

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO

Patricia Cristiane da Cunha Xavier

**PREVISÃO DO PREÇO DO CIMENTO PORTLAND E DO AÇO NO RIO
GRANDE DO SUL POR MEIO DE MODELAGEM *WAVELET-ARIMA***

Santa Maria, RS
2017

Patricia Cristiane da Cunha Xavier

**PREVISÃO DO PREÇO DO CIMENTO PORTLAND E DO AÇO NO RIO GRANDE
DO SUL POR MEIO DE MODELAGEM *WAVELET-ARIMA***

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção, Área de Concentração em Gerência da Produção, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Mestre em Engenharia de Produção**.

Orientador: Prof. Dr. Adriano Mendonça Souza

Santa Maria, RS
2017

Ficha catalográfica elaborada através do Programa de Geração Automática da Biblioteca Central da UFSM, com os dados fornecidos pelo(a) autor(a).

Xavier, Patricia Cristiane da Cunha
PREVISÃO DO PREÇO DO CIMENTO PORTLAND E DO AÇO NO RIO
GRANDE DO SUL POR MEIO DE MODELAGEM WAVELET-ARIMA /
Patricia Cristiane da Cunha Xavier.- 2017.
74 f. ; 30 cm

Orientador: Adriano Mendonça Souza
Dissertação (mestrado) - Universidade Federal de Santa
Maria, Centro de Tecnologia, Programa de Pós-Graduação em
Engenharia de Produção, RS, 2017

1. Transformada Wavelet de Haar 2. Análise
Multiresolução 3. Threshold 4. Modelos de Séries
Temporais I. Souza, Adriano Mendonça II. Título.

Patricia Cristiane da Cunha Xavier

PREVISÃO DO PREÇO DO CIMENTO PORTLAND E DO AÇO NO RIO GRANDE DO SUL POR MEIO DE MODELAGEM WAVELET- ARIMA

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção, Área de Concentração em Gerência da Produção, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Mestre em Engenharia de Produção**.


Aprovada em 21 de agosto de 2017:



Adriano Mendonça Souza, Dr. (UFSM)
(Presidente/Orientador)



Maria Lúcia Pozzatti Flores, Dr^a. (UNIPAMPA)



Luciane Flores Jacobi, Dr^a. (UFSM)

Santa Maria, RS
2017

DEDICATÓRIA

A minha avó Neli Júlia, pelo zelo, amor incondicional, investimento nos meus estudos e orações para que todos os meus sonhos se concretizem.

AGRADECIMENTOS

A realização deste trabalho foi possível graças à compreensão e dedicação de várias pessoas. Agradeço a todos que de alguma forma contribuíram para a conclusão deste estudo, em especial:

- a Deus, por ser minha fortaleza e socorro bem presente;
- à minha avó, Neli Júlia Cunha, pelo apoio e incentivo;
- à minha filha, Helena da Cunha Cúrcio, que é meu orgulho, minha vida e meu amor infinito, pela compreensão nos momentos em que estive ausente;
- à minha mãe, Sonia Cunha, que deu o suporte, para o tempo render;
- à minha tia, Emília Cunha, a quem eu devo desde sempre, a investidura no meu saber;
- ao meu orientador, Prof. Dr. Adriano Mendonça Souza, sempre empenhado no sucesso de seus orientandos;
- à Prof^a. Dr^a. Janice Rachelli, por ser desde sempre minha referência de profissional, pelas muitas horas de instrução, ferramentas e positivismo;
- à Prof^a. Dr^a. Alice Kozakevicius, que doou seu tempo desmedidamente para a concretização deste sonho e com conselhos me conduziu a transpor barreiras e acreditar em mim;
- ao Prof. Dr. Júlio Cezar Mairesse Siluk, coordenador do PPGEP, por acreditar em mim;
- ao meu mestre Prof. Hélio Ricardo Friedrich, como um pedido de perdão por não ter tido a maturidade na oportunidade dada;
- aos meus colegas, em especial a Jean Cauê Huppés e Vitor Hugo de Souza;
- aos meus alunos, todos amados, que fizeram acreditar a cada dia que eu podia e merecia escrever, em especial a Leonam Hemann e Diego Dariva;

Emprego a todos, meu apreço e consideração.

Muito obrigada!

Aprender é a única coisa que a mente
nunca se cansa, nunca tem medo e
nunca se arrepende.

(Leonardo da Vinci)

RESUMO

PREVISÃO DO PREÇO DO CIMENTO PORTLAND E DO AÇO NO RIO GRANDE DO SUL POR MEIO DE MODELAGEM *WAVELET*- ARIMA

AUTORA: Patricia Cristiane da Cunha Xavier

ORIENTADOR: Adriano Mendonça Souza

Este trabalho apresenta um estudo sobre o comportamento dos preços de duas das mais importantes *commodities* do setor da construção civil no estado do Rio Grande do Sul. A partir deste, faz-se uma análise de cada uma das séries compostas por dados mensais de preços do cimento Portland e do aço em R\$/Kg, cujo objetivo é prever o preço de ambos no referido estado, por meio da modelagem *wavelet*-ARIMA, para auxiliar no planejamento do desempenho do setor. Afetadas diretamente pela instabilidade econômica, a indústria cimenteira e do aço recuam. Intensificam-se as quedas no nível de atividade em relação ao mês anterior, perdas dos postos de trabalho ganham força no Rio Grande do Sul. Nesse sentido, foram utilizados métodos de modelagem Autorregressivos Integrados de Médias Móveis – ARIMA em dados cujo período analisado está compreendido entre setembro/1995 a dezembro/2016, totalizando 256 observações referentes ao preço mensal do cimento Portland e um total de 128 observações, para o preço mensal do aço no período de fevereiro/2007 a setembro/2017 cuja fonte é a Câmara Brasileira da Indústria da Construção. Posteriormente, a série cujos dados são preços mensais em R\$/Kg das referidas *commodities* foi decomposta por *wavelets* de Haar em aproximações e detalhes. Tais aproximações são decompostas em diferentes níveis, e os detalhes já são em si coeficientes *wavelets* que, junto das funções escalas, irão reconstruir a nova série de dados. A decomposição por *wavelets* foi utilizada como forma de buscar uma melhor acurácia de previsão, ajustando os fatores de aproximação e detalhes ao modelo autorregressivo integrado de médias móveis. Assim, conclui-se que, no Rio Grande do Sul, o modelo que melhor representa a série do preço do aço é um ARIMA (1,1,0), e o modelo que melhor representa a série do preço do cimento Portland é um ARIMA (1,1,1).

Palavras-chave: Cimento. Aço. Previsão. Preço. Modelos. *Wavelet*.

ABSTRACT

PRICE FORECASTING OF PORTLAND CEMENT AND STEEL IN RIO GRANDE DO SUL USING THE *WAVELET*- ARIMA MODEL

AUTHOR: Patricia Cristiane da Cunha Xavier

ADVISOR: Adriano Mendonça Souza

This study is a price analysis of two important civil engineering *commodities* in Rio Grande do Sul. It is based on the examination of each series composed of monthly prices of Portland cement and steel in order to predict their prices in the state using the *wavelet*-ARIMA model in order to improve the sector development. Directly affected by the economic instability, the cement and steel industry decelerates. The activity level diminishes if compared to the previous month and there is less work opportunities mainly in Rio Grande do Sul. Considering this, this study focuses on the role of these materials in the civil engineering market, aiming at predicting the cement and the steel prices in Rio Grande do Sul. ARIMA methods were used which represent the stochastic process that generates the time series data whose source is the Brazilian Construction Industry. The analyzed period goes from September/1995 to December/2016, equals 256 observations of the Portland cement monthly price and 128 observations of the steel monthly price from February/2007 to September/2017. After this, the original data series is decomposed by Haar *wavelets* in approximations and details. The approximations are decomposed in different levels, the details are *wavelets* coefficients that with the scale functions reconstruct the new data series. In every discretization process there is loss. Thus, the transformed series is more accurate than the original, that is, the *wavelets* decomposition is used as a way to achieve a more accurate prediction when the approximation factors and the details are adjusted to the autoregressive integrated moving average. Finally, in Rio Grande do Sul, the model that best represents the steel price series is an ARIMA (1,1,0) and the model that best represents the Portland cement price series is a ARIMA (1,1,1).

Keywords: Cement. Steel. Forecast. Price. Models. *Wavelet*.

LISTA DE ILUSTRAÇÕES

Figura 1 -	Representação Gráfica da <i>wavelet</i> de Morlet.....	19
Figura 2 -	Comportamento das <i>wavelets</i>	20
Figura 3 -	Filtragem de um sinal S em primeiro nível de aproximações e detalhes.....	28
Figura 4 -	Ilustração do processo de decomposição de um sinal S em quatro níveis pela análise de multirresolução em que C_j e D_j , para $j=1, 2, 3, 4$ são ditos aproximações e detalhes respectivamente.....	28
Figura 5 -	Reconstrução de um sinal S.....	31
Figura 6 -	Fluxograma do ciclo iterativo de Box e Jenkins.....	36
Figura 7 -	Fluxograma com as etapas da pesquisa.....	44
Figura 8 -	Série original de preços mensais do cimento Portland no RS no período de setembro/1995 a dezembro/2016.....	45
Figura 9 -	Correlogramas FAC e FACP da série original de preços do cimento Portland em R\$/Kg (a), e em 1ª diferença (b).....	46
Figura 10	Resíduos da série diferenciada de preços mensais do cimento Portland no RS e Previsão segundo modelo ARIMA(1, 1, 1).....	48
Figura 11 -	Sinal original do preço do cimento decimado por função escala..	49
Figura 12 -	Decomposição da série de preços mensais do cimento por Haar em 7 níveis de aproximações (C) e detalhes (D).....	50
Figura 13 -	Decomposição da série preço do cimento Portland em cinco níveis.....	53
Figura 14 -	Decomposição da série preço do cimento Portland em cinco níveis com threshold em D_2	55
Figura 15 -	Comparação entre vetor_cimento (série original) e vetor_cimento reconstruído com <i>threshold</i>	56
Figura 16 -	Série original de preços do aço CA 50a d=10mm no RS no período de fevereiro/2007 a setembro/2017.....	58
Figura 17 -	Função de autocorrelação (FAC) e Função de autocorrelação parcial (FACP) da série de preços mensais do aço no RS.....	59
Figura 18 -	Resíduos da série diferenciada de preços mensais do aço no RS e Previsão segundo modelo ARIMA(1, 1, 0).....	60
Figura 19 -	Sinal original do aço decimado por funções escala.....	61
Figura 20 -	Decomposição da série de preços mensais do aço por Haar em 6 níveis de aproximações (C) e detalhes (D).....	61
Figura 21 -	Decomposição da série de preços do aço em R\$/Kg em cinco níveis	64
Figura 22 -	Decomposição da série preço do aço em cinco níveis com threshold em D_2	65
Figura 23 -	Comparação entre série original de preços do aço e vetor aço reconstruído por TWD inversa de Haar com <i>threshold</i>	66

LISTA DE TABELAS

Tabela 1	-	Formulação algorítmica para decomposição por TWD Haar.....	42
Tabela 2	-	Modelos concorrentes para a série preço do cimento Portland em R\$/Kg no estado do Rio Grande do Sul.....	47
Tabela 3	-	Modelos concorrentes para o sinal S reconstruído.....	53
Tabela 4	-	Formulação Algorítmica para a diferença entre sinal original e sinal reconstruído do preço do cimento Portland.....	54
Tabela 5	-	Modelos concorrentes para S reconstruído com <i>threshold</i>	57
Tabela 6	-	Modelos concorrentes para a série preço do aço em R\$/Kg no estado do Rio Grande do Sul.....	59
Tabela 7	-	Modelos concorrentes para a variável aço por <i>wavelet</i> -ARIMA sem <i>threshold</i>	64
Tabela 8	-	Formulação Algorítmica para a diferença entre vetor série original e vetor aço reconstruído por TWD inversa de Haar.....	66
Tabela 9	-	Modelos concorrentes para a variável aço por <i>wavelet</i> -ARIMA com <i>threshold</i>	67
Tabela 10	-	Modelos concorrentes para as variáveis cimento e aço por ARIMA e <i>wavelet</i> -ARIMA sem <i>threshold</i> e com <i>threshold</i>	67

LISTA DE ABREVIATURAS E SIGLAS

AIC	Akaike's Information Criterion
AMR	Análise de Multirresolução
AR	Autorregressivo
ARMA	Autorregressivo de Médias Móveis
ARFIMA	Autorregressivo Fracionário Integrado de Médias Móveis
ARIMA	Autorregressivo Integrado de Médias Móveis
BIC	Bayesian Information Criterion
CBIC	Câmara Brasileira da Indústria do Cimento
FAC	Função de Autocorrelação
FACP	Função de Autocorrelação Parcial
FGV	Fundação Getúlio Vargas
FIERGS	Federação das Indústrias do Estado do Rio Grande do Sul
GARCH	Generalização do Modelo Autorregressivo para a Heterocedasticidade Condicional (ARCH)
IBRE	Instituto Brasileiro de Economia
IPA	Índice de Preços ao Produtor Amplo
INDA	Instituto Nacional dos Distribuidores do Aço
MA	Médias Móveis
MAPE	Erro Percentual Absoluto Médio
MMQ	Método dos mínimos quadrados
RMSE	Raiz do Erro Médio Quadrático
SARIMA	Médias Móveis Integradas Autorregressivas Sazonais
SINDUSCON	Sindicato da Indústria da Construção Civil
SNIC	Sindicato Nacional da Indústria do Cimento
ST	Séries Temporais
TW	Transformada <i>Wavelet</i>
TWD	Transformada <i>Wavelet</i> Discreta
U-Theil	<i>U</i> estatística de Theil

LISTA DE SÍMBOLOS

a, a_0, j	Parâmetros de dilatação
b, b_0, k	Parâmetros de translação ou compressão
C_j	Coefficientes de aproximação
$D_j, D_{j,k}$	Coefficientes de detalhes ou coeficientes <i>wavelets</i>
$f(t)$	Função qualquer
$L^2(R)$	Espaço das funções quadrado integráveis
V_m	Subespaços da Análise Multirresolução
W_m	Subespaços <i>Wavelets</i>
$\phi(t)$	Funções escalas
$\ \cdot \ $	Norma
$\langle \cdot \rangle$	Produto Interno
\oplus	Soma ortogonal
$\hat{\phi}(u)$	Transformada de Fourier
$\psi_{j,k}(t), \psi_{a,b}(t)$	Funções <i>wavelets</i> comprimidas e transladadas
$\psi(t)$	Função wavelet de Haar
λ	Limiar de corte ou <i>Threshold</i>
thr_λ^S	<i>Soft Threshold</i>
S	Vetor cujas coordenadas são obtidas pela reconstrução de um sinal por TWD inversa
n	Número de previsões realizadas
d	Número de diferenciações
p	Ordem do modelo autorregressivo
q	Ordem do modelo média móvel
B	Operador de retardo
a_t	Ruído Branco
$\sigma_{a_t}^2$	Variância do resíduo
θ	Parâmetro autorregressivo de média móvel
ϕ	Parâmetro autorregressivo
Φ	Parâmetro sazonal
d	Parâmetro de valor fracionário
μ	Média
$N(0, \sigma^2)$	Distribuição Normal
Z_t	Série original ou valor real no instante t
\hat{Z}_t	Valor previsto no instante t
$\hat{\sigma}_{p,q}^2$	Estimador de Máxima Verossimilhança
N	Tamanho da amostra
R	Conjunto dos números reais

SUMÁRIO

1	INTRODUÇÃO	13
1.1	PROBLEMA.....	15
1.2	OBJETIVOS.....	16
1.1.1	Objetivo Geral	16
1.1.2	Objetivos Específicos	16
1.3	JUSTIFICATIVA.....	17
1.4	ESTRUTURA DO TRABALHO.....	18
2	REVISÃO BIBLIOGRÁFICA	19
2.1	TRANSFORMADA <i>WAVELET</i> DE HAAR.....	19
2.1.1	Transformada <i>Wavelet</i> Discreta	23
2.1.2	Análise Multirresolução e Filtragem (AMR)	27
2.1.3	Limiarização da <i>Wavelet</i> (<i>Threshold</i>)	29
2.1.4	Processo de Reconstrução	30
2.2	MODELOS DE SÉRIES TEMPORAIS.....	32
3	MATERIAIS E MÉTODOS	41
3.1	COLETA DE DADOS.....	41
3.2	ANÁLISE.....	41
3.3	ETAPAS METODOLÓGICAS DO ESTUDO.....	43
4	RESULTADOS E DISCUSSÕES	45
4.1	MODELAGEM DA SÉRIE PREÇO DO CIMENTO PORTLAND EM R\$/Kg.....	45
4.2	MODELAGEM DA SÉRIE PREÇO DO AÇO CA 50A D=10mm EM R\$/Kg.....	57
5	CONCLUSÃO	69
	REFERÊNCIAS	70

1 INTRODUÇÃO

O custo da produção industrial de insumos da construção civil está atrelado a despesas com combustíveis, derivados do petróleo em geral e energia elétrica, o que acarreta efeitos em cadeia com a variação nos preços desses insumos. A saber, há uma gama muito grande de produtos, insumos da construção civil, com características e condições de mercado distintas.

Neste estudo, porém, trata-se das *commodities* cimento e aço, que apresentam características e condições de mercado bastante parecidas.

Assim como a indústria cimenteira desempenha um papel importante no desenvolvimento econômico e financeiro, a produção do aço é um forte indicador do estágio de desenvolvimento econômico de um país (SINDICATO NACIONAL DA INDÚSTRIA DO CIMENTO (SNIC), 2016; AÇO BRASIL, 2016).

Há uma retração nas indústrias de *commodities* cimento e aço ao serem, afetadas diretamente pela instabilidade econômica e, conseqüentemente, o setor da construção civil também sofre essa retração.

O impacto da crise econômica do país atingiu os setores de edificação residencial, comercial, industrial e de obras de infraestrutura. O mercado de cimento fechou o volume de vendas em outubro de 2016 com 4,63 milhões de toneladas, enquanto no mesmo mês do ano anterior o volume atingia 5,65 milhões de toneladas. Verifica-se, portanto, uma queda de 18,1% (CIMENTOONLINE, 2016).

Com o aço não foi diferente. Segundo o Instituto Nacional dos Distribuidores do Aço - INDA (2016), as importações seguiam em queda, dificultadas pela falta de crédito e baixa demanda interna. Mas, no primeiro trimestre de 2017, o consumo do aço apresentou um crescimento de 5% em relação ao mesmo período de 2016, suprido pelo aumento das importações, que foi de 73,1%, enquanto as vendas internas tiveram uma queda de 0,5%, já que a produção apresentou um crescimento de 10,9% devido às exportações, que tiveram crescimento de 17,4%.

Segundo dados do Sindicato da Indústria da Construção Civil (SINDUSCON-RS), em 2017, houve uma queda de 0,73% no preço do aço, juntamente com o concreto, cuja queda foi de 0,65%, que tem como ingrediente fundamental o cimento (AÇO BRASIL, 2017). Embora o cimento seja um importante insumo da construção civil, a participação no custo desse setor é cerca de 3,83%, contra 45,72% de mão-de-obra e 50,45% de outros materiais (Fundação Getúlio Vargas, FGV, 2012).

Ainda, dados da FGV e do Instituto Brasileiro de Economia – IBRE (2012) mostram que a participação do cimento nos índices de preços é da ordem de 0,68%, sendo 0,50% do Índice de Preços ao Produtor Amplo – IPA.

Historicamente, a implantação da indústria cimenteira, no Brasil, deu-se no final do século XIX, alavancada pelo crescente desenvolvimento do país. Nessa época, arcando com tarifas de importação que extrapolavam custos, empreendedores brasileiros instalaram a referida indústria. Entretanto, a produção efetiva de cimento nacional deu-se a partir de 1926, embora com dificuldades de competição com o preço do cimento europeu, devido aos elevados custos de produção, empresários de várias partes do país conseguiram mostrar aos brasileiros que por aqui se produzia cimento de qualidade.

O grande estímulo para o setor cimenteiro surgiu a partir da década de 60, com o início das grandes obras de infraestrutura, como a ponte Rio-Niterói. De fato, a indústria cimenteira desempenha um papel importante no desenvolvimento econômico e financeiro.

No Rio Grande do Sul, a indústria cimenteira chegou em 1947. Nesse ano, foi inaugurada em São Leopoldo, a primeira unidade de fabricação de cimento, movida pelos investimentos significativos no setor de construção civil, obras públicas monumentais e forte ritmo de aceleração de crescimento.

A produção mundial do aço, por sua vez, teve seu marco nos anos 50 e 60. Nessa época, a demanda era de 6% ao ano, devido, principalmente, ao crescimento da procura por bens de consumo em geral, eletrodomésticos, carros etc. Atualmente, o mercado do aço é bastante afetado pelo comércio internacional. O crescimento da China é fator importante no desempenho do setor siderúrgico, visto que a demanda externa é um aspecto relevante no mercado brasileiro.

No atual estágio de desenvolvimento da sociedade, é impossível imaginar o mundo sem o uso do aço, visto que o aço é um indicador do estágio de desenvolvimento de um país. Em 2014, o Brasil foi responsável pela produção de 33,9 milhões de toneladas de aço bruto, o que colocou o país em 9ª posição no ranking mundial e em primeiro lugar na América Latina (AÇO BRASIL, 2014).

Segundo dados preliminares da indústria e estimativas de mercado, ambos fornecidos pelo sindicato nacional da indústria do cimento, as vendas internas de cimento em junho 2016 somaram 4,7 milhões de toneladas. Em relação a junho de 2015, houve uma queda de 14,7%. Comparando o primeiro semestre de 2015 com

igual período em 2016, as vendas acumuladas caíram 14%, alcançando 28 milhões de toneladas. Além disso, as vendas acumuladas apresentaram queda de (jul/15 a jun/16), 12,6% sobre igual período anterior (jul/14 a jun/15), atingindo apenas 60,2 milhões de toneladas.

Estudos divulgados pela Câmara Brasileira da Indústria do Cimento - CBIC (2008) indicam que o peso do preço do cimento no custo de uma obra residencial gira em torno de 7% a 9%, excedendo 3% a 5% os percentuais na participação de custo nesse setor como um todo, enquanto o preço do aço fica em 4,6%. Vale lembrar que este insumo está presente desde a base até os acabamentos e promove segurança e estabilidade estrutural na obra (AÇO BRASIL, 2014). Ainda, no Brasil, a indústria do cimento e artefatos de cimento participa com 2,55% da composição do macrossetor da construção, gera emprego e renda a aproximadamente 140 mil trabalhadores e movimenta em torno de R\$ 3 bilhões/ano.

Conforme exposto, juntos, cimento (2,55%) e aço (7%), correspondem a aproximadamente 10% do custo da obra final. E a previsibilidade se justifica, inclusive, como um instrumento de controle desse impacto em obras públicas, a fim de coibir superfaturamentos intencionais ou aditivos de valores que poupados geram menos “surpresas” na etapa de conclusão de uma obra.

Nesse sentido, verifica-se a importância dessas *commodities* para o setor da construção civil, assim como sua evolução em termos de preço, problemática desta pesquisa.

1.1 PROBLEMA

Diante da importância de tais *commodities*, as indústrias do cimento e do aço estão vivenciando oscilações no mercado, o que permite o questionamento sobre os valores que o cimento e o aço possam ter em determinadas escalas temporais.

No Rio Grande do Sul, a situação do setor da construção civil mostra quedas no nível de atividade em relação ao mês anterior, perda dos postos de trabalho e capacidade operacional, uma vez que a ociosidade no estado foi superior à do país em abril/2016. Em maio/2016, seguiu sinalizando queda de dois pontos percentuais. A compra de matérias-primas não contraria essa tendência, pois decaiu 4,2% em comparação com o que ocorreu no país como um todo (FIERGS, 2016).

Segundo a Aço Brasil (2017), os segmentos automotivos, produtos químicos, máquinas, equipamentos, eletrodomésticos, autopeças, siderurgia, entre outros, que representam 48,5% da produção e 68,5% das exportações da indústria do país, operam em média com 50% de ociosidade e nos dois últimos anos demitiram mais de 500 mil colaboradores. Ainda, em dados divulgados pelo SNIC (2016), o Rio Grande do Sul participa com 3,2% da produção de cimento do país, o que corresponde a 2,2 milhões de toneladas, ou seja, produz apenas 62,9% de suas necessidades. No intuito de fornecer subsídios para que sejam tomadas decisões no gerenciamento da produção de cimento e do aço no Rio Grande do Sul, e minimização dos impactos sociais e econômicos, a busca por uma melhor acurácia da previsão dos preços desses importantes *commodities*, é o que norteia essa pesquisa. Sendo assim, problematiza-se o estudo com a seguinte pergunta de pesquisa:

O uso da transformada *wavelet* discreta de Haar associada à modelagem ARIMA torna acurada a previsão de preços para as commodities cimento e aço?

Dessa forma, o tema da pesquisa é o uso da modelagem matemática, quer seja por meio dos modelos ARIMA ou Transformada *wavelet* discreta de Haar associada a ARIMA para previsão de preços do cimento Portland e do aço. Esse tema é relevante, pois, por meio do uso dessas duas metodologias de modelagem, busca-se a acurácia nas previsões.

1.2 OBJETIVOS

1.1.1 Objetivo Geral

Determinar o melhor método de previsão para o preço do cimento Portland e do aço no Rio Grande do Sul, por meio dos modelos ARIMA e *wavelet*-ARIMA.

1.1.2 Objetivos Específicos

- Ajustar os modelos de classe geral ARIMA para os dados do cimento Portland, o mais usado na construção civil, e do aço;

- Transformar os dados de cada uma das séries, utilizando a Transformada *wavelet* discreta de Haar, com o intuito de obter maior precisão ao modelo de previsão encontrado pelo ARIMA;
- Gerar subsídios para a tomada de decisão em função das previsões realizadas.

1.3 JUSTIFICATIVA

Ao acompanhar a literatura, percebem-se as diversas aplicações das *wavelets* e seus filtros passa-alta e passa-baixa, associadas com técnicas de previsão, bem como a existência de uma gama de formas de combinar modelos e filtros. Nessa perspectiva, pesquisadores, tais como Homsy, Portugal e Araújo (2000), têm apontado que não há um só esquema para a tomada de decisão, demonstrando o dinamismo dessa área ao se utilizar séries temporais, ou transformada *wavelet* discreta de Haar com séries temporais.

Assim, elege-se *wavelets* discretas porque são utilizadas para a decomposição e filtragem de qualquer série temporal. Sua aplicabilidade, nesse aspecto, advém do fato de que não provocam redundâncias de coeficientes entre escalas (frequências) (MEYER, 1990; DAUBECHIES, 1992).

Também, devido à capacidade da transformada *wavelet* produzir uma boa melhora em ambos os domínios de frequência e de tempo, opta-se pela sua utilização para decompor e reconstruir a série de preços de modo a buscar melhores valores de previsão, por meio de modelagem da série de Haar reconstruída.

Makridakis, Wheelwright e Hyndman (1998) afirmam que realizar previsões para os setores produtivos da economia é importante para auxiliar, tanto o governo como as empresas que compõem esse setor, na determinação dos recursos necessários.

Nesse sentido, busca-se um método que possibilite a previsão do preço do cimento Portland e do aço em estudo, de forma acurada. Por isso, testa-se o método de previsão convencional ARIMA e o método composto *wavelet*-ARIMA.

1.4 ESTRUTURA DO TRABALHO

No primeiro capítulo, consta uma introdução em que se aborda o tema, destacam-se a problemática, o objetivo geral e os objetivos específicos e a justificativa desta pesquisa. A revisão bibliográfica é apresentada no capítulo 2, em que, na primeira sessão, trata-se sobre a fundamentação teórica acerca das *wavelets*, como definições, equações, parâmetros, escalas, teoremas e suas aplicações. Ainda no capítulo 2, na segunda seção, consta o embasamento teórico para desenvolvimento da metodologia Box-Jenkins, critérios penalizadores, estatísticas de previsão adotadas nesta pesquisa, modelos concorrentes e determinação do melhor modelo. Em seguida, no capítulo 3, mostram-se materiais e métodos propostos para fazer a previsão do preço do cimento Portland e do preço do aço no Rio Grande do Sul, ou seja, a aplicação das técnicas apresentadas no capítulo 2. No capítulo 4, apresentam-se os resultados e discussões acerca da pesquisa. E, como fechamento, no último capítulo desta dissertação, apresenta-se a conclusão.

2 REVISÃO BIBLIOGRÁFICA

Dividido em duas seções, este capítulo aborda definições e principais trabalhos que discorrem sobre a temática apresentada. A transformada *wavelet* de Haar, uma das transformadas *wavelets* discretas, é usada para decompor e reconstruir a série objeto desta pesquisa, com o intuito de obter uma melhor acurácia na previsão. Os modelos de séries temporais obtidos pela metodologia Box-Jenkins sob análise dos critérios penalizadores e das estatísticas de previsão.

2.1 TRANSFORMADA WAVELET DE HAAR

Wavelet significa pequena onda, onduleta ou ondaleta (MORETTIN, 2014).

As *wavelets* podem ser suaves ou não, simétricas ou não e podem ter expressões matemáticas simples ou não (MORETTIN, 2014). Na Figura 1, ilustra-se essa definição com a *wavelet* de Morlet, uma transformada *wavelet* contínua.

Figura 1 - Representação Gráfica da *wavelet* de Morlet



Fonte: Adaptado de Devleker, 2016.

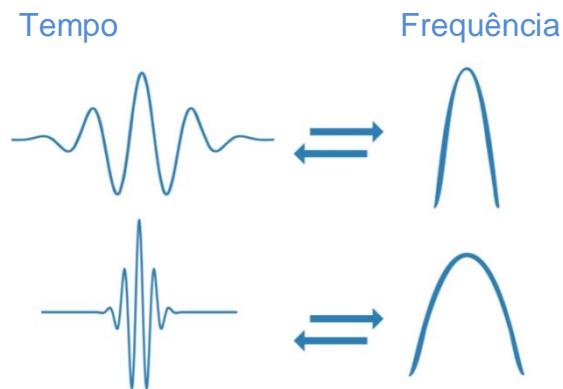
As *wavelets* são uma classe de funções usadas para reescrever uma determinada imagem ou sinal ao mesmo tempo em posição e escala, ou ainda, as *wavelets* são uma ferramenta matemática para decompor hierarquicamente funções. E, segundo Stollnitz, Derose e Salesin (1995), são técnicas para representar níveis de detalhes presentes em uma curva.

A partir de uma *wavelet*, podem-se formar outras *wavelets*, chamadas “filhas”, por translação ou contração da *wavelet* dada inicialmente. Ou seja, o procedimento

de análise adota uma função protótipo *wavelet*, chamada de “*wavelet* analisadora” ou “*wavelet* mãe”, que vem do fato de funções com diferentes tamanhos serem usadas no processo da transformada e todas serem originadas de uma *wavelet* principal.

Qualitativamente, duas características são exigidas para uma função $\psi(\cdot)$, chamada mãe: a oscilação e o decaimento rápido no tempo. E todas as funções usadas como núcleos da transformação correspondem a versões comprimidas e expandidas dessa mesma onda mãe, como apresentado na Figura 2.

Figura 2 - Comportamento das *wavelets*



Fonte: Adaptado de Devleker, 2016.

Uma larga variedade de *wavelets* pode ser usada, cada uma delas com diferentes compromissos entre o grau de compactividade da base de funções e o grau de suavidade das formas de onda.

A transformada *wavelet* discreta passa o sinal no domínio do tempo (ou espaço) por vários filtros denominados passa-alta e passa-baixa, que filtram as altas e baixas frequências do sinal. Esse procedimento é repetido cada vez que uma parte do sinal que corresponde a algumas frequências é removido do sinal.

Este estudo se desenvolve a partir de um caso particular de transformada *wavelet* discreta: transformada *wavelet* discreta de Haar, que é a *wavelet* mais simples da família Daubechies. As funções de Haar são de suporte compacto, ou seja, são definidas em um intervalo fechado e limitado, não são continuamente diferenciáveis e anulam-se completamente fora de um intervalo finito de tempo.

Assim, um dos critérios usados para definir se uma função pode ser uma *wavelet* é provar que ela é oscilatória (onda=*wave*), ou melhor, que seu valor médio no domínio temporal é nulo.

Matematicamente falando, as funções $\psi(t)$ e $\psi_{j,k}(u)$ devem satisfazer o seguinte critério:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (1)$$

Resultado de um teorema que traz como hipótese a condição de admissibilidade, equação (1):

$$C = \int_{-\infty}^{+\infty} \frac{|\hat{\varphi}(u)|^2}{|u|} du < \infty \quad (2)$$

Em que $\hat{\varphi}(u)$ é a transformada de Fourier de $\psi(t)$. Ou ainda,

$$\int_{-\infty}^{\infty} |\psi(t)| dt < \infty \quad (3)$$

Então,

$$f(u) = \frac{1}{c} \iint |j|^{2p-3} \tilde{f}_{j,k}(u) \psi_{j,k}(u) dj dk \quad (4)$$

Em que $f(u)$ é função do espaço $L^2(R)$, j e k são parâmetros de dilatação e compressão, respectivamente.

Donde segue também que:

$$\lim_{u \rightarrow 0} \hat{\varphi}(u) = 0 \quad (5)$$

se $\hat{\varphi}(u)$ é contínua, então $\hat{\varphi}(0) = 0$.

A transformada *wavelet* (TW) pode produzir uma boa representação local do sinal em ambos os domínios de frequência e de tempo. Isso justifica sua utilização em uma série de preços que geralmente apresentam características complexas, como a não estacionaridade e a volatilidade. Tan et al. (2010) propõem um método de previsão de preços baseado em transformada *wavelet* combinado com ARIMA e modelos GARCH. Por transformada *wavelet*, a série de preços é decomposta e reconstruída em uma série de aproximação e algumas séries detalhes. Em seguida, cada subsérie pode ser prevista separadamente por um modelo de série temporal

adequado. A previsão final é obtida por compor os resultados previstos de cada subsérie. Uma série “malcomportada”, ruidosa, é decomposta e reconstruída pela TWD inversa e dá origem a uma nova série com “melhor comportamento”.

Uma nova técnica de previsão do preço marginal de eletricidade associada à transformada *wavelet* é proposta por Kim, Yu e Song (2002). O resultado do estudo indica claramente que a proposta da transformada *wavelet* abordada pode ser considerada como um eficaz meio para previsão de preços.

Para Voronin e Partanen (2014), a combinação de *wavelets*, redes neurais e modelos ARIMA, ambos no domínio do tempo e *wavelets* no domínio das frequências, caracteriza uma significativa melhora e exatidão na previsão de preço e demanda de eletricidade.

Semelhantemente, Conejo (2005) propôs associar as técnicas de previsão por ARIMA, baseada na transformada *wavelet*, decompôs a série de preços, constituindo um conjunto de séries, com melhor comportamento e, então, devidamente ajustada pelos modelos ARIMA, reconstruiu a série de preços, prevendo os preços futuros.

As funções *wavelet*, têm despertado grande interesse na comunidade científica em geral. Uma vez que as transformadas *wavelet* (TW) se diferenciam de outras “ferramentas matemáticas”, como a transformada de Fourier, pela habilidade de decompor funções hierarquicamente, permitindo analisá-las em diferentes regiões de seu domínio e diferentes escalas ao mesmo tempo (SILVEIRA; KOZAKEVICIUS, 2016). Ainda, segundo os pesquisadores citados, as *wavelets* têm suas raízes na teoria das aproximações e processamento de sinais. Por isso têm sido exploradas e utilizadas nas mais diferentes aplicações e áreas do conhecimento científico.

As *wavelets* separam dados em seus diferentes componentes frequenciais e extraem cada componente com uma resolução adequada à sua escala. Daí sua vantagem sobre a análise de Fourier, que analisa o sinal como um todo, o que acarreta uma representação mais pobre para sinais que apresentam descontinuidades e variações bruscas. Pode-se afirmar que

uma *wavelet* é uma idealização matemática de uma nota musical... da mesma forma que as notas são colocadas em partituras,... pode-se imaginar a utilização de notas matemáticas na representação de certas entidades matemáticas, tais como funções e sinais (GROSSMANN; MORLET, 1984).

Os trabalhos pioneiros do engenheiro Jean Morlet e do físico Alex Grossmann descortinaram uma nova e prolífera área na Análise Funcional, as “*Ondalettes*”, cujas consequências práticas não param de crescer (OLIVEIRA, 2007).

2.1.1 Transformada *Wavelet* Discreta

A Transformada *Wavelet* Discreta trata-se de ferramenta matemática para análise, decomposição e representação de uma função em diferentes níveis de resolução ou escalas. Tal função, que pode ser um sinal, uma imagem ou um conjunto de dados, é então representada por uma curva média, e em cada escala mais refinada são, então, acrescentados detalhes complementares que, somados ao padrão médio, reconstroem exatamente a função original (STOLLNITZ; DEROSE; SALESIN, 1995).

As TW mostram-se mais eficientes em relação à Transformada de Fourier quando os sinais a serem analisados são não periódicos, uma vez que as *wavelets* servem como base para representação de outras funções (VIDAKOVIC; MUELLER, 1991).

Transformadas *wavelets* uni (1D) e bidimensionais (2D) são utilizadas em diversas áreas de pesquisa, como processamento digital de sinais, diversos métodos para resolução numérica de equações diferenciais parciais, compressão de imagens e análise estatística (BAYER; KOZAKEVICIUS, 2010).

Nas últimas décadas, uma das famílias de funções *wavelet* mais difundidas e utilizadas para a obtenção de transformadas ortonormais são as de Daubechies (1992) e Fraizer (1999), cujas propriedades fundamentais são indispensáveis para muitas aplicações e serão exploradas por meio da *wavelet* de Haar.

Essa base de funções foi definida segundo Kozakevicius et al. (2015), por Haar em 1910, antes mesmo da teoria de *wavelets* ter sido estruturada por Daubechies, que reconheceu e classificou as funções de Haar como funções *wavelets* ortonormais.

Daubechies, em 1992, construiu um conjunto de bases ortonormais de *wavelets* suaves com suporte compactos, que são alicerces das aplicações atuais de *wavelets*.

De forma geral, as *wavelets* são definidas por:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), a, b \in \mathbb{R}, a > 0, \quad (6)$$

Em que a representa um parâmetro de dilatação binária; 2^j mede o grau de compressão ou escala, b é um parâmetro de translação diádica e $k2^{-j}$ que determina a localização da *wavelet* no tempo. Nota-se que:

- Quando $|a| < 1$, $\psi_{a,b}(t)$ tem menor suporte no domínio do tempo e corresponde a altas frequências, ou seja, $\psi_{a,b}(t)$ é uma versão comprimida de $\psi(t)$, *wavelet* mãe;
- Quando $|a| > 1$, $\psi_{a,b}(t)$ tem maior suporte, espaço de tempo, do que $\psi(t)$ e corresponde às baixas frequências.

Portanto, as *wavelets* têm larguras e tempos adaptados às suas frequências (DEBNATH, 2002).

A transformada *wavelet* discreta (TWD) decompõe um sinal discreto em diferentes níveis de resolução. Semelhantemente, quando na representação de um objeto, procura-se uma escala na qual os detalhes de interesse possam ser bem representados. Essa busca por uma função que mapeie dados leva a compreender a TW de Haar como uma transformação linear de espaço vetorial para espaço vetorial, assumindo que qualquer conjunto de dados pode ser considerado um vetor (SILVEIRA; KOZAKEVICIUS, 2016).

Um modelo matemático adequado para formalizar a representação por escala é a análise de multirresolução. Há uma estreita relação entre o problema de amostragem pontual de um sinal e a ideia de escala. Decorre sua relação intrínseca com as *wavelets*, já que ambas estão definidas em $V = L^2(\mathbb{R})$ (GOMES; VELHO; GOLDENSTEIN, 1997).

Além disso, por meio das relações entre diferentes escalas, é apresentada a estrutura de análise de multirresolução associada a todas as funções *wavelets* ortonormais da família de Daubechies. A vantagem de se trabalhar com bases ortogonais é permitir a reconstrução perfeita do sinal original a partir dos coeficientes da transformada (MORETTIN, 2014).

A análise multirresolução (AMR) fornece um método eficiente de construir bases *wavelets* ortogonais já que a ideia fundamental é representar uma função (ou sinal) $f(t)$ como sendo um limite de sucessivas aproximações, cada qual sendo uma versão melhor da função $f(t)$. Estas correspondem a diferentes níveis de resolução.

A característica chave dessa análise é decompor um espaço de funções inteiro em subespaços individuais $V_m \subset V_{m+1}$ de forma que V_{m+1} seja o espaço de todas as funções reescaladas em V_m com $m \in Z$.

Entre as *wavelets* de Daubechies, a *wavelet* de Haar é a mais simples, com apenas um momento nulo e suporte no intervalo $[0,1]$.

A função escala $\phi(t)$, ou função de Haar, escolhida para gerar a análise multirresolução, é definida por:

$$\phi(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{caso contrário} \end{cases} \quad (7)$$

e $\phi(t) \in V = L^2(R)$ é constante por partes em intervalos $[k, k + 1)$.

Assim, a função *wavelet* de Haar (ou base de Haar) é definida como:

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & \text{caso contrário} \end{cases} \quad (8)$$

Em que $\psi(t) \in W_0 \subset W_j$ subespaços dois a dois ortogonais.

Em processamento de sinais, os dados são representados por um número finito de valores. Assim, serão consideradas versões discretas da *wavelet* (6), assumindo que a e b , parâmetros de dilatação e translação, respectivamente, tenham valores inteiros. Para tanto, fixam-se duas constantes positivas a_0 e b_0 , e define-se:

$$\psi_{j,k}(t) = a_0^{-j} \psi(a_0^{-j} t - kb_0), \quad j, k \in Z \text{ e para } \psi \in L^2(R), \quad (9)$$

Obtêm-se os coeficientes *wavelets* discretos $\langle f(t), \psi_{j,k}(t) \rangle$, cuja transformada *wavelet* discreta é dada pela equação:

$$\begin{aligned} (TWD)(j, k) &= \langle f(t), \psi_{j,k}(t) \rangle \\ &= \int_{-\infty}^{\infty} f(t) \overline{\psi_{j,k}(t)} dt \\ &= a_0^{-j} \int_{-\infty}^{\infty} f(t) \overline{\psi_{j,k}(a_0^{-j} t - kb_0)} dt \end{aligned} \quad (10)$$

$f(t)$ e $\psi_{j,k}(t)$ contínuas.

Essa equação também pode ser obtida da discretização dos parâmetros a e b em que $a = a_0^{-j}$ e $b = kb_0$, com j, k inteiros, de uma representação contínua:

$$(TWC)(a, b) = \int_{-\infty}^{\infty} f(t) \overline{\psi_{a,b}(t)} dt = \langle f(t), \psi_{a,b}(t) \rangle \quad (11)$$

Segundo Pozzebom (2009), a TWD representa uma função por meio de um conjunto mensurável de coeficientes *wavelets*, que correspondem a pontos em uma grade bidimensional de pontos discretos no domínio escala-tempo, indexados por j e k . Se o conjunto $\{\psi_{j,k}(t)\}$ definido por (8) é completo em $L^2(R)$ para alguma escolha de $\psi(t)$, a e b , então, pode-se expressar qualquer $f(t) \in L^2(R)$ como:

$$f(t) = \sum_{j,k=-\infty}^{\infty} \langle f(t), \psi_{j,k}(t) \rangle \psi_{j,k}(t) \quad (12)$$

Ou ainda,

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{j,k} \psi_{j,k}(t) \quad (13)$$

Pode-se afirmar que (13) é uma série de *wavelets* de $f(t)$, e os coeficientes de *wavelets* são dados por:

$$c_{j,k} = \langle f, \psi_{j,k} \rangle = \int_{-\infty}^{\infty} f(t) \psi_{j,k}(t) dt \quad (14)$$

A ideia é considerar dilatações (compressões) e translações de uma única função $\psi(t)$, de modo a cobrir R . Computacionalmente, para melhor eficiência, são utilizados $a_0 = 2$ e $b_0 = 1$, ou seja, $\psi_{j,k}$ é obtida por uma dilatação binária 2^{-j} e uma translação diádica de $k2^j$ (MORETTIN, 2014).

Assim,

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}} \psi(2^{-j}t - k) \quad (15)$$

forma uma base ortonormal de $L^2(R)$. Tem-se, portanto, uma base ortonormal de *wavelets*.

Tais funções $\psi(t)$ geram espaços complementares àqueles expandidos por funções escalas, chamadas de funções *wavelets* (DEBNATH, 2002). Chama-se de espaços *wavelets*, $W_m \in L^2(R)$, complemento ortogonal de V_m no espaço V_{m+1} , com $m \in Z$, desde que $V_m \subset V_{m+1}$ e $\{V_m\}$ seja uma análise multirresolução. Tem-se então,

$$\begin{aligned}
V_{m+1} &= V_m \oplus W_m \\
&= (V_{m-1} \oplus W_{m-1}) \oplus W_m \\
&= V_0 \oplus W_0 \oplus W_1 \oplus \dots \oplus W_m \\
&= V_0 \oplus (\oplus_{m=0}^m W_m)
\end{aligned} \tag{16}$$

Sendo $V_m \perp W_m$, e $W_m \perp W_{m'}$, para $m \neq m'$. A saber, a soma ortogonal é representada por \oplus . E os espaços *wavelets* W_m são formados por detalhes do sinal na escala V_m (POZZEBON, 2009).

Nesta pesquisa, os conjuntos de dados, preço do cimento Portland e preço do aço, são transformados em vetores. Os vetores, seguindo esse processo de decomposição, terão cada ponto ou sinal decomposto em coeficientes denominados de aproximações e detalhes. O que vem gerar a análise multirresolução, como já mencionado anteriormente.

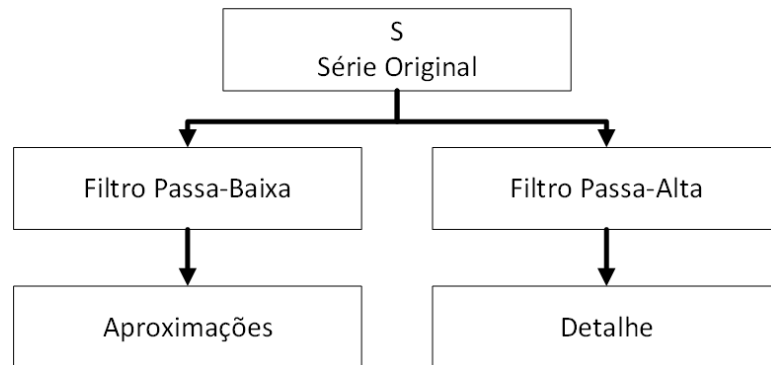
2.1.2 Análise Multirresolução e Filtragem (AMR)

Filtragem é um processo que apresenta de forma prática a realização da análise multirresolução (AMR). Tal processo, por meio de filtros passa-alta e passa-baixa faz a filtragem de um sinal, compondo versões do sinal original com coeficientes de funções *wavelets* e funções escala, respectivamente.

Como em muitos sinais, o destaque fica por conta das componentes de baixa frequência, que são as aproximações, pois fornecem características como a identidade desses sinais; enquanto as componentes de alta frequência, fornecem os detalhes.

Segue, na Figura 3, o processo de filtragem da série em seu nível mais básico, ou primeiro nível da obtenção de aproximações e detalhes.

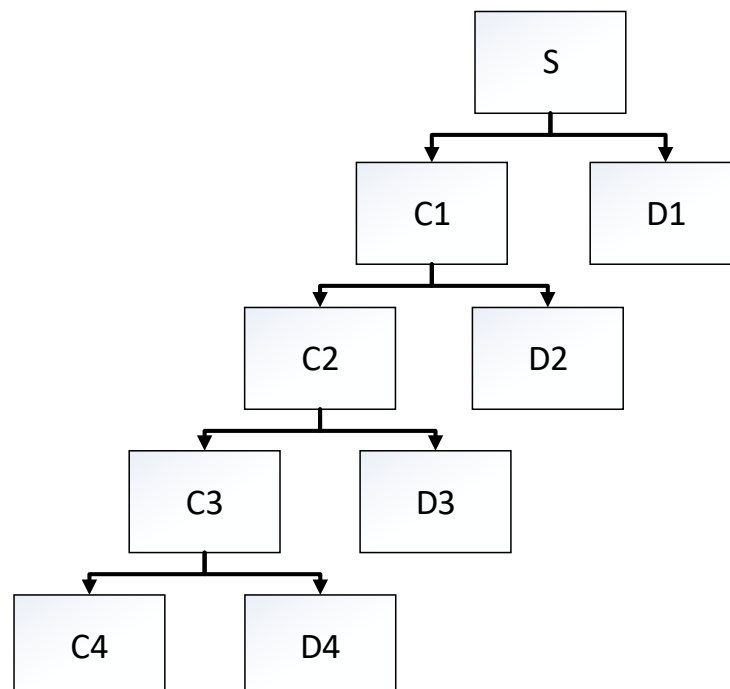
Figura 3 - Filtragem de um sinal S em primeiro nível de aproximações e detalhes



Fonte: Adaptado de Pozzebon (2009).

A decomposição pode ocorrer como um processo iterativo, com sucessivas decomposições das aproximações. Assim, são obtidos diferentes níveis de aproximações e detalhes, conforme apresentado na Figura 4, a seguir.

Figura 4 - Ilustração do processo de decomposição de um sinal S em quatro níveis pela análise de multirresolução em que C_j e D_j , para $j=1, 2, 3, 4$ são ditos aproximações e detalhes respectivamente



Fonte: Adaptado de Pozzebon (2009).

Como o processo é iterativo, pode-se decompor o sinal até um número satisfatório de níveis, levando-se em conta a natureza desse sinal. Nesta pesquisa, em que se analisam os preços mensais do cimento Portland e do aço em R\$/kg, o número de decomposições é limitado pelo nível apontado pela potência de 2^j , a saber, fatorando 256, observações, na potência de base 2, $2^j = 2^8$ tem-se $j = 8$.

Os preços mensais do aço, são compostos por 128 observações, tem-se assim, dimensão $2^j = 2^7$. Observa-se que, no processo de filtragem, ocorre uma diminuição do número de amostras a cada nível, ou seja, há uma redução na dimensão do vetor que representa os dados, pelo fato destes serem tomados 2 a 2.

2.1.3 Limiarização da *Wavelet (Threshold)*

A filtragem de sinais é uma das aplicações da TWD e baseia-se no corte (*Threshold*) ou encolhimento dos coeficientes *wavelets* (detalhes) segundo algum critério (PERLIN, 2010). Dados obtidos a partir de medições ou manipulados por aplicativos computacionais armazenados digitalmente sofrem interferência em alguma etapa do processo, a qual é denominada ruído (KOZAKEVICIUS; BAYER, 2014).

Estrategicamente, no intuito de eliminar ou até mesmo reduzir o ruído presente em determinado sinal, é que a inserção de um *threshold* em determinado nível de detalhes pode vir a contribuir para melhorar ainda mais as estatísticas dos modelos, por meio de cortes. Para tanto, seguem-se alguns passos para efetivar a filtragem (PERLIN, 2010):

- a) transformada *wavelet* discreta de Haar no sinal original;
- b) corte dos detalhes conforme a estratégia escolhida;
- c) transformada *wavelet* discreta inversa de Haar (reconstrução), com a utilização dos novos coeficientes *wavelets* (detalhes) obtidos.

Ainda, a obtenção do limiar de corte λ , proposta por Donoho e Johnstone (1994) denominado *Universal Threshold*, é dada por:

$$\lambda = \hat{\sigma} \sqrt{2 \log(T)} \quad (17)$$

Em que $\hat{\sigma}$ é o estimador do desvio-padrão do ruído e T a dimensão do vetor de entrada.

Nesse sentido, desejou-se verificar quais dos $D_{j,k}$, ditos coeficientes *wavelets*, são puramente aleatórios (ruídos) e quais representam alguma causa especial de comportamento. Seguindo então ao passo de escolha do *threshold* mais adequado ao conjunto de dados. A operação *threshold*, reduz ou anula os $D_{j,k}$ menores do que certo λ (KOZAKEVICIUS; BAYER, 2014).

Segundo Morettin (2004), há duas propostas para efetuar tais cortes, denominadas de *hard* e *soft threshold*. Este segundo, por apresentar vieses maiores, foi o escolhido para limiarizar os vetores detalhes, obtidos pela equação:

$$thr_{\lambda}^S(D_{j,k}) = \begin{cases} 0, & se |D_{j,k}| \leq \lambda \\ sign(D_{j,k})(|D_{j,k}| - \lambda), & se |D_{j,k}| > \lambda \end{cases} \quad (18)$$

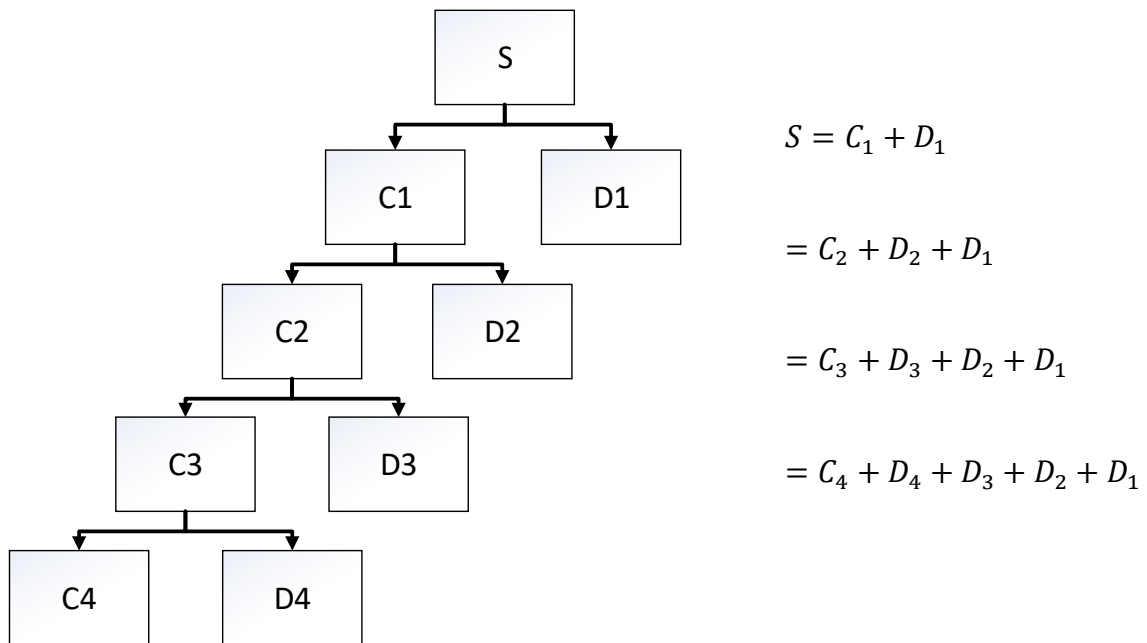
De acordo com Kozakevicius e Bayer (2014), os métodos de cortes, *hard* e *soft threshold* não são os únicos possíveis, outros tantos podem ser encontrados na literatura, como em Gao e Bruce (1997); Gao (1998); Antoniadis (2007); Smith, Agaian e Akopian (2008).

2.1.4 Processo de Reconstrução

Após realizar a decomposição do vetor original em detalhes e aproximações, este é reconstruído usando as funções de Haar que geraram a análise multirresolução (AMR).

A manipulação matemática que efetua a reconstrução dos sinais é chamada de Transformada *wavelet* discreta inversa. Há diferentes formas de reconstruir o sinal original a partir de seus coeficientes. Na Figura 5, apresenta-se um desses processos de reconstrução.

Figura 5 – Reconstrução de um sinal S



Fonte: Adaptado de Varanis (2014) e Pozzebon (2009)

Vale salientar que C_1 e D_1 referem-se a aproximações e detalhes no nível 1, respectivamente. Dessa forma, D_1 (detalhe no nível 1 de decomposição) é o primeiro coeficiente *wavelets* obtido e primeira metade a ser retirada da sequência de decomposições. Parte-se, então, para uma nova decomposição, nível 2, em que decompõem-se C_1 em C_2 e D_2 , aproximações e detalhes no nível 2, obtendo, assim, com D_2 , o segundo coeficiente *wavelet*. Conforme a Figura 5, usa-se, a cada nível, a aproximação do nível em questão, acrescentada dos detalhes que foram obtidos nos níveis anteriores em sequência.

2.2 MODELOS DE SÉRIES TEMPORAIS

Previsão nada mais é do que uma estimativa quantitativa sobre eventos futuros baseados em informações de períodos passados e recentes (STEFFEN, 2010).

Embora os dados de séries temporais surjam nas mais diversas áreas do conhecimento, algumas características são comuns a esses tipos de dados, como a dependência entre observações vizinhas, a análise e a modelagem dessa dependência. Vale lembrar que há fatores complicadores nesse processo, como presença de tendências e variações sazonais ou cíclicas que podem ser difíceis de estimar ou remover (EHLERS, 2007).

Um modelo de previsão para a demanda de cimento no Brasil por meio da análise de séries temporais é apresentado por Gomes e Lopes (2008), que compara diferentes métodos e indica que, apesar de alguns valores estarem bem próximos dos valores reais, há picos que não se comportam como o esperado, influenciados, talvez, por mudanças econômicas, políticas e cambiais.

A metodologia Box-Jenkins (1976), apontada por Marchezan e Souza (2010), mostra-se eficiente para prever preço das culturas de soja, arroz e milho, captando efeitos sazonais nos preços e descrevendo o comportamento das séries. Vários modelos são encontrados e testados, mas, por critérios penalizadores, AIC (Akaike Information Criteria), BIC (Schwarz Bayesian Criteria) e Erro Quadrado Médio, confirma-se o melhor modelo para cada série, ou seja, para cada tipo de grão.

Uma série temporal é um conjunto de observações ordenadas no tempo, ou ainda, é uma realização de um processo estocástico. Os dados contidos na referida série devem estar equidistantes, apresentando uma forte dependência serial, isto é, o valor de uma observação mostra-se correlacionado com outro valor em instantes equiespaçados de tempo, e o interesse é analisar e modelar essa dependência. Para Junges e Fontana (2011), estas podem ser classificadas em: discretas, contínuas, univariadas e multivariadas.

Representa-se uma série temporal da seguinte forma: Z_t , $t = 1, 2, 3, \dots, n$, em que n representa o tamanho da série e Z_t o conjunto de observações correlacionadas entre si, seguindo-se uma ordem cronológica.

A previsão é uma das principais razões que popularizam os modelos Box-Jenkins (1976), genericamente conhecidos por ARIMA (Auto Regressive Integrate

Moving Averages) e ditos modelos de memória curta, que, segundo Werner e Ribeiro (2003), são modelos matemáticos que captam o comportamento da correlação seriada ou autocorrelação entre os valores da série temporal e, com base nesses comportamentos, possibilitam realizar previsões futuras. Como a maior parte dos procedimentos estatísticos foi desenvolvida para analisar observações independentes, o estudo de séries temporais requer o uso de técnicas específicas.

As previsões, em muitos casos, principalmente de curto prazo, são melhores que as obtidas com base nos modelos econométricos tradicionais. Uma vez que, quanto mais longe no futuro, maior é a incerteza associada à previsão (BARROS, 2004).

Segundo Nunes et al. (2000), o procedimento de aplicar ARIMA em subséries extraídas da série original por meio das *wavelets* com seus filtros passa-baixa e passa-alta contribui positivamente para a qualidade de previsão de séries de tempo.

Reis e Silva (2004) observaram um ganho em curtíssimo prazo por meio da modelagem via *wavelets* que combinou somente padronização e diferenciação. Sugerem, ainda, que a análise de multirresolução via *wavelets* pode contribuir para uma melhora considerável na exatidão das previsões de carga realizadas por arquiteturas neurais.

Para Lima (2016), pode-se dividir a análise de séries temporais seguindo dois enfoques: no domínio do tempo ou no domínio da frequência. A saber, se a série temporal for analisada no domínio do tempo, será dita paramétrica, visto que envolve um número finito de parâmetros. Caso a análise ocorra no domínio da frequência, será dita não paramétrica, já que envolve um número infinito de parâmetros.

Designa-se o modelo ARIMA - Autorregressivo Integrado de Médias Móveis pela notação ARIMA (p, d, q), cujos termos p, d e q, inteiros positivos, indicam o número de termos autorregressivos; o número de diferenciações para que a série se torne estacionária e o número de termos de médias móveis, respectivamente. São, ainda, casos particulares do ARIMA: o modelo ARMA(p, q) - Autorregressivo de Médias Móveis, o modelo autorregressivo AR(p) e o modelo de médias móveis MA(q), ambos utilizados para séries temporais estacionárias com número de diferenças nulo (MORETTIN, 2004).

Os modelos estacionários descrevem a série que não possui tendência. Interpretam uma série temporal como uma realização de um vetor aleatório do

tamanho das séries disponíveis, seguindo o princípio da parcimônia e da construção de modelos, por meio de um ciclo iterativo. Os modelos estacionários são:

Modelos autorregressivos de ordem p – AR(p): correspondem à defasagem da série Z_t , denotados por AR(p), em que p indica a ordem do modelo, μ é a média, ϕ é o parâmetro autoregressivo e $t - p$ indica o número de defasagens, conforme a equação (19).

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (19)$$

A identificação do modelo adequado começa pelos valores dos parâmetros, ou seja, o modelo será considerado estacionário se $|\phi_p| < 1$, satisfazendo a condição de estacionaridade. A condição de estacionaridade do AR(p) estabelece que todas as raízes devem cair fora do círculo unitário.

Modelo de médias móveis de ordem q – MA(q): é a combinação do ruído branco a_t , do período atual com os ocorridos anteriormente. A notação do modelo é MA(q), na qual q indica a ordem do modelo e θ é o parâmetro de média móvel, dado pela equação (20).

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (20)$$

Dessa forma, identificando a condição de estacionaridade, o processo apresenta ordem finita. Portanto, será sempre estacionário.

Quanto à condição de invertibilidade, é imposta a restrição de que o parâmetro em módulo é menor que 1, ou seja, $|\theta_q| < 1$.

Modelo autorregressivo de médias móveis – ARMA(p,q): é a união dos modelos vistos anteriormente, os autorregressivos e o de médias móveis. É utilizado em séries encontradas na prática, para as quais não se desejam números muito grandes de parâmetros. A equação (21) representa um ARMA(p,q).

$$Z_t = \mu + \underbrace{\phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t}_{AR(p)} - \underbrace{\theta_1 a_{t-1} - \dots - \theta_q a_{t-q}}_{MA(q)} \quad (21)$$

Percebe-se que as condições de estacionariedade de um processo ARMA(p,q) são as mesmas que as de um processo AR(p), e as condições de invertibilidade são as mesmas que as de um processo MA(q).

Para Pyndyck (2004), o processo está em equilíbrio quando se mantém ao redor de um nível médio constante, caso contrário, o processo é não estacionário.

Modelo autorregressivo integrado de médias móveis – ARIMA (p, d, q): os modelos estacionários e os não estacionários diferenciam-se pelo número de diferenças necessárias para a estacionarização de uma série, representada pela letra d, se d=0, a série é estacionária; se d=1 ou d=2, a série é não estacionária. Segundo Morettin (2004), a primeira diferença busca estabilizar o nível da série e, com a segunda diferença, a variabilidade da série.

Uma generalização desse tipo de modelo, proposta por Granger e Joyeux, em que o parâmetro d assume não só valores inteiros, mas também graus de diferenciação fracionários.

Os modelos que apresentam essa propriedade permitem o estudo de séries caracterizadas por longas dependências temporais, intitulados ARFIMA(p,d,q), em que F significa “fractional”.

O processo Z_t é um ARFIMA(p,d,q) se este é a solução da equação de diferenças, conforme equação (22).

$$\phi(B)(1 - B)^d Z_t = \theta(B)a_t \quad (22)$$

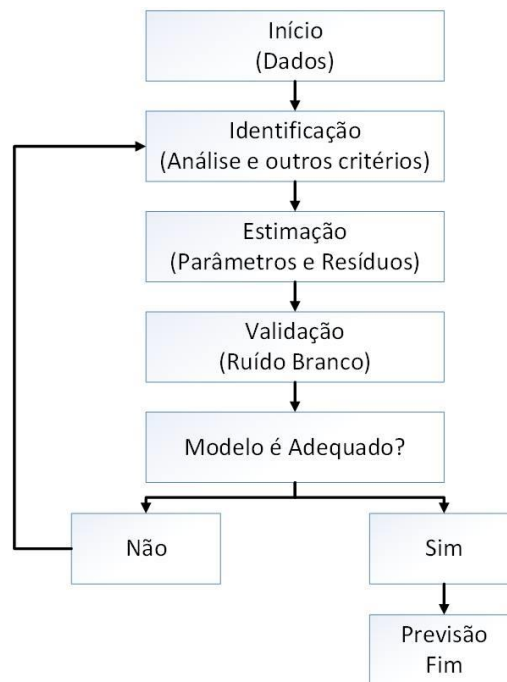
O termo $(1 - B)^d$ é definido pela expansão binomial, dada pela equação (23).

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k \quad (23)$$

Em que B é operador de retardo, $\{a_t\}$ é um processo Ruído Branco, $\phi(B)$ e $\theta(B)$ são polinômios em B de graus p e q respectivamente, e, $E(a_t) = 0, \sigma_\varepsilon^2 > 0$.

Na Figura 6, é possível observar a estratégia utilizada para a construção de modelos, pela metodologia Box-Jenkins, baseada no ciclo iterativo formado pelas etapas: identificação, estimação, validação e adequação do modelo ou diagnóstico.

Figura 6 - Fluxograma do ciclo iterativo de Box e Jenkins



Fonte: Adaptado de Morettin e Toloí (2004).

O ciclo iterativo é composto pela identificação do modelo, estimação e diagnóstico. Na identificação do modelo, analisa-se a estacionariedade da série a partir da média e variância. Em seguida, analisam-se os componentes tendência, sazonalidade e ciclo, chegando à análise de dados por meio da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP).

A estimação realizada, por meio do Método dos Mínimos Quadrados (MMQ) ou da Máxima Verossimilhança, analisa a significância dos parâmetros do componente autorregressivo, do componente de médias móveis e da variância, sejam eles modelos estacionários ou não estacionários.

Após as etapas de identificação e estimação, chega-se à adequação do modelo, por meio de critérios de seleção do modelo adequado, podendo-se realizar as previsões. Caso o modelo se apresente inadequado, deve-se retornar à etapa da estimação de novos parâmetros. Se a FAC apresentar um decaimento lento, quase linear, é um indicativo de não estacionariedade, o que sugere um novo tipo de modelo dito de memória longa, $ARFIMA(p,d,q)$, em que d assume valores fracionários. E, se série apresentar componente sazonal, utiliza-se o filtro $SARIMA(p,d,q)(P,D,Q)_s$, no qual p e q referem-se respectivamente, às ordens autorregressiva e de média móvel e d , ao número de diferenças; enquanto que P , D

e Q referem-se, respectivamente, à ordem autorregressiva sazonal, diferença sazonal e de média móvel sazonal (VICINI; SOUZA, 2007). O componente sazonal mostra flutuações ocorridas em subperíodos de tamanho s , ocorrendo, por exemplo, trimestral ou quadrimestral e assim por diante, em períodos inferiores a um ano (FERRAZ, 1999).

O modelo é avaliado pela qualidade do resíduo, que é a sequência de variáveis aleatórias, não correlacionadas entre instantes diferentes, simétrica e identicamente distribuídas, com média zero e variância constante. Uma vez verificado se o resíduo é Ruído Branco, significa que o modelo captou todas as informações do processo representado por $RB \sim (0, \sigma_{a_t}^2)$. Tal avaliação é feita por meio do correlograma da função de autocorrelação (FAC) e função de autocorrelação parcial (FACP), que servem para verificar se os dados são ou não autocorrelacionados.

Na etapa de identificação do modelo, Morettin e Tolo (2004) recomendam encontrar vários modelos significativos, para que, posteriormente, possa ser escolhido o melhor por meio dos critérios penalizadores AIC e BIC, que buscam encontrar modelos parcimoniosos, ou seja, aqueles com menor número de parâmetros. De outro modo, dados dois ou mais modelos com ajustes bons, escolhe-se aquele com menor número de parâmetros, visando à melhor compreensão do processo, haja vista a simplicidade na estrutura. Já modelos com excesso de parâmetros ajustam-se bem aos dados amostrais, mas não atingem desenvolver o comportamento da série como um todo. A saber, aplica-se o conceito da parcimônia a qualquer problema de modelagem estatística e não apenas à modelagem de séries temporais por ARIMA.

Assim, o AIC sugere escolher o modelo cujas ordens p e q minimizam o critério. Daí decorre ser dito critério penalizador de parâmetros.

Existem vários indicadores para avaliar e comparar o desempenho dos modelos de previsão que possibilitam a comparação dos modelos de previsão de forma imparcial. A saber, os critérios de seleção mais usados são o AIC e o BIC.

Em geral, esses indicadores avaliam a capacidade de precisão, a adequação dos modelos (DIEBOLD; LOPEZ, 1996).

Na equação (24), demonstra-se como é encontrado o AIC, em que $\hat{\sigma}_{p,q}^2$ é o estimador de máxima verossimilhança de σ^2 para um modelo ARMA(p,q) e N é o

tamanho da amostra, p : número de parâmetros autorregressivos e q : número de parâmetros de médias móveis e \ln é o logaritmo neperiano. E o minimizar significa identificar as ordens p e q que equilibrem o comportamento da série analisada.

Critérios de Seleção:

$$\text{Seleção do modelo: } AIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{2(p+q)}{N} \quad (24)$$

Akaike (1970) e Schwarz (1978) sugerem minimizar o BIC, que, em um modelo ARMA, é dado pela equação (25):

$$\text{Seleção do modelo: } BIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + (p + q) \frac{\ln N}{N} \quad (25)$$

Em que $\hat{\sigma}_{p,q}^2$ é a estimativa de máxima verossimilhança da variância residual do modelo ARMA(p, q).

Para avaliação dos erros de previsão, são utilizadas as medidas: raiz do erro médio quadrático (RMSE), erro percentual médio absoluto (MAPE) e a estatística U-Theil. O cálculo das referidas medidas se obtém a partir das equações (26), (27) e (28).

Critérios de Previsão:

Seleção de previsão: RMSE – Raiz do erro médio quadrático.

$$RMSE = \sqrt{\frac{\sum (Z_t - \hat{Z}_t)^2}{n}} \quad (26)$$

Na equação (26), n corresponde ao número de previsões realizadas, Z_t representa o valor real no instante t e \hat{Z}_t representa o valor previsto no instante t .

O motivo da utilização desses critérios é avaliar a capacidade preditiva dos modelos perante os dados originais.

O MAPE indica o valor médio do erro percentual absoluto das previsões sobre um conjunto de dados. É calculado a partir das previsões um passo à frente, geradas pelo modelo matemático estimado (RUSSO, 2002).

Algebricamente, é apresentado na equação (27) a estatística MAPE como método de seleção de previsão:

$$MAPE(\%) = \frac{\sum \left(\frac{Z_t - \hat{Z}_t}{Z_t} \right)}{n} \times 100 \quad (27)$$

Em que n é o número de observações, Z_t é o valor observado no instante t e \hat{Z}_t é o valor previsto no instante t .

Uma das vantagens do MAPE é a facilidade de entendê-lo, uma vez que é dado em percentual. Já a desvantagem é que, se o valor real for pequeno, qualquer discrepância na previsão fará o MAPE aumentar consideravelmente.

A estatística U de Theil mede o quanto o método de previsão é melhor se comparado à previsão ingênua (método de previsão com um passo à frente, em que só é considerado o último dado verificado), atuando como parâmetro de avaliação. Quanto mais próximo de zero, melhor o desempenho do modelo e, conseqüentemente, os erros de previsão são menores (SOUZA; SAMOHYL; MIRANDA, 2008).

A referida estatística é dada pela expressão conforme a equação (28).

$$U = \sqrt{\frac{\sum_{t=1}^{n-1} \left(\frac{\hat{Z}_{t+1} - Z_{t+1}}{Z_t} \right)^2}{\sum_{t=1}^{n-1} \left(\frac{Z_{t+1} - Z_t}{Z_t} \right)^2}} \quad (28)$$

Em que n é o número de observações; Z_t e Z_{t+1} são os valores observados nos instantes t e $t + 1$, respectivamente; \hat{Z}_{t+1} é o valor previsto para o instante $t + 1$.

A estatística de U -Theil é uma medida relativa de precisão que compara os resultados de previsão com dados históricos mínimos. É decisiva na determinação da acurácia com valores entre 0 e 1, considera-se previsão perfeita quanto mais próxima de zero for o valor de U estatística de Theil, medidas de $U \geq 1$ não agradam, ou seja, consideram-se previsão ruim. Vale lembrar, também, que todas as medidas de capacidade preditiva citadas anteriormente podem ser estendidas a erros de previsão k passos à frente.

As referidas estatísticas indicam, por meio de seus valores, ao avaliador, o melhor modelo entre os concorrentes para então fazer a previsão.

Nesta pesquisa, as estatísticas de previsão serão utilizadas em dois momentos distintos: visando obter o melhor modelo para prever com até três passos

à frente, por meio do método de modelagem autorregressivo integrado de médias móveis (ARIMA) aplicado à série original, e no ARIMA aplicado à série de dados decompostos e reconstruídos pela transformada *wavelet* de Haar.

3 MATERIAIS E MÉTODOS

Esta pesquisa pode ser classificada como exploratória, descritiva, bibliográfica e quantitativa do ponto de vista dos procedimentos que são adotados (GIL, 2007). O suporte teórico e a combinação do uso de *wavelets* com seus filtros em séries temporais tornam possível comparar modelos de previsão e determinar com maior precisão os modelos que melhor preveem os preços do cimento Portland e do aço no Rio Grande do Sul. Para consecução desta pesquisa, optou-se pelo uso dos softwares MATLAB® e E-Views 9.5 *Student Version*.

3.1 COLETA DE DADOS

Para efeitos de estudo, os dados utilizados nesta pesquisa foram obtidos através do site da Câmara Brasileira da Indústria da Construção-CBIC (<http://www.cbicdados.com.br/menu/materiais-de-construcao/>), e referem-se à evolução do valor médio do cimento Portland - em R\$/kg com valores mensais, no período de setembro/1995 a dezembro/2016, totalizando 256 observações mensais, e a evolução do valor médio do aço CA 50 de 10 mm - em R\$/kg com valores mensais, no período de fevereiro/2007 a setembro/2017, totalizando 128 observações. Justificam-se períodos diferentes em razão da divulgação dos dados no referido site. Não foram encontrados registros da evolução do valor médio do aço em período anterior a fevereiro/2007.

3.2 ANÁLISE

Foram empregados métodos de modelagem de séries temporais (BOX-JENKINS,1970), por meio do modelo ARIMA – Autorregressivos Integrados de Médias Móveis, que representam o processo estocástico gerador da série temporal.

Foram traçados os gráficos de cada uma das séries originais a fim de verificar a estacionariedade de ambas, identificar se há comportamento tendencioso, existência ou não de sazonalidade e, ainda, se há presença de *outliers*.

Na sequência, geraram-se as funções de autocorrelações (FAC) e as funções de autocorrelações parciais (FACP), por meio dos correlogramas. Foram

identificados os possíveis filtros, AR(p), MA(q), ARMA(p,q), no caso das séries serem estacionárias.

Seguiu-se, então, ao ajuste dos modelos da classe geral ARIMA à variável em estudo, obtendo-se diversos modelos concorrentes, os quais deverão apresentar a característica de ruído branco.

Entre os modelos concorrentes, a indicação do melhor modelo se deu pelo menor valor obtido pelos critérios penalizadores AIC e BIC.

De forma a contribuir à tomada de decisão, foram calculadas as estatísticas de previsão RMSE e U-Theil de cada modelo. A partir disso, escolheu-se aquele que será utilizado para as previsões do preço do cimento Portland e do preço do aço no estado do Rio Grande do Sul em até três passos à frente.

As previsões obtidas em cada uma das etapas serão comparadas a fim de obter o melhor modelo, utilizando os modelos ARIMA, *wavelet*-ARIMA com *threshold* e sem *threshold*.

Para utilizar a ferramenta matemática, transformada *wavelet* de Haar, descrita no capítulo 2 desta pesquisa, há necessidade que ambos os conjuntos de dados sejam da ordem de 2^j , condição para a decomposição por Haar.

Satisfeita a condição de ter-se conjuntos com 2^j dados, em que j indica o número de níveis de decomposição que será possível chegar. Isso permite iniciar a etapa de obtenção das aproximações e detalhes, no domínio do tempo e frequência, cujo algoritmo completo para j níveis de decomposição é dado a seguir.

Tabela 1 – Formulação algorítmica para decomposição por TWD Haar

Algoritmo 1: TWD de Haar 1D

tamanho = linspace (1, 128, 128);

for i = 1:length(tamanho)

C1(i) = (vetor_cimento((i*2) - 1) + vetor_cimento(i*2))/2.0;

end

for i = 1:length(tamanho)

D1(i) = vetor_cimento((i*2) - 1) - C1(i);

end

Além disso, é possível testar a representatividade dos dados procedendo da seguinte forma: fazer a decomposição da primeira metade dos dados da série original por meio da transformada *wavelet* discreta de Haar, que nível a nível reduz a dimensão do vetor (dados) calculando a média aritmética 2 a 2, ao que se denominam filtros. Assim, obtêm-se componentes de detalhes e aproximações (ou escalas) que nos permitem recompor a série. A decomposição via *wavelets* é possível com a ajuda do *software* MATLAB®.

A TWD inversa de Haar, aplicada a esses dados, produzirá uma nova série que será modelada pela metodologia Box-Jenkins, e será aferido sua capacidade preditiva, visto que a modelagem ARIMA será realizado na série reconstruída. Para a tomada da primeira metade dos dados, chamados “janela”, a fim de prever com maior precisão, desloca-se a “janela” tantos passos quantos foram determinados, (2 a 2, 3 a 3,...), operação garantida porque as *wavelets* são funções continuamente diferenciáveis. Procede-se novamente testando esses dados, comparando-os com o modelo obtido a partir da série original e com as sucessivas previsões das “janelas” anteriores.

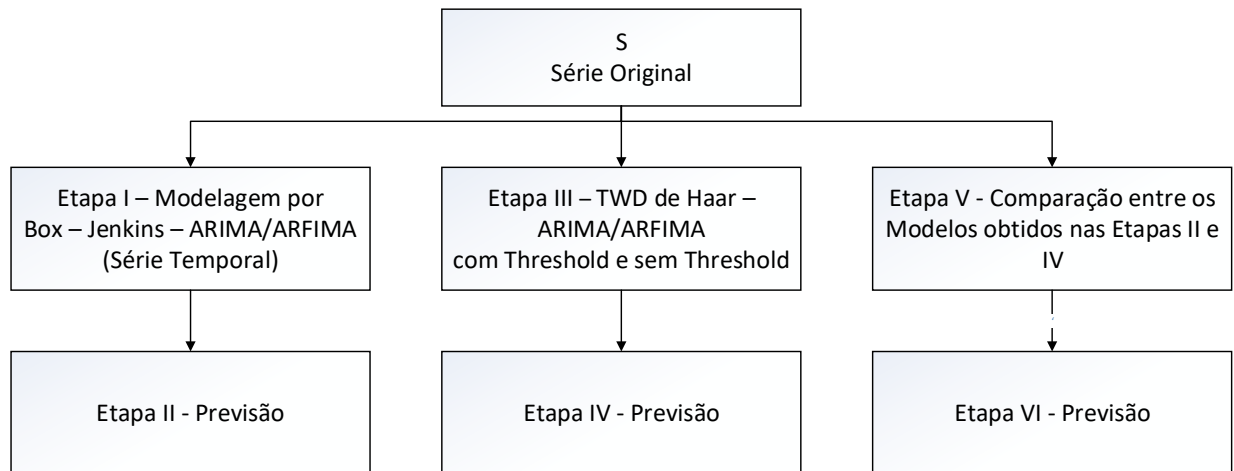
Como antes mencionado, com os sinais decompostos e reconstruídos pela transformada *wavelet* de Haar com e sem *threshold*, escolhe-se a partir dos critérios penalizadores AIC, BIC e estatísticas como RMSE e U-Theil, o melhor modelo. As previsões por meio de *wavelet*-ARIMA com e sem *threshold*, para o preço do cimento Portland e para o preço do aço, serão comparadas com o modelo obtido pelo método usual ARIMA na série original a fim de constatar a acurácia do processo.

Essas previsões visam auxiliar gestores e empresários da indústria do cimento, futuros empreendedores clientes e demais interessados nesse mercado. Além disso, inibe superfaturamentos e evita aditivos de valores inesperados sejam em obras civis, sejam em obras públicas, visto que nessas últimas as proporções são ainda maiores em razão da má gestão da administração pública.

3.3 ETAPAS METODOLÓGICAS DO ESTUDO

É abordada, neste tópico, a descrição dos procedimentos metodológicos adotados a fim de cumprir os objetivos propostos. De forma sintetizada, na Figura 7, apresentam-se as quatro etapas consideradas nesta pesquisa.

Figura 7 - Fluxograma com as etapas da pesquisa



Fonte: Autora

A Etapa I, modelagem pela metodologia Box-Jenkins, é utilizada para encontrar os modelos significativos para a previsão das séries constantes nesta pesquisa. Nessa etapa, analisam-se comportamento, tendência, sazonalidade, ruídos e parcimônia, bem como as estatísticas para seleção do melhor modelo. Para isso, leva-se em conta o menor valor obtido entre os modelos concorrentes.

A ferramenta *wavelets* é introduzida na etapa III, em que se fez a decomposição dos dados, sem *threshold* e com *threshold*, reconstruindo os dados na sequência para posterior aplicação da metodologia Box-Jenkins.

Na quinta etapa, comparam-se os modelos obtidos em cada uma das etapas II e IV a fim de indicar o melhor entre estes para prever de forma acurada os preços do cimento Portland e do aço no estado do Rio Grande do Sul, cumprindo, assim, a Etapa VI.

4 RESULTADOS E DISCUSSÕES

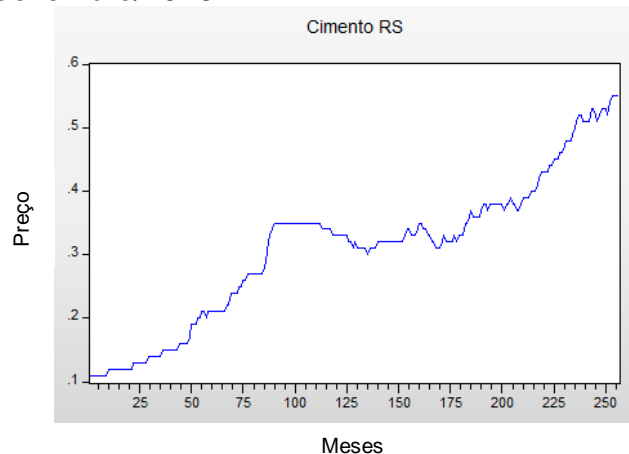
Neste capítulo, compara-se, sob forma de estudo de casos, previsões relativas a dois diferentes métodos de modelagem de séries de tempo, os quais consistem na aplicação da metodologia Box-Jenkins (ARIMA), tanto na forma usual como por um procedimento auxiliar baseado na transformada *wavelet* discreta de Haar e sua inversa. O procedimento inicial consiste na análise por metodologia Box-Jenkins das séries de preços mensais do cimento Portland e do aço em R\$/kg, separadamente. Após, o procedimento consiste no emprego da metodologia Box-Jenkins mediante transformação prévia das séries (ou sinais), assim denominadas no universo das *wavelets* com e sem o uso de limiares (do inglês, *thresholds*).

4.1 MODELAGEM DA SÉRIE PREÇO DO CIMENTO PORTLAND EM R\$/Kg

Para a aplicação da metodologia Box-Jenkins (ARIMA), foram utilizados dados mensais do preço do cimento Portland em R\$/kg no estado do Rio Grande do Sul, no período de setembro/1995 a dezembro/2016.

Para tanto a partir da inserção dos valores da série original supramencionada no software EViews 9.5 *Student Version*, foi obtido um gráfico de sequência com 256 observações que podem ser visualizadas na Figura 8, permitindo-se a inferência sobre o comportamento da variável de interesse. A escala utilizada no eixo das abscissas refere-se aos meses e no eixo das ordenadas aos preços que está na escala 1:0,1.

Figura 8 - Série original de preços mensais do cimento Portland no RS no período de setembro/1995 a dezembro/2016

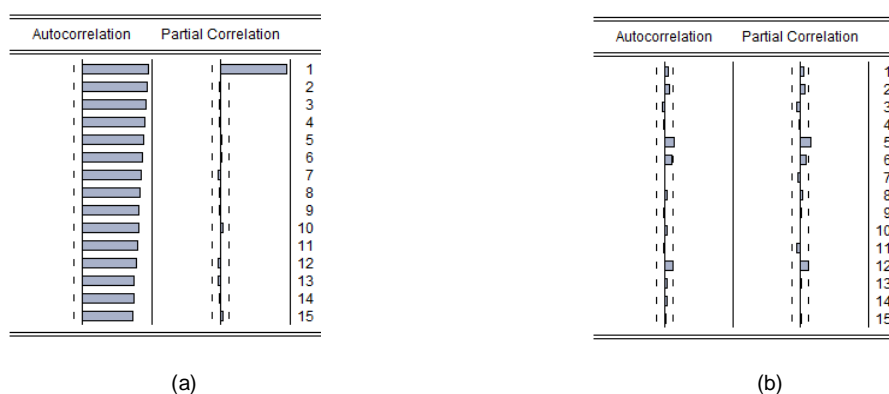


Fonte: Autora

Portanto, a partir de análise da Figura 8, foi possível verificar que a série gerada com dados do preço do cimento Portland em R\$/kg, no estado do Rio Grande do Sul, tem comportamento não estacionário. Percebeu-se, ainda, que o preço do cimento Portland no RS apresentou crescimento aparentemente linear: tendência crescente. Conjuntamente, através dessa análise, verificaram-se alguns indícios de variações sazonais. No entanto, isso pode ser contornado segundo Morettin (2004), por meio de um componente determinista.

Houve necessidade de recorrer ao correlograma da série original e em primeira diferença, para obtenção do indicativo do filtro que melhor representa o conjunto de dados. A partir do correlograma da série original, Figura 9, o decaimento lento que é observado no gráfico da FAC é característico da não estacionariedade das séries em questão, o que confirma os pressupostos apresentados anteriormente na análise gráfica da série de preços do cimento Portland em R\$/Kg no estado do Rio Grande do Sul, Figura 8. Mas é por meio do gráfico da FACP, que se obtém o indicativo quanto ao modelo que melhor representa a série. O fato de a FACP possuir apenas um lag significativo remete a um autorregressivo de ordem 1, AR(1). Nesse sentido, a partir da comparação entre os resultados demonstrados na Figura 9, que a série em 1ª diferença identifica uma sequência de variáveis aleatórias, não correlacionadas entre si e identicamente distribuídas, com média zero e variância constante, ou seja, $RB \sim N(0, \sigma^2)$. Portanto, percebe-se a obtenção do ruído branco, resultado importante nesse tipo de análise, tal como destaca Morettin (2004).

Figura 9 – Correlogramas FAC e FACP da série original de preços do cimento Portland em R\$/Kg (a), e em 1ª diferença (b)



Fonte: Autora

A seguir, apresenta-se o ajuste dos modelos da classe geral ARIMA em relação a cada uma das variáveis em estudo. Em tais modelos, ditos concorrentes, foram comparados e analisados os respectivos resíduos, e a obtenção ou não de ruído branco. Para essa finalidade, recorreu-se, assim, às estatísticas de previsão, conforme é apresentado na Tabela 2. Com base em Marchezan e Souza (2010), destaca-se em negrito o melhor resultado para a série de acordo com as medidas de previsão RMSE e U-Theil consideradas.

Tabela 2 - Modelos concorrentes para a variável preço do cimento Portland em R\$/kg no estado do Rio Grande do Sul

MODELO	PARÂMETROS	RMSE	U-Theil	AIC	BIC
ARIMA(1,1,0)	$\phi_1 = 0,122163$	0,0064	0,8844	-7,2496	-7,22193
ARIMA(1,1,1)	$\phi_1 = 0,983861$ $\theta_1 = -0,935014$	0,0062	0,7683	-7,2984	-7,25680
ARIMA(2,1,2)	$\phi_2 = 0,961470$ $\theta_2 = -0,87631$	0,0062	0,7879	-7,2651	-7,24346
SARIMA (1,0,0)(1,0,0) ₁₂	$\phi_1 = 0,999806$ $\Phi_1 = 0,189993$	0,0062	0,00956	-7,2292	-7,18774

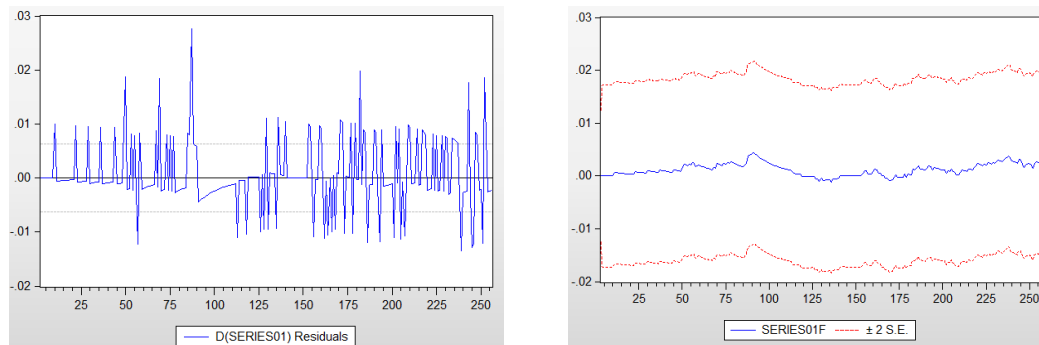
Fonte: Autora

Para a identificação dos modelos significativos para a representação da série do preço do cimento Portland em R\$/kg, faz-se a análise criteriosa dos modelos concorrentes, constantes na Tabela 2, que deverão apresentar os menores valores de AIC e BIC, ditos critérios penalizadores, e análise das estatísticas de previsão contribuem para a escolha do melhor modelo.

Dentre esses modelos, baseando-se em Box e Luceño (1997), optou-se pelo que apresentou melhores estatísticas de previsão, RMSE e U-Theil, juntamente com os critérios penalizadores AIC e BIC de menor valor sinalizando o modelo mais propício para as previsões desejadas. Levou-se em consideração o modelo mais parcimonioso, facilitando a escolha do modelo mais adequado para a série em estudo, tornando possíveis as previsões necessárias.

Por esse caminho, o objetivo desta análise visa encontrar o modelo que descreva adequadamente a série de preços de maneira que resultem implicações compatíveis com o esperado, acredita-se que se justifica a escolha do modelo: ARIMA(1,1,1) para a série preço do cimento no Rio grande do Sul, por meio da checagem dos gráficos dos resíduos e obtenção do ruído branco, indicando-se que esse modelo captou todas as informações do processo, permitindo a adequada previsibilidade para os preços do cimento Portland, em R\$/kg, o que pode ser observado na Figura 10.

Figura 10 – Resíduos da série diferenciada de preços mensais do cimento Portland no RS e Previsão segundo modelo ARIMA(1, 1, 1)



Fonte: Autora

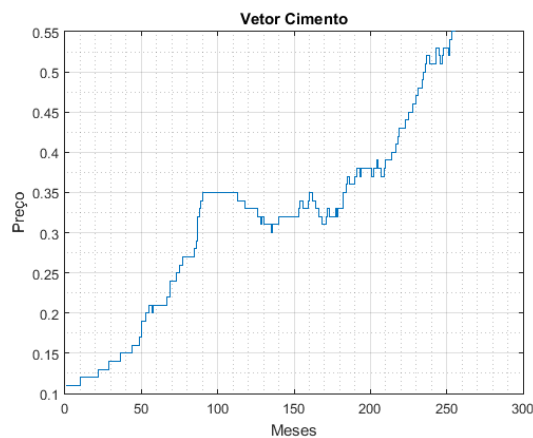
As referidas estatísticas indicam ao avaliador, a partir de seus valores, o melhor modelo dentre os concorrentes para a elaboração da previsão.

Foram analisados os modelos concorrentes, Tabela 2, cuja indicação do como melhor modelo para prever o preço do cimento Portland no RS, deu-se a partir dos menores valores dos critérios penalizadores AIC e BIC e às estatísticas de previsão RMSE e U-Theil. Os primeiros definem os modelos parcimoniosos e mostram os menores valores, os segundos “dizem” o quanto o modelo explica do esperado. Assim, o modelo representativo para a série de preços do cimento Portland no estado do Rio Grande do Sul é um ARIMA(1,1,1). Cumpre-se assim, uma das etapas deste estudo.

Sendo assim, no que se refere à análise dos dados por um procedimento auxiliar, baseado na transformada *wavelet* discreta de Haar, aqui denominado *wavelet*-ARIMA, o processo de decomposição em aproximações e detalhes, por

transformada *wavelet* discreta de Haar, para a série de preços do cimento Portland em R\$/kg, dá-se a partir do vetor cujas coordenadas são os dados originais, o qual chamaremos de $S = \text{vetor_cimento}$, neste estudo. Assim, descreve-se $\text{vetor_cimento} = [0.1100, 0.1100, 0 \dots 0.5500, 0.5500, 0.5500]$, cuja representação gráfica por processo de decimação segue como na Figura 11, quando a série original é aproximada por uma função escala. A saber, é comum tratar uma série como um sinal, que pode apresentar ruídos ou não, tal como sugere (TAN et al.,2010).

Figura 11 - Sinal original do preço cimento decimado por função escala



Fonte: Autora

Como mencionado no capítulo 2, o processo de decomposição do vetor cimento (Figura 11), segue conforme o exposto na Figura 3, por meio de médias aritméticas sucessivas das coordenadas do vetor 2 a 2, para as aproximações e diferenças para a obtenção dos coeficientes *wavelets* (detalhes). Assim, a cada nível, foram obtidos vetores aproximação, identificados por C_j e vetores detalhes ou coeficientes *wavelets*, identificados por D_j , ($j = 1,2,3, \dots$), em que o sinal é definido por uma expansão na base de Haar, contendo inicialmente 256 amostras. Na sequência, apresentam-se, na Figura 12, aproximações para o sinal, que dependem do número de coeficientes mantidos conforme a estratégia da análise multirresolução.

E é a partir do vetor cimento = S , reconstruído, que se dá início à metodologia Box-Jenkins na busca pelo modelo mais acurado, procedendo conforme a Figura 6, seção 2.2 desta dissertação. A saber, tem-se uma representação multirresolução da relação 1:1 de S com todos os vetores obtidos nos níveis de decomposição

ordenados em blocos. De onde, segundo Silveira e Kozakevicius (2015), Morettin (2014), Reis e Silva (2004):

$$S \leftrightarrow [C_7 D_7 D_6 D_5 D_4 D_3 D_2 D_1] \quad (29)$$

Assim, na Figura 12, pode-se conferir graficamente o processo de decomposição em 7 níveis de resolução da série de preços mensais do cimento Portland em R\$/kg.

Figura 12 – Decomposição da série de preços mensais do cimento por Haar em 7 níveis de aproximações (C) e detalhes (D) (continua)

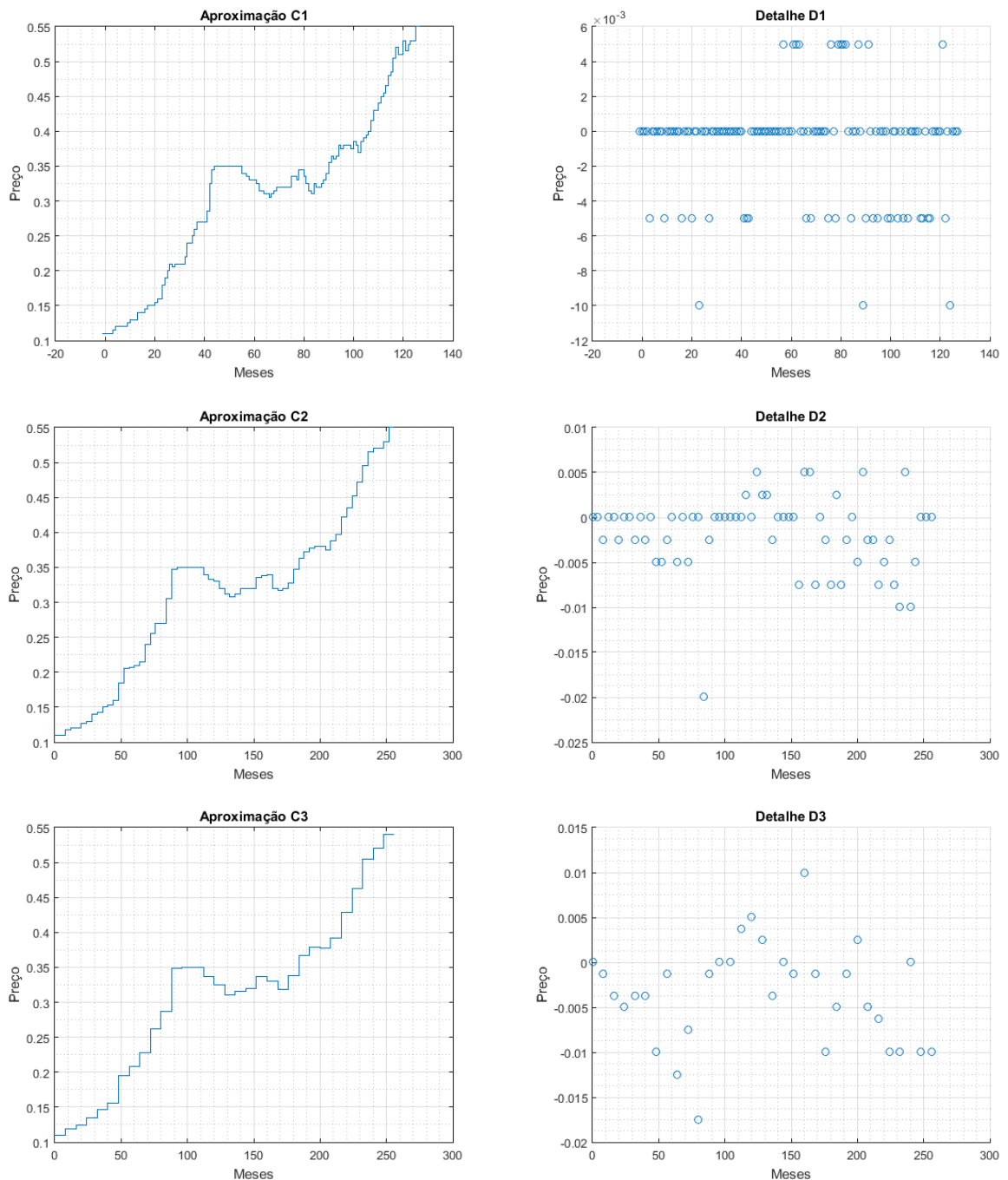


Figura 12 – Decomposição da série de preços mensais do cimento por Haar em 7 níveis de aproximações (C) e detalhes (D) (continuação)

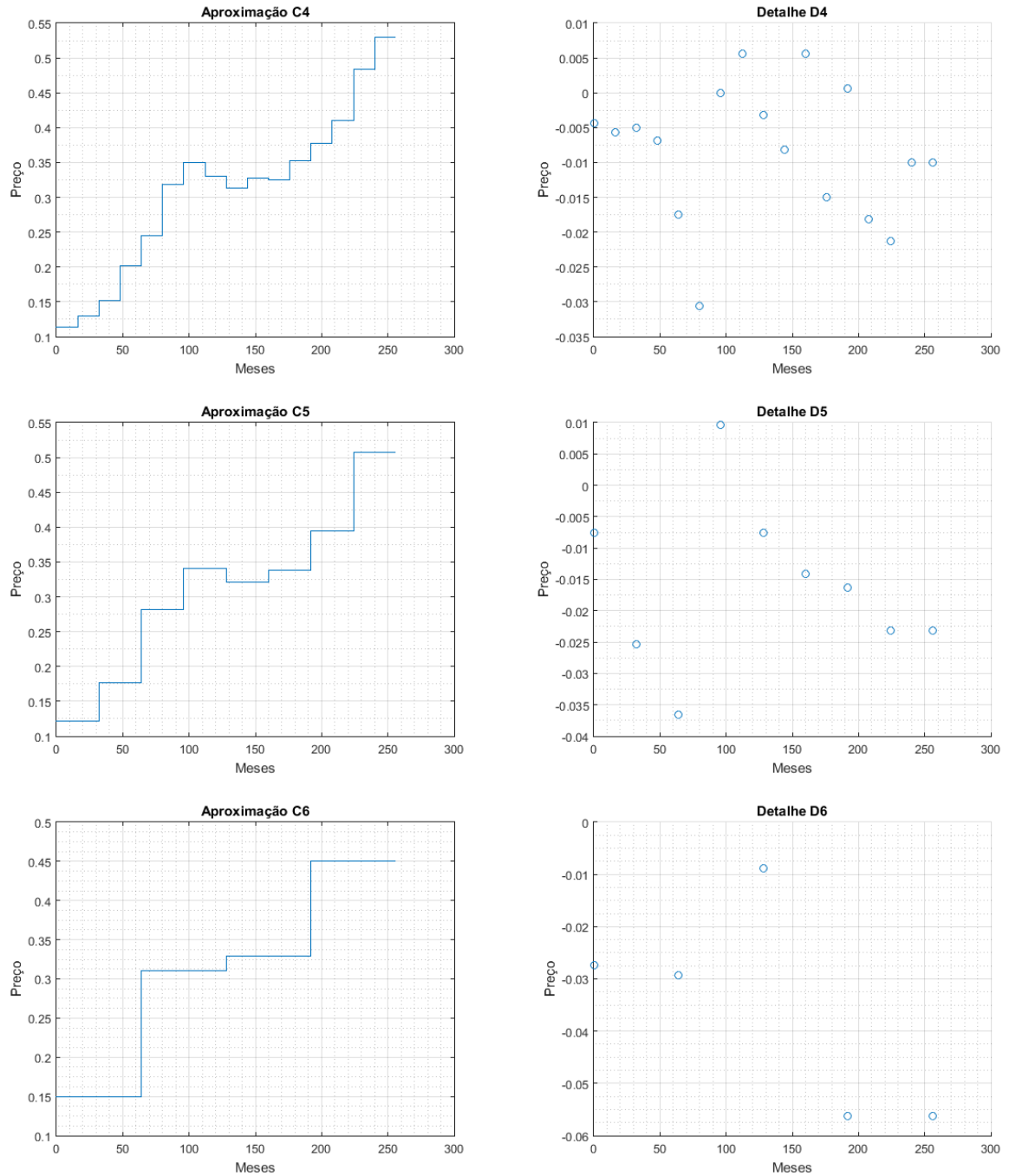
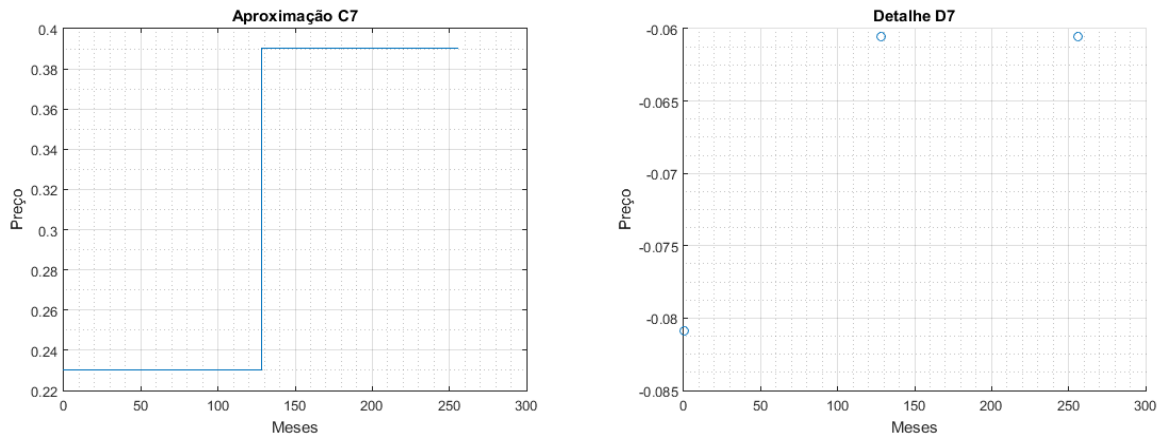


Figura 12 – Decomposição da série de preços mensais do cimento por Haar em 7 níveis de aproximações (C) e detalhes (D) (conclusão)



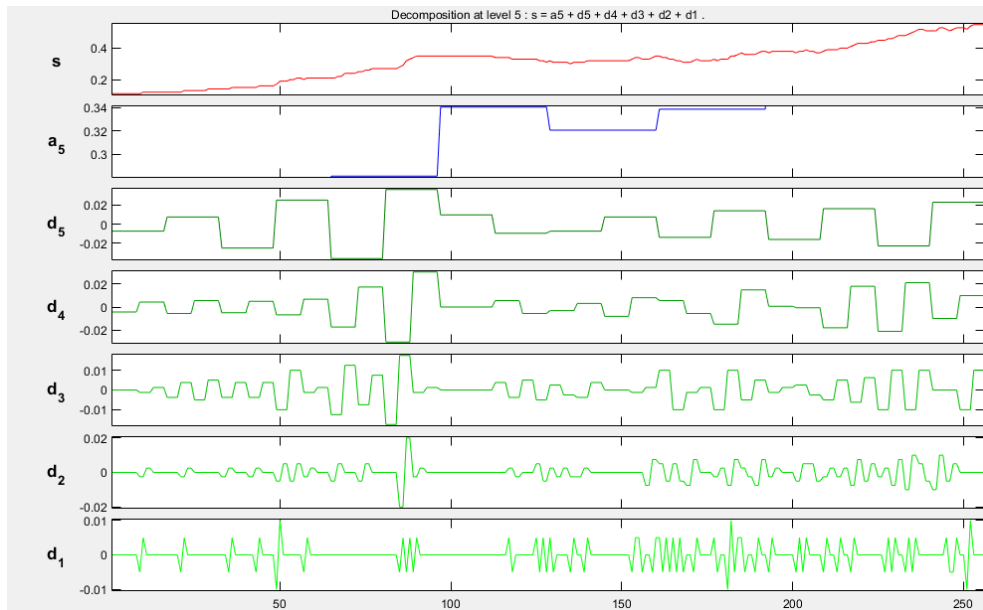
Conforme se observa no conjunto de gráficos apresentados na Figura 12, o processo de dilatação e translação confere nível a nível a utilização da *wavelet* de Haar na análise da série de preços do cimento Portland em R\$/kg. A escolha dessa *wavelet*-mãe encontra subsídios em Kozakevicius e Bayer (2014). Visto que a frequência é importante para a análise dos dados, neste procedimento será considerado, no processo de reconstrução do vetor S , a decomposição até o nível 5, uma vez que o nível mais fino como se pode observar em C7, gráfico da aproximação no nível 7, Figura 12, remeteria a valores não tão significativos, como um vetor constante, cujas coordenadas sejam todas nulas. Observa-se, ainda, que a dimensão dos vetores de complemento vão sendo reduzidos à metade, conforme a Equação (30):

$$\dim S = 256$$

$$\dim C_1 = \frac{\dim S}{2} = 128 \quad (30)$$

A Figura 13, obtida a partir de componente computacional do *software* MATLAB® R2015a, ilustra o que foi mencionado anteriormente a respeito da dimensão dos vetores aproximação, e a nova decomposição em cinco níveis. A cada passagem de nível é possível se perceber a modificação do sinal original pelos filtros passa-alta e passa-baixa da TWD de Haar.

Figura 13 - Decomposição da série preço do cimento Portland em cinco níveis



Fonte: Autora

Ao analisar a Figura 13, percebe-se a perfeita reconstrução do sinal representada por S em uma escala com amplitude maior. O nível de detalhamento escolhido permite o estudo de seus componentes de duas formas diferentes como uma representação multirresolução tal como sugere Silveira e Kozakevicius (2015), conforme abordado nesta pesquisa, ou por meio da convolução de S (abordagem para estudos posteriores), com o filtro passa-baixa para aproximação C e com o filtro passa-alta para detalhe D , seguido por decimação diádica, ou seja, a cada saída do filtro passa-baixa, descarta-se uma delas, determinando assim o conjunto de coeficientes *wavelets* (REIS e SILVA, 2004).

Nesse caso, ao utilizar a metodologia Box-Jenkins em S , segundo os mesmos critérios apontados para a série original, a busca pelo modelo com mais parcimônia e menores AIC e BIC, em que as estatísticas de previsão RMSE e U-Theil também foram apuradas, resultou no que pode ser observado na Tabela 3, na qual, dentre os modelos significativos, com base em Marchezan e Souza (2010) o escolhido encontra-se em negrito.

Tabela 3 - Modelos concorrentes para o sinal S reconstruído

(continua)

MODELO	PARÂMETROS	RMSE	U-Theil	AIC	BIC
ARIMA(1,1,0)	$\phi_1 = 0,122163$	0,00641	0,0088	-7,2496	-7,22193
ARIMA(1,0,0)	$\phi_1 = 0,999867$	0,00641	0,0097	-7,2025	-7,1748

Tabela 3 - Modelos concorrentes para o sinal S reconstruído

					(conclusão)
ARIMA(2,0,0)	$\phi_1 = 1,1221$	0,00642	0,0096	-7,2098	-7,1682
	$\phi_2 = -0,122$				
SARIMA(1,0,0)(1,0,0) ₁₂	$\phi_1 = 0,999806$	0,00646	0,0095	-7,2292	-7,18774
	$\Phi_1 = 0,189993$				
SARIMA(1,0,0)(1,0,0) ₂	$\phi_1 = 0,9998$	0,00641	0,0096	-7,2147	-7,1732
	$\Phi_1 = 0,14054$				

Fonte: Autora

A fim de promover uma escolha adequada, levam-se em conta as estatísticas de seleção de previsão RMSE e U-Theil, escolhe-se o modelo ARIMA(1,1,0).

Atribui-se a reconstrução quase perfeita do sinal pela transformada *wavelet* discreta inversa de Haar ao fato de alguns modelos constantes na Tabela 3, terem apresentado as mesmas estatísticas. Ou seja, estatísticas com os mesmos valores de parâmetros da Tabela 2. Isso pode ser confirmado a partir do algoritmo com funções específicas do *software* MATLAB e os resultados são apresentados na Tabela 4, uma vez que a diferença entre o sinal reconstruído e o sinal original é da ordem de 10^{-15} .

Tabela 4 - Formulação Algorítmica para a diferença entre sinal original e sinal reconstruído do preço do cimento Portland

Algoritmo 2: S – vetor_cimento

```
s = vetor_cimento(1:256);
ls = length(s);
[c,l] = wavedec(s,5,'haar');
a0 = waverec(c,l,'haar');
err = norm(s-a0)
err = 3.7897e-15
```

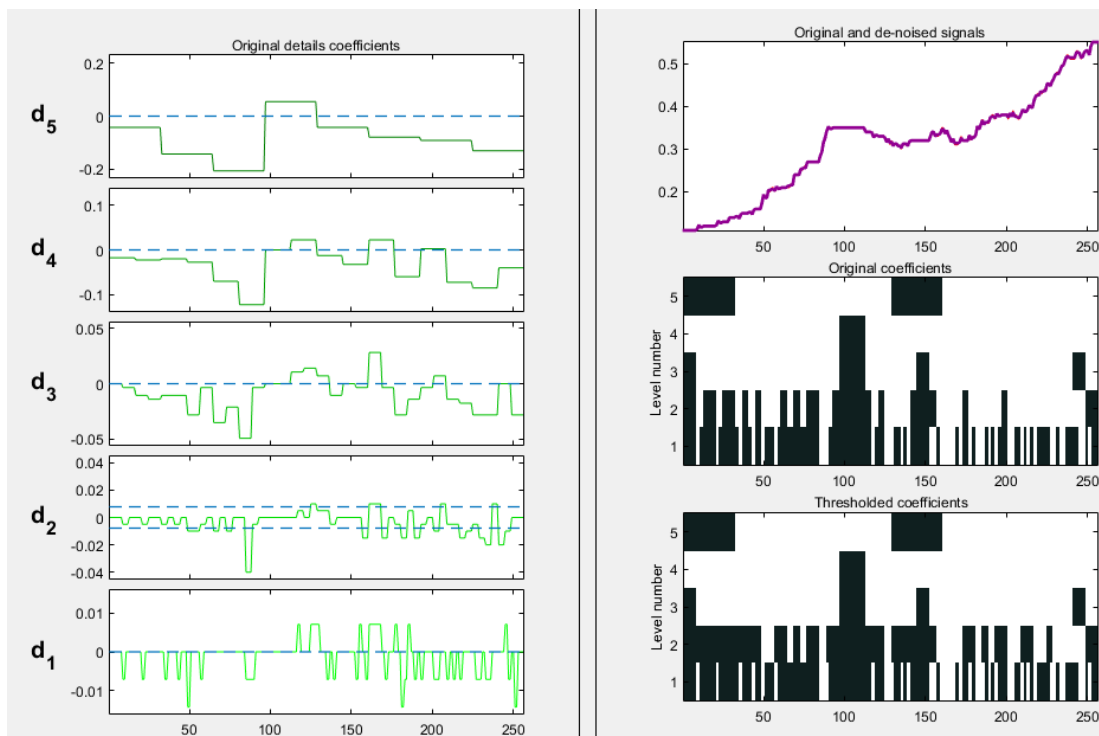
Fonte: Autora

A partir dos resultados obtidos até então, retoma-se, nesse ponto, a inspeção gráfica feita, nível a nível, na Figura 12. A partir disso, percebe-se que uma possível interferência ou limiarização (*thresholding*) dos dados entre os níveis 2 e 3 pode vir a proporcionar ganhos em termos da acurácia do modelo para previsão, conforme

exposto na revisão bibliográfica desta dissertação em Tan et al. (2010) e Kim, Yu e Song (2002).

Tal como Silveira e Kozakevicius (2010), a estratégia de limiarização (*thresholding*) necessita da obtenção prévia de um limiar de corte λ (*threshold value*) que aplicado aos coeficientes *wavelets*, detalhes, possa suavizar o comportamento do sinal analisado. Nesse sentido, na Figura 14, pode-se observar o comportamento dos detalhes ao receber um *threshold* cujo $\lambda = 0,00798$ em D_2 , quando $\hat{\sigma} = std(D_2) = 0,0042$.

Figura 14 - Decomposição da série preço do cimento Portland em cinco níveis com *threshold* em D_2

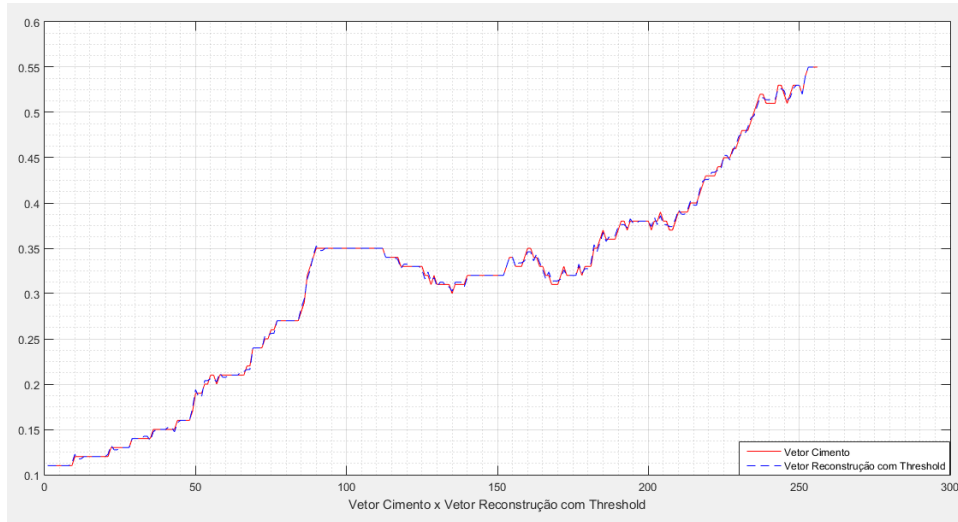


Fonte: Autora

Constatou-se, a partir dos resultados obtidos aplicando-se a série em estudo no software MATLAB, apresentados na Figura 14, que a estratégia de aplicar um *threshold* no vetor detalhes nível 2, impactou positivamente nos demais coeficientes *wavelets*. Obtendo-se, com isso, ondas mais suaves, ou ainda, com menos ruídos, em conformidade com o encontrado na literatura. A reconstrução do vetor_cimento com *threshold* pode ser observada na Figura 15, na qual apresenta-se, também, o

comparativo entre o vetor_cimento original e o vetor_cimento decomposto com limiarização (*threshold*) em D_2 .

Figura 15 - Comparação entre vetor_cimento (série original) e vetor_cimento reconstruído com *threshold*



Fonte: Autora

A partir da análise dos resultados expressos na Figura 15, percebe-se que há pouca dispersão do conteúdo das frequências originais do vetor_cimento. Ou seja, entende-se que os dados reconstruídos se ajustaram bem aos dados originais. Outras considerações poderiam ser feitas, de cunho matemático, mas não são foco desta pesquisa.

Obtém-se, a partir do vetor_cimento, representado na Figura 15, uma nova seleção de modelos, em que, a partir da metodologia Box-Jenkins, analisou-se o comportamento da série, quantificaram-se os parâmetros e as estatísticas, como os critérios penalizadores AIC e BIC e a obtenção ou não de ruído branco, procedendo integralmente como descrito na primeira etapa desta pesquisa. A partir disso, foi estimado o modelo que represente melhor a série reconstruída com *threshold*, por verificação dos modelos significativos obtidos com a ajuda do *software* E-Views, obtendo-se os resultados apresentados na Tabela 5. Novamente, salienta-se que o modelo escolhido como melhor, com base no que define Morettin (2014), foi destacado em negrito na referida tabela.

Tabela 5 - Modelos concorrentes para o sinal S reconstruído com *threshold*

MODELO	PARÂMETROS	RMSE	U-Theil	AIC	BIC
ARIMA(1,1,0)	$\phi_1 = 0,122163$	0,0064	0,0097	-7,2496	-7,2219
ARIMA(1,0,0)	$\phi_1 = 0,999867$	0,0064	0,0097	-7,2008	-7,1731
ARIMA(2,0,0)	$\phi_1 = 1,1221$ $\phi_2 = -0,122$	0,0064	0,0097	-7,2098	-7,1682
SARIMA(1,0,0)(1,0,0) ₁₂	$\phi_1 = 0,99980$ $\Phi_1 = 0,1916$	0,0064	0,0095	-7,2286	-7,1871

Fonte: Autora

Analogamente ao descrito no processo de reconstrução sem *threshold*, recorre-se ao RMSE e U-Theil para selecionar a melhor previsão neste caso. Para tanto, o modelo escolhido foi o ARIMA(1,1,0) com menores valores para AIC e BIC obtidos a partir da plotagem da série preço do cimento no *software* EViews 9.5 *Student Version*.

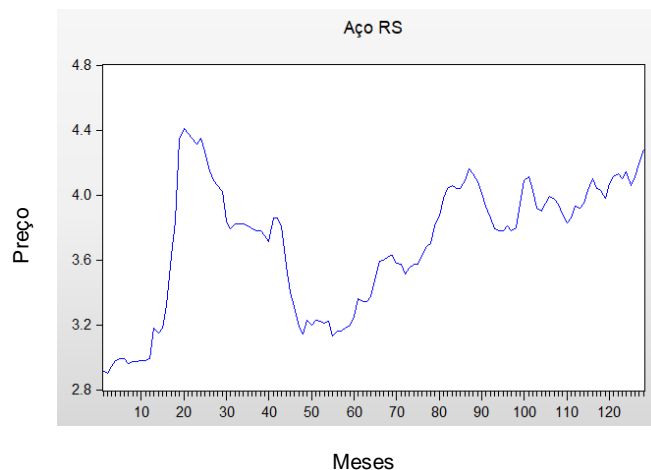
4.2 MODELAGEM DA SÉRIE PREÇO DO AÇO EM R\$/Kg

Com base nos objetivos deste estudo, neste momento, serão elaborados os modelos para a série de preços do aço, em R\$/Kg, no Rio Grande do Sul. Nesse sentido, convém salientar que os dados foram obtidos junto a Câmara Brasileira da Indústria da Construção – CBIC. A abordagem para a obtenção da previsão quanto à evolução do preço do aço no Rio Grande do Sul em R\$/kg ocorre de maneira um pouco diferente em relação ao período analisado. Levou-se em conta, o período disponibilizado em tabela a partir de fevereiro de 2007 no sítio da CBIC. Período, de previsão, diferente da avaliação dos modelos significativos para a *commodity* cimento, no qual foi avaliado o período disponibilizado em tabela, pela CBIC, a partir de setembro de 1995, por conveniência para execução da análise optou-se pelo término em dezembro de 2016. Portanto, a investigação quanto ao comportamento da série original do preço do aço é feita a partir do gráfico de sequência, Figura 16, com 128 observações obtidas no período de fevereiro/2007 a setembro/2017. A

partir do gráfico é possível se presumir a estacionariedade, possível tendência e/ou sazonalidade que esta série possa vir a apresentar.

Os modelos matemáticos, ARIMA, captam o comportamento da correlação seriada, ou autocorrelação entre os valores da série temporal. E, a partir destes, será possível fazer previsões futuras, objetivo desta pesquisa. A seguir, na Figura 16, é possível avaliar por inspeção visual o comportamento da série original da evolução do preço do aço.

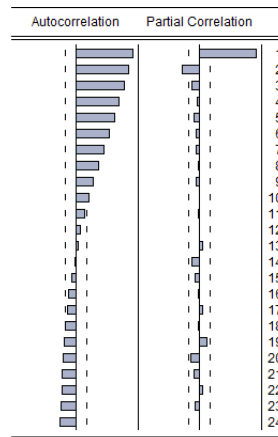
Figura 16 - Série original de preços mensais do aço CA 50a d=10mm no RS no período de fevereiro/2007 a setembro/2017



Fonte: Autora

Observando o gráfico obtido para a série original, Figura 16, verifica-se um comportamento não estacionário com uma tendência crescente. Percebe-se, além disso, que a série apresenta períodos alternados de altas e baixas variações, característica de uma série aditiva. No entanto, como é perceptível, há necessidade de recorrer a FAC e FACP a fim de confirmar tais suposições. Para tal finalidade utilizou-se o *software* EViews 9.5 *Student Version* a fim de gerar a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP), apresentadas na Figura 17, e, a partir destas, perceber o filtro que mais se adequa à série de preços do aço especificado anteriormente.

Figura 17 - Função de autocorrelação (FAC) e Função de autocorrelação parcial (FACP) da série de preços mensais do aço no RS



Fonte: Autora

Ao se observar os resultados expressos na Figura 17, pode-se verificar que a série decaiu para zero antes do lag 12. Com base nisso, é possível, portanto, afirmar que a série possui um comportamento estacionário. Na FACP, observam-se dois lags significativos o lag 1 e o lag 2, o que indica um modelo autorregressivo de ordem 2. Seguiu-se a etapa de estimação dos parâmetros e resíduos, a fim de obter modelos concorrentes para a série de preços do aço no RS, Tabela 6, em que a indicação do melhor modelo teve como medida de seleção aquela que apresentasse menores valores de AIC e BIC, e das estatísticas U-Theil e RMSE, tal como sugere SOUZA et al.,(2010).

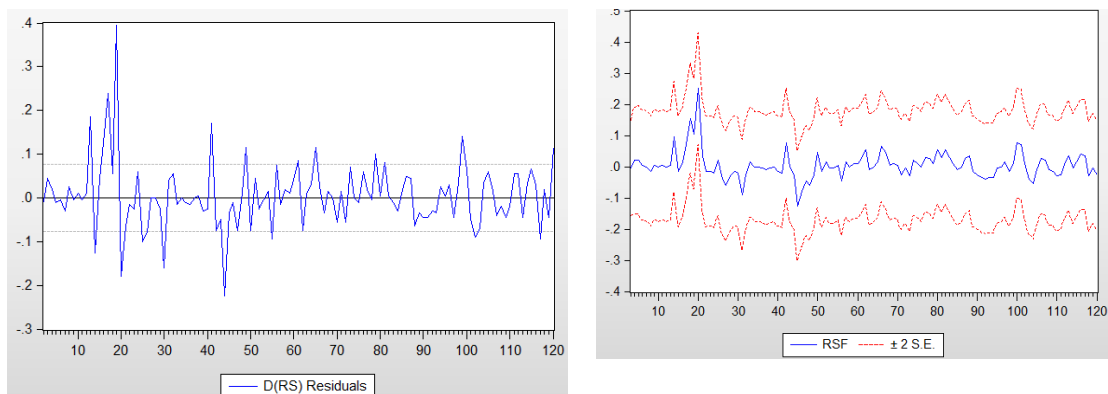
Tabela 6 - Modelos concorrentes para a variável preço do aço em R\$/Kg no estado do Rio Grande do Sul

MODELO	Parâmetros	RMSE	U-Theil	AIC	BIC
ARIMA(1,1,0)	$\phi_1=0,4928$	0,7514	0,5830	-2,3129	-2,2681
ARIMA(0,1,1)	$\theta_1=0,3727$	0,7787	0,6769	-2,2347	-2,1899
ARFIMA(1,1,0)	$\phi_1=0,8832$ $d=0,4956$	0,0750	0,0100	-2,2365	-2,1697
ARIMA(1,0,1)	$\phi_1=0,9995$ $\theta_1=0,3727$	0,07514	0,5830	-2,1593	-2,0924

Fonte: Autora

A partir disso, avaliam-se os modelos testados na Tabela 9. Percebe-se que os critérios AIC e BIC fornecem valores próximos entre os modelos testados. Por esse motivo, com base em Souza et al. (2010), houve a necessidade de recorrer a outras estatísticas como U-Theil e RMSE, definindo ARIMA(1,0,1) como o melhor modelo para a série em questão.

Figura 18 – Resíduos da série diferenciada de preços mensais do aço no RS e Previsão segundo modelo ARIMA(1, 1, 0)



Fonte: Autora.

Pode-se afirmar que a metodologia Box-Jenkins possibilita realizar previsões de forma eficiente, caracterizando-se como uma ferramenta eficaz para séries com o tipo de comportamento como a estudada neste trabalho. Foram encontrados modelos concorrentes que apresentaram bons resultados. Entretanto, ao utilizar critérios penalizadores e de previsão para efetuar a escolha do melhor modelo, optou-se pelo modelo ARIMA(1,1,0), com média zero e variância constante, ainda que os valores previstos por tais modelos fossem representativos da realidade, o que pode ser comprovado pelas estatísticas de previsão, com base em Souza et al., (2010).

A seguir, para a busca do modelo acurado inicia com o processo de decomposição do vetor em sete níveis cujas coordenadas são os preços mensais do aço no estado do Rio Grande do Sul, avaliando como esses dados se comportam até o nível mais fino dessa decomposição, conforme será apresentado na Figura 20. Antes disso, na Figura 18, tem-se o vetor aço original decimado por funções escala, estágio anterior à decomposição, obtida na inserção dos dados no MATLAB.

Figura 19 - Sinal original do aço decimado por funções escala



Fonte: Autora

Comparando-se os dados nas Figuras 16 e 19, observa-se que as características do sinal original foram mantidas, seguindo-se à decomposição em 6 níveis de resolução da série de preços mensais do aço em R\$/kg no Rio Grande do Sul, apresentada no conjunto de gráficos da Figura 20.

Figura 20 - Decomposição da série de preços mensais do aço por Haar em 6 níveis de aproximações (C) e detalhes (D)

(continua)

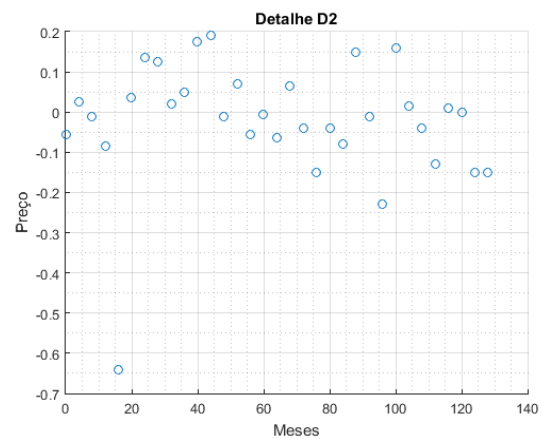
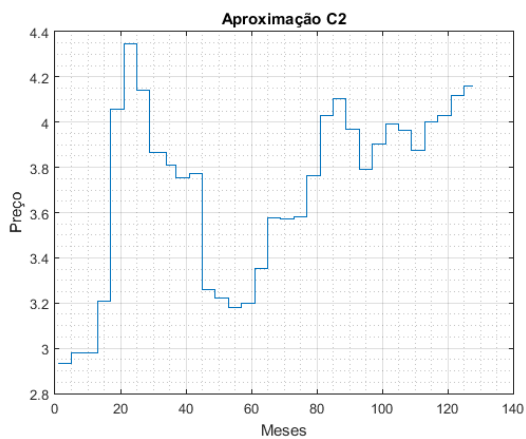
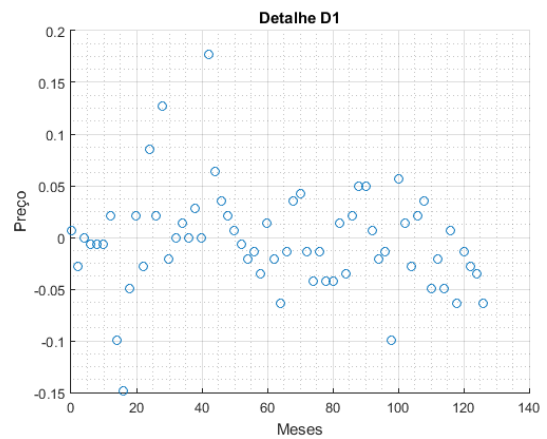
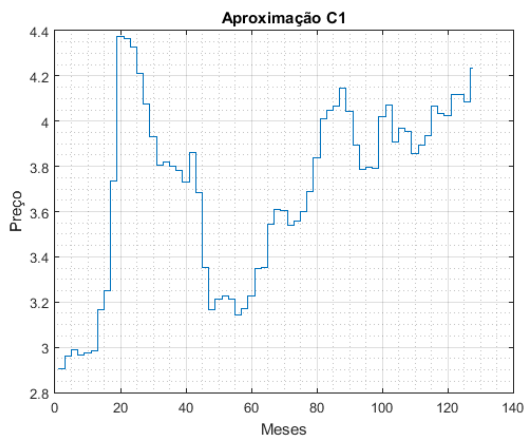


Figura 20 - Decomposição da série de preços mensais do aço por Haar em 6 níveis de aproximações (C) e detalhes (D) (continuação)

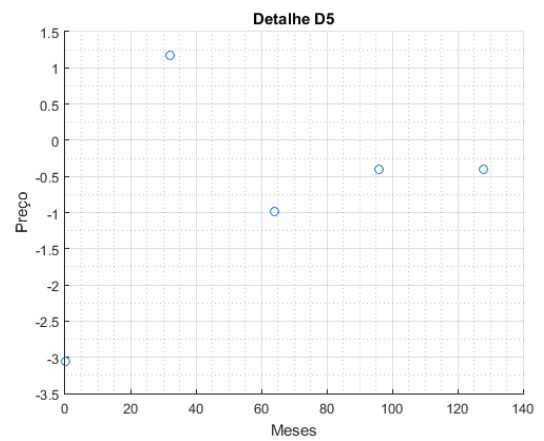
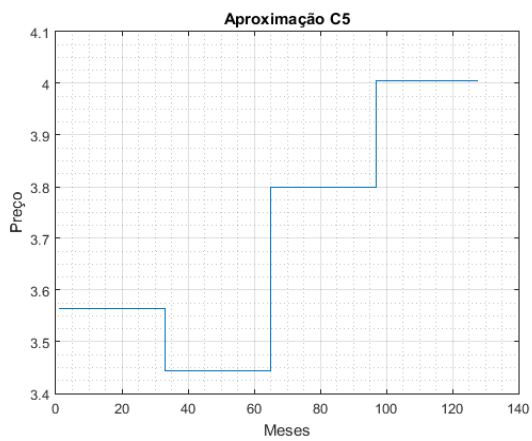
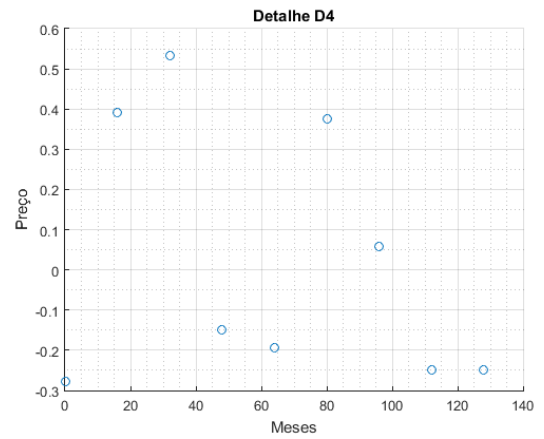
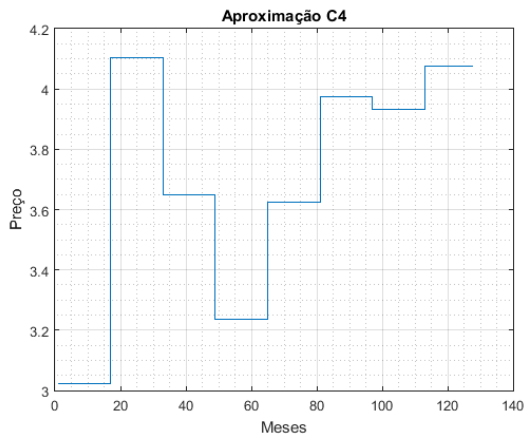
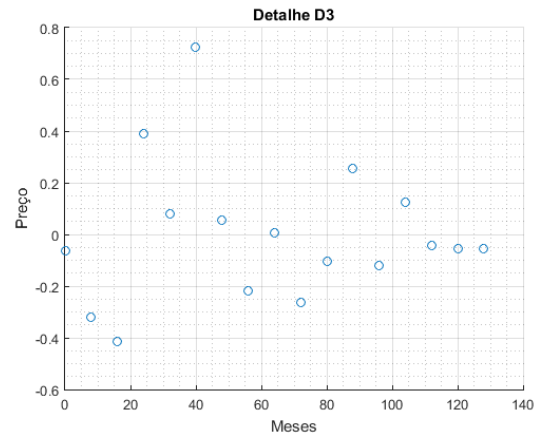
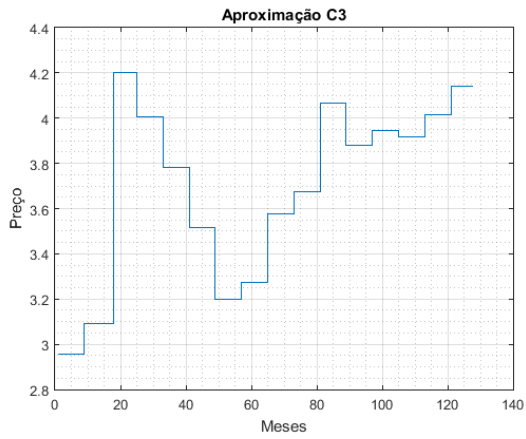
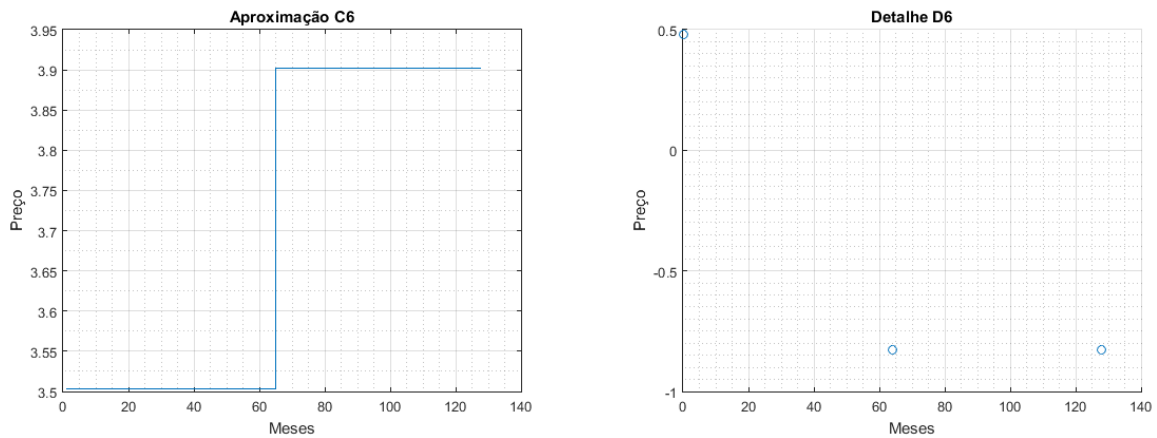


Figura 20 - Decomposição da série de preços mensais do aço por Haar em 6 níveis de aproximações (C) e detalhes (D) (conclusão)



Fonte: Autora.

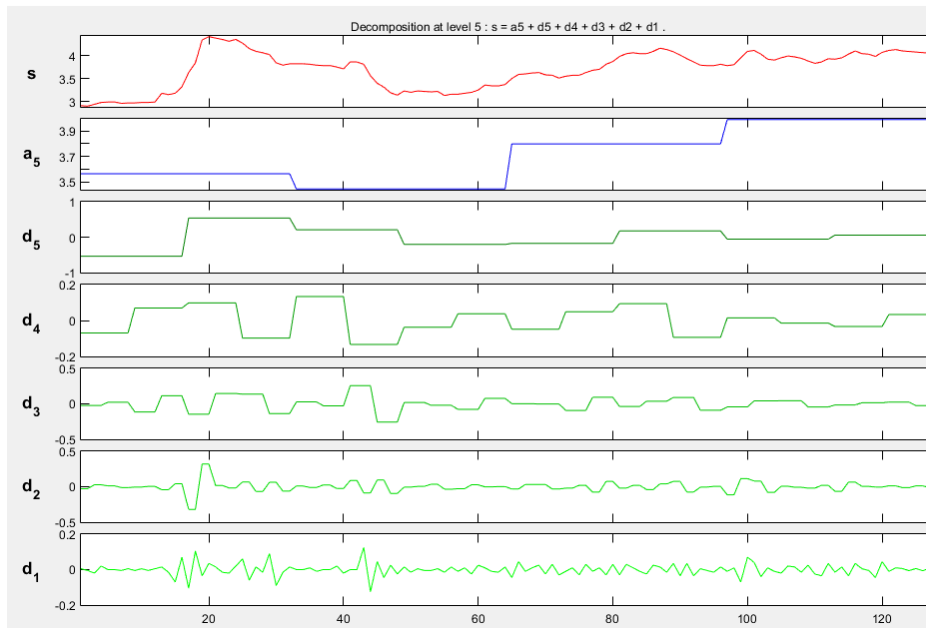
Recorre-se, aqui, ao mesmo procedimento efetuado com a série de preços do cimento Portland, optando-se por um processo de reconstrução do vetor aço em um nível anterior ao nível 6, como apresentado na Figura 20, em que há também redução no dimensionamento do vetor original.

A seguir, na Figura 21, obtida por processo computacional via *software* MATLAB, apresenta-se graficamente, a redução no dimensionamento do sinal por meio de seus filtros e decomposição em cinco níveis, $j = 5$. O vetor S é representado nesse processo como:

$$S = [C_5, D_5, D_4, D_3, D_2, D_1] \quad (31)$$

Por reconhecimento computacional, o MATLAB trata aproximações como a_5 . Leia-se: $a_5 = C_5$.

Figura 21 - Decomposição da série de preços do aço em R\$/Kg em cinco níveis



Fonte: Autora.

A partir dos resultados expressos na Figura 21, é perceptível a suavização das ondas nível a nível e a reconstrução do sinal S. Nesse sentido, novamente escolheu-se o mesmo nível de detalhamento do vetor_cimento para efeitos de comparação no intuito de evidenciar o fato de que essa escolha é arbitrária. Após, seguiu-se à busca do modelo que melhor representa o sinal do aço reconstruído pela TWD inversa de Haar, Tabela 7, no qual se apresenta entre os testados, os mais significativos.

Tabela 7 - Modelos concorrentes para a variável aço por *wavelet*-ARIMA sem *threshold*

MODELO	PARÂMETROS	RMSE	U-THEIL	AIC	BIC
ARIMA(2,1,0)	$\phi_1 = -0,0893$ $\phi_2 = 0,6738$	0,2390	0,6504	0,7415	0,8106
SARIMA(1,0,0)(1,0,0) ₂	$\phi_1 = 0,8658$ $\Phi_1 = 0,6415$	0,2003	0,6528	0,7587	0,8222
SARIMA(2,1,0)(1,1,0)₂	$\phi_1 = -0,0533$ $\phi_2 = 0,8699$ $\Phi_1 = 0,2963$	0,1747	0,5881	0,7121	0,8060

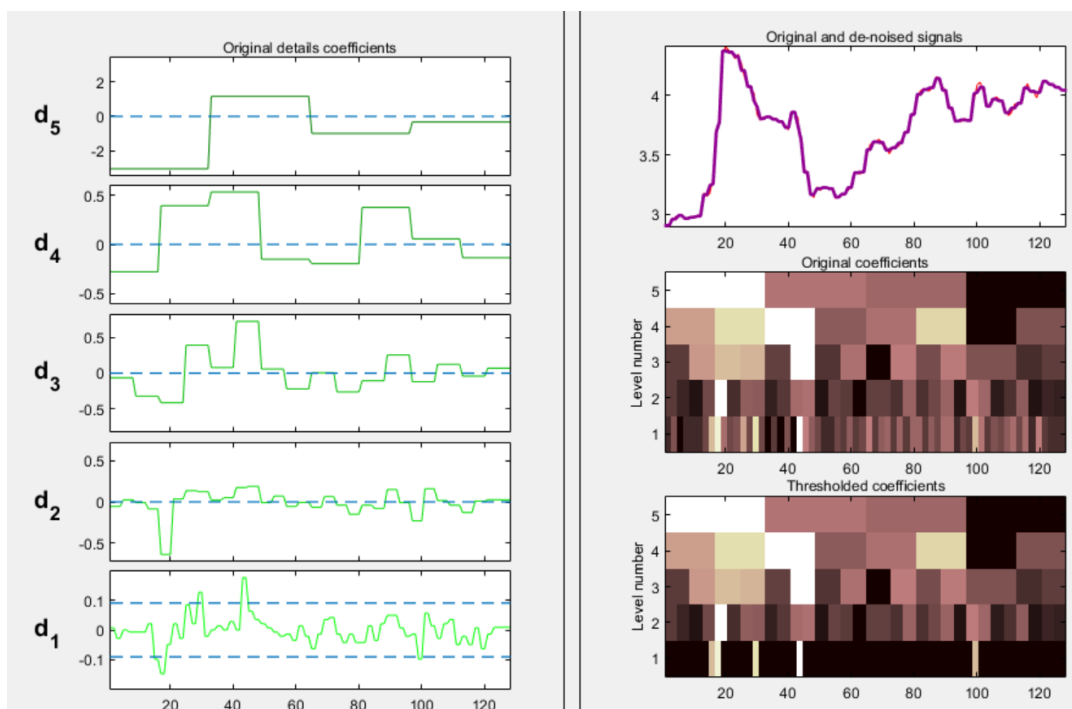
Fonte: Autora

Após a análise, conclui-se que o modelo que melhor se ajustou ao vetor_ aço reconstruído, apontando, segundo as estatísticas AIC, BIC, U-Theil e RMSE e levando-se em conta o princípio da parcimônia, o modelo SARIMA(2,1,0)(1,1,0)₂.

Pode-se observar, na Figura 22, a limiarização do sinal com um *threshold* aplicado em D_2 . O comportamento dos detalhes a partir do nível 2, foi modificado por um *threshold* cujo $\lambda = 0,2856$, quando $\hat{\sigma} = std(D_2) = 0,1503$.

A Estratégia de aplicar um *threshold* tem como objetivo eliminar possíveis ruídos que possam interferir na análise acurada do modelo procurado. Com o auxílio do *software* MATLAB, chegou-se aos gráficos apresentados a seguir.

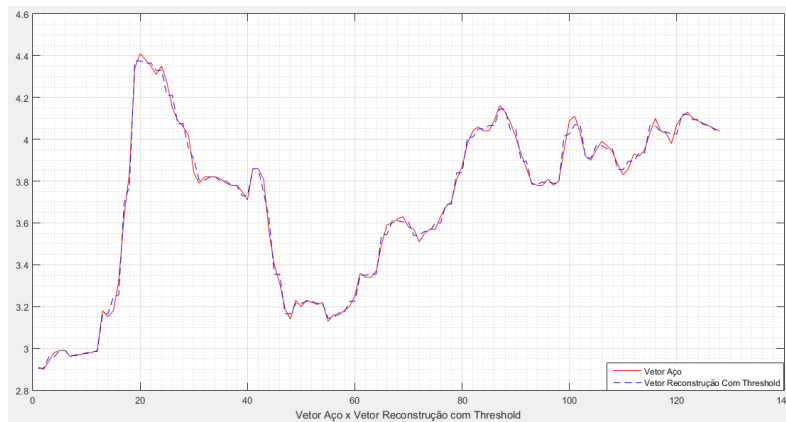
Figura 22 - Decomposição da série preço do aço em cinco níveis com *threshold* em D_2



Fonte: Autora

A partir da verificação de resultados expressos na Figura 22, de fato, obtém-se com o corte (*threshold*) no vetor detalhe do nível 2, uma melhora significativa em termos de suavização das ondas do sinal preço do aço no Rio Grande do Sul. A seguir, na Figura 23, apresenta-se o comparativo entre os vetores: vetor_ aço e vetor aço reconstruído com *threshold*.

Figura 23 - Comparação entre série original de preços do aço e vetor aço reconstruído por TWD inversa de Haar com *threshold*



Fonte: Autora

A partir de observação dos resultados a Figura 23, percebeu-se que o vetor reconstruído com *threshold* atendeu à proposta de reconstrução, se ajustando bem ao vetor série original de preços do aço, cujas coordenadas são os preços mensais do aço no Rio Grande do Sul, apresentando uma diferença da ordem de 10^{-14} , como se pode verificar na Tabela 8, em que se apresenta a formulação algorítmica para obtenção da referida diferença.

Tabela 8 - Formulação Algorítmica para a diferença entre vetor série original e vetor aço reconstruído por TWD inversa de Haar

Algoritmo 3: S – vetor_aço

```
s = vetor_aco(1:128);
ls = length(s);
[c,l] = wavedec(s,5,'haar');
a0 = waverec(c,l,'haar');
err = norm(s-a0)
err = 2.7921e-14
```

Fonte: Autora

Na sequência, estimaram-se modelos, para a tomada de decisão, buscando entre os modelos significativos, o que apresente melhor estatística de previsão e menores valores de AIC e BIC, tal como prevê Marchezan e Souza (2010). Os resultados obtidos estão expostos na Tabela 9.

Tabela 9 - Modelos concorrentes para a variável aço por *wavelet*-ARIMA com *threshold*

MODELO	PARÂMETROS	RMSE	U-THEIL	AIC	BIC
ARIMA(1,0,0)	$\phi_1=0,9997$	0,0939	0,01266	-1,7969	-1,7321
ARFIMA(2,1,0)	$\phi_1 = 0,4999$ $\phi_2 = 0,3789$ $d = 0,48588$	0,0883	0,0111	-1,9452	-1,8794
ARFIMA(1,1,1)	$\phi_1 = 0,9108$ $\theta_1 = -0,2733$ $d = 0,4982$	0,5262	0,07347	-1,8599	-1,7702
ARIMA(2,1,0)	$\phi_1 = -0,088$ $\phi_2 = 0,6604$	0,2279	0,6802	0,7334	0,7486

Fonte: Autora

Depois de feita a análise da Tabela 9, indica-se o modelo **ARFIMA(2,1,0)** como o escolhido, tendo como base o princípio da parcimônia em conjunto com os demais critérios já antes mencionados, primando pela previsão tida como ótima, U-Theil = 0,0111.

A comparação entre os diferentes métodos é realizada na Tabela 10, sendo eles ARIMA e *wavelet*-ARIMA, visando mostrar a empregabilidade e melhoria proporcionada pela transformada *wavelet* discreta de Haar, a cada uma das séries objeto desta pesquisa.

Tabela 10 - Modelos concorrentes para as variáveis cimento e aço por ARIMA e *wavelet*-ARIMA sem *threshold* e com *threshold*

(continua)

MODELO	PARÂMETROS	RMSE	U-THEIL	AIC	BIC
Cimento					
ARIMA(1,1,1)	$\phi_1 = 0,9838$	0,0062	0,0095	-7,2983	-7,2568
ARIMA	$\theta_1 = -0,9350$				
ARIMA(1,1,0)	$\phi_1 = 0,12216$	0,0088	0,0064	-7,2496	-7,2219
<i>wavelet</i> -ARIMA s/Threshold					
ARIMA(1,1,0)	$\phi_1 = 0,999867$	0,0064	0,0097	-7,2496	-7,2219
<i>wavelet</i> -ARIMA c/Threshold					

Tabela 10 - Modelos concorrentes para as variáveis cimento e aço por ARIMA e *wavelet*-ARIMA sem *threshold* e com *threshold*

(conclusão)

MODELO	PARÂMETROS	RMSE	U-THEIL	AIC	BIC
Aço					
ARIMA(1,1,0)	$\phi_1 = 0,4928$	0,7514	0,5830	-2,3129	-2,2681
ARIMA	$\theta_1 = 38544$				
SARIMA(2,1,0)(1,1,0) ₂	$\phi_1 = -0,0533$	0,1747	0,5881	0,7121	0,8060
<i>wavelet</i> -ARIMA s/Threshold	$\phi_2 = 0,8699$				
	$\Phi_1 = 0,2963$				
ARFIMA(2,1,0)	$\phi_1 = 0,4999$	0,0883	0,0111	-1,9452	-1,8794
<i>wavelet</i> -ARIMA c/Threshold	$\phi_2 = 0,3789$				
	$d = 0,48588$				

Fonte: Autora

5 CONCLUSÃO

Ao longo deste estudo, foi observado que a combinação dos modelos ARIMA às funções *wavelets* discretas de Haar para obtenção do melhor modelo é uma técnica que pode ser utilizada para diferentes tipos de dados. De modo geral é possível afirmar que a escolha do modelo para previsão do preço das *commodities* cimento e aço no Rio Grande do Sul foi baseada no princípio da parcimônia, e nos menores valores de AIC e BIC. Assim como, os valores das estatísticas RMSE e U-Theil.

Atribui-se que as pequenas diferenças encontradas entre as estatísticas dos modelos selecionados nas séries do cimento e do aço se devem a reconstrução obtida pela TWD inversa, com diferença entre o sinal original e o reconstruído ter sido mínima.

Fato este que está relacionado à escolha do limiar de corte, no qual optou-se de forma arbitrária pelo limiar universal, conforme exposto na literatura. Neste caso, o parâmetro de corte foi calculado tomando como base, componentes do espectro de frequências dos coeficientes *wavelets*.

Os resultados desta investigação permitiram chegar às seguintes conclusões em relação aos modelos: ARIMA(1,1,1) para a *commodity* cimento Portland, modelo obtido sem interferência da TWD de Haar, e da mesma forma, ARIMA(1,1,0) para a *commodity* aço.

Finaliza-se, destacando que os resultados encontrados para as duas *commodities* vem ao encontro dos objetivos propostos neste estudo.

Deixa-se a sugestão para futuras pesquisas, a investigação com interferência de outros limiares e/ou interferência em diferentes níveis de detalhes.

REFERÊNCIAS

- AKAIKE, H. Statistical predictor identification. **Annals of the Statistical Mathematics**. 22, p. 203-217, 1970.
- ANTONIADIS, A. Wavelet methods in statistics: some recent developments and their applications. **Statistics Surveys**, 1, p. 16–55, 2007.
- BARROS, M. **Processos estocásticos**. Rio de Janeiro: Papel Virtual, 2004.
- BAYER, F. M.; KOZAKEVICIUS, A. J. SPC-threshold: Uma proposta de limiarização para filtragem adaptativa de sinais. **Trends in Applied and Computational Mathematics**, v. 12, n. 2, p. 112-132, 2010.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis forecasting and control**. Edição revisada. Holden-Day, San Francisco, 1970.
- BOX, G. E. P. e LUCEÑO, A. **Statistical control by monitoring and feedback adjustment**.: Wiley Interscience, New York, 1997.
- CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO - CBIC
Disponível em: <http://www.cbicdados.com.br/menu/materiais-de-construcao/cimento>
Acesso em: 16 dez. 2015 e 14 nov. 2016.
- CONEJO, A. J. et al. Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models. **IEEE Transactions on Power Systems**, 20, 1035-1042, 2005.
- CONSULTEC. **A Indústria do cimento no Brasil**: aspectos de seus custos e Desenvolvimento, 1967.
Disponível em: http://www.snic.org.br/pdf/historia_do_cimento_no_brasil.pdf
Acesso em: 15 mar. 2016.
- DAUBECHIES, I. Ten Lectures on Wavelets. **Society for Industrial and Applied Mathematics**, Philadelphia, 1992, v. 1, p. 232.
- DEBNATH, L. **Wavelet Transforms and Their Applications**. Birkhauser: Boston, 2002.
- DEVLEKER, K. **Understanding wavelets, Part 1**: What are wavelets. 2016. 1 diapositivo. Disponível em: <<https://www.mathworks.com/videos/understanding-wavelets-part-1-what-are-wavelets-121279.html>>. Acesso em: 11 ago. 2016.
- DIEBOLD, F. X.; LOPES, J. A. Forecast Evaluation and Combination. **Handbook of Statistics Statistical Methods in Finance**, v. 14, 1996.
- DONOHO, D. L.; JOHNSTONE, I. M. Ideal spatial adaptation via wavelet shrinkage. **Biometrika**, v. 81, p. 425-455, 1994.

EHLERS, R. S. **Análise de Séries Temporais**. Departamento de Estatística, UFPR, 2007. Disponível em: <http://leg.est.ufpr.br/ehlers/notas>. Acesso em: 17 nov. 2016.

FERRAZ, M. I. F., SÁFADI, T.; LAGES, G. Uso de modelos de séries temporais na previsão de séries de precipitação pluviiais no município de Lavras – MG. **Revista Brasileira de Agrometeorologia**, v. 7, n. 2, p. 259-267, 1999.

FGV – Fundação Getúlio Vargas e IBRE – Instituto Brasileiro de Economia. Disponível em: <<http://portalibre.fgv.br/>>. Acesso em: 15 nov. 2016.

FRAIZER, M. W. **An Introduction to wavelets through linear algebra**. New York: Springer, 1999.

FIERGS – Federação das Indústrias do Rio Grande do Sul. 2016. Disponível em: <http://www.fiergs.org.br> . Acesso em: 19 nov. 2016.

GAO, H.-Y.; BRUCE, A. G. Waveshrink with firm shrinkage. **Statistica Sinica**, v. 7, p. 855-874, 1997.

GAO, H.-Y. *Wavelet* shrinkage denoising using non-negative garrote. **Journal of Computational and Graphical Statistics**, v. 7, p. 469-488, 1998.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2007.

GOMES, J.; VELHO, L.; GOLDENSTEIN, S. **Wavelets: Teoria, Software e Aplicações**. Rio de Janeiro: Conselho Nacional de Desenvolvimento Científico e Tecnológico, Instituto de Matemática Pura e Aplicada, 1997.

GOMES, N. R. M.; LOPES, C. M. C. Estudo de modelos de séries temporais para a demanda de produção de cimento. In: 19º SINAPE, 2010, São Pedro/SP. **Anais...** São Pedro/SP: Universidade Presbiteriana Mackenzie, PIBIC/MackPesquisa, 2008.

GROSSMANN, A; MORLET, J. Decomposition of hardy functions into square integrable *wavelets* of constant shape. **SIAM JOURNAL OF MATHEMATICAL ANALYSIS**, v. 15, n. 4, p. 723–736, 1984.

Instituto Aço Brasil. Disponível em: <http://www.acobrasil.org.br/site2015/> . Acesso em: 05 dez. 2016 e 14 maio 2017.

Instituto Nacional dos Distribuidores do Aço – INDA. Disponível em: <http://www.inda.org.br/> Acesso em: abril/2016 e agosto/2016.

JUNGES, A. H.; FONTANA, D. C. Modelo Agrometeorológico-espectral de estimativa de rendimento de grãos de trigo no Rio Grande do Sul. **Revista Ceres**, Viçosa, v. 58, n. 1, p. 9-16, 2011.

KIM, C. L; YU, I.; SONG, Y. H. Prediction of system marginal price of electricity using *wavelet* transform analysis. **Energy Conversion and Management**, v. 43 n. 14, p. 1839-1851, 2002.

KOZAKEVICIUS, A. J. et al. URL query string anomaly sensor designed with the bidimensional Haar *wavelet* transform. **International Journal of Information Security**, v. 14, n. 6, p. 561-581, 2015.

KOZAKEVICIUS, A. J.; BAYER, F. M. Filtragem de sinais via limiarização de coeficientes wavelet. **Ciência e Natura**, v. 36, Ed. Especial, p. 37-51, 2014.

LIMA, A. V. L. **Aplicação e performance da análise de intervenção em séries de emissão de gases poluentes na Cidade do México**. 2016. 58 p. Dissertação (Mestrado em Engenharia de Produção)—Universidade Federal de Santa Maria, Santa Maria, RS, 2016.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. J. **Forecasting methods and applications**. 3. ed. New York: John Wiley e Sons, 1998.

MARCHEZAN, A.; SOUZA, M. A. Previsão do preço dos principais grãos produzidos no Rio Grande do Sul. **Ciência Rural**, v. 40, n. 11, 2010.

MEYER, Y. **Ondelettes: ondelettes et opérateurs I**. Cambridge: Hermann, 1990.

MORETTIN, P. A. **Ondas e ondaletas: da análise de Fourier à análise de ondaletas de séries temporais**. 2. ed. São Paulo: EDUSP, 2014.

MORETTIN, P. A.; TOLOI, C. M. **Análise de Séries Temporais**. São Paulo: Blucher, 2004.

NUNES, R. N. et al. *Wavelets* e previsões de séries de tempo: uma análise empírica. In: SALÃO DE INICIAÇÃO CIENTÍFICA, 12., 2000 Porto Alegre/RS. **Anais...** Porto Alegre/RS: Universidade Federal do Rio Grande do Sul, 2000. p. 11-15.

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

PERLIN, T. **Um detector de anomalias de tráfego de rede baseado em wavelets**. 2010. 127 p. Dissertação (Mestrado em Computação)—Universidade Federal de Santa Maria, Santa Maria, RS, 2010.

POZZEBON, G. G. **Transformada Wavelet e Redes Neurais Artificiais na Análise de Sinais Relacionados à Qualidade da Energia Elétrica**. Dissertação (Mestrado em Engenharia Elétrica)—Universidade Federal de Santa Maria, Santa Maria, RS, 2009.

PYNDYCK, R. S.; RUBINFELD, D. L. **Econometria: modelos e previsões**. 4. ed. Rio de Janeiro: Elsevier, 2004.

REIS, A. J.; SILVA, A. P. A. Aplicação da transformada *wavelet* discreta na previsão de carga a curto prazo via redes neurais. **Revista Controle e Automação**, v. 15, n. 1, Jan./Fev./Mar. 2004.

RUSSO, S. L. **Gráficos de Controle para Variáveis Não-conformes Autocorrelacionadas**. Tese de Doutorado (Programa de Pós-graduação em Engenharia de Produção)–Universidade Federal de Santa Catarina, Florianópolis/SC, 2002.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**. v. 6, n. 2, p. 461-464, 1978.

SILVEIRA, T. L. T.; KOZAKEVICIUS, A. J.; RODRIGUES, C. R. Awake/sleep scoring through *wavelet* analysis associated to decision tree algorithms. In: XXX SIMPÓSIO SUL DE MICROELETRÔNICA. 2015, Santa Maria/RS: Sociedade Brasileira de Computação. Santa Maria/RS. **Anais...**, 2015.

SILVEIRA, T. L. T.; KOZAKEVICIUS, A. J. **Transformada *wavelet* de Haar: conceitos, formulações e aplicações: apostila**. Santa Maria: Programa de Pós-graduação em Informática, 2016. 59 p.

SINDUSCON – Sindicato da Indústria da Construção Civil no Estado do Rio Grande do Sul. 2017. Disponível em: <http://www.sinduscon-rs.com.br/> . Acesso em: 14 nov. 2016.

SMITH, C. B.; AGAIAN, S.; AKOPIAN, D. A *wavelet* denoising approach using polynomial *threshold* operators. **IEEE Signal Processing Letters**, v. 15, p. 906-909, 2008.

SNIC – Sindicato Nacional da Indústria do Cimento. 2016. Disponível em: <http://www.snic.org.br/> . Acesso em: 25 out. 2016.

SOUZA, F. M. **Modelos Box-Jenkins aplicados à previsão de demanda de leitos hospitalares**. 2006. 83 p. Dissertação (Programa de Pós-Graduação em Estatística e Modelagem Quantitativa)–Universidade Federal de Santa Maria, Santa Maria, RS, 2006.

SOUZA, F. M.; ALMEIDA, S. G.; FELICIANI, A. V.; SOUZA, A. M. Uso de modelo de memória longa: Previsão de equipamentos para a agroindústria. **IJIE – Iberoamerican Journal of Industrial Engineering**, v.2, n.1, p. 137 – 153, jul. 2010.

SOUZA, F. M.; ALMEIDA, S. G.; GUARNIERI, J. P.; SOUZA, A. M.; LOPES, L. F. D. Previsão do consumo de cimento no estado do Rio Grande Do Sul. **Pesquisa Operacional para o Desenvolvimento**, [s.l.], v. 2, n. 1, fev. 2010.

SOUZA, G. P.; SAMOBYL, R. V.; MIRANDA, R. G. de. **Métodos Simplificados de Previsão Empresarial**. Rio de Janeiro: Ciência Moderna Ltda. 2008.

STEFFEN, D. **Intervalo “Bootstrap” para previsões de séries temporais obtidas pelo método theta**. 2010. 116 p. Dissertação (Programa de Pós-Graduação em Métodos Numéricos em Engenharia)-Universidade Federal do Paraná, Paraná, 2010.

STOLLNITZ, E. J.; DEROSE, T. D.; SALESIN, D. H. *Wavelets for Computer Graphics: A primer – part 1*. **IEEE Computer Graphics and Applications**, v. 15, p. 76-84, 1995.

VICINI, L.; SOUZA, A. M. Geração de subsídios para a tomada de decisão na cadeia produtiva da bovinocultura do Brasil. **Gestão da Produção, Operações e Sistemas**. v. 2, n. 4, p. 49-64, 2007.

VIDAKOVIC, B.; MUELLER, P. **Wavelets for kids**: a tutorial Introduction. Durham: Institute of Statistics and Decision Sciences Technical report, Duke University, 1991.

VORONIN, S.; PARTANEN, J. Forecasting electricity price and demand using a hybrid approach based on *Wavelet* Transform, ARIMA and Neural Networks. **International Journal Energy Research**, v. 38, n. 5, p. 626-637, 2014.

WERNER, L.; RIBEIRO, J. L. D. Previsão de demanda: uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais. **Revista Gestão & Produção**, v. 10, n. 1, p. 47-67.