no semiárido paraibano / Ricardo Luis Barreto de Sales Reis. –  
2021.  
168 f. : il. color.
- Orientador: Adriano Rolim da Paz.  
Co-orientador: Carlos Ruberto Fragoso Jr.  
Dissertação (Mestrado em Recursos Hídricos e Saneamento) – Universidade  
Federal de Alagoas. Centro de Tecnologia. Maceió, 2015.
- Bibliografia: f. 158-164.  
Anexos: f. 165-168.
1. Precipitação (Meteorologia) - Previsão. 2. Índices climáticos. 3. Redes neurais  
artificiais. 4. Variabilidade climática - Teleconexões. I. Título.

CDU: 556.18

*Aos meus pais, Maria José e Ronaldo Luis  
As minhas irmãs, Marília e Mayara  
As pessoas que tiveram compreensão e paciência durante  
a elaboração deste trabalho*

## **AGRADECIMENTOS**

Agradeço a Deus e a Nossa Senhora por terem me segurado durante esse mestrado, ao qual não me deixaram desistir, mesmo quando não acreditava que fosse conseguir atingir o objetivo final.

Agradeço a minha mãe primeiramente pelo seu amor, pelo apoio, pela confiança, pelo cuidado e esforço que sempre teve para proporcionar a mim e a minhas irmãs a melhor qualidade de vida.

Agradeço ao meu pai por ter me direcionado sempre no melhor caminho, por sempre me aconselhar, por ter estado sempre disposto a me ajudar no que precisasse em qualquer momento.

Agradeço as minhas irmãs pelas grandes reflexões que sempre temos, pelas idéias compartilhadas, pelos conselhos mútuos, pelas cobranças que sempre me fazem acordar para o que falta ser feito, enfim por estarem por perto.

Agradeço a minha avó por estar sempre rezando por mim, intercedendo por mim, pela fé imensurável dela que uni a família.

Agradeço aos meus tios e tias (que são muitos), e em especial ao meu tio Ubirajara a quem sempre me deu apoio, que confiou desde o início na minha jornada, por sempre estar disposto para que eu compartilhasse os conhecimentos de cada etapa no curso, de cada matéria, de cada aula, enfim, das nossas idéias.

Agradeço aos professores Ruberto, Marllus, Antônio Netto, Karina, ao qual proporcionaram grandes aulas e além de tudo foram mais que compreensivos com os problemas dos alunos e sempre demonstraram dispostos a ajudar.

Agradeço em especial ao professor Adriano da Paz pela a grande paciência e serenidade de ter me orientado neste trabalho, por ter confiado em mim durante o projeto, por ter passado a sua experiência, e por me ajudar a encontrar um direcionamento numa área do conhecimento

Agradeço pôr fim aos meus amigos que tive nesse mestrado, ao qual criei fortes vínculos, que me ajudaram, que me escutaram e se importaram comigo, com quem eu sabia que podia e tenho certeza que sempre poderei contar, Angélica, Altair, Andressa, Juliana, Danilo, Diogo. Agradeço em especial ao meu grande amigo Henrique, que foi muito importante durante essa jornada, ao qual sempre estive presente, sempre solícito, sempre irmão.

Agradeço enfim a todos que participaram e caminharam junto comigo nesse período do mestrado.

*Não sabendo que era impossível, ele foi lá  
e fez.*

## RESUMO

Avalia-se recentemente, de forma mais intensa, a existência de uma ligação muito forte entre a circulação em grande escala e variáveis climáticas tendo um efeito considerável sobre os padrões da precipitação em todo o mundo através das teleconexões. Cada vez mais, tem sido alcançado significativo progresso no desenvolvimento de sistemas de previsão hidrológica de longo prazo (desde uma semana a alguns meses de antecedência) com o uso de índices climáticos. Desta forma, foi realizado na Bacia do Rio Piancó, um estudo entre eventos de precipitação e os índices climáticos, destacando suas teleconexões através de ferramentas de correlação e do pré-processamento de sinais. Em segunda parte, com as principais teleconexões selecionadas, foi possível desenvolver modelos de previsão da precipitação de longo prazo usando Redes Neurais Artificiais (RNA), sendo aplicado diferentes arranjos de RNA para a avaliação da qualidade da previsão nessa bacia. Observa-se que o Oceano Pacífico carrega grande influência sobre a precipitação na região, demonstrada através das correlações de seus índices climáticos, indicando fortes teleconexões. Os diversos arranjos de RNA, proporcionaram boa habilidade para uma previsão razoável com até 11 meses de antecedência, em alguns casos, com o uso de índices, até então, pouco abordados em estudos na região. Constata-se também que a distribuição do conjunto de dados para treinamento da RNA, tem forte influência sobre a qualidade da previsão.

**Palavras-chave:** Previsão de Precipitação; Índices Climáticos; Rede Neurais Artificiais; Teleconexões.

## ABSTRACT

It is estimated recently more intensively, the existence of a strong link between large-scale circulation and climatic variables having a considerable effect on rainfall patterns around the world through teleconnection. Increasingly, it has been significant progress in the development of long-term hydrological forecast systems (from one week to a few months in advance) with the use of climate indices. Thus, it was held in the Piencó River Basin, a study between precipitation events and climate indices, highlighting its teleconnection through correlation tools and signal preprocessing. In the second part, with the main selected teleconnection it was possible to develop long-term rainfall forecasting models using Artificial Neural Networks (ANN), applied different ANN arrangements for evaluating the forecast quality in this basin. It is observed that the Pacific Ocean carries great influence on rainfall in the region, demonstrated through the correlations of its climate indices, indicating strong teleconnection. The various ANN arrays, provide good skill for a reasonable prediction with up to 11 months in advance, in some cases, the use of indices, so far, less considered in studies in the region. It also notes that the distribution of the data set for the ANN training, has a strong influence on the quality of prediction.

**Key-word:** Precipitation forecast; Climate indices; Artificial Neural Networks; Teleconnections.



## SUMARIO

<b>1 INTRODUÇÃO .....</b>	<b>18</b>
<b>2 OBJETIVOS .....</b>	<b>21</b>
2.1 OBJETIVO GERAL.....	21
2.2 OBJETIVOS ESPECÍFICOS .....	21
<b>3 REFERENCIAL TEÓRICO .....</b>	<b>22</b>
3.1 PREVISÃO HIDROLÓGICA.....	22
3.1.1 Conceito e Aspectos Gerais.....	22
3.1.2 Previsão Hidrológica de Longo Prazo .....	24
3.2 TELECONEXÕES CLIMÁTICAS.....	27
3.2.1 Visão Geral.....	27
3.2.2 Relações com a variabilidade climática na América do Sul (ENOS).....	29
3.2.3 Relações com Precipitação no Nordeste do Brasil .....	32
3.2.4 Modos de Variabilidade e seus Índices Climáticos .....	34
3.3 PROCESSAMENTO DE SINAIS .....	38
3.3.1 Visão Geral pela Transformada de Fourier .....	38
3.3.2 Transformada de Wavelets (Ondeletas) .....	40
3.3.3 Análise de Wavelets Cruzada e Coerência .....	46
3.4 REDES NEURAIS ARTIFICIAIS .....	47
3.4.1 Conceitos e etapas de aplicação.....	47
3.4.2 Definição das Variáveis de Previsão .....	50
3.4.3 Identificação e Seleção das Variáveis Preditoras .....	52
3.4.4 Arquitetura da RNA e Divisão de Dados .....	54
3.4.5 Treinamento e Validação.....	60
<b>4 METODOLOGIA.....</b>	<b>68</b>
4.1 ÁREA DE ESTUDO .....	70
4.1.1 Localização.....	70
4.1.2 Caracterização Fisiográfica .....	71
4.1.3 Caracterização climática hidrológica.....	72
4.2 COLETA DE DADOS E ANALISE PRELIMINAR .....	75
4.2.1 Dados de Precipitação .....	75

4.2.2 Dados de Índices Climáticos .....	77
4.3 ANÁLISES DE TELECONEXÕES .....	79
4.3.1 Correlação de Pearson e de Spearman.....	79
4.3.2 Correlação por Wavelets Cruzada e Coerência .....	80
4.3.3 Critérios para Definição das Variáveis Preditoras.....	81
4.4 PREVISÃO DE PRECIPITAÇÃO COM REDES NEURAS ARTIFICIAIS .....	82
4.4.1 Aspectos gerais da aplicação .....	82
4.4.2 Variáveis de Entrada.....	82
4.4.3 Divisão e Pré-processamento dos Dados.....	83
4.4.4 Treinamento e critérios de parada .....	85
4.5 MEDIDAS DE DESEMPENHO.....	86
<b>5 RESULTADOS E DISCURSÃO.....</b>	<b>87</b>
5.1 CORRELAÇÕES PRECIPITAÇÃO X INDICES CLIMATICOS (TELECONEXÕES) .	87
5.2 VARIÁVEIS PREDITORAS E COMBINAÇÕES .....	100
5.3 PREVISÃO.....	108
5.3.1 RNA com 100 anos de dados .....	109
5.3.1.1 Resultados referentes ao 1º grupo de variáveis predictoras.....	109
5.3.1.2 Resultados referentes ao 2º grupo de variáveis predictoras.....	112
5.3.2 RNA com 63 anos de dados .....	116
5.3.2.1 Resultados referentes ao 1º grupo de variáveis predictoras.....	116
5.3.2.2 Resultados referentes ao 2º grupo de variáveis predictoras.....	120
5.3.3 RNA com 61 anos de dados .....	124
5.3.3.1 Resultados referentes ao 1º grupo de variáveis predictoras.....	124
5.3.3.2 Resultados referentes ao 2º grupo de variáveis predictoras.....	128
5.3.4 RNA com 57 anos de dados .....	132
5.3.4.1 Resultados referentes ao 1º grupo de variáveis predictoras.....	132
5.3.4.2 Resultados referentes ao 2º grupo de variáveis predictoras.....	136
5.3.5 RNA com 61 anos e todos os dados .....	140
5.3.5.1 Resultados referentes ao 1º grupo de variáveis predictoras.....	140
5.4 ANÁLISE DO EFEITO DA DISTRIBUIÇÃO DE DADOS .....	144
5.4.1 Análise dos dados com 100 anos.....	145
5.4.2 Análise dos dados com 63 anos.....	147
5.4.3 Análise dos dados com 61 anos.....	149

5.4.4 Análise dos dados com 57 anos.....	151
5.4.5 Análise de todos os índices com 61 anos .....	153
<b>6 CONCLUSÃO.....</b>	<b>155</b>
<b>7 RECOMENDAÇÕES.....</b>	<b>157</b>
<b>REFERÊNCIAS BIBLIOGRAFICAS .....</b>	<b>158</b>
<b>ANEXOS .....</b>	<b>165</b>

## LISTA DE FIGURAS

Figura 3.1 – Influências do El Nino/La Niña no globo, onde as cores indicam as áreas com as mesmas influencias climáticas. Fonte: CPC (2003) .....	31
Figura 3.2 – Representação esquemática dos sistemas atmosféricos na baixa e alta troposfera atuantes na América do Sul. Fonte: Reboita et al. (2010) .....	32
Figura 3.3 – Circulação superficial sobre o Atlântico equatorial e parte Pacífico oriental, de Março de 1958-1997. (a) Temperatura da superfície do mar, com espaçamento isotérmico de 2°C e com área acima de 28 ° C (cinza); (b) pressão do nível do mar, com espaçamento isobárico de 2mb, e com área inferior a 1012 mb (cinza); (c) a direção do vento resultante e velocidade, com espaçamento de 2 ms-1; (d) divergência com espaçamento isolinhas de 5 x 10-6s-1, e a convergência sombreada. A linha sólida em negrito limita o “Nordeste” do Brasil, e F e Q denotam as estações de Fortaleza e Quixeramobim. Fonte: Kucharski1 et al. (2007) .....	34
Figura 3.4 – A wavelet-mãe de Morlet. Fonte: Santos et al. (2013) .....	43
Figura 3.5 – A wavelet de Morlet (a) e a representação de sua formação através de um senoide e um envelope gaussiano (b) .....	44
Figura 3.6 – (a) Diferentes tipos de wavelets-mãe, (b) translações e (c) dilatações da base e (d) processo de correlação da transformada de Wavelets .....	45
Figura 3.7 – a) Série composta pelas frequências 7,5, 30, 60 e 120 Hz. b) Espectro de potência Wavelet. c) Espectro de potência global Wavelet. Fonte: Santos et al. (2013) .....	46
Figura 3.8 – Modelo de uma Rede Neural Artificial e detalhe do funcionamento de um dos neurônios. Fonte: Bravo (2010) .....	50
Figura 3.9 – Tipos de modelos de RNA de três camadas com três variáveis de entrada e uma de saída: (a) Multilayer Feedforward (MLP); (b) Rede Neural Recorrente (RNN). Fonte: Paz et al. (2011) .....	56
Figura 3.10 – Representação do critério de parada do treinamento por validação cruzada. Fonte: Philip e Joseph (2003) .....	60
Figura 3.11 – Ilustração da atividade de aprendizagem, a propagação (entradas) e a retro propagação (erro). Fonte: Nastos et al. (2013) .....	62
Figura 3.12 – Pontos de mínimos locais associados a função erro. Fonte: Silva et al. 2010 ...	63
Figura 3.13 – Interpretação geométrica da Regra Delta. Fonte: Silva et al. (2010) .....	65
Figura 4.1 – Fluxograma das etapas do estudo.....	70
Figura 4.2 – Mapa de Localização da Bacia. Fonte: Autor .....	72
Figura 4.3 – Mapa de relevo da Bacia. Observação para os vales muito abertos em forma de “U”. Fonte: Autor .....	73

Figura 4.4 – Precipitação média anual para a bacia do rio Piancó no período de 1911 a 2011.Fonte: Autor.....	74
Figura 4.5 – precipitação média mensal para a bacia do Piancó. Fonte: Autor.....	75
Figura 4.6 – Vazão média mensal do posto da cidade de Piancó. Fonte: Autor .....	75
Figura 4.7 – Localização dos Postos Pluviométricos dentro da área de estudo e a quantidade de dados disponíveis. Fonte: Autor .....	76
Figura 4.8– Regiões delimitadas para os trimestres mais chuvosos no Brasil. Destaque para a região onde está inserida a Bacia do Rio Piancó. Fonte: Adaptado de Rao e Hada (1990) ....	77
Figura 4.9 – Precipitação acumulada no trimestre FMA na bacia do Rio Piancó.....	78
Figura 5.1 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Pacífico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulada. 94	94
Figura 5.2 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Pacífico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulada.....	95
Figura 5.3 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Pacífico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulada.....	96
Figura 5.4 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Atlântico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulada.....	98
Figura 5.5 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Atlântico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulada.....	99
Figura 5.6 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Ártico e Antártico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulada .....	100
Figura 5.7 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos Globais e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulada .....	100
Figura 5.8 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 1º grupo de dados de 100 anos para o conjunto de treinamento, testes, validação e global... 110	110
Figura 5.9 – Melhor ajuste entre as simulações do 3º arranjo para os conjuntos de treinamento, validação e verificação dos dados .....	112
Figura 5.10 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 3º Arranjo do 1º grupo de variáveis preditoras de 100 anos.....	113

Figura 5.11 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 2º grupo de dados de 100 anos para o conjunto de treinamento, testes, validação e global...	114
Figura 5.12 – Melhor ajuste entre as simulações do 9º arranjo para os conjuntos de treinamento, validação e verificação dos dados .....	117
Figura 5.13 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 9º Arranjo do 2º grupo de variáveis preditoras de 100 anos.....	117
Figura 5.14 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 1º grupo de dados de 63 anos para o conjunto de treinamento, testes, validação e global.....	118
Figura 5.15 – Melhor ajuste entre as simulações do 7º arranjo para os conjuntos de treinamento, validação e verificação dos dados .....	121
Figura 5.16 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 7º Arranjo do 1º grupo de variáveis preditoras de 63 anos.....	121
Figura 5.17 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 2º grupo de dados de 63 anos para o conjunto de treinamento, testes, validação e global.....	122
Figura 5.18 – Melhor ajuste entre as simulações do 3º arranjo para os conjuntos de treinamento, validação e verificação dos dados .....	124
Figura 5.19 – Melhor ajuste entre as simulações do 7º arranjo para os conjuntos de treinamento, validação e verificação dos dados .....	125
Figura 5.20 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 1º grupo de dados de 61 anos para o conjunto de treinamento, testes, validação e global.....	126
Figura 5.21 – Melhor ajuste entre as simulações do 11º arranjo para os conjuntos de treinamento, validação e verificação dos dados .....	129
Figura 5.22 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 11º Arranjo do 1º grupo de variáveis preditoras de 61 anos.....	129
Figura 5.23 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 2º grupo de dados de 61 anos para o conjunto de treinamento, testes, validação e global.....	130
Figura 5.24 – Melhor ajuste entre as simulações do 12º arranjo para os conjuntos de treinamento, validação e verificação dos dados .....	133
Figura 5.25 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 12º Arranjo do 2º grupo de variáveis preditoras de 61 anos.....	133
Figura 5.26 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 1º grupo de dados de 57 anos para o conjunto de treinamento, testes, validação e global.....	134
Figura 5.27 – Melhor ajuste entre as simulações do 5º arranjo para os conjuntos de treinamento, validação e verificação dos dados .....	137
Figura 5.28 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 5º Arranjo do 1º grupo de variáveis preditoras de 57 anos.....	137

Figura 5.29 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 2º grupo de dados de 57 anos para o conjunto de treinamento, testes, validação e global.....	138
Figura 5.30 – Melhor ajuste entre as simulações do 7º arranjo para os conjuntos de treinamento, validação e verificação dos dados .....	141
Figura 5.31 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 7º Arranjo do 2º grupo de variáveis preditoras de 57 anos.....	141
Figura 5.32 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do grupo com todos os dados de 61 anos para o conjunto de treina, testes, validação e global..	142
Figura 5.33 – Melhor ajuste entre as simulações do 18º arranjo para os conjuntos de treinamento, validação e verificação dos dados .....	145
Figura 5.34 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 18º Arranjo do 2º grupo de variáveis preditoras de 57 anos.....	145
Figura 5.35 – Gráfico comparativo entre os dois grupos de variáveis de 100 anos referente a média da correlação de Pearson de todos os arranjos para cada topologia testada, de acordo com a quantidade de neurônios e a distribuição para o conjunto de treina, valida, verifica global .	147
Figura 5.36 – Gráfico comparativo entre os dois grupos de variáveis de 63 anos referente a média da correlação de Pearson de todos os arranjos para cada topologia testada, de acordo com a quantidade de neurônios e a distribuição para o conjunto de treinamento, validação, verificação e global .....	149
Figura 5.37 – Gráfico comparativo entre os dois grupos de variáveis de 61 anos referente a média da correlação de Pearson de todos os arranjos para cada topologia testada, de acordo com a quantidade de neurônios e a distribuição para o conjunto de treinamento, validação, verificação e global .....	151
Figura 5.38 – Gráfico comparativo entre os dois grupos de variáveis de 57 anos referente a média da correlação de Pearson de todos os arranjos para cada topologia testada, de acordo com a quantidade de neurônios e a distribuição para o conjunto de treinamento, validação, verificação e global .....	153
Figura 5.39 – Gráfico do efeito da distribuição para o grupo de variáveis com todos os dados referente a média da correlação de Pearson de todos os arranjos para cada topologia testada de acordo com a quantidade de neurônios para o conjunto de treina, valida, verifica e global ....	155
Figura 1 – Distribuição dos dados para os conjuntos de treinamento, validação e verificação para ao intervalo de 100 anos .....	165
Figura 2 – Distribuição dos dados para os conjuntos de treinamento, validação e verificação para ao intervalo de 63 anos .....	166
Figura 3 – Distribuição dos dados para os conjuntos de treinamento, validação e verificação para ao intervalo de 61 anos .....	167
Figura 4 – Distribuição dos dados para os conjuntos de treinamento, validação e verificação para ao intervalo de 57 anos .....	168

## LISTA DE TABELAS

Tabela 3.1 – Levantamento de técnicas e variáveis explicativas de estudos de previsão de precipitação de longo prazo nos anos de 1978 a 1999 no NEB .....	26
Tabela 3.2 – Levantamento de Índices climáticos e características .....	37
Tabela 4.1 – Valores médios mensais de dados meteorológicos na estação de São Gonçalo..	74
Tabela 4.2 – Levantamento de Índices climáticos e características .....	79
Tabela 4.3 – Levantamento de Índices climáticos e características .....	84
Tabela 5.1 – Correlação Índices Climáticos do Pacífico.....	89
Tabela 5.2 – Correlação Índices Climáticos do Atlântico .....	91
Tabela 5.3 – Correlação Índices Climáticos do Índico, Ártico e Antártico.....	92
Tabela 5.4 – Correlação Índices Climáticos Globais .....	92
Tabela 5.5 – Variáveis preditoras com 100 anos (1910-2009) .....	101
Tabela 5.6 – Variáveis preditoras com 63 anos (1948-2010) .....	102
Tabela 5.7 – Variáveis preditoras com 61 anos (1950-2010) .....	103
Tabela 5.8 – Variáveis preditoras com 57 anos (1951-2008) .....	104
Tabela 5.9 – Variáveis preditoras formada por todos os índices com 61 anos (1950-2010)	105
Tabela 5.10 – Combinações do 1º grupo de variáveis para cada intervalo de dados .....	106
Tabela 5.11 – Combinações do 2º grupo de variáveis para cada intervalo de dados .....	107
Tabela 5.12 – Combinações das variáveis do grupo formado com todos os índices .....	108
Tabela 5.13 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 100 anos.....	111
Tabela 5.14 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 100 anos .....	111
Tabela 5.15 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 100 .....	115
Tabela 5.16 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 100 anos .....	115
Tabela 5.17 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 63 anos.....	119



Tabela 5.18 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 63 anos .....	119
Tabela 5.19 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 63 anos.....	123
Tabela 5.20 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 63 anos .....	123
Tabela 5.21 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 61 anos.....	127
Tabela 5.22 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 61 anos .....	127
Tabela 5.23 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 61 anos.....	131
Tabela 5.24 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 61 anos .....	131
Tabela 5.25 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 57 anos.....	135
Tabela 5.27 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 57 anos.....	139
Tabela 5.28 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 57 anos.....	139
Tabela 5.29 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do grupo com todos os dados de 61 anos.....	143
Tabela 5.30 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do grupo com todos os dados de 61 anos.....	143

# 1 INTRODUÇÃO

A atmosfera em constante movimento, encontram-se em estado de permanente interação por meio de intercâmbio de fluxos de água líquida ou vapor, radiação eletromagnética e trocas de calor, constituindo-se o que se denomina sistema climático (Varejão-Silva, 2006). Dentro desse complexo sistema, podemos considerar a precipitação como principal fator climático atuante no regime hidrológico, principalmente em regiões tropicais, pois a alteração dessa variável apresenta impacto relativo diferenciado sobre as saídas do regime, como o nível e a vazão de um rio numa bacia hidrográfica (Tucci & Braga, 2003).

A variabilidade natural dos processos climáticos é um fator importante dominando os padrões de precipitação em diversas regiões do globo (Giannini et al., 2003; Rousseau et al., 2014; Chithra e Thampi, 2015) e em escalas globais e regionais (Sheffield e Wood, 2008). Conseqüentemente, o impacto dessa variabilidade na precipitação, mesmo dentro de limites esperados, pode ter reflexos significativos nas mais diversas atividades humanas, como o controle da necessidade de irrigação na agricultura (Meza, 2008), planejamento urbano (Silveira e Sartori, 2010; Moore et al, 2005), impactos socioeconômicos para o turismo (Olya e Alipour, 2015), operação de reservatórios para múltiplos propósitos (Bravo, 2010), e os recursos globais de água para muitos outros usos.

Apesar de existirem muitos fatores físicos importantes que afetam o regime de precipitação, tem sido mostrado que existe uma ligação entre a circulação em grande escala e variáveis climáticas (Cavalcanti e Ambrizzi, 2009), tendo um efeito considerável sobre os padrões da precipitação em todo o mundo através das teleconexões (Ropelewski e Halpert, 1987, 1989). Estas teleconexões são ligações de eventos hidrometeorológicos com outras características geofísicas como certos padrões de circulação atmosférica, às vezes separados espacialmente por distâncias em escala continental. Por exemplo, o fenômeno mais conhecido por causar tais variações climáticas é El Niño-Oscilação Sul (ENOS) (Trenberth, 1997), que mesmo sendo um fenômeno oceano-atmosfera acoplado que ocorre no Pacífico Tropical, ele é conhecido por influenciar outros oceanos (Silva, 2013), clima (Mason e Goddard, 2001) e hidrologia de outros lugares (Rao e Hada, 1990).

Como as sociedades humanas estão cada vez mais dependentes de grandes quantidades de água, uma área de conhecimento que se desenvolve bastante é a busca de relações que permita entender estes processos e obter prognósticos que diminuam o risco para a sociedade. Isto

envolve os estudos de processos climáticos interanuais e interdecadais, que são analisados através de índices climáticos.

O uso de índices climáticos para o desenvolvimento de sistemas de previsão hidrológica de longo prazo (variando de uma semana a alguns meses de antecedência) vem alcançado significativo progresso nos últimos anos. O fenômeno ENOS e as conexões climáticas em escala global passaram a ser melhor caracterizados (Cavalcanti e Ambrizzi, 2009). Esses estudos usualmente tentam explicar e prever o comportamento da precipitação, principal variável característica do regime climático, nas principais estações chuvosas do ano, normalmente bem demarcadas nas regiões tropicais. À precipitação total acumulada na estação chuvosa (precipitação sazonal) têm-se atribuído influência de fatores atmosféricos e oceânicos de larga escala, como variações na temperatura dos oceanos (Galvão, 1999).

Nas últimas décadas, várias abordagens têm sido apresentadas para a previsão de chuva. Estes métodos podem ser classificados em duas categorias: dinâmicos e empíricos. Além dos métodos estatísticos usuais utilizados (regressão, métodos multivariados, análise de periodicidades) e modelos dinâmicos (modelos de circulação global atmosférica), ou combinação deles, novos métodos vêm sendo propostos como o uso de Redes Neurais Artificiais (RNA) e Wavelets.

Ao longo dos últimos 15 anos ou mais, o uso de RNAs para a predição e previsão de variáveis de recursos hídricos tornou-se uma área de pesquisa bem estabelecida. Os fundamentos matemáticos da RNA e aplicações específicas em hidrologia, incluindo precipitação, gerou uma grande quantidade de atividade de pesquisa nesta área e levou a uma série de artigos de revisão (ASCE, 2000a; ASCE, 2000b; Maier e Dandy, 2000; Dawson e Wilby, 2001; Maier et al., 2010), que não só confirmou o potencial de RNAs para a predição e previsão de variáveis de recursos hídricos, mas também identificou uma série de desafios que precisavam ser abordados, a fim de garantir que RNAs tornassem-se uma abordagem de modelagem madura ao lado de outras abordagens modeladoras de variáveis hidrológicas.

Além do uso de RNAs, desenvolvimentos ainda mais recentes levaram a um renovado esforço para projetar melhores estruturas de previsão hidrológica para dados não-estacionários e outros sistemas com variáveis ambientais, auxiliados por transformadas matemáticas (Nourani et al., 2014; Labat, 2005). Dentre as transformações matemáticas mais usadas recentemente para a análise tempo-frequência de sinais, está a Transformada de Wavelet (TW). Estudos como Partal e Kisi (2007), Sang (2013) e He et al. (2015), são alguns exemplos onde wavelets têm sido

utilizados para a análise de dados climáticos e hidrológicos para melhorar a eficiência da previsão.

Algumas regiões têm recebido maior atenção dos pesquisadores, por apresentarem maior previsibilidade ou conexões com a variabilidade climática em outras regiões do planeta. Entre estas regiões estão o Nordeste do Brasil (NEB) (Uvo et al., 1998; Hastenrath, 2011), o Sahel (Palmer, 1986; Liu et al., 2012), o leste e o sul da África (Mason, 1996; Nicholson e Grist, 2003), partes da Índia (Sahai et al., 2000; Pal e Al-Tabbaa, 2009; Srivastava et al., 2010) e da Austrália (Mekanik et al., 2013; He et al., 2015). Os resultados para elas obtidos são importantes não apenas para aplicação local, mas também funcionam como referência comparativa no contexto amplo da pesquisa sobre variabilidade climática global.

Na região do NEB, temos o semiárido brasileiro, que tem como principal característica a alta variabilidade espacial e temporal do clima, que está associada ao elevado grau de vulnerabilidade a que este ambiente está submetido (Souza Filho, 2003). Estando localizada numa região semiárida, a bacia hidrográfica do Rio Piancó, objeto de estudo, é uma das mais importantes do estado da Paraíba por conter a maior reserva hídrica do estado (Sistema Coremas Mãe d'Água) e por ser foco de importantes demandas hídricas no estado da Paraíba e Rio Grande do Norte. Como em toda região semiárida que tem forte carência de água, a bacia do Piancó é fonte de diversos conflitos pelo uso da água, sendo um dos principais problemas da região em estudo a escassez dos recursos hídricos devido à crescente demanda por água e a multiplicidade de seu uso (Melo, 2004). Outro grande conflito de uso é com o vizinho estado do Rio Grande do Norte com a questão da vazão liberada pelo sistema de reservatórios Coremas Mãe d'Água para perenização do rio Piranhas-Açu (Lima, 2004).

Um melhor conhecimento dos mecanismos envolvidos nas interações climáticas e hidrológicas em escalas de bacias hidrográficas pode se tornar uma ferramenta valiosa na gestão de áreas como a do rio Piancó, onde é evidente uma deficiência em infraestrutura hídrica e nos aparatos governamentais, tanto em condições severas de secas ou cheias. Desta forma, será realizado na Bacia do Rio Piancó, um estudo entre eventos de precipitação e os índices climáticos, destacando suas teleconexões através de ferramentas de correlação e do processamento de sinais com uso de Wavelets, e compor através desse estudo uma aplicação de diferentes arranjos de RNA para a avaliação da previsão da precipitação de longo prazo nessa bacia.

## **2 OBJETIVOS**

### **2.1 OBJETIVO GERAL**

Avaliar as previsões de precipitação de longo prazo na bacia do rio Piencó através de diferentes arranjos de Redes Neurais Artificiais com o uso de componentes climáticas.

### **2.2 OBJETIVOS ESPECÍFICOS**

- a) Analisar as correlações entre eventos hidrológicos e índices climáticos, elencando os principais componentes;
- b) Identificar as teleconexões existentes na tentativa de obter qual a melhor correlação considerando o tempo de defasagem entre as variáveis preditoras (componentes climáticos) e sua possível influência no trimestre mais chuvoso;
- c) Avaliar o efeito das técnicas de seleção das variáveis preditoras em relação ao desempenho das previsões.
- d) Analisar o efeito das distribuições dos dados na Rede Neural Artificial e de suas variações topológicas.

## 3 REFERENCIAL TEÓRICO

### 3.1 PREVISÃO HIDROLÓGICA

#### 3.1.1 Conceito e Aspectos Gerais

A previsão hidrológica é um dos objetivos mais importantes da Engenharia de Recursos Hídricos e pode ser definida como a estimativa em um tempo futuro específico, ou em um determinado intervalo de tempo, de algumas variáveis hidrológicas (Lettenmaier e Wood, 1993 apud Bravo, 2010).

Nas regiões tropicais, o escoamento depende essencialmente da precipitação, e sua previsibilidade está vinculada à previsibilidade da precipitação e dos processos meteorológicos que a determinam (Galvão, 1999). Os métodos de previsão de precipitação mais utilizados pertencem a dois grandes grupos: métodos dinâmicos e métodos empíricos/estatísticos.

A previsão de precipitação com métodos dinâmicos é realizada utilizando modelos de circulação geral (MCG), com o objetivo de representar, com equações físicas do fenômeno, os processos que compõem o sistema climático. Esses métodos podem simular o comportamento acoplado do oceano e da atmosfera, no caso dos Modelos Acoplados de Circulação Geral (MACG), ou apenas as condições atmosféricas, no caso dos Modelos de Circulação Geral da Atmosfera (MCGA) (Bravo, 2010). Os MCGs geram uma gama de variáveis meteorológicas, que iniciam de valores de variáveis centrais ou básicas, como temperatura, umidade específica e componentes do vento; e variáveis derivadas como precipitação, nebulosidade, etc (Garcia-Moya, 2010). Esses modelos discretizam o globo em elementos de 100 a 1000 km de lado e de 1 a 5 km de altura. A topografia e os processos físicos do sistema em cada elemento são valores médios (Tucci e Braga, 2003).

Uma das desvantagens do uso de MCGs é a necessidade de grandes esforços computacionais e de modelagem à medida que se tenta um aumento de sua resolução horizontal, além do alto custo associado, o que enfatiza que a simulação de um MCG não é ainda uma tarefa simples (Garcia-Moya, 2010). Segundo Molion (2008), os MCGs também apresentam uma limitação, pois ainda não representam a complexa interação entre os processos físicos diretos (forcings) e de realimentação (feedback) adequadamente. O mesmo autor destaca ainda que um dos principais problemas é a representação dos processos físicos com escala espacial inferior à da grade tridimensional dos MCGs, como formação/desenvolvimento de nuvens e precipitação,

transportes verticais turbulentos de calor e de umidade dentro e para fora da camada limite planetária, transporte de calor pelas correntes marinhas, que precisam ser parametrizados, ou seja, representados por algoritmos matemáticos empíricos que, assumem-se, descrever a física dos processos satisfatoriamente, onde pequenas variações nas condições iniciais podem causar grandes perturbações na condição final. Garcia-Moya (2010) também reconhece que há muitos fenômenos na atmosfera cujos efeitos sobre as variáveis atmosféricas são muito importantes e cuja escala típica é muito menor do que as escalas de resolução de modelos MCGs, e que para isso deve-se estabelecer algumas hipóteses sobre o seu funcionamento.

Os métodos empíricos são baseados em relações de observação da variável prevista Y com as variáveis predictoras X. A lógica por trás dessa abordagem é descobrir características espaciais e temporais relevantes nos arquivos históricos de variáveis de precipitação Y e das predictoras X, e usá-los para prever os valores futuros de Y (Bravo, 2010). O método convencional de modelagem empírica tenta assumir uma relação estatística entre o preditor e as variáveis de previsão, que pode ser ajustada por alguma função determinada por regressão ou outros métodos de otimização, sem a definição precisa de uma relação física entre eles, onde a capacidade de previsibilidade da relação é restringida ao intervalo existente do conjunto de dados (He et al., 2015). As relações construídas usando essa abordagem estatística podem servir como um modelo de previsão, mas não para explicar ou descrever o funcionamento do sistema (Tucci e Braga, 2003). As análises de regressão múltipla (Bretherton et al., 1992; Mekanik et al., 2013) e as redes neurais artificiais (Paz et al., 2011; Deo e Sahin, 2015) são exemplos de métodos empíricos utilizados na previsão de precipitação.

Os modelos empíricos se caracterizam por seu baixo custo computacional e de modelagem, quando comparado aos MCGs, e por serem desenvolvidos para previsões específicas do local. No entanto, as desvantagens são a falta de significado físico dos fenômenos e a dependência de dados históricos disponibilizados para o ajuste (Bravo, 2010).

Pensando na utilização em manejo de recursos hídricos, esses métodos podem mais adequadamente ser classificados segundo o tipo de informação que fornecem. A previsão da precipitação pode ser dada através de valores absolutos (quantidade de precipitação) ou representados por níveis distintos de umidade (ou seca) do período de tempo. O primeiro modo é uma previsão de precipitação quantitativa onde o preditando é um valor numérico contínuo, ou seja, ele pode assumir qualquer valor (dentro de uma faixa razoável). Por sua vez, a previsão por níveis divide a amplitude de variabilidade da precipitação em várias faixas, ou categorias,

e emitem a previsão prescrevendo em qual (ou quais) categoria há maior probabilidade de ocorrência do total esperado de precipitação (Wilks, 2006). Por exemplo, pode ser previsto um total de precipitação de 300 mm para o mês seguinte (previsão quantitativa), ou pode ser previsto que a precipitação do mês seguinte irá estar acima do normal ou muito úmido (previsão qualitativa-categórica). A escolha da previsão quantitativa ou categórica depende de vários fatores, incluindo o horizonte desejado de tempo, os efeitos da previsão, os dados disponíveis, bem como o método de previsão utilizado.

Quanto ao intervalo de tempo entre o cálculo da previsão e o evento previsto, chamado de horizonte de tempo, este é amplamente classificado como previsão de curto ou de longo prazo. Previsões de curto prazo correspondem a horizontes de tempo de algumas horas a alguns dias, geralmente chamadas de “previsão do tempo”, e utilizadas principalmente para a previsão de cheias, de alerta e controle (Moore et al., 2005), operação de rio/reservatório (Bravo et al., 2010), mitigação de desastres naturais (Silveira e Sartori, 2010) e previsão de condições adversas para a agricultura, como a geada e granizo (Calanca, 2014).

Previsões de longo prazo correspondem a horizontes temporais que prescrevem de várias semanas, meses ou anos e são chamadas de “previsão climática”, usadas principalmente para o planejamento e otimização de recursos hídricos (Galvão, 1999; Nilsson et al., 2006), tais como alocação de água para irrigação agrícola (Meza, 2008), operação de reservatórios (Hsieh et al., 2003; Bravo et al., 2007), avaliação e implementação de medidas contra secas e inundações (Changnon e Vonnahme, 2003) e turismo (Olya e Alipour, 2015).

### 3.1.2 Previsão Hidrológica de Longo Prazo

A previsão hidrológica de longo prazo se caracteriza por estimar um evento com meses de antecedência. A previsão de precipitação sazonal, foco deste trabalho, é um subtipo de previsão de longo prazo focada sobre os padrões de precipitação para o período de chuvas sazonais (Galvão, 1999).

A previsão sazonal é fundada no princípio de que os processos de grande escala evoluem lentamente, fornecendo alguma previsibilidade do clima em escala de tempo sazonal (Villarini e Serinaldi, 2012). Esse tipo de previsão comumente é realizada por estatísticas sazonais. No entanto, atualmente, com o melhor conhecimento das conexões climáticas e com os avanços na simulação numérica, ampliaram-se largamente também as possibilidades de uso de MCGs (Tucci et al., 2007). Mesmo assim, o que tem motivado a exploração persistente de teleconexões



regionais/globais sobre os padrões de precipitação sazonal em diversas regiões do globo desde a época de Walker<sup>1</sup>, foi o sucesso razoável alcançado pelas abordagens empíricas, o que resultou em um grande número de preditores, bem como uma variedade de técnicas estatísticas (Mason, 1996; Shukla et al., 2011).

Para entender melhor como se deu o avanço dessas técnicas ao longo dos anos, podemos tomar como exemplo o levantamento feito por Galvão (1999), onde em sua tese ele apresenta uma lista parcial dos principais estudos de previsão da precipitação sazonal no norte-nordeste do Brasil e as respectivas técnicas usadas.

Tabela 3.1 – Levantamento de técnicas e variáveis explicativas de estudos de previsão de precipitação de longo prazo nos anos de 1978 a 1999 no NEB.

<b>Referência</b>	<b>Técnica*</b>	<b>Variáveis explicativas</b>
Girardi e Teixeira (1978)	Periodicidades	Própria série
Nobre et al. (1980)	Periodicidades	Própria série
Mesquita e Moretin (1984)	Periodicidades	Própria série
Hastenrath et al. (1984), Hastenrath (1990c), Greischar e Hastenrath (1998)	RL, ADL, RNA	Chuva da pré-estação, TSM e ventos
Kane e Trivedi (1986, 1988), Kane (1997)	Periodicidades	Própria série
Silva (1988)	AF	Própria série
Harvey e Souza (1987), Tavares e Souza (1988), Brasil e Souza (1993)	MST	Própria série
Brito et al. (1991), Mello (1997)	RL	Chuva da pré-estação
Ward e Folland (1991), Colman et al. (1997)	RL, ADL	TSM
Repelli e Alves (1994)	MCG	–
Repelli e Alves (1994)	Análise subjetiva	Atmosfera e oceanos
Xavier e Xavier (1995), Xavier et al. (1998)	MLG, RL	TSM, ENOS, ventos, atividade solar
Fernandes (1995)	RNA e MST	Própria série
Gang et al. (1996)	RNA	Própria série
Repelli e Alves (1996a)	ACC	TSM
Sperber e Palmer (1996), Brankovic e Palmer (1997), Cavalcanti et al. (1997)	MCG	–
Freitas e Billib (1997)	RNA	TSM
Graham (1997)	MCG	–
Cavalcanti et al. (1998)	MCG	–
Evans et al. (1998)	MCG	–

Fonte: Galvão (1999)

\*ACC: análise de correlações canônicas; ADL: análise discriminante linear; AF: análise de frequência; MCG: modelo de circulação global; MLG: modelos lineares generalizados; MST: modelos univariados de séries temporais; RL: regressão linear; RN: rede neural.

<sup>1</sup> Reconhecido climatologista inglês, responsável por descrever as oscilações dos parâmetros atmosféricos entre lugares muito distantes e desvendar a atuação do El Niño no início do século.

Nestes últimos anos, ao lado de estudos com o uso de MCGs, houve também uma maior contribuição de trabalhos com a aplicação de modelos de RNA para previsão hidrológica de longo prazo, incluindo a precipitação sazonal, onde mais recentemente essa técnica passou a ser auxiliada pelo pré-processamento de sinais com uso de transformadas matemáticas, mais especificadamente a Transformada de Wavelets (Sang, 2013; Santos et al., 2013).

As variáveis de entrada mais utilizadas por modelos empíricos de previsão de longo prazo de precipitação são: a temperatura na superfície do mar (TSM) (Iseri et al., 2005; Aboot e Marohasy, 2014), antecedentes da própria precipitação (Philip e Joseph, 2003; Nastos et al. 2013), as variáveis relacionadas aos padrões de circulação atmosférica (Silverman e Dracup, 2000) e os índices climáticos (Kumar et al., 2007; Paz et al., 2011; Villarini e Serinaldi, 2012). Outros trabalhos utilizando essas variáveis já foram mencionados também na Tabela 3.1.

O tempo de antecedência das previsões de precipitação de longo prazo normalmente compreendem um período de 1 a 9 meses (Georgakakos e Krysztofowicz, 2001 apud Tucci et al., 2007). Silverman e Dracup (2000) e Bravo (2010) em seus trabalhos por exemplo, partiram de variáveis de entrada com até 9 meses de antecedência. Outros trabalhos buscaram um horizonte menor de tempo como o de He et al. (2015), 1 mês; Aboot e Marohasy (2014) e Peng et al. (2014), até 3 meses; Paz et al. (2011) e Nastos et al. (2013), até 4 meses. E outros superaram esse tempo como Iseri et al. (2005) e Srivastava et al. (2010) que testaram horizontes de até 12 meses.

Um dos focos dos mais visados pelas previsões sazonais de precipitação é no planejamento agrícola, por ser particularmente vulnerável à variabilidade climática interanual e à sua imprevisibilidade (Calanca, 2014). A maior parte dos esforços nesta direção tem procurado caracterizar as relações entre a variabilidade de precipitação na forma como é expressa na previsão, ou seus preditores (índices climáticos, por exemplo), e a produção agrícola ou o ciclo vegetativo das culturas (Meza, 2008; Srivastava et al., 2010). No semi-árido do NEB por exemplo, algumas metodologias têm sido propostas para determinar a melhor época de semeadura de culturas dependentes de chuva (Repelli e Alves, 1996 apud Galvão, 1999; Araújo et al., 1998). Em virtude da grande variabilidade interanual da precipitação na região, a aplicabilidade destes resultados ainda tem sido limitada. A prática corrente na região, executada pelos núcleos estaduais de meteorologia e recursos hídricos do Nordeste é de disseminar aos usuários, a partir da pré-estação chuvosa, a previsão de precipitação.

Para o caso do abastecimento de água humano, simulam-se cenários hidrológicos possíveis para os reservatórios, considerando o prognóstico da previsão, identificando riscos de colapso no armazenamento e suprimento hídrico. A previsão de precipitação de longo prazo pode ainda diminuir a incerteza da avaliação econômica de algumas *commodities* relacionadas com a água, como o planejamento da formação do preço da energia hidrelétrica e o gerenciamento dos conflitos da água (Galvão, 1999; Tucci e Braga, 2003).

É importante ressaltar que a previsão de precipitação de longo prazo é uma área ainda em desenvolvimento na hidrologia, tendo apresentado grandes avanços ao longo dos últimos anos (Bravo, 2010). Isto vem ocorrendo graças ao maior número de dados coletados e ao uso de computadores de maior capacidade de armazenamento e de processamento. Assim, aumentou-se gradativamente a compreensão mundial das relações entre as variáveis climáticas e os eventos hidrológicos, como os períodos de seca ou de inundação. Um grande número de trabalhos tem sido feito nesse sentido, propondo novas metodologias de previsão de precipitação, conforme já foi apresentado nos parágrafos anteriores.

## 3.2 TELECONEXÕES CLIMÁTICAS

### 3.2.1 Visão Geral

O clima não é somente uma variável que marca o ritmo das estações do ano, é um fenômeno sistêmico, dinâmico e natural produzido pelo sistema Sol-Atmosfera-Terra (Molion, 2005).

A variabilidade climática de uma região é definida atualmente como a composição de variações em várias escalas temporais e espaciais. Os sistemas atmosféricos na escala de tempo que atuam sobre determinada região constituem as componentes que definem o padrão climático local. Adicionalmente, o conhecimento da frequência e intensidade de atuação destes sistemas atmosféricos, em um determinado período de tempo, é importante para a caracterização da variabilidade climática local (Nunes e Lombardo, 1995).

Um grande avanço para o entendimento da variabilidade climática deu-se com a consideração do oceano e de sua interação com a atmosfera como importante organizador dos padrões climáticos pelo globo. O monitoramento da ocorrência de anomalias da temperatura da superfície do mar (TSM) passou a ser de fundamental importância, já que essas quando surgem,

modulam o comportamento atmosférico através da modificação dos campos de pressão e, conseqüentemente, as variações das intensidades e direções dos ventos (Molion, 2005).

Desde o início do século XX, a comunidade científica tem registrado a observação de oscilações atmosféricas e oceânicas que estariam relacionadas ao comportamento climático em áreas remotas (Silva, 2013). Atualmente são conhecidos fenômenos atmosféricos que apresentam variabilidade marcada nas escalas intrasazonal, sazonal, interanual e decadal. Muitos estudos indicam a associação entre a ocorrência destes fenômenos localizados e a variabilidade climática em regiões remotas tais como indicaram os estudos de Ropelewski e Halpert (1987), Aragão (1998), Mason e Goddard (2001) e Nicholson e Grist (2003).

À influência entre as condições físicas localizadas em climas de regiões remotas dá-se o nome de teleconexão climática. Segundo Cavalcanti e Ambrizzi (2009), as teleconexões climáticas referem-se a um padrão recorrente e persistente de anomalias de uma determinada variável, que podem persistir por várias semanas ou meses e, algumas vezes, podem se tornar dominantes por vários anos consecutivos. Elas explicam, também, como as anomalias de uma determinada região são associadas às anomalias em regiões remotas. Já Carleton (2003) coloca as teleconexões em termos de relações estatísticas entre o clima e as variáveis meteorológicas em regiões geograficamente distantes. Tais relações podem ser positivas ou negativas e estar presentes por algumas décadas e ausentes em outras, além de que as regiões conectadas podem estar ou não na mesma faixa latitudinal. Na escala temporal, o sinal das teleconexões pode ser intersazonal, interanual ou interdecadal.

As teleconexões se manifestam por padrões recorrentes em grande escala na pressão atmosférica e nas variações dos ventos. Esses padrões climáticos agem conjuntamente em escalas espaciais e temporais diferentes, dessa forma, criam mecanismos de retroalimentação entre si. Contudo, são observadas interconexões e definições de alguns padrões que podem reforçar ou enfraquecer outras componentes, dependendo de suas escalas, porém, elas não se eliminam (Kayano e Andreolli, 2009 apud Silva, 2013).

Christofoletti (1979), faz abordagem geossistêmica ao explicar que os sistemas possuem elementos ou unidades, e estes se encontram inter-relacionados, uns dependendo dos outros, através de ligações que denunciam os fluxos. Este mesmo autor destaca a importância dos mecanismos de realimentação (feedback), onde mecanismos subsequentes podem voltar a exercer influências sobre os mecanismos antecedentes. Dentro desta perspectiva teórica, o

sistema climático opera através das inter-relações das unidades que o compõe, sendo que a alteração em uma destas unidades pode influenciar as demais.

A partir do conhecimento das interações climáticas induzidas pelas teleconexões, é possível obter-se alguma habilidade preditiva nas localidades afetadas. No entanto, esta capacidade preditiva difere para cada padrão de teleconexão. Em geral, nos trópicos as teleconexões costumam persistir por semanas ou meses e, portanto, a previsibilidade nas escalas sazonais e mensais é mais factível (Santos e Franco, 2011).

### 3.2.2 Relações com a variabilidade climática na América do Sul (ENOS)

A existência da forte interação entre o oceano e a atmosfera permitiu o entendimento de fenômenos originados do acoplamento entre estes dois sistemas. Desta forma, os principais padrões climáticos envolvem os oceanos Índico, Atlântico e Pacífico, com maior destaque para os dois últimos em relação à variabilidade climática global (Cavalcanti e Ambrizzi, 2009).

O Oceano Pacífico, devido às suas dimensões, permite maior interação oceano-atmosfera que os demais. Além disso, uma parte considerável de sua superfície encontra-se na zona tropical, exatamente a que recebe maior quantidade de energia solar. É razoável admitir, portanto, que a área tropical do Oceano Pacífico seja o palco de interações e acomodações capazes de interferir significativamente na variabilidade interanual do clima em escala muito mais ampla, com um período de acomodação muito maior (Varejão-Silva, 2006). Das principais configurações de teleconexões que se formam no Pacífico tem-se o El Niño Oscilação Sul (ENOS), que mais influencia nos padrões de anomalias extremas de precipitação em diversas localidades do globo (Trenberth, 1997).

O ENOS é resultado do acoplamento entre o fenômeno oceânico (El Niño/La Niña) e o fenômeno atmosférico (Oscilação Sul). O El Niño é um fenômeno oceânico interanual, caracterizado pelo aquecimento anormal das águas do Pacífico Equatorial, que provoca o enfraquecimento dos ventos alísios na região equatorial. Já o La Niña, apresenta comportamento contrário, caracterizando-se pelo resfriamento das águas do Pacífico e, conseqüentemente, no aumento dos ventos alísios. Dinamicamente relacionada está a Oscilação Sul (OS) com a troca de grandes quantidades de energia e massa na direção zonal (leste-oeste) da Bacia do Pacífico Equatorial e Tropical Sul, com dois centros de ação, um sobre a Indonésia e outro sobre o Taiti. Estes dois centros fazem parte da circulação de Walker, no sentido zonal,

com movimentos ascendentes e descendentes em áreas mais e menos aquecidas, respectivamente (Pinto, 2005; Varejão-Silva, 2006).

Os eventos de El Niño e La Niña podem durar aproximadamente de 6 a 18 meses por fase (Molion, 2008), totalizando cerca de 3 a 7 anos por ciclo completo (Varejão-Silva, 2006). Eles frequentemente começam a se formar durante o período de junho a agosto, alcançando o pico entre dezembro e abril. O arrefecimento do processo ocorre entre maio e julho (Pinto, 2005). As anomalias climáticas relacionadas são persistentes e duram vários meses, principalmente na atmosfera tropical. Chuvas acima do normal ocorrem no Peru, Equador e Ilhas do Pacífico central e leste. Existem também anomalias em latitudes extratropicais, como as temperaturas acima do normal no Alasca, Sudeste da Ásia, Sul e Sudeste do Brasil e chuvas acima da norma no Sudeste e Sul do Brasil, Uruguai e Norte da Argentina. (Aragão, 1998). A seguir, na Figura 3.1, temos as principais teleconexões do ENOS com várias regiões do globo.

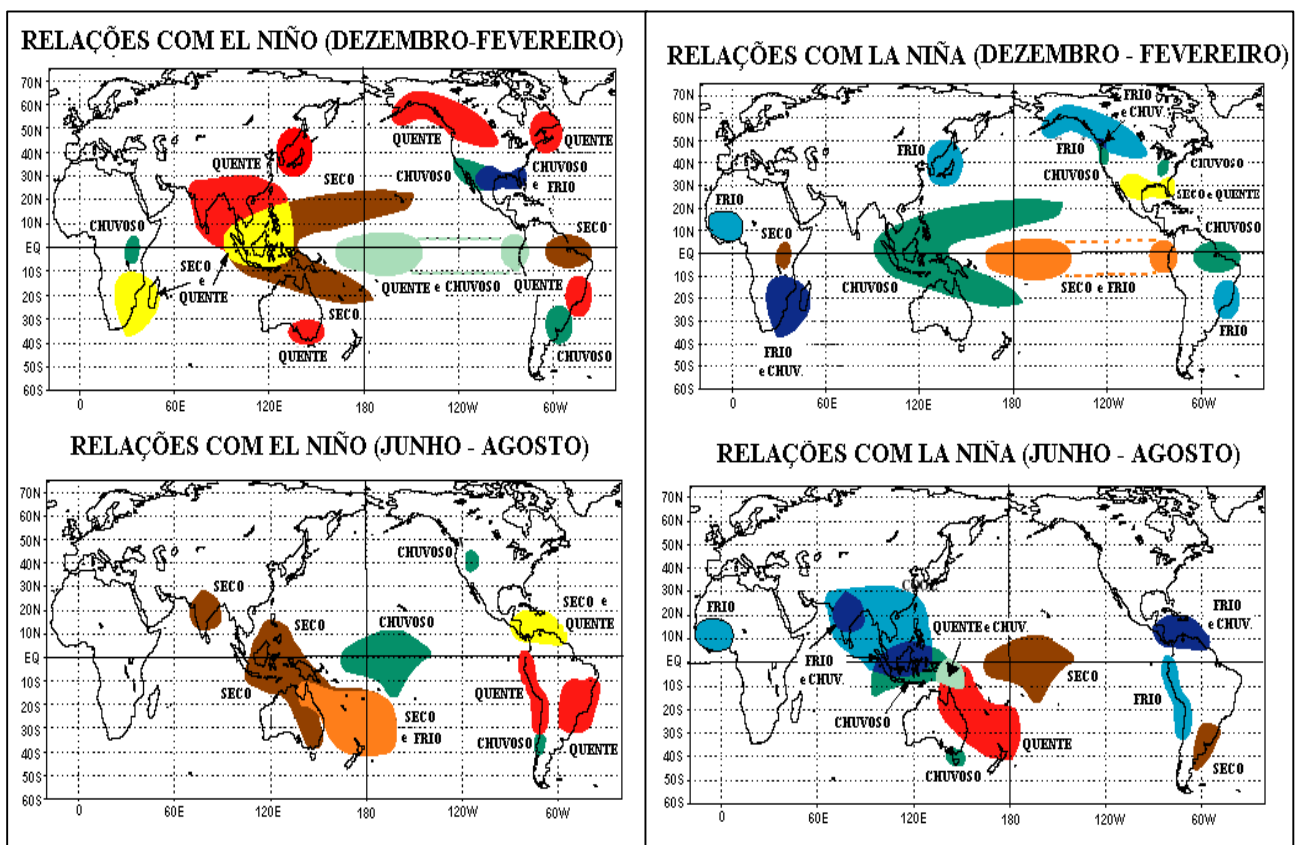


Figura 3.1 – Influências do El Niño/La Niña no globo, onde as cores indicam as áreas com as mesmas influencias climáticas. Fonte: CPC (2003).

Grimm (2003) apud Pinto (2005) observou que durante eventos de El Niño a precipitação e circulação das monções de verão da América do Sul são alteradas tanto por perturbações de grande escala associados ao ENOS como pelo aquecimento anômalo da superfície sobre o

sudeste do Brasil durante a primavera. Os mesmos eventos também parecem modular a frequência de eventos extremos de precipitação sobre regiões costeiras do sudeste do Brasil durante o verão, em associação com a intensificação do jato subtropical sobre o Atlântico Oeste. Em anos de El Niño, as frentes frias têm maior frequência sobre o Rio Grande do Sul e Uruguai, pois o jato subtropical se intensifica devido ao aumento do gradiente de temperatura norte-sul, bloqueando o avanço das frentes para as latitudes mais baixas (Cavalcanti,1996). A grande extensão latitudinal e as variadas formas de relevo da América do Sul permitem o desenvolvimento e a atuação de diferentes sistemas atmosféricos, que contribuem para a existência de 8 regimes de precipitação ao longo do continente (Reboita et al., 2010).

As teleconexões de grande escala, como a exemplo do ENOS, revelam flutuações de sistemas e componentes do tempo e do clima importantes para o estudo de previsões sazonais, pois as teleconexões que atuam em médio e longo prazo, ou seja, na escala da variabilidade climática, são observadas em escala intrassazonal, sazonal e anual, interanual, decadal a multidecadal. Esses modos em grande escala (ENOS, TSM do Atlântico Tropical) interagem entre si e, condicionam partes de outras componentes de curto prazo, ou seja, na escala do ritmo climático, como os Vórtices Ciclônicos de Altos Níveis – VCANs (verificado sobre o Nordeste Brasileiro), os Complexos Convectivos de Mesoescala – CCMs (observados em toda a América do Sul, principalmente no verão) e as Massas de Ar Continentais e Oceânicas (Nascimento Júnior, 2014). A Figura 3.2 mostra, de maneira ilustrativa, a localização dos sistemas atmosféricos atuantes na América do Sul, considerando-se a baixa e alta troposfera.

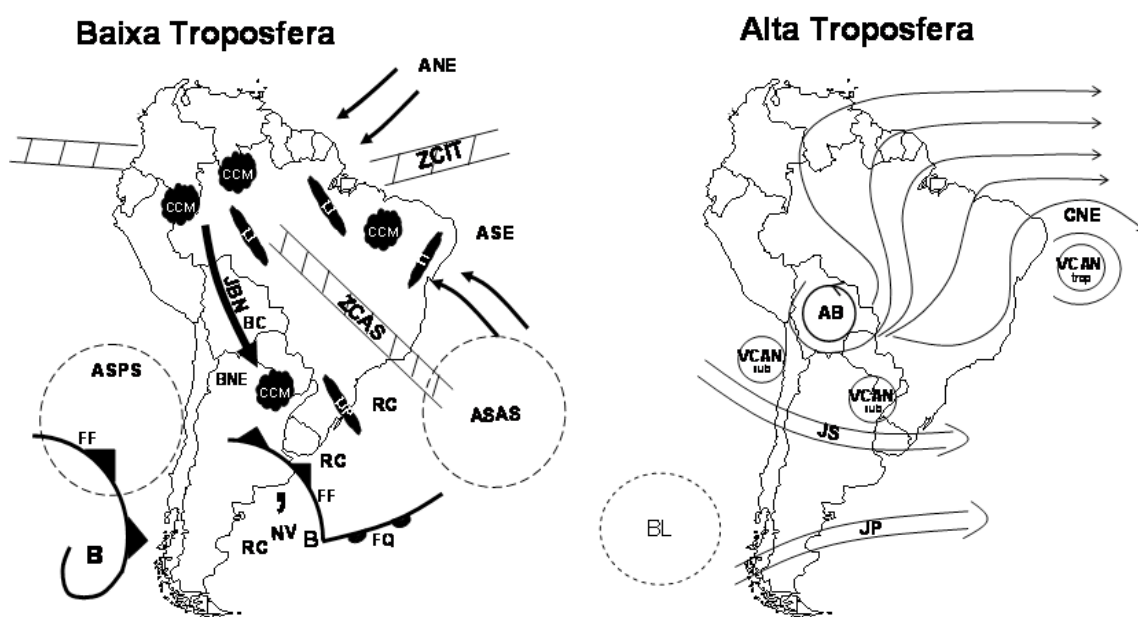


Figura 3.2 – Representação esquemática dos sistemas atmosféricos na baixa e alta troposfera atuantes na América do Sul. Fonte: Reboita et al. (2010).

Pode-se observar na Figura 2 que em relação à baixa troposfera tem-se: ANE - ventos alísios de nordeste, ASE - ventos alísios de sudeste, ASAS - anticiclone subtropical do Atlântico Sul, ASPS - anticiclone subtropical do Pacífico Sul, B - baixa pressão, BC - baixa do Chaco (região de baixas térmicas), BNE - região de baixas térmicas no noroeste argentino, CCM - complexo convectivo de mesoescala, FF - frente fria, FQ - frente quente, JBN - jato de baixos níveis a leste dos Andes, LI - linha de instabilidade tropical, LIP - linha de instabilidade pré-frontal, NV - nuvem vírgula, RC - regiões ciclogênicas, ZCAS - zona de convergência do Atlântico Sul e ZCIT - zona de convergência intertropical. Já em relação à alta troposfera tem-se: AB - alta da Bolívia, BL - região de bloqueios atmosféricos, CNE - cavado do nordeste do Brasil, JS jato subtropical, JP jato polar, VCAN - sub vórtices ciclônicos de altos níveis subtropicais, VCAN vórtices ciclônicos de altos níveis tropicais. É importante destacar que a AB e o CNE são sistemas que se configuram no verão e desaparecem no inverno. Além disso, no verão o JP e o JS se fundem em apenas um jato (Reboita et al., 2010).

Entende-se que os VCANs, CCMs e as Massas de Ar são inerentes na análise rítmica, devido à sua escala temporal ser curta e rápida. Entretanto, infere-se que é no estudo da variabilidade, em escala intrasazonal à multidecadal, que é possível observar esses padrões e modos de oscilação e, portanto, entender melhor seus efeitos na previsão da precipitação sazonal (Nascimento Júnior, 2014). O mesmo autor ainda coloca como importante que é preciso considerar, também, que esses modos de variabilidade interagem entre si e não se eliminam. Eles podem criar forçamentos e enfraquecimentos inclusive daqueles que atuam em escalas inferiores e diferentes.

### 3.2.3 Relações com Precipitação no Nordeste do Brasil

Os mecanismos dinâmicos que produzem chuvas no Nordeste do Brasil (NEB) podem ser classificados em mecanismos de grande escala, responsáveis por cerca de 30% a 80% da precipitação observada dependendo do local, e mecanismos de meso e micro escalas, que completam os totais observados (Molion e Bernado, 2002).

As relações entre os padrões anômalos de temperatura de superfície do mar TSM do Atlântico com o clima do Nordeste foram inicialmente abordadas nos artigos de Serra (1941) apud Marengo (2008). Fatores dinâmicos associados a reduções de chuva têm relação fundamentalmente com o efeito dos Oceanos tropicais (Atlântico e Pacífico), especialmente o contraste térmico entre os oceanos Atlânticos tropical do Norte e do Sul, que determina a



posição média da ZCIT e, portanto, de chuvas na região norte do Nordeste, incluído o semi-árido (Nobre e Molion, 1988 apud Molion e Bernado, 2002).

Em anos de seca no Nordeste brasileiro, a ZCIT fica bloqueada mais ao norte de sua posição normal. O Nordeste brasileiro fica, então, debaixo de uma região de subsidência que inibe a precipitação. Em anos chuvosos, ao contrário, a ZCIT move-se até cerca de 5°S e torna-se intensa com o aumento da convergência. As chuvas (secas) no Nordeste brasileiro, portanto, coincidem com o posicionamento da ZCIT mais para o sul (norte) (Uvo et al., 1998). Existem fortes evidências, observacionais, teóricas e resultados de modelos de circulação geral da atmosfera, que as condições oceânicas e atmosféricas sobre a Bacia do Atlântico Tropical influem fortemente na variabilidade interanual do clima sobre as Américas (Marengo, 2008; Hastenrath, 2011; Souza Filho, 2003). Na figura 3.3, Kucharski et al. (2007) demonstram as intensidades de variáveis climáticas do Oceano Atlântico tropical.

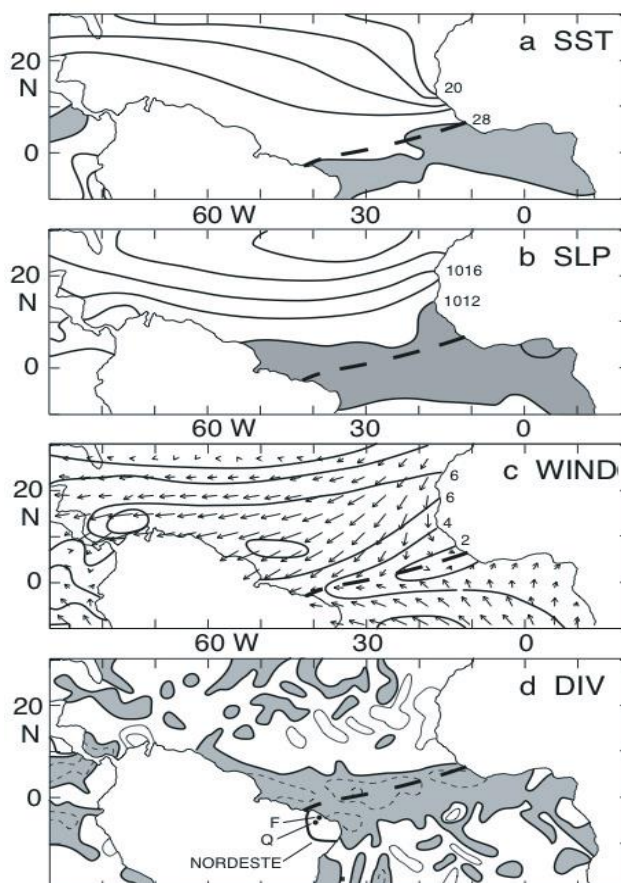


Figura 3.3 – Circulação superficial sobre o Atlântico equatorial e parte Pacífico oriental, de Março de 1958-1997. (a) Temperatura da superfície do mar, com espaçamento isotérmico de 2°C e com área acima de 28 ° C (cinza); (b) pressão do nível do mar, com espaçamento isobárico de 2mb, e com área inferior a 1012 mb (cinza); (c) a direção do vento resultante e velocidade, com espaçamento de 2 ms<sup>-1</sup>; (d) divergência com espaçamento isolinhas de 5 x 10<sup>-6</sup>s<sup>-1</sup>, e a convergência sombreada. A linha sólida em negrito limita o “Nordeste” do Brasil, e F e Q denotam as estações de Fortaleza e Quixeramobim.

Fonte: Kucharski et al. (2007).

O efeito do Oceano Pacífico Tropical especialmente durante anos de El Niño já foi identificado, desde o artigo de Walker (1928) apud Hastenrath (2011) em um estudo pioneiro, como sendo especialmente ligado a reduções drásticas de chuva durante a quadra chuvosa. Desde então, vários estudos que buscam um melhor entendimento dos mecanismos que causam as anomalias climáticas sobre o NEB têm focalizado no papel que o fenômeno ENOS exerce em tais anomalias (Gershunov e Barnett, 1998 apud Da Silva e Galvêncio, 2011; Giannini et al, 2004).

Essa ligação especial também já foi plenamente documentada para casos de estudo de 1983 e 1998 nas pesquisas desenvolvidas pelo CPTEC (Centro de Previsão de Tempo e Estudos Climáticos) e os centros estaduais de meteorologia da região (Marengo, 2008). Porém, anos secos no NEB podem ocorrer quando o El Niño não é identificado. Portanto, há outros fatores, além do ENOS que podem afetar as chuvas nestas regiões (Molion e Bernado, 2002).

Tratando-se de circulações de menor escala, segundo Ferreira e Mello (2005), sabe-se que a precipitação no NEB é também afetada pelo acoplamento de vários sistemas atmosféricos de várias escalas quase periódicos, como bandas de nebulosidade associadas a frentes frias, os Vórtices Ciclônicos de Ar Superior, os Distúrbios de Leste e as brisas terrestre e marítima, e, que podem ser modificados pelas características fisiográficas da região e pelas anomalias atmosféricas de escala planetária já comentadas, destacam-se o dipolo do Atlântico e o ENSO, que modificam a frequência, distribuição espacial e intensidade desses sistemas (Aragão, 1998). Esses fenômenos atuam em sub-regiões distintas e se superpõem em algumas sub-regiões, nas mesmas épocas ou em épocas diferentes (Molion e Bernado, 2002).

#### 3.2.4 Modos de Variabilidade e seus Índices Climáticos

Os modos de variabilidade dos fenômenos climáticos podem também ser expressos por meio de índices climáticos, os quais apresentam parte importante da oscilação observada. Um índice climático é um valor que pode ser usado para descrever o estado e as alterações no sistema climático. O clima de um lugar descreve o estado da atmosfera média durante um período prolongado de, por exemplo, meses ou anos. Alterações no clima acontecem muito mais lentamente do que para o tempo, já que este pode mudar bruscamente de um dia para outro. Os primeiros índices climáticos foram definidos a mais de um século atrás, como a Oscilação do Atlântico Norte (NAO), um dos primeiros padrões de teleconexão detectado (Hurrell, 1995).

Os índices climáticos permitem análises estatísticas, tais como comparar séries de tempo, meios de estimativas e a identificação de outliers e tendências. A associação entre os valores dos

índices climáticos e a variabilidade climática remota se baseia nas explicações teóricas sobre a dinâmica da atmosfera, como destacado por Walker (1924) e Rossby (1938) apud Santana (2013).

Há uma variedade de índices descritos na literatura uma vez que cada um se baseia em determinadas variáveis e descreve apenas determinados aspectos do clima. Desta maneira, cada um deles é definido por uma equação que utiliza variáveis que são determinantes no sistema climático de uma região atmosférica. Essas variáveis são: pressão, TSM, precipitação, velocidade dos ventos, radiação solar, temperatura do ar e cobertura de gelo, entre outros tipos.

Os índices climáticos mais simples são os de cálculos dos valores médios e extremos, tendências lineares e desvios padrão das séries temporais mais longas da(s) variável(s) de interesse. Os índices com base na temperatura do ar e precipitação são calculados a partir de dados medidos durante um longo período de tempo e, embora os resultados correspondam a uma localidade particular, é possível, dependendo da homogeneidade do lugar, extrapolá-los para áreas maiores (Quiroz, 2010). No entanto, não é comum, pois como a temperatura do ar ou eventos de precipitação abrangem características de áreas pequenas e altamente variáveis, de modo que as características de seus índices não poderiam ser extrapoladas para áreas maiores (Quiroz, 2010).

Os índices mais comuns são construídos a partir da pressão atmosférica. Quando a maioria dos meteorologistas faz menção aos índices climáticos, estão se referindo a padrões de teleconexão associados com as variações na pressão atmosférica em grande escala entre dois pontos da terra. Na literatura estes eventos são referidos como "gangorras" barométricas, como é usual em que uma extremidade a pressão é elevada, enquanto que na outra ocorre o oposto (Da Silva e Galvício, 2011). A mesma consideração de importância se tem para a TSM, onde medem a intensidade de suas anomalias, analisando a variação de temperatura para grandes células em diferentes regiões do oceano, ou a diferença entre células de duas regiões (Quiroz, 2010).

A Tabela 3.2 mostra um levantamento de diversos índices, pelo site <http://www.esrl.noaa.gov/psd/data/climateindices/list> do NOAA (National Oceanic and Atmospheric Administration), que são representativos dos diferentes padrões de variabilidade climática divididos por região oceânica e para o globo.

Tabela 3.2 – Levantamento de Índices climáticos e características

<b>Região</b>	<b>Índice</b>	<b>Descrição</b>	<b>Variáveis</b>
PACÍFICO	SOI	Southern Oscillation Index (Índice de Oscilação Sul)	Pressão
	SOI signal	Southern Oscillation Index Signal (Sinal do Índice de Oscilação Sul)	Pressão
	SOI noise	Southern Oscillation Index Noise (Ruído do Índice de Oscilação Sul)	Pressão
	NOI	Northern Oscillation Index (Índice de Oscilação do Norte)	Pressão
	WP	Western Pacific (Pacífico Ocidental)	Pressão
	NP	North Pacific Pattern (Padrão do Pacífico Norte)	Pressão
	PDO	Pacific Decadal Oscillation (Oscilação Decadal do Pacífico)	TSM
	TNI	Trans-Niño Index (Índice Trans-Niño)	TSM
	CTI	Cold Tongue Index (Índice Língua Fria do Pacífico)	TSM
	NINO 1+2	Extreme Eastern Tropical Pacific (Extremo Leste do Pacífico Tropical)	TSM
	NINO 3	Eastern Tropical Pacific (Leste do Pacífico Tropical)	TSM
	NINO 3.4	East Central Tropical Pacific (Centro Leste do Pacífico Tropical)	TSM
	NINO 4	Central Tropical Pacific (Pacífico Tropical Central)	TSM
	ONI	Oceanic Nino Index (Índice Oceânico do Nino)	TSM
	TP	Tropical Pacific (Tropical do Pacífico)	TSM
	TPI (IPO)	Tripole Interdec. Pacific Oscillation (Tripolo Interdecadal do Pacífico)	TSM
	PMM (sst)	Pacific Meridional Mode (Modo do Pacífico Meridional - TSM)	TSM
	PMM (wind)	Pacific Meridional Mode (Modo do Pacífico Meridional - Vento)	Velocidade do vento
	PNA	Pacific North American (Oscilação Pacífico-América do Norte)	Altura Geopotencial
	EPO	East Pacific Oscillation (Oscilação do Pacífico Leste)	Altura Geopotencial
MEI	Multivariate ENSO Index (Índice do Multivariado do ENOS)	Pressão, TSM, Vento, Nebulosidade	

Tabela 3.2 – Levantamento de Índices climáticos e características

ATLÂNTICO	NAO (NOAA)	North Atlantic Oscillation (Oscilação do Atlântico Norte) versão do NOAA	Pressão
	NAO (Jones)	North Atlantic Oscillation (Oscilação do Atlântico Norte) versão de Jones et al. (1997)	Pressão
	TNA	Tropical Northern Atlantic (Atlântico Tropical do Norte)	TSM
	TSA	Tropical Southern Atlantic (Atlântico Tropical do Sul)	TSM
	AMO	Atlantic Multidecadal Oscillation (Oscilação Multidecadal do Atlântico)	TSM
	AMO smoothed	Atlantic Multidecadal Oscillation (Oscilação Multidecadal do Atlântico) suavizado	TSM
	ATI	Atlantic Tripole Index (Índice do Tripolo do Atlântico)	TSM
	STA	South Tropical Atlantic (Tropical do Atlântico Sul)	TSM
	NTA	North Tropical Atlantic (Tropical do Atlântico Norte)	TSM
	EA	East Atlantic (Atlântico Leste)	Pressão
	AMM (sst)	Atlantic Meridional Mode (Modo do Atlântico Meridional - TSM)	TSM
	AMM (wind)	Atlantic Meridional Mode (Modo do Atlântico Meridional - Vento)	Velocidade do vento
ÍNDICO	OID	Indian Ocean Dipole (Dipolo do Oceano Índico)	TSM
ÁRTICO	AO (NAM)	Arctic Oscillation/ Northern Annular Mode (Oscilação do Ártico)	Altura Geopotencial
ANTÁRTICO	AAO (SAM)	Antarctic Oscillation/ Southern Annular Mode (Oscilação do Antártico)	Altura Geopotencial
	TPI	Trans Polar Index (Índice Trans Polar)	Altura Geopotencial
GLOBAL	QBO	Quasi-Biennial Oscillation (Oscilação Quasi-Biennial)	Velocidade do vento
	GMLOT	Global Mean Land-Ocean Temperature (Índice de Temperatura Global Média Terra-Oceano)	TSM
	GLAAM	Globally Int. Angular Momentum (Momento Angular Globalmente Integrado)	Rotação da Terra
	Solar Flux	Solar Flux (10.7cm) (Fluxo Solar)	Fluxo Solar

### 3.3 PROCESSAMENTO DE SINAIS

#### 3.3.1 Visão Geral pela Transformada de Fourier

O processamento de sinais consiste no método de analisar sinais do mundo real (representados por uma sequência de números) usando ferramentas matemáticas, podendo assim extrair informações desses sinais e/ou realizar transformações para torná-los mais apropriados para alguma aplicação específica (Oliveira, 2007).

A análise de séries temporais, ou análise de sinais, consiste basicamente em buscar similaridades entre um sinal (ou série histórica ou função) e funções matemáticas conhecidas, comumente carregadas de significado físico. Em matemática, o grau de semelhança entre dois sinais,  $x(t)$  e  $y(t)$ , pode ser indicado utilizando-se o produto escalar  $I_{XY}$  definido na equação 1 (Labat, 2005).

$$I_{XY} = \int_{-\infty}^{\infty} x(t)y(t)dt \quad (\text{Equação 1})$$

Se  $I_{XY}$  é igual a zero, as duas funções são ditas como sendo ortogonais. Qualquer sinal  $x(t)$  pode então ser projetado e analisadas utilizando um conjunto de funções  $y(t, \{p_i\}_{i=1, \dots, N})$ , em que  $\{p_i\}_{i=1, \dots, N}$  é um conjunto de parâmetros característicos. Desta forma, pode-se rescrever a equação 1 da seguinte maneira:

$$I_{XY}(p_i) = \int_{-\infty}^{\infty} x(t)y(t, p_i)dt \quad (\text{Equação 2})$$

De uma forma inversa,  $x(t)$  pode ser expandido (recomposto novamente) em termos de uma superposição linear do conjunto de funções  $y(t, \{p_i\})$ . Se as funções de  $y(t, \{p_i\})$  são ortogonais entre si, o sinal  $x(t)$  é, em seguida, caracterizado pelos valores do produto escalar entre o sinal  $x(t)$  e a função base de projeção  $y(t, \{p_i\})$  (Labat, 2005). Uma análise física bem conhecida baseia-se nestas propriedades, a Transformada de Fourier.

A análise de Fourier é um caso particular de análise de sinais, em que se busca aproximar um determinado sinal (ou determinadas séries históricas ou função), designado por  $f(t)$ , a uma combinação linear de componentes senoidais (senos e cossenos), em diferentes frequências ( $\omega$ ). Nessa situação particular, é feito uso da fórmula de Euler (Equação 3) como função base da

transformação, na qual o sistema de exponenciais complexas é relacionado com o sistema de senos e cossenos (Oliveira, 2007).

$$\mathbf{y}(t, \mathbf{p}_i) \Rightarrow \mathbf{y}(t, \boldsymbol{\omega}) = e^{i\boldsymbol{\omega}t} = \mathbf{cos} \boldsymbol{\omega}t + i \mathbf{sin} \boldsymbol{\omega}t \quad (\text{Equação 3})$$

A chamada Transformada de Fourier (TF) pode ser vista como uma projeção (transformação) de uma função (série temporal) do domínio tempo ( $t$ ) para o domínio da frequência ( $\boldsymbol{\omega}$ ) com o uso da equação 3.

$$\mathbf{F}(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} \mathbf{f}(t) e^{i\boldsymbol{\omega}t} dt \quad (\text{Equação 4})$$

O produto escalar  $\mathbf{F}(\boldsymbol{\omega})$  é a amplitude espectral do sinal  $\mathbf{f}(t)$  de acordo com a frequência  $\boldsymbol{\omega}$ . Esse produto corresponde à convolução do sinal através da base ortogonal trigonométrica definido por  $\{\exp(i\boldsymbol{\omega}_n t), n \in \mathbb{R}\}$ , onde  $\boldsymbol{\omega}_n$  é uma frequência discreta ( $\boldsymbol{\omega}_n = 2\pi n/T$ ).

Portanto a TF pode ser considerada como um limite de uma combinação linear infinita de ondas senoidais com um dado sinal. Dentro desta infinita combinação de ondas, teríamos como resultado da transformada, os coeficientes de Fourier  $\mathbf{F}(\boldsymbol{\omega})$ . Para melhor entendimento, suponha que  $\mathbf{f}(t)$  represente uma série temporal qualquer e  $\mathbf{y}(t, \boldsymbol{\omega})$  represente uma função filtro que tem o papel de identificar e selecionar o período de cada componente oscilatória que faça parte de  $\mathbf{f}(t)$ , a potência espectral dessa componente é justamente os coeficientes  $\mathbf{F}(\boldsymbol{\omega})$ . Essa análise é bastante usada em grandes aplicações no tratamento de sinais estacionários (Bolzan, 2006).

Entretanto, apesar de suas aplicações, podemos observar pela equação 4, que uma das deficiências da análise via TF de um sinal contínuo  $\mathbf{f}(t)$  é que ela não apresenta um caráter local. Todo o sinal, desde o começo dos tempos ( $-\infty$ ) até o fim dos tempos ( $+\infty$ ) e levado em consideração na integração. Desta forma, a TF representa um “comportamento global médio” do sinal (Oliveira, 2007). Esse não seria um problema se o sinal não mudasse durante o tempo (sinais estacionários), mas acontece que muitos sinais interessantes contêm muitas características não estacionárias ou transitórias (fluxos, tendências, mudanças repentinas, início e final de um evento). Essas características geralmente são a parte mais importante do sinal.

Bolzan (2006) destaca que a TF não fornece uma análise temporal, apenas frequencial, não fazendo distinção alguma sobre se alguma das funções senos foi somada ou não a série. Por esse motivo, tem sido utilizada a chamada transformada de Fourier por janelas (TFJ). Nesse

caso, uma janela de observação é deslocada no domínio do tempo e a transformada de Fourier da porção “visível” do sinal é calculada para cada posição da janela (Oliveira, 2007).

De uma maneira geral, a repetição infinita de funções seno e cosseno é bem adequada com grandes aplicações para o tratamento de dados que envolvam processos periódicos fixos (sinais estacionários). No entanto, este tratamento por TF não é bem adequado para os dados que envolve processos transitórios (Gurgel, 2013). Concentrando-se em hidrologia, flutuações de sinal são altamente não-estacionárias e processos físicos muitas vezes operam sob uma grande gama de escalas que variam de um dia para várias décadas, como a exemplo da precipitação e as vazões do rio (Sang, 2013).

O principal problema com a TJJ é que uma vez definido o tamanho para a janela, esta permanece constante para todas as frequências. Pode acontecer que alguns sinais precisem de um enfoque mais flexível, onde o tamanho da janela possa mudar com o objetivo de detectar o conteúdo local da frequência. Ou seja, nas frequências baixas existem tão poucas oscilações dentro da janela que a localização da frequência é perdida, enquanto que nas frequências altas existem tantas oscilações que a localização do tempo é perdida. Finalmente, a TJJ cai na suposição que o sinal pode ser decomposto em componentes senoidais, como já mencionado anteriormente (Oliveira, 2007; Santos, 2013).

### 3.3.2 Transformada de Wavelets (Ondeletas)

A maioria das séries temporais das ciências da Terra são geralmente não-estacionárias e apresentam tendências ao valor médio, mudanças na variabilidade para determinados períodos (Oliveira, 2007). Além disso, muitas séries temporais hidrológicas, tais como precipitação, apresentam eventos irregularmente distribuídos com potência não-estacionária ao longo de várias frequências. Assim, sua estrutura intrínseca temporal não é bem representada pela superposição de uma frequência de poucos componentes como a derivada de uma análise de Fourier (Santos, 2013).

Uma das ferramentas capaz de analisar séries-temporais obtidas de qualquer sistema físico é a Transformada de Wavelets (TW), criada por volta do início da década de 1980, por um pesquisador francês chamado Jean Morlet (Oliveira, 2007). Enquanto que a TF e a TJJ tem uma desvantagem significativa apresentando apenas os resultados no domínio da frequência ou com informações de tempo dentro de uma janela fixa, em contraste, a TW permite o uso de intervalos de tempo longos quando é necessária uma informação mais precisa de baixa



frequência, e intervalos mais curtos quando a informação de alta frequência é de interesse (Sang, 2013; Nourani et al., 2014).

O entendimento da TW se baseia de maneira análoga a TF, onde primeiramente observando a equação da TF (equação 4), pode-se constatar que cada função base – as exponenciais complexas ( $e^{-i\omega t}$ ) – é obtida a partir de simples dilatações na frequência ( $\omega$ ). Partindo dessa ideia, Morlet percebeu a necessidade de desenvolver uma outra função matemática base,  $\Psi$ , que além de ter duração finita (tem um início e um fim), pudesse ser capaz de dilatar ou comprimir a TW, eliminando o problema da janela fixa da TFJ. E para essa função base cobrir todo o eixo dos reais, a melhor maneira dada seria através de translações desta base em todo o comprimento da série temporal (Bolzan, 2006).

Reunindo estas duas propriedades, dilatação e translação, em um único protótipo de função base, obtemos as chamadas funções wavelets (ondeletas), onde as dilatações e as translações são dadas por duas variáveis nomeadas  $a$  e  $b$ , respectivamente. Portanto, o termo wavelet refere-se a um conjunto de funções com forma de pequenas ondas geradas por dilatações,  $\Psi(t) \rightarrow \Psi(at)$ , e translações,  $\Psi(t) \rightarrow \Psi(t + b)$ , de uma função base geradora simples  $\Psi(t)$ , a wavelet-mãe (equação 5). Logo, através da wavelet-mãe pode ser geradas as chamadas wavelet filhas (Labat, 2005).

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad (\text{Equação 5})$$

onde  $a$  e  $b$  são reais e  $a > 0$ . O termo  $a^{-1/2}$  corresponde a um fator de normalização para a energia da wavelet em escala dilatada permanecer igual à energia inicial da wavelet  $\Psi(t)$  (Labat, 2005).

Desta forma, a definição de uma transformada wavelet considerando um sinal contínuo  $f(t)$  é dada por:

$$\mathbf{W}_f(\mathbf{a}, \mathbf{b}) = \int_{-\infty}^{\infty} f(t) \Psi_{a,b}^*(t) dt \quad (\text{Equação 6})$$

onde os parâmetros  $a$  e  $b$  variam continuamente em  $\mathbf{R}$  e  $\Psi_{a,b}^*(t)$  é a função wavelet ou wavelet-mãe; o asterisco (\*) simboliza seu conjugado complexo. Os dois parâmetros  $a$  e  $b$ , correspondem às informações de escala e tempo respectivamente.

Formalmente, uma função wavelet  $\Psi(t)$  deve apresentar duas condições, a de admissibilidade (equação 7) e a de ter energia unitária (equação 8) (Oliveira, 2007):

$$\int_{-\infty}^{\infty} \Psi(t) dt = 0 \quad (\text{Equação 7})$$

$$\int_{-\infty}^{\infty} |\Psi(t)|^2 dt = 1 \quad (\text{Equação 8})$$

As condições apresentadas acima garantem que a função wavelet é de quadrado integrável e que pertence ao espaço  $L^2(\mathbb{R})$ . A primeira condição sugere que a função wavelet  $\Psi(t)$  apresenta características oscilatórias ao longo do eixo temporal e que possui média nula. A condição de admissibilidade também garante a existência da transformada wavelet inversa, que é a reconstrução da função em sua formatação original. A segunda condição está relacionada com a localização da energia (concentrada) da função wavelet em certa região finita do espaço. Esta condição de energia concentrada é a principal característica que diferencia as análises de sinais utilizando as funções wavelets e análise de funções utilizando as funções harmônicas (Análise de Fourier), cujas funções trigonométricas possuem energia infinita, ou seja, estas funções são limitadas apenas entre  $-\infty$  e  $\infty$  (Gurgel, 2013; Bolzan, 2006). Um exemplo de um “pacote” de onda, de duração finita e com uma frequência específica, é mostrada na figura 3.4.

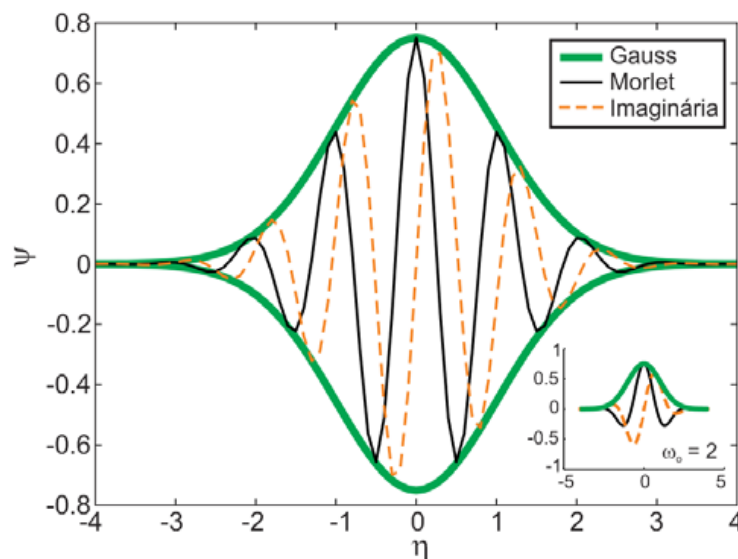


Figura 3.4 – A wavelet-mãe de Morlet.  
Fonte: Santos et al. (2013)

A wavelet de Morlet pode ser descrita pela equação 9:

$$\Psi_0(\eta) = \pi^{-0,25} e^{i\omega_0\eta} e^{-0,5\eta^2} \quad (\text{Equação 9})$$

onde  $\Psi_0(\eta)$  é o valor da wavelet no tempo adimensional  $\eta$ , e  $\omega_0$  é a frequência adimensional, igual a 6 neste caso para satisfazer uma condição de admissibilidade. No detalhe da figura 3.4 pode ser visto a wavelet de Morlet com  $\omega_0 = 2$ .

Na prática, a wavelet de Morlet mostrada na figura 3.4 é definida como um produto de uma onda exponencial complexa  $[\exp(i\omega_0\eta)]$  comprimida por um envelope gaussiano  $[\pi^{-0,25}\exp(-0,5\eta^2)]$  (Santos et al. 2013). A figura 3.5 dá uma demonstração mais clara didaticamente.

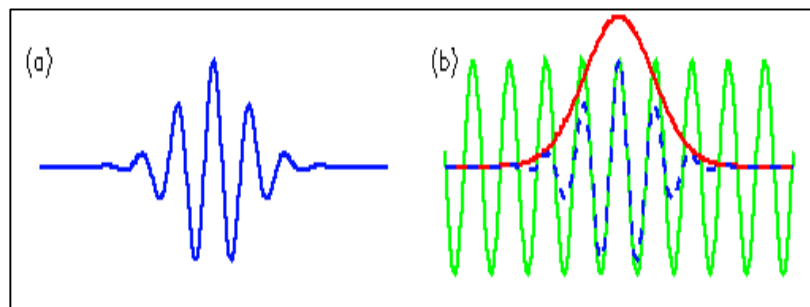


Figura 3.5 – A wavelet de Morlet (a) e a representação de sua formação através de um senoide e um envelope gaussiano (b). Fonte: Torrence e Compo (1998)

A principal propriedade da TW, que é derivada a partir do suporte compacto da sua função base, é o de proporcionar uma localização da escala de tempo de processos. Isto está em contraste com as funções harmônicas (seno e cosseno) usadas como função base das clássicas de análise de Fourier (Nourani et al., 2014).

A TW procura por correlações entre o sinal (série temporal) e a função wavelet, ou seja, ela tenta “ver” onde melhor se encaixa a wavelet-mãe comprimindo-a ou dilatando-a em vários intervalos de tempo (Figura 3.6d). Este cálculo é feito em diferentes escalas de  $a$  (Figura 3.6c) e localmente em torno do tempo  $b$  (Figura 3.6b). O resultado é um coeficiente de wavelet dentro de um mapa de contorno conhecido como escalograma ou periodograma da wavelet (Bolzan, 2006; Nourani et al., 2014).

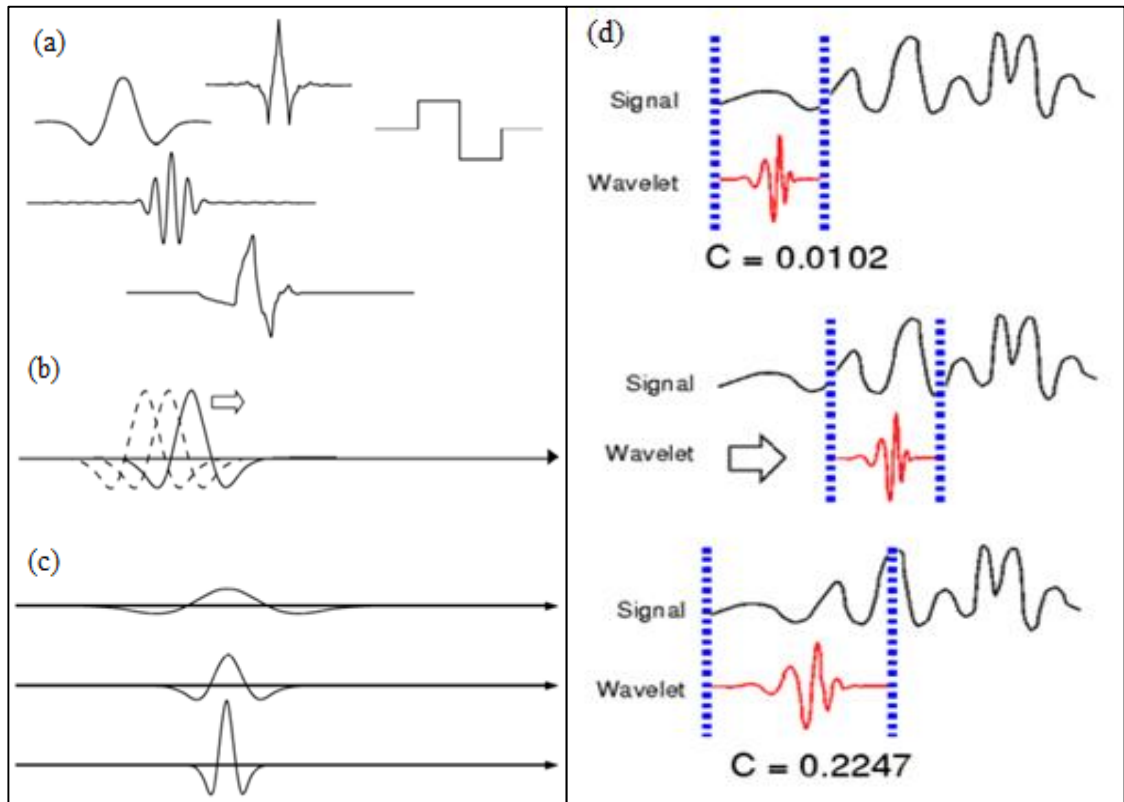


Figura 3.6 – (a) Diferentes tipos de wavelets-mãe, (b) translações e (c) dilatações da base e (d) processo de correlação da transformada de wavelets. Fonte: Software Matlab® (2014).

O que comumente se aceita é que a função ondeleta a ser utilizada possua um formato característico próximo das características encontradas na série-temporal. A função ondeleta de Morlet é a mais utilizada para analisar as várias frequências pertencentes aos sistemas hidrológicos e mostrar uma relação entre tempo e frequência (Santos, 2013; Nourani et al., 2014; Sang, 2013).

Santos et al. (2013) apresenta um excelente exemplo de como estas transformadas (Wavelet e Fourier) se diferem (Figura 3.7) mostrando uma série composta por várias frequências (7,5, 30, 60 e 120 Hz) e o seu respectivo espectro de potência Wavelet, respectivamente, na Figura 3.7a e Figura 3.7b.

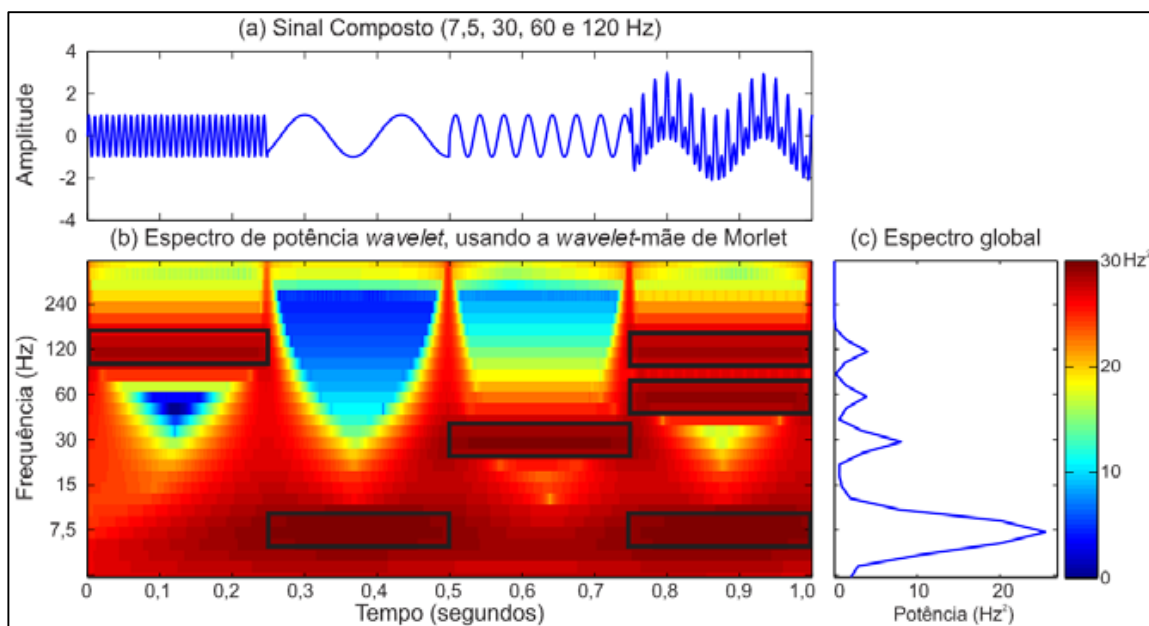


Figura 3.7 – a) Série composta pelas frequências 7,5, 30, 60 e 120 Hz. b) Espectro de potência Wavelet. c) Espectro de potência global Wavelet. Fonte: Santos et al. (2013).

Uma análise utilizando o espectro de Fourier forneceria apenas algum resultado parecido com o mostrado na Figura 3.7c, onde todas as informações referentes ao momento no qual os picos ocorreram seriam perdidas. Já o espectro de potência Wavelet (Figura 3.7b) mostra as informações de quando as referidas frequências ocorreram; por exemplo, a frequência 120 Hz ocorreu de 0 a 0,25s e depois de 0,75 a 1s. Desta forma seria possível notar que até os primeiros 75 s, ocorreram três tipos de frequência (7,5, 30 e 120 Hz) em momentos diferentes (Santos et al., 2013). Esta figura seria o mapa de contorno mencionado anteriormente, onde no eixo z a potência estaria representada, identificando onde as frequências ocorrem por meio de regiões mais altas, destacadas por um retângulo em negrito. Esse retângulo seria a identificação do nível de confiança desejado (95%) para uma boa correlação entre a wavelet-mãe e o sinal.

Comparativamente, a análise de wavelet é um método mais poderoso de análise de séries temporais. O emprego da wavelet em hidrologia, pode fornecer uma abordagem eficaz para a investigar a influência de padrões climáticos em séries temporais hidrológicas (Da Silva e Galvêncio, 2011). A análise de wavelet tem sido amplamente aplicada em recursos hídricos desde que os primeiros estudos apareceram na década de 90, e despontou nesses últimos 10 anos com surgimento de vários artigos como se tem mostrado nas revisões de Labat (2005), Labat (2010), Sang (2013) e Nourani et al. (2014).

### 3.3.3 Análise de Wavelets Cruzada e Coerência

Em muitos processos da natureza é interessante saber detalhes de sua ocorrência em termos de sua intensidade, bem como a correlação entre as grandezas que deram origem ao processo. Em meteorologia, por exemplo, muitas variáveis estão envolvidas, simultaneamente, na origem de fenômenos climáticos. A análise de Coerência Wavelet (TWC) é uma ferramenta matemática advinda do desenvolvimento da teoria das wavelets e que possibilita detectar a correlação e a semelhança entre eventos temporais, em termos de tempo-escala, frequência e energia (Oliveira, 2007). Antes de definir matematicamente a coerência wavelet é necessário falar sobre a análise de Wavelet Cruzada.

Para examinar a relação e a semelhança entre dois sinais foi criado o Espectro Wavelet Cruzada (XWT), caracterizada pelo módulo e pela fase da transformada wavelet obtida usando valores complexos. A fim de se obter uma visualização da correlação e semelhança entre dois sinais ou séries temporais, define-se o Espectro Wavelet Cruzada  $XW_{f,g}(a,b)$  entre dois sinais como sendo (Torrence e Compo, 1998):

$$XW_{f,g}(a, b) = W_f^*(a, b) \cdot W_g(a, b) \quad (\text{Equação 10})$$

onde  $W_f(a,b)$  e  $W_g(a,b)$  são os coeficientes das transformadas wavelet individual (conforme a equação 8) para cada sinal  $f(t)$  e  $g(t)$ , respectivamente; o asterisco (\*) simboliza o conjugado complexo. Dessa forma, o termo  $XW_{f,g}(a,b)$  da equação 10 expressa o produto dos coeficientes da transformada wavelet em uma dada escala  $a$ , na vizinhança temporal  $t = b$ . Nesse sentido, os coeficientes de  $XW_{f,g}(a,b)$  revelam quando existe um grau elevado de correlação entre  $f(t)$  e  $g(t)$ , para intervalos diferentes (especificados pela escala) na vizinhança de  $t = b$ .

Uma forma de enfatizar a correlação local dos sinais e é através da Coerência Wavelet que é introduzida por (Torrence e Compo, 1998):

$$WC_{f,g}(a, b) = \frac{|S(k^{-1}XW_{f,g}(a, b))|^2}{S(k^{-1}|W_f(a, b)|^2) \cdot S(k^{-1}|W_g(a, b)|^2)} \quad (\text{Equação 11})$$

Na equação 11 o fator  $k^{-1}$  é introduzido para normalizar a densidade de energia e  $S$  representa um operador de suavização em tempo e escala para a wavelet usada. A equação 11 possibilita representar a correlação entre os dois sinais por meio de um diagrama (mapa de cores ou periodograma) em função do tempo-frequência no qual é possível a visualização de eventos

coincidentes sobre as escala e frequências, em cada instante de tempo dos sinais (Labat et al., 2005).

A Coerência Wavelet apresenta valores entre 0 e 1. Para valores de  $WC_{f,g}(a,b)$  próximos de 1 a correlação é alta. Para valores de  $WC_{f,g}(a,b)$  próximos de 0 a correlação é baixa. Para melhor compreensão, essa equação pode ser comparada analogamente e similarmente com a equação da correlação de Pearson, onde a coerência entre os sinais é medida pela covariância. Uma das vantagens em ambas as aplicações das correlações é que na representação gráfica, através do escalograma, é dada a informação da frequência temporal e dos intervalos dos períodos em que ocorre maior correlação.

## 3.4 REDES NEURAIS ARTIFICIAIS

### 3.4.1 Conceitos e etapas de aplicação

Redes Neurais Artificiais (RNA) são modelos computacionais inspirados no sistema nervoso biológico e do próprio cérebro humano. Possuem a capacidade de aquisição e manutenção do conhecimento (baseado em informações) e podem ser definidas como um conjunto de unidades de processamento, caracterizadas por neurônios artificiais, que são interligados por um grande número de interconexões (sinapses artificiais), sendo as mesmas representadas por vetores/matrizes de pesos sinápticos (Silva et al. 2010).

Os modelos matemáticos baseados em RNAs se apresentam como uma solução relativamente recente para o problema de prever a saída de sistemas complexos, utilizados em uma ampla gama de campos, tais como a ciência ambiental, o reconhecimento de padrões, o sensoriamento remoto e a gestão financeira (Dawson e Wilby, 2001). Esses modelos se demonstram particularmente úteis em situações nas quais as relações dos processos físicos ainda não são completamente compreendidas. Os modelos RNAs podem, ainda, substituir os modelos matemáticos e estatísticos convencionais ou se associar a eles (Govindaraju e Rao, 2000).

Embora os primeiros trabalhos com RNAs tenham sido publicados há mais de 50 anos, tal tema começou a ser fortemente pesquisado a partir do início dos anos de 1990, tendo ainda um potencial de pesquisa imenso até hoje (Silva et al., 2010). E foi nesse início de 1990, que os modelos de RNAs começaram a ser utilizados com sucesso em áreas relacionadas à hidrologia, tais como modelagem chuva-vazão, previsão de vazões, modelagem de água subterrânea,

qualidade da água, gestão da água, previsão de precipitação, séries temporais hidrológicas e operação de reservatório (Govindaraju e Rao, 2000).

Por sua vez, nesta última década, houve um aumento muito expressivo na aplicação de modelos de RNA como uma abordagem mais apropriada para previsão hidrológica de longo prazo, incluindo a precipitação sazonal (Nourani et al., 2014; Mekanik et al., 2013). Esse aumento ocorreu devido a tais modelos serem capazes de lidar com a não-linearidade envolvida nos processos hidrometeorológicos e com grande quantidade de dados. Uma série de trabalhos de revisão sobre os fundamentos matemáticos da RNA e aplicações específicas em hidrologia, com ênfase na sua utilização para prever variáveis hidrológicas, contribuíram para esse crescimento no início desse século (ASCE, 2000a; ASCE, 2000b; Maier e Dandy, 2000; Govindaraju e Rao, 2000; Dawson e Wilby, 2001).

Desde então, RNAs têm sido estudadas para a previsão de precipitação sazonal em muitas partes do mundo, incluindo a Grécia (Nastos et al., 2013, Moustiris et al., 2011), Japão (Iseri et al., 2005), Estados Unidos (Silverman e Dracup, 2000), Irã (Morid et al., 2007; Karamouz et al. 2008), vários trabalhos na Índia (Kumar et al., 2007; Srivastava et al., 2010; Shukla et al., 2011), e mais recentemente em crescimento na Austrália (Mekanik et al., 2013; Aboot e Marohasy, 2014; Deo e Sahin, 2015; He et al., 2015). No Brasil, exemplos de estudo desse tipo são os trabalhos de Bravo (2010), Paz et al. (2011), Costa (2014).

Os sistemas que utilizam redes neurais buscam sua principal propriedade, a habilidade de aprender a partir de seu ambiente e, com isso, extrair conhecimento. Dentro disso, as características mais relevantes envolvidas com a aplicação de RNAs são (Haykin, 2001): adaptação por experiência; capacidade de aprendizado; habilidade de generalização; organização de dados; tolerância a falhas; armazenamento distribuído.

Modelos de RNAs tem um conjunto de elementos intrínsecos chamados neurônios artificiais, as unidades fundamentais de processamento. Eles são dispostos em camadas de processamento e interconectados entre si por ligações (sinapses). A força da ligação entre cada dois neurônios é medida pelo peso sináptico, e os pesos de todas as ligações constituem os parâmetros a serem definidos por meio de um processo de otimização denominado de treinamento (Paz et al., 2011). Cada único neurônio da camada intermediária se acumula, através de uma soma ponderada (Equação 12), de todas as informações enviadas pelos neurônios interconectados na camada anterior. A quantidade total de contribuições recebidas é chamado de "potencial de ativação". Este valor de potencial é aplicado em uma função de ativação para obter uma única resposta ou



saída (Figura 3.8). Esta função de ativação é uma função característica de todos os neurônios, com exceção dos neurônios de entrada.

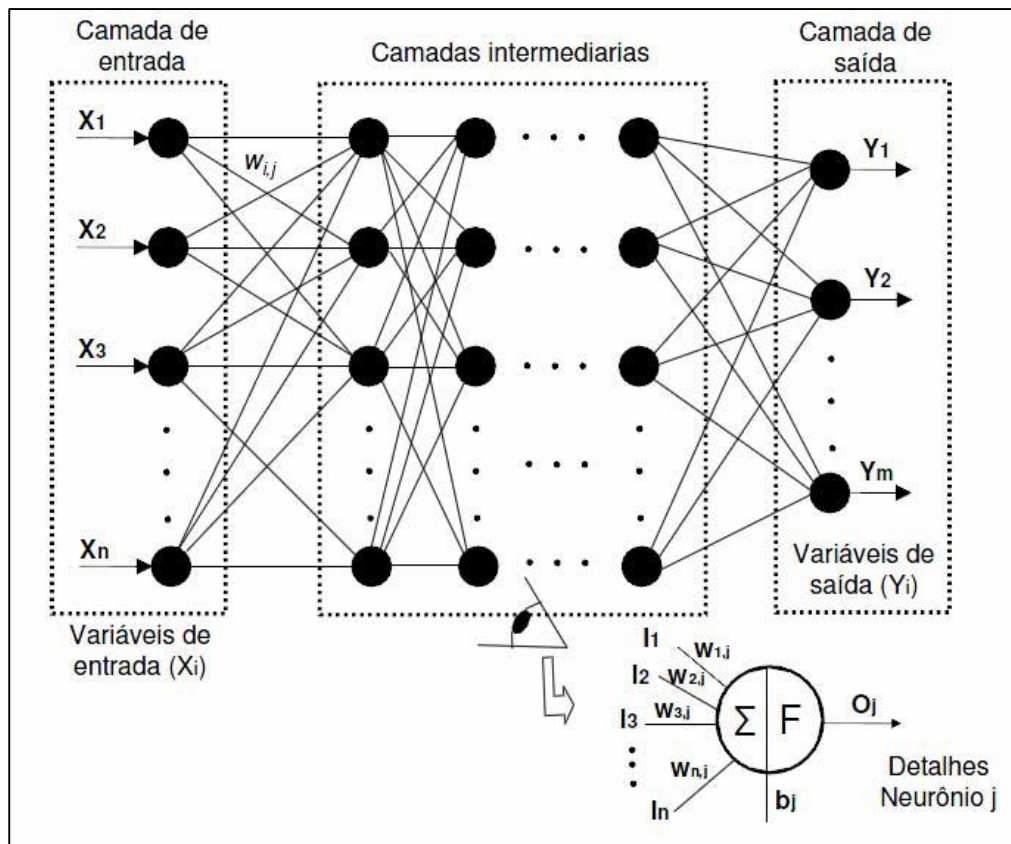


Figura 3.8 – Modelo de uma Rede Neural Artificial e detalhe do funcionamento de um dos neurônios. Fonte: Bravo (2010).

$$O_j = F \left( \sum_{i=1}^n w_{i,j} \cdot I_i - b_j \right) \quad (\text{Equação 12})$$

onde  $O_j$  é a saída do neurônio  $j$ ,  $F$  é a função de ativação (não linear),  $w_{i,j}$  são os pesos das ligações entre os neurônios  $j$  e seus neurônios interconectados,  $I_i$  são os sinais de entradas e  $b_j$  é o limiar de ativação do neurônio (bias).

Assim, pode-se resumir o funcionamento de um neurônio artificial por meio dos seguintes passos (Silva et al. 2010):

- apresentação de um conjunto de valores que representam as variáveis de entrada do neurônio;
- multiplicação de cada entrada do neurônio pelo seu respectivo peso sináptico;
- obtenção do potencial de ativação produzido pela soma ponderada dos sinais de entrada, subtraindo-se o limiar de ativação;

- d) aplicação de uma função de ativação apropriada, tendo-se como objetivo limitar a saída do neurônio;
- e) compilação da saída a partir da função de ativação neural em relação ao seu potencial de ativação.

Como visto na Figura 3.8, tipicamente a topologia de um modelo de RNA tem uma camada de entrada, uma ou mais camadas escondidas (intermediárias) e uma camada de saída. Na maioria dos modelos de RNA, a camada de entrada recebe as variáveis de entrada, por exemplo, as variáveis de previsão, fornecendo a base de informações do modelo. Através da camada oculta essa informação é processada e passada para a camada de saída, onde os produtos finais da RNA, por exemplo, os valores preditos, são gerados.

Segundo Paz et al. (2011), os passos no desenvolvimento de um modelo de RNA de previsão podem ser resumidos como:

- a) definição das variáveis de previsão (preeditando);
- b) identificação e seleção das variáveis de entrada (preditoras);
- c) definição da arquitetura da RNA;
- d) divisão dos dados;
- e) treinamento (definição de pesos da RNA e bias);
- f) validação ou avaliação de desempenho.

Esses passos são discutidos nas seções seguintes.

### 3.4.2 Definição das Variáveis de Previsão

Ao construir um modelo RNA de previsão, a primeira questão diz respeito à definição do preeditando ou variável de previsão. Esta definição está intimamente relacionada com o objetivo da previsão, mas também é dependente dos dados disponíveis, já que as relações orientadas aos dados devem ser construídas com base no conjunto de dados históricos (Paz et al., 2011). Em termos de previsão de precipitação sazonal, as variáveis de previsão usuais são a precipitação total ao longo de toda a estação chuvosa ou a precipitação acumulada durante um período de tempo específico não período úmido (Srivastava et al., 2010), que pode ser por exemplo a precipitação mensal no início da estação chuvosa (Paz et al., 2011; Aboot e Marohasy, 2014) ou a precipitação acumulada em um dado trimestre (Bravo, 2010; Costa, 2014). Também há o caso de se trabalhar com a média da estação chuvosa ou de um período específico (Nastos et

al., 2013). De qualquer forma, é importante analisar as características climáticas da região de estudo, tais como a sazonalidade da precipitação e a precipitação média mensal.

Para fornecer a previsão de precipitação específica em uma determinada região (por exemplo, uma bacia ou uma área intensamente utilizada para a agricultura), os registros de precipitações disponíveis das estações pluviométricas localizadas dentro e em torno desta área podem ser combinados em uma única variável que representaria a precipitação média regional, sendo considerada como a variável de previsão, como pode ser observado nos trabalhos de Srivastava et al. (2010) para toda a Índia e Bravo (2010) para uma bacia afluente ao reservatório de Furnas em Minas Gerais. Esses registros também podem ser divididos em microrregiões como apresentado em Costa (2014) em uma bacia tropical/semiárida no NEB ou por About e Marohasy (2014) em uma região de Queensland na Austrália.

Devido à grande variabilidade espacial da precipitação, a abordagem de usar a média acumulada pode levar a uma variável que melhor representa o regime de precipitação sobre a região de interesse, em comparação ao uso de registros de precipitação isolados de uma estação de medição, que poderia conter particularidades locais nos dados (Paz et al., 2011). Mesmo assim, há trabalhos que preferam realizar previsões para cada estação pluviométrica, dentro de cada região, como o estudo de Mekanik et al. (2013) no sul da Austrália.

O modelo de previsão com RNA pode ser concebido para prever mais do que uma variável de tempo, se necessário. Neste caso, a camada de saída do modelo RNA teria dois ou mais neurônios. Por exemplo, um modelo de RNA poderia ser desenvolvido para prever precipitação mensal em fevereiro, março e abril simultaneamente. No entanto, quando um modelo único de RNA é desenvolvido para prever diversas variáveis ao mesmo tempo, a representação da relação individual subjacente entre os preditores e cada preditando específico pode tornar-se enfraquecida em favor de representar a relação global entre os preditores e todos os preditandos simultaneamente (Paz et al., 2011). Desta forma, tais autores destacam que o procedimento mais eficiente seria o de desenvolver um modelo específico de RNA para a previsão de cada uma das variáveis de interesse.

Se o modelo de previsão de precipitação é desenvolvido para fins agrícolas, a variável de previsão também pode ser escolhida de acordo com a sua utilidade para o campo de decisões relacionadas às culturas e regiões específicas, tendo em conta as características climáticas da região e sua relação com as práticas de cultivo (Meza, 2008). A exemplo disto, o estudo de

Srivastava et al. (2010) decidiu por prever a precipitação do mês de agosto na Índia, pois é quando inicia a fase crítica de crescimento do arroz, regulando a sua produção/produktividade da principal cultura da agricultura indiana.

Finalmente, o resultado deste passo pode ser composto por um conjunto de potenciais variáveis de previsão, a partir do qual o preditor final é ainda escolhido de acordo com o poder de predição encontrado para as variáveis de previsão selecionadas, bem como de acordo com o desempenho dos modelos desenvolvidos.

### 3.4.3 Identificação e Seleção das Variáveis Predictoras

Um dos passos mais importantes no processo de desenvolvimento de um modelo de RNA de previsão é a determinação de um conjunto adequado das variáveis predictoras. No entanto, nesta tarefa é geralmente dada pouca atenção na modelagem de RNA e a maioria das entradas são determinadas numa base ad-hoc ou usam, a priori, apenas algum sistema conhecido. Isto pode resultar na inclusão de muito poucos ou demasiados dados na entrada do modelo, ambas as situações são indesejáveis (Maier et al., 2010). Essa seleção das variáveis de entrada tem como preocupação decidir qual a informação no passado e/ou presente deve ser usada para prever o valor futuro de uma determinada variável.

Os preditores representam as variáveis de entrada para o modelo, isto é, a informação a partir da qual a rede irá produzir as previsões. Este conceito pode dar origem a uma ideia errada de incluir o maior número de variáveis de previsão quanto possível com o objetivo de fornecer o máximo de informações para a RNA. Uma restrição geral para o número de variáveis de entrada de um modelo de RNA está relacionada com a quantidade de registros dos dados disponíveis (Dawson e Wilby, 2001). O uso de muitas variáveis preditivas com poucos registros de dados pode levar a RNA ao super ajustamento (overfitting), onde a RNA irá reproduzir bem as saídas para qual o modelo foi desenvolvido, mas não terá a capacidade de generalização, ou seja, não será capaz de representar bem um conjunto de dados independentes (Paz et al., 2011). Além de que, o uso demasiado de variáveis pode trazer informações redundantes, aumentando também a probabilidade de overfitting, enquanto que não fornecem qualquer informação adicional para o modelo (Maier et al., 2010).

O ideal seria que apenas as variáveis com possível e razoável relação física com a variável preditanda devessem ser selecionadas como potenciais preditores. Apesar da modelagem com RNAs não exigir a existência de tal relação, essa observação é importante para limitar as

potenciais variáveis de previsão para as relações com o fenômeno físico subjacente. Na verdade, este é recomendado para qualquer método de previsão estatística (Wilks, 2006).

A escolha das variáveis de entrada requer um estudo detalhado das condições físicas e meteorológicas que mais influenciam a variável hidrológica que será prevista (Bravo, 2010). Para a previsão da precipitação sazonal, tanto as informações coletadas na área de previsão quanto as relacionadas a regiões remotas, podem ser usadas como preditores, pois depende de fenômenos complexos, de envergadura continental e mundial, que interagem alterando o comportamento normal do sistema (Mason et al., 1996; Shukla et al., 2011). O primeiro caso compreenderia as condições meteorológicas antecedentes no tempo na área de previsão, tais como temperatura e a própria precipitação. Por exemplo, a precipitação acumulada num dado mês pode ser utilizada como entrada para prever a precipitação no mês seguinte (Nastos et al., 2013).

O segundo tipo de dados utilizados como preditores baseia-se no pressuposto de teleconexões climáticas existentes, ou seja, que uma anomalia climática em uma parte da Terra pode levar a outra anomalia climática em outra região distante (Silverman e Dracup, 2000). As variáveis relacionadas com as teleconexões são comumente utilizadas como preditores para a previsão sazonal da precipitação, e compreendem a temperatura da superfície do mar e variáveis relacionadas com padrões de circulação atmosférica, como as anomalias de pressão atmosférica a altura de 700 mb, e os índices climáticos (Kumar et al., 2007; Morid et al., 2007).

O uso de índices climáticos é preferível em relação ao uso de medidas diretas, visto que os índices são representativos da variação da intensidade, ao longo do tempo, dos modos de variabilidade atmosférica e de fenômenos climáticos. Além de que, acabam sendo relativamente mais bem descritos e compreendidos tornando-se mais plausível para explicar a relação física entre os preditores e a precipitação na área de interesse (Aboot e Marohasy, 2014).

Mesmo se apenas as variáveis fisicamente relacionadas às condições de precipitação da região de estudo fossem selecionadas, vários preditores potenciais poderiam ser identificados. Dentro disso, particularmente é importante identificar o tempo de antecedência mais adequado para cada variável de previsão (Paz et al., 2011).

Há uma grande variedade de técnicas para realizar a tarefa de definir as melhores variáveis predictoras. Um procedimento tradicional usado para qualquer método de previsão estatística é a regressão seletiva, através de técnicas como a seleção para a frente, a seleção por etapas ou

eliminação para trás (Mason et al., 1996; Wilks, 2006), ou métodos mais complexos, como o uso de transformadas matemáticas para extração das características principais dos dados como a técnica *Partial Mutual Information* (PMI) (He et al., 2015), que pode detectar a dependência não linear entre variáveis preditoras e preditando. Outras metodologias utilizadas para definir variáveis de entrada de modelos de RNA incluem a análise de componentes principais (Shukla et al., 2011) e algoritmos de Ranqueamento Fuzzy (Srivastava et al., 2010).

Maier et al. (2010) comentam sobre o uso de opções baseadas em abordagens Ad-Hoc nos trabalhos de previsão hidrológica com RNA, onde o desenvolvedor do modelo seleciona quais combinações de entrada do modelo devem ser testadas. Também há uma abordagem por etapas, onde as variáveis de entradas são sistematicamente acrescentadas (processo construtivo) ou removidas (processo de poda), ou uma abordagem global, onde um algoritmo de otimização global, como um algoritmo genético, é usado para selecionar a combinação de variáveis que maximiza o desempenho do modelo.

Uma abordagem mais simples e prática para selecionar variáveis de previsão é o uso de testes de correlação, que através de uma medida de significância estatística pode identificar as relações lineares fortemente causais dos potenciais preditores (Dawson e Wilby, 2001; Wilks, 2006). Há também a utilização de medidas de dependência estatísticas não-lineares, como a correlação de Spearman usada em Paz et al. (2011) e Costa (2014). As análises de correlação têm sido amplamente utilizadas para a seleção das variáveis de entrada para os modelos de RNAs com resultados satisfatórios (Kumar et al., 2007; Bravo, 2010), apesar de algumas análises apresentarem a limitação de apenas avaliar as relações lineares (Maier et al., 2010).

Analisando a correlação entre o preditando e cada variável de previsão individualmente e considerando diferentes tempos de espera, pode-se indicar aproximadamente as variáveis preditoras adequadas e seu tempo de antecedência (Kumar et al., 2007; Bravo et al., 2010). Por exemplo, dado que um índice climático mensal foi selecionado como potencial indicador, a análise de correlação é útil para estimar o poder preditivo de cada mês desse índice específico e definir qual deles usar (Shukla et al., 2011).

#### 3.4.4 Arquitetura da RNA e Divisão de Dados

Uma vez que as variáveis de entrada e saída da RNA são estabelecidas, o próximo passo é a definição do tipo de arquitetura da RNA e a configuração de sua topologia. A arquitetura da rede define a forma como os seus diversos neurônios estão dispostos, uns em relação aos outros,

assim como suas formas de interligação entre eles e a constituição de suas camadas. Já a topologia de uma rede neural, considerando determinada arquitetura, pode ser definida como sendo diferentes composições estruturais que esta poderá assumir, como o número de camadas ocultas, o número de neurônios por camada, e as funções de ativação de camadas ocultas e de saída (Silva et al., 2010).

Considerando a forma na qual as entradas são apresentadas à rede, existem duas classes de arquiteturas: Redes Recorrentes (RNN) (feedback) e redes de alimentação adiante ou direta (feedforward), sendo que esta última subdivide-se em duas classes, as arquiteturas do tipo camada única e do tipo múltiplas camadas, que são comumente denominadas de MLP (MultiLayer Feedforward Perceptron) (Silva et al., 2010).

As redes de alimentação direta se caracterizam pela propagação dos sinais em apenas uma direção a partir da camada de entrada, passando pela camada intermediária (se houver), até a camada de saída e não vice-versa. No caso da camada única, há a presença apenas da camada de entrada e da camada de saída, onde se conta somente a camada de saída, pois lá é onde é realizada a computação. A de múltiplas camadas se diferencia pela presença de uma ou mais camadas intermediárias, cujos nós computacionais são chamados correspondentemente de neurônios ocultos ou unidades ocultas (Figura 3.9a). A função dos neurônios ocultos é intervir entre a entrada e a saída da rede de uma maneira útil (Haykin, 2001).

As redes recorrentes se distinguem das redes de alimentação direta por terem pelo menos um laço de realimentação (feedback), apresentando conexões entre neurônios de uma mesma camada ou camadas anteriores (Figura 3.9b). A presença de laços de realimentação tem um impacto profundo na capacidade de aprendizagem da rede e no seu desempenho (Haykin, 2001).

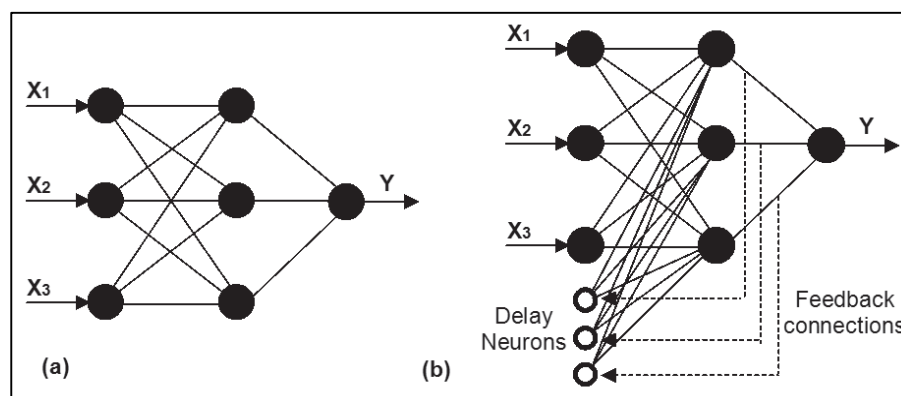


Figura 3.9 – Tipos de modelos de RNA de três camadas com três variáveis de entrada e uma de saída: (a) Multilayer Feedforward (MLP); (b) Rede Neural Recorrente (RNN). Fonte: Paz et al. (2011).

Os modelos MLP são amplamente utilizados na previsão de variáveis hidrometeorológicas (ASCE, 2000; Govindaraju e Rao, 2000; Maier et al., 2010). A utilização de funções de ativação não lineares nas camadas intermediárias de uma MLP intervém entre a entrada e a saída da rede de maneira muito útil, pois fornece a RNA a capacidade de capturar a complexidade e a não linearidade inerente nos sistemas a serem modelados (Maier et al., 2010). Há vários exemplos da utilização desse modelo para previsão de precipitação, como Philip e Joseph (2003), Kumar et al. (2007), Srivastava et al. (2010), Paz et al. (2011), Shukla et al. (2011), Mekanik et al. (2013) e Nastos et al. (2013).

Enquanto as arquiteturas de rede de alimentação direta são as mais populares entre os pesquisadores na previsão de variáveis hidrometeorológicas, as redes neurais recorrentes também têm recebido alguma atenção, tal como apresentado por Aboot e Marohasy (2014) e Karamouz et al. (2008). A existência de um mecanismo de feedback em redes recorrentes, torna mais simples para uma rede neural, a priori, a modelagem de sistemas altamente dinâmicos com atrasos de tempo (Maier et al., 2010).

Definida a arquitetura da RNA, parte-se então para a sua topologia, ou seja, para a definição do número ótimo de camadas intermediárias e o número de neurônios em cada uma delas. Os métodos para essa definição podem ser classificados em três tipos: global, passo a passo, ou ad-hoc (Maier et al., 2010). No primeiro método, a estrutura de um modelo de RNA em termos de camadas ocultas e quantidade de neurônios é encontrada pela utilização de métodos globais com base em evolução competitiva como por exemplo, algoritmos genéticos.

No segundo método, há um processo de poda ou adição, onde a topologia da RNA estará definida inicialmente para um grande número de neurônios na camada oculta (processo de poda), ou então inicialmente um pequeno número de neurônios (processo de adição). Desta forma, um neurônio ou conexão vai sendo removido ou adicionado dependendo do processo (poda/adição), sucessivamente, após cada avaliação de treinamento e desempenho (Maier et al., 2010). Este procedimento continua até que uma configuração com desempenho satisfatório seja conseguida com o número ideal de neurônios nas camadas ocultas. Um exemplo claro de utilização desse método é encontrado no trabalho de Deo e Sahin (2015), onde ele inicia com um número máximo de 50 neurônios testados e decresce até 12; ou o estudo de Sousa e Sousa (2010), que simularam uma rede com um número de 5, 10 e 15 neurônios na camada intermediária.



A terceira abordagem também relaciona um processo de tentativa e erro, na qual o modelador seleciona uma estrutura de rede com base na experiência e/ou intuição, classificado então como ad-hoc (Maier et al., 2010). Tradicionalmente o processo de experimentação sistemática é usado para a busca de melhores resultados, ou com uso do processo de poda/adição ou de maneira ad-hoc (Maier e Dandy, 2000; Maier et al., 2010).

A utilização de um grande número de camadas escondidas para uma rede MLP não é recomendada, pois segundo Silva et al. (2010), cada vez que o erro médio durante o treinamento é utilizado para atualizar os pesos das sinapses das camadas imediatamente anteriores, ele se torna menos útil ou preciso. A única camada que tem uma noção precisa do erro cometido pela rede é a camada de saída. A última camada oculta recebe uma estimativa sobre o erro. A penúltima camada oculta recebe uma estimativa da estimativa, e assim por diante. Praticamente todos os trabalhos citados neste tópico fazem uso de somente uma camada intermediária, onde destaca-se uma observação para o trabalho de Kumar et al. (2007), que também faz uso de topologias com duas camadas intermediárias.

Por fim, na etapa final da escolha da topologia são selecionadas as funções de ativação das camadas intermediárias e de saída da RNA. Existe um grande número de funções de ativação, sendo as mais utilizadas nos modelos de previsão (Bravo, 2010): a sigmoide logística, a tangente hiperbólica e a linear, apresentadas a seguir pelas equações 13, 14 e 15 respectivamente.

$$\mathbf{log\ sig}(n) = \frac{1}{1 + e^{-n}} \quad \text{(Equação 13)}$$

$$\mathbf{tan\ sig}(n) = \frac{2}{1 + e^{-2n}} - 1 \quad \text{(Equação 14)}$$

$$\mathbf{linear}(n) = n \quad \text{(Equação 15)}$$

Essas funções de ativação são aplicadas a todos os neurônios de uma camada da RNA. As camadas intermediárias recebem as funções de ativação não lineares, onde a mesma função de ativação é geralmente adotada para todos os neurônios de uma determinada camada. Também em geral, é determinada para a última camada (camada de saída) o uso da função linear.

A maioria dos modelos de RNA para previsões adotam a função sigmoide logística na camada oculta como a exemplo de Silverman e Dracup (2000), Kumar et al. (2007), Paz et al. (2011),

Shukla et al. (2011) e Costa et al. (2014). Outros preferem o uso da função tangente hiperbólica como mostrado em Nilsson et al. (2006), Karamouz et al. (2008), Bravo (2010) e Mekanik et al. (2013). Alguns trabalhos usam essas duas funções de ativação, uma para cada modelo distinto (Sousa e Sousa 2010), ou ambas simultaneamente em camadas distintas (Deo e Sahin, 2015).

Para alcançar o melhor desempenho de um modelo de RNA, o conjunto de dados tanto de variáveis preditoras quanto de previsão devem ser divididos e padronizados. Tipicamente a divisão dos dados é feita em pelo menos dois conjuntos: treinamento e validação. O conjunto de treinamento é utilizado durante o processo de calibração dos parâmetros da RNA, ou seja, para o ajuste dos pesos das conexões, enquanto que o conjunto de validação é utilizado para avaliar a extensão em que o modelo de RNA pode generalizar, ou seja, a habilidade do modelo (Paz et al., 2010; Bravo, 2010).

Idealmente, a divisão dos dados deve ser feita em três conjuntos: treinamento, verificação e validação. Essa divisão é feita para usar a validação cruzada ou interrupção precoce, um método comum e prático para determinar o melhor momento para parar o processo de treinamento. Na utilização deste método, os valores dos pesos são ajustados com base no conjunto de treinamento afim de diminuir o erro da saída, mas esse desempenho também é avaliado paralelamente pelo conjunto de verificação. Conforme o treinamento progride, o ajuste eficaz dos pesos cresce, e o erro, tanto no conjunto de treinamento como no de verificação, diminui gradualmente a medida que a rede consegue generalizar as características a partir dos dados de treinamento. A proporção que o treinamento progride mais e além de um ponto, a rede ajusta os pesos mais do que o necessário e começa a ajusta-los também ao ruído dos dados de treinamento. Em consequência disto, enquanto o erro ao longo do conjunto de treinamento ainda está a diminuir, o erro sobre o conjunto de verificação começa a aumentar (Figura 3.10) por conta da influência do ruído. Então, a rede pára o treinamento evitando o overfitting (Silva et al., 2010; Maier et al., 2010). Portanto, a localização de mínimo do erro do conjunto de verificação determina quando a rede atinge um ajuste ótimo de pesos. A interrupção do treinamento antes desse mínimo, pode fazer com que a rede sofra underfitting, ou seja, a rede não consegue captar características suficiente dos dados de treinamento.

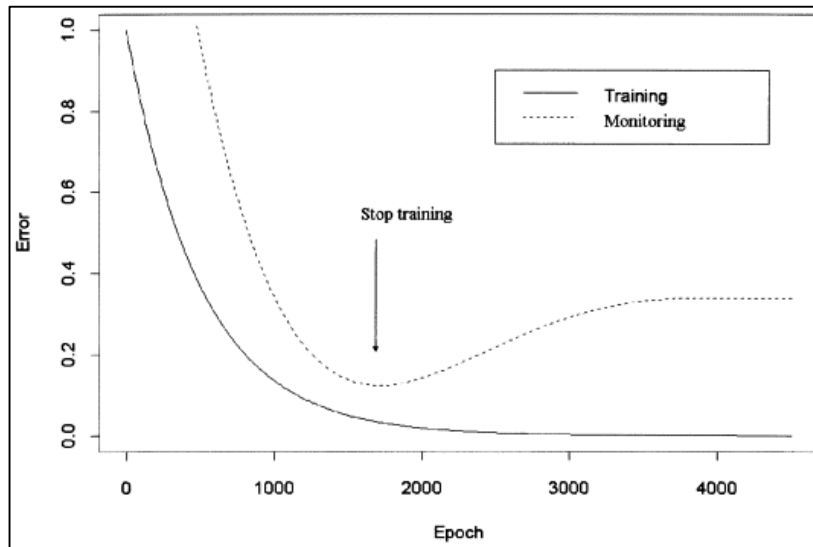


Figura 3.10 – Representação do critério de parada do treinamento por validação cruzada.  
 Fonte: Philip e Joseph (2003).

A divisão das séries de dados para treinamento e validação (e conjunto de verificação se houver) deve ser feita de tal maneira que cada conjunto ainda seja representativo de toda a população (Maier e Dandy, 2000; Maier et al., 2010). Isto pode ser testado verificando se alguns parâmetros estatísticos, como a média e o desvio padrão, são similares entre os conjuntos de dados (Paz et al. 2010; Bravo, 2010).

Normalmente, o conjunto de treinamento é composto aleatoriamente com cerca de 60% a 90% das amostras do conjunto total. Já o conjunto de verificação, a sua composição está entre 10% e 40% do conjunto total de amostras (Silva et al., 2010). A maioria dos trabalhos de previsões, usam acima de 60% dos dados disponíveis para treinamento, enquanto que dividem o restante entre verificação (caso haja) e validação, como por exemplo em Bravo (2010) que entre as 53 amostras do conjunto de dados disponíveis, dividiu em 35, 8 e 10 (66%, 15% e 19%) dados para treinamento, verificação e validação respectivamente; Paz et al. (2011) que de 52 amostras dividiu em 38, 7 e 7 (73%, 13% e 13%) dados; Aboot e Marohasy (2014) com 120 anos de dados dividiu sua série em 102 e 18 (85% e 15%) dados somente para treinamento e validação respectivamente; Nastos et al. (2013) que entre 111 anos de dados mensais, dividiu 100 e 11 (90% e 10%) dados; Kumar et al. (2007) com 33 anos de dados, dividiu em 23 e 10 (70% e 30%) dados.

Deo e Sahin (2015) seguiu uma abordagem ad-hoc e não se preocupou com as propriedades estatísticas dos subconjuntos de dados, dividindo sua série de 98 anos de dados (1915-2012) onde deixou os 91 (93%) primeiros anos para treinamento (isto é, 1915-2005) e os 7 anos

restante (2006-2012) para validação. Há também abordagens que dividem os dados de maneira a obter maior probabilidade de homogeneidade entre os parâmetros estatísticos dos conjuntos, usando critérios baseados em Partial Mutual Information (PMI), como usado por Iseri et al. (2005), que dividiu seus 97 dados de entrada em 64, 22 e 11 (66%, 23% e 11%); e critérios baseados em Self Organizing Map (SOM) como usado por Srivastava et al. (2010) onde dividiu seus 105 dados em 67, 25 e 13 (64%, 24% e 12%).

Finalmente, antes de iniciar os procedimentos de treinamento e validação, os dados de entrada devem ser padronizados, como, por exemplo, usando a Equação 16.

$$x_{new} = 0.1 + \left( \frac{x_{orig} - x_{min}}{x_{max} - x_{min}} \right) \cdot 0.8 \quad (\text{Equação 16})$$

onde  $x_{new}$  é o valor padronizado da variável de entrada  $x$  no intervalo  $[0,10; 0,90]$ ,  $x_{orig}$  é o valor original da variável de entrada em sua unidade apropriada, e  $x_{min}$  e  $x_{max}$  são os valores máximos e mínimo originais das variáveis de entrada em sua unidade apropriada, respectivamente.

O intervalo para o qual as variáveis de entrada são padronizadas deve estar relacionado com os limites assintóticos das funções de ativação. Por exemplo, as saídas da função sigmóide logística estão entre 0 e 1, sendo que os dados são geralmente dimensionados no intervalo  $[0,1, 0,9]$ . Através da padronização, dados de entrada medidos em diferentes escalas de unidades receberão a mesma consideração durante o processo de treinamento. Ela impede que uma variável de entrada com grandes magnitudes domine o processo, e é também útil para melhorar a eficiência e eficácia do treinamento (Dawson e Wilby, 2001). Outra vantagem da utilização da padronização no intervalo  $[0,10; 0,90]$  durante a modelagem hidrológica é que eventos extremos (cheias ou estiagens) mais severos que os existentes no conjunto de treinamento podem ser obtidos (Hsu et al., 1995 apud Paz et al., 2011).

### 3.4.5 Treinamento e Validação

O treinamento de um modelo de RNA consiste da aplicação de um conjunto de passos ordenados com o intuito de ajustar os pesos de seus neurônios. Assim, tal processo de ajuste, também conhecido como algoritmo de aprendizagem, realiza o ajuste dos pesos com o objetivo de minimizar alguma medida de desempenho com base nos erros entre a saída estimada e observada, onde o erro quadrático médio é o mais utilizado nessa etapa (Maier e Dandy, 2000; Maier et al., 2010).

O algoritmo de aprendizagem de rede neural MLP mais popular é o backpropagation (ou retropropagação de erros), que consiste em duas fases (Haykin, 2001):

- a) Fase forward (propagação adiante) – um padrão de atividade (vetor de entrada) se propaga pela rede, da camada de entrada até a camada de saída e finalmente um conjunto de saídas é produzido como resposta real da rede; Durante esse passo os pesos sinápticos da rede não são alterados;
- b) Fase backward (propagação reversa) – a resposta real é subtraída de uma resposta desejada (alvo) para produzir um sinal de erro. O erro se propaga na direção contrária ao fluxo de dados, indo da camada de saída até a primeira camada intermediária, ajustando os pesos sinápticos de cada neurônio de cada camada.

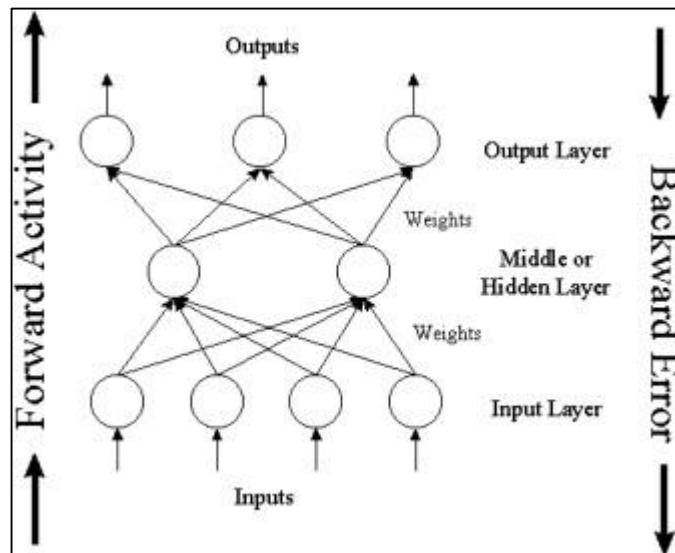


Figura 3.11 – Ilustração da atividade de aprendizagem, a propagação (entradas) e a retropropagação (erro). Fonte: Nastos et al. (2013).

A determinação da combinação de valores dos pesos da RNA que minimizam o erro do treinamento não é um problema simples. À medida que cada combinação dos valores de peso é realizada, resulta-se geralmente num erro diferente no modelo, calculado por uma função erro  $E(w)$  que cria uma superfície de erro para o espaço dos pesos (Figura 3.12). Encontrar os pesos correspondentes ao mínimo global nessa superfície de erro da medida de desempenho (função erro) é o objetivo do treinamento da RNA (Maier et al., 2010; Silva et al. 2010).

Em virtude da superfície de erro ser produzida geralmente por uma complicada forma matemática que partem das funções que definem as medidas de desempenho, há então a possibilidade de que o processo de aprendizado direcione a matriz de pesos da rede para um

ponto de mínimo local (ótimo local), que pode não corresponder aos valores mais apropriados para a generalização de resultados comparado ao ótimo global.

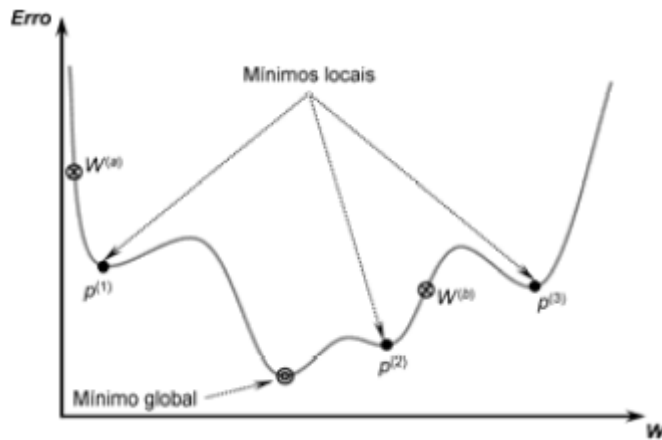


Figura 3.12 – Pontos de mínimos locais associados a função erro. Fonte: Silva et al. 2010.

Pode-se observar na Figura 3.12, que o grau de dificuldade em encontrar o valor do peso ou a combinação de valores dos pesos que resultam no menor erro do modelo, é afetado pela "rugosidade" de superfície de erro. Segundo Maier et al. (2010) o grau de rugosidade de uma superfície de erro, além de ser afetado pelo tipo de função erro, é também afetado pela quantidade de pesos na RNA, entre outras coisas, o número de neurônios no modelo. À medida que o número de pesos do modelo de RNA aumenta, o mesmo acontece com o tamanho do espaço de busca e, geralmente, o número de mínimos locais.

O treinamento da RNA é geralmente baseado em métodos de busca local (métodos de gradiente), que podem ser divididos em duas classes: de primeira ordem e de segunda ordem. Os métodos de primeira ordem são baseados em modelos lineares (gradiente descendente), enquanto que os métodos de segunda ordem são baseados em modelos quadráticos, tal como o método de Newton. Em ambos os casos, várias técnicas iterativas podem ser usadas durante o treinamento para minimizar uma medida de desempenho (Maier e Dandy, 2000; Maier et al., 2010). Os pesos são ajustados em cada iteração de acordo com uma equação geral deste tipo (Bravo, 2010):

$$w_{k+1} = w_k + \eta_k \cdot D_k \quad (\text{Equação 17})$$

onde  $w_{k+1}$  é o vetor dos pesos na iteração  $k+1$ ,  $w_k$  é o vetor dos pesos na iteração  $k$ ,  $\eta_k$  é o tamanho do passo na iteração  $k$  e  $D_k$  é um vetor que define a direção de busca na iteração  $k$ .

A principal diferença entre os dois métodos distintos de treinamento RNA é a forma como se calcula o  $D_k$ . O método principal utilizado em áreas relacionadas com hidrometeorologia tem sido o método da Regra Delta Generalizada (ou backpropagation padrão) (Maier e Dandy, 2000; Maier et al., 2010). Essencialmente, este é um método que iguala  $D_k$  ao valor negativo da primeira derivada (gradiente de primeira ordem) da função erro, com o objetivo de minimizar uma dada medida de desempenho (equação 4). Nesse algoritmo, o tamanho do passo é conhecido como a taxa de aprendizagem, que é geralmente adotada como um valor fixo (Silverman e Dracup, 2000), embora algumas versões do algoritmo permitam a mudança dinâmica dessa taxa em cada iteração incluindo mais um parâmetro, chamado de taxa de momento (Haykin, 2001). Abaixo nota-se a igualdade entre a equação 17 e 20.

$$\Delta \mathbf{w} = -\eta \cdot \nabla E(\mathbf{w}) \quad (\text{Equação 18})$$

$$\Delta \mathbf{w} = -\eta \cdot \frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} \quad (\text{Equação 19})$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \eta \frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} \quad (\text{Equação 20})$$

onde,  $w$  é o vetor dos pesos;  $\Delta w$  é a variação do vetor dos pesos;  $\nabla E(w)$  é o gradiente de primeira ordem da função erro; e  $\eta$  é a taxa de aprendizagem.

A taxa de aprendizagem exprime o quão rápido o processo de treinamento de uma rede estará sendo conduzido rumo a sua convergência (estabilização). A escolha de  $\eta$  deve ser realizada com cautela, pois quando este parâmetro tem um valor pequeno, o processo pode ser lento e tem uma chance maior de ser preso em ótimos locais. Por sua vez, um valor muito grande pode causar instabilidade no treinamento da RNA (Paz et al. 2010). Normalmente se adotam valores pertencentes ao intervalo compreendido entre 0 e 1 (Silva et al. 2010).

A interpretação geométrica do processo de convergência do método de gradiente descendente frente aos passos de atualização de  $w_{(k+1)}$ , rumo ao ponto de minimização  $w^*$  da função erro  $E(w)$ , é dada pela Figura 3.13.

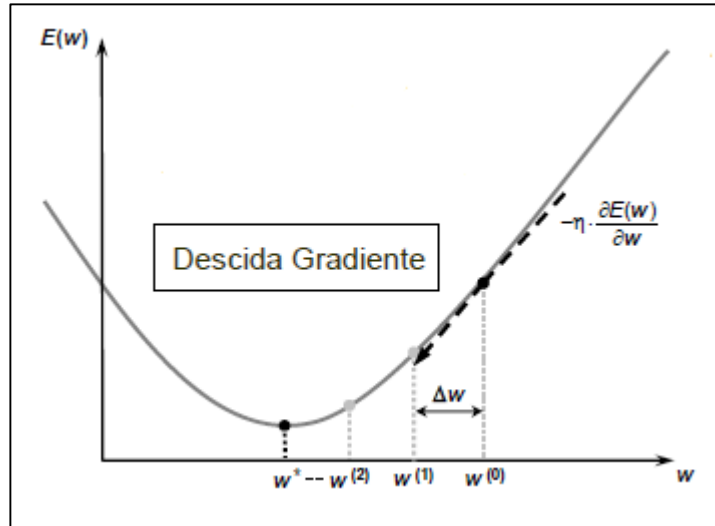


Figura 3.13 – Interpretação geométrica da Regra Delta. Fonte: Silva et al. (2010).

O algoritmo backpropagation padrão ajusta os valores das matrizes de pesos da rede MLP em relação à direção oposta do gradiente da função erro da medida de desempenho. Entretanto, a utilização desse algoritmo na prática tende a convergir muito lentamente, exigindo um elevado esforço computacional devido a um maior número de iterações (Partal e Kisi, 2007). Mesmo assim, os métodos de primeira ordem e suas variações ainda são os mais utilizados nos treinamentos de RNAs para previsões hidrológicas como apresentado há 15 anos por Maier e Dandy (2000) e também recentemente na revisão de Maier et al. (2010), o que refletiu em uma gama de publicações dentro de previsões de precipitação de longo prazo com o uso de índices climáticos, onde tem-se como exemplo em Silverman e Dracup (2000), Philip e Joseph (2003), Iseri et al. (2005), Srivastava et al., (2010), Shukla et al. (2011) e Nastos et al. (2013).

Os outros métodos de busca local são os de gradiente de segunda ordem que usam o inverso da matriz Hessiana para definir a direção de busca  $D_k$  em cada iteração (Maier e Dandy, 2000; Bravo, 2010). O método mais utilizado dentro dessa abordagem é de Levenberg-Marquardt que é uma aproximação do método de Newton (Haykin, 2001). Ele consegue melhorar o método de Gauss-Newton por meio de utilização de uma taxa de atualização variável (equação 22).

$$\Delta w = -[\nabla^2 E(w) + \mu \cdot I]^{-1} \cdot \nabla E(w) \quad \text{(Equação 21)}$$

$$w_{k+1} = w_k - [\nabla^2 E(w) + \mu \cdot I]^{-1} \cdot \nabla E(w) \quad \text{(Equação 22)}$$



onde,  $\nabla^2 E(w)$  é uma matriz Hessiana (matriz de derivadas de segunda ordem);  $\nabla E(w)$  é a matriz Jacobiana (matriz de derivadas de primeira ordem);  $\mu$  é o parâmetro ajusta a taxa de convergência do algoritmo; e  $I$  é a matriz identidade.

Esses métodos são um aperfeiçoamento do algoritmo de aprendizagem backpropagation com o objetivo de tornar o processo de convergência mais eficiente. Eles tendem a exigir um maior esforço computacional devido ao armazenamento e processamento, em comparação com os métodos de primeira ordem. No entanto, as técnicas de segunda ordem têm uma maior capacidade para escapar de mínimos locais em relação aos métodos de primeira ordem. Além de reduzir consideravelmente o número de iterações no processo de treinamento da rede (Chattopadhyay e Chattopadhyay, 2008).

A utilização dos métodos de segunda ordem para o treinamento de modelos de RNAs para previsões hidrológicas vem aumentando significativamente na última década (Maier et al., 2010) e isso pode ser notado também dentro das previsões de precipitação de longo-prazo com índices climáticos, tais como o uso do próprio método de Levenberg-Marquardt utilizado nos trabalhos de Chattopadhyay e Chattopadhyay (2008) e Mekanik et al. (2013); o método *Scaled Conjugate Gradient* (SCGM) desenvolvido por Moller (1993) e utilizado por Paz et al. (2011), Costa (2014) e He et al. (2015); e o método de Broyden-Fletcher-Goldfarb-Shanno (BFGS) usado por Deo e Sahin (2015).

É importante ressaltar novamente que o uso de um método de busca local em processo de treinamento da RNA não garante encontrar o ótimo global do problema de otimização se a superfície de erro for consideravelmente muito rugosa. Esta tendência de convergência fica condicionada à posição em que os pesos foram iniciados, pois esse treinamento é baseado somente em informações de gradiente (inclinação). Porém, teoricamente, a exemplo do artigo de Shukla et al. (2011), esse problema pode ser resolvido repetindo o treinamento da RNA várias vezes, com diferentes condições iniciais. Se os mesmos pesos são obtidos, pode ser evidência suficiente para considerar ter encontrado o verdadeiro ótimo global.

Em nenhum dos casos previamente apresentados, o conjunto de validação é utilizado no processo de treinamento, mas apenas para avaliar a habilidade do modelo de RNA de poder generalizar os dados. Após o treinamento, os pesos ajustados são fixos e, em seguida, o modelo de RNA é executado utilizando os dados de validação definidos como entrada. Este procedimento é chamado de validação da RNA e é um processo de simulação.

Várias medidas de desempenho são usadas para avaliar a precisão de modelos de RNA com base no conjunto de dados de validação. Como visto, geralmente as avaliações são feitas através de uma métrica quantitativa do erro entre os valores previstos e observados. Entre os tipos de métricas de erro existentes para medidas de desempenho, Maier et al. (2010) destacam que as mais usadas nos modelos de RNA para previsão hidrológica são as de Erros Quadráticos, Absolutos, Relativos e de Diferença entre Produtos.

As métricas de erros quadráticos baseiam-se nos quadrados das diferenças entre os valores de saída reais e modelados. As medidas de desempenho comumente empregadas pertencentes a esta categoria é a soma do erro quadrático (RSE), o erro médio quadrático (RMSE) e a eficiência Nash-Sutcliffe (NS). Uma característica destas métricas é que eles tendem a ser dominados por erros com altas magnitudes. Alternativamente, os erros absolutos podem ser usados, os quais são baseados nas diferenças absolutas entre as saídas reais e modeladas e incluem medidas tais como a soma do erro absoluto (EA), a soma do erro médio dos absolutos (MAE). Enquanto os erros absolutos fornecem informações sobre a magnitude do erro, eles não fornecem informações sobre as tendências do modelo em termos gerais de subestimação ou superestimação da previsão da série temporal. Este problema pode ser contornado considerando a soma total ou a soma média dos erros sem tomar seus valores absolutos, resultando em um erro sistemático (viés estatístico - bias) total (TBIAS) e na média do erro (MBIAS). A fim de permitir que os desempenhos de modelos com saídas de diferentes magnitudes sejam comparados mais facilmente, as métricas de erro relativo, tal como o erro médio relativo (MRE), a raiz quadrada do erro médio normalizado (NMSE) e o erro sistemático (bias) médio normalizado (NMBE) podem ser utilizados. E finalmente, uma medida empírica do erro entre as saídas reais e as modeladas pode ser obtida utilizando as estatísticas de erro de diferença de produto-momento, da qual o coeficiente de correlação de Pearson é o mais bem conhecido (Krause et al., 2005; Wilks, 2006; Maier et al., 2010).

Desta forma, a maioria dos trabalhos utiliza em suas análises de desempenho medidas que abrangem comumente cada tipo, ou pelo menos a maioria dos tipos de métricas citadas acima. Entre elas, as mais usadas para avaliar a precisão na previsão quantitativa não probabilística em modelos de redes neurais são (Paz et al., 2010): o erro quadrático médio (MSE - Equação 23), o erro médio absoluto (MAE - Equação 24), o erro médio relativo (MRE - Equação 25), o índice de eficiência de Nash-Sutcliffe (NS - Equação 26) e o coeficiente de correlação de Pearson (R - Equação 27). Além disso, a forma típica de uma medida de desempenho para avaliar a

habilidade de previsão, normalmente chamado de *Skill Score* (SS), é apresentado na Equação 28.

$$MSE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \quad (\text{Equação 23})$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - O_i| \quad (\text{Equação 24})$$

$$MRE = \frac{1}{N} \sum_{i=1}^N \left| \frac{P_i - O_i}{O_i} \right| \quad (\text{Equação 25})$$

$$NS = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (\text{Equação 26})$$

$$R = \frac{\sum_{i=1}^N (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^N (P_i - \bar{P})^2}} \quad (\text{Equação 27})$$

$$SS = \frac{PMA - PMA_{ref}}{PMA_{perf} - PMA_{ref}} \quad (\text{Equação 28})$$

onde  $P_i$  e  $O_i$  são os valores previstos e observados dos preditandos na amostra  $i$ , respectivamente;  $\bar{O}$  é o valor médio observado do preditando;  $N$  é o número total de amostras;  $PMA$ ,  $PMA_{ref}$  e  $PMA_{perf}$  são, respectivamente, o valor de desempenho de uma medida de precisão da previsão, da previsão de referência e da previsão perfeita (previsão hipotética sem erros, onde  $P_i$  é igual a  $O_i$ ).

Como pode ser visto na Equação 23-27, medidas de precisão de desempenho são em função do erro (diferença  $P_i - O_i$ ). Assim, reduzir o erro, maior é a precisão do modelo. Nesse sentido, o MSE, MAE e MRE são iguais a zero, quando as previsões perfeitas são utilizadas, e seu valor aumenta, com erros maiores de previsão. Por sua vez, NS e R tem seus valores iguais a um quando previsões perfeitas são utilizadas. No caso R de Pearson que tem seus valores variando de -1 a 1, os valores menores que zero já não correspondem a um bom desempenho na previsão, pois ao final da previsão, a expectativa é que  $P_i$  seja igual a  $O_i$ , e não o inverso.

## 4 METODOLOGIA

A metodologia deste trabalho consiste em análise de teleconexões a fim de identificar relações entre índices climáticos e precipitação na área de estudo, para em seguida desenvolver modelos de previsão usando redes neurais artificiais (Figura 4.1).

Primeiramente foi definida a variável relacionada à precipitação que constitui aquela a ser prevista, que no caso é o total acumulado no trimestre mais chuvoso. Paralelamente, em relação as variáveis preditoras, foi feito um levantamento dos dados mensais dos principais índices climáticos, os quais correspondem a modos de variabilidade climática em regiões oceânicas no planeta. Estes índices foram separados em grupos com mesma quantidade de dados disponível (por exemplo, índices com 100 anos, 60 anos...).

Para a identificação das melhores variáveis preditoras, foram feitas as correlações entre o intervalo de tempo disponível dos valores mensais (defasados de 2 a 12 meses) de cada índice coletado com o mesmo intervalo de tempo (trimestre) da precipitação acumulada.

A fim de conseguir captar as relações tanto de linearidade, não linearidade e não estacionariedade entre o preditando e os preditores, foram feitas duas abordagens de correlações: via Pearson/Spearman (linearidade/não linearidade) e via Wavelets Cruzada/Coerência (não estacionariedade). Cada abordagem criou o seu próprio grupo de potenciais preditores com níveis de confiança de 95%. No entanto, a validade física dos fenômenos climáticos representados pelos índices, resulta da vital importância de uma discussão sobre as teleconexões, captadas pelas correlações, existentes entre as variáveis.

Após o estudo anterior, é utilizada uma rede neural artificial com arquitetura do tipo progressiva de três camadas (uma camada de entrada, uma camada intermediária e uma camada de saída) para previsão da precipitação. Os potenciais preditores são agrupados novamente em relação à quantidade de dados disponível, sendo para cada grupo definidas algumas combinações entre as variáveis preditoras para a entrada da RNA. Para cada combinação, os dados são divididos (treinamento-validação-teste) segundo diferentes distribuições e testadas diferentes quantidades de neurônios na camada de intermediária. Todos os modelos de RNA desenvolvidos são avaliados usando medidas de desempenho.

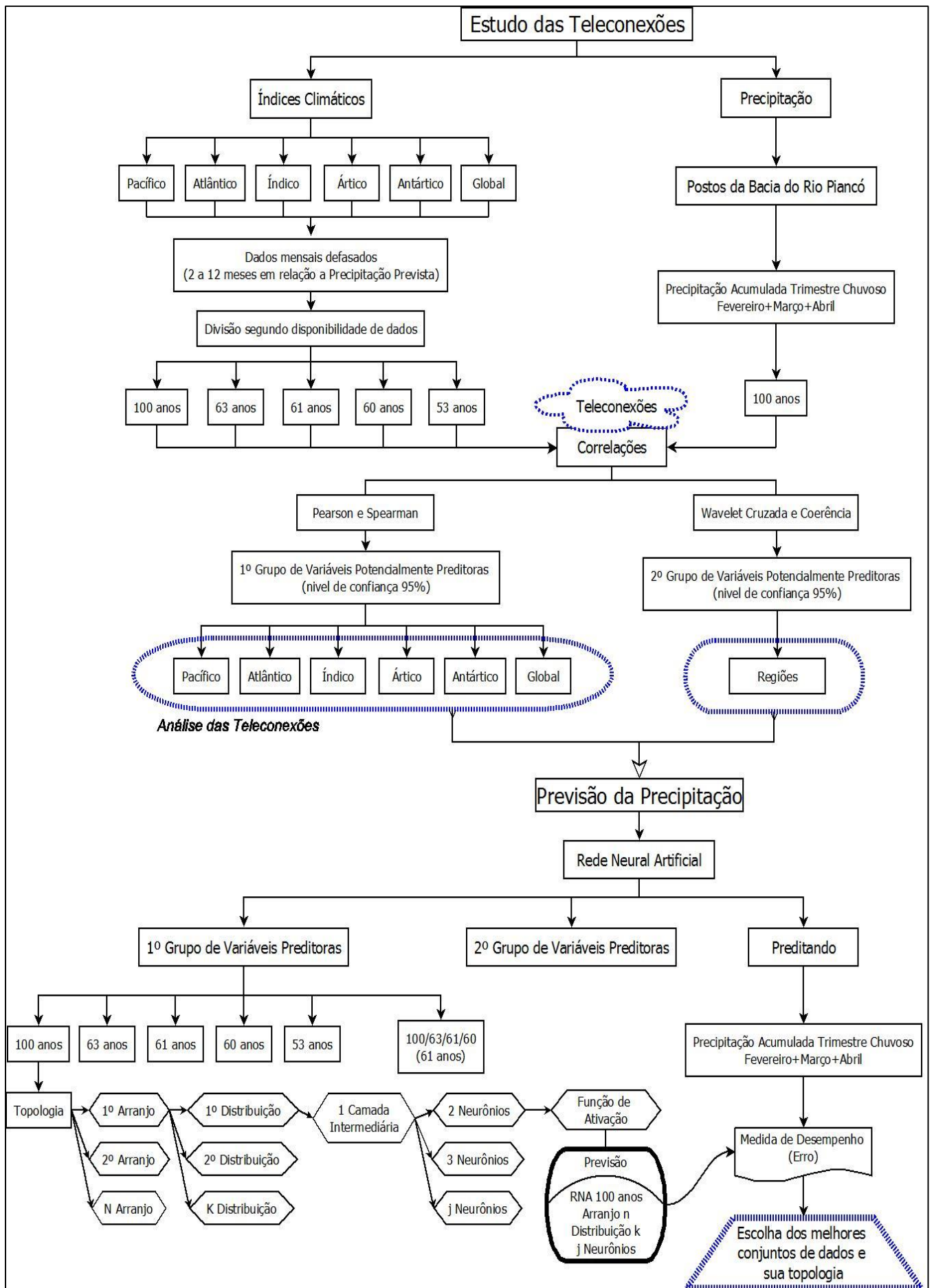


Figura 4.1 – Fluxograma das etapas do estudo.

## 4.1 ÁREA DE ESTUDO

### 4.1.1 Localização

A área da bacia hidrográfica do rio Piancó, é de 9.228 km<sup>2</sup> e foi delimitada a partir das cartas digitalizadas da Superintendência de Desenvolvimento do Nordeste – SUDENE em escala de 1:100.000. O seu perímetro, comprimento da linha do divisor de águas que a delimita, medido na mesma base cartográfica, é de 540 km (SCIENTEC,1997).

A bacia do rio Piancó corresponde a uma das sete sub-bacias do Rio Piranhas em território paraibano. Localiza-se no Sudoeste do Estado da Paraíba, entre os paralelos 6°43'51'' e 7°58'15'' Sul e meridianos 37°27'41'' e 38°42'49'' a Oeste de Greenwich. O rio Piancó apresenta suas nascentes na Serra do Umbuzeiro no município de Santa Inês recebendo significativas contribuições de seis cursos d'água na sua margem esquerda e de mais quinze na sua margem direita, desaguando finalmente no rio Piranhas no município de Pombal. Nesta bacia está inserida a maior reserva hídrica do Estado, o sistema constituído de dois reservatórios Coremas – Mãe d'Água, com capacidade de acumulação de aproximadamente 1,37 bilhões de m<sup>3</sup> (Lima, 2014).

A região da bacia a ser estudada fica justamente a montante do reservatório Coremas – Mãe d'Água, a qual possui uma área de 4650 km<sup>2</sup>, praticamente abrangendo mais da metade de toda a bacia (Figura 4.2). A delimitação da bacia hidrográfica de estudo foi definida como sendo a área contribuinte ao posto fluviométrico de Piancó (código 37340000), que se encontra imediatamente a montante do reservatório citado. A região de estudo possui grande potencial agrícola, grande parte da sua área possui terras próprias para o cultivo de lavouras, estando atualmente em implantação diversos projetos de irrigação (Melo, 2004; Lima, 2014).

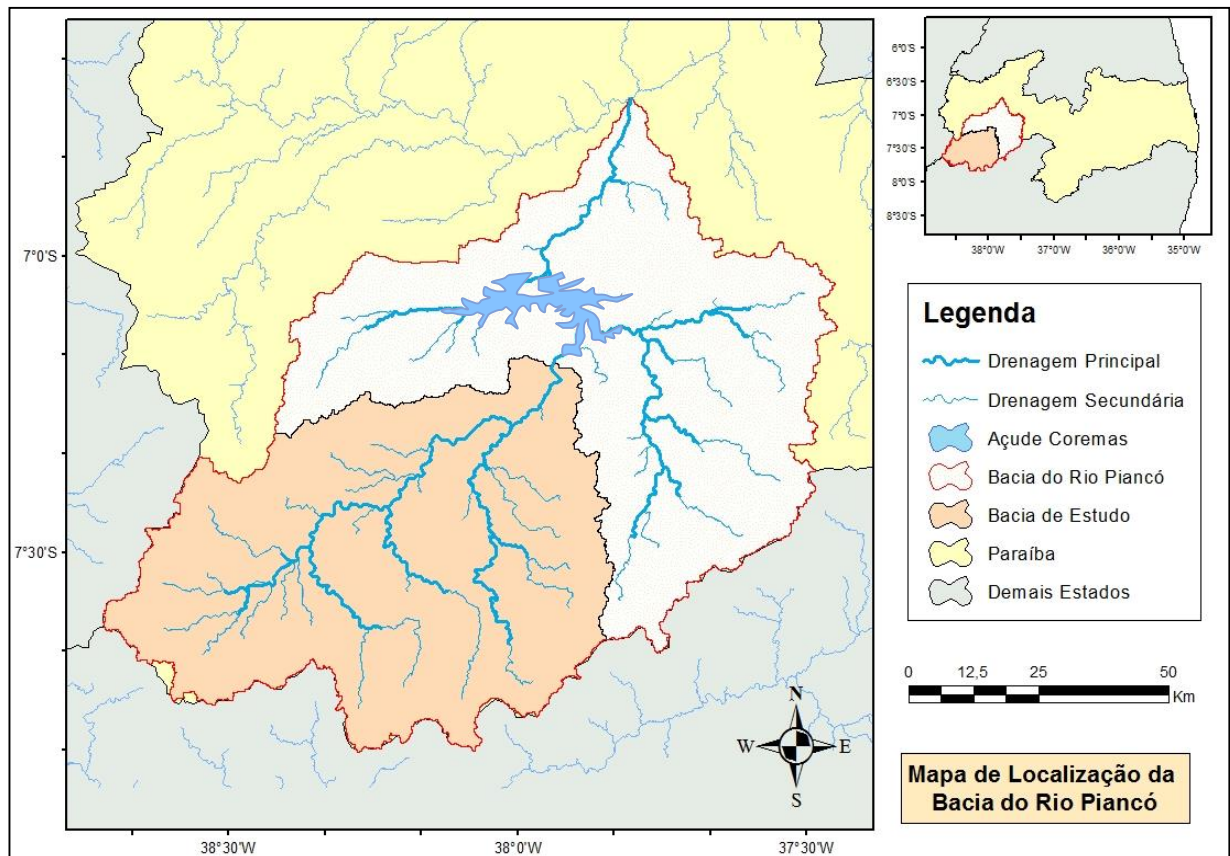


Figura 4.2 – Mapa de Localização da Bacia. Fonte: Autor

#### 4.1.2 Caracterização Fisiográfica

O relevo assume três conformações. As encostas que são divisores de água nas faixas norte, oeste e sul da bacia, são longas e apresentam-se ora côncavas ora convexas. Os altiplanos dos divisores que ocorrem no sul da bacia, nos limites com o Estado de Pernambuco, são dotados de relevo plano, suave ondulado, ondulado e às vezes forte ondulado. Por último, a depressão central onde se formam as planícies de terrenos aluviais e vales muito abertos em forma de “U” (Figura 4.3). Verifica-se que apenas 5% da área da bacia encontram-se em altitudes bastante elevadas e, o restante corresponde a altitudes que variam de 790 a 260 m, caracterizando, portanto, um relevo de moderado a forte. Apresenta uma altitude mediana igual a 461m e a altitude média da bacia igual a 485 m. A vegetação natural ou nativa é do tipo xerófila, pertencente ao bioma caatinga (Melo, 2004).

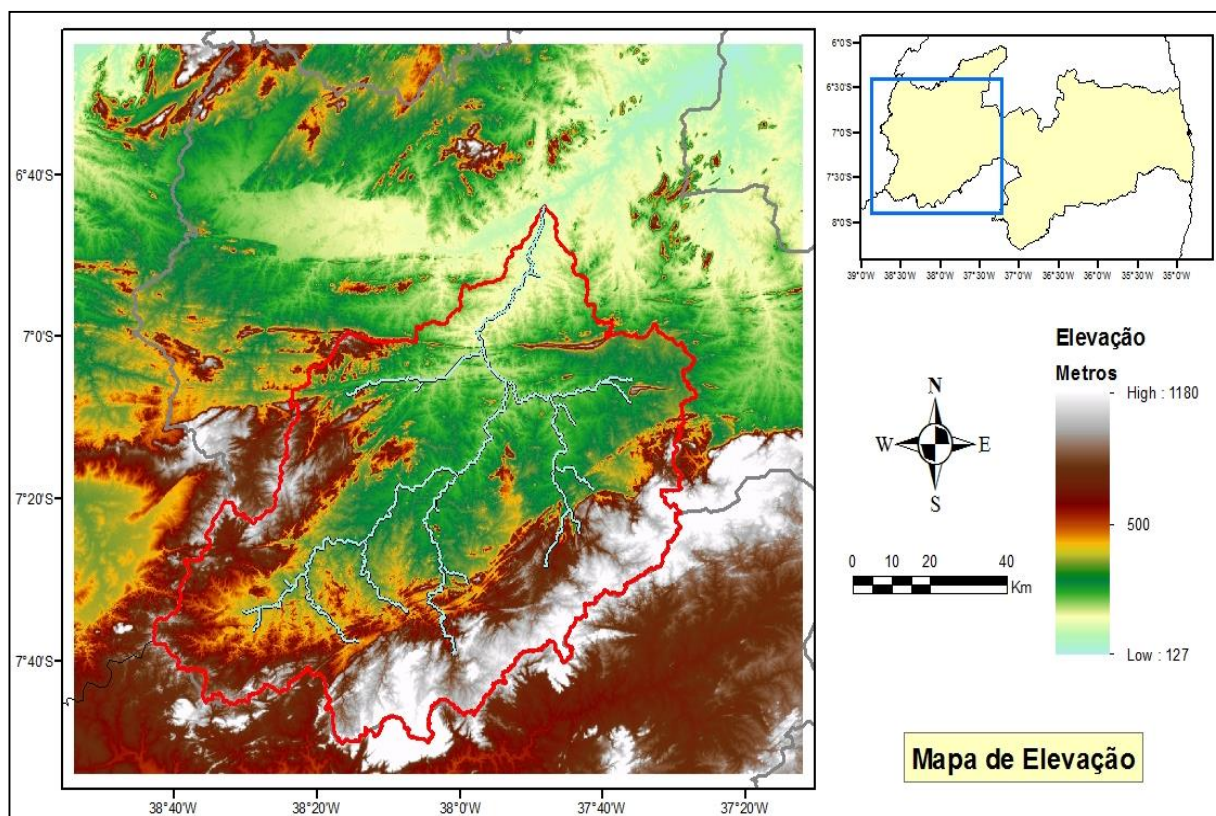


Figura 4.3 – Mapa de relevo da Bacia. Observação para os vales muito abertos em forma de “U”.  
Fonte: Autor

#### 4.1.3 Caracterização climática hidrológica

Em termos climatológicos, a bacia hidrográfica está inserida no denominado “Polígono das Secas”, constituindo um tipo semiárido quente e seco, segundo a classificação de Köeppen. Na região das cabeceiras, próximo ao município de Triunfo domina o clima do tipo Awig, isto é, clima tropical úmido com inverno seco e variações de temperatura média mensal desprezíveis ao longo do ano, conforme classificação de Köeppen. Enquanto as demais partes da bacia são classificadas como BSwb’, ou seja, quente e úmido com chuvas de verão a outono, onde as maiores precipitações ocorrem nos meses de fevereiro, março e abril e a estação seca vai de maio a dezembro (Lima, 2014).

A temperatura média anual é superior a 24°C e a amplitude térmica anual menor que 4°C. As temperaturas mais elevadas ocorrem nos meses mais secos, ou seja, outubro a janeiro e as menos elevadas entre abril e julho (SCIENTEC,1997).

A umidade relativa do ar média anual é de 64% na foz da bacia, enquanto nas cabeceiras chega a 72%. Observa-se que os meses mais úmidos são março, abril e maio quando a umidade atinge 81% em Triunfo e os meses mais secos são outubro e novembro, atingindo o valor de 60%.



A insolação diária alcança uma média de 7,3 horas nas circunvizinhanças da serra de Triunfo e 8,7 horas no restante da bacia. A velocidade média do vento na bacia, os valores médios alcançam 2,8 m/s. Com base no tanque classe A, instalado em Coremas, onde adota-se um Coeficiente do Tanque (Kp) igual a 0,75, os dados de evaporação revelam um total anual superior a 2.900 mm, condição, própria das zonas semi-áridas de latitudes tropicais. Na Tabela 4.1 encontram-se os dados climáticos na estação de São Gonçalo, o único representativo e disponível na bacia (SCIENTEC,1997).

Tabela 4.1 – Valores médios mensais de dados meteorológicos na estação de São Gonçalo.

Dados climáticos	Jan	Feb	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Temperatura (°C)	28,6	27	27,2	26,8	26,6	25,4	25,5	26,3	27,6	28,3	28,8	28,7
Umidade (%)	62	69	74	74	72	68	62	59	57	56	57	58
Insolação (horas)	8,7	8,1	7,6	7,9	8,4	8,4	8,8	9,5	9,4	9,7	9	9,1
Veloc. Vento (m/s)	2,6	2,3	2	2	2,5	2,7	3	2,8	3,3	3,3	3,2	3,3

Fonte: Lima (2014).

A pluviosidade da bacia do Piancó é caracterizada por uma distribuição de chuvas bastante irregular, com médias anuais de 800 mm, mínimas podendo chegar a 370mm e máximas até 1800mm (Figura 4.4).

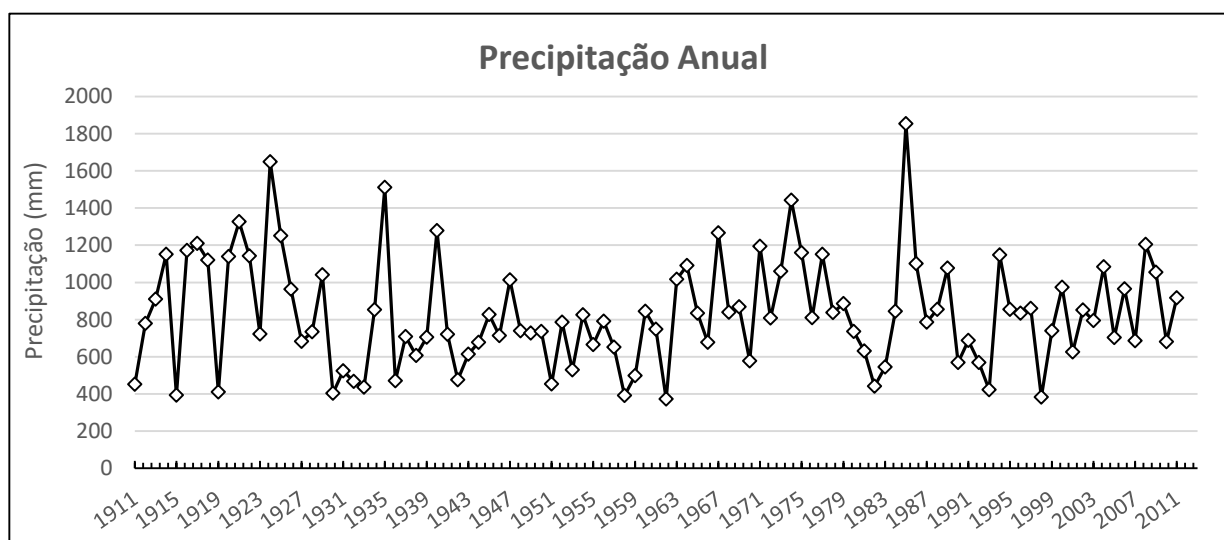


Figura 4.4 – Precipitação média anual para a bacia do rio Piancó no período de 1911 a 2011.

Fonte: Autor

No geral, pode-se destacar a presença de apenas duas estações: a “seca”, cujo clímax é de setembro a dezembro, e a “chuvosa”, restrito a um período de 3 a 4 meses por ano. Geralmente esse período chuvoso que começa de janeiro encerrando-se em meados de maio, tem nos meses

de fevereiro, março e abril a precipitação que representa cerca de 60% do valor total anual (Figura 4.5).

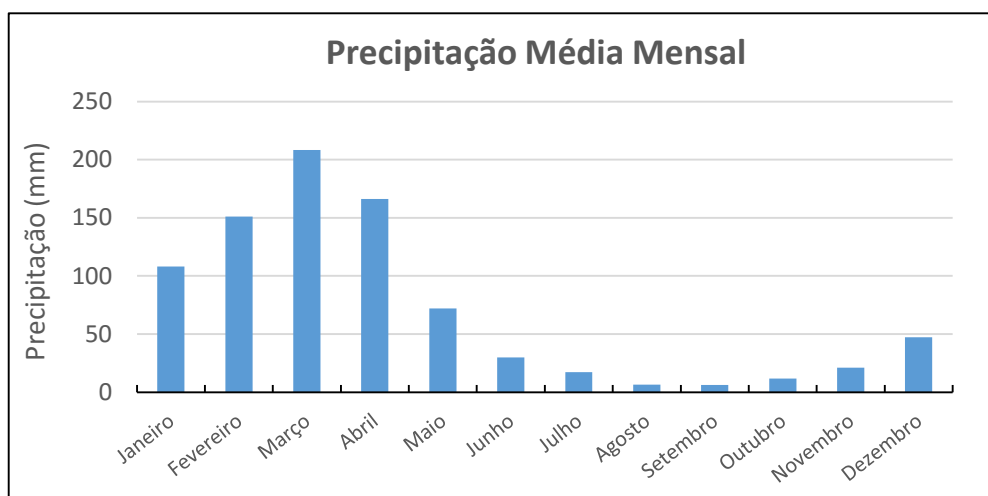


Figura 4.5 – precipitação média mensal para a bacia do Piancó. Fonte: Autor

O regime fluviométrico na bacia é característico de regiões semi-áridas. Conforme as precipitações ocorridas na bacia, observam-se períodos curtos de descargas mais elevadas (meses de fevereiro à abril) e períodos longos de baixos escoamentos (6 a 7 meses de baixa precipitação) (SCIENTEC,1997).

Na Figura 4.6, encontram-se os dados de fluviometria média mensal ( $m^3/s$ ) do posto fluviométrico de Piancó, que se localiza à montante do açude Coremas no rio Piancó nas coordenadas geográficas  $07^{\circ} 12' S$  e  $37^{\circ} 56' W$ , e que serviu como exutório na delimitação da região de estudo. Este posto opera desde maio de 1962.

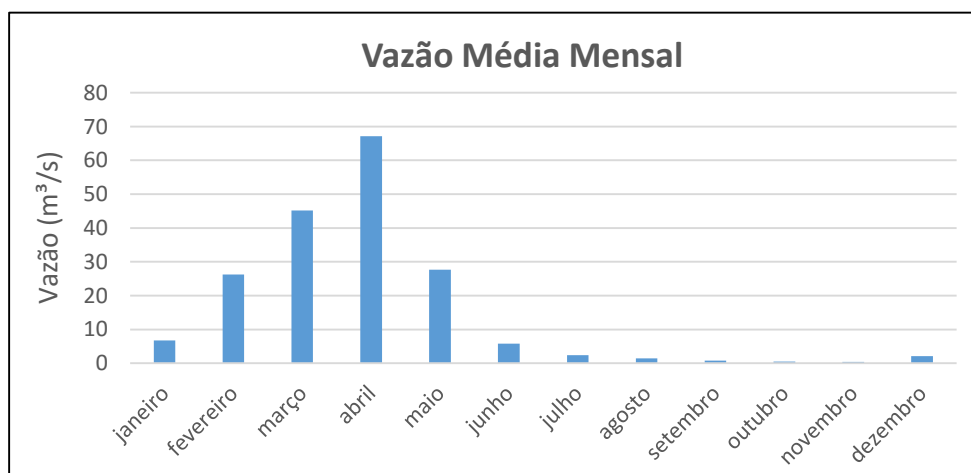


Figura 4.6 – Vazão média mensal do posto da cidade de Piancó. Fonte: Autor

## 4.2 COLETA DE DADOS E ANÁLISE PRELIMINAR

### 4.2.1 Dados de Precipitação

Os dados de precipitação utilizados foram obtidos de três diferentes fontes: Agência Nacional de Águas (ANA), Superintendência do Desenvolvimento do Nordeste (SUDENE) e Agência Executiva de Gestão das Águas do Estado da Paraíba (AESA), totalizando 13 estações pluviométricas com diferentes períodos de dados dentro de um intervalo máximo de 101 anos entre 1911 até 2011 (Figura 4.7).

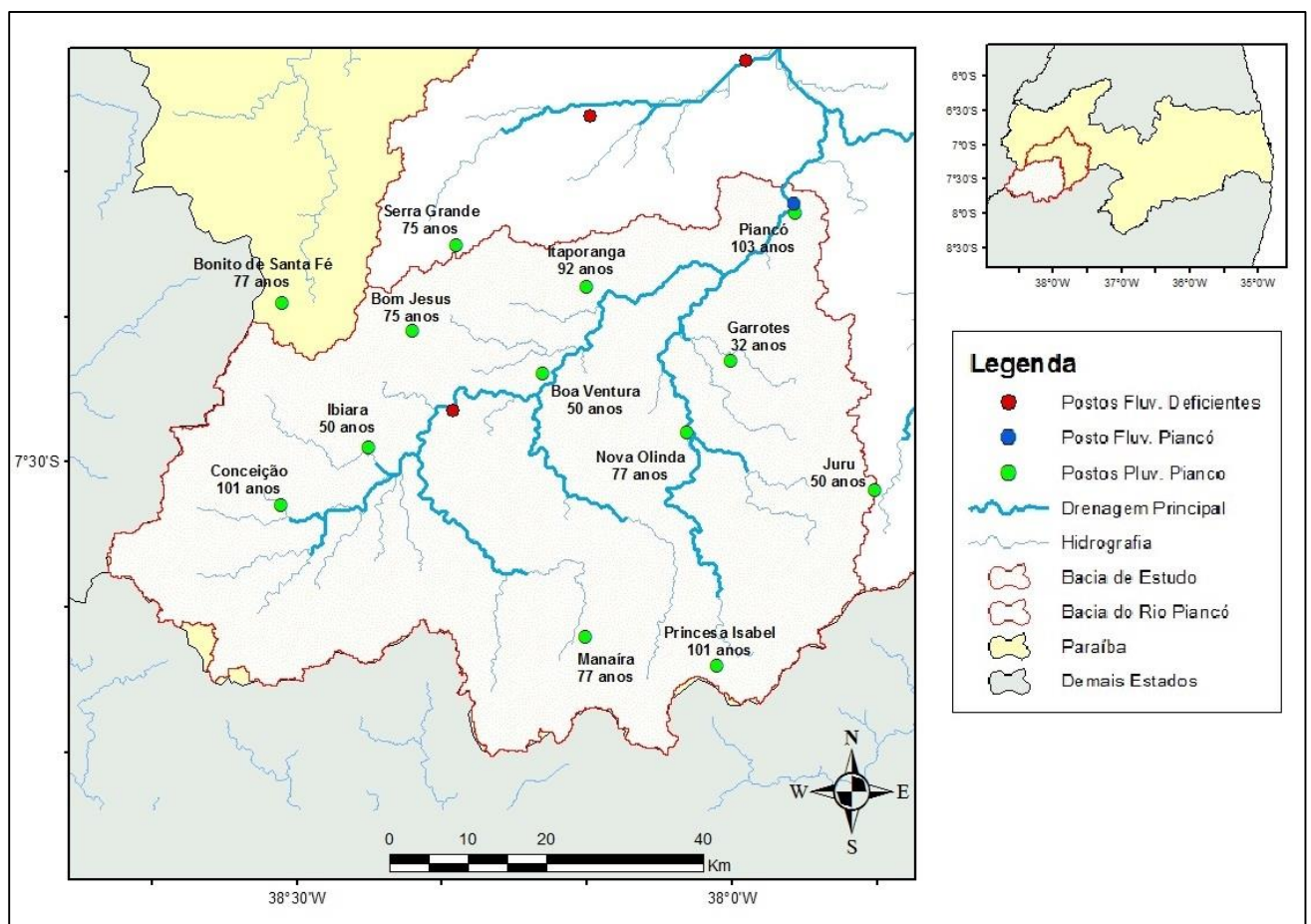


Figura 4.7 – Localização dos Postos Pluviométricos dentro da área de estudo e a quantidade de dados disponíveis. Fonte: Autor

Os dados são correspondentes as séries de precipitações diárias, onde, a priori, todas têm seus dados obtidos do site do HIDROWEB, do qual a ANA é responsável por operar. Todos os postos no site têm seus dados disponíveis até o ano de 1993, com exceção do Posto Piancó que tem dados até o ano de 2013. A SUDENE possui dados desses mesmos postos correspondentes ao período de 1986 a 1993, do qual serviram para corrigir as falhas que foram encontradas em alguns anos nos dados fornecidos pela ANA. Para o preenchimento do intervalo faltante, foram

usadas as informações disponibilizados pela AESA, que possui dados para todos os postos a partir do período imediatamente correspondente de 1994 até 2011.

Testes de consistência foram realizados nas series de precipitação através do software PrePlu (Collischonn, 2001). As series mensais de precipitações foram obtidas a partir das series diárias. Foi criado um gráfico Gantt para avaliar a situação da distribuição das falhas mensais nas series dos postos. O mesmo tipo de gráfico também foi criado para avaliar somente a situação dos meses correspondentes ao trimestre mais chuvoso.

A variável de previsão deste estudo corresponde à precipitação sazonal, sendo usada a média acumulada do trimestre mais chuvoso. A decisão de se usar a média acumulada visa obter uma variável que melhor representa o regime de precipitação sobre a região de interesse. O total acumulado nos meses de Fevereiro, Março e Abril (FMA) somam 62% da precipitação anual. Tal sazonalidade também é evidente em relação ao Nordeste, como é mostrado no artigo de Rao e Hada (1990) e reportado na Figura 4.8.

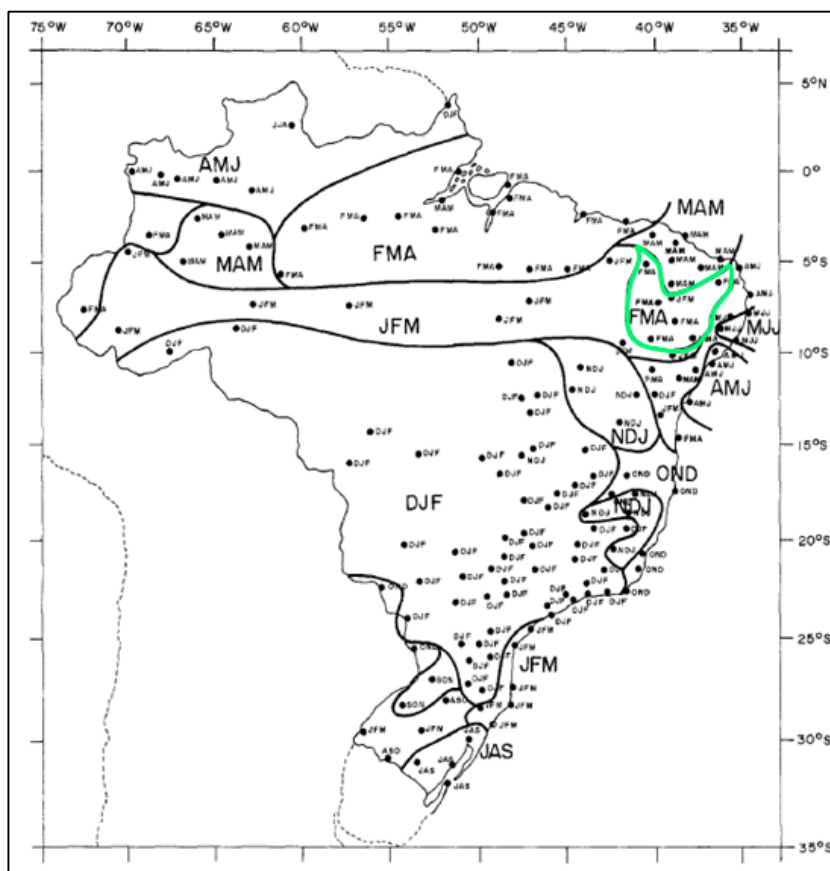


Figura 4.8 – Regiões delimitadas para os trimestres mais chuvosos no Brasil. Destaque para a região onde está inserida a Bacia do Rio Piancó. Fonte: Adaptado de Rao e Hada (1990)

Por fim, verificada a presença de uma sazonalidade bem definida na região durante a caracterização da área de estudo, foram obtidas as series anuais do trimestre mais chuvoso da região, compreendido entre os meses de Fevereiro, Março e Abril (FMA). A série de precipitação acumulada no trimestre FMA na bacia do rio Piancó é apresentada na Figura 4.9.

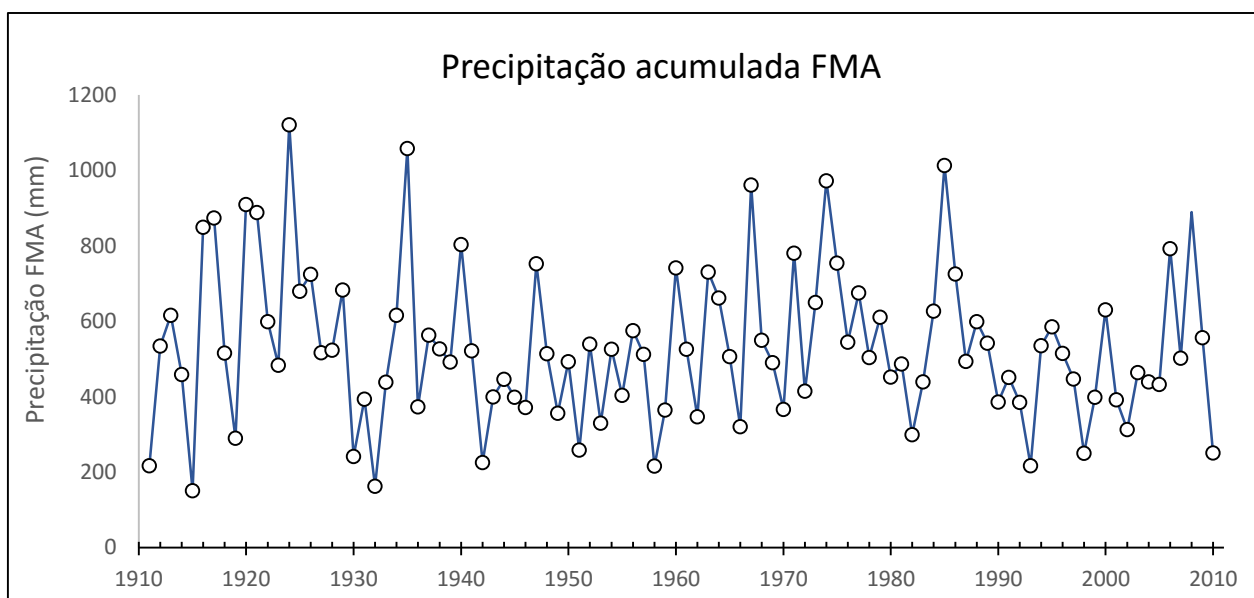


Figura 4.9 – Precipitação acumulada no trimestre FMA na bacia do Rio Piancó.

Observa-se na figura 4.9 que a precipitação FMA apresenta valores típicos entre 400mm e 600mm. Em alguns anos onde ocorrem eventos extremos, a precipitação apresenta valores superiores a 1000mm e inferiores a 250mm.

#### 4.2.2 Dados de Índices Climáticos

As séries históricas mensais referentes aos índices climáticos foram coletadas com o propósito de atingir diferentes partes dos oceanos, variando tanto em componente climática (TSM, pressão, vento, entre outros) como também em suas abordagens para o cálculo dos índices, aumentando as chances de encontrar pontos teleconectivos. Os dados foram obtidas junto ao Joint Institute for the Study of the Atmosphere and Ocean (JISAO<sup>2</sup>), The National Center for

<sup>2</sup> <http://jisao.washington.edu/>

Atmospheric Research (NCAR<sup>3</sup>), Climatic Research Unit (CRU<sup>4</sup>) e National Oceanic and Atmospheric Administration (NOAA<sup>5</sup>).

Foram selecionados os indicadores climáticos que tinham suas séries completas com pelo menos 30 anos e que estivesse dentro do intervalo entre o ano de 1910 até 2010. Desta forma, foram coletados dados para diferentes regiões oceânicas e agrupados conforme sua disposição temporal (Tabela 4.2). Os índices coletados a seguir já foram mencionados na revisão com suas devidas informações.

Tabela 4.2 – Levantamento de Índices climáticos e características

<b>Período</b>	<b>Índices</b>	<b>Intervalo</b>
<b>100 Anos</b>	NAO, SOI <sub>signal</sub> , SOI <sub>noise</sub> , PDO, GMLOTI, TNI, CTI, AMO, AMO <sub>suavizado</sub>	1910-2009
<b>92 Anos</b>	TPI	1910-2006
<b>63 Anos</b>	TNA, TSA, NOI, AMM <sub>sst</sub> , AMM <sub>wind</sub> , PMM <sub>sst</sub> , PMM <sub>wind</sub> , QBO, Solar Flux, AAO	1948-2010
<b>61 Anos</b>	NINO1+2, NINO3, NINO 3.4, NINO 4, PNA, EA, MEI, MEI <sub>rank</sub> , ONI, AO, OID, EPO	1950-2010
<b>60 Anos</b>	TP, ATI	1948-2007
	STA, NTA	1951-2010
<b>53 Anos</b>	GLAAM	1958-2010

A variedade de índices climáticos utilizados neste estudo tem o propósito de compreender se as anomalias ocorrentes em outros pontos do globo. Desta forma, além dos índices tradicionais mais utilizados para esse tipo de estudo, também foram selecionados vários outros índices usualmente menos pesquisados na literatura ao estudar a precipitação no NEB, mas que podem estar contabilizando características significativas de algum padrão climático que tenha uma teleconexão com a região analisada. Abordagem semelhante é possível encontrar nos trabalhos de Silverman e Dracup (2000), Bernado e Molion (2004) e Santana (2013).

<sup>3</sup> <http://www.cgd.ucar.edu>

<sup>4</sup> <http://www.cru.uea.ac.uk/cru>

<sup>5</sup> <http://www.esrl.noaa.gov/psd/data/climateindices/list>

### 4.3 ANALISES DE TELECONEXÕES

No contexto de se buscar teleconexões para previsão de precipitação sazonal, as correlações entre variáveis preditoras e previstas ocorrem de forma defasada. Ou seja, foram feitas correlações dos índices climáticos para cada série de seus meses, estando defasados de 2 a 12 meses em relação ao início do trimestre mais chuvoso. Desta forma, nesse estudo, foram selecionados para as correlações os meses entre Fevereiro a Dezembro do ano anterior ao da precipitação acumulada a ser prevista.

Por exemplo, a série da precipitação acumulada do trimestre FMA, correspondente ao período de 1930 a 2010, seria correlacionada com a série dos valores mensais de Dezembro dos índices climáticos no período de 1929 a 2009. Depois, para o mesmo período, seriam correlacionados os valores da série de Novembro e assim por diante até Fevereiro. Baseado nos valores de correlação são escolhidos os meses que caracterizam as variáveis preditoras.

#### 4.3.1 Correlação de Pearson e de Spearman

As correlações foram realizadas através de rotinas computacionais desenvolvidas em ambiente Matlab, com o auxílio de planilhas eletrônicas do Excel®.

Os dados dos índices são organizados em grupos nas planilhas de acordo com o intervalo de tempo disponível. Para cada índice, foi correlacionada a série temporal de cada mês com a precipitação acumulada trimestral FMA. A correlação ocorreu de forma que a série FMA do ano Y até o ano X, era correlacionada com a série do mês de Dezembro do ano Y-1 até o ano X-1. Essa correlação ocorreu com diferentes tempos de espera, até chegar a defasagem, partindo de 2 meses, até 12 meses, ou seja, finalizando no mês de Fevereiro.

O coeficiente de Pearson foi escolhido por ser a medida mais utilizada para correlações no meio científico. Também chamado de coeficiente de correlação produto-momento, esse coeficiente é medida do grau e do sinal da correlação dada pela covariância entre duas variáveis aleatórias  $x$  e  $y$ , resultando em uma medida numérica de associação linear existente entre elas. Sua fórmula, expressa na equação 27, já foi apresentada no item 3.4.5.

Considerando que a RNA também carrega um caráter não-linear (devido a sua função de ativação), o coeficiente de Spearman também foi utilizado para encontrar relações desse tipo. Nos casos em que os dados não formam uma nuvem comportada, com alguns pontos bem

distantes dos demais, ou em que parece existir uma relação crescente ou decrescente num formato de curva, o coeficiente de correlação por postos de Spearman é mais apropriado (Wilks, 2006).

Este é um método não-paramétrico que usa somente os postos, e não faz quaisquer suposições. Essencialmente tudo o que faz é calcular o coeficiente de correlação de Pearson nos postos. A sua fórmula é (Wilks, 2006):

$$r = 1 - \frac{\sum_i d_i^2}{(n^3 - n)} \quad (\text{Equação 29})$$

em que  $n$  é o número de pares  $(x_i \text{ e } y_i)$  e,

$$d_i = (\text{posto de } x_i) - (\text{posto de } y_i) \quad (\text{Equação 30})$$

Note que se os postos de  $x$  forem exatamente iguais aos postos de  $y$ , então todos os  $d_i$  serão zero e  $r$  será igual a 1.

#### 4.3.2 Correlação por Wavelets Cruzada e Coerência

A Transformada de Wavelet é utilizada para investigação da relação entre os sinais de precipitação FMA e dos índices climáticos mensais com defasagem. O procedimento para correlação da variável prevista com cada preditor mensal é o mesmo utilizado no item anterior. A diferença se encontra justamente na formulação matemática da correlação.

Essa técnica de correlação é utilizada no intuito de se captar as relações entre as componentes não estacionárias que naturalmente estão presentes em ambas as séries hidrometeorológicas (precipitação e índices climáticos). Como ambos os sinais (preditor e preditando) são caracterizados por processos altamente variáveis que podem ocorrer em várias frequências (sazonal, interanual, decadal, multidecadal), o uso dessa técnica permite encontrar as correlações dentro de uma gama de escalas (frequência) em qualquer intervalo da série. Este método proporciona efetivamente novos insights sobre a influência de cada índice climático em diferentes escalas temporais (Labat et al., 2010).

Antes de iniciar o processo de correlação, ambos os sinais originais (preditor e preditando) são projetados (transformados) do domínio do tempo  $\{F_{FMA(t)}$  e  $F_{\text{Índice,mês}(t)}\}$  para o domínio tempo-frequência  $\{W_{FMA(t,\omega)}$  e  $W_{\text{Índice,mês}(t,\omega)}\}$ , através da Transformada de Wavelets Contínua (CWT).



Após a transformação, a representação das diferentes frequências desses sinais é dada por uma função wavelet-mãe que possua um formato característico próximo das características encontradas na série-temporal. Neste caso, a função wavelet de Morlet é utilizada, pois sugere melhores adaptações os dados meteorológicos e/ou hidrológicos (Labat, 2005; Torrence e Compo, 1998).

Dois métodos de avaliação são utilizados para avaliar o grau de associação entre os dois sinais CWT  $\{W_{FMA(t,\omega)}$  e  $W_{Índice,mês(t,\omega)}\}$ , são eles: Wavelets Coerência (WTC) e Wavelet Cruzada (XWT). A WTC é um método para análise da coerência (medida através da covariância) e atraso de fase entre duas series temporais como uma função do tempo e frequência, onde pelo fato de seu cálculo ser análogo ao de Pearson, seus coeficientes de correlação variam de 0 a 1. A XWT mede a força de associação entre duas séries temporais através do espectro de potência da wavelet, onde é possível destacar os principais períodos de interferência de uma serie em relação a outra para diferentes frequências temporais (Labat, 2005; Gurgel, 2013).

Os resultados da WTC e XWT são dados através de um gráfico chamado escalograma, onde torna-se possível a visualização de eventos correlacionados, juntamente com seus coeficientes e níveis de significância, sobre as escalas de frequências, em cada instante de tempo para os sinais.

Para execução das correlações, foram utilizadas rotinas em ambiente Matlab®, criadas por Torrence e Compo (1998), e disponibilizadas no domínio <http://paos.colorado.edu/research/wavelets>. Como esse processo de correlação envolve muitas variáveis preditoras, ou seja, para cada índice climático tem-se 11 meses para testar cada um deles, essas rotinas foram modificadas afim de ler todos os dados diretamente do Excel® e executar as correlações para todos eles automaticamente.

#### 4.3.3 Critérios para Definição das Variáveis Preditoras

A definição das variáveis preditoras é de acordo com o valor da estatística de teste representada pelo valor-p. A priori são observados os índices mensais que obtenham um nível de confiança acima de 95% ( $p < 0,05$ ), dando mais destaque a medida que esse nível aumentar. Também serão observados os índices com nível de confiança a partir de 90% ( $p < 0,1$ ), que possuam algum destaque em seu comportamento mensal.

Essa definição acontecerá para os dois métodos, no qual com relação a análise feita pelo uso da Wavelets, haverá um contorno em negrito no gráfico do escalograma demarcando a região onde o nível de confiança é de 95% para um determinado intervalo de tempo.

## 4.4 PREVISÃO DE PRECIPITAÇÃO COM REDES NEURAIS ARTIFICIAIS

### 4.4.1 Aspectos gerais da aplicação

É utilizada uma rede neural com arquitetura do tipo progressiva (feedforward) de três camadas (uma camada de entrada, uma camada intermediária e uma camada de saída).

Após o estudo anterior, as variáveis preditoras definidas são novamente reorganizadas formando grupos de acordo com a quantidade de dados disponíveis da mesma forma que foi apresentado na Tabela 4.3 do item 4.2.2, separando os índices dentro de cada grupo de acordo com o mês de antecedência para a previsão. Para cada grupo existem combinações entre seus índices, os quais servem de entrada para a rede.

Caso haja um grupo formado pelas variáveis com 60 anos (ver Tabela 4.3), por seus dados apresentarem intervalos diferentes, o grupo ficará condicionado a um intervalo comum, nesse caso específico, com 57 anos (1951-2007). Um grupo extra também é formado com os índices mais bem correlacionados de cada um dos outros grupos, no qual o ficaram restrito ao mesmo intervalo de tempo em comum entre todos eles.

Para cada grupo é testada uma variação de topologias para cada combinação das variáveis de entrada. A priori, há uma variação da quantidade de dados distribuídos para treinamento-validação-teste. Para cada distribuição, há uma variação no número de neurônios na camada intermediária. Seguindo a linha do fluxograma apresentado na figura 4.1, a rede se encerra com uma única saída que é a precipitação prevista do acumulado FMA.

### 4.4.2 Variáveis de Entrada

Após serem formados grupos com os meses dos índices climáticos mais influentes na variabilidade da precipitação acumulada FMA, para cada grupo procurou-se estabelecer diferentes combinações entre as variáveis, de maneira a ter uma maior quantidade de arranjos de maneira a encontrar aquele que tenha um melhor desempenho na previsão. A fim de não haver no mesmo arranjo, variáveis que acabem fornecendo o mesmo tipo de informação, foi

usada uma tabela de correlação entre as variáveis juntamente com uma análise de cluster, sendo o suficiente para selecionar as variáveis de cada arranjo sem correr o risco de haver autocorrelações.

O número máximo e mínimo de variáveis em cada arranjo é de 5 e 2, respectivamente. A quantidade de arranjos varia de acordo com o grupo formado. Os tipos de arranjos entre as variáveis dentro de cada grupo são colocados da seguinte forma:

1º tipo de arranjo – Formado com o melhor índice de cada mês;

2º tipo de arranjo – Formado somente com índices do mesmo mês.

#### 4.4.3 Divisão e Pré-processamento dos Dados

Os registros disponíveis foram divididos em três grupos para o treinamento, teste e validação. Essa divisão foi feita de modo a utilizar o procedimento de validação cruzada. A divisão de dados foi realizada com o objetivo de distribuir os eventos da precipitação (FMA) de todas as intensidades ao longo das três séries, bem como procurar média semelhante e desvio padrão entre eles. São feitas no máximo 5 tipos de divisão de dados. Como há a possibilidade de se formarem grupos com diferentes quantidades de dados, a exemplo da Tabela 4.3 do item 4.2.2, a quantidade de tipos de divisão diminui para os grupos com menos dados.

Dessa forma, as distribuições de dados dentro de cada grupo seguem conforme a Tabela 4.3.

Tabela 4.3 – Levantamento de Índices climáticos e características

<b>Dados</b>	<b>Divisão</b>	<b>Medidas</b>	<b>Treino</b>	<b>Validação</b>	<b>Verificação</b>
<b>100 anos</b>	<i>1ª</i> <i>Distribuição</i>	Quant.	46	27	27
		Porcent.	46%	27%	27%
		Média	525.0	547.5	525.8
		Desvio	207.5	197.3	211.0
	<i>2ª</i> <i>Distribuição</i>	Quant.	50	25	25
		Porcent.	50%	25%	25%
		Média	525.8	533.2	540.3
		Desvio	217.3	180.2	206.2
	<i>3ª</i> <i>Distribuição</i>	Quant.	60	20	20
		Porcent.	60%	20%	20%
		Média	525.5	540.1	539.8
		Desvio	202.5	209.1	213.2
	<i>4ª</i> <i>Distribuição</i>	Quant.	66	17	17
		Porcent.	66%	17%	17%
		Média	528.1	533.2	541.5

		Desvio	201.3	207.3	222.9
		Quant.	70	15	15
	5 <sup>a</sup>	Porcent.	70%	15%	15%
	<i>Distribuição</i>	Média	530.5	543.7	522.4
		Desvio	211.5	216.7	162.4
<b>63 anos</b>		Quant.	35	14	14
		Porcent.	56%	22%	22%
	1 <sup>a</sup>	Média	517.1	517.3	523.2
	<i>Distribuição</i>	Desvio	182.1	186.1	186.0
		Quant.	41	11	11
		Porcent.	65%	17%	17%
	2 <sup>a</sup>	Média	522.0	512.4	511.7
	<i>Distribuição</i>	Desvio	183.8	184.0	183.4
		Quant.	45	9	9
		Porcent.	71%	14%	14%
	3 <sup>a</sup>	Média	518.6	514.2	522.3
	<i>Distribuição</i>	Desvio	182.0	186.0	191.4
<b>61 anos</b>		Quant.	33	14	14
		Porcent.	54%	23%	23%
	1 <sup>a</sup>	Média	519.16	523.52	525.42
	<i>Distribuição</i>	Desvio	185.86	184.44	186.49
		Quant.	39	11	11
		Porcent.	64%	18%	18%
	2 <sup>a</sup>	Média	520.39	526.26	521.22
	<i>Distribuição</i>	Desvio	186.15	186.47	183.15
		Quant.	43	9	9
		Porcent.	70%	15%	15%
	3 <sup>a</sup>	Média	521.99	520.00	521.30
	<i>Distribuição</i>	Desvio	185.88	187.33	183.16
<b>57 anos</b>		Quant.	29	14	14
		Porcent.	52%	25%	25%
	1 <sup>a</sup>	Média	533.82	531.93	532.73
	<i>Distribuição</i>	Desvio	184.08	191.84	188.04
		Quant.	35	11	11
		Porcent.	63%	20%	20%
	2 <sup>a</sup>	Média	530.07	533.82	532.12
	<i>Distribuição</i>	Desvio	184.82	186.84	183.97
		Quant.	41	8	8
		Porcent.	73%	14%	14%
	3 <sup>a</sup>	Média	531.65	532.29	536.38
	<i>Distribuição</i>	Desvio	184.55	183.32	183.94

Para todas as topologias das redes há somente uma camada intermediária, onde nessa camada, para cada distribuição, é simulada a RNA com 2 a 10 neurônios, a fim de buscar a distribuição e a quantidade de neurônios que apresentem os melhores resultados para a previsão. É usada a função de ativação sigmoide logística, usada por vários autores como mostrado na revisão, na camada intermediária para todas as topologias. Após a divisão dos dados, é realizado o pré-

processamento das variáveis de entrada e saída da rede segundo Maier e Dandy (2000). Os valores das variáveis são normalizados no intervalo [0,1, 0,9] pela equação 17, antes do processo de treinamento. No Anexo 1 se encontram os gráficos da divisão dos dados.

#### 4.4.4 Treinamento e critérios de parada

No treinamento da RNA, o algoritmo de aprendizagem utilizado é o backpropagation, mais comum no uso da rede neural MLP. O método de busca apresentado é o método do Gradiente Conjugado Escalado (SCGM - Scaled Conjugate Gradient Method), que é classificado como um método de busca local de segunda ordem (Paz et al., 2011). De acordo com o estudo de Chattopadhyay e Chattopadhyay (2008), esse método foi considerado o mais eficaz para previsão de precipitação, após um estudo comparativo com outros métodos (método de Levenberg-Marquardt e método do Momentum).

O algoritmo SCGM, desenvolvido por Moller (1993), combina a abordagem do algoritmo de Levenberg-Marquardt com a aproximação do gradiente conjugado. Em SCGM, o tamanho do passo é ajustado em cada uma das iterações. Uma busca é feita ao longo da direção do gradiente conjugado para determinar o tamanho do passo de aprendizagem, o que minimiza a função de desempenho ao longo dessa linha. O método SCGM pode passar para o mínimo de uma função quadrática N-dimensional em N passos. A direção de busca em cada uma das iterações é determinada por:

$$D_{k+1} = -\nabla E(w_{k+1}) - \beta \cdot \nabla E(w_k) \quad (\text{Equação 31})$$

$$\beta = \frac{[\nabla E(w_{k+1}) - \nabla E(w_k)] \cdot \nabla E(w_{k+1})}{[\nabla E(w_k)]^2} \quad (\text{Equação 32})$$

onde  $D_{k+1}$  é um vetor que define a direção de busca na iteração  $k+1$ ;  $w$  é o vetor dos pesos;  $\Delta w$  é a variação do vetor dos pesos;  $\nabla E(w)$  é o gradiente da função erro; e  $\beta$  é a taxa de aprendizagem. Como medida de desempenho a ser otimizada pelo treinamento, considerou-se o erro quadrático médio (MSE).

Com relação aos critérios de parada, fica definido um erro quadrático médio  $\varepsilon_1 = 10^{-6}$ . Também será considerado o erro entre duas interações consecutivas, no qual o algoritmo converge quando esse valor for  $\varepsilon_2 = 10^{-6}$  ( $\varepsilon_2 = \varepsilon_1^k - \varepsilon_1^{k-1}$ ). É definido um número máximo de interações igual a 10000 (dez mil) caso a precisão especificada se torne inalcançável. Com relação à

validação cruzada, após ser encontrado um ponto de mínimo, e os dados de verificação começarem a apresentar maiores erros, são realizadas mais 500 iterações a fim de verificar se não há outro ponto de mínimo. Caso não haja, é considerado o ponto anteriormente encontrado como um valor de mínimo, no qual ajusta melhor os pesos para cada amostra. O processo de treinamento ocorre em rotinas no ambiente Matlab®.

#### 4.5 MEDIDAS DE DESEMPENHO

Como medidas de desempenho para avaliação os resultados das previsões de precipitação, são utilizados os valores do coeficiente de correlação de Pearson, do erro médio absoluto (MAE) e do índice de eficiência de Nash-Sutcliffe (NS), sendo desta forma analisados por tipos diferentes métricas, referente a Diferença entre Produtos, Erros Absolutos e Erros Quadráticos respectivamente.

As medidas são aplicadas para cada conjunto de dados (treinamento-verificação-validação) e também de forma global. São realizados 10 treinamentos para cada topologia, de forma independente, ou seja, as respostas são armazenadas em planilha Excel®, e logo em seguida, todas as topologias são reinicializadas e a rede novamente treinada. A verificação final da topologia se constará após serem calculados média e o desvio-padrão dos 10 treinamentos, a fim de verificar o desempenho e a oscilação da mesma.

## 5 RESULTADOS E DISCURSÃO

### 5.1 CORRELAÇÕES PRECIPITAÇÃO X INDICES CLIMATICOS (TELECONEXÕES)

Serão apresentados primeiramente as correlações dos índices climáticos de acordo com as regiões oceânicas as quais são contabilizadas características significativas de algum padrão climático que tenha uma teleconexão com a precipitação acumulada FMA da região de estudo.

A região do Oceano Pacífico abrange a maior quantidade de índices climáticos coletados (Tabela 5.1). Ao todo, foram 18 índices nessa região, dos quais 12, apresentaram correlações com níveis de confiança de pelo menos 95% ao longo dos meses. Destes melhores índices correlacionados, cinco ( $SOI_{noise}$ , NINO 3, NINO 3.4, ONI e NOI) se destacam por apresentarem correlações com níveis de confiança superiores a 99%, com defasagem de 3 a 11 meses. Nitidamente é percebida a forte influência da região do Ninos sobre o regime da precipitação estudada, no qual as correlações se estendem por vários meses antecedentes. Em alguns índices surgem boas correlações de forma isolada em determinado mês, no qual nos outros meses desse índice não há bons resultados, como é o caso do  $SOI_{noise}$  e do EPO.

Vale ressaltar que as regiões de análise de grande parte dos índices do Oceano Pacífico acabam se sobrepondo, nos quais deixam as suas correlações com tendências parecidas ao longo dos meses, como é o caso dos índices MEI, ONI e TP em relação a região dos Ninos. Pode-se notar o quanto eles conseguem manter a coerência de correlações entre si. Mesmo assim, também vale ressaltar, que cada índice contabiliza de maneira única determinado tipo de oscilação presente naquela região, pois cada um tem a sua maneira de medir. E dentro disso, algumas dessas medidas acabam por enaltecer determinado tipo de anomalia que ocorre nessa região.

É demonstrado que grande parte dos índices no Pacífico tem maiores correlações nos meses de menor antecedência em relação ao trimestre chuvoso da bacia de estudo, e que vão perdendo “força” a medida que aumenta a antecedência. Esse fato é bem nítido para os índices da região dos Ninos, onde o NINO 3 merece destaque por conseguir manter-se regular em suas correlações a medida que aumenta a antecedência. Outro ponto de destaque é para o intervalo de maior influência que o NINO 1+2 em dissonância com os outros dessa região, pois ele apresenta maior poder teleconectivo em um intervalo onde os outros já perdem a influência.

Tabela 5.1 – Correlação entre os Índices Climáticos do Pacífico Precipitação de FMA

Antec. (Meses)		2	3	4	5	6	7	8	9	10	11	12
Índice	*p	DEZ	NOV	OUT	SET	AGO	JUL	JUN	MAI	ABR	MAR	FEV
SOI <sub>signal</sub>	R	0.07	0.13	0.09	0.04	0.04	0.03	0.07	0.01	0.02	-0.02	-0.14
	Rs	0.09	0.14	0.10	0.07	0.07	0.01	0.06	0.01	0.01	0.01	-0.11
SOI <sub>noise</sub>	R	-0.16	-0.08	-0.13	-0.15	-0.03	-0.18	-0.30	-0.04	-0.09	-0.27	-0.07
	Rs	-0.03	-0.01	-0.05	-0.09	0.02	-0.15	-0.26	-0.02	-0.08	-0.28	-0.06
PDO	R	-0.03	-0.03	-0.10	-0.12	-0.12	-0.07	-0.15	-0.08	0.03	-0.04	0.07
	Rs	-0.05	-0.03	-0.15	-0.13	-0.13	-0.10	-0.13	-0.08	-0.01	-0.07	0.08
TNI	R	0.04	0.00	-0.05	-0.08	-0.08	-0.08	-0.08	-0.10	-0.08	-0.04	0.01
	Rs	0.06	0.01	-0.02	-0.05	-0.05	-0.07	-0.08	-0.09	-0.07	-0.03	0.02
CTI	R	-0.23	-0.21	-0.17	-0.11	-0.13	-0.17	0.00	-0.09	-0.03	-0.01	0.04
	Rs	-0.25	-0.21	-0.16	-0.07	-0.12	-0.15	-0.14	-0.11	-0.02	-0.04	0.06
MEI	R	-0.35	-0.30	-0.28	-0.29	-0.22	-0.23	-0.22	-0.16	-0.10	0.01	0.07
	Rs	-0.33	-0.28	-0.27	-0.24	-0.23	-0.22	-0.26	-0.26	-0.16	0.00	0.04
NINO 1+2	R	-0.28	-0.25	-0.27	-0.23	-0.25	-0.21	-0.28	-0.28	-0.26	-0.25	-0.04
	Rs	-0.25	-0.21	-0.23	-0.16	-0.22	-0.17	-0.31	-0.27	-0.27	-0.26	-0.10
NINO 3	R	-0.38	-0.35	-0.29	-0.28	-0.28	-0.29	-0.33	-0.31	-0.19	-0.09	0.05
	Rs	-0.37	-0.35	-0.26	-0.26	-0.24	-0.22	-0.31	-0.35	-0.25	-0.12	0.02
NINO 3.4	R	-0.38	-0.35	-0.28	-0.27	-0.25	-0.26	-0.29	-0.25	-0.11	0.02	0.08
	Rs	-0.36	-0.33	-0.27	-0.26	-0.25	-0.24	-0.30	-0.29	-0.12	-0.01	0.05
NINO 4	R	-0.37	-0.28	-0.26	-0.25	-0.20	-0.17	-0.18	-0.13	-0.07	0.05	0.04
	Rs	-0.36	-0.28	-0.28	-0.28	-0.21	-0.20	-0.20	-0.17	-0.09	-0.03	0.00
ONI	R	-0.39	-0.35	-0.31	-0.27	-0.26	-0.27	-0.27	-0.21	-0.09	0.02	0.07
	Rs	-0.36	-0.32	-0.29	-0.25	-0.24	-0.25	-0.28	-0.23	-0.09	-0.01	0.05
PNA	R	0.02	-0.19	-0.02	0.03	0.07	-0.11	-0.03	-0.05	0.11	0.09	-0.01
	Rs	0.02	-0.14	0.07	-0.03	0.05	-0.11	-0.05	-0.06	0.06	0.03	0.00
TP	R	-0.36	-0.30	-0.28	-0.25	-0.24	-0.22	-0.24	-0.23	-0.15	-0.08	0.04
	Rs	-0.31	-0.27	-0.24	-0.22	-0.19	-0.16	-0.24	-0.24	-0.18	-0.10	0.02
EPO	R	0.00	0.05	-0.21	0.07	0.04	-0.13	-0.20	-0.27	-0.09	-0.18	0.06
	Rs		0.00	-0.17	0.09	0.06	-0.10	-0.23	-0.22	-0.18	-0.12	0.07
NOI	R	0.18	0.05	0.30	0.23	0.06	-0.05	0.33	0.22	0.21	0.17	-0.02
	Rs	0.24	-0.01	0.28	0.16	-0.04	-0.05	0.29	0.22	0.16	0.19	0.01
PMM <sub>sst</sub>	R	0.06	0.05	-0.02	-0.01	-0.05	0.00	-0.09	-0.02	0.03	0.07	-0.03
	Rs	-0.02	0.01	-0.08	-0.08	-0.08	-0.03	-0.13	-0.07	0.00	0.05	-0.05
PMM <sub>wind</sub>	R	0.06	-0.01	-0.17	-0.13	-0.09	-0.10	-0.11	-0.18	-0.18	-0.01	-0.07
	Rs	0.10	-0.06	-0.22	-0.17	-0.12	-0.13	-0.08	-0.22	-0.21	0.06	-0.02
IPO	R	-0.28	-0.22	-0.17	-0.17	-0.19	-0.14	-0.15	-0.07	-0.08	-0.06	0.00
	Rs	-0.24	-0.21	-0.18	-0.17	-0.18	-0.13	-0.18	-0.11	-0.08	-0.05	0.01

\* Identificação dos Coeficientes de Pearson (R) e de Spearman (Rs).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul – positivo) e 99% (amarelo).

Na região do Oceano Atlântico foram coletados 11 índices climáticos, dos quais 6 apresentaram correlações com níveis de confiança de pelo menos 95% ao longo dos meses (Tabela 5.2). Destes melhores índices correlacionados, apenas um, o AMO, apresenta pelo menos uma correlação superior a 99% de confiança, com uma defasagem de 3 meses. A versão suavizada do AMO, consegue captar a influência do mesmo com níveis de confiança superiores de 95%



sobre a precipitação de estudo durante todos os meses defasados. Por ser um índice que capta as oscilações multidecadais, a sua variabilidade acaba permanecendo quase que constante ao longo dos meses, de modo que consegue estender também a sua correlação.

Os índices TNA e TSA, se comportam de maneira a se contrabalancearem, pois devido a suas regiões de abrangência (ao Norte e Sul do Atlântico, respectivamente), proporcionam correlações com sinais opostos. Ou seja, o índice TSA representa as anomalias de temperatura das águas da região ao sul do Atlântico, em um quadrante abaixo da linha do Equador. De maneira análoga, é tratado também o índice TNA ao norte do Atlântico. A correlação positiva do índice TSA refere-se ao fato de que ao aquecer essa região, acontece uma formação de uma zona de baixa pressão que tenderá a deslocar a ZCIT para o sul, proporcionando mais chuvas. Um aquecimento maior ao norte, provoca o seu deslocamento para a direção do mesmo, diminuindo as chuvas nas proximidades do Nordeste, trazendo uma correlação negativa para esse índice. Os índices STA e NTA, também representam regiões ao sul e norte, mas com dimensões e posições um pouco diferentes, onde evidentemente, também trazem as mesmas tendências, só que com outras intensidades.

Dentre os meses quantificados pelo índice NAO, o mês de novembro (3 meses de antecedência) salta com uma correlação negativa bem expressiva ( $p < 0,05$ ) destacando-se devido a inferioridade dos outros valores mensais do mesmo. A influência da NAO na América do Sul ainda não está muito bem definida. Mesmo assim, esse mês de destaque com a precipitação de estudo, coloca alguma evidência de teleconexão.

Tabela 5.2 – Correlação entre Índices Climáticos do Atlântico Precipitação de FMA

Antec. (Meses)		2	3	4	5	6	7	8	9	10	11	12
Índice	*p	DEZ	NOV	OUT	SET	AGO	JUL	JUN	MAI	ABR	MAR	FEV
NAO	R	0.07	-0.21	0.01	0.02	-0.05	0.02	0.02	0.01	0.03	-0.09	-0.07
	Rs	0.07	-0.18	-0.06	-0.02	-0.08	0.03	0.00	-0.01	0.02	-0.07	-0.11
TNA	R	-0.25	-0.19	-0.25	-0.19	-0.15	-0.13	-0.09	-0.01	0.04	-0.03	0.04
	Rs	-0.27	-0.16	-0.21	-0.21	-0.17	-0.15	-0.10	0.00	0.06	-0.01	0.01
TSA	R	0.21	0.24	0.20	0.32	0.31	0.30	0.26	0.20	0.20	0.13	0.16
	Rs	0.21	0.21	0.14	0.18	0.23	0.27	0.18	0.14	0.11	0.07	0.09
AMO <sub>uns</sub>	R	-0.22	-0.26	-0.21	-0.17	-0.14	-0.10	-0.14	-0.10	-0.13	-0.23	-0.18
	Rs	-0.25	-0.27	-0.19	-0.18	-0.13	-0.10	-0.13	-0.08	-0.09	-0.21	-0.15
AMO <sub>sno</sub>	R	-0.23	-0.23	-0.22	-0.22	-0.23	-0.23	-0.23	-0.23	-0.23	-0.23	-0.23
	Rs	-0.21	-0.21	-0.21	-0.21	-0.21	-0.22	-0.22	-0.23	-0.23	-0.23	-0.23
ATI	R	-0.23	-0.20	-0.22	-0.14	-0.15	-0.15	-0.20	-0.15	-0.07	-0.20	-0.10
	Rs	-0.25	-0.22	-0.20	-0.17	-0.13	-0.13	-0.20	-0.12	-0.07	-0.19	-0.08
STA	R	0.22	0.18	0.18	0.20	0.24	0.23	0.22	0.20	0.17	0.15	0.15
	Rs	0.21	0.15	0.10	0.11	0.16	0.18	0.13	0.14	0.10	0.08	0.10
NTA	R	-0.28	-0.22	-0.23	-0.20	-0.16	-0.13	-0.07	-0.02	0.00	0.02	0.01
	Rs	-0.30	-0.24	-0.23	-0.22	-0.16	-0.16	-0.10	-0.03	-0.02	0.02	-0.02
EA	R	0.11	-0.09	-0.11	-0.10	-0.04	-0.13	-0.09	-0.01	-0.08	-0.03	-0.20
	Rs	-0.03	-0.12	-0.13	-0.12	-0.05	-0.10	-0.10	-0.03	-0.11	-0.08	-0.20
AMM <sub>sst</sub>	R	-0.15	-0.18	-0.21	-0.21	-0.22	-0.21	-0.11	-0.07	-0.04	-0.11	-0.08
	Rs	-0.19	-0.21	-0.18	-0.16	-0.18	-0.19	-0.13	-0.08	-0.03	-0.06	-0.07
AMM <sub>wind</sub>	R	-0.09	0.06	-0.21	-0.12	-0.17	-0.14	-0.25	-0.24	0.14	-0.14	-0.09
	Rs	-0.08	0.03	-0.23	-0.09	-0.12	-0.10	-0.26	-0.23	0.19	-0.18	-0.04

\* Identificação dos Coeficientes de Pearson (R) e de Spearman (Rs).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul – positivo) e 99% (amarelo).

Com relação as outras regiões oceânicas (Tabela 5.3), não houve grandes destaques passíveis de teleconexões. Somente o índice AO, da região Ártica, apresenta uma correlação positiva de destaque ( $p < 0,005$ ) em um de seus meses (Junho – 8 meses de antecedência). Na região Antártica o seu análogo AAO, não esboçou boas correlações.

Tabela 5.3 – Correlação os Índices Climáticos do Índico, Ártico e Antártico e a Precipitação de FMA

Antec. (Meses)			2	3	4	5	6	7	8	9	10	11	12
Região	Índice	*p	DEZ	NOV	OUT	SET	AGO	JUL	JUN	MAI	ABR	MAR	FEV
ÍNDICO	IND	R	-0.14	-0.07	-0.02	0.00	0.01	-0.01	-0.03	-0.04	-0.03	0.01	0.01
		Rs	-0.16	-0.08	-0.06	-0.04	-0.02	-0.03	-0.04	-0.06	-0.05	-0.05	-0.04
ÁRTICO	AO	R	0.05	-0.01	0.05	0.07	-0.04	0.10	0.26	0.06	-0.19	-0.05	-0.06
		Rs	0.05	-0.04	0.08	0.04	-0.06	0.08	0.26	-0.02	-0.16	0.03	-0.05
ANTART.	AAO	R	0.10	0.01	0.04	-0.03	0.01	-0.12	0.08	0.10	0.05	0.01	-0.03
		Rs	0.13	0.03	0.11	0.00	-0.05	-0.10	0.11	0.12	0.04	0.03	-0.01
	TPI	R	0.04	-0.05	-0.14	-0.09	-0.10	-0.04	-0.12	0.15	0.13	0.05	-0.02
		Rs	0.08	-0.03	-0.16	-0.07	-0.08	-0.01	-0.09	0.16	0.07	0.09	-0.03

\* Identificação dos Coeficientes de Pearson (R) e de Spearman (Rs).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul – positivo) e 99% (amarelo).

Dos 4 índices globais analisados, dois tiveram correlações significativas ( $p < 0,05$ ), GLAMM e Solar Flux (Tabela 5.4). Nitidamente, ficou clara uma relação existente entre o fluxo solar, ou no caso, as manchas solares e a precipitação da região de estudo. Essa correlação negativa estaria associada as secas ocorrente na região. Pelo fato do ciclo de manchas solares se renovarem a cada 11 anos em média, a variabilidade ao longo dos meses se manteria baixa, estendendo a correlação deste índice em todos os meses. Outro destaque interessante está para as correlações significativas do índice GLAAM durante alguns meses. Como este índice, em essência, é uma medida do saldo líquido dos ventos em todo o planeta, as anomalias ocorrentes neste índice, estão bastante associadas a anomalias do ENSO, pois este tem maior influência sobre o bloqueio e o enfraquecimento dos mesmos, principalmente na região do Equador, e conseqüentemente pode evidenciar algum padrão climático formado não só pelo ENSO mas também pelo conjunto com outros modos de variabilidade que exerçam influência sobre o Nordeste.

Tabela 5.4 – Correlação dos Índices Climáticos Globais e a Precipitação de FMA

Antec. (Meses)			2	3	4	5	6	7	8	9	10	11	12
Índice	*p		DEZ	NOV	OUT	SET	AGO	JUL	JUN	MAI	ABR	MAR	FEV
QBO	R		-0.05	-0.10	-0.11	-0.10	-0.07	-0.06	-0.04	-0.02	0.00	0.02	0.00
	Rs		-0.01	-0.02	-0.02	-0.04	-0.02	-0.03	-0.01	0.04	0.08	0.10	0.09
GMLOTI	R		-0.14	-0.10	-0.16	-0.11	-0.15	-0.10	-0.12	-0.09	-0.10	-0.10	-0.07
	Rs		-0.13	-0.12	-0.18	-0.11	-0.14	-0.08	-0.12	-0.13	-0.12	-0.12	-0.10
GLAAM	R		-0.44	-0.32	-0.19	-0.23	-0.23	-0.28	-0.29	-0.29	-0.28	-0.19	-0.14
	Rs		-0.41	-0.26	-0.10	-0.15	-0.17	-0.28	-0.24	-0.23	-0.22	-0.15	-0.07
Solar Flux	R		-0.33	-0.33	-0.35	-0.32	-0.29	-0.29	-0.26	-0.21	-0.29	-0.28	-0.26
	Rs		-0.34	-0.33	-0.30	-0.31	-0.28	-0.26	-0.24	-0.21	-0.25	-0.27	-0.26

\* Identificação dos de Pearson (R) e de Spearman (Rs).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul - positivo) e 99% (amarelo).

Sendo analisadas anteriormente as possíveis teleconexões existentes através dos coeficientes de Pearson e Spearman, a seguir serão apresentados os gráficos dos escalogramas correspondentes as melhores correlações, calculada pelas técnicas de Wavelet Cruzada e Wavelet Coerência, dos valores mensais dos índices para cada região oceânica.

A região do Oceano Pacífico, sob essa nova análise de Wavelets, propôs uma maior quantidade de índices climáticos selecionados, em relação ao método anterior, total de 18, com boas correlações em várias frequências de tempo (Figura 5.1, 5.2, 5.3).

Destes melhores índices mensais, destaque para a região dos Ninos, ao qual demonstrou forte correlação para o período de tempo de 5-6 anos para alguns meses e para outros encontrava evidências entre o período de 12-13 anos. Destaque para o NINO 4 que apresenta praticamente correlações somente para o período de 12-13 anos. Como a maioria dos índices escolhidos do oceano Pacífico optam por captar mais as oscilações da região Tropical, ao qual está inserida a região do Ninos, grande parte dos índices selecionados demonstram suas correlações para os mesmos períodos de tempo destacados acima. Por essa análise, foi possível destacar outros índices que não foram bem correlacionados no método anterior, mas que demonstraram boas correlações com níveis de confiança de pelo menos 95%. Nesse caso, temos o índice PMM que tanto para sua parte referente a oscilação da temperatura do oceano quanto a dos ventos, demarcam na maioria dos meses, correlações para o período de 12-13 anos. O índice PNA, nessa mesma linha, também demonstrou boas correlações para o período citado, com meses apresentando maiores tendências para o período 15-16 anos. Detalhe para o índice PDO ao qual, em ambos os métodos, não obteve boas correlações com a precipitação da área de estudo.

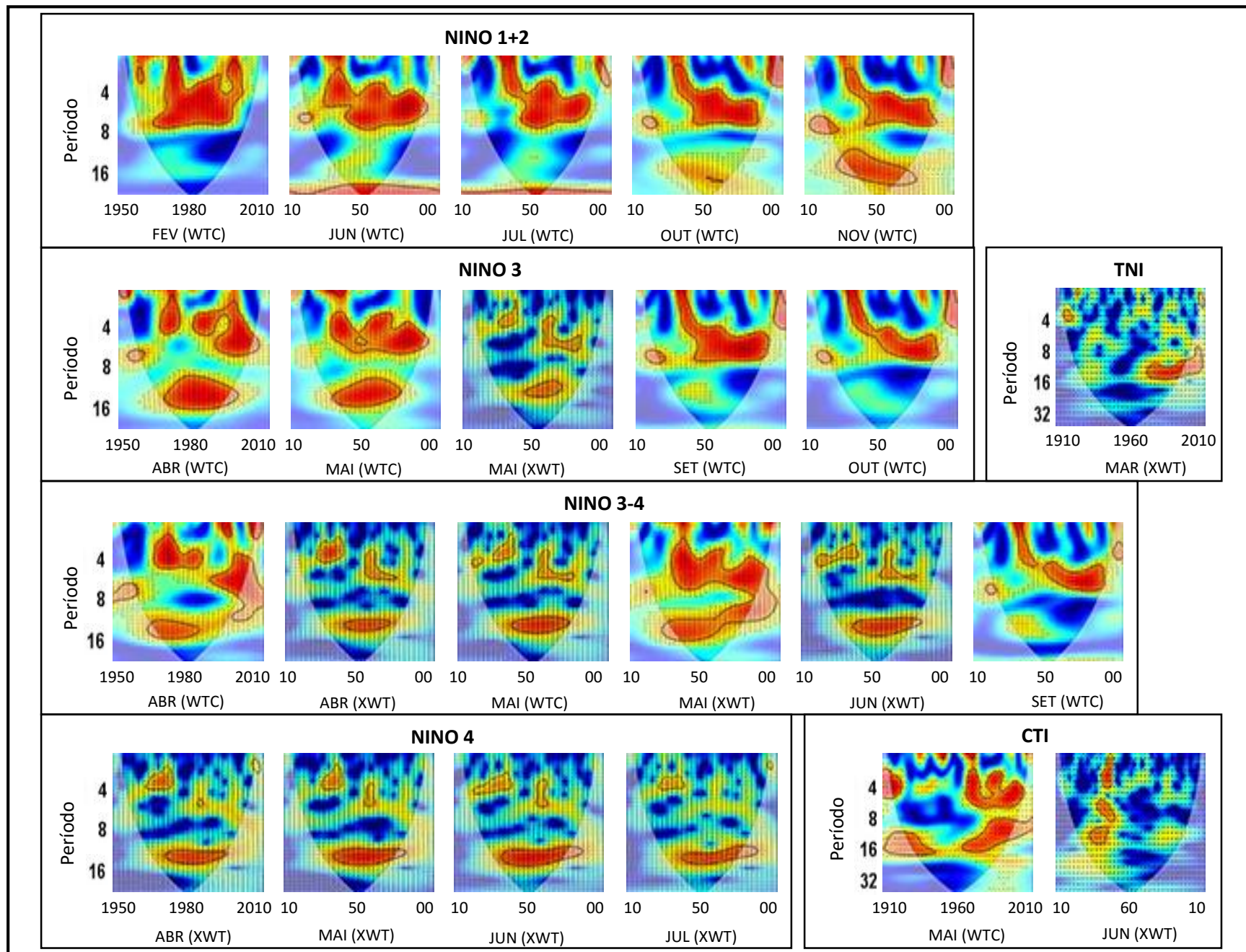


Figura 5.1 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Pacífico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulado.

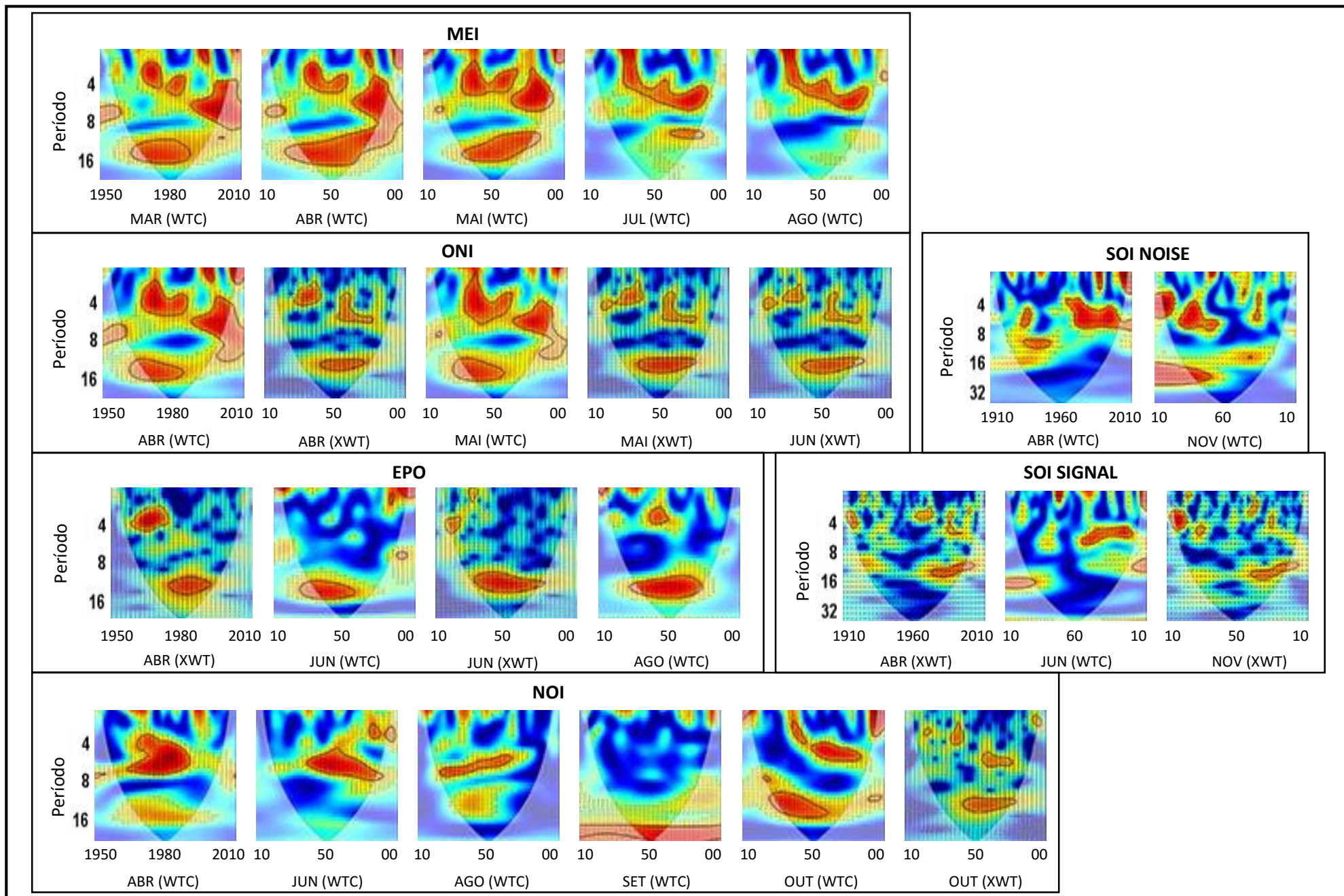


Figura 5.2 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Pacífico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulado

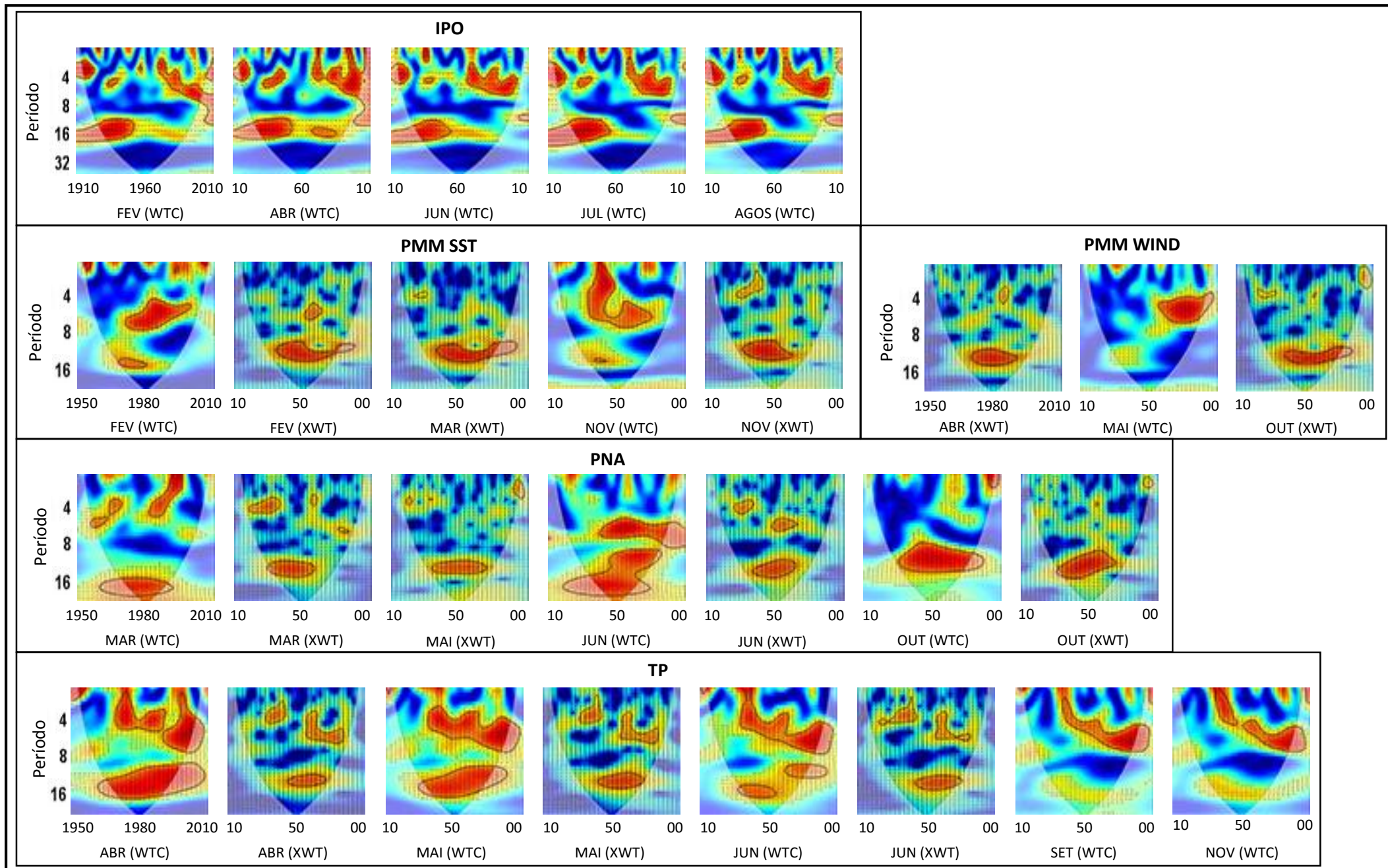


Figura 5.3 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Pacífico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulado

Na região do Atlântico, foi possível selecionar todos os 11 índices avaliados, com boas correlações para pelo menos um de seus meses, como foi o caso de índice  $AMO_{smo}$  (Figura 5.5). Ao contrário do outro método, esse índice não obteve boas correlações em suas imagens, no qual foi considerado apenas o escalograma da série do mês de novembro como o mais representativo. A versão original do AMO, também obteve uma fraca correlação ao longo dos meses, no qual os melhores escalogramas foram dos meses com maior e menor antecedência, fevereiro e novembro, com uma correlação muito demarcada no período de 4-6 anos e 15-16 anos, respectivamente.

Em geral, a maioria dos índices nessa análise demonstraram maiores correlações para o período entre 8 e 16 anos. Os índices com maior destaque foram o TSA e STA, ao qual abrangendo regiões que se interceptam no Atlântico Tropical, apresentaram correlações para o período acima de 16 anos (Figura 5.4).

Dos 4 índices globais analisados por Wavelets, as melhores correlações foram para os índices QBO e o Fluxo Solar (Figura 5.7). Ambos tiveram correlações com nível de confiança acima 95%, ao qual foram bem delimitadas para o período correspondente a 11-12 anos. Com relação ao Fluxo Solar, as correlações de sinal negativo apresentada pelo método anterior, com as tendências apresentadas pela análise de Wavelets, pode indicar a influência solar na diminuição dos totais de precipitação do trimestre analisado. Alguns autores têm tentado atribuir as secas nordestinas a causas externas ao sistema terra-oceano-atmosfera. Sampaio Ferraz (1950) apud Molion e Bernado (2002), por exemplo, notou periodicidades aparentes de 12 anos na série pluviométrica de Fortaleza, e sugeriu que as secas estavam relacionadas com manchas solares, fazendo uma previsão de que haveria uma seca severa no máximo solar seguinte (1957/58). Essa previsão foi acertada, no qual o máximo de manchas foi o maior registrado até aquela data e a seca foi uma das mais severas. O outro índice melhor correlacionado, a Oscilação Quase-Bienal (QBO), é caracterizada por períodos de alternância entre ventos de leste (anomalia negativa) e de oeste (anomalia positiva) na estratosfera tropical (Barnston e Livezey, 1989). Medina e Nunes (2005) verificaram que a anomalia positiva do vento zonal estratosférico, favorece as condições de valores mais elevados de precipitação na região Nordeste, do que no período de anomalia negativa. Essa influência nos campos de superfície dos ventos pode ser explicada pela perturbação que a QBO provoca na célula de Walker, o que altera a circulação global (Barnston e Livezey, 1989). Por ser afetado pelo fenômeno ENOS, os resultados aqui apresentados não são climatologicamente conclusivos, no qual, mais estudos sobre a QBO e seus impactos na região do semiárido devem ser realizados.



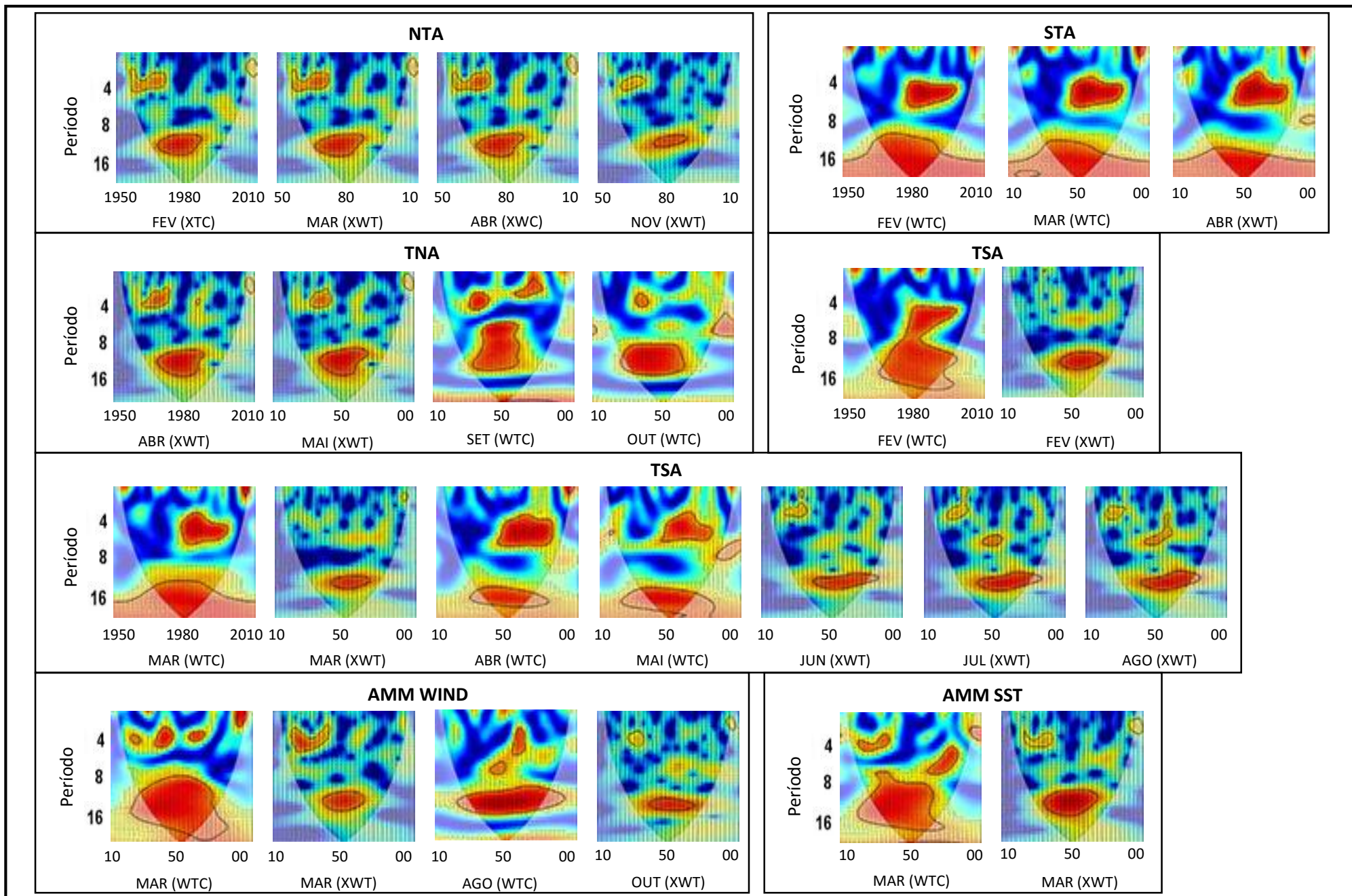


Figura 5.4 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Atlântico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulado

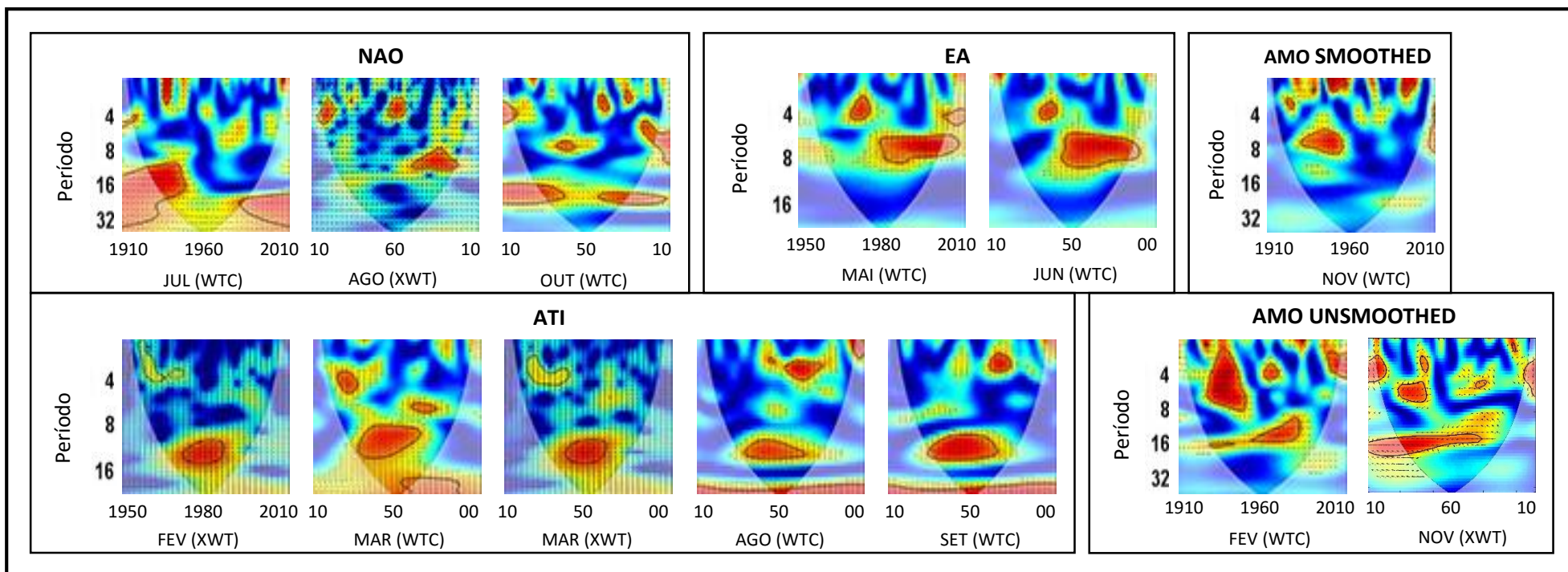


Figura 5.5 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Oceano Atlântico e a Precipitação de FMA. Nível de confiança de 95% delimitado na área circulado

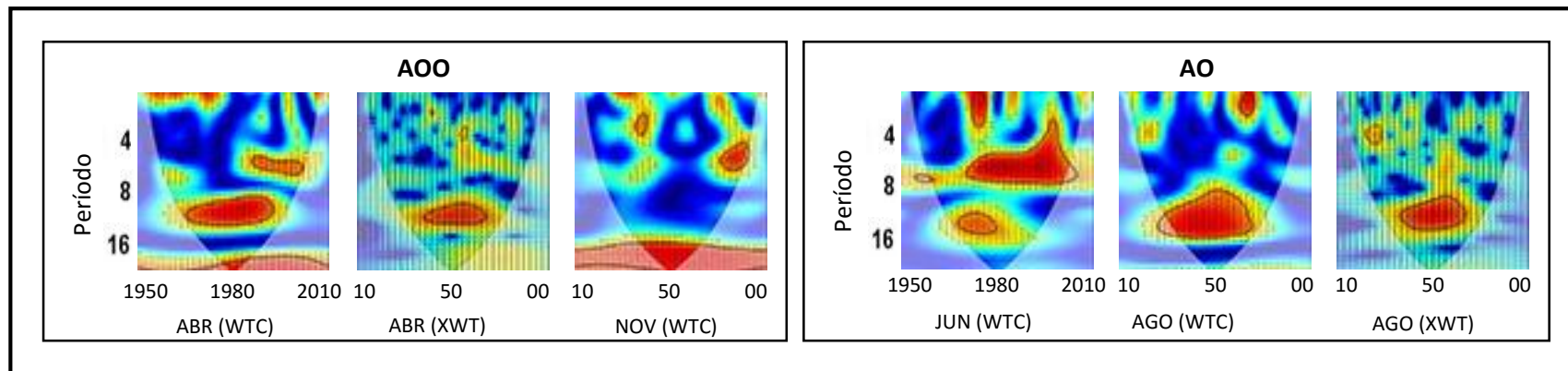


Figura 5.6 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos do Ártico/Antártico e a Precipitação de FMA.

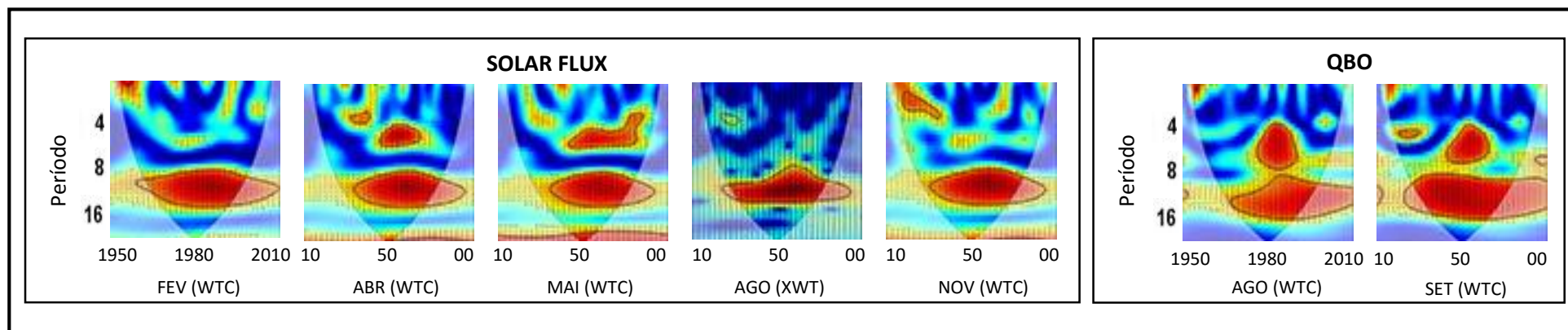


Figura 5.7 – Melhores resultados das análises de Wavelets Coerência (WTC) e Wavelets Cruzada (XWT) entre os índices climáticos Globais e a Precipitação de FMA.

## 5.2 VARIÁVEIS PREDITORAS E COMBINAÇÕES

Dentre os índices analisados para cada região, para a definição das variáveis predictoras, foi feita novamente a separação dos mesmos de acordo com o maior intervalo de tempo em comum, formando-se assim grupos distintos com a mesma quantidade de dados para a entrada do modelo de previsão. Desta forma, os grupos formados foram de acordo com as tabelas 5.5, 5.6, 5.7, 5.8 e 5.9.

Avaliando-se os grupos de maneira conjunta, percebe-se que o 2º grupo detém o maior número de variáveis predictoras em relação ao 1º grupo. O significativo aumento de variáveis 2º grupo, foi devido a técnica de seleção na qual consegue captar correlações em várias frequências para diferentes intervalos de tempo na série. Percebe-se também que em relação ao 1º grupo, a maioria dos índices apresentam níveis de significância de 95%, com exceção do intervalo de 57 anos. Quanto ao 2º grupo, a maioria dos selecionados apresentam uma maior correlação destacados pela técnica de Wavelet Coerência.

Tabela 5.5 – Variáveis predictoras com 100 anos (1910-2009)

	Índices	Tipo*	Mês	Antecedência		Índices	Tipo*	Mês	Antecedência
<b>1º Grupo</b>	NAO	r	NOV	3 Meses	<b>2º Grupo</b>	SOI	X	NOV	3 Meses
	AMO uns	r/rs				SOI noise	C		
	CTI	r/rs				AMO smo	C		
	AMO smo	r/rs				NAO	C	OUT	4 Meses
	IPO	r/rs				NAO	X	AGO	6 Meses
	AMO smo	r/rs	SET	5 Meses		NAO	C	JUL	7 Meses
	IPO	r	AGO	6 Meses		IPO	C	JUN	8 Meses
	SOI noise	r/rs	JUN	8 Meses		SOI	C		
	AMO smo	r/rs	MAR	11 Meses		SOI signal	C		
	SOI noise	r/rs				IPO	C	CTI	X
AMO uns	r/rs	CTI			C	MAI	9 Meses		
AMO smo	r/rs			SOI noise	C	ABR	10 Meses		
				SOI signal	X	MAR	11 Meses		
				IPO	C				
				TNI	X	FEV	12 Meses		
				AMO uns	C				
				IPO	C				

\* Identificação do tipo de correlação: Pearson (r), Spearman (rs), Wavelet Coerência (C) e Cruzada (X).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul - positivo) e 99% (amarelo). No caso da Wavelets a avaliação é feita no escalograma.

Tabela 5.6 – Variáveis predictoras com 63 anos (1948-2010)

1º Grupo	Índices	Tipo	Mês	Antecedência
	Solar Flux	r/rs	NOV	3 Meses
	Solar Flux	r	OUT	4 Meses
	NOI	r/rs		
	TNA	r		
	TSA	r	SET	5 Meses
	Solar Flux	r/rs		
	TSA	r	AGO	6 Meses
	Solar Flux	r/rs		
	TSA	r/rs	JUL	7 Meses
	Solar Flux	r/rs		
TSA	r	JUN	8 Meses	
Solar Flux	r			
NOI	r			
AMM_wind	rs			
AMM_wind	r/rs	MAI	9 Meses	
Solar Flux	r/rs	MAR	11 Meses	

2º Grupo	Índices	Tipo	Mês	Antecedência
	Solar Flux	XC	NOV	3 Meses
	PMM_wind	XC		
	AAO	C		
	PMM_wind	X	OUT	4 Meses
	AMM_wind	X		
	NOI	XC		
	TNA	C		
	QBO	C	SET	5 Meses
	NOI	C		
	TNA	C		
	Solar Flux	XC	AGO	6 Meses
	QBO	C		
	AMM_wind	C		
	NOI	C		
TSA	X			
NOI	C	JUN	8 Meses	
TSA	X			
Solar Flux	XC	MAI	9 Meses	
PMM_wind	C			
TSA	C			
TNA	X			
Solar Flux	XC	ABR	10 Meses	
PMM_wind	X			
NOI	XC			
TSA	C			
TNA	X			
AAO	XC			
PMM_sst	X	MAR	11 Meses	
AMM_wind	XC			
AMM_sst	XC			
TSA	XC			
Solar Flux	XC	FEV	12 Meses	
PMM_sst	XC			
TSA	XC			

\* Identificação doo tipo de correlação: Pearson (r), Spearman (rs), Wavelet Coerência (C) e Cruzada (X).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul - positivo) e 99% (amarelo). No caso da Wavelets a avaliação é feita no escalograma.

Tabela 5.7 – Variáveis predictoras com 61 anos (1950-2010)

1º Grupo	Índices	Tipo	Mês	Antecedência	2º Grupo	Índices	Tipo	Mês	Antecedência
	NINO 3	r/rs					NINO 1+2	C	NOV
NINO 3.4	r				NINO 1+2	C			
NINO 4	r/rs		NOV	3 Meses	NINO 3	C	OUT	4 Meses	
MEI	r/rs				PNA	XC			
ONI	r				NINO 3	C	SET	5 Meses	
NINO 1+2	r				NINO 3.4	C			
NINO 3	r/rs		OUT	4 Meses	MEI	C			
NINO 3.4	r/rs				AO	XC	AGO	6 Meses	
NINO 4	r/rs				EPO	C			
MEI	r/rs				NINO 1+2	C			
ONI	r/rs				NINO 4	X	JUL	7 Meses	
NINO 3	r/rs				MEI	C			
NINO 3.4	r/rs		SET	5 Meses	NINO 1+2	C			
NINO 4	rs				NINO 3.4	X			
MEI	r				NINO 4	X			
ONI	r				ONI	X	JUN	8 Meses	
NINO 3	r		AGO	6 Meses	EPO	XC			
NINO 3.4	r				AO	C			
ONI	r				PNA	XC			
NINO 3	r				EA	C			
NINO 3.4	r		JUL	7 Meses	NINO 3	XC			
ONI	r				NINO 3.4	XC			
NINO 1+2	r/rs				NINO 4	X	MAI	9 Meses	
NINO 3	r				MEI	C			
NINO 3.4	r/rs		JUN	8 Meses	ONI	XC			
MEI	rs				PNA	X			
ONI	r/rs				EA	C			
AO	r/rs				NINO 3	C			
NINO 1+2	r/rs				NINO 3.4	XC			
NINO 3	rs		MAI	9 Meses	NINO 4	X	ABR	10 Meses	
NINO 3.4	r/rs				MEI	C			
MEI	rs				ONI	XC			
EPO	r				EPO	X			
NINO 1+2	r/rs		ABR	10 Meses	MEI	C			
NINO 1+2	rs		MAR	11 Meses	PNA	XC	MAR	11 Meses	
					NINO 1+2	C	FEV	12 Meses	

\* Identificação do tipo de correlação: Pearson (r), Spearman (rs), Wavelet Coerência (C) e Cruzada (X).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul - positivo) e 99% (amarelo). No caso da Wavelets a avaliação é feita no escalograma.

Tabela 5.8 – Variáveis predictoras com 57 anos (1951-2008)

1º Grupo	Índices	Tipo	Mês	Antecedência	2º Grupo	Índices	Tipo	Mês	Antecedência
	TP-SST-EOF	r/rs	NOV	3 Meses		NTA	X	NOV	3 Meses
	TP-SST-EOF	r	OUT	4 Meses		TP-SST-EOF	C	SET	5 Meses
	NTA	r/rs				ATI	C		
	STA	r	JUL	7 Meses		TP-SST-EOF	C	AGO	6 Meses
ATI	r/rs	JUN	8 Meses	ATI		C	JUN	8 Meses	
				TP-SST-EOF		XC	MAI	9 Meses	
				STA		C	ABR	10 Meses	
				NTA		X			
				TP-SST-EOF		XC	MAR	11 Meses	
				STA		C			
				NTA		X			
				ATI		XC	FEV	12 Meses	
				STA		C			
				NTA		X			
				ATI	X				

\* Identificação doo tipo de correlação: Pearson (R), Spearman (Rs), Wavelet Coerência (C) e Cruzada (X).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul - positivo) e 99% (amarelo). No caso da Wavelets a avaliação é feita no escalograma.

Tabela 5.9 – Variáveis predictoras formada por todos os índices com 61 anos (1950-2010)

1º Grupo	Índices	Tipo	Mês	Antecedência		
	1º Grupo	SOI	r/rs	NOV	3 Meses	
SOI signal		r/rs				
CTI		r/rs				
IPO		r/rs				
Solar Flux		r/rs				
NINO 3		r/rs				
NINO 3-4		r/rs				
NINO 4		r/rs				
MEI		r/rs				
ONI		r				
NTA		rs				
TNA		r	OUT			4 Meses
NOI		r/rs				
AMM_wind		rs				
PMM_wind		r				
Solar Flux	r					
NINO 1+2	r					
NINO 3	r/rs					
NINO 3-4	r/rs					
NINO 4	r/rs					
MEI	r/rs					
ONI	r/rs					
TSA	r	SET		5 Meses		
Solar Flux	r/rs					
NINO 3	r/rs					
NINO 3-4	r/rs					
NINO 4	rs					
MEI	r					
ONI	r					

1º Grupo	Índices	Tipo	Mês	Antecedência	
	1º Grupo	NAO	rs	AGO	6 Meses
TSA		r			
Solar Flux		r/rs			
NINO 3		r			
NINO 3-4		r			
ONI		r			
TSA		r/rs	JUL	7 Meses	
Solar Flux		r			
NINO 3		r			
NINO 3-4		r			
ONI		r			
NAO		r/rs			JUN
AMO uns		r/rs			
CTI		r			
NOI		r/rs			
AMM_wind	r/rs				
Solar Flux	r/rs				
NINO 1+2	r/rs				
NINO 3	r				
NINO 3-4	r/rs				
MEI	rs				
ONI	r/rs				
AO	r/rs				
AMM_wind	r/rs	MAI	9 Meses		
PMM_wind	rs				
NINO 1+2	rs				
NINO 3	r/rs				
NINO 3-4	r/rs				
MEI	rs				
EPO	r				
Solar Flux	r	ABR	10 Meses		
NINO 1+2	r/rs				
SOI noise	rs	MAR	11 Meses		
Solar Flux	r/rs				
NINO 1+2	rs				

\* Identificação doo tipo de correlação: Pearson (r), Spearman (rs), Wavelet Coerência (C) e Cruzada (X).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul - positivo) e 99% (amarelo). No caso da Wavelets a avaliação é feita no escalograma.

Dentro de cada grupo, procurou-se estabelecer diferentes combinações entre as variáveis, de maneira a ter uma maior quantidade de tentativas acerca de encontra aquela que tenha um melhor desempenho na previsão. A seguir se encontra as tabelas 5.10, 5.11 e 5.12 com os arranjos formados para cada grupo de preditores.



Tabela 5.10 – Combinações do 1º grupo de variáveis para cada intervalo de dados.

100 ANOS			
<b>1ª Arranjo</b>			
NAO	r	NOV	3 Meses
AMO	r/rs		
CTI	r/rs		
<b>2ª Arranjo</b>			
NAO	r	NOV	3 Meses
AMO	r/rs		
<b>3ª Arranjo</b>			
NAO	r	NOV	3 Meses
SOI	r/rs	JUN	
SOI	r/rs	MAR	
<b>4ª Arranjo</b>			
NAO	r	NOV	3 Meses
SOI	r/rs	JUN	
AMO	r/rs	MAR	
<b>5ª Arranjo</b>			
IPO	r	AGO	6 Meses
SOI	r/rs	JUN	
AMO	r/rs	MAR	
<b>6ª Arranjo</b>			
IPO	r	AGO	6 Meses
SOI	r/rs	MAR	
AMO	r/rs		
<b>7ª Arranjo</b>			
SOI	r/rs	JUN	8 Meses
SOI	r/rs	MAR	
AMO	r/rs		
<b>8ª Arranjo</b>			
SOI	r/rs	MAR	11 Meses
AMO	r/rs		
<b>9ª Arranjo</b>			
SOI	r/rs	MAR	11 Meses
AMO	r/rs		

63 ANOS			
<b>1ª Arranjo</b>			
Solar	r/rs	OUT	4 Meses
NOI	r/rs		
TNA	r		
<b>2ª Arranjo</b>			
TSA	r	SET	5 Meses
Solar	r/rs		
<b>3ª Arranjo</b>			
TSA	r	AGO	6 Meses
Solar	r/rs		
<b>4ª Arranjo</b>			
TSA	r/rs	JUL	7 Meses
Solar	r/rs		
<b>5ª Arranjo</b>			
Solar	r/rs	NOV	3 Meses
NOI	r/rs	OUT	
TNA	r		
<b>6ª Arranjo</b>			
NOI	r/rs	OUT	4 Meses
Solar	r/rs	SET	
TSA	r/rs	JUL	
<b>7ª Arranjo</b>			
TSA	r	SET	5 Meses
Solar	r/rs	SET	
NOI	r	JUN	
AMM <sub>wind</sub>	r/rs	MAI	
<b>8ª Arranjo</b>			
Solar	r	JUN	8 Meses
NOI	r		
AMM <sub>wind</sub>	rs		
<b>9ª Arranjo</b>			
AMM <sub>wind</sub>	r/rs	MAI	9 Meses
Solar	r/rs	MA	

61 ANOS			
<b>1ª Arranjo</b>			
NINO 3	r/rs	NOV	3 Meses
AO	r/rs	JUN	
EPO	r	MAI	
NINO 1+2	r/rs	ABR	
<b>2ª Arranjo</b>			
ONI	r/rs	OUT	4 Meses
AO	r/rs	JUN	
EPO	r	MAI	
NINO 1+2	r/rs	ABR	
<b>3ª Arranjo</b>			
NINO 3.4	r/rs	SET	5 Meses
AO	r/rs	JUN	
EPO	r	MAI	
NINO 1+2	r/rs	ABR	
<b>4ª Arranjo</b>			
ONI	r	AGO	6 Meses
AO	r/rs	JUN	
EPO	r	MAI	
NINO 1+2	r/rs	ABR	
<b>5ª Arranjo</b>			
NINO 3.4	r	JUL	7 Meses
AO	r/rs	JUN	
EPO	r	MAI	
NINO 1+2	r/rs	ABR	
<b>6ª Arranjo</b>			
NINO 3	r	JUN	8 Meses
AO	r/rs	JUN	
EPO	r	MAI	
<b>7ª Arranjo</b>			
NINO 3	rs	MAI	9 Meses
AO	r/rs	JUN	
EPO	r	MAI	
<b>8ª Arranjo</b>			
AO	r/rs	JUN	8 Meses
EPO	r	MAI	
NINO 1+2	r/rs	ABR	
<b>9ª Arranjo</b>			
AO	r/rs	JUN	8 Meses
EPO	r	MAI	
NINO 1+2	rs	MA	
<b>10ª Arranjo</b>			
ONI	r	SET	5 Meses
AO	r/rs	JUN	
MEI	rs	MAI	
EPO	r		
<b>11ª Arranjo</b>			
AO	r/rs	JUN	8 Meses
MEI	rs	MAI	
EPO	r		

57 ANOS			
<b>1ª Arranjo</b>			
TP	r/rs	NOV	3 Meses
NTA	r/rs	OUT	
STA	r	JUL	
<b>2ª Arranjo</b>			
TP	r	OUT	4 Meses
NTA	r/rs		
STA	r	JUL	
<b>3ª Arranjo</b>			
TP	r	OUT	4 Meses
NTA	r/rs		
<b>4ª Arranjo</b>			
NTA	r/rs	OUT	4 Meses
STA	r	JUL	
<b>5ª Arranjo</b>			
TP	r/rs	NOV	3 Meses
STA	r	JUL	
ATI	r/rs	JUN	
<b>6ª Arranjo</b>			
TP	r	OUT	4 Meses
STA	r	JUL	
ATI	r/rs	JUN	
<b>7ª Arranjo</b>			
STA	r	JUL	7 Meses
ATI	r/rs	JUN	

\* Identificação doo tipo de correlação: Pearson (r), Spearman (rs), Wavelet Coerência (C) e Cruzada (X).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul - positivo) e 99% (amarelo). No caso da Wavelets a avaliação é feita no escalograma.

Tabela 5.11 – Combinações do 2º grupo de variáveis para cada intervalo de

100 ANOS				
<b>1ª Arranjo</b>				
SOI	X			3 Meses
SOI noise	C	NOV		
AMO smo	C			
<b>2ª Arranjo</b>				
NAO	C	OUT		4 Meses
NAO	X	AGO		
NAO	C	JUL		
<b>3ª Arranjo</b>				
NAO	X	AGO		6 Meses
IPO	C			
NAO	C	JUL		
<b>4ª Arranjo</b>				
NAO	X	AGO		6 Meses
IPO	X	JUN		
SOI signal	X	ABR		
TNI	X	MAR		
<b>5ª Arranjo</b>				
NAO	C	JUL		7 Meses
CTI	C	MAI		
SOI noise	X	ABR		
AMO uns	C	FEV		
<b>6ª Arranjo</b>				
CTI	X	JUN		8 Meses
SOI signal	X	ABR		
TNI	X	MAR		
<b>7ª Arranjo</b>				
CTI	C	MAI		9 Meses
TNI	X	MAR		
AMO uns	C	FEV		
<b>8ª Arranjo</b>				
SOI noise	C	ABR		10 Meses
SOI signal	X			
<b>9ª Arranjo</b>				
AMO uns	X	FEV		12 Meses
IPO	C			

63 ANOS				
<b>1ª Arranjo</b>				
PMM <sub>wind</sub>	XC	NOV		3 Meses
QBO	C	SET		
AMM <sub>wind</sub>	C	AGO		
NOI	C	JUN		
<b>2ª Arranjo</b>				
Solar Flux	XC	NOV		3 Meses
AAO	C			
AMM <sub>wind</sub>	X	OUT		
NOI	XC			
<b>3ª Arranjo</b>				
QBO	C	SET		5 Meses
NOI	C			
TNA	C			
<b>4ª Arranjo</b>				
Solar Flux	XC			6 Meses
QBO	C			
AMM <sub>wind</sub>	C	AGO		
NOI	C			
TSA	X			
<b>5ª Arranjo</b>				
NOI	C	JUN		7 Meses
TSA	X			
<b>6ª Arranjo</b>				
Solar Flux	XC	MAI		8 Meses
PMM <sub>wind</sub>	C			
TNA	X			
<b>7ª Arranjo</b>				
PMM <sub>wind</sub>	C	MAI		8 Meses
TSA	C			
TNA	X			
<b>8ª Arranjo</b>				
Solar Flux	XC	ABR		10 Meses
PMM <sub>wind</sub>	X			
NOI	XC			
<b>9ª Arranjo</b>				
NOI	XC	ABR		10 Meses
TSA	C			
AAO	XC			
<b>10ª Arranjo</b>				
PMM <sub>sst</sub>	X	MAR		11 Meses
AMM <sub>wind</sub>	XC			
TSA	XC			
<b>11ª Arranjo</b>				
Solar Flux	XC	FEV		12 Meses
PMM <sub>sst</sub>	XC			
TSA	XC			
<b>12ª Arranjo</b>				
PMM <sub>wind</sub>	C	MAI		9 Meses
Solar Flux	XC	ABR		
NOI	XC	ABR		
AMM <sub>wind</sub>	XC	MAR		
TSA	XC	FEV		

61 ANOS				
<b>1ª Arranjo</b>				
NINO 1+2	C	NOV		3 Meses
PNA	XC	OUT		
<b>2ª Arranjo</b>				
NINO 1+2	C	OUT		4 Meses
PNA	XC			
<b>3ª Arranjo</b>				
PNA	XC	OUT		4 Meses
NINO 3	C	SET		
EPO	C	AGO		
<b>4ª Arranjo</b>				
MEI	C	AGO		6 Meses
AO	XC			
<b>5ª Arranjo</b>				
PNA	XC	OUT		4 Meses
EPO	C	AGO		
MEI	X	JUL		
<b>6ª Arranjo</b>				
AO	XC	AGO		6 Meses
EPO	XC	JUN		
AO	C			
PNA	X	MAI		
<b>7ª Arranjo</b>				
NINO 1+2	C			8 Meses
EPO	XC	JUN		
PNA	XC			
<b>8ª Arranjo</b>				
ONI	X			8 Meses
PNA	XC	JUN		
EA	C			
<b>9ª Arranjo</b>				
EPO	XC			8 Meses
PNA	XC	JUN		
EA	C			
<b>10ª Arranjo</b>				
AO	C	JUN		8 Meses
PNA	X	MAI		
EA	C			
EPO	X	ABR		
<b>11ª Arranjo</b>				
ONI	XC	MAI		9 Meses
PNA	X			
EA	C			
<b>12ª Arranjo</b>				
NINO 3.4	XC	MAI		9 Meses
PNA	X			
<b>13ª Arranjo</b>				
EPO	X	ABR		10 Meses
PNA	XC	MAR		
<b>14ª Arranjo</b>				
PNA	XC	OUT		4 Meses
PNA	XC	JUN		
PNA	X	MAI		
PNA	XC	MAR		

57 ANOS				
<b>1ª Arranjo</b>				
NTA	X	NOV		3 Meses
TP	C			
<b>2ª Arranjo</b>				
ATI	C	SET		5 Meses
TP	C			
<b>3ª Arranjo</b>				
ATI	C	AGO		6 Meses
TP	XC	JUN		
<b>4ª Arranjo</b>				
STA	C	ABR		10 Meses
TP	XC			
<b>5ª Arranjo</b>				
STA	C	MAR		11 Meses
ATI	XC			
<b>6ª Arranjo</b>				
STA	C	FEV		12 Meses
NTA	X			
<b>7ª Arranjo</b>				
ATI	C	AGO		6 Meses
TP	XC	MAI		
STA	C	ABR		
<b>8ª Arranjo</b>				
TP	XC	ABR		10 Meses
STA	C	MAR		
ATI	X	FEV		

\* Identificação doo tipo de correlação Wavelet Coerência (C) e Cruzada (X).

Tabela 5.12 – Combinações das variáveis do grupo formado com todos os índices.

Todos os Índices 61 ANOS				
<b>1ª Arranjo</b>				
NINO 3-4	r/rs	NOV	3 Meses	
IPO	r/rs	OUT		
Solar Flux	r	SET		
TNA	r	OUT		
TSA	r	SET		
<b>2ª Arranjo</b>				
NINO 3-4	r/rs	NOV	3 Meses	
Solar Flux	r	OUT		
TSA	r	SET		
<b>3ª Arranjo</b>				
Solar Flux	r	OUT	4 Meses	
TSA	r	SET		
NOI	r/rs	JUN		
EPO	r	MAI		
<b>4ª Arranjo</b>				
TSA	r	SET	5 Meses	
NOI	r/rs	JUN		
EPO	r	MAI		
<b>5ª Arranjo</b>				
SOI	r/rs	NOV	3 Meses	
Solar Flux	r/rs	NOV		
NTA	rs			
<b>6ª Arranjo</b>				
SOI signal	r/rs	NOV	3 Meses	
Solar Flux	r/rs	NOV		
NTA	rs			
<b>7ª Arranjo</b>				
CTI	r/rs	NOV	3 Meses	
Solar Flux	r/rs	NOV		
NTA	rs			
<b>8ª Arranjo</b>				
IPO	r/rs	NOV	3 Meses	
Solar Flux	r/rs	NOV		
NTA	rs			
<b>9ª Arranjo</b>				
Solar Flux	r/rs	NOV	3 Meses	
NINO 3-4	r/rs	NOV		
NTA	rs			
<b>10ª Arranjo</b>				
AMM <sub>wind</sub>	rs	OUT	4 Meses	
PMM <sub>wind</sub>	r	OUT		
Solar Flux	r	OUT		
<b>11ª Arranjo</b>				
TNA	r	OUT	4 Meses	
PMM <sub>wind</sub>	r	OUT		
Solar Flux	r	OUT		
NINO 1+2	r	OUT		
<b>12ª Arranjo</b>				
TNA	r	OUT	4 Meses	
Solar Flux	r	OUT		
MEI	r/rs			
<b>13ª Arranjo</b>				
AMM <sub>wind</sub>	rs	OUT	4 Meses	
PMM <sub>wind</sub>	r	OUT		
Solar Flux	r	OUT		
NINO 1+2	r	OUT		
<b>14ª Arranjo</b>				
Solar Flux	r	OUT	4 Meses	
NINO 3-4	r/rs	OUT		
<b>15ª Arranjo</b>				
TNA	r	OUT	4 Meses	
PMM <sub>wind</sub>	r	OUT		
Solar Flux	r	OUT		
<b>16ª Arranjo</b>				
AMM <sub>wind</sub>	rs	OUT	4 Meses	
Solar Flux	r	OUT		
MEI	r/rs			
<b>17ª Arranjo</b>				
AMM <sub>wind</sub>	rs	OUT	4 Meses	
Solar Flux	r	OUT		
ONI	r/rs			
<b>18ª Arranjo</b>				
TSA	r	SET	5 Meses	
Solar Flux	r/rs	SET		
NINO 3	r/rs			
<b>19ª Arranjo</b>				
TSA	r	SET	5 Meses	
Solar Flux	r/rs	SET		
NINO 3-4	r/rs			
<b>20ª Arranjo</b>				
Solar Flux	r/rs	SET	5 Meses	
NINO 3-4	r/rs	SET		
<b>21ª Arranjo</b>				
Solar Flux	r/rs	SET	5 Meses	
NINO 3	r/rs	SET		
<b>22ª Arranjo</b>				
TSA	r	SET	5 Meses	
NINO 3	r/rs	SET		
<b>23ª Arranjo</b>				
TSA	r	SET	5 Meses	
NINO 3-4	r/rs	SET		
<b>24ª Arranjo</b>				
NAO	rs	AGO	6 Meses	
TSA	r	AGO		
Solar Flux	r/rs	AGO		
NINO 3-4	r	AGO		
<b>25ª Arranjo</b>				
NAO	rs	AGO	6 Meses	
TSA	r	AGO		
Solar Flux	r/rs	AGO		
<b>26ª Arranjo</b>				
NAO	rs	AGO	6 Meses	
Solar Flux	r/rs	AGO		
NINO 3-4	r	AGO		
<b>27ª Arranjo</b>				
TSA	r	AGO	6 Meses	
Solar Flux	r/rs	AGO		
<b>28ª Arranjo</b>				
TSA	r	AGO	6 Meses	
NINO 3-4	r	AGO		
<b>29ª Arranjo</b>				
TSA	r/rs	JUL	7 Meses	
Solar Flux	r	JUL		
NINO 3-4	r	JUL		
<b>30ª Arranjo</b>				
TSA	r/rs	JUL	7 Meses	
NINO 3-4	r	JUL		
<b>31ª Arranjo</b>				
Solar Flux	r	JUL	7 Meses	
NINO 3-4	r	JUL		
<b>32ª Arranjo</b>				
NAO	r/rs	JUN	8 Meses	
CTI	r	JUN		
Solar Flux	r/rs	JUN		
<b>33ª Arranjo</b>				
CTI	r	JUN	8 Meses	
AMM <sub>wind</sub>	r/rs	JUN		
Solar Flux	r/rs	JUN		
<b>34ª Arranjo</b>				
AMO uns	r/rs	JUN	8 Meses	
Solar Flux	r/rs	JUN		
NINO 3-4	r/rs	JUN		
<b>35ª Arranjo</b>				
NAO	r/rs	JUN	8 Meses	
NOI	r/rs	JUN		
Solar Flux	r/rs	JUN		
<b>36ª Arranjo</b>				
AMM <sub>wind</sub>	r/rs	JUN	8 Meses	
Solar Flux	r/rs	JUN		
NINO 3	r	JUN		
<b>37ª Arranjo</b>				
Solar Flux	r/rs	JUN	8 Meses	
NINO 3-4	r/rs	JUN		
AO	r/rs	JUN		
<b>38ª Arranjo</b>				
AMO uns	r/rs	JUN	8 Meses	
CTI	r	JUN		
Solar Flux	r/rs	JUN		
AO	r/rs	JUN		
<b>39ª Arranjo</b>				
AMO uns	r/rs	JUN	8 Meses	
Solar Flux	r/rs	JUN		
NINO 3-4	r/rs	JUN		
AO	r/rs	JUN		
<b>40ª Arranjo</b>				
AMM <sub>wind</sub>	r/rs	MAI	9 Meses	
PMM <sub>wind</sub>	rs	MAI		
NINO 1+2	rs	MAI		
<b>41ª Arranjo</b>				
AMM <sub>wind</sub>	r/rs	MAI	9 Meses	
NINO 1+2	rs	MAI		
EPO	r	MAI		
<b>42ª Arranjo</b>				
AMM <sub>wind</sub>	r/rs	MAI	9 Meses	
EPO	r	MAI		
<b>43ª Arranjo</b>				
PMM <sub>wind</sub>	rs	MAI	9 Meses	
NINO 1+2	rs	MAI		
<b>44ª Arranjo</b>				
Solar Flux	r	ABR	10 Meses	
NINO 1+2	r/rs	ABR		
<b>45ª Arranjo</b>				
SOI noise	rs	MAR	11 Meses	
Solar Flux	r/rs	MAR		
NINO 1+2	rs	MAR		
<b>46ª Arranjo</b>				
Solar Flux	r	ABR	10 Meses	
NINO 1+2	r/rs	ABR		
SOI noise	rs	MAR		
<b>47ª Arranjo</b>				
NINO 1+2	r/rs	ABR	10 Meses	
SOI noise	rs	MAR		
Solar Flux	r/rs	MAR		
<b>48ª Arranjo</b>				
Solar Flux	r	ABR	10 Meses	
SOI noise	rs	MAR		
NINO 1+2	rs	MAR		

\* Identificação do tipo de correlação: Pearson (r), Spearman (rs).

\*\* As cores identificam os níveis de significância de 90% (verde), 95% (vermelho – negativo e azul - positivo) e 99% (amarelo). No caso da Wavelets a avaliação é feita no escalograma.

### 5.3 PREVISÃO

Diante de cada grupo de variáveis preditoras e de acordo com as técnicas de seleção das mesmas, foram simulados para cada arranjo nesses grupos diferentes topologias de redes neurais, alterando-se a distribuição dos conjuntos dados e a quantidade de neurônios em sua camada oculta. Para avaliar o grau de estabilidade de cada topologia testada em cada um dos arranjos, foram realizadas 10 simulações para cada topologia, onde a média e o desvio-padrão das medidas de desempenho foram calculados para análise da oscilação desses resultados, visto que, as RNAs apresentam valores distintos a cada treinamento.

Considerando a gama de resultados encontrados, foram selecionados para cada arranjo, apenas a topologia que obteve globalmente as melhores medidas de desempenho, e também de forma mais homogênea para os três conjuntos de dados (treinamento, validação e verificação), afim de condicionar maior segurança ao modelo.

Os resultados a seguir foram organizados através de gráficos das médias e desvios das medidas de desempenho dos conjuntos de dados de cada arranjo. Também é apresentada uma tabela com os valores das medidas de desempenho, sintetizando a informações de cada topologia escolhida para o referido arranjo, contendo a antecedência de previsão, a distribuição entre os conjuntos de dados, a quantidade de variáveis de entrada ( $N$ ) e o número de neurônios da camada intermediária da RNA em relação a quantidade de entrada ( $N + i$ ).

Dentre estes resultados selecionados para os arranjos, houve a escolha daquele considerado como o melhor modelo do grupo. Desse modelo, foi colocado em destaque o melhor ajuste da previsão entre suas 10 simulações, representando-a através do gráfico de dispersão e de série temporal dos valores observados e calculados da precipitação.

### 5.3.1 RNA com 100 anos de dados

#### 5.3.1.1 Resultados referentes ao 1º grupo de variáveis predictoras

Na série de gráficos apresentados na Figura 5.8, observa-se que os modelos mais homogêneos em relação ao coeficiente de Pearson são os do 3º, 4º, 7º e 8º arranjo, como os valores mais elevados para o 3º e 7º arranjo. O coeficiente NS, apresenta-se com valores baixos para todos os modelos ( $NS < 0,3$ ), além da alta oscilação entre os conjuntos na maioria dos arranjos e do elevado desvio. Apesar disso, o NS acaba se apresentando também mais homogêneo e de certa forma mais elevado para o 3º e 7º arranjo, como maiores desvios para o 3º arranjo. Os valores de MAE demonstram pouca oscilação, com valores mais baixos para os conjuntos de verificação, não havendo grandes diferenças entre o 3º e 7º arranjos.

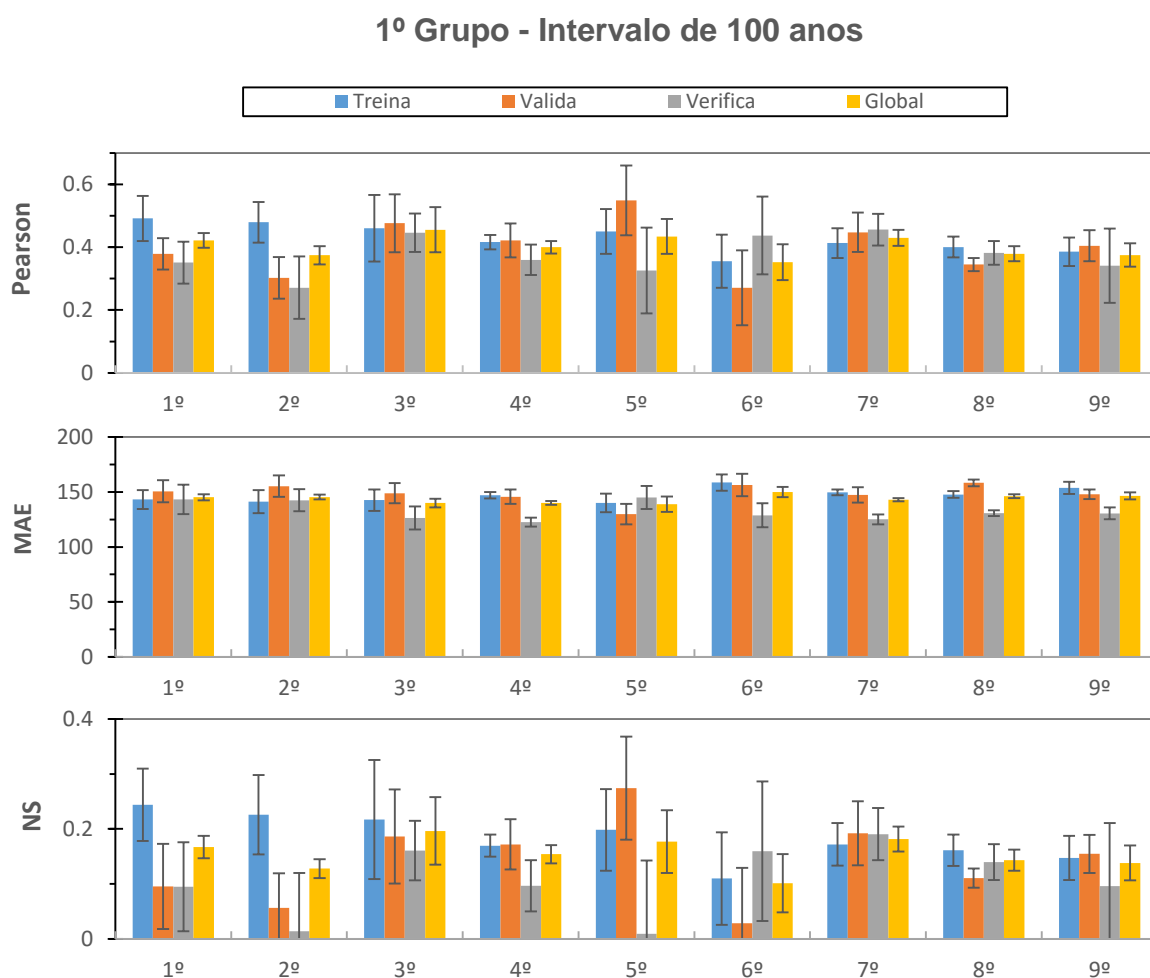


Figura 5.8 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 1º grupo de dados de 100 anos para o conjunto de treinamento, testes, validação e global.

Tabela 5.13 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 100 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	46-27-27	3	-1	0.49	0.38	0.35	0.42	143	151	143	145	0.24	0.10	0.09	0.17
2º	3 Meses	46-27-27	2	2	0.48	0.30	0.27	0.37	141	155	142	145	0.23	0.06	0.01	0.13
3º	3 Meses	46-27-27	4	1	0.46	0.48	0.45	0.46	143	149	126	140	0.22	0.19	0.16	0.20
4º	3 Meses	46-27-27	3	-1	0.42	0.42	0.36	0.40	147	146	123	140	0.17	0.17	0.10	0.15
5º	6 Meses	60-20-20	3	3	0.45	0.55	0.33	0.43	140	130	145	139	0.20	0.27	0.01	0.18
6º	6 Meses	46-27-27	3	5	0.36	0.27	0.44	0.35	159	156	129	150	0.11	0.03	0.16	0.10
7º	8 Meses	50-25-25	3	-1	0.41	0.45	0.46	0.43	150	147	125	143	0.17	0.19	0.19	0.18
8º	11 Meses	50-25-25	2	0	0.40	0.34	0.38	0.38	148	158	131	146	0.16	0.11	0.14	0.14
9º	11 Meses	50-25-25	2	1	0.39	0.40	0.34	0.38	154	148	130	146	0.15	0.15	0.10	0.14

Tabela 5.14 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 100 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	46-27-27	3	-1	0.07	0.05	0.07	0.02	9	10	13	3	0.07	0.08	0.08	0.02
2º	3 Meses	46-27-27	2	2	0.06	0.07	0.10	0.03	10	10	10	2	0.07	0.06	0.11	0.02
3º	3 Meses	46-27-27	4	1	0.11	0.09	0.06	0.07	10	9	10	4	0.11	0.09	0.05	0.06
4º	3 Meses	46-27-27	3	-1	0.02	0.05	0.05	0.02	3	6	4	2	0.02	0.05	0.05	0.02
5º	6 Meses	60-20-20	3	3	0.07	0.11	0.14	0.06	8	9	11	7	0.07	0.09	0.13	0.06
6º	6 Meses	46-27-27	3	5	0.08	0.12	0.12	0.06	7	10	11	5	0.08	0.10	0.13	0.05
7º	8 Meses	50-25-25	3	-1	0.05	0.06	0.05	0.03	3	7	5	2	0.04	0.06	0.05	0.02
8º	11 Meses	50-25-25	2	0	0.03	0.02	0.04	0.02	3	3	3	2	0.03	0.02	0.03	0.02
9º	11 Meses	50-25-25	2	1	0.05	0.05	0.12	0.04	6	4	5	3	0.04	0.03	0.11	0.03

Na tabela 5.13, com o detalhamento da topologia, podemos perceber que houve uma maior concentração de modelos na distribuição 46-27-27 (treinamento-validação-verificação), seguido da distribuição 50-25-25. Pode-se perceber também, que não houve a presença das distribuições 66-17-17 e 70-15-15. Apesar de uma maior variação, podemos destacar que boa parte dos modelos selecionados para cada arranjo apresenta uma topologia com igual ou menos neurônios na camada intermediária em relação a quantidade de variáveis de entrada, ou seja, não sendo preciso aumentar significativamente o número de neurônios na camada intermediária para se haver um melhor ajuste desses modelos.

Destacados o 3º e 7º arranjo como aqueles com melhores desempenho dentre os demais, percebe-se a priori que há uma diferença considerável entre a antecedência de previsão, onde essa antecedência é provocada principalmente pelas variáveis de entrada. Em quase todos os arranjos há a presença de pelo menos dois entre índices  $SOI_{noise}$  de junho,  $SOI_{noise}$  de março e  $AMO_{smo}$  de março. O 7º arranjo é formado somente por estes 3 índices, enquanto que o 3º arranjo, além desses índices, há a adição do NAO de novembro. A adição deste índice acaba provocando um razoável aumento no desempenho (Pearson e NS) do conjunto de treinamento, aumentando consequentemente o desempenho global. Em troca, provoca um leve aumento nos desvios nos conjuntos de treinamento e também no de validação, o que acaba aumentando um pouco o desvio global (Tabela 5.14). Apesar de aumentar levemente os desvios e a antecedência, este arranjo proporciona a possibilidade de ajustes melhores em suas simulações, justamente por conseguir manter mais constante o desempenho do conjunto de verificação e possibilitar o aumento dos outros conjuntos, caso que não ocorre no 7º arranjo. De acordo com isso, o melhor ajuste deste arranjo é apresentado nos gráficos da figura 5.9 e 5.10.

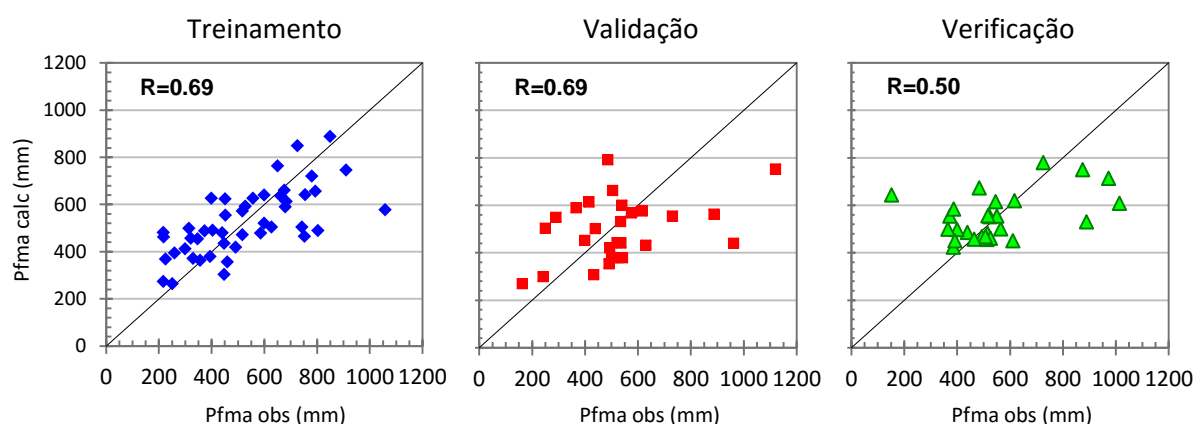


Figura 5.9 – Melhor ajuste entre as simulações do 3º arranjo para os conjuntos de treinamento, validação e verificação dos dados.

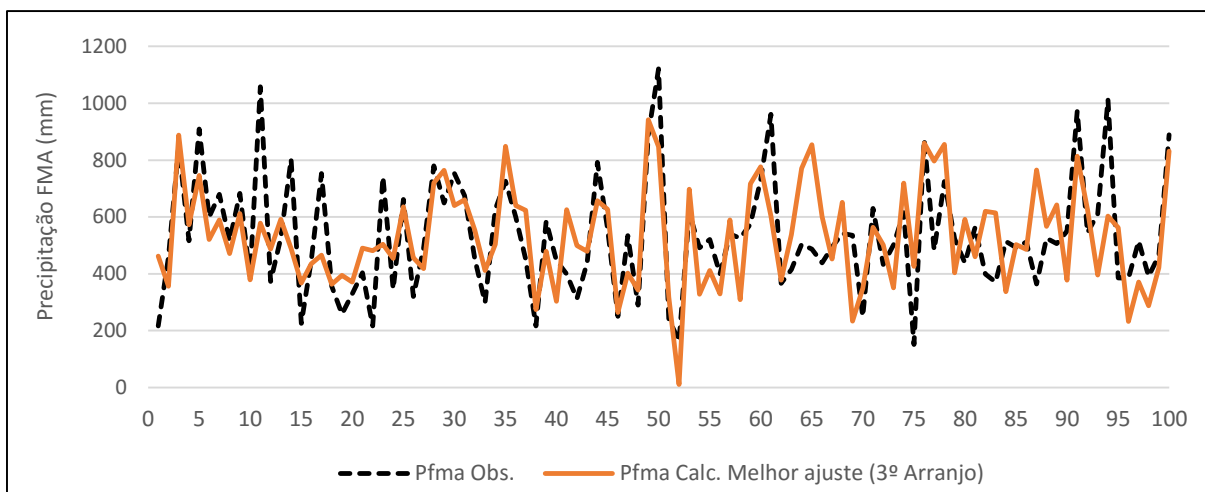


Figura 5.10 – Série temporal da precipitação observada e do melhor ajuste de previsão do 3º Arranjo do 1º grupo de variáveis predictoras de 100 anos.

### 5.3.1.2 Resultados referentes ao 2º grupo de variáveis predictoras

Na série de gráficos apresentados na Figura 5.11, observa-se a priori que os valores dos coeficientes de Pearson para as melhores topologias dos arranjos selecionado pelas variáveis desse grupo, são consideravelmente menores que o 1º grupo. Além da alta oscilação entre os conjuntos de dado, há um alto desvio em praticamente todos os modelos. Percebe-se que os valores do conjunto de validação se sobressaem em relação aos outros. Como não há a presença de um modelo mais homogêneo entre os conjuntos que também ofereça maiores valores em suas medidas de desempenho, temos o 1º e o 9º arranjo que possibilitam um maior desempenho global para a previsão, e diferentemente dos outros, conseguem manter todos seus conjuntos nos patamares mais altos.

A média dos valores de NS para esse grupo apresenta-se ainda mais baixos para todos os modelos com 5 arranjos apresentando valores negativos para os conjuntos de verificação, além dos altos desvios. Os valores de NS negativos indicam que a variância residual, entre os dados observados e calculados, é maior que a resíduos entre os dados calculados e a média dos dados observados. Ou seja, indica que a maioria dos dados estimados estão muito próximo da média dos dados observados (diminuindo o denominador da equação), ocasionando altos resíduos na diferença entre cada dado estimado com o seu dado observado (aumentando o numerador da equação). Graficamente, isso também reflete o quanto os dados se distanciam da reta 1:1. Como essa medida de desempenho é bastante sensível a esse tipo situação, a presença de alguns dados fora dessa reta, e com valores próximos a média observada, diminuem consideravelmente essa



medida, podendo atingir valores negativos. Os valores mais altos nesta medida são novamente dos conjuntos de validação para maioria dos modelos, o que pode indicar um melhor ajuste a reta 1:1. Os valores mais homogêneos são os do 2º e 4º arranjo, no entanto são globalmente mais baixos que os demais. Analisando o melhor desempenho global, com os valores menos críticos do conjunto de verificação, teríamos o 1º e o 9º arranjo com os melhores modelos. Os valores de MAE se demonstram com um pouco mais oscilação em relação ao 1º grupo, tanto entre os conjuntos, quanto para seus desvios, com um patamar mais elevado para o conjunto de treinamento em quase todos os arranjos.

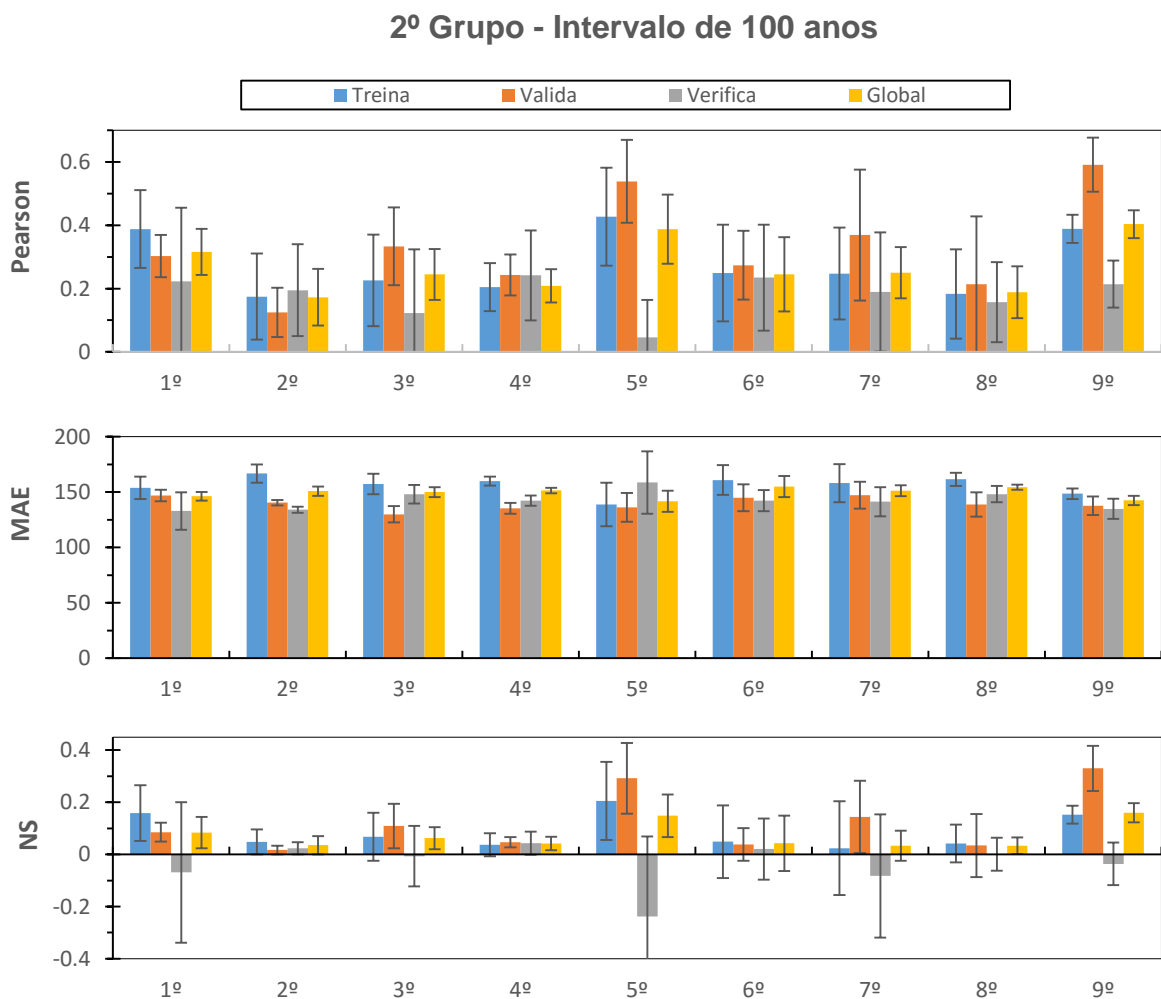


Figura 5.11– Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 2º grupo de dados de 100 anos para o conjunto de treinamento, testes, validação e global.

Tabela 5.15 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 100 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	46-27-27	3	3	0.39	0.30	0.22	0.32	154	147	133	146	0.16	0.09	-0.07	0.08
2º	4 Meses	46-27-27	3	-1	0.17	0.12	0.19	0.17	167	140	134	151	0.05	0.02	0.02	0.04
3º	6 Meses	60-20-20	3	-1	0.23	0.33	0.12	0.24	157	130	148	150	0.07	0.11	-0.01	0.06
4º	6 Meses	60-20-20	4	-2	0.20	0.24	0.24	0.21	160	135	142	151	0.04	0.05	0.04	0.04
5º	7 Meses	66-17-17	4	2	0.43	0.54	0.05	0.39	139	136	159	142	0.21	0.29	-0.24	0.15
6º	8 Meses	66-17-17	4	3	0.25	0.27	0.23	0.25	161	145	142	155	0.05	0.04	0.02	0.04
7º	9 Meses	50-25-25	3	2	0.25	0.37	0.19	0.25	158	147	141	151	0.02	0.14	-0.08	0.03
8º	10 Meses	60-20-20	3	4	0.18	0.21	0.16	0.19	162	139	148	154	0.04	0.03	0.00	0.03
9º	12 Meses	50-25-25	2	2	0.39	0.59	0.21	0.40	149	138	135	142	0.15	0.33	-0.04	0.16

Tabela 5.16 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 100 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	46-27-27	3	3	0.12	0.07	0.23	0.07	10	5	17	4	0.11	0.04	0.27	0.06
2º	4 Meses	46-27-27	3	-1	0.14	0.08	0.14	0.09	8	2	3	4	0.05	0.02	0.02	0.04
3º	6 Meses	60-20-20	3	-1	0.14	0.12	0.20	0.08	9	7	8	4	0.09	0.09	0.12	0.04
4º	6 Meses	60-20-20	4	-2	0.08	0.06	0.14	0.05	4	5	5	3	0.04	0.02	0.04	0.03
5º	7 Meses	66-17-17	4	2	0.15	0.13	0.12	0.11	20	13	28	10	0.15	0.14	0.31	0.08
6º	8 Meses	66-17-17	4	3	0.15	0.11	0.17	0.12	13	12	9	10	0.14	0.06	0.12	0.11
7º	9 Meses	50-25-25	3	2	0.15	0.21	0.19	0.08	17	12	13	5	0.18	0.14	0.24	0.06
8º	10 Meses	60-20-20	3	4	0.14	0.21	0.13	0.08	6	11	7	2	0.07	0.12	0.06	0.03
9º	12 Meses	50-25-25	2	2	0.04	0.09	0.07	0.04	5	9	9	4	0.03	0.09	0.08	0.04

Na tabela 5.15, com o detalhamento da topologia, podemos notar que houve uma maior variação das distribuições dos dados nos melhores modelos de cada arranjo. Diferentemente do 1º grupo de variáveis de 100 anos, nesse 2º grupo houve um aumento de modelos com distribuições com mais dados no conjunto de treinamento (com 60 e 66 dados). Mas, igual ao 1º grupo, não houve a presença da distribuição 70-15-15. Também houve uma maior variação no número de neurônios na camada intermediária, onde percebe-se que a maioria dos modelos selecionados para cada arranjo apresenta uma topologia com mais neurônios na camada oculta em relação a quantidade de variáveis de entrada, ou seja, nesse grupo foi preciso haver um aumento do número de neurônios na camada intermediária para se obter um melhor ajuste do modelo.

Destacados o 1º e 9º arranjo como aqueles com melhores desempenho dentre os demais, novamente percebe-se a priori que há uma diferença considerável entre a antecedência de previsão, além da quantidade de variáveis de entrada. Outra consideração é na uniformidade mensal das variáveis, ou seja, no 1º arranjo todos os índices (SOI, SOI<sub>noise</sub> e AMO<sub>smo</sub>) são do mês de novembro, enquanto que no 9º arranjo todos são do mês de fevereiro (IPO e AMO<sub>uns</sub>). Uma característica coincidente entre os dois arranjos é nas regiões de seus índices, no qual o IPO (conhecido como Tripolo Interdecadal do Pacífico) mede as oscilações de temperatura no norte, centro e sul do Pacífico, onde as regiões de oscilação do centro e sul ao qual o IPO contabiliza são muito próximas das mesmas regiões do índice SOI, que apesar de medir a pressão, esta é diretamente influenciada pela temperatura da superfície. Outra relação ao está no uso do índice AMO nesses arranjos, onde apesar de atuarem na mesma região do Atlântico, um mede a oscilação do sinal original (AMO<sub>uns</sub>), enquanto que o outro mede essa oscilação de forma suavizada (AMO<sub>smo</sub>).

Nesse contexto, percebe-se que as características captadas pelo 9º arranjo aumentam as medidas de desempenho consideravelmente no conjunto de validação, mantendo as demais sem alteração, conseqüentemente aumentando seu desempenho também de forma global. Esse arranjo também possibilita menores desvios em todas as medidas para todos os conjuntos (Tabela 5.16).

Analisando as simulações de ambos os arranjos, o 9º arranjo proporciona a possibilidade de melhores ajustes em suas simulações. Ainda assim, devido aos valores de NS muito baixo para todos os arranjos, em todas as simulações do 9º arranjo a maioria dos dados se concentraram muito próximo a média da precipitação observada (530 mm). Como já foi destacado, apenas o conjunto de validação proporcionou melhor ajuste em relação aos outros. De acordo com isso, o melhor ajuste deste arranjo é apresentado nos gráficos da figura 5.12 e 5.13.

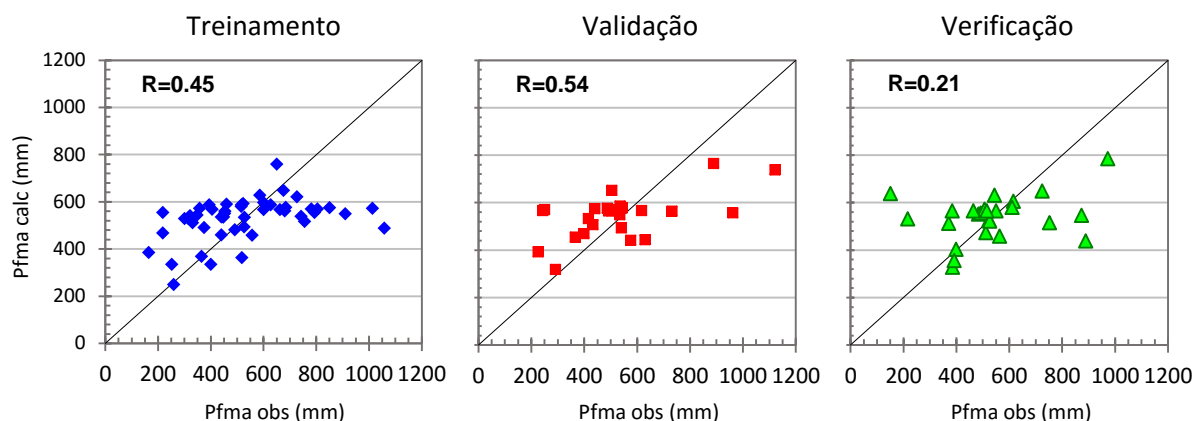


Figura 5.12 – Melhor ajuste entre as simulações do 9º arranjo para os conjuntos de treinamento, validação e verificação dos dados.

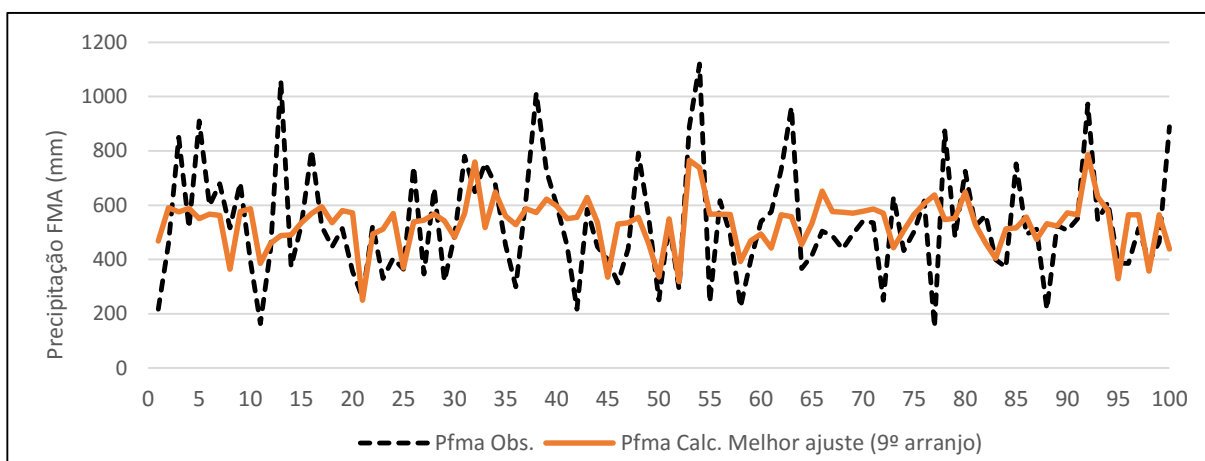


Figura 5.13– Serie temporal da precipitação observada e do melhor ajuste de previsão do 9º Arranjo do 2º grupo de variáveis predictoras de 100 anos.

### 5.3.2 RNA com 63 anos de dados

#### 5.3.2.1 Resultados referentes ao 1º grupo de variáveis predictoras

Na série de gráficos apresentados na Figura 5.14, observa-se que não há modelos tão homogêneos entre seus conjuntos de dados em relação ao coeficiente de Pearson. Mas os valores dessa medida de desempenho em seus conjuntos são significativamente mais elevados em relação ao grupo de variáveis com 100 anos. Podemos destacar o 2º e 7º arranjo como aqueles obtiveram tanto um alto desempenho global, quanto também conseguiram manter mais

homogêneos os seus conjuntos de dados, apesar do alto desvio no conjunto de verificação do 2º arranjo.

O coeficiente NS, apresenta-se bastante oscilante entre os conjuntos em todos os arranjos, e média sem valores negativos, mas com altos desvio que evidenciam alguns arranjos com simulações que obtiveram valores negativos em seus conjuntos. Podemos observar que o 7º arranjo, apesar do alto desvio no conjunto da verificação, acaba se demonstrando mais homogêneo que os demais com um valor global entre os maiores do grupo (NS=0,56).

Os valores de MAE demonstram mais oscilação que o grupo de variáveis de 100 anos, com patamares significativamente menores, abaixo dos 140mm, com valores mais altos para os conjuntos de verificação. O 2º e 7º arranjo apresentam os valores de seus conjuntos de verificação praticamente nos mesmos patamares que os demais, mas com valores globais baixos (Tabela 5.17).

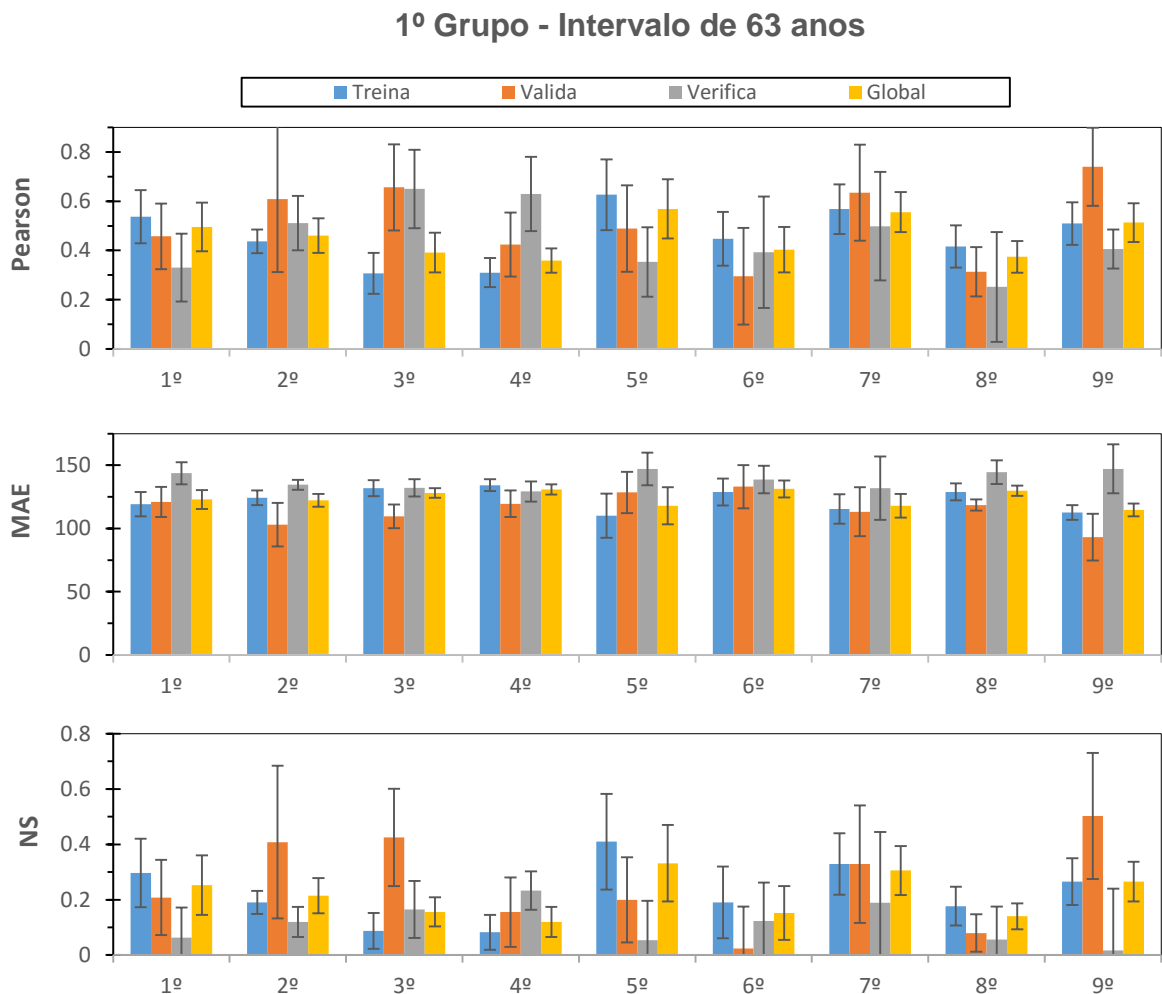


Figura 5.14 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 1º grupo de dados de 63 anos para o conjunto de treinamento, testes, validação e global.

Tabela 5.17 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 63 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	4 Meses	45-9-9	3	1	0.54	0.46	0.33	0.50	119	121	144	123	0.30	0.21	0.06	0.25
2º	5 Meses	41-11-11	2	0	0.44	0.61	0.51	0.46	124	103	135	122	0.19	0.41	0.12	0.21
3º	6 Meses	41-11-11	2	0	0.31	0.66	0.65	0.39	132	110	132	128	0.09	0.43	0.16	0.16
4º	7 Meses	41-11-11	2	3	0.31	0.42	0.63	0.36	134	120	129	131	0.08	0.16	0.23	0.12
5º	3 Meses	45-9-9	3	5	0.63	0.49	0.35	0.57	110	129	147	118	0.41	0.20	0.05	0.33
6º	4 Meses	41-11-11	3	1	0.45	0.29	0.39	0.40	129	133	139	131	0.19	0.02	0.12	0.15
7º	5 Meses	41-11-11	4	3	0.57	0.64	0.50	0.56	115	113	132	118	0.33	0.33	0.19	0.31
8º	8 Meses	41-11-11	3	-1	0.42	0.31	0.25	0.37	129	119	144	130	0.18	0.08	0.06	0.14
9º	9 Meses	45-9-9	2	1	0.51	0.74	0.41	0.51	113	93	147	115	0.27	0.50	0.02	0.27

Tabela 5.18 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 63 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	4 Meses	45-9-9	3	1	0.11	0.13	0.14	0.10	10	12	9	8	0.12	0.14	0.11	0.11
2º	5 Meses	41-11-11	2	0	0.05	0.30	0.11	0.07	6	17	4	5	0.04	0.28	0.05	0.06
3º	6 Meses	41-11-11	2	0	0.08	0.17	0.16	0.08	6	9	7	4	0.06	0.18	0.10	0.05
4º	7 Meses	41-11-11	2	3	0.06	0.13	0.15	0.05	5	11	8	4	0.06	0.13	0.07	0.05
5º	3 Meses	45-9-9	3	5	0.14	0.18	0.14	0.12	17	16	13	15	0.17	0.15	0.14	0.14
6º	4 Meses	41-11-11	3	1	0.11	0.20	0.23	0.09	11	17	11	7	0.13	0.15	0.14	0.10
7º	5 Meses	41-11-11	4	3	0.10	0.20	0.22	0.08	12	19	25	9	0.11	0.21	0.26	0.09
8º	8 Meses	41-11-11	3	-1	0.09	0.10	0.22	0.06	7	5	9	4	0.07	0.07	0.12	0.05
9º	9 Meses	45-9-9	2	1	0.09	0.16	0.08	0.08	6	19	19	5	0.08	0.23	0.22	0.07

Na tabela 5.17, com o detalhamento da topologia, podemos perceber que houve uma grande concentração de modelos na distribuição 41-11-11 (6 situações), seguido da distribuição 45-9-9 e sem a presença da distribuição 35-14-14. Apesar de uma maior variação, podemos destacar que boa parte dos modelos selecionados para cada arranjo apresenta uma topologia com mais neurônios na camada oculta em relação a quantidade de variáveis de entrada, ou seja, nesse grupo há um aumento do número de neurônios na camada intermediária para se obter um melhor ajuste do modelo. Um detalhe a ser destacado na tabela 5.18, se deve a presença de altos desvios para os arranjos desse grupo de variáveis preditoras.

Destacados o 2º e 7º arranjo como aqueles com melhores desempenho dentre os demais, nota-se que os dois mantem a mesma antecedência de previsão, com a diferença da quantidade de variáveis de entrada. Em todas as combinações há a presença do índice do Fluxo Solar, onde quase sempre está acompanhado do índice TSA. O 2º arranjo faz apenas uma combinação entre esses dois índices para o mês de setembro, enquanto que o 7º arranjo, além desses índices adiciona o NOI de junho e  $AMM_{wind}$  de maio. Ao que se demonstra, a adição desses índices no 7º arranjo aumenta o conjunto de treinamento, dando uma maior homogeneidade ao modelo com um valor global maior. Essa afirmação se torna mais clara, se observarmos também o 3º e 4º arranjo, pois estes possuem apenas mesmos os índices do 2º arranjo alterando-se os meses para agosto e julho respectivamente, notando-se que seus conjuntos de treinamento obtêm medidas de desempenho menores. Apesar de no 7º arranjo haver melhores medidas de desempenho de forma mais homogênea, este também oferece maiores desvios em suas simulações.

Analisando as simulações de ambos os arranjos, o 7º arranjo proporciona a possibilidade de melhores ajustes em suas simulações. Deve-se destacar que as simulações dos conjuntos de validação de ambos os arranjos, conseguem estimar muito bem um dos pontos de maiores picos de precipitação observada, elevando dessa forma seus coeficientes de Pearson. De acordo com essas considerações, o melhor ajuste do 7º arranjo é apresentado nos gráficos da figura 5.15 e 5.16.

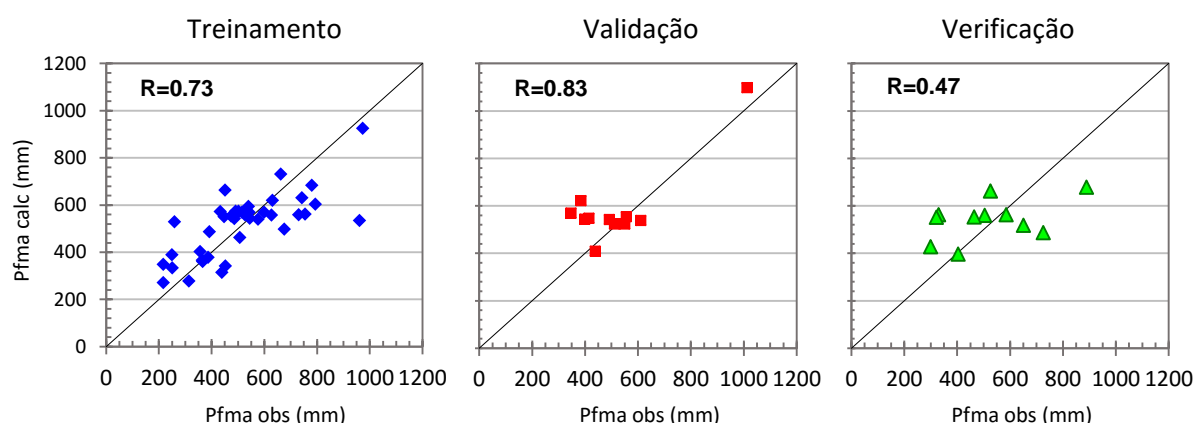


Figura 5.15 – Melhor ajuste entre as simulações do 7º arranjo para os conjuntos de treinamento, validação e verificação dos dados.

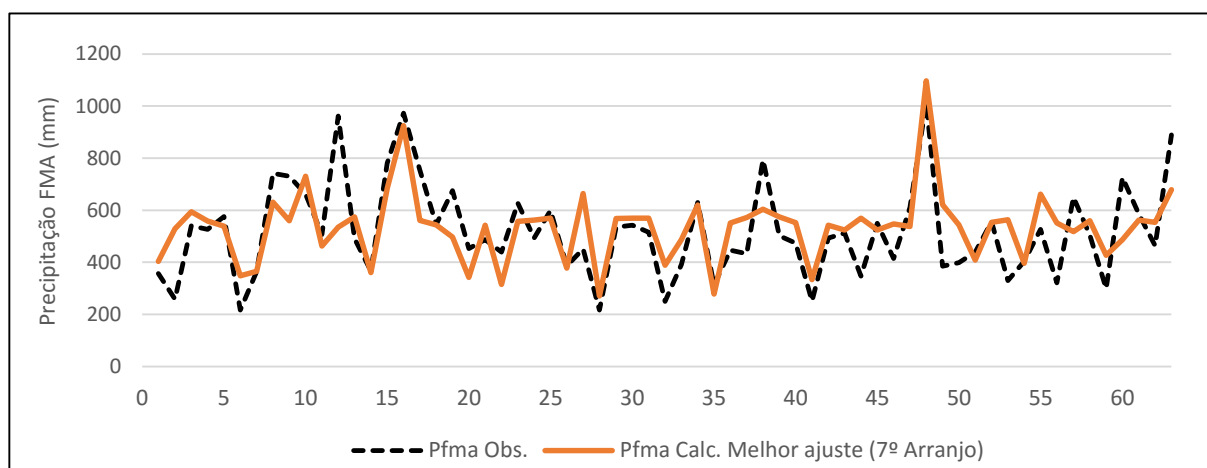


Figura 5.16 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 7º Arranjo do 1º grupo de variáveis predictoras de 63 anos.

### 5.3.2.2 Resultados referentes ao 2º grupo de variáveis predictoras

Na série de gráficos apresentados na Figura 5.17, observa-se que em relação ao coeficiente de Pearson há uma alta oscilação dos conjuntos entres os modelos escolhidos para os arranjos, com valores globais um pouco menores em relação ao 1º grupo de variáveis de 63 anos. Também nota-se que houve um grande aumento nos desvios para os arranjos desse grupo. Os modelos que conseguiram obter o melhor desempenho global mantendo também os níveis de seus conjuntos também em patamares mais elevados, foram o 3º e 4º arranjo.

Em média os valores de NS para esse grupo apresenta-se com valores mais baixos em relação ao grupo anterior, inclusive com 4 arranjos apresentando valores negativos para os conjuntos



de verificação, além dos altos desvios presentes na maioria dos modelos. Mantendo a linha de raciocínio do parágrafo anterior, os modelos que demonstraram melhor desempenho nesse índice foram o 3º e 4º arranjo.

Os valores de MAE demonstram praticamente a mesma oscilação que o 1º grupo de variáveis de 63 anos, com valores mais altos para os conjuntos de verificação. O 3º e 4º arranjo apresentam os valores de seus conjuntos de verificação praticamente nos mesmos patamares que os demais, com destaque para o alto desvio no 3º arranjo, mas ao todo ambos tiveram os valores dos outros conjuntos mais baixos (Tabela 5.20).

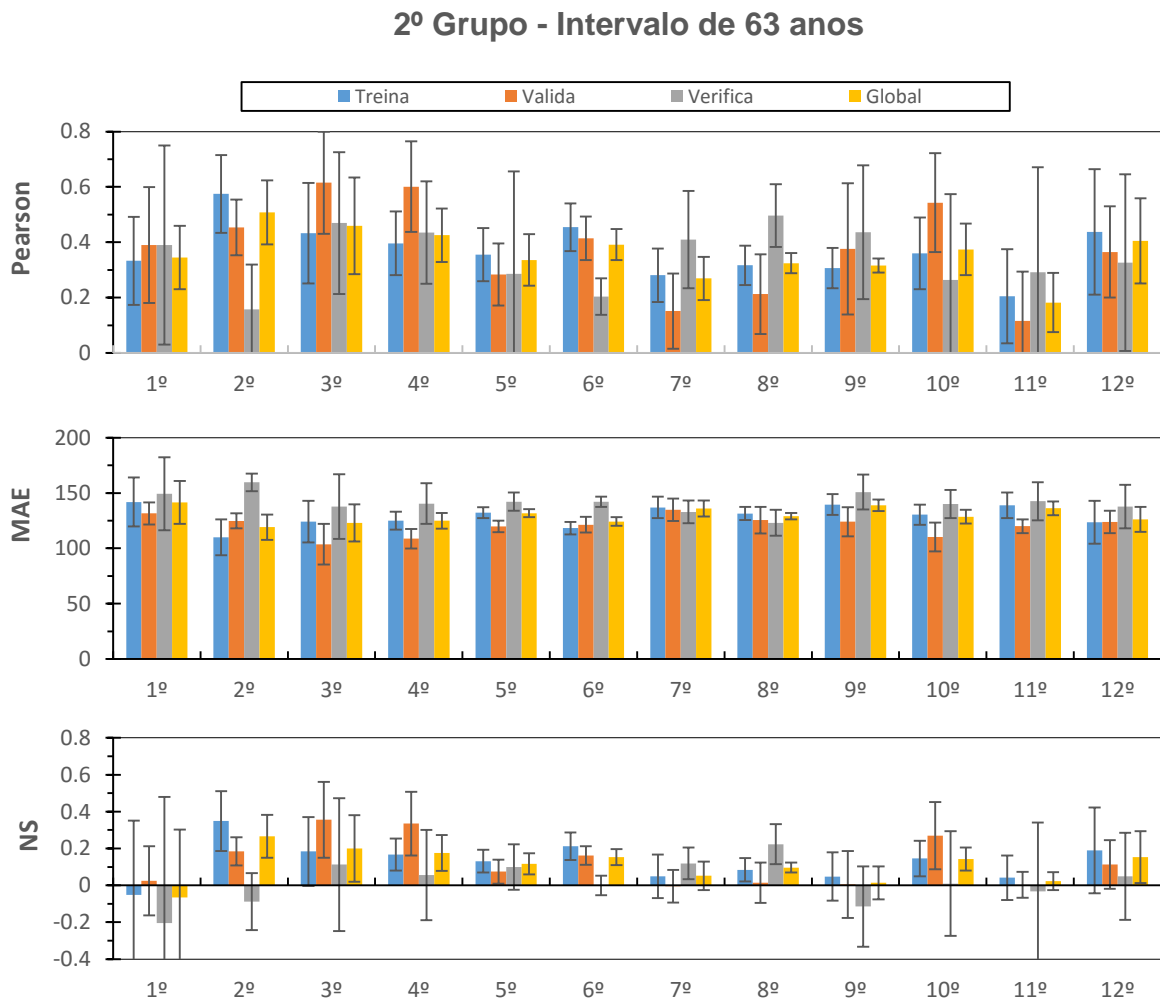


Figura 5.17 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 2º grupo de dados de 63 anos para o conjunto de treinamento, testes, validação e global.

Tabela 5.19 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 63 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	41-11-11	4	5	0.33	0.39	0.39	0.34	142	132	149	141	-0.05	0.02	-0.20	-0.07
2º	3 Meses	45-9-9	4	-1	0.57	0.45	0.16	0.51	110	125	160	119	0.35	0.18	-0.09	0.27
3º	5 Meses	41-11-11	3	3	0.43	0.61	0.47	0.46	124	104	138	123	0.18	0.36	0.11	0.20
4º	6 Meses	41-11-11	5	-3	0.40	0.60	0.44	0.43	125	109	140	125	0.17	0.33	0.06	0.18
5º	7 Meses	41-11-11	2	0	0.35	0.28	0.29	0.34	132	120	142	132	0.13	0.07	0.10	0.12
6º	8 Meses	35-14-14	3	0	0.45	0.41	0.20	0.39	118	121	142	124	0.21	0.16	0.00	0.15
7º	8 Meses	41-11-11	3	4	0.28	0.15	0.41	0.27	137	135	133	136	0.05	-0.01	0.12	0.05
8º	10 Meses	41-11-11	3	4	0.32	0.21	0.50	0.32	132	125	123	129	0.08	0.01	0.22	0.10
9º	10 Meses	45-9-9	3	5	0.31	0.38	0.44	0.32	140	124	151	139	0.05	0.00	-0.11	0.01
10º	11 Meses	41-11-11	3	0	0.36	0.54	0.26	0.37	130	110	140	129	0.15	0.27	0.01	0.14
11º	12 Meses	41-11-11	3	-1	0.20	0.12	0.29	0.18	139	120	143	136	0.04	0.00	-0.03	0.02
12º	9 Meses	41-11-11	5	2	0.44	0.36	0.33	0.40	124	124	138	126	0.19	0.11	0.05	0.15

Tabela 5.20 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 63 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	41-11-11	4	5	0.16	0.21	0.36	0.11	22	10	33	19	0.40	0.19	0.68	0.37
2º	3 Meses	45-9-9	4	-1	0.14	0.10	0.16	0.12	16	7	8	11	0.16	0.08	0.15	0.12
3º	5 Meses	41-11-11	3	3	0.18	0.18	0.26	0.17	19	18	29	17	0.19	0.21	0.36	0.18
4º	6 Meses	41-11-11	5	-3	0.12	0.16	0.19	0.10	8	9	18	7	0.09	0.17	0.24	0.10
5º	7 Meses	41-11-11	2	0	0.10	0.11	0.37	0.09	5	5	8	4	0.06	0.06	0.12	0.06
6º	8 Meses	35-14-14	3	0	0.09	0.08	0.07	0.06	6	7	5	4	0.07	0.05	0.05	0.04
7º	8 Meses	41-11-11	3	4	0.10	0.14	0.18	0.08	10	10	10	7	0.12	0.09	0.09	0.08
8º	10 Meses	41-11-11	3	4	0.07	0.14	0.11	0.04	6	12	12	3	0.06	0.11	0.11	0.03
9º	10 Meses	45-9-9	3	5	0.07	0.24	0.24	0.03	9	13	16	5	0.13	0.18	0.22	0.09
10º	11 Meses	41-11-11	3	0	0.13	0.18	0.31	0.09	9	13	13	6	0.10	0.18	0.28	0.06
11º	12 Meses	41-11-11	3	-1	0.17	0.18	0.38	0.11	12	6	17	6	0.12	0.07	0.37	0.05
12º	9 Meses	41-11-11	5	2	0.23	0.16	0.32	0.15	19	10	20	11	0.23	0.13	0.24	0.14

Na tabela 5.19, com o detalhamento da topologia, podemos perceber que houve novamente uma grande concentração de modelos na distribuição 41-11-11 (9 situações), com 2 presenças da distribuição 45-9-9 e apenas uma da distribuição 35-14-14. Já na camada intermediária, novamente houve uma maior variação no número de neurônios, com uma maior frequência de topologias com mais neurônios na camada oculta em relação a quantidade de variáveis de entrada, com 4 situações usando de 4 ou 5 neurônios a mais para um melhor ajuste no modelo.

Destacados o 3º e 4º arranjo como aqueles com melhores desempenho dentre os demais, nota-se que a antecedência de previsão é de 5 e 6 meses respectivamente, e a diferença da quantidade de variáveis de entrada é de duas a mais para o 4º arranjo. Não há diferenças significativas das medidas de desempenho entres esses arranjos, apenas uma relativa superioridade nas medidas do 3º arranjo, mas também com maiores desvios. Ambos têm em comum os índices NOI e QBO ambos de setembro para o 3º arranjo e agosto para o 4º arranjo, além dos índices TNA de setembro no 3º arranjo e do TSA de agosto no 4º arranjo, ao qual medem a TSM acima e abaixo da região equatorial no Atlântico, com suas oscilações atuando como gangorras. Como estes índices citados são coincidentes entre os arranjos, provavelmente a adição dos índices de Fluxo Solar AMMwind (ambos de agosto), no 4º arranjo, trouxeram uma estabilidade maior ao modelo, diminuído os desvios sem alterar relativamente o desempenho.

Analisando as simulações de ambos os arranjos, os dois proporcionam a possibilidade de bons ajustes em suas simulações. Priorizando o melhor ajuste, foi mais frequente encontra-lo nas simulações do 3º arranjo, sendo selecionado como o mais viável, além de possuir menos variáveis de entrada. De acordo com essas considerações, o melhor ajuste do 3º arranjo é apresentado nos gráficos da figura 5.18 e 5.19.

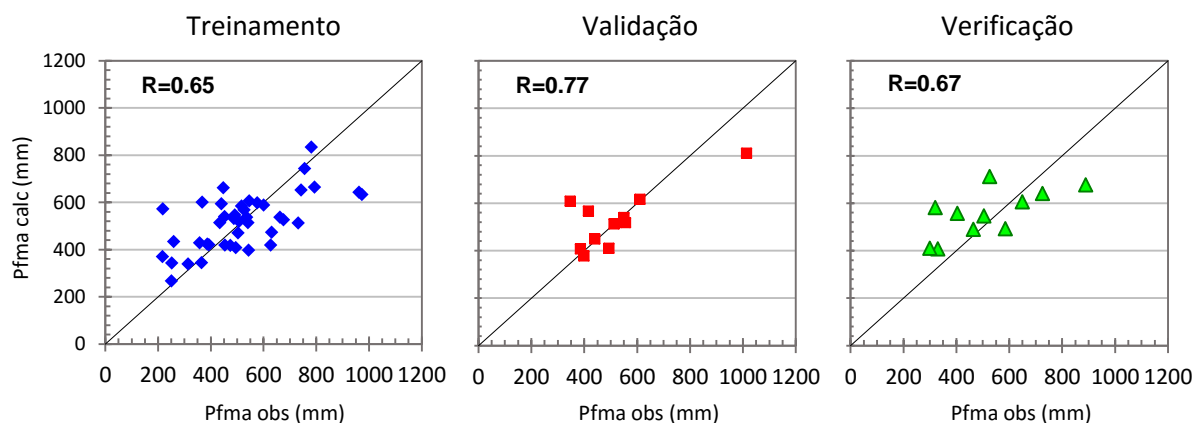


Figura 5.18 – Melhor ajuste entre as simulações do 3º arranjo para os conjuntos de treinamento, validação e verificação dos dados.

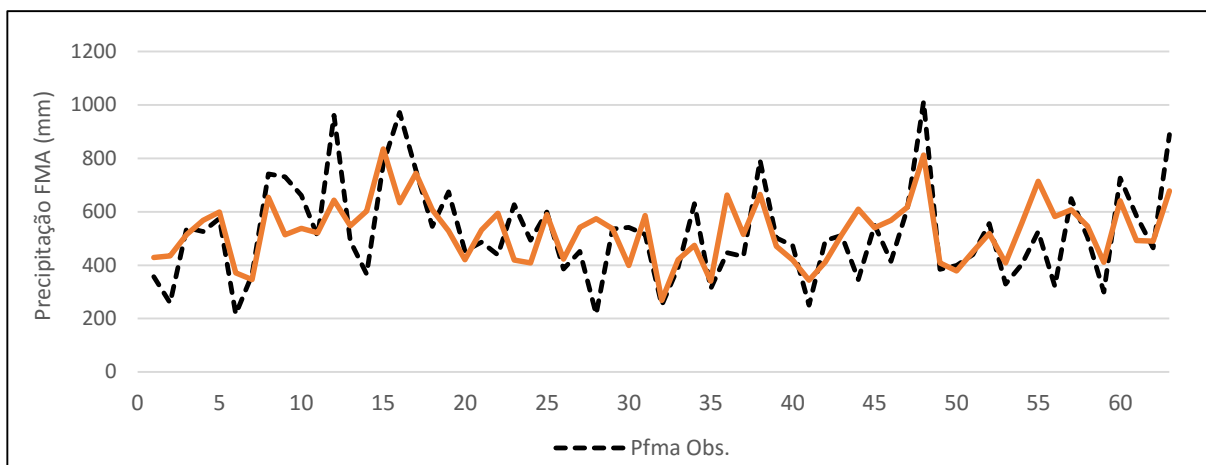


Figura 5.19 – Melhor ajuste entre as simulações do 7º arranjo para os conjuntos de treinamento, validação e verificação dos dados.

### 5.3.3 RNA com 61 anos de dados

#### 5.3.3.1 Resultados referentes ao 1º grupo de variáveis predictoras

Na série de gráficos apresentados na Figura 5.20, observa-se que em relação ao coeficiente de Pearson, a maioria dos modelos escolhidos apresentam valores maiores para os seus conjuntos em relação aos grupos anteriores, com valores globais em torno de 0,5, e sem a presença de desvios tão altos como visto também anteriormente. Os valores do conjunto de validação são os maiores para todos os modelos. Seguindo a linha de maior desempenho global, com maior parcimônia entre os conjuntos, podemos destacar o 1º, 3º, 5º e 11º arranjo como os de melhor desempenho.

Os valores de NS para esse grupo de variáveis se demonstram com desempenhos melhores do que os outros grupos já avaliados, com algumas exceções no caso do 2º, 4º, 6º e 7º arranjo, que apresentaram valores negativos para os conjuntos de verificação, além dos altos desvios. Como melhores resultados para essa medida, temos novamente o 1º, 3º, 5º e 11º arranjo, onde o 1º arranjo se demonstra mais homogêneo e o 11º arranjo como maior desempenho global, além de um alto desvio no conjunto de verificação.

Com relação aos valores de MAE, nota-se uma nítida diminuição dos patamares nos conjuntos de treinamento e validação, com a maioria abaixo de 130mm. Os valores de verificação demonstram-se mais elevados que os outros conjuntos para todos os modelos, com os casos mais críticos referentes ao 2º, 4º e 6º arranjo. Os valores dos conjuntos mais baixos estão presentes no 5º e 11º arranjo.

### 1º Grupo - Intervalo de 61 anos

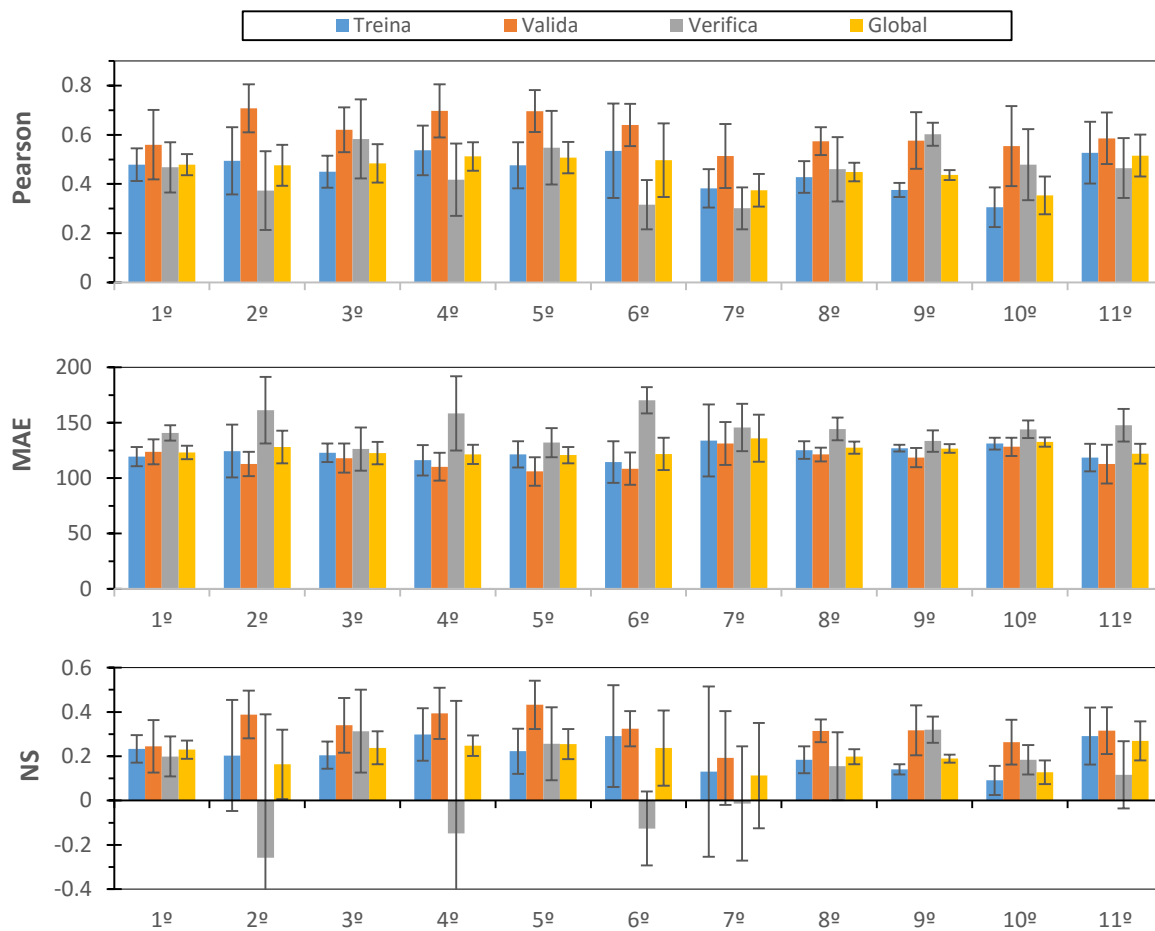


Figura 5.20 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 1º grupo de dados de 61 anos para o conjunto de treinamento, testes, validação e global.

Tabela 5.21 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 61 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	43-9-9	4	-2	0.48	0.56	0.47	0.48	119	124	141	123	0.23	0.24	0.20	0.23
2º	4 Meses	43-9-9	4	3	0.49	0.71	0.37	0.48	124	113	161	128	0.20	0.39	-0.26	0.16
3º	5 Meses	43-9-9	4	-2	0.45	0.62	0.58	0.48	123	118	126	123	0.21	0.34	0.31	0.24
4º	6 Meses	43-9-9	4	1	0.54	0.70	0.42	0.51	116	110	158	121	0.30	0.39	-0.15	0.25
5º	7 Meses	43-9-9	4	4	0.48	0.70	0.55	0.51	121	106	132	121	0.22	0.43	0.26	0.26
6º	8 Meses	43-9-9	3	0	0.54	0.64	0.32	0.50	114	108	170	122	0.29	0.32	-0.13	0.24
7º	9 Meses	33-14-14	3	4	0.38	0.51	0.30	0.37	134	131	146	136	0.13	0.19	-0.01	0.11
8º	8 Meses	43-9-9	3	2	0.43	0.57	0.46	0.45	125	121	144	128	0.18	0.31	0.16	0.20
9º	8 Meses	43-9-9	3	-1	0.38	0.58	0.60	0.44	127	119	133	127	0.14	0.32	0.32	0.19
10º	5 Meses	43-9-9	4	-2	0.31	0.55	0.48	0.35	131	128	144	133	0.09	0.26	0.18	0.13
11º	8 Meses	43-9-9	3	5	0.53	0.59	0.46	0.52	118	113	148	122	0.29	0.32	0.12	0.27

Tabela 5.22 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 61 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	43-9-9	4	-2	0.07	0.14	0.10	0.04	9	11	7	6	0.06	0.12	0.09	0.04
2º	4 Meses	43-9-9	4	3	0.14	0.10	0.16	0.08	24	11	30	15	0.25	0.11	0.65	0.16
3º	5 Meses	43-9-9	4	-2	0.07	0.09	0.16	0.08	8	13	20	10	0.06	0.12	0.19	0.07
4º	6 Meses	43-9-9	4	1	0.10	0.11	0.15	0.06	14	13	33	9	0.12	0.12	0.60	0.05
5º	7 Meses	43-9-9	4	4	0.09	0.08	0.15	0.06	12	13	13	7	0.10	0.11	0.16	0.07
6º	8 Meses	43-9-9	3	0	0.19	0.09	0.10	0.15	19	15	12	15	0.23	0.08	0.17	0.17
7º	9 Meses	33-14-14	3	4	0.08	0.13	0.09	0.07	32	19	21	21	0.38	0.21	0.26	0.24
8º	8 Meses	43-9-9	3	2	0.06	0.06	0.13	0.04	8	6	10	5	0.06	0.05	0.15	0.03
9º	8 Meses	43-9-9	3	-1	0.03	0.12	0.05	0.02	3	9	10	4	0.02	0.11	0.06	0.02
10º	5 Meses	43-9-9	4	-2	0.08	0.16	0.14	0.08	5	8	8	4	0.07	0.10	0.07	0.05
11º	8 Meses	43-9-9	3	5	0.13	0.10	0.12	0.09	13	17	15	9	0.13	0.11	0.15	0.09

Na tabela 5.21, com o detalhamento da topologia, podemos perceber que entre os modelos houve praticamente a presença de apenas da distribuição 43-9-9, com exceção do 7º arranjo obteve seu melhor desempenho com um modelo de distribuição 33-14-14. Já na camada intermediária, novamente houve uma maior variação no número de neurônios, com mais da metade das melhores topologias escolhidas para cada arranjo possuindo mais neurônios na camada oculta em relação a quantidade de variáveis de entrada.

Foram destacados o 5º e o 11º arranjo como aqueles com melhores desempenho dentre as medidas. Nota-se que a antecedência de previsão é de 7 e 8 meses respectivamente, e a diferença da quantidade de variáveis de entrada é de uma a mais para o 5º arranjo. Aparentemente, apesar do 5º arranjo apresentar melhores medidas de desempenho para os conjuntos de validação e verificação, o 11º arranjo obtém o valor mais elevado em seu conjunto de treinamento, que por possui mais dados que os demais conjuntos, acaba obtendo um desempenho global melhor. Ambos não apresentam diferenças significativas em seus desvios (Tabela 5.22).

Avaliando os índices de ambos os arranjos, eles possuem em comum os dois índices que também estão presente em todos os outros arranjos, no caso, o índice AO de junho e EPO de maio. A maioria dos arranjos desse grupo tem algumas características coincidentes entre si, devido a alta correlação entre seus índices, principalmente pela maioria abranger as regiões dos Ninos no Pacífico. Apesar disso, os índices dentro de cada arranjo são os menos correlacionados possíveis entre si. As variáveis do 5º arranjo são formadas por dois índices da região dos Ninos, NINO 3.4 de julho e NINO1+2 de abril, além dos índices AO de junho e EPO de maio. Enquanto que o 11º arranjo é formado apenas pelo índice MEI de maio, que reúne vários fatores climáticos na sua contabilização, também atuando na região dos Ninos. A priori, a combinação do MEI com os índices AO e EPO, trouxe o melhor desempenho entre os arranjos com uma maior antecedência de previsão.

Analisando as simulações de ambos os arranjos, os dois proporcionam a possibilidade de bons ajustes em suas simulações. Priorizando o melhor ajuste, foi possível encontra-lo nas simulações do 11º arranjo, sendo selecionado como o mais viável, além de possuir menos variáveis de entrada, também possui maior antecedência de previsão. De acordo com essas considerações, o melhor ajuste do 11º arranjo é apresentado nos gráficos da figura 5.21 e 5.22.

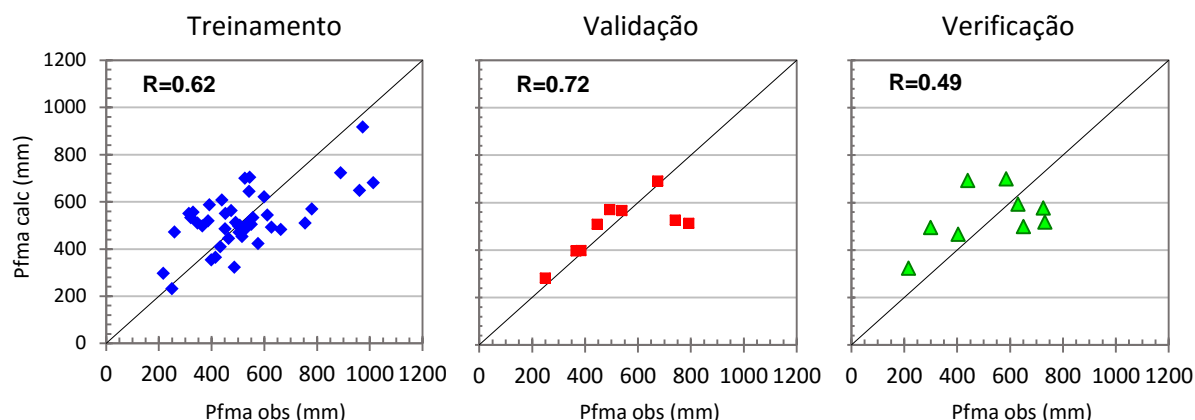


Figura 5.21 – Melhor ajuste entre as simulações do 11º arranjo para os conjuntos de treinamento, validação e verificação dos dados.

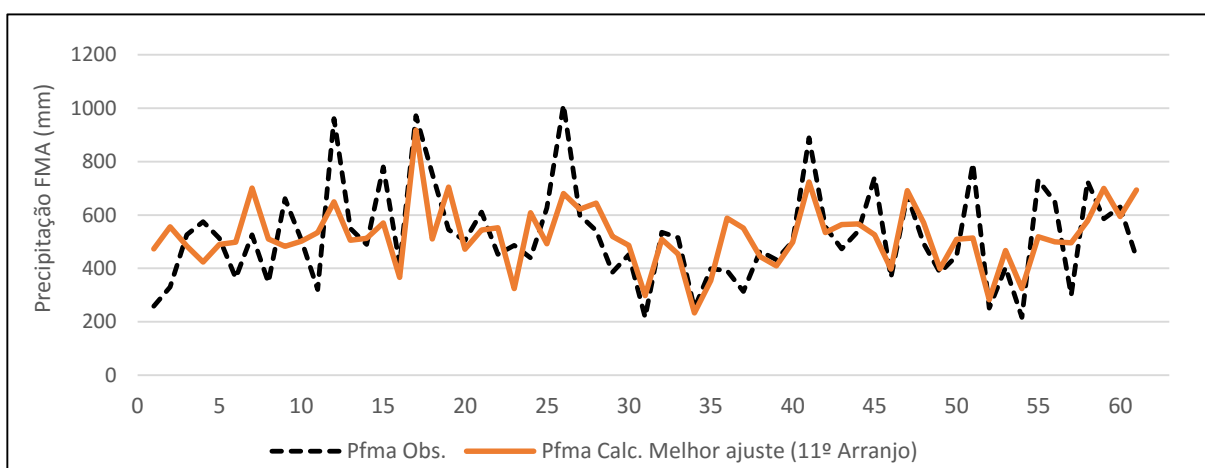


Figura 5.22 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 11º Arranjo do 1º grupo de variáveis predictoras de 61 anos.

### 5.3.3.2 Resultados referentes ao 2º grupo de variáveis predictoras

Na série de gráficos apresentados na Figura 5.23, com relação ao coeficiente de Pearson, nota-se que os valores dos conjuntos de verificação estão significativamente mais baixos em relação ao 1º grupo de variáveis de 61 anos, inclusive com maiores desvios. Dentro dessa alta oscilação entre os conjuntos, as medidas de validação para a maioria dos arranjos conseguem alcançar altos valores, mesmo assim, analisando-se de forma global a maioria se mantém com valores abaixo de 0,4. Apenas o 8º e o 12º arranjo conseguem manter os maiores valores globais do grupo, com as medidas de seus conjuntos também em níveis mais altos.



Com relação as medidas de NS para esse grupo de variáveis, os valores se encontram em níveis mais baixos comparados ao grupo anterior, além de apresentar para a maioria de seus arranjos, conjuntos de verificação e até de validação com valores negativos. A maioria dos desvios nessas medidas, não se encontram tão elevadas quanto em outros grupos já analisados. Mesmo assim, é de se destacar nesse grupo que há arranjos que ultrapassaram bastante essas medidas para o conjunto de verificação, como o caso do 10º, 11º e 13º arranjo. As melhores medidas para esse coeficiente, se encontram no 8º e 12º arranjo, com os valores mais altos para o ultimo citado.

Os valores de MAE, a maioria dos arranjos não apresenta muita oscilação entre os conjuntos, apenas uma leve superioridade nas medidas de verificação para maioria dos arranjos, com destaque para o alto desvio nesse conjunto no 10º e 11º arranjo. Os valores mais baixos dos conjuntos, estão presentes no 12º arranjo.

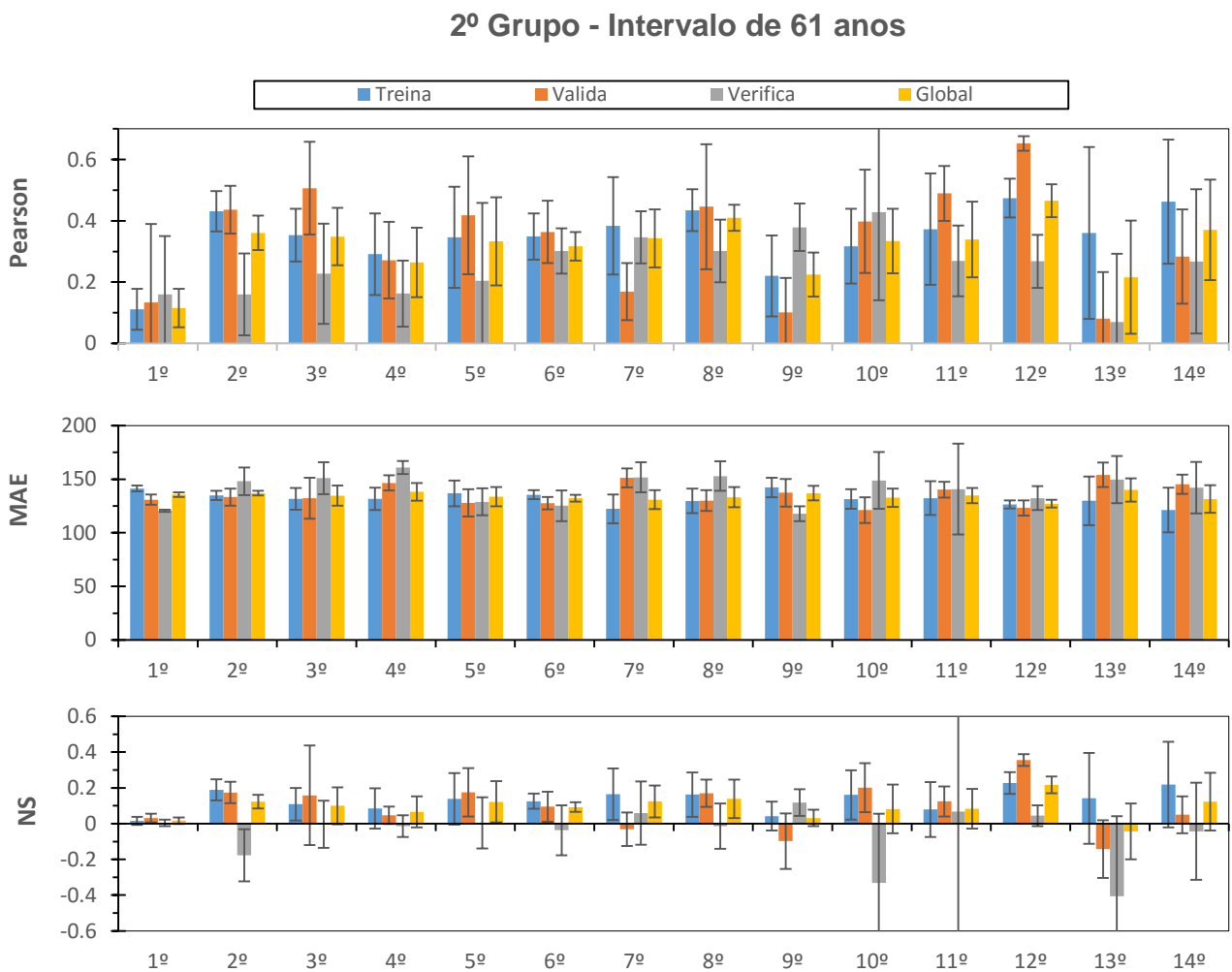


Figura 5.23 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 2º grupo de dados de 61 anos para o conjunto de treinamento, testes, validação e global.

Tabela 5.23 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 61 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	39-11-11	2	0	0.11	0.13	0.16	0.12	141	131	121	136	0.02	0.03	0.00	0.02
2º	4 Meses	39-11-11	2	3	0.43	0.44	0.16	0.36	135	133	148	137	0.19	0.17	-0.18	0.12
3º	4 Meses	43-9-9	3	3	0.35	0.51	0.23	0.35	132	132	151	135	0.11	0.16	0.00	0.10
4º	6 Meses	43-9-9	2	3	0.29	0.27	0.16	0.26	132	147	161	138	0.09	0.05	-0.01	0.07
5º	4 Meses	39-11-11	3	4	0.35	0.42	0.20	0.33	137	128	129	134	0.14	0.17	0.00	0.12
6º	6 Meses	39-11-11	4	0	0.35	0.36	0.30	0.32	136	128	125	132	0.13	0.10	-0.04	0.09
7º	8 Meses	43-9-9	3	2	0.38	0.17	0.35	0.34	122	151	152	131	0.17	-0.03	0.06	0.12
8º	8 Meses	43-9-9	3	2	0.43	0.45	0.30	0.41	130	130	153	133	0.16	0.17	-0.01	0.14
9º	8 Meses	39-11-11	3	5	0.22	0.10	0.38	0.22	142	137	118	137	0.04	-0.10	0.12	0.03
10º	8 Meses	43-9-9	4	5	0.32	0.40	0.43	0.33	131	121	149	133	0.16	0.20	-0.33	0.08
11º	9 Meses	39-11-11	3	1	0.37	0.49	0.27	0.34	132	140	141	135	0.08	0.12	0.07	0.08
12º	9 Meses	33-14-14	2	0	0.47	0.65	0.27	0.47	127	123	132	127	0.23	0.36	0.04	0.22
13º	10 Meses	33-14-14	2	5	0.36	0.08	0.07	0.22	130	154	150	140	0.14	-0.14	-0.41	-0.04
14º	4 Meses	33-14-14	4	3	0.46	0.28	0.27	0.37	121	145	142	132	0.22	0.05	-0.04	0.12

Tabela 5.24 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 61 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	39-11-11	2	0	0.07	0.26	0.19	0.06	3	5	1	2	0.02	0.02	0.02	0.02
2º	4 Meses	39-11-11	2	3	0.07	0.08	0.13	0.06	4	8	13	2	0.06	0.06	0.15	0.04
3º	4 Meses	43-9-9	3	3	0.09	0.15	0.16	0.09	10	19	15	9	0.09	0.28	0.13	0.10
4º	6 Meses	43-9-9	2	3	0.13	0.12	0.11	0.11	10	7	6	8	0.11	0.05	0.06	0.09
5º	4 Meses	39-11-11	3	4	0.16	0.19	0.25	0.14	12	13	13	9	0.14	0.14	0.14	0.12
6º	6 Meses	39-11-11	4	0	0.08	0.10	0.07	0.05	4	6	14	3	0.04	0.08	0.14	0.03
7º	8 Meses	43-9-9	3	2	0.16	0.09	0.08	0.09	13	9	14	9	0.14	0.09	0.18	0.09
8º	8 Meses	43-9-9	3	2	0.07	0.20	0.10	0.04	11	10	14	9	0.12	0.08	0.13	0.11
9º	8 Meses	39-11-11	3	5	0.13	0.11	0.08	0.07	9	13	7	7	0.08	0.16	0.07	0.05
10º	8 Meses	43-9-9	4	5	0.12	0.17	0.29	0.11	9	12	26	9	0.14	0.14	0.39	0.14
11º	9 Meses	39-11-11	3	1	0.18	0.09	0.12	0.12	16	7	42	7	0.15	0.08	0.72	0.11
12º	9 Meses	33-14-14	2	0	0.06	0.02	0.09	0.05	4	7	11	4	0.06	0.03	0.06	0.05
13º	10 Meses	33-14-14	2	5	0.28	0.15	0.22	0.19	23	11	22	11	0.25	0.16	0.45	0.16
14º	4 Meses	33-14-14	4	3	0.20	0.15	0.24	0.16	21	9	24	13	0.24	0.10	0.27	0.16

Na tabela 5.23, com o detalhamento da topologia, podemos perceber que entre os modelos houve uma maior variação das distribuições entre os arranjos. Diferentemente do 1º grupo de variáveis de 61 anos, a distribuição mais frequente foi a 39-11-11, seguida da distribuição 43-9-9 e 33-14-14. Também houve uma maior variação no número de neurônios na camada intermediária, com o destaque que nesse grupo não teve entre os melhores modelos de RNA, topologias com menos neurônios na camada intermediária em relação a quantidade de variáveis de entrada. Ou seja, nesse grupo foi preciso haver um aumento do número de neurônios na camada intermediária para se obter um melhor ajuste na maioria dos modelos.

Destacados o 8º e 12º arranjo como aqueles com melhores desempenho dentre os demais, observa-se que a antecedência de previsão é de 8 e 9 meses respectivamente, e a diferença da quantidade de variáveis de entrada é de uma a mais para o 8º arranjo. Globalmente o 12º arranjo se demonstra com melhor desempenho entre suas medidas, onde apresenta os maiores valores do grupo para os conjuntos de treinamento e validação, somente com o coeficiente de Pearson relativamente mais baixo para o conjunto de verificação em relação ao 8º arranjo. Ambos os arranjos têm em comum a presença do índice PNA (junho e maio, respectivamente para 8º e 12º arranjo), que também está presente em praticamente todos os outros arranjos. Outro índice relativamente em comum é o índice ONI e NINO3.4 (junho e maio, respectivamente para 8º e 12º arranjo), onde o ONI usa os valores do NINO3.4 para o cálculo de seu índice. A maior diferença é a adição no 8º arranjo do índice EA (maio) do oceano Atlântico.

Analisando as simulações de ambos os arranjos, o 12º arranjo proporciona a possibilidade de melhores ajustes em suas simulações. Além disso, este arranjo tem a vantagem de possuir menos variáveis de entrada, como também uma maior antecedência de previsão. De acordo com essas considerações, o melhor ajuste do 12º arranjo é apresentado nos gráficos da figura 5.24 e 5.25.

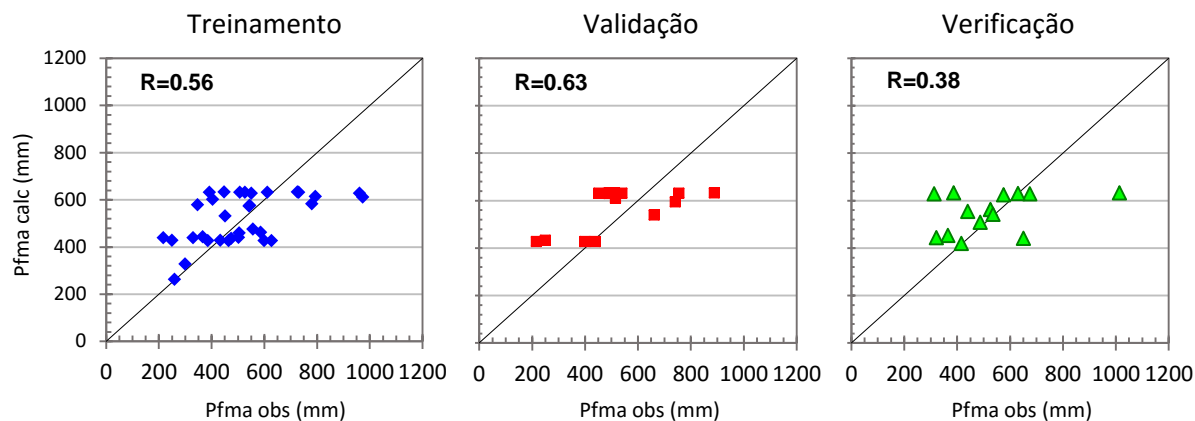


Figura 5.24 – Melhor ajuste entre as simulações do 12º arranjo para os conjuntos de treinamento, validação e verificação dos dados.

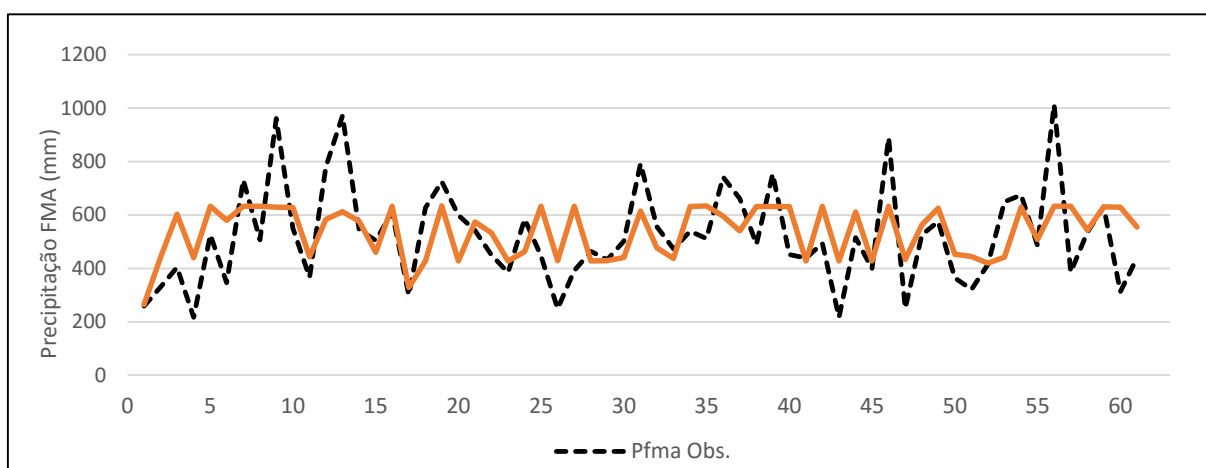


Figura 5.25 – Série temporal da precipitação observada e do melhor ajuste de previsão do 12º Arranjo do 2º grupo de variáveis predictoras de 61 anos.

### 5.3.4 RNA com 57 anos de dados

#### 5.3.4.1 Resultados referentes ao 1º grupo de variáveis predictoras

Na série de gráficos apresentados na Figura 5.26, analisando o coeficiente de Pearson, observa-se nitidamente que a maioria dos modelos apresentam seus conjuntos de verificação em níveis muito abaixo em relação aos outros conjuntos, que permanecem em patamares em torno de 0,6 para esse coeficiente, elevando assim seus níveis globais. Uma exceção acontece no 3º arranjo ao qual se demonstra mais homogêneo, mas globalmente o mais baixo do grupo. O 5º e o 6º

arranjo se apresentam com bons níveis globais, em torno de 0,5, mantendo valores razoáveis para o conjunto de verificação se comparado aos outros arranjos do grupo.

Os valores de NS para esse grupo de variáveis também obteve péssimos resultados para o conjunto de verificação, com a maioria dos valores negativos, com exceção do 3º arranjo. Em contrapartida, os outros conjuntos obtiveram bons resultados para esse coeficiente, elevando assim os seus níveis globais, com exceção novamente do 3º arranjo. O 6º arranjo se apresenta com os valores menos críticos em relação a verificação e com melhor desempenho global.

Com relação aos valores de MAE, os níveis dos conjuntos de treinamento e validação estão em patamares menores para todos os arranjos se comparado com o conjunto de verificação. O que reflete as diferenças encontradas nos outros coeficientes. O 6º arranjo se apresenta como o mais homogêneo dentre os outros.

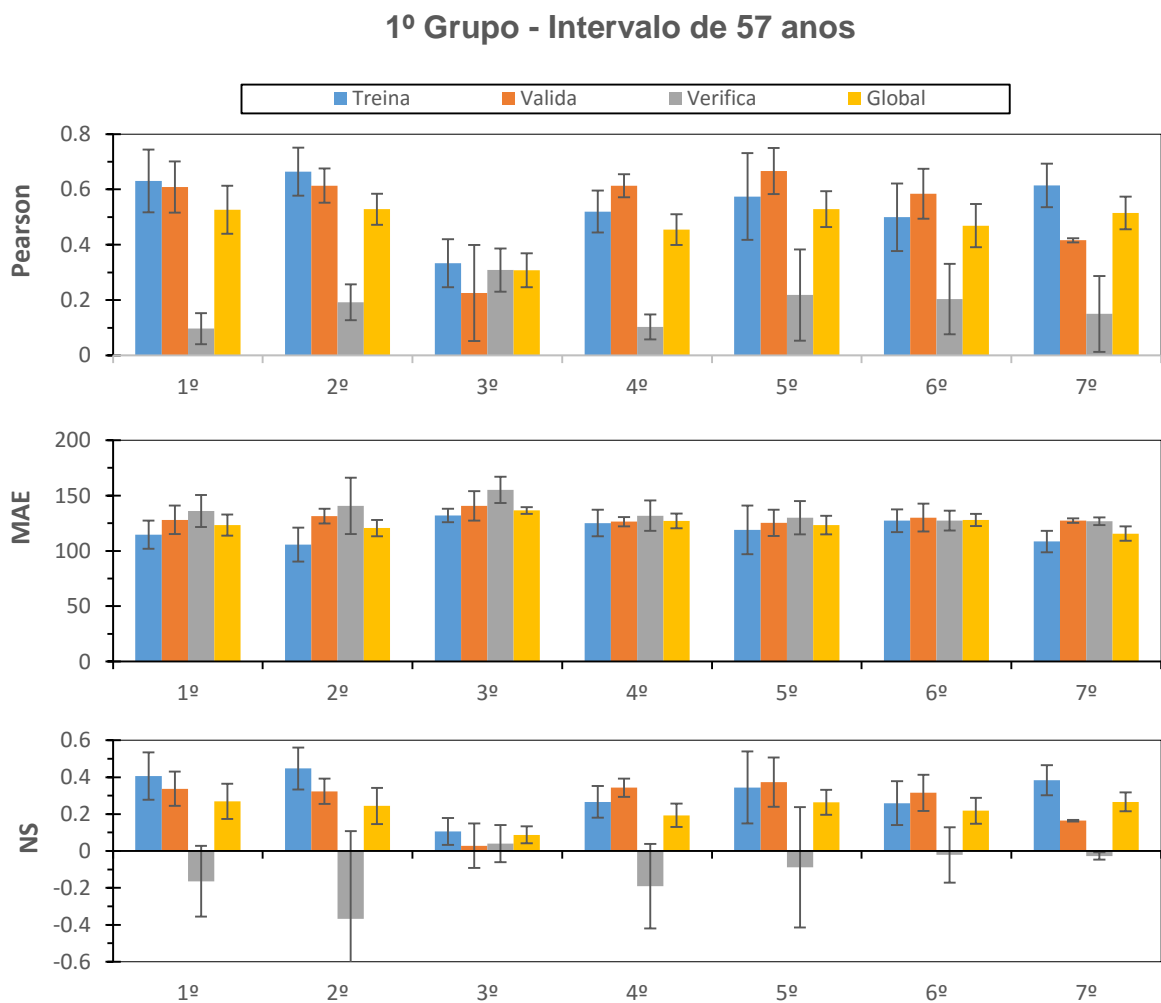


Figura 5.26 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 1º grupo de dados de 57 anos para o conjunto de treinamento, testes, validação e global.

Tabela 5.25 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 57 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	29-14-14	3	2	0.63	0.61	0.10	0.53	115	128	136	123	0.41	0.34	-0.16	0.27
2º	4 Meses	29-14-14	3	0	0.66	0.61	0.19	0.53	106	131	141	121	0.45	0.32	-0.37	0.24
3º	4 Meses	41-8-8	2	2	0.33	0.23	0.31	0.31	132	141	155	137	0.11	0.03	0.04	0.09
4º	4 Meses	29-14-14	2	1	0.52	0.61	0.10	0.45	125	126	132	127	0.27	0.34	-0.19	0.19
5º	3 Meses	29-14-14	3	0	0.57	0.67	0.22	0.53	119	125	130	123	0.34	0.37	-0.09	0.26
6º	4 Meses	29-14-14	3	1	0.50	0.58	0.20	0.47	127	130	127	128	0.26	0.32	-0.02	0.22
7º	7 Meses	35-11-11	2	0	0.61	0.42	0.15	0.52	108	127	127	116	0.38	0.16	-0.03	0.27

Tabela 5.26 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 1º grupo de dados de 57 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	29-14-14	3	2	0.11	0.09	0.06	0.09	13	13	14	10	0.13	0.09	0.19	0.09
2º	4 Meses	29-14-14	3	0	0.09	0.06	0.07	0.06	15	7	25	7	0.11	0.07	0.48	0.10
3º	4 Meses	41-8-8	2	2	0.09	0.17	0.08	0.06	6	13	12	3	0.07	0.12	0.10	0.05
4º	4 Meses	29-14-14	2	1	0.08	0.04	0.05	0.06	12	4	14	7	0.09	0.05	0.23	0.06
5º	3 Meses	29-14-14	3	0	0.16	0.08	0.17	0.06	22	12	15	8	0.19	0.13	0.33	0.07
6º	4 Meses	29-14-14	3	1	0.12	0.09	0.13	0.08	10	13	9	6	0.12	0.10	0.15	0.07
7º	7 Meses	35-11-11	2	0	0.08	0.01	0.14	0.06	10	2	4	6	0.08	0.00	0.02	0.05

Na tabela 5.25, com o detalhamento da topologia, podemos perceber que houve uma maior concentração de modelos na distribuição 29-14-14 (5 situações), e apenas uma presença das distribuições 35-11-11 e 41-8-8. Não houve uma grande variação no número de neurônios na camada intermediária para esse grupo em relação aos outros grupos, onde a variação ficou bem distribuída entre 0 e 2 a mais neurônios na camada oculta em relação a quantidade de variáveis de entrada. Dessa forma, observa-se que nesse grupo não teve entre os melhores modelos de RNA, topologias com menos neurônios na camada intermediária em relação a quantidade de variáveis de entrada.

Destacados o 5º e 6º arranjo como aqueles com melhores desempenho dentre os demais, observa-se que a antecedência de previsão é de 3 e 4 meses respectivamente, com a mesma quantidade de variáveis de entrada. Globalmente o 5º arranjo se demonstra com melhor desempenho entre suas medidas, com todos os seus conjuntos apresentando maiores valores em relação ao 6º arranjo. Mas também, em comparação, apresenta desvios mais elevados em seus conjuntos. Os arranjos formados nesse grupo são bastante idênticos em relação aos índices usados, onde a maior variação acontece entre os meses desses índices. No caso do 5º e 6º arranjo, ambos têm em comum dois índices com os mesmos meses, STA de julho e ATI de junho. O índice TP também faz parte de ambos, mas com o mês diferente, novembro para o 5º arranjo e outubro para o 6º arranjo. Essa diferença além de ser entre os meses, é também vista entre as correlações deste índice com a precipitação observada, onde o TP de novembro apresenta uma correlação mais alta, inclusive com não sendo somente por Pearson, mas também por Spearman. Este detalhe pode ter atribuído um melhor desempenho para o 5º arranjo em relação ao 6º.

Analisando as simulações de ambos os arranjos, o 5º arranjo proporciona a possibilidade de melhores ajustes em suas simulações. Vale destacar que ambos os arranjos obtiveram um fraco ajuste nas simulações de seus conjuntos verificação, algo que já era esperado devido ao péssimo desempenho nos valores de NS. De acordo com essas considerações, o melhor ajuste encontrado para o 5º arranjo é apresentado nos gráficos da figura 5.27 e 5.28.

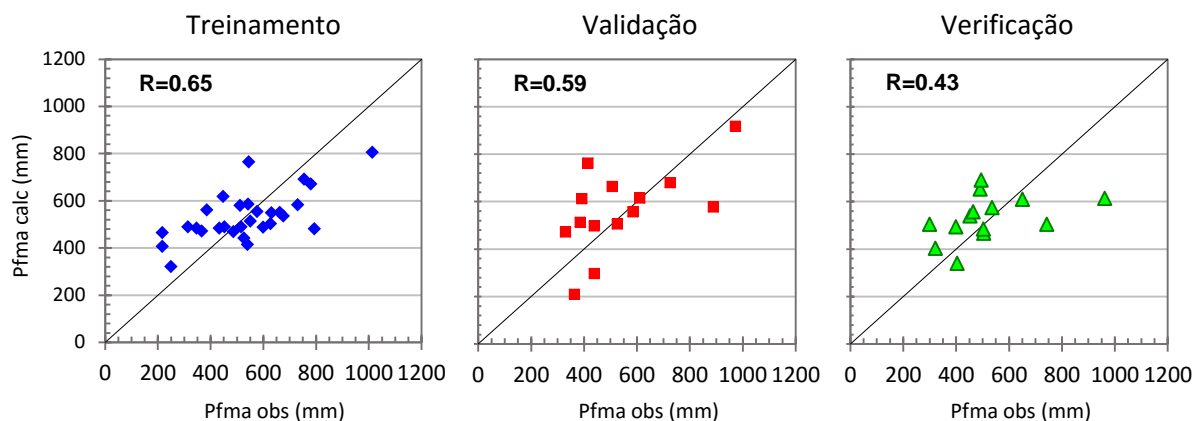


Figura 5.27 – Melhor ajuste entre as simulações do 5º arranjo para os conjuntos de treinamento, validação e verificação dos dados.

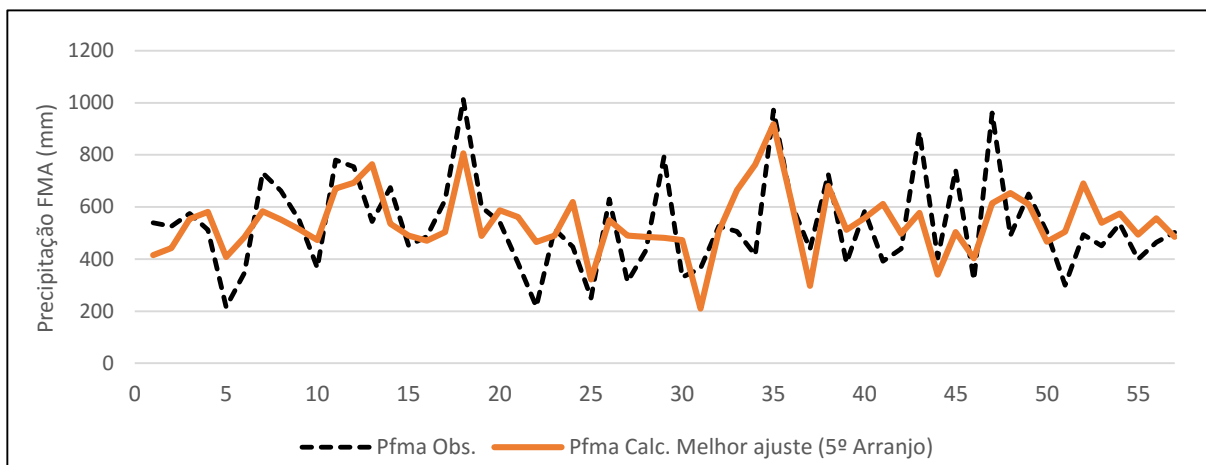


Figura 5.28 – Série temporal da precipitação observada e do melhor ajuste de previsão do 5º Arranjo do 1º grupo de variáveis preditoras de 57 anos.

#### 5.3.4.2 Resultados referentes ao 2º grupo de variáveis preditoras

Na série de gráficos apresentados na Figura 5.29, analisando o coeficiente de Pearson, observa-se novamente, como constatado no 1º grupo de variáveis de 57 anos, que boa parte dos modelos apresentam seus conjuntos de verificação em níveis muito abaixo em relação aos outros conjuntos, mas com a presença de mais exceções, sendo o caso do 1º, 2º, 3º e 7º arranjo. Dentre estes citados, o 3º arranjo se demonstra mais homogêneo, mantendo os níveis de os seus conjuntos entre patamares mais elevados em relação aos outros modelos. Nota-se também para esse grupo o alto desvio para praticamente todos os modelos, com destaque para o conjunto de validação.



Os valores de NS para esse grupo de variáveis, também obteve p $\acute$ ssimos resultados para o conjunto de verifica $\c$ o, como constatado no 1 $^{\circ}$  grupo de vari $\acute$ veis de 57 anos, com a maioria dos valores negativos, com exce $\c$ o do 1 $^{\circ}$  e 7 $^{\circ}$  arranjo. Mas, diferentemente do outro 1 $^{\circ}$  grupo, os outros conjuntos n $\acute$ o obtiveram bons resultados para esse coeficiente, diminuindo assim, significativamente, os seus n $\acute$ veis globais. Apenas o 1 $^{\circ}$  e 7 $^{\circ}$  arranjo se apresentam com os valores menos cr $\acute$ ticos em rela $\c$ o a verifica $\c$ o e com melhor desempenho global.

Com rela $\c$ o aos valores de MAE, os n $\acute$ veis dos conjuntos de treinamento e valida $\c$ o est $\acute$ o em patamares menores em praticamente todos os arranjos se comparado com o conjunto de verifica $\c$ o, com maior destaque para o 4 $^{\circ}$  e 8 $^{\circ}$  arranjo. O que reflete as diferen $\c$ as encontradas pelos outros coeficientes. O 6 $^{\circ}$  arranjo se apresenta como o mais homog $\acute$ neo dentre os outros.

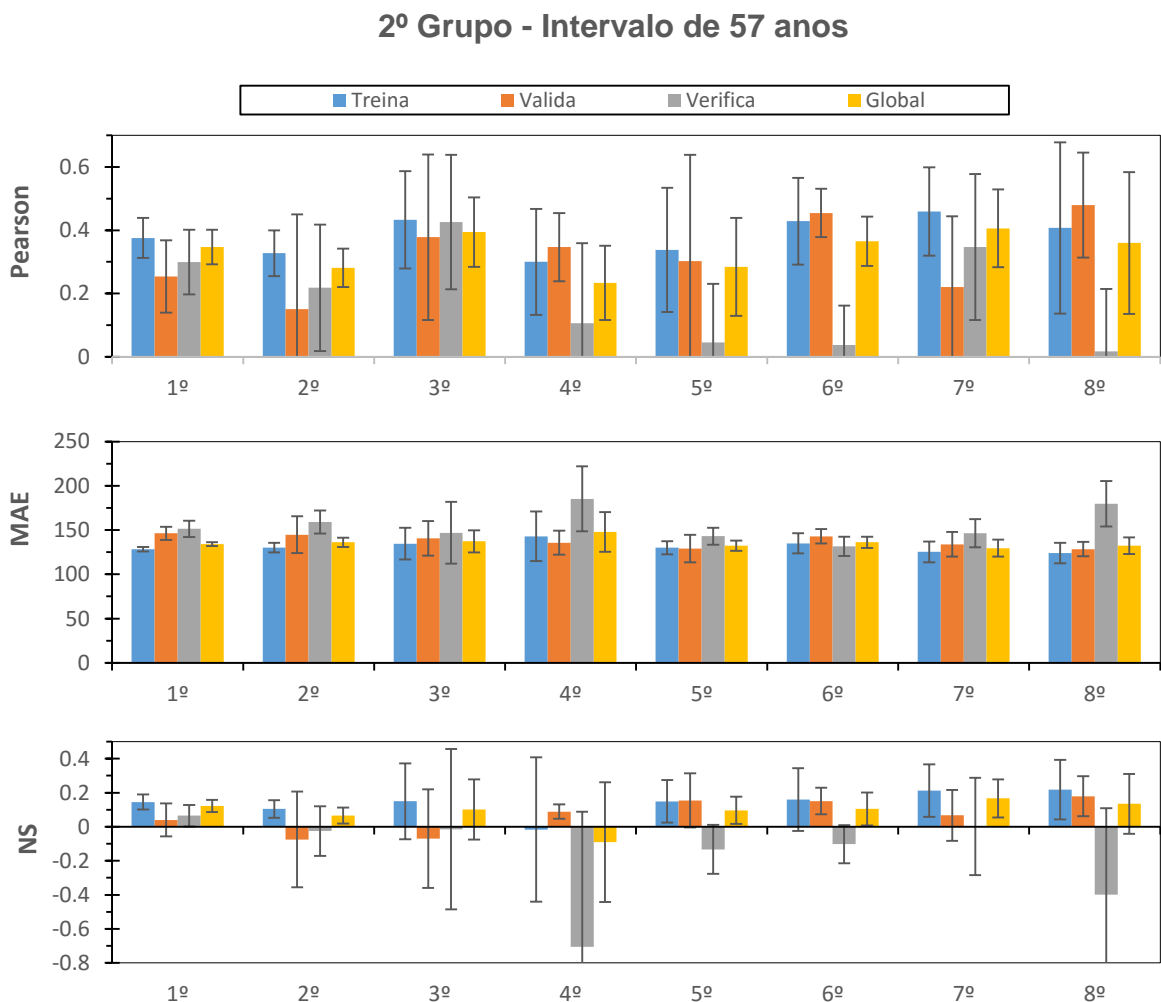


Figura 5.29 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do 2 $^{\circ}$  grupo de dados de 57 anos para o conjunto de treinamento, testes, valida $\c$ o e global.

Tabela 5.27 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 57 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	41-8-8	2	1	0.38	0.25	0.30	0.35	128	146	151	134	0.14	0.04	0.07	0.12
2º	5 Meses	41-8-8	2	4	0.33	0.15	0.22	0.28	130	145	159	136	0.10	-0.07	-0.02	0.07
3º	6 Meses	41-8-8	2	4	0.43	0.38	0.43	0.39	135	141	147	137	0.15	-0.07	-0.01	0.10
4º	10 Meses	41-8-8	2	3	0.30	0.35	0.11	0.23	143	136	185	148	-0.02	0.09	-0.71	-0.09
5º	11 Meses	35-11-11	2	0	0.34	0.30	0.04	0.28	130	129	143	132	0.15	0.15	-0.13	0.10
6º	12 Meses	29-14-14	2	4	0.43	0.45	0.04	0.37	135	143	132	136	0.16	0.15	-0.10	0.10
7º	6 Meses	41-8-8	3	-1	0.46	0.22	0.35	0.41	125	134	146	129	0.21	0.07	0.00	0.17
8º	10 Meses	41-8-8	3	-1	0.41	0.48	0.02	0.36	124	128	180	132	0.22	0.18	-0.40	0.13

Tabela 5.28 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do 2º grupo de dados de 57 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
1º	3 Meses	41-8-8	2	1	0.06	0.11	0.10	0.05	3	7	9	2	0.04	0.10	0.06	0.04
2º	5 Meses	41-8-8	2	4	0.07	0.30	0.20	0.06	5	21	13	5	0.05	0.28	0.15	0.05
3º	6 Meses	41-8-8	2	4	0.15	0.26	0.21	0.11	18	20	35	13	0.22	0.29	0.47	0.18
4º	10 Meses	41-8-8	2	3	0.17	0.11	0.25	0.12	28	14	37	22	0.42	0.04	0.79	0.35
5º	11 Meses	35-11-11	2	0	0.20	0.34	0.19	0.15	8	16	9	6	0.13	0.16	0.14	0.08
6º	12 Meses	29-14-14	2	4	0.14	0.08	0.12	0.08	11	8	11	6	0.18	0.08	0.11	0.10
7º	6 Meses	41-8-8	3	-1	0.14	0.22	0.23	0.12	12	14	16	10	0.15	0.15	0.29	0.11
8º	10 Meses	41-8-8	3	-1	0.27	0.17	0.20	0.22	12	8	26	10	0.17	0.12	0.51	0.18

Na tabela 5.27, com o detalhamento da topologia, podemos perceber que diferentemente do outro grupo, uma maior concentração de modelos na distribuição 41-8-8 (5 situações), e apenas uma presença das distribuições 35-11-11 e 29-14-14. Na camada intermediária, houve uma maior variação no número de neurônios, com mais da metade das melhores topologias escolhidas para cada arranjo possuindo mais neurônios na camada oculta em relação a quantidade de variáveis de entrada.

Destacados o 3º e 7º arranjo como aqueles com melhores desempenho dentre os demais, nota-se que os dois mantem a mesma antecedência de previsão, 6 meses, com a diferença da quantidade de variáveis de entrada, com uma a mais para o 7º arranjo. Globalmente o 7º arranjo se demonstra com melhor desempenho entre os três coeficientes, comparado ao 3º arranjo, onde também apresenta um desvio relativamente igual para Pearson e menor desvio para as medidas de MAE e NS.

Como no grupo anterior, os arranjos formados nesse grupo acabam sendo bastante idênticos pois não há muita variação em relação aos índices usados, mas diferentemente do 1º grupo de variáveis com 57 anos, há uma amplitude muito maior entre os meses usados para esses índices. No caso do 3º e 7º arranjo, o índice ATI de agosto é em comum para ambos. O índice TP também faz parte de ambos, mas com o mês diferente, junho para o 3º arranjo e maio para o 7º arranjo. A maior diferença fica por conta da variável a mais presente no 7º arranjo, ou seja, o índice STA de abril. A adição deste índice pode ter provocado melhores desempenhos neste arranjo, principalmente com relação aos coeficientes de MAE e NS, diminuindo também seus desvios.

Analisando as simulações de ambos os arranjos, o 7º arranjo proporciona a possibilidade de melhores ajustes em suas simulações. Vale destacar que a maioria das simulações em ambos os arranjos concentraram seus valores estimados próximo a média observada da precipitação (em torno de 530 mm), algo que já era esperado devido ao péssimo desempenho nos valores do coeficiente NS. De acordo com essas considerações, o melhor ajuste encontrado para o 7º arranjo é apresentado nos gráficos da figura 5.30 e 5.31.

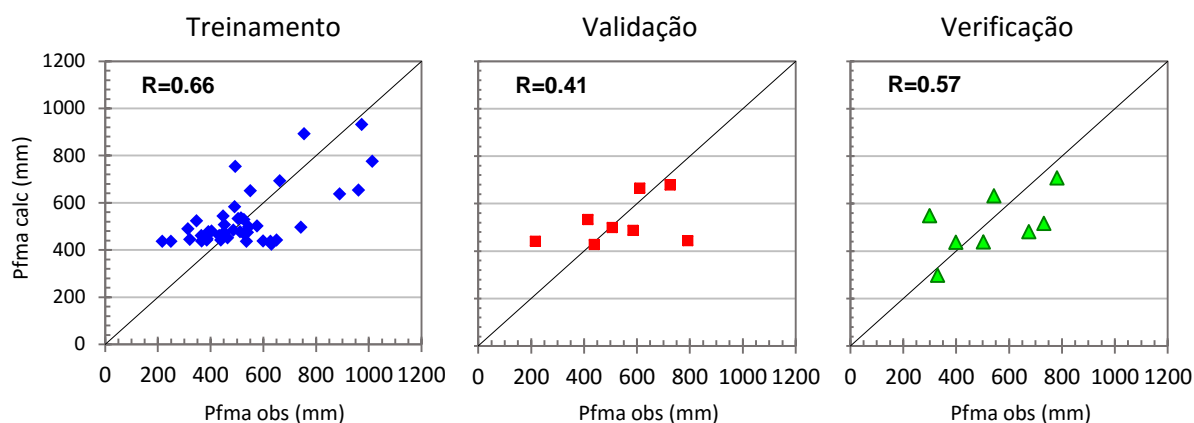


Figura 5.30 – Melhor ajuste entre as simulações do 7º arranjo para os conjuntos de treinamento, validação e verificação dos dados.

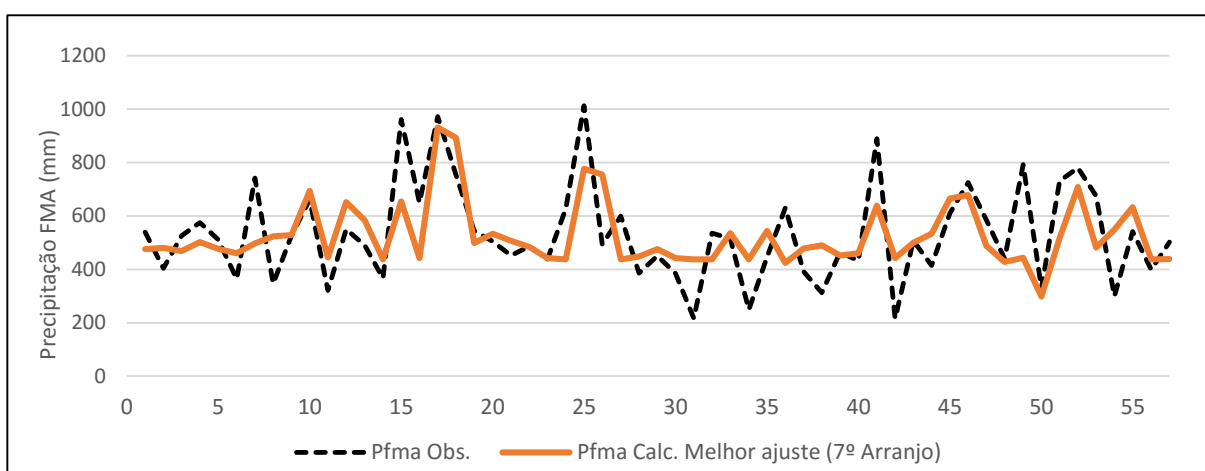


Figura 5.31 – Série temporal da precipitação observada e do melhor ajuste de previsão do 7º Arranjo do 2º grupo de variáveis predictoras de 57 anos.

### 5.3.5 RNA com 61 anos e todos os dados

#### 5.3.5.1 Resultados referentes ao 1º grupo de variáveis predictoras

Na série de gráficos apresentados na Figura 5.32, diante grande quantidade de arranjos formados para esse grupo (49 arranjos), foram selecionados somente os 10 arranjos que apresentaram os melhores modelos de previsão com relação as medidas de desempenho.

Analisando o coeficiente de Pearson, observa-se a priori que os arranjos formados nesse grupo apresentam ótimos valores para todos os modelos, com praticamente os conjuntos de dados atingindo patamares acima de 0,5 nesse coeficiente. Nota-se um maior destaque para o conjunto de verificação, com a maioria atingindo patamares acima de 0,65. Os desvios nesse grupo também se demonstram relativamente mais acentuados. Dentre as medidas desse coeficiente, o

3º arranjo obtém o melhor desempenho global, com as medidas de todos os seus conjuntos entre as mais elevadas.

Os valores do coeficiente NS para esse grupo também obteve bons resultados para todos os modelos seleccionados, com a maioria dos patamares de seus níveis globais atingindo valores acima de 0,25. Nota-se que a maioria dos modelos possuem melhores resultado para os conjuntos de verificação, enquanto que os outros os melhores resultados estão no conjunto de validação. Deve-se destacar as presenças de altos desvios para os melhores modelos dos arranjos seleccionados nesse grupo.

Com relação aos valores de MAE, os níveis globais de praticamente todos modelos estão abaixo de 120mm, com a maioria apresentando baixos desvios. Destaque para o conjunto de verificação do 18º e 19º arranjo, ao qual apresentam valores muito mais baixos que os demais, 92 e 85mm respectivamente.

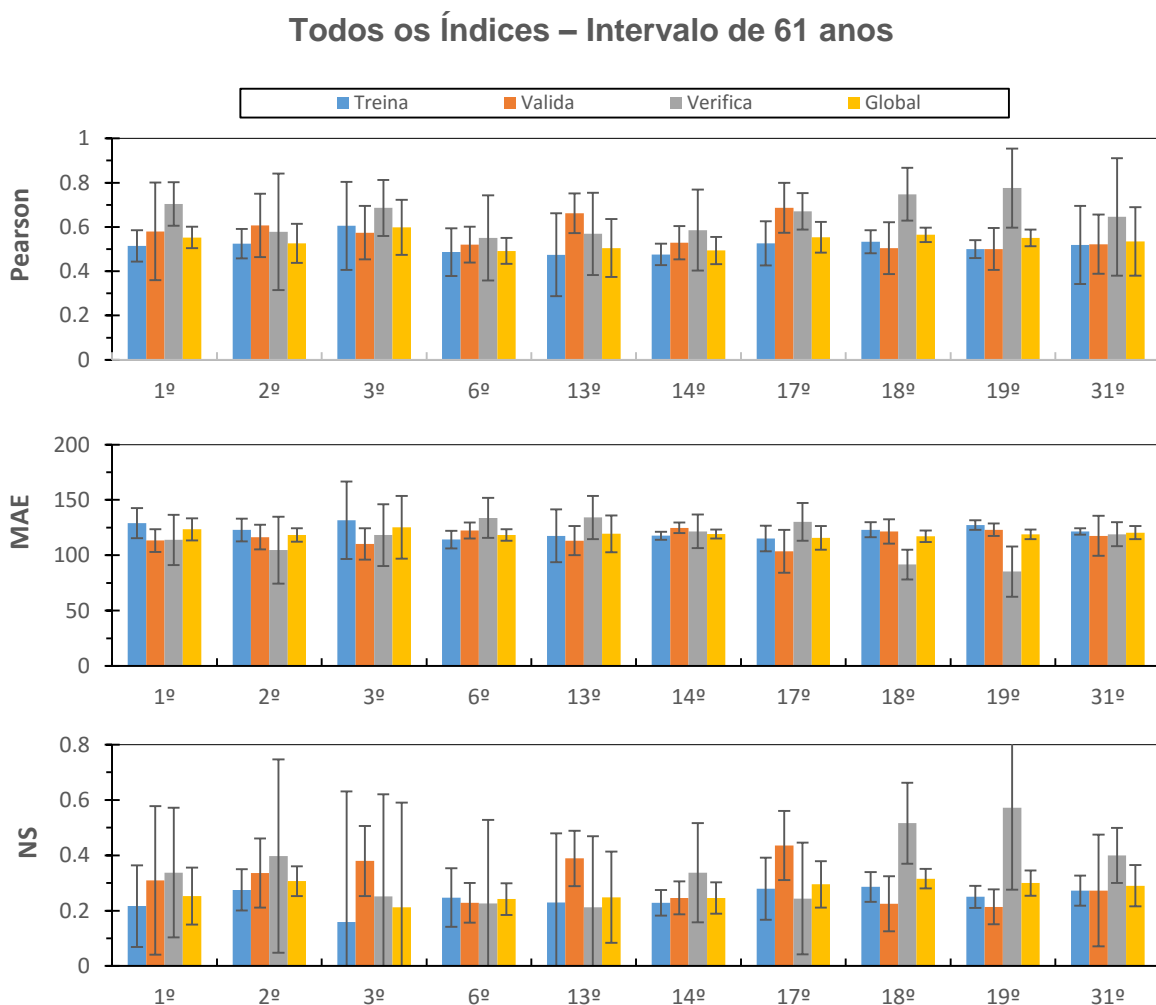


Figura 5.32 – Gráfico das medidas de desempenho das melhores topologias de cada arranjo do grupo com todos os dados de 61 anos para o conjunto de treinamento, testes, validação e global.

Tabela 5.29 – Média das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do grupo com todos os dados de 61 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
<b>1º</b>	3 Meses	39-11-11	5	4	0.51	0.58	0.70	0.55	129	113	114	123	0.22	0.31	0.34	0.25
<b>2º</b>	3 Meses	39-11-11	3	1	0.52	0.61	0.58	0.53	123	116	105	118	0.28	0.34	0.40	0.31
<b>3º</b>	4 Meses	43-9-9	4	1	0.61	0.57	0.69	0.60	132	110	118	125	0.16	0.38	0.25	0.21
<b>6º</b>	3 Meses	43-9-9	3	1	0.49	0.52	0.55	0.49	114	122	134	118	0.25	0.23	0.23	0.24
<b>13º</b>	4 Meses	43-9-9	4	2	0.47	0.66	0.57	0.50	117	113	134	119	0.23	0.39	0.21	0.25
<b>14º</b>	4 Meses	43-9-9	4	2	0.48	0.53	0.59	0.49	118	125	122	119	0.23	0.25	0.34	0.25
<b>17º</b>	4 Meses	43-9-9	3	2	0.53	0.69	0.67	0.55	115	104	130	116	0.28	0.44	0.24	0.30
<b>18º</b>	5 Meses	39-11-11	3	-1	0.53	0.50	0.75	0.56	123	122	92	117	0.29	0.22	0.52	0.32
<b>19º</b>	5 Meses	39-11-11	3	-1	0.50	0.50	0.78	0.55	127	123	85	119	0.25	0.21	0.57	0.30
<b>31º</b>	7 Meses	43-9-9	2	1	0.52	0.52	0.65	0.53	121	118	119	120	0.27	0.27	0.40	0.29

Tabela 5.30 – Desvio-padrão das medidas de desempenho das simulações para as melhores topologias de RNA de cada arranjo do grupo com todos os dados de 61 anos.

Variáveis		Topologia			Pearson				MAE				Nash-Sutcliffe			
Arranjo	Antec.	Divisão	Entra (N)	Inter. (N+i)	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global	Treina	Valida	Verifica	Global
<b>1º</b>	3 Meses	39-11-11	5	4	0.06	0.11	0.10	0.05	3	7	9	2	0.04	0.10	0.06	0.04
<b>2º</b>	3 Meses	39-11-11	3	1	0.07	0.30	0.20	0.06	5	21	13	5	0.05	0.28	0.15	0.05
<b>3º</b>	4 Meses	43-9-9	4	1	0.15	0.26	0.21	0.11	18	20	35	13	0.22	0.29	0.47	0.18
<b>6º</b>	3 Meses	43-9-9	3	1	0.17	0.11	0.25	0.12	28	14	37	22	0.42	0.04	0.79	0.35
<b>13º</b>	4 Meses	43-9-9	4	2	0.20	0.34	0.19	0.15	8	16	9	6	0.13	0.16	0.14	0.08
<b>14º</b>	4 Meses	43-9-9	4	2	0.14	0.08	0.12	0.08	11	8	11	6	0.18	0.08	0.11	0.10
<b>17º</b>	4 Meses	43-9-9	3	2	0.14	0.22	0.23	0.12	12	14	16	10	0.15	0.15	0.29	0.11
<b>18º</b>	5 Meses	39-11-11	3	-1	0.17	0.17	0.20	0.22	12	8	26	10	0.17	0.12	0.51	0.18
<b>19º</b>	5 Meses	39-11-11	3	-1	0.06	0.11	0.10	0.05	3	7	9	2	0.04	0.10	0.06	0.04
<b>31º</b>	7 Meses	43-9-9	2	1	0.07	0.30	0.20	0.06	5	21	13	5	0.05	0.28	0.15	0.05

Na tabela 5.29, com o detalhamento da topologia, podemos perceber que os 10 melhores modelos selecionados para esse grupo, houve uma maior concentração de modelos na distribuição 43-9-9 (6 situações), seguido da distribuição 39-11-11 e sem a presença da distribuição 33-14-14. Já na camada intermediária, houve uma maior frequência de topologias com mais neurônios na camada oculta em relação a quantidade de variáveis de entrada. Mas sem muita variação em relação a quantidade, onde houve 4 situações que só foi preciso um neurônio a mais, e outras 3 situações em que seus modelos se ajustaram melhor com 2 neurônios a mais.

Diante dos bons resultados apresentados por todos os modelos selecionados, foram destacados o 4º arranjo, por obter o melhor desempenho com relação a Pearson entre os arranjos; o 18º arranjo por obter tanto um bom desempenho com relação a Pearson como também para o coeficiente NS; e o 32º arranjo pela sua maior antecedência, além de apresentar boas medidas de desempenho nos coeficientes.

Analisando os índices destes 3 arranjos, nota-se todos possuem a mesma quantidade de variáveis de entrada, onde o 18º arranjo contém dois índices coincidentes em relação ao 3º e 31º arranjo, sendo eles o TSA e o Fluxo Solar respectivamente. No qual o TSA é o do mesmo mês (setembro) para o 3º e 18º, enquanto que o Fluxo Solar é de meses diferentes para o 18º e 31º (setembro e junho respectivamente). Essa defasagem de meses em relação ao índice do Fluxo Solar é pouco perceptível para os seus valores devido a variação deste índice ocorrer de forma praticamente decadal, sendo então muito próximos os valores entre setembro e junho. Desta forma, pode-se considerar que o 18º consegue conter parte das características do 3º e do 31º arranjo, no qual provavelmente lhe conferiu permanecer mais parcimonioso com seus desempenhos globais entre os melhores do grupo para todos os coeficientes.

Analisando as simulações destes arranjos, os três proporcionam a possibilidade de bons ajustes em suas simulações. Priorizando o melhor ajuste, foi possível encontra-lo nas simulações do 18º arranjo, sendo selecionado como o mais viável. De acordo com essas considerações, o melhor ajuste do 18º arranjo é apresentado nos gráficos da figura 5.33 e 5.34.

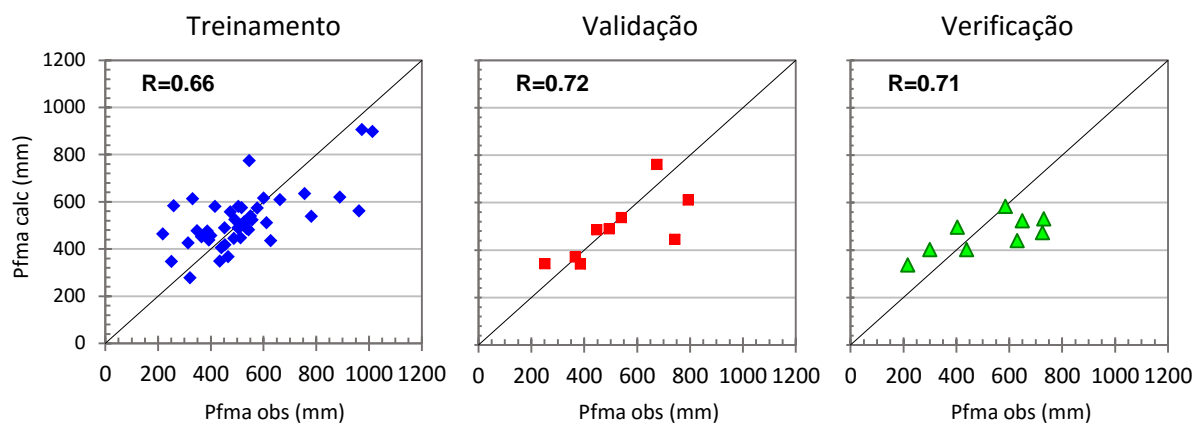


Figura 5.33 – Melhor ajuste entre as simulações do 18º arranjo para os conjuntos de treinamento, validação e verificação dos dados.

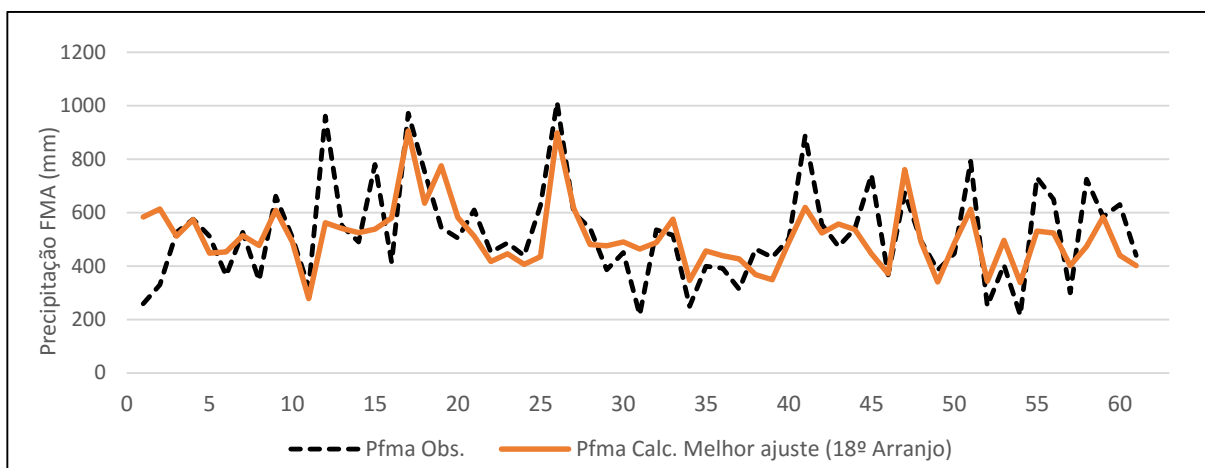


Figura 5.34 – Serie temporal da precipitação observada e do melhor ajuste de previsão do 18º Arranjo do 2º grupo de variáveis predictoras de 57 anos.

## 5.4 ANÁLISE DO EFEITO DA DISTRIBUIÇÃO DE DADOS

Nesse item foi analisado o desempenho dos grupos diante da variação das topologias testadas para os seus arranjos. Os gráficos gerados correspondem a média dos valores de Pearson para todos os arranjos de cada grupo, sendo demonstrada a média global e a média por conjunto de dados (treinamento, validação e verificação), variando-se o número de neurônios e a distribuição dos dados.



#### 5.4.1 Análise dos dados com 100 anos

Analisando os gráficos da Figura 5.35, com relação as distribuições, para os dois grupos, observa-se uma nítida diminuição dos valores de desempenho do conjunto de verificação à medida que se aumenta a quantidade de dados do conjunto de treinamento. Em média, o desempenho dos conjuntos de verificação consegue chegar a valores negativos para as distribuições com maiores quantidades de dados no conjunto de treinamento. Ou seja, a distribuição 46-27-27 para esse grupo de dados de 100 anos, acaba conferindo a cada conjunto as mesmas características que estão presentes em toda a série de precipitação, onde dessa forma, fez com que os conjuntos convergissem de maneira homogênea apresentando os mesmos desempenhos. A medida que aumenta-se os dados de treinamento, mesmo mantendo as características estatísticas (como a média e desvio-padrão) iguais aos outros conjuntos, ainda assim, acaba transferindo-se também outras características além das que foram mantidas. Dessa forma, os dados de treinamento ficam mais condicionados a desempenhos isolados devido a variáveis de entrada que favorecem mais as características presentes nesse conjunto e que não estão presentes nos outros.

Comparando-se os desempenhos dos dois grupos, o 2º grupo obteve resultados muito abaixo em relação ao 1º grupo. Ao qual pode-se destacar que as variáveis selecionadas somente pela avaliação dos escalogramas das correlações por Wavelets, não conseguem através de seus arranjos fornecerem bons preditores para a previsão por Redes Neurais Artificiais.

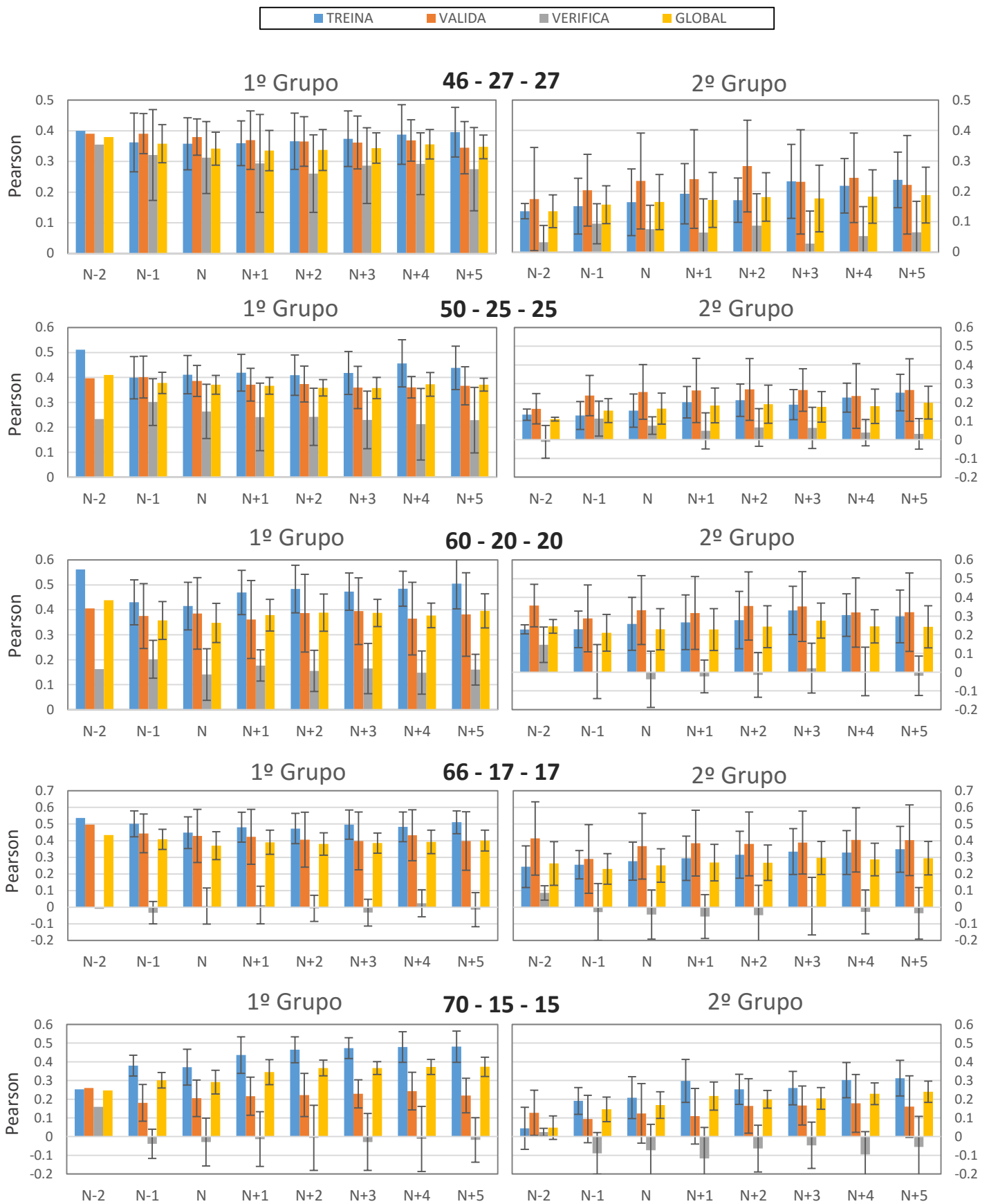


Figura 5.35 – Gráfico comparativo entre os dois grupos de variáveis de 100 anos referente a média da correlação de Pearson de todos os arranjos para cada topologia testada, de acordo com a quantidade de neurônios e a distribuição para o conjunto de treinamento, validação, verificação e global.

#### 5.4.2 Análise dos dados com 63 anos

Comparando-se os desempenhos dos dois grupos, o 2º grupo obteve em média resultados um pouco inferiores em relação ao 1º grupo, principalmente para o conjunto de verificação, com destaque para a distribuição 45-9-9. Nota-se também, que para ambos os grupos, os melhores resultados do conjunto de verificação estão presentes na distribuição 41-11-11, onde a diferença para os outros grupos também é menor em relação as outras distribuições.

Diferentemente dos dois grupos de variáveis com intervalo de 100 anos, para os dados com 63 anos, observa-se um aumento dos valores de desempenho do conjunto de validação a medida que se aumenta a quantidade de dados do conjunto de treinamento, onde provavelmente na distribuição 41-11-11 aconteça uma maior homogeneidade dos conjuntos entre as características captadas na série de precipitação.

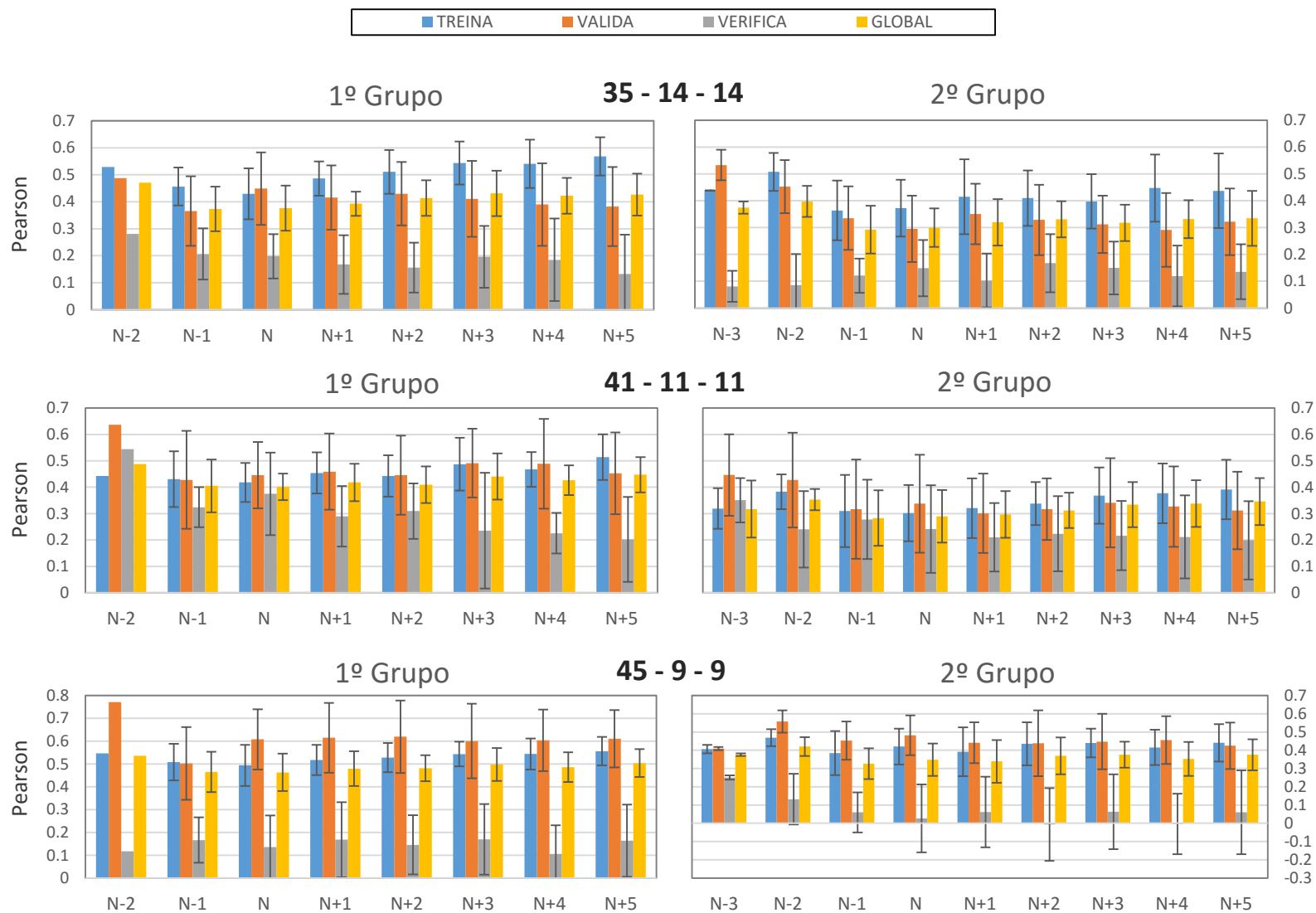


Figura 5.36 – Gráfico comparativo entre os dois grupos de variáveis de 63 anos referente a média da correlação de Pearson de todos os arranjos para cada topologia testada, de acordo com a quantidade de neurônios e a distribuição para o conjunto de treinamento, validação, verificação e global.

### 5.4.3 Análise dos dados com 61 anos

Comparando-se os desempenhos dos dois grupos, nota-se que nitidamente o 2º grupo obteve resultados muito abaixo em relação ao 1º grupo, principalmente se tratando do conjunto de verificação, que para a distribuição 43-9-9 obtém os piores valores, enquanto que o 1º grupo obtém melhores.

Coincidindo com as mesmas tendências do grupo com intervalo de 63 anos, a medida que se aumenta a quantidade de dados do conjunto de treinamento, acontece um aumento dos valores de desempenho do conjunto de validação, sendo maiores na distribuição 43-9-9, que por também apresentar os melhores resultados para o conjunto de verificação, eleva o desempenho global em relação as outras. Vale destacar os baixos desvios apresentados no 1º grupo de variáveis em relação ao desempenho entre os seus arranjos, provavelmente devido as suas combinações haver sempre na maioria a participação dos mesmos índices.

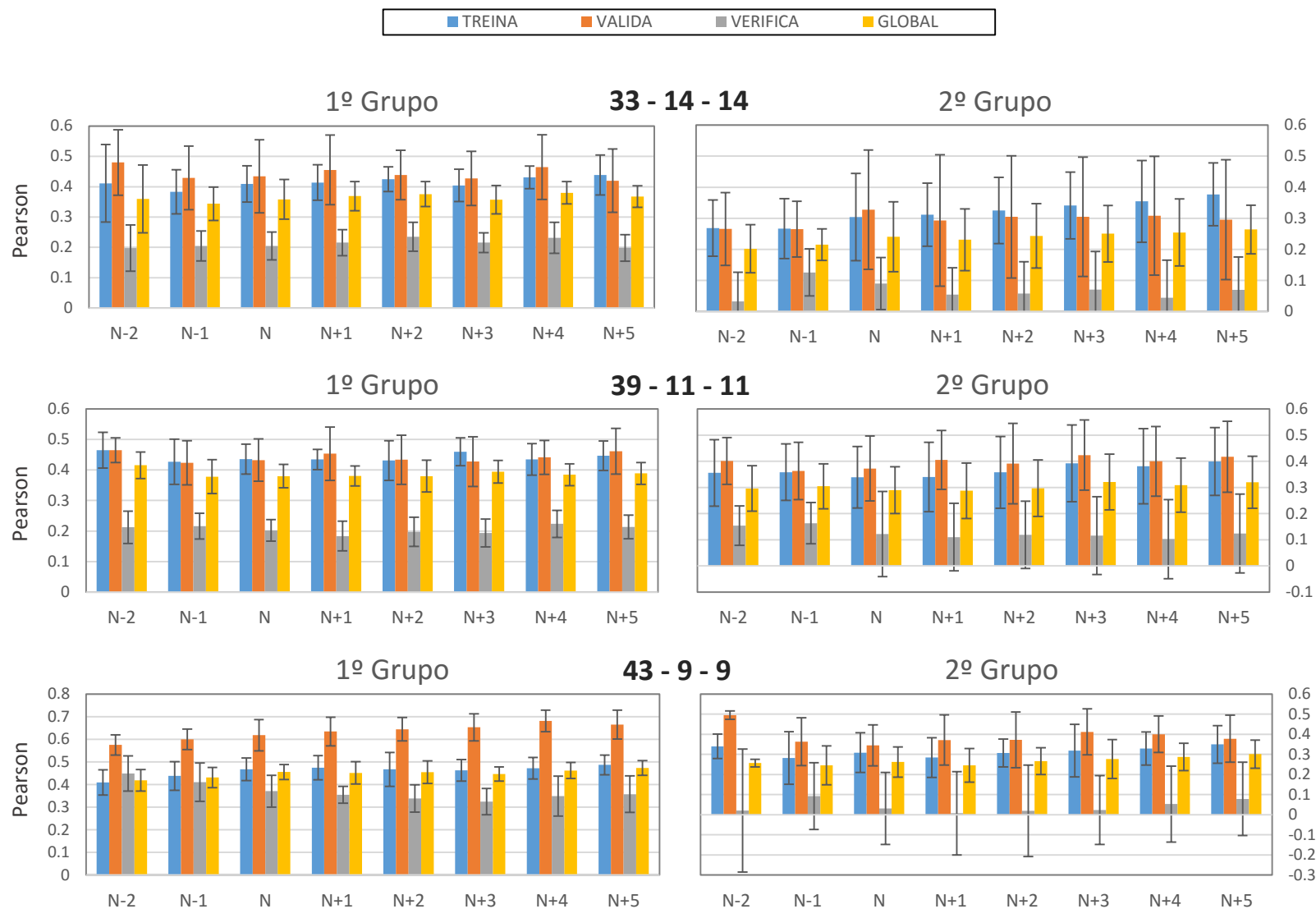


Figura 5.37 – Gráfico comparativo entre os dois grupos de variáveis de 61 anos referente a média da correlação de Pearson de todos os arranjos para cada topologia testada, de acordo com a quantidade de neurônios e a distribuição para o conjunto de treinamento, validação, verificação e global.

#### 5.4.4 Análise dos dados com 57 anos

Comparando-se os desempenhos dos dois grupos, nota-se que as variáveis para estes intervalos de dados foram as que mais prejudicaram o desempenho do conjunto de verificação, chegando a obter valores negativos para a maioria das distribuições. O melhor desempenho para o 1º grupo encontra-se com menor número dados de treinamento, na distribuição 29-14-14, onde em contraste, para o 1º grupo encontra-se com maior número dados de treinamento na distribuição 41-8-8.

Coincidindo com as mesmas tendências do grupo com intervalo de 100 anos, para ambos os grupos deste intervalo de 57 anos, há um aumento do desempenho do conjunto de treinamento a medida que também há um aumento da quantidade de seus dados.

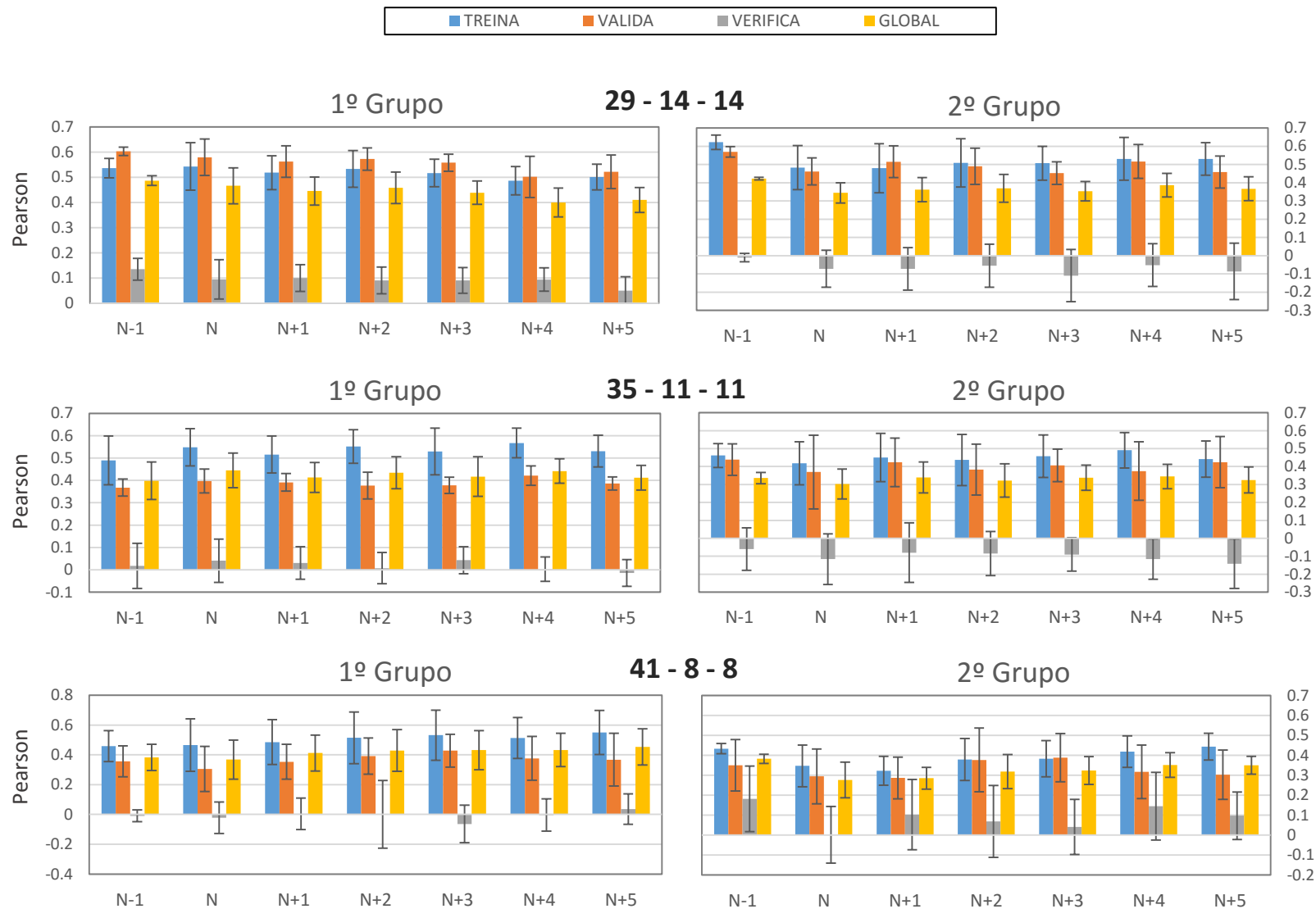


Figura 5.38 – Gráfico comparativo entre os dois grupos de variáveis de 57 anos referente a média da correlação de Pearson de todos os arranjos para cada topologia testada, de acordo com a quantidade de neurônios e a distribuição para o conjunto de treinamento, validação, verificação e global.



#### 5.4.5 Análise de todos os índices com 61 anos

Os resultados apresentados para esse grupo contendo todos os índices, manteve seus valores de desempenho mais estáveis entre as distribuições para todos os conjuntos de dados. O conjunto de treinamento obteve uma leve superioridade em seus desempenhos para todas as distribuições, destacando-se também, que um há aumento no mesmo a medida que é adicionado neurônios na camada intermediária. Por fim, nota-se que há um relativo aumento dos desvios na distribuição 43-9-9.

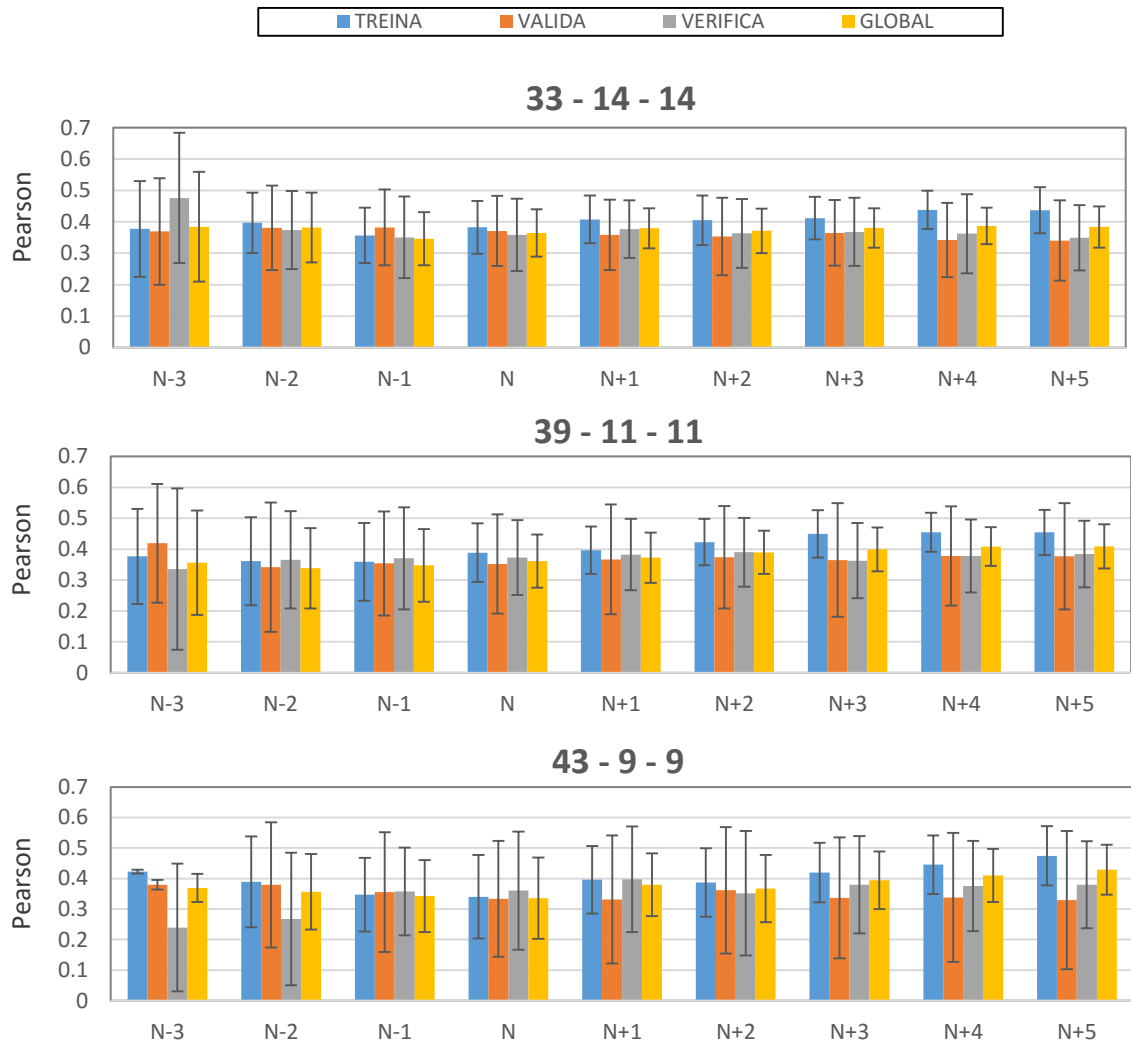


Figura 5.39 – Gráfico do efeito da distribuição para o grupo de variáveis com todos os dados referente a média da correlação de Pearson de todos os arranjos para cada topologia testada de acordo com a quantidade de neurônios para o conjunto de treinamento, validação, verificação e global.

## 6 CONCLUSÃO

Diante do estudo sobre teleconexões, constatou-se que há uma grande quantidade de índices climáticos que possuem boas correlações, algumas chegando aos maiores níveis de defasagem mensal testado, considerando os dois métodos de análises de correlações para definição das variáveis predictoras. Grande partes dos índices são muito correlacionáveis entre si, devido a abrangência de uma mesma região para contabilização de suas medidas.

O oceano Pacífico foi o que conteve a maior quantidade de indicadores climáticos com ótimas correlações, muito devido a participação da região do El Niño e, por também grande parte dos índices avaliados contabilizarem, em diferentes intensidades, oscilações do fenômeno ENOS. A maioria das correlações possuindo sinais negativos, indicam que as anomalias de precipitação para o trimestre FMA foram negativas durante os eventos da fase quente do ENOS (El Niño).

No oceano Atlântico, a maioria dos índices avaliados, por captarem as oscilações da região intertropical, conseguiram manter boas correlações com a precipitação observada, com bons níveis defasagem mensal em relação a mesma.

A regiões polares (Ártico e Antártico) com seus índices, apresentam baixa correlação pelo método de Pearson/Spearman, mas o uso da Wavelets mostra que há mais correlações possíveis além das captadas pelo outro método para essas regiões.

Os índices globais mostraram relações até então pouco estudadas como as correlações presentes no Fluxo Solar, que em ambas a técnicas esteve muito presente em praticamente todos os meses com altas correlações para Pearson/Spearman e a confirmação de que o período de renovação das manchas solares tem alguma relação com a precipitação observada devido à forte indicação no uso da Wavelets, delimitando o período de 11 anos.

A natureza do tipo de mecanismo de teleconexão envolvido, ainda é um ponto a ser debatido em outros trabalhos. Em todo caso, estes conhecimentos podem revelar e colaborar na expressão das tendências dos dados em variadas regiões e na determinação de impactos e padrões de variabilidade. No entanto, estes fenômenos e padrões não devem ser entendidos e interpretados como fenômenos de natureza linear e/ou como sistemas de causa x efeito, mas como produto de multicausas e escalas, com mecanismos diversos e teleconexões, que ainda não foram verificadas ou totalmente estudadas, e que devido sua complexidade não nos permite

afirmar se a repercussão de determinado impacto é fruto puro e único de um fenômeno natural, longe de querer naturalizar (no sentido sócio-espacial) os impactos relacionados a este caráter.

As escolhas das variáveis preditoras foi reflexo das abordagens utilizadas para sua seleção, ou seja, com o aumento da perspectiva dada pelo uso da Wavelets, foi possível a escolha de mais índices com uma quantidade maior de meses comparado a uma avaliação fornecida pelo simples valor numérico dado pela correlação de Pearson/Spearman.

A exclusão de índices correlacionados entre si nas combinações dos arranjos para a entrada a previsão, diminuiu consideravelmente a quantidade das mesmas, fornecendo quase o mesmo número de arranjos para ambos os grupos em cada intervalo de tempo. Além disso, diminuiu também a quantidade de variáveis dentro de cada arranjo.

Quanto as previsões, o 1º grupo de variáveis, selecionados pela correlação de Pearson/Spearman, apresentou os melhores desempenhos em todas os intervalos de tempo comparado ao 2º grupo. No entanto, as variáveis presentes no intervalo de 63 anos conseguiram manter bons desempenhos para ambas as técnicas.

O uso de modelos de RNA se mostrou como uma ferramenta promissora, apresentando previsões de precipitação razoáveis quanto ao horizonte de tempo, podendo eleger arranjos com até 9 meses de antecedência.

A constatação da influência distribuição da quantidade de dados divididos em cada conjunto (treinamento, validação e verificação), sobre a qualidade das previsões, foi bem destacada quando se isolou os índices climáticos de acordo os intervalos de tempo de toda a sua série fornecida. Essa influência não foi presente quando se usou todos índices climáticos na formação dos arranjos.

## 7 RECOMENDAÇÕES

Após as definições dos melhores arranjos para cada grupo de variáveis, é interessante a aplicação de uma melhor medida de desempenho baseada em tabelas de contingência, nas quais poderiam identificar o quanto o modelo consegue prever categorias de eventos restritos a intervalos previamente definidos de acordo com a utilização do resultado da previsão. Onde, dependendo da utilização, poderá ser mais importante saber a qualidade da previsão para dois resultados possíveis (evento dicotômico) ao invés de uma quantidade numérica exata do mesmo, ou seja, por exemplo, se a precipitação superará um certo valor ou não, se será um trimestre de seca ou não, podendo ampliar para casos em que o evento a ser previsto possua três ou mais resultados possíveis. Desta forma, as medidas de desempenho poderiam ser obtidas a partir da informação resumida em tabelas de contingência, facilitando o uso das previsões durante o processo de tomada de decisão.

É aconselhável a aplicação de técnicas com critérios baseados em Partial Mutual Information (PMI) ou Self Organizing Map (SOM) tanto para a combinação das variáveis preditoras na formação dos arranjos, quanto para a abordagens de distribuição de dados dos conjuntos.

Com relação ao treinamento, é recomendável o uso de um método de busca global ao invés de local para uma melhor otimização do resultado, pois se a superfície de erro for consideravelmente muito rugosa, os pesos ajustados poderão não garantir encontrar o ótimo global do problema. Com isso, a parada do treinamento da rede para evitar o overfitting pode ser prejudicada, causando o underfitting, ou seja, a rede não consegue captar características suficiente dos dados de treinamento. Os métodos de busca global mais recomendado seriam os algoritmos evolucionários, onde os algoritmos genéticos são bem indicados.

## REFERÊNCIAS BIBLIOGRÁFICAS

ANA - Agência Nacional de Águas. Disponível em <<http://www.ana.gov.br>>. Acessado em Maio 2014.

ASCE (Task Committee on Application of Artificial Neural Networks in Hydrology). 2000a. **Artificial neural networks in hydrology. I: Preliminary concepts**. Journal of Hydrologic Engineering, v. 5(2), p. 115-123.

ASCE (Task Committee on Application of Artificial Neural Networks in Hydrology). 2000b. **Artificial neural networks in hydrology. II: Hydrologic applications**. Journal of Hydrologic Engineering, v. 5(2), p. 124-137.

AESA – Agência Executiva de Gestão das Águas do Estado da Paraíba. 2014. Disponível em <<http://www.aesa.pb.gov.br>>. Acessado em Maio 2014.

ABBOT, J.; MAROHASY, J. 2014. **Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks**. Atmospheric Research, Volume 138, 1 March 2014, Pages 166-178.

ARAGÃO, J. O. R. 1998. **O impacto do ENSO e do dipolo do Atlântico no nordeste do Brasil**. Bulletin Institut Français Études Andines, v.27 n.3, p.839-844.

ARAÚJO, V.P.A.; GALVÃO, C.O.; PEDROSA, L.N.; NAVARRO, E.F. 1998. **Um sistema de informações para zoneamento agroclimático**. In: Simpósio Internacional Sobre Gestão de Recursos Hídricos, Gramado – RS.

BARNSTON, A. G.; LIVEZEY, R. E. 1989. **A closer look at the effect of the 11-year solar cycle and the QBO on the Northern Hemisphere 700mb height and extratropical North American surface temperature**. *J. Climate*, 2, 1295–1313.

BERNARDO, S. O.; MOLION, L. C. B. 2004. **Variabilidade Temporal e Espacial na Precipitação da Costa Leste do Nordeste no Período 1992-97**. In: XIII Congresso Brasileiro de Meteorologia, Fortaleza-CE. Anais XIII Congresso Brasileiro De Meteorologia. Rio de Janeiro: Sociedade Brasileira de Meteorologia.

BRAVO, J. M.; COLLISCHONN, W.; PILAR, J. V.; TUCCI, C. E. M. 2007. **Influência da capacidade de regularização de reservatórios nos benefícios da previsão de vazão de longo prazo**. Anais do XVII Simpósio Brasileiro de Recursos Hídricos, São Paulo, SP.

BRETHERTON, C. S., SMITH, C., & WALLACE, J. M. 1992. **An intercomparison of methods for finding coupled patterns on climate data**. Journal of Climate, 5, 541–560.

BOLZAN, M.J.A. 2006. **Transformada em ondeleta: uma necessidade**. Rev. Bras. Ensino Fís. [online]. vol.28, n.4, pp. 563-567.

CARLETON, A.M. 2004. **Atmospheric teleconnections involving the Southern Ocean**. J. Geophys. Res. 2003, 108, 8080.

CALANCA, P. 2014. Weather Forecasting Applications in Agriculture. Encyclopedia of Agriculture and Food Systems, 437-449.

CAVALCANTI, I, F. A. AMBRIZZE, T. 2009. **Teleconexões e suas influências no Brasil**. In: CAVALCANTI, I, F. A.; FERREIRA, N. J.; SILVA, M, G, A, J; DIAS, M. A. F. S (Org). Tempo e clima no Brasil. Oficina de Textos, p. 318-335.

Chattopadhyay, S.; Chattopadhyay, G. 2008. **Comparative study among different neural net learning algorithms applied to rainfall time series**. Meteorol. Appl., 15 (2), pp. 273–280.

CHITHRA N R.; THAMPI, S. G. 2015. **Detection and Attribution of Climate Change Signals in Precipitation in the Chaliyar River Basin, Kerala, India**. Aquatic Procedia (4), pp 755 – 763.

CHRISTOFOLETTI, A. 1979. **Análise de Sistemas em Geografia**. São Paulo: Ed. Hucitec, Ed. da Universidade de São Paulo, 106p.

COLLISCHONN, W. **Simulação hidrológica de grandes bacias**. 2001. Tese (Doutorado em Engenharia de Recursos Hídricos e Saneamento Ambiental) – Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas. 270p.

CRU. 2014. **Climatic Research Unit**. Disponível em: <http://www.cru.uea.ac.uk/cru>. Acesso: novembro de 2014.

DAWSON, C. W., & WILBY, R. L. 2001. **Hydrological modelling using artificial neural networks**. Progress in Physical Geography, 25(1), 80–108.

DA SILVA, D.F.; GALVÍNCIO J. D. 2011. **Estudo da Influência da Oscilação Decadal do Pacífico no Nordeste do Brasil**. Revista Brasileira de Geografia Física (4), pp. 665-676.

DEO, R. C.; ŞAHIN, M. 2015. **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, Volumes 161–162, Pages 65-81.

FERREIRA, A. G.; MELLO, N. G. S. 2005. **Principais sistemas atmosféricos atuantes sobre a Região nordeste do Brasil e a influência dos oceanos Pacífico e Atlântico no clima da região**. Revista Brasileira de Climatologia, Vol. 1, No 1.

GALVÃO, C. O. 1999. **Aplicabilidade em recursos hídricos da previsão de precipitação de longo prazo no nordeste do Brasil**. Tese de doutorado, Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas, Porto Alegre, Brasil, 153 p.

GARCIA-MOYA, J. A. Z. 2010. **Los modelos numéricos de predicción del tiempo**. TETHYS, Revista de Meteorologia.

GIANNINI, A.; SARAVANAN, R.; CHANG, P. 2004. **The preconditioning role of Tropical Atlantic variability in the development of the ENSO teleconnection: implications for the prediction of Nordeste rainfall**. Climate Dynamics, v.22, p. 839-855.

GOVINDARAJU, R.S.; RAO, A. R. 2000. **Artificial Neural Networks in Hydrology**. Water Science and Technology Library, Springer, Netherlands.

GURGEL V. C. **Aplicação de Técnicas Wavelets em Análise De Séries Temporais para Detecção de Correlações**. Trabalho de Conclusão de Curso. Angicos - UFERSA. 62p

HASTENRATH S. 2011. **Exploring the climate problems of Brazil's Nordeste: a review.** Climatic Change, v. 112, pp 243–251.

HAYKIN, S. 2001. **Redes Neurais – Princípios e prática.** Ed. Bookman, Porto Alegre – RS.

HE, X.; HUADE G.; JIANXIN Q. 2015. **A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall.** Journal of Hydrology, Volume 527, Pages 88-100.

HURRELL, J. W. 1995. **Decadal trends in the North Atlantic Oscillation and relationships to regional temperature and precipitation.** Science, 269, pp 676-679.

ISERI, Y.; DANDY, G. C.; MAIER, H. R.; KAWAMURA, A.; JINNO, K. 2005. **Medium term forecasting of rainfall using artificial neural networks.** In: MODSIM 2005 International congress on modelling and simulation. Modelling and Simulation Society of Australia and New Zealand, Org.: Zerger A, Argent R. M. December , Melbourne, Australia, pp 1834–1840.

JISAO. 2014. Joint Institute for the Study of the Atmosphere and Ocean. Disponível em: <http://jisao.washington.edu>. Acesso: novembro de 2014.

JONES, P. D.; T. JONSSON; WHEELER D. 1997. **Extension to the north Atlantic oscillation using early instrumental pressure observations from Gibraltar and South-west Iceland.** International Journal of Climatology, vol. 17, 1433-1450

KARAMOUZ, M., RAZAVI, S., & ARAGHINEJAD, A. 2008. **Long-lead seasonal rainfall forecasting using time-delay recurrent neural networks: a case study.** Hydrological Processes, 22, 229–241.

KRAUSE, P.; BOYLE, D. P.; BASE1, F. 2005. **Comparison of different efficiency criteria for hydrological model assessment.** Advances in Geosciences. European Geosciences Union, 5, pp 89–97.

KUCHARSKI, F.; POLZIN, D.; HASTENRATH, S. **Teleconnection mechanisms of Northeast Brazil droughts: modeling and empirical evidence.** Revista Brasileira de Meteorologia, v.23, n.2, 115-125, 2008

KUMAR, D. N., REDDY, M. J., & MAITY, R. 2007. **Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques.** Journal of Intelligent Systems, 16(4), 307–322.

LABAT, D. 2005. **Recent advances in wavelet analyses: Part 1 – A review of concepts.** J. Hydrol. 314, pp289–311.

LABAT, D. 2010. **Cross wavelet analyses of annual continental freshwater discharge and selected climate indices.** Journal of Hydrology, 385, pp269–278.

LIMA, C. A. G. 2004. **Análise e sugestões para diretrizes de uso das disponibilidades hídricas superficiais da bacia hidrográfica do rio Piancó, situada no estado da Paraíba.** 2004. Tese (Programa de Doutorado em Recursos Naturais) – UFCG, Campina Grande-PB.



LIU, W. T.; XIE, X.; KATSAROS, K. B. **Observation of oceanic origin of Sahel precipitation from space**. *Remote Sensing of Environment* 123 (2012) 593–599

MAIER, H. R., & DANDY, G. C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environmental Modelling & Software*, Volume 15 edição (1), 101–124.

MAIER, H. R.; JAIN, A.; DANDY, G. C.; SUDHEER, K. P.; 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environ. Model. Softw.* 25, 891–909.

MANSON, S.J.; GODDARD, L. 2001. Probabilistic precipitation anomalies associated with ENSO. *Bulletin of the American Meteorological Society*, v.82, n.4, p.619-638.

MASON, S. J., JOUBERT, A. M., COSIJN, C. E CRIMP. S.J. 1996. Review of seasonal forecasting techniques and their applicability to southern Africa. *Climatology Research Group, University of the Witwatersand, South Africa*, 7p.

MARENGO, J. A. Vulnerabilidade, impactos e adaptação à mudança do clima no semi-árido do Brasil. *Parcerias Estratégicas*. Brasília, 27, p. 149-176, 2008.

MEDINA B. L.; NUNES A. B. 2011. Oscilação quase-bienal em fase neutra de enso e sua influência na américa do sul: estudo de caso. *IV Encontro Sul-Brasileiro de Meteorologia*, 11 a 15 de Abril, Pelotas –RS.

MEKANIK, F.; IMTEAZ, M. A.; GATO-TRINIDAD, S.; ELMAHDI, A. 2013. Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes *J. Hydrol.*, 503, pp. 11–21

MEZA F. J.; HANSEN, J. W.; OSGOOD D. 2008: Valor Econômico da Sazonal Clima Previsões da Agricultura: Revisão da ex-ante As avaliações e recomendações para futuras pesquisas. *J. Appl. Meteor. Climatol.* , 47 (5), 1269-1286.

MOLION, L.C.B. 2005. Aquecimento Global, El Niños, Manchas Solares, Vulcões e Oscilação Decadal do Pacífico. *Revista Climanálise*.

MOLION, L.C.B. 2008. Aquecimento global: uma visão crítica. *Revista Brasileira de Climatologia*. agosto, p.08-24.

MOLION, L.C.B.; BERNARDO, S.O. 2002. Uma revisão da dinâmica das chuvas no Nordeste Brasileiro. *Revista Brasileira Meteorologia*, 17(1), pp 1-10.

MOORE, R. J.; BELL, V. A.; JONES, D. A. 2005. Forecasting for flood warning. *Comptes Rendus Geoscience*, n. 337, p. 203-217.

MORID, S., SMAKHTIN, V., & BAGHERZADEH, K. 2007. Drought forecasting using artificial neural networks and time series of drought indices. *International Journal of Climatology*, 27, 2103–2111.

Moustris, K. P. et al. 2011. Precipitation Forecast Using Artificial Neural Networks in Specific Regions of Greece. *Water Resour Manage*, 25, pp1979–1993.

NASTOS P. T. et al. 2013. Rain intensity forecast using artificial neural networks in Athens, Greece Atmos. Res., 119, pp. 153–160.

NASCIMENTO JÚNIOR L. 2014. Aspectos observacionais das teleconexões climáticas no contexto da variabilidade. Observatorium: Revista Eletrônica de Geografia, v.6, n.16, p. 46-58.

NICHOLSON S. E.; GRIST, J. P. 2008. The Seasonal Evolution of the Atmospheric Circulation over West Africa and Equatorial Africa. Department of Meteorology, The Florida State University, Tallahassee, Florida, vol. 16, no. 7.

NILSSON, P.; UVO, C. B.; BERNDTSSON, R. 2006. Monthly runoff simulation: comparison of a conceptual model, neural networks and a combination of them. Journal of Hydrology, v. 321, p. 344-363.

NOAA. 2014. National Oceanic Atmospheric Administration. National Weather Service. Climate Prediction Center. Teleconnections. Disponível em: [http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily\\_ao\\_index/teleconnections.shtml](http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/teleconnections.shtml). Acesso: novembro de 2014.

NCAR. 2014. The National Center for Atmospheric Research. Disponível em: <http://www.cgd.ucar.edu>. Acesso: novembro de 2014.

NOURANI, V.; BAGHANAM, A. H.; ADAMOWSKI, J.; KISI, O. 2014. Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review, Journal of Hydrology, Volume 514, Pages 358-377.

NUNES, L.H.; LOMBARDO, M.A. 1995. A questão da variabilidade climática – Uma reflexão crítica. Revista do Instituto Geológico, v.16, n.1/2, 21-31.

OLIVEIRA H.M. 2007. Análise de Sinais para Engenheiros: Uma abordagem via Wavelets Editora Brasport, RJ. Série SBrT/Brasport. 238p

OLYA H. G. T.; ALIPOUR H. 2015. Risk assessment of precipitation and the tourism climate index. Tourism Management, Volume 50, Pages 73-80.

PARTAL, T., KIŞI, Ö. 2007. Wavelet and neuro-fuzzy conjunction model for precipitation forecasting, Journal of Hydrology, Volume 342, Issues 1–2, Pages 199-212.

PAZ, A. R.; UVO, C. B.; BRAVO, J. M.; COLLISCHONN, W.; ROCHA, H. R. 2010. Seasonal precipitation forecast based on artificial neural networks, Capítulo X, Computational Methods Applied to Agricultural Research: Advances and Applications. Org.: Prado, H. A.; Luiz, A. J. B.; Chaib, H., 26 p.

PHILIP N. S.; JOSEPH, K.B. 2003. A neural network tool for analysing trends in rainfall. Comput. Geosci. 29 (2), pp. 215-223

PINTO, E. J. A. 2005. Estudo de indicadores climáticos para a previsão de longo termo de vazões na bacia do Alto São Francisco. Tese (doutorado) – Universidade Federal de Minas Gerais, Departamento de Engenharia Sanitária e Ambiental, Departamento de Engenharia Hidráulica e Recursos Hídricos. 269f

QUIROZ, M. C. J. 2010. Indicadores Climaticos – Uma maneira para identificar la variabilidade climática a escala global. Elaboración de un boletín con información hidroclimática de los mares de México.

RAO, V.B., HADA, K. 1990. Characteristics of Rainfall over Brazil: Annual Variations and Connections with the Southern Oscillations. *Theor. Appl.Climatol.* 42, 81-91.

REBOITA M. S., GAN M. A., DA ROCHA R. P., AMBRIZZI T. 2010. Regimes de Precipitação na América do Sul: Uma Revisão Bibliográfica. *Revista Brasileira de Meteorologia*, 25(2):185-204.

ROPELEWSKI, C. F., HALPERT, M. S. 1987. Global and Regional Scale Precipitation Patterns Associated with the El-Nino Southern Oscillation. *Monthly Weather Review*, v. 115, p. 1606-1626.

ROPELEWSKI, C. F., HALPERT, M. S. 1989. Precipitation Patterns Associated with the High Index Phase of the Southern Oscillation. *Journal of Climate*, v. 2, p. 268-284.

SANTANA, F. M. 2013. Investigação da Influência da Variabilidade Natural do Clima no Comportamento das Ondas na Costa do Ceará. UFRJ/COPPE, Rio de Janeiro, XXI, 203 p

SANTOS, C. A. G.;FREIRE, P. K. M. M; TORRENCE, C. 2013. A Transformada Wavelet e sua Aplicação na Análise de Séries Hidrológicas. *Revista Brasileira de Recursos Hídricos* Volume 18 n.3 –Jul/Set,271-280.

SANTOS, I. A. & FRANCO, N. J. N. 2011. ANOMALIAS DA PRECIPITAÇÃO NO SUL DO BRASIL E AS TELECONEXÕES. XVII Congresso Brasileiro de Agrometeorologia. SESC Centro de Turismo de Guarapari, Guarapari – ES.

SANG, Y. F. 2013. A review on the applications of wavelet transform in hydrology time series analysis. *Atmos. Res.* 122, 8–15.

SCIENTEC – Associação para Desenvolvimento da Ciência e Tecnologia. 1997. Plano Diretor de Recursos Hídricos do Estado da Paraíba. Governo do Estado da Paraíba.

SILVA, N. I., SPATTI, D., & FLAUZINO, R. 2010. Redes Neurais Artificiais para engenharia e Ciências Aplicadas. São Paulo: Artliber Editora. 400p

SILVA, E.R.L.G. 2013. Associação da variabilidade climática dos oceanos com a vazão de rios da Região Norte do Brasil. 2013. Dissertação (Mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo. 182 p.

SILVERMAN, D., & DRACUP, J. A. 2000. Artificial Neural Networks and long-range precipitation prediction in California. *Journal of Applied Meteorology*, Volume39, 57–SHEFFIELD, J.; WOOD, E.F. 2008. Project changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. *Climate Dynamics*.v.31, p.79-105. 66.

SHUKLA, R.P. ET AL. 2011. Prediction of Indian summer monsoon rainfall using Niño indices: a neural network approach *Atmos. Res.*, 102 1-2, pp. 99–109

SILVEIRA, R. D.; SARTORI, M. G. B. 2010. Relação entre tipos de tempo, eventos de precipitação extrema e inundações no espaço urbano de São Sepé – RS. *Revista Brasileira de Climatologia*. Ano 6, Volume 7, Setembro, p.1-19.

SOUZA FILHO, F. A. 2003. Variabilidade e mudanças de clima no semi-árido. In: *Clima e Recursos Hídricos*. Org.: Tucci C. M. e Braga. B. Coleção ABRH Vol. 9, Porto Alegre, RS, pp. 65-77.

SOUSA, W. S.; SOUSA, F. A. S. 2010. Rede neural artificial aplicada à previsão de vazão da Bacia Hidrográfica do Rio Piancó. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.14, n.2, p.173–180.

SRIVASTAVA, G.; PANDA, S. N.; MONDAL, P.; LIU, J. 2010. Forecasting of rainfall using ocean-atmospheric indices with a fuzzy neural technique, *Journal of Hydrology*, Volume 395, Issues 3–4, 15, Pages 190-198,

TORRENCE, C.; COMPO, G. P. 1998. A practical guide to wavelet analysis. *Bulletin American Meteorology Soc.* 79, 61-78.

TRENBERTH, K. E. 1997: A definição de El Niño. *Buli. Amer. Met. Soc.* , 78, 2771-2777.

TUCCI, C. E. M; BRAGA, B. 2003. *Clima e Recursos Hídricos no Brasil*. Coleção ABRH, p. 348.

TUCCI, C. E. M. Modelos hidrológicos. 2007. Porto Alegre: Ed. Da Universidade/UFRGS/ABRH. 700p.

UVO, C. R. B.; REPELLI, C. A.; ZEBIAK, S.; KUSHNIR, Y. 1998. The relationship between tropical Pacific and Atlantic SST and Northeast Brazil monthly precipitation. *Journal of Climate*, v.11, p.551-562.

VAREJÃO-SILVA, M. A. 2006. *Meteorologia e Climatologia*. Versão Digital 2. Recife, 449p.

WILKS, D. S. 2006. *Statistical methods in the atmospheric sciences*. Burlington, Massachusetts: Academic Press, 630 p.

## ANEXOS

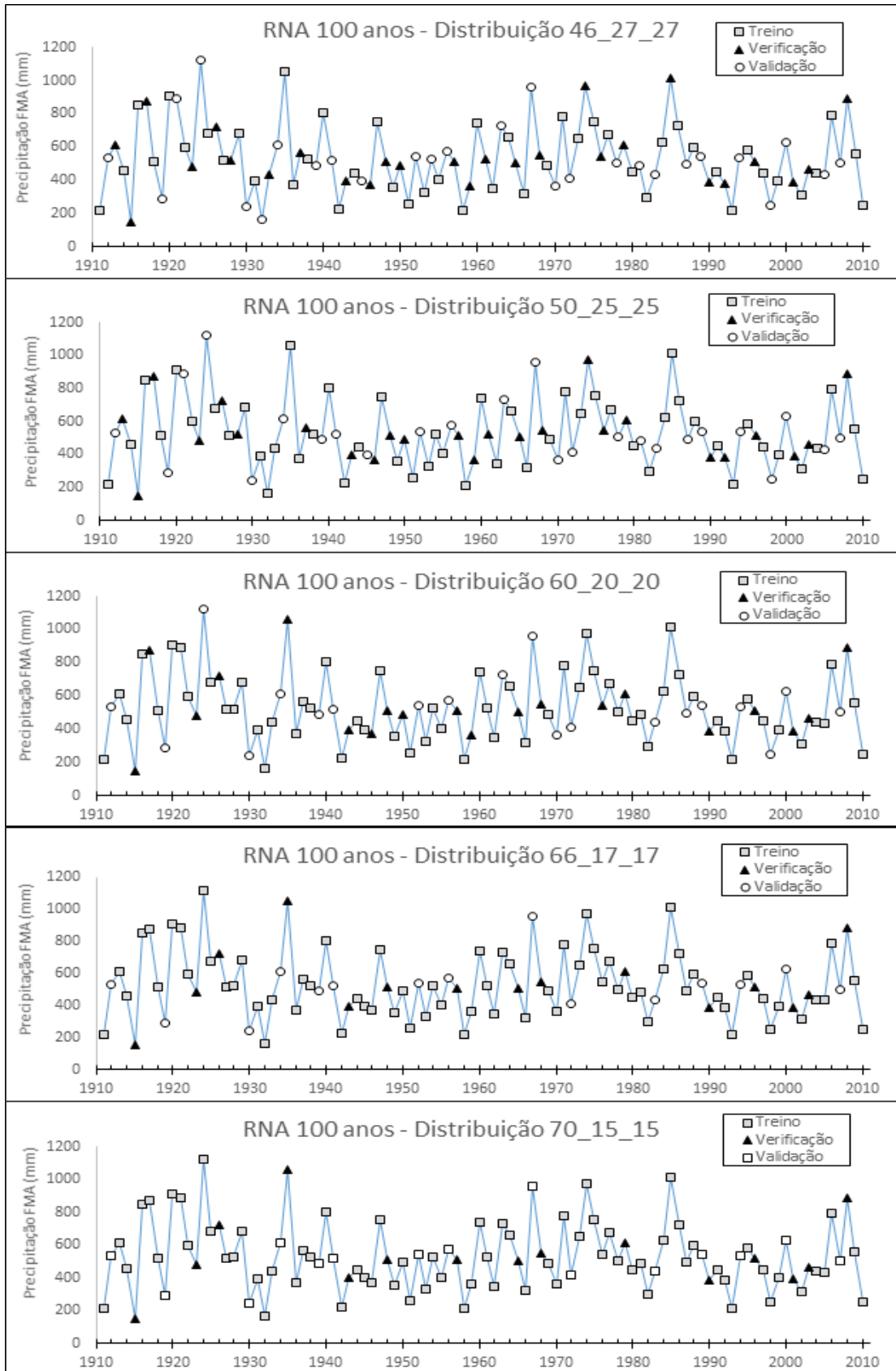


Figura 1 – Distribuição dos dados para os conjuntos de treinamento, validação e verificação para ao intervalo de 100 anos

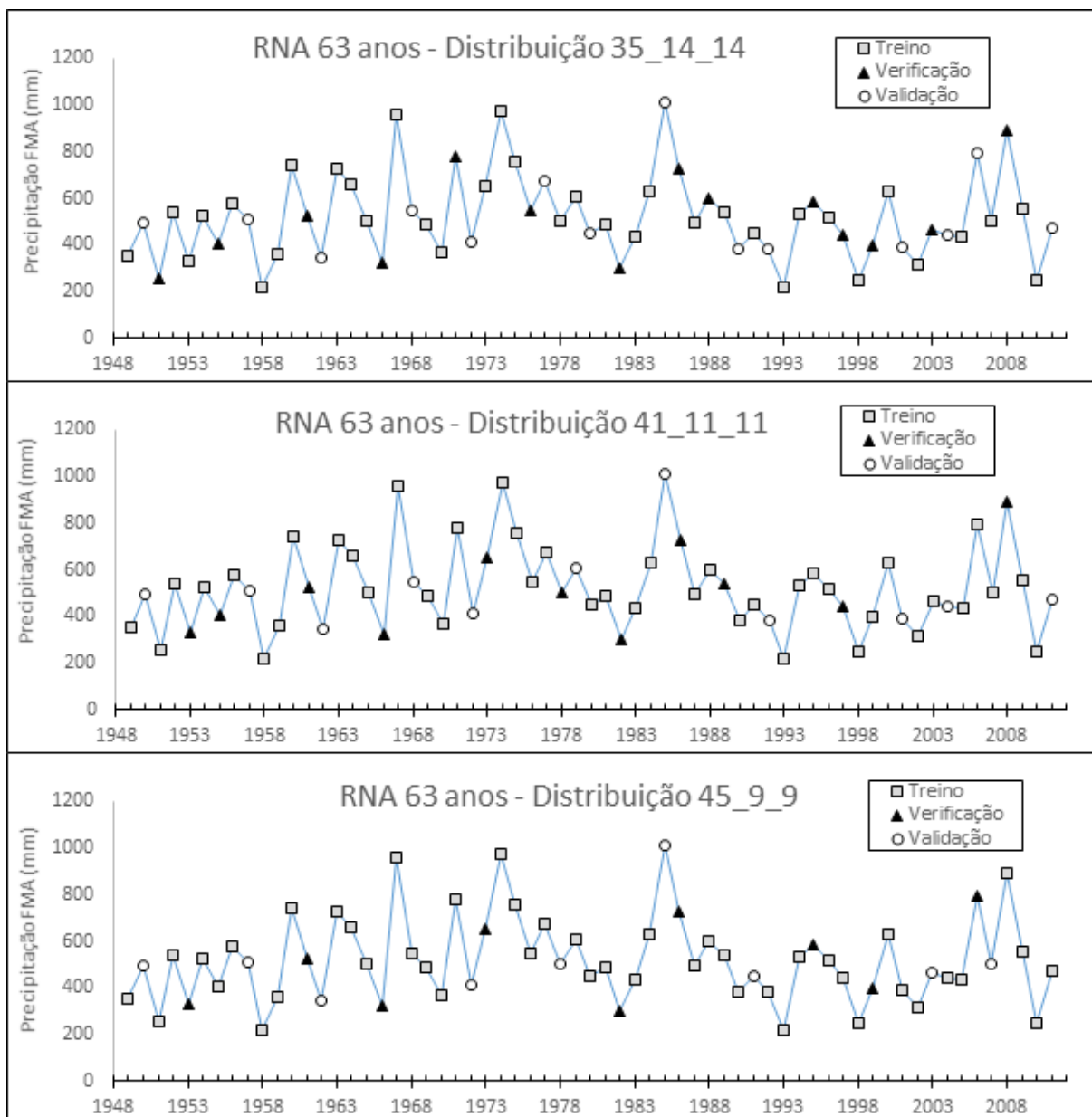


Figura 2 – Distribuição dos dados para os conjuntos de treinamento, validação e verificação para ao intervalo de 63 anos

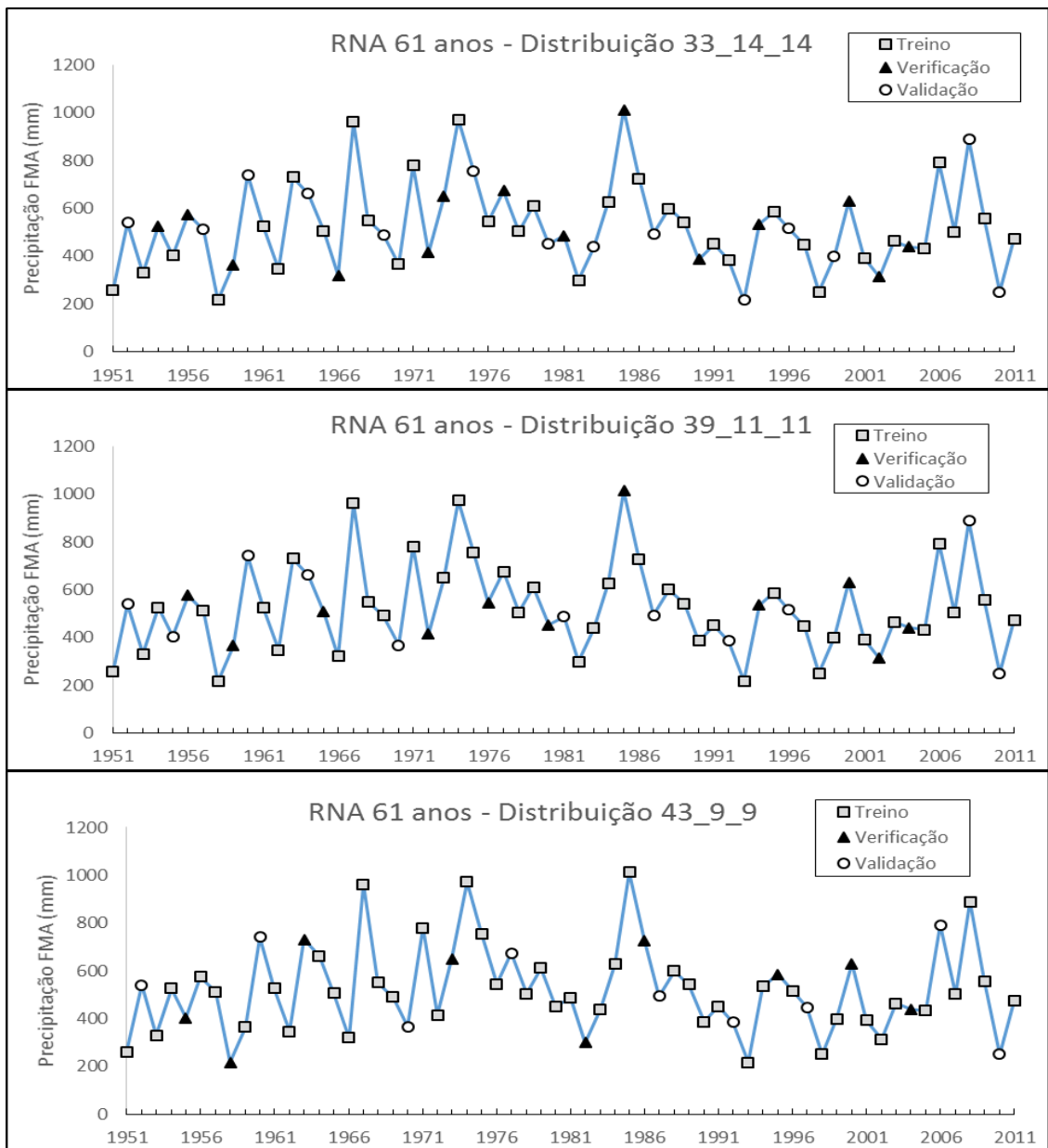


Figura 3 – Distribuição dos dados para os conjuntos de treinamento, validação e verificação para ao intervalo de 61 anos

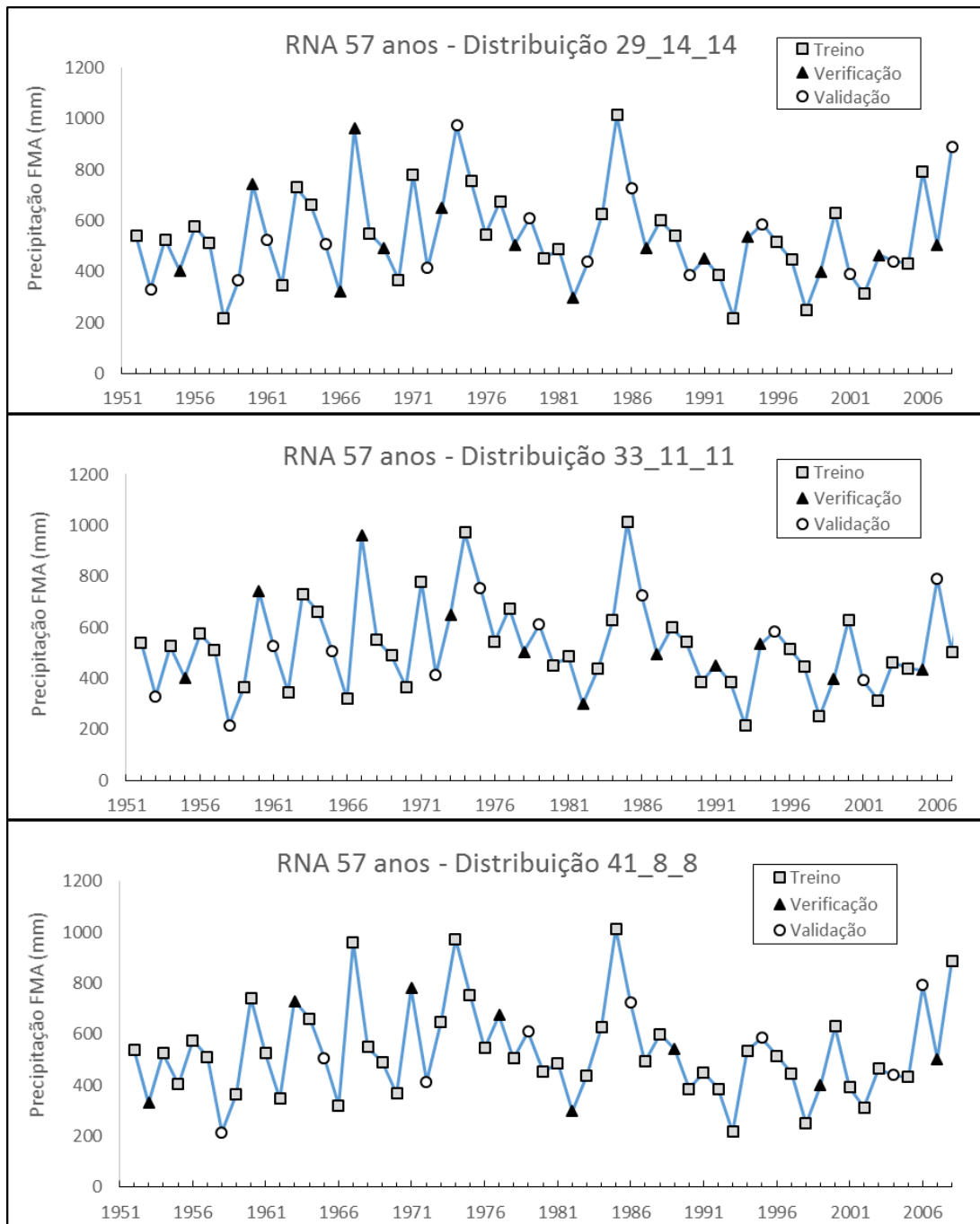


Figura 4 – Distribuição dos dados para os conjuntos de treinamento, validação e verificação para ao intervalo de 57 anos