



MODELO HÍBRIDO EMD-RNA DE PREVISÃO DE COTAS FLUVIAIS MÉDIAS
DIÁRIAS EM MANAUS A PARTIR DE INFORMAÇÕES CLIMÁTICAS DE
LARGA ESCALA

Jorge Augusto Pimentel Filho

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Oceânica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Oceânica.

Orientador: Claudio Freitas Neves

Rio de Janeiro
Setembro de 2019

MODELO HÍBRIDO EMD-RNA DE PREVISÃO DE COTAS FLUVIAIS MÉDIAS
DIÁRIAS EM MANAUS A PARTIR DE INFORMAÇÕES CLIMÁTICAS DE
LARGA ESCALA

Jorge Augusto Pimentel Filho

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM
CIÊNCIAS EM ENGENHARIA OCEÂNICA.

Examinada por:

Prof. Claudio Freitas Neves, Ph.D.

Prof. Paulo Cesar Colonna Rosman, Ph.D.

Prof. Luiz Pereira Calôba, Dr. Ing.

Prof. José Paulo Soares de Azevedo, Ph.D.

Prof. Isimar de Azevedo Santos, D.Sc.

Prof. Márcio Cataldi, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

SETEMBRO DE 2019

Pimentel Filho, Jorge Augusto

Modelo Híbrido EMD-RNA de Previsão de Cotas Fluviais Médias Diárias em Manaus a Partir de Informações Climáticas de Larga Escala/ Jorge Augusto Pimentel Filho. – Rio de Janeiro: UFRJ/COPPE, 2019.

XVI, 175 p.: il.; 29,7 cm.

Orientador: Claudio Freitas Neves

Tese (doutorado) – UFRJ/ COPPE/ Programa de Engenharia Oceânica, 2019.

Referências Bibliográficas: p.143-150.

1.Modelo Híbrido. 2. EMD-RNA. 3. Amazônia. I. Neves, Claudio Freitas. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Oceânica. III. Título.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

MODELO HÍBRIDO EMD-RNA DE PREVISÃO DE COTAS FLUVIAIS MÉDIAS
DIÁRIAS EM MANAUS A PARTIR DE INFORMAÇÕES CLIMÁTICAS DE
LARGA ESCALA

Jorge Augusto Pimentel Filho

Setembro/2019

Orientador: Claudio Freitas Neves

Programa: Engenharia Oceânica

Este trabalho comparou desempenhos entre modelos híbridos Decomposição Empírica Modal (EMD)-Rede Neural Artificial (RNA) e modelos RNAs convencionais, na previsão de níveis extremos anuais que penalizam rotineiramente a cidade de Manaus, na margem esquerda do rio Negro. Diversos estudos relataram ganhos significativos de desempenho usando modelos híbridos EMD-RNA. No entanto, a transitoriedade das funções intrínsecas modais e a queda de desempenho em previsão, comparando com os estágios de treinamento e de reanálise, foram desvantagens observadas, justificando a presente investigação. Primeiro, a dinâmica tropical foi revisada, revelando peculiaridades como a dualidade entre auto-afinidade e escalas hierárquicas. Para lidar com tais particularidades, aplicou-se a Análise de Flutuação Destendenciada, mostrando relações fractais entre as TSMs, a Radiação de Onda Longa na Amazônia e os níveis em Manaus. Adicionalmente, utilizou-se a Coerência Wavelet, obtendo-se informações multi-escala em espectros tempo-freqüência. As análises possibilitaram selecionar preditores e suscitaram uma discussão sobre as causas da intensificação dos eventos críticos em Manaus nas últimas décadas. Os ganhos com os modelos híbridos foram superiores a 40% e 35% nas previsões anuais de máximas e de mínimas, respectivamente, mas revelaram uma maior variância dos resíduos, encorajando novas investigações.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

FORECASTING DAILY MEAN WATER STAGES IN MANAUS VIA HYBRID
EMD-ANN MODEL USING LARGE SCALE CLIMATIC TIME SERIES

Jorge Augusto Pimentel Filho

September/2019

Advisor: Claudio Freitas Neves

Department: Ocean Engineering

This work compared the performance of Hybrid Empirical Mode Decomposition (EMD)-artificial neural network (ANN) models and conventional ANN models in forecasting annual extreme water stages that routinely penalizes Manaus, on the left bank of Negro River. Several studies reported significant performance gains using Hybrid EMD-ANN models. Nevertheless, the transitoriness of Intrinsic Mode Functions (IMFs) and the forecasting performance decline, comparing with training and hindcasting stages, where observed drawbacks, yielding the present investigation. First, an integrated oceanographic, meteorological and Amazon hydrological review is conducted, exploring their relations to define the predictors of the models. The subtleties of the tropical environment produce an atmospheric characterized by the duality between self-affinity and hierarchic scales. To deal with such peculiarities, the Detrended Fluctuation Analysis (DFA) was applied, revealing fractal relations between the SSTs, the Long Wave Radiation (OLR) on the Amazon and the levels in Manaus. In addition, Wavelet Coherence (WTC) was used to provide multi-scale insight via time-frequency spectra. Those analysis enabled the predictors selection and encouraged a discussion about the causes of intensification of events in Manaus in the last decades. Hybrid models gains where superior to 40% and 35% in high and low annual peaks forecasting, respectively, but revealed higher residuals variance, encouraging further investigations.

Sumário

1.	Introdução	1
2.	Motivação	5
3.	O Ambiente Tropical e a Ação Antrópica	9
3.1.	Circulação geral da Atmosfera	11
3.2.	El Niño-Oscilação Sul – ENSO e a Amazônia.....	14
3.3.	A TSM do Atlântico (dipolo), a Zona de Convergência Intertropical e a Amazônia	17
3.4.	Ondas Equatoriais de Leste	24
3.5.	Variabilidade Intrasazonal	25
3.6.	Degelo nos Andes e desmatamento - relações com o clima e hidrologia.....	28
3.7.	Análise dimensional nos trópicos.....	31
3.8.	Hierarquia ou auto-afinidade?	34
	Conceitos de hierarquia, auto similaridade e auto-afinidade	35
4.	Métodos de análise de dados usados	40
4.1.	Contextualização	41
4.2.	Wavelet e a coerência espectral.....	42
4.3.	A EMD e a HHT.....	45
4.4.	EMD.....	45
	Problema dos contornos.....	46
	Mixagem entre Modos.....	48
	Cruzamento de Envoltórias.....	51
4.5.	A DFA.....	52
4.6.	Análise de Componentes Principais – PCA e Regressão Linear Parcial de Mínimos Quadrados - PLS.....	54
5.	Redes Neurais	57
5.1.	Conceitos básicos	57
	Topologia	59
	Treinamento	60
	Função de ativação	62
5.2.	Parametrização, generalização e convergência	63
	Contextualização	64
	Norma	66
	Condicionamento de Sistemas Lineares	68
	Condicionamento e ranqueamento em RNAs <i>feedforward</i>	69

Multicolinearidade, ranqueamento e condicionamento.....	72
5.3. Estratégias de Generalização.....	78
Parada forçada.....	78
Regularização.....	80
6. Revisão Bibliográfica - Modelos Híbridos.....	84
7. Metodologia.....	88
7.1 Coerência Espectral Wavelet.....	89
7.2 Análise fractal nos dados.....	90
7.3 EMD.....	91
Problema dos Contornos.....	91
Mixagens entre modos.....	91
Cruzamento de Envoltórias.....	92
7.4 A PCA e a PLS.....	92
7.5 RNAs e cenários.....	93
7.6 Cenários comparativos entre modelos EMD-RNA e modelos convencionais RNA 94	
7.7 Cenários complementares de simulação.....	96
8. Dados.....	98
9. Aplicações e Resultados.....	104
9.1 Análise Fractal.....	105
9.2 Coerência espectral.....	108
9.3 Resultados comparativos dos Modelos Híbridos e modelos RNA convencionais.....	117
Seleção do melhor modelo convencional C1:.....	118
Escolha dos melhores modelos híbridos PLS, PCA e C2.....	120
9.4 Condicionamento e Generalização dos modelos.....	127
9.5 Resultados Comparativos entre Modelo Híbrido e Modelo de Galdino & Clarke 131	
10. Discussão.....	133
11. Conclusões e Recomendações.....	138
12. Bibliografia.....	143
I. Movimento Browniano Fractal e a auto-afinidade.....	151
Caminhar aleatório e o movimento Browniano.....	152
Movimento Browniano Fractal.....	153
II. Caracterização da dinâmica atmosférica tropical.....	155

a) Análise dimensional nos trópicos.....	155
Análise de dados observacionais.....	158
b) Hierarquia ou auto-afinidade?.....	160
Hierarquia	161
Quantificação da auto-afinidade	162
c) Ruído $1/f$ e eventos tipo pulso na variabilidade da atmosfera tropical	168
Caracterização $1/f$	169
Comportamento pulsante	171
Escala temporal mais longa	175

Lista de Figuras

- Figura 2-1. Mapa ilustrando a bacia amazônica, a bacia do rio Negro, a bacia do rio Solimões, a hidrografia principal e alguns postos fluviométricos de interesse..... 6
- Figura 2-2. (A) Curva-chave ajustada a partir de apenas 24 medições de descarga em Manaus; (B) Curva-chave em escala bi-logarítmica; (C) desvios relativos (%) das vazões estimadas ao longo das cotas; (D) Desvios relativos (%) das vazões estimadas ao longo do tempo..... 7
- Figura 3-1. Ilustração de como seria a circulação geral atmosférica em célula única: linhas tracejadas representam ventos superficiais dos polos para o equador; convergência na atmosfera inferior no equador representada por setas vermelhas; cor vermelha indica ar mais quente e azul, ar mais frio;..... 12
- Figura 3-2. Ilustração da circulação geral atmosférica com cinturões de baixa (verde) e de alta pressão (amarela) se alternando e com respectivas células de circulação, com tramos defletidos pelo efeito de Coriolis. 13
- Figura 3-3. Células de Walker durante episódio de El Niño (acima) e em situação normal (abaixo). 14
- Figura 3-4. Fases do ENSO, ilustrando as interações entre a atmosfera e o Pacífico Tropical Sul: Acima, La Niña, com espelhamento da circulação atmosférica no oceano, aprofundamento da termoclina a oeste e ressurgência a leste; ao centro, condição normal e abaixo, condição de El Niño, com espalhamento de água quente superficial pelo Pacífico..... 16
- Figura 3-5. Valores de OLR em janeiro (acima) e julho (abaixo) provenientes do ERBE – NASA *Earth Radiation Budget Experiment*, correspondente a médias mensais do período de 1985 a 1986. Adaptado de (ACKERMAN, 2013). 18
- Figura 3-6. Perfis idealizados de temperatura potencial em ordem decrescente de estabilidade (crescente de convecção), sendo A o mais estável (menor convecção) e F o menos estável (maior convecção). A linha vermelha indica os mínimos de θ_e – Adaptado de ASPLINDER (1976). 20
- Figura 3-7. Identidade pluviométrica da região amazônica e partes das regiões nordeste e centro-oeste do Brasil, sobreposta a isoietas médias anuais; Fonte: <http://www.snirh.gov.br/frst/>..... 22

Figura 3-8. Croquis ilustrando como seria um mosaico temporal-longitudinal de imagens de satélite com nuvens em região da ZCIT destacando deslocamento de ondas equatoriais de leste – seta vermelha destaca o deslocamento das nuvens.	24
Figura 3-9. Figura auto-afim com fator de escala horizontal $\delta_x = 4$ e fator de escala vertical $\delta_y = 2$	36
Figura 3-10. Estruturas fractais geradas com diferentes expoentes de Hurst H : para $H = 0,5$ a série não guarda memória temporal; na medida em que H aumenta, a série passa a possuir maior memória temporal; para $H > 1$, as séries não são mais estacionárias e são movimentos Brownianos fractais (MBFs), com $H_{MBF} = H_{RGF} - 1$	38
Figura 4-1. (Acima) Série NINO1+2 durante a primeira iteração da EMD com envoltórias, destacando-se os extremos dos contornos; (abaixo) detalhes dos trechos de contorno da série sem pontos extremos para a interpolação das envoltórias no início e no fim da série (linhas tracejadas vermelhas).	47
Figura 4-2. Exemplo de ocorrência de mixagem entre modos na EMD, com eixos verticais em graus Celsius e eixo horizontal em dias: as cinco primeiras IMFs de cima para baixo apresentaram mixagens entre modos vizinhos; a causa provável são efeitos intermitentes (pulsantes) em escala intrasazonal no Pacífico Oeste que influenciam a TSM no Pacífico Leste, mormente causados pela OMJ.	49
Figura 4-3. CEMD de série diária de níveis de Itapéua, no rio Solimões, com eixos verticais em cm e eixos horizontais em dias: (acima) máximos e mínimos destacando presença intermitente de frequências intrasazonais; (abaixo) detalhe mostrando as envoltórias (a de máximos é roxa e a de mínimos é verde) aderidas ao sinal (azul) graças ao uso dos pseudo-extremos ao longo da série.	50
Figura 4-4. Computação do VMQ das flutuações em relação à tendência linear (A), quadrática (B) e cúbica (C). A linha vermelha tracejada é o ajuste à tendência, com oito janelas de tamanho 1.000. A distância entre as linhas sólidas vermelhas e a tendência vermelha tracejada é de 1 VMQ. Adaptado de IHLEN (2012).	53
Figura 5-1. Esquema de um neurônio e de uma RNA <i>multi-layer Perceptron</i> – MLP. .	59
Figura 5-2. Ilustração de superfície de erro em função dos pesos w_1 e w_2 , com dois caminhos distintos de ajuste: círculos, iterações levando a um mínimo local; triângulos, iterações levando à solução – adaptado de SERGIU (2011).	61
Figura 5-3. Exemplos de funções de ativação: (a) linear; (b) linear com limiar nos extremos; (c) limiar; (d) sigmoidal.	63

Figura 5-4. Funções compostas $A(x, y)$ e $B(x, y)$ utilizando-se a TANSIG como função de ativação.	76
Figura 5-5. Funções compostas $C(x, y)$ e $D(x, y)$ utilizando-se a TANSIG como função de ativação.	77
Figura 5-6. Ilustração das distribuições normalizadas de probabilidades condicionais do conjunto de amostras X dados os modelos, em ordem de complexidade crescente M_1 , M_2 e M_3 . Em destaque, um exemplo X^* . Adaptada de BISHOP (1995).....	82
Figura 7-1. Fluxograma da modelagem híbrida proposta, a saber: Passo (1) composto da sincronização correlacional dos preditores NATL, SATL e NINO1+2 com o alvo e truncamento das séries nas datas correlatas ao dia da previsão do alvo; Passo (2) constitui-se da EMD dos vetores do passo (1), num exemplo no qual obtiveram-se 11 IMFs no total; Passo (3) é a análise pela PCA ou pela PLS.....	94
Figura 7-2. Nova estratégia de amostragem para a validação cruzada k -fold, com $k = 2$: eixo vertical em cm e eixo horizontal em dias; em laranja estão os pontos de treinamento e em azul, os dois subconjuntos de testes, pegando-se mínimos sorteados e os 59 pontos antecedentes.....	96
Figura 8-1. Mapa destacando regiões de interpolação de séries de interesse da pesquisa e bacias hidrográficas estudadas, além do posto HIDROWEB/ANA 14990000, cuja série de níveis médios diários foi alvo dos modelos testados.	99
Figura 8-2. Séries temporais de TSM do Atlântico e do Pacífico em °C e de OLR em W/m^2 – marcadores em laranja são série média mensal e curva azul é série interpolada diariamente. Destaque em NINO4 mostrando interpolação diária por spline cúbica.	101
Figura 8-3. Série de níveis médios diários em cm da estação HIDROWEB-ANA 1499000, destacando-se trecho alvo do estudo em verde e níveis alvos da modelagem com marcadores vermelhos, sendo 10 testes com níveis máximos e dez com níveis mínimos.....	102
Figura 8-4. Na sequência: série NATL, IMFs 1 e 2, IMF1+2, IMFs 3 a 4 e resíduo da EMD.....	103
Figura 8-5. Na sequência: série SATL, IMFs 1 e 2, IMF1+2, IMFs 3 a 5 e resíduo da EMD.....	103
Figura 8-6. Na sequência: série NINO1+2, IMFs 1 e 2, IMF1+2, IMFs 3 a 6 e resíduo da EMD.....	103

Figura 9-1. Escalogramas bilogaritmos das DFAs dos sinais de TSM, de OLR e de MAN destacando-se trechos auto-afins, respectivas retas ajustadas e derivadas: eixos horizontais em dias, eixos verticais nas respectivas unidades (TSM em °C, OLR em W/m ² , níveis em cm).....	107
Figura 9-2. Coerência Espectral Wavelet ao longo dos períodos em dias (eixos verticais a esquerda) entre MAN e (a) NATL, (b) SATL, (c) TROP, (d) NINO1+2, (e) NINO3, (f) NINO3.4, (g) NINO4, (h) OLR1, ; (i) OLR2, (j) OLR3; a escala inferior é dos quadrados da coerência; as setas indicam defasagens como frações angulares dos períodos dos eixos verticais em um círculo trigonométrico; sinais em fase têm ângulo zero e sinais anti-correlatos têm ângulo inicial igual a π ; os marcadores vermelhos apontam máximos (quadrados) e mínimos (triângulos) anuais de MAN entre os dez mais extremos desde 1902 (eixos centrais em m); os contornos delimitados por linhas pretas demarcam significância de 5% sobre o ruído vermelho.	110
Figura 9-3. Evolução temporal das parcelas de coerência Wavelet em períodos maiores que 128 dias (eixos verticais esquerdos) sobreposta a níveis máximos (quadrados) e mínimos (triângulos) em MAN entre os dez mais extremos desde 1902 (eixos verticais direitos em m): tendências de aumento da coerência da TSM e OLR com níveis acompanhada de ocorrência de picos extremos em MAN, especialmente níveis máximos anuais.	112
Figura 9-4. O mesmo que na Figura 9-2 para coerências entre as séries de OLR e NATL, SATL e NINO3.4.....	114
Figura 9-5. O mesmo da Figura 9-3 para coerências entre as séries OLR e NATL, SATL e NINO3.4.	115
Figura 9-6. Distribuição de resíduos dos modelos PLSN5, PCAN3 e C2N3 ao longo de 1.000 simulações nas dez previsões de máximas anuais: eixo horizontal em dias após truncamento das séries e eixo vertical em cm; retângulos azuis delimitam quartis centrais; linhas vermelhas são as medianas; linhas roxas tracejadas delimitam estimativas dos quartis centrais dos resíduos nos testes das RNAs; linhas azuis são as médias dos resíduos dos modelos híbridos e linhas cinzas são médias dos resíduos do modelo convencional C1N3.	124
Figura 9-7. Distribuição de resíduos dos modelos PLSN18, PCAN15 e C2N18 ao longo de 1.000 simulações nas dez previsões de mínimas anuais: eixo horizontal em dias após truncamento das séries e eixo vertical em cm; retângulos azuis delimitam quartis centrais; linhas vermelhas são as medianas; linhas roxas tracejadas delimitam	

estimativas dos quartis centrais dos resíduos nos testes das RNAs; linhas azuis são as médias dos resíduos nos modelos híbridos e linhas cinzas são médias dos resíduos do modelo convencional C1N3.	125
Figura 9-8. Cotogramas ao longo dos 60 dias de previsão de alguns testes, destacando-se os dias 15, 30, 45 e 60 das simulações – eixos verticais em cm.	127
Figura 9-9. Matrizes de correlação dos Jacobianos dos modelos $N = 3$ na previsão da mínima de 31/10/2002.	129
Figura 9-10. Desempenho do modelo C1N8 (esquerda) e de C2N8 (direita) até o momento da parada antecipada, seis épocas após a inflexão das curvas de desvios médios quadráticos dos conjuntos de validação.	130
Figura 9-11. Matrizes de correlação dos Jacobianos dos modelos C1N8 e C2N8 com a validação cruzada <i>k-fold</i>	131
Figura II-1. Mapa de contornos de escala de auto afinidade τ_s em seção transversal latitude-longitude com média móvel espacial de 5° longitudinais. Em destaque, regiões aproximadas dos dados de alimentação do modelo híbrido do presente estudo: TSM do Atlântico Norte em laranja ($5^\circ\text{-}20^\circ\text{N}$, $60^\circ\text{-}30^\circ\text{O}$); TSM do Atlântico Sul em marrom ($0^\circ\text{-}20^\circ\text{S}$, $30^\circ\text{O}\text{-}10^\circ\text{L}$); SOI em azul (entre Darwin, na Austrália e Tahiti); El Niño1+2($0\text{-}10^\circ\text{S}$, $90^\circ\text{-}80^\circ\text{O}$), El Niño 3 ($5^\circ\text{N}\text{-}5^\circ\text{S}$, $150^\circ\text{-}90^\circ\text{O}$), El Niño 4 ($5^\circ\text{N}\text{-}5^\circ\text{S}$, $160^\circ\text{L}\text{-}150^\circ\text{O}$) em vermelha; El Niño 3.4 ($5^\circ\text{N}\text{-}5^\circ\text{S}$, $170\text{-}120^\circ\text{O}$) em verde, com linha tracejada. Adaptado de YANO & NISHI (1989).	166
Figura II-2. Variabilidade como função da janela temporal para ruído branco (linha verde), ruído $1/f$ (azul) e ruído Browniano (vermelha). Adaptado de YANO <i>et al.</i> (2004).....	170
Figura II-3. Trecho de série temporal superficial do navio Kexue mostrando dois eventos de convecção caracterizados por pulsos nos instantes $t = 30.600$ min e $t = 30.750$ min nas séries de temperatura do ar T_{ar} ($^\circ\text{C}$), umidade relativa q (g/Kg), velocidade do vento $ v_x $ (m/s) e precipitação P (mm/h). Nota-se a relação de quedas repentinas na temperatura e umidade com pulsos nas séries de vento e de precipitação. Adaptado de YANO <i>et al.</i> (2004).	172
Figura II-4. Localização de um pulso pelo elemento principal (em vermelho) e pelos elementos adjacentes (em azul) no espectro Wavelet discretizado pela escala característica de duração dos eventos (eixo j) e pelo endereçamento do evento no tempo (eixo i). Adaptado de YANO <i>et al.</i> (2004).	173

Figura II-5. As mesmas séries da Figura II-3 recompostas com os eventos tipo pulso de escala 0,5-4 hs..... 174

Lista de Nomenclatura

ANA	Agência Nacional de Águas;
AMO	Oscilação Multidecadal do Atlântico;
AO	Oscilação do Ártico;
ATN	Atlântico Tropical Norte;
ATS	Atlântico Tropical Sul;
C1	Modelos de redes neurais convencionais;
C1MAX	Modelos de redes neurais convencionais para previsão de cotas máximas;
C1MIN	Modelos de redes neurais convencionais para previsão de cotas mínimas;
C2	Modelos híbridos EMD-RNA;
CEMD	Decomposição Empírica Modal Compacta;
CGB	Cobertura Máxima de Gelo no Mar Báltico;
COARE	Coupled Ocean-Atmosphere Response Experiment;
CPRM	Serviço Geológico do Brasil;
CWT	Transformada Wavelet Contínua;
DFA	Análise de Flutuação Destendenciada;
EEMD	Decomposição Empírica Modal em Grupo
EMD	Decomposição Empírica Modal;
ENSO	El Niño-Oscilação Sul;
ET	Evapotranspiração;
GPU	Unidade de Processamento Gráfico;
HAS	Análise Espectral de Hilbert;
HHT	Transformada de Hilbert-Huang;
HIDROWEB	Base de dados hidrológicos integrante do SNIRH;
ISC	Spline Cúbica;
k-cv	Validação Cruzada <i>k-fold</i> ;
LSA	<i>Large-Scale Array</i> ;
MAN	Níveis do rio Negro em Manaus;
MBF	Movimento Browniano Fractal;
MLP	<i>Back-Propagation Multi-Layer Perceptron</i> ;
MP	Modelo Probabilístico de Galdino & Clarke;
NAO	Oscilação do Atlântico Norte;

NATL	Temperatura da Superfície do Mar no Atlântico Norte (5-20°N) (60°-30°O);
NEB	Nordeste Brasileiro;
NINO1+2	Temperatura da Superfície do Mar no Pacífico (0-10°S) (90°-80°O);
NINO3	Temperatura da Superfície do Mar no Pacífico (5°N-5°S) (150° -90°O);
NINO3.4	Temperatura da Superfície do Mar no Pacífico (5°N-5°S) (170-120°O);
NINO4	Temperatura da Superfície do Mar no Pacífico (5°N-5°S) (160°L-150°O);
OLR	Radiação de Onda Longa;
<i>OLR1</i>	Radiação de Onda Longa (3°N-10,5°S) (80,6°-50,6° W);
OLR2	Radiação de Onda Longa (8,6° - 12,4° S) (65,6° - 60° W);
OLR3	Radiação de Onda Longa (2,9°N - 12,4° S) (75 ° - 60° W);
OMJ	Oscilação de Madden-Julian;
P	Precipitação;
PCA	Análise de Componentes Principais;
PDO	Oscilação Decadal do Pacífico;
PLS	Regressão Linear Parcial de Mínimos Quadrados;
PSM	Pressão na Superfície do Mar;
RNA	Rede Neural Artificial;
SAMS	Sistema de Monções Sul-Americano;
SATL	Temperatura da Superfície do Mar do Atlântico Sul (0°-20°S) (30°O-10°L);
SNIRH	Sistema Nacional de Informações sobre Recursos Hídricos;
SOI	Índice de Oscilação Sul;
STFT	Transformada de Fourier por Janelas;
TOGA	<i>Tropical Ocean and Global Atmosphere;</i>
TROP	Temperatura da Superfície do Mar do Atlântico Tropical (10°N-10°S) (0°-360°);
TSM	Temperatura da Superfície do Mar;
VIT	Variabilidade Intrazonal Tropical;
VMQ	Valor Médio Quadrático;
WTC	Coerência Wavelet;
ZACAS	Zona de Convergência Sul-Americana;
ZCA	Zona de Confluência dos Alísios;
ZCIT	Zona de Convergência Intertropical;

1. Introdução

Redes Neurais Artificiais – RNAs, são sistemas de processamento desvinculados de algoritmos convencionais, compostos por unidades nodais (neurônios) capazes de aprender experimentalmente através de “sinapses”, de forma análoga ao cérebro humano. Os nós são dispostos em camadas e interligados conferindo às RNAs um paralelismo inerente, que faz delas uma alternativa para solução de problemas complexos, envolvendo não-estacionariedade e não-linearidade, com alto desempenho.

Desde sua criação, as RNAs passaram por diversas fases, destacando-se a criação do algoritmo *back-propagation* (RUMELHART *et al.*, 1986), permitindo solução de problemas não lineares, e o surgimento recente das GPUs – *Graphic processing units*, permitindo desenvolvimento do paralelismo no processamento e o processamento de modelos mais complexos, colocando as RNAs no protagonismo da Inteligência Artificial.

As vantagens inerentes às RNAs são intimamente relacionadas a fatores como a arquitetura adotada para a rede (a quantidade de camadas intermediárias e dos seus respectivos nós), à escolha da função de ativação e à forma de computá-la à escolha do algoritmo de treinamento e das estratégias de generalização da RNA.

Na presente pesquisa, avaliou-se uma abordagem distinta com RNAs, capaz de modificar seus desempenhos independente dos fatores citados no parágrafo anterior: a adoção da Decomposição Empírica Modal – EMD, no pré-processamento dos vetores de entrada de RNAs *feedforward*, *back-propagation multi-layer Perceptron* – MLP. A EMD é parte integrante da HHT (HUANG *et al.*, 1998) e consiste num processo iterativo de peneiramento de sinais capaz de separar modos por faixas de frequência de forma empírica e desatrelada de bases matemáticas pré-definidas, fazendo com que os dados revelem por si suas características, inclusive não-linearidades e não-estacionariedades. Esses modos resultantes são chamados funções intrínsecas modais – IMFs.

Os modelos propostos, denominados híbridos EMD-RNA, são estudados na literatura há cerca de uma década, havendo trabalhos descrevendo ganhos expressivos de desempenho contrapostos por outros relatando problemas relacionados a transitoriedade das IMFs (mudam com o acréscimo de poucas amostras à série) ou relatando desempenhos bem

inferiores daqueles observados em reanálise e em treinamento em previsões feitas com os modelos híbridos. A investigação realizada explorou a capacidade de previsão de tais modelos, comparando seu desempenho com modelos convencionais RNA.

A problemática da previsão de níveis extremos máximos e mínimos anuais em Manaus foi tomada como motivação, no capítulo 2. Os modelos híbridos foram testados em previsões com 60 dias de antecipação dos picos de níveis anuais em Manaus. A cidade está situada na confluência dos rios Solimões e Negro, na bacia amazônica. Tendo experimentado um crescimento populacional expressivo nas últimas décadas, Manaus sofre as consequências da falta de infraestrutura, sobretudo de saneamento, associada às baixas declividades características da região, fatores que amplificam os impactos das cheias e das secas registradas no porto de Manaus. A série disponível de níveis médios diários do rio Negro, desde 1902, é um registro ininterrupto da dinâmica tropical processada pela bacia que abriga a maior floresta tropical do mundo. Tal série, apresentou nos últimos anos, uma evidente tendência de intensificação dos extremos, registrando-se recordes de máximas e mínimas cotas anuais.

Testar o desempenho dos modelos híbridos em um cenário com tendências de intensificação de eventos críticos foi uma oportunidade para comparar seu desempenho com outros modelos. Tirando proveito das características do ambiente tropical e das dimensões e da capacidade de regularização da bacia, foram utilizados como vetores de entrada dos modelos propostos sinais que traduzem teleconexões oceanográficas, atmosféricas e hidrológicas de larga escala. Para embasar a escolha dos preditores dos modelos, o capítulo 3 apresentou uma revisão bibliográfica sobre o ambiente tropical e a ação antrópica, começando pelo papel desestabilizador do calor latente na atmosfera tropical e suas relações com padrões de temperatura da superfície do mar TSM, de regiões específicas das bacias dos oceanos Atlântico e Pacífico. As relações da hidrologia e climatologia amazônicas com feições de larga escala dos trópicos, incluindo El Niño, o dipolo do Atlântico, a variabilidade interanual no Pacífico e a Zona de Convergência Intertropical – ZCIT, serão revisadas na literatura.

Buscou-se ainda, uma descrição na literatura, das particularidades e sutilezas da dinâmica tropical, começando pela sua análise dimensional, que evidencia incertezas entre o regime de ondas atmosféricas e o equilíbrio dinâmico entre o calor latente e a advecção vertical.

Adicionalmente, a dualidade entre auto-afinidade e separação de escalas foi revisada em estudos envolvendo dados atmosféricos tropicais. Os espectros cheios resultantes dessa dinâmica possuem características de ruídos $1/f$ e com alguns picos expressando sinais hierárquicos importantes e serão revisados na literatura.

As sutilezas do ambiente tropical requerem ferramentas apropriadas para explorar características da dinâmica tropical e justificar a utilização dos padrões de TSM como preditores de níveis em Manaus. Isso vai além de estabelecer relações observacionais entre a TSM das bacias oceânicas, o calor latente e a hidrologia na Amazônia. Assim, no Capítulo 4, realizou-se uma breve contextualização situando as ferramentas de análise de dados a serem utilizadas. A EMD foi introduzida, destacando-se seu caráter empírico, suas vantagens e seus principais problemas. A Análise Destendenciada de Flutuações, livre tradução para Detrended Fluctuation Analysis – DFA, foi revisada como alternativa para revelar relações fractais comuns aos dados de interesse, e destacar trechos auto-afins de variabilidade tropical e escalas marcantes dessa dinâmica. Além disso a Coerência Espectral Wavelet – WTC, foi discutida como ferramenta para a avaliação das relações fenomenológicas numa visão multi-escala dos espectros no domínio tempo-frequência. Foram apresentadas ainda duas técnicas de pré-processamento de dados pretendendo-se remediar possíveis problemas de ranqueamento e mal-condicionamento nas RNAs empregadas: a Análise de componentes principais – PCA e a Regressão Linear Parcial de Mínimos Quadrados – PLS.

Coube ainda uma revisão sobre RNAs no Capítulo 5, explorando conceitos básicos como topologia, treinamento e funções de ativação, além de aspectos mais profundos sobre parametrização, generalização e convergência, citando fatores que possam influenciar o desempenho dos modelos propostos e que ajudaram a definir estratégias para testá-los.

Os modelos híbridos EMD-RNA, presentes há cerca de uma década na literatura, foram revisados no Capítulo 6. Trabalhos, sobretudo com aplicações em hidrologia, que mostraram melhorias expressivas de desempenho foram confrontados com outros que apontam problemas do método como a transitoriedade das IMFs e os desempenhos significativamente inferiores de tais modelos em previsões, comparados aos desempenhos em reanálise.

No Capítulo 7 as metodologias para avaliação das relações fractais e fenomenológicas entre dados de TSM, de radiação de onda longa sobre a Amazônia – OLR e dos níveis em Manaus foram definidas, utilizando-se, respectivamente a DFA e a WTC. Também foram apresentadas metodologias para pré-processamento de dados de entrada dos modelos com a EMD, a PCA e a PLS. As estratégias para lidar com os problemas da EMD foram apresentadas. Os cenários de simulação foram definidos, discriminando as famílias de modelos híbridos a serem analisadas.

No Capítulo 8 foram apresentadas as séries de dados utilizadas na presente pesquisa.

O Capítulo 9 apresentou as aplicações e resultados da pesquisa. Foram apresentadas primeiro a análise fractal e a WTC dos dados. Como tais etapas justificam a utilização dos dados de TSM como preditores dos modelos, os resultados da DFA e da WTC foram comentados nesse capítulo, estendendo-se a discussão às possíveis causas das tendências presentes nas séries usadas. Na sequência, os resultados comparativos dos modelos híbridos e convencionais foram mostrados. A seleção do melhor modelo convencional e a escolha dos melhores modelos híbridos foi realizada. Adicionalmente, foram exibidos resultados destacando aspectos da generalização e condicionamento dos modelos híbridos. Os resultados da modelagem híbrida foram confrontados aos de um modelo probabilístico adaptado para Manaus.

A discussão dos resultados foi feita do Capítulo 10 e as conclusões e recomendações estão no Capítulo 11.

2. Motivação

A ocupação da região amazônica se iniciou em 1616, quando os colonizadores portugueses construíram um forte e assentaram a vila de Santa Maria de Belém, perto da boca do rio Amazonas. A partir desse ponto, iniciou-se o movimento rio acima, na direção Oeste, em terras formalmente pertencentes à Coroa Espanhola, de acordo com o Tratado de Tordesilhas.

O forte de São José do Rio Negro foi construído em 1669 na margem esquerda do rio Negro, perto de sua confluência com o rio Solimões, no coração da floresta Amazônica, 1.600 km a montante de Belém, com o propósito de estabelecimento das fronteiras do Império Português na América do Sul. Ao redor do forte, cresceu uma vila que viria a se tornar a cidade de Manaus em 1850, capital da Província. Acessível apenas por barcos, a cidade ganhou importância depois de 1870, com a ascensão do ciclo da borracha: o teatro de Manaus, um dos mais importantes do Brasil e cartão postal de Manaus, foi construído, e a energia elétrica foi estabelecida, bem no meio da floresta.

O monopólio da borracha, exercido por Brasil e Peru, se quebrou em 1913, quando sementes de seringueira foram contrabandeadas para a Ásia. A cidade entrou em um período longo de declínio econômico, permanecendo isolada do resto do país e permanecendo acessível apenas por navegação, até o estabelecimento da Zona Franca de Manaus, em 1970.

Após a criação da Zona Franca de Manaus, a população cresceu de 300.000, em 1970, para 1,4 milhões de habitantes, em 2010, atingindo uma população estimada de 2,2 milhões de habitantes em 2018. Manaus é hoje a sétima cidade mais populosa do Brasil, porém com uma renda per capita mensal de R\$ 850,00, bem abaixo da média brasileira de R\$1.268,00, em 2017. Em 50 anos, a população, portanto, cresceu sete vezes, levando a cidade à condição de metrópole, com todos os problemas sociais, econômicos e de infraestrutura relacionados.

A cidade é rotineiramente atingida pela cheia anual do rio Negro, causando mais problemas e prejuízos econômicos aos habitantes, especialmente aos ribeirinhos, não só do rio Negro, mas também dos igarapés, normalmente sem acesso a saneamento básico.

Extensas áreas são atingidas devido às baixas declividades topográficas. Por outro lado, secas severas causam distúrbios no porto de Manaus, suspendendo operações e impactando a economia local, afetando o transporte regional e isolando comunidades, LOPES *et al.* (2016). Durante os eventos de seca, a baixa umidade do solo reduz a evapotranspiração, favorecendo a ocorrência de incêndios florestais naturais ou criminosos, convertendo a floresta de sumidouro em fonte de emissões de gases do efeito-estufa.

Por estar na confluência entre os rios Negro e Solimões, a série MAN reflete a variabilidade hidrológica de ambas bacias hidrográficas. A Figura 2-1 ilustra a bacia a montante de MAN e a localização da estação código HIDROWEB/ANA 14990000. As estações Moura (14840000), no rio Negro, e Itapéua (13150000) e Manacapuru (14100000), no Solimões são também destacadas.

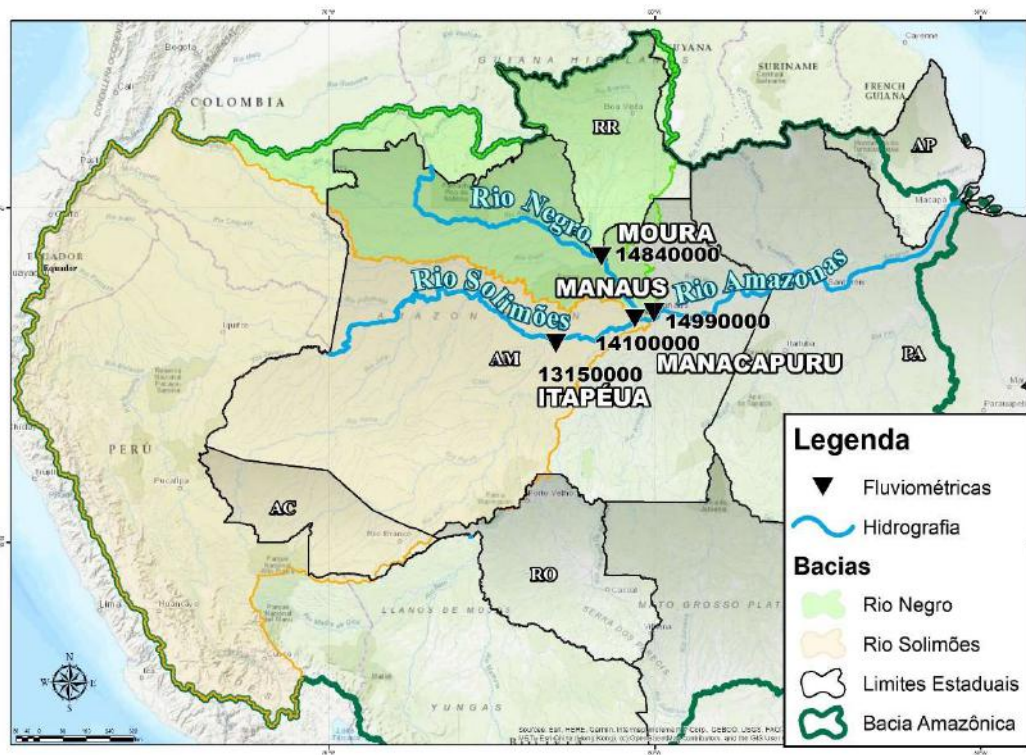


Figura 2-1. Mapa ilustrando a bacia amazônica, a bacia do rio Negro, a bacia do rio Solimões, a hidrografia principal e alguns postos fluviométricos de interesse.

Com a intensificação e recorrência dos níveis extremos anuais em Manaus, as tendências da série podem impactar a previsibilidade dos modelos existentes. No capítulo 9 (Tabela 9-2) foram elencados eventos extremos de níveis máximos e mínimos anuais de MAN no

período do presente estudo, revelando a ocorrência de sete dos maiores níveis máximos anuais desde 1989 e cinco níveis mínimos com probabilidade de ocorrência menor que 10%, desde 1995.

A extensão do histórico de níveis MAN e da bacia de drenagem a montante, de dimensões sinópticas, fazem dessa série um importante registro das interações entre a dinâmica atmosférica tropical e a hidrologia da bacia. Porém, devido às baixas declividades na região e à proximidade do posto de Manaus em relação à confluência entre o Negro e o Solimões, os efeitos de remanso acrescentam incertezas elevadas nas estimativas de vazões feitas relacionando níveis e medições de descargas realizadas. A Figura 2-2 ilustra a tentativa de ajuste de relação cota-descarga (curva-chave) em Manaus, evidenciando elevados desvios relativos das vazões estimadas em relação às medições. Por esse motivo, foram feitas poucas medições (apenas 24 no ajuste mostrado) naquele posto.

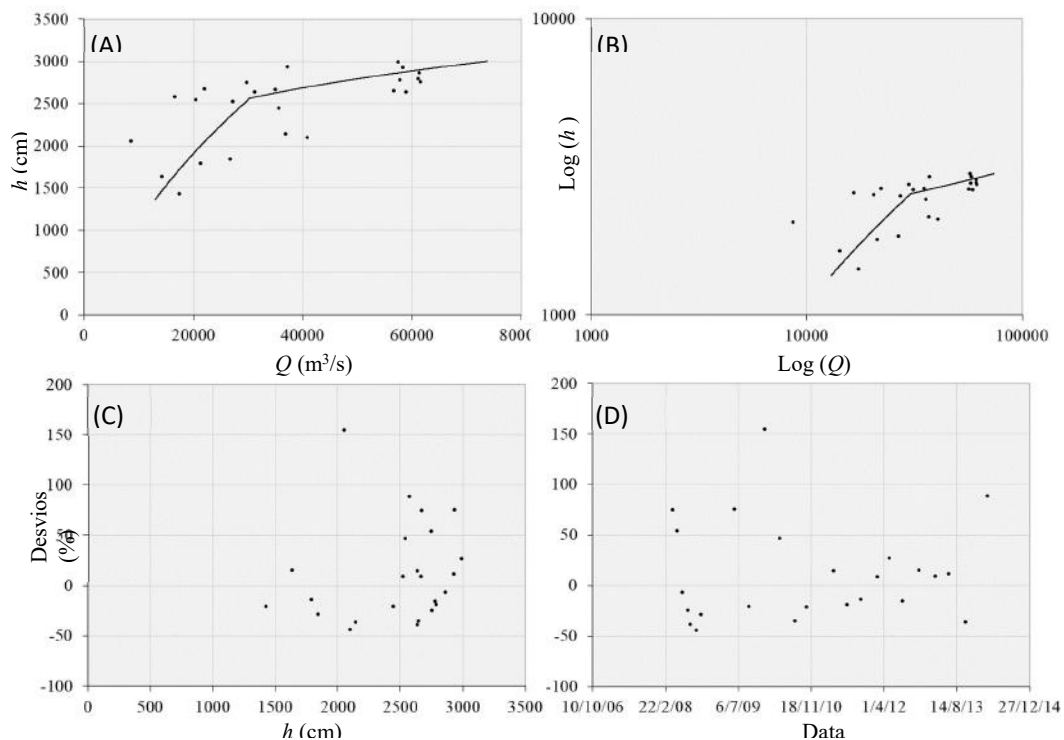


Figura 2-2. (A) Curva-chave ajustada a partir de apenas 24 medições de descarga em Manaus; (B) Curva-chave em escala bi-logarítmica; (C) desvios relativos (%) das vazões estimadas ao longo das cotas; (D) Desvios relativos (%) das vazões estimadas ao longo do tempo.

Outra forma de caracterizar o efeito de remanso na bacia é observando as relações fenomenológicas entre os postos da Figura 2-1. Manaus possui hidrograma defasado em 11 dias em relação a Itapéua, a aproximadamente 390 km a montante, no rio Solimões.

Já Moura, mesmo estando cerca de 250 km a montante no rio Negro, tem hidrograma defasado de 10 dias em relação a Manaus. Isso demonstra o remanso no rio Negro causado pelo Solimões.

Assim, apesar de uma abordagem em termos de vazões pudesse ser mais apropriada para descrever o problema, os desvios nas séries assim geradas comprometeriam a análise, sendo preferível fazer uso da extensa série de Manaus em vez de buscar postos com séries de vazões geradas, com incertezas inerentes ao processo de geração.

Existem diversos modelos na literatura para a previsão dos picos anuais de cheia em Manaus. Eles tiram proveito da série centenária de níveis médios diários do rio Negro na estação fluviométrica do porto de Manaus – MAN, bem como da capacidade de regularização da enorme bacia, que se traduz em uma série com grande auto-correlação, usualmente com uma subida e uma descida lentas. Uma revisão sobre tais modelos é apresentada em PIMENTEL *et al.* (2011).

Entender o papel integrado dos oceanos, da atmosfera e da floresta na dinâmica tropical pode ser o primeiro passo para explicar a intensificação dos eventos críticos em Manaus e, a partir de tal avaliação, modelar e prever padrões futuros, considerando mudanças climáticas de escalas planetárias. A presente pesquisa explorou alguns dos principais mecanismos forçantes do clima e da hidrologia amazônicos: o papel do calor latente e da umidade na dinâmica tropical, a TSM do Atlântico e do Pacífico, a presença de eventos pulsantes na composição da variabilidade tropical (ao contrário de uma atmosfera dominada por ondas), a dualidade entre auto-afinidade e hierarquia de escalas nos trópicos e os impactos do uso do solo na Amazônia.

Depois de melhor explorar tais aspectos, os modelos híbridos EMD-RNA foram testados para a previsão de níveis em Manaus. Seu desempenho foi comparado com os modelos RNA convencionais, para avaliar os ganhos com a EMD, e com um modelo probabilístico de Galdino & Clarke – MP, adaptado por PIMENTEL *et al.* (2011) para se ter ideia da previsibilidade proporcionada pelos modelos híbridos no caso de Manaus.

3. O Ambiente Tropical e a Ação Antrópica

A meteorologia tropical engloba o estudo de fenômenos atmosféricos em escalas espaço-temporais diversas, como tempestades tropicais, relâmpagos, ciclones tropicais, monções, El Niño, ondas de Rossby equatoriais, oscilações de Madden-Julian – OMJ, jatos de leste, vórtices ciclônicos de altos níveis, etc. Ocorrem interações importantes para o estudo do tempo e do clima tropicais, muitas desconhecidas, entre as várias escalas de circulação, desde a convecção de cúmulos tropicais, passando pela mesoescala até a larga escala.

A circulação atmosférica nos trópicos guarda diferenças marcantes em relação à região extratropical. Em latitudes mais altas, a energia potencial zonal associada ao gradiente latitudinal de temperatura é a principal fonte de perturbações de larga escala, sendo o calor latente e o aquecimento por radiação fontes secundárias de energia. Nos trópicos, a estocagem de energia potencial é pequena devido aos gradientes de temperatura muito baixos. A liberação de calor latente parece ser a fonte primária para perturbações originadas na zona equatorial. Ela ocorre em grande parte no interior de sistemas de nuvens convectivas, que estão geralmente imersos em sistemas de larga escala. Além de infringir uma resposta na circulação atmosférica local, o calor diabático associado à precipitação tropical pode induzir a respostas remotas através de excitação de ondas equatoriais.

Sob fraco efeito de Coriolis em baixas latitudes, não há formulação teórica simples análoga à quase-geostrofia para entender a circulação nos trópicos. Em vez disso, faz-se necessário estudar as interações entre as escalas locais de convecção nas torres de calor, e as circulações de mesoescala e de larga escala. Adicionalmente, a distribuição de calor latente está intimamente relacionada aos padrões de TSM, que por sua vez, sofrem influência da circulação atmosférica (HOLTON, 2004).

Uma descrição mais completa da atmosfera tropical englobaria alguns conceitos:

- o balanço de energia nos trópicos, caracterizando a maior quantidade de energia incidente na região tropical, a menor variabilidade sazonal do fluxo de energia naquela região, além das parcelas que são absorvidas pela atmosfera ou pelas superfícies, sendo reemitidas como radiação de ondas longas, calor sensível

(condução e convecção seca) e calor latente (convecção de umidade), este último processo, o mais significativo e de interesse para a presente pesquisa;

- O protagonismo do calor latente na troca de energia entre a superfície e a atmosfera, pois ao ser transferida da superfície para a atmosfera pelas nuvens de convecção profunda, a energia força o movimento ascendente e as feições de movimento de larga escala da circulação geral dos trópicos;
- Ocupando boa parte da região equatorial, os oceanos fornecem parcela significativa da umidade advectada. A extensão e localização da convecção depende de diversos fatores como o fluxo de calor sensível, capaz de alterar a estabilidade da atmosfera (LAING & JENNI-LOUISE, 2016);
- A umidade, que é propriedade abundante na atmosfera tropical, tem efeito de desestabilização vertical, possibilitando convecção profunda e formação de nuvens, pois ao ascender, o ar úmido resfria adiabaticamente e se o resfriamento for suficiente para saturar a parcela de ar, formam-se nuvens pela condensação do vapor, liberando calor latente que aquece a parcela de ar, aumentando sua fluabilidade, favorecendo assim a instabilidade.
- Esse fenômeno é denominado convecção úmida e para comparar seus critérios de estabilidade com os presentes em uma atmosfera seca, conceitos correlatos tais como a taxa de perda adiabática, a temperatura potencial, a taxa de perda adiabática saturada e a temperatura potencial equivalente seriam discutidos;

Tais conceitos ajudariam a clarificar o papel da TSM em certas localidades do Atlântico e do Pacífico no clima tropical e, conseqüentemente, na hidrologia em Manaus. Boas revisões são feitas em HOLTON (2004) e WALLACE & HOBBS (2006). Optou-se por uma abordagem mais focada na caracterização das relações fenomenológicas entre os dados de TSM de tais oceanos e a série alvo a ser estudada.

A revisão apresentada a seguir focou nas feições da região tropical responsáveis pela conexão entre as TSM e Manaus e nas mudanças em curso na ocupação da bacia pelo homem que poderiam alterar tais relações. Assim, primeiramente, revisaram-se as feições gerais da atmosfera tropical. Depois, descreveram-se El Niño e o Dipolo do Atlântico e citaram-se obras que evidenciam suas relações com a climatologia amazônica e com a vazão do rio Amazonas. A ZCIT, as ondas Equatoriais e a variabilidade Intrasazonal

foram revisadas. Na sequência, revisaram-se estudos sobre relações entre o degelo dos Andes, o desmatamento na Amazônia e o clima e hidrologia da bacia, buscando dar suporte à identificação de possíveis alterações nas relações entre oceanos, clima e níveis em Manaus. Por fim, apresentaram-se uma revisão da análise dimensional, da dualidade entre hierarquia e auto-afinidade e do papel dos eventos pulsantes na construção de padrões auto-afins nos trópicos.

3.1. Circulação geral da Atmosfera

A importância da convecção profunda na introdução de energia na atmosfera tropical e em sua desestabilização foi ressaltada. Para entender as relações entre tais fenômenos e a série histórica de níveis em Manaus, deve-se pensar em como a energia advinda dos oceanos é transportada para a região da bacia amazônica e como influencia o clima de chuvas da região. Obviamente, as chuvas que alimentam a bacia a montante do posto em Manaus, são resultados da ação conjunta de fenômenos de diversas escalas espaciais e temporais, algumas locais, outras remotas.

Segundo CAVALCANTI *et al.* (2009), essas interações com fenômenos distantes, também chamadas de teleconexões, baseia-se em três processos principais: as ondas atmosféricas, a continuidade de massa e a mudança de fase da água. O papel do calor latente na condensação dentro das torres de calor foi citado. Mas é necessário entender como a circulação geral nos trópicos influencia as conexões entre a TSM e a hidrologia em Manaus. As feições gerais marcantes da circulação nos trópicos são descritas a seguir.

Com o excesso de energia nos trópicos em relação aos polos, o esperado seriam padrões de circulação gerando células de troca direta de energia entre as zonas mais aquecidas, com correntes ascendentes nos trópicos, e as mais frias, com correntes descendentes nos polos. Nesse caso, as feições gerais da circulação atmosférica seriam bem representadas pela Figura 3-1.

Porém, ao deslocar-se, a parcela de ar é defletida devido à conservação do momentum angular, pelo efeito de Coriolis. As deflexões compensam as diferenças de velocidades superficiais da Terra ao longo da trajetória do ar em movimento, devidas às variações no raio de rotação.

Como a intensidade do efeito de Coriolis aumenta com a latitude, as massas de ar que ascenderam na região equatorial e estão se deslocando para os polos na Figura 3-1, são defletidas para a esquerda no hemisfério sul e para direita no hemisfério norte. Assim, ao atingir as latitudes médias, o ar na alta atmosfera já não ruma para os polos, mas sim para leste nos trópicos. De forma semelhante, os tramos da circulação que rumam dos polos para o equador, junto à superfície, são defletidos e giram, formando ventos fortes de leste, fazendo da Antártida o local com os ventos mais fortes do planeta.

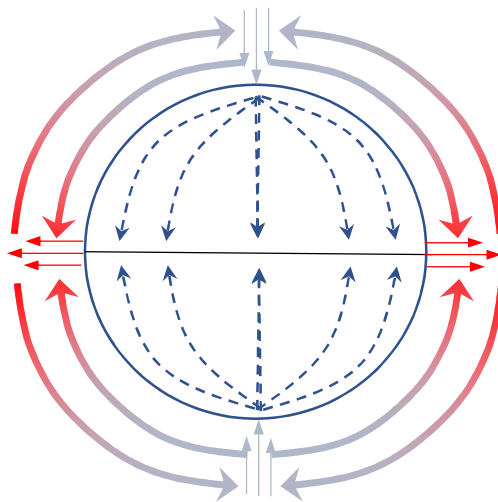


Figura 3-1. Ilustração de como seria a circulação geral atmosférica em célula única: linhas tracejadas representam ventos superficiais dos polos para o equador; convergência na atmosfera inferior no equador representada por setas vermelhas; cor vermelha indica ar mais quente e azul, ar mais frio;

Adicionalmente, conforme ruma para os polos, forma-se um cinturão de alta pressão ao redor dos 30° norte e sul, com o ar descendente. Ele é comprimido, torna-se mais quente e mais capaz de evaporar água. Junto à superfície, o ar diverge para norte e para o sul. Muitos desertos localizam-se sob influência desse ar, ao longo dos 30° de latitude.

O ar movendo-se de volta para o equador, na parte inferior dessa célula de circulação, também é defletido, formando os ventos alísios, soprando sempre de leste. Nas proximidades do equador, há uma zona de baixa pressão, para a qual esses ventos convergem e o ar ascende. Trata-se da ZCIT.

A circulação geral da atmosfera é constituída por ventos e sistemas de pressões que formam células de circulação como a descrita na região tropical. Ao todo, são sete

cinturões de altas e baixas pressões alternados com três células de circulação de vento entre eles. A Figura 3-2 ilustra os cinturões de pressão intercalados e as células de circulação resultantes. De maior importância para a presente pesquisa é a célula localizada na região tropical, denominada Célula de Hadley.

Até agora, o modelo de circulação geral apresentado não considera os efeitos da presença dos continentes e supõe o sol incidindo a pino no equador, semelhante ao modelo conceitual “aqua-planet”, usado em simulações numéricas simplificadas da circulação geral atmosférica. Boas revisões sobre o tema são recorrentes na literatura (WALLACE & HOBBS, 2006) (MARSHAL & PLUMB, 2008).

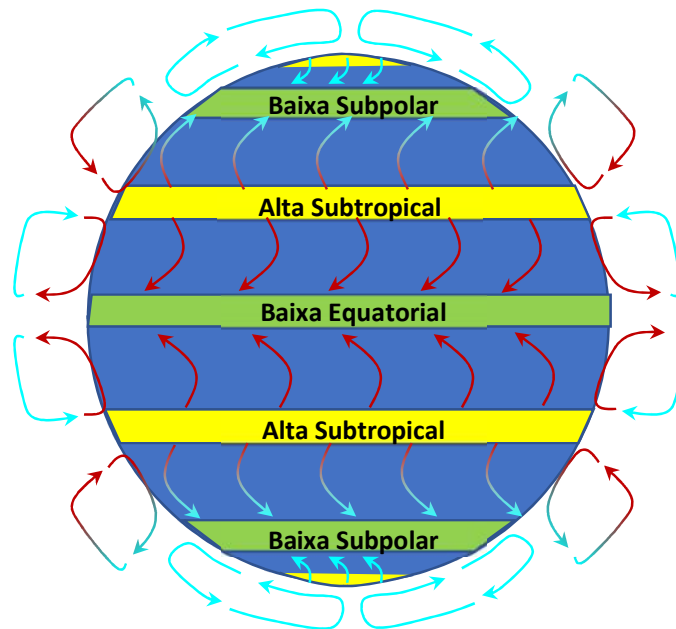


Figura 3-2. Ilustração da circulação geral atmosférica com cinturões de baixa (verde) e de alta pressão (amarela) se alternando e com respectivas células de circulação, com tramos defletidos pelo efeito de Coriolis.

Outra importante feição da circulação geral que afeta sobremaneira os trópicos são as células de Walker. São células verticais-zonais de circulação causadas pelas diferenças de temperatura entre oceanos e continentes e marcadas por ascendências, acima dos continentes e na porção oeste dos oceanos, e por subsidências, acima das partes orientais dos oceanos. A parte superficial das células correspondem aos alísios. O escoamento em baixos níveis tende a ser de ventos leste, mas com o efeito de Coriolis negligenciável em

baixas latitudes, são as diferenças de temperaturas superficiais entre oceanos e continentes e o movimento da célula de Hadley que determinam esse movimento.

Essa circulação figura como uma das principais conexões entre a circulação atmosférica e a circulação oceânica nos trópicos. A variabilidade de tais células por influência do El Niño-Oscilação Sul (ENSO) modula inclusive padrões de pressão na superfície do mar – PSM, e de TSM no Atlântico Tropical e, conseqüentemente, modifica as chuvas na região de interesse da presente pesquisa. Configura-se assim, uma teleconexão entre as bacias do Pacífico e do Atlântico. As células de Walker são ilustradas na Figura 3-3, em duas situações distintas: a primeira sob atuação de um episódio de El Niño; a segunda, em uma situação normal.

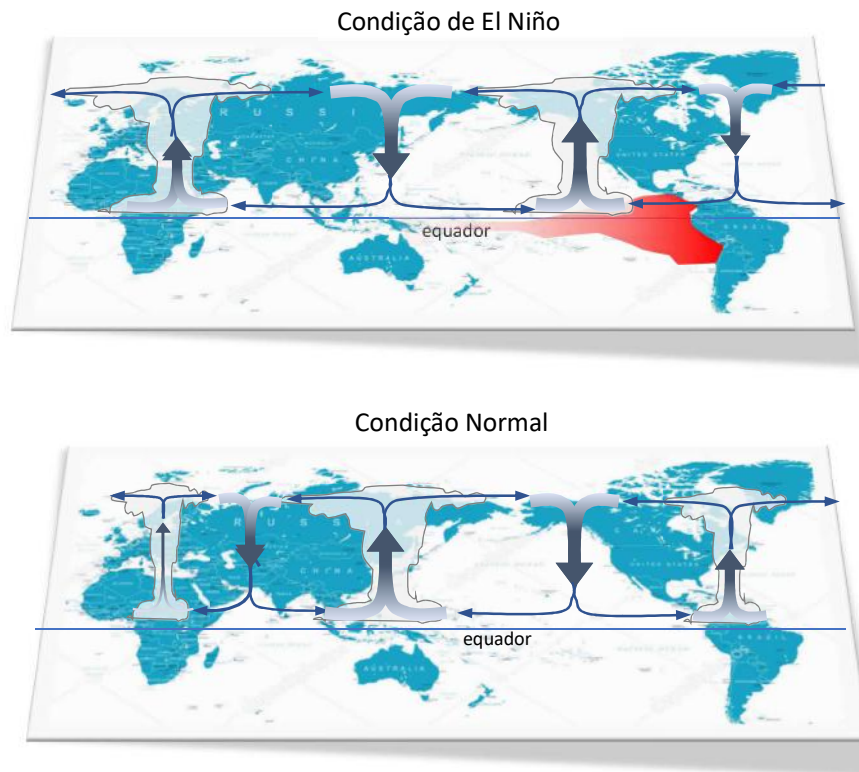


Figura 3-3. Células de Walker durante episódio de El Niño (acima) e em situação normal (abaixo).

3.2. El Niño-Oscilação Sul – ENSO e a Amazônia

A conexão entre a circulação de Walker e a dinâmica no Pacífico é explicada pelo ENSO. Trata-se de ciclo natural no Pacífico, envolvendo padrões de temperatura, ventos e nuvens, capaz de influenciar o clima global, causando eventos extremos como secas ou

chuvas excessivas. Assim, o ENSO é a principal fonte de variabilidade climática interanual global (CAVALCANTI *et al.*, 2009).

O ENSO oscila entre três estágios: La Niña, neutro e El Niño. Algumas vezes, a mesma fase pode ocorrer por dois ou mais anos consecutivos, sendo que os episódios ocorrem em intervalos irregulares de dois a sete anos (VILLWOCK, 2002). Formalmente, El Niño – EN e La Niña – LN, são caracterizados por anomalias de TSM na região central do Pacífico Tropical superiores a +0,5°C (EN) ou inferiores a -0,5°C (LN), respectivamente. Quando as anomalias duram menos que cinco meses, são classificadas como condições de EN ou LN. Persistindo a anomalia, tem-se um episódio de EN ou LN.

A condição de La Niña se estabelece quando a circulação oceânica espelha a célula de Walker sobre o Pacífico Sul, como ilustrado na Figura 3-4. Assim, na atmosfera, no Pacífico Oeste observam-se baixas pressões, umidade e ar quente, com correntes ascendentes, ventos leste junto à superfície do mar, ventos oeste em níveis elevados e correntes descendentes com alta pressão, ventos secos e mais frios na costa leste do Pacífico Sul.

Em condições normais, a célula de Walker é capaz de forçar a circulação do Pacífico, empilhando água no oeste, tornando a termoclina mais rasa no leste. Assim, as águas profundas e geladas afloram no Pacífico Leste, com subsidência de águas quentes na costa oeste. De fato, essa fase normal ocorre de forma semelhante no Atlântico.

Já o EN, geralmente ocorre quando há um enfraquecimento da célula de Walker, observando-se queda de pressão sobre o Taiti e sobre o Pacífico central e leste e elevação da pressão sobre o Oceano Índico, Austrália e Indonésia. Há enfraquecimento dos alísios, podendo ocorrer até inversão dos ventos superficiais, colapsando a célula de Walker.

Com o enfraquecimento dos alísios, a inclinação da termoclina no Pacífico não mais se sustenta e uma língua de água quente se espalha do Pacífico oeste e do Índico para o Pacífico leste, causando chuvas em áreas tipicamente secas da costa oeste da América do Sul e secas a leste do Pacífico.

As oscilações características do Pacífico, cuja bacia possui maior dimensão zonal, diferenciam-se das ocorridas na bacia do Atlântico, meridionalmente disposta. Apesar de tais diferenças, essas oscilações estão relacionadas.

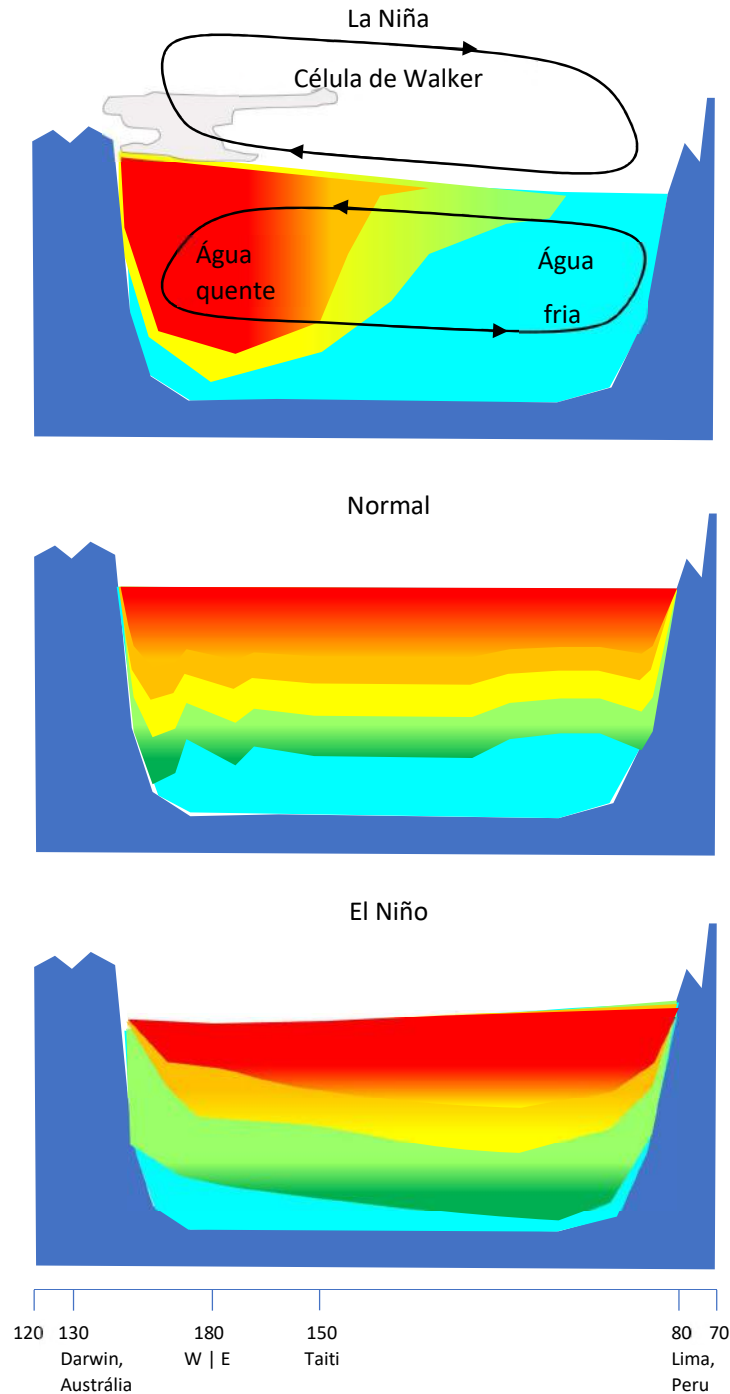


Figura 3-4. Fases do ENSO, ilustrando as interações entre a atmosfera e o Pacífico Tropical Sul: Acima, La Niña, com espelhamento da circulação atmosférica no oceano, aprofundamento da termoclina a oeste e ressurgência a leste; ao centro, condição normal e abaixo, condição de El Niño, com espalhamento de água quente superficial pelo Pacífico.

A influência do ENSO e do padrão de dipolo do Atlântico tropical na hidrologia Amazônica é largamente abordada na literatura. A análise de verões e outonos austrais para anos com ocorrência de EN e dipolo quente (anomalias positivas no Atlântico tropical Norte e negativas no Atlântico tropical Sul) mostraram correlações com anomalias negativas de precipitação estendendo-se para leste sobre a Amazônia centro-oeste (SOUZA *et al.*, 2000).

LOPES *et al.* (2016) mostraram haver uma tendência significativa de secas mais intensas nas sub-bacias do sul da Amazônia, altamente correlata com a TSM do Atlântico tropical, ao estudar relações de índices de seca extraídos dos dados existentes de descarga na Amazônia e as anomalias do Pacífico e do Atlântico.

As relações entre chuvas na Amazônia e as anomalias de TSM do Pacífico e do Atlântico tropicais são estudadas em RONCHAIL *et al.*, (2002), mostrando significativa relação entre as chuvas no nordeste da Amazônia com as TSMs tropicais. Essa relação, segundo os autores, é forte com o EN durante todo o ano, sendo a relação com a TSM do Atlântico, observada mais no outono. O estudo mostra também o aumento das anomalias com combinações específicas de condições oceânicas: EN e TSM fria no Atlântico sul juntas estão associadas a anomalias negativas de chuvas muito fortes no norte da Amazônia. Mas interessante também é que as análises mostraram que as anomalias de chuvas, mesmo no NE da Amazônia, não estão associadas só às anomalias de TSM tropicais.

Uma revisão sobre teleconexões e a influência do EN e da TSM do Atlântico no clima do Norte e Nordeste do Brasil são apresentadas em (REBOITA & SANTOS, 2014). As conclusões mostram acoplamento entre as anomalias de TSM do Atlântico Tropical e do Pacífico, com ocorrência de déficits de pluviosidade em eventos de EN.

3.3. A TSM do Atlântico (dipolo), a Zona de Convergência Intertropical e a Amazônia

A ZCIT é uma das feições mais marcantes da meteorologia tropical. Trata-se de uma região relativamente estreita de forte convecção profunda de cúmulos nimbos com alta relação com a convergência dos tramos superficiais norte e sul da célula de circulação de Hadley. Ela é considerada como principal gerador de precipitação na região tropical.

Com o calor latente sendo a principal fonte de transmissão de energia entre as superfícies e atmosfera, nuvens de convecção profunda são características de regiões onde a precipitação excede a evaporação. Estão concentradas sobre porção tropical dos continentes e nas bacias oceânicas aquecidas (região em torno dos 10° de latitude) podem ser identificadas por mínimos nos padrões de Radiação de Ondas Longas – OLR, pois os topos de tais nuvens possuem temperatura menor que a da vizinhança. Esse aglomerado de nuvens ocorre na ZCIT.

Por outro lado, existem regiões, de 10° a 40° de latitude, nas quais a evaporação excede a precipitação, alimentando as células de convecções tropicais. Em tais regiões observam-se valores elevados de OLR. A Figura 3-5 ilustra dados de OLR do Experimento ERBE – NASA *Earth Radiation Budget Experiment*.

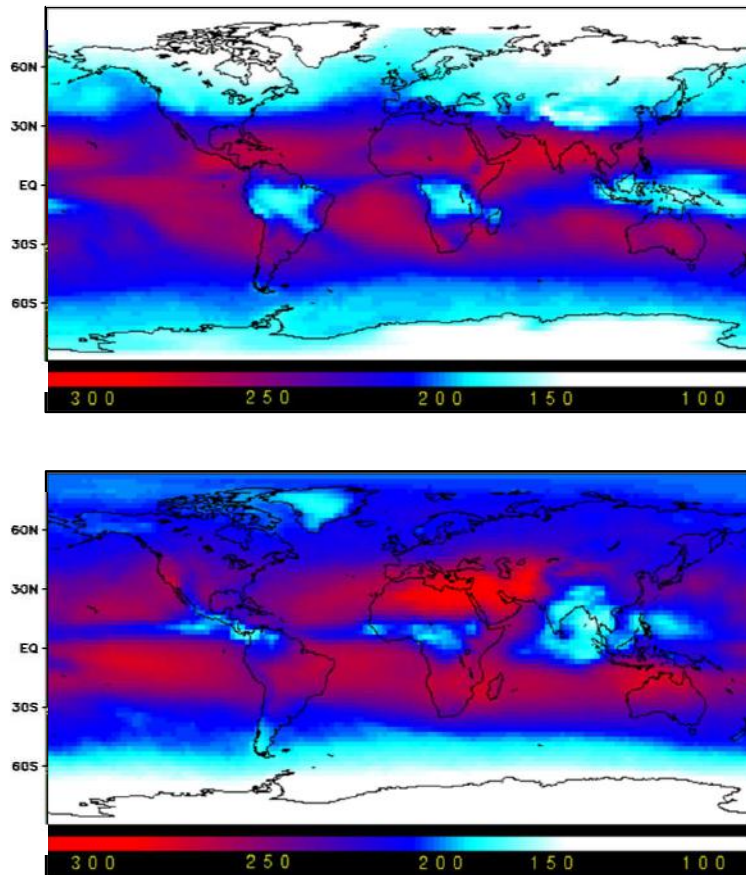


Figura 3-5. Valores de OLR em janeiro (acima) e julho (abaixo) provenientes do ERBE – NASA *Earth Radiation Budget Experiment*, correspondente a médias mensais do período de 1985 a 1986. Adaptado de (ACKERMAN, 2013).

A umidade, que é propriedade abundante na atmosfera tropical, tem efeito de desestabilização vertical, possibilitando convecção profunda e formação de nuvens. Ao ascender, o ar úmido resfria adiabaticamente. Se o resfriamento for suficiente para saturar a parcela de ar, formam-se nuvens pela condensação do vapor, liberando calor latente que aquece a parcela de ar, aumentando sua fluabilidade, favorecendo assim a instabilidade.

Esse fenômeno é denominado convecção úmida. As mudanças na umidade são responsáveis, quase exclusivamente, pelas mudanças na energia da troposfera tropical através da convecção, que carrega calor latente para cima. A distribuição da energia estática na troposfera expressa por perfis verticais de temperatura potencial equivalente – θ_e , pode ser associada a modos característicos das condições de nuvens cúmulos observadas.

Em eventos de baixa convecção, os perfis verticais apresentam mínimos de energia na média troposfera. Se a convecção aumenta, os perfis associados apresentam-se mais homogêneos, reduzindo gradualmente a estabilidade e tendo mínimos em maiores altitudes. Mínimos mais acentuados e mais baixos estão associados a movimentos descendentes de ar seco.

A Figura 3-6 ilustra perfis idealizados de θ_e correspondentes a estados de formação de nuvens associadas à convecção. Os perfis vão da maior estabilidade (menor convecção) em A para a menor estabilidade (maior convecção) em F. Nota-se o aumento da altitude do ponto de mínima θ_e de A para F (linha vermelha) (ASPLINDER, 1976).

Deve-se assim, destacar a inconsistência da descrição da ZCIT simplesmente como uma parte da célula de Hadley onde, por conservação, o ar é forçado a ascender e retornar aos polos (HOLTON, 2004). Esse modelo simplificado não condiz com os perfis verticais de temperatura potencial equivalente mostrados na Figura 3-6, que revelam estabilidade condicional acima da meia troposfera, cerca de 600 hPa. A forma plausível pela qual o calor pode ser transportado da superfície para a alta troposfera seria de forma pseudo-adiabática, no interior das torres de calor. Nessa visão seriam necessárias milhares de

células convectivas agindo simultaneamente para justificar o balanço de transporte de calor ao longo da ZCIT.

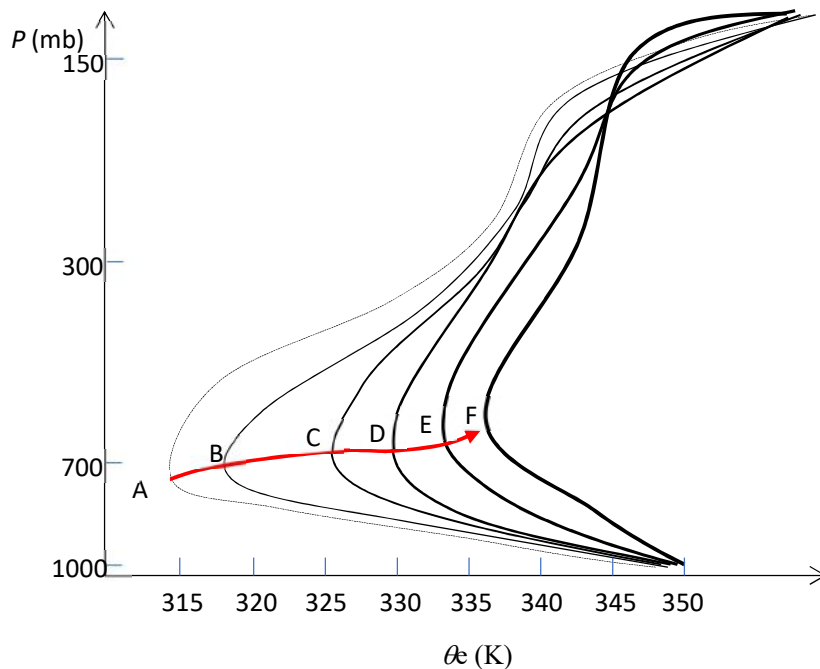


Figura 3-6. Perfis idealizados de temperatura potencial em ordem decrescente de estabilidade (crescente de convecção), sendo A o mais estável (menor convecção) e F o menos estável (maior convecção). A linha vermelha indica os mínimos de θ_e – Adaptado de ASPLINDER (1976).

Além da cobertura vigorosa de nuvens, caracterizada por baixos valores de OLR, como já visto, outras características dessa zona são as correntes ascendentes significativas e o papel importante do calor latente liberado na condensação. Essa faixa de nuvens é composta por aglomerados de cúmulos nimbo de diferentes escalas. Na verdade, a literatura cita um conjunto de variáveis meteorológicas na faixa equatorial que podem caracterizar a ZCIT (CAVALCANTI *et al.*, 2009): a Zona de Confluência dos Alísios – ZCA, a região do cavado equatorial, as áreas de máxima TSM e de máxima convergência de massas e a banda de máxima cobertura de nuvens convectivas, caracterizada pelas imagens de OLR. Essas variáveis podem apresentar pequenas defasagens entre si, porém com altas correlações.

A localização da ZCIT muda ao longo do tempo, migrando em direção ao hemisfério no qual é verão. Assim, ela migra sazonalmente, de uma faixa ao redor dos 14° N entre agosto e setembro, para uma região em torno dos 2°S, de março a abril. Sua migração,

associada ao regime de ventos alísios e ao seu enfraquecimento ou fortalecimento, é determinante para as chuvas no NE brasileiro – NEB, e influencia também as chuvas no Norte do Brasil.

Vários autores citados por CAVALCANTI *et al.* (2009) mostraram o fortalecimento do transporte de umidade e o aumento da convecção no norte da América do Sul devidos à confluência dos ventos e a convergência de massas em baixos níveis associados à ZCIT. Além disso, por ter relação com a circulação geral atmosférica, a ZCIT pode ser afetada por sistemas de larga escala e teleconexões.

CAVALCANTI *et al.* (2009) citaram a relação de variações no ciclo sazonal e outras perturbações de escalas temporais menores da ZCIT com o regime de chuvas no NEB. Um ano mais chuvoso está relacionado a um deslocamento mais para ao sul da ZCIT durante o verão. Da mesma forma, o início da migração também afeta o regime de chuvas.

Outros trabalhos citados por CAVALCANTI *et al.* (2009) relacionam a TSM do Atlântico com o posicionamento da ZCIT e com o regime de chuvas no NEB. O posicionamento da ZCIT é influenciado, sobretudo, pelo dipolo norte-sul de TSM no Atlântico tropical, sendo a região de máxima cobertura de nuvens correlata àquela de maior TSM.

O modo simétrico ao equador de variabilidade de TSM do Atlântico, chamado de dipolo do Atlântico Tropical, caracteriza-se por anomalias de sinais opostos de TSM nas bacias norte e sul do Atlântico (HASTENRATH & HELLER, 1977). Com o dipolo, ocorrem gradientes meridionais de temperatura que afetam sobremaneira a ZCIT e seu deslocamento. Em consequência, as chuvas no Atlântico Equatorial, na parte norte do Nordeste do Brasil, até a parte central da Amazônia são moduladas por deslocamentos da ZCIT.

Para ilustrar os diversos padrões de chuva amazônicos, a identidade pluviométrica (DE PESSOA, 2014) das regiões norte, nordeste e centro-oeste do Brasil são ilustradas na Figura 3-7. A identidade pluviométrica é uma forma gráfica intuitiva de representação de padrões de chuva, permitindo rápida visualização espaço-temporal. Cada posto tem os padrões de ocorrência de chuva diária representados por um esquema binário no qual células correspondentes a determinado dia são pintadas quando ocorre chuva, formando

uma matriz de visualização na qual cada ano ocupa uma coluna com os dias (células) contados de cima para baixo em cada coluna, ao longo de determinado ano.

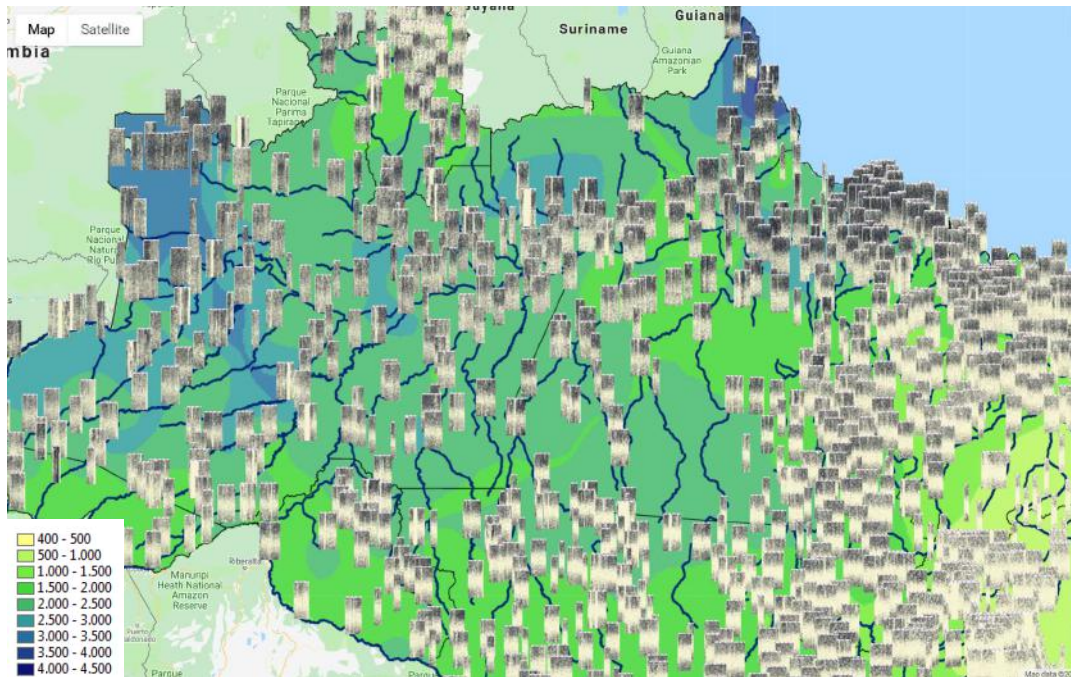


Figura 3-7. Identidade pluviométrica da região amazônica e partes das regiões nordeste e centro-oeste do Brasil, sobreposta a isoietas médias anuais; Fonte: <http://www.snirh.gov.br/frst/>.

Em anos nos quais a TSM da bacia do Atlântico Tropical Sul – ATS, possui anomalia positiva de março a maio, entre o equador e 15°S, e a TSM do Atlântico Tropical Norte – ATN, entre 12°N e 20° N, está mais fria do que o de costume, ocorrem pressões mais baixas do que a média no ATS e mais altas que a média no ATN, deslocando a confluência dos alísios mais para o sul e, junto com eles, a ZCIT, causando precipitações acima da média no NEB e no Norte.

As anomalias de TSM e de PSM no Atlântico também têm sido associadas ao El Niño-Oscilação Sul (ENSO). Assim, através de teleconexões atmosféricas, os gradientes de TSM no Atlântico Tropical são influenciados pelo ENSO (NOBRE & SHUKLA, 1996), causando alterações nos padrões dos alísios e nas chuvas.

Diversos autores propuseram mecanismo de explicação para o impacto do ENSO no aumento da TSM do Atlântico Norte através da intensificação da célula de Walker e enfraquecimento da célula de Hadley sobre a região (sobretudo de 10°N a 20°N), com

enfraquecimento dos alísios de NE e, por consequência, aumento da TSM. GIANNINI *et al.* (2000) tratam da competição entre a alta de PSM do Atlântico subtropical e a ZCIT do Pacífico Leste. A circulação atmosférica na região é moldada por essa competição, com alterações em padrões de chuva causadas pelas mudanças nos ventos superficiais e nas TSMs, por consequência.

A intensificação e aumento da frequência das cheias em Manaus são relacionados ao fortalecimento da circulação de Walker, devida ao aquecimento do Atlântico e resfriamento do Pacífico, BARICHIVICH *et al.* (2018). Os autores destacam a importância e a qualidade da série de níveis médios diários de Manaus, citando outros autores, por drenar grande parte da bacia, por sua extensão centenária e por não conter falhas, ao contrário dos registros de chuva na bacia. Segundo eles, as fortes secas observadas na região central e sul corroboram resultados de simulações de mudanças climáticas. Porém, a tendência mais proeminente na bacia foi de fortes cheias. O aquecimento do Atlântico intensifica a circulação de Walker, que possui ramo ascendente sobre a Amazônia, aumentando assim a convecção profunda, o que causaria o aumento das cheias. Um mecanismo plausível para explicar o aquecimento do Atlântico seria o afluxo de águas mornas provenientes do oceano Índico via corrente das Agulhas. O estudo destaca ainda que a tendência de resfriamento deve cessar nos próximos anos. O Atlântico, porém, provavelmente continuará esquentando nas próximas décadas devido à variabilidade natural da Atlantic Multidecadal Oscillation – AMO, por ação antrópica e pela atuação da corrente das Agulhas.

Por fim, os impactos da água doce do Amazonas na TSM do Atlântico e em modos importantes de larga escala do clima ainda pouco conhecidos, são explorados por JAHFER *et al.* (2017), através de simulações com modelo integrando oceano e atmosfera em cenário com supressão total da vazão do Amazonas. Resultados apontaram influências em modos principais de forçantes climáticos de larga com aquecimento do Atlântico Norte, resfriamento do Atlântico sul e impactos no Pacífico observados no EN. Tais resultados levam a reflexão da importância da conexão entre a Amazônia e o Atlântico, em ciclo que engloba a água doce da bacia, correspondendo a mais de 50% do total de *run-off* do Atlântico.

3.4. Ondas Equatoriais de Leste

A convecção profunda na zona tropical força distúrbios ondulatórios através da liberação de calor latente. A taxa de aquecimento diabático nas torres de convecção provoca campos de velocidades verticais proporcionais e, por conservação da massa, convergência em níveis baixos (ciclones) e divergência em altos níveis (anticiclones), com máximas velocidades na zona convectiva (HOLTON, 2004).

A interação de eventos de larga escala e de escala local se caracteriza, portanto, pela convergência em larga escala nos baixos níveis, umidificando o ar que é elevado nas torres onde o calor latente desestabiliza a atmosfera promovendo a convecção livre. A alteração na estrutura vertical da atmosfera pelo calor latente é ilustrada na Figura 3-6. A energia potencial gerada pelo aquecimento diabático é convertida de pronto em velocidade. Nos trópicos não há armazenamento de energia potencial como nos fenômenos envolvendo massas de ar nos subtrópicos.

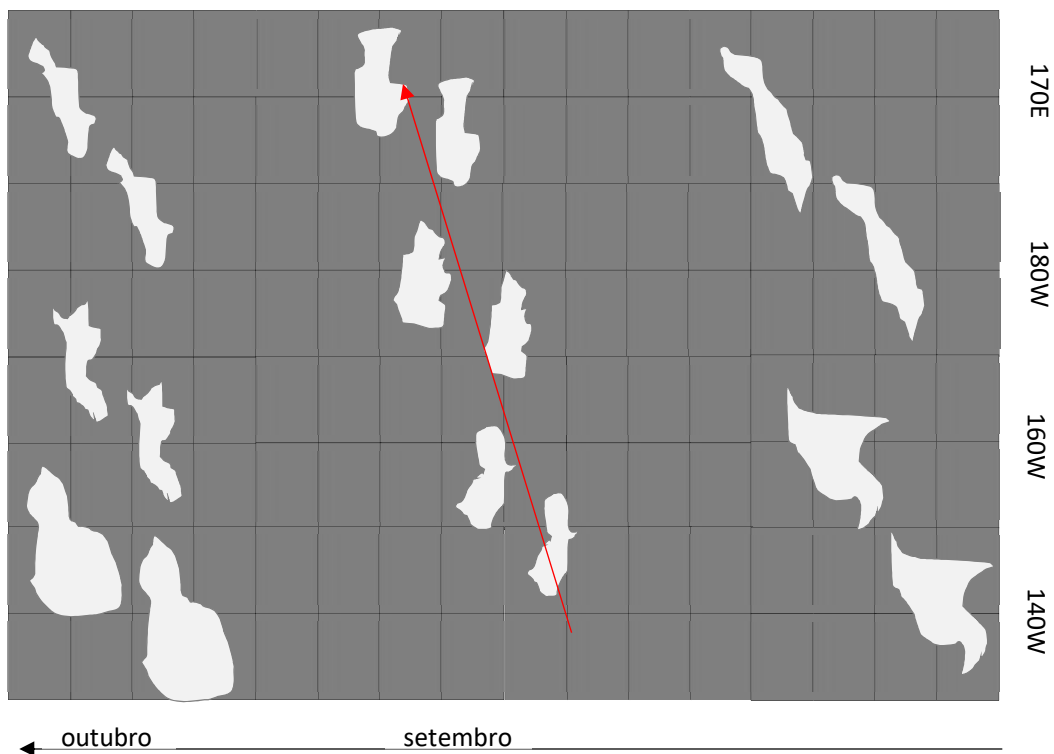


Figura 3-8. Croquis ilustrando como seria um mosaico temporal-longitudinal de imagens de satélite com nuvens em região da ZCIT destacando deslocamento de ondas equatoriais de leste – seta vermelha destaca o deslocamento das nuvens.

As ondas equatoriais modificam a variabilidade da cobertura de nuvens na ZCIT, com uma propagação em direção ao oeste, velocidade característica de 8 a 10 m/s. Tal marcha para oeste fica caracterizada em seções tempo-longitude de imagens de satélite da região, com diagonais indicando o deslocamento para oeste. A Figura 3-8 ilustra como as ondas Equatoriais de leste apareceriam retratadas em um mosaico de imagens de satélite temporal-longitudinal. A seta vermelha destaca a declividade no mosaico, indicando movimento em direção ao oeste dos grupos de nuvens convectivas.

3.5. Variabilidade Intrasazonal

A importância dos eventos da banda semanal à interanual na dinâmica climática sul-americana está evidenciada no número de publicações que tratam de padrões intrasazonais e suas influências nas monções sul-americanas. O presente item revisou as relações entre os eventos climáticos intrasazonais originados no Pacífico Oeste e o clima na América do Sul, em especial na bacia amazônica, e sua relação com os sinais de TSM de larga escala do Pacífico Leste. Buscou-se assim, caracterizar como essa parte importante da variabilidade tropical é captada e reproduzida nas simulações de previsão hidrológica em Manaus.

A identificação da variabilidade de larga escala no sistema de monções sul-americano – SAMS, foi feita por CARVALHO *et al.* (2010), procedendo a análise de funções ortogonais de dados de precipitação pentadal, umidade específica, temperatura do ar e ventos meridionais e zonais a 850 hPa. O trabalho focou nas anomalias intrasazonais, caracterizadas pela primeira função empírica ortogonal EOF, usada como um índice de larga escala, mostrando variabilidade na faixa de 30-90 dias e na segunda EOF, na faixa de 15-20 dias, mais relacionada com a Zona de Convergência Sul-Americana – ZACAS. Os autores destacaram a consistência entre os picos de 30-90 dias da primeira EOF e a oscilação de Madden-Julian – OMJ, conhecida por ter papel importante na modulação da convecção no leste da América do Sul durante o verão. Os autores destacaram ainda o papel central dos trens de ondas de Rossby na variabilidade da ZACAS, tendo sido demonstrado que alguns desses trens estão associados à propagação da OMJ.

Na região amazônica, eventos intrasazonais de precipitação definidos a partir de uma quantidade substancial de estações pluviométricas foram relacionados a índices caracterizadores da OMJ (MAYTA, 2018), mostrando coerência sobretudo com índices

relativos à convecção. Os eventos intrasazonais foram definidos a partir de valores extremos do primeiro componente principal (PC1) da análise EOF de dados de chuva filtrados da região. Assim, os autores avaliaram a relação da precipitação na região com a OMJ ao longo do ciclo sazonal, identificando uma média de 5 eventos por ano, sendo cerca de 25% dos eventos relacionados também às ondas de Rossby e outras oscilações além da OMJ. Outro achado foi a maior importância relativa da variabilidade intrasazonal durante o inverno do HS, quando a contribuição de precipitação relativa devida a tais efeitos pode atingir os 50% do total.

Sendo assim, passa-se a explorar as feições de circulação e de convecção profunda da atmosfera tropical que se organizam em ondas planetárias com períodos de 30-90 dias e velocidade média para leste de 5 m/s, cruzando o oceano Índico Tropical e o Pacífico Oeste e Central. Trata-se da oscilação de Madden-Julian – OMJ. Revisões robustas sobre suas características gerais, traços observacionais em diversas escalas, efeitos em diversas regiões do planeta e previsibilidade são feitas em (ZHANG, 2005) e (KAYANO *et al.*, 2009).

Na escala intrasazonal, a OMJ é considerada o componente com mais influência nas anomalias de precipitação nos trópicos e subtropicais, sobretudo nas regiões de estações chuvosas curtas (KAYANO *et al.*, 2009), como no caso do NEB, ou na bacia amazônica no inverno do HS, como já citado (MAYTA 2018), ou ainda, modulando o SAMS, como também já citado. Importante para a presente pesquisa é também a possibilidade da OMJ interferir em fenômenos de frequências mais baixas, como o ENSO.

A OMJ, como descrito por (KAYANO *et al.*, 2009) foi inicialmente identificada como coerência espectral cruzada entre pressão à superfície, temperatura em vários níveis e vento zonal. Mas pode ser identificada como padrões coerentes de convecção profunda, circulação atmosférica e outros parâmetros, sendo que os padrões de circulação de larga-escala aparecem acoplados às zonas mais restritas de convecção. Padrões de OLR associados à OMJ, por exemplo, apontam anomalias mais fortes no oceano Índico e no Pacífico Oeste e pouco sensíveis no Pacífico Leste e América do Sul.

Outra característica geral da OMJ apontada em (ZHANG, 2005) é a variabilidade dos padrões, não só entre eventos, mas dentro de uma mesma ocorrência, ao longo da marcha

para leste. A oeste, a OMJ é fortemente acoplada à convecção profunda, propagando-se lentamente, na faixa dos 5 m/s. Já mais a leste, não há tal acoplamento e a OMJ exibe propagação bem mais rápida, podendo chegar aos 30-35 m/s. Ainda sobre a propagação, os eventos de OMJ podem ser divididos em três classes, exibindo ramos para norte ou propagando-se apenas para leste.

Quanto à importância da OMJ na previsibilidade climática, ainda existem limitações na representatividade da OMJ, com variância intrasazonal na maioria dos modelos numéricos globais mais fraca do que a observada (KAYANO *et al.*, 2009). Mesmo assim, a previsibilidade com modelos é maior na presença de amplitudes significativas da OMJ. Com as limitações existentes em modelos numéricos globais, houve a motivação para modelos empíricos.

A OMJ, foi apontada como responsável por grande parte da variabilidade interanual característica do ENSO. podendo modular sua intensidade, fase e duração. Essa interação ocorre tanto na superfície como sub-superficialmente no Pacífico. De maior interesse para o presente estudo é saber como os ventos em baixos níveis associados à OMJ podem alterar os padrões de TSM do Pacífico, fazendo-se presente na previsão de níveis em Manaus.

Evidências observacionais dos forçantes de ondas de Kelvin oceânicas pela OMJ e anomalias de TSM de eventos individuais de ENSO no Pacífico Leste foram usados com o propósito de estudar relações das anomalias interanuais da atividade sazonal da OMJ com anomalias interanuais na TSM equatorial do Pacífico leste no período de 1980 a 1999 (ZHANG & GOTTSCHALCK, 2002). A variabilidade interanual da OMJ importante para o ENSO foi medida a partir de quantificadores da interação ar-oceano no Pacífico, interação na qual as ondas de Kelvin têm papel importante.

Dados de ventos superficiais de reanálise de modelo global foram usados para compor um índice capaz de mensurar a variabilidade interanual dos forçantes das ondas de Kelvin, o atrito com o vento equatorial zonal associado a OMJ. Tal variabilidade foi então relacionada às anomalias de TSM no Pacífico Leste, associadas ao ENSO, no período citado.

Cada evento quente característico do ENSO estudado teve relação distinta com a variabilidade interanual dos forçantes das ondas de Kelvin associadas à OMJ. Contudo, para os seis eventos ENSO quentes o aumento das anomalias de TSM é quase linearmente relacionado ao crescimento da intensidade da tensão de atrito com o vento, precursora das ondas. Portanto, existem semelhanças entre tais eventos, com forte transferência de energia do vento para a superfície do Pacífico Oeste, seja no início, no desenvolvimento ou na fase madura de quase todos os eventos quentes de ENSO, sendo tal atividade muito fraca imediatamente após os picos de aquecimento.

Para todos os eventos quentes, uma sequência similar de acontecimentos ocorreu, inicialmente com anomalias de TSM surgidas no Pacífico Leste, seguidas de forte transferência de energia para as ondas de Kelvin por eventos intrasazonais de vento da OMJ. Na medida em que a TSM aumentou em toda a bacia (anomalias positivas), os forçantes das ondas foram intensificados. Seguiram-se picos de TSM a leste, com enfraquecimento dos forçantes de ondas, que se tornaram muito fracos após os picos de TSM.

Resultados correlacionais das anomalias interanuais de TSM com forçantes das ondas de Kelvin pela OMJ foram apresentados em grade de longitude versus defasagem em meses. Mais detalhes a respeito dos níveis de confiança adotados, graus de liberdade e número de amostras independentes estimado da autocorrelação devem ser consultados, destacando que cada evento ENSO extremo foi tomado como independente. Correlações foram estabelecidas com níveis de confiança superiores a 95% e com os forçantes do vento antecipando as anomalias de TSM por períodos de seis a doze meses. Os níveis de confiança foram mais altos para as regiões do Niño12 e Niño3 do que no Pacífico Central e Oeste.

3.6. Degelo nos Andes e desmatamento - relações com o clima e hidrologia

A bacia amazônica sofreu alterações importantes na cobertura do solo nas últimas décadas: o derretimento dos Andes e o desmatamento, sobretudo na região sul da bacia.

As geleiras dos Andes, que correspondem hoje a mais de 99% das geleiras tropicais (KASER, 1999), vêm diminuindo desde o início do século 18, exceto por alguns períodos de estabilização ou ganho. Tal redução, drástica nas três últimas décadas, atingiu

principalmente as geleiras abaixo dos 5.400 m, grande maioria dos Andes, projetando-se seu desaparecimento nas próximas décadas RABATEL *et al.* (2013).

Os impactos hidrológicos do acelerado derretimento das geleiras da Cordillera Blanca, no Peru, na cabeceira da bacia do rio Santa foram estudados por BARAER *et al.* (2012), usando descargas diárias de nove estações. Os autores usaram simulação hidrológica baseada no recuo das geleiras como base para interpretar as tendências encontradas nas séries de vazões. Apesar de desaguar no Pacífico e de ter tamanho muito pequeno comparada a bacia de contribuição de interesse da presente pesquisa, os resultados mostraram alterações significativas nas vazões de estiagem, um aumento da sazonalidade.

Mesmo na bacia do rio Santa Branca, pequena comparada às proporções da área de drenagem da série MAN, a influência maior no balanço hídrico da água de derretimento acontece na estação seca. Durante a estação chuvosa, o fluxo de calor latente na vertical fornece energia extra e aumenta a ablação em 33% na Cordillera Blanca, (MARK, 2003). Porém, com 70% a 90% da chuva concentrada nesse período, o derretimento extra torna-se menos importante frente às vazões geradas pela precipitação. Resultados de outro estudo sugerem aumento de 1,6% na vazão específica das regiões mais cobertas por geleiras e contribuição em média 60% maior da contribuição subsuperficial durante a seca, MARK *et al.* (2010).

Existem muitas incertezas associadas ao papel do degelo dos Andes: quantificação da espessura das massas de gelo e de sua evolução, alterações nas coberturas de nuvens, neve precipitada na zona de acumulação das geleiras e balanço entre produção subsuperficial da bacia e proveniente do derretimento. Apesar disso, as preocupações com impactos na agricultura, geração hidroelétrica e com outros aspectos socioeconômicos são perceptíveis para a região de cabeceira do Amazonas (VUILLE *et al.*, 2018). Porém, tomando os 1.920 km² de área correspondente a 99,6% de geleiras tropicais na América do Sul estimada em 2006 em FRANCOU, B. & VINCENT (2010), se todo o derretimento anual ocorresse no intervalo de um mês, seria necessária uma ablação seguramente maior que 1 m para gerar vazão da ordem de 1% da média observada no rio Solimões em Manacapuru, maior que 100.000 m³/s, próximo e a montante de Manaus. Mas as maiores ablações citadas por RABATEL *et al.* (2013) no período de 1976-2010 são de 0,76 m por ano, o que produziria vazões menores que 1% da vazão média em Manacapuru.

As evidências apontam para uma contribuição significativamente maior da vazão proveniente das chuvas sobre a bacia em relação à água de degelo. Assim, possíveis alterações na coerência entre MAN e as séries de TSM seriam mais bem explicadas pelo desmatamento ocorrido, sobretudo na porção sul da Amazônia. Maior ênfase é dada para as alterações na porção sudeste da floresta em DAVIDSON *et al.* (2012) e na região sudoeste em LIMA *et al.* (2014).

As taxas de evapotranspiração – ET, e de infiltração são sabidamente maiores em florestas tropicais do que em áreas onde a floresta foi substituída. Além disso, parte significativa da precipitação – P vem da reciclagem de água de ET da própria floresta (GIAMBELLUCA, 2002). Na Amazônia, a parcela de reciclagem da ET é de cerca de dois terços e a umidade restante vem do oceano Atlântico, DAVIDSON *et al.* (2012).

Com o desmatamento, a resposta hidrológica de uma bacia dependerá das alterações no balanço entre precipitação – P e ET. Desmatamentos em regiões menores não devem causar maiores interações com o clima e alterações dos padrões de chuva, reduzindo a ET e aumentando a vazão e o risco de cheias conseqüentemente. Porém, a substituição de áreas maiores de floresta desencadeia interações complexas entre o clima e a bacia, podendo levar à redução de P, ao ponto de diminuir o escoamento, DAVIDSON *et al.* (2012).

Ainda segundo DAVIDSON *et al.* (2012), para que possíveis alterações em vazões de grandes bacias fossem desmembradas das respostas da bacia relacionadas à variabilidade normal da P, seriam necessárias séries extensas e alterações significativas. A complexidade da dinâmica tropical, com variabilidade em diversas escalas dificulta ainda mais essa tarefa. Adicionalmente, as alterações hidrológicas precisariam ser diferenciadas daquelas relacionadas diretamente às mudanças climáticas externas à bacia, como o fortalecimento da circulação de Walker descrito por BARICHIVICH *et al.* (2018).

Assim, não é tarefa fácil identificar os fatores citados como causadores de alterações hidrológicas em rios amazônicos. Nos rios Tocantins, COSTA *et al.* (2003) e Araguaia, COE *et al.* (2011), ambos no cerrado brasileiro, aumentos das descargas foram relacionados ao desmatamento. Previsões com modelos que consideram as reduções no fluxo de calor latente e aumentos no calor sensível devido ao desmatamento e

consequente redução da ET, acoplando efeitos locais do desmatamento às forçantes climáticas, apontam para diversos cenários futuros em termos hidrológicos na região amazônica.

Como exemplo, a total substituição da floresta por grama apontou para aumento da temperatura de 2° C sobre a área modificada, além da indução de um padrão de dipolo na precipitação, com redução de 7.9% a oeste e aumento de 8.3% a leste da bacia (LLOPART *et al*, 2018). Segundo os autores, o dipolo de P em tais condições foi achado por diversos outros autores. Outro estudo analisou as bacias dos rios Juruá, Purus e Madeira, que sofreram alterações no uso do solo nas últimas décadas, considerando diversos cenários de desmatamento, com e sem acoplamento com a resposta climática ao desmatamento (LIMA *et al.*, 2014). O cotejo dos balanços hídricos anual e sazonal revelou aumentos nas vazões proporcionais aos desmatamentos em cenários sem resposta climática para as três bacias. Com a introdução da resposta climática, o aumento do desmatamento produziu reduções progressivas e uniformes de P nas bacias dos rios Purus e Juruá, com diminuição das vazões médias anuais. Já no rio Madeira, houve aumento de P ao sul e diminuição de P no restante da bacia e a queda maior da ET compensou a perda de P, aumentando as vazões. O comportamento diferenciado do Madeira foi atribuído à cobertura vegetal diferente, com parte da área composta de savana.

A complexidade das relações entre o desmatamento, o clima e a hidrologia aumenta se considerarmos que as áreas alteradas vão progressivamente recuperando o equilíbrio energético original (GIAMBELLUCA, 2002).

3.7. Análise dimensional nos trópicos

A análise dimensional constitui uma valiosa ferramenta para identificar nas equações governantes os termos dominantes no cenário de interesse.

No âmbito da presente pesquisa, o trabalho de YANO & BONAZZOLA (2009) constituiu referência importante, na medida em que revelou sutilezas da dinâmica atmosférica tropical. Tais informações inspiraram a utilização da DFA, como ferramenta de análise fractal, e da Coerência Wavelet, para análise de correlações multi-escalas em séries não lineares e não estacionárias do interesse da presente pesquisa, permitindo-se mais a frente,

a evidenciação das relações entre as séries de TSM e de OLR com a dinâmica de níveis em Manaus. Assim, maiores detalhes sobre a análise dimensional e a análise de dados observacionais realizada por YANO & BONAZZOLA (2009) podem ser consultadas no ANEXO II.

YANO & BONAZZOLA (2009) realizaram análise dimensional da dinâmica atmosférica tropical abrangendo a mesoescala espacial, a escala sinóptica e a planetária. Fixando a magnitude de ventos horizontais, foram realizadas estimativas para os termos da equação da termodinâmica em diferentes escalas horizontais de movimentos, destacando-se as sutilezas e incertezas inerentes à análise dimensional.

Para as diferentes escalas espaciais, dois regimes foram cogitados por YANO & BONAZZOLA (2009) para a dinâmica: o primeiro, marcado pelo equilíbrio dinâmico entre a advecção vertical de temperatura potencial e a taxa de calor diabático na termodinâmica; o segundo, consistindo em uma dinâmica de ondas equatoriais lineares moduladas pela convecção de calor, estabelecido pelo equilíbrio entre a advecção vertical de temperatura potencial e o termo de variação local em uma escala de tempo majorada, também na equação termodinâmica. O equilíbrio dinâmico desembocaria na eliminação das ondas de gravidade, com uma estratificação neutra, um cenário, portanto, qualitativamente distinto da dinâmica tropical de larga-escala dominada por ondas equatoriais, descrito tradicionalmente na literatura.

Os autores concluíram por uma maior adequação do equilíbrio dinâmico na escala sinóptica. Conforme a escala espacial L aumenta (de 1.000 a 3.000 km), no entanto, há incertezas entre os dois regimes relacionadas à sensibilidade dos parâmetros à L , com possibilidade das ondas equatorias comandarem a dinâmica subitamente. Para escalas maiores que 3.000 km, ainda segundo os autores, teoricamente, o regime de ondas equatoriais moduladas pela advecção é esperado. Porém, análise de dados indicou que a condição balanceada pode ser satisfeita para escalas acima de $L \sim 10^3$ km. YANO & BONAZZOLA (2009) destacam o caso da Oscilação de Madden-Julian, que pode ser interpretada à luz da dinâmica balanceada ao invés de por ondas equatoriais acopladas à convecção, mesmo com escala característica bem acima dos 3.000 km.

Adicionalmente, os autores realizaram análise de dados observacionais. O conjunto de dados usados é oriundo da malha de larga-escala *Large-Scale Array* – LSA, durante o experimento *Tropical Ocean and Global Atmosphere* – TOGA, *Coupled Ocean-Atmosphere Response Experiment* – COARE, *Intensive Observing Period* (IOP, 1/10/1992 a 28/02/1993). Os dados coletados possuem resolução espacial de 1° e temporal de 6h, com variação total de 40° em longitude e 20° em latitude e período de 4 meses, permitindo assim escalas de promediação espacial de 2° a 20° , correspondendo a $\Delta x \approx 200 - 2.000$ km, e temporal de 12h a 4 meses.

A relação R entre a derivada local e da advecção vertical de temperatura potencial da equação termodinâmica foi utilizada por YANO & BONAZZOLA (2009). As médias móveis de R ao longo das escalas foram calculadas e então, o valor médio quadrático de tal resultado foi tomado ao longo das escalas espaciais e temporais. O numerador e o denominador decresceram com o aumento das escalas espacial e temporal, porém, em taxas distintas. Valores baixos de R indicam equilíbrio termodinâmico.

Os resultados mostraram valores maiores de R no domínio para Δt pequeno, independente de Δx , o que, segundo os autores, pode estar relacionado erros de medição. Mas R rapidamente decresceu para janelas de promediação maiores que alguns dias. Para períodos de promediação Δt acima de 70 dias, a relação R apresentou leve decréscimo se mantida a escala espacial horizontal, sugerindo a eliminação rápida de variações de baixa frequência relacionadas às torres de convecção, em janelas temporais maiores e uma relevância maior que a prevista para o regime de dinâmica balanceada. No entanto, a análise dos dados foi limitada a escalas de 2.000 km, podendo revelar tendências mais aderentes às previsões da análise dimensional para escalas espaciais maiores.

A tendência apontada, pelos resultados da análise dimensional, de transição do regime de dinâmica balanceada para dinâmica dominada pelas ondas de gravidade entre 1.000 e 3.000 km não foi observada na análise dos dados. O parâmetro R deveria aumentar para escalas horizontais maiores, indicando assim a tendência a transição para regime de dinâmica de ondas apontado pela análise dimensional. Porém, R mostrou-se quase independente de Δx para Δt menor que 40 dias.

Os autores ressaltaram ainda que a persistência da variabilidade nas diversas escalas, característica de ruídos $1/f$ associados a comportamentos pulsantes na atmosfera tropical pode explicar a invariância de R ao longo das escalas temporais e ao longo das escalas espaciais para períodos menores que 70 dias.

3.8. Hierarquia ou auto-afinidade?

Como visto, a análise dimensional realizada por YANO & BONAZZOLA (2009) apontou, teoricamente, os domínios nos quais a atmosfera tropical seria dominada por ondas ou pelo equilíbrio dinâmico. Foram encontradas, porém, incertezas (ligadas às estimativas de alguns parâmetros) nos limites da transição teórica entre tais regimes na faixa de 1.000 a 3.000 km. Adicionalmente, a análise de dados observacionais realizada pelos autores apontou invariância de R ao longo das escalas temporais e ao longo das escalas espaciais para períodos menores que 70 dias, o que pode ser característica de ruídos $1/f$ associados a comportamentos pulsantes na atmosfera tropical.

Considerando a importância da caracterização da dinâmica tropical e seus regimes no estabelecimento de relações entre o calor transferido dos oceanos para a atmosfera e a hidrologia em Manaus, apresenta-se base teórica a respeito da hierarquização de escalas envolvidas na dinâmica tropical. Os espectros de energia envolvidos nesse processo possuem características marcantes de auto-afinidade, apesar de haver certo grau de hierarquia e separação de escalas em regiões nas quais destacam-se alguns sinais marcantes.

O leitor deve estar familiarizado com os conceitos básicos de geometria fractal, a fim de compreender as principais características das séries temporais estudadas nessa pesquisa, pois elas apresentam, como muitos outros fenômenos naturais, complexidade, auto-afinidade e auto-similaridade, que tornam sua tratativa como fractais interessante, permitindo a caracterização de suas dimensionalidades fractais e quebras de escalas. O conceito de fractal e de dimensão fractal pode ser consultado em MANDELBROT (1983). A dimensão fractal D , relaciona-se ao expoente de Hurst H , ambos sendo descritos em MANDELBROT (1983) e FAVARETTO (2004). Já o conceito de Movimento Browniano Fractal – MBF é abordado em (HENRIQUES, 2008). Uma revisão mais detalhada sobre tais conceitos é apresentada no ANEXO I, apresentando-se a seguir a

conceituação de auto-afinidade, auto-similaridade e a revisão de algumas pesquisas caracterizando aspectos fractais importantes da atmosfera tropical.

Conceitos de hierarquia, auto similaridade e auto-afinidade

No âmbito da presente pesquisa, define-se hierarquia de escalas como a existência de separação de escalas no espectro de energia de determinado parâmetro. Em tal condição, é possível discernir picos energéticos em determinadas frequências, possivelmente associados a forçantes importantes do fenômeno de interesse analisado.

Em bandas distintas dos sinais hierárquicos, no entanto, frequentemente é difícil separarem-se frequências características, ficando o espectro de energia cheio e caracterizado pela ocorrência de trechos com inclinação característica das relações de variabilidade presentes ao longo das escalas consideradas. Essa relação fractal caracteriza a auto-afinidade. Quando há semelhança fractal isotrópica, o termo auto-similaridade é usado. Quando a relação é anisotrópica, como é o caso de séries temporais, a relação é auto-afim.

Um conjunto fractal delimitado D é auto-similar com respeito à escala δ se D é a união de N subconjuntos não sobrepostos D_1, \dots, D_N , cada qual congruente ao conjunto $D_{*\delta}$ obtido pela transformação de similaridade definida

$$\begin{aligned} \delta(D) : D &\rightarrow D_{*\delta} \\ \mathbf{x} = (x, y) &\mapsto \mathbf{x}' = (\delta x, \delta y) \end{aligned}$$

onde $0 < \delta < 1$. Congruente significa que o conjunto de pontos D_i é idêntico ao conjunto $D_{*\delta}$ após translações e rotações YANO & NISHI (1989).

Existem também fractais anisotrópicos, caso das séries temporais, nos quais as semelhanças se dão em escalas distintas nas diferentes direções. Estes últimos fractais são denominados auto-afins. Considerando $f(t)$, uma função de uma variável e unívoca, uma transformação auto-afim associa, para um ponto $(t, f(t))$, um novo ponto $(\delta_1 t, \delta_2 f(t))$, no qual δ_1 e δ_2 não são iguais. Portanto, a auto-afinidade refere-se ao vetor $\delta = (\delta_1, \delta_2)$. A

auto-afinidade estatística, significando que o conjunto é estatisticamente auto-afim, é um termo mais apropriado.

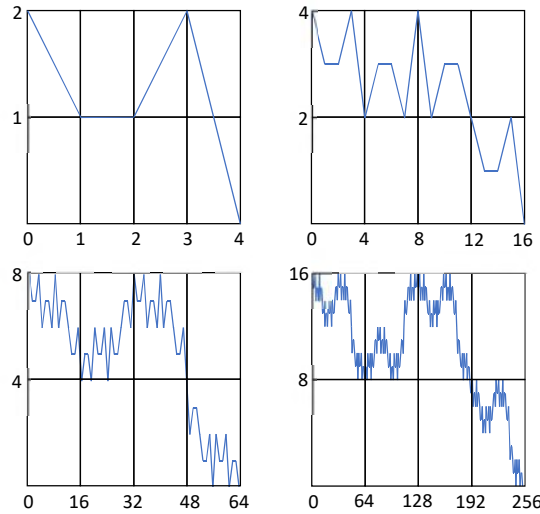


Figura 3-9. Figura auto-afim com fator de escala horizontal $\delta_x = 4$ e fator de escala vertical $\delta_y = 2$.

A Figura 3-9 ilustra um fractal auto-afim construído com fatores de escala de afinidade $\delta_x = 4$ e $\delta_y = 2$, na horizontal e vertical, respectivamente.

Se uma função $G[f(t)]$ não muda sua forma quando submetida à transformação:

$$G[\delta_2 f(\delta_1 t)] = \gamma G[f(t)] \quad (3-1)$$

Sendo γ a razão de escala, então a série temporal $\{f(t)\}$ é auto-afim em termos de $G[f(t)]$. Se pegarmos o espectro de potência $P(f)$ dessa função, sendo f a frequência, a relação acima pode ser reescrita como:

$$\delta_2 P\left(\frac{1}{\delta_1} f\right) = P(f) \quad (3-2)$$

Sob a transformação auto-afim definida pelo vetor $\delta = (\delta_1, \delta_2)$, a equação (3-2) mostra que o espectro de potência não muda a sua forma, sendo δ_1 o fator de reescalonamento de

tempo e δ_2 o de amplitude, $\delta_1 \neq \delta_2$. A única solução possível para a equação algébrica (3-2) é:

$$P(f) \propto f^{-\alpha} \quad (3-3)$$

Isto significa que uma transformação auto-afim deve ter um espectro de potência proporcional à potência inversa α da frequência, sendo α uma constante positiva. Em outras palavras: a variabilidade auto-afim é construída pela relação com frequências mais altas com taxa constante de redução das amplitudes, conforme crescem as frequências.

A auto-afinidade pode ser expressa pelas transformações:

$$t \mapsto \lambda t \quad (3-4)$$

$$f \mapsto \lambda^H f \quad (3-5)$$

sendo H o expoente de Hurst. Combinando as transformações acima, tem-se:

$$\lambda^H f(t) \cong f(\lambda t) \quad (3-6)$$

$$f(t) \cong \lambda^{-H} f(\lambda t) \quad (3-7)$$

A propriedade de auto-afinidade é expressa por (3-7) sendo uma igualdade estatística (expressa pelo símbolo \cong), significando que uma ou mais partes do perfil são semelhantes ao todo. O expoente de Hurst H é está intimamente ligado às características geométricas do perfil, sintetizando toda a informação relacionada ao seu reescalonamento (FAVARETTO, 2004).

Em perfis fractal, a dimensão fractal D relaciona-se com o expoente de Hurst H por (MANDELBROT, 1983):

$$D = 2 - H \quad (3-8)$$

A relação acima determina que, na medida em que H diminui ($0 \leq H \leq 0,5$), D se aproxima de 2 e o perfil analisado vai se tornando rugoso e preenchendo o espaço. Se H aumenta

($0,5 \leq H$), a dimensão fractal diminui, suavizando o perfil. Com $H = 0,5$, tem-se $D = 1,5$, correspondendo ao ruído branco.

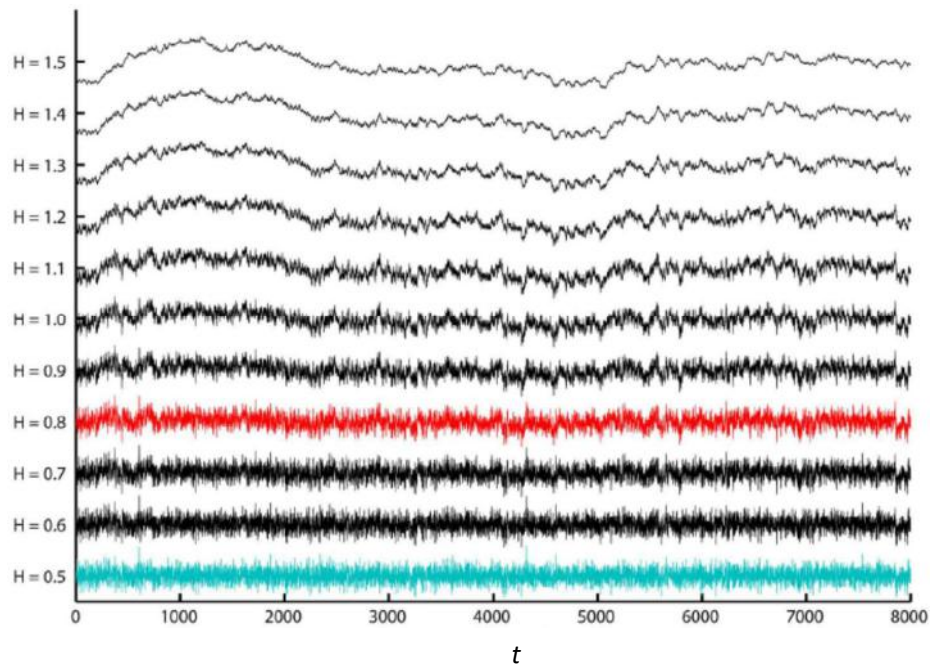


Figura 3-10. Estruturas fractais geradas com diferentes expoentes de Hurst H : para $H = 0,5$ a série não guarda memória temporal; na medida em que H aumenta, a série passa a possuir maior memória temporal; para $H > 1$, as séries não são mais estacionárias e são movimentos Brownianos fractais (MBFs), com $H_{MBF} = H_{RGF} - 1$.

A Figura 3-10 ilustra como o expoente de Hurst H influencia a geometria de perfis auto-afins, alterando-a desde as séries tipo ruídos ($0 \leq H \leq 1$) às do tipo caminhar randômico – livre tradução de *random walk*, ($1 \leq H$). A série tem, assim, dependência temporal crescente para $0,5 < H$, apresentando autocorrelações com intervalos cada vez maiores. Para $0 \leq H < 0,5$, as séries apresentam estrutura anti-correlacionada, ou seja, se há uma tendência de crescimento em um instante, será, provavelmente, sucedida por uma tendência de declínio. Para o ruído branco, $H = 0,5$, a série tem estrutura cuja dependência temporal esvanece em período muito curto.

Análises implementadas por YANO & NISHI (1989) exploraram a presença ou não de serapção de escalas em dados de radiação de ondas longas – OLR da NOAA. Os autores justificaram a escolha de tal parâmetro para estudar a dualidade entre hierarquia e auto-afinidade na atmosfera tropical pela disponibilidade de séries mais longas e pela OLR

relacionar-se com a atividade convectiva, apesar das interferências de temperaturas da superfície.

Os espectros de OLR apresentados por YANO & NISHI (1989), além de demonstrarem o caráter misto da variabilidade na atmosfera tropical entre auto-afinidade e hierarquia, evidenciaram, através da presença dos picos, a relação entre fenômenos de diversas escalas e sistemas convectivos nos trópicos. Tais características foram ressaltadas em escalogramas entre o comprimento representativo da variabilidade $L(t)$ com a janela de promediação temporal t para OLR ao longo do equador, com intervalos de 45° de longitude. Em tais escalogramas, alguns picos espectrais foram caracterizados por desvios em relação às linhas sólidas que representam os ajustes de mínimos quadrados da variabilidade. Por fim, o mapa de contornos de escala de auto-afinidade τ_s em seção transversal latitude-longitude com média móvel espacial de 5° longitudinais mostraram a distribuição espacial da dualidade entre auto-afinidade e hierarquia na atmosfera tropical refletindo-se nos dados de OLR. Destacaram-se as regiões de origens dos dados a serem usados na presente pesquisa para prever os níveis em Manaus. Mais detalhes sobre YANO & NISHI (1989), incluindo a definição de $L(t)$, de τ_s e o mapa com interpolação espacial dessa propriedade são apresentados no ANEXO II. Tal mapa foi adaptado, sobrepondo-se as regiões principais de promediação das séries de interesse da presente pesquisa.

A análise do ruído $1/f$ e dos eventos tipo pulso em dados superficiais da atividade convectiva tropical por YANO *et al.* (2004) sugerem ser os eventos pulsantes elementos chave para interpretar a dualidade da variabilidade tropical, consistindo tanto de auto-afinidade como de hierarquia de escalas. Segundo os autores, os pulsos associados às correntes verticais convectivas, ventos de oeste em escala intrasazonal associados à OMJ e eventos de El Niño possuem, a grosso modo, durações características, ainda que contribuam para espectros de bandas muito mais largas, satisfazendo a uma lei de auto-afinidade, particularmente em frequências intermediárias (longe dos limites de resolução e dos limites impostos pelo tamanho das séries). Os autores ressaltaram que nem todos os eventos tipo pulso podem ser associados a ruídos $1/f$, como é o caso da precipitação analisada, com picos mais definidos e caudas curtas. A não adequação da precipitação pode ser fruto de limitações da medição, pois tais eventos podem ser muito localizados espaço-temporalmente.

Ainda segundo os autores, o comportamento auto-afim da variabilidade convectiva da atmosfera tropical, manifestando-se como ruído $1/f$ possui implicações na hipótese tradicional de quase-equilíbrio em parametrizações convectivas. Segundo tal hipótese, a resposta da convecção profunda tropical a qualquer forçante de larga escala ocorreria em escalas temporais muito mais curtas, estando assim a convecção quase em equilíbrio com os forçantes. Para validar essa hipótese fisicamente, tal resposta mais curta deveria ser detectável, manifestando-se como uma queda abrupta na variabilidade convectiva com o aumento das escalas de promediação, o que não foi observado.

YANO *et al.* (2004) estenderam a análise a escalas mais curtas, de 20min a 3hs, tendo como referência o tempo característico de resposta da convecção, estabelecido em outras pesquisas. Os espectros analisados mostraram o comportamento de ruído $1/f$ terminando na escala de 1h. Para escalas menores, a temperatura e a umidade relativa indicaram transição para declividade menos acentuada, típica de ruído branco. Assim, não se podem caracterizar tais rupturas de escala em 1h como indicativas das escalas de ajuste da convecção, pois elas implicam em maior variabilidade para escalas maiores, contrariando o esperado segundo a hipótese de quase equilíbrio físico da convecção tropical. Os autores propõem a visão baseada em eventos pulsantes caracterizando ruídos $1/f$ inclusive em escalas maiores que a horária como alternativa ao quase-equilíbrio convectivo tropical.

4. Métodos de análise de dados usados

Neste capítulo revisam-se os métodos de análise de dados usados no presente trabalho. A análise de dados foi usada na caracterização das relações fenomenológicas existentes entre as séries de TSM do Atlântico e Pacífico e a série histórica de níveis em Manaus e no pré-processamento desses dados de TSM para alimentação dos modelos híbridos EMD-RNA a serem discutidos.

Inicialmente, duas técnicas espectrais são tratadas: a Wavelet e a HHT. A Wavelet foi usada na análise de coerência espectral dos dados de TSM com a hidrologia em Manaus. Objetiva-se também contextualizar e justificar o uso da EMD, parte integrante da HHT nos modelos híbridos. Em seguida, a análise destendenciada de flutuações é revisada, para subsidiar a caracterização de semelhanças fractais entre os dados de TSM e os níveis em Manaus. Por fim, apresentam-se a análise de componente principal, livre tradução de *Principal Component Analysis* - PCA, e a regressão linear parcial de mínimos quadrados,

livre tradução de *Partial Least Square Regression* - PLS, usadas também no pré-tratamento de dados como forma de mitigar possíveis problemas de multicolinearidade e o mal-condicionamento dos modelos híbridos.

4.1. Contextualização

O pré-processamento de dados, com determinação de parâmetros adequados à construção de um modelo, e o pós-processamento de resultados modelados para verificação da representatividade do fenômeno descrito, são duas aplicações importantes da análise de dados e mais especificamente do processamento de sinais (HUANG & ATTOH-OKINE, 2005).

As características dos dados a serem analisados são determinantes na escolha da ferramenta de análise. Fatores como disponibilidade de dados (extensão dos registros), não estacionariedade e não linearidade dos processos precisam ser contemplados no processamento de sinais.

No passado, limitações em ferramentas de análises acarretaram em imposições de hipóteses limitantes de periodicidade e estacionariedade nos estudos, sobretudo em componentes de longo termo. Técnicas baseadas na Transformada de Fourier foram largamente utilizadas no passado. Tais aplicações basearam-se em análises harmônicas dos sinais de interesse, tentando estabelecer relações das componentes de médio e longo prazo com outros sinais e conseguir alguma previsibilidade dos fenômenos estudados.

As técnicas tradicionais baseadas na Transformada de Fourier tendem a dispersar a energia do espectro em diferentes frequências, levando a erros de interpretação, sobretudo nas baixas frequências (RAMACHANDRA & HSU, 2008). Além disso, geralmente as hipóteses de estacionariedade e linearidade podem ser inapropriadas no tratamento de fenômenos. Por outro lado, a análise só no domínio da frequência ou só no domínio do tempo também pode ser grande limitante.

Com o passar do tempo, surgiram alternativas a essas limitações. Para sanar a perda da dimensão temporal e permitir a tratativa da não-estacionariedade, na Transformada de Fourier por Janelas – STFT, sigla em inglês para *Short Time Fourier Transform*, o

domínio é dividido em janelas temporais, calculando-se a transformada para cada uma delas.

Sendo assim, a resolução obtida fica atrelada à janela escolhida. Além disso, a escolha de base de funções ortogonais pré-definidas dificulta descrições de fenômenos não-lineares, caracterizados por variações de frequência dentro de um mesmo período. De interesse para a presente pesquisa são a Wavelet e a HHT, ambas descritas a seguir. São ambas técnicas de análise tempo-frequência, e serão revisadas buscando embasar o estudo da coerência espectral Wavelet e contextualizar a escolha da EMD para a alimentação dos modelos híbridos com RNAs. Após revisar as técnicas que envolvem transformações, a PCA e a PLS serão revisada.

4.2. Wavelet e a coerência espectral

A Coerência Espectral Wavelet será utilizada para caracterizar as relações fenomenológicas entre os dados oceânicos das bacias do Atlântico e do Pacífico e a série de níveis de Manaus.

Como não foi usada no modelo híbrido, tema central da tese, uma revisão aprofundada sobre a transformada Wavelet foge ao escopo da presente pesquisa. Assim, busca-se, tão somente, revisar trabalhos nos quais o espectro cruzado e a coerência espectral Wavelet tenham sido utilizados e apresentar características gerais da transformada Wavelet, comparando-a com a HHT.

A Wavelet também utiliza a divisão do domínio em janelas, de forma semelhante à STFT, porém, o comprimento da janela se adapta às características da função a ser analisada. Sendo assim, a STFT pode ser considerada como um caso particular da Wavelet. (PINTO, 2009).

Outra diferença em relação à STFT está na base de decomposição do sinal, não senoidal no caso da Wavelet e definida a partir da ondeleta mãe.

O problema da não-estacionariedade é superado pela Wavelet. Variando-se parâmetros da base, séries com energia distribuída de forma não estacionária em várias frequências podem ser analisadas. No entanto, a desvantagem está na escolha da base matemática a

priori. Assim, se um conjunto de wavelets inapropriado for escolhido, a decomposição do sinal pode levar a resultados desvinculados da realidade, com baixa resolução temporal em baixas frequências e com resolução de frequência baixa em altas frequências.

Para séries temporais, a transformada Wavelet Contínua - CWT, é definida como a convolução da série com as bases ou Wavelets locais, conforme a equação (4-1), sendo $x(t)$ a série temporal, ψ_s a Wavelet e t o tempo.

$$W_{x,\psi}(s,t) = (x(t) * \psi_s(t)) \quad (4-1)$$

Assim, a função Wavelet é definida no espaço tempo-frequência, com média zero e com resolução Δt e $\Delta \omega$, no tempo e na frequência. Através da manipulação dos parâmetros de escala (s), a base pode ser deformada ou trasladada, adaptando-se a resolução no tempo e na frequência.

Com a convolução, o sinal é decomposto em modos com variabilidade temporal, dando origem a um espaço tempo-frequência. Como outros processos de filtragem, a transformação sofre de efeitos de contorno, que podem ser caracterizados por um cone de influência, variando com a resolução adequada em cada banda de frequência.

A energia espectral Wavelet é definida por $|W_{x,\psi}|^2$. Já o espectro cruzado entre duas séries temporais X e Y é definido como:

$$W_{XY}(s,t) = W_{X,\psi}(s,t) W_{Y,\psi}^*(s,t) \quad (4-2)$$

O asterisco denota o conjugado complexo. A energia do espectro Wavelet cruzado é definida como $|W_{XY}(s,t)|$.

Com a fase local de $W_{X,\psi}(s,t)$ e $W_{Y,\psi}(s,t)$ sendo dada pelos respectivos argumentos complexos, pode-se estimar a diferença de fase (defasagem) local entre os sinais. Para estimar a diferença de fase entre os sinais é preciso trabalhar com uma média e um intervalo de confiança. Também devem-se estabelecer intervalos de confiança para a

energia do espectro cruzado. GRINSTED *et al.* (2004) utilizaram a média circular das fases em regiões com significância maior que 5% e fora da influência dos contornos para o cálculo das defasagens.

Enquanto os espectros cruzados revelam áreas com energia comum, a coerência espectral Wavelet – WTC pode ser vista como um coeficiente de correlação localizado no espaço tempo-frequência, definido como (JEVREJEVA *et al.*, 2003):

$$C_{XY}(s,t) = \frac{\left| S\left(s^{-1}W_{XY}(s,t)\right) \right|^2}{S\left(s^{-1}\left|W_{X,\psi}(s,t)\right|^2\right) \cdot S\left(s^{-1}\left|W_{Y,\psi}(s,t)\right|^2\right)} \quad (4-3)$$

O operador S , definido pela equação (10) de GRINSTED *et al.* (2004), é usado para suavização na interpolação do escalograma tempo-frequência. Os resultados podem ser sensíveis a tal operador. A defasagem dos sinais é achada como no espectro cruzado. Para revelar regiões de coerência espectral que não são mera coincidência, a significância dos espectros Wavelet dos sinais pode ser obtida com a hipótese nula de que os mesmos foram gerados por processos estacionários com determinada distribuição paramétrica. Essas distribuições são como fatias no espaço tempo-frequência, com determinada resolução temporal. Então, os intervalos com determinado nível de significância podem ser estimados. Mais detalhes sobre a significância e sobre a suavização no espaço e na frequência podem ser vistos em (TORRENCE & COMPO, 1998).

Relações no domínio do tempo-frequência foram estabelecidas entre a Oscilação do Ártico -AO e a área de cobertura máxima de gelo no mar Báltico utilizando o espectro cruzado Wavelet e a coerência espectral por GRINSTED *et al.* (2004). Os autores buscaram, além de evidenciar as conexões entre as séries, explorar suas relações de fase, buscando evidenciar causa-efeito.

A coerência espectral e o espectro cruzado Wavelet foram aplicados a dados atmosféricos e oceanográficos de larga escala e hidrologia em escala continental por LABAT (2010). O autor utilizaram uma estimativa da descarga fluvial dos cinco continentes entre 1876 e 1994, baseada em 221 séries temporais das bocas dos principais rios do mundo, obtendo seus espectros cruzados com a Oscilação do Atlântico Norte – NAO, com a AO, com o

índice de TSM do Pacífico NINO3.4, com o Índice de Oscilação Sul – SOI e com a Oscilação Decadal do Pacífico. A investigação focou em fenômenos de escalas interanuais a decadais. No caso da América do Sul, as vazões foram compostas, principalmente pelos rios Amazonas, Orinoco e Paraná. Os resultados revelaram coerência espectral relacionando as descargas sul-americanas tanto ao SOI como ao NINO3.4 na faixa de variabilidade interanual de 2-8 anos. A NAO, com maior influência ao norte da América do Sul, mostrou-se fator principal na variabilidade interanual entre 1920 e 1935 das descargas. Além disso na faixa multidecadal, houve coerência entre as vazões sul-americanas e a PDO.

4.3. A EMD e a HHT

A HHT (HUANG *et al.*, 1998) surge como método de análise espectral para dados não estacionários e não lineares composto por duas etapas: a EMD e a análise espectral de Hilbert – HSA, sigla em inglês para *Hilbert Spectral Analysis*.

O caráter adaptativo da EMD origina um conjunto de Funções Intrínsecas Modais – IMFs, sigla em inglês para *Intrinsic Mode Functions*, sem expressão analítica pré-definida, de classe L^2 e que possuem Transformadas de Hilbert. Tal característica faz com que os dados “falem por si”, contemplando não-linearidade e não-estacionariedade e resultando em um espectro com resoluções tempo-frequência adequadas.

A Transformada de Hilbert não foi objeto da presente pesquisa e sua revisão mais detalhada fugiria ao escopo, podendo-se consultar BARNHART (2011) e em HUANG *et al.* (1998). Importantes aplicações da HHT no estudos de séries temporais hidrológicas estão em RAMACHANDRA & HSU (2008).

4.4. EMD

O método consiste na identificação empírica dos modos oscilatórios intrínsecos pelas suas escalas temporais características para decompor o sinal adequadamente. Isso é feito através de um processo de peneiramento iterativo descrito abaixo, sendo $x(t)$ o sinal a ser decomposto:

- a) busca dos máximos locais M_i , $i=1, 2, \dots$, e dos mínimos locais m_k , $k=1, 2, \dots$, da série;

- b) interpolação da envoltória superior $M(t)$ e da inferior $m(t)$ do sinal, respectivamente pelos pontos M_i e m_k ;
- c) cálculo de $e(t) = (M(t)+m(t))/2$;
- d) substituição de $x(t)$ por $x(t) - e(t)$;
- e) retorno ao passo (a), parando quando $x(t)$ for uma IMF $\varphi(t)$;
- f) remoção da IMF $\varphi(t)$ obtida do sinal, substituindo $x(t)$ por $x(t) - \varphi(t)$, retornado a (a) se o resultado tiver mais que um extremo.

Duas condições precisam ser satisfeitas por uma IMF (HUANG *et al.* 1998): que o número de extremos e o número de zeros sejam iguais ou difiram no máximo por um (critério usado para a verificação de parada do passo (e) e passagem para (f)); em qualquer ponto, a média local das envoltórias de máximos e mínimos deve ser zero.

Caso as envoltórias encontradas nos passos (a) e (b) fossem simétricas em relação ao valor zero, seria o caso de um sinal com amplitude modulada (AM). Portanto o peneiramento é uma forma iterativa de remoção da assimetria entre as envoltórias, no qual baixas frequências vão sendo extraídas do sinal até que o resultado seja uma IMF AM. Mas, como a frequência instantânea pode variar em cada IMF, pode-se dizer que a EMD é uma decomposição AM/FM do sinal (RATO *et al.*, 2008).

Ao mesmo tempo em que o caráter empírico da EMD a confere adaptabilidade necessária ao tratamento de sinais reais, também a priva de formalidade matemática. Sendo assim, pequenas modificações no algoritmo podem interferir no resultado. A seguir, apresenta-se uma revisão dos principais problemas da EMD e algumas alternativas achadas na literatura.

Problema dos contornos

Trata-se da impossibilidade de interpolação das envoltórias da etapa (b) da EMD nos contornos da série. Entende-se por contorno, o trecho inicial e final da série temporal. Os comprimentos das envoltórias a serem interpoladas não coincidirão com a série original, mas com as séries M_i , $i=1, 2, \dots, I$ para a envoltória de máximos e m_k , $k=1, 2, \dots, K$ para a envoltória de mínimos.

Tal dificuldade de interpolação é também chamada de efeito de contorno e afeta todos os métodos de análise de sinais (HUANG & WU, 2008). O problema dos contornos é o mais sério para a aplicação pretendida na presente pesquisa e é ilustrado na Figura 4-1. Foram pesquisadas diversas alternativas de solução para o problema.

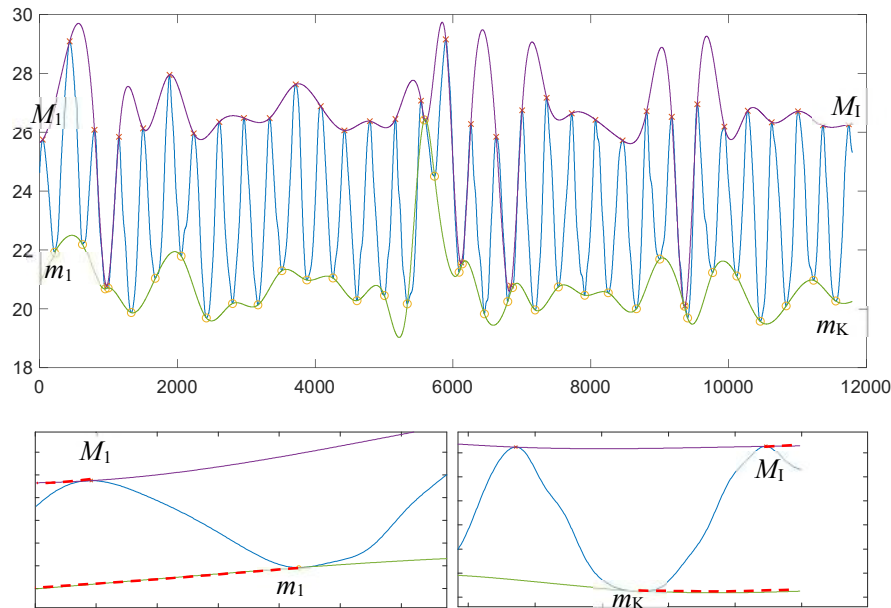


Figura 4-1. (Acima) Série NINO1+2 durante a primeira iteração da EMD com envoltórias, destacando-se os extremos dos contornos; (abaixo) detalhes dos trechos de contorno da série sem pontos extremos para a interpolação das envoltórias no início e no fim da série (linhas tracejadas vermelhas).

No caso da EMD, a interpolação das envoltórias da etapa (b) é feita ajustando-se splines cúbicas, uma para o conjunto de máximos e outra para o de mínimos. Como os extremos não coincidirão com os contornos da série, haverá a necessidade de extrapolação das envoltórias (splines), requerendo ao menos um ponto de apoio além dos extremos para extrapolar cada curva. Essa operação sempre introduzirá sinais espúrios na EMD.

Um método misto de extensão dos extremos foi adotado por ZHANG *et al.* (2010), no qual utilizou-se espelhamento dos extremos para extensão da envoltória em altas frequências e ajuste de polinômio por mínimos quadrados para extrapolação em baixas frequências. Os resultados mostraram ser vantajosa a utilização seletiva dos dois métodos, quando comparados à utilização clássica de apenas um método de extensão dos extremos durante todo o peneiramento da EMD.

A extrapolação das envoltórias de extremos nos contornos pode ser feita utilizando-se a tendência linear dos conjuntos de máximos e de mínimos nas proximidades dos extremos, como apresentado por YANG *et al.* (2015).

Já em TIANLU *et al.* (2013), as informações das envoltórias são completadas utilizando-se valores extremos de partes similares da série. Procede-se uma busca de máxima similaridade, tendo como base as ondeletas contendo os extremos das regiões de contorno. Assim, supondo que se queira completar o início da série e que o primeiro extremo seja um máximo M_1 , a ondeleta entre M_1 e o segundo máximo M_2 será tomada como base, varrendo-se a série de máximo em máximo, buscando trecho de mínimas diferenças quadradas. As informações à esquerda da ondeleta selecionada são usadas nesse caso no ponto inicial da série, apoiando a nova interpolação da envoltória. Os autores usaram três índices para medir o desempenho da metodologia: o primeiro relacionado às mudanças de energia entre o somatório de IMFs e o sinal original; o segundo, para medir correlação entre modos achados e componentes conhecidas em um sinal sintético; o terceiro índice é relativo ao tempo de processamento computacional. Segundo os autores, o método melhora a acurácia da EMD.

Mixagem entre Modos

Outra grande dificuldade durante a EMD são as mixagens de frequências entre modos. Elas se caracterizam pela ocorrência de escalas muito distintas em uma mesma IMF ou pela ocorrência de escalas similares em IMFs distintas. Como consequência, essa mistura de escalas se reproduz no domínio do tempo-frequência e as IMFs podem perder significado físico (WU & HUANG, 2009). Tal situação é ilustrada na Figura 4-2.

Existem duas causas principais para essa mistura: a primeira é quando o sinal a ser decomposto caracteriza-se por possuir espectro de energia cheio, como é o caso dos ruídos $1/f$ característicos do clima tropical e vistos nessa revisão; a segunda é a presença de fenômenos intermitentes no sinal, por exemplo os eventos com características de pulsos, também observados com frequência nos trópicos.

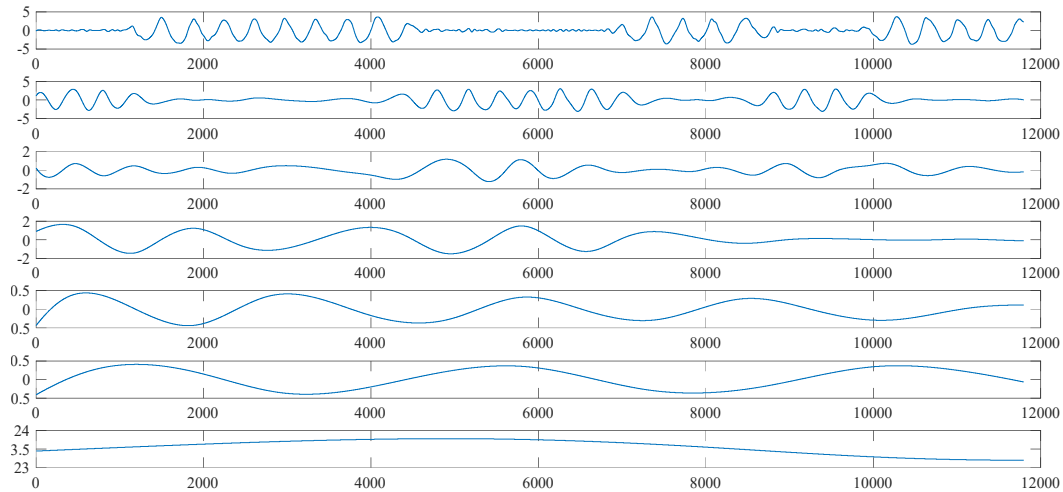


Figura 4-2. Exemplo de ocorrência de mixagem entre modos na EMD, com eixos verticais em graus Celsius e eixo horizontal em dias: as cinco primeiras IMFs de cima para baixo apresentaram mixagens entre modos vizinhos; a causa provável são efeitos intermitentes (pulsantes) em escala intrasazonal no Pacífico Oeste que influenciam a TSM no Pacífico Leste, mormente causados pela OMJ.

Para lidar com o problema de separação de escalas, surgiu a decomposição empírica modal em conjunto, livre tradução de *Ensemble Empirical Mode Decomposition – EEMD* (WU & HUANG, 2009). Nela, as IMFs resultantes são definidas como as médias de um conjunto de repetições da EMD, cada qual aplicada ao sinal a ser decomposto acrescido de um ruído branco de amplitude finita. A introdução do ruído preenche o domínio tempo-frequência, de forma que bandas distintas do sinal são projetadas sobre o fundo de referência assim criado. Os efeitos do ruído branco são estatisticamente removidos com a promediação de um conjunto suficientemente grande de tentativas, já que os resultados individuais são distintos entre si.

Outra abordagem buscou reduzir a mixagem de modos introduzindo pseudo-extremos nas séries. Trata-se da decomposição empírica modal compacta, livre tradução para *Compact Empirical Mode Decomposition – CEMD* (FAN *et al.*, 2012). Relembrando que a mixagem de modos é causada pela distribuição irregular de extremos em séries temporais, quer seja pela presença de eventos pulsantes ou pela coexistência de frequências distintas, mas não o bastante para serem peneiradas separadamente.

A CEMD foi pensada para o caso de pulsos de alta frequência surfando oscilações de frequência mais baixa. De forma geral, a introdução de pseudo-extremos age fazendo com que as envoltórias adiram nas oscilações de menor frequência e se descolem nos extremos

dos pulsos que se desejam extrair. Porém, a tarefa de separar os pulsos da oscilação de baixa frequência pode ser complicada em sinais reais.

A Figura 4-3 ilustra a CEMD da série de níveis diários do posto de Itapéua, localizada no rio Solimões, a montante de Manaus. O objetivo é destacar as dificuldades de utilização da metodologia quando os pulsos se apresentam distribuídos de forma irregular, ora perto dos picos sazonais, ora na fase de enchimento da calha, por vezes isolados, e em outras, agrupados. Pode-se especular que alguns desses pulsos sejam reflexos hidrológicos de eventos de escala intrasazonal. Outros, no entanto, podem ser consequência do efeito de remanso no rio Solimões, por conta das baixas declividades da bacia.

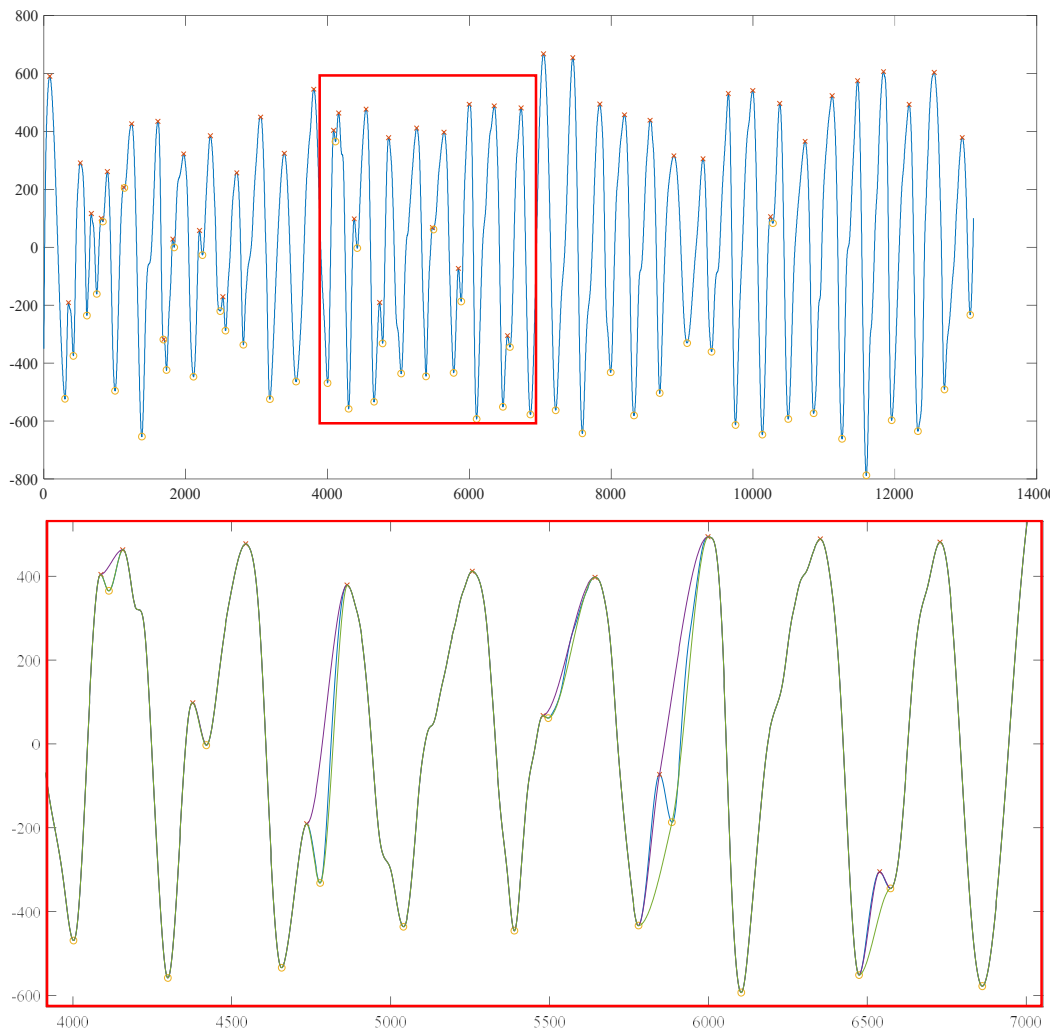


Figura 4-3. CEMD de série diária de níveis de Itapéua, no rio Solimões, com eixos verticais em cm e eixos horizontais em dias: (acima) máximos e mínimos destacando presença intermitente de

frequências intrasazonais; (abaixo) detalhe mostrando as envoltórias (a de máximos é roxa e a de mínimos é verde) aderidas ao sinal (azul) graças ao uso dos pseudo-extremos ao longo da série.

Cruzamento de Envoltórias

Quando as envoltórias interpoladas de máximos ou de mínimos se cruzam com o sinal durante o peneiramento, ou ainda pior, quando elas se cruzam entre si, sinais espúrios são introduzidos no processo, contaminando as IMFs e comprometendo sua interpretação física. Fatores como a preservação de derivadas mais altas na interpolação via spline aliada à irregularidade na distribuição dos extremos da série (causada por pulsos ou por interação entre frequências próximas) podem ocasionar o problema. Ainda, dificuldades de extensão das envoltórias pelo problema dos contornos também podem acarretar o problema de cruzamento descrito.

Assim, as abordagens apresentadas para mitigação da mistura de modos e do problema dos contornos contribuem indiretamente para evitar o cruzamento de envoltórias. Em certos casos, isso ocorreu através da proposição de novas interpolações para as envoltórias para resolver os contornos (YANG *et al.*, 2015), (TIANLU *et al.*, 2013) e a mistura de modos (FAN *et al.*, 2012).

Outras pesquisas foram dedicadas mais especificamente ao problema dos cruzamentos de envoltórias. A interpolação spline cúbica - ISC, largamente utilizada, é descrita em ZHU *et al.* (2013) como causadora do cruzamento de envoltórias, por conservar a suavidade da curva garantindo a existência da primeira e segunda derivadas. Além disso, segundo os autores, a ISC é um método de interpolação global, fazendo com que mudanças locais interfiram em toda a curva, o que não seria adequado para sinais não-estacionários. Assim, os autores propuseram interpolar as envoltórias usando polinômios cúbicos Hermitianos por trechos de interpolação, definindo suas derivadas de forma a tornar a curva interpolada menos sinuosa.

Com finalidade semelhante, KOKES & NGHIEN (2012) propuseram usar a Spline Cúbica Forçada, livre tradução de *Constrained Cubic Spline*, descrita em (KRUGER, n.d.). Essa parece uma opção promissora para a EMD, podendo melhorar diversos aspectos durante o peneiramento. Porém, não foram encontradas ferramentas disponíveis para sua utilização no MATLAB.

4.5. A DFA

A Análise Destendenciada de Flutuações, livre tradução para *Detrended Fluctuation Analysis* – DFA foi desenvolvida para lidar com a análise de séries temporais de intervalos de batimentos cardíacos. Tais séries são altamente não-estacionárias, apesar de se falar em ritmo senoidal ao descrevê-las. A DFA propõe-se a distinguir as tendências causadas por ruídos externos, também não estacionários, das flutuações intrínsecas ao sistema em análise.

De forma geral, as flutuações causadas por estímulos não correlatos ao sistema podem ser separadas (decompostas) dos estímulos intrínsecos ao sistema, podendo-se visualizar as diferenças nas propriedades correlacionais dessas duas classes de estímulos. Para tanto, supõe-se de forma plausível que as flutuações oriundas da dinâmica do sistema, complexo e não-linear, apresentem correlações de longo alcance. Outras respostas devem originar um tipo distinto de flutuação, com escala temporal característica relacionada ao estímulo, apesar de altamente não-estacionária também. Esses ruídos externos podem ser tratados como tendências, podendo ser separados das flutuações mais sutis ocasionadas pelo sistema (PENG *et al.*, 1995).

Cabe ressaltar que os ruídos, apesar de serem tratados como tendências externas, com escalas diferenciadas, tem importância na caracterização do sinal.

A DFA é uma análise do valor médio quadrático – VMQ, da variabilidade em relação à tendência da série, analisada por escalas. Suas etapas são enumeradas a seguir.

- 1) A série discreta de dados $B(k)$ é deslocada pelo valor da média $\langle B \rangle$ e integrada por soma cumulativa:

$$y(k) = \sum_{i=1}^k [B(i) - \langle B \rangle]$$

- 2) Então, a série é segmentada em janelas não sobrepostas de comprimento Δt e, em cada janela, uma regressão polinomial $y_{\Delta t}(k)$ é ajustada à série (geralmente linear, quadrática ou cúbica), como ilustra a Figura 4-4;
- 3) Por último, o resíduo médio quadrático $F(\Delta t)$ é calculado

$$F(\Delta t) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_{\Delta t}(k)]^2} \quad (4-4)$$

sendo N o número total de pontos da série. Nota-se que $F^2(\Delta t)$ pode ser encarado como a média da soma do quadrado dos resíduos achados em janelas de escala Δt . A escolha da ordem n do polinômio de regressão é denotada como DFA_n . A falta da especificação de n refere-se a DFA_1 , nesse caso. A capacidade de eliminação de tendências da série depende da escolha da ordem do polinômio. Tendências de ordem n no perfil $y(k)$ (ou $n - 1$ na série original) são eliminadas.

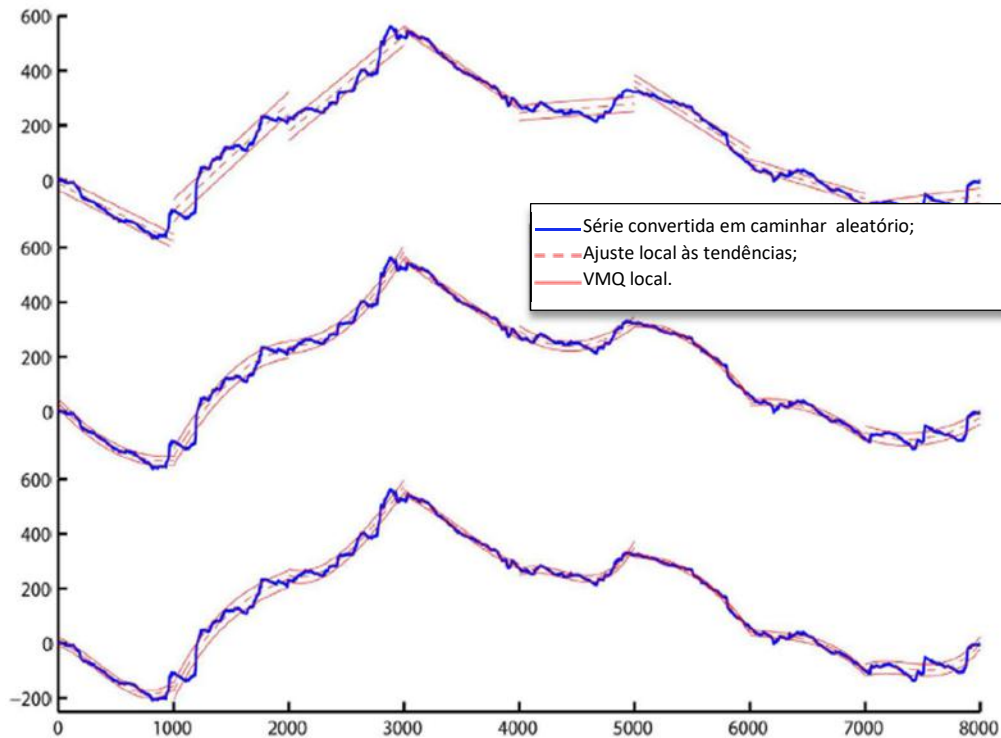


Figura 4-4. Computação do VMQ das flutuações em relação à tendência linear (A), quadrática (B) e cúbica (C). A linha vermelha tracejada é o ajuste à tendência, com oito janelas de tamanho 1.000. A distância entre as linhas sólidas vermelhas e a tendência vermelha tracejada é de 1 VMQ. Adaptado de IHLEN (2012).

Esse procedimento caracteriza a auto-afinidade através da medida da dispersão do resíduo das flutuações integradas em relação ao regressor $y_{\Delta t}(k)$ em diferentes resoluções

(tamanhos de janelas). Se o escalonamento $F(\Delta t)$ versus Δt for uma lei de potência, espera-se ajuste linear para uma plotagem bilogarítmica (“log-log”):

$$F(\Delta t) = C(\Delta t)^\alpha \Rightarrow \ln(F(\Delta t)) = \alpha \ln(\Delta t) + \ln(C)$$

Sendo C uma constante. O expoente de escalonamento α pode ser estimado por ajuste linear por mínimos quadrados, sendo uma medida da auto-correlação na série e uma estimativa do expoente de Hurst H .

4.6. Análise de Componentes Principais – PCA e Regressão Linear Parcial de Mínimos Quadrados - PLS

A PCA e a PLS são revisadas, destacando-se sua capacidade de remediar problemas de ranqueamento e de condicionamento em sistemas lineares. No entanto, limitando-se ao escopo do presente trabalho, pré-requisitos teóricos da Regressão Múltipla Linear, por exemplo, devem ser buscados.

A característica mais marcante da PLS é que ela combina características da PCA e da regressão múltipla, sendo especialmente útil em casos nos quais se tenham múltiplos preditores com determinado grau de colinearidade entre alguns deles, ou quando o número de amostras disponíveis para a regressão seja insuficiente para prevenir a multicolinearidade. O problema da multicolinearidade relacionado às RNAs *feedforward* é revisado a parte.

Dado um conjunto de l variáveis dependentes com n amostras, compondo uma matriz $\{\mathbf{Y} \in \mathbb{R}^{n \times l}\}$ e um conjunto de k preditores armazenados em $\{\mathbf{X} \in \mathbb{R}^{n \times k}\}$. A PLS busca a predição de \mathbf{Y} usando \mathbf{X} , escolhendo um subconjunto de regressores ótimos para explicar a covariância entre \mathbf{X} e \mathbf{Y} . Para os fins pretendidos na presente pesquisa, podem-se concentrar esforços em problemas multivariados, porém com apenas uma variável dependente, sendo \mathbf{Y} um vetor nesse caso.

Considerando-se primeiramente o problema da multicolinearidade, a PCA é uma abordagem que projeta os preditores através das direções ortogonais sucessivas de máxima variabilidade de \mathbf{X} . Assim, podem-se eliminar os preditores de menor

contribuição e tomar componentes ortogonais principais para a regressão de Y . A ortogonalidade garante a eliminação da colinearidade.

Porém, a composição assim feita baseia-se nas características apenas dos preditores, não garantindo que os componentes principais sejam bons preditores. A PLS generaliza a PCA ao procurar por conjuntos de componentes frutos da decomposição simultânea de X e Y , considerando a condição de que tais componentes expliquem o máximo possível da covariância entre preditores e alvos.

Assim, na PLS, X e Y são decompostos como um produto de um conjunto comum de fatores ortogonais que representam as projeções nas direções principais e um conjunto de vetores que representam os cossenos dos ângulos de projeção. Cada fator é originado por uma combinação linear de variáveis independentes originais.

Ao maior autovalor, corresponderá o principal autovetor, que define a direção no espaço das variáveis independentes X responsável por descrever a máxima variância em Y . Na sequência, a variância restante das amostras (não explicada pelo primeiro componente principal) será explicada por um segundo componente principal ortogonal ao primeiro. A projeção prossegue até o limite no qual o número de componentes principais é igual ao número de variáveis independentes.

Da forma descrita, um conjunto de componentes principais será encontrado para os preditores e outro conjunto separado será achado para os alvos, conforme descrito por GELADI & KOWALSKI (1986). Após o processo, o conjunto de preditores pode ser reduzido a um subconjunto capaz de descrever grande parte da variabilidade das variáveis dependentes.

Um exemplo de utilização conjunta da PLS e de RNAs é vista em YU *et al.* (2017). Os autores apresentaram um modelo híbrido PLS-RNA para predição da pontuação de preferências dos consumidores por sabores de chás verdes prontos para beber. No modelo híbrido, as relações entre os componentes dos sabores de chás (preditores) e as preferências dos consumidores foi inicialmente estabelecida pela PLS. Depois, o resíduo do ajuste foi usado como saída no treinamento de uma RNA tendo os mesmos preditores. Os resultados mostraram desempenhos superiores do modelo RNA e do modelo híbrido

PLS-RNA em relação ao modelo linear PLS, porém de forma similar, considerando erro médio quadrático e coeficiente de determinação R^2 .

5. Redes Neurais

Esse item revisa conceitos e aspectos importantes das RNAs, embasando a investigação sobre os impactos do pré-processamento de dados via EMD. Para tanto, primeiro apresentam-se de forma geral, os conceitos básicos, com a topologia, o processo de treinamento de RNAs e suas funções de ativação. Em sequência, aspectos relacionando parametrização, generalização e convergência das RNAs são revisados, considerando uma contextualização histórica da inteligência artificial – IA, e os aspectos inerentes ao superdimensionamento, ao mal condicionamento e ao mal ranqueamento nos algoritmos de interesse. Por fim, as estratégias de generalização adotadas, quais sejam, parada antecipada e regularização, são revisadas.

5.1. Conceitos básicos

A utilização de RNAs ganhou vulto a partir da década de 80, como alternativa à computação algorítmica tradicional, sobretudo a partir da descrição do algoritmo *back-propagation* (RUMELHART *et al.*, 1986) permitindo solução de problemas não linearmente separáveis.

Em *data mining*, os esquemas *feedforward*, *back-propagation multi-layer Perceptron* – MLP, são mais comumente usados, existindo muitas outras topologias. Nesta pesquisa, as aplicações serão voltadas para modelagem de séries com RNAs *feedforward* MLP.

RNAs são sistemas de processamento paralelo e distribuído, compostos de unidades nodais capazes de aprender experimentalmente, através de sinapses, de forma análoga ao cérebro humano.

A computação neural não é feita com algoritmos convencionais, o que fez das RNAs uma alternativa para solução de problemas envolvendo funções de qualquer natureza. Os nodos são distribuídos em uma ou mais camadas interligadas por conexões, na maioria dos modelos, associadas a pesos. Os pesos propiciam o armazenamento de conhecimento pela rede, pois vão sendo modificados de acordo com os erros estimados para cada nó. A informação apresentada à rede em uma camada de entrada vai sendo transferida pelas conexões e é ponderada pelos pesos atribuídos aos neurônios.

A Figura 5-1 (superior) ilustra a estrutura interna de um neurônio. Nela, $x_i, i = 1, 2, \dots, k$, são as entradas do neurônio, $w_i, i = 1, 2, \dots, k$, são os pesos sinápticos, sendo w_0 um valor limiar e y é a combinação linear das entradas ponderadas pelos respectivos pesos e por w_0 . O valor da saída y sofre ainda uma transformação, geralmente não linear, pela função de ativação, acrescentando a complexidade necessária para aproximação da solução.

Na parte inferior da Figura 5-1 apresenta-se um esquema geral de uma RNA, com uma camada de entrada, uma intermediária e uma de saída. Repare que as conexões são entre nós de diferentes camadas. No caso, cada nó de uma camada conecta-se com todos os nós da camada subsequente.

O paralelismo inerente à estrutura das RNAs faz delas uma alternativa atraente na solução de problemas complexos, possibilitando desempenho superior aos modelos convencionais (BRAGA *et al.*, 1998), sobretudo em problemas envolvendo não linearidade e não estacionariedade. Mas, de forma geral, tal vantagem está intimamente relacionada à arquitetura adotada para a rede: a quantidade de camadas intermediárias e de nós nelas. A escolha da função de ativação e a forma de computá-la também influenciam. Além desses fatores, a escolha do algoritmo de treinamento influencia no desempenho da RNA.

A solução de problemas com RNAs passa inicialmente pela fase de aprendizagem ou treinamento da rede, na qual as informações de exemplos apresentados à rede, necessárias para a representação das saídas, são armazenadas automaticamente nos pesos dos nodos. Ainda segundo BRAGA *et al.* (1998), a capacidade de aprender através de exemplos e de generalizar a informação aprendida são os atrativos principais das RNAs. No entanto, as redes fazem mais do que mapear relações entre entradas e saídas, extraindo informações não explicitadas nos exemplos.

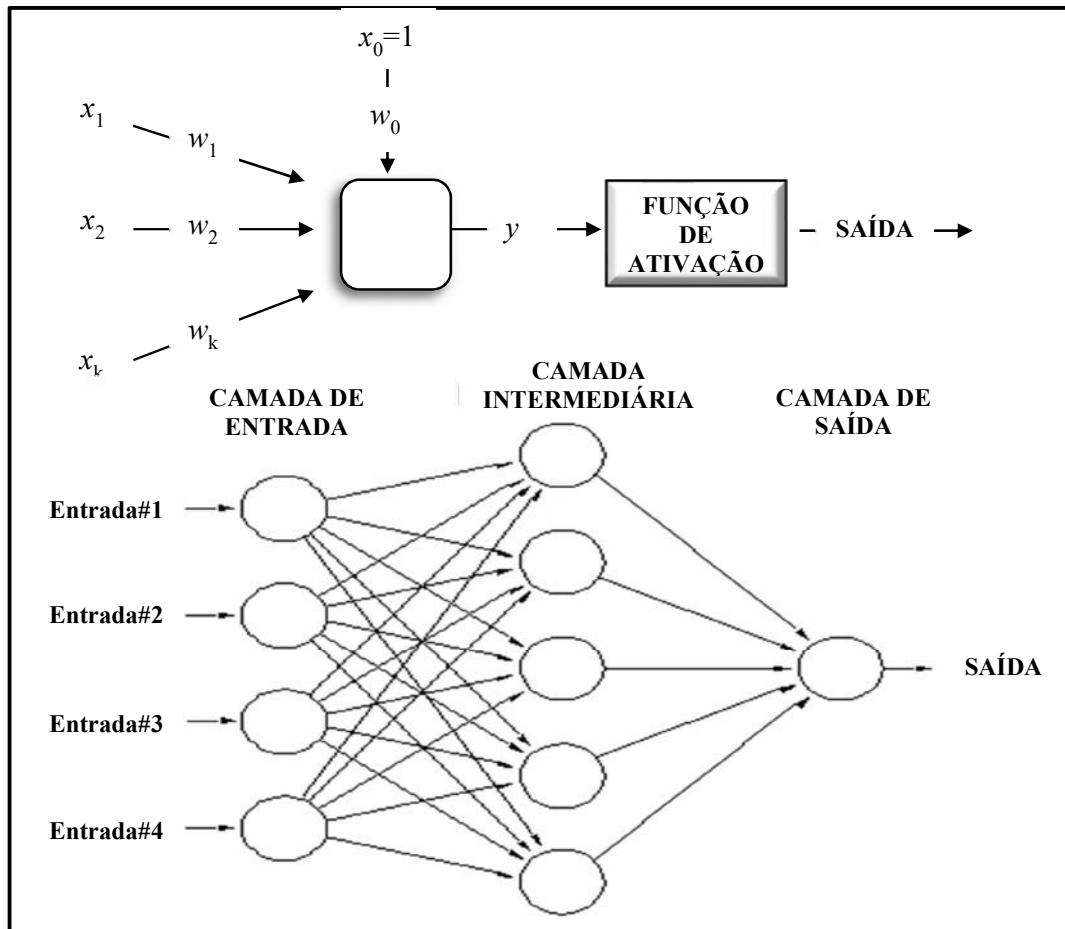


Figura 5-1. Esquema de um neurônio e de uma RNA *multi-layer Perceptron* – MLP.

Topologia

Fatores como o número de exemplos disponíveis para o treinamento, quantidade de ruído presente no sinal, complexidade da função a ser aprendida e distribuição estatística dos dados de treinamento influenciam na arquitetura da rede, cuja configuração mais adequada é obtida de forma experimental.

O interesse da presente pesquisa está nas redes *feedforward*, nas quais os neurônios são organizados em camadas interligadas sem, no entanto, haver realimentação e ligação entre neurônios de uma mesma camada.

CYBENKO (1989) demonstrou formalmente pela primeira vez que uma RNA *feedforward* de uma camada intermediária pode aproximar qualquer função contínua. Enquanto duas camadas intermediárias são suficientes para aproximar qualquer função matemática. A utilização de duas ou mais camadas pode facilitar o treinamento da rede. Porém, um

número grande de camadas não é apropriado, pois o erro medido durante o treinamento propaga-se da saída para camadas anteriores, sendo menos útil e preciso com muitas camadas.

A quantidade de nós nas camadas intermediárias é definida empiricamente, mas depende da quantidade de exemplos de treinamento, da quantidade de ruído nos exemplos, da complexidade da série a ser aprendida e da distribuição estatística dos dados de treinamento (BRAGA *et al.*, 1998). Como será discutido especificamente ao falar-se de estratégias de generalização, um número de conexões muitas vezes menor que a quantidade de exemplos de treinamento evita o *overfitting* (reprodução de ruídos do sinal modelado). Por outro lado, com um número excessivamente baixo de nós a rede pode não aproximar satisfatoriamente o fenômeno simulado.

A definição da topologia é feita, portanto, de forma que a rede se adeque da melhor forma possível ao problema, considerando o conjunto de exemplos disponível e a necessidade de aproximação com boa generalização. A rede, no entanto, deve ser capaz de se adaptar aos dados de saída sem ser excessivamente flexível a ponto de reproduzir ruídos presentes devidos a erros de medição da série alvo.

Treinamento

De forma resumida, treinamento é a etapa durante a qual os pesos das sinapses são ajustados baseando-se na comparação de exemplos de entradas apresentados à rede e suas respectivas saídas.

O treinamento de redes MLP foi um empecilho à disseminação das RNAs, até o surgimento do algoritmo *back-propagation*, o mais utilizado. Ele é constituído de duas fases: *forward* e *backward*. Na primeira, uma resposta da rede é obtida para determinada entrada. Na segunda, tal resposta é usada para estimar os erros nas camadas intermediárias a partir da saída correspondente. A regra Delta é usada na estimativa dos erros intermediários para ajuste dos pesos.

Descrevê-la foge ao escopo do trabalho, mas cabe citar que ela requer a utilização de funções de ativação contínuas, diferenciáveis e, geralmente, não decrescentes, sendo chamadas de funções semi-lineares. (BRAGA *et al.*, 1998). Isso ocorre pela necessidade

do cálculo do gradiente do erro em relação aos pesos na regra Delta, ou seja, a diferenciação da função de ativação.

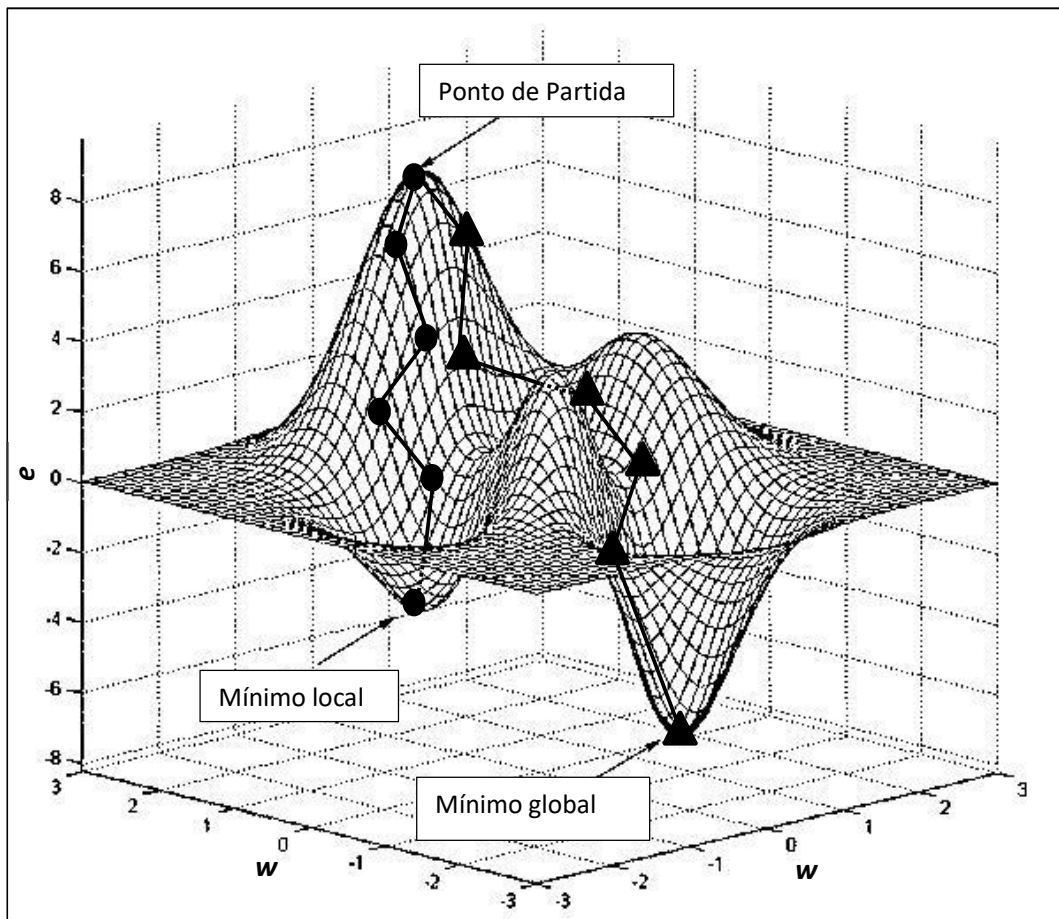


Figura 5-2. Ilustração de superfície de erro em função dos pesos w_1 e w_2 , com dois caminhos distintos de ajuste: círculos, iterações levando a um mínimo local; triângulos, iterações levando à solução – adaptado de SERGIU (2011).

Visualizar um espaço N-dimensional, sendo N o número de conexões (e de pesos) na RNA, seria uma tarefa impossível. A Figura 5-2 ilustra o treinamento com duas conexões, originando um espaço bidimensional. São mostrados dois caminhos hipotéticos do algoritmo, sendo a dimensão vertical o erro e as demais, pesos a serem ajustados. Cada ponto denota uma iteração na busca do mínimo.

Pela regra Delta, o treinamento segue ajustes dos pesos na direção do gradiente do erro, buscando um mínimo com a máxima taxa de ajuste. Mas na ilustração, o caminho com

círculos levou a um mínimo local. Esse é um dos maiores problemas no treinamento de redes MLP. A escolha de parâmetros do treinamento pode modificar muito sua eficiência, mas segue como uma tarefa pouco conhecida.

Algumas técnicas existem para acelerar o treinamento e reduzir incidência de mínimos locais. Uma opção é a adição de termo à fórmula de ajuste de erro (termo momentum), aumentando a velocidade de aprendizado, evitando platôs e mínimos locais;

O treinamento de redes MLP via *back-propagation* pode esbarrar ainda no problema do *overfitting* já mencionado. Algumas alternativas são encerrar o treinamento quando o erro começa a subir, usar uma penalização para os valores dos pesos através de uma técnica de regularização ou podar pesos da rede (*pruning*). O problema do *overfitting*, a parada usando a validação cruzada e a regularização serão revisadas em mais detalhes a frente.

Função de ativação

A não linearidade no tratamento de problemas é importante característica de RNAs, fazendo com que soluções se aproximem mais de dados reais. As componentes presentes em RNAs responsáveis por introdução da não linearidade no processo são portas de limiar quadráticas e funções de ativação. Vamos nos ater às funções de ativação, pois as portas de limiar não lineares são menos comuns.

As funções de ativação devem ser diferenciáveis. Isso porque na regra Delta o ajuste de pesos segue na direção do gradiente do erro, calculado com a derivada da função de ativação. Muitas funções não lineares satisfazem essa condição, mas são, em geral, difíceis de implementar computacionalmente.

O custo computacional da diferenciação da função de ativação é alto, exigindo esforços para aumentar a eficiência em detrimento da acurácia. Sanar tal problema é tema alvo de diversas pesquisas na área de RNAs (MUÑOZ *et al.*, 2008). A sigmoideal é largamente usada como função de ativação. A Figura 5-3 ilustra alguns tipos de funções de ativação.

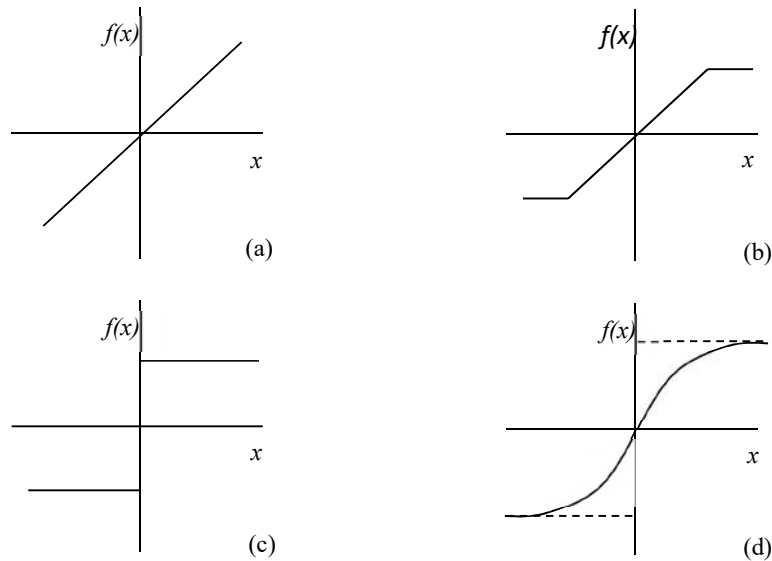


Figura 5-3. Exemplos de funções de ativação: (a) linear; (b) linear com limiar nos extremos; (c) limiar; (d) sigmoidal.

Por fim, a escolha da função sigmoidal como função de ativação ocasiona um dos principais problemas do *back-propagation*. A derivada da sigmoidal se aproxima de zero quando a saída do nodo se aproxima de zero ou de um. O ajuste de pesos passa a ser insuficiente, já que depende do valor dessa derivada. As estratégias para contornar tal problema envolvem, basicamente, forçar o erro a assumir valores muito altos, adicionar valor constante à derivada ou usar uma função de erro não linear. Tal característica é apontada como uma possível causadora de mal-condicionamento do Jacobiano da função objetivo, como apresentado mais à frente na revisão.

5.2. Parametrização, generalização e convergência

Pesquisas que relacionam a parametrização de RNAs, sua capacidade de generalização e aspectos de convergência são revisadas nesse item.

Busca-se base teórica para caracterizar possíveis mudanças na capacidade de generalização e na convergência do modelo híbrido. De antemão, sabe-se que a EMD causará aumento do número de parâmetros do modelo. Abordagens para a mitigação de problemas de generalização e convergência serão destacadas.

A revisão deve embasar a utilização de um número restrito de alternativas para a parametrização, generalização e convergência, observando as características da ferramenta usada nos testes, o MATLAB. Assim, os aspectos levantados aqui levarão em consideração a viabilidade de implantação no MATLAB e deverão permitir uma quantidade plausível de cenários a serem simulados para testes do modelo híbrido.

Contextualização

A capacidade de aproximação universal das RNAs *feedforward* foi discutida. Garantido um nível mínimo de complexidade da rede, elas podem aproximar qualquer função. No entanto, um aumento excessivo da complexidade da RNA pode proporcionar excelentes resultados na interpolação do conjunto de treinamento sem garantir acurácia em testes.

Se o número de amostras para o treinamento da RNA tendesse ao infinito e fosse livre de viés, os parâmetros convergiriam de forma a guiar o risco empírico (testes) e o risco real (associado ao erro de generalização) de forma semelhante para um mínimo. Com um número finito de amostras, e com tendências nos dados, o risco empírico pode não refletir a capacidade de generalização da rede, fazendo os parâmetros convergirem para uma solução com viés.

Assim, a complexidade da série alvo, com não-linearidades, tendências, ruídos, etc., e a disponibilidade de dados para acomodar um aumento de parâmetros do modelo são fatores que influenciam na capacidade de generalização das RNAs.

Os termos *overfitting* ou *overtraining* são usados para caracterizar a situação na qual o modelo apresenta sensível queda de desempenho na validação, tendo se ajustado excessivamente ao conjunto de treinamento, de forma a reproduzir partes do alvo compostas por ruídos ou erros de medição. A tradução para tais termos é geralmente superdimensionamento, o que remete à adoção de um número exagerado de parâmetros no modelo e às dificuldades de acomodá-los usando uma quantidade limitada de amostras, de forma a garantir generalização do modelo.

Cabe aqui uma breve contextualização do atual grau de desenvolvimento e utilização das RNAs de forma a melhor exprimir a conceituação feita acima. Essa necessidade se

justifica considerando os desafios que serão lançados com as modificações produzidas pela EMD na parametrização e generalização das RNAs.

As RNAs tiveram graus de participação e importância inconstantes desde seu surgimento no cenário da inteligência artificial – IA. Algum tempo depois de surgirem, a comoção e as expectativas a respeito das possibilidades criadas com as RNAs foram amortecidas pelas limitações no grau de complexidade dos modelos impostas pela disponibilidade de recursos computacionais.

Com o aumento da capacidade de processamento e com a variedade e quantidade de dados hoje disponíveis, as possibilidades de aplicação da IA se multiplicaram. Dentre os avanços computacionais, destaca-se o desenvolvimento do paralelismo no processamento vinculado às GPUs – *Graphic processing units*, permitindo o processamento de modelos mais complexos. Já as informações para alimentar tais modelos são referentes ao movimento de *Big Data* e estão disponíveis, inclusive via internet, na forma de mapas, planilhas, textos, figuras, áudios, etc.

Dessa vez, as RNAs passaram a ser protagonistas da IA. Com capacidade de processamento e informações à disposição, modelos mais complexos tornam-se cada dia mais promissores. Segundo NEYSHABUR (2017), o termo aprendizado profundo - *Deep Learning*, refere-se ao treinamento de modelos tipicamente complexos e superdimensionados, sendo o termo intercambiável com o conceito de RNA, já que muitos modelos da área são baseados em modificações em diferentes tipos de RNAs. O autor diz ainda que o interesse no aprendizado profundo – DL, reaflorou por volta de 2006, obtendo enorme sucesso prático desde então, causando grande defasagem entre o entendimento teórico sobre o assunto e os avanços práticos.

No DL, com o excesso de parâmetros, a função objetivo passa a ter vários mínimos, muitos dos quais não generalizam bem o problema. Sendo assim, ao escolher o mínimo errado leva a perda de generalização. Diferentes estratégias de generalização, tais como ajustes nas taxas de aprendizado, regularização, critérios de parada e até os critérios de escolha das amostras para validação passam a conduzir a diferentes resultados, provenientes de diferentes mínimos.

Expressar a complexidade em termos de número de parâmetros ou por qualquer indicador que seja função apenas da arquitetura, sem levar em conta fatores como a complexidade das amostras e a possível utilização de estratégias de generalização é limitar a capacidade das RNAs aos estágios anteriores de desenvolvimento da IA.

Cada parâmetro utilizado contribuirá para o aumento da variância do resíduo. Utilizar parâmetros cuja contribuição para o ajuste seja menor que o aumento que ele causa na variância terá um efeito negativo. Esse aumento da variância do resíduo independente da importância do parâmetro é um resultado conhecido, expresso, por exemplo, por (SJÖBERG & LJUNG, 1995). Alguns parâmetros são nesse contexto supérfluos, podendo causar ajuste excessivo às características irrelevantes do conjunto de treinamento, causando perda de generalização.

Existe ainda a possibilidade de prejuízo da convergência do modelo por mal condicionamento do Jacobiano da função objetivo.

A seguir, alguns conceitos são apresentados à luz da revisão bibliográfica de forma a embasar a discussão sobre parametrização, generalização e convergência. Busca-se assim, relacionar o *overfitting* com a utilização de parâmetros supérfluos e não só com o número excessivo de parâmetros. A convergência pode também ser relacionada à utilização de uma parametrização inadequada.

Norma

Norma é um conceito para medir tamanho, sendo uma função $\|\cdot\|$ de um espaço vetorial ν em \mathbb{R} , ou seja, que associa um vetor a um número real (norma vetorial) ou a uma matriz (norma matricial). Essa função satisfaz às seguintes propriedades:

$$(i) \quad \|\mathbf{v}\| \geq 0 \quad \therefore \|\mathbf{v}\| = 0 \Leftrightarrow \mathbf{v} = \mathbf{0};$$

$$(ii) \quad \|\alpha\mathbf{v}\| = |\alpha| \|\mathbf{v}\|;$$

$$(iii) \quad \|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Sendo, \mathbf{u} e \mathbf{v} dois elementos de ν e $\alpha \in \mathbb{R}$. Alguns exemplos de normas vetoriais, $\|\cdot\|: \mathbb{R}^n \rightarrow \mathbb{R}$, são:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$$

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$$

$$\|\mathbf{x}\|_\infty = \max |x_i| \quad (i = 1 : n)$$

Para o operador linear \mathbf{A} , definido no espaço vetorial v , $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$, a norma do operador linear é definida como:

$$\|\mathbf{A}\| = \max_{\mathbf{x} \in v - \{0\}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \quad (5-1)$$

Ou, introduzindo um vetor unitário \mathbf{u} em relação à norma $\|\cdot\|_p$, tal que $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|_p$, tem-se:

$$\|\mathbf{A}\|_p = \max \left\{ \|\mathbf{A}\mathbf{u}\|_p \mid \|\mathbf{u}\|_p = 1 \right\}; \quad (5-2)$$

Todas as normas condicionadas atendem à condição de consistência:

$$\|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \|\mathbf{x}\| \quad (5-3)$$

De maior interesse para o caso de algoritmos de mínimos quadrados é a norma induzida. Essa função também é chamada de norma espectral ou euclidiana e formalmente corresponde ao maior valor singular da matriz \mathbf{A} . Assim:

$$\|\mathbf{A}\|_2 = \max \sigma_i \quad \therefore \sigma_i = \sqrt{\lambda_i} \quad (5-4)$$

Sendo: $\begin{cases} \sigma_i : \text{valores singulares de } \mathbf{A}; \\ \lambda_i : \text{autovalores de } \mathbf{A}^T \mathbf{A} \rightarrow \det(\mathbf{A}^T \mathbf{A} - \lambda_i I) = 0. \end{cases}$

Se a matriz é simétrica então seus valores singulares são iguais aos autovalores:

Condicionalmento de Sistemas Lineares

O condicionalmento de um sistema linear fornece um indicativo de quão acurada a solução será após a aproximação numérica. De forma geral, dado um sistema $\mathbf{Ax} = \mathbf{b}$, seu condicionalmento indica como os erros e flutuações em \mathbf{b} serão transformados pela matriz \mathbf{A} , afetando a solução \mathbf{x}^* . Assim, com um número de condicionalmento elevado, mesmo pequenos erros em \mathbf{b} poderão causar grandes variações na solução.

Apesar de ser uma característica da matriz, tomada antes de se contabilizarem erros de arredondamento, questões relativas à representação numérica e operações de pontos flutuantes podem, em um sistema mal condicionalmento, prejudicarem ainda mais a acurácia. Além disso, se o número de condicionalmento é elevado, o erro relativo pode ser muito mais elevado que o resíduo relativo.

Definindo o resíduo \mathbf{r} e o erro \mathbf{e} como:

$$\mathbf{r} = \mathbf{Ae} \therefore \mathbf{e} = \mathbf{A}^{-1}\mathbf{r}$$

Nesse caso, as seguintes relações são satisfeitas:

Da consistência expressa em (5-3):

$$\|\mathbf{r}\| \leq \|\mathbf{A}\|\|\mathbf{e}\| \wedge \|\mathbf{e}\| \leq \|\mathbf{A}^{-1}\|\|\mathbf{r}\|$$

O que implica, rearranjando:

$$\frac{\|\mathbf{r}\|}{\|\mathbf{A}\|} \leq \|\mathbf{e}\| \leq \|\mathbf{A}^{-1}\|\|\mathbf{r}\| \quad (5-5)$$

Aplicando a consistência da norma, dessa vez às equações $\mathbf{b} = \mathbf{Ax}^*$ e $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$:

$$\|\mathbf{b}\| \leq \|\mathbf{A}\|\|\mathbf{x}^*\| \wedge \|\mathbf{x}^*\| \leq \|\mathbf{A}^{-1}\|\|\mathbf{b}\| \Rightarrow \frac{1}{\|\mathbf{A}^{-1}\|\|\mathbf{b}\|} \leq \frac{1}{\|\mathbf{x}^*\|} \leq \frac{\|\mathbf{A}\|}{\|\mathbf{b}\|} \quad (5-6)$$

Combinando (5-5) e o lado direito de (5-6), obtém-se:

$$\frac{1}{\|\mathbf{A}\|\|\mathbf{A}^{-1}\|} \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} \leq \underbrace{\frac{\|\mathbf{e}\|}{\|\mathbf{x}^*\|}}_{E_r} \leq \|\mathbf{A}\|\|\mathbf{A}^{-1}\| \underbrace{\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}}_{R_r}$$

Assim, o termo $C(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$ condiciona a relação entre erro relativo $E_r = \frac{\|\mathbf{e}\|}{\|\mathbf{x}^*\|}$ e o

resíduo relativo $R_r = \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}$ do sistema. Quando esse número de condicionamento $C(\mathbf{A})$ é

grande, o erro relativo pode tornar-se consideravelmente maior que o resíduo relativo. Nesse caso, \mathbf{A} é dita mal condicionada e não há confiabilidade na solução numérica. Já com $C(\mathbf{A}) \approx 1$, o erro relativo fica próximo ao resíduo relativo.

Condicionamento e ranqueamento em RNAs *feedforward*

Pode-se então passar a descrever as dificuldades decorrentes do mal condicionamento e de deficiência no ranqueamento do Hessiano e Jacobiano de RNAs *feedforward*. Tais questões são abordadas por SAARINEN *et al.* (1994) usando a decomposição em valores singulares, um processo que fatora uma matriz $\{\mathbf{A} \in \mathbb{R}^{l \times q} \mid l \geq q\}$ na forma $\{\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T \mid \mathbf{U}^T \mathbf{U} = \mathbf{I}_l \wedge \mathbf{V}^T \mathbf{V} = \mathbf{I}_q\}$. A matriz $\Sigma \in \mathbb{R}^{l \times q}$ é diagonal, com elementos $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_q \geq 0$, dados por (5-4). O valor do condicionamento da matrix, $C(\mathbf{A})$, considerando a norma euclidiana, de interesse para problemas de mínimos quadrados, é sem demonstração formal:

$$C(\mathbf{A}) = \frac{\sigma_1}{\sigma_r} \quad (5-7)$$

Para um ranqueamento $\{r < q \mid \sigma_{r+1} = \dots = \sigma_q = 0 \wedge \sigma_r > 0\}$, sendo $q-r$ a deficiência de ranqueamento de \mathbf{A} . A origem de (5-7) fica clara se lembrarmos que $C(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$ e que, para a norma euclidiana, tem-se (5-4). Na prática, ressalta-se que não se espera valores singulares iguais a zero, mas muito pequenos, complicando a determinação de r . Mais comum é a ocorrência de $\sigma_1 \gg \sigma_{r+1} > \dots > \sigma_q$. Sob tais condições, \mathbf{A} apresentará características de uma matriz com deficiência de ranqueamento, no mínimo de grau $q-r$.

SAARINEN *et al.* (1994) exploram exemplos nos quais há excesso de amostras, situação recorrente em DL. A pesquisa foi voltada para redes *feedforward* de reconhecimento de padrões para as quais diversos algoritmos se mostram lentos, tentando explicar o baixo desempenho considerando o ranqueamento e o condicionamento.

Para o caso de interesse das RNAs, seguindo a nomenclatura adotada por SAARINEN *et al.* (1994), seja $g(\mathbf{w})$ a função resíduo entre a função de aproximação $F(\varphi_i, \mathbf{w})$ e a função alvo y_i .

$$g_i(\mathbf{w}) = F(\varphi_i, \mathbf{w}) - y_i \quad (5-8)$$

A função de aproximação combina o vetor de entrada φ_i , que é p -dimensional, com o vetor de parâmetros (pesos e de limiares reais) $\mathbf{w} \in \mathbb{R}^N$, sendo N o número de parâmetros, a serem determinados durante o treinamento para os padrões ($i = 1, 2, \dots, n$). Deseja-se achar os valores de \mathbf{w} que minimizam a norma de g . No caso dos mínimos quadrados, a norma é euclidiana.

$$\|g(\mathbf{w}^*)\|^2 = \min_{\mathbf{w} \in \mathbb{R}^N} \sum_{i=1}^m g_i(\mathbf{w})^2$$

Tomando a função $\phi(\mathbf{w})$:

$$\phi(\mathbf{w}) = \frac{1}{2} \|g(\mathbf{w})\|^2 \quad (5-9)$$

Diversos algoritmos buscarão minimizar essa função pesquisando a direção e o tamanho do ajuste a cada iteração, de forma que $\phi(\mathbf{w} + \alpha \mathbf{v}) < \phi(\mathbf{w})$, sendo \mathbf{v} a direção de busca e α o tamanho do ajuste. Nesse processo, os pontos críticos \mathbf{w}^* correspondem a $\nabla\phi(\mathbf{w}^*) = 0$, sendo:

$$\nabla\phi(\mathbf{w}) = J(\mathbf{w})^T g(\mathbf{w}) \quad (5-10)$$

O primeiro termo do lado direito é a transposta do Jacobiano de $g(\mathbf{w})$. O Jacobiano é dado por $J_{ij} = \partial g_i / \partial w_j$. Outra matriz a ser definida é o Hessiano de $g(\mathbf{w})$:

$$H(\mathbf{w}) = J(\mathbf{w})^T J(\mathbf{w}) + \sum_{i=1}^m g_i(\mathbf{w}) H_i(\mathbf{w}) \quad (5-11)$$

Um ponto crítico \mathbf{w}^* precisa ainda que os autovalores de $H(\mathbf{w}^*)$ sejam não negativos para que seja um mínimo local. Na presente pesquisa o algoritmo Levenberg-Marquardt será utilizado por ser o mais rápido dentre os algoritmos *backpropagation* disponíveis no MATLAB, sendo opção altamente recomendada dentre os algoritmos supervisionados, apesar do maior consumo de memória computacional. Uma limitação do LM é o uso obrigatório do erro médio quadrático como função de performance, como visto acima.

Assim como os métodos quase-Newton, o LM foi desenvolvido almejando o desempenho de segunda ordem sem ter que computar a matriz Hessiana H , aproximando-a através da seguinte expressão de atualização:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - [J^T J + \mu \mathbf{I}]^{-1} J^T g \quad (5-12)$$

Quando o escalar μ é zero, ou quando o resíduo é zero e $J(\mathbf{w}^*)$ possui ranqueamento total, LM exhibe convergência quadrática, como Newton, usando a aproximação de H . Quando μ é grande, LM se torna gradiente-descendente, com passo pequeno e convergência linear. O método de Newton é mais rápido e mais acurado perto dos mínimos. Por isso, busca-se aproximá-lo quando possível, reduzindo μ depois de cada iteração que reduza o resíduo. Mas μ é aumentado depois de um passo que aumente g .

Os aspectos da convergência local e global de diversos métodos são abordados por SAARINEN *et al.* (1994). Para o algoritmo LM, localmente, a convergência depende do tamanho do resíduo, que influencia o termo de momentum μ . Na prática, esse termo é

ajustado para garantir melhor convergência global. Mas deve-se notar a forte influência do Jacobiano na convergência global e, portanto, do ranqueamento e condicionamento de $J(\mathbf{w}^*)$.

Multicolinearidade, ranqueamento e condicionamento

Imaginemos, em referência à Copa do Mundo de Futebol da Rússia, que um treinador de uma das seleções escalasse seu time com três atacantes. Além disso, imaginemos que o meio-campo fosse composto por jogadores ofensivos e que os laterais do time avançassem ao ataque constantemente. Em um primeiro momento, qualquer espectador ou comentarista poderia apostar na capacidade ofensiva do time descrito.

Mas, provavelmente, o que se observaria em campo seriam diversos jogadores tentando desempenhar funções semelhantes durante a partida. O time teria, eventualmente, dificuldades de ocupar todo o campo, se concentrando em determinados setores. Haveria problemas de entrosamento e deficiências em setores como defesa e criação. O resultado? Não se pode prever se o time conseguiria tirar proveito de sua capacidade ofensiva, podendo falhar em outras tarefas.

A multicolinearidade é uma condição na qual ocorre significativa correlação ou interdependência entre preditores ou variáveis independentes de um sistema. Quando a relação entre preditores é alta, torna-se difícil fazer ajustes confiáveis dos parâmetros, resultando em conclusões incorretas sobre as relações entre os preditores e as séries alvos. Além disso, a multicolinearidade torna mais difícil estimar a importância relativa de cada variável independente na explicação do comportamento da variável dependente.

Existem diversas formas de identificar a multicolinearidade em um sistema linear. Dentre elas:

- i. inspeção da matrix de correlações entre preditores, pois valores elevados indicam colinearidade;
- ii. cálculo do Fator de Inflação da Variância, livre tradução de *Variation Inflation Factor* – VIF, que quantifica a multicolinearidade em análises de regressão por mínimos quadrados;

- iii. análise dos valores singulares da matriz de coeficientes ou dos autovalores da matriz de correlação, sendo valores próximos de zero ou valores altos de condicionamento, indicativos fortes de multicolinearidade.

Outros sinais da multicolinearidade são, por exemplo, instabilidades nos coeficientes ao dividir-se as amostras em partes ou mudanças dramáticas no modelo ao adicionar-se ou retirar-se alguma variável.

A inspeção das correlações é um procedimento que dispensa explicações.

Seja R_k^2 o coeficiente de determinação do k -ésimo preditor φ_k , ($k = 1, 2, \dots, p-1$) quando o mesmo é regredido linearmente a partir dos demais preditores, havendo p preditores no total. O VIF do k -ésimo preditor é definido da seguinte forma:

$$\text{VIF}_k = \frac{1}{1 - R_k^2} \quad \therefore k = [1, p-1] \quad (5-13)$$

Quando $R_k^2 \approx 0$ tem-se $\text{VIF}_k \approx 1$, ou seja, o k -ésimo preditor não possui relação linear com os demais preditores. Por outro lado, $\lim_{R_k^2 \rightarrow 1} \text{VIF}_k = \infty$ quando a variável “independente” em questão é linearmente relacionada com as demais. Assim, como sugerido pelo nome, o VIF quantifica o aumento da variância da estimativa do coeficiente do k -ésimo preditor pela colinearidade com os demais preditores. Valores de VIF excedendo 5 a 10 indicam multicolinearidade, que pode acarretar má estimação dos coeficientes. Uma descrição mais completa do VIF é feita em MURRAY *et al.* (2012).

A multicolinearidade pode ser identificada ainda através da análise dos valores singulares da matriz de coeficientes ou dos autovalores da matriz de correlação. O ranqueamento e condicionamento do Jacobiano foram citados como fatores importantes na convergência global, inclusive ao utilizar-se o algoritmo *feedforward* LM.

Assim, parece crucial para a presente pesquisa, saber como e quando a multicolinearidade pode afetar os sistemas de equações oriundos dos algoritmos *feedforward* LM a serem utilizados. Para então, sugerir caminhos a serem seguidos para remediar possíveis

problemas de convergência, inconsistência entre resíduos e erros e outras dificuldades trazidas pela multicolinearidade.

Ainda sobre o trabalho de SAARINEN *et al.* (1994), os autores descrevem o fluxo de informações em uma rede *feedforward* com duas camadas intermediárias, para mostrar que ela pode facilmente apresentar mal condicionamento do Jacobiano devido a redundâncias na parametrização do espaço nas proximidades de mínimos locais.

Nós da camada de entrada, nós das camadas intermediárias, nós da camada de saída, limiares dos nós e conexões ponderadas por pesos entre eles. Cada nó com várias entradas e apenas uma saída. A RNA *feedforward* sendo caracterizada por fluxo avante apenas e sem conexões entre nós da mesma camada.

Após a soma ponderada por pesos dos sinais convergentes das conexões de entrada de cada nó da primeira camada intermediária, o resultado é transformado pela função de excitação. No caso das camadas intermediárias, funções sigmoidais, por exemplo, podem ser usadas como ativação. Na camada de saída, a função de ativação adotada é comumente linear. Mas os autores não consideraram a ativação nos nós da camada de saída na descrição realizada dos balanços de fluxo por nó, segundo eles, sem perda de generalização, pois a sigmoideal sim desempenha papel central na ocorrência da multicolinearidade.

Sobre as funções de ativação, elas são responsáveis pelo aumento da capacidade de representação das RNAs, pois introduzem a não-linearidade na modelagem. Mas junto com a maior representatividade vem o aumento da complexidade do espaço solução da função a ser otimizada, que deixa de ser convexo.

Mapeando o fluxo *feedforward* desde as entradas até a totalização de somas ponderadas nas saídas e considerando funções de ativação sigmoidais, pode-se mostrar o Jacobiano de forma explícita para uma RNA *feedforward* de duas camadas. Ao explicitar os elementos do Jacobiano dessa forma, SAARINEN *et al.* (1994) mostraram que uma linha de J não pode ser combinação linear exata de outras linhas. Sendo assim, quaisquer dependências lineares entre linhas ou camadas precisam ter outra origem.

Explorando as propriedades da sigmoideal, os autores buscaram explicações alternativas para a ocorrência de relações lineares entre linhas de J . Isso foi feito observando o comportamento de funções compostas de sigmoideais, que aparecem na explicitação dos termos do Jacobiano. Eles consideraram a função sigmoideal de ativação e sua derivada:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad \therefore \quad \sigma'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} \quad (5-14)$$

E as funções compostas:

$$\begin{aligned} A(x, y) &= \sigma(x + y) - \sigma(x); \\ B(x, y) &= \sigma'(x + y) - \sigma'(x); \\ C(x, y) &= \sigma'(x + y)\sigma'(x); \\ D(x, y) &= \sigma(x + y)\sigma'(x); \end{aligned} \quad (5-15)$$

A característica mais importante dos termos de (5-15) é que para valores grandes de x e y , tais funções compostas assumem valores próximos a zero, o que explica a quase dependência linear entre colunas do Jacobiano. Consideremos a mesma análise com a tangente sigmoideal, função de ativação padrão do algoritmo a ser usado no MATLAB. Assim:

$$\sigma(x + y) = \frac{2}{1 + e^{-2(x+y)}} - 1 \quad \therefore \quad \sigma'(x + y) = \frac{4e^{-2(x+y)}}{(e^{-2(x+y)} + 1)^2} \quad (5-16)$$

Para a tangente sigmoideal – TANSIG, ocorre situação semelhante ao considerarmos as funções compostas A até D em (5-15). As Figuras Figura 5-4 e Figura 5-5 ilustram tais funções compostas com a adoção de TANSIG. Percebe-se, assim como no caso testado por SAARINEN *et al.* (1994), a ocorrência de valores nulos ou patamares nas funções A, B, C e D para valores maiores de x e y .

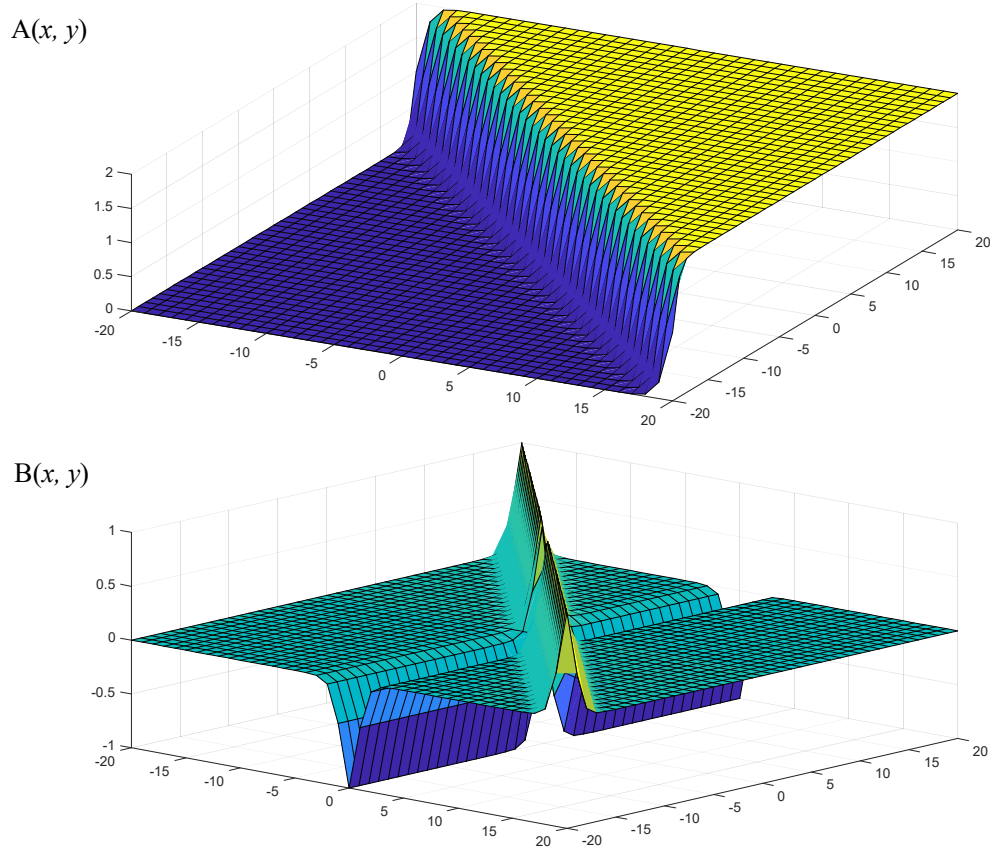


Figura 5-4. Funções compostas $A(x, y)$ e $B(x, y)$ utilizando-se a TANSIG como função de ativação.

Os autores apresentaram três proposições e suas demonstrações a respeito do condicionamento e ranqueamento do Jacobiano da rede com duas camadas intermediárias. Tais proposições são citadas abaixo, devido a relevância do problema de condicionamento e ranqueamento na presente pesquisa, omitindo-se as demonstrações:

- i. Seja \mathbf{B} uma matriz extraída de \mathbf{A} , constituída por colunas de \mathbf{A} . Então $C(\mathbf{B}) \leq C(\mathbf{A})$;
- ii. Seja $\mathbf{A} = [\mathbf{x} \ \mathbf{y}] \in \mathbb{R}^{n \times 2}$, com as colunas \mathbf{x} e \mathbf{y} satisfazendo $\cos(\theta) = 1 - \varepsilon$ para $\varepsilon \in (0, 1)$, então:

$$C^2(\mathbf{A}) \geq \frac{1}{4\varepsilon(2-\varepsilon)} \left(\frac{\|\mathbf{y}\|^2}{\|\mathbf{x}\|^2} + \frac{\|\mathbf{x}\|^2}{\|\mathbf{y}\|^2} + 2 \right)$$

- iii. Seja $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \cdots \mathbf{a}_p]$ uma matriz n por p com colunas normalizadas, tal que $\mathbf{a}_i^T \mathbf{a}_j = 1 - \varepsilon_{ij} \forall \{i, j\} \subset [1, p]$, com $\varepsilon_{ii} = 0$ para $i \in [1, p]$. Suponha $\varepsilon_{ij} \leq \varepsilon < 1$. Então, \mathbf{A} possui um valor singular maior ou igual a $\sqrt{p(1-\varepsilon)}$ enquanto os demais valores singulares não são maiores que $\sqrt{p\varepsilon}$.

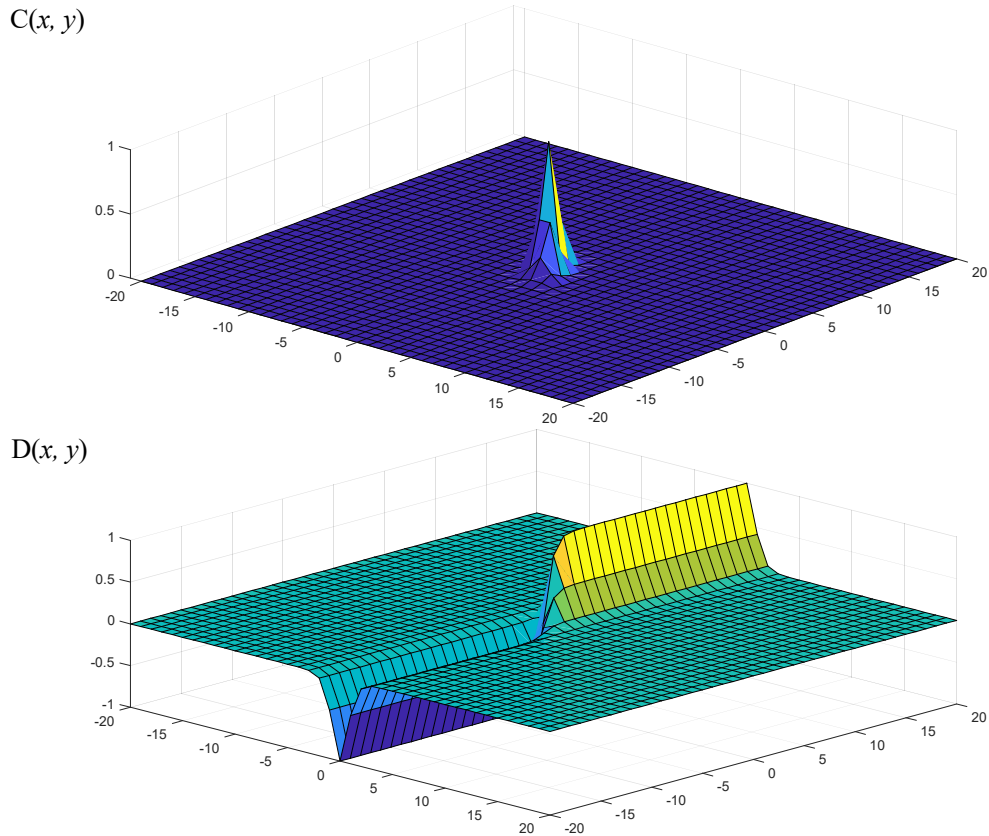


Figura 5-5. Funções compostas $C(x, y)$ e $D(x, y)$ utilizando-se a TANSIG como função de ativação.

As proposições fornecem um limite inferior para o número de condicionamento quando duas colunas são quase colineares, sendo $C(J)$ ao menos $O(\varepsilon^{-1/2})$. Também fica evidente que se p colunas de J são quase paralelas e se as colunas de J são normalizadas, pode-se esperar uma deficiência de ranqueamento de, no mínimo $p - 1$, com valores singulares respectivos $O(p\varepsilon)^{1/2}$ e um valor singular $O(p^{1/2})$, implicando nas restrições numéricas de convergência, instabilidades e de incompatibilidade entre resíduos e erros discutidas anteriormente.

Os autores mostraram assim que as deficiências de ranqueamento e o mal condicionamento são características inerentes ao Jacobiano dos *feedforward* de duas camadas intermediárias. Espera-se severidade menor de tais problemas em redes de uma camada, como as que serão utilizadas nessa pesquisa. Os efeitos da EMD na multicolinearidade no Jacobiano são desconhecidos. Com vetores de predição possuindo modos marcantes da variabilidade tropical em comum, tais efeitos merecem investigação na presente pesquisa.

Estratégias de ranqueamento e condicionamento das matrizes foram adotadas, estudando-se seus efeitos nas simulações.

5.3. Estratégias de Generalização

Postas as relações entre a parametrização, o condicionamento do Jacobiano das RNAs e a generalização, passam-se a descrever algumas abordagens buscando aumento da capacidade de generalização das redes.

Parada forçada

Uma das táticas para evitar o *overfitting* é a parada forçada do treinamento. Para implementá-la, o conjunto disponível de dados tem que ser dividido em três subconjuntos. Assim, nem todos os dados poderão ser usados para computar o gradiente da função objetivo e para atualização dos pesos e vieses no treinamento. Um segundo conjunto, chamado de conjunto de validação, é desentranhado dos dados e fica destinado ao monitoramento da evolução do erro durante o treinamento. Os pontos do conjunto de validação não são usados no ajuste. Há ainda um terceiro conjunto, de testes, também independente do treinamento.

Normalmente, os erros no conjunto de validação diminuem numa fase inicial e começam a crescer numa segunda fase, caracterizando o *overfitting*. As técnicas de parada forçada prematura usam um critério para interromper o treinamento e retornar os valores obtidos dos pesos e vieses perto do mínimo valor de erro estimado através do conjunto de validação. Um critério recorrente é limitar o número de iterações sem melhora do erro de validação.

O conjunto de testes, também não usado durante o treinamento, auxilia na comparação de diferentes modelos. A comparação do desempenho dos testes e da validação permite constatar a conveniência da divisão dos conjuntos, considerando o grau de independência necessário em relação ao treinamento, mas preservando a representatividade dos dados.

Ao falarem da busca por mínimos e sua relação com o *overfitting*, SJÖBERG & LJUNG (1995) comparam os benefícios da parada antecipada aos da regularização. Segundo os autores, parar a busca por mínimos antecipadamente tem efeitos semelhantes aos da regularização (a ser apresentada mais adiante), pois a complexidade do modelo aumenta com as iterações (épocas) do treinamento.

Ao formalizarem matematicamente uma teoria estatística assintótica para o excesso de ajuste no treinamento (*overtraining*, sinônimo de *overfitting*) e sua relação com a validação cruzada em redes estocásticas classificadoras *feedforward* com treinamento baseado na divergência de Kullback-Leibler, AMARI *et al.* (1996) citam três fases do superdimensionamento relacionado o número de exemplos n com o número de parâmetros N .

Segundo os autores, quando n é menor, porém próximo a N , os exemplos podem ser memorizados e o superdimensionamento é observado, sobretudo quando $n \approx N$. Nessa fase, a aplicação da validação cruzada simplesmente repartindo os dados disponíveis sacrifica um conjunto de treinamento já deficiente. Uma fase intermediária ocorre quando $n > N$, na qual a validação cruzada passa, em geral, a ser efetiva, segundo os autores, apesar de ser difícil a construção de uma teoria geral nessa fase. Na fase assintótica, n é grande o suficiente e a teoria assintótica da estatística se aplica, sendo os parâmetros estimados normalmente distribuídos com boa aproximação, ao redor dos valores reais.

O trabalho de AMARI *et al.* (1996) despertou interesse durante a presente pesquisa não só por tratar da validação cruzada e da parada antecipada, mas por falar da relação entre a estimativa do erro do conjunto de validação e a real capacidade de generalização da rede nas diversas fases citadas. Não se pode ter acesso ao real ponto ótimo de parada, o que faz com que, em certas situações, possíveis ganhos com a parada antecipada do treinamento sejam menores que as perdas no treinamento pelo sacrifício de elementos dos conjuntos de validação e teste.

Os autores fornecem uma resposta analítica para a melhor divisão do conjunto original com vistas à validação cruzada. Segundo eles, quando o número N de parâmetros é grande, a melhor estratégia é usar quase todas as amostras t para treinamento, deixando apenas $n/\sqrt{2N}$ elementos para a validação.

Foram realizados testes com redes *feedforward* classificadora de três camadas com $N \sim 100$, sem fornecer mais detalhes sobre os dados usados nas simulações. Foram usadas funções sigmóides na camada intermediária e funções *softmax* na camada de saída. Com $n > 30N$, os resultados teóricos representaram bem os testes, indicando ser a validação cruzada não necessária, pois a incerteza no instante de parada causava mais problemas na generalização do que não usar a parada. Para $n < 30N$, em uma faixa intermediária, o treinamento excessivo certamente ocorre e a validação cruzada melhora a generalização.

Outros trabalhos falaram de métodos de estimação do erro para avaliação do desempenho de classificadores. RODRÍGUEZ *et al* (2010) apresentaram uma análise de sensibilidade do teste de validação cruzada *k-fold* – *k-cv*. Os autores compararam o viés e a variância para diferentes valores de k e apresentaram a variância decomposta segundo fontes devidas a sensibilidade a mudanças do conjunto de treinamento e a mudanças nas partições dos subconjuntos desentranhados. Os autores recomendaram a utilização do *k-cv*. Já (KOHAVI, 1995) comparou o *k-cv* ao *bootstrap*, analisando variância e viés para experimento com grande número de amostras reais.

Os trabalhos citados acima apenas evidenciam a importância da escolha do método de estimação dos desvios obtidos. Cada tipo de amostragem penaliza de forma distinta o conjunto de treinamento e generaliza em grau distinto os desvios. Tais observações servem mesmo que a parada antecipada não seja feita.

Regularização

A definição do nível adequado de complexidade dos modelos parece ser um problema para as redes *feedforward* a serem usadas nessa pesquisa. O superdimensionamento e suas consequências foi discutido e a revisão apontou que a parada antecipada possui a desvantagem de sacrificar exemplos para os conjuntos de validação e de testes. O problema da indeterminação do instante ótimo para a parada também foi apontado.

Apresenta-se a regularização Bayesiana como alternativa para a generalização dos modelos em BISHOP (1995). Apesar de haver comparações entre os benefícios da parada antecipada e da regularização na bibliografia pesquisada, SJÖBERG & LJUNG (1995), algumas vantagens da regularização são destacadas:

- Problemas com abordagem convencional, utilizando função objetivo baseada na minimização do erro ou do erro médio quadrático, são caso geral da regularização;
- desobriga a utilização de conjuntos de validação e teste, evitando penalizar o conjunto de treinamento como ocorre na validação cruzada, podendo comportar o uso de modelos mais complexos a serem adequados durante a simulação;

Ainda em BISHOP (1995) uma revisão da base teórica da regularização, partindo do teorema de Bayes é apresentada. Considerando um conjunto de amostras X e um conjunto de modelos M , supostos ordenados por grau de complexidade crescente (aumento do número de parâmetros), segundo o teorema de Bayes, tem-se:

$$p(M_i | X) = \frac{p(X | M_i) p(M_i)}{p(X)} \quad (5-17)$$

Pode-se supor, de forma razoável, não haver preferência para nenhum modelo a priori, atribuindo-se igual valor para todas as probabilidades $p(M_i)$. O valor de $p(X)$ não depende dos modelos adotados, mas apenas do conjunto de amostras. Assim, resta avaliar $p(X | M_i)$ para permitir a comparação entre diferentes modelos, dado um conjunto de amostras.

A Figura 5-6, adaptada de BISHOP (1995), ilustra as distribuições normalizadas condicionais de um conjunto de amostras X , dados três modelos M_1 , M_2 e M_3 , ordenados do menos para o mais complexo. Nota-se que o aumento da complexidade acarreta numa cobertura mais extensa do conjunto de exemplos descritos pelos modelos. Modelos mais flexíveis seriam capazes de reproduzir uma gama maior de funções. No entanto, também aumenta o espalhamento da distribuição e penalização de $p(X | M_i)$. Assim, a regularização objetiva probabilidades mais elevadas em extensões maiores para modelos

com complexidade adequada, como ilustrado com o exemplo X^* , apontando maior probabilidade para o modelo M_2 .

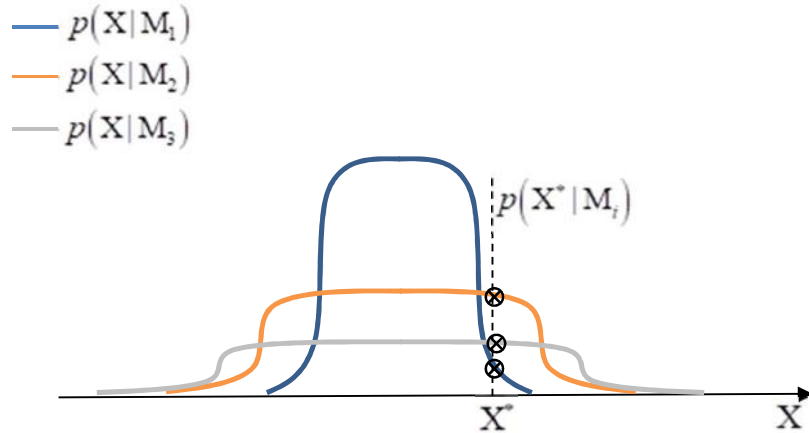


Figura 5-6. Ilustração das distribuições normalizadas de probabilidades condicionais do conjunto de amostras X dados os modelos, em ordem de complexidade crescente M_1 , M_2 e M_3 . Em destaque, um exemplo X^* . Adaptada de BISHOP (1995).

O processo de treinamento convencional das redes *feedforward* de interesse na presente pesquisa objetiva a minimização da soma dos desvios quadráticos (norma euclidiana). Partindo da equação (5-9), para a regularização, um termo relativo a soma dos quadrados dos pesos da rede é adicionado e a função objetivo torna-se:

$$E_F = \frac{1}{N} \sum_{i=1}^N (g_i(\mathbf{w}))^2$$

Para a regularização, um termo relativo a soma dos quadrados dos pesos da rede é adicionado e a função objetivo torna-se:

$$F = (1-\gamma)E_F + \gamma E_w \therefore \begin{cases} E_F = \frac{1}{N} \sum_{i=1}^N (g_i(\mathbf{w}))^2 \\ E_w = \frac{1}{n} \sum_{j=1}^n w_j^2 \end{cases} \quad (5-18)$$

Sendo E_w o termo relativo a soma dos quadrados dos pesos e γ um parâmetro tal que quanto maior γ mais suave será a resposta da rede. Em outras palavras, quanto maior tal parâmetro, maior o efeito de minimização dos pesos e de regularização. Por outro lado,

valores baixos de γ farão o algoritmo priorizar a minimização de desvios. O desenvolvimento e derivação da equação (5-18), utilizando o teorema de Bayes e usando tratamento estatístico pode ser acompanhada em BISHOP (1995).

A otimização do parâmetro de regularização é descrita em (FORESEE & HAGAN (1997). Detalhar o processo extrapola o escopo da presente pesquisa, cabendo ressaltar seus efeitos de suavização através da minimização dos pesos e de ajuste da complexidade com eliminação de parâmetros não efetivos na minimização dos desvios.

6. Modelos Híbridos

O pré-processamento de dados acoplado à utilização de RNAs não é assunto inexplorado. A EMD e outras técnicas de decomposição espectral foram combinadas às RNAs para previsões em diversas áreas, incluindo a hidrologia.

Alguns autores exploraram o paralelismo e a capacidade de generalização das RNAs de forma conjunta com as técnicas modernas de análise de sinais, especialmente a Wavelet e a EMD. Em tais obras, a decomposição Wavelet ou a EMD foram usadas no pré-processamento das entradas de RNAs.

Escoamentos superficiais (*runoff*) foram antecipados usando a decomposição Wavelet de preditores (dados de chuva da bacia e de escoamento), associada à RNA em NOURANI *et al.* (2009), permitindo, segundo os autores, previsão de feições do escoamento de curto e longo termos um dia à frente devido à decomposição multi-escala realizada. A bacia hidrográfica do estudo localizava-se no noroeste do Irã e possuía 75 km². As séries de precipitações e de escoamentos superficiais usadas são diárias e com extensão de 21 anos. Os resultados do modelo híbrido (utilizando o pré-processamento via Wavelet) foram comparados aos resultados de modelos convencionais com RNAs. Os resultados mostraram efeitos diferentes do pré-processamento via Wavelet entre as fases de calibração e validação dos modelos. Na calibração o desempenho melhorou progressivamente com o aumento do número de modos resultantes da decomposição Wavelet. No entanto os autores relataram que o aumento do número de parâmetros devido à decomposição Wavelet fez com que houvesse um nível ótimo de decomposição Wavelet a partir do qual o desempenho em validação caiu. Apesar da partição realizada, separando 25% do conjunto de amostras para testes, os autores não especificaram se a decomposição Wavelet foi usada sobre toda a série de escoamentos superficiais ou mesmo como o conjunto de testes foi escolhido.

O mesmo foi feito para modelagem de escoamento médios diários em dois rios não perenes em bacias semiáridas em ADAMOWSKI & SUN (2010) a partir das próprias séries de vazões. As bacias dos dois rios localizavam-se na ilha de Ciprus, sendo um com 67 km² e o outro com 63 km², ambos com dados entre 1965 e 2007. Os resultados foram comparados com os de uma RNA normal na previsão com um e três dias de antecedência,

mostrando maior acurácia do modelo híbrido. Novamente, apesar da utilização de conjunto de testes, não houve maiores detalhes sobre o método de escolha dos testes. Além disso, subintende-se que as decomposições tenham sido realizadas sobre as séries completas dos preditores.

Previsões de vazão diária com antecipação de um, três, cinco e sete dias foram feitas com modelo híbrido Wavelet - RNA mostrando melhoras de resultados nas comparações com RNA normal em SANTOS & DA SILVA (2013). A previsão foi feita a partir da decomposição dos próprios dados de vazão diária naturalizada da UHE Sobradinho, no rio São Francisco, com bacia de 498.968 km², com série diária de 18.900 valores (1931 a 1982) de duas formas distintas: a primeira separando o sinal apenas em altas e baixas frequências (chamadas de detalhes e aproximação); a segunda decomposição foi em múltiplos níveis. As componentes resultantes da decomposição usadas em cada previsão foram definidas por tentativa e erro. Os autores destacaram que a técnica foi importante na remoção de defasagens na previsão diária, frequentemente observada em modelos com RNAs convencionais. A validação cruzada foi usada para generalização dos modelos. Mais uma vez, a decomposição do sinal parece ter sido feita sobre o conjunto total de amostras.

Previsões de vazões mensais via RNAs autoregressivas – ARMA, alimentadas com IMFs para prever tais modos um passo à frente, mostraram resultados superiores que as RNAs ARMA sem a EMD para previsão da vazão (KISI *et al.*, 2014). No estudo foram usados 32 anos de vazões mensais em um posto do rio Kayseri, na Turquia, com 15.581,6 km².

De forma similar, também para dados mensais médios ISMAIL & SHABRI (2014) utilizaram a EMD em conjunto com uma modificação da Máquina de Vetor Suporte, livre tradução de *Support Vector Machine* – SVM, SVM de mínimos quadrados, destinada à solução de problemas lineares. A bacia do estudo, no rio Selangor, na Malásia, possuía 2.200 km² e a série de dados de vazão usada possuía extensão de janeiro de 1960 a dezembro de 2006. Os resultados foram comparados com os de modelo SVM de mínimos quadrados sem o pré-processamento via EMD, mostrando, segundo os autores, melhor desempenho do modelo híbrido.

A reanálise de dados de séries longas de vazões diárias de uma estação com 115 anos e dados, no rio Potomac, em Maryland, USA, com área de drenagem de 24.996 km², e de outra estação com 80 anos de dados, no rio Clark, em Montana, USA, com área de drenagem de 23.317 km², foram estudadas usando modelos EMD-RNA, explorando os efeitos da decomposição como pré-processamento, da inicialização aleatória dos pesos da RNA e a importância do uso de medidores de desempenho e testes adequados ao modelo em NAPOLITANO *et al.* (2011). As RNAs tiveram como alvo os modos da EMD. Os autores relataram a transitoriedade das IMFs como fator importante do desempenho dos modelos híbridos, sobretudo em sinais nos quais originem-se modos de alta frequência muito energéticos.

Destaca-se o uso do termo reanálise referindo-se à modelagem de dados apoiada em valores futuros da série de alguma forma. No caso de NAPOLITANO *et al.* (2011), a EMD foi feita antes da separação do conjunto de testes. Como o objetivo era prever valores da série original a partir da previsão dos modos, se houvesse a separação antes da decomposição, as RNAs seriam treinadas com conjunto de modos possivelmente desvinculado daquele obtido para os testes, inviabilizando a aplicação. Por isso o termo transitoriedade foi aplicado na presente pesquisa, referindo-se às mudanças ocorridas nas IMFs mesmo diante da adição de poucos valores à série de dados decomposta.

Por fim, ZHANG *et al.* (2015) questionaram o melhor desempenho dos modelos híbridos relatado na literatura, testando-os em previsões de vazões mensais. Os modelos híbridos testados usaram ou a decomposição Wavelet ou a EMD ou a Análise Espectral Singular, livre tradução de *Singular Spectrum Analysis* - SSA em dados de entrada de RNAs e de modelos ARMA – sigla em inglês para *Autoregressive Moving Average*. Os testes foram em dados mensais de duas estações do rio Yangtze, na China, tendo 624 e 756 meses de extensão. Os autores dividiram os testes em reanálise e previsão, livres traduções para *hindcast* e *forecast*. Os resultados confirmaram a superioridade dos modelos híbridos para reanálise, porém mostraram desempenho bem inferior de tais modelos nas previsões, quando comparados aos modelos RNA convencionais e aos modelos ARMA.

A diferença nos regimes de reanálise e previsão presentes nos experimentos de ZHANG *et al.* (2015) está no momento no qual as decomposições (Wavelet ou EMD) foram implementadas: nos experimentos de reanálise, as decomposições ocorreram antes do

desmembramento da série em subconjuntos de treinamento e de teste; já para os experimentos de previsão, os autores primeiro desentranharam o conjunto de testes e então realizaram as decomposições do conjunto de treinamento resultante.

Se a decomposição da série, seja por Wavelet ou via EMD, for feita antes da definição de tais grupos, os modos resultantes conterão informações futuras à data de previsão. Já em previsão, os modos resultantes da decomposição não se apoiarão em valores futuros à data do alvo.

De forma mais geral, pode-se classificar como reanálise, qualquer modelagem que tenha como entrada partes da série que contenham informações futuras dos alvos num modelo auto-regressivo. Nesse caso, a solução do modelo se baseiaria em informações futuras dos alvos.

Mas por decomporem apenas o conjunto de treinamento (via Wavelet ou EMD), os modelos utilizados por de ZHANG *et al.* (2015) sofrem, em regime de previsão, com a transitoriedade inerente aos modos. Se os preditores fossem séries distintas do alvo, com picos correlacionais defasados o bastante em relação à série alvo, os modos contariam com informações correlatas a períodos posteriores ao instante da previsão. Dessa forma, as IMFs carreariam mais informações futuras, evitando-se o problema da transitoriedade das IMFs.

7. Metodologia Adotada

Esse item traz as metodologias adotadas para o cumprimento dos objetivos da presente pesquisa, relembrando:

- objetivo secundário de estudo das relações entre os dados de TSM das bacias do Atlântico e do Pacífico com o histórico de níveis em Manaus, de forma a justificar sua utilização como preditores nos modelos de interesse;
- objetivo principal de previsão de níveis em Manaus através de modelos híbridos EMD-RNA e comparação do desempenho de tais modelos com modelos convencionais RNA.

Com maior quantidade de energia solar incidente e menor variabilidade sazonal dessa energia, com baixos gradientes de temperatura e pressão impossibilitando a estocagem de energia potencial, viu-se que nos trópicos a liberação de calor latente no interior das torres de calor acopla efeitos convergentes e divergentes horizontais com efeitos locais do calor diabático. Tais efeitos, com comportamentos ora de onda ora de pulsos, se sobrepõe a um fundo de pequenas variações zonais e assim os fenômenos tropicais compõe espectros de energia cheios, com características de ruído rosa, observando-se a dualidade entre hierarquia e auto-afinidade ou auto-similaridade.

Além disso, a distribuição do calor latente, desestabilizador da atmosfera tropical, está intimamente relacionada à TSM, que também é afetada pela atmosfera. Foram revisados trabalhos evidenciando as relações entre as TSM do Atlântico e do Pacífico Leste entre si e com o clima na região de interesse do estudo. Fenômenos tropicais de diferentes escalas foram discutidos. Resta, no entanto, explorar os efeitos de tais fenômenos na hidrologia da bacia em estudo, numa visão multi-escala, que permita revelar a complexidade inerente dos espectros descritos acima.

Explorando a análise fractal e as relações fenomenológicas entre a TSM das duas bacias oceânicas e o cotograma em Manaus, buscou-se justificar a utilização das TSMs como preditores dos níveis em Manaus através do modelo híbrido EMD-RNA e fomentar a discussão sobre o desempenho do modelo.

O presente capítulo traz, primeiramente, a coerência espectral Wavelet como metodologia aplicada para estudo das relações fenomenológicas entre as TSM e o cotagrama em Manaus. Em seguida, a DFA é definida como metodologia de análise fractal dos dados. As soluções particulares para os problemas encontrados na EMD são apresentadas logo após. Em sequência, a PLS é definida como estratégia para sanar possíveis ocorrências de multicolinearidade nos jacobianos das RNAs. Por fim, discutem-se detalhes das RNAs utilizadas na modelagem e os cenários a serem simulados para avaliação do desempenho dos modelos híbridos propostos.

7.1 Coerência Espectral Wavelet

A coerência espectral Wavelet propicia uma visão no domínio tempo-frequência das relações fenomenológicas entre as séries de TSM caracterizadoras do dipolo do Atlântico e do El Niño no Pacífico Leste com os níveis em Manaus.

O aplicativo MATLAB possui ferramenta para a análise Wavelet e apresentação dos espectros cruzados e da coerência espectral, obtendo-se escalograma tempo-frequência com contornos de correlação cruzada ao fundo e vetores de defasagem sobrepostos. Esta ferramenta referencia os trabalhos de GRINSTED *et al.* (2004) e TORRENCE & COMPO (1998). A coerência espectral é computada usando-se a ondeleta Morlet analítica. A fase do espectro cruzado é usada para identificar a defasagem entre os sinais.

O aplicativo apresenta ainda um cone de influência no espaço tempo-frequência, delimitando o efeito dos contornos. Para áreas nas quais a coerência espectral exceda 0,5, são plotadas setas para indicar a defasagem entre os sinais. A angulação das setas corresponde à defasagem local num círculo trigonométrico unitário. Assim, uma seta apontando para cima corresponde a um quarto de período de defasagem. O tamanho da defasagem depende, assim, do período no qual está localizada a seta.

Foram usadas configurações de interpolação e de decomposição Wavelet padrões. A priori, tal escolha deve ser satisfatória, uma vez que foram confrontadas séries mensais de TSM interpoladas diariamente com séries diárias de níveis. Dessa forma, as escalas principais de resolução foram intrasazonais, sazonais e interanuais, até cerca de seis anos. Os possíveis efeitos de falta de resolução ficaram em partes dos escalogramas de menor interesse.

Quanto à significância no traçado dos contornos de coerência espectral, foram feitas como indicado em TORRENCE & COMPO (1998). Uma distribuição paramétrica foi ajustada às fatias verticais do escalograma, tomando-se 5% de significância para a coerência.

Uma dificuldade inerente à interpretação dos resultados da ferramenta é que tanto a TSM do Atlântico Norte – NATL como as TSM do Pacífico, apresentaram anti-correlações em faixas de frequência mais energéticas em relação a Manaus, série alvo. As fases são indicadas nos escalogramas conforme o círculo trigonométrico. Assim, sinais correlatos têm fases indicadas por ângulos positivos a partir do zero trigonométrico. Sinais anti-correlatos possuem fases contadas a partir de 180°.

7.2 Análise fractal dos dados

A descrição apoiada na geometria fractal revelou características da variabilidade das séries, possíveis escalas hierárquicas e a auto-afinidade presentes nos dados de interesse da pesquisa. Além de destacar tais características em cada série, a utilização de uma ferramenta de análise fractal permitiu explorar as semelhanças fractais entre os dados.

No ANEXO II, a escala temporal auto-afim foi definida por trechos retos nos escalogramas de OLR de YANO & NISHI (1989). Algumas escalas destacaram na forma de desvios abruptos nos escalogramas. Além disso, um mapa de contornos de auto-afinidade da OLR, adaptado com a sobreposição das regiões de origem dos dados de TSM usados na presente pesquisa foi apresentado no ANEXO II.

A DFA foi usada para caracterizar trechos com inclinações semelhantes entre as séries e pontos de quebras característicos de escalas notáveis ou separações de escalas do ambiente tropical. Viu-se que, a grosso modo, os espectros de energia apresentados em YANO & NISHI (1989), YANO *et al.* (2001) e YANO *et al.* (2004) ajustam-se a uma potência negativa, implicando na ausência de frequências dominantes que caracterizam diferentes escalas. Mas algumas feições marcantes da dinâmica tropical produziram picos característicos evidenciados pela DFA.

O MBF é o modelo mais estudado de variabilidade auto-afim (HENRIQUES, 2008). Segundo MANDELBROT (1983), a característica básica do MBF é que a extensão de

interdependência entre seus incrementos é infinita, ou seja, mesmo amostragens distantes entre si, exibem interdependências, diferente do que ocorre em outros processos estocásticos e o expoente de Hurst H generaliza a relação entre a variabilidade no MBF e a magnitude do intervalo de tempo, podendo ser essa relação linear ou não.

Na análise aqui proposta com a DFA, no entanto, o foco principal não foi só nos valores de H de cada escalograma, mas na comparação de tais valores entre séries, permitindo o cotejo das características fractais das séries de TSM e de suas alterações de leste para oeste, do Atlântico para o Pacífico. A série de níveis em Manaus também foi analisada.

7.3 EMD

As séries de TSM foram decompostas usando a EMD antes de alimentar os modelos híbridos. A revisão mostrou o problema dos contornos e os demais problemas decorrentes do caráter empírico e adaptativo da EMD. A seguir apresentam-se as alternativas adotadas para mitigar tais efeitos.

Problema dos Contornos

Optou-se por implementar um esquema de tratamento dos contornos similar ao de TIANLU *et al.* (2013), utilizando-se como base de pesquisa 1 ½ ondeleta do início e do fim da série para completar seus respectivos extremos. O acréscimo de meia ondeleta foi feito considerando-se a presença de pulsos intermitentes junto aos picos principais da série, permitindo, assim, intervalos de comparação mais representativos das tendências mais relevantes das séries temporais.

No entanto, o esquema não faz sentido em baixas frequências, por não haver períodos bastantes para a comparação. Portanto, ao extrapolar envoltórias em conjuntos de extremos com menos que quatro elementos, optou-se pelo espelhamento dos pontos extremos usando as ondeletas vizinhas aos contornos.

Mixagens entre modos

As metodologias da EEMD e da CEMD foram testadas nos dados de interesse buscando mitigar o problema das misturas de frequências entre modos. Mas a presença de espectros cheios e de pulsos como no caso da OMJ, além de patamares de máximos e mínimos fizeram com que a separação correta das frequências não fosse observada.

No caso da EEMD, existe um custo computacional de implementação muito elevado, sem garantia de resultados satisfatórios. Quanto maiores os ruídos introduzidos, maior o número de iterações a fim de neutralizá-los.

Já a CEMD, quando aplicada aos sinais utilizados nesta tese, precisou de um processo de tentativa e erro para melhor definir quais frequências necessitariam ser extraídas separadamente com a definição dos pseudo-extremos, também sem garantia de sucesso.

Como os dados de TSM utilizados foram interpolados a partir de dados mensais médios, a mixagem de modos ocorreu geralmente entre a primeira IMF, relativa a efeitos intrasazonais, e a segunda IMF, que deveria conter apenas a sazonalidade. Para evitar custos computacionais e de implementação das metodologias de mitigação de mixagens de modos, adotou-se a soma das duas primeiras IMFs das TSMs sempre que tal problema for identificado.

Cruzamento de Envoltórias

Não foram adotados métodos exclusivos para a mitigação do cruzamento de envoltórias. As interpolações de envoltórias de ZHU *et al.* (2013) e KOKES & NGHIEN (2012) precisariam de soluções desenvolvidas na presente pesquisa, acarretando em custos de implantação e computacionais. Como o problema não se mostrou decisivo para o desenvolvimento da pesquisa, tais metodologias não foram empregadas.

Cabe lembrar que a interpolação suave dos contornos ajuda a reduzir possíveis problemas de cruzamento de envoltórias gerados por splines ajustadas sem apoios no início e no fim das séries.

7.4 A PCA e a PLS

A PCA e PLS foram testadas no pré-processamento de dados de TSM. A pesquisa bibliográfica destacou a possibilidade de ocorrência de multicolinearidade no jacobiano de redes *feedforward*. A colinearidade pode ocorrer, já que, além de influências mútuas entre as temperaturas das duas bacias oceânicas, existem algumas escalas hierárquicas comuns nos dois sinais e que poderão originar IMFs similares.

Assim, a PCA foi usada para usada para transformar o conjunto de estimadores em um novo conjunto linearmente independente, organizado do componente principal de maior para o de menor relação com a variabilidade dos preditores. Já a PLS transformou o conjunto de estimadores em um novo conjunto linearmente independente, organizado do componente principal de maior relação com a variabilidade de níveis em Manaus para o de menor variabilidade.

7.5 RNAs e cenários

Finalmente, após a decomposição dos dados de TSM com a EMD e sua transformação pela PCA ou PLS, os mesmos alimentaram RNAs *feedforward* de duas camadas com funções de ativação sigmoidais nos neurônios da camada intermediária e funções de ativação lineares nos neurônios da camada de saída. As redes foram treinadas com algoritmo LM *backpropagation*. A ferramenta *nftool* do aplicativo MATLAB foi usada para construção e simulações dos modelos.

A regularização Bayesiana foi usada no algoritmo LM, permitindo o auto-ajuste do parâmetro de regularização dos pesos. Espera-se que a utilização conjunta da PCA ou PLS e da regularização consiga eliminar parâmetros supérfluos para o ajuste, mitigando a multicolinearidade e o superdimensionamento nos modelos.

Em cada uma das aplicações, os dados foram separados em conjunto de treinamento e de testes, sempre da seguinte forma: conjunto de treinamento compreendendo dados selecionados com intervalos fixos de 15 dias; conjunto de testes compreendendo os demais dados. Com as dimensões da bacia em questão, tal frequência de amostragem foi capaz de descrever satisfatoriamente o fenômeno, pois os picos e depressões da curva ficaram bem resolvidos. Os resultados de cada simulação puderam assim ser comparados a um conjunto fixo de testes, permitindo avaliar a efetividade do estimador nos modelos híbridos e no modelo convencional.

Buscando melhor generalização dos resultados de cada família de modelo, foram testadas seis arquiteturas diferenciadas pelo número N de neurônios na camada intermediária sendo $N=(3, 5, 8, 12, 15, 18)$. Cada modelo assim composto foi simulado 1.000 vezes, permitindo avaliação de seu comportamento médio.

7.6 Cenários comparativos entre modelos EMD-RNA e modelos convencionais

RNA

Os resultados dos modelos híbridos foram comparados aos modelos convencionais RNA alimentados diretamente com os dados de TSM. A Figura 7-1 ilustra o fluxograma da modelagem.

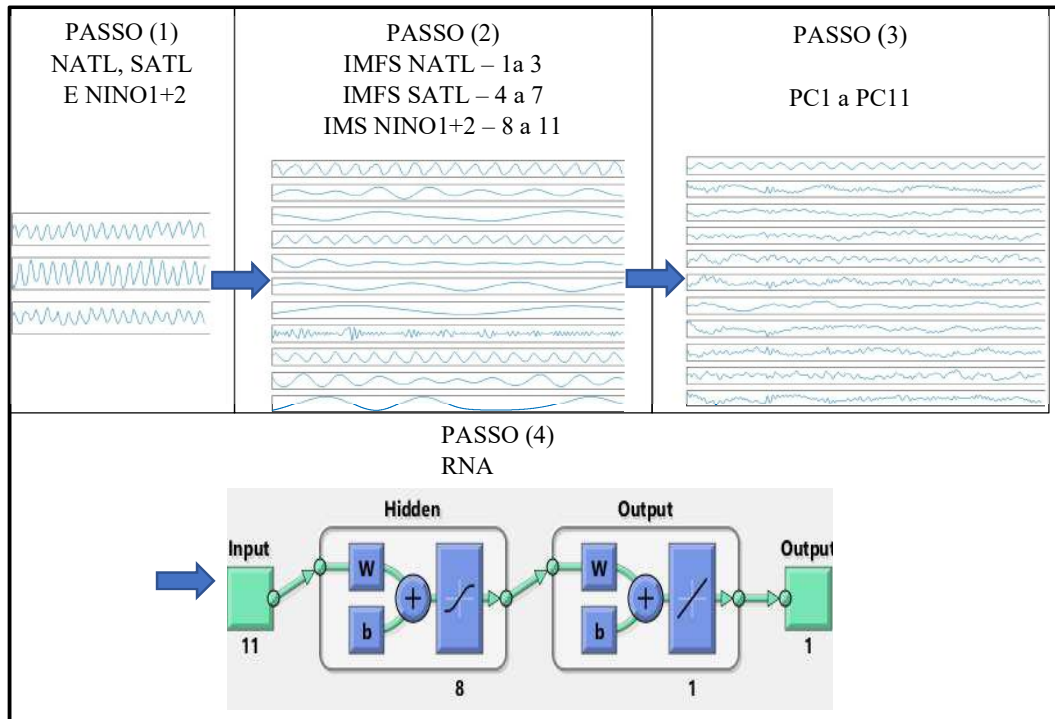


Figura 7-1. Fluxograma da modelagem híbrida proposta, a saber: Passo (1) composto da sincronização correlacional dos preditores NATL, SATL e NINO1+2 com o alvo e truncamento das séries nas datas correlatas ao dia da previsão do alvo; Passo (2) constitui-se da EMD dos vetores do passo (1), num exemplo no qual obtiveram-se 11 IMFs no total; Passo (3) é a análise pela PCA ou pela PLS.

Com a adoção da EMD em séries de TSM correlatas, esperam-se famílias de IMFs resultantes com alta correlação entre diferentes séries, em bandas de frequências semelhantes. O passo (3) da Figura 7-1 busca evitar problemas decorrentes de mal-ranqueamento e mal-condicionamento do jacobiano das RNAs. A proposta é projetar o conjunto total de IMFs de todos os postos de TSM em uma base ortogonal, sendo pela PLS ou pela PCA, e comparar o desempenho dos modelos assim alimentados com os demais.

Assim, houve quatro famílias de modelos testados, cada qual caracterizada por uma metodologia de pré-processamento de dados de TSM:

- PLS, com modelos cujos dados de TSM foram pré-processados com a EMD seguida da PLS, compreendendo os passos (1) a (4) da Figura 7-1;
- PCA, com modelos cujos dados de TSM foram pré-processados com a EMD seguida da PCA, compreendendo os passos (1) a (4) da Figura 7-1;
- C2, com modelos cujos dados de TSM foram pré-processados com a EMD, compreendendo os passos (1), (2) e (4) da Figura 7-1;
- C1, com modelos convencionais, alimentados com as séries de TSM sem pré-processamento, compreendendo os passos (1) e (4) da Figura 7-1.

O fato é que existem mais possibilidades de cenários, considerando-se os parâmetros da EMD e das RNAs. Variando-se os critérios de parada da EMD, por exemplo, obtém-se um conjunto diferente de IMFs. O mesmo vale para a escolha ou não de método de mitigação dos problemas da EMD. Na RNA, pode-se fixar o parâmetro de regularização ou deixar que se ajuste internamente. O número de nós da camada intermediária pode ser modificado, buscando-se o melhor ajuste.

Para evitar um número impraticável de simulações, apresentaram-se, anteriormente, metodologias adotadas para as questões da EMD. Já a generalização dos modelos foi feita através da escolha auto-ajustável do parâmetro de regularização e sem eliminar componentes principais frutos da PLS ou da PCA. A parada forçada não foi utilizada. Portanto, resultaram quatro famílias de modelos segundo a metodologia de pré-processamento, cada qual com seis configurações da camada intermediária dependendo do número de nós $N = (3, 5, 8, 12, 15, 18)$, totalizando 24 modelos por teste, cada qual simulado 1.000 vezes em cada uma das previsões.

A identificação dos diversos cenários de simulação foi feita pela sigla da respectiva família de modelo seguida de N e o número de nós na camada intermediária. Assim, PLSN18 identifica o modelo híbrido com 18 nós na camada intermediária e cujos vetores NATL, SATL e NINO1+2 foram pré-processados com a EMD e com a PLS.

Além disso, foram conduzidas investigações sobre a generalização dos modelos, usando as matrizes de auto-correlação dos Jacobianos e os números de condicionamento, permitindo comparação entre modelos híbridos e convencionais.

7.7 Cenários complementares de simulação

Além do elevado número de simulações objetivando comparar modelos EMD-RNA e os modelos RNA convencionais (da ordem de $5,0 \times 10^5$ simulações), foram realizadas simulações complementares, apenas em caráter experimental, como sugestão de investigação futura, buscando explorar a utilização da validação cruzada com amostragem *k-fold* estratificada, com $k = 2$, associada à utilização da parada antecipada para generalizar os modelos. Para tanto, foram simulados os modelos C1N8 e C2N8, na previsão da mínima anual de 31/10/02, tendo sido essa escolha arbitrária.

A estratificação se refere à escolha de dois subconjuntos (“dobras”) amostrados do conjunto de treinamento em cada simulação, porém com características específicas (em determinado estrato) que os qualifiquem para a validação. No caso em apreço, em cada simulação, foram desentranhados dois conjuntos ($k = 2$) de pontos semelhantes, cada qual pegando uma mínima anual sorteada aleatoriamente e os 59 dias que a antecedem. Os 30 pontos que sucedem os conjuntos de validação também não foram usados para treinamento, para aumentar a independência entre validação e treinamento. A Figura 7-2 ilustra a amostragem proposta.

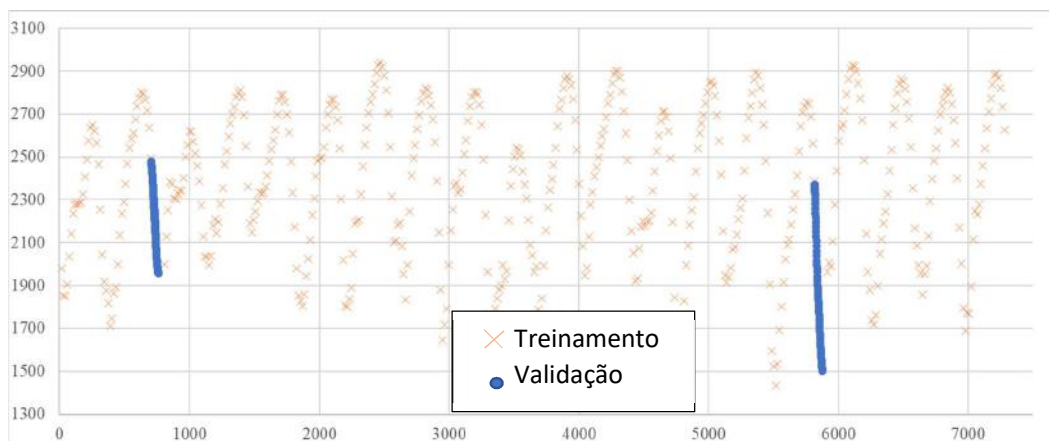


Figura 7-2. Nova estratégia de amostragem para a validação cruzada *k-fold*, com $k = 2$: eixo vertical em cm e eixo horizontal em dias; em laranja estão os pontos de treinamento e em azul, os dois subconjuntos de testes, pegando-se mínimos sorteados e os 59 pontos antecedentes.

Diferente da técnica de amostragem proposta nos testes comparativos entre modelos híbridos e convencionais RNA, a validação cruzada *k-fold* mostrada na Figura 7-2 priorizou cotejar o desempenho dos modelos durante a simulação, revelando possíveis ocorrências de *overfitting* nos modelos híbridos que justifiquem as elevadas variâncias de seus desvios absolutos médios, observadas anteriormente, mesmo diante de números de condicionamento dos seus Jacobianos serem significativamente menores que dos modelos convencionais.

Os gráficos de desempenho dos modelos C1N8 e C2N8 durante as simulações com validação cruzada, as matrizes de auto-correlação, serão apresentadas, bem como seus números de condicionamento, fornecendo mais argumentos a discussão acerca da generalização dos modelos.

Por fim, foram conduzidas simulações com o modelo probabilístico de Galdino & Clarke – MP, adaptado por PIMENTEL *et al.* (2011) para previsão de níveis máximos anuais em Manaus, permitindo assim, a comparação de seu desempenho com o dos modelos híbridos que apresentaram os melhores resultados nas previsões de máximas anuais realizadas. Essa base comparativa forneceu subsídios para avaliar melhor, não só o esquema híbrido EMD-RNA mas também a utilização de dados de TSM como preditores dos níveis em Manaus.

8. Dados

As séries de dados de TSM e de OLR usadas na presente pesquisa foram disponibilizadas pela *National Oceanic and Atmospheric Administration* – NOAA. Elas são fruto de interpolações de uma grade de resolução espacial $1^\circ \times 1^\circ$ (KALNAY *et al.* 1996). A última atualização das séries para o presente fim foi feita em junho de 2017. Até aquela ocasião, a NOAA fornecia as séries de TSM, em $^\circ\text{C}$, já interpoladas espacialmente para as seguintes regiões:

- Dados médios mensais de TSM referentes ao El Niño - NINO1+2 ($0^\circ\text{-}10^\circ\text{S}$) ($90^\circ\text{-}80^\circ\text{O}$), NINO3 ($5^\circ\text{N-}5^\circ\text{S}$) ($150^\circ\text{-}90^\circ\text{O}$), NINO3.4 ($5^\circ\text{N-}5^\circ\text{S}$) ($170^\circ\text{-}120^\circ\text{O}$) e NINO4 ($5^\circ\text{N-}5^\circ\text{S}$) ($160^\circ\text{L-}150^\circ\text{O}$);
- Dados médios mensais de TSM do Atlântico Norte, Sul e Trópicos Globais, respectivamente, NATL($5^\circ\text{-}20^\circ\text{N}$) ($60^\circ\text{-}30^\circ\text{O}$), SATL($0^\circ\text{-}20^\circ\text{S}$) ($30^\circ\text{O-}10^\circ\text{L}$) e TROP ($10^\circ\text{N-}10^\circ\text{S}$) ($0^\circ\text{-}360^\circ$);

As séries NATL e SATL são sinais meridionalmente anti-correlatas e captam o dipolo do Atlântico, descrito na revisão bibliográfica, como de forte influência na ZCIT, na pluviosidade de Amazônia e do NEB.

As séries mensais de OLR (OLR1, OLR2 e OLR3) em W/m^2 possuem a mesma resolução espacial bruta que as séries de TSM e foram interpoladas para as regiões: OLR1, cobrindo a Amazônia ($3^\circ\text{N-}10,5^\circ\text{S}$) ($80,6^\circ\text{-}50,6^\circ\text{W}$); OLR2, sobre uma área antropizada da bacia do rio Madeira ($8,6^\circ\text{-}12,4^\circ\text{S}$) ($65,6^\circ\text{-}60^\circ\text{W}$) e OLR3, na porção oeste da Amazônia, a montante de MAN ($2,9^\circ\text{N - }12,4^\circ\text{S}$) ($75^\circ\text{-}60^\circ\text{W}$). Tais séries estão disponíveis desde jan/1948.

O mapa da Figura 8-1 ilustra as regiões de interpolação das séries de TSM e de OLR de interesse da pesquisa, representadas como retângulos na projeção cilíndrica. As bacias hidrográficas de interesse e a localização do posto de Manaus também são mostradas.

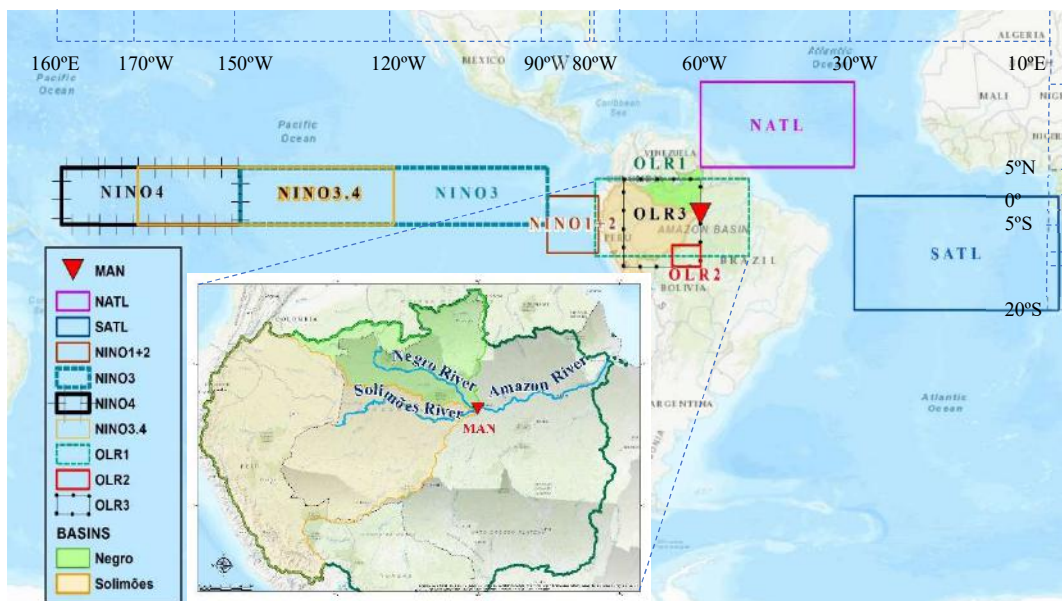


Figura 8-1. Mapa destacando regiões de interpolação de séries de interesse da pesquisa e bacias hidrográficas estudadas, além do posto HIDROWEB/ANA 14990000, cuja série de níveis médios diários foi alvo dos modelos testados.

Para os estudos de coerência espectral e para as simulações com RNAs, buscaram-se períodos comuns entre os dados de TSM, OLR e a hidrologia em Manaus. Assim, foram utilizadas séries de 01/1982 a 12/2016. Após a obtenção das interpolações espaciais mensais, as séries de TSM e de OLR foram interpoladas para valores diários, para compatibilização com a resolução temporal da série MAN. Mais detalhes sobre a origem dos dados e métodos de interpolação e de consistência estão em <https://www.esrl.noaa.gov/psd/data/gridded/index.html>.

As séries são mostradas na Figura 8-2. Os valores médios mensais, representados em vermelho, sobrepõe-se às curvas interpoladas utilizando-se spline cúbica. O período de out/95 a abr/01 foi destacado na série NINO3.4 evidenciando os efeitos da interpolação que, mesmo na série mais ruidosa de TSM, não acrescentou variabilidade, aparentemente. Os gráficos revelam ainda as surpreendentes tendências de aumento das TSMs e de redução da OLR no período estudado.

Os valores de OLR2 possuem amplitude maior em relação a OLR1 e OLR3. Esse pode ser um efeito das escalas de promediação espacial, já que OLR2 foi interpolada sobre uma região menor do rio Madeira, com desmatamento nas últimas décadas. Porém pode

ser também um efeito da retirada da vegetação, que altera o balanço do fluxo de calor junto à superfície.

A série alvo dessa pesquisa constitui um valioso e extenso registro diário dos níveis no porto de Manaus, desde 15/09/1902 até os dias de hoje. A estação hidrológica de código HIDROWEB/ANA 14990000 localiza-se no rio Negro (3,13833° S; 60,027222° W), afluente do Amazonas, com área de drenagem de aproximadamente 712.000 km². A estação 1499000 é operada pelo Serviço Geológico do Brasil – CPRM e está inserida no Sistema Nacional de Informações de Recursos Hídricos - SNIRH, cabendo à Agência Nacional de Águas – ANA, organizar, implantar e gerir o SNIRH, de acordo com a sua lei de criação, lei nº 9.984, de 17 de julho de 2000.

Os dados são coletados por telemetria desde 01/06/2001 e representam níveis médios diários com referência arbitrária local, sendo medidos em centímetros. A série pode ser acessada em http://www.snirh.gov.br/hidroweb/publico/medicoes_historicas_abas.jsf. Na Figura 8-3 ilustra-se MAN. Sobrepostos à série, destacam-se o período de estudo em verde e os eventos extremos anuais que foram alvos dos testes para aferição do desempenho do modelo híbrido, sendo dez níveis máximos e dez mínimos.

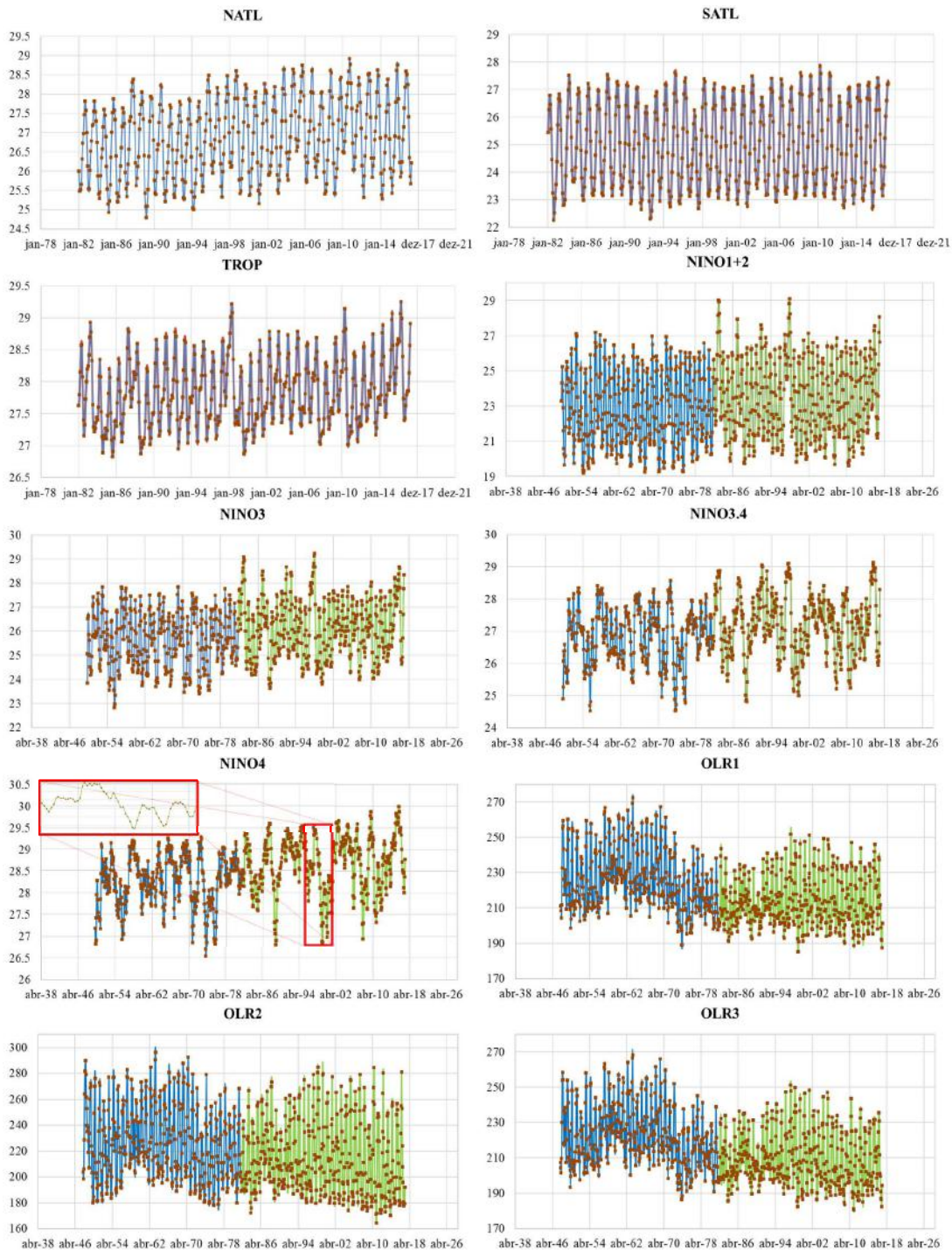


Figura 8-2. Séries temporais de TSM do Atlântico e do Pacífico em °C e de OLR em W/m²—marcadores em laranja são série média mensal e curva azul é série interpolada diariamente. Destaque em NINO4 mostrando interpolação diária por spline cúbica.

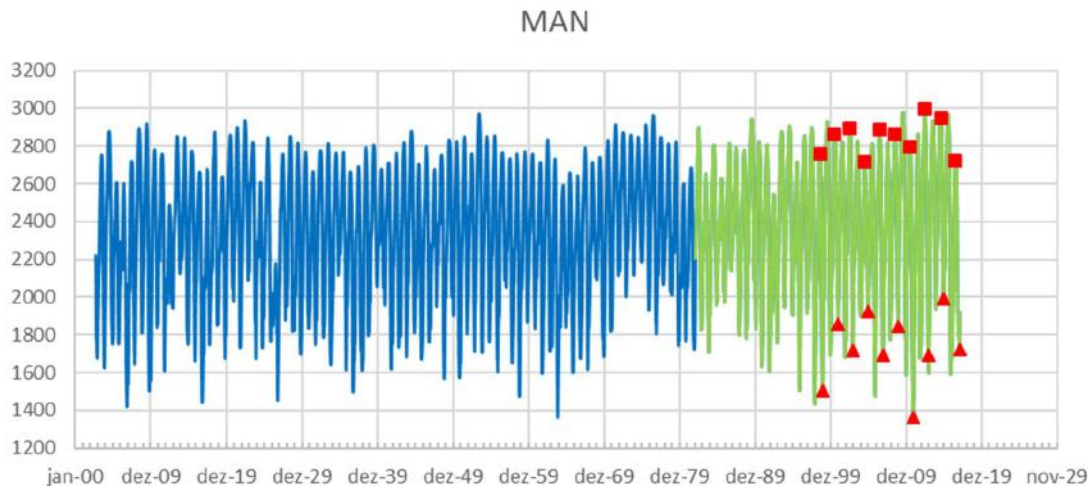


Figura 8-3. Série de níveis médios diários em cm da estação HIDROWEB-ANA 1499000, destacando-se trecho alvo do estudo em verde e níveis alvos da modelagem com marcadores vermelhos, sendo 10 testes com níveis máximos e dez com níveis mínimos.

As séries de TSM de NATL, SATL e de NINO1+2 foram escolhidas como preditores por possuírem as maiores correlações com MAN. Outras séries estudadas poderiam ser incluídas nas simulações, como a OLR. No entanto, com o aumento do número de preditores, aumentariam também os parâmetros das RNAs, aumentando o custo computacional e podendo agravar problemas de superdimensionamento. Além disso, a correlação existente entre as séries estudadas poderia aumentar ocorrência de multicolinearidade no jacobiano das RNAs e mal condicionamento, como aponta a literatura.

Para ilustrar os resultados da EMD, apresentam-se as séries preditoras, as respectivas IMFs e os resíduos do peneiramento nas Figuras Figura 8-4, Figura 8-5 e Figura 8-6. Em todos os casos, houve mixagem de frequências até a escala sazonal, pela intermitência de pulsos de frequências maiores, mormente na faixa intrasazonal. Para não prejudicar o treinamento das RNAs, as IMFs 1 e 2 foram somadas.

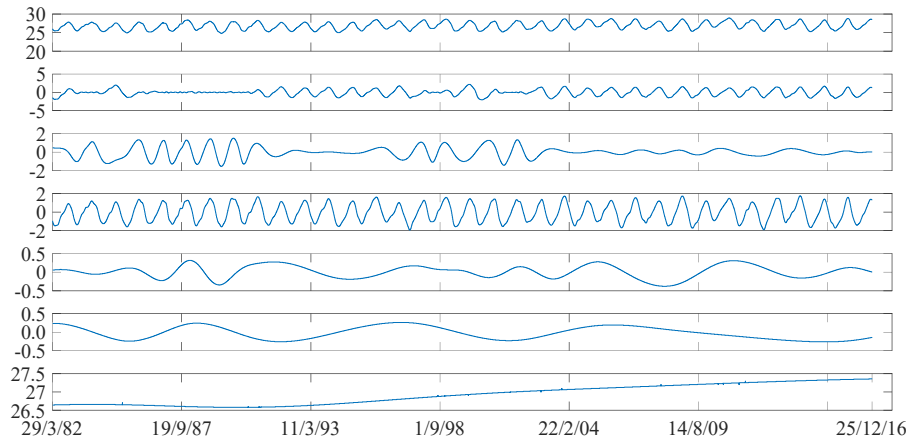


Figura 8-4. Na seqüência: série NATL, IMFs 1 e 2, IMF1+2, IMFs 3 a 4 e resíduo da EMD.

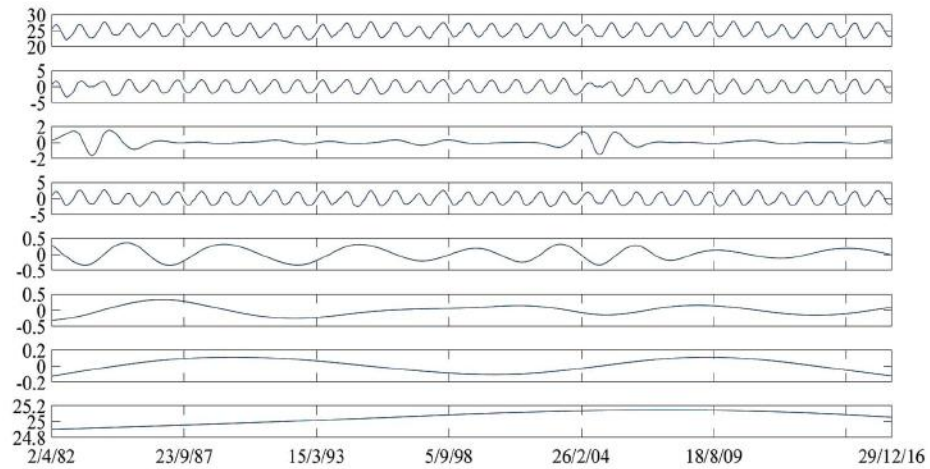


Figura 8-5. Na seqüência: série SATL, IMFs 1 e 2, IMF1+2, IMFs 3 a 5 e resíduo da EMD.

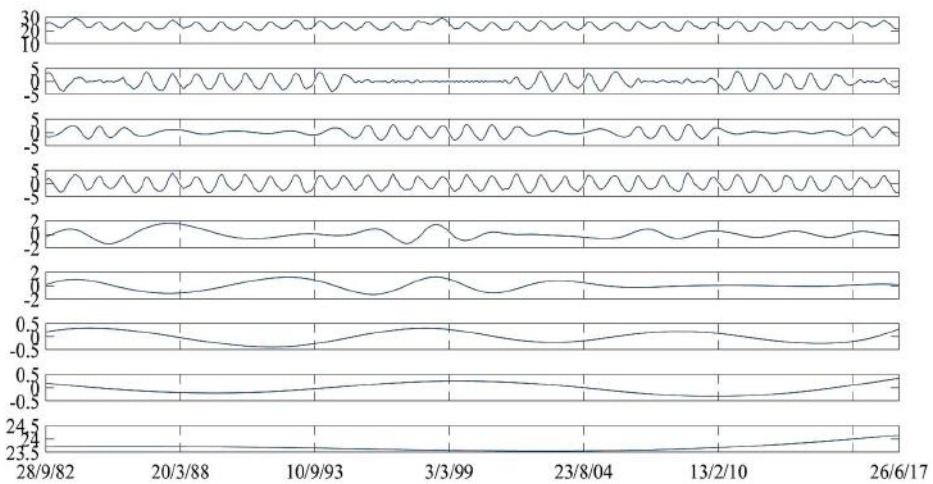


Figura 8-6. Na seqüência: série NINO1+2, IMFs 1 e 2, IMF1+2, IMFs 3 a 6 e resíduo da EMD.

9. Aplicações e Resultados

Os estudos de análise fractal e de coerência espectral são apresentados, evidenciando relações fenomenológicas e semelhanças fractais entre as séries de TSM e de OLR e a hidrologia em Manaus. Tais resultados são comentados no presente item, objetivando embasar e justificar a utilização de tais dados na alimentação dos modelos híbridos.

Em seguida, o desempenho dos modelos híbridos será apresentado em relação ao modelo convencional RNA, considerando os cenários descritos na metodologia. Com base na disponibilidade de dados de TSM dos preditores escolhidos, 20 testes foram montados para os 24 modelos em questão, sendo dez para prever cotas máximas anuais e dez para prever cotas mínimas anuais, todos tendo as séries NATL, SATL e NINO1+2 como preditores e MAN como alvo.

Nas simulações dos 20 testes, cabe lembrar, os conjuntos de treinamento e de testes das RNAs foram desmembrados de forma semelhante, amostrando-se, sequencialmente, um ponto para treinamento e 15 para testes, ao longo das séries estudadas. Essa amostragem buscou um equilíbrio entre o nível de independência desejado entre os conjuntos de testes e de treinamento e o desfalque causado nas amostras ao desentranhar-se o conjunto de testes.

A regularização Bayesiana foi utilizada para generalizar os modelos. Portanto, o conjunto de testes foi utilizado tão somente para aferir discrepâncias entre as estimativas de desvios feitas com tal conjunto das simulações pretendidas. Assim, pretendeu-se identificar possíveis sinais de superdimensionamento e de mal-condicionamento, comparando tais características entre modelos.

O desempenho dos 24 modelos propostos foi cotejado estudando-se a distribuição dos resíduos em 1.000 simulações de cada modelo em cada um dos 20 testes. Os melhores modelos convencionais C1 para máximas e para mínimas – C1MAX e C1MIN foram selecionados e confrontados com os modelos híbridos. Os melhores modelos híbridos serão apresentados.

Serão apresentadas ainda, duas simulações adicionais feitas com configurações semelhantes de arquitetura entre um modelo convencional C1 e um modelo híbrido C2, utilizando-se a validação cruzada com amostragem *k-fold* estratificada, propiciando a parada antecipada. Essas simulações complementares objetivaram um cotejo da evolução do desempenho das redes a partir de uma amostragem de conjuntos de testes mais independentes do conjunto de treinamento, mas que por outro lado, certamente desfalcam o treinamento mais severamente, impactando nos resultados.

9.1 Análise Fractal

Os padrões fractais dos sinais de TSM, de OLR e de níveis em MAN foram estudados através da DFA. Utilizaram-se escalas de análise equivalentes. Para permitir a melhor caracterização das séries, cada uma delas foi analisada ao longo de seu respectivo período máximo de dados disponíveis, adaptando-se assim as janelas máximas de análise ao tamanho de cada série.

Como a DFA é parte do objetivo secundário de caracterização de relações fenomenológicas entre os dados estudados e a hidrologia em MAN, os resultados foram comentados neste item, a fim de embasar a utilização das séries como preditores dos níveis em MAN, sendo assim também com a coerência Wavelet.

Os escalogramas resultantes da DFA dos sinais são apresentados na Figura 9-1 em escala bilogarítmica, com eixos verticais em °C para as TSM, em W/m² para OLR e em cm para MAN. Os eixos horizontais são as janelas temporais analisadas, em dias. Foram ajustadas retas, apresentando-se os respectivos coeficientes angulares em cada trecho auto-afim das séries. Intercalando tais trechos, observam-se quebras, que indicam hierarquia de certas escalas.

As escalas com períodos menores que 12 dias devem ser desconsideradas, por englobarem apenas dados de interpolação a partir de médias mensais, exceto em MAN. No entanto, as inclinações de todos os sinais foram notavelmente semelhantes nas frequências menores, refletindo a capacidade de regularização da bacia a montante de MAN.

As curvas ajustadas para as escalas entre 12 e 108 dias e entre 162 e 324 dias apresentaram similaridades notáveis entre todas as séries estudadas, com inclinações consistentes de

cerca de 1,9 e 1,4, respectivamente. Tais trechos correspondem aos fenômenos intrasazonais e sazonais, faixas importantes nas relações encontradas no estudo de coerência Wavelet. As quebras são localizadas entre 9 e 12 dias e entre 10 e 162 dias, para a escala intrasazonal e entre 324 e 486 dias, para a escala sazonal.

Há sutis reduções das influências da escala hierárquica sazonal, de leste para oeste no Pacífico, com redução das inclinações entre 100 e 400 dias. Essas reduções não são observadas entre 10 e 100 dias, o que destaca a hierarquia intrasazonal no Pacífico, relacionada à OMJ e outras oscilações de oeste.

Explorando escalas auto-afins mais lentas que 648 dias, observam-se semelhanças menores entre os sinais, com discrepâncias, inclusive entre dados da mesma bacia oceânica. Porém, há similaridade nessa escala auto-afim entre MAN NINO1+2 e NATL, com cerca de 0,4 de inclinação. Os dados TROP possuem certa similaridade com NINO3 nessa faixa. Já SATL exibe declividade mais baixa, de 0.18, que os demais sinais em escalas interanuais e mais lentas.

Por fim, a similaridade pode ser avaliada de forma resumida no gráfico (j) da Figura 9-1. Nele, percebem-se alterações de leste para oeste na dualidade tropical. Os sinais NINO3 (e), NINO3.4 (f) e NINO4 (g) apresentaram características multi-fractais nas escalas mais lentas, com relações não lineares nos gráficos. A hierarquia da escala sazonal, bem destacada com as quebras nos demais escalogramas, é praticamente imperceptível no gráfico (g) de NINO3.4, que exibe inclinações menores entre 10 e 100 dias e maiores acima dos 400 dias, reforçando a maior importância do Pacífico nas oscilações de grande período e menor reflexo da sazonalidade em tais séries.

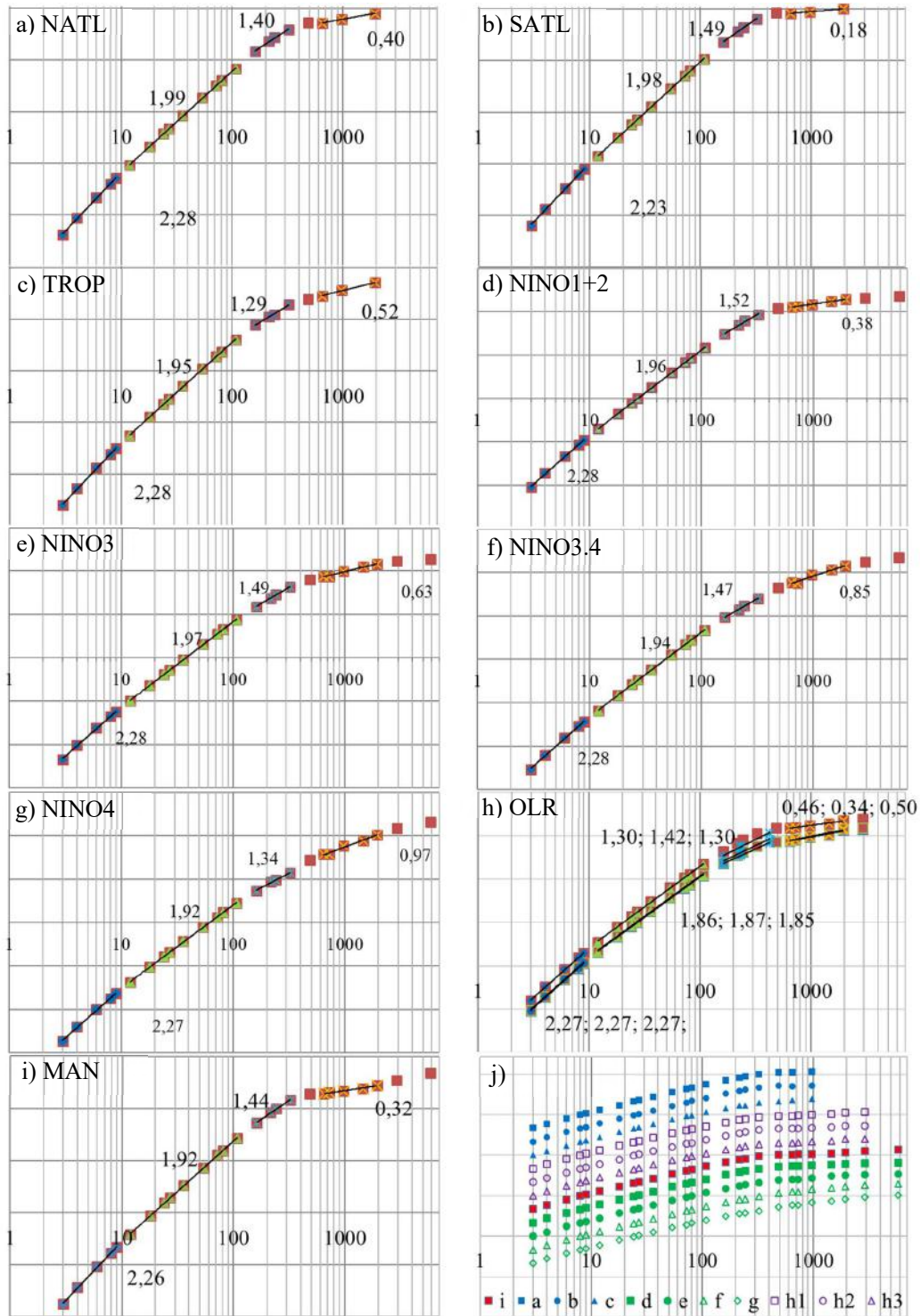


Figura 9-1. Escalogramas bilogarítmicos das DFAs dos sinais de TSM, de OLR e de MAN destacando-se trechos auto-afins, respectivas retas ajustadas e derivadas: eixos horizontais em dias, eixos verticais nas respectivas unidades (TSM em °C, OLR em W/m², níveis em cm).

9.2 Coerência espectral

A seguir, apresentam-se os escalogramas de coerência espectral Wavelet de MAN com as séries de TSM e de OLR na Figura 9-2. Os contornos interpolam o quadrado da coerência espectral ao longo do tempo (eixo horizontal em dias) e dos períodos (eixo vertical em dias) correspondentes à gama de frequências resultante da transformada Wavelet contínua de cada par de sinais.

As linhas contínuas pretas correspondem a significância de 5% relativa à hipótese nula de ajuste de distribuição paramétrica ao longo dos períodos, considerando os fenômenos quase estacionários. Já as setas indicam as defasagens locais como ângulos em um círculo trigonométrico indicando frações dos períodos indicados nos eixos verticais: sinais em fase têm ângulo zero e sinais anti-correlatos têm ângulo inicial igual a π .

Assim, os níveis em Manaus têm distribuição sazonal com notável correlação com a temperatura no Atlântico Sul -SATL, e nos Trópicos Globais – TROP, com defasagens indicadas por ângulos no primeiro quadrante, de cerca de 75° (SATL) e 65° (TROP). A defasagem indicada na faixa sazonal de NATL é semelhante àquela observada em SATL, porém, com ângulo contado a partir de π , uma vez que NATL possui correlação negativa com Manaus. A mesma referência inicial de defasagem vale para dados de TSM do Pacífico (d), (e), (f) e (g), todos com correlações negativas com MAN. As defasagens na faixa sazonal de MAN em relação aos sinais são similares às defasagens globais dos correlogramas, como resumido na Tabela 9-1.

Tabela 9-1. Coeficientes de correlação global e respectivas defasagens entre os níveis MAN e os sinais estudados.

	α	Δt
NATL	-0,82046	72
SATL	0,876045	76
TROP	0,650817	64
NINO1+2	-0,83947	254
NINO3	-0,73417	221
NINO3.4	-0,54821	187
NINO4	-0,39869	138
OLR1	-0,82739	108
OLR2	-0,8598	120
OLR3	-0,81471	114

Sobreposto aos espectros está cotograma contendo os níveis máximos anuais (quadrados) e os mínimos anuais (triângulos) com permanência inferior a 10% nos respectivos histogramas de picos anuais de Manaus desde 1902, resumidos na Tabela 9-2. A ocorrência de picos anuais de baixa permanência é maior que a esperada no domínio analisado. Como os escalogramas possuem extensões temporais relativamente pequenas, é difícil buscar explicações, na coerência espectral, para o elevado número de extremos de MAN.

Os escalogramas apontam redução gradual da coerência espectral com Manaus na faixa sazonal no Pacífico, de leste para oeste, de NINO1+2 para NINO4, resultado condizente com a análise fractal, Figura 9-1. Curioso observar que na faixa sazonal, as defasagens de MAN com a TSM também se reduzem gradualmente no Pacífico, de leste para oeste.

Outro fato notável são as correlações presentes em todos os escalogramas na faixa de 128 a 200 dias. Essa banda é relativamente mais lenta que aquela relacionada na literatura à OMJ e a eventos oscilatórios originados no Pacífico Oeste, com duração de 10-100 dias. Porém, como os dados de TSM e de OLR foram interpolados a partir de médias mensais, não se pode esperar caracterização de coerência espectral em faixas de períodos menores que 128 dias. As correlações de 128 a 200 dias foram mais fortes entre MAN e as TSMs do Atlântico e aumentaram gradualmente para oeste no Pacífico, enquanto a importância da hierarquia sazonal dos dados de TSM em MAN diminuíram em direção ao Pacífico Oeste, como adiantado na análise fractal.

Observa-se nos escalogramas uma sutil tendência de aumento das parcelas de energia contidas nos períodos maiores que 128 dias ao longo do tempo. Tal tendência é explorada a seguir, explorando possíveis relações com a intensificação dos picos anuais de níveis em MAN.

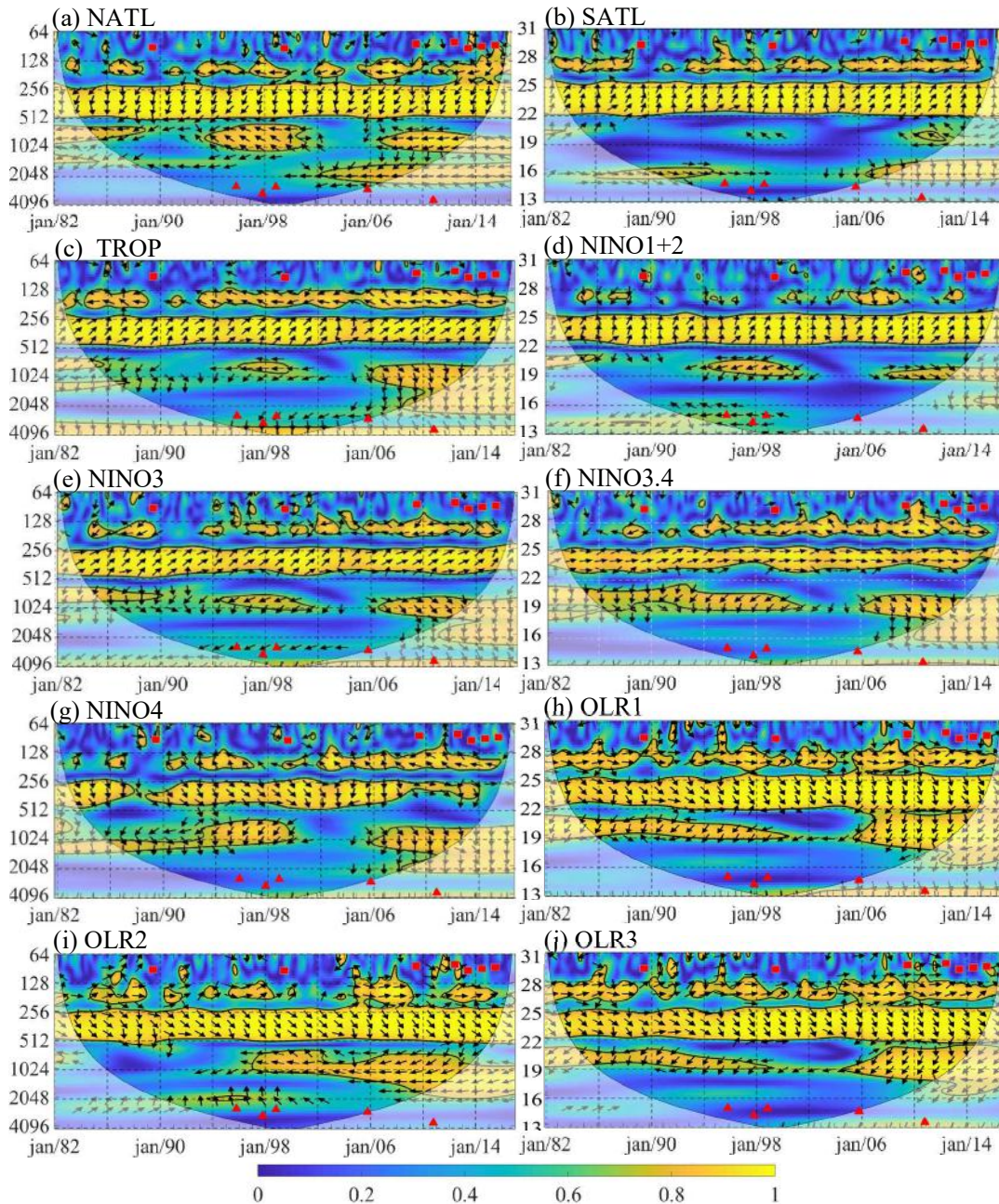


Figura 9-2. Coerência Espectral Wavelet ao longo dos períodos em dias (eixos verticais a esquerda) entre MAN e (a) NATL, (b) SATL, (c) TROP, (d) NINO1+2, (e) NINO3, (f) NINO3.4, (g) NINO4, (h) OLR1, ; (i) OLR2, (j) OLR3; a escala inferior é dos quadrados da coerência; as setas indicam defasagens como frações angulares dos períodos dos eixos verticais em um círculo trigonométrico; sinais em fase têm ângulo zero e sinais anti-correlatos têm ângulo inicial igual a π ; os marcadores vermelhos apontam máximos (quadrados) e mínimos (triângulos) anuais de MAN entre os dez mais extremos desde 1902 (eixos centrais em m); os contornos delimitados por linhas pretas demarcam significância de 5% sobre o ruído vermelho.

Tabela 9-2. Resumo dos extremos anuais com permanência abaixo de 10% entre jan/82 e dez/16.

ordem	data	máximo (m)	P>(%)	ordem	data	mínimo (m)	P<(%)
1	28/05/2012	29,97	1%	1	24/10/2010	13,64	1%
2	07/07/2009	29,77	2%	4	4/11/1997	14,34	3%
4	29/06/2015	29,66	4%	8	25/10/2005	14,75	7%
6	03/07/2014	29,5	5%	10	30/10/1998	15,03	9%
7	3/7/1989	29,42	6%	12	30/10/1995	15,06	10%
9	13/06/2013	29,33	8%				
10	23/06/1999	29,3	9%				

Assim, a Figura 9-3 ilustra a evolução temporal das parcelas de coerência Wavelet em períodos maiores que 128 dias entre os níveis em MAN e os sinais de TSM e de OLR de 1982 a 2016, excluindo-se os primeiro e o último ano dos escalogramas da Figura 9-2, para reduzir efeitos dos contornos das séries. Como os espectros foram escalonados de forma semelhante, os valores de coerência abaixo dos 128 dias foram somados em cada passo de tempo e divididos pelo total de células nessas posições do espectro. De novo, níveis extremos anuais foram sobrepostos. As tendências lineares ajustadas sugerem o aumento da coerência entre MAN e os outros sinais nas escalas resolvíveis.

Uma feição interessante na Figura 9-3 é a redução de WTC notada entre 1998 e 2008, após a qual existe aumento de ocorrência de níveis máximos anuais extremos. NATL, NINO3.4, NINO4 e OLR2 não apresentaram tal característica semelhante a uma oscilação.

Olhando novamente os escalogramas da Figura 9-2, percebe-se que o aumento de coerência espectral não se deu na escala sazonal, a assinatura principal das relações fenomenológicas entre os níveis em MAN e os demais dados estudados. As análises da coerência espectral e da evolução dos níveis extremos não nos permitem, por si só, saber se as relações fenomenológicas entre os níveis em MAN e os demais dados está se alterando definitivamente, mesmo pelo curto período do estudo. Mas pode suscitar uma discussão mais profunda das possíveis causas das alterações.

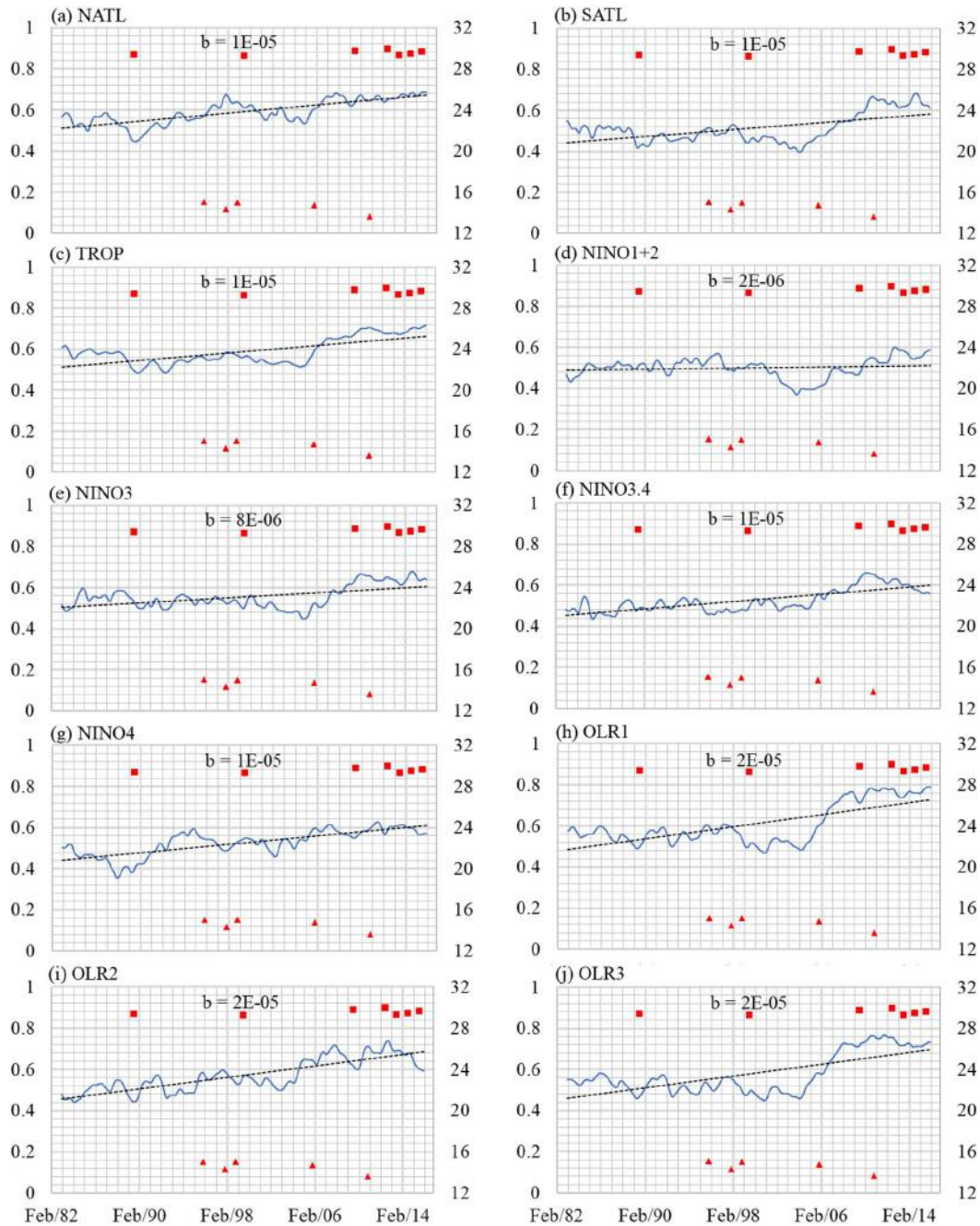


Figura 9-3. Evolução temporal das parcelas de coerência Wavelet em períodos maiores que 128 dias (eixos verticais esquerdos) sobreposta a níveis máximos (quadrados) e mínimos (triângulos) em MAN entre os dez mais extremos desde 1902 (eixos verticais direitos em m): tendências de aumento da coerência da TSM e OLR com níveis acompanhada de ocorrência de picos extremos em MAN, especialmente níveis máximos anuais.

Outros fenômenos não abordados interferem nas relações estudadas. Como exemplo, sabe-se que o rio Solimões causa efeito de remanso em postos muito a montante de Manaus, como em Moura, no rio Negro, código HIDROWEB/ANA 14840000, que apresenta pico correlacional de 96% com Manaus, com 10 dias de atraso. Ou seja, com

as baixas inclinações da região, não é só a hidrologia a montante de Manaus que molda os níveis usados no presente estudo. O remanso adiciona complexidade às análises e às previsões com os modelos híbridos, pois ocorrem apenas sob certas condições hidrodinâmicas, alterando amplitudes em MAN e defasagens em relação aos sinais estudados.

A uma primeira vista, o aumento da coerência Wavelet entre MAN e todos os sinais de TSM sugere a entrada em ação de um forçante externo na faixa de frequências mais lentas que 128 dias. O fato de a coerência entre MAN e os sinais OLR ter aumentado também corrobora com a ideia de uma excitação que não havia antes nos períodos mais lentos. Assim, uma possível explicação seria a tendência de aquecimento do Atlântico com o vazamento das Agulhas e o consequente fortalecimento da circulação de Walker, introduzindo energia comum a todos os sinais avaliados em baixas frequências.

A revisão bibliográfica discutiu possíveis efeitos, na hidrologia e no clima, do desmatamento em florestas tropicais. Para entender se a conversão de mais de 600.000 km² de florestas em pastagens (ANA, 2007) estão associadas às mudanças de clima e da hidrologia na Amazônia, foram estudadas as coerências Wavelet entre os sinais OLR1, OLR2 e OLR3 e NATL, SATL e NINO3.4 na Figura 9-4, e suas evoluções temporais em baixas frequências são exibidas na Figura 9-5, nos mesmo formato das Figuras Figura 9-2 e Figura 9-3.

As fortes relações entre MAN e TSM são também observadas entre OLR e TSM, com maior coerência na faixa sazonal com o Atlântico e maior coerência em escalas mais lentas e na escala intrasazonal com o Pacífico. Outro fato importante sobre a sazonalidade é que a defasagem média entre MAN e a TSM do Atlântico foi de cerca de 70 dias. Já entre MAN e a OLR, a defasagem foi de cerca de 120 dias. Os mínimos anuais de OLR (máxima cobertura de nuvens) ocorrem mais faseados com os mínimos anuais de NATL (máximos de SATL). Já os máximos anuais de OLR (mínima cobertura de nuvem) ocorrem antes do máximo anual de NATL (mínimo de SATL). Uma explicação plausível seria a reciclagem de chuva realizada pela floresta, que contribui com grande parte da OLR. Assim, uma componente de maior frequência aparece e deforma os picos mínimos, prolongando-os. Conseqüentemente, as fases das coerências espectrais entre OLR e a TSM do Atlântico na faixa sazonal devem ser lidas no sentido oposto do ciclo

trigonométrico, indicando que o pico das OLRs ocorre cerca de 50 dias antes das TSMs do Atlântico.

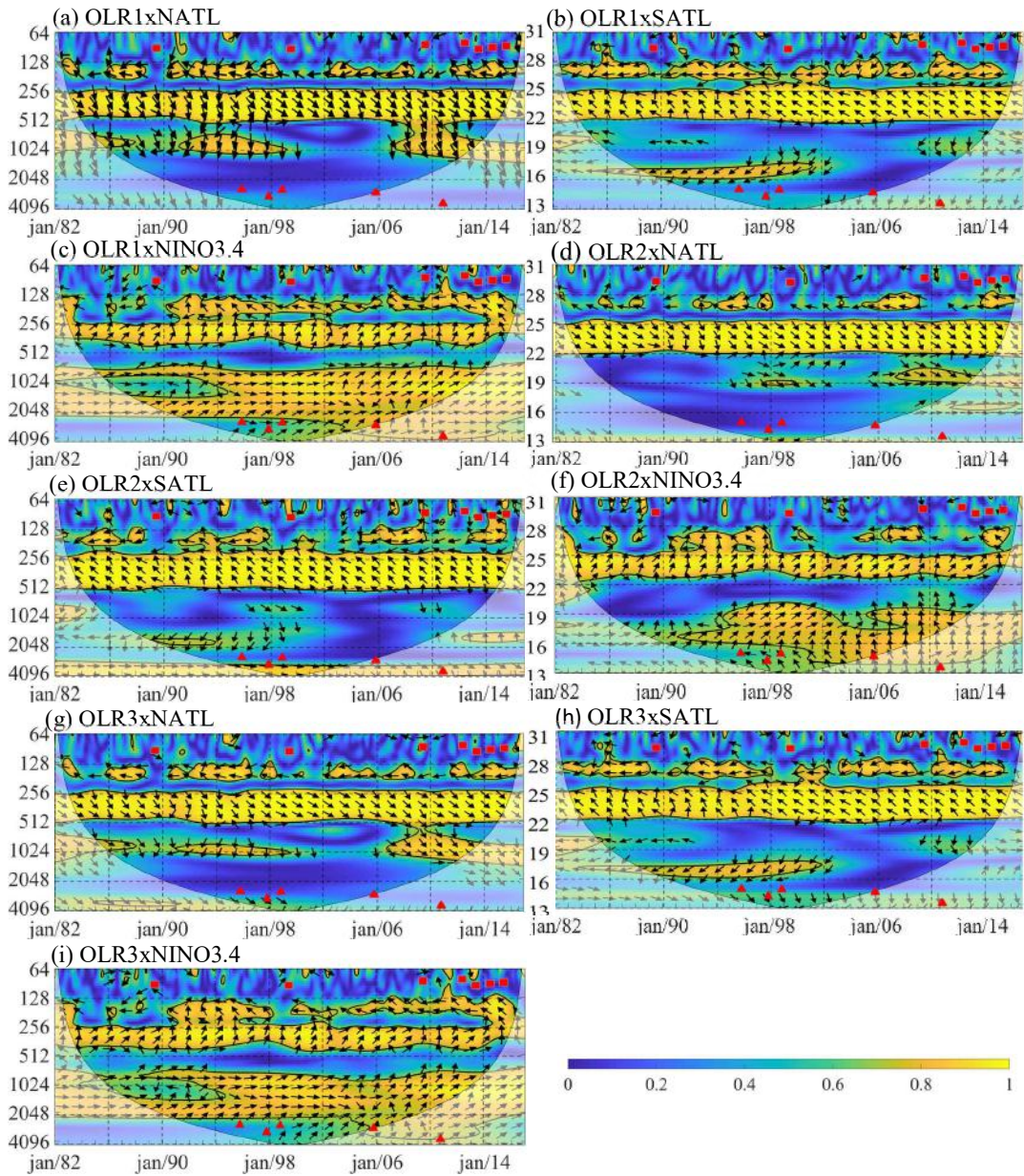


Figura 9-4. O mesmo que na Figura 9-2 para coerências entre as séries de OLR e NATL, SATL e NINO3.4.

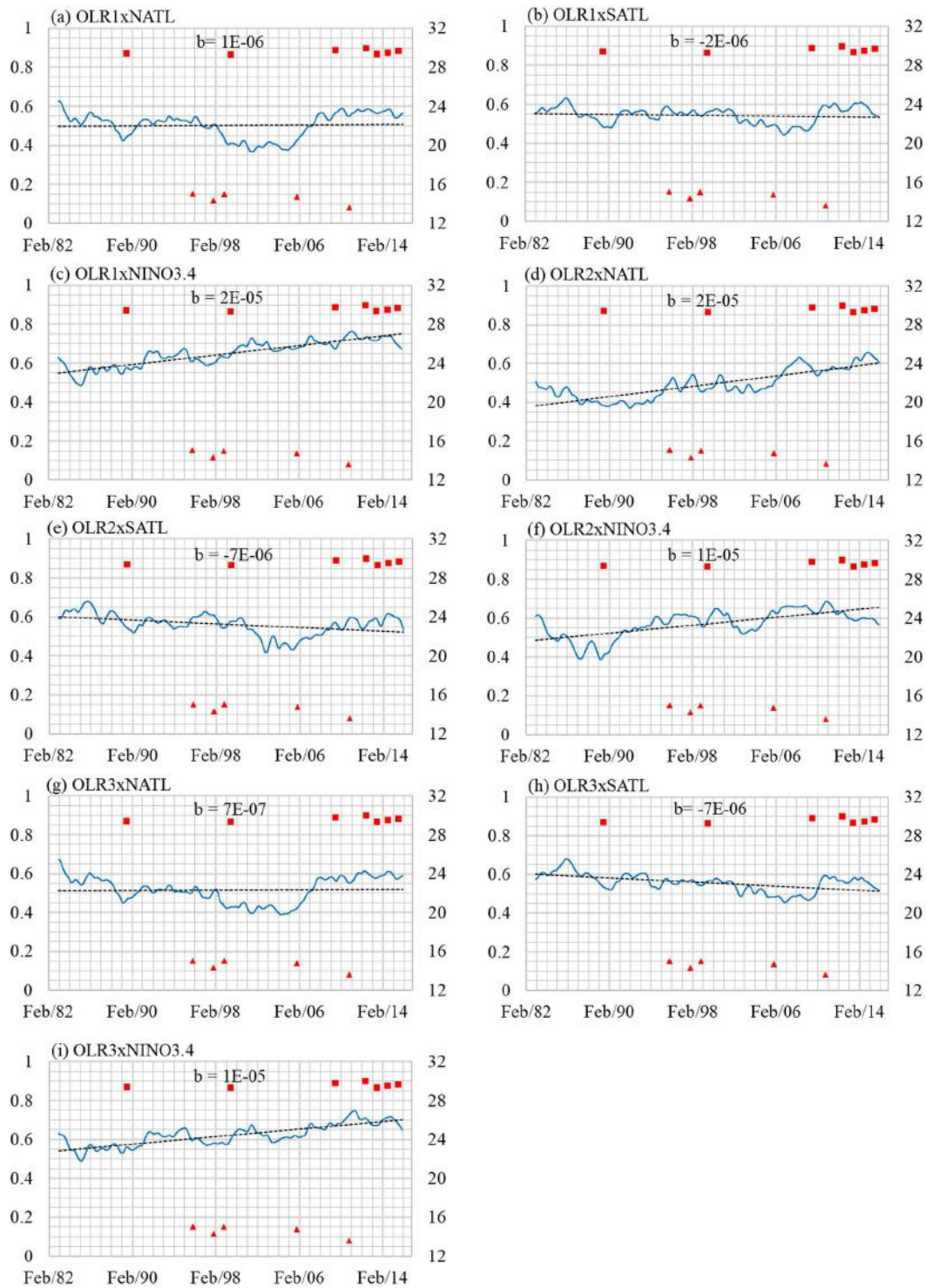


Figura 9-5. O mesmo da Figura 9-3 para coerências entre as séries OLR e NATL, SATL e NINO3.4.

A correspondente evolução da coerência Wavelet, no entanto, revela tendências menos definidas que as achadas na Figura 9-3. O teste não paramétrico de Mann-Kendall foi aplicado para a hipótese nula de ausência de tendências lineares monotônicas com

significância de 5% nas séries de evolução da coerência em baixas frequências estudadas. Adicionalmente, o Teste de Inclinação de Sen - livre tradução de *Sen slope test*, foi usado para estimar a magnitude das tendências. Tais procedimentos são largamente usados na literatura e recomendados pela Organização Mundial de Meteorologia e são descritos em MOREIRA & NAGUETTINI (2016).

Os resultados são resumidos na Tabela 9-3, destacando-se em vermelho os casos de redução da coerência espectral ou de não descarte da ausência de tendência. A hipótese nula de ausência de tendência na série é rejeitada se o valor absoluto de Z exceder o valor correspondente para a significância α escolhida de 5%, $|Z| > Z_{\alpha}/2$, sendo Z a estatística padronizada das tendências. Já *p-val* é estimado pela probabilidade cumulativa da distribuição normal ajustada à estatística da tendência analisada. A magnitude das tendências é apontada por (b), do Teste de Inclinação de Sen.

Tabela 9-3. Resultados do Teste de Mann-Kendall para análise de tendências na evolução da coerência Wavelet (H) e os coeficientes do teste de inclinação de Sen (b): (Z) é a estatística normalizada das tendências e (p-val) é a probabilidade cumulativa da distribuição normal associada.

	Z	p-val	b	H
MANxNATL	6,89	5,64E-12	1,35E-05	1
MANxSATL	2,71	6,71E-03	1,01E-05	1
MANxTROP	4,50	6,86E-06	1,20E-05	1
MANxNINO1+2	1,31	1,89E-01	2,65E-06	0
MANxNINO3	2,18	2,91E-02	7,40E-06	1
MANxNINO3.4	10,60	0,00E+00	1,18E-05	1
MANxNINO4	7,29	3,15E-13	1,33E-05	1
MANxOLR1	3,74	1,85E-04	1,94E-05	1
MANxOLR2	12,34	0,00E+00	1,89E-05	1
MANxOLR3	3,60	3,21E-04	1,77E-05	1
OLR1xNATL	0,68	4,95E-01	1,89E-06	0
OLR1xSATL	-0,84	4,04E-01	-1,60E-06	0
OLR1xNINO3.4	10,58	0,00E+00	1,69E-05	1
OLR2xNATL	7,24	4,34E-13	1,90E-05	1
OLR2xSATL	-3,70	2,19E-04	-6,04E-06	1
OLR2xNINO3.4	8,02	1,11E-15	1,38E-05	1
OLR3xNATL	0,02	9,87E-01	5,98E-08	0
OLR3xSATL	-3,74	1,82E-04	-8,01E-06	1
OLR3xNINO3.4	12,70	0,00E+00	1,33E-05	1

Com exceção da coerência entre MAN e NINO1+2, denotada por MANxNINO1+2, observaram-se tendências de aumento das coerências entre MAN e os sinais de TSM com

confiança estatística superior a 95%. No entanto, com exceção de OLR2xNATL, os demais dados de OLR apresentaram reduções das coerências com NATL e SATL ou não apresentaram tendências considerando a significância adotada.

A coerência mais persistente e com tendências crescentes entre a OLR e NINO3.4 para as faixas mais lentas chamou atenção e corrobora com as inclinações maiores encontradas nos sinais de TSM do Pacífico, mormente à oeste, na Figura 9-1. Concomitantemente, ocorreram tendências de diminuição entre as WTC de OLR2xSATL e de OLR3xSATL. Isso leva a questionamento sobre a relação entre o aquecimento do Atlântico e os eventos extremos em Manaus. MAN está mais intimamente ligada com a TSM do Atlântico do que a OLR1 e a OLR3. Apesar da OLR antecipar o escoamento superficial, a bacia amazônica e o Atlântico Sul estão em um ciclo ocasionado pelo afluxo de água doce do rio, que corresponde a cerca de 50% do total de contribuição de *runoff* naquele oceano. Os impactos da água doce do Amazonas na TSM do Atlântico e em modos importantes de larga escala do clima ainda são pouco conhecidos, JAHFER *et al.* (2017).

A OLR estudada também apresentou WTC crescente com Manaus com tendências mais fortes que as observadas com o Atlântico.

9.3 Resultados comparativos dos Modelos Híbridos e modelos RNA convencionais

O desempenho dos 24 modelos propostos foi cotejado através da distribuição dos resíduos em 1.000 simulações de cada modelo para cada um dos 20 testes. Em cada teste, as séries foram truncadas 60 dias antes da data do extremo máximo ou mínimo anual a ser previsto, transladando-se os dados de TSM de acordo com as defasagens ótimas apontadas pelos respectivos correlogramas com MAN até o truncamento. Em uma aplicação real dos modelos, as simulações seriam refeitas mediante aquisição de novos valores até se observar a inflexão do resultado nos 60 dias posteriores. As datas das máximas e mínimas cotas anuais, alvos dos testes são resumidas na Tabela 9-4.

Tabela 9-4. Datas das máximas e mínimas cotas anuais em MAN objetos dos testes: as séries de TSM e de MAN foram truncadas 60 dias antes; as séries de TSM foram trasladadas de acordo com defasagens ótimas apontadas pelos correlogramas; então, as previsões foram carregadas até as datas indicadas.

TESTE	MÁXIMAS	MÍNIMAS
T1	05/07/98	30/10/98
T2	26/06/00	11/12/00
T3	27/06/02	31/10/02
T4	12/06/04	17/11/04
T5	09/06/06	25/10/06
T6	20/06/08	28/10/08
T7	11/06/10	24/10/10
T8	28/05/12	22/10/12
T9	03/07/14	28/10/14
T10	15/06/16	13/12/16

Cabe ressaltar que nos 20 testes realizados, a estratégia de generalização adotada foi a regularização Bayeziana. Foram utilizados subconjuntos de testes idênticos em todos os modelos, tomando-se um ponto a cada 15 dias para treinamento e os demais pontos para testes. Apesar dessa estratégia não propiciar perturbações e variabilidade nos testes, eles não serão usados em estratégias de parada, mas apenas com intuito de vislumbrar as discrepâncias entre estimativas de desvios e os desvios de fato obtidos com as simulações.

Seleção do melhor modelo convencional C1:

Primeiramente, os melhores modelos convencionais C1 para máximas e para mínimas cotas anuais – C1MAX e C1MIN, foram selecionados comparando-se os módulos dos resíduos médios nas 1.000 simulações, aos 15, 30, 45 e 60 dias à frente das datas de truncamento, sendo o dia 60, correspondente ao pico de cota máxima ou mínima anual. Para tanto, utilizaram-se matrizes de cruzamento dos módulos dos resíduos médios dos modelos em cada um dos testes. A matriz de cruzamento do teste1, correspondente à previsão da cota máxima anual de 1998 é apresentada na Tabela 9-5. Os modelos RNA convencionais C1 com número de nós na camada intermediária $N = (3, 5, 8, 12, 15, 18)$ são comparados, computando-se 1, quando o desempenho do modelo de uma determinada coluna foi superior (menor módulo do resíduo médio) e zero, caso contrário. Na última linha, são totalizadas as ocasiões em que tal modelo foi superior. Repetindo-se essa análise para todos os testes de máxima, o desempenho relativo de cada modelo C1 foi computado nos dias 15, 30, 45 e 60 além do truncamento. O mesmo procedimento foi realizado para os testes de mínimas anuais.

Assim, os desempenhos dos modelos convencionais C1 foram resumidos na Tabela 9-6. As linhas de cima, totalizam as ocasiões em que o respectivo modelo C1 da coluna foi superior, comparado a cada um dos outros modelos C1 ao longo de todos os testes,

diferenciando-se os dez testes de previsão de máximas anuais (esquerda) dos dez testes de previsão de mínimas anuais (direita). As linhas de baixo representam a ponderação entre os pontos de cada modelo e os pesos P15, P30, P45 e P60, aplicados aos respectivos dias de simulação a frente do truncamento das séries.

Cada peso é a relação entre a soma dos quadrados dos resíduos médios no respectivo dia de simulação pela soma total dos quadrados dos resíduos médios em todos os dias de simulação, ao longo dos testes, distinguindo-se, mais uma vez, testes de previsão de cotas máximas dos de mínimas. A ponderação com valores quadráticos dá importância crescente ao desempenho dos modelos de 15 para 60 dias à frente do truncamento das séries, conforme aumenta a dificuldade de ajuste do modelo.

Tabela 9-5. Matriz de cruzamento de módulos dos resíduos médios dos modelos C1 com N = (3, 5, 8, 12, 15, 18) nós na camada intermediária, nos dias 15, 30, 45 e 60 à frente do truncamento das séries, para simulação da máxima anual de 06/05/1998: células com valor 1 indicam que modelo da respectiva coluna teve resíduo menor; caso contrário, o valor da célula é nulo.

TESTE1	N	3	3	3	5	5	5	5	8	8	8	8	12	12	12	12	15	15	15	15	18	18	18	18	M	M	M	M			
CI	d	15	30	45	60	15	30	45	60	15	30	45	60	15	30	45	60	15	30	45	60	15	30	45	60	15	30	45	60		
N	d	RES	48	96	127	178	48	96	127	178	50	98	130	182	48	96	127	178	48	96	127	178	55	103	137	195	50	98	129	182	
3	15	48	0			1					0				0						1					0					
3	30	96		0			1				0				0						1					0					
3	45	127			0			0				0				0					0				0				0		
3	60	178				0					0					0					0				0				0		
5	15	48	0			0					0				0						0				0				0		
5	30	96		0			0				0				0						0				0				0		
5	45	127			1			0				0				0					0				0				0		
5	60	178				1			0			0				0					0				0				0		
8	15	50	1				1				0				1						1				0			1			
8	30	98		1				1				0			1						1				0			1			
8	45	130			1				1				0			1					1				0			1			
8	60	182				1				1				0			1				1				0				1		
12	15	48	1				1					0				0					1				0			0			
12	30	96		1				1					0			0					1				0			0			
12	45	127			1				1					0							1				0			0			
12	60	178				1				1					0						1				0				0		
15	15	48	0				1				0				0						0				0			0			
15	30	96		0				1				0				0					0				0			0			
15	45	127			1				1				0			0					0				0			0			
15	60	178				1				1				0			0				0				0			0			
18	15	55	1				1				1				1						1				0			1			
18	30	103		1				1				1				1					1				0			1			
18	45	137			1				1				1				1				1				0			1			
18	60	195				1				1				1				1			1				0				1		
1	15	50	1				1				0				1						1				0			0			
M	30	98		1				1				0			1						1				0			0			
M	45	129			1				1				0			1					1				0			0			
M	60	182				1				1				0			1				1				0			0			
TOTAL:			4	4	6	6	6	6	5	5	1	1	1	1	3	3	3	3	3	5	5	4	4	0	0	0	0	2	2	2	2

Os pesos de ponderação foram utilizados como alternativa a índices como o coeficiente de correlação ou à soma dos módulos dos resíduos porque, ao mesmo tempo que é importante ter boa semelhança entre a série alvo MAN e a saída das RNAs, deve-se observar a acurácia na previsão da cota máxima ou mínima anual.

Em vermelho, na Tabela 9-6, destacam-se as pontuações dos melhores modelos C1 considerando os critérios adotados. Em máximas, C1N3 foi escolhido. Em mínimas, houve desempenho semelhante entre C1N5 e C1N8.

Tabela 9-6. Acima: Soma, ao longo dos testes, das matrizes de cruzamento dos modelos C1 nos dias 15, 30, 45 e 60 à frente do truncamento das séries para as diversas configurações da camada intermediária N = (3, 5, 8, 12, 15, 18); Abaixo: pontuações dos modelos C1 ponderadas por pesos P15, P30, P45 e P60 obtidos pela relação entre a soma dos quadrados dos resíduos nos respectivos dias pela soma total dos quadrados dos resíduos.

MÁXIMAS ANUAIS							
dia	N=3	N=5	N=8	N=12	N=15	N=18	Média
15	43	38	38	36	32	22	28
30	39	33	37	42	35	24	28
45	41	31	27	32	41	32	34
60	38	28	28	38	35	33	32
PESOS	N=3	N=5	N=8	N=12	N=15	N=18	Média
P15 = 0,23	9,99	8,82	8,82	8,36	7,43	5,11	6,50
P30 = 0,20	7,80	6,60	7,40	8,40	7,00	4,80	5,60
P45 = 0,20	8,24	6,23	5,43	6,43	8,24	6,43	6,83
P60 = 0,23	8,57	6,31	6,31	8,57	7,89	7,44	7,22
TOTAL:	34,60	27,97	27,97	31,76	30,57	23,78	26,15
MÍNIMAS ANUAIS							
dia	N=3	N=5	N=8	N=12	N=15	N=18	Média
15	27	33	35	36	39	36	31
30	40	25	39	34	36	31	33
45	32	36	40	32	30	31	37
60	27	43	32	36	37	29	34
PESOS	N=3	N=5	N=8	N=12	N=15	N=18	Média
P15 = 0,12	3,10	3,79	4,02	4,13	4,47	4,13	3,56
P30 = 0,21	8,46	5,29	8,25	7,19	7,62	6,56	6,98
P45 = 0,24	7,82	8,79	9,77	7,82	7,33	7,57	9,04
P60 = 0,29	7,75	12,34	9,18	10,33	10,62	8,32	9,76
TOTAL:	27,12	30,21	31,22	29,47	30,04	26,58	29,33

Escolha dos melhores modelos híbridos PLS, PCA e C2

Os cenários híbridos ótimos foram selecionados de forma semelhante à adotada na seleção dos modelos convencionais. Porém, cada um dos modelos híbridos foi comparado aos melhores modelos C1 escolhidos anteriormente. Como houve praticamente um empate entre C1N5 e C1N8 na previsão de mínimas anuais, esses dois modelos convencionais foram usados na avaliação dos modelos híbridos a seguir.

Assim, os modelos PLSN5, PCAN3 e C2N3 apresentaram melhores desempenhos em comparação aos modelos C1 em previsões de máximas e PLSN18, PCAN15 e C2N18 foram mais bem-sucedidos nas previsões de mínimas. Houve diferenças de desempenho entre arquiteturas ótimas dos modelos PLS, PCA e C2, apontadas nas tabelas Tabela 9-7, Tabela 9-8 e Tabela 9-9, entre previsões de cotas máximas e mínimas anuais.

Tabela 9-7. Resumo do desempenho dos modelos PLSxC1, comparando módulos dos resíduos médios (M) e módulo da mediana dos resíduos (Me) a 15, 30, 45 e 60 dias à frente do truncamento das séries, com pontos representando vezes em que determinado modelo PLS superou o modelo C1 indicado; P15, P30, P45 e P60 obtidos pela relação entre a soma dos quadrados dos resíduos nos respectivos dias pela soma total dos quadrados dos resíduos.

COTAS MÁXIMAS PLSxC1N3																
DIA	PLSN3		PLSN4		PLSN5		PLSN6		PLSN8		PLSN12		PLSN15		PLSN18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
15	6	6	7	6	8	8	8	8	7	8	6	6	6	6	7	7
30	7	7	7	7	7	7	8	7	6	7	5	5	6	5	6	4
45	6	7	6	6	7	6	6	6	4	6	4	4	2	2	3	3
60	6	6	5	5	7	5	6	7	5	7	3	4	2	2	2	2
PESOS	PLSN3		PLSN4		PLSN5		PLSN6		PLSN8		PLSN12		PLSN15		PLSN18	
P15=0,081	0,49	0,49	0,57	0,49	0,65	0,65	0,65	0,65	0,57	0,65	0,49	0,49	0,49	0,49	0,57	0,57
P30=0,154	1,08	1,08	1,08	1,08	1,08	1,08	1,24	1,08	0,93	1,08	0,77	0,77	0,93	0,77	0,93	0,62
P45=0,290	1,75	2,04	1,75	1,75	2,04	1,75	1,75	1,75	1,16	1,75	1,16	1,16	0,58	0,58	0,87	0,87
P60=0,473	2,84	2,84	2,37	2,37	3,31	2,37	2,84	3,31	2,37	3,31	1,42	1,89	0,95	0,95	0,95	0,95
TOTAL:	6,15	6,45	5,76	5,68	7,08	5,84	6,47	6,79	5,03	6,79	3,84	4,32	2,94	2,79	3,31	3,01
COTAS MÍNIMAS PLSxC1N5																
DIA	PLSN3		PLSN4		PLSN5		PLSN6		PLSN8		PLSN12		PLSN15		PLSN18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
15	5	5	6	5	6	6	7	7	7	7	7	7	7	7	7	7
30	5	4	6	5	6	5	6	6	6	6	7	7	7	7	7	8
45	3	4	5	5	4	5	3	3	4	4	7	7	5	5	5	5
60	4	3	5	5	4	4	4	4	3	3	4	4	4	4	5	5
PESOS	PLSN3		PLSN4		PLSN5		PLSN6		PLSN8		PLSN12		PLSN15		PLSN18	
P15=0,096	0,48	0,48	0,58	0,48	0,58	0,58	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67
P30=0,233	1,17	0,93	1,40	1,17	1,40	1,17	1,40	1,40	1,40	1,40	1,63	1,63	1,63	1,63	1,63	1,87
P45=0,290	0,87	1,16	1,45	1,45	1,16	1,45	0,87	0,87	1,16	1,16	2,03	2,03	1,45	1,45	1,45	1,45
P60=0,381	1,52	1,14	1,90	1,90	1,52	1,52	1,52	1,52	1,14	1,14	1,52	1,52	1,52	1,52	1,90	1,90
TOTAL:	4,04	3,72	5,33	5,00	4,66	4,72	4,47	4,47	4,37	4,37	5,86	5,86	5,28	5,28	5,66	5,89
COTAS MÍNIMAS PLSxC1N8																
DIA	PLSN3		PLSN4		PLSN5		PLSN6		PLSN8		PLSN12		PLSN15		PLSN18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
15	5	5	7	6	7	7	7	7	7	7	8	8	7	7	7	8
30	5	4	6	5	6	5	6	6	6	6	7	7	7	7	7	8
45	3	4	5	5	4	5	3	3	5	5	7	7	5	5	5	5
60	4	3	5	5	4	4	4	4	3	3	4	4	4	4	5	4
PESOS	PLSN3		PLSN4		PLSN5		PLSN6		PLSN8		PLSN12		PLSN15		PLSN18	
P15=0,096	0,48	0,48	0,67	0,58	0,67	0,67	0,67	0,67	0,67	0,67	0,77	0,77	0,67	0,67	0,67	0,77
P30=0,233	1,17	0,93	1,40	1,17	1,40	1,17	1,40	1,40	1,40	1,40	1,63	1,63	1,63	1,63	1,63	1,87
P45=0,290	0,87	1,16	1,45	1,45	1,16	1,45	0,87	0,87	1,45	1,45	2,03	2,03	1,45	1,45	1,45	1,45
P60=0,381	1,52	1,14	1,90	1,90	1,52	1,52	1,52	1,52	1,14	1,14	1,52	1,52	1,52	1,52	1,90	1,52
TOTAL:	4,04	3,72	5,43	5,10	4,76	4,81	4,47	4,47	4,66	4,66	5,95	5,95	5,28	5,28	5,66	5,61

Tabela 9-8. Resumo do desempenho dos modelos PCAxC1, comparando módulos dos resíduos médios (M) e módulo da mediana dos resíduos (Me) a 15, 30, 45 e 60 dias à frente do truncamento das séries, com pontos representando vezes em que determinado modelo PCA superou o modelo C1 indicado; P15, P30, P45 e P60 obtidos pela relação entre a soma dos quadrados dos resíduos nos respectivos dias pela soma total dos quadrados dos resíduos.

COTAS MÁXIMAS PCAxCIN3												
DIA	PCAN3		PCAN5		PCAN8		PCAN12		PCAN15		PCAN18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
15	6	6	6	7	7	7	5	6	5	5	6	6
30	7	7	7	6	5	5	6	6	6	6	6	6
45	6	6	8	5	4	5	4	4	3	3	3	3
60	6	8	6	6	5	6	3	3	3	3	2	2
PESOS	PCAN3		PCAN5		PCAN8		PCAN12		PCAN15		PCAN18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
P15=0,084	0,51	0,51	0,51	0,59	0,59	0,59	0,42	0,51	0,42	0,42	0,51	0,51
P30=0,150	1,05	1,05	1,05	0,90	0,75	0,75	0,90	0,90	0,90	0,90	0,90	0,90
P45=0,288	1,73	1,73	2,31	1,44	1,15	1,44	1,15	1,15	0,87	0,87	0,87	0,87
P60=0,477	2,86	3,82	2,86	2,86	2,39	2,86	1,43	1,43	1,43	1,43	0,95	0,95
TOTAL:	6,15	7,10	6,73	5,80	4,88	5,65	3,91	3,99	3,62	3,62	3,23	3,23
COTAS MÍNIMAS PCAxCIN5												
DIA	PCAFN3		PCAFN5		PCAFN8		PCAFN12		PCAFN15		PCAFN18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
15	5	5	7	6	6	6	7	7	7	7	7	7
30	4	6	7	6	6	6	7	7	8	8	8	8
45	4	4	5	5	5	5	8	8	7	7	6	6
60	4	4	4	4	4	4	4	4	5	6	5	6
PESOS	PCAFN3		PCAFN5		PCAFN8		PCAFN12		PCAFN15		PCAFN18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
P15=0,0895	0,47	0,47	0,66	0,57	0,57	0,57	0,66	0,66	0,66	0,66	0,66	0,66
P30=0,229	0,91	1,37	1,60	1,37	1,37	1,37	1,60	1,60	1,83	1,83	1,83	1,83
P45=0,294	1,18	1,18	1,47	1,47	1,47	1,47	2,35	2,35	2,06	2,06	1,76	1,76
P60=0,383	1,53	1,53	1,53	1,53	1,53	1,53	1,53	1,53	1,91	2,29	1,91	2,29
TOTAL:	4,09	4,55	5,27	4,94	4,94	4,94	6,15	6,15	6,46	6,85	6,17	6,55
COTAS MÍNIMAS PCAxCIN8												
DIA	PCAFN3		PCAFN5		PCAFN8		PCAFN12		PCAFN15		PCAFN18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
15	5	6	7	7	7	7	8	8	7	8	7	7
30	4	5	7	6	6	6	7	7	8	8	8	8
45	4	4	5	5	5	5	8	8	7	8	6	6
60	4	4	4	4	4	4	4	4	5	6	5	6
PESOS	PCAFN3		PCAFN5		PCAFN8		PCAFN12		PCAFN15		PCAFN18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
P15=0,0895	0,47	0,57	0,66	0,66	0,66	0,66	0,76	0,76	0,66	0,76	0,66	0,66
P30=0,229	0,91	1,14	1,60	1,37	1,37	1,37	1,60	1,60	1,83	1,83	1,83	1,83
P45=0,294	1,18	1,18	1,47	1,47	1,47	1,47	2,35	2,35	2,06	2,35	1,76	1,76
P60=0,383	1,53	1,53	1,53	1,53	1,53	1,53	1,53	1,53	1,91	2,29	1,91	2,29
TOTAL:	4,09	4,42	5,27	5,04	5,04	5,04	6,24	6,24	6,46	7,24	6,17	6,55

Após a seleção das configurações ótimas dos modelos híbridos, estudaram-se as distribuições dos resíduos de tais modelos. A Figura 9-6 Figura 9-7 ilustram a estatística não paramétrica dos resíduos ao longo dos testes T1 a T10. Os retângulos azuis delimitam, aos 15, 30, 45 e 60 dias de simulação, os quartis centrais da distribuição dos resíduos nas 1.000 simulações com cada modelo.

Tabela 9-9. Resumo do desempenho dos modelos C2xC1, comparando módulos dos resíduos médios (M) e módulo da mediana dos resíduos (Me) a 15, 30, 45 e 60 dias à frente do truncamento das séries, com pontos representando vezes em que determinado modelo C2 superou o modelo C1 indicado; P15, P30, P45 e P60 obtidos pela relação entre a soma dos quadrados dos resíduos nos respectivos dias pela soma total dos quadrados dos resíduos

COTAS MÁXIMAS C2xC1N3												
DIA	C2N3		C2N5		C2N8		C2N12		C2N15		C2N18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
15	6	4	8	8	5	5	5	5	6	6	7	7
30	7	6	6	6	4	4	5	5	6	6	6	7
45	6	6	5	3	4	4	3	3	4	4	5	5
60	6	6	6	4	4	3	3	3	3	3	3	3
PESOS	C2N3		C2N5		C2N8		C2N12		C2N15		C2N18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
P15=0,072	0,433	0,289	0,577	0,577	0,361	0,361	0,361	0,361	0,433	0,433	0,505	0,505
P30=0,143	1,003	0,860	0,860	0,860	0,573	0,573	0,717	0,717	0,860	0,860	0,860	1,003
P45=0,290	1,740	1,740	1,450	0,870	1,160	1,160	0,870	0,870	1,160	1,160	1,450	1,450
P60=0,494	2,967	2,967	2,967	1,978	1,978	1,483	1,483	1,483	1,483	1,483	1,483	1,483
TOTAL:	6,143	5,856	5,854	4,285	4,072	3,578	3,431	3,431	3,937	3,937	4,299	4,442
COTAS MÍNIMAS C2xC1N5												
DIA	C2N3		C2N5		C2N8		C2N12		C2N15		C2N18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
15	4	3	6	6	8	8	7	7	7	7	6	6
30	7	5	6	6	5	5	6	6	6	6	6	7
45	4	4	4	4	5	4	4	3	6	5	5	5
60	3	4	3	3	4	4	4	4	5	5	7	7
PESOS	C2N3		C2N5		C2N8		C2N12		C2N15		C2N18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
P15=0,094	0,375	0,281	0,562	0,562	0,750	0,750	0,656	0,656	0,656	0,656	0,562	0,562
P30=0,222	1,555	1,111	1,333	1,333	1,111	1,111	1,333	1,333	1,333	1,333	1,333	1,555
P45=0,288	1,150	1,150	1,150	1,150	1,438	1,150	1,150	0,863	1,725	1,438	1,438	1,438
P60=0,397	1,190	1,586	1,190	1,190	1,586	1,586	1,586	1,586	1,983	1,983	2,776	2,776
TOTAL:	4,270	4,128	4,235	4,235	4,884	4,597	4,725	4,438	5,697	5,410	6,109	6,331
COTAS MÍNIMAS C2xC1N8												
DIA	C2N3		C2N5		C2N8		C2N12		C2N15		C2N18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
15	5	4	7	7	8	8	8	8	7	7	6	6
30	7	5	6	6	5	5	6	6	6	6	6	7
45	4	4	4	4	5	4	5	4	7	7	5	5
60	3	4	3	3	4	4	4	4	5	5	7	7
PESOS	C2N3		C2N5		C2N8		C2N12		C2N15		C2N18	
	M	Me	M	Me	M	Me	M	Me	M	Me	M	Me
P15=0,094	0,468	0,375	0,656	0,656	0,750	0,750	0,750	0,750	0,656	0,656	0,562	0,562
P30=0,222	1,555	1,111	1,333	1,333	1,111	1,111	1,333	1,333	1,333	1,333	1,333	1,555
P45=0,288	1,150	1,150	1,150	1,150	1,438	1,150	1,438	1,150	2,013	2,013	1,438	1,438
P60=0,397	1,190	1,586	1,190	1,190	1,586	1,586	1,586	1,586	1,983	1,983	2,776	2,776
TOTAL:	4,364	4,222	4,329	4,329	4,884	4,597	5,107	4,819	5,985	5,985	6,109	6,331

As linhas vermelhas são as medianas dos resíduos. As linhas azuis são as médias dos resíduos dos modelos híbridos e as cinzas são as médias dos resíduos dos modelos convencionais de melhor desempenho. As retas horizontais roxas tracejadas são estimativas das delimitações dos quartis centrais dos desvios feitas a partir do conjunto de testes das RNAs. A discrepância entre os resíduos estimados pelos testes e os verificados é considerável.

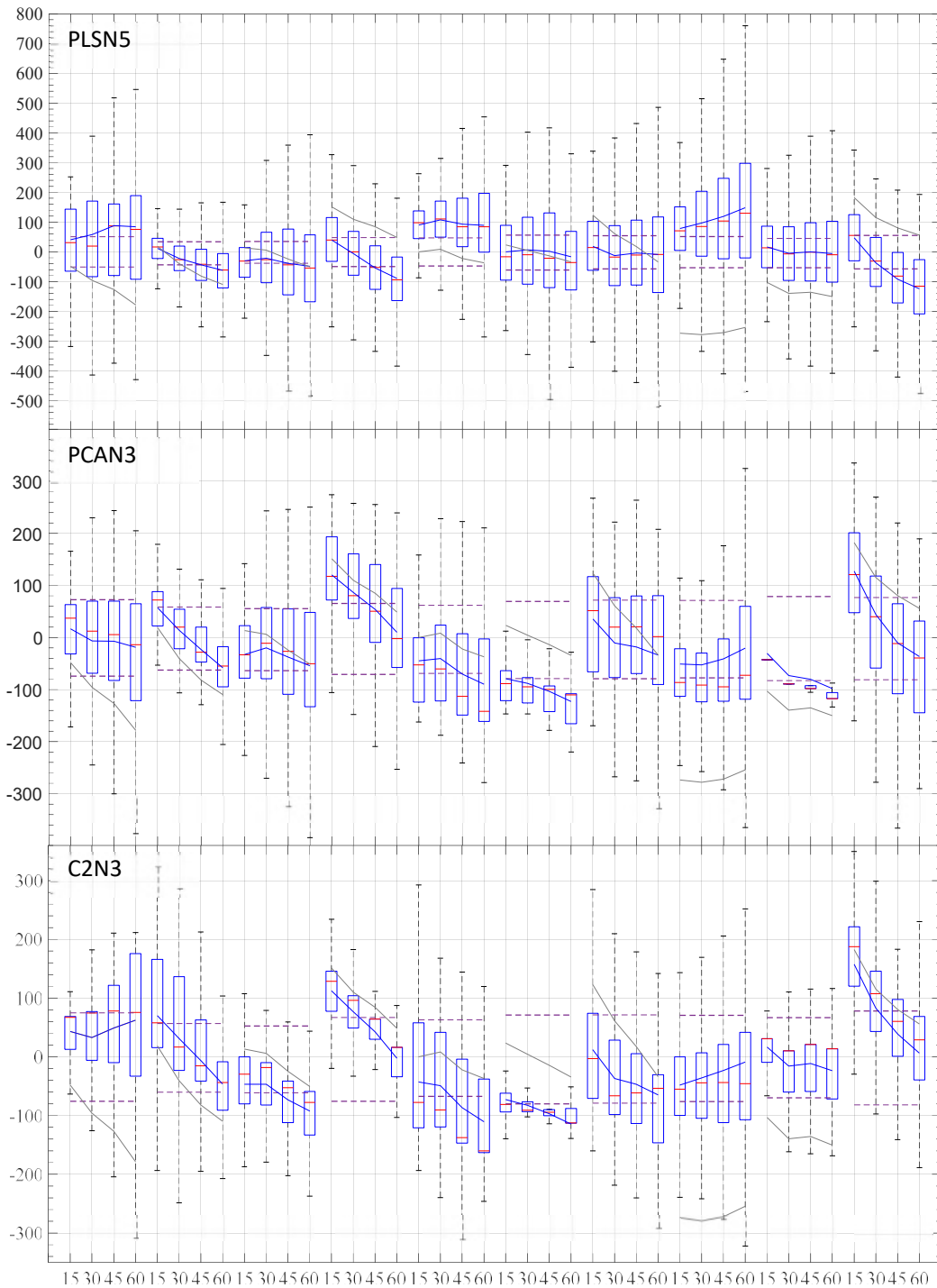


Figura 9-6. Distribuição de resíduos dos modelos PLSN5, PCAN3 e C2N3 ao longo de 1.000 simulações nas dez previsões de máximas anuais: eixo horizontal em dias após truncamento das séries e eixo vertical em cm; retângulos azuis delimitam quartis centrais; linhas vermelhas são as medianas; linhas roxas tracejadas delimitam estimativas dos quartis centrais dos resíduos nos testes das RNAs; linhas azuis são as médias dos resíduos dos modelos híbridos e linhas cinzas são médias dos resíduos do modelo convencional C1N3.

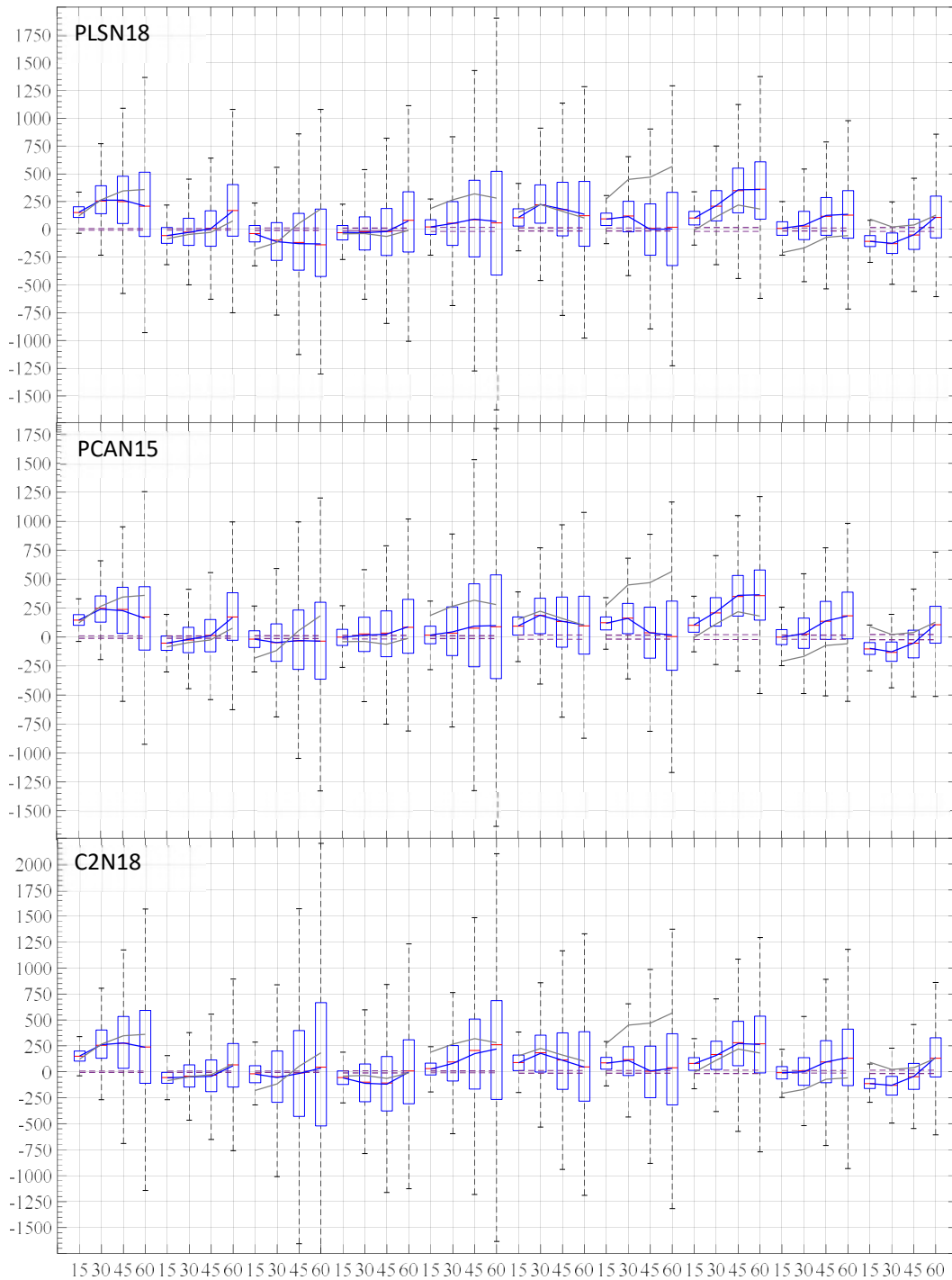


Figura 9-7. Distribuição de resíduos dos modelos PLSN18, PCAN15 e C2N18 ao longo de 1.000 simulações nas dez previsões de mínimas anuais: eixo horizontal em dias após truncamento das séries e eixo vertical em cm; retângulos azuis delimitam quartis centrais; linhas vermelhas são as medianas; linhas roxas tracejadas delimitam estimativas dos quartis centrais dos resíduos nos testes das RNAs; linhas azuis são as médias dos resíduos nos modelos híbridos e linhas cinzas são médias dos resíduos do modelo convencional C1N3.

Tabela 9-10. Resumo das médias dos módulos dos resíduos dos modelos testados com N = (4, 5, 8, 12, 15, 18) nós na camada intermediária em previsões de cotas máximas e mínimas aos 15, 30 45 e 60 dias após truncamento das séries.

COTAS MÁXIMAS							COTAS MÍNIMAS						
C1							C1						
	N=3	N=5	N=8	N=12	N=15	N=18		N=3	N=5	N=8	N=12	N=15	N=18
15d	93,692	95,241	94,622	90,651	92,680	92,771	15d	133,869	131,393	134,579	133,128	144,776	142,824
30d	85,886	89,139	82,383	75,568	76,626	76,923	30d	174,267	173,508	171,004	170,417	183,361	178,851
45d	85,914	90,068	80,721	87,551	88,535	89,401	45d	174,033	174,009	177,034	176,864	179,164	176,545
60d	95,255	99,116	95,424	101,381	102,893	103,788	60d	191,068	191,521	194,604	192,104	186,784	188,159
PLS							PLS						
	N=3	N=5	N=8	N=12	N=15	N=18		N=3	N=5	N=8	N=12	N=15	N=18
15d	64,588	37,279	36,217	35,264	34,819	28,687	15d	127,385	114,389	84,284	80,774	74,053	71,821
30d	48,213	37,409	71,174	71,491	78,359	75,474	30d	178,712	162,940	141,094	130,553	121,644	120,066
45d	47,292	53,501	88,851	107,951	129,098	121,569	45d	222,418	190,801	162,075	136,285	123,251	122,496
60d	56,847	67,773	104,816	141,988	173,328	164,321	60d	266,981	258,542	220,959	176,209	149,013	140,143
PCA							PCA						
	N=3	N=5	N=8	N=12	N=15	N=18		N=3	N=5	N=8	N=12	N=15	N=18
15d	59,596	40,544	50,398	47,714	40,892	24,749	15d	131,563	99,639	81,215	73,160	65,511	66,196
30d	43,614	42,833	78,935	78,049	71,622	62,786	30d	178,893	148,741	131,611	119,975	109,503	109,751
45d	44,613	56,655	98,687	111,696	117,968	107,932	45d	207,776	171,771	158,577	127,334	111,370	116,507
60d	54,336	70,003	114,695	141,706	153,898	147,095	60d	252,137	238,595	213,406	167,074	133,205	131,325
C2							C2						
	N=3	N=5	N=8	N=12	N=15	N=18		N=3	N=5	N=8	N=12	N=15	N=18
15d	62,239	31,822	53,486	36,609	26,515	17,623	15d	129,746	97,262	72,932	73,142	70,942	68,356
30d	49,029	41,277	84,311	68,818	60,592	47,684	30d	170,027	146,282	125,716	111,665	112,046	113,125
45d	47,464	57,133	104,847	99,886	103,808	93,448	45d	209,733	193,921	146,561	133,294	116,283	116,128
60d	53,134	76,807	123,298	132,747	148,878	138,884	60d	258,542	254,149	187,492	178,877	144,996	116,989

A análise dos módulos dos resíduos médios de todas as simulações para cada configuração de modelo é exibida na Tabela 9-10. Em vermelho estão os melhores desempenhos de cada modelo. Já a Tabela 9-11 resume os ganhos percentuais obtidos com cada modelo híbrido em relação aos melhores modelos convencionais. Foram observadas reduções superiores a 40%, em previsões de máximas, e a 35%, em previsões de mínimas, nos resíduos médios com a modelagem híbrida.

Tabela 9-11. Resumo das reduções percentuais médias de módulos dos desvios dos modelos híbridos em relação aos modelos convencionais aos 15, 30, 45 e 60 dias à frente do truncamento das séries, diferenciando-se previsões de cotas máximas das previsões de cotas mínimas.

COTAS MÁXIMAS							COTAS MÍNIMAS						
PLSN3xC1							PLSN18xC1						
	N=3	N=5	N=8	N=12	N=15	N=18		N=3	N=5	N=8	N=12	N=15	N=18
15d	31%	32%	32%	29%	30%	30%	15d	46%	45%	47%	46%	50%	50%
30d	44%	46%	41%	36%	37%	37%	30d	31%	31%	30%	30%	35%	33%
45d	45%	47%	41%	46%	47%	47%	45d	30%	30%	31%	31%	32%	31%
60d	40%	43%	40%	44%	45%	45%	60d	27%	27%	28%	27%	25%	26%
PCAN3xC1							PCAN18xC1						
	N=3	N=5	N=8	N=12	N=15	N=18		N=3	N=5	N=8	N=12	N=15	N=18
15d	36%	37%	37%	34%	36%	36%	15d	51%	50%	51%	50%	54%	54%
30d	49%	51%	47%	42%	43%	43%	30d	37%	37%	36%	36%	40%	39%
45d	48%	50%	45%	49%	50%	50%	45d	33%	33%	34%	34%	35%	34%
60d	43%	45%	43%	46%	47%	48%	60d	31%	31%	33%	32%	30%	30%
C2N3xC1							C2N18xC1						
	N=3	N=5	N=8	N=12	N=15	N=18		N=3	N=5	N=8	N=12	N=15	N=18
15d	34%	35%	34%	31%	33%	33%	15d	49%	48%	49%	49%	53%	52%
30d	43%	45%	40%	35%	36%	36%	30d	35%	35%	34%	34%	38%	37%
45d	45%	47%	41%	46%	46%	47%	45d	33%	33%	34%	34%	35%	34%
60d	44%	46%	44%	48%	48%	49%	60d	39%	39%	40%	39%	37%	38%

Por fim, na figura MMMM ilustram-se alguns cotagramas nos domínios de previsões dos testes realizados. Os marcadores indicam os valores aos 1, 15, 30, 45 e 60 dias de simulação.

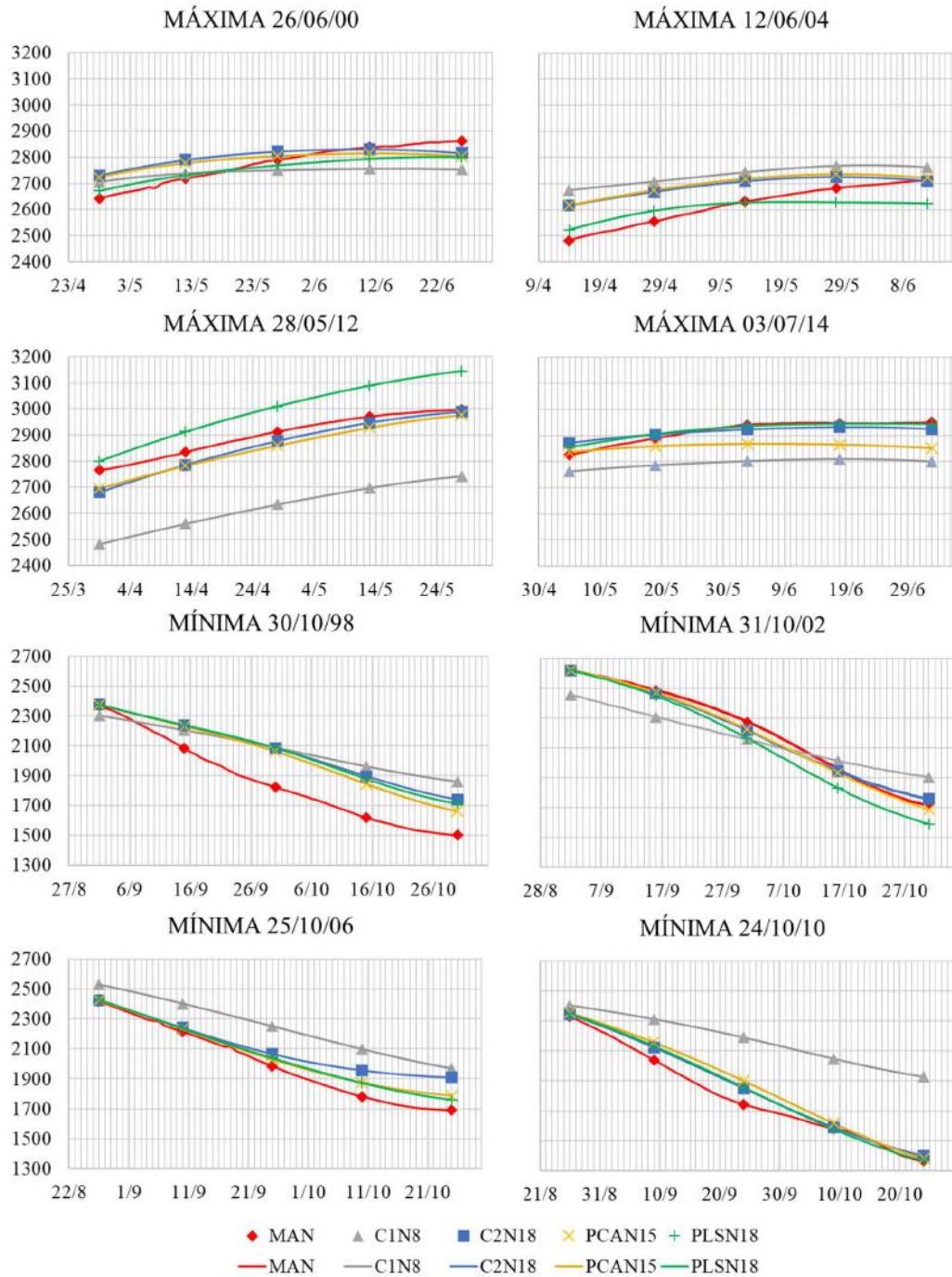


Figura 9-8. Cotagramas ao longo dos 60 dias de previsão de alguns testes, destacando-se os dias 15, 30, 45 e 60 das simulações – eixos verticais em cm.

9.4 Condicionamento e Generalização dos modelos

Aprofundando-se a investigação dos efeitos da EMD, da PLS e da PCA no pré-processamento das RNAs, apresentam-se as matrizes de auto-correlação dos Jacobiano dos modelos dos modelos C1N3, C2N3, PLSN3 e PCAN3 na Figura 9-9. O caso ilustrado corresponde à previsão de cota mínima de 31/10/02. Os resultados correspondentes às RNAs com três nós na camada intermediária foram arbitrariamente escolhidos para exibição.

A Tabela 9-12 resume o número de parâmetros e os números de condicionamento dos Jacobianos e dos seus transpostos para diversas configurações de arquitetura das RNAs, também para o teste de previsão de mínima anual de 31/10/02.

Tabela 9-12. Modelos simulados na previsão da mínima de 31/10/02, números de parâmetros para as diferentes configurações de arquitetura das RNAs *feedforward* correspondentes e condicionamento dos Jacobianos $C(J)$ e de seus transpostos $C(J^*)$.

MODELO	PARÂMETROS	C(J)	C(J*)
C1N3	16	1,42 E03	1,42 E03
C2N3	37	345	345
PLSN3	37	209	209
PCAN3	37	182	182
C1N5	26	2,85 e21	1,67 e22
C2N5	61	745	745
PLSN5	61	469	469
PCAN5	61	487	487
C1N8	41	5,48 E23	8,72 E23
C2N8	97	1,43 E3	1,43 E3
PLSN8	97	1,06 E3	1,06 E3
PCAN8	97	1,79 E3	1,79 E3
C1N12	61	3,48 E24	6,68 e24
C2N12	145	1,67 E03	1,67 E03
PLSN12	145	1,57 E03	1,57 E03
PCAN12	145	1,57 E03	1,57 E03
C1N18	91	1,62 E25	1,44 E25
C2N18	217	4,26 E03	4,26 E03
PLSN18	217	3,08 E03	3,08 E03
PCAN18	217	3,28 E03	3,28 E03

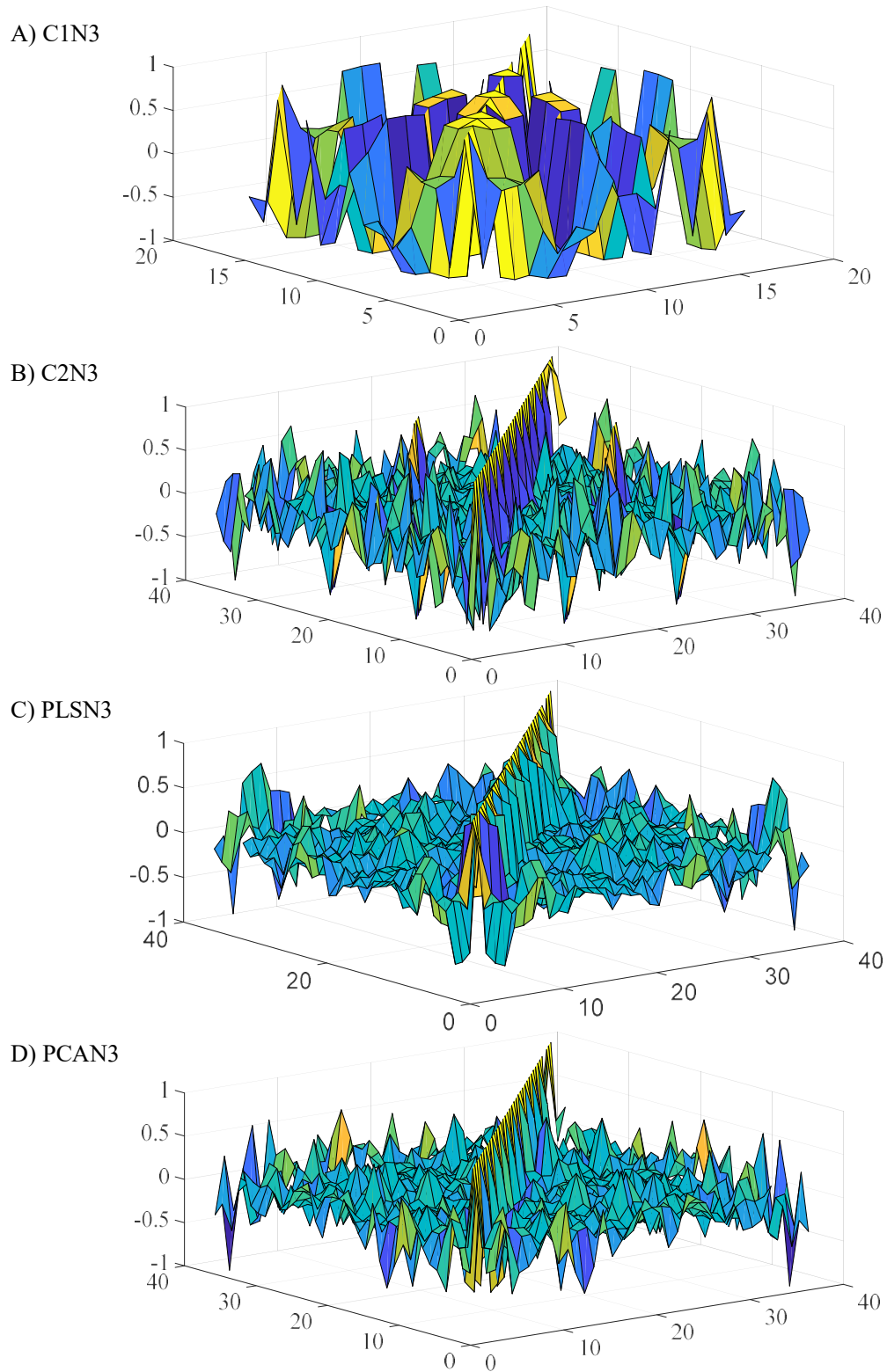


Figura 9-9. Matrizes de correlação dos Jacobianos dos modelos N=3 na previsão da mínima de 31/10/2002.

Para melhor explorar o comportamento dos modelos híbridos, considerando a discrepância entre os números de condicionamento menores obtidos e as estimativas ruins de desvios feitas pelos conjuntos de testes, sugerem-se simulações complementares usando os modelos C1N8 e C2N8, na previsão da mínima anual de 31/10/02 com uma técnica diferente de amostragem para validação.

A validação cruzada *k-fold* estratificada, com $k = 2$, foi utilizada com a parada antecipada para generalizar os modelos. Os critérios de seleção das dobras estratificadas foram apresentados em detalhes no item 7.7 da metodologia e destinaram-se a conferir maior independência entre os conjuntos de treinamento e de validação, como ilustrado na Figura 7-2. Com isso, pretende-se inspecionar melhor os desempenhos dos modelos e identificar possíveis sinais de *overfitting* nos modelos híbridos que justifiquem as elevadas variâncias nos seus desvios absolutos médios, mesmo com números de condicionamento consideravelmente menores em relação aos modelos C1.

Os respectivos desempenhos do modelo convencional (esquerda) e do modelo híbrido (direita) são ilustrados na Figura 9-10 até as paradas antecipadas, seis épocas após as inflexões das curvas de desvios médios quadráticos dos conjuntos de validação.

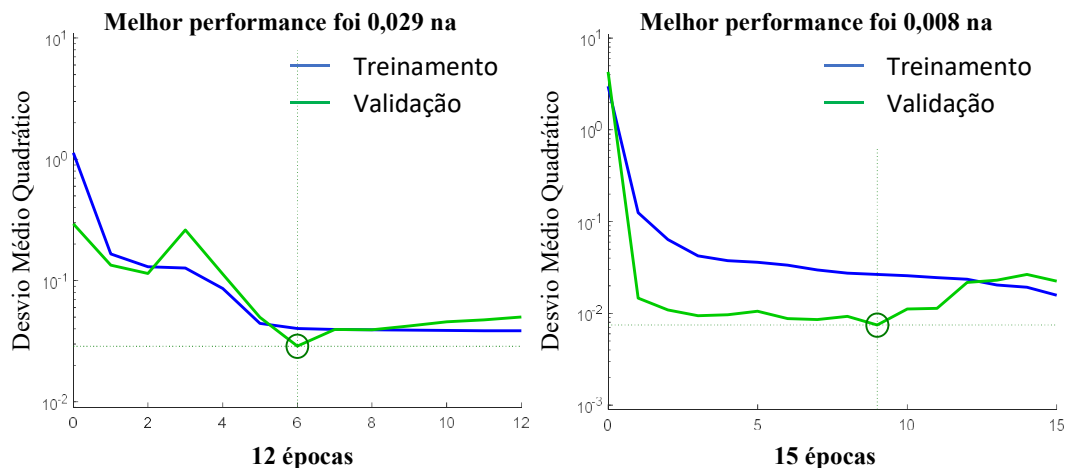


Figura 9-10. Desempenho do modelo C1N8 (esquerda) e de C2N8 (direita) até o momento da parada antecipada, seis épocas após a inflexão das curvas de desvios médios quadráticos dos conjuntos de validação.

Para quantificar o condicionamento nos testes propostos, utilizaram-se as matrizes de auto-correlação dos Jacobianos novamente, ilustradas na Figura 9-11.

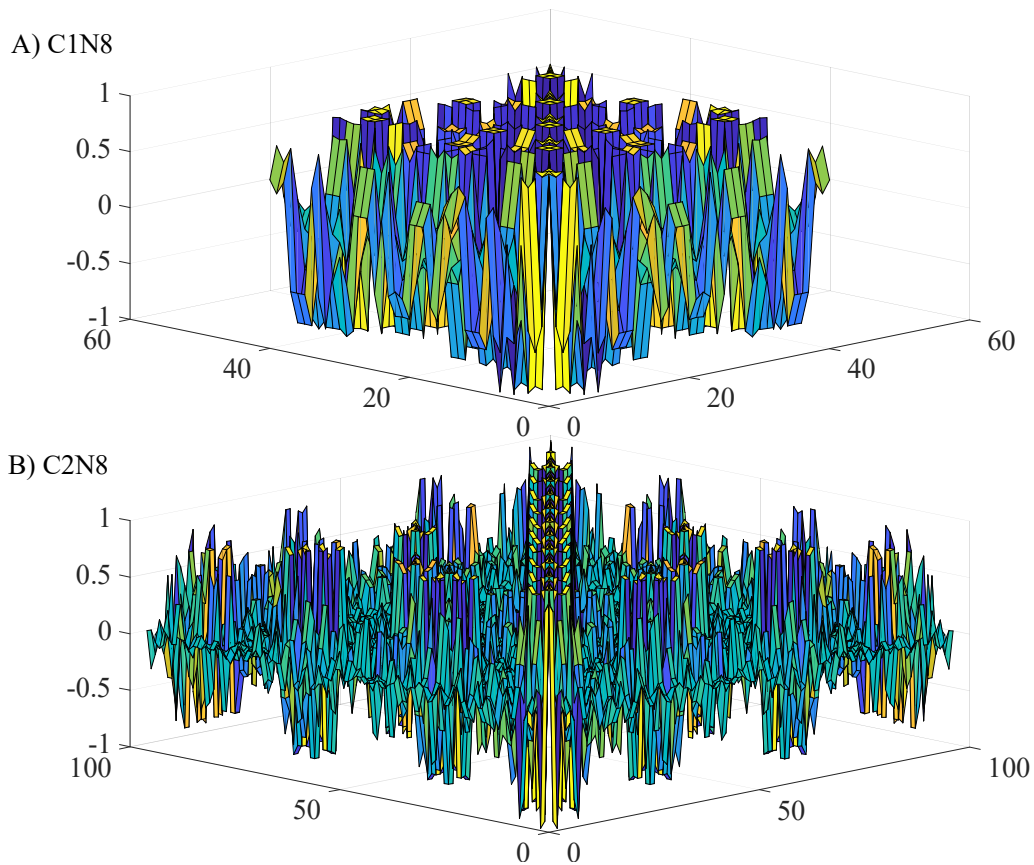


Figura 9-11. Matrizes de correlação dos Jacobianos dos modelos C1N8 e C2N8 com a validação cruzada *k-fold*.

Por fim, os números de parâmetros totais/efetivos e os respectivos números de condicionamento dos Jacobianos e de seus transpostos para os modelos C1N8 e C2N8, achados em simulações com a validação cruzada *k-fold* são resumidos na Tabela 9-13.

Tabela 9-13. Número de parâmetros totais/efetivos e condicionamentos dos jacobianos e de seus transpostos para os modelos estudados com a parada antecipada *k-fold*.

	PARÂMETROS	$C(J)$	$C(J^*)$
C1N8	41/4	3,87 E13	3,87 E13
C2N8	97/42	6,10 E09	6,10 E09

9.5 Resultados Comparativos entre Modelo Híbrido e Modelo de Galdino & Clarke

O modelo probabilístico de Galdino & Clarke – MP, adaptado por PIMENTEL *et al.* (2011) para previsão de níveis em Manaus foi utilizado como base complementar para avaliação do desempenho dos modelos híbridos em previsões de cotas máximas anuais.

O MP faz previsões a partir de datas de referência usando valores das cotas de tais datas ao longo do histórico para o cálculo das probabilidades. Para o presente estudo comparativo, as datas de previsão do MP escolhidas foram 01/04, 16/04 e 16/05 de cada ano. Considerando que o dia mais provável de ocorrência da máxima anual é 15/06, em média, as previsões antecipariam os picos em 75, 60 e 30 dias. Porém, em muitos anos do histórico, a máxima anual ocorreu antes de 15/06. Nesses casos, a previsão do MP torna-se mais fácil.

O desvio médio obtido pelo MP na previsão de máximas de 1919 a 2016 foi de 54,44 cm. Nos mesmos anos para os quais o modelo híbrido foi testado, anos pares de 1998 a 2016, o desvio absoluto médio de MP foi de 53,93 cm, contra 53,13 cm, para C2N3, modelo híbrido de melhor desempenho em termos de desvio absoluto médio.

10. Discussão

O comparativo dos módulos dos resíduos médios (M) e dos módulo da mediana dos resíduos (Me) a 15, 30, 45 e 60 dias à frente do truncamento das séries mostrou que os modelos PLSN5, PCAN3 e C2N3 apresentaram melhores desempenhos em comparação aos modelos C1 em previsões de máximas e PLSN18, PCAN15 e C2N18 foram mais bem-sucedidos nas previsões de mínimas (Tabelas Tabela 9-7, Tabela 9-8 e Tabela 9-9).

Em previsão de máximas, o modelo PLSN5 superou C1 em oito ocasiões, aos 15 dias, e em sete ocasiões, aos 60 dias. Em previsões de mínimas, o modelo PLSN5 superou C1 em sete ocasiões (considerando médias dos resíduos), ou oito ocasiões (considerando medianas dos resíduos), aos 15 dias, e em cinco ocasiões, aos 60 dias. Desempenhos comparativos semelhantes foram verificados para os modelos PCA. Para C2, no entanto, observaram-se melhores desempenhos em relação a C1 em seis ocasiões, aos 15 dias e em seis ocasiões (considerando médias dos resíduos) ou sete ocasiões (considerando medianas dos resíduos).

Assim, evidenciaram-se os efeitos da projeção dos preditores com a PLS e com a PCA. Tais métodos têm caráter linear na determinação das direções de maior variabilidade dos alvos e dos preditores. Tal caráter linear pode ter sido responsável por garantir desempenhos melhores próximo ao truncamento das séries aos 15 dias, com piora dos resultados mais acentuada aos 60 dias. Já os modelos C2 apresentaram desempenhos menores que os modelos PLS e PCA nas proximidades do truncamento das séries, mas mantiveram melhor a qualidade da solução até os 60 dias.

Tais números também evidenciam as distinções físicas entre previsões de cotas máximas e mínimas anuais, apontando melhores desempenhos nas previsões de máximas e justificando o tratamento diferenciado de tais eventos durante a análise.

Durante o período de cotas baixas, a hidrodinâmica naturalmente sofre maior influência da morfologia da calha do rio. Além disso, com o escoamento confinado em uma calha de menor área, as variações de vazão se traduzem mais rapidamente em variações de níveis. Outras acelerações como efeitos de remanso também podem se fazer sentir mais.

Por isso as previsões em cotas baixas carregam mais incertezas associadas ao escoamento e ao corpo d'água.

A Tabela 9-11 apontou reduções superiores a 40%, em previsões de máximas, e a 35%, em previsões de mínimas nos resíduos médios com a modelagem híbrida em relação aos modelos convencionais.

Já os cotagramas dos domínios de previsão ilustraram, além dos melhores ajustes com os modelos híbridos, maiores correlações entre suas previsões e as cotas medidas avante do truncamento das séries. A EMD aumentou a verossimilhança entre as previsões e o fenômeno simulado.

A comparação entre os desempenhos dos modelos híbridos e do modelo probabilístico de Galdino & Clarke – MP, deve levar em consideração as diferenças na montagem dos cenários de simulação dessas ferramentas. O MP faz previsões das máximas a partir de datas de referência. Nos anos nos quais a cheia ocorre antes de 15/06, portanto, há menor antecipação da previsão, que tende a se tornar mais fácil.

Por outro lado, os modelos híbridos foram aplicados truncando-se as séries 60 dias antes dos picos anuais. Considerando a auto-correlação em MAN, presente graças a regularização da bacia, a qualidade das previsões com a modelagem híbrida aumentaria se as séries fossem truncadas a menos de 60 dias dos picos.

Portanto, considerando a diferença de complexidade entre as previsões feitas pelos modelos híbridos propostos e pelo MP, a equivalência de desvios absolutos entre eles credencia a estratégia adotada de modelagem hidrológica a partir de dados oceanográficos e climatológicos, pois, apesar da equivalência de acurácia na previsão das máximas, há garantia de cerca de sessenta dias de antecipação com os modelos híbridos. A previsão do MP pode ocorrer sem antecipação em relação ao pico, comprometendo sua utilidade operacional.

O estudo da distribuição de resíduos absolutos dos modelos híbridos e dos valores previstos pelos conjuntos de testes das RNAs objetivou proporcionar melhor compreensão dos efeitos da adoção da EMD, de PLS e de PCA na modelagem. Os

modelos híbridos exibiram variâncias elevadas caracterizadas pelas alturas dos retângulos azuis, que delimitaram os quartis centrais dos desvios absolutos. Outra característica marcante foram as discrepâncias entre as estimativas de desvios, feitas a partir dos conjuntos de testes das RNAs dos modelos híbridos (retas tracejadas roxas), e as obtidas nas previsões, apontadas pelos retângulos azuis.

Por outro lado, as boas estimativas de desvios para os conjuntos de testes (conjuntos de 15 elementos intercalados por amostras usadas para treinamento) evidenciam a potencialidade da ferramenta para diversas aplicações: preenchimento de lacunas em séries temporais e ajustes de curvas-chave são alguns exemplos práticos.

A revisão mostrou a multicolinearidade inerente às redes *feedforward* utilizadas na presente pesquisa. Essa característica é causada pelos patamares presentes nas funções de ativação sigmoidais e é reforçada ao utilizarem-se preditores com elevada correlação mútua. A PLS e a PCA foram assim adotadas após a EMD, no pré-processamento de dados, compondo duas famílias extras de modelos híbridos, a fim de melhorar o ranqueamento e o condicionamento dos jacobianos dos modelos híbridos.

A Figura 9-9 mostrou maiores correlações entre coeficientes do jacobiano do modelo convencional C1N3 do que dos modelos híbridos com mesmo número de nós na camada intermediária, nas previsões da mínima de 2002. Isso se traduz em números de condicionamento dos Jacobianos dos modelos C1 expressivamente maiores em relação aos modelos híbridos na Tabela 9-12 para diversas configurações da camada intermediária. Entre os modelos híbridos, houve pouca redução da colinearidade após aplicação da PLS ou da PCA em relação a C2.

Mesmo com esse efeito da EMD de redução expressiva da multicolinearidade e do número de condicionamento, deve haver outra causa para as altas variâncias encontradas em seus resíduos. Como visto, o número de condicionamento do Jacobiano traduz a relação entre o erro relativo e o resíduo relativo. Com $C(J)$ muito maiores nos modelos convencionais, as estimativas feitas a partir do conjunto de testes dos modelos híbridos deveriam ser melhores.

Tal fato remete à discussão do item Estratégias de Generalização, no capítulo 5, no qual benefícios da parada antecipada e da regularização Bayesiana foram comparados, destacando-se as incertezas presentes nos estimadores de desvios das RNAs em função da relação entre o número de amostras n e o número de parâmetros N .

Os trabalhos citados discutiram a importância da escolha do método de amostragem para testes e de estimação dos desvios obtidos. Cada tipo de amostragem penaliza de forma distinta o conjunto de treinamento e generaliza em grau distinto os desvios, mesmo que a parada antecipada não seja utilizada, como na presente pesquisa.

A EMD aumenta o número de parâmetros das RNAs nos modelos híbridos. Como consequência, aumenta a dificuldade de estimação dos desvios pelo conjunto de testes. Além disso, tem-se aumento da variância dos resíduos (retângulos azuis) em relação aos modelos convencionais, cujos retângulos demarcando os quartis centrais dos desvios seriam muito pequenos comparados aos dos modelos híbridos e, portanto, não foram exibidos aqui.

Porém, essa explicação não parece totalmente satisfatória quando se analisaram modelos convencionais com o mesmo número de parâmetros, aumentando-se N em $C1$. Mesmo assim, a variância dos modelos convencionais permaneceu mais controlada e com menores discrepâncias nas estimativas dos erros. Apesar disso, os valores médios dos desvios dos modelos híbridos mostraram-se melhores.

Assim, propuseram-se simulações complementares para tentar entender melhor as variâncias elevadas dos desvios dos modelos híbridos e as inconsistências entre desvios estimados em testes e desvios achados nas simulações. Os resultados preliminares utilizando a parada antecipada com validação *k-fold* estratificada revelaram no próprio conjunto de validação, de forma mais realista, as dificuldades inerentes às simulações propostas em Manaus, tanto com modelos híbridos como com os convencionais.

A nova estratégia proposta, no entanto, sacrificaria sobremaneira o conjunto de treinamento e, certamente impactaria negativamente os resultados.

A nova estratégia de amostragem para validação revelou diferenças mais acentuadas entre o desempenho em treinamento e em validação, sobretudo nos modelos híbridos. O condicionamento do modelo híbrido C2N8 mostrou-se, mais uma vez, melhor que do modelo convencional C1N8, considerando a matriz de auto-correlação do Jacobiano e o número de condicionamento do Jacobiano e de seu transposto.

O gráfico de desempenho com a parada antecipada via validação cruzada *k-fold* de C2N8 comparado a C1N8 revelou discrepâncias maiores dos desvios médios quadráticos em validação em relação ao treinamento. Apesar dessa discrepância, os mesmos gráficos revelaram desempenho estimado maior para o modelo híbrido.

As diferenças de comportamento dos modelos híbridos em testes e em resultados já fora ressaltada na literatura sobre modelos híbridos em ZHANG *et al.* (2015). Porém, o problema ganha uma nova perspectiva diante da quantidade de simulações conduzidas na presente pesquisa, que revelaram, outras características positivas dos modelos híbridos e resultados melhores em termos de desvios absolutos médios.

11. Conclusões e Recomendações

A modelagem híbrida EMD-RNA foi abordada na presente pesquisa. O problema dos níveis extremos anuais que assolam a cidade de Manaus foi tomado como motivação na investigação dos modelos. Localizada no coração da Amazônia brasileira, na confluência entre os rios Negro e Solimões, a série de níveis histórica de Manaus tem sua variabilidade intrinsecamente relacionada às características da dinâmica tropical.

Foi realizada uma revisão sobre o ambiente tropical, evidenciando o papel crucial do calor latente na desestabilização da atmosfera tropical e suas relações com padrões de TSM nas bacias oceânicas do Atlântico e do Pacífico. As feições da circulação de larga escala nos trópicos, a ZCIT, o El Niño, o dipolo do Atlântico, as oscilações intrasazonais e seus efeitos na variabilidade interanual do El Niño e do dipolo do Atlântico foram exploradas, bem como suas relações com a climatologia e a hidrologia na bacia amazônica.

A revisão bibliográfica apontou ainda as sutilezas da dinâmica tropical, expressas nas incertezas entre o regime de ondas e o equilíbrio dinâmico entre o calor latente e a advecção vertical. Mais ainda, a dualidade entre hierarquia de escalas (visão tradicional da meteorologia) e a auto-afinidade foi observada em diversos estudos envolvendo dados atmosféricos nos trópicos. Como consequência, espectros de energia cheios e sem separações de escalas, com características de ruídos $1/f$ e com alguns picos expressando sinais hierárquicos importantes são esperados no ambiente de escopo.

Com a TSM relacionada ao calor latente e às chuvas na Amazônia, a presente pesquisa buscou explorar mais as relações fenomenológicas dessas séries com os níveis em Manaus. Para tanto, foram apresentadas as bases teóricas da DFA e da WTC e foram revisadas pesquisas que utilizaram tais técnicas em dados hidrológicos e atmosféricos. A EMD, parte integrante da HHT também foi apresentada, destacando seu caráter empírico e os problemas relacionados ao processo.

Em seguida, a base teórica sobre RNAs *feedforward* foi revisada, destacando-se sua característica intrínseca de mal condicionamento e ranqueamento do Jacobiano relacionando-o com a multicolinearidade. Efeitos do desentranhamento dos conjuntos de

validação e testes e aspectos sobre a generalização foram apresentados, comparando-se a parada forçada à Regularização Bayesiana.

De posse da base teórica sobre a DFA e a WTC, exploraram-se as relações fenomenológicas entre as séries de TSM e a hidrologia em Manaus, para utilização das séries de TSM como preditores na modelagem, considerando as especificidades tropicais revisadas. A DFA mostrou o caráter fractal das séries e a dualidade tropical expressa através de segmentos auto-afins intercalados por quebras indicando a hierarquia de certas escalas. a Coerência espectral Wavelet entre MAN e a TSM de regiões do Atlântico e do Pacífico, entre MAN e a OLR de três regiões distintas da Amazônia e entre as séries OLR e as séries de TSM das bacias oceânicas.

Mesmo não sendo utilizadas como preditores nos modelos, as séries de OLR ajudaram a entender as sutilezas do ambiente tropical e as relações da hidrologia com as bacias do Atlântico e do Pacífico.

A DFA revelou similaridades notáveis nas escalas intrasazonais e sazonais entre todos os dados, com reduções sutis na sazonalidade no Pacífico, de leste para oeste, sem redução das inclinações nas escalas mais rápidas, destacando a hierarquia intrasazonal no Pacífico, sobretudo com a MJO. Foram observadas características multifractais no Pacífico em escalas mais lentas, com relações não lineares nos escalogramas, reduzindo a hierarquia sazonal de leste para oeste em comparação ao Atlântico e à MAN. Em compensação, a variabilidade aumenta mais para escalas mais lentas no Pacífico, antecipando sua importância em tais frequências, tanto para a OLR como para MAN.

A Coerência Wavelet complementou a análise DFA com uma visão multi-escala variável. Foram confirmadas as notáveis correlações entre as séries estudadas na escala sazonal entre os dados, progressivamente menores para oeste na TSM do pacífico. A defasagem sazonal de MAN foi de cerca de 70 dias em relação à TSM do Atlântico, chegando a cerca de 250 dias em relação a NINO1+2 e diminuindo para leste. As correlações de 128 a 200 dias foram mais presentes entre MAN e TROP, no Atlântico e aumentaram gradualmente entre MAN e a TSM para oeste no Pacífico (MJO e outras oscilações de oeste), também corroborando com a DFA. Entre OLR e a TSM, observaram-se também relações sazonais mais fortes com o Atlântico e intrasazonais mais fortes com o Pacífico. A defasagem

média na faixa sazonal entre OLR e o Atlântico é menor que entre OLR e MAN, possivelmente porque as mínimas de OLR sofrem influência da floresta.

Por fim, a evolução da coerência Wavelet nos períodos maiores que 128 dias foi estudada, revelando tendências positivas com significância de 5% entre MAN e a TSM e mais fortes entre MAN e a OLR. Já a coerência das séries de OLR não apresentou tendências com confiança estatística ou exibiu quedas da coerência com o Atlântico e aumento com o Pacífico com confiança superior a 95%. Essas características das evoluções da WTC entre as bacias oceânicas, com MAN apresentando aumento da coerência com o Atlântico e a OLR com maior coerência com o Pacífico, sugerem um ciclo entre a bacia amazônica e o Atlântico Sul propiciado pela descarga de água doce do Amazonas.

Não foram observados possíveis efeitos da ação antrópica na região de promediação da OLR2, na WTC. Considerando a extensão da floresta, uma possível explicação é que o desmatamento em curso ainda não foi capaz de afetar os padrões de chuvas, representados pela OLR.

Com as relações entre a TSM e MAN evidenciadas, as séries NATL, SATL e NINO1+2 foram tomadas como preditores. Foram construídas quatro famílias de modelos. Os modelos híbridos foram compostos por tais preditores decompostos pela EMD, no caso de C2 e com projeção das IMFs no caso dos cenários PLS e PCA. Foram realizadas dez previsões de máximas anuais e dez de mínimas anuais em MAN. Cada modelo foi simulado 1.000 vezes em cada teste, permitindo o estudo da distribuição dos resíduos absolutos.

Os modelos convencionais de melhor desempenho, C1N3 em máximas e C1N5 e C1N8 em mínimas, foram comparados aos modelos híbridos. Os melhores modelos híbridos foram PLSN5, PCAN3 e C2N3, em máximas, e PLSN18, PCAN15 e C2N18, em mínimas. Os modelos híbridos apresentaram melhoras de desempenho em termos de diminuição dos desvios absolutos médios ao longo de todas as simulações superiores a 40%, em previsões de máximas, e a 35%, em previsões de mínimas.

No comparativo com o modelo probabilístico já validado, os desempenhos dos modelos híbridos ótimos foram equivalentes, em termos de desvios absolutos médios, mesmo

considerando a maior complexidade da previsão apenas a partir de dados climáticos de larga escala e com antecipação fixa de 60 dias.

A distribuição dos resíduos médios dos melhores modelos híbridos mostrou, no entanto, uma característica negativa em comum: apesar dos números de condicionamento dos Jacobianos de tais modelos serem significativamente menores em comparação aos modelos tradicionais C1, eles apresentaram variâncias consideráveis e discrepâncias maiores entre os resíduos estimados pelos conjuntos de testes e os verificados nas simulações.

Os modelos convencionais C1 com os mesmos números de parâmetros não apresentaram tais características, que sugerem superdimensionamento. Para melhor explorar tais atributos da modelagem híbrida com a EMD, sugerem-se testes complementares com outras técnicas de amostragem para a validação, podendo-se avaliar a parada forçada como estratégia de generalização.

Como recomendação para investigação complementar, sugerem-se simulações com estratégias diferentes de amostragens para a validação e generalização dos modelos. Duas simulações utilizando a validação cruzada *k-fold* com parada antecipada foram realizadas, comparando os modelos C1N8 e C2N8 na previsão da mínima do dia 31/10/2002. Novas estratégias de amostragem podem revelar melhor as discrepâncias de comportamentos entre treinamento e testes, presente nos modelos híbridos.

Apesar disso, tais estratégias podem sacrificar o treinamento e comprometer os resultados. De qualquer forma, a extensa investigação conduzida na presente pesquisa revela mais sobre os modelos híbridos do que apenas os excelentes desempenhos em treinamento e as inconsistências em validação relatadas na literatura.

Espera-se que esses testes complementares ajudem a entender os motivos da grande variância presente nas distribuições de desvios absolutos médios dos modelos híbridos, apesar de seus números de condicionamento consideravelmente menores e das reduções de vieses maiores que 40% em previsões de máximas e maiores que 35%, em previsões de mínimas, em relação aos modelos convencionais.

Adicionalmente, sugere-se que o potencial da modelagem híbrida EMD-RNA seja mais explorado no âmbito da gestão da rede hidrometeorológica brasileira. A modelagem híbrida constitui alternativa atraente para a otimização de ajuste de relações cotas-descargas sob efeito de remanso, em séries com alterações morfodinâmicas de calha ou em séries com alterações causadas por desnivelamento de réguas de medição de níveis. Com as dimensões continentais do Brasil e com uma matriz energética em expansão na direção de áreas remotas da Amazônia, esses são problemas recorrentes em estudos hidrológicos.

O efeito de remanso é comum na região amazônica e dificulta o ajuste das curvas-chaves. Já o isolamento de muitos postos hidrológicos pela floresta agrava a ocorrência de alterações nos cotogramas causadas por movimentações das réguas ou por mudanças morfodinâmicas das calhas. A expectativa é que a EMD ajude as RNAs a lidarem com tais efeitos na medida em que eles sejam separados em IMFs.

Em regiões remotas, são também comuns as lacunas nas séries por interrupção das medições. Assim, outra importante aplicação sugerida para os modelos EMD-RNA é no preenchimento de lacunas em séries hidrológicas, seja via dados oceanográficos, meteorológicos ou hidrológicos disponíveis.

Muito mais do que as características da modelagem EMD-RNA, foram exploradas relações fenomenológicas importantes no ambiente tropical, permitindo uma abordagem capaz de lidar com as sutilezas desse ambiente e integrando a oceanografia, a meteorologia e a hidrologia. Diante das mudanças climáticas em curso, entende-se que a abordagem foi capaz de incorporar tendências em séries de larga escala para explicar os eventos extremos recorrentes na última década nos escoamentos da bacia amazônica.

12. Bibliografia

- ACKERMAN, S., 2013. *Cooperative Institute For Meteorological Satellite Studies*.
[Online]
Disponível em: <http://cimss.ssec.wisc.edu/wxwise/homerbe.html>
[Acesso em: 04 04 2017].
- ADAMOWSKI, J. & SUN, K., 2010. Development of a coupled Wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, jun., Volume 390, pp. 85-91.
- AMARI, S.; MURATA, N.; MÜLLER, K. R.; FINKE, M.; YANG, A. H., 1996. *Statistical Theory of Overtraining - Is Cross-Validation Asymptotically Effective?*.
[Online]
Disponível em: <http://dblp.uni-trier.de/db/conf/nips/nips1995.html>
[Acesso em: 25 5 2018].
- ASPLINDER, C. I., 1976. A Classification of the Structure of the Tropical Atmosphere and Related Energy Fluxes. *Journal of Applied Meteorology*, abr., Volume 15, pp. 692-697.
- BARAER, M.; MARK, BG.; MCKENZIE, JM.; CONDOM, T.; BURY, J.; HUH, K.; PORTOCARRERO, C.; J., GÓMEZ; RATHAY, S., 2012. Glacier recession and water resources in Peru's Cordillera Blanca. *Journal of Glaciology*, 58(207).
- BARICHIVICH, J. et al., 2018. Recent intensification of Amazon flooding extremes driven by strengthened Walker circulation. *Science Advances*, 4(9).
- BARNHART, B. L., 2011. *The Hilbert-Huang Transform: theory, applications, development*. [Online]
Disponível em: <http://ir.uiowa.edu/cgi/viewcontent.cgi?article=2751&context=etd>
[Acesso em: 19 10 2018].
- BISHOP, C. M., 1995. *Bayesian methods for neural networks*. [Online]
Disponível em: <https://research.microsoft.com/apps/pubs/default.aspx?id=67154>
[Acesso em: 27 9 2018].
- BRAGA, A. D. P., DE CARVALHO, A. P. D. L. F. & LUDEMIR, T. B., 1998. *Fundamentos de Redes Neurais Artificiais*. 1 ed. ed. Rio de Janeiro: DCC/IM, COPPE/Sistemas, NCE/UFRJ.
- CARVALHO, L. M. V. et al., 2010. *Moisture transport and intraseasonal variability in the South America monsoon system*. [Online]

Disponível em:

https://www.researchgate.net/publication/225393858_Moisture_transport_and_interseasonal_variability_in_the_South_America_monsoon_system

[Acesso em: 7 10 2018].

CAVALCANTI, I. F., FERREIRA, N. J., DIAS, M. A. F. D. S. & SILVA, M. G. A. J.

D. S., 2009. *Tempo e Clima no Brasil*. São Paulo: Oficina de Textos.

COE, M.T.; LATRUBESSE, E.M.; FERREIRA, M.E.; AMSLER, M.L., 2011. The effects of deforestation and climate variability on the streamflow of the Araguaia River, Brazil. *Biogeochemistry*. Volume 105, pp. 1-3.

COSTA, M., BOTTA, A. & CARDILLE, J., 2003. Effects of large-scale changes in

land cover on the discharge of the Tocantins River, Southeastern Amazonia. *J*

Hydrol, Volume 283, pp. 206-217.

CYBENKO, G., 1989. Approximation by superpositions of a sigmoidal function.

Mathematics of Control, Signals, and Systems, 2(4), pp. 303-314.

DAVIDSON, E. A.; ARAÚJO, A. C.; ARTAXO, P.; BALCH, J. K.; BROWN, I. F.; BUSTAMANTE, M. M. C.; COE, M. T.; DEFRIES, R. S.; KELLER, M.; LONGO, M.; MUNGER, J. W.; SCHROEDER, W.; SOARES F., BRITALDO S.; SOUZA, C.; WOFSEY, S. C., 2012. The Amazon basin in transition. *Nature*, , 481(7381), pp. 321-328.

DE PESSOA, J. A., 2014. Pluviometric ID: precipitation characteristics at a glance.

Atmos. Sci. Let, Volume 15, pp. 288-291.

FAVARETTO, A. B., 2004. *Estimativa do Exponente de Hurst de Séries Temporais de Chuvas do Estado de São Paulo Usando as transformadas de Fourier, Wavelet e Análise R/S*. Rio Claro(SP): Universidade Estadual Paulista.

FORESEE, F. D. & HAGAN, M. T., 1997. *Gauss-Newton approximation to Bayesian learning*. [Online]

Disponível em: <http://hagan.ecen.ceat.okstate.edu/icnn97a.pdf>

[Acesso em: 30 9 2018].

FRANCOU, B. & VINCENT, C., 2010. *Les glaciers `a l`'epreuve du climat*. marseille: IRD.

GELADI, P. & KOWALSKI, B. R., 1986. Partial Least Square Regression: a Tutorial.

Analytica Chimica Acta, Volume 185, pp. 1-17.

GIAMBELLUCA, T. W., 2002. Hydrology of altered tropical forest. *Hydrological Processes*, Volume 16, pp. 1665-1669.

- GIANNINI, A., KUSHNIR, Y. & CANE, M. A., 2000. Interannual Variability of Caribbean Rainfall, ENSO, and the Atlantic Ocean*. *Journal of Climate*, 13(2), pp. 297-311.
- GIRARDI, C. & GIRARDI, R. D. M., 2001. *Fundação joaquim Nabuco*. [Online] Disponível em: <http://www.fundaj.gov.br/geral/nesa/cta.pdf> [Acesso em 11 set. 2016].
- GRINSTED, A., MOORE, J. C. & JEVREJEVA, S., 2004. Application of the cross Wavelet transform and Wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, Volume 11, pp. 561-566.
- HASTENRATH, S. & HELLER, L., 1977. Dynamics of climatic hazards in northeast Brazil. *Quarterly Journal of the Royal Meteorological Society*, 103(435), pp. 77-92.
- HENRIQUES, M. C., 2008. *Ondeletas e o Movimento Browniano Fracionário: Aplicação à Caracterização de Poços de Petróleo*. Natal(RN): UFRN.
- HOLTON, J. R., 2004. *An introduction to Dynamic Meteorology*. 4ª ed. Seattle(Washington D.C.): Elsevier Academic Press.
- HUANG, N. & ATTOH-OKINE, N. O., 2005. *Hilbert-Huang Transform in Engineering*. 1ª ed. Boca Raton, FL: Taylor&Francis Group.
- HUANG, N. E.; SHENG, Z.; LONG, S. R.; WU, M. C.; SHIH, H. H.; ZHENG, Q.; YEN, N. C.; TUNG, C. C.; LIU, H. H., 1998. *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*. [Online] Disponível em: <http://rspa.royalsocietypublishing.org/content/454/1971/903> [Acesso em: 18 10 2018].
- HUANG, N. E. & WU, Z., 2008. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46(2).
- IHLEN, E. A. F., 2012. Introduction to multifractal detrended fluctuation analysis in Matlab. *frontiers in PHYSIOLOGY*, 04 jun, Volume 3, pp. 1-18.
- ISMAIL, S. & SHABRI, A., 2014. Empirical Mode Decomposition for River Flow Forecasting. *Australian Journal of Basic and Applied Sciences*, 30 ago, Volume 8, pp. 8-15.
- JAHFER, S., VINAYACHANDRAN, P. N. & NANJUNDIAH, R. S., 2017. Long term impact of Amazon river runoff on northern hemispheric climate. *Scientific Reports*, Volume 7, pp. 1-9.

- JEVREJEVA, S., Moore, J. C. & GRINSTED, A., 2003. Influence of the Arctic Oscillation and El Nino-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach. *Journal of Geophysical Research*, Volume 108.
- KASER, G., 1999. A review of the modern fluctuations of tropical glaciers. *Global and Planetary Change*, 22(1), pp. 93-103.
- KAYANO, M. T., JONES, C. & DIAS, P. L. S., 2009. Variabilidade Intrasazonal. In: In: et al. eds. *Tempo e Clima no Brasil*. 1ª ed. São Paulo: Oficina de Textos.
- KOHAVI, R., 1995. *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. Stanford, CA, s.n.
- KOKES, J. & NGHIEN, N., 2012. *Using constrained cubic spline instead of natural cubic spline to eliminate overshoot and undershoot in Hilbert Huang Transform*. [Online]
Disponível em: <http://ieeexplore.ieee.org/document/6228658>
[Acesso em: 25 10 2018].
- KRUGER, C. J. C., n.d. *Constrained Cubic Spline Interpolation for Chemical Engineering Applications*. [Online]
Disponível em: www.korf.co.uk/spline.pdf
[Acesso em: 25 10 2018].
- LABAT, D., 2010. Cross Wavelet analyses of annual continental freshwater discharge and selected climate indices. *Journal of Hydrology*, pp. 269-278.
- LAING, A. & JENNI-LOUISE, E., 2016. *COMET MetEd*. [Online]
Disponível em:
http://www.meted.ucar.edu/tropical/textbook_2nd_edition/index.htm
[Acesso em: 07 02 2017].
- LIMA, L. S.; FILHO, B. S. SOARES; RODRIGUES, HERMANN; COE, MICHAEL T.; CUADRA, S. V.; DIAS, LÍVIA C. P.; COSTA, MARCOS HEIL; ROLFS, P. H., 2014. Feedbacks between deforestation, climate, and hydrology in the Southwestern Amazon: implications for the provision of ecosystem services. *Landscape Ecology*, , 29(2), pp. 261-274.
- LOPES, A. V., CHIANG, J. C. H., THOMPSON, S. A. & DRACUP, J. A., 2016. Trend and uncertainty in spatial-temporal patterns of hydrological droughts in the Amazon basin. *Geophysical Research Letters*, 43(7), pp. 3307-3316.
- MANDELBROT, B. B., 1983. *The Fractal Geometry of Nature*. 1ª ed. New York: W.H. FREEMAN AND COMPANY.

- MARK, B. G. & SELTZER, G. O., 2003. Tropical glacier meltwater contribution to stream discharge: a case study in the Cordillera Blanca, Peru. *Journal of Glaciology*, , 49(165), pp. 271-281.
- MARK, BRYAN; BURY, JEFFREY; MCKENZIE, JEFFREY; FRENCH, ADAM; BARAER, MICHEL, 2010. Climate Change and Tropical Andean Glacier Recession: Evaluating Hydrologic Changes and Livelihood Vulnerability in the Cordillera Blanca, Peru. *Annals of the Association of American Geographers*, , 100(4), p. 794–805.
- MARSHAL, J. & PLUMB, A., 2008. *Atmosphere, Ocean and Climate Dynamics: An Introductory Text*. 1^a ed. Burlington(MA): Academic Press.
- MAYTA, V. R. C., AMBRIZZI, T., ESPINOZA, J. C. & DIAS, P. L. S., 2018. The role of the Madden–Julian oscillation on the Amazon Basin intraseasonal rainfall variability. *International Journal of Climatology*.
- MOREIRA, J. & NAGUETTINI, M., 2016. Detecção de Tendências Monotônicas Temporais e Relação com Erros dos Tipos I e II: Estudo de Caso em Séries de Precipitações Diárias Máximas Anuais do Estado do Acre. *Revista Brasileira de Meteorologia*, 31(4), pp. 394-402.
- MURRAY, L.; NGUYEN, H.; LEE, Y. F.; REMMENGA, M. D.; SMITH, D. W., 2012. *Variance Inflation Factors in Regression Models with Dummy Variables*. Kansas, Kansas City, USA, Conference on Applied Statistics in Agriculture.
- NAPOLITANO, G., SERINALDI, F. & SEE, L., 2011. Impact of EMD decomposition and random initialisation of weights in ANN hindcasting of daily stream flow series: An empirical examination. *Journal of Hydrology*, Volume 406, pp. 199-214.
- NEYSHABUR, B., 2017. *Implicit Regularization in Deep Learning*. s.l., Ph.D. thesis, Toyota Technological Institute, Chicago, Illinois, USA.
- NOBRE, P. & SHUKLA, J., 1996. Variations of Sea Surface Temperature, Wind Stress, and Rainfall over the Tropical Atlantic and South America. *Journal of Climate*, 12 out, Volume 9, pp. 2464-2479.
- NOURANI, V., KOMASI, M. & MANO, A., 2009. A Multivariate ANN-Wavelet Approach for Rainfall-runoff Modeling. *Water Resource Manager*, 3 fev., Volume 23, pp. 2877-2894.

- PENG, C. K., HAVLIN, S., STANLEY, H. E. & GOLDBERGER, A. L., 1995. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, 5(1), pp. 82-87.
- PIMENTEL, J., WANDERLEY, R. & GLASHESTER, S., 2011. Análise da Aplicabilidade do Modelo de Cheias de Galdino & Clarke em Manaus. *10º SILUSBA*, sep, p. 20.
- PINTO, R. M. D. F., 2009. *Novas Abordagens ao Estudo de Sinais Biométricos: Análise em Tempo-Frequência e Transformada de Hilbert-Huang*. Lisboa, Portugal: Tese de D.Sc., Universidade de Lisboa.
- RABATEL, A.; FRANCOU, B.; SORUCO, A.; GOMMEZ, J.; CÁCERES, B.; CEBALLOS, J. L., 2013. Current state of glaciers in the tropical Andes: a multi-century perspective on glacier evolution and climate change. *The Cryosphere*, Volume 7, pp. 81-102.
- RAMACHANDRA, A. R. & HSU, E.-C., 2008. *Hilbert-Huang Transform analysis of hydrological and environmental time series*. Dordrecht: Springer.
- REBOITA, M. S. & SANTOS, I. A., 2014. Influência de Alguns Padrões de Teleconexão na Precipitação no Norte e Nordeste do Brasil. *Revista Brasileira de Climatologia*, 15(1), pp. 28-48.
- RODRÍGUEZ, J. D., PÉREZ, A. & LOZANO, J. A., 2010. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), pp. 569-575.
- RONCHAIL, J.; COCHONNEAU, G.; MOLINIER, M.; GUYOTT, J.; CHAVES, A. G. M.; V., GUIMARAES; OLIVEIRA, E., 2002. Interannual rainfall variability in the Amazon basin and sea-surface temperatures in the equatorial Pacific and the tropical Atlantic Oceans. *International Journal of Climatology*, 22(13), pp. 1663-1686.
- RUMELHART, D. E., HINTON, G. E. & WILLIAMS, R. J., 1986. Learning representations by back-propagating errors. *Nature*, 323(6088), p. 533-536.
- SAARINEN, S., BRAMLEY, R. & CYBENKO, G., 1994. Ill-Conditioning in Neural Network Training Problems. *SIAM journal on Scientific and Statistical Computing*, pp. 1-23.
- SANTOS, C. A. G. & DA SILVA, G. B. L., 2013. Daily streamflow forecasting using a Wavelet transform and artificial neural network hybrid models. *Hydrological Sciences Journal*, maio, Volume 1, pp. 37-41.

- SERGIU, C., 2011. *CODE PROJECT*. [Online]
Disponível em: <http://www.codeproject.com/Articles/175777/Financial-predictor-via-neural-network>
[Acesso em 8 set. 2016].
- SJÖBERG, J. & LJUNG, L., 1995. Overtraining, regularization and searching for a minimum, with application to neural networks. *International Journal of Control*, 62(6), pp. 1391-1407.
- SOUZA, E. B.; KAYANO, M. T.; TOTA, J.; PEZZI, L.; FISCH, G.; NOBRE, C., 2000. On the influences of the El Niño, La Niña and Atlantic Dipole Pattern on the Amazonian Rainfall during 1960-1998. *Acta Amazonica*, abr/jun, 30(2), pp. 305-318.
- TIANLU, M., ZENGLI, L. & ZHI, Z., 2013. *Inhibition of Envelope Correction Method to EMD End Effect*. Paris, França, Atlantis Press, pp. 2097-2100.
- TORRENCE, C. & COMPO, G. P., 1998. A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, jan, Volume 79.
- VILLWOCK, A., 2002. NOAA site offers ENSO information. *Eos, Transactions American Geophysical Union*, 83(37), pp. 407-407.
- VUILLE, M.; CAREY, M.; HUGGEL, C.; BUYTAERT, W.; RABATEL, A.; JACOBSEN, D.; SORUCO, A.; VILLACÍS, M.; YARLEQUE, C.; TIMM, O.; CONDOM, T.; SALZMANN, N.; SICART, J. E., 2018. Rapid decline of snow and ice in the tropical Andes – Impacts, uncertainties and challenges ahead. *Earth-Science Reviews*, Volume 176, pp. 195-213.
- WALLACE, J. M. & HOBBS, P. V., 2006. *Atmospheric Science*. s.l.:s.n.
- WU, Z. & HUANG, N. E., 2009. ENSEMBLE EMPIRICAL MODE DECOMPOSITION: A NOISE-ASSISTED DATA ANALYSIS METHOD. *Advances in Adaptive Data Analysis*, 1(01), pp. 1-41.
- YANG, X., CHENG, G. & LIU, H., 2015. Improved Empirical Mode Decomposition Algorithm of Processing Complex Signal for IoT Application. *International Journal of Distributed Sensor Networks*, pp. 1-8.
- YANO, J. I., BLENDER, R., ZHANG, C. & FRAEDRICH, K., 2004. 1/f noise and pulse-like events in the tropical atmospheric surface variabilities. *Quarterly Journal of the Royal Meteorological Society*, Volume 130, pp. 1697-1721.

- YANO, J. I. & BONAZZOLA, M., 2009. Scale Analysis for Large-Scale Tropical Atmospheric Dynamics. *Journal of the Atmospheric Science*, Jan., Volume 66, pp. 159-172.
- YANO, J. I., FRAEDRICH, K. & BLENDER, R., 2001. Tropical Convective Variability as $1/f$ Noise. *Journal of Climate*, Feb., Volume 14, pp. 3608-3616.
- YANO, J. I. & NISHI, N., 1989. The Hierarchy and Self-Affinity of the Time Variability within the Tropical Atmosphere Inferred from the NOAA OLR Data. *Journal of Meteorological Society of Japan*, Oct., 67(5), pp. 771-787.
- YU, P., LOW, M. Y. & ZHOU, W., 2017. Development of a partial least squares-artificial neural network (PLS-ANN) hybrid model for the prediction of consumer liking scores of ready-to-drink green tea beverages. *Food Research International*, Oct., Volume 103, pp. 68-75.
- ZHANG, C., 2005. Madden-Julian Oscillation. *Reviews of Geophysics*, 43(2).
- ZHANG, C. & GOTTSCHALCK, J., 2002. SST Anomalies of ENSO and the Madden-Julian Oscillation in the Equatorial Pacific*. *Journal of Climate*, 15(17), pp. 2429-2445.
- ZHANG, Q., ZHU, H. & SHEN, L., 2010. *A new method for mitigation of end effect in empirical mode decomposition*. [Online]
Disponível em: <https://www.researchgate.net/publication/224134959>
[Acesso em 21 out 2019].
- ZHANG, X., PENG, Y., ZHANG, C. & WANG, B., 2015. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. *Journal of Hydrology*, Sep., Volume 530, pp. 137-152.
- ZHU, W., ZHAO, H., XIANG, D. & CHEN, X., 2013. A flattest constrained envelope approach for empirical mode decomposition.. *PLOS ONE*, 8(4).

I. Movimento Browniano Fractal e a auto-afinidade

Vivemos num mundo no qual o determinismo e o indeterminismo se mesclam para gerar fenômenos complexos aos quais somos capazes de descrever sem total certeza e resolução, mas ao mesmo tempo, sem ignorá-los completamente (HENRIQUES, 2008).

O Movimento Browniano Fractal – MBF, é um modelo de processo estocástico no qual essas características determinísticas e indeterminísticas estão presentes, com oscilações de alta frequência associadas à correlações de longo alcance, essas últimas devidas à própria definição do movimento. As feições geométricas desse modelo podem ser, de certa forma, caracterizadas pelo expoente de Hurst, quantificando-se a probabilidade de tendências e anti-tendências do sinal.

Assim, ainda segundo HENRIQUES (2008), podem-se prever certos comportamentos estatísticos de séries auto-afins ou auto-similares. Esse tipo de processo é uma evolução do movimento Browniano clássico, estudado por Einstein e formalizado por Wiener, no qual a aleatoriedade reina absoluta. O movimento Browniano está presente nas nuances da natureza descritas pela mecânica quântica, organizadas por eventos probabilísticos e que constroem um mundo macroscópico para o qual o determinismo pareceu ser a visão científica mais plausível até certa época.

Cabe aqui uma observação sobre as diferenças entre aleatoriedade e imprevisibilidade por falta de resolução (espacial ou temporal) ou por deficiência na extensão da série (espacial ou temporal). Os fenômenos reais carregam não linearidades e não estacionariedade e, muitas vezes, interdependência de um número tal de variáveis que os tornam impossíveis de serem descritos deterministicamente, frente às limitações de resolução ou de extensão de amostragem.

Como em um MBF, nos fenômenos complexos do interesse dessa pesquisa, não se consegue prever exatamente o futuro, mas pode-se, na medida em que as séries se adensam e se prolongam, melhorar as relações ente efeitos de comportamentos atuais e passados com os futuros. Espera-se que as técnicas de análise de dados discutidas nessa

pesquisa, aplicadas aos dados de extensão espaço-temporal consideráveis, ajudem a separar as características determinísticas da imprevisibilidade. E que ainda ajudem a ressaltar características comuns inerentes a tais séries e à hidrologia de Manaus.

O entendimento do MBF e do ruído Gaussiano Fractal, são cruciais para o entendimento de tais ferramentas e da análise fractal. Dessa forma os dados utilizados na presente pesquisa puderam ser comparados entre si e descritos pela geometria fractal, revelando características estatísticas, possíveis escalas hierárquicas ou a auto-afinidade presentes nas séries temporais.

Caminhar aleatório e o movimento Browniano

Um caminhar aleatório descreve passos aleatórios em intervalos uniformes de tempo. Assim, o caminhar aleatório é também denominado ruído Gaussiano – RG, na literatura. Já o movimento Browniano – MB, pode ser descrito como processo cujos incrementos independentes possuem distribuição (HENRIQUES, 2008):

$$B(t_2) - B(t_1) \sim N(0, \sigma^2), \quad (0 \leq t_1 < t_2) \quad (\text{I-1})$$

na qual $N(\mu, \sigma^2)$ expressa uma distribuição normal de média μ e variância σ^2 . O MB é classificado como processo Gaussiano relacionado de forma trivial ao RG: os incrementos do MB formam o RG e, conseqüentemente, a soma acumulada de um RG forma um MB. Segue de (I-1) (HENRIQUES, 2008):

$$E\left\{[B(t_2) - B(t_1)]^2\right\} = \sigma^2 |t_2 - t_1| \quad (\text{I-2})$$

Ou seja, a variância do incremento de um movimento Browniano é proporcional à magnitude do deslocamento. Além disso, $B(0) = 0$ e $E[B(t)] = 0$. A auto-afinidade significa, usando a relação (3-7) com $H = 1/2$ e considerando $t_1 = 0$ e $B(t_1) = 0$, para $\lambda > 0$:

$$B(t) \cong \frac{1}{\sqrt{\lambda}} B(\lambda t) \quad (\text{I-3})$$

Sendo a igualdade acima estatística, significando que os dois termos possuem a mesma distribuição, sendo estatisticamente indistinguíveis.

Movimento Browniano Fractal

A expressão (I-2) pode ser generalizada com a inclusão do expoente de Hurst H como segue:

$$E\left\{[B(t_2) - B(t_1)]^2\right\} = \sigma^2 |t_2 - t_1|^{2H}, \quad 0 < H < 1 \quad (\text{I-4})$$

Esse processo Gaussiano é conhecido como movimento Browniano fractal ou fracionário – MBF. Nele, a variância dos incrementos não apresenta, necessariamente, relação linear com o tempo, ficando a exceção para $H = 1/2$, para o qual (I-4) reduz-se ao movimento Browniano comum.

Segundo HENRIQUES (2008), a importância do MBF aumentou com Mandelbrot nas décadas de 1960 e 1970, mas sua ideia já havia sido considerada por Kolmogorov em 1940. O MBF já foi e continua sendo bastante usado para modelar variações de índices financeiros, estudar difusões anômalas e em computação gráfica, como modelo padrão para geração de paisagens fractais.

Segundo MANDELROT (1983), a característica básica do MBF é que a extensão de interdependência entre seus incrementos é infinita, ou seja, mesmo amostragens distantes entre si, exibem interdependências, diferente do que ocorre em outros processos estocásticos.

Desenvolvendo (I-4) tem-se:

$$E[B(t_2)B(t_1)] = \frac{\sigma^2}{2} \left(|t_2|^{2H} + |t_1|^{2H} - |t_2 - t_1|^{2H} \right) \quad (\text{I-5})$$

HENRIQUES (2008) demonstrou que para qualquer $s > 0$:

$$E[B_H(st_2)B_H(st_1)] = E[s^H B_H(t_2)s^H B_H(t_1)]$$

ou seja, $B_H(st)$ e $s^H B_H(t)$ são dois processos Gaussianos com mesma média e mesma covariância. Portanto, podemos afirmar que:

$$B_H(t) \cong \frac{1}{s^H} B_H(st) \quad (\text{I-6})$$

generalizando (I-3) para o MBF. Convencionou-se dizer que os incrementos de B_H são estatisticamente auto-afins de acordo com o parâmetro H . Pode-se generalizar isso para incrementos a partir de qualquer tempo t :

$$B_H(t_1 + t_2) - B_H(t_1) \cong \frac{1}{s^H} [B_H(t_1 + st_2) - B_H(t_1)]$$

O MBF é o único processo Gaussiano, auto-afim e com incrementos estacionários conhecido. A exemplo do que ocorre com o MB e o RG, o ruído Gaussiano fractal é obtido através das diferenças do MBF e, conseqüentemente, o MBF é obtido da soma acumulada do respectivo ruído.

II. Caracterização da dinâmica atmosférica tropical

A seguir apresentam-se revisões detalhadas de trabalhos que contribuíram para a caracterização da dinâmica atmosférica tropical e ajudaram a identificar escalas hierárquicas e auto-afins comuns entre os sinais de interesse da presente pesquisa. Em resumo, após a análise dimensional realizada por YANO & BONAZZOLA (2009) apontar os domínios teóricos nos quais a atmosfera tropical seria dominada por ondas ou pelo equilíbrio, tais previsões não se confirmaram na análise de dados observacionais realizada pelos autores. Não foram encontrados indícios da transição entre regimes na faixa de 1.000 a 3.000 km na análise de dados, apesar das incertezas associadas à análise dimensional. Adicionalmente, a análise de dados observacionais realizada pelos autores apontou invariância de regime ao longo das escalas temporais e ao longo das escalas espaciais, para períodos menores que 70 dias, o que pode ser característica de ruídos $1/f$ associados a comportamentos pulsantes na atmosfera tropical.

Assim, além de YANO & BONAZZOLA (2009), revisam-se trabalhos que tratam da dualidade entre hierarquia e auto-afinidade e da caracterização da dinâmica tropical como ruído, associada à presença de eventos pulsantes em tal ambiente. No capítulo 3, discutiram-se os conceitos de auto-similaridade e auto-afinidade, o conceito de fractal e de dimensão fractal D e de sua relação com o expoente de Hurst H . Mais detalhes sobre tais conceitos podem ser consultados em MANDELBROT (1983) e FAVARETTO (2004). Já o conceito de Movimento Browniano Fractal – MBF foi revisado no ANEXO I.

a) Análise dimensional nos trópicos

A análise dimensional constitui uma valiosa ferramenta para identificar nas equações governantes os termos dominantes no cenário de interesse.

YANO & BONAZZOLA (2009) realizaram análise dimensional da dinâmica atmosférica tropical abrangendo a mesoescala espacial, a escala sinóptica e a planetária. Para permitir uma análise abrangente em escalas, os casos mais simples foram considerados, sem anisotropia horizontal. Fixando a magnitude de ventos horizontais, foram realizadas estimativas para os termos da equação da termodinâmica em diferentes escalas

horizontais de movimentos, destacando-se as sutilezas e incertezas inerentes à análise dimensional.

Para as diferentes escalas espaciais, dois regimes foram cogitados por YANO & BONAZZOLA (2009) para a dinâmica: o primeiro, marcado pelo equilíbrio dinâmico entre a advecção vertical de temperatura potencial e a taxa de calor diabático na termodinâmica; o segundo, consistindo em uma dinâmica de ondas equatoriais lineares moduladas pela convecção de calor, sendo estabelecida pelo equilíbrio entre a advecção vertical de temperatura potencial e o termo de variação local em uma escala de tempo majorada, também na equação termodinâmica. O equilíbrio dinâmico desembocaria na eliminação das ondas de gravidade, com uma estratificação neutra, um cenário, portanto, qualitativamente distinto da dinâmica tropical de larga-escala dominada por ondas equatoriais, descrito tradicionalmente na literatura.

YANO & BONAZZOLA (2009), ao adimensionalizarem o momentum, obtiveram um parâmetro fruto da multiplicação da escala advectiva de tempo L/U pelo fator $\Omega(L/a)y$ (oriundo da aproximação no plano β do parâmetro de Coriolis). Esse parâmetro foi chamado pelos autores de adimensional β , sendo:

$$\beta = \frac{2\Omega L^2}{aU} \therefore \begin{cases} \Omega: \text{velocidade angular da Terra;} \\ a: \text{raio da Terra.} \end{cases}$$

O adimensional β foi usado por YANO & BONAZZOLA (2009) como elo entre as escalas espaciais horizontais para avaliação sistemática da ordem dos demais parâmetros, comparando-os com dados observacionais nos diversos regimes. A alternativa do termo de atrito ou o termo de pressão serem dominantes no momentum seria mais plausível em uma análise focada na camada limite da troposfera, o que não é o caso.

Os autores enfatizaram as sutilezas associadas à análise dimensional da atmosfera tropical de larga-escala, que pode conduzir a resultados inconclusivos. A sensibilidade quadrática do parâmetro adimensional β em relação à escala espacial L torna a análise dimensional de larga-escala sutil nos trópicos. YANO & BONAZZOLA (2009) consideraram três distintas possibilidades para β , balizadas pela escolha da relação L^2/U em β :

$\beta \ll 1$, $\beta \sim 1$ e $\beta \gg 1$. A magnitude de β foi determinada pela escala horizontal L , havendo pouca margem para modificação da velocidade horizontal do escoamento U .

A partir de β , a análise de YANO & BONAZZOLA (2009) forneceu estimativas das flutuações da temperatura potencial $\delta\theta$, da máxima taxa de calor diabático possível δQ e das escalas de tempo possíveis. A comparação com dados observacionais excluiu certos regimes e forneceu restrições à análise.

Os autores mostraram ser a dinâmica balanceada uma possibilidade interessante para escalas espaciais menores que $L \sim 10^3$ km e $U \sim 10$ m/s, sendo plausível também para escalas sinópticas. Mas segundo os autores, a não divergência assintótica deve também ser cogitada na presença de forte estratificação vertical, tendo fundamento teórico e observacional, o que levaria a velocidade vertical a zero na ordem dominante (não divergência). O balanço entre calor diabático e advecção vertical seria recobrado com a expansão assintótica da velocidade vertical. O calor diabático requerido para equilibrar a advecção vertical aumenta excessivamente com a redução da escala L , a menos que a não-divergência assintótica seja considerada.

Foram destacadas pelos autores as sutilezas da análise, que não permitem excluir a possibilidade que as ondas de gravidade quebrem o equilíbrio entre a advecção e o calor diabático na escala sinóptica. A distinção entre os dois regimes ficou dentro das incertezas da análise dimensional, apesar da estimativa de flutuação da temperatura potencial para dinâmica de ondas, $\delta\theta \sim 3$ K, ser elevada frente às observações. A sensibilidade de β às escalas horizontais faz com que os regimes de ondas e de equilíbrio sejam separados por fator 3 em larga escala espacial (de 1.000 a 3.000 km), prevendo-se assim que as ondas dominem subitamente com aumento discreto de L .

Porém, a análise de dados indicou que a condição balanceada é satisfeita para escalas acima de $L \sim 10^3$ km. YANO & BONAZZOLA (2009) destacaram o caso da Oscilação de Madden-Julian, que pode ser interpretada à luz da dinâmica balanceada ao invés de por ondas equatoriais acopladas à convecção, mesmo com escala característica bem acima dos 3.000 km.

Sendo assim, a análise dimensional é uma estimativa grosseira do balanço dominante. Contudo, fixando a magnitude de ventos horizontais, foram realizadas estimativas para a escala temporal, para as variações de temperatura e para o calor diabático em diferentes escalas horizontais de movimentos.

Análise de dados observacionais

Foram feitas duas análises por YANO & BONAZZOLA (2009), ambas, dependentes das escalas temporais Δt e espaciais horizontais Δx dos dados. O conjunto de dados usados é oriundo da malha de larga-escala *Large-Scale Array* – LSA, durante o experimento *Tropical Ocean and Global Atmosphere* – TOGA, *Coupled Ocean-Atmosphere Response Experiment* – COARE, *Intensive Observing Period* (IOP, 1/10/1992 a 28/02/1993). Os dados foram processados pela *Colorado State University*. Os autores ressaltaram a possibilidade de contaminação dos dados pela dinâmica local dos cúmulos nos trópicos, entre outros fatores.

Para permitir as análises, YANO & BONAZZOLA (2009) obtiveram inicialmente médias móveis temporais e espaciais dos dados ao longo das escalas, considerando-se as escalas horizontais de longitude e de latitude semelhantes, $\Delta x \equiv \Delta y$. Os dados coletados possuem resolução espacial de 1° e temporal de 6h, com variação total de 40° em longitude e 20° em latitude e período de 4 meses, permitindo assim escalas de promediação espacial de 2° a 20°, correspondendo a $\Delta x \approx 200 - 2.000$ km, e temporal de 12h a 4 meses.

Então, os autores estimaram a ordem de magnitude da variável analisada pelo valor médio quadrático para cada escala de promediação Δx e Δt .

A primeira análise realizada por YANO & BONAZZOLA (2009) consistiu na caracterização da velocidade do vento horizontal médio quadrático $\langle |\mathbf{v}_H|^2 \rangle^{1/2} (\Delta x, \Delta t)$ a 850 hPa, ao longo das escalas espaciais e temporais de promediação. A segunda análise realizada, de maior interesse para a presente pesquisa consistiu na comparação entre o termo de variação local e o termo de advecção vertical da temperatura potencial na equação termodinâmica, utilizando-se o parâmetro R dado, em determinada escala, por:

$$R = \left\{ \frac{\langle (\partial\theta/\partial t)^2 \rangle}{\langle [\omega(\partial\theta/\partial p)]^2 \rangle} \right\}^{1/2} (\Delta x, \Delta t) \quad (\text{II-1})$$

Sendo: $\begin{cases} \partial\theta/\partial t & \text{a variação local da temperatura potencial;} \\ \omega(\partial\theta/\partial p) & \text{o termo de advecção vertical de temperatura potencial.} \end{cases}$

A relação R foi calculada pelos autores para 850, 500 e 250 hPa, revelando resultados semelhantes aos da análise com a velocidade, com tendência decrescente de R com a pressão. O numerador e o denominador decresceram com o aumento das escalas espacial e temporal, porém, em taxas distintas.

Valores baixos de R indicam equilíbrio termodinâmico. Os resultados mostraram valores maiores de R no domínio para as menores escalas temporais, independente das escalas espaciais horizontais. Os autores ressaltaram a possibilidade dos valores relativamente maiores de R para Δt pequenos estarem associados a erros de medição. Mas R rapidamente decresceu para janelas de promediação maiores que alguns dias. Para períodos de promediação Δt acima de 60 dias, a relação R apresentou leve decréscimo se mantida a escala espacial horizontal Δx , sugerindo a eliminação rápida de variações de baixa frequência relacionadas às torres de convecção, em janelas temporais maiores.

YANO & BONAZZOLA (2009) relataram que a tendência apontada pelos resultados da análise dimensional, de transição do regime de dinâmica balanceada para dinâmica dominada pelas ondas de gravidade entre 1.000 e 3.000 km, não foi observada na análise dos dados. O parâmetro R deveria aumentar para escalas horizontais maiores, indicando assim a tendência a transição para regime de dinâmica de ondas apontado pela análise dimensional. Porém, R mostrou-se quase independente de Δx para Δt menor que 40 dias.

A discrepância mais marcante entre os resultados da análise dimensional e os dados foi o fato de R ter apresentado tendência de decréscimo com aumento de Δx para Δt superior a 70 dias, sugerindo uma relevância maior que a prevista para o regime de dinâmica balanceada. No entanto, a análise dos dados foi limitada a escalas de 2.000 km, podendo

revelar tendências mais aderentes às previsões da análise dimensional para escalas espaciais maiores.

Os autores ressaltaram ainda que a persistência da variabilidade nas diversas escalas, característica de ruídos $1/f$ associados a comportamentos pulsantes na atmosfera tropical pode explicar a invariância de R ao longo das escalas temporais e ao longo das escalas espaciais para períodos menores que 70 dias.

Ainda sobre o parâmetro R , sua adequação para caracterizar o tipo de dinâmica atmosférica nos trópicos pode ser contestada, considerando-se as dificuldades na hierarquização de escalas inerente aos processos tropicais. Os autores ressaltaram que a persistência da variabilidade nas diversas escalas, característica de ruídos $1/f$ associados a comportamentos pulsantes na atmosfera tropical pode explicar a invariância de R ao longo das escalas temporais e ao longo das escalas espaciais para períodos menores que 70 dias. A formação de espectros cheios, tipo ruído, a partir de eventos pulsantes é discutida a seguir.

b) Hierarquia ou auto-afinidade?

Análises implementadas por YANO & NISHI (1989) exploraram a presença ou não de separação de escalas e hierarquização em dados de radiação de ondas longas – OLR da NOAA. Os autores justificaram a escolha de tal parâmetro para estudar a dualidade entre hierarquia e auto-afinidade na atmosfera tropical pela disponibilidade de séries mais longas e pela OLR relacionar-se com a atividade convectiva, apesar das interferências de temperaturas da superfície.

Inicialmente os autores reanalisaram a hierarquia na atmosfera tropical em seção tempo x longitude. Em seguida, os espectros de potência foram estudados e ajustaram-se, grosso modo, a uma potência -1 da frequência, implicando assim na ausência de frequências dominantes que caracterizariam diferentes escalas. Aprofundando a investigação, a análise fractal foi usada para medir a mudança da variabilidade com o tamanho das janelas de promediação temporal das séries de OLR.

Algumas feições marcantes da dinâmica tropical, já estudadas sob a visão tradicional de hierarquização de escalas, produzem picos característicos nos espectros de potência e

feições diferenciadas na análise fractal. A saber: a variação intrasazonal, o ciclo anual e o ENSO.

Hierarquia

A hierarquia da variabilidade na atmosfera tropical foi reexaminada por YANO & NISHI (1989) usando dados de OLR da NOAA, com extensão de 1.944 dados diários de janeiro de 1979 a abril de 1984, em grade de $2,5^\circ \times 2,5^\circ$. Os dados foram divididos em três grupos e apresentados em seções com interpolações tempo \times longitude: período de 1 ano (1981), com dados diários; período de três anos, com promediação de três dias (1980 a 1982); período de 5 anos, com promediação de 5 dias.

Os autores apontaram distúrbios sinóticos de 10-20 dias de período (período observado em um ponto fixo), propagando-se para leste, de 60° E a 180° E, com velocidade que leva cerca de 30 dias para circundar o globo como a menor escala temporal discernível na seção de 1 ano. Esses distúrbios foram chamados de superaglomerado de nuvens – livre tradução para *super cloud-clusters*, sendo organizados por aglomerados de nuvens de centenas de quilômetros, formando os distúrbios de milhares de quilômetros detectados na análise.

Os distúrbios sinóticos de 10-20 dias são modulados por um sinal com período de 30-60 dias, um conjunto de nuvens ao qual YANO & NISHI (1989) referem-se como variabilidade intrasazonal tropical – VIT, de escala da ordem de 10^4 km propagando-se para leste, com velocidade menor que os superaglomerados.

Em escalas mais lentas, YANO & NISHI (1989) apontaram um ciclo anual evidente no Atlântico (0° E – 45° E) e no Pacífico leste (90° W – 30° W), bem como uma abrupta intrusão no Pacífico leste de atividade convectiva vinda do Pacífico oeste, um El Niño de 1982-1983, com período de 3 a 10 anos, acompanhado de variabilidade global de igual escala, compondo um ENSO.

Segundo os autores, apesar de identificada em locais e em ocasiões específicas, essa hierarquização não é sempre clara. Já a auto-similaridade já fora demonstrada em vários padrões horizontais de nuvens tropicais: em nuvens sobre o oceano Índico; em nuvens na

ZCIT, em escalas de ~ 1 km a 10^3 km; em imagens de satélite LANDSAT, descendo até escala de ~ 50 m. A estrutura fractal foi demonstrada também tridimensionalmente em imagens de radar. As referências podem ser vistas em YANO & NISHI (1989). A quantificação citada buscou estudar a auto-afinidade, de forma similar, através utilização da geometria fractal.

Da mesma forma, a auto-afinidade fora reconhecida há muito na climatologia e na hidrologia em diversas escalas, implicando na presença de fractais temporais e espaciais comuns na atmosfera tropical. Isso, segundo YANO & NISHI (1989), corrobora com a visão da dinâmica atmosférica como turbulência bidimensional, com ausência de separação de escalas e a energia transitando por uma cascata de vórtices com auto-afinidade e auto-similaridade.

Quantificação da auto-afinidade

Para quantificação da auto-afinidade e da similaridade da dinâmica tropical, YANO & NISHI (1989) utilizaram tanto a visão tradicional de espectros de energia de séries temporais de OLR como também estudaram a auto-afinidade através da variabilidade das séries ao longo das escalas.

Os espectros de OLR apresentados por YANO & NISHI (1989), além de demonstrarem o caráter misto da variabilidade na atmosfera tropical entre auto-afinidade e hierarquia, evidenciaram, através da presença de picos, a relação entre fenômenos de diversas escalas e sistemas convectivos nos trópicos. Todos os fenômenos notáveis nos espectros, por sua vez, possuem relação com a TSM.

De forma geral, os espectros de cinco anos de OLR com intervalos de 45° de longitude apresentados pelos autores exibiram potência decaindo com o inverso da frequência elevada a um expoente positivo. Essa característica não é uma exceção, sendo achada em vasta quantidade de séries temporais físicas e está associada à auto-afinidade da variabilidade tropical.

No entanto, diversos picos se destacaram nos espectros: o ciclo anual, entre 45°W e 90°E ; um pico largo de 1-5 anos entre 180°W e 90°W , devido ao ciclo anual e ENSO, mesmo destacando-se a insignificância estatística de eventos com durações maiores que um ano

nessa análise, mas tendo em mente a ocorrência de ENSO em 1982-1983; finalmente, o pico de 30-60 dias (ou 20-80) próximo a 90°E.

Tais sinais característicos foram apontados em diversos outros estudos conforme citado por YANO & NISHI (1989). Destaca-se, no entanto, que, com exceção do pico anual, os demais não são estreitos o bastante para serem classificados como periódicos, nos espectros tratados pelos autores.

Para a análise da auto-afinidade da variabilidade da radiação OLR, YANO & NISHI (1989) postularam:

$$L(t) = \sum_{j=1}^{N/t-1} |\bar{\phi}_{j+1}(t) - \bar{\phi}_j(t)| \quad (\text{II-2})$$

sendo $L(t)$ a magnitude da variabilidade, N o número total de dias da amostra e $\bar{\phi}_j(t)$ o j -ésimo valor ($t \times j$ -ésimo dia) da promediação de janela t da OLR, definido como:

$$\bar{\phi}_j(t) = \sum_{i=1}^t \frac{\phi_{(j-1)t+i}}{t} \quad \left(j = 1, \dots, \frac{N}{t} \right) \quad (\text{II-3})$$

O valor da janela temporal t foi tomado como cada divisor possível de N , sendo no caso da análise $N = 2^3 \times 3^5 = 1.944$ dados diários de OLR da NOAA, com extensão de janeiro de 1979 a abril de 1984.

Se a variabilidade é auto-afim, então o comprimento $L(t)$ deve decrescer à taxa constante conforme t é aumentado à taxa constante. Ou seja, conforme a série é “alisada” com o aumento da janela temporal de promediação (t), a variabilidade diminui. Em perfis fractal, a dimensão fractal D relaciona-se com o expoente de Hurst H por (MANDELBROT, 1983):

$$D = 2 - H \quad (\text{II-4})$$

levando a uma relação entre a potência negativa da variabilidade e a dimensão fractal.

$$L(t) \propto t^{1-D} \quad (\text{II-5})$$

sendo D a dimensão fractal.

YANO & NISHI (1989) mostraram que a potência inversa do espectro está relacionada com o expoente de Hurst por:

$$2H = \alpha - 1 \quad (\text{II-6})$$

Assim, chega-se a uma relação entre a dimensão D e a potência inversa α do espectro:

$$\alpha = 5 - 2D \quad (\text{II-7})$$

Os autores realizaram análise da auto-afinidade através das relações de comprimento de variabilidade $L(t)$ da OLR com a janela de promediação t para os dados ao longo do equador, com intervalos de 45° de longitude. Observou-se ajuste bom de $L(t)$ para escalas temporais menores, tendendo a desviar ou se espalhar conforme t cresceu.

A dimensão fractal D foi o parâmetro ajustado da equação (II-5) por YANO & NISHI (1989), buscando o menor erro para o domínio $1 \leq t \leq t_{\max}$, com t_{\max} variando de 3 a 1.944 dias. Em escala bilogarítmica, o ajuste produziu linhas retas, cujos trechos de melhor ajuste representam a faixa de extensão da escala temporal auto-afim τ_s . Tais trechos de melhores ajustes foram limitados arbitrariamente pelos autores, pelo ponto no qual houve desvio em relação ao ajuste de $\delta \text{Log}_{10} L(t) = 1/10$. Acima desse limite ocorreram desvios abruptos dos valores calculados em relação à reta ajustada.

Cabe ressaltar que o conceito de escala temporal auto-afim τ_s usado por YANO & NISHI (1989) difere da auto-afinidade definida na presente pesquisa. Enquanto τ_s caracteriza o trecho no até o qual não existem desvios notáveis no escalograma de $L(t)$, a auto-afinidade pode ser identificada em escalas maiores que τ_s , com a variabilidade assumindo novas relações de semelhança além desse ponto.

Dentre tais desvios, os autores destacaram o decréscimo de $L(t)$, entre 0° e 45° W, por volta de 200 dias, representando o pico anual achado nos espectros. Já em 90° E, enquanto a oscilação de 30-60 dias foi notável no espectro, $L(t)$ apareceu como uma linha reta no intervalo aproximado de 1-10 dias, definindo assim a escala temporal auto-afim τ_s e decrescendo com taxa constante abruptamente para escalas temporais maiores. Não houve variação da declividade discernível que pudesse ser associada ao período de 30-60 dias. Essa diferença entre a detecção do ciclo anual e da variação de 30-60 dias na análise de auto-afinidade pode estar associada à maior definição dos picos anuais nos espectros em relação aos intrasazonais.

Por fim, segundo os autores, na região de atividade do ENSO, de 135° W a 90° W, houve um decréscimo abrupto de $L(t)$ próximo à escala de 200 dias. Mas o período de 2-5 anos do ENSO não foi evidenciado, talvez pela limitação de extensão da série de dados.

Outro resultado de YANO & NISHI (1989) relevante para a presente pesquisa foi a construção de um mapa de contornos de auto-afinidade da OLR a partir dos resultados acima, aplicando-se média móvel de 5° de longitude para suavizar efeitos do limiar arbitrário de $\delta \text{Log}_{10} L(t) = 1/10$. Portanto, optou-se pela adaptação do mapa de contornos de auto-afinidade nesse texto, buscando evidenciar o comportamento predominante nas regiões de origem das séries de TSM usadas como dados de alimentação do modelo híbrido para previsão hidrológica em Manaus.

Os autores destacam de antemão que valores relativamente largos de escala de auto-afinidade τ_s são achados em latitudes maiores que 15° N a 30° N. Como os dados de OLR dessas latitudes envolvem outras quantidades físicas além da temperatura do topo das nuvens, isso torna a interpretação de tal resultado complexa, sendo conveniente concentrar-se em latitudes mais baixas, sobretudo de 15° S a 15° N.

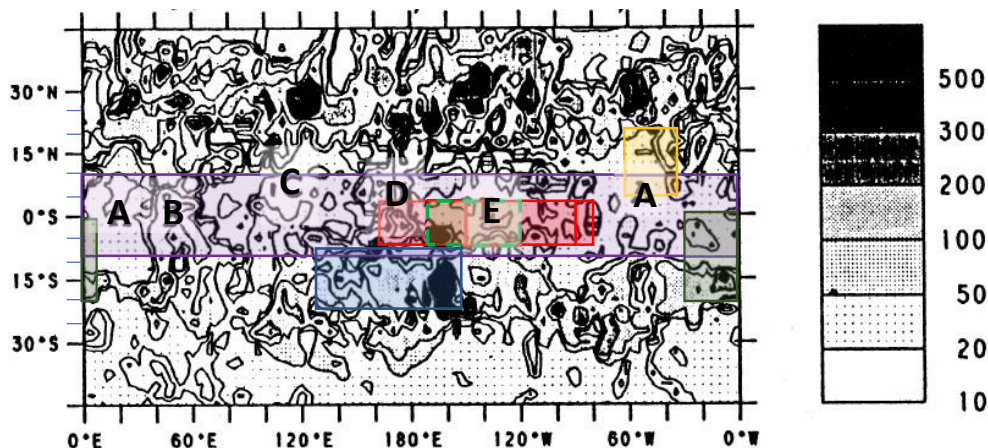


Figura II-1. Mapa de contornos de escala de auto afinidade τ_s em seção transversal latitude-longitude com média móvel espacial de 5° longitudinais. Em destaque, regiões aproximadas dos dados de alimentação do modelo híbrido do presente estudo: TSM do Atlântico Norte em laranja (5-20°N, 60°-30°O); TSM do Atlântico Sul em marrom (0°-20°S, 30°O-10°L); SOI em azul (entre Darwin, na Austrália e Tahiti); El Niño1+2(0-10°S, 90°-80°O), El Niño 3 (5°N-5°S, 150°-90°O), El Niño 4 (5°N-5°S, 160°L-150°O) em vermelha; El Niño 3.4 (5°N-5°S, 170-120°O) em verde, com linha tracejada. Adaptado de YANO & NISHI (1989).

Essa área de maior interesse foi dividida em cinco regiões pelos autores, A até E, de acordo com a magnitude de τ_s : as regiões A, C e E, nas quais τ_s é relativamente baixa (20-50 dias) e as regiões B e D, nas quais τ_s é relativamente alta, maior que 100 dias. As regiões de auto-afinidade relativamente menor correspondem àquelas nas quais domina alguma periodicidade.

A Região A está localizada sobre o Atlântico, sofrendo influência do marcante ciclo anual. Em tal região localizam-se os dados de TSM do Atlântico Sul e Norte, doravante referidos, respectivamente como SATL e NATL. Boa parte dos dados de TSM do Atlântico Tropical – TROP, também foram coletados nas regiões A, C e E. Assim, tais regiões de baixas escalas τ_s estão sob influência de ciclos anuais pronunciados.

A área de coleta dos dados de TSM do El Niño (NINO1+2, NINO3 e NINO3.4) no Pacífico Sul possui grande interseção com a área E, de relativa baixa escala de auto-afinidade τ_s . Tal região está sob influência predominante dos ciclos anual e de 2-5 anos do El Niño. Já a série Niño 4 é oriunda de uma área que, além de ter interseção com a região E, se estende pela região D, de maiores valores relativos da escala τ_s .

Já a oscilação de 30-60 dias é importante na região **C**, do Oceano Índico leste à Oceania.

Segundo YANO & NISHI (1989), as regiões **A** e **E** possuem propriedades similares: a declividade de $L(t)$ é razoavelmente baixa com $D \approx 1$ de $t \approx \tau_s$ a $t \approx 200$ dias. Há um decréscimo abrupto de $L(t)$ acima de $t \approx 200$ dias nas duas regiões. Assim, a variabilidade é auto-afim abaixo dos 200 dias, acima dos quais é caracterizada pelo ciclo anual tanto em **A** quanto em **E**. Por outro lado, na região **C** a variabilidade continua auto-afim, a grosso modo, para $t \geq \tau_s$, com uma dimensão fractal maior ($D \approx 2,5$) do que para $t \leq \tau_s$, não havendo decréscimo abrupto de $L(t)$ que indique escala temporal destacada, que possa remeter a periodicidade marcante.

Outra conclusão importante da análise de auto-afinidade é que uma região com periodicidade dominante nem sempre possui escala de auto afinidade pequena. YANO & NISHI (1989) destacam que a região **C** em particular, estendendo-se de 80° E a 160° E ao longo do equador, correspondendo, a grosso modo, à região de atividade da oscilação de 30-60 dias. Ainda segundo os autores, a VIT se estende longitudinalmente às regiões **B** e **D**, nas quais τ_s é grande.

Outro relato importante dos autores sobre a relação entre a dinâmica e a hierarquização de escalas é que a região **D** parece ser de interação entre El Niño e VIT. A propagação para leste da VIT geralmente se limita a oeste da linha da data, porém penetra esporadicamente no Pacífico leste, o que pode ser uma componente sazonal. Isso, ocasionalmente desencadeia um evento ENSO. Pode-se especular que a interação entre ENSO e VIT leve a auto-afinidade mais pronunciada. Algo similar pode ocorrer na região **B**, por interação entre o ciclo anual e a VIT.

Ainda pode-se especular se a distribuição não homogênea de auto-afinidade em regiões de atividade da oscilação de 30-60 dias está relacionada a possíveis variações ao longo de sua propagação para leste. De fato, YANO & NISHI (1989), mostraram, ao interpolar a OLR em seção transversal tempo-longitude, que tanto a oscilação de 30-60 dias como os superaglomerados de nuvens ocasionalmente se enfraquecem ao redor dos 120° E (região **C**), com os superaglomerados terminando nessa região de τ_s relativamente elevada.

Apesar das precauções requeridas ao usar-se τ_s , o desvio abrupto de $L(t)$ sugere uma escala característica de 20 dias para os superaglomerados de nuvens.

c) Ruído 1/f e eventos tipo pulso na variabilidade da atmosfera tropical

Na discussão sobre a variabilidade na atmosfera tropical anterior, vimos que a variabilidade auto-afim é construída por relação com taxa constante de redução das amplitudes conforme crescem as frequências, possuindo assim um espectro de potência proporcional a expoente negativo α (constante e positivo) da frequência, comportamento descrito pela relação:

$$S(f) \propto f^{-\alpha} \quad (\text{II-8})$$

Tal característica confirmou-se no estudo dos espectros de OLR apresentado em YANO & NISHI (1989).

No caso do estudo da OLR na atmosfera tropical, espectros com variabilidade auto-afim a grosso modo, mas com picos característicos em algumas regiões foram revelados: sazonalidade do El Niño, sazonalidade do Atlântico, oscilações de 30-60 dias e superaglomerados foram identificados, ou no estudo dos espectros, ou no estudo da relação entre o comprimento de variabilidade e a janela temporal, ou no mapa de contornos de escala de auto-afinidade da Figura II-1.

O comportamento misto entre auto-afinidade e a presença de eventuais picos característicos é observado em diversos outros estudos meteorológicos citados em YANO *et al.* (2004), contrariando a visão tradicional da meteorologia que presume escalas características para os fenômenos individuais. Assim, a descrição mais precisa da variabilidade auto-afim torna-se justificável no contexto da presente pesquisa.

Uma característica marcante é $L(t)$ se ajustando à equação (II-5) com dimensão fractal $D \approx 2$ para janelas temporais menores que a escala temporal auto-afim τ_s . Nesse caso a equação (II-7) fornece $\alpha \approx 1$, apesar dessa relação não ser tão evidente nos espectros de OLR. Esse comportamento aproximadamente de ruído 1/f (f sendo a frequência) é achado

por YANO *et al.* (2004) em temperaturas superficiais tropicais, umidade, velocidade do vento (para períodos entre 1 hora e 10 dias) e tensão de atrito com o vento (até períodos de 2 anos).

A origem do ruído $1/f$ pode ser relacionada à presença de eventos do tipo pulsante com natureza altamente intermitente, sendo comuns na variabilidade atmosférica e localizados no tempo e no espaço. Apesar de não serem pulsos tão estreitos, tais fenômenos atmosféricos são mais bem caracterizados como pulsos do que como ondas.

Caracterização $1/f$

As características únicas do ruído $1/f$ ficam evidentes quando se comparam séries possuindo espectro de potência $\sim f^{-\alpha}$ com diferentes expoentes: em um ruído branco teríamos $\alpha = 0$; em um movimento Browniano, teríamos $\alpha = 2$.

Séries que possuem frequências dominantes, com coerência de fase não possuem, necessariamente, as características apresentadas acima. Mas tais séries não seriam consideradas ruídos. YANO *et al.* (2004) utilizaram uma medida da variabilidade para uma dada escala temporal Δt para evidenciar as características $\sim f^{-\alpha}$ em séries temporais reais:

$$\left\langle \left| \overline{x(t + \Delta t)}^{\Delta t} - \overline{x(t)}^{\Delta t} \right|^2 \right\rangle \quad (\text{II-9})$$

Sendo $\langle * \rangle$ referente à promediação ao longo de toda a extensão da série e a sobrebarra referente à média móvel na janela Δt . A Figura II-2 ilustra a variabilidade versus janela temporal em escalas logarítmicas para diversos valores de α .

Percebe-se que a variabilidade do ruído branco se reduz com o aumento do intervalo de tempo Δt . Por outro lado, o ruído Browniano apresenta aumento de variabilidade para janelas de promediação maiores, revelando não-estacionariedade em oposição à estacionariedade do ruído branco. O ruído $1/f$ apresenta características de uma mistura do

ruído branco e do Browniano, retendo o mesmo grau de variabilidade em todas as escalas temporais, com períodos dominados por pulsos incorporados a outros com menos ruídos.

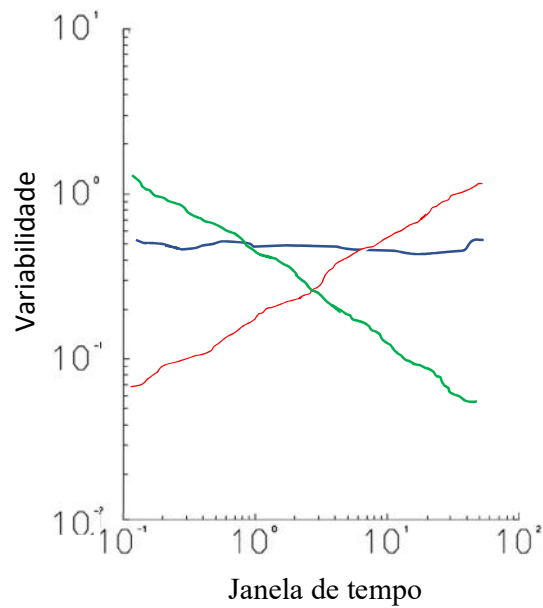


Figura II-2. Variabilidade como função da janela temporal para ruído branco (linha verde), ruído $1/f$ (azul) e ruído Browniano (vermelha). Adaptado de YANO *et al.* (2004)

YANO *et al.* (2004) adaptaram dados meteorológicos tropicais de superfície da base TOGA-COARE, com período de novembro de 1992 a fevereiro de 1993, fornecidos pelo Centro de Processamento/ Arquivamento e Distribuição de Dados da Universidade do Estado da Califórnia – COARE-MET, por site eletrônico¹. Tais dados haviam sido usados previamente em YANO *et al.* (2001) na caracterização da variabilidade convectiva tropical como um ruído $1/f$, com intervalos de 6-12 horas e foram adaptados para escalas mais longas e mais curtas e usados em nova caracterização espectral da variabilidade de temperatura T , umidade específica q , velocidade do vento $|\mathbf{V}_H|$ e a taxa de precipitação P .

Para estender a análise a escalas mais longas, os autores usaram também ventos superficiais de bases oriundas de boias no equador, a 165° E de 1987 a 1993, além de dados de 1992 a 1999 da mesma região, oriundos de outras bases.

¹ pelo endereço <http://www.coaps.fsu.edu/COARE>

Os resultados da análise espectral de dados em escalas mais curtas de medição (1min de intervalo) de duas estações específicas foram apresentados para temperatura, umidade relativa, velocidade do vento e taxa de precipitação. A declividade $1/f$ foi observada nos espectros, com exceção no de taxa de precipitação, desde a escala temporal horária (0,5-5 horas), até a escala intrasazonal (30-60 dias).

O espectro de precipitação apresentou-se com declividade mais suave, característica de ruído branco. Apesar da precipitação ser consequência direta da atividade convectiva dos cúmulos, estações isoladas não representam corretamente a precipitação de um sistema convectivo. O caráter comumente localizado da precipitação pode exigir promediação espacial apropriada para revelar melhor seu caráter $1/f$, ressaltaram YANO *et al.* (2004).

Comportamento pulsante

Os eventos convectivos de comportamento pulsante aparecem incorporados a um cenário de fundo de fase calma. Tal cenário se aproxima de equilíbrio permanente nas escalas temporais mais curtas dos dados da meteorologia superficial tropical analisados por YANO *et al.* (2004). Esses pulsos são caracterizados por quedas repentinas na umidade e na temperatura da camada limite ocasionadas por correntes convectivas descendentes, sendo desencadeados de forma intermitente, seguindo a característica geral de ruído $1/f$. Eles apareceram associados a rajadas de vento e precipitação durante períodos de convecção ativa nas séries do navio de pesquisa Kexue.

Nem todas as faixas de eventos pulsantes de baixa duração estudadas em YANO *et al.* (2004) podem ser identificadas nas séries de interesse dessa pesquisa, pois os dados de TSM são médias mensais e a série hidrológica de Manaus é apresentada em valores de níveis médios diários. A Figura II-3 ilustra bem alguns eventos convectivos reproduzidos em dados meteorológicos superficiais, mostrando alterações em correntes horizontais associadas a quedas repentinas de temperatura e de umidade, bem como a eventos de chuva. A relevância da apresentação de tais resultados nesta obra se justifica, mesmo que os pulsos específicos presentes nos trechos ilustrados nessa figura possuam escalas menores que a resolução das séries de interesse dessa pesquisa, pois os eventos pulsantes aparecerão em outras faixas de frequência de interesse e, conforme relatado por YANO & NISHI (1989), distúrbios sinóticos de 10-20 dias, aglomerados de nuvens convectivas se fundem em superaglomerados e influenciam a VIT.

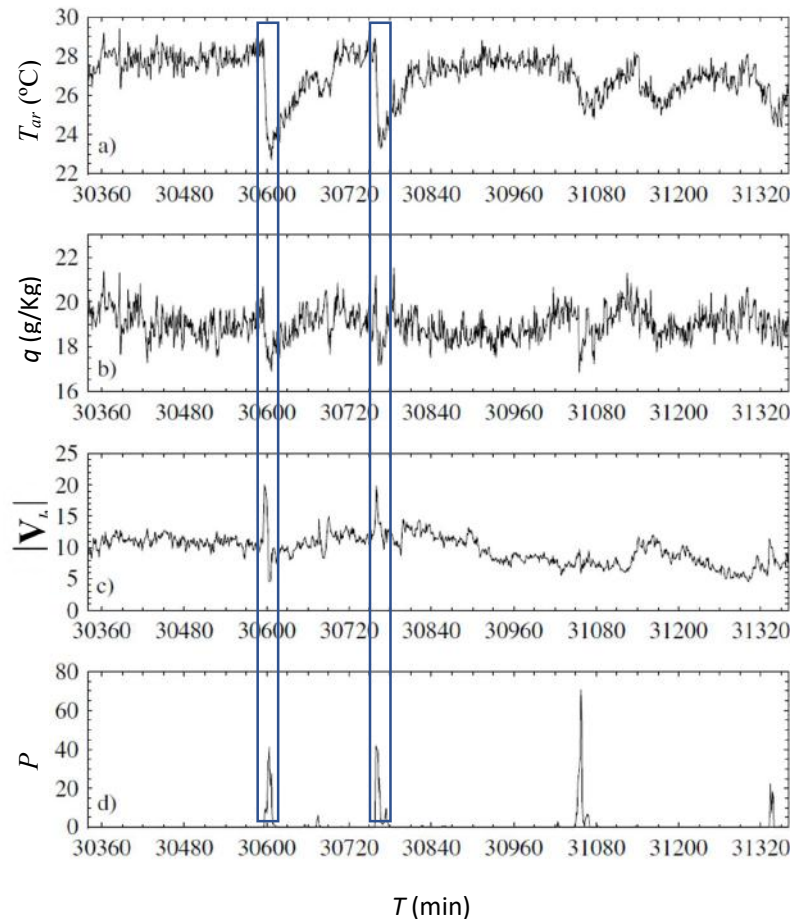


Figura II-3. Trecho de série temporal superficial do navio Kexue mostrando dois eventos de convecção caracterizados por pulsos nos instantes $t = 30.600$ min e $t = 30.750$ min nas séries de temperatura do ar T_{ar} (°C), umidade relativa q (g/Kg), velocidade do vento $|v_v|$ (m/s) e precipitação P (mm/h). Nota-se a relação de quedas repentinas na temperatura e umidade com pulsos nas séries de vento e de precipitação. Adaptado de YANO *et al.* (2004).

Para avaliar a contribuição desses eventos pulsantes associados à convecção YANO *et al.* (2004) aplicaram um método de extração de pulsos às séries meteorológicas superficiais. O método se vale da estrutura localizada das ondeletas tipo Meyer para extrair pulsos em escalas específicas das séries. Os pulsos são indicados por máximos locais nos valores absolutos dos coeficientes no espaço Wavelet, lembrando que a decomposição Wavelet pode ser interpretada como uma correlação entre o sinal e a ondeleta.

Segundo os autores, a natureza da decomposição permite que os eventos sejam representados pelas suas escalas características e durações. Quando um máximo é

localizado, os elementos no espaço Wavelet adjacentes são extraídos junto com o modo no qual se localiza o máximo. A identificação de um pulso no espectro Wavelet é ilustrada na Figura II-4.

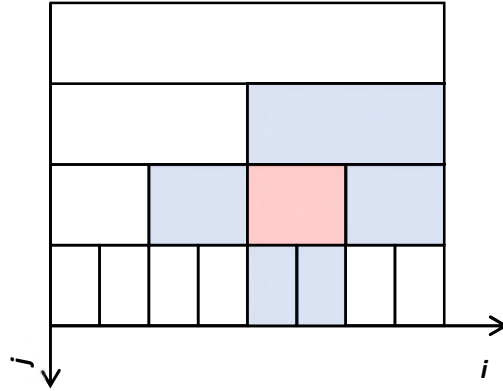


Figura II-4. Localização de um pulso pelo elemento principal (em vermelho) e pelos elementos adjacentes (em azul) no espectro Wavelet discretizado pela escala característica de duração dos eventos (eixo j) e pelo endereçamento do evento no tempo (eixo i). Adaptado de YANO *et al.* (2004).

A extração de eventos pulsantes de escalas arbitrariamente definidas de 0,5-4 horas foi realizada para as séries da Figura II-3 por YANO *et al.* (2004). Os pulsos encontrados recompuseram as séries mostradas na Figura II-5. Os eventos destacados em azul na Figura II-3 foram bem reproduzidos. O ruído de menores escalas foi eliminado da série de temperatura do ar.

Segundo os autores, o caráter pulsante menos evidente nas séries de umidade relativa (q) e de velocidade do vento $|\mathbf{V}_h|$ refletiram-se em séries recompostas mais ruidosas para esses parâmetros na Figura II-5. Apesar disso, ainda se pode perceber o caráter pulsante associado aos eventos convectivos. Apenas um dos eventos de precipitação foi identificado, por outro lado. Ele localiza-se por volta de $t = 31.080$ min. Todos os outros eventos de precipitação não são resolvidos pela escala definida de 0,5-4hs.

Os espectros de potência das séries de ondeletas extraídas de escala temporais 0,5-4hs foram apresentados por YANO *et al.* (2004). Observou-se potência em bandas muito mais largas que as da extração, tanto em altas como em baixas frequências. Essa contribuição

espectral é possível, segundo os autores, porque os aglomerados de ondeletas retiradas estendem-se muito além das escalas de extração.

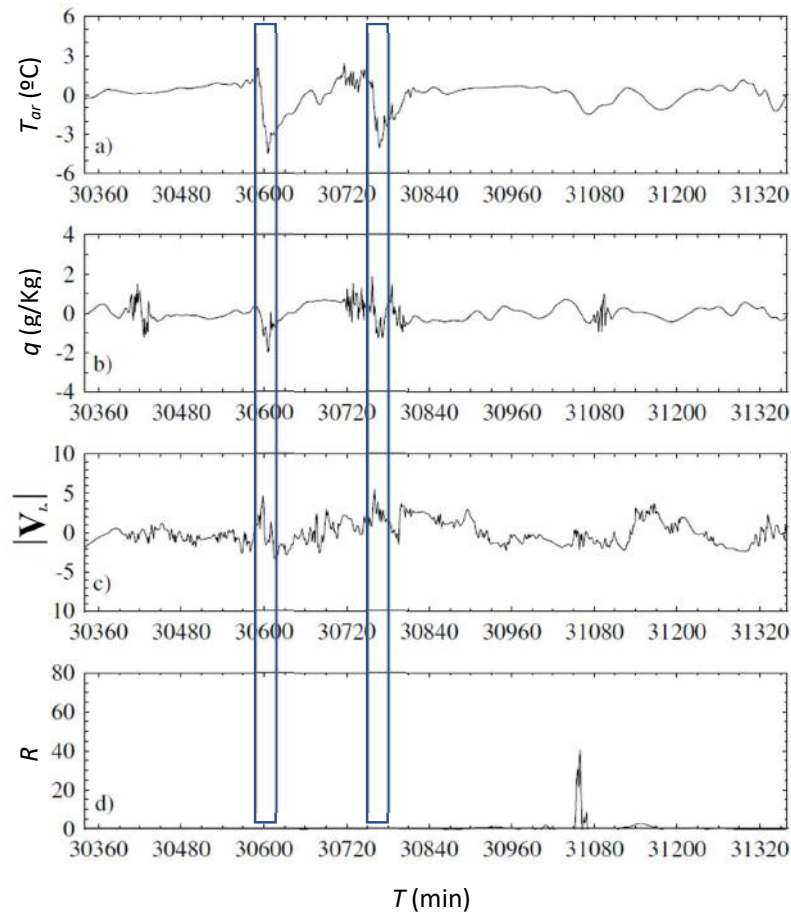


Figura II-5. As mesmas séries da Figura II-3 recompostas com os eventos tipo pulso de escala 0,5-4 hs.

Os espectros resultantes se aproximam dos originais para frequências maiores que $0,1 \text{ h}^{-1}$, tendo-se como principais conclusões os fatos de serem espectros muito mais abrangentes em frequências que os pulsos extraídos e que seguem, de forma geral, formato de ruído $1/f$ para períodos de 1-10hs para as três primeiras variáveis. Os autores especularam que tais eventos pulsantes associados à atividade convectiva constituem a base do ruído $1/f$.

Para verificar se o mesmo ocorre em escalas temporais maiores, o procedimento de extração foi repetido na escala 32-128hs por YANO *et al.* (2004). Os resultados foram espectros com bandas ainda mais largas, reforçando a especulação dos autores, de serem

os eventos pulsantes a base do ruído $1/f$, apesar das declividades espectrais apontarem para valores de $\alpha > 1$, especialmente para a umidade relativa. Tais resultados são importantes na caracterização geral da atmosfera tropical e no estabelecimento de relações fenomenológicas entre os dados usados na presente pesquisa para modelar a hidrologia em Manaus.

A correlação de eventos tipo pulso contendo defasagens foi estudada pelos autores. Os resultados mostraram não haver correlação entre os eventos. Por falta de correlação o espectro da extração pode ser representado pela soma dos espectros de eventos individuais. Ou seja, os eventos individuais também formam espectros $1/f$, constituindo ingredientes individuais para a formação do espectro total.

Escalas temporais mais longas

Para analisar escalas mais longas, YANO *et al.* (2004) utilizaram dados zonais de tensão de atrito do vento no equador a 165° E. Os dados horários originais, com acurácia de 0,2 m/s em módulo e $0,3^\circ$ em direção, foram promediados em janelas de 5 dias, totalizando 512 valores em aproximadamente 7 anos, de 1987 a 1993. Também foram examinados dois conjuntos de dados adicionais da região de 160° E de 1992 a 1999, um consistindo em médias de cinco dias e outro de médias semanais de ventos superficiais zonais.

O espectros de tensão de atrito do vento obtido por YANO *et al.* (2004) também revelou o caráter $1/f$, relacionado à auto-afinidade em escalas temporais intrasazonais. O espectro assemelha-se ao ruído $1/f$ de 10 a 500 dias de período, apesar de alguns picos se apresentarem. Os dados de ventos zonais também apresentaram tais características, com picos menos pronunciados, no entanto.

Os autores demonstraram que o comportamento tipo pulsante também domina a dinâmica tropical em escalas intrasazonais, estando associado a ventos de oeste da oscilação de Madden-Julian – OMJ. A extração de tais eventos foi realizada por YANO *et al.* (2004) em escalas de 20-160 dias e 40-160 dias. Os autores apresentam ainda os espectros resultantes da extração na escala 40-160 dias, destacando o fato surpreendente que a estrutura do espectro da série original seja preservada após a extração, mesmo até 50 dias, apenas perdendo variância em alguns intervalos.