



Rafael Moraes de Souza

**Modelagem de Séries Periódicas via Estruturas
PAR(p) Utilizando Encolhimento Wavelet**

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

Julho de 2013



Rafael Moraes de Souza

**Modelagem de Séries Periódicas via Estruturas
PAR(p) Utilizando Encolhimento Wavelet**

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. José Francisco Moreira Pessanha
UERJ

Profa. Karla Tereza Figueiredo Leite
UEZO

Profa. Luciana Conceição Dias Campos
UFJF

Prof. Rogério Silva de Mattos
UFJF

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico

Rio de Janeiro, 22 de julho de 2013

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

Rafael Morais de Souza

Graduou-se em Ciências Econômicas em 2006, com ênfase em Econometria Espacial e obteve o título de Mestre em Economia Aplicada em 2008 (Modelos Econométricos com Insumo-Produto), ambos pela Universidade Federal de Juiz de Fora. Atualmente, Doutorando em Engenharia Elétrica (Métodos de Apoio à Decisão) pela Pontifícia Universidade Católica do Rio de Janeiro.

Ficha Catalográfica

Souza, Rafael Morais de

Modelagem de séries periódicas via estruturas PAR(p) utilizando encolhimento wavelet / Rafael Morais de Souza; orientador: Reinaldo Castro Souza. – 2013.

105 f. ; 30 cm

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

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Séries temporais. 3. Encolhimento wavelet. 4. Modelo periódico. I. Souza, Reinaldo Castro. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

*Aos meus pais, Tadeu e Imaculada, à minha noiva Flaviane,
ao meu irmão Daniel, aos meus avós, Joaquim e Santinha e
à minha tia Gina pelo amor incondicional e pela dedicação
de sempre.*

Agradecimentos

A Deus pela constante presença.

Ao Prof. Dr. Reinaldo Castro Souza, pela orientação, pelos conselhos, ideias e dedicação.

Aos professores membros da banca pelas contribuições nas diversas etapas da elaboração deste trabalho.

A todos os meus amigos de PUC-Rio, em especial, Luiz Albino, Moisés, Guilherme, Keila, Pedro Guilherme e Fernando, pelas interações e ajudas nos estudos e na elaboração deste trabalho.

Aos meus grandes amigos de república, por terem me recebido com grande carinho. Certamente, tornaram este doutorado mais fácil.

A todos os professores, funcionários e alunos do programa de pós-graduação, por formarem um ótimo ambiente de estudo e pesquisa.

À PUC-Rio, por toda a estrutura e à CAPES e ao CNPq pelo apoio financeiro.

A todas as pessoas que participaram direta ou indiretamente para a realização deste trabalho.

Resumo

Souza, Rafael Moraes de; Souza, Reinaldo Castro (Orientador). **Modelagem de séries periódicas via estruturas PAR(p) utilizando encolhimento wavelet.** Rio de Janeiro, 2013. 105p. Tese de Doutorado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese apresenta uma modelagem para séries temporais que possuem uma estrutura de autocorrelação que depende não somente do intervalo de tempo entre as observações, mas também do período observado. Esta modelagem consiste na combinação entre as metodologias de encolhimento wavelets e modelos auto-regressivos periódicos – PAR(p). As wavelets vêm sendo utilizadas na literatura de séries temporais como um procedimento auxiliar de pré-processamento da série em análise; e o modelo PAR(p) tem sua importância reconhecida devido ao seu emprego em séries hidrológicas mensais. Especificamente, no setor elétrico brasileiro, pode-se observar a predominância da energia de origem hidrelétrica. Uma das principais características das matrizes de energia com esta composição é a forte dependência do padrão de precipitação e das condições hidrológicas futuras, o que torna a série de vazão muito irregular e difícil de ser modelada. Portanto, a geração hidrelétrica pode ser considerada uma variável estocástica e pequenos avanços na modelagem estocástica de vazões permitem um melhor planejamento da operação do sistema. Assim, esta tese teve como objetivos obter melhores previsões e geração de cenários a partir do modelo PAR(p) e avançar na modelagem estocástica de vazões. Os principais resultados obtidos indicaram que a combinação das metodologias obteve melhor desempenho do que apenas a modelagem PAR (p), tanto em casos simulados, quanto no emprego a séries reais.

Palavras-chave

Séries Temporais; Encolhimento Wavelet; Modelo Periódico.

Abstract

Souza, Rafael Morais de; Souza, Reinaldo Castro (Advisor). **Modeling of periodic series via PAR(p) structures utilizing wavelet shrinkage**. Rio de Janeiro, 2012. 105p. PhD Thesis – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis presents an approach to modeling time series that have an autocorrelation structure that depends not only on the time interval between the observations, but the observed period. This approach consists of a combination between wavelet shrinkage and periodic autoregressive models – PAR(p). Wavelets have been used in the literature of time series as an auxiliary procedure for preprocessing the series under analysis, and the PAR(p) model has its importance recognized due to its use in hydrological series monthly. Specifically, in the Brazilian electricity sector, it is possible to observe the predominance of hydroelectric energy. One of the main characteristics of energy matrices with this composition is the strong dependence on rainfall patterns and future hydrological conditions, which turn the streamflow series very irregular and difficult to be modeled. Therefore, hydroelectric generation can be considered a stochastic variable and small advances in stochastic modeling of streamflows allow for better planning of the system's operation. Thus, this thesis aimed to improve forecasting and scenario generation with the PAR(p) model and to contribute to stochastic modeling of flow. The main results indicated that the combination of methodologies achieves better results than only the PAR(p) modeling, both in simulated cases and in real time series.

Keywords

Time Series; Wavelet Shrinkage; Periodic Models.

Sumário

1. Introdução	12
1.1. Objetivo do trabalho	13
1.2. Relevância do trabalho	15
1.3. Planejamento do setor elétrico	16
1.4. Organização	20
2. Metodologias	21
2.1. Metodologia de Box & Jenkins	21
2.2. Modelo PAR(p)	24
2.2.1 Geração de séries sintéticas no setor elétrico	29
2.2.2. Geração de séries sintéticas com o modelo PAR(p)	29
2.3. Utilização de wavelets	33
2.4. Teoria wavelet	33
2.4.1. Domínio do tempo x domínio da frequência	34
2.4.2. Wavelets	37
2.4.2.1. Função wavelet	38
2.4.2.2. Transformadas wavelet contínua e discreta	39
2.4.2.3. Decomposição wavelet	42
2.4.2.4. Encolhimento wavelet	44
2.4.2.4.1. Regras de limiarização	46
2.4.2.4.2. Limiares	47
3. Metodologia proposta	49
3.1. Encolhimento wavelet	50
3.2. Testes de resíduos	51
3.2.1. Ljung-Box	52
3.2.2. Jarque-Bera	53
3.2.3. Teste ARCH	53
3.3. Aplicação dos testes e modificação da proposta inicial para metodologia proposta	54
3.4. Comparação das metodologias	57
3.5. Aplicação a casos simulados	58
3.6. Aplicação às séries de ENAs e avaliação de desempenho do modelo	59
3.6. 1. Testes de média	60
3.6. 2. Testes de variância	61
3.6. 3. Testes de aderência	62
3.6. 4. Análise de sequências	63
4. Análise dos resultados	66
4.1. Resultados da aplicação às séries dos casos simulados	66
4.2. Resultados da aplicação às séries de ENA	75
5. Conclusões	87
Referências bibliográficas	90

Anexos	96
Anexo 1 – Curvas de funções wavelets	96
Anexo 2 – Testes de resíduos (p-valores) para diferentes setups da série temporal completa referente ao caso 1	97
Anexo 3 – Testes de resíduos (p-valores) para diferentes setups da série temporal completa referente ao caso 2	100
Continuação	101
Anexo 4 – Wavelets utilizadas no encolhimento das séries de meses referentes ao caso 3 e testes de resíduos (p-valores)	103
Anexo 5 – Wavelets utilizadas no encolhimento das séries de meses referentes ao caso 4 e testes de resíduos (p-valores)	103
Anexo 6 – Wavelets utilizadas no encolhimento das séries de meses da ENA Sudeste/Centro-Oeste e testes de resíduos (p-valores)	104
Anexo 7 – Wavelets utilizadas no encolhimento das séries de meses da ENA Sul e testes de resíduos (p-valores)	104
Anexo 8 – Wavelets utilizadas no encolhimento das séries de meses da ENA Nordeste e testes de resíduos (p-valores)	105
Anexo 9 – Wavelets utilizadas no encolhimento das séries de meses da ENA Norte e testes de resíduos (p-valores)	105

Lista de figuras

Figura 1 - Filtros do modelo ARIMA (p,d,q)	22
Figura 2 - Estágios da metodologia Box & Jenkins	24
Figura 3 - Resolução no plano tempo-frequência pela STFT	36
Figura 4 - Resolução no plano tempo-frequência pela análise wavelet	38
Figura 5 - Exemplo de uma decomposição wavelet de nível 2	44
Figura 6 - Proposta inicial	50
Figura 7 - Metodologia proposta	56
Figura 8 - Sequência Negativa	64
Figura 9 - Séries históricas de energia natural afluyente (ENA)	76
Figura 10 - Médias e desvios-padrão médios dos cenários	80
Figura 11 - Testes para o subsistema Sudeste/Centro-Oeste	81
Figura 12 - Testes para o subsistema Sul	82
Figura 13 - Testes para o subsistema Nordeste	83
Figura 14 - Testes para o subsistema Norte	84
Figura 15 - MAPE das previsões out-of-sample dos modelos PAR(p) e WS + PAR(p) para os anos de 2006 a 2010	86

Lista de tabelas

Tabela 1 - Características utilizadas para a elaboração dos setups.....	51
Tabela 2 - Sequência negativa	64
Tabela 3 - Wavelets utilizadas no encolhimento das séries de meses referentes ao caso 1 e testes de resíduos.....	67
Tabela 4 - Parâmetros reais e estimados e MAPEs dos modelos (caso 1)	68
Tabela 5 - Wavelets utilizadas no encolhimento das séries de meses referentes ao caso 2 e testes de resíduos.....	69
Tabela 6 - Parâmetros reais e estimados e MAPEs dos modelos (caso 2)	70
Tabela 7 - Parâmetros reais e estimados e MAPEs dos modelos (caso 3)	71
Tabela 8 - Parâmetros reais e estimados e MAPEs dos modelos (caso 4)	72
Tabela 9 - RMSE e MAPE da previsão <i>out-of-sample</i> (caso 1).....	73
Tabela 10 - RMSE e MAPE da previsão <i>out-of-sample</i> (caso 2).....	73
Tabela 11 - RMSE e MAPE da previsão <i>out-of-sample</i> (caso 3).....	74
Tabela 12 - RMSE e MAPE da previsão <i>out-of-sample</i> (caso 4).....	75
Tabela 13 - RMSE e MAPE dos ajustes das séries de ENA	78
Tabela 14 - Análise das sequências negativas (PAR(p)).....	85
Tabela 15 - Análise das sequências negativas (WS + PAR(p)).....	85

Introdução

Esta tese apresenta uma metodologia para séries temporais que possuem uma estrutura de autocorrelação que depende não somente do intervalo de tempo entre as observações, mas também do período observado. Este é o caso de, por exemplo, séries hidrológicas mensais, que apresentam um comportamento periódico de suas propriedades probabilísticas, *e.g.* média, variância, assimetria e estrutura de correlação. A modelagem proposta para tal é baseada na combinação do encolhimento de ruídos wavelets com modelos auto-regressivos periódicos de ordem p (PAR(p)). Com isso, o objetivo é realizar um pré-tratamento das séries para efetuar a modelagem PAR(p) de uma forma mais apurada ao serem utilizadas séries tratadas.

Os modelos auto-regressivos periódicos (*periodic autoregressive* – PAR(p)), introduzidos por Thomas e Fiering (1962), são modelos utilizados para séries temporais que apresentam uma estrutura de autocorrelação que depende não somente da defasagem entre as observações, mas também da estação do ano. Um exemplo didático de sazonalidade ou de correlação que varia periodicamente é destacado em Hipel e McLeod (1994). Um exemplo citado pelos autores é o de que, em uma determinada localidade do hemisfério Norte, a temperatura mensal para janeiro ao longo dos anos pode flutuar com variância constante em torno de uma média de -5°C . A maneira na qual a temperatura de janeiro está correlacionada com dezembro e novembro, assim como no janeiro anterior, tende a permanecer a mesma ao longo dos anos.

A importância de se utilizar o modelo PAR(p) é reconhecida devido ao seu emprego em séries hidrológicas mensais. De fato, na literatura pode ser observada a ampla aplicação do modelo à hidrologia. Thomas e Fiering (1962) utilizam os modelos para o problema de simulação de vazões/ENAs. Mondal e Wasimi (2006) propõem novas técnicas para esta simulação e as aplicam a uma série mensal do Rio Ganges. Maceira *et al* (2006) discutem a modelagem PAR(p) no contexto do setor elétrico brasileiro. Souza *et. al.* (2012) propõem a identificação

da ordem e a geração dos cenários via Bootstrap para utilizá-los na Programação Dinâmica Estocástica (PDE) do setor elétrico brasileiro.

Por sua vez, wavelets são funções matemáticas que separam dados em suas diferentes componentes frequenciais e extraem cada componente de acordo com sua resolução de escala adequada. Com isto, o objetivo é transformar a informação de uma série em coeficientes que possam ser manipulados, analisados ou até mesmo utilizados para reconstruir a série original (OLIVEIRA, 2007).

As wavelets vêm sendo utilizadas na literatura de séries temporais, combinadas com outros tipos de modelos preditivos e resultando em relevantes ganhos em termos de modelagem. Neste contexto, a teoria wavelet consiste em um procedimento auxiliar de pré-processamento da série em análise, que pode ser realizado, geralmente, de duas maneiras: por meio da decomposição ou por meio do encolhimento de ruídos da série temporal a ser modelada. Arino e Vidakovic (1995) e Arino (1995), propuseram a decomposição das séries temporais em subséries para se adequar um método preditivo distinto para a modelagem individual de cada uma delas, e posteriormente utilizaram as previsões das subséries, com a finalidade de se obterem as previsões da série temporal original. Donoho e Johnstone (1993, 1994, 1998), Donoho *et al* (1993, 1995) e Johnstone e Silverman (1996) propuseram um procedimento para se obter, à partir de séries originais de tempo, séries filtradas, que são possíveis pontos de partida à aplicação de metodologias de previsão.

1.1

Objetivo do trabalho

Assim, esta tese consiste no desenvolvimento de uma combinação entre as metodologias de encolhimento wavelets e modelos auto-regressivos periódicos de forma a respeitar as propriedades estatísticas em todas as fases da modelagem, isto é, tanto na etapa do encolhimento wavelet (no que diz respeito à sua aplicação e à diminuição de ruídos presentes nas séries), quanto na etapa de modelagem $PAR(p)$. Além do objetivo geral, o trabalho teve como objetivos específicos: 1) obter melhores previsões e geração de cenários a partir do modelo $PAR(p)$; e 2) avançar no que diz respeito à modelagem estocástica de vazões.

A metodologia se dá em duas etapas: inicialmente é realizado o encolhimento wavelet para remover ruídos das séries. Em seguida, com as séries filtradas (menos ruidosas) a modelagem $PAR(p)$ é utilizada para a previsão ou a geração de cenários. Para a aplicação da mesma, foram utilizados dois tipos de séries: 1) quatro séries temporais mensais simuladas com 1.200 observações, realizações de diferentes processos $PAR(p)$; e 2) as séries de energia natural afluentes (ENAs) utilizadas para o planejamento da operação do setor elétrico brasileiro.

Diversos são os trabalhos que destacam os ganhos obtidos com as combinações entre decomposição e/ou encolhimento wavelet e modelos preditivos, dentre os quais podem-se citar: 1) Fayal (2008), Liu *et al* (2010), Catalão *et al* (2011) Teixeira Júnior *et al* (2011) e Teixeira Júnior *et al* (2012) e Guan *et al* (2013) combinaram decomposição wavelet e redes neurais artificiais; 2) Arino (1995), Schlüter e Deuschle (2010) e Li *et al* (2011) utilizaram decomposição wavelet e modelos ARIMA; 3) Campisi-Pinto *et al* (2012) utilizaram encolhimento wavelet e redes neurais artificiais; 4) Rocha (2008) empregou o encolhimento wavelet juntamente com uma combinação de previsões feita a partir do método de suavização exponencial e do modelo ARIMA; 5) Alrumaih e Al-Fawzan (2002) testaram encolhimento wavelet com os seguintes métodos preditivos: regressão linear, médias móveis simples, amortecimento exponencial e modelos ARIMA; e 6) Liu *et al* (2007) e Schlüter e Deuschle (2010) combinaram encolhimento wavelet e modelos ARIMA.

A combinação de métodos aqui apresentada, por sua vez, contempla o encolhimento de ruídos wavelets e os modelos $PAR(p)$, e como parte dos avanços metodológicos incorporados nesta tese, podem-se citar:

- i) a realização de uma combinação de wavelets com modelos auto-regressivos periódicos, ainda não contemplada na literatura;
- ii) a utilização de testes de resíduos para validar os encolhimentos wavelets, garantido que apenas ruídos sejam eliminados e não sejam perdidas informações relevantes à modelagem das séries;
- iii) no intuito de verificar a adequabilidade do método proposto, é realizada a modelagem de algumas séries simuladas, permitindo assim observar e comparar o desempenho do mesmo em um ambiente controlado.

Com isso, o método aqui proposto amplia a potencialidade dos modelos $PAR(p)$. Cabe ressaltar que a escolha pela aplicação às séries hidrológicas mensais neste setor se deu por dois motivos: 1) o setor elétrico brasileiro já utiliza o modelo $PAR(p)$; e 2) uma pequena variação na modelagem das vazões pode causar grande impacto em termos de valores para o setor, uma vez que o país ocupa a nona posição mundial em termos de geração energética, produzindo um total de aproximadamente 470 TWh/ano (OECD, 2012).

1.2

Relevância do trabalho

Esta tese visa contribuir para a melhoria da previsão de séries temporais e da geração de cenários a partir do modelo $PAR(p)$. Em outras palavras, propõe-se uma nova abordagem para levantar e discutir pontos de melhorias, atingindo menores índices de erros e respeitando os testes estatísticos necessários.

Especificamente, com relação à aplicação às séries reais, deve-se destacar que o Setor Elétrico Brasileiro é amplamente dependente das vazões e todo seu planejamento está baseado em modelos que utilizam as mesmas de forma previstas e/ou simuladas. O modelo $PAR(p)$ é a ferramenta utilizada neste setor para a geração dos cenários de afluência que são utilizados posteriormente no cálculo de otimização estocástica (MARCATO, 2002).

Além disso, o sistema de planejamento da expansão, operação e precificação do setor elétrico brasileiro tem sido questionado ao longo dos últimos anos quanto à sua eficácia na correta representação dos cenários sintéticos de ENAs/vazões e, como consequência, podem-se destacar distúrbios no preço “justo” do PLD (Preço de Liquidação das Diferenças). Em junho de 2010, a ANEEL expediu um despacho (despacho nº 2.654, de 24 de Junho) determinando a revisão das previsões de vazões do Programa Mensal de Operação (PMO) o que resultou em recálculo do PLD. Este recálculo apresentou aumento significativo em algumas semanas e redução expressiva em outras. (FERREIRA, 2013).

Como contribuições desta tese podem-se citar: a realização da combinação dos métodos; os ganhos obtidos na previsão e na geração de cenários; e, com relação à aplicação específica no setor elétrico brasileiro, pode-se observar que o método proposto foi superior ao $PAR(p)$ em vários aspectos. Neste ponto é

importante destacar que pequenos avanços na modelagem estocástica de vazões podem contribuir para a melhoria do planejamento da operação do sistema.

1.3

Planejamento do setor elétrico

A energia de origem hidrelétrica predomina na matriz elétrica brasileira. Uma das principais características de matrizes energéticas com esta composição é a forte dependência dos regimes de chuvas e das condições hidrológicas futuras. Assim, a geração hidrelétrica pode ser considerada uma variável estocástica de modo que a cada instante, deve-se determinar quanto deplecionar dos reservatórios e/ou qual o nível de geração hidrotérmica ótimo. É importante observar que, uma tomada de decisão hoje terá consequências no futuro e, desta forma, o problema da operação do Sistema Integrado Nacional pode ser considerado um problema complexo que, segundo Toscano (2009), pode ser tratado como um problema de otimização de grande porte, dinâmico, estocástico, interconectado, não-linear e que requer simplificações como a divisão deste em horizontes diferentes de planejamento (longo prazo, médio prazo e curto prazo).

Cabe ao Operador Nacional do Sistema Elétrico (ONS) a responsabilidade do planejamento da operação do sistema elétrico brasileiro. Segundo Luna *et al* (2006), o objetivo do modelo de planejamento energético a cargo da ONS é minimizar o custo total de operação por meio de decisões de geração (térmica ou hidráulica), intercâmbio entre regiões e cortes de carga, sendo estas variáveis dependentes das vazões no futuro. Do mesmo modo, se faz necessário o estudo das vazões e, também o estudo e desenvolvimento de modelos capazes de prever e/ou simular séries hidrológicas.

Segundo Tolmasquim (2011), com a crise de racionamento do ano de 2001 ficou clara a necessidade de se modificar a estrutura institucional do setor elétrico brasileiro, que não foi capaz de atingir os três objetivos de um serviço público, em especial com relação à prestação de serviços de energia elétrica como a confiabilidade de suprimento, modicidade tarifária e universalidade. De acordo com Torres e Almeida (2003) seria necessário que, para que houvesse o suprimento adequado da energia elétrica, agentes privados e públicos participassem em diversas frentes, cada qual em seu papel (investimentos por

parte dos agentes privados e políticas de desenvolvimento e questões inerentes ao planejamento da expansão, à regulação, à concessão e ao licenciamento ambiental do setor por parte dos agentes públicos).

De tal forma, o Novo Modelo em vigor visa atender aos três objetivos de um serviço público de eletricidade já citados (confiabilidade de suprimento, modicidade tarifária e universalidade) e, para isso, foram tomadas quatro grandes medidas: 1) modificação do modo de contratação de energia por parte das distribuidoras; 2) retomada no planejamento no setor; 3) criação de programas efetivos de universalização e; 4) reorganização institucional¹ (TOLMASQUIM, 2011).

Com relação ao planejamento do setor, por meio da realização do Plano Decenal de Energia, objetiva-se orientar ações e decisões para que haja o equilíbrio entre as projeções de crescimento econômico do Brasil e a consequente necessidade de expansão de oferta de energia elétrica e, por meio dos resultados dos leilões, é determinada a oferta para a elaboração do Plano de Operação, que condiciona o programa de operação. Realizada a operação, é feito o registro da produção e dos fluxos elétricos que servem de base para a análise do desempenho do sistema (ONS) e para o processo de contabilização e liquidação das transações de energia no mercado de curto prazo (Câmara de Comercialização de Energia Elétrica – CCEE) (TOLMASQUIM, 2011).

No planejamento do setor elétrico brasileiro, devem-se observar duas funções principais: 1) o planejamento da expansão, que trata da ampliação da capacidade instalada do sistema visando atender à demanda futura de forma segura e econômica, levando-se em consideração a projeção da demanda de energia elétrica e os recursos energéticos e tecnológicos disponíveis e; 2) o planejamento da operação que, tendo em vista a estratégia de expansão, tem o objetivo de atender a demanda de forma a minimizar o custo de operação

¹ Ficou estabelecido um conjunto de agentes institucionais, com competências e atribuições distintas e que podem ser classificados em três níveis: 1) agentes que executam atividades de governo: Conselho Nacional de Política Energética (CNPE), Ministério de Minas e Energia (MME) e Comitê de Monitoramento do Setor Elétrico (CMSE); 2) agentes que executam atividades regulatórias: Agência Nacional de Energia Elétrica (ANEEL) e; 3) entidades de direito privado que executam atividades especiais: Empresa de Energia Elétrica (EPE) que realiza o planejamento de médio e longo prazo, Câmara de Comercialização de Energia Elétrica (CCEE) que viabiliza as atividades de comercialização e Operador Nacional do Sistema Elétrico (ONS) que coordena o SIN (TOLMASQUIM, 2011).

seguindo um nível de segurança pré-determinado (FORTUNATO, 1990) (TOLMASQUIM, 2011).

Essas funções são complementares e, institucionalmente, o planejamento da expansão ocorre no âmbito da Empresa de Pesquisa Energética (EPE), enquanto o planejamento da operação ocorre no ONS que exerce as atividades de coordenação do Sistema Interligado Nacional (SIN)².

A energia de origem hidrelétrica predomina na matriz elétrica brasileira. Este fato decorre, principalmente, da extensa superfície territorial do país que possui muitos planaltos e rios. Segundo EPE (2011), a participação da energia hidrelétrica no total da capacidade instalada no Brasil no ano de 2010 foi de 71,21% com 80.703 MW, seguido pelo setor termelétrico, com 26,20% do total (29.689 MW) e pelas energias nuclear e eólica com, respectivamente, 1,77% e 0,82% do total da capacidade instalada.

Devido à distância entre as grandes centrais hidrelétricas e os principais centros de carga, existem intensos fluxos de energia entre as diversas regiões do país que possuem regimes pluviométricos diferentes. Estes fluxos se devem ao despacho hidrelétrico, cujo objetivo principal é utilizar os estoques de água dos reservatórios das usinas hidrelétricas de forma eficiente, considerando as condições hidrológicas em cada região. Assim, o controle das vazões é de fundamental importância para evitar desperdício e obter o máximo de energia contida, o que exige um esforço do setor elétrico no sentido de desenvolver ferramentas e técnicas de modelagem de operação de reservatórios.

Devido a essa complexidade, o planejamento da operação é realizado por meio de uma série de modelos de cálculos³, tendo em vista a minimização do custo de operação. As variáveis de decisão são a geração hidro e termelétrica, em que apenas a hidrologia é considerada como variável estocástica. O acoplamento entre os modelos é feito por meio da função de custo futuro da operação energética.

Na etapa do planejamento da operação de médio prazo (cinco anos), utiliza-se o modelo Newave. Este modelo objetiva a minimização do custo de

² O SIN abrange quase todo o território do Brasil e compreende sete subsistemas: Sul, Sudeste/Centro-Oeste, Nordeste, Norte, Acre, Rondônia e Manaus, sendo que os quatro primeiros correspondem às regiões geoeletricas atuais e os três últimos correspondem a sistemas isolados que serão integrados ao SIN no horizonte do Plano Decenal de Energia (TOLMASQUIM, 2011).

³ Para maiores detalhes ver Tolmasquim (2011).

operação e, para isso, utilizam-se de séries hidrológicas sintéticas ou do conjunto de séries históricas⁴ geradas pelo módulo GEVAZP (Modelo de Geração de Séries Sintéticas de Energia e Vazão) em uma estrutura paralela (pente), obtendo o custo futuro que é utilizado nos outros módulos.

O Newave representa o parque hidrelétrico de forma agregada e o cálculo da política de operação se baseia na programação dinâmica dual estocástica. O modelo apresenta quatro módulos computacionais⁵: 1) cálculo do sistema equivalente; 2) cálculo do modelo estocástico de energias afluentes; 3) cálculo da política de operação hidrotérmica e; 4) simulação da operação. Os cenários fornecidos pelo módulo GEVAZP, são utilizados nos módulos 1 e 2 citados.

A função de custo futuro obtida no modelo Newave é usada pelo modelo Decomp em estudos de curto prazo (até 12 meses). Neste modelo, são definidas as metas de geração por usina. Os cálculos deste módulo também estão sujeitos às afluições estocásticas e, como resultado desta etapa do planejamento obtêm-se os custos marginais de operação. Segundo Maceira *et al* (2006), os cenários hidrológicos podem ser representados por meio de uma árvore de afluições, com probabilidades de ocorrência associadas a cada ramo.

O último módulo do planejamento da operação ocorre no Dessem-Pat. Neste, o despacho é calculado com o horizonte de até duas semanas, sendo que os dois primeiros dias são discretizados em intervalos de trinta minutos, não existindo a coordenação entre o despacho central e os regionais.

No planejamento da expansão também são utilizados os módulos Newave e Decomp para a tomada de decisão. Estes módulos são utilizados no cálculo do custo marginal de operação (CMO). Este custo é comparado com o custo marginal de expansão (CME) que é obtido com base nos resultados dos leilões de energia nova. No caso destes custos serem iguais, deve haver uma expansão na capacidade instalada⁶.

Esta tese tem como foco o módulo estocástico que serve de base aos módulos de planejamento Newave e Decomp. Por depender das condições hidrológicas futuras, a geração hidrelétrica pode ser considerada uma variável

⁴ A série histórica é uma sequência de valores mensais de uma vazão natural observada no passado e a série sintética é uma amostra dentre as possíveis afluições, produzida por um modelo estocástico de vazões (TOLMASQUIM, 2011).

⁵ Para maiores detalhes ver CEPEL (2000).

⁶ Para maiores detalhes ver Tolmasquim (2011).

estocástica e, a cada instante, deve-se determinar quanto deplecionar dos reservatórios. Dessa forma, o foco será especificamente nas séries de ENAs, de periodicidade mensal. Esta energia corresponde à estimativa da energia que pode ser gerada com todas as vazões afluentes a cada um dos reservatórios que compõem aquele reservatório equivalente. Cada subsistema do SIN é representado por um reservatório equivalente que é composto pelas usinas de sua região.

1.4

Organização

Para apresentar o modelo, sua construção e suas aplicações à série temporal simulada e às séries de ENAs utilizadas para o planejamento do setor elétrico brasileiro, esta tese está organizada da seguinte forma: além da introdução, no segundo capítulo, apresentam-se as metodologias utilizadas: 1) a metodologia PAR(p) e, 2) a teoria wavelet.

No terceiro capítulo, é apresentada a metodologia proposta e são realizadas as explicações de como são feitas a aplicação e a avaliação da mesma. No quarto capítulo são expostos os resultados referentes à aplicação da metodologia proposta a alguns casos simulados (onde são realizadas previsões para os meses de um ano à frente) e às séries reais (onde são gerados cenários para o planejamento do setor elétrico).

No quinto capítulo são apresentados todos os resultados obtidos. Finalizando, no sexto capítulo têm-se as conclusões finais da tese.

2

Metodologias

Neste capítulo, estão dispostas as metodologias que compõem a combinação proposta no capítulo 3. Inicialmente, é apresentada a metodologia $PAR(p)$, começando pela metodologia de Box & Jenkins, onde se observam os principais conceitos e abordagens, a formulação básica dos principais modelos e os estágios fundamentais da modelagem e, em seguida, descreve-se o modelo $PAR(p)$, seus passos: identificação, estimação e verificação e sua utilização para a geração de séries sintéticas no setor elétrico. A parte seguinte trata da teoria wavelet, em que é realizada uma discussão entre domínio do tempo e domínio da frequência; posteriormente, a análise de Fourier *short time* é brevemente exposta; e por fim, é apresentada a metodologia wavelet, de forma a se destacar, na parte final, o encolhimento wavelet e seus vários limiares e regras de limiarização.

2.1

Metodologia de Box & Jenkins⁷

Nesta seção são apresentados os principais conceitos e abordagens da metodologia Box & Jenkins, bem como a formulação básica dos principais modelos e os estágios fundamentais da modelagem, elaborados para o ajuste de modelos estocásticos da família auto-regressivo integrado de médias móveis (*autoregressive integrated moving average* – $ARIMA(p,d,q)$) a séries temporais.

De acordo com Souza e Camargo (2004), o fundamento teórico da modelagem Box & Jenkins se baseia na passagem de um processo ruído branco⁸ por filtros lineares, denominado Teoria Geral de Sistemas Lineares. Na Figura 1, é possível observar a passagem do ruído branco pelo conjunto dos sucessivos filtros associados aos parâmetros dos modelos $ARIMA(p,d,q)$ que representam os sistemas estimulados pela série a_t que geram a série temporal Z_t . Eles são,

⁷ Para maiores detalhes, ver: Box *et al* (1994), Hamilton (1994) e Morettin e Toloí (2006).

⁸ Um processo ruído branco é um processo cujas variáveis aleatórias são independentes, identicamente distribuídas de média zero e variância constante.

respectivamente, o filtro de médias móveis (parâmetro q), o filtro auto-regressivo estacionário (parâmetro p) e o filtro de integração não-estacionário (parâmetro d).

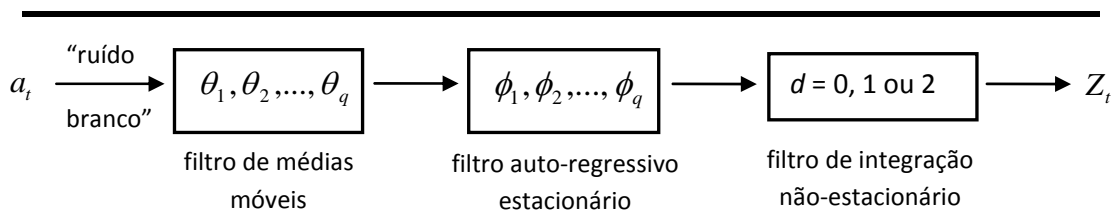


Figura 1 - Filtros do modelo ARIMA (p,d,q)

Fonte: adaptado de Souza e Camargo (2004).

Porém, a análise de séries temporais segue no caminho inverso ao modelo ARIMA(p,d,q). Existe uma série temporal Z_t gerada pelo sistema em questão, no qual o objetivo é associar um modelo que corresponda aos processos internos que geraram esta série. Assim, estabelecendo-se um modelo para esta representação, a série a_t corresponde à sequência de ruídos que resultam ao se extrair de Z_t os valores obtidos com o modelo ajustado a essa série.

Em resumo, segundo Souza e Camargo (2004), se $X_t = \Delta^d Z_t$ for estacionária (em que Z_t é a série original, Δ^d é o operador de diferença e X_t é a série estacionária após a aplicação das diferenças), a mesma pode ser modelada por um ARMA (p,q), contudo, se X_t é uma diferença de grau d de Z_t , pode-se dizer que Z_t segue um modelo ARIMA (p,d,q):

$$\phi(B)\Delta^d Z_t = \theta(B)a_t \quad (1)$$

Em que p é a ordem de $\phi(B)$, d é o número de diferenças que torna a série estacionária e q é a ordem de $\theta(B)$. Os parâmetros p e q representam a quantidade de parâmetros referentes aos comprimentos de defasagem em que se observam valores significativos das autocorrelações e que correspondem às particularidades do sistema gerador das séries a serem explicadas pelo modelo. A expressão (1) também pode ser escrita da seguinte maneira:

$$Z_t = \alpha + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2)$$

Onde Z_t é uma série temporal, α é a sua média, $\phi_1 \dots \phi_p$ são os parâmetros auto-regresivos, $\theta_1 \dots \theta_q$ são os parâmetros de médias móveis, e a_t é a série de ruídos independentes com média zero e variância σ_a^2 .

O conceito de Box & Jenkins baseia-se em duas ideias básicas: o princípio da parcimônia (deve-se escolher um modelo com o menor número de parâmetros possíveis para uma representação matemática apropriada) e a construção de modelos através de um ciclo iterativo.

Na Figura 2 é possível observar os estágios da modelagem Box & Jenkins. Inicialmente, nas etapas de especificação e identificação, a ordem do modelo deve ser escolhida através de tentativas, aplicando as funções de autocorrelação e autocorrelação parcial amostrais. Em seguida, deve-se proceder para a estimação dos parâmetros ϕ e θ por meio do uso dos estimadores de máxima verossimilhança ou de suas aproximações. Posteriormente, devem ser realizados testes estatísticos para os resíduos com a finalidade de verificar a adequação do modelo. Nesta etapa, é verificado se as hipóteses assumidas nas etapas anteriores são atendidas. Por fim, pode-se utilizar o modelo definido para a realização da previsão ou controle.

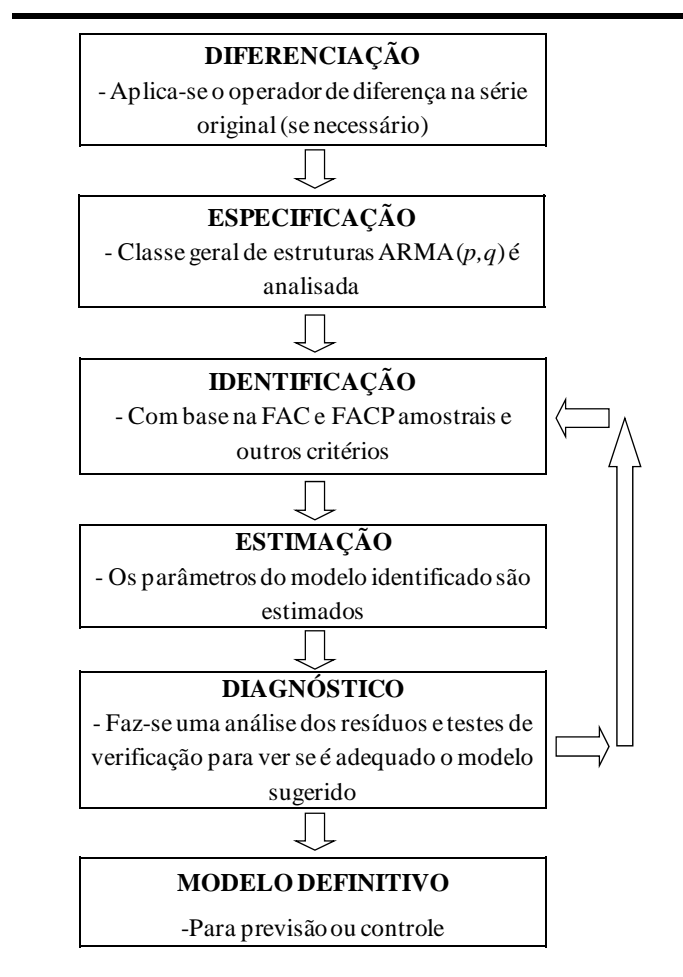


Figura 2 - Estágios da metodologia Box & Jenkins

Fonte: adaptado de Souza e Camargo (2004).

Na próxima seção, é introduzido o modelo $PAR(p)$, que consiste em uma classe de formulações auto-regressivas cujos parâmetros têm comportamento periódico e p representa um vetor $p = (p_1 \ p_2 \ \dots \ p_n)$ onde cada elemento representa a ordem de cada período.

2.2

Modelo $PAR(p)$ ⁹

Segundo Hipel e McLeod (1994), algumas séries temporais têm uma estrutura de autocorrelação que depende não somente do intervalo de tempo entre as observações, mas também do período observado. Maceira (1989) destaca que

⁹ Para mais detalhes, ver Hipel e McLeod (1994).

este é o caso, por exemplo, das séries hidrológicas mensais, que apresentam um comportamento periódico de suas propriedades probabilísticas, *e.g.* a média, a variância, a assimetria e a estrutura de correlação.

Para este tipo de séries, Kavvas e Delleur (1975) mostraram, por meio de resultados analíticos e empíricos, que diferenciações sazonais ou não-sazonais na série, embora muito eficientes na remoção de periodicidades, distorcem o espectro original, tornando o ajuste pelo modelo ARMA impraticável ou impossível. Além disso, de acordo com McKerchar e Delleur (1974), a capacidade de previsão do modelo sazonalmente diferenciado pode ser debilitado pelo fato de não ser levado em consideração a variação sazonal nos desvios padrão sazonal. Em adição, a diferenciação não-sazonal não preserva a estrutura sazonal na previsão.

Séries temporais com tais características podem ser analisadas por meio de formulações auto-regressivas cujos parâmetros têm comportamento periódico. Estes modelos são chamados de auto-regressivos periódicos e são referenciados por PAR(p), sendo, para o caso mensal, p um vetor $p = (p_1 \ p_2 \ \dots \ p_{12})$, em que cada elemento representa a ordem de cada período.

O modelo PAR (p) pode ser representado da seguinte maneira:

$$\left(\frac{Z_t - \mu_m}{\sigma_m} \right) = \phi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \phi_2^m \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) + \dots + \phi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + a_t^{(m)} \quad (3)$$

Em que: Z_t é a série sazonal de período $s = 12$ (no caso mensal); t é o índice de tempo, $t = 1, 2, \dots, sN$, função do ano T ($T = 1, 2, \dots, N$) e do período m ($m = 1, 2, \dots, s$); μ_m é a média sazonal de período m ; σ_m é o desvio-padrão sazonal de período m ; p_m é a ordem do operador auto-regressivo do período m . Neste caso, a ordem muda de acordo com o período, ao contrário do AR(p) e; $a_t^{(m)}$ é a série de ruídos independentes com média zero e variância $\sigma_a^{2(m)}$. A estratégia analítica desta metodologia é o ajuste de um modelo auto-regressivo de ordem p para cada um dos períodos das séries em análise, no caso de periodicidade mensal, um modelo para cada mês.

A primeira etapa da modelagem Box & Jenkins consiste na identificação dos modelos PAR(p), em que são identificadas as ordens p_m mais apropriadas

dos operadores autoregressivos de cada período. Esta etapa é realizada por meio da análise das funções de autocorrelação (FAC) e autocorrelação parcial (FACP). Sendo a correlação entre Z_t e Z_{t-k} representada por $\rho^m(k)$, de tal forma que t corresponde ao período m :

$$\rho^m(k) = E \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right], k = 0, 1, 2, 3, \dots \quad (4)$$

Multiplicado ambos os lados de (3) por $\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right)$ e tomando o valor esperado, a estrutura de dependência temporal da série pode ser descrita pelo conjunto de funções de autocorrelação $\rho^m(k)$ ¹⁰ dos períodos $m=1, 2, \dots, s$:

$$\begin{aligned} \rho^m(k) &= E \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] = \phi_1^m E \left[\left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] + \dots + \\ &+ \phi_{p_m}^m E \left[\left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] + E \left[a_t^{(m)} \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \end{aligned} \quad (5)$$

Fixando-se m e variando k de 1 a p^m em (5) obtêm-se para cada período um conjunto de equações periódicas de Yule-Walker. Para um período m qualquer:

$$\begin{bmatrix} 1 & \rho^{m-1}(1) & \rho^{m-1}(2) & \dots & \rho^{m-1}(p_m-1) \\ \rho^{m-1}(1) & 1 & \rho^{m-2}(1) & \dots & \rho^{m-2}(p_m-2) \\ \rho^{m-1}(2) & \rho^{m-2}(1) & 1 & \dots & \rho^{m-3}(p_m-3) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho^{m-1}(p_m-1) & \rho^{m-2}(p_m-2) & \rho^{m-3}(p_m-3) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1^m \\ \phi_2^m \\ \phi_3^m \\ \vdots \\ \phi_{p_m}^m \end{bmatrix} = \begin{bmatrix} \rho^m(1) \\ \rho^m(2) \\ \rho^m(3) \\ \vdots \\ \rho^m(p_m) \end{bmatrix} \quad (6)$$

¹⁰ O estimador $\hat{\rho}^m(k)$ é dado pela seguinte equação: $\hat{\rho}^m(k) = \frac{\hat{\gamma}^m(k)}{\hat{\sigma}_m \hat{\sigma}_{m-k}}$, com

$$\hat{\gamma}^m(k) = \frac{1}{N} \sum_{i=1}^N \left(z_{(i-1)12+m} - \hat{\mu}_m \right) \left(z_{(i-1)12+m-k} - \hat{\mu}_{m-k} \right), \hat{\sigma}_m = \sqrt{\left(\frac{1}{N} \sum_{i=1}^N \left(z_{(i-1)12+m} - \hat{\mu}_m \right)^2 \right)}, \text{ e}$$

$$\hat{\mu}_m = \frac{1}{N} \sum_{i=1}^N z_{(i-1)12+m}, \text{ para } m = 1, 2, \dots, 12.$$

Considerando ϕ_{kj} o j -ésimo parâmetro auto-regressivo de um processo de ordem k , ϕ_{kk} é o último parâmetro deste processo e, assim, as equações periódicas de Yule-Walker podem ser reescritas da seguinte maneira:

$$\begin{bmatrix} 1 & \rho^{m-1}(1) & \rho^{m-1}(2) & \dots & \rho^{m-1}(k-1) \\ \rho^{m-1}(1) & 1 & \rho^{m-2}(1) & \dots & \rho^{m-2}(k-2) \\ \rho^{m-1}(2) & \rho^{m-2}(1) & 1 & \dots & \rho^{m-3}(k-3) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho^{m-1}(k-1) & \rho^{m-2}(k-2) & \rho^{m-3}(k-3) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1}^m \\ \phi_{k2}^m \\ \phi_{k3}^m \\ \vdots \\ \phi_{kk}^m \end{bmatrix} = \begin{bmatrix} \rho^m(1) \\ \rho^m(k2) \\ \rho^m(k3) \\ \vdots \\ \rho^m(kk) \end{bmatrix} \quad (7)$$

O conjunto de valores ϕ_{kk}^m , $m = 1, 2, \dots, s$ é denominado autocorrelação parcial do período m . Cada coeficiente de autocorrelação parcial de ordem k coincide com o último parâmetro de um modelo auto-regressivo de mesma ordem. Assim, em um processo auto-regressivo de ordem p_m , a função de autocorrelação parcial ϕ_{kk}^m é diferente de zero para k menor ou igual a p_m e zero para k maior que p_m .

Deste modo, a identificação do modelo PAR(p) fundamenta-se em determinar as ordens mais apropriadas dos operadores auto-regressivos de cada período p_m , $m = 1, 2, \dots, s$, devendo ser feito por meio da obtenção das estimativas $\hat{\phi}_{kk}^m$, $k = 1, \dots, N/4$ e substituição das autocorrelações pelos respectivos valores amostrais em (7). Se a ordem do operador auto-regressivo em um período qualquer m for p_m , então $\hat{\phi}_{kk}^m$ para $k > p_m$ terá distribuição aproximadamente normal com média zero e variância N^{-1} , ou seja, aproximação de Quenouille (MACEIRA *et al*, 2006).

É importante observar que existe uma discussão na literatura com relação à melhor maneira de identificação da ordem dos parâmetros auto-regressivos. Em um primeiro critério (utilizado pelo setor elétrico brasileiro), deve-se procurar a maior ordem i para cada período m de tal forma que todas as estimativas de $\hat{\phi}_{kk}^m$ para $k > i$ não sejam mais significativas (CEPEL, 2000, MACEIRA *et al*, 2006).

Porém, Stedinger (2001) critica este procedimento de identificação tradicional, uma vez que as defasagens intermediárias não-significativas são consideradas, fato que pode resultar em uma estrutura não parcimoniosa. O autor

sugere que a identificação deve ser realizada com base no critério de que para cada período m , a maior ordem i é procurada, de forma que todas as estimativas de $\hat{\phi}_{kk}^m$ para $k < i$ sejam significativas (neste trabalho, denominado critério dois).

Um terceiro critério para a identificação, abordado por Oliveira (2010), foi a utilização de técnicas de reamostragem *Bootstrap* para definir a ordem do modelo. Os resultados obtidos mostraram que a técnica proposta obteve resultados semelhantes aos obtidos pelo critério dois e propiciou bons resultados quanto à geração de cenários.

Para modelos auto-regressivos, segundo Hipel e McLeod (1994), as estimativas obtidas pelo método dos momentos são em geral tão eficientes quanto às obtidas pela máxima verossimilhança¹¹. Dessa forma, os parâmetros ϕ_i^m podem ser obtidos por meio da substituição em (6), de $\rho^{m-j}(k)$, $j = 0, \dots, (p_m - 1)$, $k = 1, \dots, p_m$ pelas suas estimativas amostrais.

É possível notar que podem ser efetuadas as estimações dos parâmetros do modelo para o m -ésimo período independentemente dos parâmetros de qualquer outro período e, cada um dos m sistemas resultantes pode ser solucionado por Decomposição de Cholesky. Por fim, as estimativas de $\sigma_a^{2(m)}$ podem ser obtidas a partir da substituição dos ρ_i^m e dos ϕ_i^m por seus estimadores na seguinte expressão:

$$\sigma_a^{2(m)} = 1 - \phi_1^m \rho_1^m - \phi_2^m \rho_2^m - \dots - \phi_{p_m}^m \rho_{p_m}^m \quad (8)$$

A etapa de verificação do modelo consiste em testar a adequação dos modelos, verificando, se as hipóteses assumidas durante as etapas anteriores são atendidas. Nesta etapa, são utilizados critérios e testes estatísticos para verificar a hipótese nula dos resíduos estimados serem ruído branco. Em outras palavras, não rejeitar a hipótese nula dos testes, equivale a dizer que o modelo foi capaz de explicar satisfatoriamente o comportamento da série de forma que os erros não apresentem nenhuma estrutura de correlação. Este fato consiste em um indicador de eficiência explicativa do modelo.

¹¹ Maiores detalhes a respeito do método de momentos e do método de máxima verossimilhança podem ser observados em Casella e Berger (2010).

2.2.1

Geração de séries sintéticas no setor elétrico¹²

A utilização de critérios probabilísticos nas diversas atividades do planejamento da operação gerou a necessidade do uso da modelagem probabilística de aflúências aos aproveitamentos hidrelétricos ou a subsistemas. A simulação da operação energética do sistema para diversos cenários de aflúências gera os índices de risco para o suprimento de energia, *e.g.* pode-se observar a demanda de energia e o risco do não atendimento adequado em um determinado período.

Porém, o que existe na prática é um único cenário, que é o registro de aflúências observado no passado (série histórica), o que se configura insuficiente para estimar índices com incertezas aceitáveis. A série histórica segue um processo estocástico, sendo apenas uma de suas realizações. Assim, estima-se um modelo com base nesta série e, por meio deste modelo, tenta-se obter novas realizações do processo estocástico que o gerou. Estas realizações são as séries sintéticas de aflúências que são diferentes do histórico, porém são estatisticamente equiprováveis. Desta forma, é possível extrair toda a informação contida na série histórica, possibilitando a avaliação de riscos e incertezas referentes ao sistema hidrelétrico.

Assim, nesta fase do planejamento da operação, primeiramente é realizada a modelagem probabilística de ENAs por meio dos modelos $PAR(p)$, com o objetivo de se obterem as características básicas da série histórica para, em seguida, produzir séries sintéticas de aflúências diferentes da série histórica, mas igualmente prováveis.

2.2.2

Geração de séries sintéticas com o modelo $PAR(p)$ ¹³

Nesta seção são apresentados os conceitos, as formulações matemáticas e estatísticas utilizadas para a geração de cenários hidrológicos (séries hidrológicas sintéticas) no contexto do setor elétrico brasileiro.

¹² Baseado em Maceira *et al* (2006).

¹³ Baseado em CEPEL (2000).

Como já destacado, o valor observado da série histórica no instante t pode ser interpretado como um valor amostrado da distribuição de probabilidades associada à variável aleatória do processo estocástico em t . Como não são disponíveis outras séries temporais geradas por este processo estocástico, o objetivo é ajustar o modelo $PAR(p)$ que tenha sido o gerador da série histórica e, a partir dele, poder gerar séries sintéticas que representem as séries temporais possíveis de serem amostradas pelo processo. O modelo PAR ajustado deve permitir então, que se façam tantos sorteios quantos forem necessários para o problema em questão. Desta forma, cada sorteio estará associado a uma série temporal, denominada série sintética.

Manipulando a equação do modelo (3) para isolar o valor das vazões e, de posse de (8), chega-se à seguinte forma:

$$Z_t = \mu_m + \phi_1^m \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \phi_2^m \sigma_m \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) + \dots + \phi_{p_m}^m \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m \sigma_a^{(m)} E_t \quad (9)$$

Ou:

$$Z_t = \mu_m + \phi_1^m \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \phi_2^m \sigma_m \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) + \dots + \phi_{p_m}^m \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m \sqrt{1 - \phi_1^m \rho_1^m - \phi_2^m \rho_2^m - \dots - \phi_{p_m}^m \rho_{p_m}^m} E_t \quad (10)$$

Onde: E_t é independente com média zero e variância unitária. Esta expressão é conveniente para a geração de séries sintéticas, pois permite um esquema de geração sequencial no qual a cada instante t , Z_t é obtido como uma função de valores $Z_{t-\tau}$, $\tau = 1, \dots, p_m$ e de um sorteio E_t . É preciso arbitrar alguns valores iniciais $Z_{t-\tau}$ para iniciar o processo de geração. Estimativas para esses valores iniciais podem ser as médias mensais ou os últimos valores amostrais. Os primeiros valores gerados devem ser desprezados para não condicionar a série sintética aos valores iniciais.

Para obter-se um Z_t positivo é necessário que:

$$Z_t = \mu_m + \phi_1^m \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \phi_2^m \sigma_m \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) + \dots + \phi_{p_m}^m \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m a_t > 0 \quad (11)$$

Colocando a expressão (11) em função de a_t , tem-se:

$$a_t > - \left(\frac{\mu_m}{\sigma_m} \right) - \phi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) - \phi_2^m \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) - \dots - \phi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \quad (12)$$

Chamando o lado direito da equação (12) de Δ , tem-se:

$$a_t > \Delta \quad (13)$$

Portanto, Δ é função apenas do primeiro e segundo momentos (média e variância) do período m e dos coeficientes auto-regressivos e é dada por:

$$\Delta = - \left(\frac{\mu_m}{\sigma_m} \right) - \phi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) - \phi_2^m \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) - \dots - \phi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \quad (14)$$

Muitos pesquisadores assumem que os resíduos a_t apresentam distribuição Normal e, uma possível não-normalidade pode ser corrigida pela transformação Box-Cox. O modelo de geração de séries sintéticas deve ser aplicado diretamente à série temporal original e deve ser capaz de lidar com resíduos que apresentam um forte coeficiente de assimetria. Para isso foi adotado o ajuste de uma distribuição Lognormal com três parâmetros aos resíduos mensais a_t (Maceira e Bezerra, 1997). Assim, a variável ξ_t segue uma distribuição Normal com média zero e variância igual a $\sigma_{\xi_t}^{2(m)}$:

$$\xi_t = \ln(a_t - \Delta) \quad (15)$$

Os parâmetros Δ e $\sigma_{\xi_t}^{2(m)}$ são estimados de forma a preservar os momentos dos resíduos, por meio das seguintes relações (CHABERNEAU, 1978):

$$\sigma_{\xi}^{(m)} = \sqrt{\ln(\theta)} \quad (16)$$

$$\Delta = \frac{-\sigma_a^{(m)}}{\sqrt{(\theta-1)}} \quad (17)$$

Onde:

$$\theta = 1 + \frac{\sigma_a^{2(m)}}{(-\Delta)^2} \quad (18)$$

$\sigma_{\xi}^{2(m)}$ pode ser obtido a partir de (16) e (18).

Com a finalidade de gerar vazões mensais multivariadas, os resíduos assumidos como espacialmente não correlacionados, ξ_t , podem ser transformados em resíduos espacialmente correlacionados, W_t , da seguinte maneira:

$$W_t = D\xi_t \quad (19)$$

Onde D é uma matriz quadrada de dimensão igual ao número de subsistemas equivalentes e, sua estimação pode ser feita por:

$$DD^T = \hat{U} \quad (20)$$

Segundo Penna (2009), visando preservar dependências espaciais entre os subsistemas, faz-se \hat{U} sendo a estimativa da matriz de covariâncias $\left[E \left(\begin{smallmatrix} \hat{\xi}_t \\ \hat{\xi}_t^T \end{smallmatrix} \right) \right]$ entre as ENAs dos subsistemas, *e.g.* o elemento u_{12} é a covariância espacial de *lag* zero entre os resíduos dos subsistemas 1 e 2. A matriz D pode ser estimada

adotando-se D como uma matriz triangular inferior ou por decomposição espectral (fatoração de Cholesky).

Utilizando-se destes cenários gerados, o objetivo é calcular estratégias ótimas de operação para diversos cenários hidrológicos.

2.3

Utilização de wavelets

A utilização das wavelets no contexto da combinação de métodos proposta ocorre da seguinte forma: utiliza-se o encolhimento wavelet para remover ruídos existentes nas séries originais para, em seguida, à série transformada, aplicar a modelagem $PAR(p)$ conforme já explicitado. Com isso, espera-se obter uma modelagem mais adequada para explicar o comportamento da série em análise, já que o encolhimento wavelet permite amenizar as influências desproporcionais, de um ou mais pares de observações contaminados por ruídos, na estimativa da FAC e da FACP. As influências espúrias podem levar a valores espúrios e à consequente identificação errônea da verdadeira ordem do processo gerador dos dados. Consequentemente, a utilização das wavelets possibilita também o melhor ajuste dos modelos e a obtenção de ganhos preditivos.

2.4

Teoria wavelet

A teoria wavelet é a parte da metodologia proposta utilizada para realizar o pré-processamento dos dados. Inicialmente, as funções wavelets são empregadas para transformar a informação de uma série em coeficientes que podem ser agregados em subséries de aproximação e detalhes. A eliminação de coeficientes contaminados de ruído ocorre nos coeficientes de detalhes (maiores frequências), a partir de determinado valor (limiar) e da utilização de uma regra de limiarização, tornando, assim, nulos os coeficientes abaixo de um determinado valor. Os coeficientes menos ruidosos obtidos nesta fase devem ser somados de forma a se chegar à série menos ruidosa.

Para o entendimento do encolhimento de ruídos wavelets, inicialmente, esta seção realiza uma discussão entre domínio do tempo e domínio da frequência;

em seguida, a análise de Fourier *short time* é brevemente exposta; e por fim, é apresentada a metodologia wavelet, destacando o encolhimento wavelet e seus vários limiares e regras de limiarização.

2.4.1

Domínio do tempo x domínio da frequência

De acordo com Chatfield (1996), as ferramentas naturais de serem utilizadas para a análise de um processo através do tempo – chamada de análise no domínio do tempo – são as funções de autocovariância e de autocorrelação. A função de autocovariância entre Z_t e Z_{t-k} , representada por $\gamma(k)$, é definida da seguinte maneira:

$$\gamma(k) = E[(Z_t - \mu)(Z_{t-k} - \mu)], \quad k = 0, 1, 2, 3, \dots \quad (21)$$

A função de autocorrelação (expressa inicialmente na equação (4) para a identificação do modelo PAR(p)) pode ser caracterizada, de outra forma, pela divisão da autocovariância para o período k pela função de autocovariância do período 0:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \quad k = 0, 1, 2, 3, \dots \quad (22)$$

Por outro lado, a ferramenta natural para considerar as propriedades de frequência de uma série temporal é a chamada função densidade espectral. Qualquer tipo de análise ou inferência relacionado a esta função são chamadas de análise no domínio da frequência (CHATFIELD, 1996).

A função densidade espectral tem como característica a representação de informações das influências de frequências sobre o processo. Normalmente, a análise espectral é realizada quando o interesse principal é a busca por periodicidades ocultas na série. A partir de um número suficiente de frequências diferentes com um número suficiente de amplitudes e fases, pode-se modelar qualquer processo estacionário, pela seguinte relação:

$$Z_t = \sum_{j=1}^K R_j \cos(\omega_j t + \theta_j) + a_t \quad (23)$$

Onde ω_j é a frequência j , R_j e θ_j são, respectivamente, a amplitude e a fase relacionadas a esta frequência e a_t é um termo de erro aleatório.

Como $\cos(\omega_j t + \theta_j) = \cos(\omega_j t) \cos(\theta_j) - \sin(\omega_j t) \sin(\theta_j)$, tem-se:

$$Z_t = \sum_{j=1}^K (c_j \cos \omega_j t + b_j \sin \omega_j t) + a_t \quad (24)$$

Em que $c_j = R_j \cos \theta_j$ e $b_j = -R_j \sin \theta_j$. Desta forma, as séries podem ser representadas em termos de contribuições que ocorrem em diferentes escalas temporais, ou frequências características, representadas por funções seno e cosseno.

Supondo Z_t , um processo estocástico estacionário com média zero e função de autocovariância que satisfaça uma condição de independência assintótica, onde valores bastante separados no tempo sejam pouco dependentes, isto é, considerando que as autocovariâncias sejam absolutamente somáveis $\left(\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty \right)$. Nestas condições, a função densidade espectral é definida como a transformada de Fourier de $\gamma(k)$ (Morettin e Toloi, 2006), como segue:

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-i\omega k}, \quad -\infty \leq \omega \leq \infty \quad (25)$$

Onde a equação (25) é uma representação da equação (24) por meio de números complexos.

Assim, uma função de autocovariância e sua respectiva transformada de Fourier são duas formas de representar uma mesma informação. A função de autocovariância mostra a evolução temporal de uma série e esconde a informação sobre as frequências, já a transformada de Fourier dá informações sobre frequências e esconde informações temporais.

A Transformada de Fourier leva à escolha entre tempo e frequência. Pode-se então determinar as frequências presentes em um determinado sinal, porém não se pode ter a ideia de quando elas ocorrem. Essa limitação pode ser superada por meio da utilização da técnica Short-Time Fourier Transform (STFT). Ela permite analisar um sinal por meio de uma “janela” curta, que delimita uma parte do sinal aproximadamente estacionária.

$$STFT(\tau, \omega) = \int_{-\infty}^{+\infty} \gamma(k) g(k - \tau) e^{-i\omega k} d\omega \quad (26)$$

Em que: $\gamma(k)$ está previamente janelado por $g(k)$ em torno do tempo τ .

No entanto, por meio da Figura 3, é possível observar um problema envolvido no uso desta técnica: a janela definida é constante para todo o período de análise. Assim, quanto menor a janela, melhor pode-se observar as componentes de alta frequência do sinal, porém não se consegue captar as componentes de baixa frequência e, quanto maior a janela, melhor pode-se enquadrar as baixas frequências, todavia observa-se menos das altas frequências.

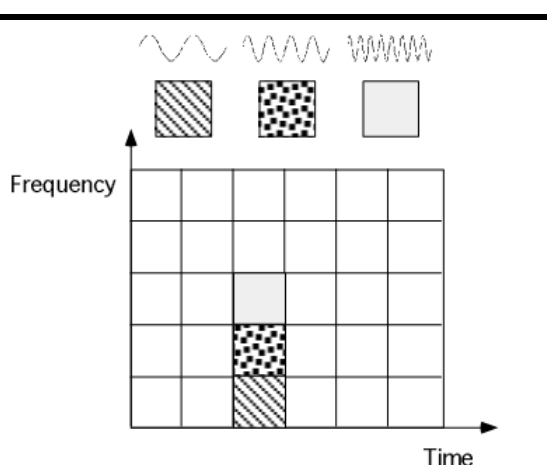


Figura 3 - Resolução no plano tempo-frequência pela STFT

Fonte: Graps (1995).

2.4.2

Wavelets

Wavelets são funções matemáticas que separam dados em suas diferentes componentes frequenciais e extraem cada componente de acordo com sua resolução de escala adequada. Com isto, o objetivo é transformar a informação de um sinal em coeficientes que possam ser manipulados, analisados ou até mesmo utilizados para reconstruir o sinal original. As wavelets possuem vantagens em relação à análise de Fourier já que esta última analisa o sinal como um todo, fato que leva a uma representação mais pobre para sinais que possam conter descontinuidades e variações bruscas (OLIVEIRA, 2007).

O uso de dilatações e translações pode ser considerado como o princípio mais geral na construção das wavelets. Por meio das dilatações, elas podem distinguir as características locais de um sinal em diferentes escalas e, a partir das translações, elas podem cobrir toda a região na qual o sinal é estudado. Além disso, a propriedade de localidade das wavelets leva às suas vantagens sobre a transformada de Fourier na análise de sinais não-estacionários (LIMA, 2002).

Por meio da Figura 4 é possível observar que, na análise realizada com wavelets, tem-se uma baixa resolução na frequência e alta no tempo para frequências mais baixas e uma resolução alta na frequência e baixa no tempo para frequências mais altas. Isto se deve ao fato de sinais com componentes de baixa frequência ter alterações mais lentas no domínio temporal e sinais com componentes em alta frequência ter rápidas alterações no domínio temporal.

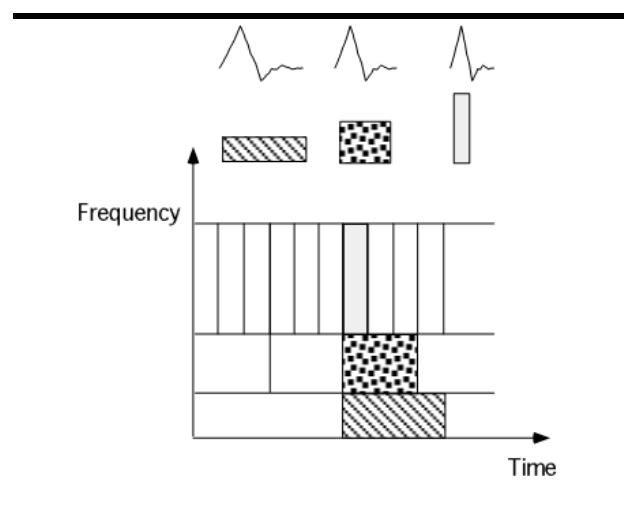


Figura 4 - Resolução no plano tempo-frequência pela análise wavelet

Fonte: Graps (1995).

Assim, foi desenvolvida a transformada wavelet como uma alternativa à STFT e, em seu procedimento de análise, deve-se adotar uma função protótipo wavelet, denominada wavelet analisadora ou wavelet-mãe. Esta análise deve ser feita por meio da multiplicação da série por esta função (wavelet). Todas as funções utilizadas para realizar esta transformação correspondem a versões comprimidas ou expandidas de uma mesma wavelet-mãe (OLIVEIRA, 2007).

No contexto de séries temporais, a Teoria Wavelet consiste em um procedimento auxiliar de pré-processamento da série em análise. Este pré-processamento pode ser realizado, geralmente, de duas maneiras: por meio da decomposição ou pelo encolhimento de ruídos da série temporal a ser modelada.

Nas próximas seções deste capítulo, primeiramente são formalizadas a função wavelet e as transformadas wavelet contínua e discreta para, em seguida, serem apresentados a decomposição wavelet e o encolhimento de ruídos wavelet, ambos, os procedimentos já estão sendo aplicados em séries temporais.

2.4.2.1

Função wavelet

Segundo Kubrusly (2001), os espaços de Hilbert são definidos como espaços lineares completos munidos com um produto interno $\langle \cdot ; \cdot \rangle$. Um espaço de

Hilbert H é separável se e somente se inclui uma base (ortonormal ou ortogonal) contável.

Assim, de acordo com Kubrusly (2001), o espaço de funções quadraticamente integráveis sobre o espaço dos reais, denotado por $L^2(R)$, munido de um produto interno $\langle \cdot ; \cdot \rangle$, é um exemplo de espaços de Hilbert separáveis. De tal modo, o espaço de Hilbert $L^2(R)$ são todas as funções definidas em R de energia finita, ou seja:

$$L^2(R) = \left\{ f : X \rightarrow R : \int_R |f(\cdot)|^2 \partial(x) < \infty \right\} \quad (27)$$

Uma wavelet é uma função $\psi(x) \in L^2(R)$ que gera uma base ortonormal $\{\psi_{j,k}(x)\}_{j,k \in \mathbb{Z}}$ para o espaço de Hilbert $L^2(R)$ ¹⁴. As funções wavelets devem ser de energia finita, *i.e.* $\int_{-\infty}^{+\infty} \psi^2(t) dt < \infty$ e o seu valor médio no domínio temporal deve ser nulo, *i.e.* $\int_{-\infty}^{+\infty} \psi(t) dt = 0$.

2.4.2.2

Transformadas wavelet contínua e discreta¹⁵

A transformada wavelet contínua de uma série $Z(t)$ é o produto interno entre esta série e várias versões transladadas e escalonadas da wavelet-mãe sobre todo o tempo. Formalmente tem-se que:

$$TWC(j, k) = \frac{1}{\sqrt{|j|}} \int_{-\infty}^{+\infty} Z(t) \psi\left(\frac{t-k}{j}\right) dt, \text{ onde } j \neq 0 \text{ e } k \in R \quad (28)$$

¹⁴ Uma coleção $\{\psi_j(x)\}_{j \in \mathbb{Z}}$ é uma base ortonormal do espaço de Hilbert $L^2(R)$, se atende às seguintes propriedades: 1) Ortogonalidade: $\langle \psi_j(x), \psi_l(x) \rangle = 0, \forall j \neq l$; 2) Normalidade: $\|\psi_j(x)\| = 1, \forall j \in \mathbb{Z}$ e; 3) Completeza: Se $f(x) \in L^2(R)$ e $\langle f(x), \psi_j(x) \rangle = 0$, então: $f(x) = 0$.

¹⁵ Mais detalhes em Daubechies (1992), Graps (1995), Mallat (1999), Lima (2002) e Oliveira (2007).

Nesta equação (28), uma wavelet-mãe ψ é utilizada para criar uma família de wavelets $\psi(\text{escala}, \text{translação}, t) = \psi_{j,k}(t) = \frac{1}{\sqrt{|j|}} \psi\left(\frac{t-k}{j}\right)$ onde é possível observar a propriedade de escala das wavelets por meio de $\psi_j(t) = \frac{1}{\sqrt{|j|}} \psi\left(\frac{t}{j}\right)$ e a propriedade de deslocamento por meio de $\psi_k(t) = \psi(t-k)$.

No entanto, a utilização da transformada contínua não seria viável, pois teriam que ser realizadas infinitas translações e escalonamentos. Dessa forma, foram introduzidas as wavelets discretas, que não são transladas nem escalonadas continuamente, e sim em intervalos discretos.

De acordo com Levan e Kubrusly (2003), um elemento $\psi(t) \in L^2(R)$ com o usual produto interno $\langle \cdot ; \cdot \rangle$ e norma $\| \cdot \|$ é chamado wavelet ortonormal se a sequência de funções representada na equação (29) consiste em uma base ortonormal para $L^2(R)$,

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \quad (29)$$

Neste caso, j e k são inteiros arbitrários. O fator de escala j faz a dilatação e o índice de localização k define a posição da wavelet-mãe ψ para gerar wavelets, e.g. wavelets da família Daubechies (Daubechies, 1988). Se ψ é uma wavelet, então $\psi_{j,k}$ também o será para qualquer $j, k \in \mathbb{Z}$.

Em outras palavras, uma $L^2(R)$ wavelet é uma função $\psi(t) \in L^2(R)$ que gera uma base ortonormal wavelet duplamente indexada $\{\psi_{j,k}(t)\}_{(j,k) \in \mathbb{Z}}$ para o espaço de funções $L^2(R)$. Assim, de acordo com o Teorema da Série de Fourier, cada função $Z(t)$ possui uma expansão (única) por série de Fourier em H , dado por:

$$Z(t) = \sum_{j,k \in \mathbb{Z} \times \mathbb{Z}} \langle Z(t), \psi_{j,k}(t) \rangle \psi_{j,k}(t) \quad (30)$$

Segundo Kubrusly (2001), em relação a uma base ortonormal com indexação dupla de um espaço de Hilbert H , a expansão por série de Fourier de um vetor $x \in H$ possui a propriedade do Teorema de Fubini, em que a ordem dos índices do somatório se torna indiferente (tal resultado se deve à convergência incondicional da série de Fourier). Isto é:

$$Z(t) = \sum_{j,k \in \mathbb{Z} \times \mathbb{Z}} \langle Z(t), \psi_{j,k}(t) \rangle \psi_{j,k}(t) = \sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \langle Z(t), \psi_{j,k}(t) \rangle \psi_{j,k}(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \langle Z(t), \psi_{j,k}(t) \rangle \psi_{j,k}(t) \quad (31)$$

É possível reescrever a expansão por série de Fourier de uma função $Z(t)$ em H , em termos de funções wavelet e escala^{16,17}:

$$Z(t) = \sum_{k \in \mathbb{Z}} a_{j_0,k} \phi_{j_0,k} + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k} \quad (32)$$

As transformadas wavelets discretas são definidas como o produto interno da função $Z(t) \in L^2(R)$ por alguma wavelet $\psi_{j,k}$ ou escala $\phi_{j_0,k}$. Estas transformadas wavelets são chamadas coeficientes de detalhes, $d_{j,k}$, e de aproximação, $a_{j_0,k}$, e, formalmente, podem ser representadas da seguinte maneira:

$$d_{j,k} = TWD_{\psi}(j, k) = \langle Z(t), \psi_{j,k}(t) \rangle = \int_R Z(t) \psi_{j,k}(t) d(t), \forall (j, k) \in \mathbb{Z} \times \mathbb{Z} \quad (33)$$

$$a_{j_0,k} = TWD_{\phi}(j_0, k) = \langle Z(t), \phi_{j_0,k}(t) \rangle = \int_R Z(t) \phi_{j_0,k}(t) d(t), \forall (k) \in \mathbb{Z} \quad (34)$$

As famílias $\{d_{j,k}\}_{\substack{(j,k) \in \mathbb{Z} \times \mathbb{Z} \\ j \geq j_0}}$ e $\{a_{j_0,k}\}_{k \in \mathbb{Z}}$ são chamadas, de forma alternativa, coeficientes wavelets (ou coeficientes de Fourier) e mensuram a associação linear

¹⁶ Um caso particular de função wavelet é a chamada função escala, que pode ser representada da seguinte maneira: $\phi_{j_0,k}(t) = 2^{j_0/2} \phi(2^{j_0} t - k)$, $j_0, k \in \mathbb{R}$. Para um j_0 fixo, são feitas translações em k , gerando uma base ortonormal para $L^2(R)$.

¹⁷ Alguns exemplos de funções escala e wavelets podem ser vistas no Anexo 1.

entre alguma função $Z(t)$ e alguma função wavelet ($\psi_{j,k}$ ou $\phi_{j_0,k}$) para cada $(j,k) \in Z \times Z$ e $j_0 \in Z$ (fixo).

2.4.2.3

Decomposição wavelet

A sequência infinita de subespaços $V_j(\phi)$, $j \in Z$ de $L^2(R)$ é definida como Análise Multirresolução Wavelet (porque usa base ortonormal wavelet), com relação à função escala $\phi(t) \in L^2(R)$, se as seguintes condições são mantidas (LEVAN e KUBRUSLY, 2003):

- i. $V_j(\phi) \subset V_{j+1}(\phi)$, $\forall j \in Z$;
- ii. $\bigcap_{j \in Z} V_j(\phi) = \{0\}$;
- iii. $\bigcup_{j \in Z} V_j(\phi) = L^2(R)$;
- iv. $v(t) \in V_m(\phi) \Leftrightarrow v[2(t)] \in V_{m+1}(\phi)$, para algum $m \in Z$ e;
- v. $\{\phi[(t)-n], n \in Z\}$ é uma base ortonormal do subespaço $V_0(\phi) \subset L^2(R)$.

Um espaço de aproximação $V_j(\phi)$ pode ser expandido em termos de subespaços de detalhes $W_k(\psi)$, em que $-\infty \leq s \leq j-1$ (LEVAN e KUBRUSLY, 2003):

$$V_j(\phi) = \bigoplus_{s=-\infty}^{j-1} W_s(\psi) \quad (35)$$

Conforme Kubrusly e Levan (2002) mostram, considerando o resultado da equação (35), é possível decompor o espaço $L^2(R)$ em termos dos subespaços de detalhes $W_s(\psi)$ para os níveis no intervalo $j_0 \leq s \leq +\infty$ e de aproximação $V_{j_0}(\phi)$ para o nível j_0 como mostram as expressões a seguir:

$$L^2(R) = \left[\bigoplus_{s=-\infty}^{j_0-1} W_s(\psi) \right] \oplus \left[\bigoplus_{s=j_0}^{+\infty} W_s(\psi) \right] \quad (36)$$

$$L^2(R) = V_{j_0}(\phi) \oplus \left[\bigoplus_{s=j_0}^{+\infty} W_s(\psi) \right] \quad (37)$$

Os subespaços de aproximação e detalhes de $L^2(R)$ são disjuntos e ortogonais entre si (Levan e Kubrusly, 2003). Assim, cada função $Z(t)$ em $L^2(R)$ possui uma única representação sob forma de decomposição ortogonal, em termos de projeções ortogonais, dada por:

$$Z(t) = Z_{V_{j_0}(\phi)}(t) + Z_{V_{j_0}(\psi)}(t) + Z_{V_{j_0+1}(\psi)}(t) + \dots + Z_{V_{+\infty}(\psi)}(t) \quad (38)$$

De posse da decomposição ortogonal descrita em (38) e da família ortonormal wavelet $\left\{ \phi_{j_0,k}(t), \psi_{j,k}(t) \right\}_{\substack{(j,k) \in \mathbb{Z} \times \mathbb{Z} \\ j \geq j_0}}$, de base de espaço $L^2(R)$, cada projeção ortogonal em (38) tem uma única representação sob a forma de série de Fourier com base wavelet ortonormal (Kubrusly, 2001). Com isso, tem-se:

$$Z(t) = \sum_{k \in \mathbb{Z}} a_{j_0,k} \phi_{j_0,k}(t) + \sum_{k \in \mathbb{Z}} d_{j_0,k} \psi_{j_0,k}(t) + \sum_{k \in \mathbb{Z}} d_{j_0+1,k} \psi_{j_0+1,k}(t) + \dots + \sum_{k \in \mathbb{Z}} d_{+\infty,k} \psi_{+\infty,k}(t) \quad (39)$$

$$Z(t) = \sum_{k \in \mathbb{Z}} a_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{+\infty} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(t) \quad (40)$$

Em que: $\{a_{j_0,k}\}_{k \in \mathbb{Z}}$ são os coeficientes de aproximação e $\{d_{j,k}\}_{\substack{(j,k) \in \mathbb{Z} \times \mathbb{Z} \\ j \geq j_0}}$ são

os coeficientes de detalhes.

Uma decomposição wavelet de 2 camadas, por exemplo, pode ser visualizada na Figura 5, em que a série (S) é decomposta, no nível 1, em coeficientes de aproximação (A_1) e de detalhe (D_1), e no nível 2, os coeficientes de aproximação são decompostos em coeficientes de aproximação de nível 2 (A_2) e coeficientes de detalhe de nível 2 (D_2) resultando nos componentes A_2 , D_2 e D_1 .

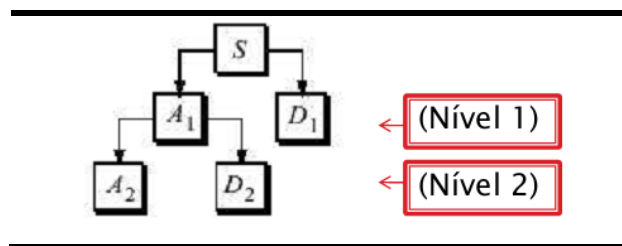


Figura 5 - Exemplo de uma decomposição wavelet de nível 2

Fonte: Elaboração própria.

2.4.2.4

Encolhimento wavelet

O encolhimento de sinais wavelet não deve ser visto como suavização (embora seja tratado algumas vezes na literatura como sinônimos). Enquanto a suavização remove altas frequências e mantém as baixas, o encolhimento objetiva remover os ruídos existentes e preservar qualquer sinal presente, independente de seu conteúdo (TASWELL, 2000).

Por meio do modelo (41) pode-se entender a ideia do encolhimento wavelet. Este pode ser visto como um modelo de regressão no tempo e o método de eliminação de ruído pode ser considerado uma estimação não paramétrica de uma função, usando bases ortogonais. O alvo da regressão não paramétrica é estimar a função desconhecida $f = (f_1, f_2, \dots, f_n)$, com o menor erro quadrático médio. Então, o objetivo é estimar uma função $\hat{f} = \hat{f}(y_1, y_2, \dots, y_n)$ a partir dos dados observados (MORETTIN, 1997):

$$y_t = f_t + \varepsilon_t, \quad t = 1, \dots, n \quad (41)$$

Onde: y_t é o sinal original, f_t é um sinal determinístico de interesse e ε_t é um choque estocástico $\varepsilon_t \sim N(0, \sigma^2)$.

De acordo com Taswell (2000), este procedimento não necessita de quaisquer tipos de suposições sobre a natureza do sinal e permite descontinuidades no mesmo. O encolhimento wavelet trabalha com a ideia de que os ruídos estão presentes nos detalhes, assim, devem ser considerados no processo de encolhimento apenas os coeficientes de detalhes.

O encolhimento consiste na multiplicação de cada membro de $\{d_{j,k}\}_{(j,k) \in \{j\}_{j_0}^{+\infty} \times Z}$ por um número real pertencente ao intervalo $[0,1]$, respeitando a regra de limiarização e o respectivo limiar. Dessa forma, pode ser gerada uma nova família de coeficientes de detalhes, denotada por $\{\tilde{d}_{j,k}\}_{(j,k) \in \{j\}_{j_0}^{+\infty} \times Z}$. A expansão em (40) pode ser reescrita da seguinte maneira:

$$\tilde{Z}(t) = \sum_{k \in Z} a_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{+\infty} \sum_{k \in Z} \tilde{d}_{j,k} \psi_{j,k}(t) \quad (42)$$

Em que: $\tilde{Z}(t)$ é a nova série temporal (filtrada), $\{a_{j_0,k}\}_{k \in Z}$ são os coeficientes de aproximação, $\phi_{j_0,k}$ é a função escala utilizada, $\{\tilde{d}_{j,k}\}_{\substack{(j,k) \in Z \times Z \\ j \geq j_0}}$ são os coeficientes de detalhes menos ruidosos e $\psi_{j,k}$ é a função wavelet.

O procedimento de encolhimento do ruído consiste nos três estágios seguintes: 1) escolher a base wavelet e o nível de decomposição a serem utilizados e realizar a transformada wavelet discreta da série original para a obtenção dos coeficientes wavelets, que são contaminados pelo ruído; 2) a eliminação dos coeficientes contaminados de ruído que se encontram em cada nível de resolução e são os de frequências mais altas (chamados coeficientes de detalhes), é feita a partir de certo valor (limiar) e a utilização de uma regra de limiarização, tornando, assim, nulos os coeficientes abaixo de um determinado valor. Várias escolhas são possíveis. Nesta fase, são obtidos os coeficientes menos ruidosos; e 3) realizar a transformada wavelet inversa dos coeficientes do estágio anterior para se chegar à série menos ruidosa (MORETTIN, 1997).

Como existem vários tipos de regras de limiarização e limiares propostos na literatura, as próximas seções têm como objetivo introduzi-los. É importante ressaltar que só são apresentados aqueles utilizados nesta tese¹⁸.

¹⁸ Maiores detalhes sobre regras de limiarização e limiares, ver Donoho e Johnstone (1994, 1998), Donoho *et al* (1993), Morettin (1997) e Gutierrez (2002).

2.4.2.4.1

Regras de limiarização

A eliminação dos coeficientes contaminados de ruído, resultantes da transformação wavelets da série original, é realizada por meio de uma regra de limiarização. Como resultados da utilização desta regra, os coeficientes abaixo de um determinado valor (η) se tornam nulos e os demais assumem certo valor a depender da regra utilizada. Várias são as regras de limiarização destacadas na literatura. Nesta tese são utilizadas as regras do limiar duro (*hard threshold*) e do limiar suave (*soft threshold*). Gutierrez (2002) destaca que estas regras são amplamente utilizadas devido a sua simplicidade e desempenho.

Na regra do limiar duro, cada elemento da família $\{d_{j,k}\}_{(j,k) \in \{j\}_{j_0}^{+\infty} \times \mathbb{Z}}$ de coeficientes de detalhes é preservado ou cancelado como a seguir:

$$\tilde{d}_{j,k} = \begin{cases} d_{j,k}, & \text{se } |d_{j,k}| > \eta \\ 0, & \text{se } |d_{j,k}| \leq \eta \end{cases} \quad (43)$$

Neste caso: $d_{j,k}$ é o coeficiente de detalhe com ruído, $\tilde{d}_{j,k}$ é o coeficiente sem o ruído e η é o limiar. É possível observar que apenas os coeficientes de detalhes que são, em valores absolutos, menores ou iguais ao limiar (que é necessariamente um valor real positivo) são cancelados, enquanto os outros coeficientes de detalhes permanecem inalterados.

Já pela regra do limiar suave, os elementos da família $\{d_{j,k}\}_{(j,k) \in \{j\}_{j_0}^{+\infty} \times \mathbb{Z}}$ de coeficientes de detalhes podem ser atenuados ou cancelados, de acordo com a seguinte regra:

$$\tilde{d}_{j,k} = \begin{cases} \text{sign}(d_{j,k})(|d_{j,k}| - \eta), & \text{se } |d_{j,k}| > \eta \\ 0, & \text{se } |d_{j,k}| \leq \eta \end{cases} \quad (44)$$

Onde: $d_{j,k}$ é o coeficiente de detalhe com ruído, $\tilde{d}_{j,k}$ é o coeficiente sem o ruído e η é o limiar. É possível observar que os coeficientes de detalhes que são, em valores absolutos, menores ou iguais ao limiar são cancelados, enquanto os outros coeficientes de detalhes são atenuados por $\text{sign}(d_{j,k})(|d_{j,k}| - \eta)$. Algumas formas de se determinar o limiar η são apresentadas na próxima seção.

2.4.2.4.2

Limiares

Como já ressaltado, existem vários tipos de cálculos propostos na literatura para os parâmetros de limiarização η . São apresentados apenas aqueles que são utilizados na tese: o universal, o minimax e o estimador não viesado do risco de Stein (*Estimation Stein's Unbiased Risk Estimate* – SURE¹⁹).

O limiar universal não depende da escala j , isso significa que para todo $j \in Z$, o limiar universal η assumirá sempre o mesmo valor (é um limiar global). Assim, de acordo com Donoho e Johnstone (1998), dado uma série $Z(t)$, o cálculo deste limiar é dado por:

$$\eta = \eta_j = \hat{\sigma}_Z (2 \log T)^{\frac{1}{2}} \quad (45)$$

Onde, $\hat{\sigma}_Z$ é a variância da série $Z(t)$ e T é o tamanho da série. Segundo Morettin (1997), este limiar leva a estimativas que subestimam a série original, já que há a tendência de eliminar muitos coeficientes, especialmente em frequências muito elevadas.

Outro limiar global (que é utilizado para a análise em todos os níveis) é o Minimax, proposto por Dohono e Johnstone (1994). Este limiar, diferentemente do universal, leva a bons resultados quando há saltos abruptos e picos agudos na série original. Este parâmetro é obtido pela minimização de um limite superior teórico do risco assintótico. O objetivo é estimar uma série aproximada com o

¹⁹ Para maiores detalhes sobre o SURE e como ele é calculado, ver Donoho e Johnstone (1993), Johnstone e Silverman (1996) e Morettin (1997).

menor erro médio quadrático. Para que uma função f (ver equação 41) atinja o risco Minimax, tem-se que:

$$R(F) = \inf_{\hat{f}} \sup_f R(\hat{f}, f) \quad (46)$$

Onde: $R(\hat{f}, f) = \left(\frac{1}{T}\right) E \|\hat{f} - f\|^2$ e $\|\hat{f} - f\|^2 = \sum_{t=0}^{T-1} f_t^2$. f pertence a algum espaço funcional específico F . O valor limiar Minimax pode ser aproximado numericamente.

Um limiar não global é o SURE, que depende do nível j , isto é, para cada escala há um limiar η_j . Donoho e Johnstone (1993) propuseram uma técnica de seleccionar um limiar, pela minimização do SURE, em cada nível de resolução. Se no nível j houver η_j coeficientes, a definição do limiar é feita da seguinte forma:

$$\eta_j = \arg \min_{0 \leq \eta \leq \sqrt{2 \log(N_j)}} SURE(y_j, \eta) \quad (47)$$

Neste capítulo foram apresentados: o modelo $PAR(p)$ e como são realizados seus procedimentos de identificação, estimação e verificação; e a teoria wavelet, com ênfase no encolhimento wavelet e em seus respectivos limiares e regras de limiarização. No próximo capítulo será apresentada a metodologia proposta, baseado nas metodologias deste capítulo; e serão descritas as séries utilizadas para realizar as aplicações e como será feita a avaliação e a comparação da modelagem proposta com outros modelos.

3

Metodologia proposta

Para alcançar os objetivos descritos anteriormente, o presente trabalho utiliza uma metodologia que consiste em uma combinação de encolhimento wavelet com o modelo $PAR(p)$. Assim, neste capítulo, é apresentada a referida metodologia e é dada uma explicação de como são realizadas a aplicação e a comparação da mesma.

As etapas da proposta inicial são mostradas na Figura 6. O primeiro passo consiste em submeter uma série temporal ao encolhimento wavelet. Dada a possibilidade de se realizarem diversos encolhimentos wavelets por meio da variação do número de camadas de decomposição; da família wavelet; do limiar; e da regra de limiarização, obtêm-se várias séries filtradas e suas correspondentes séries de resíduos. Como o objetivo desta etapa é o de remover apenas ruídos, os resíduos correspondentes às séries encolhidas devem ser verificados por meio de testes estatísticos.

Em seguida, deve-se utilizar uma das séries menos ruidosas (cujos valores extraídos atenderam aos testes de resíduos) para ser ajustada por meio de um modelo $PAR(p)$, conforme exposto na seção 2.2, para que posteriormente, possa ser realizada uma previsão ou uma geração de cenários.

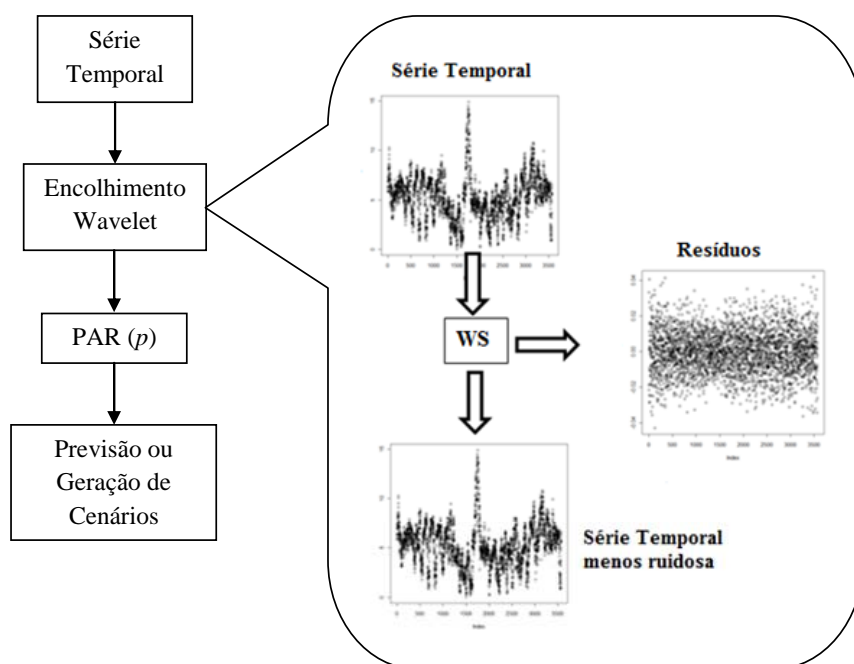


Figura 6 - Proposta inicial

Fonte: Elaboração própria.

Os passos da metodologia proposta estão dispostos neste capítulo, na seguinte ordem: 1) inicialmente, é destacado o encolhimento wavelet; 2) em seguida, são apresentados os testes para os resíduos gerados pelo encolhimento wavelet; 3) a modelagem PAR é destacada; e 4) por fim, explicita-se como é realizada a comparação desta abordagem com o modelo PAR(p) clássico.

3.1

Encolhimento wavelet

O primeiro estágio desta modelagem consiste no tratamento dos dados. Ou seja, realiza-se o encolhimento de ruídos por meio das wavelets, conforme já destacado na seção 2.4. Na parte destacada da Figura 6, referente à etapa “Encolhimento Wavelet”, é possível observar que à série temporal original, é aplicado o encolhimento wavelet representado por “WS” e, como resultados, obtêm-se uma série filtrada (ou menos ruidosa) e uma série de resíduos.

Diversos tipos de encolhimento baseados em diferentes *setups* podem ser realizados para gerar várias séries encolhidas e suas respectivas séries candidatas a

serem ruídos. Somente uma série encolhida cujos resíduos atendem aos testes de ruídos pode ser utilizada na modelagem $PAR(p)$.

A construção dos *setups* mencionados pode ser observada na Tabela 1. Ela consiste na combinação entre o número de camadas de decomposição wavelet (que, neste caso, pode ser de 2 a 4), a família wavelet utilizada (Haar, Daubechies, Coiflet ou Symlet), o tipo de limiar (duro ou suave) e uma das regras de limiarização (universal, minimax ou SURE). Um *setup*, por exemplo, pode ser obtido por uma decomposição de quatro camadas de uma família symlet⁷, e utilizando do limiar suave e da regra de limiarização minimax para o encolhimento wavelet.

Tabela 1 - Características utilizadas para a elaboração dos setups

Camadas	Família	Limiar	Regra de Limiarização
2	haar	duro	universal
3	daubechies 2	suave	minimax
4	daubechies 3		SURE
	⋮		
	daubechies 45		
	coiflet 1		
	coiflet 2		
	⋮		
	coiflet 5		
	symlet 1		
	symlet 2		
	⋮		
	symlet 8		

Fonte: Elaboração própria.

3.2

Testes de resíduos²⁰

Ao realizar os diversos encolhimentos wavelets segundo *setups* diferentes, obtêm-se várias séries filtradas e suas correspondentes séries de resíduos. Como o objetivo é o de remover apenas ruídos, às séries candidatas, devem ser aplicados

²⁰ Para maiores detalhes sobre testes de resíduos, ver Fernandes (2010).

testes estatísticos para verificar se as mesmas são ruídos. Desta maneira, garante-se que nenhum tipo de informação importante para descrever o processo gerador das séries seja perdido.

Com este objetivo, são realizados três testes: 1) o teste de Ljung-Box para testar a hipótese de erros descorrelatados; 2) o teste Jarque-Bera para verificar a normalidade dos erros; e 3) o teste ARCH para averiguar se variância do erro não segue um processo tipo ARCH/GARCH. Estes testes são expostos a seguir e, em seguida, é realizada uma avaliação dos mesmos quando aplicados em séries que seguem um processo $PAR(p)$.

3.2.1

Ljung-Box

O teste Ljung-Box verifica se as primeiras m autocorrelações de uma série temporal são estatisticamente significativas, isto é, investiga se algum subconjunto das autocorrelações é não nulo. Nesta tese, o objetivo é verificar as séries de ruídos obtidas por meio dos vários *setups* de encolhimento wavelets. As hipóteses nula e alternativa para este teste são:

$$\begin{aligned} H_0 : \rho(1) = \rho(2) = \dots = \rho(m) = 0 \\ H_1 : \text{pelo menos um dos } \rho' s \neq 0 \end{aligned} \quad (48)$$

Onde H_0 é a hipótese nula e H_1 é a hipótese alternativa. Neste caso, $\rho(k)$ é a autocorrelação de ordem k . Para testar a hipótese nula, a estatística de teste é calculada da seguinte forma:

$$LB(m) = T(T+2) \sum_{k=1}^m \hat{\rho}^2(k) / (T-k) \sim \chi^2(v) \quad (49)$$

Em que T é o tamanho da série e m é o número de *lags* sendo testados. Rejeitar a hipótese nula implica na existência de autocorrelação nos resíduos, ou seja, a série testada não é composta apenas por ruídos.

3.2.2

Jarque-Bera

No teste Jarque-Bera, a hipótese de normalidade do termo aleatório é investigada por meio de dois coeficientes associados a distribuições de probabilidade: os coeficientes de assimetria e de curtose. Neste caso, é investigado se a série de resíduos é oriunda de uma variável aleatória com distribuição normal, sendo que sob normalidade, a assimetria é igual a zero e a curtose é igual a três. As hipóteses nula e alternativa para este teste são:

$$\begin{aligned} H_0 : S = 0, K = 3 \\ H_1 : \text{caso contrário} \end{aligned} \quad (50)$$

Rejeitar a hipótese nula significa que os resíduos não seguem distribuição normal. Sob hipótese nula, H_0 , os resíduos seguem uma distribuição normal; a estatística Jarque-Bera possui distribuição qui-quadrado com dois graus de liberdade e a estatística de teste pode ser descrita da seguinte maneira:

$$JB = (T/6)(\hat{S} - 0)^2 + (T/24)(\hat{K} - 3)^2 \sim \chi^2(2) \quad (51)$$

Onde T é o número de observações e \hat{S} e \hat{K} são, respectivamente, os coeficientes de assimetria e de curtose da série.

3.2.3

Teste ARCH

O teste ARCH verifica se a variância do erro segue um processo tipo ARCH/GARCH (heterocedasticidade condicional), como pode ser visto no exemplo da equação (52) de um processo GARCH(1,1).

$$\begin{aligned} E(\varepsilon_t^2 | Y_{t-1}) = \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \text{ (GARCH(1,1))} \\ \alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1 \end{aligned} \quad (52)$$

Neste caso, é testado se há a presença de mudanças na variância condicional dos resíduos, isto é, se a magnitude dos resíduos está relacionada à magnitude de resíduos passados. As hipóteses nula e alternativa para este teste são:

$$\begin{aligned} H_0 &: \text{não existe efeito ARCH até a ordem } p \\ H_1 &: \text{existe efeito ARCH} \end{aligned} \quad (53)$$

Para testar se a série de ruídos resultante do encolhimento wavelet possui efeito ARCH(q), deve-se estimar:

$$\hat{\varepsilon}_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \hat{\varepsilon}_{t-i}^2 + \nu_t \quad (54)$$

Em que $\hat{\varepsilon}_t$ são os resíduos do encolhimento wavelet. Sob a hipótese nula de que não existe efeito ARCH até ordem q nos resíduos, tem-se que:

$$TR^2 \sim \chi^2(q) \quad (55)$$

T é o número de observações e R^2 é o coeficiente de determinação da regressão dos resíduos.

3.3

Aplicação dos testes e modificação da proposta inicial para metodologia proposta

Inicialmente, a aplicação dos testes de resíduos foi realizada, seguindo o método proposto, às séries de resíduos resultantes da aplicação de diversos encolhimentos wavelets com diferentes *setups* a dois tipos de séries: 1) séries temporais mensais simuladas com 1.200 observações²¹, realizações de diferentes processos PAR(p), e 2) séries de ENAs utilizadas para o planejamento do setor elétrico brasileiro.

²¹ Maiores detalhes sobre as séries simuladas são destacados na seção 3.5.

Os resultados indicaram que nenhuma das séries de resíduos atendeu aos testes citados. Devido à grande quantidade de *setups* propostos, somente alguns resultados²² são dispostos nos Anexos 2 e 3, onde é possível observar que as séries testadas sempre rejeitaram pelo menos uma das hipóteses nulas dos testes descritos.

Assim sendo, optou-se por realizar o encolhimento wavelet de outra forma. Como pode ser observado na Figura 7, na metodologia proposta, o encolhimento é então aplicado às séries temporais de valores referentes a cada mês. Dessa maneira, podem-se encontrar diferentes *setups* para os diferentes meses. Isto caracteriza a mudança da metodologia inicial (Figura 6) para a metodologia proposta (Figura 7).

Novamente, depois de realizado o encolhimento, as séries de ruídos são testadas e as séries filtradas referentes aos ruídos aprovados de cada mês são reorganizadas em uma única série e se tornam candidatas a serem utilizadas no modelo $PAR(p)$.

²² Todos os demais resultados encontram-se à disposição com o autor.

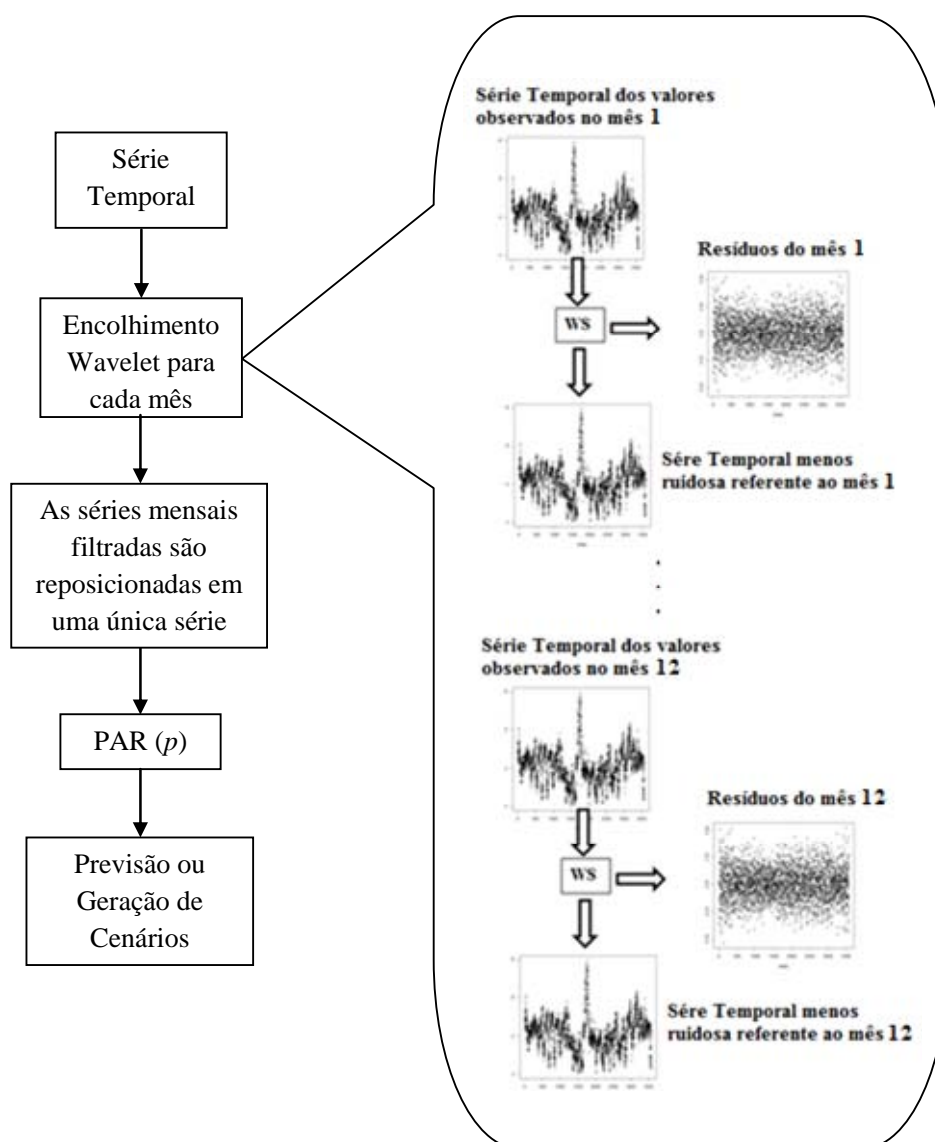


Figura 7 - Metodologia proposta

Fonte: Elaboração própria.

Com isso, o objetivo é que este método possa ser aplicado em séries temporais que tenham comportamento periódico das suas propriedades probabilísticas, como por exemplo, a média e a variância, obtendo valores mais apurados para estes momentos amostrais citados, além de melhorar a identificação e a estimação dos parâmetros auto-regressivos do modelo, já que muitos resíduos são removidos antes destas etapas da modelagem.

Conforme destacado, deve ser utilizada uma série filtrada referente a cada uma das 12 séries de meses para a modelagem $PAR(p)$. Assim, o modelo $PAR(p)$ transformado pode ser representado da seguinte maneira:

$$\left(\frac{\tilde{Z}_t - \tilde{\mu}_m}{\tilde{\sigma}_m} \right) = \varphi_1^m \left(\frac{\tilde{Z}_{t-1} - \tilde{\mu}_{m-1}}{\tilde{\sigma}_{m-1}} \right) + \varphi_2^m \left(\frac{\tilde{Z}_{t-2} - \tilde{\mu}_{m-2}}{\tilde{\sigma}_{m-2}} \right) + \dots + \varphi_{p_m}^m \left(\frac{\tilde{Z}_{t-p_m} - \tilde{\mu}_{m-p_m}}{\tilde{\sigma}_{m-p_m}} \right) + a_t^{(m)} \quad (56)$$

Em seguida, os procedimentos para identificação, estimação e verificação do modelo descrito na equação (56) devem ser realizados assim como o descrito na seção 2.2. É importante observar que a escolha dos melhores setups é realizada por meio da verificação das funções de autocorrelação dos resíduos dos modelos e por meio da comparação dos valores de erro percentual médio absoluto (Mean Absolute Percentual Error – MAPE) atribuído a cada um dos modelos resultantes dos diferentes setups. Isto é, dado que as FACs dos resíduos de determinados modelos apresentam valores estatisticamente nulos, observam-se quais os menores MAPEs para determinar as melhores séries filtradas a serem utilizadas para cada mês.

3.4

Comparação das metodologias

Esta tese compara o método proposto (Encolhimento Wavelet + PAR(p)) à modelagem PAR(p). Adicionalmente, a utilização de médias móveis simples (com janelas de três e de cinco observações) para o encolhimento de ruídos e o posterior ajuste pelo PAR(p) também foi realizada apenas para termos de comparação.

Inicialmente, são analisadas as funções de autocorrelação dos resíduos dos modelos. Rejeitar a hipótese nula de que os resíduos de um modelo ajustado não são correlacionados significa constatar que este não é adequado para descrever a série temporal e deve ser modificado²³. Em seguida, para avaliar o ajuste dos modelos, são avaliados os critérios de erro percentual médio absoluto (MAPE) e de raiz do erro médio quadrático (*root mean square error* – RMSE) dos mesmos no período *in-sample*, isto é, na parte da amostra utilizada para estimar os parâmetros.

Como a aplicação da metodologia acontece em dois tipos de séries – simulada e real – algumas medidas são levadas em consideração especificamente

²³ Em todas as utilizações de médias móveis, a hipótese nula de resíduos descorrelacionados sempre foi rejeitada. Por isso, estes resultados não são retratados nesta tese.

para a comparação dos modelos ajustados às séries simuladas e outras, especificamente, para a mesma comparação só que em modelos ajustados às séries reais.

Nos casos simulados (seção 3.5), também é realizada a comparação entre as médias, os desvios-padrão, os parâmetros auto-regressivos estimados pelos modelos e os valores reais utilizados para a simulação. Além disso, é feita uma avaliação da capacidade preditiva dos modelos no período *out-of-sample*, isto é, na parte da amostra utilizada para validar os mesmos.

No caso das séries reais (seção 3.6) são gerados cenários de acordo com a seção 3.6.3 e, para examinar os modelos, são realizados os testes de aderência destacados na seção 3.6.4.

3.5

Aplicação a casos simulados

Inicialmente, para a verificação da metodologia proposta, é realizada a simulação de algumas séries temporais mensais com 1.200 realizações de um processo PAR(p). Este processo possui um parâmetro auto-regressivo para cada mês e foi obtido da seguinte maneira:

$$Z_t = \mu_m + \sigma_m \phi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \sigma_m a_t^{(m)} \quad (57)$$

Em que: Z_t é a série sazonal de período $s = 12$; t é o índice de tempo, $t = 1, 2, \dots, 1.200$, função do ano T ($T = 1, 2, \dots, 100$) e do período m ($m = 1, 2, \dots, 12$); μ_m é a média sazonal de período m ; σ_m é o desvio-padrão sazonal de período m ; ϕ_1^m é o parâmetro auto-regressivo de período m ; e $a_t^{(m)}$ é a série de erros independentes para cada m , com média zero e variância igual a um.

Em cada um dos casos foram simulados 12.012 valores, sendo que os primeiros 10.800 foram retirados, para não condicionar a série sintética aos valores iniciais, os 1.200 seguintes foram mantidos e utilizados para a modelagem. Os últimos 12 valores foram utilizados para o período de teste *out-of-*

sample. Os valores das médias, dos desvios-padrão e dos parâmetros auto-regressivos foram arbitrados, e a primeira observação do mês de dezembro (a ser utilizada para a obtenção do primeiro valor do mês de janeiro) é a própria média arbitrada para o mesmo.

Assim, para a obtenção de Z_t , sorteia-se um valor para $a_t^{(m)}$, que segue uma distribuição normal (0,1), utilizam-se os valores arbitrados para μ_m , σ_m e ϕ_1^m , e o valor Z_{t-1} obtido no mês anterior. Os valores de μ_m , σ_m e ϕ_1^m são apresentados nas Tabelas 4, 6, 7 e 8 juntamente com seus respectivos valores estimados.

3.6

Aplicação às séries de ENAs e avaliação de desempenho do modelo²⁴

Como já destacado, os cenários de ENA são gerados pelo modelo PAR(p) e são utilizados pelo Newave no planejamento de médio prazo do setor elétrico. Segundo Marcato (2002), a execução destes programas ocorre em conjunto. Desta forma, nesta tese, não é possível utilizar o conjunto de cenários gerados no Newave²⁵.

Os cenários hidrológicos sintéticos construídos com base no método proposto devem ser capazes de reproduzir as propriedades estatísticas da série histórica original. Segundo CEPEL (2000), a utilidade de um modelo pode ser medida por sua capacidade de reproduzir distribuições de probabilidade de

²⁴ Para mais informações, ver CEPEL (2000), Maceira *et al* (2006), Penna (2009), Campos (2010), Oliveira (2010) e Ferreira (2013).

²⁵ Ferreira (2013) se refere ao modelo PAR(p) utilizado nesta tese como PAR(p) *mainstream*. Em seu trabalho, o autor cita a possibilidade da entrada de novas usinas em operação com o decorrer dos anos, aumentando a capacidade de geração e produtividade. Neste caso, a cada mês em que novas unidades geradoras hidrelétricas se tornam operacionais, caracteriza-se uma nova configuração hidrelétrica do sistema, *i.e.*, a série histórica de ENA sofre um incremento para todo o período de análise. Desta forma, é caracterizada, no contexto do NEWAVE, a configuração, que é uma série histórica de energia correspondente a um dado período de tempo fixo. Para que o Setor Elétrico Brasileiro seja capaz de incorporar os efeitos de futuras usinas que entrarão em operação no horizonte de estudo, o PAR(p) deve ser estimado com base nas diferentes configurações de ENAs simultaneamente. Em outras palavras, deve haver um “cruzamento” entre as configurações do parque gerador, caracterizando o PAR(p) interconfigurações, onde se ajusta um modelo auto-regressivo de ordem p para cada um dos meses e cada uma das configurações das séries hidrológicas históricas. No caso do PAR(p) interconfigurações, devem ser estimados tantos modelos quanto o número de meses em que ocorre a mudança de configuração.

variáveis aleatórias relevantes ao processo. Neste trabalho, a representação de períodos críticos, as correlações seriais, entre outros critérios, devem ser levados em conta para esta avaliação. Assim, com a finalidade de avaliar as características das séries geradas pelo método proposto, devem ser aplicados alguns testes estatísticos neste conjunto de séries sintéticas.

3.6.1

Testes de média

Uma das premissas fundamentais para avaliar a adequação do modelo é a preservação das médias históricas nos cenários. Segundo Casella e Berger (2010), suponha duas populações independentes e que seguem distribuição Normal (no caso de não serem Normais, se aplicam as condições do Teorema Central do Limite). A primeira tem média μ_1 e variância σ_1^2 e a segunda tem média μ_2 e variância σ_2^2 . O Teste t é utilizado para testar a igualdade das médias de duas amostras diferentes de tamanhos n_1 e n_2 , respectivamente e admite as seguintes hipóteses:

$$\begin{aligned} H_0 : \mu_1 &= \mu_2 \\ H_1 : \mu_1 &\neq \mu_2 \end{aligned} \quad (58)$$

Onde H_0 é a hipótese nula e H_1 é a hipótese alternativa. Para testar $H_0 : \mu_1 = \mu_2$ (as médias históricas são estatisticamente iguais às médias dos cenários) a estatística de teste é calculada da seguinte forma:

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (59)$$

Onde \bar{X}_1 , \bar{X}_2 , s_1^2 e s_2^2 são as médias e variâncias amostrais, respectivamente, das duas amostras testadas.

Se $H_0 : \mu_1 = \mu_2$ for verdadeira, t_0 é distribuída como $t_{n_1+n_2-2}$. Assim, se $|t_0| > t_{\alpha/2, n_1+n_2-2}$ (em que α é o nível de significância), a hipótese nula deve ser rejeitada.

3.6.2

Testes de variância

O procedimento para testar a igualdade de variâncias entre a série histórica e os cenários sintéticos é o teste de Levene. Este procedimento é utilizado, pois ele não requer o pressuposto de normalidade das amostras (ALMEIDA e ELLIAN, 2008).

Suponha $k \geq 2$ amostras aleatórias independentes entre si. A amostra i representa n_i variáveis aleatórias independentes e identicamente distribuídas com distribuição G_i de média μ_i e variância σ_i^2 , com G_i , μ_i e σ_i^2 desconhecidos. A hipótese de igualdade de variâncias é testada contra a hipótese alternativa de que pelo menos uma das variâncias não é igual às demais:

$$\begin{aligned} H_0 : \sigma_1^2 &= \dots = \sigma_k^2 \\ H_1 : \sigma_i^2 &\neq \sigma_j^2, \text{ para } i = 1, \dots, k \text{ e } j = 1, \dots, k \text{ e } i \neq j \end{aligned} \quad (60)$$

Os desvios absolutos nas variáveis X_{ij} com relação às médias amostrais

dos grupos $\bar{X}_i = \frac{\sum_{j=1}^{n_i} X_{i,j}}{n_i}$ por $Z_{ij} = \left| X_{ij} - \bar{X}_i \right|$, $i = 1, \dots, k$, $j = 1, \dots, n_i$ e estatística de teste é calculada da seguinte forma:

$$W_0 = \left(\frac{n-k}{k-1} \right) \frac{\sum_{i=1}^k n_i \left(\bar{Z}_{i.} - \bar{Z}_{..} \right)^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} \left(Z_{ij} - \bar{Z}_{i.} \right)^2} \quad (61)$$

$$\bar{Z}_{i\cdot} = \frac{\sum_{j=1}^{n_i} Z_{ij}}{n_i}, \quad \bar{Z}_{\cdot\cdot} = \frac{\sum_{i=1}^k n_i \cdot \bar{Z}_{i\cdot}}{n} \quad \text{e} \quad n = \sum_{i=1}^k n_i.$$

Onde

rejeita a hipótese nula se $W_0 > F_{(k-1, n-k), (1-\alpha)}$. Em que $F_{(k-1, n-k), (1-\alpha)}$ representa o quantil de ordem $1-\alpha$ da distribuição $F_{(k-1, n-k)}$ e α é o nível de significância do teste.

3.6.3

Testes de aderência

Os testes de aderência²⁶ são testes não-paramétricos que têm como objetivo verificar a forma de determinada distribuição de probabilidade. A ideia destes testes é determinar se certa distribuição postulada é razoável em uma amostra, *i.e.*, diz a respeito ao grau de concordância entre a distribuição da amostra e da população da qual esta foi extraída.

Um destes testes é o teste de Kolmogorov-Smirnov. Ele consiste em especificar a distribuição de frequência acumulada que ocorreria segundo uma distribuição teórica e compará-la com a distribuição de frequência acumulada observada. A teórica representa o que seria esperado sob H_0 . É importante observar que este teste é aplicado a variáveis aleatórias contínuas. Considerando $F_1(x)$ e $F_2(x)$ como as funções de distribuição acumuladas das variáveis aleatórias referentes às amostras 1 e 2, respectivamente, o teste de hipóteses apropriado é o seguinte:

$$\begin{aligned} H_0 : F_1(x) &= F_2(x) \\ H_1 : F_1(x) &\neq F_2(x) \end{aligned} \tag{62}$$

Outro teste de aderência a ser observado é o teste Qui-Quadrado, que pode ser empregado em variáveis aleatórias discretas para avaliar a eficiência do ajuste da distribuição, *i.e.* avaliar o quanto a frequência observada está próxima da

²⁶ Maiores detalhes podem ser encontrados em Conover (1971), em Siegel e Castellan Jr. (2006), entre outros.

frequência esperada. Neste teste, as observações são agrupadas em categorias que podem ser de duas ou mais. O teste visa estabelecer o grau de correspondência entre as variáveis observadas e esperadas em cada uma das categorias e admite as seguintes hipóteses:

$$\begin{aligned} H_0: & \text{não há diferença entre as frequências observadas e as esperadas} \\ H_1: & \text{há diferença entre as frequências observadas e as esperadas} \end{aligned} \quad (63)$$

Neste trabalho, o teste Qui-Quadrado é utilizado na variável aleatória comprimento de sequências. Esta variável será definida na próxima seção para duas categorias.

3.6.4

Análise de sequências

Segundo CEPEL (2000), as principais características da série observada devem ser preservadas no modelo de geração de séries sintéticas. Assim, pode-se medir a utilidade de um modelo pela sua capacidade em reproduzir distribuições de probabilidade de variáveis aleatórias ao processo. No caso em estudo, os períodos críticos, como as secas registradas no histórico, devem ser reproduzidos.

Assim, as variáveis aleatórias introduzidas nesta seção estão relacionadas à representação de períodos críticos. Desta forma, é utilizado o conceito de sequência negativa. Uma sequência negativa pode ser definida como um período de tempo em que as aflúências estão continuamente abaixo de valores pré-determinados, precedidas e sucedidas por valores acima destes limites. Geralmente, utilizam-se como limites pré-determinados, as médias mensais.

Para ilustrar o conceito de sequência negativa, deve ser observada a Figura 8, onde a linha contínua representa uma série de ENA e a linha pontilhada representa um limite pré-determinado. Os intervalos $[t_1, t_2]$ e $[t_3, t_4]$ correspondem às sequências negativas.

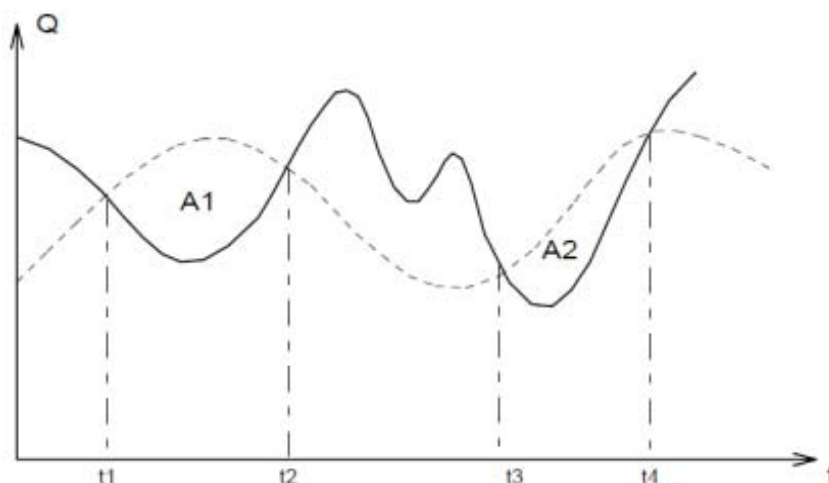


Figura 8 - Sequência Negativa

Fonte: CEPEL (2000).

A partir deste conceito, podem ser criadas três variáveis de interesse: comprimento, soma e intensidade de sequência. Estas variáveis são definidas da seguinte maneira:

Tabela 2 - Sequência negativa

Variável Aleatória	Descrição	Cálculo
Comprimento de Sequência	Corresponde ao comprimento de intervalos $(t_1 - t_2)$ e $(t_3 - t_4)$	$C = (t_2 - t_1)$
Soma de Sequência	Corresponde à área abaixo do limite durante a sequência. Na figura anterior, equivalem às áreas A1 e A2	$S = \sum_{i=1}^{t_2} (Z_i - \mu_i)$
Intensidade de Sequência	Corresponde ao valor médio abaixo do limite, isto é, à soma de sequência dividida pelo respectivo comprimento de sequência	$I = \frac{S}{C} = \frac{\sum_{i=1}^{t_2} (Z_i - \mu_i)}{(t_2 - t_1)}$

Fonte: Oliveira (2010)

Estes cálculos devem ser realizados para cada sequência negativa encontrada na série histórica, resultando assim em três amostras referentes às três variáveis aleatórias mencionadas na Tabela 2. Os mesmos cálculos devem ser realizados para as séries sintéticas, obtendo-se assim, outras três amostras das mesmas variáveis. Isso possibilita testar a hipótese das variáveis serem

provenientes das mesmas distribuições, por meio dos testes de aderência descritos no item anterior. A variável comprimento de sequência é avaliada pelo teste Qui-Quadrado e as variáveis soma e intensidade de sequência devem ser avaliadas pelo teste de Kolmogorov-Sminorv.

Porém, os índices mais relevantes para o planejamento estão associados a valores extremos das distribuições, *e.g.* o período crítico que corresponde à pior situação hidrológica em todo o histórico. Este fato é decorrente da existência de reservatórios plurianuais, pouco sensíveis a variações de curto prazo (CEPEL, 2000). Desta forma, também são estabelecidos os índices do tipo “máximo” sobre as três variáveis de sequência negativa estabelecidas.

De acordo com CEPEL (2000), já que a série histórica possui apenas um valor, faz mais sentido falar em “tipicidade” do valor histórico em relação à distribuição dos valores gerados do que falar em aderência de distribuições. Isto é, deseja-se saber qual a probabilidade de a amostra histórica ser sorteada, uma vez que o modelo de geração adotado é o verdadeiro.

O desempenho do modelo pode ser avaliado pela proporção de índices gerados maiores ou menores do que o índice histórico. Caso esta proporção seja muito pequena, isso é um indício de que a observação histórica é atípica para o modelo considerado. Para esta análise podem ser consideradas as seguintes variáveis: máximo comprimento de sequência, máxima soma de sequência e máxima intensidade de sequência.

4

Análise dos resultados

Este capítulo apresenta os resultados obtidos nesta tese. Ele está disposto da seguinte maneira: 1) inicialmente, são apresentados os resultados referentes à aplicação da metodologia proposta aos casos simulados; 2) em seguida, os resultados da aplicação às séries de ENAs; e 3) por fim, as considerações finais dos resultados encontrados são realizadas.

4.1

Resultados da aplicação às séries dos casos simulados

Para avaliar os resultados obtidos pela metodologia proposta, inicialmente, são geradas quatro séries simuladas, segundo quatro processos estocásticos diferentes, gerados a partir da seção 3.5 e dos valores destacados nas Tabelas 4, 6, 7 e 8. É realizada uma comparação entre os valores dos parâmetros reais dos processos estocásticos que geraram as séries simuladas e os parâmetros obtidos pela modelagem $PAR(p)$ clássica e a metodologia proposta wavelet shrinkage + $PAR(p)$ (WS + $PAR(p)$). Com isso, o objetivo é observar qual metodologia consegue estimar os parâmetros de forma mais precisa dentro de diferentes contextos.

Todas as séries simuladas seguem um processo $PAR(p)$ com um parâmetro auto-regressivo em cada mês. Nesta seção, o critério de identificação adotado é sempre o critério dois, citado na seção 2.2, em que a identificação deve ser realizada com base no critério de que para cada período m , a maior ordem i é procurada, de forma que todas as estimativas de $\hat{\phi}_{kk}^m$ para $k < i$ sejam significativas.

Inicialmente é apresentado o caso simulado 1, cujos parâmetros (definidos pelo autor) estão dispostos na Tabela 4 juntamente com seus respectivos parâmetros estimados. Em primeiro lugar, na Tabela 3, são apresentados os melhores *setups* encontrados para cada série referente a cada um dos meses. É possível observar o número de camadas da decomposição wavelet, a família

wavelet utilizada, o tipo de limiar e a regra de limiarização para cada um dos meses. A melhor regra de limiarização, para todos os casos, foi a suave e o limiar local SURE sempre foi utilizado, com exceção do nono mês no qual o limiar global Minimax foi utilizado.

Às series de resíduos resultantes de cada *setup*, foram aplicados os três testes mencionados na seção 3.2. O p-valor de cada um destes testes pode ser observado por meio da Tabela 3. Os resultados obtidos indicam que em todos os casos, a hipótese nula não pôde ser rejeitada a 5% de significância. As respectivas hipóteses nulas são: 1) Ljung-Box: erros descorrelatados; 2) Jarque-Bera: normalidade dos resíduos; e 3) teste ARCH: a variância dos erros não segue um processo do tipo ARCH/GARCH.

Tabela 3 - Wavelets utilizadas no encolhimento das séries de meses referentes ao caso 1 e testes de resíduos

Mês	Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
1	4	Daubechies32	suave	SURE	0.11	0.38	1.00
2	4	Daubechies42	suave	SURE	0.39	0.50	0.58
3	4	Symlet2	suave	SURE	0.08	0.08	0.29
4	3	Symlet7	suave	SURE	0.19	0.20	0.58
5	3	Symlet3	suave	SURE	0.08	0.19	0.11
6	4	Daubechies42	suave	SURE	0.08	0.50	0.38
7	2	Symlet2	suave	SURE	0.05	0.06	0.15
8	4	Daubechies34	suave	SURE	0.06	0.50	0.74
9	3	Symlet3	suave	Minimax	0.18	0.11	0.19
10	4	Daubechies37	suave	SURE	0.11	0.45	0.71
11	3	Daubechies12	suave	SURE	0.07	0.12	0.46
12	4	Daubechies45	suave	SURE	0.05	0.33	0.95

Fonte: Elaboração própria.

Os parâmetros definidos pelo autor para o caso simulado 1 estão dispostos na Tabela 4 juntamente com os respectivos parâmetros estimados pela metodologia proposta ($WS + PAR(p)$) e pelo modelo $PAR(p)$. Os valores destacados são melhores quando comparados aos valores obtidos pelo outro modelo. É possível observar que as médias estimadas pela metodologia proposta foram sempre iguais ou melhores que as estimadas pelo modelo $PAR(p)$, quando comparadas aos parâmetros reais.

Nos meses em que os desvios-padrão do modelo $PAR(p)$ ficaram muito maiores que os verdadeiros parâmetros, o modelo $WS + PAR(p)$ sempre faz uma melhor estimação, dado que o encolhimento diminui a dispersão dos dados. Já

quando ocorre o contrário (desvios-padrão menores), e metodologia proposta apresenta um desempenho menor, subestimando estes parâmetros.

Conforme já destacado, todos os meses são realizações de processos estocásticos com apenas um parâmetro auto-regressivo para cada mês. Assim, para efeitos de comparação, sempre que um modelo estimou mais de um parâmetro auto-regressivo, ele foi considerado de pior desempenho. Este foi o caso do modelo $PAR(p)$ para o mês 2. No geral, o modelo $PAR(p)$ estimou mais parâmetros auto-regressivos de forma mais precisa se comparado ao modelo $WS + PAR(p)$, porém obteve um MAPE maior em 10 dos 12 meses.

Tabela 4 - Parâmetros reais e estimados e MAPEs dos modelos (caso 1)

Mês	Média			Desvio-Padrão			Phi 1			Phi 2		MAPE	
	Real	WS+ PAR(p)	PAR(p)	Real	WS+ PAR(p)	PAR(p)	Real	WS+ PAR(p)	PAR(p)	WS+ PAR(p)	PAR(p)	WS+ PAR(p)	PAR(p)
1	5	5.09	5.09	2	2.16	2.55	0.7	0.67	0.71	0.00	0.00	23.42	25.58
2	7	7.11	7.11	3	2.84	3.07	0.3	0.37	0.52	0.00	-0.23	52.78	56.20
3	7	6.91	6.90	1.5	1.36	1.71	-0.3	-0.43	-0.42	0.00	0.00	16.57	22.02
4	8	7.98	7.98	1.8	1.71	2.06	0.6	0.42	0.51	0.00	0.00	17.94	21.47
5	5	5.04	5.04	2	2.33	2.60	0.7	0.65	0.62	0.00	0.00	32.87	30.88
6	7	6.68	6.67	3	3.15	3.38	0.3	0.45	0.46	0.00	0.00	12.80	14.12
7	7	7.17	7.17	1.5	1.06	1.45	-0.3	-0.23	-0.24	0.00	0.00	12.17	17.04
8	8	8.31	8.32	1.8	1.71	1.96	0.6	0.52	0.51	0.00	0.00	16.38	18.74
9	5	5.66	5.68	2	1.48	2.42	0.7	0.59	0.64	0.00	0.00	2.89	1.89
10	7	7.27	7.27	3	2.82	3.01	0.3	0.40	0.44	0.00	0.00	30.98	31.95
11	7	7.20	7.21	1.5	1.07	1.57	-0.3	-0.13	-0.19	0.00	0.00	12.93	18.74
12	8	8.05	8.05	1.8	2.06	2.36	0.6	0.39	0.45	0.00	0.00	22.65	25.74

Fonte: Elaboração própria.

Para a realização do caso simulado 2, foram mantidos os mesmos parâmetros auto-regressivos, porém os valores das médias e dos desvios-padrão do processo gerador dos dados foram aumentados (ver Tabela 6). Na Tabela 5 estão dispostos os *setups* utilizados para os encolhimentos wavelet e os testes aplicados aos seus respectivos ruídos, mostrando que foram extraídos somente resíduos das séries de dados. Neste caso, houve a utilização de maior variedade de famílias wavelet (Haar, Daubechies, Coiflets e Symlets) nos melhores *setups* em relação ao caso 1.

Tabela 5 - Wavelets utilizadas no encolhimento das séries de meses referentes ao caso 2 e testes de resíduos

Mês	Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
1	4	Daubechies2	suave	SURE	0.15	0.14	0.15
2	2	Daubechies25	suave	SURE	0.13	0.50	0.24
3	4	Haar	suave	SURE	0.80	0.25	0.49
4	3	Coiflet2	suave	SURE	0.25	0.20	0.17
5	4	Symlet3	suave	SURE	0.62	0.19	0.27
6	3	Daubechies33	suave	Minimax	0.62	0.43	0.83
7	3	Daubechies9	suave	SURE	0.41	0.16	0.48
8	3	Daubechies44	suave	SURE	0.17	0.11	0.38
9	4	Daubechies25	suave	SURE	0.20	0.19	0.29
10	3	Haar	suave	Minimax	0.38	0.18	0.36
11	4	Daubechies39	suave	SURE	0.18	0.29	0.24
12	3	Coiflet2	suave	Minimax	0.45	0.22	0.11

Fonte: Elaboração própria.

Os parâmetros definidos pelo autor para o caso simulado 2 e os seus respectivos parâmetros estimados estão dispostos na Tabela 6. Como no caso 1, os valores destacados são melhores em comparação aos valores estimados pelo outro modelo para os mesmos parâmetros. Novamente, comparando às médias reais, as médias estimadas pela metodologia proposta foram sempre iguais ou melhores que as estimadas pelo modelo $PAR(p)$.

O número de ocorrências em que o desvio-padrão mensal estimado pelo $PAR(p)$ é melhor do que o estimado pelo modelo $WS + PAR(p)$ aumentou de cinco no caso 1, para sete no caso 2. Porém, o número de vezes em que os parâmetros auto-regressivos foram melhores (para o $PAR(p)$) diminuiu em decorrência da estimação de parâmetros auto-regressivos de segunda ordem em quatro dos meses analisados. Isto pode ser explicado pela influência desproporcional de um ou mais pares de observações na estimativa da FACP, que geram valores espúrios e levam à identificação errônea da verdadeira ordem do processo gerador dos dados. No caso do $WS + PAR(p)$, pode-se dizer que a realização do encolhimento diminuiu a influência destes valores.

Tabela 6 - Parâmetros reais e estimados e MAPEs dos modelos (caso 2)

Mês	Média			Desvio-Padrão			Phi 1			Phi 2		MAPE	
	Real	WS+ PAR(p)	PAR(p)	Real	WS+ PAR(p)	PAR(p)	Real	WS+ PAR(p)	PAR(p)	WS+ PAR(p)	PAR(p)	WS+ PAR(p)	PAR(p)
1	8	8.10	8.11	3	3.48	3.74	0.7	0.55	0.58	0.00	0.00	26.20	28.23
2	10	10.36	10.36	3	3.00	3.20	0.3	0.33	0.34	0.00	0.00	26.11	28.12
3	8	8.35	8.35	3	2.80	3.04	-0.3	-0.11	-0.10	0.00	0.00	36.61	39.64
4	10	10.35	10.35	3	3.38	3.60	0.6	0.43	0.43	0.00	-0.20	26.98	27.61
5	8	7.63	7.62	3	3.14	3.44	0.7	0.56	0.58	0.00	0.00	33.07	36.39
6	10	9.77	9.76	3	1.91	2.93	0.3	0.24	0.15	0.00	0.24	18.85	27.70
7	8	7.94	7.94	3	2.83	3.06	-0.3	-0.29	-0.25	0.00	0.00	42.21	46.89
8	10	10.16	10.16	3	2.81	3.07	0.6	0.45	0.41	0.00	0.00	23.31	25.98
9	8	8.84	8.84	3	3.37	3.64	0.7	0.55	0.55	0.00	0.00	34.84	37.71
10	10	10.26	10.26	3	2.06	3.01	0.3	0.36	0.49	0.00	-0.25	19.73	28.20
11	8	7.84	7.83	3	2.40	2.66	-0.3	-0.24	-0.25	0.00	0.00	33.67	37.65
12	10	9.88	9.86	3	2.43	3.54	0.6	0.52	0.49	0.00	-0.24	23.05	31.36

Fonte: Elaboração própria.

Foi observado nos casos 1 e 2 que: 1) a média estimada pelo WS + PAR(p) foi sempre igual ou superior em relação ao outro modelo; e 2) em casos de desvios-padrão superestimados em relação ao parâmetro real, o método WS + PAR(p) se mostrou superior em termos de ajuste para este parâmetro. Assim, no caso simulado 3, apenas os desvios-padrão foram modificados. Anteriormente, no caso 2, os desvios-padrão arbitrados para o processo gerador dos dados foram no valor de 3 e agora, no caso 3, estes valores passaram a ser de 3,5. Os *setups* utilizados para os encolhimentos wavelet e os testes aplicados aos seus respectivos ruídos são apresentados no 4.

Na Tabela 7 é possível observar que a maioria dos parâmetros foi mais bem estimada pelo método proposto. As médias sempre foram iguais às estimadas pelo PAR(p) ou mais apuradas. 10 dos desvios-padrão mensais foram mais bem estimados pelo WS + PAR(p) e, mesmo o modelo PAR(p) não realizando nenhuma estimativa incorreta da quantidade de parâmetros auto-regressivos, o WS + PAR(p) se mostrou superior em 7 meses e igual em 2 meses. O MAPE do método proposto foi sempre maior, mostrando o melhor ajuste do mesmo.

Comparando os casos 2 e 3, foi observado que, aumentando os desvios-padrão, a tendência do modelo PAR(p) em superestimar estes parâmetros aumenta. É importante observar que isto não foi possível de ser comprovado por meio de várias simulações deste mesmo processo, pois o método proposto nesta tese não é totalmente automatizado, sendo necessária a análise humana na escolha do melhor *setup* para cada encolhimento.

Tabela 7 - Parâmetros reais e estimados e MAPEs dos modelos (caso 3)

Mês	Média			Desvio-Padrão			Phi 1			MAPE	
	Real	WS+ PAR(p)	PAR(p)	Real	WS+ PAR(p)	PAR(p)	Real	WS+ PAR(p)	PAR(p)	WS+ PAR(p)	PAR(p)
1	8	8.50	8.53	3.5	3.90	5.05	0.7	0.70	0.73	68.81	88.58
2	10	10.02	10.02	3.5	3.25	3.46	0.3	0.40	0.43	38.85	41.37
3	8	7.80	7.79	3.5	3.65	3.94	-0.3	-0.30	-0.31	24.00	27.87
4	10	10.23	10.24	3.5	4.01	4.17	0.6	0.57	0.56	30.60	32.37
5	8	7.94	7.94	3.5	3.56	3.75	0.7	0.57	0.57	45.92	48.95
6	10	10.03	10.04	3.5	3.79	3.93	0.3	0.20	0.20	56.21	58.60
7	8	7.30	7.30	3.5	4.02	4.22	-0.3	-0.31	-0.32	0.87	2.57
8	10	10.58	10.59	3.5	3.53	3.82	0.6	0.49	0.50	44.06	48.11
9	8	8.19	8.20	3.5	3.62	4.76	0.7	0.68	0.69	38.73	55.45
10	10	10.33	10.33	3.5	3.49	3.72	0.3	0.38	0.40	29.18	31.36
11	8	7.73	7.73	3.5	3.20	3.41	-0.3	-0.30	-0.31	21.29	24.28
12	10	10.32	10.33	3.5	3.45	4.47	0.6	0.40	0.43	39.25	51.49

Fonte: Elaboração própria.

No caso 4 (Tabela 8), apenas as médias do processo gerador dos dados são maiores que as do processo. Todos os demais parâmetros são iguais. Os melhores *setups* estão no Anexo 5. É possível comprovar que, nos 4 casos analisados, as médias estimadas pelo método proposto foram sempre iguais ou melhores que as médias estimadas pelo PAR(p). No caso 4, todos os desvios-padrão foram melhores quando estimados pelo WS + PAR(p), deixando claro que em situações que os desvios-padrão da série temporal são maiores que os desvios-padrão de seu processo gerador, o método proposto se mostrou melhor.

Com relação aos parâmetros auto-regressivos, foi possível observar, que no caso 4, o modelo PAR(p) se mostrou de melhor desempenho, porém, no geral, o método proposto foi capaz de reduzir a influência desproporcional de um ou mais pares de observações na estimativa da FACP, fato que leva à identificação errônea da verdadeira ordem do processo gerador dos dados.

Tabela 8 - Parâmetros reais e estimados e MAPEs dos modelos (caso 4)

Mês	Média			Desvio-Padrão			Phi 1			MAPE	
	Real	WS+ PAR(p)	PAR(p)	Real	WS+ PAR(p)	PAR(p)	Real	WS+ PAR(p)	PAR(p)	WS+ PAR(p)	PAR(p)
1	20	19.58	19.58	3.5	3.37	4.45	0.7	0.57	0.57	12.25	15.97
2	25	24.62	24.61	3.5	3.51	3.71	0.3	0.28	0.31	10.94	11.54
3	20	20.26	20.26	3.5	3.49	3.77	-0.3	-0.27	-0.29	13.96	15.10
4	25	24.78	24.77	3.5	3.48	4.49	0.6	0.44	0.48	10.07	13.35
5	20	19.55	19.53	3.5	3.10	4.30	0.7	0.63	0.66	11.01	14.13
6	25	25.42	25.42	3.5	3.52	3.68	0.3	0.41	0.39	9.51	10.01
7	20	20.00	20.00	3.5	3.48	3.67	-0.3	-0.34	-0.34	13.64	14.39
8	25	25.44	25.45	3.5	3.41	3.64	0.6	0.47	0.47	9.82	10.51
9	20	19.95	19.95	3.5	3.92	4.13	0.7	0.60	0.59	14.01	14.84
10	25	25.02	25.02	3.5	3.57	3.77	0.3	0.21	0.21	11.59	12.16
11	20	20.02	20.02	3.5	3.47	3.61	-0.3	-0.33	-0.33	14.22	14.77
12	25	25.39	25.39	3.5	3.47	3.58	0.6	0.51	0.51	9.54	9.89

Fonte: Elaboração própria.

As previsões *out-of-sample*, para 12 períodos a frente, realizadas para os quatro casos simulados estão dispostos nas Tabelas 9, 10, 11 e 12. Além dos RMSEs e dos MAPEs, são apresentados os valores de diferença relativa, *i.e.* o quanto o método proposto foi melhor (ou pior) que o modelo PAR(p) para a previsão do mês em questão.

No caso 1 (Tabela 9), é possível observar que a previsão um passo à frente é mais bem realizada pelo PAR(p) no mês 1, sendo 75,47% melhor que o método proposto. No total dos meses, o RMSE e o MAPE do PAR(p) foram melhores em apenas dois e três meses²⁷, respectivamente, mostrando o ganho do método proposto. Porém, se a magnitude das diferenças entre as estatísticas for considerada, é possível dizer que o modelo PAR(p) apresentou melhor desempenho.

²⁷ Esta diferença entre o número de meses em que o RMSE e o MAPE são melhores (ou piores) se dá devido ao arredondamento das casas decimais.

Tabela 9 - RMSE e MAPE da previsão *out-of-sample* (caso 1)

Mês	RMSE			MAPE			Diferença Relativa
	PAR(<i>p</i>) (1)	WS + PAR(<i>p</i>) (2)	Dif (1) - (2)	PAR(<i>p</i>) (3)	WS + PAR(<i>p</i>) (4)	Dif (3) - (4)	
1	0.32	0.55	-0.23	5.37%	9.42%	-4.05%	-75.47%
2	0.75	0.75	0.00	11.59%	11.61%	-0.02%	-0.13%
3	0.01	0.00	0.01	0.15%	0.07%	0.08%	55.10%
4	1.03	1.02	0.01	11.42%	11.37%	0.05%	0.43%
5	2.83	2.83	0.00	36.04%	36.00%	0.04%	0.11%
6	1.03	1.02	0.01	13.36%	13.30%	0.06%	0.45%
7	0.21	0.22	-0.01	2.88%	2.93%	-0.05%	-1.62%
8	0.47	0.46	0.01	5.94%	5.90%	0.04%	0.62%
9	0.34	0.32	0.02	6.34%	5.95%	0.38%	6.05%
10	1.20	1.20	0.00	14.19%	14.20%	0.00%	-0.03%
11	1.35	1.35	0.00	23.10%	22.99%	0.11%	0.48%
12	0.38	0.38	0.00	4.99%	4.97%	0.01%	0.26%

Fonte: Elaboração própria.

No caso 2 (Tabela 10), observa-se que a previsão um passo à frente é mais bem realizada pelo método proposto, sendo 31,93% melhor que o PAR(*p*). No total dos meses, o RMSE e o MAPE do PAR(*p*) foram melhores em apenas dois e três meses, respectivamente. Na previsão deste caso, o método proposto se mostrou superior, assim como na estimação dos parâmetros do processo gerador dos dados.

Tabela 10 - RMSE e MAPE da previsão *out-of-sample* (caso 2)

Mês	RMSE			MAPE			Diferença Relativa
	PAR(<i>p</i>) (1)	WS + PAR(<i>p</i>) (2)	Dif (1) - (2)	PAR(<i>p</i>) (3)	WS + PAR(<i>p</i>) (4)	Dif (3) - (4)	
1	2.00	1.36	0.64	51.35%	34.96%	16.40%	31.93%
2	1.49	1.33	0.15	18.06%	16.19%	1.86%	10.31%
3	0.25	0.24	0.02	2.89%	2.71%	0.18%	6.33%
4	0.32	0.18	0.14	3.18%	1.79%	1.39%	43.73%
5	2.26	2.33	-0.06	22.68%	23.32%	-0.64%	-2.81%
6	1.11	1.15	-0.04	10.18%	10.51%	-0.33%	-3.21%
7	5.23	5.22	0.01	39.76%	39.68%	0.08%	0.21%
8	3.88	3.88	0.00	27.63%	27.65%	-0.02%	-0.08%
9	7.40	7.39	0.01	45.60%	45.54%	0.06%	0.13%
10	0.04	0.04	0.00	0.37%	0.37%	0.00%	0.11%
11	1.66	1.65	0.01	17.48%	17.40%	0.08%	0.48%
12	0.25	0.23	0.01	2.45%	2.32%	0.13%	5.38%

Fonte: Elaboração própria.

No caso 3 (Tabela 11), a previsão um passo à frente realizada pelo método proposto se mostrou 64,51% melhor que o $PAR(p)$. No total dos meses, o RMSE e o MAPE do método proposto tiveram um desempenho pior, respectivamente, em quatro e cinco meses do total, mesmo assim, com valores muito pequenos se comparados aos demais meses. Na previsão deste caso, o método proposto foi superior, assim como na estimação dos parâmetros do processo gerador dos dados deste caso.

Tabela 11 - RMSE e MAPE da previsão *out-of-sample* (caso 3)

Mês	RMSE			MAPE			Diferença Relativa
	$PAR(p)$ (1)	WS + $PAR(p)$ (2)	Dif (1) - (2)	$PAR(p)$ (3)	WS + $PAR(p)$ (4)	Dif (3) - (4)	
1	3.12	1.11	2.01	31.53%	11.19%	20.34%	64.51%
2	2.71	2.22	0.50	31.38%	25.63%	5.75%	18.33%
3	4.99	4.79	0.20	40.52%	38.92%	1.60%	3.95%
4	3.13	3.03	0.10	23.90%	23.15%	0.75%	3.13%
5	1.52	1.58	-0.05	24.29%	25.15%	-0.86%	-3.53%
6	1.66	1.67	-0.01	19.93%	20.01%	-0.08%	-0.39%
7	3.26	3.27	-0.01	30.81%	30.90%	-0.09%	-0.29%
8	0.33	0.34	-0.01	2.99%	3.08%	-0.09%	-3.01%
9	1.27	1.25	0.01	18.24%	18.06%	0.18%	1.01%
10	8.15	8.15	0.00	373.30%	373.23%	0.07%	0.02%
11	1.57	1.57	0.00	25.47%	25.49%	-0.02%	-0.09%
12	1.48	1.47	0.01	16.71%	16.63%	0.08%	0.47%

Fonte: Elaboração própria.

No caso 4 (Tabela 12), as previsões um passo e dois passos à frente realizadas pelo método proposto se mostraram, respectivamente, 21,65% e 24,20% melhor que o $PAR(p)$. Neste caso, o método proposto obteve melhor desempenho em 11 meses na previsão *out-of-sample*. Somente perdeu no mês 10 nas estatísticas RMSE e MAPE. Vale destacar que a estimação dos parâmetros média e desvio-padrão deste método proposto foram melhores, fato este que justifica o encolhimento wavelet das séries referentes a cada um dos meses, com o intuito de se apurar melhor estes parâmetros e, conseqüentemente, obter melhores resultados para a previsão.

Tabela 12 - RMSE e MAPE da previsão *out-of-sample* (caso 4)

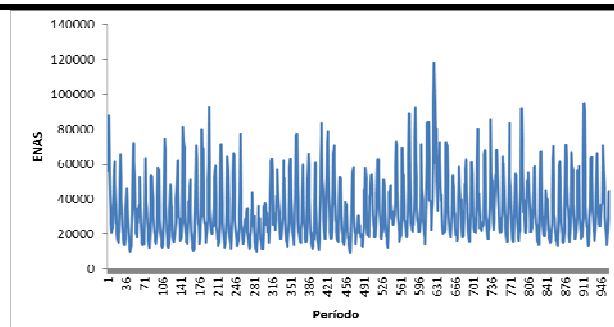
Mês	RMSE			MAPE			Diferença Relativa
	PAR(p) (1)	WS + PAR(p) (2)	Dif (1) - (2)	PAR(p) (3)	WS + PAR(p) (4)	Dif (3) - (4)	
1	3.17	2.48	0.69	16.83%	13.19%	3.64%	21.65%
2	0.47	0.35	0.11	1.89%	1.43%	0.46%	24.20%
3	1.93	1.88	0.04	8.76%	8.56%	0.20%	2.28%
4	5.39	5.34	0.05	17.92%	17.75%	0.17%	0.92%
5	9.48	9.43	0.05	32.74%	32.58%	0.16%	0.49%
6	2.58	2.58	0.01	9.23%	9.21%	0.02%	0.22%
7	6.13	6.12	0.00	44.12%	44.11%	0.01%	0.02%
8	6.72	6.71	0.01	35.85%	35.82%	0.03%	0.09%
9	8.05	8.05	0.00	67.70%	67.67%	0.03%	0.05%
10	5.77	5.78	-0.01	18.74%	18.76%	-0.02%	-0.09%
11	3.91	3.91	0.00	24.29%	24.28%	0.01%	0.02%
12	0.11	0.10	0.00	0.42%	0.41%	0.00%	0.93%

Fonte: Elaboração própria.

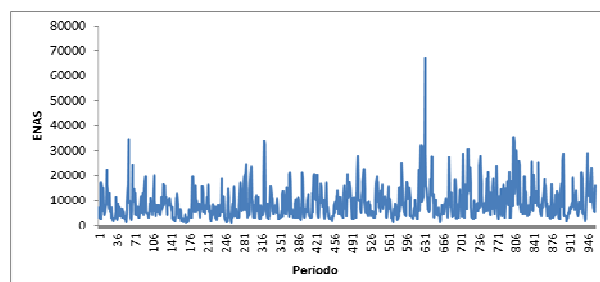
4.2

Resultados da aplicação às séries de ENA

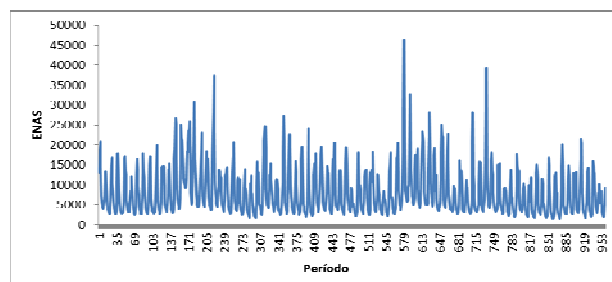
Nesta tese as séries temporais utilizadas são as séries de ENA referentes aos subsistemas Sudeste/Centro-Oeste, Sul, Nordeste e Norte. A base de dados se encontra em ONS (2012). Os dados utilizados são mensais e o horizonte de estudo inicia-se em janeiro de 1931 e termina em dezembro de 2010, sendo que a unidade de medida é o Megawatt médio (MWmed). As séries históricas podem ser observadas por meio da Figura 9, em que a série referente ao subsistema Sudeste/Centro-Oeste possui valores absolutos maiores que as demais e a série de ENA do subsistema Sul é, aparentemente, a mais irregular, embora seja a que apresenta valores absolutos menores. O método proposto apresenta maiores ganhos relativos nestes dois casos citados (vide Tabela 13).



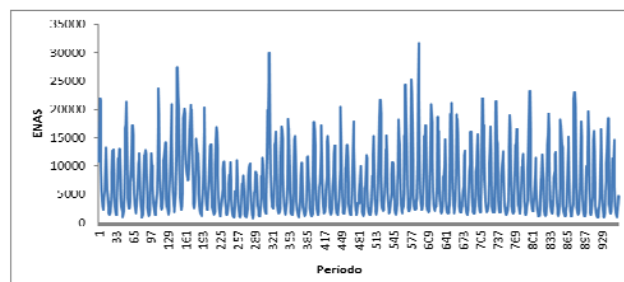
(a) Sudeste/Centro-Oeste



(b) Sul



(c) Nordeste



(d) Norte

Figura 9 - Séries históricas de energia natural afluyente (ENA)

Fonte: ONS (2012).

Nesta seção, o critério de identificação adotado é sempre o critério um (utilizado pelo setor elétrico brasileiro), citado na seção 2.2, em que se deve procurar a maior ordem i para cada período m de tal forma que todas as estimativas de $\hat{\phi}_{kk}^m$ para $k > i$ não sejam mais significativas, sempre respeitando um limite o número de termos auto-regressivos igual a seis em cada mês.

Conforme destacado, os melhores *setups* encontrados para as séries referentes a cada um dos meses são compostos pelo número de camadas da decomposição wavelet, a família wavelet utilizada, o tipo de limiar e a regra de limiarização para cada um dos meses. Estas informações referentes às séries de ENA juntamente com o p-valor da série de resíduos de cada *setup* podem ser vistos nos Anexos 6, 7, 8 e 9. As hipóteses nulas de: erros descorrelatados; normalidade dos resíduos; e variância dos erros não segue um processo do tipo ARCH/GARCH e não foram rejeitadas, em nenhum dos casos, a 5% de significância.

Na Tabela 13, são comparados os ajustes dos modelos às séries por meio das estatísticas RMSE e MAPE, além da diferença relativa²⁸ entre os modelos (quando positivo, indica que o método proposto foi melhor e quando negativo indica que este foi pior). Os resultados obtidos indicam que o método proposto obtém melhor ajuste na maioria dos meses analisados. Somente nos meses de junho do subsistema Sudeste/Centro-Oeste e de julho do subsistema Sul, o PAR(p) obteve melhor ajuste, sendo que, no primeiro caso, este se mostrou como a menor diferença relativa entre todos os meses da série referente ao Sudeste/Centro-Oeste.

Em média, o WS + PAR(p) se mostrou 2,67% melhor no subsistema Sudeste/Centro-Oeste, 3,32% no Sul, 2,36% no Nordeste e 2,01% no Norte. Os maiores ganhos relativos ocorreram nos meses de abril do subsistema Sudeste/Centro-Oeste (5,42% - o MAPE foi 0,66% menor), janeiro do subsistema Sul (10,46% - o MAPE foi 4,96% menor), julho do subsistema Nordeste (5,85% - o MAPE foi 0,31% menor) e maio do subsistema Norte (2,85% - o MAPE foi 0,33% menor). As perdas relativas ocorreram somente nos meses junho do Sudeste/Centro-Oeste (0,99% - o MAPE foi 0,10% maior) e julho do Sul (0,91% - o MAPE foi 0,40% maior).

É importante observar que as diferenças entre as medidas RMSE para os subsistemas Nordeste e Norte se mostraram pequenas, porém estas diferenças para os subsistemas Sudeste/Centro-Oeste e Sul foram maiores, devido aos elevados valores de ENA do Sudeste/Centro-Oeste e devido ao ganho relativo do WS + PAR(p) no caso do Sul ter sido maior que nos outros casos.

²⁸ Segundo o MAPE.

Tabela 13 - RMSE e MAPE dos ajustes das séries de ENA

Subsistema	Mês	RMSE			MAPE			Diferença Relativa
		PAR(p) (1)	WS + PAR(p) (2)	Dif (1) - (2)	PAR(p) (3)	WS + PAR(p) (4)	Dif (3) - (4)	
Sudeste/ Centro- Oeste	Janeiro	51.12	49.03	2.09	18.37	17.66	0.71	3.88%
	Fevereiro	56.30	53.21	3.08	20.28	19.31	0.97	4.80%
	Março	46.84	46.41	0.43	16.77	16.44	0.33	1.95%
	Abril	29.43	28.22	1.20	12.17	11.51	0.66	5.42% *
	Maiο	22.19	21.91	0.28	10.37	10.24	0.13	1.21%
	Junho	23.41	23.59	-0.19	10.26	10.36	-0.10	-0.99%
	Julho	13.85	13.54	0.32	7.68	7.53	0.15	1.97%
	Agosto	15.76	15.36	0.40	9.15	8.86	0.29	3.16%
	Setembro	22.45	22.26	0.20	13.14	12.85	0.30	2.24%
	Outubro	30.86	29.90	0.96	17.64	17.06	0.58	3.31%
	Novembro	29.96	29.02	0.94	14.15	13.67	0.48	3.41%
	Dezembro	36.60	36.04	0.57	14.88	14.63	0.25	1.71%
	Média	31.56	30.71	0.86	13.74	13.34	0.40	2.67%
Sul	Janeiro	41.45	37.54	3.91	47.45	42.48	4.96	10.46% **
	Fevereiro	41.74	40.73	1.01	44.62	42.75	1.87	4.20%
	Março	30.66	29.17	1.50	33.85	31.58	2.27	6.71% *
	Abril	40.28	39.75	0.54	46.08	45.07	1.01	2.19%
	Maiο	61.83	60.40	1.43	82.38	80.13	2.25	2.73%
	Junho	50.92	49.76	1.15	53.55	52.13	1.42	2.65%
	Julho	50.07	50.47	-0.41	43.93	44.32	-0.40	-0.91%
	Agosto	62.29	60.79	1.50	68.99	67.27	1.72	2.49%
	Setembro	56.16	53.82	2.34	53.28	50.80	2.48	4.65%
	Outubro	56.24	56.00	0.24	49.17	48.83	0.34	0.69%
	Novembro	43.75	43.23	0.52	43.31	42.85	0.46	1.06%
	Dezembro	39.29	38.26	1.03	48.09	46.69	1.40	2.91%
	Média	47.89	46.66	1.23	51.22	49.58	1.65	3.32%
Nordeste	Janeiro	28.94	28.21	0.73	19.52	19.04	0.48	2.48%
	Fevereiro	38.63	37.94	0.68	27.46	26.97	0.49	1.78%
	Março	37.92	36.95	0.97	23.54	22.96	0.58	2.45%
	Abril	34.23	33.27	0.96	27.74	26.92	0.82	2.97%
	Maiο	18.09	18.09	0.00	14.84	14.81	0.03	0.18%
	Junho	7.29	7.18	0.11	8.41	8.17	0.23	2.77%
	Julho	4.49	4.39	0.10	5.33	5.02	0.31	5.85% *
	Agosto	3.38	3.35	0.03	4.42	4.36	0.05	1.22%
	Setembro	4.66	4.62	0.05	6.62	6.56	0.06	0.97%
	Outubro	8.98	8.80	0.18	12.82	12.53	0.29	2.29%
	Novembro	20.41	19.93	0.48	23.21	22.59	0.62	2.68%
	Dezembro	28.92	28.23	0.69	25.55	24.84	0.70	2.74%
	Média	19.66	19.24	0.41	16.62	16.23	0.39	2.36%
Norte	Janeiro	26.15	25.87	0.28	21.97	21.70	0.28	1.26%
	Fevereiro	29.99	29.42	0.57	22.54	22.03	0.51	2.25%
	Março	22.39	21.84	0.55	15.02	14.63	0.39	2.60%
	Abril	20.47	20.03	0.45	13.65	13.30	0.35	2.56%
	Maiο	14.12	13.78	0.35	11.40	11.08	0.33	2.85%
	Junho	7.67	7.52	0.15	9.00	8.80	0.20	2.23%
	Julho	3.76	3.73	0.02	5.95	5.88	0.07	1.15%
	Agosto	2.71	2.69	0.02	4.94	4.88	0.06	1.28%
	Setembro	3.91	3.88	0.03	7.42	7.38	0.04	0.54%
	Outubro	6.86	6.76	0.10	12.71	12.46	0.25	1.95%
	Novembro	11.75	11.54	0.21	17.73	17.24	0.50	2.80%
	Dezembro	21.02	20.53	0.49	23.14	22.52	0.62	2.67%
	Média	14.23	13.97	0.27	13.79	13.49	0.30	2.01%

Fonte: Elaboração própria.

* indica que a diferença relativa é maior que 5%.

** indica que a diferença relativa é maior que 10%.

As médias e os desvios-padrão médios dos cenários gerados²⁹ pelos modelos $PAR(p)$ e $WS + PAR(p)$, e as médias e os desvios-padrão das séries históricas, são apresentados na Figura 10. As séries em azul correspondem às médias e aos desvios-padrão referentes a cada mês do ano, calculados a partir da série histórica. Como esta análise foi feita para um período de 60 meses (ou 5 anos), os picos e vales da série em azul são exatamente iguais por 5 vezes (as 12 médias obtidas pelas séries históricas são repetidas 5 vezes e os desvios-padrão obtidos pelas séries históricas são repetidos 5 vezes). O objetivo é que as séries geradas tenham as propriedades estatísticas iguais às das séries históricas e, na Figura 10, é realizada apenas uma comparação visual entre médias e desvios-padrão dos cenários obtidos por $PAR(p)$ e $WS + PAR(p)$ e médias mensais e desvios-padrão mensais da série histórica.

Foram gerados 200 cenários para o horizonte de 60 meses. Observa-se que nos quatro subsistemas, a média histórica e a média dos cenários gerados pelos dois modelos têm um comportamento muito similar. Com relação aos desvios-padrão, vale lembrar (seção 4.1) que eles são sempre estimados em menor magnitude pela metodologia proposta em relação à estimação pelo $PAR(p)$, fato que não prejudica os desvios-padrão dos cenários. É possível observar que estes parâmetros gerados pelo $WS + PAR(p)$ são mais próximos dos históricos em muitos dos instantes de tempo observados, *e.g.*, na Figura 10(h), os desvios-padrão gerados pelo $WS + PAR(p)$ reproduzem melhor o histórico, principalmente em seus picos.

²⁹ Seguindo a metodologia apresentada na seção 2.2.2.

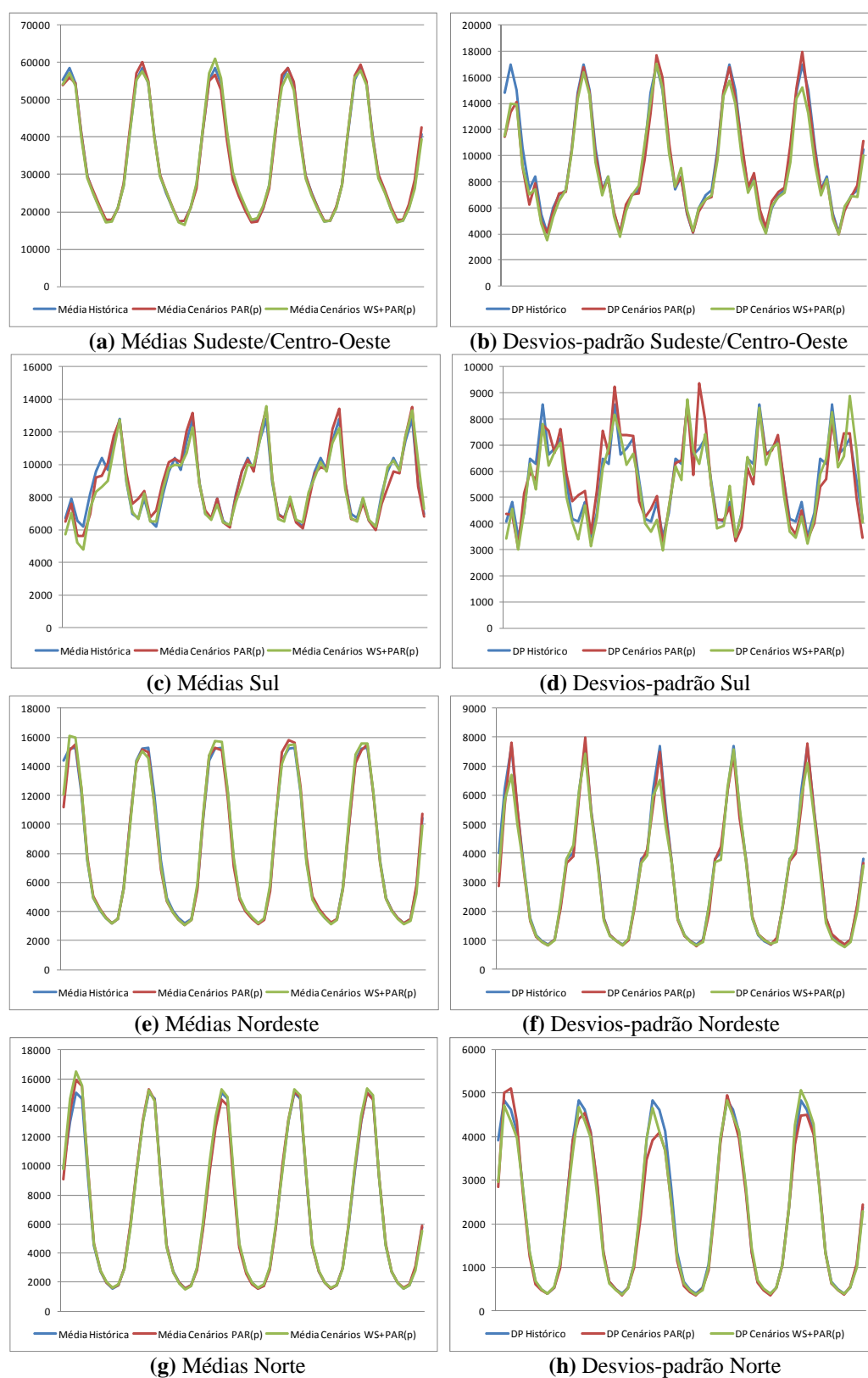


Figura 10 - Médias e desvios-padrão médios dos cenários

Fonte: Elaboração própria.

Em seguida, são realizados os testes de média para verificar se as médias dos cenários gerados são estatisticamente iguais às médias do histórico; A evolução dos p-valores é analisada nas Figuras 11 a 14. Testa-se, por meio da estatística t , a hipótese nula da média dos 200 elementos de cada um dos 60 períodos da matriz de cenários ser igual à média das observações de cada um dos meses da série histórica. Analogamente ao realizado para os testes de médias, o Teste de Levene verifica a igualdade de variâncias. Os resultados apresentados nas Figuras 11 a 14 indicam que as variâncias dos cenários são estatisticamente iguais às variâncias históricas para a maior parte dos períodos.

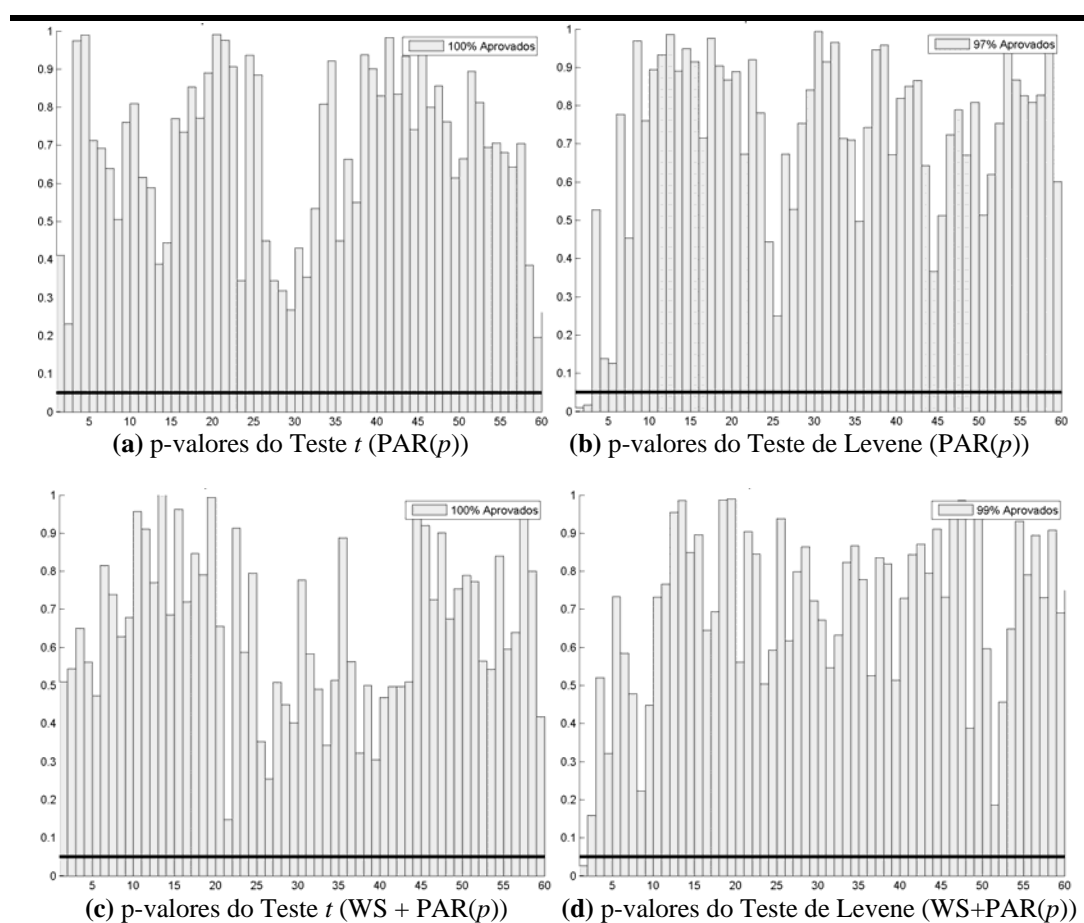


Figura 11 - Testes para o subsistema Sudeste/Centro-Oeste

Fonte: Elaboração própria.

Obs: A linha escura indica o nível de significância de 5%.

Os cenários gerados pelo $PAR(p)$ obtiveram melhores resultados na reprodução da média histórica. Os índices de não rejeição da hipótese nula de igualdade de médias são de, pelo menos, 99% para os cenários gerados pelo $PAR(p)$ frente a 95% para o $WS + PAR(p)$. É possível observar nas Figuras 12(a), 12(c), 13(a), 13(c) e 14(c) que a reprodução da média histórica no início da geração dos cenários é mais difícil, porém, à medida que o tempo passa, os p-valores ficam sempre acima do valor ao nível de significância de 5%.

Já os cenários gerados pelo $WS + PAR(p)$ obtiveram melhores resultados na reprodução da variância histórica. Os índices de não rejeição da hipótese nula de igualdade de variâncias são de, no mínimo, 99% para os cenários gerados pelo $WS + PAR(p)$, enquanto o outro modelo obteve índices de, pelo menos 97%.

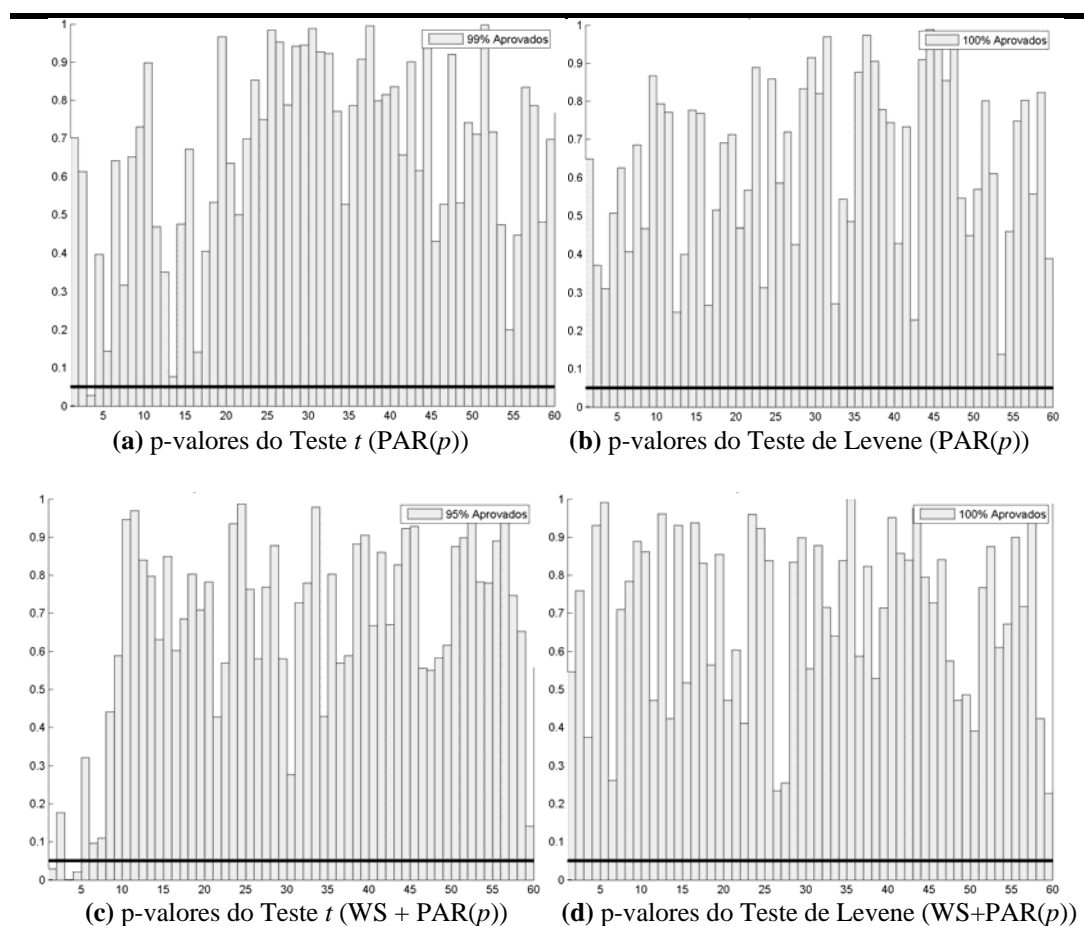


Figura 12 - Testes para o subsistema Sul

Fonte: Elaboração própria.

Obs: A linha escura indica o nível de significância de 5%.

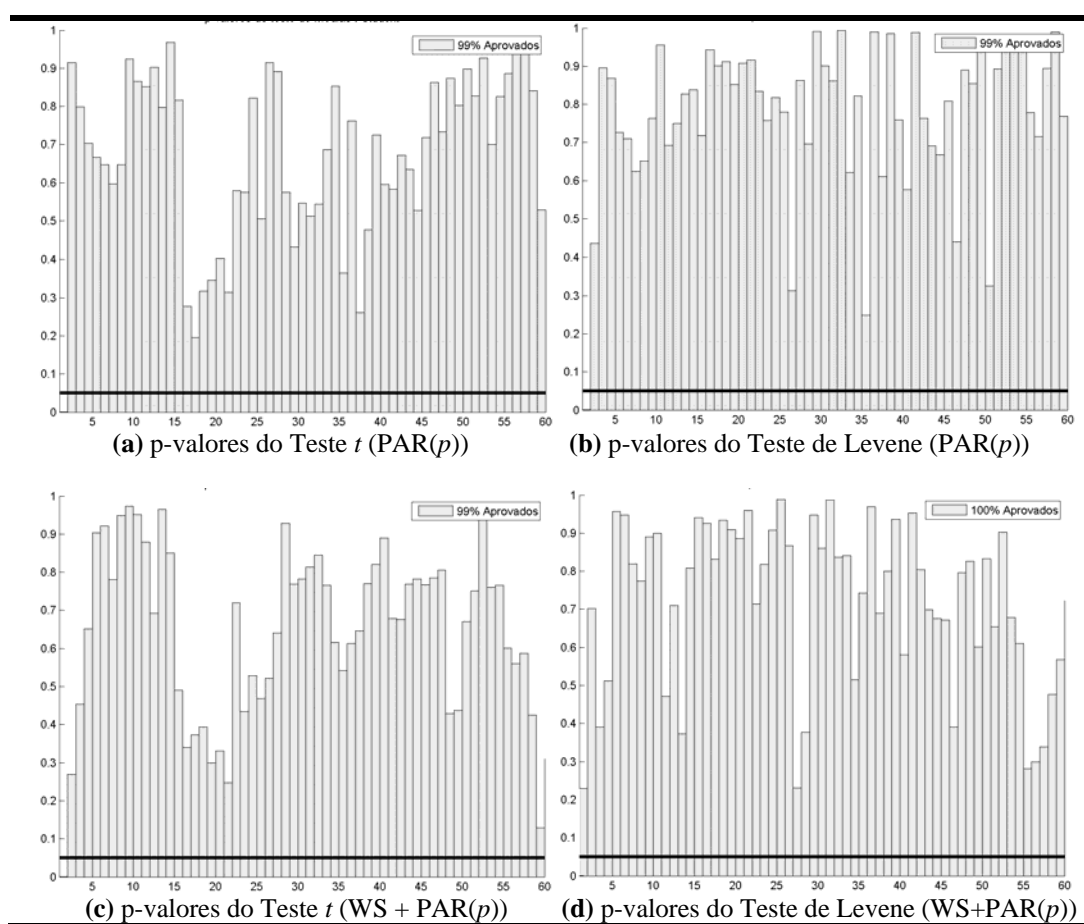


Figura 13 - Testes para o subsistema Nordeste

Fonte: Elaboração própria.

Obs: A linha escura indica o nível de significância de 5%.

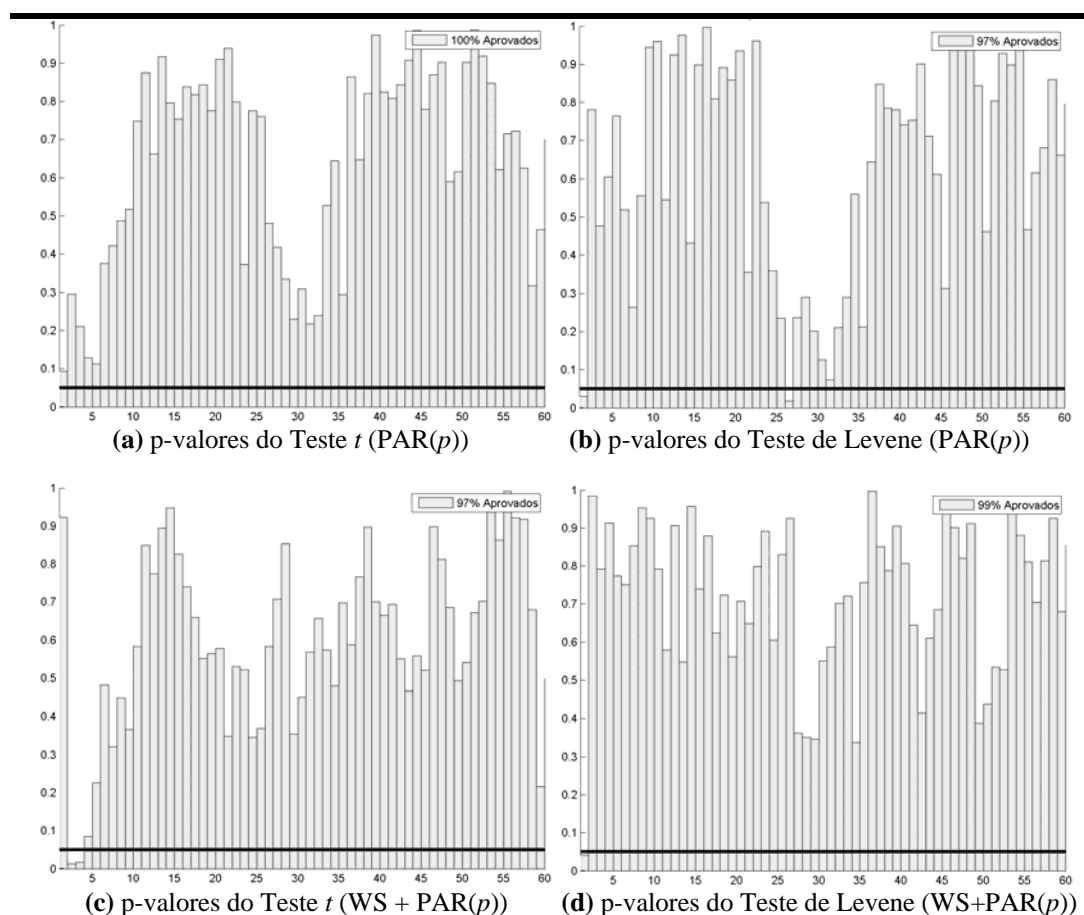


Figura 14 - Testes para o subsistema Norte

Fonte: Elaboração própria.

Obs: A linha escura indica o nível de significância de 5%.

Com o objetivo de verificar a capacidade dos cenários em reproduzir os períodos críticos de seca observados na série original, utiliza-se o conceito de sequências, destacado na seção 3.6.4.4, para a definição das seguintes variáveis aleatórias: comprimento, soma e intensidade de sequência. Com isso, é possível observar se os cenários reproduzem as distribuições de probabilidades de cada uma destas variáveis, comparando-as com a sua respectiva distribuição histórica. Para tal finalidade, são utilizados os testes de aderência de Kolmogorov-Smirnov para a soma e a intensidade e Qui-Quadrado para o comprimento.

No teste de Kolmogorov-Smirnov, a análise é feita com base no p-valor e no teste Qui-Quadrado, ela é realizada por meio da estatística de teste. Não rejeitar as hipóteses nulas significa obter p-valores superiores ao nível de significância considerado de 5% e valores de teste menores que o valor crítico tabelado de 3,84 (segundo o número de graus de liberdade e o nível de significância). Verificando

as Tabelas 14 (PAR(p)) e 15 (WS + PAR(p)), pode-se concluir que ambos os modelos chegaram aos mesmos resultados e só não foram capazes de reproduzir a variável intensidade para o subsistema Sul, com os respectivos p-valores: 0,02 e 0,01; e a variável comprimento para o subsistema Norte, com os respectivos valores de estatísticas de teste: 8,15 e 7,26 (neste ponto, vale destacar que a metodologia proposta apresentou uma menor distância em relação ao valor crítico).

Tabela 14 - Análise das sequências negativas (PAR(p))

Subsistema	Comprimento	Soma	Intensidade
	Valor crítico: 3,84	p-valor mín: 0,05	p-valor mín: 0,05
Sudeste/Centro-Oeste	0.20	0.20	0.11
Sul	0.15	0.59	0.02
Nordeste	1.49	0.94	0.73
Norte	8.15	0.33	0.24

Fonte: Elaboração própria.

Tabela 15 - Análise das sequências negativas (WS + PAR(p))

Subsistema	Comprimento	Soma	Intensidade
	Valor crítico: 3,84	p-valor mín: 0,05	p-valor mín: 0,05
Sudeste/Centro-Oeste	0.26	0.17	0.18
Sul	0.35	0.64	0.01
Nordeste	1.81	0.86	0.68
Norte	7.26	0.29	0.33

Fonte: Elaboração própria.

Além disso, para o caso do subsistema Sudeste/Centro-Oeste, em que os valores de ENA são maiores, fazendo com que pequenos ganhos em termos de modelagem tenham grandes impactos, as médias dos cenários gerados foram utilizadas como previsão pontual³⁰ para o período de 2006 a 2010 e, estes valores foram comparados aos valores reais ocorridos. Esta comparação é realizada por

³⁰ De acordo com Sharma (2000), uma alternativa para as previsões pontuais é simular uma variedade de cenários plausíveis de sequências de resultados que podem ocorrer no futuro. Em seu trabalho, este autor simulou uma variedade de sequências de chuvas possíveis de ocorrer, para serem utilizadas na política operacional ótima do reservatório de Warragamba dam, localizado perto de Sydney (Austrália). Desta forma, é possível atribuir probabilidades às previsões, permitindo a avaliação de riscos e incertezas referentes ao sistema hidrelétrico. A média destas previsões probabilísticas é equivalente à previsão pontual.

meio do MAPE. Na Figura 15, pode-se observar que o método proposto é melhor que o PAR(p) em 11 meses, só perdendo no mês de fevereiro.

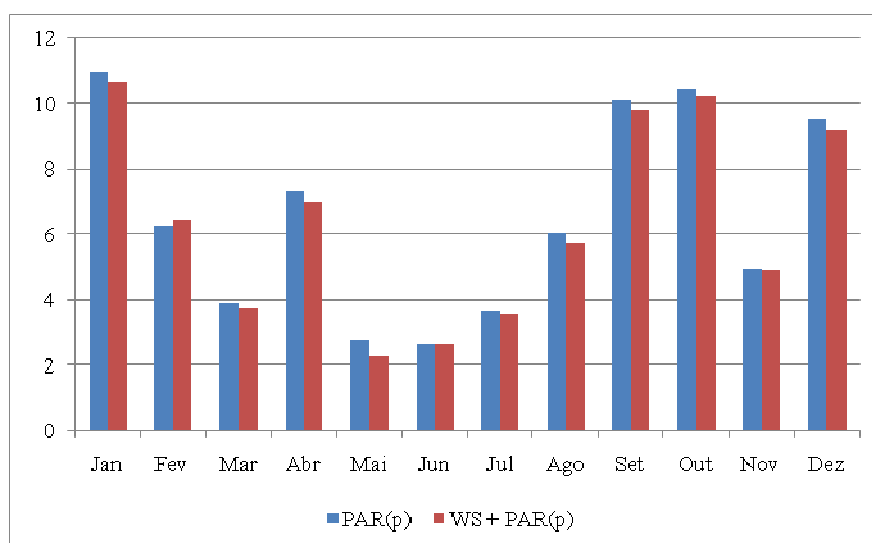


Figura 15 - MAPE das previsões out-of-sample dos modelos PAR(p) e WS + PAR(p) para os anos de 2006 a 2010

Fonte: Elaboração própria.

Foi possível observar que a aplicação do WS + PAR(p) aos casos simulados obteve melhores resultados, em comparação ao modelo PAR(p), no que diz respeito à estimação dos parâmetros e à previsão *out-of-sample*. Quando aplicado às séries reais, os ajustes obtidos pelo WS + PAR(p) também foram superiores. A estratégia de comparação adotada para as séries reais foi a de simular cenários, assim como realizado no setor elétrico brasileiro. Os cenários obtidos por ambos os modelos foram muito semelhantes, valendo destacar uma pequena vantagem para os cenários gerados pelo PAR(p) na reprodução da média histórica e uma pequena vantagem para o WS + PAR(p) na reprodução do desvio-padrão histórico. Ao utilizar as médias dos cenários para a previsão da série de ENA do subsistema Sudeste, o método proposto também apresentou ganhos.

Conclusões

Esta tese desenvolveu uma combinação de encolhimento wavelet e modelos auto-regressivos periódicos com o objetivo de obter melhores previsões e geração de cenários a partir do modelo $PAR(p)$; e avançar no que diz respeito à modelagem estocástica de vazões. Assim como as combinações de decomposição wavelet e/ou encolhimento wavelet e modelos preditivos já realizadas na literatura, era de se esperar que o método proposto fosse mais apropriado para explicar o comportamento das séries de tempo analisadas e que obtivesse ganhos de previsão.

Para alcançar tais objetivos, a metodologia proposta foi aplicada a dois tipos de séries: 1) séries temporais mensais simuladas com 1.200 observações, realizações de diferentes processos $PAR(p)$, e 2) séries de ENAs utilizadas para o planejamento do setor elétrico brasileiro. Foram empregados dois diferentes critérios para a identificação da ordem dos modelos e diferentes tipos de análises foram realizados em cada um dos tipos de séries.

Os resultados obtidos indicaram que, por meio dos casos simulados, o método proposto obteve melhor ajuste e realizou a estimação das médias de uma forma mais precisa. Com relação aos desvios-padrão, sempre que estes foram superestimados pelo $PAR(p)$, o método proposto o estimou de melhor forma. Estas observações comprovam a eficiência do método proposto em estimativas mais apuradas para estes parâmetros.

No que diz respeito aos parâmetros auto-regressivos, não foi possível apontar uma ampla vantagem de algum dos modelos, porém pôde ser observado que o $PAR(p)$ realizou a identificação errônea destes parâmetros em alguns dos meses analisados. Os encolhimentos wavelets do $WS + PAR(p)$ diminuíram a influência de valores desproporcionais de um ou mais pares de observações na estimativa da FACP, reduzindo os valores espúrios e identificando a ordem correta dos processos geradores dos dados. Na previsão *out-of-sample* das séries simuladas, o método proposto foi melhor em três dos quatro casos propostos.

Com relação à aplicação às séries de ENA referentes aos subsistemas Sudeste/Centro-Oeste, Sul, Nordeste e Norte, os ajustes dos modelos foram medidos por meio das estatísticas RMSE e MAPE e foi possível observar que o método proposto obteve melhor ajuste em quase todos os meses analisados. Somente nos meses de junho do subsistema Sudeste/Centro-Oeste e de julho do subsistema Sul, o $PAR(p)$ obteve melhor ajuste, sendo que, no primeiro caso, este se mostrou como a menor diferença relativa entre todos os meses da série referente ao Sudeste/Centro-Oeste. Em média, o $WS + PAR(p)$ foi 2,67% melhor no subsistema Sudeste/Centro-Oeste, 3,32% no Sul, 2,36% no Nordeste e 2,01% no Norte.

Para o caso das séries reais, foram gerados 200 cenários para um horizonte de 60 meses. Para os quatro subsistemas, a média histórica e a média dos cenários gerados pelos dois modelos obtiveram um comportamento muito similar. Já os desvios-padrão médios gerados pelo $WS + PAR(p)$ ficaram mais próximos dos históricos em alguns dos instantes de tempo observados.

Foram realizados testes de média e de variância para verificar as respectivas hipóteses nulas: média dos cenários estatisticamente igual à média das observações e variância dos cenários estatisticamente igual à variância histórica. A evolução dos p-valores para os testes obteve valores semelhantes, apontando para uma pequena vantagem dos cenários gerados pelo $PAR(p)$ na reprodução da média histórica e uma pequena vantagem os cenários gerados pelo $WS + PAR(p)$ na reprodução da variância histórica. Ambos também atingiram resultados semelhantes na reprodução de sequências negativas.

A combinação de modelos aqui apresentada, contemplou o encolhimento de ruídos wavelets e os modelos $PAR(p)$, e como parte dos avanços metodológicos incorporados nesta tese, podem-se citar:

- i) a realização de uma combinação de wavelets com modelos auto-regressivos periódicos, ainda não contemplada na literatura;
- ii) a utilização de testes de resíduos para validar os encolhimentos wavelets, garantido que apenas ruídos fossem eliminados e não fossem perdidas informações relevantes à modelagem das séries;
- iii) foi realizada a modelagem de algumas séries simuladas, permitindo assim observar e comparar o desempenho do mesmo em um ambiente controlado;

iv) o método proposto atingiu melhores ajustes na modelagem das séries de ENA, devido ao tratamento dos dados, que por sua vez, são muito irregulares.

Com isso, a metodologia aqui proposta ampliou a potencialidade dos modelos $PAR(p)$ e como extensões para este trabalho podem-se citar: i) a aplicação a outros tipos de séries que possuam comportamento semelhante às séries estudadas nesta tese, como as séries de vento e as de radiação; ii) o desenvolvimento de uma ferramenta computacional para ajustar, de forma automática o $WS + PAR(p)$ a diversas séries resultantes de um mesmo processo gerador de dados, aumentando assim, a confiabilidade estatística da estimação dos modelos; e iii) a utilização do $PAR(p)$ interconfigurações, descrito em Ferreira (2013), no lugar do $PAR(p)$ utilizado nesta tese, possibilitando incorporar os efeitos de futuras usinas que entrarão em operação para a geração das séries sintéticas do setor elétrico brasileiro. Para realizar esta abordagem, é necessário ajustar um modelo auto-regressivo de ordem p para cada um dos meses e cada uma das configurações das séries hidrológicas históricas e estimar tantos modelos quanto o número de meses em que ocorre a mudança de configuração.

Referências bibliográficas

ALMEIDA, A.; ELLIAN, S. N. J. Modificações e alternativas aos testes de levene e de brown e forsythe para igualdade de variâncias e médias. *Revista Colombiana de Estadística*, v.31, n.2, p. 241–260, 2008.

ALRUMAIH, R.M.; AL-FAWZAN, M.A. Time Series Forecasting Using Wavelet Denoising: an Application to Saudi Stock Index. *Journal of King Saud University*, v.14, n.2, p.221-234, 2002.

ARINO, M.A. *Time series forecasts via wavelets: An application to car sales in the Spanish market*. Institute of Statistics & Decision Sciences, Duke University, 1995. 19p. (Discussion paper, 95-30).

ARINO, M.A.; VIDA KOVIC, B. *On wavelet scalograms and their applications in economic time series*. Institute of Statistics & Decision Sciences, Duke University, 1995. 19p. (Discussion paper, 95-21).

BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. *Time Series Analysis: Forecasting and Control*. 3ed. New Jersey: Prentice Hall, 1994.

CAMPISI-PINTO, S.; ADAMOWSKI, J.; ORON, G. Forecasting Urban Water Demand Via Wavelet-Denoising and Neural Network Models. Case Study: City of Siracuse, Italy. *Water Resources Management*, v.26, n.12, p.35-39, 2012.

CAMPOS, L.C.D. *Modelo Estocástico Periódico baseado em Redes Neurais*. 2010. 174f. Tese (Doutorado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2010.

CASELLA, G.; BERGER, R.L. *Inferência Estatística*. São Paulo: Cengage Learning, 2010.

CATALÃO, J.P.S.; POUSINHO, H.M.I.; MENDES, V.M.F. Short-Term Wind Power Forecasting in Portugal by Neural Networks and Wavelet Transform. *Renewable Energy*, v.36, p.1245-1251, 2011.

CENTRO DE PESQUISAS DE ENERGIA ELÉTRICA - CEPEL. *Manual de referência do modelo Newave*. Relatório Técnico, Rio de Janeiro, 2000. Disponível em: <http://www.cose.fee.unicamp.br/cose/it511/newave-decompdessem/Newave_comentado.pdf>. Acesso em: 12 mai. 2013.

CHARBENEAU, R.J. Comparison of the Two- and Three-Parameter Log Normal Distributions Used in Streamflow Synthesis. *Water Resources Research*, v.14, n.1, p.149-150, 1978.

CHATFIELD, C. *The Analysis of Time Series: An Introduction*. 5ed. Londres: Chapman & Hall, 1996.

CONOVER, W.U. *Practical Nonparametric Statistics*. 1ed. New York: John Wiley and Sons, 1971.

DAUBECHIES, I. Orthonormal Bases of Compactly Supported Wavelets. *Comm. Pure and Applied Math*, v.41, p.909-996, 1988.

DAUBECHIES, I. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, 1992.

DONOHU, D.L.; JOHNSTONE, I. M. *Adapting Unknown Smoothness via Wavelet Shrinkage*. Division of Biostatistics/Stanford University Technical Report, n.161, 1993.

DONOHU, D.L.; JOHNSTONE, I.M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, v.81, n.3, p.425-455, 1994.

DONOHU, D.L.; JOHNSTONE, I.M. Minimax Estimation via Wavelet Shrinkage. *The Annals of Statistics*, v.26, n.3, p.879-921. 1998.

DONOHU, D.L.; JOHNSTONE, I.M.; KERKYACHARIAN, G.; PICARD, D. *Wavelet Shrinkage: Asymptopia?* Stanford University Technical Report, 1993.

DONOHU, D.L.; JOHNSTONE, I.M.; KERKYACHARIAN, G.; PICARD, D. *Universal near Minimaxity of Wavelet Shrinkage*. Stanford University Technical Report, 1995.

EMPRESA DE PESQUISA ENERGÉTICA - EPE. *Balanço Energético Nacional 2011: Ano base 2010*. Rio de Janeiro: EPE, 2011. Disponível em: <<https://ben.epe.gov.br/>>. Acesso em: 24 Abr. 2012.

FAYAL, M.A.A. *Previsão de vazão por redes neurais artificiais e transformada wavelet*. 2008. 158f. Dissertação (Mestrado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

FERNANDES, C.A.C. *Notas de aula da disciplina Análise de Séries Temporais*. Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2010.

FERREIRA, P.G.C. *A Estocasticidade associada ao setor elétrico brasileiro e uma nova abordagem para a geração de aflúências via modelos periódicos gama*. 2013. 139f. Tese (Doutorado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.

FORTUNATO, L.A.M.; ARARIPE NETO, T.A.; ALBUQUERQUE, J.C.R.; PEREIRA, M.V.F. *Introdução ao planejamento de expansão e operação de sistemas de produção de energia elétrica*. Niterói: Universidade Federal Fluminense, EDUFF, 1990.

GRAPS, A. An introduction to wavelets. *IEEE Computational Science and Engineering*, v.2, n.2, p. 50-61, 1995.

GUAN, C.P.; LUH, B.; MICHEL, L.D.; WANG, Y.; FRIEDLAN, P.B. Very Short-Term Load Forecasting: Wavelet Neural Networks With Data Pre-Filtering. *IEEE Transactions on Power Systems*, v.28, n.1, p.30-41, 2013.

GUTIERREZ, C.E.C. *Eliminação do Ruído por Encolhimento de Wavelets: Uma aplicação à série de Preço Spot de Energia Elétrica do Brasil*. 2002. 84f. Dissertação (Mestrado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2002.

HAMILTON, J.D. *Time Series Analysis*. Princeton: Princeton University Press, 1994.

HIPEL, K.W.; MCLEOD, A.I. *Time Series Modelling of Water Resources and Environmental Systems*. Amsterdam, The Netherlands: Elsevier, 1994.

JOHNSTONE, I.M.; SILVERMAN, B. *Wavelet Threshold Estimators for data with Correlated Noise*. Stanford University Technical Report, 1996.

KAVVAS, M.L.; DELLEUR, J.W. Removal of periodicities by differencing and monthly mean subtraction. *Journal of Hydrology*, v. 26, n. 3-4, p.335-353, 1975.

KUBRUSLY, C.S. *Elements of Operator Theory*. Boston: Birkhäuser, 2001.

KUBRUSLY, C.S.; LEVAN, N. Dual-Shift Decomposition of Hilbert Space. *Semigroups of Operators: Theory and Application*, v.2, p.145-147, 2002.

LEVAN, N.; KUBRUSLY, C.S. A Wavelet “Time-Shift-Detail” Decomposition. *Mathematics and Computers in Simulation*, v.63, p.73-78, 2003.

LI, L.L.; LI, J.H.; HE, P.J. e WANG, C.S. *The Use of Wavelet Theory and ARMA Model in Wind Speed Prediction*. 1st International Conference on Electric Power Equipment. China, 2011.

LIMA, P.C. Wavelets: uma introdução. *Matemática Universitária*, n.33, p.13-44, 2002.

LIU, B. *Wireless Communications, Networking and Mobile Computing*. International Conference on Wireless Communications, Networking and Mobile Computing. Shanghai, China, 2007.

LIU, H.; TIAN, H.Q.; CHEN, C.; LI, Y. A Hybrid Statistical Method to Predict Wind Speed and Wind Power. *Renewable Energy*, v.35, n.8, p.1857-1861, 2010.

LUNA, I.; BALLINI, R.; SOARES, S. Técnica de identificação de modelos lineares e não-lineares de séries temporais. *Revista Controle & Automação*, v.17, n.3, p.245-256, 2006.

MACEIRA, M.E.P. *Operação Ótima de Reservatórios com Previsão de Afluências*. 1989. 153f. Dissertação (Mestrado em Engenharia Civil) – Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE), Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1989.

MACEIRA, M.E.P.; BEZERRA, C.V. *Stochastic Streamflow Model for Hydroelectric Systems*. In: Proceedings of 5th International Conference on Probabilistic Methods Applied to Power Systems. Vancouver, Canadá, 1997.

MACEIRA, M.E.P.; PENNA, D.D.J.; DAMÁZIO, J.M. Geração de Cenários Sintéticos de Energia e Vazão para o Planejamento da Operação Energética. *Cadernos do IME – Série Estatística*, v.21, p.11-35, 2006.

MALLAT, S. *A Wavelet Tour of Signal Processing*. 2ed. San Diego: Academic Press. 1999.

MARCATO, A.L.M. *Representação híbrida de sistemas equivalentes e individualizados para o planejamento da operação de médio prazo de sistemas de potência de grande porte*. 2002. 200f. Tese (Doutorado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2002.

MCKERCHAR, A.I.; DELLEUR, J.W. Simulation procedures for Box-Jenkins models. *Water Resources Research*, v.10, n.2, p.246-255, 1974.

MONDAL, M.S.; WASIMI, S.A. Generating and forecasting monthly flows of the Ganges River with PAR model. *Journal of Hydrology*, v.323, n.1-4 p.41-56, 2006.

MORETTIN, P.A. *Wavelets in Statistics*. São Paulo: Instituto de Matemática e Estatística, Universidade de São Paulo, 1997.

MORETTIN, P.A.; TOLOI, C.M.C. *Análise de Séries Temporais*. 2ed. São Paulo: Edgard Blücher, 2006.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT - OECD. 2011 Report. Retrieved February 15, 2012. Disponível em: <<http://www.oecd.org/home/>>. Acesso em: 21 fev. 2013.

OLIVEIRA, F.L.C. *Nova abordagem para geração de cenários de aflúências no planejamento da operação energética de médio prazo*. 2010. 200f. Dissertação (Mestrado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2010.

OLIVEIRA, H.M. *Análise de sinais para engenheiros: uma abordagem via wavelets*. Rio de Janeiro: Brasport, 2007.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO – ONS. Rio de Janeiro: ONS, 2012. Disponível em: <<http://www.ons.org.br/>>. Acesso em: 23 Abr. 2012.

PENNA, D.D.J. *Definição da árvore de cenários de aflúências para o planejamento da operação energética de médio prazo*. 2009. 293f. Tese (Doutorado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009.

ROCHA, V.B. *Uma abordagem de wavelets aplicada à combinação de previsões: uma análise teórica e experimental*. 2008. 155f. Dissertação (Mestrado em Ciências) – Departamento de Matemática, Universidade Federal do Paraná, Curitiba, 2008.

SCHLÜTER, S.; DEUSCHLE, C. *Using wavelets for time series forecasting: Does it pay off?* German: Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung, University of Erlangen-Nuremberg, 2010. 32p. (Discussion paper, 4)

SHARMA, A. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 – A nonparametric probabilistic forecast model. *Journal of Hydrology*, v.239, n.1-4, p.249-258, 2000.

SIEGEL, S.; CASTELLAN JR., N.J. *Estatística não-paramétrica para ciências do comportamento*. Porto alegre: Artmed, 2006.

SOUZA, R.C.; CAMARGO, M.E. *Análise e Previsão de Séries Temporais: Os Modelos ARIMA*. 2ed. Rio de Janeiro: Gráfica e Editora Regional, 2004.

SOUZA, R.C.; MARCATO, A.L.; DIAS, B.H.; OLIVEIRA, F.L. Optimal Operation of Hydrothermal Systems with Hydrological Scenario Generation through Bootstrap and Periodic Autorregressive Models. *European Journal of Operational Research*, v.222, n.3, p.606-615, 2012.

STEDINGER, J.R. *Report on the Evaluation of CEPEL's PAR Models*. Ithaca, New York: School of Civil and Environmental Engineering, Cornell University, 2001.

TASWELL, C. The what, how, and why of wavelet shrinkage denoising. *Computing in Science & Engineering*, v.2, n.3, p.12-19, 2000.

TEIXEIRA JÚNIOR, L.A.; PESSANHA, J.F.M.; SOUZA, R.C. *Análise wavelet e redes neurais artificiais na previsão da velocidade de vento*. In: Simpósio Brasileiro de Pesquisa Operacional. Ubatuba, 2011.

TEIXEIRA JÚNIOR, L.A.; PESSANHA, J.F.M.; MENEZES, M.L.; CASSIANO, K.M.; SOUZA, R.C. *Redes Neurais Artificiais e Decomposição Wavelet na Previsão da Radiação Solar Direta*. In: Simpósio Brasileiro de Pesquisa Operacional. Rio de Janeiro, 2012.

THOMAS, H.A.; FIERING, M.B. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In: Maas, A., Hufschmidt, M.M., Dorfman, R., Thomas, H.A., Marglin, S.A., Far, G.M., (Orgs.) *Design of Water Resource Systems*, p.459-463, 1962.

TOLMASQUIM, M.T. *Novo Modelo do Setor Elétrico Brasileiro*. Rio de Janeiro: Synergia, 2011.

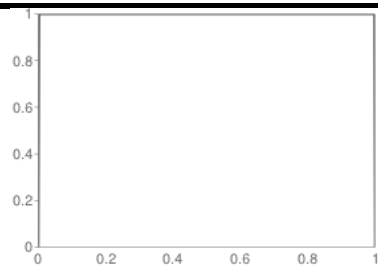
TORRES, M.E.L.; ALMEIDA, P.E.F. Cap.2 – Minas Gerais do Século XXI. In: _____. *Infra-estrutura: sustentando o desenvolvimento*. BDMG, 2003.

TOSCANO, A.E. Comparação entre os modelos NEWAVE e ODIN no planejamento energético do Sistema Interligado Nacional. 2009. 116f. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica, Universidade Estadual de Campinas, Campinas, 2009.

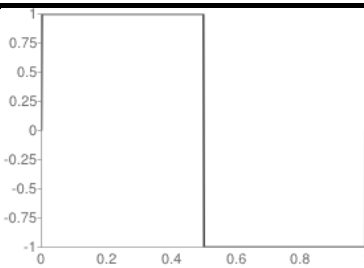
WAVELET BROWSER – *Discrete Wavelet Transform in Python. Wavelet Properties Browser*, 2012. Disponível em <http://wavelets.pybytes.com/>. Acesso em 30 abr., 2012.

Anexos

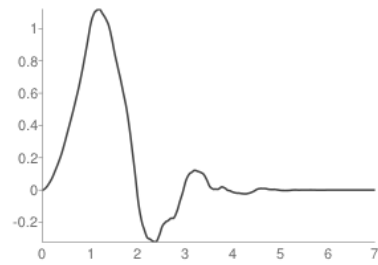
Anexo 1 – Curvas de funções wavelets



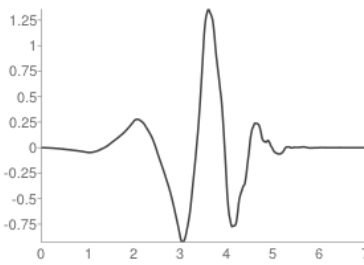
(a) Função Escala de Haar



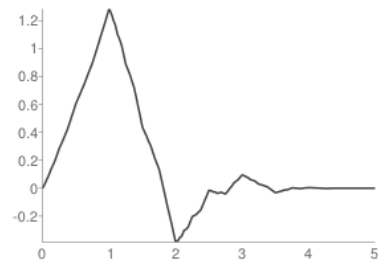
(b) Função Wavelet de Haar



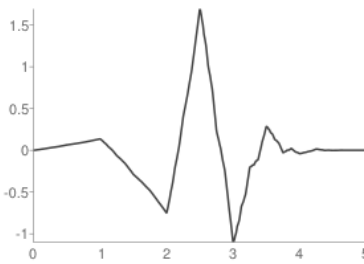
(c) Função Escala de Daubechies4



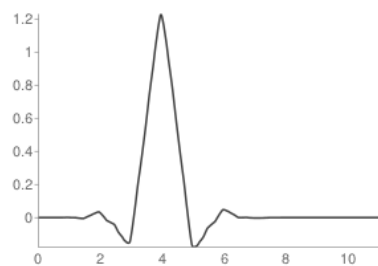
(d) Função Wavelet de Daubechies 4



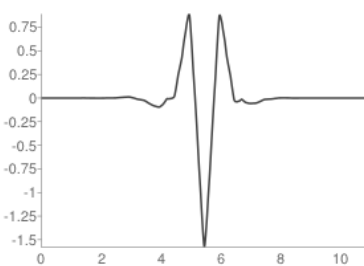
(e) Função Escala de Symlets3



(f) Função Wavelet de Symlets 3



(g) Função Escala de Coiflets 2



(h) Função Wavelet de Coiflets2

Fonte: Wavelet Browser (2012).

Anexo 2 – Testes de resíduos (p-valores) para diferentes setups da série temporal completa referente ao caso 1

Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
2	Symlet3	suave	universal	0.00	0.03	0.00
2	Haar	suave	universal	0.00	0.00	0.00
2	Coiflet2	suave	universal	0.00	0.00	0.00
2	Daubechies15	suave	universal	0.00	0.01	0.06
2	Daubechies30	suave	universal	0.00	0.32	0.00
2	Daubechies45	suave	universal	0.00	0.48	0.00
3	Symlet3	suave	universal	0.00	0.07	0.00
3	Haar	suave	universal	0.00	0.35	0.00
3	Coiflet2	suave	universal	0.00	0.00	0.00
3	Daubechies15	suave	universal	0.00	0.03	0.00
3	Daubechies30	suave	universal	0.00	0.34	0.00
3	Daubechies45	suave	universal	0.00	0.05	0.00
4	Symlet3	suave	universal	0.00	0.14	0.00
4	Haar	suave	universal	0.00	0.43	0.00
4	Coiflet2	suave	universal	0.00	0.01	0.00
4	Daubechies15	suave	universal	0.00	0.04	0.00
4	Daubechies30	suave	universal	0.00	0.08	0.00
4	Daubechies45	suave	universal	0.00	0.02	0.01
2	Symlet3	duro	universal	0.00	0.00	0.02
2	Haar	duro	universal	0.00	0.02	0.71
2	Coiflet2	duro	universal	0.00	0.01	0.32
2	Daubechies15	duro	universal	0.00	0.02	0.36
2	Daubechies30	duro	universal	0.00	0.23	0.98
2	Daubechies45	duro	universal	0.00	0.03	0.50
3	Symlet3	duro	universal	0.00	0.01	0.00
3	Haar	duro	universal	0.00	0.50	0.73
3	Coiflet2	duro	universal	0.00	0.01	0.01
3	Daubechies15	duro	universal	0.00	0.00	0.15
3	Daubechies30	duro	universal	0.00	0.05	0.18
3	Daubechies45	duro	universal	0.00	0.02	0.53
4	Symlet3	duro	universal	0.00	0.03	0.00
4	Haar	duro	universal	0.00	0.50	0.49
4	Coiflet2	duro	universal	0.00	0.01	0.01
4	Daubechies15	duro	universal	0.00	0.02	0.17
4	Daubechies30	duro	universal	0.00	0.13	0.48

Continua

Continuação

Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
4	Daubechies45	duro	universal	0.00	0.06	0.90
2	Symlet3	suave	minimax	0.00	0.00	0.00
2	Haar	suave	minimax	0.00	0.00	0.00
2	Coiflet2	suave	minimax	0.00	0.00	0.00
2	Daubechies15	suave	minimax	0.00	0.00	0.01
2	Daubechies30	suave	minimax	0.00	0.12	0.01
2	Daubechies45	suave	minimax	0.00	0.10	0.00
3	Symlet3	suave	minimax	0.00	0.00	0.00
3	Haar	suave	minimax	0.00	0.01	0.00
3	Coiflet2	suave	minimax	0.00	0.00	0.00
3	Daubechies15	suave	minimax	0.00	0.00	0.00
3	Daubechies30	suave	minimax	0.00	0.11	0.00
3	Daubechies45	suave	minimax	0.00	0.01	0.00
4	Symlet3	suave	minimax	0.00	0.00	0.00
4	Haar	suave	minimax	0.00	0.04	0.00
4	Coiflet2	suave	minimax	0.00	0.00	0.00
4	Daubechies15	suave	minimax	0.00	0.01	0.00
4	Daubechies30	suave	minimax	0.00	0.08	0.00
4	Daubechies45	suave	minimax	0.00	0.01	0.01
2	Symlet3	duro	minimax	0.00	0.50	0.00
2	Haar	duro	minimax	0.00	0.28	0.01
2	Coiflet2	duro	minimax	0.00	0.50	0.25
2	Daubechies15	duro	minimax	0.00	0.44	0.45
2	Daubechies30	duro	minimax	0.00	0.50	0.00
2	Daubechies45	duro	minimax	0.00	0.50	0.18
3	Symlet3	duro	minimax	0.00	0.45	0.17
3	Haar	duro	minimax	0.00	0.08	0.04
3	Coiflet2	duro	minimax	0.00	0.39	0.56
3	Daubechies15	duro	minimax	0.00	0.33	0.77
3	Daubechies30	duro	minimax	0.00	0.27	0.61
3	Daubechies45	duro	minimax	0.00	0.50	0.73
4	Symlet3	duro	minimax	0.00	0.50	0.08
4	Haar	duro	minimax	0.00	0.13	0.15
4	Coiflet2	duro	minimax	0.00	0.19	0.68
4	Daubechies15	duro	minimax	0.00	0.15	0.43
4	Daubechies30	duro	minimax	0.00	0.34	0.34
4	Daubechies45	duro	minimax	0.00	0.50	0.94
2	Symlet3	suave	SURE	0.00	0.00	0.00
2	Haar	suave	SURE	0.00	0.00	0.00

Continua

Continuação

Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
2	Coiflet2	suave	SURE	0.00	0.00	0.00
2	Daubechies15	suave	SURE	0.00	0.00	0.00
2	Daubechies30	suave	SURE	0.00	0.00	0.00
2	Daubechies45	suave	SURE	0.00	0.05	0.00
3	Symlet3	suave	SURE	0.00	0.00	0.00
3	Haar	suave	SURE	0.00	0.00	0.00
3	Coiflet2	suave	SURE	0.00	0.00	0.00
3	Daubechies15	suave	SURE	0.00	0.00	0.00
3	Daubechies30	suave	SURE	0.00	0.00	0.00
3	Daubechies45	suave	SURE	0.00	0.01	0.00
4	Symlet3	suave	SURE	0.00	0.00	0.00
4	Haar	suave	SURE	0.00	0.00	0.00
4	Coiflet2	suave	SURE	0.00	0.00	0.00
4	Daubechies15	suave	SURE	0.00	0.00	0.00
4	Daubechies30	suave	SURE	0.00	0.00	0.00
4	Daubechies45	suave	SURE	0.00	0.01	0.00
2	Symlet3	duro	SURE	0.00	0.00	0.00
2	Haar	duro	SURE	0.00	0.00	0.00
2	Coiflet2	duro	SURE	0.00	0.00	0.00
2	Daubechies15	duro	SURE	0.00	0.00	0.00
2	Daubechies30	duro	SURE	0.00	0.00	0.00
2	Daubechies45	duro	SURE	0.00	0.00	0.00
3	Symlet3	duro	SURE	0.00	0.00	0.00
3	Haar	duro	SURE	0.00	0.00	0.00
3	Coiflet2	duro	SURE	0.00	0.00	0.00
3	Daubechies15	duro	SURE	0.00	0.00	0.00
3	Daubechies30	duro	SURE	0.00	0.00	0.00
3	Daubechies45	duro	SURE	0.00	0.00	0.00
4	Symlet3	duro	SURE	0.00	0.00	0.00
4	Haar	duro	SURE	0.00	0.00	0.00
4	Coiflet2	duro	SURE	0.00	0.00	0.00
4	Daubechies15	duro	SURE	0.00	0.00	0.00
4	Daubechies30	duro	SURE	0.00	0.00	0.00
4	Daubechies45	duro	SURE	0.00	0.00	0.00

Fonte: Elaboração própria.

Anexo 3 – Testes de resíduos (p-valores) para diferentes setups da série temporal completa referente ao caso 2

Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
2	Symlet3	suave	universal	0.00	0.00	0.00
2	Haar	suave	universal	0.00	0.00	0.00
2	Coiflet2	suave	universal	0.00	0.00	0.00
2	Daubechies15	suave	universal	0.00	0.01	0.00
2	Daubechies30	suave	universal	0.00	0.12	0.01
2	Daubechies45	suave	universal	0.00	0.02	0.01
3	Symlet3	suave	universal	0.00	0.00	0.00
3	Haar	suave	universal	0.00	0.01	0.00
3	Coiflet2	suave	universal	0.00	0.00	0.00
3	Daubechies15	suave	universal	0.00	0.05	0.00
3	Daubechies30	suave	universal	0.00	0.49	0.01
3	Daubechies45	suave	universal	0.00	0.14	0.01
4	Symlet3	suave	universal	0.00	0.01	0.00
4	Haar	suave	universal	0.00	0.01	0.00
4	Coiflet2	suave	universal	0.00	0.01	0.00
4	Daubechies15	suave	universal	0.00	0.12	0.01
4	Daubechies30	suave	universal	0.00	0.50	0.05
4	Daubechies45	suave	universal	0.00	0.01	0.00
2	Symlet3	duro	universal	0.00	0.50	0.68
2	Haar	duro	universal	0.00	0.14	0.00
2	Coiflet2	duro	universal	0.00	0.06	0.08
2	Daubechies15	duro	universal	0.00	0.50	0.58
2	Daubechies30	duro	universal	0.00	0.50	0.10
2	Daubechies45	duro	universal	0.00	0.50	0.11
3	Symlet3	duro	universal	0.00	0.50	0.68
3	Haar	duro	universal	0.00	0.17	0.20
3	Coiflet2	duro	universal	0.00	0.07	0.39
3	Daubechies15	duro	universal	0.00	0.50	0.77
3	Daubechies30	duro	universal	0.00	0.43	0.30
3	Daubechies45	duro	universal	0.00	0.50	0.86
4	Symlet3	duro	universal	0.00	0.50	0.92
4	Haar	duro	universal	0.00	0.11	0.26
4	Coiflet2	duro	universal	0.00	0.06	0.43
4	Daubechies15	duro	universal	0.00	0.49	0.47
4	Daubechies30	duro	universal	0.00	0.13	0.37

Continua

Continuação

Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
4	Daubechies45	duro	universal	0.00	0.50	0.82
2	Symlet3	suave	minimax	0.00	0.00	0.00
2	Haar	suave	minimax	0.00	0.00	0.00
2	Coiflet2	suave	minimax	0.00	0.00	0.00
2	Daubechies15	suave	minimax	0.00	0.00	0.00
2	Daubechies30	suave	minimax	0.00	0.06	0.00
2	Daubechies45	suave	minimax	0.00	0.02	0.00
3	Symlet3	suave	minimax	0.00	0.00	0.00
3	Haar	suave	minimax	0.00	0.00	0.00
3	Coiflet2	suave	minimax	0.00	0.00	0.00
3	Daubechies15	suave	minimax	0.00	0.02	0.00
3	Daubechies30	suave	minimax	0.00	0.34	0.00
3	Daubechies45	suave	minimax	0.00	0.17	0.00
4	Symlet3	suave	minimax	0.00	0.00	0.00
4	Haar	suave	minimax	0.00	0.00	0.00
4	Coiflet2	suave	minimax	0.00	0.00	0.00
4	Daubechies15	suave	minimax	0.00	0.06	0.00
4	Daubechies30	suave	minimax	0.00	0.50	0.03
4	Daubechies45	suave	minimax	0.00	0.12	0.00
2	Symlet3	duro	minimax	0.00	0.00	0.04
2	Haar	duro	minimax	0.00	0.32	0.00
2	Coiflet2	duro	minimax	0.00	0.18	0.00
2	Daubechies15	duro	minimax	0.00	0.40	0.00
2	Daubechies30	duro	minimax	0.00	0.50	0.00
2	Daubechies45	duro	minimax	0.00	0.14	0.15
3	Symlet3	duro	minimax	0.00	0.04	0.43
3	Haar	duro	minimax	0.00	0.39	0.00
3	Coiflet2	duro	minimax	0.00	0.19	0.06
3	Daubechies15	duro	minimax	0.00	0.50	0.23
3	Daubechies30	duro	minimax	0.00	0.17	0.74
3	Daubechies45	duro	minimax	0.00	0.50	0.77
4	Symlet3	duro	minimax	0.00	0.05	0.53
4	Haar	duro	minimax	0.00	0.25	0.01
4	Coiflet2	duro	minimax	0.00	0.44	0.12
4	Daubechies15	duro	minimax	0.00	0.28	0.28
4	Daubechies30	duro	minimax	0.00	0.48	0.65
4	Daubechies45	duro	minimax	0.00	0.50	0.85
2	Symlet3	suave	SURE	0.00	0.00	0.00

Continua

Continuação

Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
2	Haar	suave	SURE	0.00	0.00	0.00
2	Coiflet2	suave	SURE	0.00	0.00	0.00
2	Daubechies15	suave	SURE	0.00	0.00	0.00
2	Daubechies30	suave	SURE	0.00	0.01	0.00
2	Daubechies45	suave	SURE	0.00	0.01	0.00
3	Symlet3	suave	SURE	0.00	0.00	0.00
3	Haar	suave	SURE	0.00	0.00	0.00
3	Coiflet2	suave	SURE	0.00	0.00	0.00
3	Daubechies15	suave	SURE	0.00	0.00	0.00
3	Daubechies30	suave	SURE	0.00	0.04	0.00
3	Daubechies45	suave	SURE	0.00	0.03	0.00
4	Symlet3	suave	SURE	0.00	0.00	0.00
4	Haar	suave	SURE	0.00	0.00	0.00
4	Coiflet2	suave	SURE	0.00	0.00	0.00
4	Daubechies15	suave	SURE	0.00	0.00	0.00
4	Daubechies30	suave	SURE	0.00	0.07	0.00
4	Daubechies45	suave	SURE	0.00	0.01	0.00
2	Symlet3	duro	SURE	0.00	0.00	0.00
2	Haar	duro	SURE	0.00	0.00	0.00
2	Coiflet2	duro	SURE	0.00	0.00	0.00
2	Daubechies15	duro	SURE	0.00	0.00	0.00
2	Daubechies30	duro	SURE	0.00	0.00	0.00
2	Daubechies45	duro	SURE	0.00	0.00	0.00
3	Symlet3	duro	SURE	0.00	0.00	0.00
3	Haar	duro	SURE	0.00	0.00	0.00
3	Coiflet2	duro	SURE	0.00	0.00	0.00
3	Daubechies15	duro	SURE	0.00	0.00	0.00
3	Daubechies30	duro	SURE	0.00	0.00	0.00
3	Daubechies45	duro	SURE	0.00	0.00	0.00
4	Symlet3	duro	SURE	0.00	0.00	0.00
4	Haar	duro	SURE	0.00	0.00	0.00
4	Coiflet2	duro	SURE	0.00	0.00	0.00
4	Daubechies15	duro	SURE	0.00	0.00	0.00
4	Daubechies30	duro	SURE	0.00	0.00	0.00
4	Daubechies45	duro	SURE	0.00	0.00	0.00

Fonte: Elaboração própria.

Anexo 4 – Wavelets utilizadas no encolhimento das séries de meses referentes ao caso 3 e testes de resíduos (p-valores)

Mês	Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
1	4	Coiflet1	suave	Minimax	0.30	0.32	0.20
2	3	Daubechies13	suave	SURE	0.23	0.50	0.83
3	4	Symlet4	suave	SURE	0.47	0.50	0.15
4	4	Daubechies38	suave	SURE	0.41	0.46	0.55
5	3	Daubechies40	suave	SURE	0.17	0.17	0.73
6	4	Daubechies31	suave	SURE	0.12	0.50	0.61
7	4	Coiflet4	suave	SURE	0.25	0.41	0.17
8	4	Daubechies15	suave	SURE	0.54	0.15	0.18
9	4	Symlet8	suave	Minimax	0.18	0.14	0.37
10	4	Daubechies38	suave	SURE	0.12	0.50	0.83
11	2	Daubechies20	suave	SURE	0.12	0.19	0.49
12	3	Symlet4	suave	Minimax	0.68	0.14	0.32

Fonte: Elaboração própria.

Anexo 5 – Wavelets utilizadas no encolhimento das séries de meses referentes ao caso 4 e testes de resíduos (p-valores)

Mês	Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
1	3	Daubechies6	suave	Minimax	0.14	0.27	0.33
2	4	Daubechies20	suave	SURE	0.38	0.49	0.11
3	4	Daubechies17	suave	SURE	0.32	0.37	0.40
4	3	Symlet7	suave	Minimax	0.43	0.17	0.59
5	4	Symlet8	suave	Minimax	0.29	0.50	0.75
6	2	Daubechies42	suave	SURE	0.14	0.44	0.97
7	3	Daubechies12	suave	SURE	0.15	0.41	0.91
8	4	Daubechies13	suave	SURE	0.42	0.14	0.47
9	3	Daubechies15	suave	SURE	0.59	0.50	0.39
10	4	Daubechies38	suave	SURE	0.16	0.14	0.26
11	3	Daubechies28	suave	SURE	0.21	0.50	0.68
12	3	Daubechies8	suave	SURE	0.43	0.36	0.16

Fonte: Elaboração própria.

Anexo 6 – Wavelets utilizadas no encolhimento das séries de meses da ENA Sudeste/Centro-Oeste e testes de resíduos (p-valores)

Mês	Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
Janeiro	4	Daubechies5	suave	SURE	0.93	0.50	0.49
Fevereiro	4	Daubechies8	suave	SURE	0.05	0.35	0.10
Março	4	Daubechies7	suave	SURE	0.43	0.23	0.63
Abril	4	Daubechies4	suave	SURE	0.48	0.14	0.51
Maio	4	Symlet6	suave	SURE	0.43	0.13	0.22
Junho	2	Daubechies11	suave	Minimax	0.11	0.50	0.26
Julho	4	Daubechies37	suave	SURE	0.06	0.22	0.85
Agosto	4	Daubechies6	suave	SURE	0.11	0.19	0.60
Setembro	4	Symlet2	suave	SURE	0.10	0.26	0.33
Outubro	4	Daubechies12	suave	SURE	0.07	0.50	0.39
Novembro	4	Daubechies28	suave	SURE	0.10	0.50	0.29
Dezembro	4	Symlet3	suave	SURE	0.17	0.48	0.90

Fonte: Elaboração própria.

Anexo 7 – Wavelets utilizadas no encolhimento das séries de meses da ENA Sul e testes de resíduos (p-valores)

Mês	Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
Janeiro	4	Symlet2	suave	SURE	0.11	0.22	0.26
Fevereiro	4	Daubechies5	suave	SURE	0.68	0.17	0.36
Março	3	Daubechies9	suave	SURE	0.11	0.27	0.11
Abril	3	Daubechies43	suave	SURE	0.05	0.06	0.71
Maio	3	Daubechies18	suave	SURE	0.06	0.06	0.92
Junho	2	Daubechies4	suave	SURE	0.21	0.25	0.33
Julho	3	Daubechies13	suave	SURE	0.09	0.45	0.51
Agosto	4	Daubechies10	suave	SURE	0.14	0.09	0.43
Setembro	4	Symlet8	suave	SURE	0.45	0.08	0.36
Outubro	4	Daubechies4	suave	Universal	0.31	0.45	0.72
Novembro	4	Daubechies42	suave	SURE	0.08	0.10	0.07
Dezembro	4	Daubechies4	suave	SURE	0.07	0.36	0.13

Fonte: Elaboração própria.

Anexo 8 – Wavelets utilizadas no encolhimento das séries de meses da ENA Nordeste e testes de resíduos (p-valores)

Mês	Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
Janeiro	4	Daubechies7	suave	SURE	0.09	0.11	0.15
Fevereiro	3	Daubechies13	suave	SURE	0.06	0.50	0.42
Março	4	Daubechies36	suave	SURE	0.44	0.09	0.79
Abril	4	Daubechies16	suave	SURE	0.17	0.14	0.49
Maio	4	Symlet5	suave	SURE	0.06	0.50	0.64
Junho	4	Symlet4	suave	SURE	0.31	0.28	0.19
Julho	4	Symlet5	suave	Universal	0.71	0.50	0.29
Agosto	4	Daubechies12	suave	SURE	0.28	0.22	0.72
Setembro	4	Daubechies14	suave	SURE	0.10	0.50	0.60
Outubro	4	Daubechies30	suave	SURE	0.69	0.50	0.56
Novembro	4	Symlet8	suave	SURE	0.14	0.07	0.33
Dezembro	4	Daubechies9	suave	SURE	0.09	0.27	0.49

Fonte: Elaboração própria.

Anexo 9 – Wavelets utilizadas no encolhimento das séries de meses da ENA Norte e testes de resíduos (p-valores)

Mês	Camadas	Família	Limiar	Regra de Limiarização	Ljung-Box	Jarque-Bera	Teste ARCH
Janeiro	4	Daubechies27	suave	SURE	0.07	0.14	0.50
Fevereiro	2	Daubechies8	suave	SURE	0.23	0.22	0.20
Março	3	Daubechies15	suave	SURE	0.16	0.38	0.23
Abril	3	Daubechies20	suave	SURE	0.12	0.24	0.96
Maio	4	Daubechies22	suave	SURE	0.69	0.07	0.62
Junho	4	Daubechies5	suave	SURE	0.29	0.49	0.41
Julho	4	Daubechies43	suave	Universal	0.58	0.50	0.55
Agosto	4	Daubechies16	suave	SURE	0.15	0.43	0.10
Setembro	3	Daubechies2	suave	Universal	0.24	0.09	0.69
Outubro	3	Symlet2	suave	SURE	0.30	0.08	0.47
Novembro	3	Daubechies6	suave	SURE	0.28	0.22	0.89
Dezembro	3	Daubechies26	suave	SURE	0.05	0.50	0.44

Fonte: Elaboração própria.