



**Luiz Albino Teixeira Júnior**

**Combinação SSA-Wavelet de Métodos Preditivos com  
Ajuste Numérico MINIMAX, na Geração de Previsões e de  
Cenários**

**Tese de Doutorado**

Tese apresentada como requisito parcial para  
obtenção do grau de Doutor pelo Programa de Pós-  
graduação em Engenharia Elétrica do Departamento  
de Engenharia Elétrica da PUC-Rio.

Orientador: Prof. Reinaldo Castro Souza

Rio de Janeiro

Maio de 2013





**Luiz Albino Teixeira Júnior**

**Combinação SSA-Wavelet de Métodos Preditivos com Ajuste Numérico MINIMAX, na Geração de Previsões e de Cenários**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

**Prof. Reinaldo Castro Souza**  
**Orientador**

Departamento de Engenharia Elétrica – PUC-Rio

**Prof. Carlos Kubrusly**

Departamento de Engenharia Elétrica – PUC-Rio

**Profa. Marley Maria Bernardes Rebuzzi Vellasco**

Departamento de Engenharia Elétrica – PUC-Rio

**Prof. José Francisco Moreira Pessanha**

UERJ

**Prof. André Luís Marques Marcato**

UFJF

**Prof. José Eugenio Leal**

Coordenador Setorial do Centro  
Técnico Científico

Rio de Janeiro, 20 de maio de 2013



Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

### **Luiz Albino Teixeira Júnior**

Possui curso técnico em Informática Industrial, no Centro Federal de Educação Tecnológica - CEFET - (2003-2004); graduação em Matemática, nas Faculdades Integradas de Cataguases - FIC - (2002-2005); especialização em Métodos Estatísticos Computacionais, na Universidade Federal de Juiz de Fora - UFJF - (2006); mestrado em Engenharia Elétrica, com Ênfase em Métodos de Apoio à Decisão, na Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio - (2007-2009); e doutorado em Engenharia Elétrica, com Ênfase em Métodos de Apoio à Decisão, na Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio. À título de doutorado sanduiche, de fevereiro de 2012 até agosto de 2012, estagiou e desenvolveu pesquisas referentes à sua linha de pesquisa de doutorado, na Inglaterra, no Departamento de Matemática, Estatística e Computação, na The Open University - OU -, sob a orientação do professor Dr. Álvaro Eduardo Faria Júnior. Desde de setembro de 2009, faz parte de um grupo de pesquisa cujo nome é "Modelagem Estatística e Séries Temporais; Aplicações na Área de Energia", na PUC-Rio, o qual é coordenado pelo Dr. Reinaldo Castro Souza, atuando com linha de pesquisa em "Análise de Séries Temporais". Atualmente atua como professor assistente de Estatística, na Universidade da Integração Latino Americana (UNILA), em Foz do Iguaçu (PR). Tem interesse em previsão, geração de cenários e análise de séries temporais utilizando, individualmente ou de forma integrada, as seguintes metodologias: Métodos Estatísticos de Previsão, Inteligência Artificial, Combinação Híbrida de Métodos Preditivos, Teoria Wavelet, Singular Spectrum Analysis, Programação Matemática e Métodos de Amostragem.

### **Ficha Catalográfica**

Teixeira Júnior, Luiz Albino

Combinação SSA-Wavelet de métodos preditivos com ajuste numérico MINIMAX, na geração de previsões e de cenários / Luiz Albino Teixeira Júnior; orientador: Reinaldo Castro Souza. – 2013.

114 f. ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2013.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Método SSA. 3. Decomposição Wavelet. 4. Redes neurais artificiais. 5. Programação matemática multiobjetivo MINIMAX, com abordagem de programação por metas. 6. Método de simulação de quase-Monte-Carlo. 7. Previsões. I. Souza, Reinaldo Castro. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3



## Agradecimentos

Agradeço, primeiramente, a Deus, pois é o responsável maior por chegar até aqui; à minha esposa; à minha mãe; e à minha família como um todo, por todo apoio, amor e suporte em todos os momentos.

Quero expressar também a minha enorme gratidão ao meu orientador e amigo Reinaldo Castro Souza, bem como a todos os professores do Departamento de Engenharia Elétrica. Mui especialmente, agradeço demais ao meu eterno professor e amigo Carlos Silva Kubrusly por tudo, bem como ao amigo e companheiro José Francisco Pessanha. Agradeço também ao CNPq, pelo suporte financeiro, e ao professor Álvaro Eduardo Faria Júnior, por ter me orientado durante o meu estágio na Inglaterra, na *The Open University*.

Por fim, agradeço aos meus amigos: Ricardo Vella, Márcio Leone, Rafael, Álvaro Albuquerque, Iuri Steiner (o polêmico), Alcina, Ana Paiva, Márcia Guerra, dentre outros que, de forma direta ou indireta, muito me ajudaram. Deixo a todos, portanto, os meus enormes e sinceros agradecimentos por suas participações, neste processo, tão importante na minha vida.



## Resumo

Teixeira Júnior, Luiz Albino; Souza, Reinaldo Castro (Orientador). **Combinação SSA-Wavelet de Métodos Preditivos com Ajuste Numérico MINIMAX, na Geração de Previsões e Cenários**. Rio de Janeiro, 2013. 114p. Tese de Doutorado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Nesta tese de doutorado, é proposta uma combinação híbrida de métodos preditivos que agrega cinco abordagens distintas e genéricas, do ponto de vista de modelagem: método SSA; decomposição *wavelet*; redes neurais artificiais; programação matemática multiobjetivo MINIMAX, com abordagem de programação por metas; e método de simulação de quase Monte-Carlo. Para exemplificar e demonstrar a eficiência da combinação híbrida proposta, são mostrados, no Capítulo 7, os principais resultados de uma aplicação computacional, no qual é possível verificar que o seu desempenho, em termos de modelagem, foi consideravelmente superior, em relação a todas as estatísticas de aderência consideradas.

## Palavras-chave

Método SSA; decomposição *wavelet*; redes neurais artificiais; programação matemática multiobjetivo MINIMAX, com abordagem de programação por metas; método de simulação de quase-Monte-Carlo; previsões.



## Abstract

Teixeira Júnior, Luiz Albino; Souza, Reinaldo Castro (Orientador). **SSA-Wavelet Combination of Predictive Methods with MINIMAX Numerical Adjustment in Forecast and Scenarios Generation**. Rio de Janeiro, 2013. 114p. PhD Thesis - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

In this thesis, we propose a hybrid combination of predictive methods that aggregates five distinct and general approaches, from the viewpoint of modeling: SSA method; wavelet decomposition, artificial neural networks, multiobjective mathematical programming MINIMAX, with goal programming approach; quasi-Monte-Carlo simulation method. To exemplify and demonstrate the efficiency of the proposed hybrid combination are shown, in Section 7, the main results of a computer application in which you can verify that their performance, in terms of modeling, was significantly higher, compared to all considered adherence statistics.

## Keywords

SSA method; wavelet decomposition; artificial neural networks; MINIMAX multiobjective mathematical programming, with goal programming approach; quasi-Monte-Carlo simulation method; forecasts.



# SUMÁRIO

<b>1. INTRODUÇÃO</b>	<b>12</b>
1.1. Objetivo e Justificativas	15
1.2. Visão Geral da Metodologia Proposta	19
1.3. Estrutura da Tese	21
<b>2. ABORDAGENS DE PRÉ-PROCESSAMENTO DE SINAIS</b>	<b>22</b>
2.1. Método SSA	22
2.1.1. Decomposição	22
2.1.2. Reconstrução	25
2.2. TEORIA WAVELET	28
2.2.1. Vetores <i>Wavelet</i>	28
2.2.2. Transformada <i>Wavelet</i>	29
2.2.3. Expansão <i>Wavelet</i>	30
2.2.4. Funções <i>Wavelet</i>	32
<b>3. REDES NEURAIS ARTIFICIAIS</b>	<b>36</b>
3.1 Componentes Básicas de uma RNA	38
3.1.1 Funções de Ativação	38
3.1.2 Arquitetura	40
3.1.3 Treinamento Neural	41
3.2. Normalização dos Padrões	42
3.3. RNA's Multicamadas <i>Feedforward</i> , com uma Camada Escondida	44
3.4. Geração de Previsões via Redes Neurais Artificiais	47
<b>4. COMBINAÇÃO DE MÉTODOS PREDITIVOS VIA PROGRAMAÇÃO MATEMÁTICA</b>	<b>48</b>
4.1. Combinação de Métodos Preditivos	48
4.1.1. Combinação Linear de Métodos Preditivos	49
4.1.1.1. Combinação Linear de Previsões Tradicional	50
4.2. Constantes e Pesos Adaptativos	51
4.3. Ajuste Numérico via Programação Matemática	53
4.3.1. Programação Matemática	53
4.3.2. PPM Multiobjetivo MINIMAX	55
4.3.3. Programação Multiobjetivo MINIMAX, com Abordagem de Programação por Metas	56
4.3.4. Ajuste Numérico via PPM Multiobjetivo da Combinação de Previsões Proposta	57
<b>5. MÉTODO DE SIMULAÇÃO DE QUASE-MONTE-CARLO</b>	<b>59</b>
5.1. Sequências de Números Pseudo-Aleatórios	59
5.2. Sequência de Números Quase-Aleatórios	60
5.2.1. Método de Simulação de Quase-Monte-Carlo Híbrido	62
5.3. Geração de Amostras de Distribuição Normal-Padrão	64
5.3.1. Geração de Choques Estocásticos de Distribuição Normal	67



<b>6. METODOLOGIA PROPOSTA</b>	<b>69</b>
6.1. Filtragem de uma Série Temporal via Método SSA	69
6.2. Decomposição <i>Wavelet</i> de Nível $p$ de uma Série Temporal	69
6.3. Combinação das Previsões para as Componentes <i>Wavelet</i>	71
6.4. Combinação Linear de Previsões Proposta com Ajuste Numérico via PPM Multiobjetivo MINIMAX, com Abordagem de Programação por Metas	73
6.5. Geração de Cenários QMC Híbridos	75
<b>7. EXPERIMENTOS COMPUTACIONAIS</b>	<b>78</b>
7.1. Filtragem da Série Temporal via Método SSA	79
7.2. Decomposição <i>Wavelet</i> da Série Temporal Filtrada via Método SSA	83
7.3. Modelagem das Componentes <i>Wavelet</i>	87
7.4. Combinação Linear das Previsões das Componentes <i>Wavelet</i>	88
7.5. Geração dos Cenários QMC Híbrido e dos Intervalos de Confianças	94
<b>8. CONCLUSÕES</b>	<b>99</b>
<b>TRABALHOS FUTUROS</b>	<b>104</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>106</b>



## LISTA DE FIGURAS

<b>Figura 1.1</b> - Ilustração das etapas da metodologia proposta	20
<b>Figura 2.1</b> - Gráfico das 3 componentes SSA totais	27
<b>Figura 2.2</b> - Funções <i>wavelet sym 20</i>	34
<b>Figura 2.3</b> - Decomposição <i>wavelet</i> de nível 1, com a base ortogonal <i>wavelet sym 20</i> , da série temporal mensal de vazão de afluentes da região sul do Brasil	34
<b>Figura 3.1</b> - Esquematização simplificada de um neurônio biológico	36
<b>Figura 3.2</b> - Arquitetura básica de um neurônio artificial	38
<b>Figura 3.3</b> - RNA MLP <i>feedforward</i> .	46
<b>Figura 3.4</b> - Ilustração da dinâmica de previsão, 1 passo à frente, de uma série temporal, com estrutura de autodependência igual a 4, através de uma RNA.	47
<b>Figura 5.1</b> - Pares ordenados de números quase-aleatórios híbridos das dimensões 49 (ordenada) e 50 (abscissa) sobre $[0,1]^2$ .	64
<b>Figura 5.2</b> - Ilustração dos $N$ choques estocásticos filtrados via fatoração de <i>Cholesky</i> para cada uma das $D$ dimensões.	68
<b>Figura 7.1</b> - Série temporal mensal da vazão de afluentes da região Sul do Brasil.	78
<b>Figura 7.2</b> - 9 Primeiros autovetores singulares na SVD da matriz trajetória da amostra de treino da série temporal mensal de vazão de afluentes da região sul do Brasil.	79
<b>Figura 7.3</b> - Gráficos contínuos de dispersão dos pares de autovalores singulares da SVD: (2,3); (9,10); (10,11); e (14,15).	80
<b>Figura 7.4</b> - Componente SSA 3 da amostra de treino inicial série temporal mensal de vazão de afluentes da região sul do Brasil.	82
<b>Figura 7.5</b> - Funções <i>wavelet db 10</i> .	84
<b>Figura 7.6</b> - Funções <i>wavelet db 40</i> .	84
<b>Figura 7.7</b> - Funções <i>wavelet db 30</i> .	85
<b>Figura 7.8</b> - Componentes <i>wavelet</i> de aproximação e de detalhe da decomposição de nível 2, com base ortonormal <i>db 10</i> , da série temporal total mensal filtrada.	85
<b>Figura 7.9</b> - Componentes <i>wavelet</i> de aproximação e de detalhe da decomposição de nível 2, com base ortonormal <i>db 40</i> , da série temporal total mensal filtrada.	86
<b>Figura 7.10</b> - Componentes <i>wavelet</i> de aproximação e de detalhe da decomposição de nível 2, com base ortonormal <i>db 30</i> , da série temporal total mensal filtrada.	86
<b>Figura 7.11</b> - Valores de MAPE dos métodos RNA-WAVE 1, RNA-WAVE 2 e RNA-WAVE 3 e método proposto.	91
<b>Figura 7.12</b> - Valores de MAPE dos métodos RNA-WAVE 1, RNA-WAVE 2 e RNA-WAVE 3 e método proposto.	91
<b>Figura 7.13</b> - Dispersão entre as observações e as previsões do método proposto, na amostra de validações (correlação igual a 99,96%).	93
<b>Figura 7.14</b> - Dispersão entre as observações e as previsões do método proposto, na amostra de teste (correlação igual a 99,98%).	93



<b>Figura 7.15</b> - Histograma da série temporal de resíduos associados às previsões geradas pelo método proposto, na amostra de treino.	94
<b>Figura 7.16</b> - Gráfico dos limites inferiores e superiores dos intervalos de confiança (linha vermelha) e das observações (linha azul), na amostra de Validação.	95
<b>Figura 7.17</b> - Histogramas dos cenários QMC híbridos do método proposto, nos instantes 1 e 94, na amostra de validação.	96
<b>Figura 7.18</b> - Histogramas dos cenários QMC híbridos do método proposto, nos instantes 1 e 94, na amostra de teste.	97
<b>Figura 7.19</b> - Gráfico dos limites inferiores e superiores dos intervalos de confiança (linha vermelha) e das observações (linha azul), na amostra de teste.	98



## LISTA DE TABELAS

<b>Tabela 5.1</b> - Propriedades estatísticas dos números quase-aleatórios e pseudo-aleatórios.	61
<b>Tabela 5.2</b> - Números quase-aleatórios do método QMC híbrido das três primeiras dimensões ( $d=1, 2, 3$ ).	63
<b>Tabela 5.3</b> - Valores para $a_n$ e $b_n$ [fonte: KRYKOVA, 2003].	65
<b>Tabela 5.4</b> - Valores de $c_n$ e $k_j$ [fonte: KRYKOVA, 2003].	66
<b>Tabela 5.5</b> - Estatísticas dos choques estocásticos gerados de 1000 amostras simuladas de números pseudo-aleatórios e quase-aleatórios, via inversor de Moro; e parâmetros de uma distribuição de probabilidade normal-padrão (teórica) [fonte: KRYKOVA (2003)].	67
<b>Tabela 7.1</b> - Teste de independência estatística BDS (onde: $H_0$ : “independência estatística” e $H_1$ : “ $H_0$ é falsa”) para a componentes SSA 3.	83
<b>Tabela 7.2</b> - Testes de estacionariedade de <i>Dickey-Fuller</i> (onde $H_0$ : “não estacionariedade de 2º ordem” e $H_1$ : “ $H_0$ é falsa”) para a componentes SSA 3.	83
<b>Tabela 7.3</b> - Configurações das 9 RNA ‘s <i>feedforward</i> , com uma camada escondida, utilizadas na modelagem das 9 componentes <i>wavelet</i> .	88
<b>Tabela 7.4</b> - Parâmetros ótimos da combinação linear proposta.	89
<b>Tabela 7.5</b> - Configurações das RNA’ s: RNA 1, RNA 2, RNA 3, RNA-WAVE 1, RNA-WAVE 2 e RNA-WAVE 3.	90
<b>Tabela 7.6</b> - Valores de MSE dos 9 métodos preditivos implementados.	90
<b>Tabela 7.7</b> - Teste BDS para a série temporal de resíduos associados às previsões geradas pelo método proposto, na amostra de treino.	93
<b>Tabela 7.8</b> - Amplitudes médias dos intervalos de confiança dos 6 métodos preditivos, na amostra de validação e de teste.	96



## INTRODUÇÃO

Ao longo dos anos, diversos *métodos preditivos* (ou *previsores*) vêm sendo propostos com o objetivo de se projetar séries temporais (estocásticas). Em linhas gerais, podem-se agrupá-los em três classes: os estatísticos (como, por exemplo, os *modelos de Box & Jenkins* [veja SOUZA & CAMARGO (2004)]), os de inteligência artificial (com, por exemplo, as *Redes Neurais Artificiais* [veja HAYKIN (2001)]) e os de combinação (que podem ser, basicamente, classificados como *linear* ou *não linear* [veja MUBWANDARIKWA (2007)]). Existe uma gama de *métodos preditivos individuais* (ou *base*) - como, por exemplo, os modelos de *Box & Jenkins* e as *Redes Neurais Artificiais* - consagrados na Literatura que podem ser utilizados na projeção de séries de tempo. Não obstante, na maioria das vezes, os *processos estocásticos* [veja HAMILTON (1994)] associados às séries temporais são *não estacionários* [veja HAMILTON (1994)] - e, por vezes, *não diferenciáveis* [veja SOUZA & CAMARGO (1996)] - ou excessivamente *ruidosos* [veja GUTIERREZ (2002), CASSIANO *et al* (2012)], de modo que, de acordo, dentre outros, com WERNER & RIBEIRO (2006), FLORES & WHITE (1989) e FARIA & MUBWANDARIWA (2008), uma abordagem mais segura e eficiente, no processo de extrapolação preditiva fora da amostra, é a combinação de métodos preditivos individuais. É importante salientar que o termo “*combinação de métodos preditivos individuais (ou base)*” é utilizado, na Literatura, em sentido amplo, ao passo de se referir tanto à combinação de previsões ou de cenários (abordagem clássica), quanto à de *densidades preditivas bayesianas* [veja MUBWANDARIKWA (2007)] (abordagem bayesiana). Nesta tese de doutorado, no entanto, é tratado apenas no sentido de combinação de previsões e de cenários.

Em BATES & GRANGER (1969), que é considerado o artigo seminal sobre o tema, propõe-se a combinação linear de previsões oriundas de dois *modelos estatísticos clássicos* [veja HAMILTON (1994)], conforme a equação em (1.1), abaixo. Em NEWBOLD & GRANGER (1974), ampliou-se a abordagem BATES & GRANGER (1969), propondo-se a combinação linear de previsões



oriundas dos métodos preditivos individuais *Box & Jenkins*, *Holt-Winters* e *Auto-Regressão Stepwise* [veja MORETTIN (2005), HAMILTON (1994)]. Os autores concluíram que as previsões linearmente combinadas, em regra, eram mais seguras que as previsões oriundas dos referidos modelos estatísticos base. Em WINKLER & MAKRIDAKIS (1983), foram realizadas combinações lineares, postas três a três, de previsões de dez modelos estatísticos clássicos e os resultados obtidos confirmaram as conclusões de NEWBOLD & GRANGER (1974). De acordo com GRANGER (1989), *“a combinação de previsões é uma forma simples e pragmática que possibilita a produção de melhores previsões”*. De acordo com HOLLAUER *et al* (2008), algumas razões que motivam a combinação de métodos preditivos base: *“Primeiro, a pura diversificação de previsões leva à diminuição do erro diversificável. A segunda razão concerne à robustez, na medida em que, não sendo particular a especificação, não padece plenamente também de suas deficiências, além de incluir mais variáveis e informações de especificação de outros modelos, uma vez que se pesam informações de modelos que incluem outras variáveis e seus lag’s. Ademais, a experiência mostra que modelos parcimoniosos combinados possuem desempenho superior”*. FARIA & MUBWANDARIKWA (2008) salientam que a previsão combinada é uma agregadora de informações oriundas de diferentes fontes (no caso, os métodos preditivos base) sobre a flutuação estocástica da série temporal a ser modelada. Em WALLIS (2011), é feita uma revisão histórica, na qual são mencionados muitos dos principais artigos sobre combinação de métodos preditivos individuais, publicados em quarenta e dois anos, os quais atestam a sua utilização e a sua eficiência, além de mostrarem uma enorme diversidade de aplicações e abordagens.

A previsão linearmente combinada, no instante  $t$ , proposta no artigo de BATES & GRANGER (1969) é dada conforme a equação em (1.1).

$$\hat{y}_{CL,t} = \hat{y}_{1,t} \times \rho_1 + \hat{y}_{2,t} \times \rho_2 \quad (1.1)$$

Onde:  $\hat{y}_{CL,t}$  é a previsão linearmente combinada, em  $t$ ;  $\hat{y}_{1,t}$  e  $\hat{y}_{2,t}$  são, respectivamente, as previsões, em  $t$ , oriundas dos métodos preditivos individuais 1 e 2; e  $\rho_1$  e  $\rho_2$  são os pesos adaptativos lineares (a serem estimados ou



otimizados). Os valores ótimos para os pesos adaptativos  $\rho_1$  e  $\rho_2$  (denotados, respectivamente, por  $\hat{\rho}_1$  e  $\hat{\rho}_2$ ) podem ser obtidos por meio de um *problema de programação matemática* [veja RADSGALE (2004), Capítulo 4] cujo objetivo é a minimização da estatística MSE (*mean square error*), dentro da amostra, entre as previsões linearmente combinadas  $\{\hat{y}_{CL,t}\}_{t=1}^T$  e as observações da série temporal  $\{y_t\}_{t=1}^T$  de interesse, onde  $T$  é a *cardinalidade* [veja KUBRUSLY (2001)] dos conjuntos  $\{\hat{y}_{CL,t}\}_{t=1}^T$  e  $\{y_t\}_{t=1}^T$  [veja ainda TEIXEIRA Jr *et al* (2011a)]. Em outras palavras, os valores ótimos  $\hat{\rho}_1$  e  $\hat{\rho}_2$  consistem na solução do objetivo

$$\underset{\rho_1, \rho_2}{Min} \left( \frac{\sum_{t=1}^T (\hat{y}_{CL,t} - y_t)^2}{T} \right), \text{ onde } \rho_1 \text{ e } \rho_2 \text{ são as variáveis de decisão [veja Capítulo$$

4] a serem otimizadas e “Min” significa “mínimo” [veja ainda TEIXEIRA Jr (2009)].

Ainda neste contexto, em TEIXEIRA Jr *et al* (2011a), mostra-se que a programação matemática é extremamente útil, na medida em que torna factível, dentre outros, os seguintes aspectos: (i) a otimização simultânea de várias estatísticas de aderência, por meio da *programação matemática multiobjetivo* [veja Capítulo 4]; (ii) o teste de diferentes *algoritmos de otimização* [veja ZIONTS & WALLENIS (1976)]; (iii) uma maior operacionalidade na combinação dos métodos preditivos base, na medida em que não implica a necessidade de se deduzir fórmulas analíticas para obtenção dos valores dos parâmetros adaptativos - como, por exemplo, os cinco *estimadores* [veja CASSELA & BERGER (2001)] do artigo de NEWBOLD & GRANGER (1974). Nas referências TEIXEIRA Jr *et al* (2011a) e TEIXEIRA Jr *et al* (2012b), mostra-se que as combinação linear de métodos preditivos base, com ajuste numérico dos parâmetros adaptativos via programação matemática, obteve relevantes ganhos preditivos obtidos nas projeções das séries temporais utilizadas, dentro e fora da amostra.



## 1.1

### Objetivo e Justificativas

Diante das referências citadas, no Capítulo 1, bem como do fato de não haver ainda uma combinação de métodos preditivos individuais que seja considerada a melhor na projeção de séries temporais, de acordo com WERNER & RIBEIRO (2006) e WALLIS (2011), propõe-se a “*Combinação SSA-Wavelet de Métodos Preditivos com Ajuste Numérico MINIMAX, na Geração de Previsões e de Cenários*”. A metodologia proposta integra à combinação de métodos preditivos base, que é definida no Capítulo 6, quatro abordagens que incorrem, de forma relevante, de acordo com a literatura e com os experimentos realizados durante a pesquisa, em relevantes ganhos preditivos e que sejam genéricas, no sentido de propiciar à modelagem de qualquer série temporal (estacionária, não estacionária ou excessivamente ruidosa) que apresente estrutura de *autodependência linear ou não linear* [veja MORETTIN (2005), HAMILTON (1994)]: *método SSA, decomposição wavelet, redes neurais artificiais e simulação de quase-Monte-Carlo híbrido*.

Da *Teoria Wavelet* e da *Análise Espectral Singular (Singular Spectrum Analysis - SSA)* [veja Capítulo 2] eivam importantes métodos e abordagens auxiliares que consistem em realizar o pré-processamento (basicamente, *alisamento, filtragem ou decomposição* [veja GOLYANDINA *et al* (2001)]) de uma série de tempo de interesse, antes de sua efetiva modelagem. Os referidos podem ser integrados a algum previsor base (como, por exemplo, uma rede neural artificial), de forma a gerar uma nova metodologia para se projetar séries temporais. Desde que sejam utilizados de forma adequada, são imprescindíveis na obtenção de ganhos preditivos. Tal fato pode ser verificado, dentre muitas outras, nas seguintes referências: DONOHO (1993), DONOHO & JONHSTONE (1994), DONOHO *et al* (1995), BJORN (1995), ELSNER & TSONIS (1996), AUSSSEN & MURTAGH (1997), DAOUDI (1999), ZHENG *et al* (1999), GOLYANDINA *et al* (2001), ALEXANDROV & GOLYANDINA (2004), HASSANI (2007), LEI & RAN (2008), AQUINO *et al* 2009, HASSANI *et al* (2009), HASSANI *et al* (2010), TEIXEIRA Jr *et al* (2011b), MENEZES *et al* (2012), ORTEGA (2012), TEIXEIRA Jr *et al* (2012a), TEIXEIRA Jr *et al* (2012b). Em particular, nesta tese



de doutorado, primeiramente, o método SSA é utilizado para se realizar a *filtragem estatística* [veja MENEZES *et al* (2012)] da série temporal a ser modelada, visto que possibilita verificar, visualmente, quais são as *componentes* SSA [veja Seção 2.1] classificadas como ruído (estatístico) - o que não é possível em *abordagens de encolhimento de ruído wavelet* [veja GUTIERREZ (2002)]. Em seguida, são realizadas *K decomposições ortogonais wavelet* [veja Seção 2.2] da versão filtrada (via método SSA) da série temporal supracitada. Salienta-se que é utilizada a decomposição ortogonal *wavelet* (e não a decomposição via método SSA), porque os resultados dos experimentos obtidos, ao longo da pesquisa, mostraram que as componentes ortogonais *wavelet* possuem melhores padrões de comportamento, para fins de modelagem, que as componentes SSA [veja ainda PASCAL *et al* (1998), SOUZA *et al* (2012a), SOUZA *et al* (2012b), TEIXEIRA Jr *et al* (2012b), TEIXEIRA Jr *et al* (2012a)].

As *redes neurais artificiais* (RNA' s), que, de acordo com HAYKIN (2001), são classificadas em várias classes distintas, podem ser definidas, basicamente, como sistemas interligados constituídos por unidades de processamento simples, denominadas *neurônios artificiais* [veja ainda ABELÉM (1994)]. As RNA' s, enquanto métodos preditivos individuais, possuem a vantagem de não exigir que as séries temporais a serem modeladas satisfaçam a pressupostos restritivos - como, por exemplo, a *estacionariedade de 2º ordem* [veja HAMILTON (1994)]. Em particular, as *RNA' s feedforward, com uma camada escondida*, formalmente definida no Capítulo 3, podem ser utilizadas como previsores individuais e, para tal finalidade, consistem em abordagens consagradas na literatura de Projeção de Séries Temporais. Por conta de sua eficiência e de sua ampla aplicabilidade, foram escolhidas para integrar a metodologia proposta. Os referidos atributos destas RNA' s, no contexto de projeção de séries temporais, podem ser verificados, dentre outras, nas seguintes referências: ZANDONADE (1993), ABELÉM (1994), AUSSEM & MURTAGH (1997), ROCHA REIS & ALVES DA SILVA (2004), PALIT & POPOVIC (2005), CHAABENE & BEN AMMAR (2008), LEI & RAN (2008), ZERVAS *et al* (2008), AQUINO *et al* (2009), CAO *et al* (2009), CAMPOS (2010), DENG *et al* (2010), TEIXEIRA Jr *et al* (2012a), TEIXEIRA Jr *et al* (2012b), TEIXEIRA Jr



*et al* (2011b), ZHOU *et al* (2011), SOUZA *et al* (2012a), SOUZA *et al* (2012b), ZHANG & BEHERA (2012).

Os *métodos de amostragem* [veja RIPLEY (1987)] consistem em algoritmos numéricos que podem ser utilizados na geração de *choques estocásticos* [veja DIAS (2013)], os quais são utilizados na geração de cenários distribuídos de acordo com alguma distribuição probabilidade - como, por exemplo, a distribuição de probabilidade *normal-padrão* [veja CASELLA & BERGER (2010)]. Basicamente, os métodos de amostragem podem ser agrupados em duas classes: os paramétricos - como, por exemplo, o *método de simulação de Monte-Carlo* [veja NIEDERREITER (1992)] e o *método de simulação de Quase-Monte-Carlo* [veja OWEN (2000)] - e os não paramétricos - como, por exemplo, os métodos de *Jacknife* e de *bootstrap* [veja SOUZA & CAMARGO (2004)]. Em KRYKOVA (2003) e ALBUQUERQUE (2008), verifica-se que o método de simulação de *Quase-Monte-Carlo Híbrido*, integrado com o *inversor de MORO* [veja MORO (1995), Seção 5.2], possui relevantes vantagens na geração de cenários (distribuídos de acordo com uma distribuição de probabilidade normal-padrão), quando comparado com o tradicional método de simulação de Monte-Carlo (integrado com o referido inversor). De acordo com BUSSAB & MORETTIN (2002), a principal desvantagem da utilização de métodos não paramétricos, neste contexto, reside no fato da exigência de um número suficientemente grande de *resíduos estatísticos* [veja LUTKEPOHL (2006)], para garantir uma aproximação satisfatória das densidades preditivas empíricas geradas (cenários), em relação às distribuições de probabilidades reais [veja ainda DIAS (2013)].

Note que a metodologia proposta integra os pontos (1), (2), (3), (4) e (5), a seguir, que são importantes, de acordo com as referências acima e com os experimentos testados nesta pesquisa, na obtenção de ganhos preditivos e possibilitam a modelagem de qualquer série temporal com estrutura de autodependência (linear ou não linear):

(1) Dois procedimentos de pré-processamento (filtragem SSA e  $K$  decomposições ortogonais *wavelet* distintas de nível  $p$  [veja Capítulo 6]) associadas à série temporal de interesse, pois tão importante quanto a escolha dos



previsores é a forma como as observações são apresentadas aos mesmos no processo de modelagem;

(2) Combinação linear das previsões para as componentes ortogonais *wavelet* que se mostrou mais eficiente que a combinação linear tradicional, em todos os experimentos em TEIXEIRA *et al* (2012b), e em todos os testes realizados nesta pesquisa;

(3) Um método numérico de ajuste global (Programação Matemática MINIMAX, com abordagem de Programação por Metas) para obtenção dos valores ótimos para os parâmetros adaptativos da combinação proposta, pois possibilita: a otimização simultânea de todas as estatísticas de aderência que sejam consideradas relevantes pelo tomador de decisão, o teste de outros algoritmos numéricos de otimização e a não necessidade de se deduzir fórmulas analíticas para cálculo dos valores (estimados ou ótimos) para os parâmetros da combinação;

(4) Um método paramétrico de amostragem eficiente. No caso, o método de simulação QMC Híbrido, para obtenção dos cenários para as componentes ortogonais *wavelet*. Tal possibilita, de acordo com KRYKOVA (2003), rápida convergência, em relação ao tradicional Método de Monte-Carlo e aos outros métodos QMC, e, além do mais, para grandes horizontes de previsão, é o mais adequado, de acordo com DIAS (2013) e KRYKOVA (2003), na medida em que preserva as propriedades de *interdependência* e de *intradependência* [veja CHATFIELD (2004)] dos choques estocásticos simulados; e

(5) Aplicabilidade a qualquer série temporal com estrutura de autodependência (linear ou não linear). Em contrapartida, pode-se verificar que as combinações de métodos preditivos individuais propostas na Literatura de Projeção de Séries Temporais, em regra, basicamente consistem em se combinar diferentes previsores individuais ou propor abordagens alternativas para obtenção de valores para os respectivos parâmetros adaptativos tais que minimizam (ou maximizam) uma única estatística de aderência (geralmente, a estatística MSE). Além disso, geralmente, não consideram o pré-processamento das observações e, por vezes, são aplicáveis apenas a pequenas classes de séries temporais, dado que



os métodos preditivos base (como, por exemplo, o modelo de *Box & Jenkins*) exigem determinados pressupostos restritivos.

## 1.2

### Visão Geral da Metodologia Proposta

A descrição teórica da metodologia proposta é exposta, de forma detalhada, no Capítulo 6. Não obstante, em linhas gerais, é subdividida nas 5 etapas, a seguir:

1 - Filtragem via método SSA da série temporal a ser modelada. Nessa etapa, os dados temporais disponíveis são filtrados, de modo a gerar uma versão da série temporal em questão menos ruidosa (ou menos estocástica);

2 - Realização de  $K$  decomposições ortogonais *wavelet* ( $DW'$  s) de nível  $p$  da série temporal filtrada, na etapa 1, onde  $K, p \in \mathbb{N}$ , sendo que  $\mathbb{N}$  denota o conjunto dos inteiros positivos, de forma a gerar  $K \times (p+1)$  componentes ortogonais *wavelet*;

3 - Modelagem individual das  $K \times (p+1)$  componentes ortogonais *wavelet*, da etapa 2, utilizando RNA' s. Com efeito, são necessárias  $K \times (p+1)$  RNA' s para se gerar previsões para cada componente ortogonal *wavelet*, dentro e fora da amostra (no caso, nas amostras de validação e de teste);

4 - Combinação linear proposta, conforme definido na Seção 6.3, das previsões para as  $K \times (p+1)$  componentes ortogonais *wavelet*, da etapa 3. O ajuste numérico dos seus parâmetros adaptativos da combinação linear proposta ocorre por meio de um problema de programação matemática multiobjetivo MINIMAX, com abordagem de programação por metas, descrito, de forma detalhada, na Seção 6.4;

5 - Geração dos cenários por meio do método de simulação QMC híbrido, conforme descrito na Seção 6.5, para as  $K \times (p+1)$



componentes ortogonais *wavelet*, da etapa 2, e a sua combinação, utilizando a combinação linear proposta com os parâmetros ótimos obtidos, na Etapa 4, de forma a gerar os cenários QMC híbridos combinados.

Na Figura 1.1, tem-se um fluxograma que esquematiza as 5 etapas básicas gerais da metodologia proposta, considerando um exemplo, no qual são feitas duas decomposições ortogonais *wavelet* de nível 2 distintas (ou seja, utilizando duas bases ortonormais *wavelet* diferentes) da série temporal filtrada via método SSA, gerando duas componentes ortogonais *wavelet* de aproximação (A1) e duas de detalhe (D1 e D2), para cada decomposição. Por consequência, são necessárias 6 diferentes RNA's para a modelagem individual de cada uma das 6 componentes ortogonais *wavelet* geradas. Em seguida as previsões são linearmente combinadas (CL), com ajuste via Programação Matemática. Por fim, é realizada a geração dos cenários QMC Híbridos combinados utilizando a combinação linear ótima.

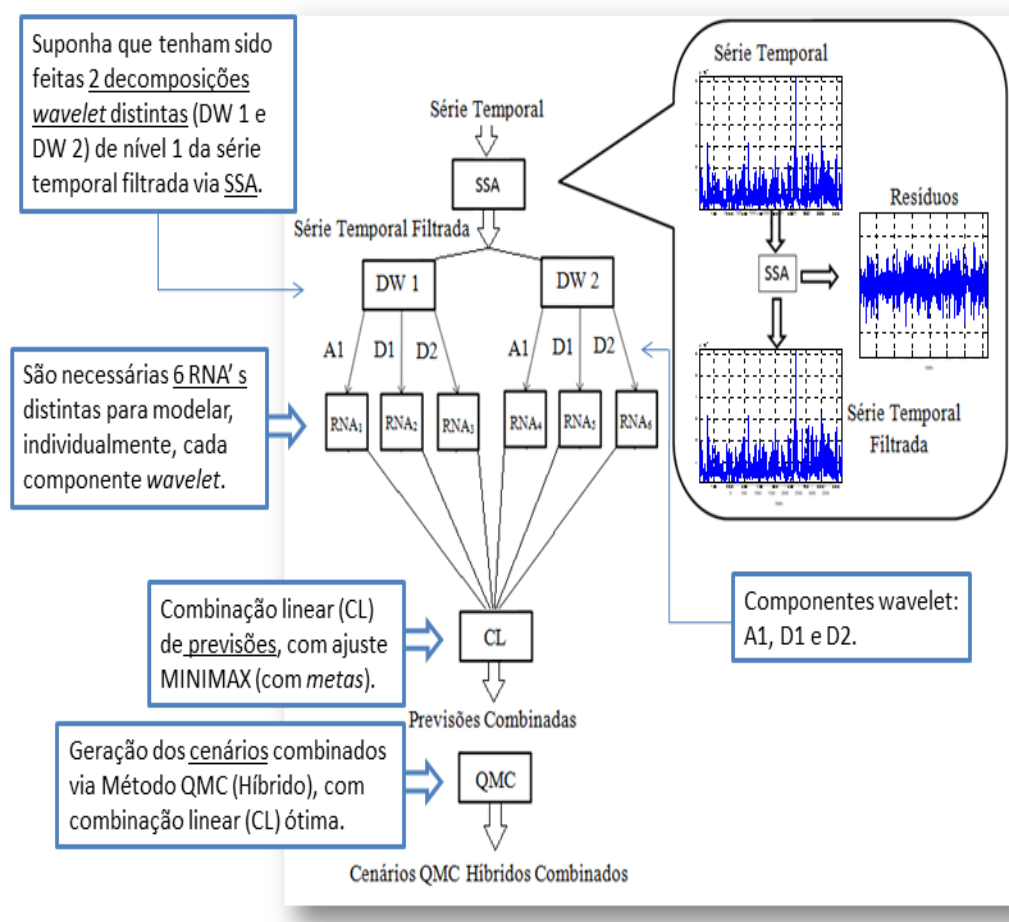


Figura 1.1 - Ilustração das etapas da metodologia proposta.



### 1.3

#### Estrutura da Tese

A presente tese de doutorado está dividida em 8 capítulos. No Capítulo 2, é feita uma sucinta fundamentação teórica sobre o método SSA e a Teoria *Wavelet*. Na Seção 2.1.1, é definida a SVD (*singular value decomposition*) de uma matriz trajetória  $X$  gerada a partir de uma série temporal  $\{y_t\}_{t=1}^T$  e, na Seção 2.1.2, a definição de média diagonal (que é utilizada para a reconstrução da série temporal supracitada). Na Seção 2.2.1, são definidos os vetores *wavelet*; na Seção 2.2.2, são definidas as transformadas *wavelet*; na Seção 2.2.3, é fundamentada e definida uma série de *Fourier*, com bases ortornormais *wavelet*; e, na Seção 2.2.4, é apresentada a definição das funções  $l^2$ -*wavelet* e são mostrados gráficos, com a finalidade de se exemplificar uma função *wavelet* e as componentes ortogonais *wavelet* para uma decomposição de nível 1. No Capítulo 3, são apresentados os elementos básicos que integram a arquitetura de uma RNA e a definição de uma RNA *feedforward*, com uma camada escondida, assim como a sua aplicabilidade na modelagem de séries temporais. No Capítulo 4, é definida, formalmente, a combinação linear de previsões proposta e são definidos conceitos básicos relativos à programação matemática (PM), especialmente os da PM multiobjetivo MINIMAX, com abordagem de programação por metas (que foi utilizada na obtenção de valores ótimos dos parâmetros adaptativos).

No Capítulo 5, são apresentados conceitos relativos ao método de simulação de quase-monte-carlo (QMC), especialmente o método QMC híbrido e o inversor de MORO, ambos utilizados na metodologia proposta. Na Seção 5.2.2, é descrito o procedimento para se obter os choques estocásticos normalmente distribuídos. No Capítulo 6, é descrita, de forma detalhada, a metodologia proposta e a sua fundamentação Matemática, baseada nos conteúdos das seções anteriores. No Capítulo 7, é realizado um experimento computacional que ilustra empiricamente a metodologia proposta e mostra a sua eficiência, na modelagem da série temporal de *vazão mensal de afluentes da região sul do Brasil* [veja CAMPOS (2010)], quando comparado com outras 8 abordagens consagradas na Literatura. No Capítulo 8, têm-se as principais conclusões da pesquisa.



## 2

# ABORDAGENS DE PRÉ-PROCESSAMENTO DE SINAIS

## 2.1

### Método SSA

O método SSA (*Singular Spectrum Analysis* - SSA) é uma técnica de processamento de sinais que pode ser utilizada, dentre outras aplicações [veja GOLYANDINA *et al* (2001)], no *encolhimento de ruído estatístico de séries de tempo estocásticas* [veja GUTIERREZ (2002)]. Sua versão básica pode ser descrita em duas partes: decomposição [Seção 2.1.1] e reconstrução [Seção 2.1.2].

### 2.1.1

#### Decomposição

Seja o trio  $(\Omega, \chi_\Omega, \rho)$  um *espaço de probabilidade* [veja JAMES (2011)], onde:  $\Omega$  é um *espaço amostral* [veja JAMES (2011)];  $\chi_\Omega$  é uma  $\sigma$ -álgebra [veja KUBRUSLY (2007)] de subconjuntos de  $\Omega$  (ou, simplesmente, *espaço evento* em  $\Omega$ ); e  $\rho$  é uma *medida de probabilidade* definida sobre  $\chi_\Omega$  [veja KUBRUSLY (2007)]. Considere que  $\{y_t\}_{t=1}^T$  seja uma *série temporal* em  $\mathbb{R}$  (sendo  $\mathbb{R}$  o conjunto dos números reais), onde  $T$  é a sua cardinalidade, definida sobre um espaço de probabilidade  $(\Omega, \chi_\Omega, \rho)$ , onde, em  $t$ :  $y_t = \mu_t + e_t$ , sendo que  $\mu_t$  é a componente determinística do ponto  $y_t$ ; e  $e_t$ , um *estado* [veja HAMILTON (1994)] da *variável aleatória* [veja JAMES (2011)]  $\varepsilon_t \sim D(0, \sigma^2)$ , onde:  $D$  é uma *distribuição de probabilidade* [veja CASELLA & BERGER (2010)] abstrata, com os parâmetros média e variância iguais, respectivamente, a 0 e  $\sigma^2$ ; e  $\{\varepsilon_t\}_{t=1}^T$ , um *processo estocástico ruído branco* [veja CHATFIELD (2004)].



Tome  $\{y_t\}_{t=1}^T$  e a interprete como um vetor  $Y_T := [y_t]_{1 \times T}$ , onde o símbolo  $:=$  significa “definido por”, no espaço real  $\mathbb{R}^T$   $T$ -dimensional. Considere o mapa *linear e invertível* [veja KUBRUSLY (2001)]  $F_L: \begin{cases} \mathbb{R}^T \rightarrow \mathbb{R}^{L \times K} \\ Y_T \in \mathbb{R}^T \rightarrow F_L(Y_T) \in \mathbb{R}^{L \times K} \end{cases}$ , definido por  $F_L(Y_T) = [X_1, \dots, X_K] = X$ , sendo que  $X_k := [y_k, \dots, y_{k+L-1}] \in \mathbb{R}^L$ , onde:  $k = 1, \dots, K$  e  $\mathbb{R}^L$  denota o espaço real  $L$ -dimensional. De acordo com YIOU *et al* (2000), a matriz  $X$  é chamada de *matriz trajetória* e o parâmetro  $L$ , de *janela* da matriz trajetória  $X$ , onde  $L$  assume um valor inteiro no intervalo  $2 \leq L \leq T$  e  $K = T - L + 1$ .

Considere  $S := XX'$ , onde  $X'$  é a matriz transposta da matriz trajetória  $X$ , e tome o par ordenado  $(\mathbb{R}^L, \langle \cdot, \cdot \rangle)$ , que é, de acordo com KUBRUSLY (2001), um exemplo de *espaço de Hilbert*, onde a forma complexa  $\langle \cdot, \cdot \rangle: \mathbb{R}^L \rightarrow \mathbb{C}$  é um *produto interno* [veja KUBRUSLY (2001), LIMA (2006)]. Pode-se verificar, baseado em KUBRUSLY (2001), que a matriz  $S$  é um *operador normal e compacto* sobre o espaço de *Hilbert*  $(\mathbb{R}^L, \langle \cdot, \cdot \rangle)$ . Com efeito, de acordo com o *Teorema Espectral* (para operadores normais e compactos sobre espaços de *Hilbert*), enunciado em KUBRUSLY (2012), a matriz  $S$  pode ser expandida por meio da seguinte soma ponderada:  $S = \sum_{\lambda_j \in \sigma(S)} \lambda_j (U_j U_j') = \sum_{j \in \Gamma_{\sigma(S)}} \lambda_j (U_j U_j')$ , onde:  $\Gamma_{\sigma(S)}$  é o conjunto de índices  $j$ ;  $\sigma(S)$  é o *espectro* [veja KUBRUSLY (2012)] do operador  $S$ ; e a sequência  $\{(U_j U_j')\}_{j \in \Gamma_{\sigma(S)}}$  é uma sequência de *matrizes compactas* [veja KUBRUSLY (2001)] sobre o espaço de *Hilbert*  $(\mathbb{R}^L, \langle \cdot, \cdot \rangle)$ , onde  $U_j$  é o *autovetor* [veja ELSNER & TSONIS (2010)] do operador  $S$  associado ao  $j$ -ésimo *autovalor*  $\lambda_j$  [veja ELSNER & TSONIS (2010)]. Além disso, pode-se verificar, baseado em ELSNER & TSONIS (2010), que  $S$  é uma *matriz semidefinida positiva*, o que implica que todos os seus autovalores são não negativos. Isto é,  $\lambda_j \geq 0, \forall j \in \Gamma_{\sigma(S)}$ .



Baseado em KUBRUSLY (2012) e em ELSNER & TSONIS (2010), tem-se que, dado que a matriz  $S$  é um operador normal, compacto e semidefinido positivo, segue que a matriz trajetória  $X$  pode ser expandida através da *decomposição em valores singulares* (*singular value decomposition* - SVD), definida em (2.1).

$$X = \sum_{j \in \Gamma_{\sigma(S)}} (\lambda_j)^{\frac{1}{2}} U_j V_j' \quad (2.1)$$

Onde:  $V_j := (X' U_j) / (\lambda_j)^{\frac{1}{2}}, \forall j \in \Gamma_{\sigma(S)}$ ; as sequências  $\left\{ (\lambda_j)^{\frac{1}{2}} \right\}_{j \in \Gamma_{\sigma(S)}}$  e  $\{U_j\}_{j \in \Gamma_{\sigma(S)}}$  são chamadas, de acordo com GOLYANDINA *et al* (2001), de *espectro singular* e de *sequência vetores singulares* da matriz trajetória  $X$ , respectivamente; e o trio  $(\lambda_j, U_j, V_j)$  é, de acordo com HASSANI (2007), é a  $j$ -ésima *autotripla* da matriz trajetória  $X$ , na SVD. Assuma, daqui em diante, a ordenação  $(\lambda_j)^{\frac{1}{2}} \geq (\lambda_{j+1})^{\frac{1}{2}} \geq 0, \forall j \in \Gamma_{\sigma(S)}$ , em (2.1).

De acordo com ELSNER & TSONIS (2010), pode-se verificar que a cardinalidade do conjunto de índices  $\Gamma_{\sigma(S)}$ , na SVD, é igual ao valor que o parâmetro  $L$  da matriz trajetória  $X$  assume, de modo que  $X = \sum_{j \in \Gamma_{\sigma(S)}} (\lambda_j)^{\frac{1}{2}} U_j V_j' = \sum_{j=1}^L (\lambda_j)^{\frac{1}{2}} U_j V_j' = \sum_{j=1}^L X_j$ , onde  $X_j := (\lambda_j)^{\frac{1}{2}} U_j V_j'$  e  $X_j$  é a  $j$ -ésima *matriz elementar* da matriz trajetória  $X$ . Considere que  $d$  seja o *posto* [veja ELSNER & TSONIS (2010)] do operador  $S$ , onde,  $d \leq L$ . Tautologicamente, obtém-se a identidade dada em (2.2).

$$X = \sum_{j=1}^L X_j = \sum_{j=1}^d X_j \quad (2.2)$$

De acordo com YIOU *et al* (2000), a contribuição da  $j$ -ésima matriz elementar  $X_j$  da matriz trajetória  $X$ , na SVD, pode ser mensurada por

$$(\lambda_j)^{\frac{1}{2}} / \sum_{j=1}^d (\lambda_j)^{\frac{1}{2}}.$$



### 2.1.2

#### Reconstrução

De acordo com GOLYANDINA *et al* (2001), a etapa de reconstrução do método SSA pode ser subdividida em duas subetapas: *agrupamento* e *média diagonal*. A subetapa de agrupamento das matrizes elementares tem como objetivo a diminuição do número de matrizes elementares, na SVD da matriz trajetória  $X$ . Para tal, é realizado um procedimento de agrupamento das matrizes elementares na sequência  $\{X_j\}_{j=1}^d$  em  $m$  grupos disjuntos, onde  $m \leq d$ , através do *método de análise gráfica dos autovetores singulares* - descrito em MENEZES *et al* (2012), CASSIANO *et al* (2012), ELSNER & TSONIS (2010) e SOUZA *et al* (2012a) -, gerando uma nova sequência de matrizes, denotada por  $\{X_{I_i}\}_{i=1}^m$ , onde

$X_{I_i} = \sum_{l=1}^{p_i} X_{I_{il}}$ , onde  $i = 1, \dots, m$ . Com efeito, obtém-se a identidade em (2.3).

$$X = \sum_{j=1}^d X_j = \sum_{i=1}^m X_{I_i} = \sum_{i=1}^m \sum_{l=1}^{p_i} X_{I_{il}} \quad (2.3)$$

Tome a união de conjuntos  $\bigcup_{i=1}^m \{X_{I_{il}}\}_{l=1}^{p_i}$  de matrizes, em (2.3), e considere

a função  $F_L^{-1}: \begin{cases} \mathbb{R}^{L \times K} \rightarrow \mathbb{R}^T \\ F_L^{-1}(X) \in \mathbb{R}^{L \times K} \rightarrow Y_T \in \mathbb{R}^T \end{cases}$ , que é o mapa inverso de  $F_L$ , tal que

$F_L^{-1}(X) = Y_T := [y_t]_{1 \times T}$ , conforme a demonstração, em (2.4).

$$\begin{aligned} F_L^{-1}(X) &= F_L^{-1}\left(\sum_{i=1}^m X_i\right) = F_L^{-1}\left(\sum_{i=1}^m \sum_{l=1}^{p_i} X_{I_{il}}\right) = \sum_{i=1}^m F_L^{-1}\left(\sum_{l=1}^{p_i} X_{I_{il}}\right) \Rightarrow \\ &\Rightarrow F_L^{-1}(X) = \sum_{i=1}^m \sum_{l=1}^{p_i} F_L^{-1}(X_{I_{il}}) = \sum_{i=1}^m \left(\sum_{l=1}^{p_i} [y_t^{I_{il}}]_{1 \times T}\right) = \sum_{i=1}^m [y_t^{I_i}]_{1 \times T} \Rightarrow \\ &\Rightarrow F_L^{-1}(X) = \sum_{i=1}^m [y_t^{I_i}]_{1 \times T} = \sum_{i=1}^m [y_t^{(i)}]_{1 \times T} = [y_t]_{1 \times T} \Rightarrow F_L^{-1}(X) = [y_t]_{1 \times T} \end{aligned} \quad (2.4)$$



A componente  $\left[ y_t^{(i)} \right]_{1 \times T}$ , em (2.4), consiste na  $i$ -ésima *componente SSA* da série temporal  $\left[ y_t \right]_{1 \times T}$  gerada, por meio do mapa  $F^{-1}$ , a partir da matriz  $X_{I_i}$ . Tome a matriz trajetória  $X$  e assumamos que  $L^* := \min\{L, K\}$  e  $K^* = \max\{L, K\}$ . Considere que  $x_{l,k}^{(i)}$  seja o elemento da  $l$ -ésima linha e  $k$ -ésima coluna, na matriz  $X_{I_i}$ . De acordo com YIOU *et al* (2000), o elemento  $y_t^{(i)}$  da  $i$ -ésima componente SSA  $\left[ y_t^{(i)} \right]_{1 \times T}$  é calculado de acordo com a *média diagonal*, definida em (2.5).

$$y_t^{(i)} = \begin{cases} \sum_{l=1}^t x_{l,t-l+1}^{(i)} / t, & 1 \leq t < L^* \\ \sum_{l=1}^{L^*} x_{l,t-l+1}^{(i)} / L^*, & L^* \leq t < K^* \\ \sum_{l=t-K^*+1}^{T-K^*+1} x_{l,t-l+1}^{(i)} / T-K+1, & K^* \leq t \leq T \end{cases} \quad (2.5)$$

Segundo HASSANI (2007) e GOLIANDYNA *et al* (2001), a partir do método de análise gráfica do perfil dos vetores singulares, as componentes SSA da série temporal  $\left[ y_t \right]_{1 \times T}$  podem ser classificadas em três categorias distintas: *tendência*, *harmônica* ou *ruído*, de modo a se obter a decomposição *ortogonal* [veja KUBRUSLY (2001)] SSA, em (2.6).

$$\left[ y_t \right]_{1 \times T} = \left[ y_t^{(tendência\_total)} \right]_{1 \times T} + \left[ y_t^{(harmônica\_total)} \right]_{1 \times T} + \left[ y_t^{(ruído\_total)} \right]_{1 \times T} \quad (2.6)$$

Onde, em relação à série temporal  $\left[ y_t \right]_{1 \times T}$ , obtêm-se as três componentes totais:

- $\left[ y_t^{(tendência\_total)} \right]_{1 \times T} := \sum_{m \in \Gamma_m} \left[ y_t^{(tendência\_m)} \right]_{1 \times T}$ , onde  $\left[ y_t^{(tendência\_total)} \right]_{1 \times T}$  é a componente SSA total de tendência e  $\left[ y_t^{(tendência\_m)} \right]_{1 \times T}$ , a  $m$ -ésima componente SSA de tendência;



- $\left[ y_t^{(harmônica\_total)} \right]_{1 \times T} := \sum_{n \in \Gamma_n} \left[ y_t^{(harmônica\_n)} \right]_{1 \times T}$ , onde  $\left[ y_t^{(harmônica\_total)} \right]_{1 \times T}$  é a componente SSA total harmônica e  $\left[ y_t^{(harmônica\_n)} \right]_{1 \times T}$ , a  $n$ -ésima componente SSA harmônica;
- $\left[ y_t^{(ruído\_total)} \right]_{1 \times T} := \sum_{p \in \Gamma_p} \left[ y_t^{(ruído\_p)} \right]_{1 \times T}$ , onde  $\left[ y_t^{(ruído\_total)} \right]_{1 \times T}$  é a componente SSA total de ruído e  $\left[ y_t^{(ruído\_p)} \right]_{1 \times T}$ , a  $p$ -ésima componente SSA de ruído; e
- $d = \# \Gamma_m + \# \Gamma_n + \# \Gamma_p$ , onde  $\#$  denota uma *medida* [veja KUBRUSLY (2007)] de cardinalidade.

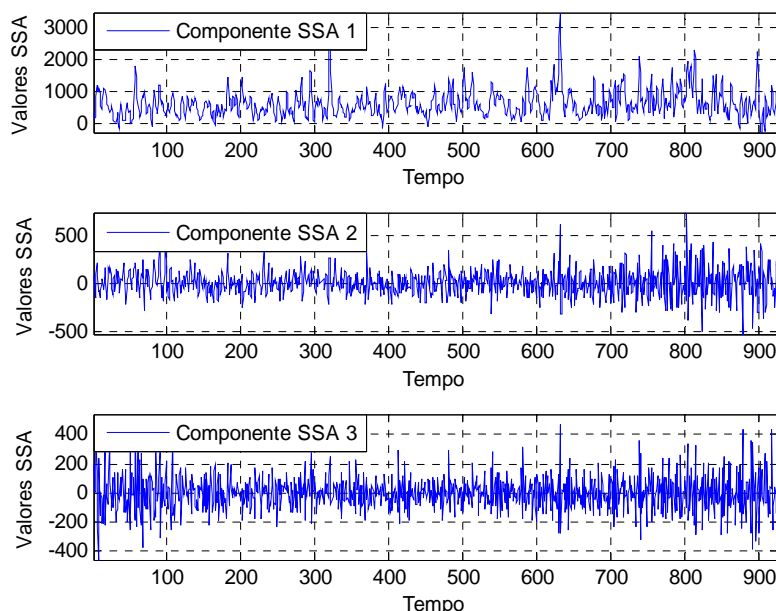


Figura 2.1 – Gráfico das 3 componentes SSA totais [fonte: SOUZA *et al* (2012a)].

Na Figura 2.1, têm-se as 3 componentes SSA totais da série temporal mensal de vazão de afluentes da usina hidrelétrica Governador Bento Munhoz, extraídas do artigo CASSIANO *et al* (2012): componente de tendência (componente SSA 1), componente harmônica (componente SSA 2) e componente de ruído (componente SSA 3). A soma das três componentes (ortogonais) SSA totais resulta na série temporal supracitada.

Para informações adicionais, concernente a aspectos teóricos e aplicados, sobre o método SSA, veja ainda as seguintes referências: BROOMHEAD &



KING (1986), VAUTARD *et al* (1992), GHIL & TARICCO (1997), SCHOELLHAMER (2001), GOLYANDINA & OSIPOV (2006), GISTATGROUP (2010), SOUZA *et al* (2012b).

## 2.2

### Teoria Wavelet

#### 2.2.1

##### Vetores Wavelet

Seja  $H$  um *espaço de Hilbert* abstrato (*complexo*, de *dimensão infinita* e, por simplicidade, *separável*) [veja KUBRUSLY (2001)]. Suponha que  $D$  e  $T$  sejam operadores (isto é, mapas *lineares* e *limitados* [veja KUBRUSLY (2001)] de  $H$  para  $H$ ) que satisfaçam aos seguintes axiomas:  $(A_0)$   $D$  e  $T$  são *shifts bilaterais de multiplicidade contável infinita* [veja KUBRUSLY (2001)] (ou, simplesmente, *shifts*) sobre  $H$ ; e  $(A_1)$   $DT \neq TD$  (condição de não comutatividade) e  $DT^2 = TD$  [veja KUBRUSLY & LEVAN (2006)].

**Definição 2.1** - Se  $D$  e  $T$  são operadores sobre  $H$  que satisfazem aos axiomas  $(A_0)$  e  $(A_1)$ , então qualquer vetor  $0 \neq \omega$  em  $H$  que gera uma base ortonormal  $\{D^m T^n \omega\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$  para  $H$  é um *vetor ortonormal wavelet* (ou, simplesmente, *vetor wavelet*) e os vetores  $\omega_{m,n} := D^m T^n \omega$ , onde  $(m,n) \in \mathbb{Z} \times \mathbb{Z}$ , são *vetores ortonormais wavelet* gerados a partir de  $\omega$ .

De acordo com DAUBECHIES (1988) e MALLAT (1998), o subespaço linear *fechado* [veja KUBRUSLY (2001)]  $W_m(\omega) := \left( \text{span}_{n \in \mathbb{Z}} \{ \omega_{m,n} \} \right)^-$  de  $H$ , onde  $m \in \mathbb{Z}$ , é chamado de subespaço de detalhe de nível  $m$ . Segundo LEVAN & KUBRUSLY (2003), a *projeção ortogonal* [veja KUBRUSLY (2001)]  $P_{\omega,m}(x) := \sum_{n \in \mathbb{Z}} \langle x, \omega_{m,n} \rangle \omega_{m,n}$  de  $x$  sobre  $W_m(\omega)$  pode ser chamada de *componente* (ou *camada*) de detalhe de nível  $m$  do vetor  $x$  em  $H$ .



**Definição 2.2** - Se  $D$  e  $T$  são operadores sobre  $H$  que satisfazem aos axiomas  $(A_0)$  e  $(A_1)$ , então qualquer vetor  $0 \neq \phi$  em  $H$  que gera um subconjunto  $\{D^m T^n \phi\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$  de  $H$  tal que  $\langle D^{m'} T^{n'} \phi, D^j T^k \phi \rangle = 0$ , sempre que  $m' = j$  e  $n' \neq k$ , para todo  $m', n', j, k \in \mathbb{Z}$ , e  $\langle D^{m'} T^{n'} \phi, D^j T^k \phi \rangle \neq 0$ , se caso contrário, é um *vetor wavelet escala* (ou, simplesmente, *vetor escada*) e os vetores  $\phi_{m,n} := D^m T^n \phi$ , onde  $(m,n) \in \mathbb{Z} \times \mathbb{Z}$ , são *vetores wavelet escala* gerados a partir de  $\phi$ .

De acordo com DAUBECHIES (1988) e MALLAT (1998), o subespaço linear fechado  $V_m(\phi) := \left( \text{span}_{n \in \mathbb{Z}} \{ \phi_{m,n} \} \right)^-$  de  $H$ , onde  $m \in \mathbb{Z}$ , é chamado de subespaço de aproximação de nível  $m$ . A projeção ortogonal  $P_{\phi,m}(x) := \langle x; \phi_{m,n} \rangle \phi_{m,n}$  de  $x$  sobre  $V_m(\phi)$  pode ser chamada, de acordo com TEIXEIRA JR *et al* (2012b), de *componente* (ou *camada*) de aproximação de nível  $m$  do vetor  $x$  em  $H$ .

## 2.2.2

### Transformada Wavelet

De acordo com OGDEN (1997), por *transformada wavelet* sobre  $H$ , entende-se como sendo o produto interno entre um vetor  $x \in H$  e um vetor *wavelet*  $\omega_{m,n} \in W_m(\omega)$  ou um vetor escala  $\phi_{m,n} \in V_m(\phi)$ , onde  $(m,n) \in \mathbb{Z} \times \mathbb{Z}$ . O conjunto de todas as transformadas *wavelet*, de acordo com MALLAT (1998), pode ser subdividido em dois grupos disjuntos: o dos *coeficientes de detalhes*, denotado por  $\{a_{m,n}\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$ ; e o dos *coeficientes de aproximação*, denotado por  $\{d_{m,n}\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$ . Dado algum  $(m,n) \in \mathbb{Z} \times \mathbb{Z}$ , segue que o coeficiente de detalhe e o de aproximação são, respectivamente, definidos como segue:  $d_{m,n} := \langle x; \omega_{m,n} \rangle$  e  $a_{m,n} := \langle x; \phi_{m,n} \rangle$ . Segundo KUBRUSLY (2001), as sequências  $\{d_{m,n}\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$  e  $\{a_{m,n}\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$  de transformadas *wavelet* podem ainda ser chamadas de *coeficientes de Fourier*.



### 2.2.3

#### Expansão Wavelet

De acordo com LEVAN & KUBRUSLY (2003), uma cadeia crescente (no sentido de relação por inclusão) de subespaços fechados  $\{V_m(\phi)\}_{m \in \mathbb{Z}}$  é uma *Análise Multirresolução Wavelet* (ou, simplesmente, *AMR Wavelet*), com vetor escala  $\phi$ , se os axiomas (a), (b), (c), (d), (e), a seguir, são satisfeitos.

- (a)  $V_m(\phi) \subset V_{m+1}(\phi)$ ,  $m \in \mathbb{Z}$ ;
- (b)  $\bigcap_{m \in \mathbb{Z}} V_m(\phi) = \{0\}$ ;
- (c)  $\left( \bigcup_{m \in \mathbb{Z}} V_m(\phi) \right)^- = H$ , ou seja,  $\bigcup_{m \in \mathbb{Z}} V_m(\phi)$  é denso [veja KUBRUSLY (2001)] em  $H$ ;
- (d)  $v \in V_m(\phi) \Leftrightarrow Dv \in V_{m+1}(\phi)$ ,  $m \in \mathbb{Z}$  (onde:  $\Leftrightarrow$  significa “se, e somente se,”); e
- (e)  $\{D^m T^n \phi\}_{n \in \mathbb{Z}}$  é uma base ortonormal do subespaço  $V_m(\phi)$ .

Em MALLAT (1998), mostra-se que, dada uma cadeia  $\{V_m(\phi)\}_{m \in \mathbb{Z}}$  de subespaços de aproximação sobre um *Hilbert*  $H$  é uma *AMR Wavelet*, segue que o subespaço de aproximação  $V_m(\phi)$ , na escala  $m$ , pode ser expandido ortogonalmente em termos dos subespaços  $V_{m-1}(\phi)$  e  $W_{m-1}(\omega)$  tal como:  $V_m(\phi) = V_{m-1}(\phi) + W_{m-1}(\omega)$ , para todo  $m \in \mathbb{Z}$ . Em LEVAN & KUBRUSLY (2003), prova-se que o espaço de aproximação  $V_{m_0}(\phi)$ , de nível  $m_0$ , pode ser ortogonalmente expandido tal como a seguinte *soma topológica* [veja KUBRUSLY (2001)]:  $V_{m_0}(\phi) = \left( \sum_{m=-\infty}^{m_0-1} W_m(\omega) \right)^-$ , para todo  $m_0 \in \mathbb{Z}$ .

Em KUBRUSLY & LEVAN (2006), é mostrado que o espaço de *Hilbert*  $H$  pode ser expandido tal como:  $H = \left( \sum_{m \in \mathbb{Z}} W_m(\omega) \right)^-$ , sendo que  $W_{m'}(\omega) \perp W_j(\omega)$  (onde o símbolo  $\perp$  significa “ortogonal”), para todo  $m' \neq j$ , onde  $m', j \in \mathbb{Z}$ . Isto é, para todo  $x \in H$ , existe uma expansão ortogonal (única) tal como



$x = \sum_{m \in \mathbb{Z}} P_{\omega, m}(x)$  sobre o espaço de *Hilbert*  $H$ , onde:  $P_{\omega, m}(x) := \langle x; \omega_{m, n} \rangle \omega_{m, n}$ , para todo  $m \in \mathbb{Z}$ , e  $\{P_{\omega, m}\}_{m \in \mathbb{Z}}$  é uma resolução de identidade *wavelet* sobre  $H$ . Dadas as somas topológicas  $V_{m_0}(\phi) = \left( \sum_{m=-\infty}^{m_0-1} W_m(\omega) \right)^-$  e  $H = \left( \sum_{m \in \mathbb{Z}} W_m(\omega) \right)^-$ , segue que  $H = \left( \sum_{m=-\infty}^{m_0-1} W_m(\omega) \right)^- + \left( \sum_{m=m_0}^{+\infty} W_m(\omega) \right)^- \Rightarrow H = V_{m_0}(\phi) + \left( \sum_{m=m_0}^{+\infty} W_m(\omega) \right)^-$  - onde o símbolo  $\Rightarrow$  significa “implica”. Isto é, para todo  $x \in H$ , existe uma expansão ortogonal (única)  $x = P_{\phi, m_0}(x) + \sum_{m=m_0}^{+\infty} P_{\omega, m}(x)$ , onde as projeções ortogonais  $P_{\phi, m_0}(x) := \sum_{n \in \mathbb{Z}} \langle x; \phi_{m_0, n} \rangle \phi_{m_0, n}$  e  $P_{\omega, m}(x) := \sum_{n \in \mathbb{Z}} \langle x; \omega_{m, n} \rangle \omega_{m, n}$  do vetor  $x$  em  $H$  são, respectivamente, as suas componentes *wavelet* de aproximação de nível  $m_0$  e de detalhe de nível  $m$ ; e  $\{P_{\phi, m_0}\} \cup \{P_{\omega, m}\}_{m=m_0}^{+\infty}$  é uma resolução de identidade *wavelet* sobre  $H$ . Portanto, qualquer vetor  $x$  em um espaço de *Hilbert*  $H$  pode ser expandido tal como em (2.7).

$$x = \sum_{n \in \mathbb{Z}} a_{m_0, n} \phi_{m_0, n} + \sum_{m=m_0}^{+\infty} \sum_{n \in \mathbb{Z}} d_{m, n} \omega_{m, n} \quad (2.7)$$

Onde:  $a_{m, n} := \langle x; \phi_{m, n} \rangle$  e  $d_{m, n} := \langle x; \omega_{m, n} \rangle$  são, respectivamente, os coeficientes de aproximação e de detalhe, onde  $(m, n) \in \mathbb{Z} \times \mathbb{Z}$ . De acordo com o *Teorema da Série de Fourier*, enunciado em KUBRUSLY (2001), a expansão definida, em (2.7), consiste na série de *Fourier* do vetor  $x$  sobre o espaço de *Hilbert*  $H$ , definida em termos de uma base ortonormal *wavelet*

$$\{\phi_{m_0, n}\}_{n \in \mathbb{Z}} \cup \{\omega_{m, n}\}_{(n, m) \in \mathbb{Z} \times \{m\}_{m=m_0}^{+\infty}}.$$



### 2.2.4

#### Funções Wavelet

De acordo com TEIXEIRA Jr *et al* (2012b), um elemento  $\omega(\cdot)$  em um espaço de Hilbert  $(l^2, \langle \cdot; \cdot \rangle)$  [veja KUBRUSLY (2001)] - onde

$l^2 := \left\{ \{ \xi_i \}_{i \in \mathbb{Z}} : \sum_{i \in \mathbb{Z}} |\xi_i|^2 < \infty \right\}$  e a forma complexa  $\langle \cdot; \cdot \rangle : l^2 \rightarrow \mathbb{C}$  é um produto interno

definido sobre o espaço linear  $l^2$  - é uma função  $l^2$ -wavelet se a sequência

$\left\{ \omega_{m,n}(\cdot) := D^m T^n = 2^{\frac{m}{2}} \omega[2^m(\cdot) - n] \right\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$  de funções é uma base ortonormal de  $l^2$

. Com efeito, de acordo com a Seção 2.2.3, qualquer função  $f(\cdot)$  em  $(l^2, \langle \cdot; \cdot \rangle)$  admite sua expansão, por meio de uma série de *Fourier*, definida em termos da base ortonormal wavelet  $\{ \omega_{m,n}(\cdot) \}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$ , definida em (2.8).

$$f(\cdot) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} \langle f(\cdot), \omega_{m,n}(\cdot) \rangle \omega_{m,n}(\cdot) \quad (2.8)$$

A projeção ortogonal de  $f(\cdot)$  sobre  $\omega_{m,n}(\cdot)$ , de acordo com LEVAN & KUBRUSLY (2003), pode ser interpretada como uma *variação de detalhes* de  $f(\cdot)$ , na escala  $2^m$  e translação unitária  $n$ . Segundo DAUBECHIES (1988), para cada  $m \in \mathbb{Z}$ , a projeção ortogonal de  $f(\cdot)$  sobre o subespaço de detalhe  $W_m(\omega)$  de  $l^2$ , onde  $W_m(\omega) := \text{span}\{ \omega_{m,n}(\cdot) \}_{n \in \mathbb{Z}}$ , é definida pela soma parcial, na escala  $2^m$ , em (2.9).

$$f_{W_m(\omega)}(\cdot) := \sum_{n \in \mathbb{Z}} \langle f(\cdot), \omega_{m,n}(\cdot) \rangle \omega_{m,n}(\cdot) \quad (2.9)$$

A projeção ortogonal  $f_{W_m(\omega)}(\cdot)$  pode ser referida como uma *componente de detalhe* da função  $f(\cdot)$  em  $(l^2, \langle \cdot; \cdot \rangle)$ , na escala  $2^m$ , de modo que  $f(\cdot)$  pode ser interpretada como a soma de suas camadas de detalhe em todas as escalas em  $\mathbb{Z}$  [veja LEVAN & KUBRUSLY (2003)]. Isto é,  $f(\cdot) = \sum_{m \in \mathbb{Z}} f_{W_m(\omega)}(\cdot)$ . De acordo com



TEIXEIRA Jr *et al* (2012b), um elemento  $\phi(\cdot)$  no espaço de *Hilbert*  $(l^2, \langle \cdot; \cdot \rangle)$  é uma *função*  $l^2$ -*wavelet* *escala* (ou, simplesmente, *função*  $l^2$ -*escala*) se a sequência  $\left\{ \phi_{m_0, n}(\cdot) l^2 := D^{m_0} T^n = 2^{\frac{m_0}{2}} \phi[2^{m_0}(\cdot) - n] \right\}_{(m_0, n) \in \mathbb{Z} \times \mathbb{Z}}$  é tal que  $\langle \phi_{l, i}; \phi_{j, k} \rangle = 0$ , sempre que  $l = j$  e  $i \neq k$ , e  $\langle \phi_{l, i}; \phi_{j, k} \rangle \neq 0$ , caso contrário. Baseado em DAUBECHIES (1988), tem-se que, para cada  $m_0 \in \mathbb{Z}$ , a projeção ortogonal de  $f(\cdot)$  sobre o subespaço de aproximação  $V_{m_0}(\omega)$  de  $l^2$ , onde  $V_{m_0}(\phi) := \text{span} \{ \phi_{m_0, n}(\cdot) \}_{n \in \mathbb{Z}}$ , é dada pela soma parcial, na escala  $2^{m_0}$ , em (2.10).

$$f_{V_{m_0}(\phi)}(\cdot) := \sum_{n \in \mathbb{Z}} \langle f(\cdot), \phi_{m_0, n}(\cdot) \rangle \phi_{m_0, n}(\cdot) \quad (2.10)$$

A projeção ortogonal  $f_{V_{m_0}(\phi)}(\cdot)$  sobre o subespaço  $V_{m_0}(\omega)$  pode ser interpretada como a *camada* (ou *componente*) *wavelet de aproximação* de uma função  $f(\cdot)$  em  $(l^2, \langle \cdot; \cdot \rangle)$ , na escala  $2^{m_0}$ . A *decomposição ortogonal wavelet*, em (2.11), de uma função  $f(\cdot)$  em  $(l^2, \langle \cdot; \cdot \rangle)$  pode ser interpretada como a soma de uma componente *wavelet* de aproximação, na escala  $2^{m_0}$ , e de todas de detalhe nas escalas em  $\{2^m\}_{m=m_0}^{+\infty}$ .

$$f(\cdot) = f_{V_{m_0}(\phi)}(\cdot) + \sum_{m=m_0}^{+\infty} f_{W_m(\omega)}(\cdot) \quad (2.11)$$

Os parâmetros  $m$  e  $n$  são chamados, respectivamente, de *parâmetro de aproximação* e de *parâmetro de detalhe*. Na Figura 2.2, têm-se os gráficos das funções *wavelet* da família de *Symlet*, com *momento nulo* [veja MORETIN (1997)] igual a 20 (*sym 20*). Na parte superior, tem-se o da *função escala sym 20* e, na inferior, o da *função wavelet sym 20*.



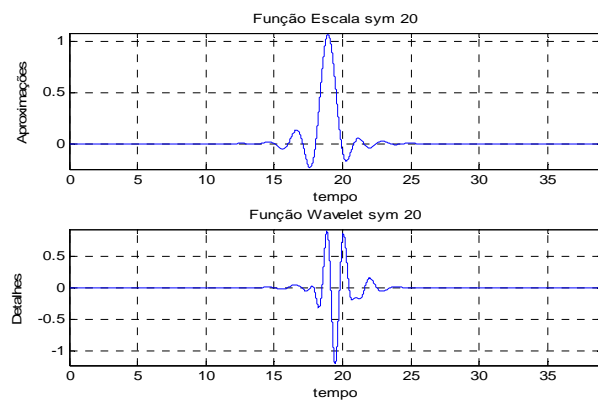


Figura 2.2 - Funções *Wavelet sym 20*.

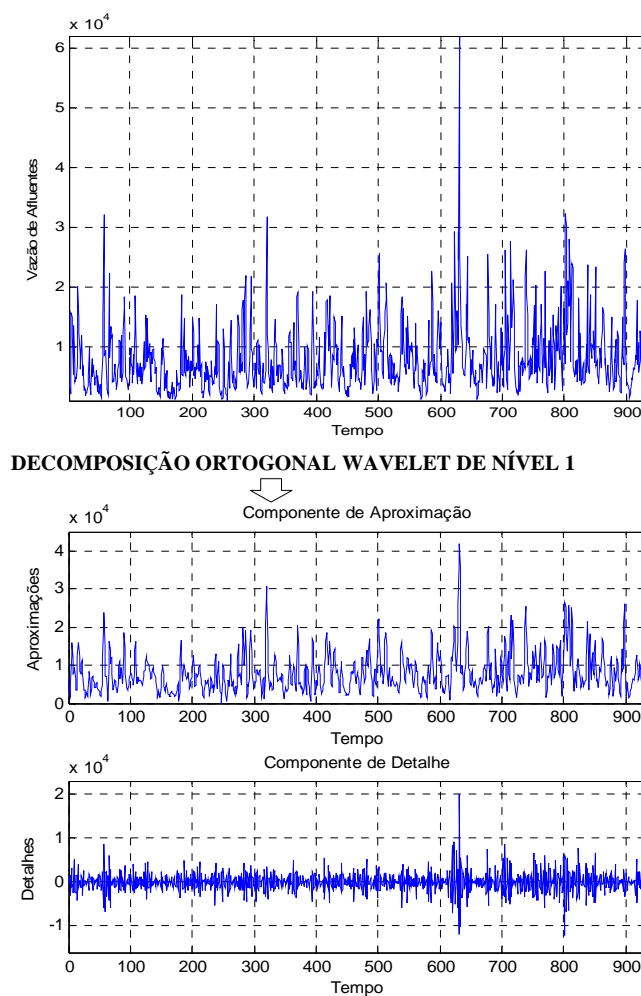


Figura 2.3 - Decomposição *wavelet* de nível 1, com a base ortogonal *wavelet sym 20*, da série temporal mensal de vazão de afluentes da região sul do Brasil.



Na Figura 2.3, têm-se as componentes ortogonais de aproximação e de detalhe da série temporal mensal de vazão de afluentes da região Sul do Brasil de uma *decomposição ortogonal wavelet de nível 1*, em termos de um subconjunto finito da base ortonormal *wavelet* gerada a partir da função *wavelet sym 20* [veja TEIXEIRA Jr *et al* (2011b)].



## REDES NEURAIS ARTIFICIAIS

De acordo com HAYKIN (2001), as *redes neurais artificiais* (RNA' s) com uma, e somente uma, camada escondida consistem em sistemas constituídos por unidades de processamento simples, chamadas de *neurônios artificiais* [veja ABELÉM (1994)] - e, de acordo com CYBENKO (1989), podem ser utilizadas para aproximar classes de *funções contínuas de suporte compacto* [veja KUBRUSLY (2001)]. Os neurônios artificiais de uma RNA são dispostos em *camadas neurais* [veja TAFNER (1996)] interligadas por um grande número de *conexões* - denominados *pesos sinápticos* [veja TAFNER (1996)], os quais ponderam os *padrões de entrada* [veja RUSSELL & NORVIG (1995)] recebidos. De acordo com TAFNER (1996), o funcionamento das RNA' s é inspirado na estrutura massivamente paralela do cérebro humano, de forma a realizar operações similares, através de conhecimento experimental obtido previamente a partir dos padrões (ou sinais), tais como: *aprendizado, associação, generalização e abstração*.

Na Figura 3.1, tem-se uma esquematização simplificada de um neurônio biológico em cuja estrutura um neurônio artificial de uma RNA é inspirado.

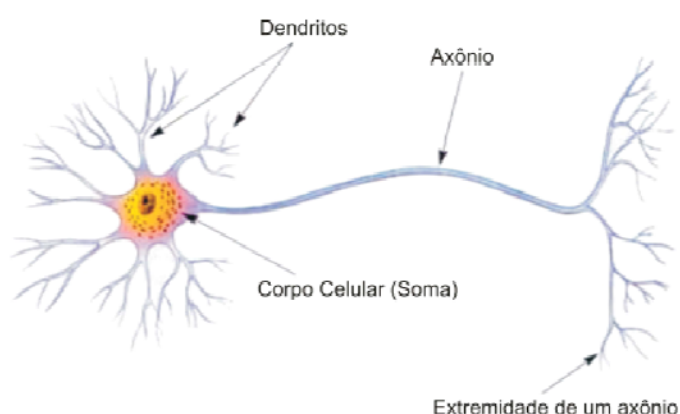


Figura 3.1 - Esquematização simplificada de um neurônio biológico.



Segundo TAFNER (1996), um neurônio biológico é constituído, basicamente, pelos seguintes elementos: um *corpo celular* (ou *soma*), que contém o núcleo da célula; diversos *dendritos*, através dos quais impulsos elétricos são recebidos; e um *axônio*, por meio do qual são enviados impulsos elétricos. As interligações entre neurônios são efetuadas através de *sinapses* (isto é, pontos de contato que são controlados por impulsos elétricos e por reações químicas) entre os dendritos e os axônios, formando, assim, uma rede neural biológica de transmissão de informações. Segundo HAYKIN (2001), um neurônio artificial é constituído pelos elementos descritos, em (1), (2), (3) e ilustrados, na Figura 3.2.

(1) Um *conjunto de pesos sinápticos*, denotado por  $\left\{\{W_{kt}\}_{k=1}^K\right\}_{t=1}^T$ , e um

*conjunto de padrões de entrada*, denotado por  $\{x_t\}_{t=1}^T$ , onde  $W_{kt}$  denota o peso sináptico associado ao padrão de entrada  $x_t$  (observado no instante  $t$ ) e ao  $k$ -ésimo neurônio artificial. Cada padrão de entrada  $x_t$  é multiplicado pelo peso respectivo sináptico  $W_{kt}$ , ou seja,  $x_t W_{kt}$ , para todo  $t \in \{1, \dots, T\}$  e para todo  $k \in \{1, \dots, K\}$ ;

(2) Um *somador* - que consiste em um operador denotado por  $\sum$  - para

realizar a operação soma dos padrões de entrada em  $\{x_t\}_{t=1}^T$  ponderados (por meio da operação de multiplicação) pelos respectivos pesos sinápticos no conjunto  $\left\{\{W_{kt}\}_{k=1}^K\right\}_{t=1}^T$ . Isto é,  $\sum_{t=1}^n x_t W_{kt} + b_k$ , onde:  $x_t$  é o

padrão de entrada, em  $t$ ;  $W_{kt}$  é o peso sináptico associado ao padrão de entrada  $x_t$  e ao  $k$ -ésimo neurônio artificial; e  $b_k$  é o intercepto (ou nível, ou *bias*), que tem a função de aumentar ou diminuir a entrada líquida da função de ativação, podendo, desse modo, ser positivo ou negativo. A

soma ponderada  $\sum_{t=1}^n x_t W_{kt} + b_k$ , de acordo com HAYKIN (2001), pode ser

chamada de *regra de propagação*; e



- (3) Uma *função de ativação*, denotada por  $\varphi$ , que é utilizada para restringir a amplitude da saída  $y_k$  de um neurônio artificial, dada a soma ponderada

$$\sum_{t=1}^n x_t W_{kt} + b_k = v_k. \text{ Isto é, } y_k = \varphi(v_k), \text{ onde } k \in \{1, \dots, K\}. \text{ Além disso,}$$

segundo HERRERA & LOZANO (1998), são utilizadas na identificação de relações lineares ou não lineares entre os padrões de entrada e os de saída em modelos neuronais.

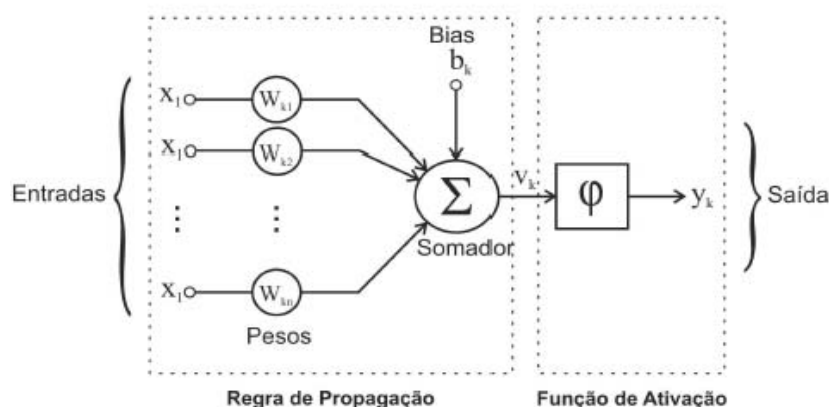


Figura 3.2 - Arquitetura básica de um neurônio artificial.

### 3.1

#### Componentes Básicas de uma RNA

De acordo ZSOLT (2006), as RNA's (básicas) podem ser classificadas de acordo com as características nominadas, a seguir: (a) *função de ativação* dos neurônios artificiais [Seção 3.1.1]; (b) a *arquitetura*, que consiste na forma como os neurônios estão interligados na estrutural neuronal [Seção 3.1.2]; e (c) O *treinamento neural* [Seção 3.1.3].

#### 3.1.1

##### Função de Ativação

De acordo com CAMPOS (2010), os quatro tipos de funções de ativação mais utilizados, em aplicações envolvendo a projeção de séries temporais, são: *linear, degrau, logística, tangente hiperbólica*.



**(a) Função Linear:**

$$y_k = \sum_{t=1}^n x_t W_{kt} + b_k. \quad (3.1)$$

A *função linear*, geralmente, é utilizada nos neurônios da camada de saída e é indicada quando as sequências de padrões de saída associados a estes não são *conjuntos limitados* [veja KUBRUSLY (2001)].

**(b) Função Degrau (ou Limiar):**

$$y_k = \begin{cases} 1, & \text{se } v_k \geq 0 \\ 0, & \text{se } v_k < 0 \end{cases}. \quad (3.2)$$

O conjunto imagem da *função degrau* (ou *limiar*) é constituído pelos valores 0 ou 1. No entanto, por vezes, é necessário que a sua imagem seja constituída pelos valores -1 ou +1, de modo que, neste caso, a função de limiar pode ser particularmente chamada de *função de sinal* [veja HAYKIN (2001)].

**(c) Função Sigmoide Logística:**

$$y_k = \frac{1}{1 + \exp\{-\alpha v_k\}}, \text{ onde } \alpha \in \mathbb{R}. \quad (3.3)$$

A *função sigmoide logística* toma valores no *intervalo limitado e fechado* [veja KUBRUSLY (2001)]  $[0,1]$  e consiste em uma *transformação monotonicamente crescente* [veja LIMA (2006)], onde  $\sum_{t=1}^n x_t W_{kt} + b_k = v_k$  é a regra de propagação do  $k$ -ésimo neurônio artificial.

**(d) Tangente Hiperbólica:**

$$y_k = \tanh(\alpha v_k), \text{ onde } \alpha \in \mathbb{R}. \quad (3.4)$$

A *função tangente hiperbólica*, denotada por  $\tanh(\cdot)$ , é utilizada quando se deseja que a saída de um neurônio artificial tome valores no intervalo limitado e fechado  $[-1,1]$ .



### 3.1.2

#### Arquitetura

A *arquitetura* de uma RNA se refere à maneira como os seus neurônios artificiais estão interconectados. De acordo com RUSSELL & NORVIG (1995), os dois tipos básicos de arquitetura de RNA' s são a *feedforward* (*não recorrente*) [letra (a)] e a *recorrente* [letra (b)].

- (a) **RNA' s, com arquitetura *feedforward*:** as RNA' s com arquitetura *feedforward* são aquelas que não possuem conexões entre neurônios da mesma camada ou entre um neurônio de uma camada com outro neurônio de uma camada anterior. Na Figura 3.3, é ilustrada uma RNA *feedforward*, com uma camada escondida (intermediária ou oculta) e um neurônio na camada de saída. Uma RNA *feedforward* possui, basicamente: um conjunto de nós de entradas, que simplesmente distribuem os padrões de entrada para os seus neurônios artificiais; uma camada de saída, cujos neurônios fornecem as saídas finais da referida RNA; e camadas intermediárias, cujas saídas dos neurônios são as entradas dos neurônios da camada seguinte [veja HAYKIN (2001)].
- (b) **RNA' s, com arquitetura *recorrente*:** as RNA' s com arquitetura *recorrente* são aquelas que possuem realimentação das saídas para as entradas, de modo que as saídas destas são determinadas pelas entradas atuais e pelas saídas anteriores [veja HAYKIN (2001)]. As estruturas das RNA' s *recorrentes*, portanto, podem apresentar interligações entre neurônios da mesma camada ou entre neurônios de camadas anteriores.

De acordo RUSSELL & NORVIG (1995), a arquitetura de uma RNA (básica) envolve, em linhas gerais, os seguintes parâmetros: padrões de entrada, número de camadas intermediárias, número de neurônios por camada (entrada, intermediária e de saída), tipo de função de ativação dos neurônios e os padrões de saída.



### 3.1.3

#### Treinamento Neural

Segundo ABELÉM (1994), o *treinamento neural* de uma RNA consiste em um processo, no qual, com o uso de um *algoritmo de otimização* [veja VANDERPLAATS (1999)] plausível, é realizado o ajuste numérico dos pesos sinápticos, de forma a minimizar (ou maximizar) uma *função-objetivo* [veja RANDSGALE (2004)]. Para tal, de acordo com HAYKIN (2001), é comum se utilizar um único conjunto de sinais como padrões de entrada e de saída, simultaneamente, o qual é denominado *amostra de treino* [veja ainda CAMPOS (2010)]. Geralmente, o treinamento de uma RNA utiliza como *objetivo* [veja RANDSGALE (2004)] a minimização da função *erro quadrático médio* (EQM) [veja KUBRUSLKY (2001)] entre os padrões de saída da RNA e os desejados.

HERRERA & LOZANO (1998) destacam que é de extrema importância a utilização de padrões (de entrada e de saída) que sejam representativos, quanto a uma *população* [veja MORETTIN & TOLOI (2006)] de interesse, e que contenham um número de sinais suficiente grande, para que se obtenha uma modelagem adequada. Segundo HAYKIN (2001), os procedimentos de treinamento neural ocorrem em dois tipos básicos:

- (a) **Treinamento supervisionado:** os padrões de treino são constituídos pelos conjuntos de entrada e de saída desejadas. Durante o processo de treinamento, as entradas são apresentadas as RNA' s e, após isso, são calculadas as saídas neuronais. Em seguida, cada saída da RNA é comparada com a respectiva saída desejada, gerando-se, assim, um sinal de erro (que é a diferença entre as duas saídas). O algoritmo de treinamento, a cada ciclo, atualiza os pesos sinápticos, baseados nos valores de erros nos padrões de treino, a fim de que as saídas da RNA sejam as mais próximas possíveis das saídas desejadas; e
- (b) **Treinamento não supervisionado:** os padrões de treinamento não contêm padrões de saída desejados, de forma que não existem comparações que acarrete a geração de sinais de erro. De acordo com PALIT & POPOVIC (2005), o processo de treinamento não supervisionado, em geral, extrai as propriedades estatísticas do conjunto



de padrões de entrada, de maneira a formar agrupamentos disjuntos constituídos de padrões similares, com base em uma *topologia* [veja KUBRUSLY (2001)].

Após o treinamento neural, é apresentado à RNA treinada um conjunto de padrões de *validação* e de *teste* [veja ABELÉM (1994), CAMPOS (2010)], que são constituídos de padrões não apresentados antes à RNA, mas que pertençam à mesma população dos padrões de treino, a fim de validá-la e testá-la. Existem vários critérios para a escolha de uma RNA [veja RUSSELL & NORVIG (1995)], mas, no caso das RNA's com treinamento supervisionado (que são as de interesse nesta pesquisa), o erro quadrático médio (nas amostras de treino, validação e teste) é o mais utilizado.

### 3.2

#### Normalização dos Padrões

De acordo TAFNER (1996), a *normalização dos padrões* de entrada e de saída de RNA's, com treinamento supervisionado, é fundamental para que suas magnitudes fiquem dentro de uma faixa de valores, a fim de evitar que os padrões com maiores magnitudes distorçam os valores dos pesos sinápticos no processo de treinamento neural e de suprimir problemas associados à *paralisia neural* [veja HAYKIN (2001)]. Neste caso, os padrões devem ser normalizados em um intervalo limitado compatível com o domínio das funções de ativação dos respectivos neurônios da estrutura neuronal. De acordo com CAMPOS (2010), as normalizações de sinais mais comuns, em aplicações envolvendo projeções de séries temporais, são as que os transformam em pontos nos intervalos limitados e fechados  $[0,1]$  ou  $[-1,1]$ ; ou, ainda, em pontos *padronizados* (ou seja, elementos de um conjunto de sinais cuja média é igual a 0 e desvio-padrão, igual a 1).

Nesta perspectiva, as normalizações de padrões *MINIMAX*, *método da faixa*, *z-score*, *sigmoidal* são, respectivamente, definidas nos itens (a), (b), (c) e (d), abaixo. Para tanto, assuma previamente que  $\{x_t\}_{t=1}^n$  seja, simultaneamente, o conjunto dos padrões de entrada e de saída de uma RNA, com treinamento supervisionado, e que  $\{x_t^m\}_{t=1}^n$  seja a sua versão normalizada.



- (a) **Método MINIMAX:** utiliza o valor máximo do conjunto  $\{abs(x_t)\}_{t=1}^n$  de sinais, o qual é denotado por  $Max\{abs(x_t)\}_{t=1}^n$ , onde  $abs(x_t)$  denota o valor absoluto de  $x_t$ . A normalização dos pontos da sequência  $\{x_t\}_{t=1}^n$  ocorre conforme a transformação definida em (3.5), que toma valores no intervalo  $[-1,1]$ .

$$x_t^m = x_t / Max\{abs(x_t)\}_{t=1}^n \quad (3.5)$$

- (b) **Método *premnmx*:** utiliza os valores máximo e mínimo do conjunto de padrões  $\{x_t\}_{t=1}^n$ , os quais são denotados, respectivamente, por  $Max\{x_t\}_{t=1}^n$  e  $Min\{x_t\}_{t=1}^n$ . A normalização de cada ponto  $x_t$  ocorre de acordo com a transformação definida em (3.6), a qual toma valores no intervalo  $[-1,1]$ .

$$x_t^m = 2 \times \left[ \left( x_t - Min\{x_t\}_{t=1}^n \right) / \left( Max\{x_t\}_{t=1}^n - Min\{x_t\}_{t=1}^n \right) \right] - 1 \quad (3.6)$$

- (c) **Método do z-score:** realiza a normalização do conjunto  $\{x_t\}_{t=1}^n$  em torno de sua média, denotada por  $\bar{x}_n$ , utilizando o seu desvio padrão, denotado por  $S_x$ , de acordo com a transformação definida em (3.7). A média de conjunto  $\{x_t^m\}_{t=1}^n$  de padrões normalizados é igual a 0 e o desvio-padrão, igual a 1.

$$x_t^m = (x_t - \bar{x}_n) / S_x \quad (3.7)$$

- (d) **Método Sigmoidal:** utiliza a média e o desvio-padrão do conjunto  $\{x_t\}_{t=1}^n$  de padrões, os quais são, respectivamente, denotados por  $\bar{x}_n$  e  $S_x$ . A normalização de cada ponto  $x_t$  em  $\{x_t\}_{t=1}^n$  ocorre de acordo com a transformação definida em (3.8), que toma valores no intervalo  $[0,1]$ .

$$x_t^m = 1 / (1 + \exp\{(x_t - \bar{x}_n) / S_x\}) \quad (3.8)$$



### 3.3

#### RNA' s Multicamadas *Feedforward*, com uma Camada Escondida

Seja  $C[a, b]$  a coleção de todas as *funções reais contínuas* [veja KUBRUSLY (2001)]  $f(\cdot):[a, b] \rightarrow \mathbb{R}$  definidas sobre o intervalo limitado e fechado  $[a, b]$ , onde  $a < b$  e  $a, b \in \mathbb{R}$ . Tome a *norma* [KUBRUSLY (2001)]

$\|\cdot\|_{\infty} : C[a, b] \rightarrow \mathbb{C}$ , definida por  $\|f(\cdot)\|_{\infty} = \sup_{x \in [a, b]} |f(x)|$ , onde  $f(\cdot) \in C[a, b]$  e

$\sup_{x \in [a, b]} |f(x)|$  denota o *supremo* [veja KUBRUSLY (2001)] da função  $f(\cdot)$  sobre

$[a, b]$ . Considere ainda que  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  seja um mapa real contínuo de  $\mathbb{R}$  em  $\mathbb{R}$ ,

tal que  $\varphi(t) = \begin{cases} 0, & t \rightarrow -\infty \\ 1, & t \rightarrow \infty \end{cases}$ , onde  $t \in \mathbb{R}$ . Em tais condições, de acordo com

SYBENKO (1989),  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  é classificado como *função sigmoidal*. Em

SYBENKO (1989), mostra-se que o espaço linear  $\text{span}\{\varphi_k\}_{k=1}^K$  sobre o corpo real

$\mathbb{R}$  (que, de acordo com KUBRUSLY (2001), é a coleção de todas as

combinações finitas  $\sum_{k=1}^K \gamma_k \varphi_k$ , onde  $\gamma_k \in \mathbb{R}$ , para todo  $k=1, \dots, K$ , e  $\varphi_k: \mathbb{R} \rightarrow \mathbb{R}$  é a

$k$ -ésima função sigmoidal da sequência  $\{\varphi_k\}_{k=1}^K$ ) é *denso* [veja KUBRUSLY

(2001)] no *espaço topológico*  $(C[a, b], \|\cdot\|_{\infty})$  [veja KUBRUSLY (2011)]. Isto é,

$(\text{span}\{\varphi_k\}_{k=1}^K)^- = (C[a, b], \|\cdot\|_{\infty})$ , onde  $(\text{span}\{\varphi_k\}_{k=1}^K)^-$  denota o *fecho* [veja

KUBRUSLY (2001)] do espaço linear  $\text{span}\{\varphi_k\}_{k=1}^K$ . Com efeito, a sequência

$\left\{ \sum_{k=1}^K \gamma_k \varphi_k \right\}_{k=1}^K$  de combinações lineares de funções sigmoidais converge para o

limite  $f(\cdot)$ , em  $(C[a, b], \|\cdot\|_{\infty})$ , quando  $K \rightarrow \infty$ . Ou, simplesmente,

$\lim_{K \rightarrow \infty} \left\| \sum_{k=1}^K \gamma_k \varphi_k(\cdot) - f(\cdot) \right\|_{\infty} = 0$ . Em outras palavras, de acordo com o *Teorema da*

*Aproximação Universal* [veja SYBENKO (1989)] e com a definição formal de

*convergência uniforme*, definida em KUBRUSLY (2001), dado que



$\left(\text{span}\{\varphi_k\}_{k=1}^K\right)^- = \left(C[a,b], \|\cdot\|_\infty\right)$ , segue que, para todo  $\varepsilon > 0$ , existe um  $K_\varepsilon \in \mathbb{Z}$  tal que:

$$K \geq K_\varepsilon \Rightarrow \left\| \sum_{k=1}^K \gamma_k \varphi_k(\cdot) - f(\cdot) \right\|_\infty \leq \varepsilon. \quad (3.9)$$

Onde o símbolo  $\Rightarrow$ , em (3.9), significa “implica”. Na Definição 3.1, a seguir, entenda o termo “*RNA multicamada feedforward*” como sendo uma *RNA feedforward*, com uma camada intermediária com  $K$  neurônios com as mesmas funções de ativação e uma camada de saída com 1 neurônio cuja função de ativação é do tipo linear.

**Definição 3.1** - Seja  $\{x_t\}_{t=1}^n$  um conjunto de padrões, onde  $x_t \in \mathbb{R}$ , para todo  $t = 1, \dots, n$ . Assuma que o conjunto  $\{W_{tk}\}_{t=1}^J$ , onde  $W_{tk} \in \mathbb{R}$ ,  $t = 1, \dots, n$  e  $k = 1, \dots, K$ , seja o conjunto de pesos sinápticos associados ao  $k$ -ésimo neurônio da camada intermediária e que  $J$  seja o tamanho da janela neural. Uma *RNA multicamada feedforward* é um mapeamento

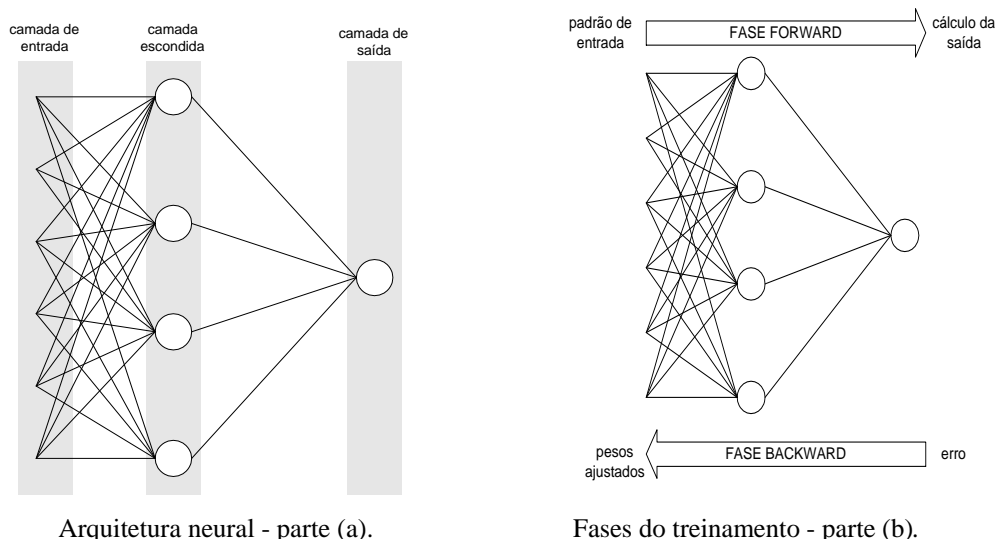
$$RNA_K(\cdot): \begin{cases} \mathbb{R}^J \rightarrow \mathbb{R} \\ (x_{t-1}, \dots, x_{t-J}) \in \mathbb{R}^J \end{cases} \xrightarrow{RNA} \in RNA(t) \in \mathbb{R} \quad \text{de } \mathbb{R}^J \text{ em } \mathbb{R}, \text{ definido pela}$$

combinação linear (finita)  $\sum_{k=1}^K \alpha_k \varphi_k(v_k) + \alpha$ , onde:  $v_k := \left( \sum_{j=1}^J x_{t-j} W_{tk} + \alpha_k \right) \in \mathbb{R}$  é a

regra de propagação do  $k$ -ésimo neurônio da camada escondida;  $\varphi_k(\cdot)$  é a  $k$ -ésima função de ativação sigmoideal da camada escondida; e  $\alpha_k \in \mathbb{R}$  e  $\alpha \in \mathbb{R}$  são, respectivamente, o  $k$ -ésimo peso sináptico e o intercepto da camada da camada saída. Dado o *Teorema da Aproximação Universal* e a Definição 3.1, verifica-se, trivialmente, que, de fato,  $RNA_K(\cdot) \in \left(\text{span}\{\varphi_k\}_{k=1}^K\right)^-$ , de forma que uma *RNA multicamada feedforward*  $RNA_K(\cdot)$  é, portanto, um aproximador de qualquer função  $f(\cdot)$  em  $C[a,b]$ , na norma  $\|\cdot\|_\infty$ .



Na Figura 3.3, têm-se: na parte (a), a ilustração de uma arquitetura de uma RNA multicamada (MLP) *feedforward*, com uma camada escondida com 4 neurônios, 1 neurônio na camada de saída e janela  $J$  igual 6; e, na parte (b), as duas fases do *algoritmo de treinamento backpropagation* [veja HAYKIN (2001), RUSSELL & NORVIG (1995)]: a fase *forward* e a fase *backward*.



Arquitetura neural - parte (a).

Fases do treinamento - parte (b).

Figura 3.3 - RNA MLP *feedforward*.

O *backpropagation* é um dos principais algoritmos de treinamento neural. O ajuste numérico dos pesos sinápticos, com o uso deste algoritmo, ocorre por meio de um processo de otimização de duas fases básicas: a *forward* e a *backward*. Na fase *forward*, calcula-se a resposta fornecida pela RNA a partir de um conjunto de padrões de entrada, que depende do tamanho da janela neural. Na fase *backward*, o erro entre a resposta da RNA e a resposta desejada é utilizado no processo de atualização dos pesos sinápticos. Ao longo do treinamento, os padrões de entrada e as respectivas respostas desejadas são apresentados à RNA, de forma a minimizar o erro quadrático médio, da amostra de treino, entre os sinais de saída da RNA e os sinais de saída desejados.



### 3.4

#### Geração de Previsões via Redes Neurais Artificiais

A previsão de valores futuros de uma série temporal, com o uso de uma RNA *feedforward* (com treinamento supervisionado), com uma camada escondida e uma camada de saída com um neurônio artificial, é possível porque as suas observações (sinais) podem ser interpretadas como pontos de alguma função de suporte compacto e, de acordo com a Seção 3.3, tais RNA's são aproximadores deste tipo de mapeamento, na topologia  $\| \cdot \|_{\infty}$ . Assim sendo, previamente, define-se o tamanho da janela neural  $L$  e do horizonte de previsão  $h$  desejado. Em uma série temporal com estrutura de dependência temporal *autoregressiva* [veja HAMILTON (1994)] (linear ou não linear), tem-se que o conjunto de padrões de entrada é constituído pelos valores passados da própria série temporal que se deseja projetar. O padrão de saída desejado, neste caso, consiste no valor a ser previsto.

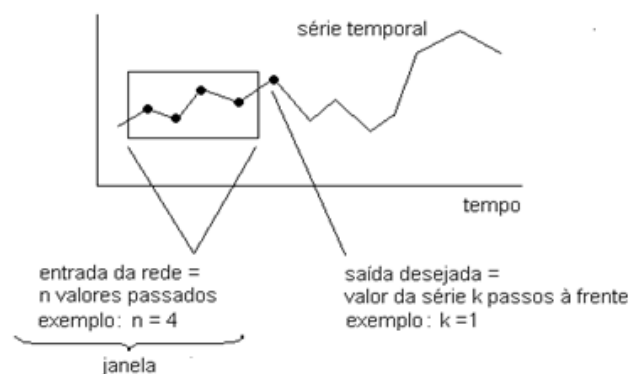


Figura 3.4 - Ilustração da dinâmica de previsão, 1 passo à frente, de uma série temporal, com estrutura de autodependência igual a 4, através de uma RNA.

Na Figura 3.4, tem-se a ilustração da dinâmica de previsão, 1 passo à frente, de uma série temporal, com estrutura de autodependência com defasagem igual a 4 (ou seja,  $L$  é igual a 4), através de uma RNA *feedforward*. Note que o vetor de padrões de entrada é constituído pelos últimos 4 sinais temporais, antes do padrão de saída a ser previsto. Para prever o sinal subsequente, a janela da RNA é deslocada um passo à frente [veja CAMPOS (2010)].



## COMBINAÇÃO DE MÉTODOS PREDITIVOS VIA PROGRAMAÇÃO MATEMÁTICA

### 4.1

#### Combinação de Métodos Preditivos

Tome uma série temporal  $\{y_t\}_{t=1}^T$  definida sobre algum espaço de probabilidade  $(\Omega, \mathcal{X}_\Omega, \rho)$  e considere que se tenha o interesse em modelá-la, a fim de se gerar previsões, dentro e fora da amostra. Assuma que, para tal, se tenha disponível um conjunto com  $M$  métodos preditivos plausíveis, o qual é denotado por  $\{\mu_m\}_{m=1}^M$ , onde  $M \in \mathbb{N} \setminus \{0,1\}$ , sendo  $\mathbb{N}$  o conjunto dos números inteiros positivos.

De acordo com FARIA & MUBWANDARIKWA (2008), existem, basicamente, duas possibilidades em se determinar um método preditivo gerador de previsões para a série temporal  $\{y_t\}_{t=1}^T$ :

1. Escolher um método preditivo em  $\{\mu_m\}_{m=1}^M$ , baseado em algum de seleção; ou
2. Escolher  $K$  métodos preditivos em  $\{\mu_m\}_{m=1}^M$  e, em seguida, combiná-los (de forma linear ou não linear).

De acordo com FLORES & WHITE (1989), no processo de combinação de métodos preditivos, é importante se considerar dois aspectos: o primeiro concerne à escolha dos  $K$  métodos preditivos base em  $\{\mu_m\}_{m=1}^M$ ; e o segundo, quanto à forma de se combiná-los. Em relação ao primeiro aspecto, os autores destacam que as previsões oriundas dos métodos preditivos base podem ser classificadas em três categorias: *objetivas*, *subjetivas* e *mistas* (isto é, as resultantes da combinação de previsões objetivas e subjetivas). As *previsões objetivas* são aquelas oriundas de previsores com base teórica estatística ou matemática; enquanto que as *previsões subjetivas* são as decorrentes de



juízo humano (geralmente, grupo focado ou opinião de especialistas [veja WERNER & RIBEIRO (2006)]). Salienta-se que, nesta pesquisa, todos os esforços se concentram somente na combinação de previsões objetivas. Quanto ao segundo aspecto, embora haja um consenso entre os pesquisadores de que a combinação de previsores individuais, em regra, incorre em ganhos preditivos na projeção de séries temporais, não existe um senso comum em se definir a melhor forma de se fazê-la [veja WALLIS (2011)].

Para a Definição 4.1, tome uma série temporal  $\{y_t\}_{t=1}^T$  em  $(\Omega, \mathcal{X}_\Omega, \rho)$  e assumamos que exista o interesse em se proceder à sua modelagem, a fim de se gerar previsões, dentro e fora da amostra, por meio de uma função de combinação de métodos preditivos base. Considere  $\nabla^K$  seja o conjunto de todas as previsões (dentro e fora da amostra) oriundas dos  $K$  métodos preditivos individuais escolhidos, pelo tomador de decisão, em  $\{\mu_m\}_{m=1}^M$ , onde  $K \leq M$ ; e tome um mapa

$$\hat{y}_C : \begin{cases} \nabla^K \rightarrow \mathbb{R} \\ \left[ \hat{y}_{t,k} \right]_{k=1}^K \in \nabla^K \rightarrow \hat{y}_{t,C} \in \mathbb{R} \end{cases} \quad \text{de combinação de previsões (genérico) que}$$

transforma, de acordo com sua definição, um vetor  $\left[ \hat{y}_{t,k} \right]_{k=1}^K \in \nabla^K$  de previsões em uma previsão combinada  $\hat{y}_{t,C} \in \mathbb{R}$ , para todo  $t=1, \dots, T+h$ , onde  $h \in \mathbb{N}$  é o horizonte de previsão.

#### 4.1.1

##### Combinação Linear de Métodos Preditivos

As *combinações lineares de métodos preditivos* (abordagem clássica e bayesiana) são vastamente citadas na literatura de projeção de séries temporais. Em CLEMEN (1989), por exemplo, é feito um detalhamento sobre as combinações lineares de previsões (abordagem clássica); enquanto que as combinações lineares de densidades preditivas (abordagem bayesiana) são detalhadas em GENEST & ZIDEK (1986), *apud in* MUBWANDARIKWA (2007). Na Seção 4.1.1.1, é definida a combinação linear de previsões tradicional; na Seção 6.3, porém, é definida e detalhada a combinação linear de previsões proposta.



#### 4.1.1.1

### Combinação Linear de Previsões Tradicional

De acordo com CLEMEN (1989), a combinação de previsões consiste em uma abordagem clássica que utiliza algum mecanismo ponderador (calculado de acordo com algum objetivo) associado às previsores a serem combinadas. Geralmente, adota-se, como objetivo, a minimização da estatística MSE (*mean square error*).

MAKRIDAKIS & WINKLER (1983) propuseram a combinação linear de previsões oriundas de  $K$  métodos em  $\{\mu_m\}_{m=1}^M$ , onde  $K \leq M$ , utilizando pesos adaptativos variantes no tempo. Os autores verificaram que as previsões combinadas foram mais acuradas que as dos previsores individuais. Em EVANS (2003), *apud in* TEIXEIRA Jr (2009), foi proposto o uso de uma constante aditiva na combinação linear de previsões. O autor mostrou, em seus experimentos, que a constante adaptativa acarretou ganhos preditivos.

Em TEIXEIRA Jr (2009), foram linearmente combinadas previsões de três métodos preditivos base: um método de *amortecimento exponencial* [veja MORETTIN & TOLOI (2006)], um modelo de *Box & Jenkins* e uma rede neural artificial. Os resultados estatísticos associados às previsões linearmente combinadas foram superiores aos das previsões dos previsores individuais. No que tange aos pesos adaptativos lineares, foram assumidos como sendo invariantes no tempo, normalizados, não negativos e com soma igual a uma unidade, preservando a sua interpretabilidade probabilística; e, para obtenção de seus valores ótimos, foi utilizado, como objetivo, o mínimo da estatística MAPE (*mean absolute percentual error*).

**Definição 4.1** - Uma função  $\hat{y}_{CL} : \begin{cases} \nabla^K \rightarrow \mathbb{R} \\ \left[ \hat{y}_{t,k} \right]_{k=1}^K \in \nabla^K \rightarrow \hat{y}_{t,CL} \in \mathbb{R} \end{cases}$  de combinação linear

de previsões para a série temporal  $\{y_t\}_{t=1}^T$  é um mapeamento definido por:

$$\hat{y}_{CL,t} = \sum_{k=1}^K \rho_{k,t} \times \hat{y}_{t,k} + \alpha. \quad (4.1)$$



Onde, em  $t$ :  $\hat{y}_{CL,t}$  é a previsão linearmente combinada;  $\hat{y}_{t,k}$ , a previsão do  $k$ -ésimo método base;  $\rho_{k,t}$ , o peso adaptativo associado linearmente à previsão  $\hat{y}_{t,k}$ ;  $\alpha$  e, constante adaptativa aditiva.

## 4.2

### Constantes e Pesos Adaptativos

No que concerne aos valores - ótimos (abordagem via programação matemática) ou estimados (abordagem estatística) - dos parâmetros adaptativos (ou seja, das *constantes* e dos *pesos adaptativos*) das combinações lineares de previsores individuais, existem, na literatura de projeções de séries temporais, muitas propostas de se obtê-los. Não obstante, não existe aquela que seja considerada a mais eficiente.

Segundo FARIA & MUBWANDARIKWA (2008), é comum a imposição das restrições  $\sum_{k=1}^K \rho_{k,t} = 1$  e  $\rho_{k,t} \geq 0$  sobre os valores que os pesos adaptativos em  $\{\rho_{k,t}\}_{k=1}^K$  podem assumir, onde  $t=1, \dots, T+h$ , a fim de preservar a sua interpretabilidade probabilística. Note que, dada a Definição 4.1, se  $\sum_{k=1}^K \rho_{k,t} = 1$  e  $\rho_{k,t} \geq 0$ , então, caso um peso adaptativo assuma valor (ótimo ou estimado) igual a 1, todos os outros pesos adaptativos assumem, por consequência, valores iguais a zero. Tal situação é um caso especial de combinação linear de previsões, visto que, em  $t$ , a previsão linearmente combinada é idêntica à previsão do previsor base associado ao peso que assume valor igual a 1.

Em seguida, têm-se algumas abordagens para a obtenção de valores para os parâmetros adaptativos de combinação linear de métodos preditivos individuais:

- (a) Proposto por GRANGER & BATES (1969), o chamado *método combinação ideal* é uma abordagem estatística que minimiza a variância dos resíduos da combinação de previsões, utilizando a método de condição de primeira ordem;



- (b) Em SMITH & MAKOV (1978), *apud in* MUBWANDARIKWA (2007), o *método de Quase-Bayes* estima os pesos adaptativos por meio de um problema de classificação utilizando a distribuição de *Dirichlet*;
- (c) Conhecido como *método de pesos de Akaike* e proposto por TURKHEIMER *et al* (2003), *apud in* TEIXEIRA Jr (2009), utiliza o critério de informação de *Akaike* (AIC) [veja MUBWANDARIKWA (2007)] para obtenção de uma fórmula analítica para o cálculo dos pesos adaptativos;
- (d) Em MUBWANDARIKWA (2007), é proposto o denominado *método de pesos por priori informativa*, que estabelece, inicialmente, aos pesos adaptativos (interpretados como variáveis aleatórias) uma distribuição de probabilidade *a priori*. À medida que os dados são observados, a distribuição de probabilidade é corrigida, isto é, é gerada uma distribuição de probabilidade *a posteriori*, através da conjugação da distribuição de probabilidade *a priori* e a *função de verossimilhança* [veja CASSELA & BERGER (2001)] condicionada às observações;
- (e) Em TEIXEIRA Jr (2009), os valores para os parâmetros adaptativos da combinação linear de previsões tradicional são obtidos numericamente, por meio de um problema de otimização, cujo objetivo é o mínimo da estatística MAPE; e
- (f) Em TEIXEIRA Jr *et al* (2011a), os parâmetros adaptativos de uma combinação linear tradicional são interpretados como *variáveis de decisão* [RAGSDALE (2004)] e os valores ótimos são obtidos meio de um PPM multiobjetivo MINIMAX, com abordagem de programação por metas.

Em síntese, pode-se verificar que os parâmetros adaptativos das combinações lineares de métodos preditivos base podem ser considerados: variantes ou invariantes no tempo; restritos ou irrestritos; e ser interpretados como variáveis aleatórias ou variáveis de decisão. Para visualizar outros exemplos, veja as seguintes referências: MAKRIDAKIS (1989), WINKLER (1989), GUPTA & WILTON (1987), BATCHELOR & DUA (1995), FLORES & WHITE (1998),



ARMSTRONG (2001), ZOU & YANG (2004), FARIA & MUBWANDARIKWA (2008), TEIXEIRA Jr *et al* (2012b)].

### 4.3

#### Ajuste Numérico via Programação Matemática

##### 4.3.1

##### Programação Matemática

De acordo com RAGSDALE (2004), a *programação matemática* consiste em um campo da matemática que acampa métodos numéricos que visam a encontrar *soluções viáveis* [veja GOLDBARG & LUNA (2000)] ótimas que correspondem a valores extremos de um ou mais objetivos, que são representados por funções e chamados de *funções-objetivo*. Ainda de acordo com autor, o termo “*matemática*”, neste contexto, é usado de em sentido amplo, abrangendo não apenas os seus elementos familiares, mas também os tópicos relacionados à *lógica matemática*. Segundo WIERZBICHI (1980), nos problemas de programação matemática (ou, simplesmente, PPM’ s), são considerados três elementos básicos: as *decisões*, as *restrições* e a *função-objetivo*. As  $m$  *decisões*, geralmente, consistem nas quantidades a serem otimizadas (isto é, minimizada ou maximizada), de acordo com algum objetivo de interesse, e podem ser denotadas pela sequência  $\{X_{l'}\}_{l'=1}^m$ , onde  $X_{l'}$  representa a  $l'$ -ésima variável de decisão, onde  $1 \leq l' \leq m$ . Por sua vez, a *função-objetivo* consiste em uma função dependente das variáveis de decisão que devem ser otimizadas e pode ser denotada, em um PPM, por  $f_{FO}(X_1, \dots, X_m)$ . As  $N$  *restrições*, em um PPM, podem ser expressas, em linhas gerais, de três maneiras: (a) restrição menor, denotada por  $f_n(X_1, \dots, X_m) < b_n$ ; restrição maior, denotada por  $f_n(X_1, \dots, X_m) > b_n$ ; e restrição igual, denotada por  $f_n(X_1, \dots, X_m) = b_n$ , onde  $n = 1, \dots, N$ . A notação  $b_n$  representa um valor numérico, supostamente conhecido e chamado de *parâmetro* de PPM, que integra a  $n$ -ésima restrição de um PPM [veja NASCIMENTO *et al* (2008)].



Os PPM's, de acordo com RAGSDALE (2004), podem ser classificados em duas classes básicas: os lineares [letra (a)]; e os não lineares [letras (b)].

- (a) **PPM's Lineares:** um PPM é classificado como *linear* se, e somente se, todas as suas restrições e a sua função-objetivo são lineares.
- (b) **PPM's não Lineares:** um PPM é classificado como *não linear* se, e somente se, a sua função-objetivo ou, ao menos, uma de suas restrições for não linear.

De acordo com RAGSDALE (2004), a formulação (genérica) de um PPM, com uma função-objetivo, pode ser dada conforme em (4.2).

$$\begin{aligned}
 &\text{Objetivo: } \underset{\{X_i\}_{i=1}^m}{\text{Min}} \quad f_{FO}(X_1, \dots, X_m) \\
 &s.a: \quad \left\{ \begin{array}{l} f_1(X_1, \dots, X_m) \leq b_1 \\ f_2(X_1, \dots, X_m) \geq b_2 \\ \cdot \\ \cdot \\ \cdot \\ f_N(X_1, \dots, X_m) \geq b_N \end{array} \right. \quad (4.2) \\
 &\text{Parâmetros: } \{b_n\}_{n=1}^N
 \end{aligned}$$

De acordo com WIERZBICHI (1980), no que concerne aos algoritmos de otimização utilizados na solução de um PPM, é importante salientar que não existem aqueles que sejam considerados universais, no sentido de terem eficácia satisfatória em todos os PPM's que permitem a sua aplicabilidade. Além disso, alguns algoritmos são aplicáveis somente à classe de PPM's lineares (como, por exemplo, o *SIMPLEX* [veja OLIVEIRA & CARRAVILLA (1998)]) e outros, apenas à classe dos não lineares (com, por exemplo, o algoritmo de *Levenberg & Marquardt* [veja GAVIN (2011)]).

Segundo ZELNY (1982), se um PPM envolve mais de uma função-objetivo, então é chamado de *problema de programação matemática multiobjetivo* (ou, simplesmente, *PPM multiobjetivo*). Com efeito, os PPM's simples (isto é, os PPM's com apenas uma função-objetivo) podem ser



interpretados como casos particulares de PPM' s multiobjetivos. De acordo com ZELNY (1974), as funções-objetivo de um PPM multiobjetivo podem ser transformadas em uma única função-objetivo, de modo que todos os algoritmos de otimização aplicáveis aos PPM' s simples também o são aos PPM' s multiobjetivos. Um caso particular são os *PPM' s multiobjetivos MINIMAX*, descrito na Seção 4.3.2.

### 4.3.2

#### PPM Multiobjetivo MINIMAX

De acordo com RAGSDALE (2004), o conceito do *PPM multiobjetivo MINIMAX* derivada da *teoria de jogos*, na qual é necessário resolver situações de conflito, com várias funções-objetivas a serem otimizados, simultaneamente. Nestes PPM' s, são comparados, para cada  $n = 1, \dots, N$ , os desvios absolutos entre uma  $n$ -ésima solução viável, denotada por  $f_n(X_1, \dots, X_m)$ , e a sua respectiva *solução objetivo* [veja GOLDBARG & LUNA (2000)], denotada por  $Objetivo_n$ . De acordo com RAGSDALE (2004), a formulação geral de um PPM multiobjetivo MINIMAX é dada em (4.3).

$$\begin{aligned}
 & \text{Objetivo: } \underset{\{X_i\}_{i=1}^m, Q}{\text{Min}} \quad Q \\
 & \text{s.a:} \\
 & \left\{ \begin{aligned} & \left( \left| Objetivo_1 - f_1(X_1, \dots, X_m) \right| / Objetivo_1 \right) \leq Q \\ & \left( \left| Objetivo_2 - f_2(X_1, \dots, X_m) \right| / Objetivo_2 \right) \leq Q \\ & \vdots \\ & \left( \left| Objetivo_N - f_N(X_1, \dots, X_m) \right| / Objetivo_N \right) \leq Q \end{aligned} \right. \quad (4.3) \\
 & \text{Parâmetros: } \{Objetivo_n\}_{n=1}^N
 \end{aligned}$$

Em (4.3), note que  $\left( \left| Objetivo_n - f_n(X_1, \dots, X_m) \right| / Objetivo_n \right)$  é o  $n$ -ésimo desvio absoluto percentual e  $Q$  é, ao mesmo tempo, uma variável de decisão e a



função-objetivo, a qual assume um valor máximo mínimo que restringe cada desvio absoluto percentual na sequência

$\left\{ \left( \left| \text{Objetivo}_n - f_n(X_1, \dots, X_m) \right| / \text{Objetivo}_n \right) \right\}_{n=1}^N$  de desvios absolutos percentuais.

Existem várias abordagens em programação matemática que podem ser utilizadas para a obtenção de cada  $\text{Objetivo}_n$ , onde  $n=1, \dots, N$ , as quais podem ser verificadas em SAWARAGI *et al* (1985) e HAIMES & HALL (1974).

### 4.3.3

#### Programação Multiobjetivo MINIMAX, com Abordagem de Programação por Metas

Na classe dos PPM's multiobjetivos MINIMAX, convém destacar o *PPM multiobjetivo MINIMAX, com programação por metas*. De acordo com RAGSDALE (2004), a sua ideia principal *a priori*, é encontrar uma solução viável ótima para cada  $\text{Objetivo}_n$ , chamada de *meta* e denotada por  $\text{Meta}_n$ . Por exemplo, se existe o interesse em se minimizar o  $n$ -ésimo objetivo, então  $\text{Min}\{\text{Objetivo}_n\} = \text{Meta}_n$ , onde  $n=1, \dots, N$ , sendo que *Min* significa “mínimo”.

Não obstante, quando a obtenção de todas as metas for, simultaneamente, inviável, é possível se obter soluções tais que minimizem os desvios absolutos percentuais entre as funções-objetivo em  $\{f_n(X_1, \dots, X_m)\}_{n=1}^N$  e as suas respectivas metas em  $\{\text{Objetivo}_n\}_{n=1}^N$ , de forma que  $\left( \left| \text{Objetivo}_n - f_n(X_1, \dots, X_m) \right| / \text{Objetivo}_n \right) \leq Q$ , para cada  $n=1, \dots, N$ . As suas  $N$  metas em  $\{\text{Meta}_n\}_{n=1}^N$  podem ser obtidas por meio da otimização individual de cada função-objetivo, conforme o PPM simples, em (4.1), onde  $\text{Objetivo}_n: \text{Min}_{\{X_l\}_{l=1}^m} f_{FO,n}(X_1, \dots, X_m)$  representa o  $n$ -ésimo objetivo, onde  $n=1, \dots, N$  [veja TEIXEIRA Jr *et al* (2011a)].

De acordo com RAGSDALE (2004), a formulação geral teórica de um programa multiobjetivo MINIMAX, com abordagem de programação por metas é dada em (4.4).



$$\begin{aligned}
&\text{Objetivo: } \underset{\{X_i\}_{i=1}^m, Q}{\text{Min}} \quad Q \\
&s.a: \\
&\left\{ \begin{aligned} &(|Meta_1 - f_1(X_1, \dots, X_m)| / Meta_1) \leq Q \\ &(|Meta_2 - f_2(X_1, \dots, X_m)| / Meta_2) \leq Q \\ &\cdot \\ &\cdot \\ &(|Meta_N - f_N(X_1, \dots, X_m)| / Meta_N) \leq Q \end{aligned} \right. \quad (4.4) \\
&\text{Parâmetros: } \{Meta_n\}_{n=1}^N
\end{aligned}$$

Por fim, destaca-se que os PPM's multiobjetivos MINIMAX, com abordagens de programação por metas, requerem algoritmos de otimização que sejam compatíveis com os problemas em questão. Neste caso, o tomador de decisão investiga e escolhe em um conjunto de soluções viáveis multiobjetivas a mais satisfatória [veja ainda ZIONTS & WALLENIS (1976) e VANKAN & MAAS (2002)].

#### 4.3.4

#### Ajuste Numérico via PPM Multiobjetivo da Combinação de Previsões Proposta

Em TEIXEIRA Jr *et al* (2011a), pode-se verificar que o ajuste numérico da combinação linear de previsões tradicional através da PPM multiobjetivo MINIMAX, com abordagem de Programação por Metas, acarretou em ganhos preditivos, em seus experimentos, em todas as estatísticas de aderência, fora da amostra. Outro ponto que fomenta a sua utilização é que, em geral, a ordenação, quanto ao desempenho em termos de modelagem, das melhores combinações de previsões varia de uma estatística de aderência para outra, de modo que a utilização de uma estatística global que, simultaneamente, considera todas as outras, torna trivial a escolha da melhor. Ademais, foram observados ganhos preditivos, fora da amostra, em outros experimentos realizados durante a pesquisa.



A solução do PPM multiobjetivo MINIMAX, com abordagem de Programação por Metas, utilizado para o ajuste numérico das constantes e dos pesos adaptativos da combinação de previsões proposta é dada em duas partes básicas:

(1) encontrar as  $L$  metas associadas às  $L$  estatísticas de aderência escolhidas pelo tomador de decisão; e

(2) ajustar os parâmetros da combinação de previsões proposta, utilizando as  $L$  metas e todas as  $L$  estatísticas de aderência adotadas no passo 1.

A formulação do PPM multiobjetivo proposto é descrito, de forma detalhada, na Seção 6.4.



## MÉTODO DE SIMULAÇÃO DE QUASE-MONTE-CARLO

De acordo com ALBUQUERQUE (2008), o método *simulação de quase-Monte-Carlo* (ou, simplesmente, *método QMC*) é uma abordagem numérica de simulação similar ao tradicional método de *simulação de Monte-Carlo* (ou, simplesmente, *método MC*), no sentido de que ambas objetivam à geração de sequências de números independentes no intervalo  $[0,1]$  cuja distribuição de probabilidade é a *uniforme* [veja CASELLA & BERGER (2010)] - geralmente denotada por  $U[0,1]$ . Segundo KRYKOVA (2003), os campos de aplicabilidade dos métodos QMC e MC são, basicamente, os mesmos [veja LAW & KELTON (2000), ACWORTH *et al* (1997), OWEN (2000), NIEDERREITER (1992), TEZUKA (1998)], no entanto o método MC simula os chamados *números pseudo-aleatórios*; enquanto que e o método QMC, os chamados *números quase-aleatórios*.

### 5.1

#### Sequências de Números Pseudo-Aleatórios

De acordo com KRYKOVA (2003), embora as sequências de números pseudo-aleatórios sejam similares às sequências de *números aleatórios* [veja RIPLEY (1987)], no sentido de serem constituídas de números independentes e de possuírem propriedades estatísticas muito similares, são geradas por meio de algoritmos determinísticos desenvolvidos para tal finalidade - como, por exemplo, o *gerador aleatório congruente* [veja PASKOV (1995)]. A geração de uma sequência de números pseudo-aleatórios depende de um valor inicial a ser sorteado no intervalo  $[0,1]$  que pode ocorrer através de um algoritmo computacional. A partir dos números pseudo-aleatórios simulados, podem-se obter, com a utilização de algoritmos *inversores* [veja ALBUQUERQUE (2008),



RIPLEY (1987)], quantis de outra distribuição de probabilidade de interesse (como, por exemplo, a *normal-padrão* [veja CASELLA & BERGER (2010)]).

## 5.2

### Sequência de Números Quase-Aleatórios

Segundo DIAS (2013), na classe das sequências de números quase-aleatórios, a sequência de *van der Corput* é a mais utilizada em casos unidimensionais ( $D=1$ ), devido à sua eficiência e à sua simplicidade de implementação. De acordo com KRYKOVA (2003), o algoritmo de geração de números quase-aleatórios de *van der Corput* segue os passos (1), (2) e (3), a seguir.

- (1) Tome o valor de índice  $n$  (que está na base decimal) associado ao  $n$ -ésimo número (a ser calculado ainda) no intervalo  $[0,1]$ . Em seguida, expanda-o em termos de alguma base numérica  $p$ . Por exemplo, considere que  $n = 4$  (que é o índice relativo ao quarto número no intervalo  $[0,1]$ ) e que  $p = 2$  (ou, seja base binária). A expansão do número 4 em termos da base numérica binária é dada por:  $4 = 1x2^2 + 0x2^1 + 0x2^0$ . Com efeito,  $4_{10} = 100_2$ , onde  $4_{10}$  e  $100_2$  denotam, respectivamente, o números 4 (na base decimal) e 100 (na base binária);
- (2) Reflita o número na base numérica  $p$ , de forma a inverter a ordem dos algarismos. No exemplo, o número 100 (ou, de forma equivalente,  $100_2$ ) é refletido tal como: 001, onde  $a_0(4) = 0$ ,  $a_1(4) = 0$  e  $a_2(4) = 1$ ; e
- (3) Baseado no referido exemplo, o quarto número no intervalo  $[0,1]$  da sequência de *van der Corput* é igual a  $1/8$ , dada a seguinte operação:  $0x2^{-1} + 0x2^{-2} + 1x2^{-3} = 1/8$ .



As etapas (1), (2) e (3), acima, podem ser generalizadas para todo  $n \in \mathbb{N}$  e para toda base numérica  $b \in \mathbb{N}$ , tal como (a) e (b), a seguir [veja ALBUQUERQUE (2008)].

- (a) Expanda o valor de índice  $n$ , em termos de uma base numérica  $p$ , de acordo a equação em (5.1). Note que é uma generalização dos itens (1) e (2).

$$n = \sum_{j=0}^m a_j(n) b^j \quad (5.1)$$

- (b) Gere o  $n$ -ésimo número (na base numérica  $p$ ) da sequência de *van der Corput*, denotado por  $\Phi_p(n)$ , de acordo com a definição em (5.2). Verifique que esta etapa é uma generalização do item (3).

$$\Phi_p(n) := \sum_{j=0}^m a_j(n) b^{-j-1} \quad (5.2)$$

Na Tabela 5.1, são comparadas as propriedades estatísticas de 1000 números pseudo-aleatórios (obtidos via gerador aleatório congruente) e 1000 números quase-aleatórios (obtidos via algoritmo de *van der Corput*, na base numérica binária), com a distribuição de probabilidade uniforme  $U[0,1]$  teórica.

Tabela 5.1 - Propriedades estatísticas dos números quase-aleatórios e pseudo-aleatórios.

	Quase-Aleatórios	Uniforme	Pseudo-Aleatórios
Valor Mínimo	9,76563e-4	0	3,27922e-4
Valor Máximo	0,998047	1	0,999982
Média	0,498866	0,5	0,512323
Mediana	0,498047	0,5	0,521946
Desvio-Padrão	0,288661	0,288675	0,291302
Variância	0,083325	0,083333	0,084877
Assimetria	2,042816e-3	0	-0,077605
Curtose	1,794803	1,8	1,773279

Pode-se verificar, na Tabela 5.1, que a sequência de números quase-aleatórios apresenta valores estatísticos mais próximos aos parâmetros teóricos da



distribuição uniforme  $U[0,1]$ , na maioria das estatísticas consideradas. Por sua vez, nas referências HALTON (1960), FAURE (1982), NIEDERREITER (1987) e SOBOL (1967), *apud in* KRYKOVA (2003), podem ser encontrados algoritmos determinísticos que geram sequências de números quase-aleatórios para os casos multidimensionais (isto é,  $D \in \mathbb{N} \setminus \{1\}$ ).

O objetivo principal, de acordo com KRYKOVA (2003), do uso das sequências de números quase-aleatórios, em casos multidimensionais, consiste em distribuí-los, 2 a 2, em um plano cartesiano, de forma a não formarem *clusters*. Ainda de acordo com a autora, a convergência, na *topologia Euclidiana* [veja KUBRUSLY (2001)], quando  $N \rightarrow \infty$ , do método QMC é computacionalmente mais rápida, em relação ao método MC, sendo na ordem de  $\ln(N)^D / N$  para o método QMC, contra  $1/\sqrt{N}$  para método de MC, onde:  $D$  é o número total de *dimensões* [veja DIAS (2013)] e  $N$ , o número de simulações em cada dimensão. De acordo com SOBOL (1967), existem outras vantagens relevantes do método de QMC sobre o MC que fomentam a sua utilização: os resultados estatísticos, em relação à distribuição de probabilidade uniforme  $U[0,1]$ , são melhores, no caso unidimensional; e, principalmente, no âmbito multidimensional, é menos propenso à formação de *clusters*.

### 5.2.1

#### Método de Simulação de Quase-Monte-Carlo Híbrido

Segundo KRYKOVA (2003), o *método Quase-Monte-Carlo Híbrido* (ou, simplesmente, *método QMC Híbrido*) consiste um algoritmo numérico utilizado na simulação de  $N$  números quase-aleatórios para cada dimensão  $d$ , onde  $d = 1, \dots, D$ . De acordo com ALBUQUERQUE (2008), o método *QMC híbrido* pode ser implementado conforme os procedimentos (1) e (2), a seguir:

1. Para a geração de  $N$  números quase-aleatórios para a primeira dimensão, pode-se utilizar, a sequência de *van der Corput* na base binária; e



2. Para cada dimensão  $d$ , onde  $2 \leq d \leq D$ , calcula-se cada elemento na

sequência  $\left\{ \left\{ \Phi_p^d(n) \right\}_{d=1}^D \right\}_{n=1}^N$ , onde:  $\Phi_p^d(n) := \sum_{j=0}^{m_n} a_j^d(n) b^{-j-1}$  é o  $n$ -ésimo

número quase-aleatório da dimensão  $d$ , dada a definição da *permutação aleatória*, em (5.3).

$$a_j^d(n) := \sum_{i=j}^{m_n} \left\{ \left( \frac{i!}{j!(i-j)!} \right) \times \left[ a_j^{d-1}(n) \bmod b \right] \right\} \quad (5.3)$$

A permutação aleatória, em (5.3), é uma operação recursiva que preserva, de acordo com DIAS (2013), as propriedades de *intraindependência* e *interindependência* [veja BARROS (2004)] das sequências numéricas quase-aleatórias sobre  $U[0,1]$ , bem como a propriedade de *baixa discrepância* [veja TEZUKA (1998)] presentes na primeira dimensão em todas as outras dimensões. Na Tabela 5.2, têm-se os 8 primeiros números quase-aleatórios, calculados via método QMC Híbrido, das três primeiras dimensões, onde, na primeira dimensão, se tem uma sequência de *van der Corput*, com base binária.

Tabela 5.2 - Números quase-aleatórios do método QMC híbrido das três primeiras dimensões ( $d=1, 2, 3$ ).

N = 8	Dimensão 1 (base 2)	Dimensão 2	Dimensão 3
1	1/2	1/2	1/4
2	1/4	1/8	3/8
3	3/4	3/4	1/2
4	1/8	5/8	1/16
5	5/8	3/8	5/8
6	3/8	1/4	7/8
7	7/8	3/4	3/4
8	1/16	7/8	1/8

Dados os 1000 pares ordenados em  $U[0,1]^2$ , das dimensões 49 e 50, na Figura 5.1, simulados através do método QMC híbrido, pode-se verificar a não ocorrência de *clusters*, dada a uniformidade na distribuição dos pontos no plano, o que evidencia a preservação da interdependência entre as dimensões 49 e 50.



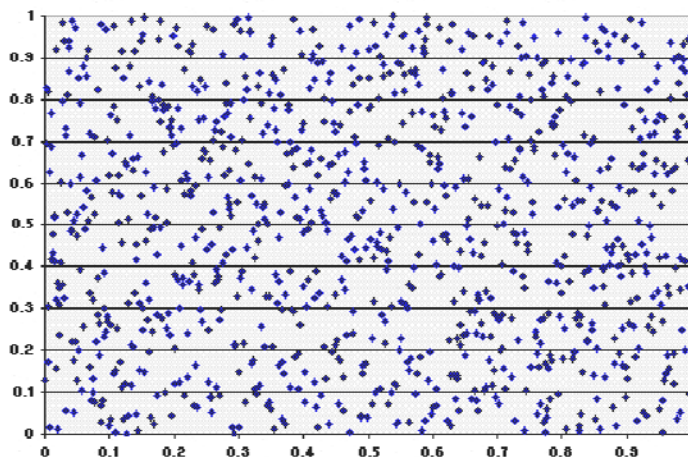


Figura 5.1 - Pares ordenados de números quase-aleatórios híbridos das dimensões 49 (ordenada) e 50 (abscissa) sobre  $[0,1]^2$ .

De acordo com KRYKOVA (2003): “As sequências de números quase-aleatórios híbridos construídas a partir permutações de van der Corput na base binária não apresentam problemas de correlação em ambientes de altas dimensões”.

### 5.3

#### Geração de Amostras de Distribuição Normal-Padrão

Tome uma sequência  $\left\{ \left\{ \Phi_p^d(n) \right\}_{d=1}^D \right\}_{n=1}^N$  de números gerados (por exemplo, através do método MC ou QMC) sobre o hipercubo  $[0,1]^D$ . Interprete cada ponto  $\Phi_p^d(n)$ , em  $[0,1]$ , como o valor de probabilidade acumulada associado ao quantil  $z_{\Phi_p^d(n)}$  de uma distribuição de probabilidade normal-padrão. Isto é:

$$\Phi_p^d(n) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{\Phi_p^d(n)}} \exp\left\{-\frac{z^2}{2}\right\} \partial z. \quad (5.4)$$

De acordo com ALBUQUERQUE (2008), um *inversor normal-padrão* consiste em um algoritmo que transforma  $\Phi_p^d(n)$  em  $z_{\Phi_p^d(n)}$  - que é interpretado como o quantil associado à probabilidade acumulada  $\Phi_p^d(n)$  -, para cada  $n$  e  $d$ . O *algoritmo de Box & Muller* [veja KRYKOVA (2003)] é um exemplo de inversor



normal-padrão, no entanto, quando utilizado na transformação de números quase-aleatórios, não é considerado o algoritmo mais apropriado, visto que não possui um desempenho satisfatório na modelagem das caudas da distribuição normal-padrão. Por outro lado, dado que o *inversor de Chebychev* [veja CHEBYCHEV (1990), *apud in* ALBUQUERQUE (2008)] possui desempenho satisfatório na modelagem das caudas de distribuições normal-padrão e o *inversor de Beasley-Springer* [BEASLEY & SPRINGER (1977)], na sua região central, segue que o *inversor de Moro* [veja MORO (1995)] se apresenta como uma solução híbrida que combina ambos os inversores tal como (1) e (2), a seguir.

1. A região central da distribuição de probabilidade normal-padrão, onde

$|U| \leq 0,42$  e  $U = \Phi_p^d(n) - 0,5$ , é modelada pelo inversor (normal-padrão)

de *Beasley-Springer*, onde  $z_{\Phi_p^d(n)} = U \left( \frac{\sum_{n=0}^3 a_n U^{2n}}{\sum_{n=0}^4 b_n U^{2n}} \right)$ , sendo que  $a_n$  e

$b_n$  são valores, na Tabela 5.3.

Tabela 5.3: Valores para  $a_n$  e  $b_n$  [fonte: KRYKOVA, 2003].

$n$	$a_n$	$b_n$
0	2,50662823884	1,00
1	-18,61500062529	-8,4735109309
2	41,39119773534	23,08336743743
3	-25,44106049637	-21,06224101826
4		3,13082909833

2. As caudas da distribuição de probabilidade normal-padrão, onde

$|U| > 0,42$ , são modeladas pelas séries truncadas de *Chebychev* (ou, simplesmente, inversor de *Chebychev*), definidas por em (5.5).

$$z_{\Phi_p^d(n)} = \begin{cases} \left( \sum_{n=0}^8 c_n T_n(x) \right) - c_0/2, & \text{se } U > 0 \\ c_0/2 - \left( \sum_{n=0}^8 c_n T_n(x) \right), & \text{se } U \leq 0 \end{cases}. \quad (5.5)$$



Onde:  $x = k_1 \left\{ 2 \ln \left[ -\ln(0,5 - |U|) \right] - k_2 \right\}$  e  $T_n(x) = \cos(n\theta)$ , onde  $\cos^{-1}(x) = \theta$ . Os valores das sequências  $\{c_n\}_{n=0}^8$  e  $\{k_j\}_{j=1}^2$  de valores, em (5.5), podem ser visualizados na Tabela 5.4, a seguir.

Tabela 5.4: Valores de  $c_n$  e  $k_j$  [fonte: KRYKOVA, 2003].

$n$	$c_n$	$k_j$
0	7,7108870705487895	
1	2,7772013533685169	0,4179886424926431
2	0,3614964129261002	4,2454686881376569
3	0,0373418233434554	
4	0,0028297143036967	
5	0,0001625716917922	
6	0,0000080173304740	
7	0,0000003840919865	
8	9,9999999129707170	

Segundo ALBUQUERQUE (2008), pode-se verificar que:  $x = -1$ , se  $\Phi_p^d(n) = 0,92$ ; e  $x = 1$ , se  $\Phi_p^d(n) = 1 - 10^{-12}$ . Dessa maneira, a partir dos números quase-aleatórios, pode-se obter, por meio do inversor de *Moro*, uma sequência de  $N$  amostras independentes de uma distribuição de probabilidade normal-padrão (também chamadas de *choques estocásticos quase-aleatórios normal-padrão*),

denotada por  $\left\{ \left\{ z_{\Phi_p^d(n)} \right\}_{d=1}^D \right\}_{n=1}^N$ .



Tabela 5.5: Estatísticas de 1000 choques estocásticos quase-aleatórios normal-padrão, via inversor de Moro, e parâmetros de uma distribuição de probabilidade normal-padrão (teórica) [fonte: KRYKOVA (2003)].

ESTATÍSTICAS	NORMAL-PADRÃO	INVERSOR DE MORO	
		<i>Pseudo-Aleatórios</i>	<i>Quase-Aleatórios</i>
<b>Média</b>	0	-0,0111	0,0002
<b>Mediana</b>	0	0,0009	0,0000
<b>Desvio-padrão</b>	1	0,9779	0,9886
<b>Variância</b>	1	0,9562	0,9772
<b>Assimetria</b>	0	0,0533	0,0097
<b>Curtose</b>	3	2,8451	2,7870

Na Tabela 5.5, são comparadas as estatísticas de 1000 choques estocásticos gerados, com o uso do inversor de Moro, a partir dos números pseudo-aleatórios (gerados por meio do algoritmo aleatório congruente) e dos números quase-aleatórios (gerados por meio do algoritmo de *van der Corput* na base binária), em relação aos parâmetros de uma distribuição de probabilidade normal-padrão teórica. Note ainda que os choques estocásticos quase-aleatórios normal-padrão obtiveram os melhores resultados, na maioria das estatísticas em análise. Outros resultados empíricos que motivam a utilização dos números quase-aleatórios híbridos, quando utilizado o inversor de *Moro* na geração de amostras normal-padrão podem ser verificados em KRYKOVA (2003) e ALBUQUERQUE (2008).

### 5.3.2

#### Geração de Choques Estocásticos de Distribuição Normal

Tome a sequência  $\left\{ \left\{ z_{\Phi_p^d(n)} \right\}_{d=1}^D \right\}_{n=1}^N$ . Uma vez escolhidos e estimados os

métodos preditivos a serem combinados (e, conseqüentemente, obtidas as respectivas previsões e os respectivos resíduos, dentro da amostra de treino), procede-se à simulação da distribuição de probabilidade empírica (ou cenários).



Para que os pontos da sequência  $\left\{ \left\{ z_{\Phi_p^d(n)} \right\}_{d=1}^D \right\}_{n=1}^N$  sejam transformados em choques estocásticos com a mesma escala de uma série temporal  $\{y_t\}_{t=1}^T$  de interesse, realiza-se sua filtragem através da *fatoração de Cholesky*, a qual é descrita, detalhadamente, em ALBUQUERQUE (2008) e HAMILTON (1994). Após isso, obtém-se uma nova sequência  $\left\{ \{a_{ij}\}_{i=1}^N \right\}_{j=1}^D$  de choques estocásticos normais, na escala de  $\{y_t\}_{t=1}^T$ , onde  $a_{ij}$  é o choque relativo ao  $i$ -ésimo cenário da dimensão (instante)  $j$  [veja Figura 5.2]

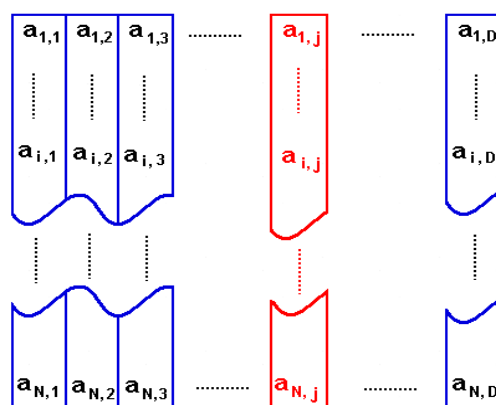


Figura 5.2 – Ilustração dos N choques estocásticos filtrados via fatoração de *Cholesky* para cada uma das D dimensões.



## METODOLOGIA PROPOSTA

### 6.1

#### Filtragem de uma Série Temporal via Método SSA

Considere uma série temporal  $\{y_t\}_{t=1}^T$  definida sobre um espaço  $(\Omega, \mathcal{X}_\Omega, \rho)$  de medida de probabilidade, a qual pode ser, equivalentemente, interpretada como um vetor  $Y_T := [y_t]_{1 \times T}$  no espaço linear real  $\mathbb{R}^T$   $T$ -dimensional. De acordo com o Capítulo 2, por meio do método de SSA, com abordagem de análise gráfica de autovetores singulares, podem-se obter as componentes SSA classificadas como ruído estatístico e, em seguida, eliminá-las, de forma a se obter o novo vetor  $[\tilde{y}_1, \dots, \tilde{y}_T]_{1 \times T}$  em  $\mathbb{R}^T$ , que consiste em uma versão filtrada da série temporal original  $Y_T := [y_t]_{1 \times T}$ . Assuma, nas seções seguintes, que a série temporal filtrada (via método SSA, com abordagem de análise gráfica de autovetores singulares) é denotada por  $\{y_{t,SSA}\}_{t=1}^T$ , ao invés de  $[\tilde{y}_1, \dots, \tilde{y}_T]_{1 \times T}$ .

### 6.2

#### Decomposição *Wavelet* de Nível $p$ de uma Série Temporal

Tome a série temporal  $\{y_{t,SSA}\}_{t=1}^T$ , que pode ser interpretada como o mapeamento  $y_{SSA}(\cdot): \{1, \dots, T\} \rightarrow \mathbb{R}$ , e considere a *função característica* [veja KUBRUSLY (2007)]  $\chi_{\{1, \dots, T\}}: \begin{cases} 1, \{1, \dots, T\} \\ 0, \mathbb{Z} \setminus \{1, \dots, T\} \end{cases}$ , onde  $\mathbb{Z}$  é o conjunto dos números inteiros. Com efeito, a *composição* [veja KUBRUSLY (2001)]  $y_{SSA}(\cdot)\chi = \tilde{y}(\cdot): \mathbb{Z} \rightarrow \mathbb{R}$  é definida por  $\tilde{y}_t = \begin{cases} y_{SSA,t}, & t \in \{1, \dots, T\} \\ 0, & t \in \mathbb{Z} \setminus \{1, \dots, T\} \end{cases}$ . Note que a função  $\tilde{y}(\cdot): \mathbb{Z} \rightarrow \mathbb{R}$  - ou, simplesmente,  $\tilde{y}(\cdot)$  - é, de fato, uma sequência infinita



$\{\tilde{y}_t\}_{t \in \mathbb{Z}}$  quadraticamente somável - ou seja,  $\sum_{t \in \mathbb{Z}} |\tilde{y}_t|^2 < \infty$  -, de maneira que  $\{\tilde{y}_t\}_{t \in \mathbb{Z}} = \tilde{y}(\cdot)$  é um vetor no espaço de *Hilbert*  $(l^2, \langle \cdot; \cdot \rangle)$  (complexo, de dimensão infinita e separável), onde  $l^2 := \left\{ \{\xi_t\}_{t \in \mathbb{Z}} : \sum_{t \in \mathbb{Z}} |\xi_t|^2 < \infty; \xi_t \in \mathbb{C}, \forall t \in \mathbb{Z} \right\}$  e a forma complexa  $\langle \cdot; \cdot \rangle : l^2 \rightarrow \mathbb{C}$  é uma norma definida sobre o espaço linear  $l^2$ .

Dado que a série temporal  $\tilde{y}(\cdot)$  é um elemento no espaço de *Hilbert*  $(l^2, \langle \cdot; \cdot \rangle)$ , segue que, de acordo com a Seção 2.2, pode ser ortogonalmente expandida, por meio da  $k$ -ésima série de *Fourier*, onde  $k = 1, \dots, K$ , definida em termos de uma  $k$ -ésima base ortonormal *wavelet*  $\{\phi_{m_0, n, k}(\cdot)\}_{n \in \mathbb{Z}} \cup \{\omega_{m, n, k}(\cdot)\}_{(m, n) \in \mathbb{Z} \times \{m\}_{m=m_0}^{+\infty}}$  de  $(l^2, \langle \cdot; \cdot \rangle)$ , em (6.1).

$$\tilde{y}(\cdot) = \sum_{n \in \mathbb{Z}} \langle \phi_{m_0, k, n}(\cdot); \tilde{y}(\cdot) \rangle \phi_{m_0, k, n}(\cdot) + \sum_{m=m_0, k}^{+\infty} \sum_{n \in \mathbb{Z}} \langle \omega_{m, n}(\cdot); \tilde{y}(\cdot) \rangle \omega_{m, n}(\cdot) \quad (6.1)$$

Onde, em  $(l^2, \langle \cdot; \cdot \rangle)$ , para cada  $k$ , têm-se:

$$\phi_{m_0, k, n}(\cdot) := D^{m_0, k} T^n \phi_k(\cdot) = 2^{\frac{m_0, k}{2}} \phi_k \left[ 2^{\frac{m_0, k}{2}} (\cdot) - n \right], \quad \omega_{m, n}(\cdot) := D^m T^n \omega_k(\cdot) = 2^{\frac{m}{2}} \omega_k \left[ 2^{\frac{m}{2}} (\cdot) - n \right],$$

$\tilde{y}_{V_{m_0, k}(\phi)}(\cdot) := \sum_{n \in \mathbb{Z}} \langle \phi_{m_0, k, n}(\cdot); \tilde{y}(\cdot) \rangle \phi_{m_0, k, n}(\cdot)$  é a componente de aproximação de nível

$m_0, k$  de  $\tilde{y}(\cdot)$ ; e  $\tilde{y}_{W_m(\omega)}(\cdot) := \sum_{n \in \mathbb{Z}} \langle \omega_{m, n}(\cdot); \tilde{y}(\cdot) \rangle \omega_{m, n}(\cdot)$  é a componente de detalhes de

nível  $m$  de  $\tilde{y}(\cdot)$ . A expansão, em (6.1), é, portanto, a  $k$ -ésima *decomposição wavelet* do vetor  $\tilde{y}(\cdot)$ .

De acordo com TEIXEIRA Jr et al (2012b), a *decomposição wavelet de nível  $p$* , onde  $p \in \mathbb{N}$ , é a  $k$ -ésima aproximação da série de *Fourier*, em (6.1), denotada por  $\tilde{y}_{approx, p, k}(\cdot)$ , e é dada, em termos genéricos, em (6.2).

$$\tilde{y}(\cdot) \cong \tilde{y}_{approx, p, k}(\cdot) = \sum_{n \in \Gamma_{V_{m_0, k}}} \langle \phi_{m_0, k, n}(\cdot); \tilde{y}(\cdot) \rangle \phi_{m_0, k, n}(\cdot) + \sum_{m=m_0, k}^{m_0, k + (p-1)} \left( \sum_{n \in \Gamma_{W_m}} \langle \omega_{m, n}(\cdot); \tilde{y}(\cdot) \rangle \omega_{m, n}(\cdot) \right) \quad (6.2)$$



Onde:  $\cong$  significa “aproximadamente”; e  $p$ , o nível das  $K$  decomposições ortogonais *wavelet*; e  $\Gamma_{V_{m_{0,k}}}$  e  $\Gamma_{W_m}$  são os subconjuntos finitos de índices incluídos em  $\mathbb{Z}$ . Salienta-se que existe uma gama de critérios para se definir os conjuntos de índices  $\Gamma_{V_{m_{0,k}}}$  e  $\Gamma_{W_m}$  [veja DAUBICHIES (1988)]. No *software* MATLAB, por exemplo, ao se realizar uma decomposição ortogonal *wavelet* de nível  $p$  de uma série temporal de interesse, o valor ótimo de  $m_{0,k}$  é obtido por meio de um problema de programação matemática cujo objetivo é o mínimo da função erro quadrático médio (EQM) entre os vetores  $\tilde{y}_{approx,p,k}(\cdot)$  e  $\tilde{y}(\cdot)$ . Isto é,

$$\underset{m_{0,k}}{Min} \left\{ \frac{\sum_{t=1}^T [\tilde{y}(t) - \tilde{y}_{approx,p,k}(t)]^2}{T} \right\}.$$

Após a realização das  $K$  decomposições ortogonais *wavelet* de nível  $p$ , obtêm-se os seguintes conjuntos de sinais *wavelet*:  $\left\{ \left\{ \tilde{y}_{V_{m_{0,k}}(\phi)}(t) \right\}_{t=1}^T \right\}_{k=1}^K$  e  $\left\{ \left\{ \tilde{y}_{W_m(\omega)}(t) \right\}_{t=1}^T \right\}_{m=m_{0,k}}^{m_{0,k}+(p-1)} \right\}_{k=1}^K$ . O próximo passo é o de realizar a modelagem individual, por meio de RNA's, de cada das  $(p+1) \times K$  componente ortogonais *wavelet* e, após isso, combinar as previsões obtidas, de acordo com a combinação proposta.

### 6.3

#### Combinação das Previsões para as Componentes *Wavelet*

Tome os conjuntos  $\left\{ \left\{ \tilde{y}_{V_{m_{0,k}}(\phi)}(t) \right\}_{t=1}^T \right\}_{k=1}^K$  e  $\left\{ \left\{ \tilde{y}_{W_m(\omega)}(t) \right\}_{t=1}^T \right\}_{m=m_{0,k}}^{m_{0,k}+(p-1)} \right\}_{k=1}^K$  de

sinais *wavelet*. Assuma que cada componente ortogonal *wavelet*, das  $K$  decomposições distintas de nível  $p$ , tenha sido modelada por uma RNA (*feedforward*; com uma camada intermediária; e um neurônio na camada de saída, com função de ativação linear) distinta. Desse modo, para cada  $k$ , os conjuntos



$\left\{\hat{y}_{V_{m_0,k}(\phi)}(t)\right\}_{t=1}^{T+h}$  e  $\left\{\left\{\hat{y}_{W_m(\omega)}(t)\right\}_{t=1}^{T+h}\right\}_{m=m_0,k}^{m_0,k+(p-1)}$  denotem, respectivamente, as previsões, dentro e fora da amostra (amostras de validação e de treino), para as componentes ortogonais *wavelet*  $\left\{\tilde{y}_{V_{m_0,k}(\phi)}(t)\right\}_{t=1}^T$  e  $\left\{\left\{\tilde{y}_{W_m(\omega)}(t)\right\}_{t=1}^T\right\}_{m=m_0,k}^{m_0,k+(p-1)}$ , onde  $h$  é o horizonte de previsão e  $m$  o parâmetro de escala que assume um valor inteiro no intervalo  $m_{0,k} \leq m \leq m_{0,k} + (p-1)$ . Considere ainda que  $\nabla_{V_{m_0}(\phi)} := \left\{\left\{\hat{y}_{V_{m_0,k}(\phi)}(t)\right\}_{t=1}^T\right\}_{k=1}^K$  e  $\nabla_{W_{m,k}(\omega)} := \left\{\left\{\hat{y}_{W_m(\omega)}(t)\right\}_{t=1}^T\right\}_{m=m_0,k}^{m_0,k+(p-1)}$ , para cada  $m$ , onde  $m_{0,k} \leq m \leq m_{0,k} + (p-1)$ .

**Definição 6.1** - A combinação de previsões (proposta)

$\hat{y}_{CL}(\cdot): \nabla_{V_{m_0}(\phi)} \cup \left\{\bigcup_{k=1}^K \nabla_{W_{m,k}(\omega)}\right\} \rightarrow \mathbb{R}$  é um mapa definido por:

$$\begin{aligned} \hat{y}_{CL}(t) = & \left\{ \sum_{k=1}^K \hat{y}_{V_{m_0,k}(\phi)}(t) \times \beta_{V_{m_0,k}(\phi)} + \alpha_{V_{m_0,k}(\phi)} \right\} \times \gamma_{V_{m_0}(\phi)} + \dots \\ & \dots + \left\{ \sum_{m=m_0,k}^{m_0,k+(p-1)} \left( \sum_{k=1}^K \hat{y}_{W_m(\omega)}(t) \times \beta_{W_m(\omega)} + \alpha_{W_m(\omega)} \right) \times \gamma_{W_m(\omega)} \right\} + \alpha \end{aligned} \quad (6.3)$$

$$\text{Onde: } \left\{ \beta_{V_{m_0,k}(\phi)}, \alpha_{V_{m_0,k}(\phi)} \right\}_{k=1}^K \cup \left\{ \left\{ \beta_{W_m(\omega)}, \alpha_{W_m(\omega)}, \gamma_{W_m(\omega)} \right\}_{k=1}^K \right\}_{m=m_0,k}^{m_0,k+(p-1)} \cup \left\{ \gamma_{V_{m_0}(\phi)}, \alpha \right\}$$

consiste no conjunto de parâmetros (pesos e constantes adaptativas) a serem otimizados, por meio de PPM multiobjetivo MINIMAX, com abordagem de Programação por Metas, o qual é descrito na Seção 6.4; e  $\hat{y}_{CL}(t)$  é a previsão linearmente combinada, em  $t$ . Portanto, o conjunto com todas as previsões combinadas, da Definição 6.1, dentro e fora da amostra, para a série temporal  $\{y_t\}_{t=1}^T$  são denotadas por  $\{\hat{y}_{CL}(t)\}_{t=1}^{T+h}$ , onde  $h$  é o horizonte de previsão.



## 6.4

### Combinação Linear de Previsões Proposta com Ajuste Numérico via PPM Multiobjetivo MINIMAX, com Abordagem de Programação por Metas

Nesta seção, é apresentada a formulação do problema de programação matemática multiobjetivo MINIMAX, com abordagem de programação por metas que foi utilizado para se obter os valores ótimos das constantes e dos pesos adaptativos da combinação de previsões proposta, na Definição 6.1.

O referido PPM multiobjetivo pode ser dividido em duas etapas básicas: (1) encontrar as  $L$  metas (por meio de PPM's simples) associadas às  $L$  estatísticas de aderência, escolhidas pelo tomador de decisão, que quantificam a o ajuste das previsões combinadas em relação à série temporal; e (2) ajustar a combinação proposta por meio do PPM multiobjetivo supracitado, utilizando as  $L$  metas e todas as  $L$  estatísticas de aderência do passo 1.

Em relação à etapa 1, a formulação teórica do PPM simples utilizada para o cálculo da cada meta  $l$  (ou valor ótimo de  $l$ -ésima estatística de aderência), onde  $1 \leq l \leq L$ , é dado abaixo.

$$FO: \underset{\alpha, \beta, \gamma}{Min} \{ Estatística_l \} = Meta_{Estatística_l}$$

Sujeito a:

$$\begin{aligned} \hat{y}_{CL}(t) = & \left\{ \sum_{k=1}^K \hat{y}_{V_{m_0,k}(\phi)}(t) \times \beta_{V_{m_0,k}(\phi)} + \alpha_{V_{m_0,k}(\phi)} \right\} \times \gamma_{V_{m_0}(\phi)} + \dots \\ & \dots + \left\{ \sum_{m=m_0,k}^{m_0,k+(p-1)} \left( \sum_{k=1}^K \hat{y}_{W_m(\omega)}(t) \times \beta_{W_m(\omega)} + \alpha_{W_m(\omega)} \right) \times \gamma_{W_m(\omega)} \right\} + \alpha \end{aligned}$$

$$Estatística_l = f_l \left( \{y_t\}_{t=1}^T, \{\hat{y}_{CL}(t)\}_{t=1}^T \right)$$

Utilizando o PPM simples acima, se obtém o conjunto  $\{Meta_{Estatística_l}\}_{l=1}^L$  de metas que são interpretadas como parâmetros no PPM multiobjetivo MINIMAX, com abordagem de programação por metas.



Note que, em relação ao valor de cada *meta*, é obtido por meio de um PPM com apenas um objetivo. Por exemplo, em relação à estatística MSE, são obtidos (utilizando algum algoritmo numérico plausível) valores ótimos para as constantes e pesos adaptativos para a combinação descrita, na Definição 6.1, de forma se obter o valor mínimo ótimo para a estatística MSE. Este valor ótimo consiste na *meta da estatística MSE*. O mesmo procedimento é realizado, individualmente, para cada uma das outras estatísticas de aderência escolhidas pelo tomador de decisão.

Em seguida, tem-se a formulação teórica do PPM multiobjetivo utilizado para obtenção dos valores ótimos para as constantes e pesos adaptativos (definitivas) da combinação de previsões proposta (que é a etapa 2). O objetivo, neste PPM, é a minimização da variável de decisão  $Q$ .

$$FO: \underset{\alpha, \beta, \gamma}{Min}\{Q\}$$

*Sujeito a:*

$$\left\{ \begin{array}{l} Estatística_1 = f_1\left(\{y_t\}_{t=1}^T, \{\hat{y}_{CL}(t)\}_{t=1}^T\right) \\ \cdot \\ \cdot \\ Estatística_L = f_L\left(\{y_t\}_{t=1}^T, \{\hat{y}_{CL}(t)\}_{t=1}^T\right) \end{array} \right.$$

$$\begin{aligned} \hat{y}_{CL}(t) = & \left\{ \sum_{k=1}^K \hat{y}_{V_{m_0,k}(\phi)}(t) \times \beta_{V_{m_0,k}(\phi)} + \alpha_{V_{m_0,k}(\phi)} \right\} \times \gamma_{V_{m_0}(\phi)} + \dots \\ & \dots + \left\{ \sum_{m=m_0,k}^{m_0,k+(p-1)} \left( \sum_{k=1}^K \hat{y}_{W_m(\omega)}(t) \times \beta_{W_m(\omega)} + \alpha_{W_m(\omega)} \right) \times \gamma_{W_m(\omega)} \right\} + \alpha \end{aligned}$$



$$\left\{ \begin{array}{l} \frac{|Estatística_1 - Meta_{Estatística_1}|}{|Meta_{Estatística_1}|} \leq Q \\ \cdot \\ \cdot \\ \cdot \\ \frac{|Estatística_L - Meta_{Estatística_L}|}{|Meta_{Estatística_L}|} \leq Q \end{array} \right.$$

Onde:  $\left\{ \left\{ \hat{y}_{V_{m_{0,k}}(\phi)}(t) \right\}_{t=1}^T \right\}_{k=1}^K$  e  $\left\{ \left\{ \hat{y}_{W_m(\omega)}(t) \right\}_{t=1}^T \right\}_{m=m_{0,k}}^{m_{0,k}+(p-1)} \right\}_{k=1}^K$  são os conjuntos de parâmetros, nos PPM's acima; e  $\left\{ \beta_{V_{m_{0,k}}(\phi)}, \alpha_{V_{m_{0,k}}(\phi)} \right\}_{k=1}^K \cup \left\{ \left\{ \beta_{W_m(\omega)}, \alpha_{W_m(\omega)}, \gamma_{W_m(\omega)} \right\}_{k=1}^K \right\}_{m=m_{0,k}}^{m_{0,k}+(p-1)} \cup \left\{ \gamma_{V_{m_0}(\phi)}, \alpha \right\}$ , o conjunto das variáveis de decisão comuns.

## 6.5

### Geração de Cenários QMC Híbridos

Tome as seguintes séries temporais de resíduos:

$$\left\{ \left\{ y_{W_m(\omega)}(t) - \hat{y}_{W_m(\omega)}(t) \right\}_{t=1}^T \right\}_{m=m_{0,k}}^{m_{0,k}+(p-1)} \right\}_{k=1}^K \text{ e } \left\{ \left\{ y_{V_{m_{0,k}}(\phi)}(t) - \hat{y}_{V_{m_{0,k}}(\phi)}(t) \right\}_{t=1}^T \right\}_{k=1}^K.$$

Para cada uma das componentes (ortogonais) *wavelet*, considere um horizonte de previsão igual a

$h$ , onde  $h \in \mathbb{N}$ , e tome a sequência  $\left\{ \left\{ z_{\Phi_p^d(n)} \right\}_{d=1}^D \right\}_{n=1}^N$  de choques estocásticos

normais-padrão independentes, gerados por meio do método de QMC híbrido - descrito no Capítulo 5 -, onde:  $D$  é o número de dimensões (que, no caso, é o horizonte de previsão, ou seja,  $h = D$ ); e  $N$  é o número de cenários em cada dimensão  $d$  onde  $d = t$ .

Em seguida, para cada uma das  $(p+1) \times K$  componentes *wavelet* modeladas acima, calcule a *matriz de Cholesky* [veja HAMILTON (1994)],



utilizando a respectiva série temporal de resíduos, de forma a possibilitar o procedimento de filtragem por *fatoração de Cholesky* [veja ALBUQUERQUE (2008)], através do qual, a sequência  $\left\{ \left\{ \mathbf{z}_{\Phi_p^d(n)} \right\}_{d=1}^D \right\}_{n=1}^N$  é transformada em outra sequência  $\left\{ \left\{ a_{ij} \right\}_{i=1}^N \right\}_{j=1}^D$  de choques estocásticos independentes de uma distribuição normal, na escala da respectiva componente *wavelet*. Destaca-se que, à medida que amostra de treino é atualizada, *matriz de Cholesky* é recalculada para o cálculo dos cenários subsequentes.

Para projeções *um passo à frente*, o  $i$ -ésimo cenário de cada componente ortogonal *wavelet*, no instante  $t$ , é dado por  $\bar{y}_i(t) = \hat{\hat{y}}_{wave}(t) + a_{it}$ , onde, em  $t$ :  $a_{it}$  é o  $i$ -ésimo choque estocástico normal; e  $\hat{\hat{y}}_{wave}(t)$  é a previsão da respectiva componente *wavelet*  $y_{wave}(t)$ ; e  $\bar{y}_{it}$  é o  $i$ -ésimo cenário normal para  $y_{wave}(t)$ . Por fim, após a geração dos  $N$  cenários, para todas as componentes ortogonais *wavelet*, no horizonte de previsão igual a  $h$ , é a realizada sua combinação tal como, em (6.5).

$$\begin{aligned} \bar{y}_{CL,i}(t) = & \left\{ \sum_{k=1}^K \bar{y}_{V_{m_{0,k}}(\phi),i}(t) \times \hat{\beta}_{V_{m_{0,k}}(\phi)} + \hat{\alpha}_{V_{m_{0,k}}(\phi)} \right\} \times \hat{\gamma}_{V_{m_{0,k}}(\phi)} + \dots \\ & + \left\{ \sum_{m=m_{0,k}}^{m_{0,k}+(p-1)} \left( \sum_{k=1}^K \bar{y}_{W_m(\omega),i}(t) \times \hat{\beta}_{W_m(\omega)} + \hat{\alpha}_{W_m(\omega)} \right) \times \hat{\gamma}_{W_m(\omega)} \right\} + \hat{\alpha} \end{aligned} \quad (6.5)$$

Onde:

$\left\{ \hat{\beta}_{V_{m_{0,k}}(\phi)}, \hat{\alpha}_{V_{m_{0,k}}(\phi)}, \hat{\gamma}_{V_{m_{0,k}}(\phi)} \right\}_{k=1}^K \cup \left\{ \left\{ \hat{\beta}_{W_m(\omega)}, \hat{\alpha}_{W_m(\omega)}, \hat{\gamma}_{W_m(\omega)} \right\}_{k=1}^K \right\}_{m=m_{0,k}}^{m_{0,k}+(p-1)} \cup \{\hat{\alpha}\}$  é o conjunto dos valores ótimos, obtidos por meio do PPM multiobjetivo descrito na Seção 6.4, para os parâmetros da combinação da Definição 6.1;  $\bar{y}_{CL,i}(t)$  é o  $i$ -ésimo cenário combinado, em  $t$ ; e,  $\bar{y}_{V_{m_{0,k}}(\phi),i}(t)$  é o  $i$ -ésimo cenário normal para  $y_{V_{m_{0,k}}(\phi)}(t)$  e  $\bar{y}_{W_m(\omega),i}(t)$ ,  $i$ -ésimo cenário normal para  $y_{W_m(\omega)}(t)$ , em  $t$ .

Para projeções *dois passo à frente ou mais*, o  $i$ -ésimo cenário de cada componente ortogonal *wavelet*, no instante  $t$ , é dado por:  $\bar{y}_i(t+1) = \bar{y}_i(t) + a_{it}$ ,



onde:  $a_{it}$  é o  $i$ -ésimo choque estocástico normal, em  $t$ ; e  $\bar{y}_i(t+1)$  é o cenário normal para a componente *wavelet*  $y_{wave}(t+1)$ , em  $t+1$ ; e  $\bar{y}_i(t)$  é o  $i$ -ésimo cenário normal para  $y_{wave}(t)$ , em  $t$ .



## EXPERIMENTOS COMPUTACIONAIS

Neste capítulo, é apresentado um detalhamento da metodologia proposta em um caso aplicado, explicitando os principais resultados empíricos e análises. Para tanto, foi realizado um experimento computacional que visou à modelagem da *série temporal mensal de vazão de afluentes da região sul do Brasil* [veja ONS], com cardinalidade igual a 936 observações, a fim de gerar previsões e de cenários, um passo à frente. A série temporal supracitada foi escolhida devido à sua relevância no setor energético brasileiro [veja CAMPOS (2010)] e por apresentar estrutura temporal com enorme complexidade, em termos de modelagem, sendo, inclusive, objeto de pesquisa de outras dissertações de mestrado e teses de doutorado.

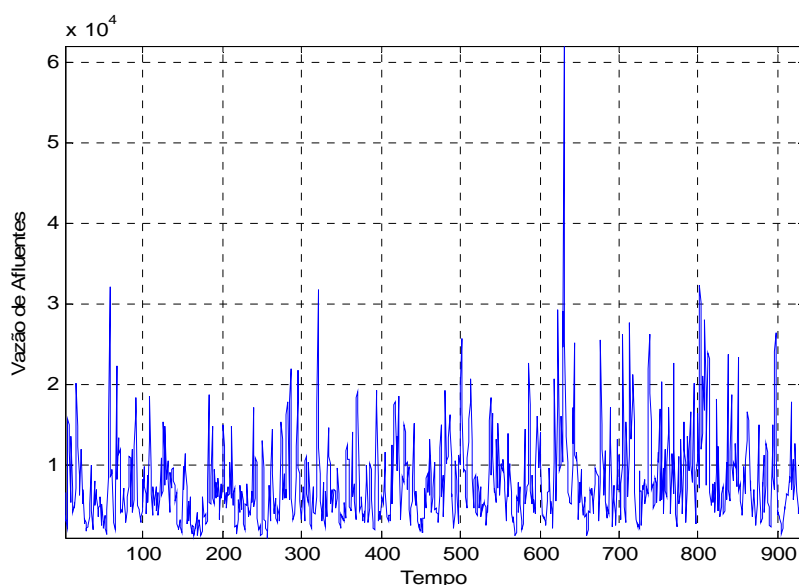


Figura 7.1 - Série temporal mensal da vazão de afluentes da região sul do Brasil.

Para efeito de comparação, foram ainda utilizados outras 8 abordagens consagradas na literatura, no experimento computacional proposto. São mostrados, a seguir, de forma detalhada, os resultados principais da modelagem, em um horizonte de previsão igual a 188 passos. Quanto às amostras utilizadas, tem-se que: 80% dos primeiros dados observados foram utilizados na obtenção do



valor ótimo do parâmetro  $L$  do método SSA, para a escolha das melhores bases ortonormais wavelet, para o treinamento das RNA's envolvidas e na otimização dos parâmetros adaptativos da combinação linear de previsões proposta; 10% dos dados observados subsequentes foram utilizados na amostra de fase validação das RNA's; e os dados remanescentes, na fase de teste.

## 7.1

### Filtragem da Série Temporal via Método SSA

Nesta etapa, foi realizada a filtragem da amostra de treino da série temporal de vazão de afluentes supracitada por meio do método SSA, com a abordagem de análise gráfica de vetores singulares na SVD. O valor para o parâmetro  $L$ , para o qual foi gerado o maior número de autovetores classificados como ruído, foi o 462 [veja CASSIANO *et al* (2012)]. Para tal, foram utilizados os *softwares* MATLAB (<http://www.mathworks.com>) e *Caterpillar SSA* (<http://www.gistatgroup.com/cat/index.html>) para obtenção de todos os resultados empíricos do método SSA neste experimento.

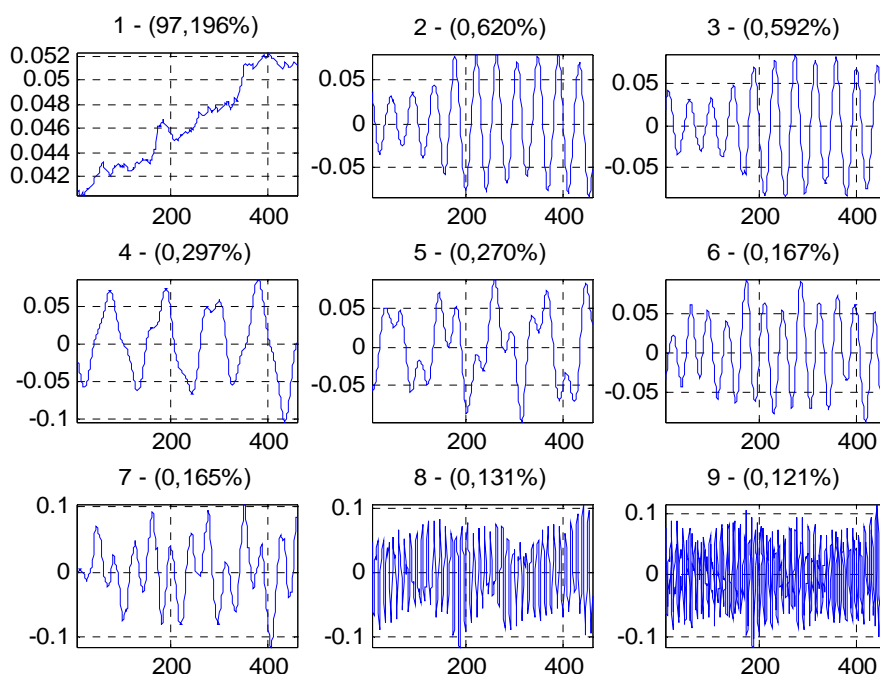


Figura 7.2 - 9 Primeiros autovetores singulares na SVD da matriz trajetória da amostra de treino da série temporal mensal de vazão de afluentes da região sul do Brasil.



Na Figura 7.2, podem-se visualizar os 9 primeiros autovetores singulares da SVD da matriz trajetória gerada a partir da amostra de treino. Da esquerda para a direita, têm-se os seguintes autovetores singulares: 1, 2 e 3, na primeira linha; 4, 5 e 6, na segunda linha; e 7, 8 e 9, terceira linha. As contribuições, em termos de valor singular, das 9 primeiras componentes SSA também podem ser verificadas, na Figura 7.2. Por exemplo, a contribuição da componente SSA 1 (que está associada ao autovetor singular 1, na SVD) é igual a 97,196%. Além disso, baseado nos gráficos dos autovetores singulares, na Figura 7.2, a componente SSA 1 foi classificada como tendência e as outras 8 componentes SSA's, como harmônicas [veja GOLYANDINA (2001)].

É importante salientar que, para cada nova observação agregada à amostra de treino inicial (que consiste em 80% das primeiras observações), no processo de geração de previsões um passo à frente, procedeu-se a uma nova decomposição SVD da nova matriz trajetória (gerada a partir desta amostra atualizada); mantendo-se, porém, o valor de janela igual a 462. Além disso, foi mantida a classificação previamente dada aos autovetores singulares, como baseado na amostra de treino inicial, de modo que as componentes SSA totais foram sempre constituídas pelas mesmas componentes SSA para todas as atualizações.

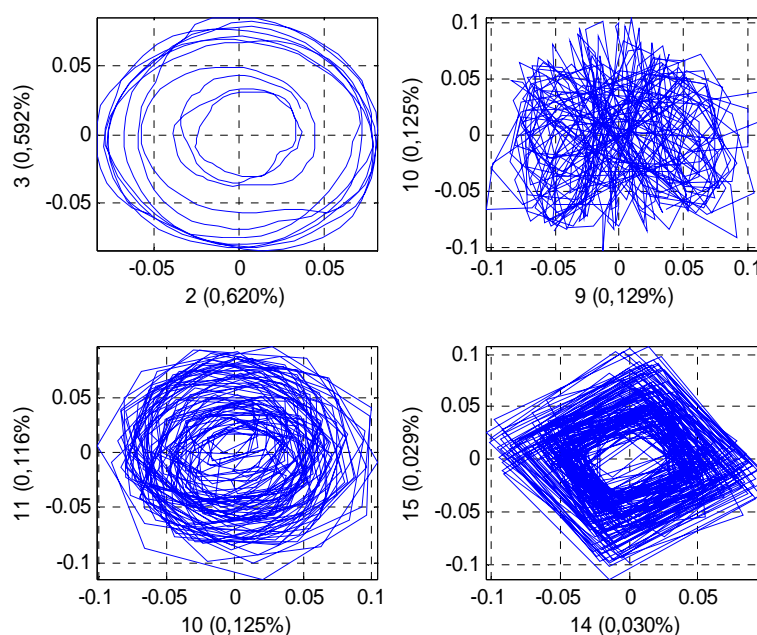


Figura 7.3 - Gráficos contínuos de dispersão dos pares de autovalores singulares da SVD: (2,3); (9,10); (10,11); e (14,15).



Na Figura 7.3, têm-se os gráficos (contínuos) de dispersão de alguns pares de autovetores singulares da SVD da matriz trajetória gerada a partir da amostra de treino (inicial) da série temporal em questão. Por meio de sua análise, podem-se verificar, ao menos, *a priori*, quais os integrantes de determinado par ordenado estão associados. Além disso, auxiliam na identificação de quais autovetores singulares são classificados como tendência, harmônica ou ruído [veja HASSANI (2007), GOLIANDINA (1997)].

Ainda na Figura 7.3, de acordo com GOLYANDINA (2001), note que os gráficos dos pares de autovetores singulares (2,3), (10,11), e (14,15), da SVD em análise, mostram que existe, em cada par, associação não linear entre os autovetores singulares que os compõem, de modo que não podem ser classificados como ruído. Por sua vez, o gráfico contínuo de dispersão dos autovetores singulares do par (9,10) mostra que estes não estão associados. Dado que o autovetor singular 10 não pode ser classificado como ruído (visto que está associado ao autovetor singular 11), existem evidências, *a priori*, que o autovetor singular 9 seja ruído. Por outro lado, o gráfico do par (2,3) evidencia, de acordo com HASSANI (2007), que os autovetores singulares 2 e 3 sejam componentes harmônicas (sendo passíveis, portanto, serem representadas por combinações lineares de funções senos ou cossenos).

Em última análise, de acordo com HASSANI (2007), pode-se verificar que os autovetores singulares dos pares (10,11) e (14,15) são componentes harmônicas, uma vez que os seus diagramas de dispersão exibem polígonos regulares. Ainda de acordo com o autor, os autovetores singulares 10 e 11 evidenciam, neste caso, um período igual a 7 meses, pois o seu gráfico de dispersão sugere a figura de um heptágono regular e a série temporal em análise é de frequência mensal; enquanto que os vetores 14 e 15 possuem período igual a 4 meses, na maneira que o seu gráfico do diagrama de dispersão sugere a figura de paralelogramo regular.

Baseado nas interpretações acima, procedeu-se, com o uso dos *softwares* Caterpillar SSA e MATLAB, à realização do reagrupamento e, posteriormente, à soma de todas as componentes SSA da amostra de treino inicial classificadas como ruído, de forma a gerar a componente SSA 3 [veja Figura 7.4].



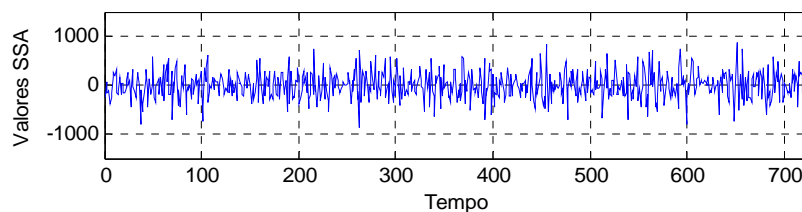


Figura 7.4 - Componente SSA 3 da amostra de treino inicial série temporal mensal de vazão de afluentes da região sul do Brasil.

Verificou-se que a componente total SSA 3 (que é dada pela soma de todas as componentes SSA da amostra de treino inicial cujos respectivos autovetores singulares foram classificados como ruído), de fato, possui, a 5% de significância, as propriedades estatísticas básicas de ruído, isto é: *independência* estatística [veja Tabela 7.1]; *estacionariedade de 2º ordem* [veja Tabela 7.2]; e *média populacional igual 0*, pois o *p-valor* obtido no teste *t* [veja CASELLA & BERGUER (2010)] foi igual a 80,23%, de forma a não se rejeitar, fortemente, a hipótese nula  $H_0$ , onde  $H_0$ : “média da componente SSA é igual a 0” e  $H_1$ : “ $H_0$  é falsa”. Os resultados do teste *t* e os testes BDS e *Dickey-Fuller*, nas Tabelas 7.1 e 7.2, abaixo, foram implementados no software *E-views* (<http://www.eviews.com/>).

Tabela 7.1 - Teste de independência estatística BDS (onde  $H_0$ : “independência estatística” e  $H_1$ : “ $H_0$  é falsa”) para a componente SSA 3.

Teste BDS			
Dimensão	Estatística BDS	Estatística Z	P-Valor
2	-0,0000412	-0,510289	0,5849
3	-0,0000917	-0,410241	0,6580
4	-0,0000566	-0,144583	0,9025
5	-0,0000244	-0,045879	0,9287
6	-0,0000302	-0,482100	0,6024

Tabela 7.2 - Testes de estacionariedade de *Dickey-Fuller* (onde  $H_0$ : “não estacionariedade de 2º ordem” e  $H_1$ : “ $H_0$  é falsa”) para a componente SSA 3.

Teste de Dickey-Fuller		
Estatística ADF	Nível	Valores Críticos
-13,22529	1%	-3,4402
	5%	-2,8651
	10%	-2,5687



A componente SSA 1 (classificada como *componente de tendência total*, visto que é definida pela soma de todas as componentes SSA da amostra de treino inicial cujos respectivos autovetores singulares foram classificados como tendência) e a componente SSA 2 (classificada como *componente harmônica total*, dado que é definida pela soma de todas as componentes SSA da amostra de treino inicial cujos respectivos autovetores singulares foram classificados como harmônico) foram somadas, de modo a gerar uma versão filtrada (via método SSA) da amostra de treino inicial da série temporal em análise.

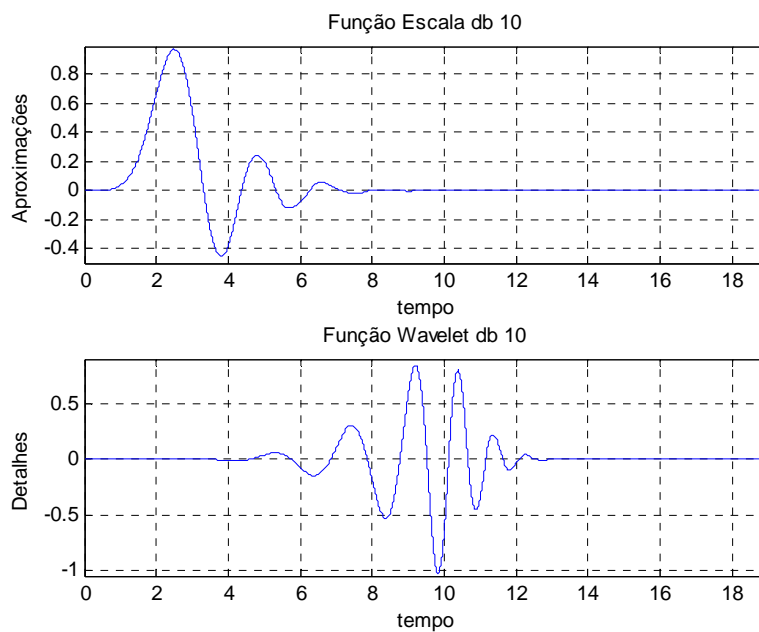
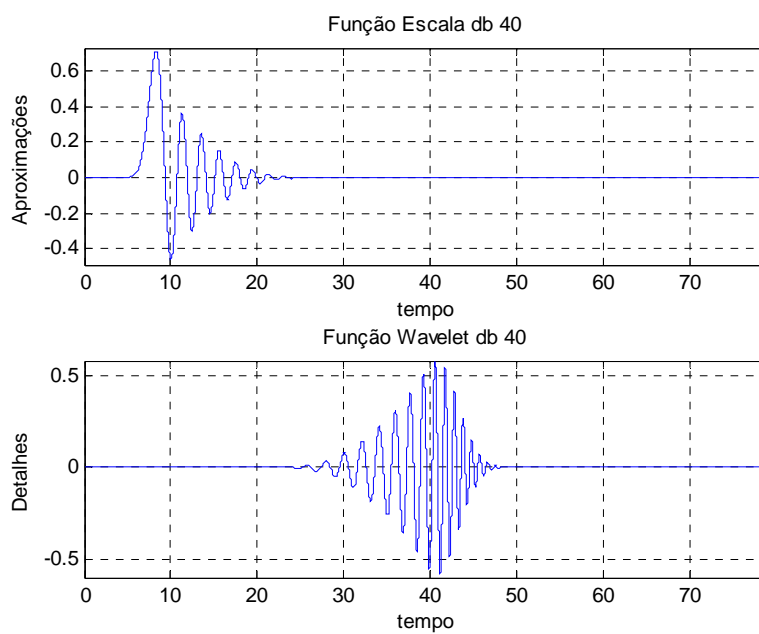
## 7.2

### Decomposição *Wavelet* da Série Temporal Filtrada via Método SSA

Nesta etapa, foram feitas as 3 diferentes decomposições *wavelet*, de nível 2, isto é, 3 decomposições ortogonais com diferentes bases ortonormais *wavelet*, com o auxílio do *software* MATLAB, da versão filtrada (via SSA) da amostra de treino inicial (ou seja, 80 % das primeiras observações) da série temporal em análise.

À medida que a amostra de treino foi sendo atualizada (ou seja, que uma nova observação foi agregada à amostra, no processo de previsão 1 passo à frente), foram recalculadas as 3 novas séries de *Fourier* (e, conseqüentemente, as 3 decomposições ortogonais *wavelet*). Neste processo, foram utilizadas as seguintes bases ortonormais *wavelet* da família de Daubechies: *db 10*, *db 40* e *db 30* [veja <http://www.mathworks.com/help/wavelet/ref/waveinfo.html>]. Utilizou-se, como critério de escolha das bases ortonormais *wavelet*, aquelas cujo erro quadrático médio entre os sinais filtrados (via SSA), da amostra de treino inicial, e o resultado da soma das 3 componentes *wavelet* era menores.



Figura 7.5 - Funções *wavelet db 10*.Figura 7.6 - Funções *wavelet db 40*.



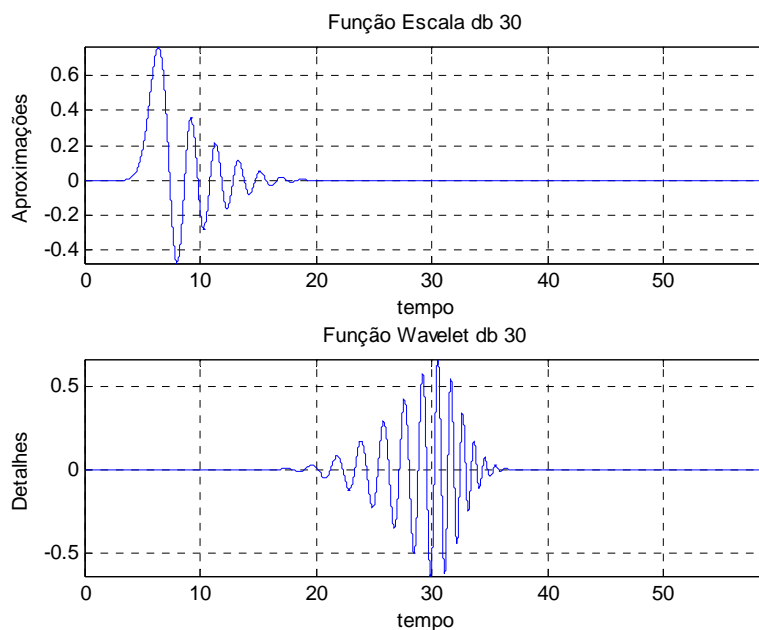


Figura 7.7 - Funções *wavelet db 30*.

Nas Figuras 7.8, 7.9 e 7.10, a seguir, têm-se, respectivamente, os gráficos das componentes ortogonais *wavelet* das decomposições da série temporal total, de nível 2, com as bases ortonormal *wavelet* da família de Daubechies *db 10*, *db 40* e *db 30*.

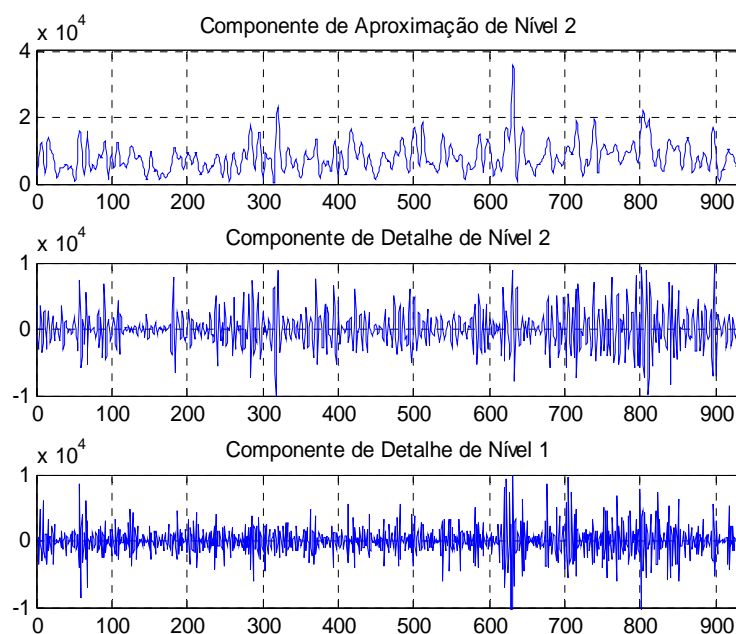


Figura 7.8 - Componentes *wavelet* de aproximação e de detalhe da decomposição de nível 2, com base ortonormal *db 10*, da série temporal total mensal filtrada total.



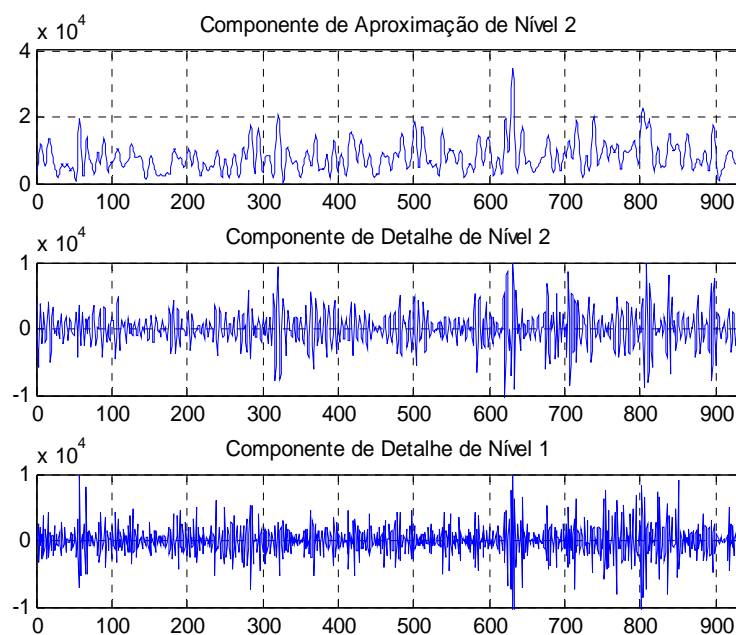


Figura 7.9 - Componentes *wavelet* de aproximação e de detalhe da decomposição de nível 2, com base ortonormal *db 40*, da série temporal total mensal filtrada total.

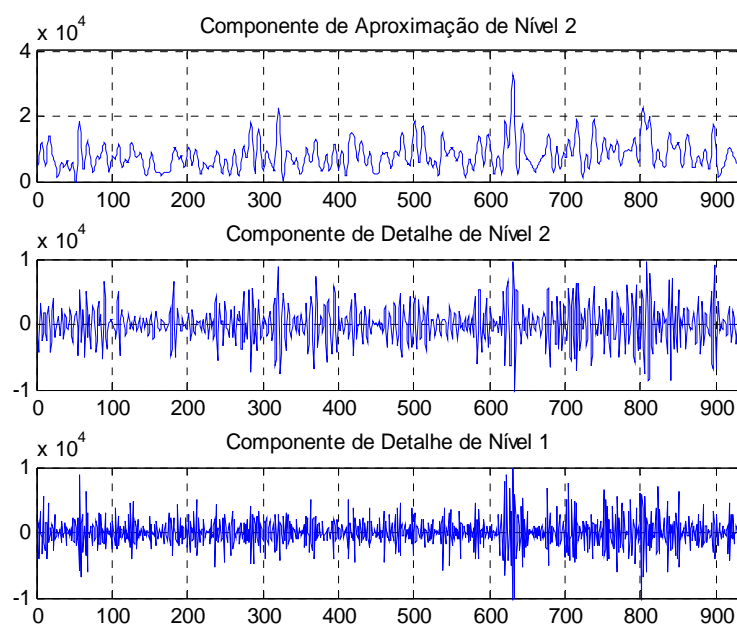


Figura 7.10 - Componentes *wavelet* de aproximação e de detalhe da decomposição de nível 2, com base ortonormal *db 30*, da série temporal total mensal filtrada total.

É importante destacar que, para as componentes ortogonais das 3 decomposições *wavelet* da série temporal total, nas Figuras 7.8, 7.9 e 7.10,



realizaram-se, inicialmente, as decomposições de 80% de suas primeiras observações da série temporal mensal filtrada (via método SSA) de vazão de afluentes da região sul do Brasil (que consiste na amostra de treino inicial). Em seguida, para cada nova observação (no processo de geração das previsões um passo à frente), a amostra de treino foi atualizada e, por consequência, calculou-se uma nova série de *Fourier* (ou seja, os coeficientes de aproximação e de detalhes foram atualizados para cada nova observação agregada na amostra de treino). Portanto, foram calculados 188 diferentes séries de *Fourier*, com bases ortonormal *wavelet* (ou seja, uma série de *Fourier* para cada nova observação agregada à amostra de treino inicial).

### 7.3

#### Modelagem das Componentes *Wavelet*

Como visto na Seção 7.2, as 3 diferentes decomposições ortogonais *wavelet*, de nível 2, da série temporal mensal de vazão de afluentes da região sul do Brasil filtrada via método SSA geram 3 componentes de aproximação e 6 componentes de detalhe. Com efeito, foram necessárias 9 RNA' s *feedforward*, com uma camada escondida, distintas para a realização da modelagem de cada componente ortogonal *wavelet*, cujas configurações se encontram na Tabela 7.3, a seguir. No processo de treinamento das 9 RNA' s, utilizou-se a amostra de treino inicial (ou seja, 80 % das primeiras observações da série temporal em questão). Salienta-se, ainda, que as 9 RNA' s foram implementadas no *software* MATLAB.



Tabela 7.3 - Configurações das 9 RNA 's *feedforward*, com uma camada escondida, utilizadas na modelagem das 9 componentes *wavelet*.

Método Preditivo	Componente <i>Wavelet</i>	Tamanho da Janela	Nº de Neurônios na CE	Método de Normalização
RNA 1	aproximação de nível 2 (db 10)	20	10	<i>Premnmx</i>
RNA 2	detalhe de nível 2 (db 10)	18	10	<i>Premnmx</i>
RNA 3	detalhe de nível 1 (db 10)	20	15	<i>Premnmx</i>
				<i>Premnmx</i>
RNA 4	aproximação de nível 2 (db 40)	22	12	<i>Premnmx</i>
RNA 5	detalhe de nível 2 (db 40)	12	12	<i>Premnmx</i>
RNA 6	detalhe de Nível 1 (db 40)	22	14	<i>Premnmx</i>
				<i>Premnmx</i>
RNA 7	aproximação de nível 2 (db 30)	18	15	<i>Premnmx</i>
RNA 8	detalhe de nível 2 (db 30)	18	15	<i>Premnmx</i>
RNA 9	detalhe de nível 1 (db 30)	22	15	<i>Premnmx</i>

Destaca-se ainda que foram utilizadas, em todas as RNA' s da Tabela 7.3, as funções de ativação *tansing*, na camada escondida (CE), e *purelin*, na camada de saída [veja HAKYN (2000)]. O algoritmo de treinamento utilizado foi o *trainlm* [veja <http://www.mathworks.com/help/nnet/ref/trainlm.html>].

## 7.4

### Combinação Linear das Previsões das Componentes *Wavelet*

Após a modelagem das 9 componentes *wavelet* [Seção 7.3], procedeu-se à combinação linear de suas previsões, conforme a abordagem descrita na Seção 6.3. O ajuste numérico dos parâmetros (isto é, dos pesos e constantes adaptativas) da combinação proposta ocorreu por meio do PPM multiobjetivo MINIMAX, com abordagem de Programação por Metas, descrito na Seção 6.4, no qual foram utilizadas as estatísticas MAPE (*mean absolute percentual error*), MSE (*mean square error*), e MAE (*mean absolute error*) da amostra de treino inicial e RMSE (*root mean square error*), como funções-objetivo, e as suas respectivas *metas*.



Os valores ótimos para os parâmetros da combinação linear proposta das previsões das componentes *wavelet* podem ser verificadas na Tabela 7.4. O PPM multiobjetivo MINIMAX proposto foi implementado no *software* AIMMS (<http://www.aimms.com/>), com uso do algoritmo de otimização CONOPT.

Tabela 7.4 - Parâmetros ótimos da combinação linear proposta.

$\beta_{V_{m_0}(\phi),1}$	$\beta_{V_{m_0}(\phi),2}$	$\beta_{V_{m_0}(\phi),3}$	$\alpha_{V_{m_0}(\phi),1}$	$\alpha_{V_{m_0}(\phi),2}$	$\alpha_{V_{m_0}(\phi),3}$	$\beta_{V_{m_0}(\phi)}$
0,078141173	1,718473842	0,285602915	4,618383436	4,617961914	4,600956777	0,518942771
$\beta_{W_{m_0}(\omega),1}$	$\beta_{W_{m_0}(\omega),2}$	$\beta_{W_{m_0}(\omega),3}$	$\alpha_{W_{m_0}(\omega),1}$	$\alpha_{W_{m_0}(\omega),2}$	$\alpha_{W_{m_0}(\omega),3}$	$\beta_{W_{m_0}(\omega)}$
0,076603617	1,951064489	0,317620234	5,871862293	5,872187962	5,8893014	0,457214133
$\beta_{W_{m_0+1}(\omega),1}$	$\beta_{W_{m_0+1}(\omega),2}$	$\beta_{W_{m_0+1}(\omega),3}$	$\alpha_{W_{m_0+1}(\omega),1}$	$\alpha_{W_{m_0+1}(\omega),2}$	$\alpha_{W_{m_0+1}(\omega),3}$	$\beta_{W_{m_0+1}(\omega)}$
0,098708549	2,006064467	0,311480366	2,726352738	2,726759666	2,743866794	0,450188292

Para efeitos de comparação, a série temporal em questão foi modelada ainda por 8 outros previsores: (a) 3 RNA' s *feedforward*, com uma camada escondida (abordagem convencional); (b) 3 RNA' s *feedforward*, com uma camada escondida, integrada com decomposição *wavelet*, de nível 2, conforme a abordagem descrita em TEXEIRA Jr *et al* (2012a), denotadas por RNA-WAVE 1, RNA-WAVE 2 e RNA-WAVE 3; e (c) pela *combinação linear* (denotada, na Tabela 7.5, por “*Comb Linear*”) - conforme a abordagem de GRANGER & NEWBOLD (1974) - e a *combinação geométrica* (denotada, na Tabela 7.5, por “*Comb Geométrica*”) - conforme a abordagem de TEIXEIRA Jr (2009) - das previsões oriundas dos métodos preditivos RNA-WAVE 1, RNA-WAVE 2 e RNA-WAVE 3. O ajuste dos parâmetros adaptativos ótimos das referidas combinações linear e geométrica de previsões foram obtidos por meio de PPM simples, o qual foi implementado no *software* AIMMS. Destaca-se ainda que o objetivo da combinação linear de previsões acima é a minimização da estatística MSE; enquanto que o objetivo da combinação geométrica é a minimização da estatística MAPE. Na Tabela 7.5, têm-se as melhores configurações da RNA 1, RNA 2 e RNA 3, RNA-WAVE 1, RNA-WAVE 2 e RNA-WAVE 3, com base na estatística MSE de treino inicial, após exaustivos testes. Destaca-se que, para as 6 RNA' s da Tabela 7.5, foram utilizadas as funções de ativação *tansing*, na camada escondida; *purelin*, na camada de saída; e o algoritmo de treinamento *trainlm*.



Para verificar as peculiaridades concernentes às bases ortonormais *wavelet* denotadas por *db40*, *sym20* e *db30*, veja <http://www.mathworks.com/help/wavelet/ref/waveinfo.html>.

Tabela 7.5 - Configurações das RNA's: RNA 1, RNA 2, RNA 3, RNA-WAVE 1, RNA-WAVE 2 e RNA-WAVE 3.

Método Preditivo	Base Wavelet	Tamanho da Janela	Nº de Neurônios na CE	Método de Normalização
RNA 1	-	15	20	MINIMAX
RNA 2	-	10	18	Sigmoidal
RNA 3	-	12	25	z-score
RNA-WAVE 1	db 40	5	3	Premmx
RNA-WAVE 2	sym 20	6	5	Premmx
RNA-WAVE 3	db 30	7	4	Premmx

Os resultados dos desempenhos, em relação às estatísticas MSE, nas amostras de treino (inicial), validação e teste, dos 9 métodos preditivos implementados para a modelagem da série temporal de interesse, podem ser verificados na Tabela 7.6.

Tabela 7.6 - Valores de MSE dos 9 métodos preditivos implementados.

Métodos Preditivos	MSE (treino)	MSE (validação)	MSE (teste)
RNA 1	2.068.357,85	4.149.883,80	8.700.588,73
RNA 2	1.715.565,42	5.521.917,49	4.073.701,22
RNA 3	2.082.508,02	4.099.053,67	9.091.625,01
RNA-WAVE 1	512.275,72	3.000.844,48	5.575.192,63
RNA-WAVE 2	598.507,81	2.975.426,91	4.689.580,95
RNA-WAVE 3	288.104,30	3.280.344,12	6.309.941,53
Comb Linear	<b>228.989,50</b>	<b>2.742.448,26</b>	<b>4.240.544,72</b>
Comb Geométrica	<b>247.537,00</b>	<b>2.646.178,57</b>	<b>4.187.472,41</b>
Método Proposto	<b>5.143,29</b>	<b>34.584,76</b>	<b>9.661,77</b>

Na Tabela 7.6, note que os métodos de combinação linear e geométrica de previsões, bem como a combinação proposta, obtiveram resultados superiores às outras 6 RNA's, nas 3 amostras, quanto à estatística MSE. Em particular, pode-se verificar que os resultados da combinação de previsões proposta foram muito superiores às combinações linear e geométrica de previsões.



Para comparar os 9 melhores métodos preditivos, da Tabela 7.6, quanto à estatística MAPE, verifique as Figuras 7.11 e 7.12, a seguir. As estatísticas *U de Theil* [veja MUBWANDARIKWA (2007)] das amostras de treino, de validação e de teste da metodologia proposta são, respectivamente, iguais a: 0,013953956, 0,0249467 e 0,019112839. Como  $U < 1$  e muito próximo a 0, significa que o método proposto foi muito superior ao *método ingênuo* [veja SOUZA & CAMARGO (2004)].

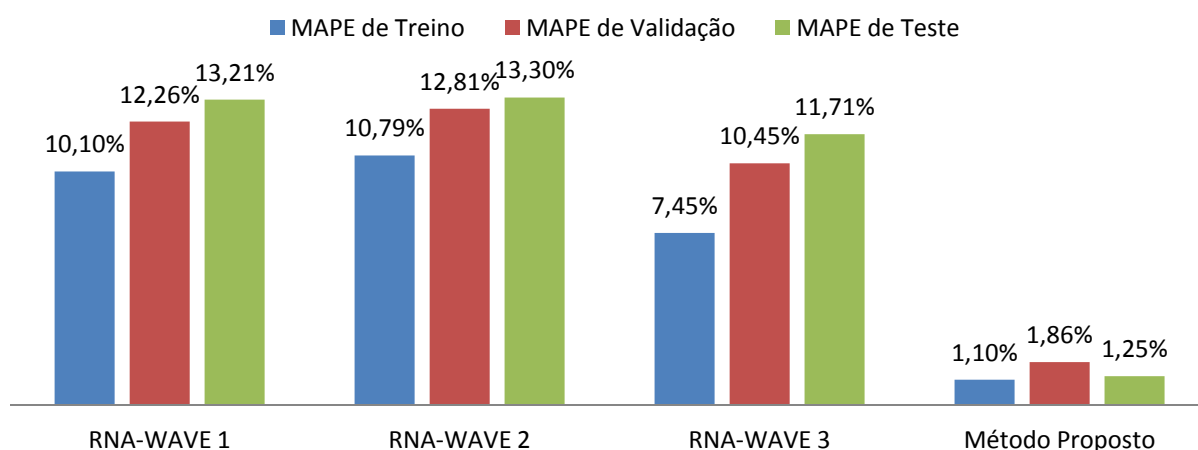


Figura 7.11 - Valores de MAPE dos métodos RNA-WAVE 1, RNA-WAVE 2 e RNA-WAVE 3 e método proposto.

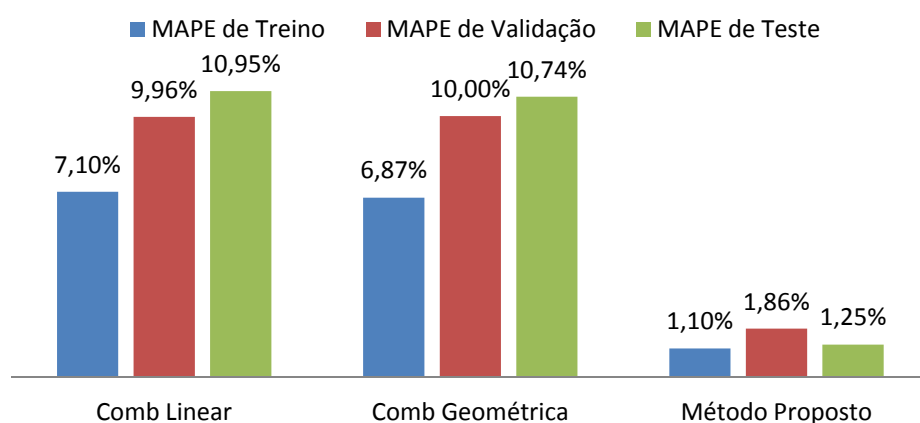


Figura 7.12 - Valores de MAPE dos métodos RNA-WAVE 1, RNA-WAVE 2 e RNA-WAVE 3 e método proposto.



Na Figura 7.11, são comparadas as RNA's RNA-WAVE 1, RNA-WAVE 2 e RNA-WAVE 3, com o Método Proposto (combinação de previsões proposta); na Figura 7.12, são comparadas as combinações lineares e geométrica de previsões, com o Método Proposto. Verifica-se, novamente, que o Método Proposto foi muito superior, nas 3 amostras, em relação a todos os outros previsores. Nas Figuras 7.13 e 7.14, tem-se uma descrição gráfica do nível de correlação entre as observações e as previsões, um passo à frente, oriundas da combinação proposta (Método Proposto), nas amostras de validação e de teste. Note que as previsões possuem alta correlação em relação às observações, corroborando o seu bom desempenho obtido no experimento.

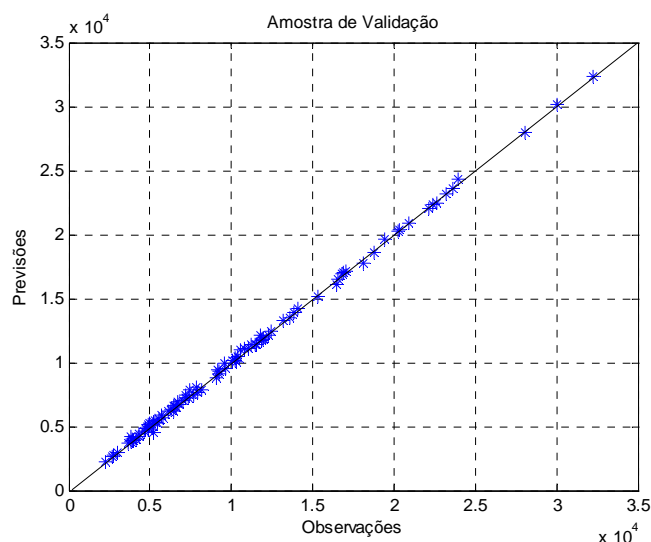


Figura 7.13 - Dispersão entre as observações e as previsões do método proposto, na amostra de validações (correlação igual a 99,96%).



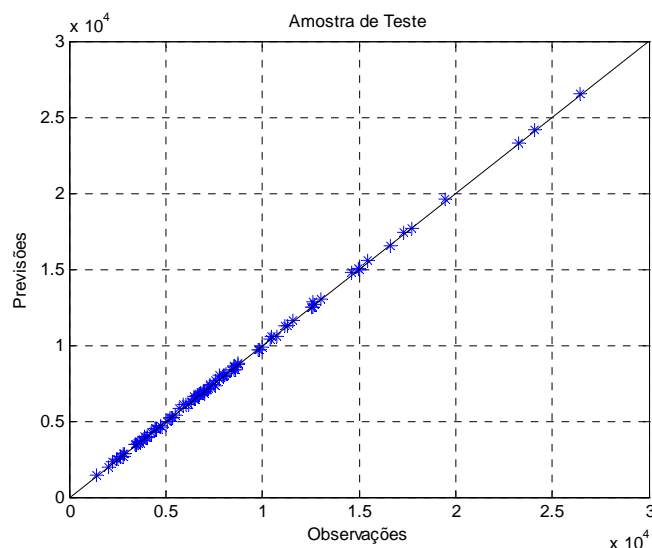


Figura 7.14 - Dispersão entre as observações e as previsões do método proposto, na amostra de teste (correlação igual a 99,98%).

Na Tabela 7.7, têm-se as estatísticas do *teste BDS* [veja HAMILTON (1994)] - cuja hipótese nula ( $H_0$ ) é “independência estatística da série temporal dos resíduos associados ao Método Proposto da amostra de treino” e a hipótese alternativa ( $H_1$ ) é “ $H_0$  é falsa” - considerando 5% de significância. Note que  $H_0$  não é rejeitada, a 5% de significância, nas 6 dimensões – o que evidencia que, do ponto de vista Estatístico, o Método Proposto foi adequadamente utilizado na modelagem de toda estrutura de autodependência presente na amostra de treino da série temporal supracitada.

Tabela 7.7 - Teste BDS para a série temporal de resíduos associados às previsões geradas pelo método proposto, na amostra de treino.

Dimensões	$p$ -valor do Teste BDS
1	91,96 %
2	89,21 %
3	87,08 %
4	85,23 %
5	83,59 %
6	91,96 %



## 7.5

### Geração dos Cenários QMC Híbrido e dos Intervalos de Confianças

Nesta última etapa, foram gerados cenários para a série temporal mensal de vazão de afluentes da região sul do Brasil, por meio do método de simulação QMC híbrido, conforme descrito na Seção 6.5. Em seguida, foram calculados os intervalos de confiança (um passo à frente), com 99 % de credibilidade, nas amostras de validação e de teste. Para efeito de comparação, foram também simulados cenários via método QMC híbrido, com o mesmo procedimento descrito na Seção 6.5 para os outros 5 melhores métodos preditivos, ou seja: RNA-WAVE 1, RNA-WAVE 2, RNA-WAVE 3, *Comb Linear*, *Comb Geométrica*.

Os métodos preditivos RNA 1, RNA 2 e RNA 3 foram desconsiderados, visto que obtiveram desempenho muito inferior aos demais, quanto ao MSE e ao MAPE. Para a avaliação de desempenho, quanto aos cenários gerados, foram consideradas as estatísticas de amplitudes médias dos intervalos de confiança, nas amostras de validação e de teste, de cada método preditivo. Na Figura 7.15, tem-se o histograma da série temporal de resíduos associados às previsões do método proposto, na amostra de treino, para a série temporal em questão. Pode-se verificar que a distribuição das frequências relativas dos resíduos apresenta simetria em torno do intervalo de classe que contém o valor zero.

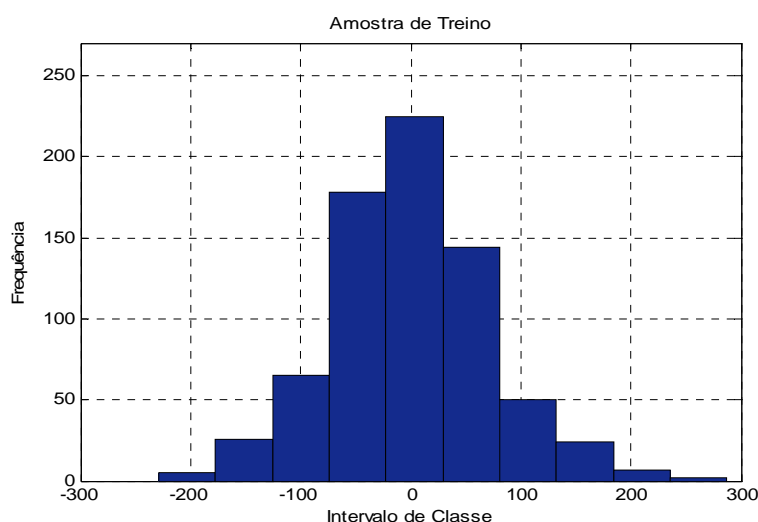


Figura 7.15 - Histograma da série temporal de resíduos associados às previsões geradas pelo método proposto, na amostra de treino.



Com efeito, dada a simetria do histograma em torno de zero, na Figura 7.15, assumiu-se que a série temporal de resíduos possuem distribuição de densidade de probabilidade normal, com média populacional igual a 0. Para as amostras de validação e de teste, foi assumido que esta condição probabilística era sempre preservada. Por equiparação, no processo de geração dos cenários dos outros 5 métodos preditivos utilizados, foi pressuposto também a condição de normalidade, com média populacional igual a 0, de suas respectivas séries temporais de resíduos, nas amostras de treino, de validação e de teste.

Na Figura 7.16, tem-se o gráfico dos limites inferior e superior dos intervalos de confiança (linha vermelha), com 99% de credibilidade, e das observações da série temporal mensal de vazão de afluentes da região sul do Brasil (linha azul), na amostra de validação.

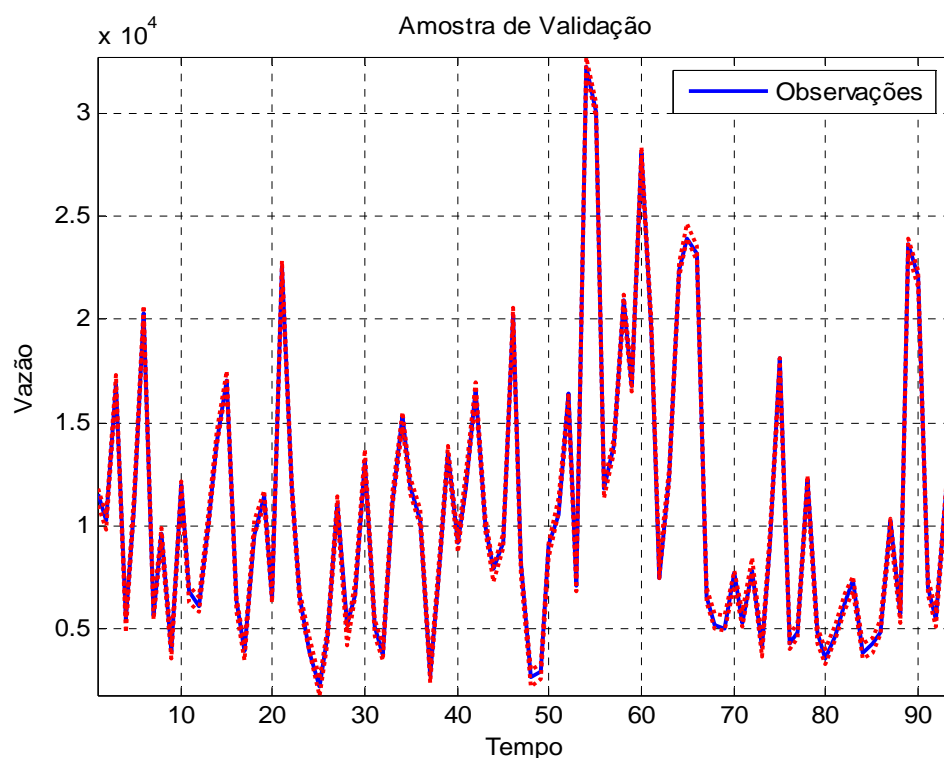


Figura 7.16 - Gráfico dos limites inferiores e superiores dos intervalos de confiança (linha vermelha) e das observações (linha azul), na amostra de validação.

Na Tabela 7.8, têm-se as amplitudes médias dos 6 métodos preditivos implementados, neste experimento, nas amostras de validação e de teste. Note que os métodos preditivos combinados possuem desempenho superior aos previsores



individuais, quanto às amplitudes médias dos intervalos de confiança, com 99% de credibilidade. Em particular, pode-se verificar que o método proposto foi muito superior aos métodos preditivos combinados.

Tabela 7.8 - Amplitudes médias dos intervalos de confiança dos 6 métodos preditivos, na amostra de validação e de teste.

Método	Amplitude Média	
	Amostra de Validação	Amostra de Teste
RNA-WAVE 1	2.967,75	3.977,94
RNA-WAVE 2	3.177,14	4.391,55
RNA-WAVE 3	3.044,77	2.965,94
Comb Linear	<b>2.878,91</b>	<b>2.886,18</b>
Comb Geométrica	<b>2.593,98</b>	<b>2.600,55</b>
Método Proposto	<b>742,00</b>	<b>815,77</b>

Portanto, a incerteza associada à análise de risco feita com o uso do método proposto. Por meio de análise visual, pode-se verificar que as observações dos 94 instantes não violam os limites inferior e superior dos intervalos de confiança. Tal fato foi verificado, de forma exaustiva, e confirmado.

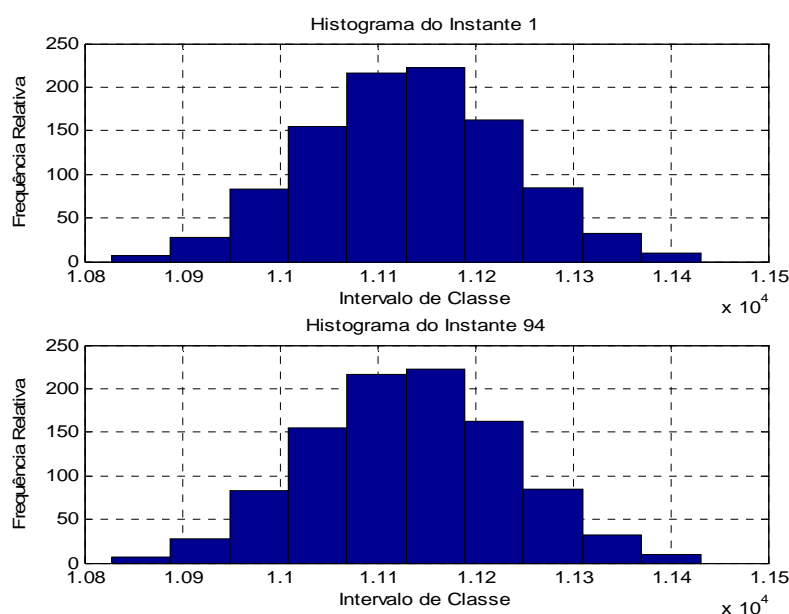


Figura 7.17 - Histogramas dos cenários QMC híbridos do método proposto, nos instantes 1 e 94, na amostra de validação.



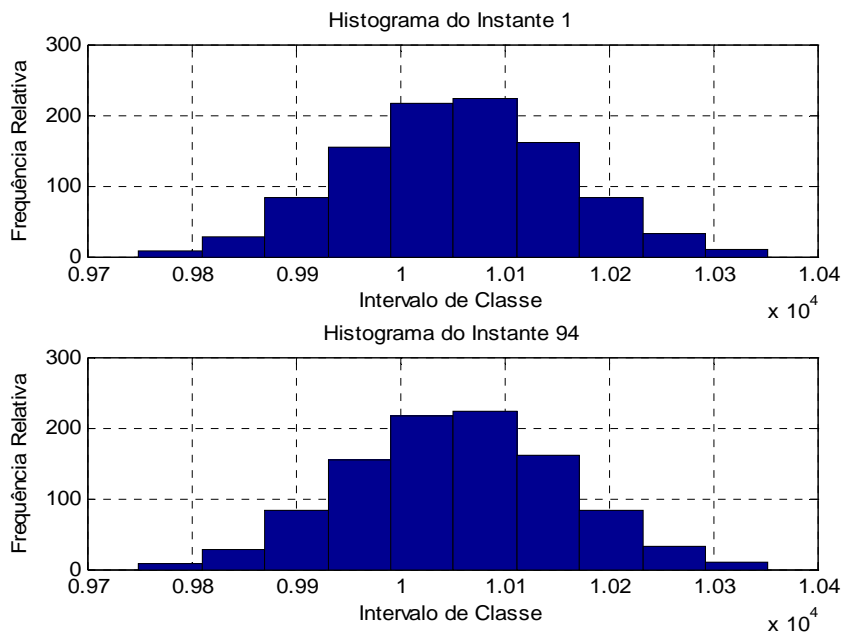


Figura 7.18 - Histogramas dos cenários QMC híbridos do método proposto, nos instantes 1 e 94, na amostra de teste.

Nas figuras 7.17 e 7.18, mostra-se, nos instantes 1 e 94, nas amostras de validação e de teste, que o método proposto preservou a condição de normalidade assumida para os cenários QMC híbridos normais gerados para as 9 componentes ortogonais *wavelet* da série temporal deste experimento. Em outras palavras, os cenários QMC híbridos combinados também são normalmente distribuídos nos referidos instantes. Convém destacar que este fato foi verificado, exaustivamente, e confirmado para os todos os 184 instantes.



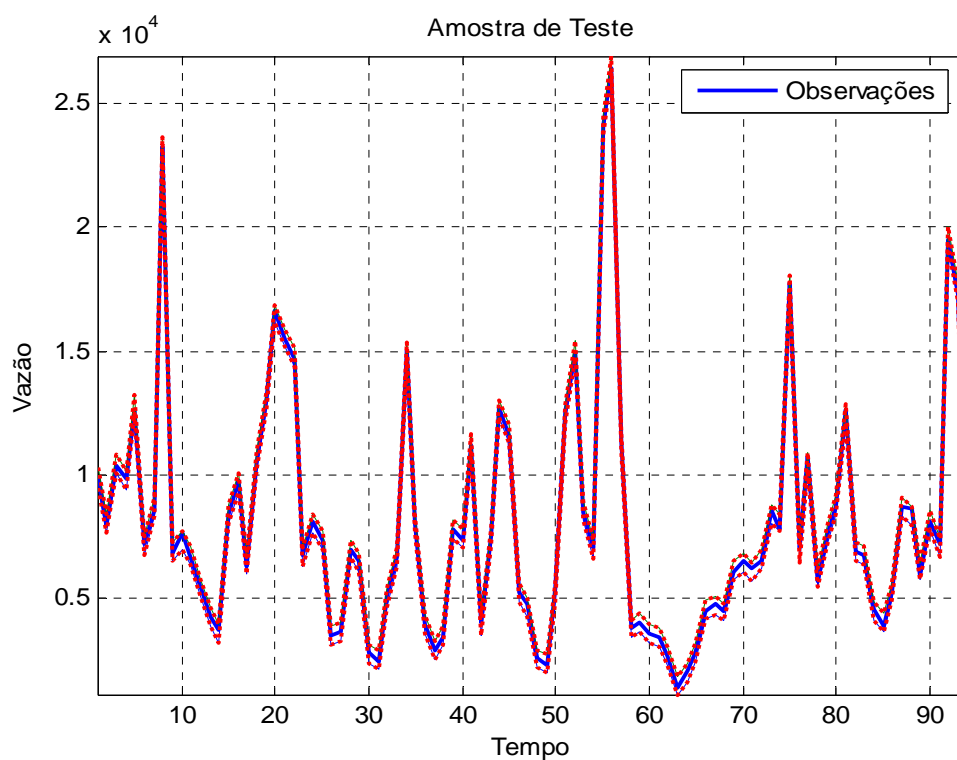


Figura 7.19 - Gráfico dos limites inferiores e superiores dos intervalos de confiança (linha vermelha) e das observações (linha azul), na amostra de teste.

Pode-se verificar, visualmente, na Figura 7.19, que as observações dos 94 instantes, na amostra de teste, estão entre os limites inferior e superior dos intervalos de confiança, com 99% de credibilidade. Tal fato foi verificado, exaustivamente, e confirmado para todos os 94 instantes, na amostra de teste.



## CONCLUSÕES

Nesta tese de doutorado, foi proposta uma metodologia híbrida para a combinação linear de métodos preditivos de séries temporais, na geração de previsões e de cenários, que utiliza, de forma integrada, alguns dos principais métodos e abordagens presentes na Literatura: *método SSA*; *Teoria Wavelet*; *redes neurais artificiais*; *programação matemática multiobjetivo MINIMAX*, *com abordagem de programação por metas*; e *método de simulação de quase-Monte-Carlo híbrido*. Para mostrar a eficiência da metodologia proposta, bem como detalhá-la em um caso aplicado, foi realizada a modelagem da série temporal mensal de vazão mensal de afluentes da região sul do Brasil, para se gerar previsões e cenários um passo à frente, dentro e fora da amostra. Como *benchmark*, a referida série temporal foi modelada, para a mesma finalidade, por meio de outros 8 métodos preditivos consagrados na Literatura, sendo 6 individuais (3 RNA's simples e 3 RNA's integradas com decomposição *wavelet*) e 2 combinados (*combinação linear*, conforme a abordagem em GRANGER & NEWBOLD (1974), e *combinação geométrica*, conforme a abordagem em TEIXEIRA Jr (2009)). Para a geração de cenários QMC híbridos associados a estes métodos preditivos, utilizou-se o procedimento descrito na Seção 6.5.

Em relação à metodologia proposta, é importante destacar alguns pontos. O método SSA, com abordagem de análise de gráfica de autovetores singulares, foi utilizado para a filtragem (ou encolhimento de ruído) da amostra de treino inicial da série temporal mensal de vazão de afluentes da região sul do Brasil, gerando uma versão menos ruidosa desta. O parâmetro  $L$  ótimo obtido (no caso,  $L=468$ ) foi o que gerou o maior número de componentes SSA classificadas como ruído, na amostra de treino inicial. Considerou-se, neste experimento, como amostra de treino inicial 80% das primeiras observações. Em seguida, foram realizadas três diferentes decomposições *wavelet* de nível 2 da amostra de treino inicial filtrada (via método SSA), gerando-se 9 componentes ortogonais *wavelet*, sendo 3 de aproximação e 6 de detalhe. É importante salientar que, à medida que um novo dado era agregado à amostra de treino, atualizando-a, foi realizado,



novamente, o procedimento de filtragem SSA - utilizando, porém, o parâmetro  $L$  ótimo (no caso,  $L=468$ ), obtido a partir da amostra de treino inicial, e a classificação dada, previamente, às componentes SSA, com base na referida. A cada nova observação agregada, foi calculada uma nova decomposição *wavelet*. Como foi considerado um horizonte de previsão igual a 188 passos à frente, foram calculadas 188 diferentes séries de *Fourier wavelet*, de modo a se obter 188 diferentes decomposições *wavelet* (de nível 2). Para a escolha das melhores bases ortonormais *wavelet*, considerou-se aquelas que, dado o nível de decomposição igual a 2 (que foi o que apresentou os melhores resultados em termos de modelagem, durante a pesquisa), forneciam o menor EQM entre os sinais da amostra de treino inicial e os respectivos sinais *wavelet* gerados em decorrência da soma das componentes de aproximação e de detalhe. Com base nisso, foram escolhidas as 3 bases ortonormais *wavelet* da família de *Daubechies*, a seguir: *db 10*, *db 40* e *db 30*.

É importante destacar também que a integração do método SSA e da decomposição *wavelet* possibilitaram a geração de componentes ortogonais *wavelet* com um padrão de comportamento temporal menos ruidoso (ou menos estocástico) e mais regular. Em particular, no tocante à regularidade de comportamento, tal fato se deve, porque cada componente ortogonal *wavelet* está associada ao parâmetro de escala  $m$  [veja Seção 2.2.4] das funções *wavelet* que a compõe e este, à frequência da respectiva componente *wavelet* [veja MORETTIN (1997)]. Por sua vez, a utilização de 3 diferentes bases ortonormais *wavelet* (no caso, *db 10*, *db 40* e *db 30*) e de 9 RNA's *feedforward*, com uma camada escondida (com diferentes configurações) possibilitou a modelagem por meio de 9 abordagens distintas, onde cada uma forneceu informações (representadas por previsões) diferentes acerca da flutuação estocástica-temporal da série temporal filtrada. Dessa maneira, ao se combinar linearmente a previsão para as componentes *wavelet*, obteve-se, para cada instante  $t$ , uma previsão combinada que, segundo FARIA & MUBWANDARIKWA (2008), pode ser interpretada como uma agregadora de informações oriundas de diferentes fontes (abordagens preditivas). Destaca-se ainda que foi utilizada a decomposição ortogonal *wavelet* (e não a decomposição via método SSA), porque os resultados dos experimentos obtidos, ao longo da pesquisa, mostraram que as componentes ortogonais *wavelet*



possuem melhores padrões de comportamento, para fins de modelagem, que as componentes SSA [veja SOUZA *et al* (2012a), SOUZA *et al* (2012b), TEIXEIRA Jr *et al* (2012b), TEIXEIRA Jr *et al* (2012a)]. Ademais, é possível classificar, visualmente, quais componentes SSA são ruído estatístico (o que torna o processo de filtragem, operacionalmente, mais rápido) [veja CASSIANO *et al* (2012), HASSANI *et al* (2009), GOLYANDINA *et al* (2001)], o que não é possível com as componentes ortogonais *wavelet*. Por último, no que tange à utilização de RNA's, isso ocorreu pelo fato de serem aplicáveis à modelagem de quaisquer séries temporais que apresentem estrutura de autodependência (linear ou não linear), além de serem eficientes e consagradas na Literatura para tal finalidade [veja Capítulo 3].

No tocante à escolha da combinação linear de previsões, tem-se que esta é mais abrangente, no sentido de se poder efetuar a combinação de valores previstos positivos ou negativos - o que não seria possível, por exemplo, com a *combinação geométrica de previsões* [veja TEIXEIRA Jr (2009)], que comporta somente a combinação de valores previstos positivos, muito embora se possam obter melhores resultados, em muitos casos, com a sua utilização. Outro ponto importante é que, na combinação geométrica de previsões, se a previsão de um método preditivo base, em  $t$ , é igual a zero, então a respectiva previsão combinada também assume valor igual a zero, o que não ocorre na combinação linear de previsões. Além disso, ao se combinar cenários normalmente distribuídos, a combinação linear preserva a sua forma distribucional, gerando-se, assim, cenários linearmente combinados também distribuídos normalmente. No processo de simulação dos cenários (via método de simulação QMC híbrido), assumiu-se distribuição normal. E, por ser considerado o melhor inversor normal-padrão (quando utilizados os números quase-aleatórios), foi utilizado o inversor de MORO para a metodologia proposta e para todos os métodos preditivos como *benchmark*.

No que tange à otimização dos parâmetros adaptativos (constantes e pesos adaptativos) da combinação linear proposta. Optou-se pela otimização via PPM multiobjetivo MINIMAX, com abordagem de programação por metas, por dois aspectos básicos: o primeiro é que não se tinha a intenção de dar preferência a nenhuma das estatísticas de aderência utilizadas na modelagem, de modo que foi



realizado um ajuste numérico global, no qual todas foram, simultaneamente, otimizadas; o segundo é que, de acordo com TEIXEIRA JR *et al* (2011a), é possível se obter ganhos preditivos, fora da amostra (no caso, nas amostras de validação e de teste), com tal abordagem. Ademais, os experimentos realizados durante a pesquisa mostraram que, quando não se obtém ganho preditivo fora da amostra, os resultados são, geralmente, equivalentes. Os parâmetros adaptativos ótimos foram obtidos a partir da amostra de treino inicial e não foram atualizados com a agregação de novos dados.

Em relação aos cenários QMC híbridos combinados (utilizando a combinação linear ótima), foi possível verificar que a sua forma distribucional (no caso, a distribuição normal) foi preservada e que a amplitude média de seus intervalos de confiança, com 99% de credibilidade, foi bem menor, evidenciando, assim, menor incerteza sobre o futuro da série temporal em análise. Verificou-se ainda, gráfico e exaustivamente, que nenhuma das observações, nas amostras de validação e de teste, violou os limites dos intervalos de confiança, mostrando que a análise de risco, sob esta óptica, foi eficiente.

Em síntese, no processo de geração de previsões, a combinação linear proposta, com ajuste numérico via PPM multiobjetivo MINIMAX, com abordagem de programação por metas, das previsões para as 9 componentes *wavelet* da versão filtrada (via SSA) da série temporal mensal de vazão de afluentes da região sul oriundas de 9 RNA's distintas gerou resultados consideravelmente melhores que os obtidos pelos outros 8 métodos preditivos, nas amostra de treino, validação e de teste, em todas as estatísticas de aderência consideradas. A utilização da combinação proposta, com os parâmetros ótimos, integrada com a simulação QMC híbrida forneceu cenários que possibilitaram a construção de intervalos de confiança, com 99% de credibilidade, eficientes, nas amostras de validação e de teste.

Pode-se ainda verificar, de forma individual, a contribuição de cada abordagem utilizada na metodologia proposta, nos artigos gerados durante a pesquisa. Em SOUZA *et al* (2012a), SOUZA *et al* (2012b), MENESES *et al* (2012), PESSANHA *et al* (2012) e CASSIANO *et al* (2012), podem ser verificadas aplicações do método SSA (com a abordagem do método proposto), na filtragem estatística de outras séries de tempo de difícil modelagem e o seu



impacto em termos de ganhos preditivos. Em TEIXEIRA Jr *et al* (2011 b), TEIXEIRA Jr *et al* (2012 a) e PESSANHA *et al* (2012), podem ser verificados os ganhos preditivos incorridos na integração da decomposição *wavelet* com redes neurais artificiais, na modelagem de outras séries temporais, como, por exemplo, a de velocidade de vento e a de radiação solar direta. Em TEIXEIRA Jr *et al* (2011 a) e TEIXEIRA Jr *et al* (2012 b), podem ser verificados os ganhos preditivos incorridos com a utilização da programação matemática multiobjetivo, com abordagem de programação por metas, na combinação linear de previsões. Em especial, em TEIXEIRA Jr *et al* (2012 b), pode-se verificar o ganho preditivo obtido quando utilizada a combinação linear de previsões de componentes ortogonais *wavelet*, com ajuste numérico via programação matemática, na geração de previsões *multi-step*. Em TEIXEIRA Jr (2009), verificam-se os ganhos preditivos acarretados com o uso do método de simulação QMC híbrido na geração de intervalos de confiança, em projeções *multi-step*. Em TEIXEIRA Jr *et al* (2013), pode-se verificar que a combinação de previsões integrada com o método de filtragem estatística SSA (com a abordagem da tese) e a decomposição *wavelet* obtiveram melhores resultados na projeção *multi-step*. Assim sendo, tem-se que cada módulo que integra a metodologia proposta foi testado, individualmente, em distintas séries temporais, horizontes de previsão e passos à frente a serem previstos.

Portanto, o método proposto se mostrou ser uma alternativa factível e eficiente à modelagem de séries temporais, dadas a sua abrangência e a sua eficiência mostrada nas pesquisas.



## TRABALHOS FUTUROS

Diante dos resultados obtidos nesta tese de doutorado, podem ser estudadas outras metodologias dela decorrentes ou similares, com a finalidade de métodos preditivos cada vez mais eficientes e mais abrangentes, na modelagem de séries temporais que apresentam estruturas temporais complexas de serem modeladas. Dentre tantas possibilidades, eis algumas:

1. Podem-se combinar, de forma não linear, as previsões para as componentes ortogonais *wavelet*, por meio de uma RNA *feedforward*, com uma camada escondida. É possível verificar que, se todos os neurônios da camada escondida tiverem função de ativação linear, tem-se uma abordagem alternativa de combinação linear de previsões. Desse modo, podem-se testar, por exemplo, diferentes padrões de entrada, diferentes valores de janela neural e algoritmos de treinamento;
2. É possível fazer testes estatísticos não paramétricos, com o auxílio do software *@Risk* [veja <http://www.palisade.com/risk/>], a fim de se verificar as distribuições de probabilidades mais prováveis para os resíduos associados às previsões das componentes *wavelet*. Desse modo, são selecionadas as  $K$  mais importantes (baseado, por exemplo, no  $p$ -valor), implementa-se o método QMC e, então, escolhem-se os inversores que transformam os números quase-aleatórios (entre 0 e 1) em amostras destas  $K$  distribuições mais prováveis. A programação matemática (simples ou multiobjetivo) pode ser utilizada para estimação dos parâmetros adaptativos. Por fim, os choques estocásticos gerados para cada componente *wavelet* e estes são combinados (utilizando a combinação ótima), de forma a gerar uma densidade preditiva combinada (e, a partir desta, calcular as estatísticas de risco e os intervalos de confiança);
3. Podem-se combinar diferentes abordagens de combinações de previsões, onde os ponderadores podem ser obtidos por meio de programação matemática (simples ou multiobjetivo). Pode-se, por exemplo, combinar



linear e não linearmente as previsões para as componentes ortogonais *wavelet*, de forma a gerar novas previsões combinadas. A abordagem de geração dos choques estocásticos QMC híbridos, da proposta 2, podem ser utilizadas nesta metodologia de combinação de combinações de métodos preditivos base. Além disso, podem-se utilizar pesos adaptativos variantes no tempo. A utilização de *algoritmos genéticos* também é interessante na obtenção dos parâmetros ótimos.

4. Em séries temporais que apresentem *volatilidade estocástica* [veja MORETTIN (2005)], a previsão da média pode ocorrer utilizando a metodologia a combinação linear de previsões, do Capítulo 6. Como a volatilidade não é observável, existe incerteza quanto aos modelos estatísticos a serem utilizados em sua modelagem. Ante isso, podem-se combinar linearmente os mais prováveis e estudar as distribuições de probabilidades mais prováveis, com o auxílio do *software @Risk*, e gerar, com isso, os cenários combinados, conforme descrito em na proposta 2, acima.
5. Modelos Bayesianas talvez possam ser mais eficientes, com o uso das combinações propostas; e do método de pré-processamento de dados via método SSA e decomposição *wavelet*, utilizando a abordagem de tratamentos proposta nesta tese de doutorado.
6. As séries temporais de velocidade de vento [TEIXEIRA Jr *et al* (2011b)], vazão mensal de afluentes [veja CAMPOS (2010)] e de radiação solar direta [veja TEIXEIRA Jr (2012b)], que são excessivamente ruidosas ou não estacionárias, são passíveis de serem modeladas com as abordagens destacadas acima.



## REFERÊNCIAS BIBLIOGRÁFICAS

- [1]. ABELÉM, A (1994). **Redes Neurais Artificiais na Previsão de Séries Temporais**. Departamento de Engenharia Elétrica. Dissertação de Mestrado. Pontifícia Universidade Católica do Rio de Janeiro.
- [2]. ACWORTH, P.; BROADIE, M.; GLASSERMAN, P. (1997). **A Comparison of Some Monte Carlo and Quasi-Monte Carlo Techniques for Option Pricing**. In P. Hellekalek and H. Niederreiter, editors, Monte Carlo and Quasi- Monte Carlo Methods in Scientific Computing, number 127 in Lecture Notes in Statistics, pp. 1-18, New York. Springer-Verlag.
- [3]. ALBUQUERQUE, A. R. (2008). **Fluxo de Caixa em Risco: Uma Nova Abordagem para o Setor de Distribuição de Energia**. Dissertação de Mestrado. PUC-RIO.
- [4]. ALEXANDROV, Th. and GOLYANDINA, N. (2004). **Thresholds Setting for Automatic Extraction of Time Series Trend and Periodical Components with the Help of the Caterpillar SSA Approach**. *Proc. IV International Conference SICPRO'05*, 25-28.
- [5]. AQUINO, R. R. B.; LIRA, M. M. S.; OLIVEIRA, J. B.; CARVALHO J.; M. A.; NETO, O. N.; ALMEIDA, G. J. (2009). **Application of Wavelet and Neural Network Models for Wind Speed and Power Generation Forecasting in a Brazilian Experimental Wind Park**, *International Joint Conference on Neural Networks*. Atlanta, USA.
- [6]. ARMSTRONG, J. S. (2001). **Combining Forecasting**. In: **ARMSTRONG, J. S. Principles of Forecasting: A Handbook for Researchers and Practitioners**. Kluwer Academic Publishers.
- [7]. AUSSEM, A. J. and MURTAGH, F. (1997). **Combining Neural Network Forecasts on Wavelet-Transformed Time Series**. *Connection Science* 9 (1): 113-121.
- [8]. BARROS, M. (2004). **Processos Estocásticos**. Rio de Janeiro - RJ: Paper Virtual Editora.
- [9]. BATCHELOR, R. and DUA, P (1995). **Forecaster Diversity and the Benefits of Combining Forecasts**. *Management Science*, v.41, p.68-75. 151.
- [10]. BATES, J. M. and GRANGER, C. W. J. (1969). **The Combining of Forecasts**. *Operational Research Quarterly*, v.20, n.4, p. 451-468.
- [11]. BEASLEY, J. D. and SPRINGER, S. G. (1997). **The Percentage Points of the Normal Distribution**. *Applied Statistics* 23, 1997, pp.118-121.



- [12]. BJORN, V. (1995). **Multiresolution Methods for Financial Time Series Prediction**. Proceedings of the IEEE/IAFE. Conference on Computational Intelligence for Financial Engineering, 97.
- [13]. BROOMHEAD, D. S. AND KING, G. P. (1986a). **Extracting Qualitative Dynamics from Experimental Data**. Physica D 20, 217-236.
- [14]. CAMPOS, L. C. D. (2010). **Modelo Estocástico Periódico baseado em Redes Neurais**. Tese de Doutorado. Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).
- [15]. CASSELA, G. and BEGER, R. (2001). **Statistical Inference**. 2º Edição ed., Duxbury.
- [16]. CASSIANO, K. M.; **TEIXEIRA Jr, L. A.**; MENESES, M. L.; PESSANHA, J. F. M. SOUZA, R. C.; (2012). **Remoção de Ruídos de Séries Temporais através da Análise Espectral Singular**. XV Simpósio de Pesquisa Operacional & Logística da Marinha.
- [17]. CAO, S.; WENG, W.; CHEN, J.; LIU, W.; YU, G.; CAO, J. (2009). **Forecast of Solar Irradiance Using Chaos Optimization Neural Networks**. Power and Energy Engineering Conference, Wuhan, March.
- [18]. CHAABENE, M. and BEN AMMAR, M. (2008). **Neuro-Fuzzy Dynamic Model with Kalman Filter to Forecast Irradiance and Temperature for Solar Energy Systems**. Renewable Energy, 33, pp. 1435-1443.
- [19]. CHATFIELD, C. (2004). **The Analysis of Time Series: an Introduction**. Chapman & Hall/CRC, 6<sup>th</sup> ed.
- [20]. CHUI, C. K. (1992). **Introduction to Wavelets**. Academic Press. San Diego.
- [21]. CLEMEN, R. T. (1989). **Combining Forecasts: A Review and Annotated Bibliography**. International Journal of Forecasting. v. 5, p.559-583.
- [22]. CYBENKO, G. (1989). **Approximation by Superposition of a Sigmoidal Function**. Math Control Signals Systems 2: 303-314.
- [23]. DAUBECHIES I. (1988). **Orthonormal Bases of Compactly Supported Wavelet**. Comm. Pure and Applied Math., 41, pp. 909-996.
- [24]. DAOUDI, K.; FRAKT, A. B.; WILLSKY, A. S. (1999). **Multiscale Autoregressive Models and Wavelets**. Transactions on Information Theory, 45, 828-845.
- [25]. DENG, F.; SU, G.; LIU, C.; WANG, Z. (2010). **Global Solar Radiation Modeling Using The Artificial Neural Network Technique**. Power and Energy Engineering Conference, Chengdu, March.



- [26]. DIAS, Marco A. G. (2013). **Quasi-Monte Carlo Simulation**. Disponível em: <[https://www.puc-rio.br/marco.ind/quasi\\_mc.html](https://www.puc-rio.br/marco.ind/quasi_mc.html)>.
- [27]. DONOHO, D. L. and JONHSTONE, I. M. (1994). **Ideal Spatial Adaptation by Wavelet Shrinkage**. *Biometric*, 81,425-455.
- [28]. DONOHO, D. L.; JONHSTONE, I. M.; KERKYACHARIAN, G.; and PICARD, D. (1995). **Wavelet Shrinkage: Asymptopia? (with discussion)** *Journal Royal Statistical Society. B*, 57:301-369.
- [29]. ELSNER J. B. and TSONIS A. A. (1996). **Singular Spectrum Analysis: a New Tool in Time Series Analysis**. Plenum Press, New York.
- [30]. FARIA, A. E. and MUBWANDARIKWA, E (2008). **Multimodality on the Geometric Combination of Bayesian Forecasting Models**. *International Journal of Statistics and Management System*, 3, 1-25
- [31]. FERNANDES, C. (2006). **Notas de aula do curso de Séries Financeiras**. Pontifícia Universidade Católica do Rio de Janeiro.
- [32]. FLORES, B. E.; and WHITE, E. M. (1998). **A Framework for the Combination of Forecasts**. *Journal Academic Marketing Science*, v.16 (3-4), p.95-103.
- [33]. FLORES, B. E. and WHITE, E. M. (1989). **Subjective versus Objective Combining of Forecasts: An Experiment**. *Journal of Forecasting*, v.8, p.331-341.
- [34]. GAVIN, H. (2011). **The Levenberg-Marquardt Method for Nonlinear Least Squares Curve-Fitting Problems**. Department of Civil and Environmental Engineering Duke University. September 28, 2011.
- [35]. GISTATGROUP (2010). **Caterpillar SSA**. Petersburg University. Department of Mathematics. Russia. Disponível no Seguinte Link: <http://www.gistatgroup.com/cat/index.html>.
- [36]. GHIL, M. AND TARICCO, C. (1997). **Advanced Spectral Analysis Methods**. In *Past and present Variability of the Solar-Terrestrial system: Measurement, Data Analysis and Theoretical Model* (Edited by G. C. Castagnoli and A. Provenzale), 137-159. IOS Press.
- [37]. GRANGER, C. W. J. (1989). **Invited Review: Combining Forecasts - Twenty Years Later**. *Journal of Forecasting*, 8, pp. 167-173.
- [38]. GRANGER, C.W.J. & RAMANATHAN, R. (1984). **Improved Methods of Forecasting**. *Journal of Forecasting*, v.3, p.197-204.
- [39]. GOLDBARG, M. and LUNA, H. P. L. (2000). **Otimização Combinatória e Programação Linear**. Campus, Brasil.



- [40]. GOLYANDINA N.; NETRUTKIN V.; ZHIGLJAVSKY A. (2001). **Analysis of Time Series Structure: SSA and Related Techniques**. Chapman & Hall / CRC. London.
- [41]. GOLYANDINA, N. and OSIPOV, E. (2006). **Caterpillar SSA Method for Analysis of Time Series with Missing Values**. Chapman & Hall / CRC. London.
- [42]. GUPTA, S. and WILTON, P. C. (1987). **Combination of Forecasts: An Extension**. Management Science. v.33, n.3, p.356-372.
- [43]. GUTIERREZ, C. E. C. (2002). **Eliminação de Ruído por Encolhimento Wavelets. Uma aplicação à série de Preço Spot de Energia Elétrica do Brasil**. Dissertação de Mestrado. Pontifícia Universidade Católica do Rio de Janeiro.
- [44]. HAIMES, Y. Y., and HALL, W. A. (1974). **Multiobjective in Water Resources Systems Analysis: The Surrogate Worth Trade-Off Method**. Wat. Resources' Res., 10, pp 615 - 624.
- [45]. HAMILTON, J. (1994). **Time Series Analysis**. Princeton University Press
- [46]. HAYKIN, S. (2001). **Redes Neurais: Princípios e Prática**. 2º ed. Porto Alegre, Bookman.
- [47]. HASSANI H. (2007). **Singular Spectrum Analysis: Methodology and Comparison**. Journal of Data Science.5, 239-257, 2007.
- [48]. HASSANI, H.; ZHIGLJAVSKY, A.; PATTERSON, K.; SOOFI, A. (2010). **A Comprehensive Causality Test Based on the Singular Spectrum Analysis**. Causality in Science, Oxford University press, Forthcoming.
- [49]. HASSANI, H.; HERAVI, S.; ZHIGLJAVSKY, A. (2009). **Forecasting European Industrial Production with Singular Spectrum Analysis**. International Journal of Forecasting. 25, pp. 103-118.
- [50]. HERRERA F. and LOZANO M. V. J. L. (1998). **Artificial Intelligence Review**. Tackling Real-coded Genetic Algorithms: Operators and Tools for Behavioral Analysis, journal, p. 265-319.
- [51]. JAMES, B. (2011). **Probabilidade: Um Curso em Nível Intermediário**. Coleção Projeto Euclides. Instituto de Matemática Pura e Aplica (IMPA).
- [52]. KUBRUSLY C. S. (2001). **Elements of Operator Theory**. Birkhäuser, Boston, 2001.
- [53]. KUBRUSLY (2007). **Measure Theory**. The First Curse.
- [54]. KUBRUSLY C. S. (2012). **Spectral Theory of Operators in Hilbert Space**. Birkhäuser, Boston, 2012.
- [55]. KUBRUSLY C. S., LEVAN, N. (2006). **Abstract Wavelets Generated by Hilbert Space Shift Operators**. Adv. Math. Sci. Appl. 14, 643-660.



- [56]. KRYKOVA, I. (2003). **Evaluating of Path-Dependent Securities with Low Discrepancy Methods**. Thesis. Faculty of the Worcester Polytechnic Institute, London (UK).
- [57]. LAW, A. M. and KELTON, W. D. (2000). **Simulation Modeling & Analysis**. McGraw-Hill Books, NY, Second Edition.
- [58]. LEVAN N. and KUBRUSLY, C. S. (2003). **A Wavelet Time-Shift Detail Decomposition**. Mathematics and Computers in Simulation. 63, 73-78.
- [59]. LEI C. AND RAN L. (2008). **Short-term wind speed forecasting model for Wind Farm Based on Wavelet Decomposition DRPT**. Nanjing. China, pp 2525-2529.
- [60]. LIMA, E. L. (2006). **Análise Real. Funções de uma Variável**. Volume 1, 8º Edição, Rio de Janeiro: IMPA.
- [61]. LUTKEPOHL, H. (2006). **New Introduction to Multiple Time Series Analysis**. Springer.
- [62]. MACEIRA, M. E. P. (1989). **Operação Ótima de Reservatórios com Previsão de Afluências**. Dissertação de Mestrado. Coordenação dos Programas de Pós-Graduação de Engenharia da Universidade Federal do Rio de Janeiro, Brasil.
- [63]. MALLAT S. (1998). **A Wavelet Tour of Signal Processing**. Academic Press, San Diego.
- [64]. MAKRIDAKIS, S. (1989). **Why Combining Works?** International Journal of Forecasting, v.5, p.601-603.
- [65]. MAKRIDAKIS, S. and WINKLER, R. L. (1983). **Averages of Forecasts: Some Empirical Results**. Management Science, v.29, n.9, p.987-996.
- [66]. MENEZES L. M.; TEIXEIRA Jr, L. A. T; CASSIANO K. M.; SOUZA, R. C.; and PESSANHA, J. F. M. (2012). **SSA, Hierarchical Clustering and ARIMA Model in Forecast of Affluent Flow**. The 32nd Annual International Symposium on Forecasting. Boston - USA .
- [67]. MORETTIN, P. A. (2006). **Wavelets in Statistics**. Institute of Mathematics and Statistics University of São Paulo, Brazil.
- [68]. MORETTIN, P. A. e TOLOI, L. M. C (2006). **Análise Séries Temporais**. 2ª Ed. ABE Projeto Fisher, Ed. Edgard Blucher.
- [69]. MORO B. **The Full Monte**. Risk, 8, nº 2, February 1995.
- [70]. MUBWANDARIKWA, E. (2007). **The Geometric Combination of Bayesian Forecasting Models**. Thesis. Department of Statistics. Open University, London (UK).



- [71]. NASCIMENTO, M. C. V.; ALÉM JUNIOR, D. J.; CHERRI, L. H.; MASSAMITSU, F. (2008). **Apresentações para Aulas de Modelagem Matemática**. São Carlos: ICMC-USP.
- [72]. NEWBOLD, P. and GRANGER, C. W. J. (1974). **Experience with Forecasting Univariate Time Series and Combination of Forecasts**. Journal Royal Statistical Society, series A, v.137, n.2, p.131-165.
- [73]. NIEDERREITER H. (1992). **Random Number Generation and Quasi-Monte Carlo Methods**. CBMS-NSF, Vol. 63. Philadelphia: SIAM.
- [74]. OGDEN, R. T. (1997). **Essential Wavelets for Statistical Applications and Data Analysis**. Birkhäuser, Boston.
- [75]. OLIVEIRA, J. F. and CARRAVILLA, M. A. (1998). **Programação Linear e Método Simplex**. Notas de Aula - FEUP. Versão 1, 1998.
- [76]. ONS - Operador Nacional do Sistema Elétrico. Site: <http://www.ons.com.br>.
- [77]. OWEN, A. B. (2000). **Monte Carlo Extension of Quasi-Monte Carlo**. Working Paper, Stanford University, 7 pp.
- [78]. PALIT A. K., POPOVIC D. (2005). **Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications**. Springer-Verlag, London.
- [79]. PASCAL, Y.; SORNETTE, D.; GHIL, M. (1998). **Data-Adaptive Wavelets and Multi-Scale SSA**. Institute of Geophysics and Planetary Physics. University of California Los Angeles at Los Angeles, CA 90095-1567.
- [80]. PASKOV S. (1995). **New Methodologies for Valuing Derivatives. Monte Carlo: Methodologies and Applications for Pricing and Risk Management**. Number 25, pp. 281-298.
- [81]. PESSANHA, J. F. M.; **TEIXEIRA Jr, L. A.**; MENESES, M. L.; SOUZA, R. C.; and CASSIANO, K. M. (2012). **Análise Espectral, Teoria Wavelet e Redes Neurais Artificiais na Previsão de Velocidade de Vento**. 57ª Reunião Anual da RBras. Piracicaba, Brasil.
- [82]. RAGSDALE, C. (2004). **Spreadsheet Modeling & Decision Analysis: A Practical Introduction to Management Science**. Fourth edition, South-Western.
- [83]. RUSSELL, S.J. and NORVIG, P. (1995). **Artificial Intelligence: A Modern Approach**. New Jersey: Prentice-Hall Inc.
- [84]. RIPLEY, B. D. (1987). **Stochastic Simulation**. John Wiley & Sons. Inc., 237 pp.
- [85]. ROCHA REIS, A.J. & ALVES DA SILVA, A.P. (2004). **Aplicação da Transformada Wavelet Discreta na Previsão de Carga de Curto Prazo via Redes Neurais**. Sba Controle & Automação vol.15 nº1, Campinas Jan. / Mar.



- [86]. SAWARAGI, Y.; NAKAYAMA, H.; and TANINO, T. (1985). **Theory of Multiobjective Optimization**. Academic Press, Inc., Orlando, Florida, U.S.A.
- [87]. SCHOELLHAMER, D. H. (2001). **Singular Spectrum Analysis for Time Series with Missing Data**. Geophys. Res. Lett 28, 3187-3190.
- [88]. SOBOL I. M. (1967). **On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals**. U.S.S.R. Computational Mathematics and Mathematics Physics, 7, pp.86-112.
- [89]. SOUZA, R. C. and CAMARGO, M. E. (1996). **Análise e Previsão de Séries Temporais: Os Modelos ARIMA**. Ijuí: SEDIGRAF.
- [90]. SOUZA, R. C.; **TEIXEIRA Jr, L. A.**; CASSIANO K. M.; MENEZES L. M.; and PESSANHA, J. F. M. (2012b). **Artificial Neural Networks Integrated with Singular Spectrum Analysis and Wavelet Theory in the Short-Term Wind Speed Forecasting**. International Workshop on Energy Efficiency for a More Sustainable World. Portugal.
- [91]. SOUZA, R. M.; **TEIXEIRA Jr, L. A.**; CASSIANO K. M.; SOUZA, R. C.; MENEZES L. M.; and PESSANHA, J. F. M. (2012a). **Streamflow Forecasting Using Singular Spectrum Analysis, ARIMA Model and Artificial Neural Network**. The 32nd Annual International Symposium on Forecasting. Boston - USA.
- [92]. YIOU P.; SORNETTE D.; and GHIL M. (2000). **Data-Adaptative Wavelets and Multi-Scale Singular Spectrum Analysis**. Physica D, 142:254-290.
- [93]. WALLIS, K. F. (2011). **Combining Forecasts - Forty Years Later**. Applied Financial Economics, 21, pp. 33-41.
- [94]. WERNER, L. & RIBEIRO, J. L. D. (2006). **Modelo Composto para Prever Demanda através da Integração de Previsões**. Produção, v. 16, n. 3, 2006.
- [95]. WIERZBICHI, A. P. (1980). **The Use of Reference Objectives in Multiobjective Optimization**. In Multiple Criteria Decision Making Theory and Application, Fandel, G. and Gal, T. (eds.), Springer-Verlag, New York, pp 469 - 486.
- [96]. WINKLER, R. L. (1989). **Combining Forecasts: A Philosophical Basis and Some Current Issues**. International Journal of Forecasting, v.5, p. 605-609.
- [97]. WINKLER, R. L. & MAKRIDAKIS, S. (1982). **The Combination of Forecasting**. Journal of the Royal Statistical Society, series A, v.146, p.150-157.
- [98]. TAFNER, M. A. (1996). **Redes Neurais Artificiais: Introdução e Princípios de Neurocomputação**. - Blumenau: EKO.



- [99]. **TEIXEIRA Jr, L. A.** (2009). **Combinação Geométrica de Métodos Preditivos**. Aplicação à Série de Consumo Residencial de Energia Elétrica. Dissertação de Mestrado, PUC-Rio.
- [100]. **TEIXEIRA Jr, L. A.**; PESSANHA, J. F. M.; and SOUZA, R. C. (2011a). **Combinação Linear de Previsões com Ajuste MINIMAX**. XIV Simpósio de Pesquisa Operacional e Logística da Marinha, Rio de Janeiro, Brasil.
- [101]. **TEIXEIRA Jr, L. A.**; PESSANHA, J. F. M.; and SOUZA, R. C. (2011b). **Análise Wavelet e Redes Neurais Artificiais na Previsão de Velocidade de Vento**. XVII Simpósio Brasileiro de Pesquisa Operacional, Ubatuba, Brasil. Disponível no site: <http://www.xliiisbpo.iltc.br/pdf/87357.pdf>.
- [102]. **TEIXEIRA Jr, L. A.**; MENESES, M. L.; PESSANHA, J. F. M. SOUZA, R. C.; CASSIANO, K. M. (2012a). **Redes Neurais Artificiais Wavelet Causal na Previsão da Radiação Solar Direta**. Simpósio Brasileiro de Pesquisa Operacional.
- [103]. **TEIXEIRA Jr, L. A.**; MENESES, M. L.; PESSANHA, J. F. M. SOUZA, R. C.; CASSIANO, K. M. (2013). **Residential Electricity Consumption Forecasting Using a Geometric Combination Approach**. International Journal of Energy and Statistics. Vol. 1, No. 2, 1-13.
- [104]. **TEIXEIRA Jr, L. A.**; MENESES, M. L.; PESSANHA, J. F. M. SOUZA, R. C.; CASSIANO, K. M. (2012b). **Modelagem Box & Jenkins, Decomposição Wavelet e Programação Matemática na Geração de Previsões**. XV Simpósio de Pesquisa Operacional & Logística da Marinha.
- [105]. TEZUKA S. (1998). **Financial Applications of Monte Carlo and Quasi-Monte Carlo Methods**. Random and Quasi-Random Point Sets, P. Hellekalek & G. Larcher, Eds., Springer-Verlag New York, pp. 303-332.
- [106]. THUILLARD M. (2001). **Wavelets in Soft Computing**. World Scientific Series in Robotics and Intelligent Systems, v.25. World Scientific Publishing.
- [107]. VANDERPLAATS, G. (1999). **Numerical Optimization Techniques for Engineering Design**. Third edition, Vanderplaats Research & Development, Inc, Colorado Springs.
- [108]. VANKAN, W.J. and MAAS, R. (2002). **Approximate Modelling and Multiobjective Optimization in Aeronautic Design**. CMMSE, National Aerospace Laboratory NLR, Alicante, Spain.
- [109]. VAUTARD P.; YIOU and GHIL M. (1992). **Singular-Spectrum Analysis**. A toolkit for Short, Noisy Chaotic Signals. Physica D, 58:95-126.
- [110]. ZANDONADE, E. (1993). **Aplicação da Metodologia de Redes Neurais em Previsão de Séries Temporais**. Dissertação de Mestrado. Departamento de Engenharia Elétrica, PUC-RJ.



- [111]. ZHANG, N. and BEHERA, P.K. (2012). **Solar Radiation Prediction Based on Recurrent Neural Networks Trained by Levenberg-Marquardt Backpropagation Learning Algorithm**. Innovative Smart Grid Technologies. Washington, January.
- [112]. ZELENY, M. (1974). **Linear Multiobjective Programming**. Springer-Verlag. New York.
- [113]. ZELENY, M. (1982). **Multiple Criteria Decision Making**. New York: McGraw-Hill.
- [114]. ZERVAS, P.L.; SARIMVEIS, H.; PALYVOS, J.A.; MARKATOS, N.C.G. (2008). **Prediction of Daily Global Solar Irradiance on Horizontal Surfaces Based on Neural-Network Techniques**. Renewable Energy, 33, pp 1796-1803.
- [115]. ZHENG, G.; STARCK, J. L.; CAMPBELL, J.; MURTAGH (1999). **The Wavelet Transform for Filtering Financial Data Streams**. Journal of Computational Intelligence in Finance, 7 18-35.
- [116]. ZHOU, H.; SUN, W.; LIU, D.; ZHAO, J.; YANG, N. (2011). **The Research of Daily Total Solar-Radiation and Prediction Method of Photovoltaic Generation Based on Wavelet-Neural Network**. Power and Energy Engineering Conference, Wuhan, March.
- [117]. ZIONTS, S. and WALLENIUS, J. (1976). **An Interactive Programming Method for Solving the Multiple Criteria Problem**. Mgmt. Sci., 22, 6, pp 632 - 663.
- [118]. ZOU, H. & YANG, Y. (2004). **Combining Time Series Models for Forecasting**. International Journal of Forecasting, v.20, n.1, p.69-84.
- [119]. ZSOLT, L. (2006). **Redes Neurais Artificiais: Fundamentos e Aplicações**. 4º Ed, Livraria de Física.