



Universidade do Estado do Rio de Janeiro

Centro de Tecnologia e Ciência

Faculdade de Engenharia

Cesar Machado Pereira

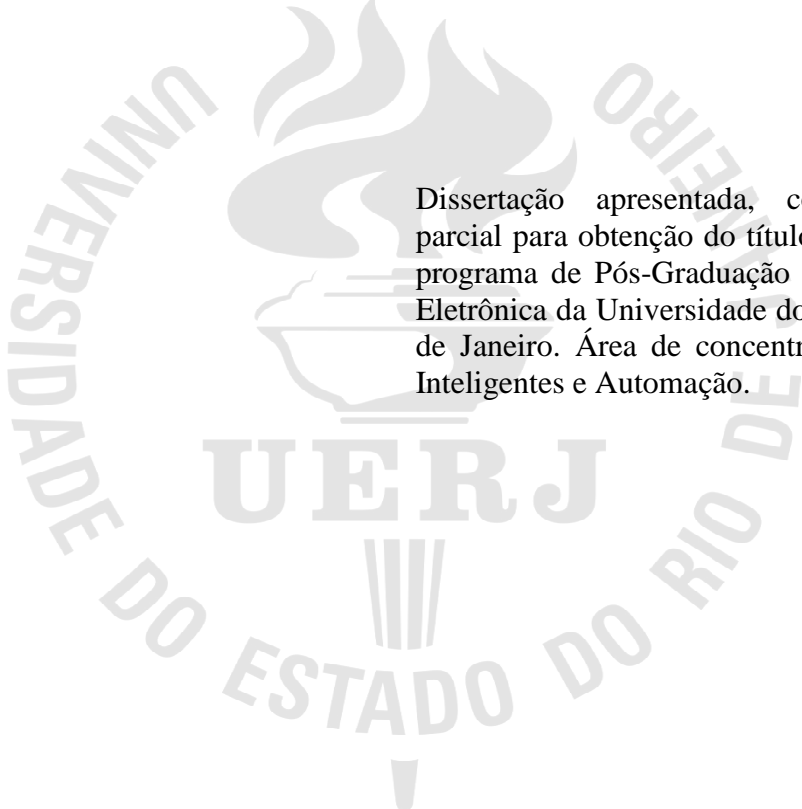
**Modelagem Fuzzy para Previsão de uma Série Temporal de Distribuição de
Energia Elétrica**

Rio de Janeiro

2015

Cesar Machado Pereira

Modelagem Fuzzy para Previsão de uma Série Temporal de Distribuição de Energia Elétrica



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao programa de Pós-Graduação em Engenharia Eletrônica da Universidade do Estado do Rio de Janeiro. Área de concentração: Sistemas Inteligentes e Automação.

Orientador: Prof. Dr. Nival Nunes de Almeida

Coorientadora: Prof^a. Dr^a. Maria Luiza Fernandes Velloso

Rio de Janeiro

2015

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CTC / B

P436 Pereira, Cesar Machado.
Modelagem fuzzy para previsão de uma série temporal de distribuição de energia elétrica / Cesar Machado Pereira. - 2015. 101 f.

Orientador: Nival Nunes de Almeida.
Coorientador: Maria Luiza Fernandes Velloso.
Dissertação (Mestrado) – Universidade do Estado do Rio de Janeiro, Faculdade de Engenharia.

1. Engenharia eletrônica. 2. Sistemas difusos – Dissertações. 3. Energia elétrica - Distribuição - Dissertações. I. Almeida, Nival Nunes. II. Universidade do Estado do Rio de Janeiro. III. Título.

CDU 681.5:621.3

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

Assinatura

Data

Cesar Machado Pereira

Modelagem Fuzzy para previsão de uma Série Temporal de Energia Elétrica

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao programa de Pós-Graduação em Engenharia Eletrônica da Universidade do Estado do Rio de Janeiro. Área de concentração: Sistemas Inteligentes e Automação.

Aprovado em: 24 de fevereiro de 2015.

Banca Examinadora:

Prof. Dr. Nival Nunes de Almeida. (Orientador)

Faculdade de Engenharia – UERJ

Prof^a. Dr^a. Maria Luiza Fernandes Velloso. (Coorientadora)

Faculdade de Engenharia – UERJ

Prof. PhD Reinaldo Castro Souza

Pontifícia Universidade Católica do Rio de Janeiro - PUC-RJ

Prof. Dr. Orlando Bernardo Filho

Faculdade de Engenharia – UERJ

Rio de Janeiro

2015

DEDICATÓRIA

Para todos que me mostraram capaz e me ajudaram a realizar mais este sonho.

AGRADECIMENTOS

Agradeço:

À minha família, Paulo Cesar de Oliveira Pereira, Jussara Trindade Machado Pereira, Juliana Machado Pereira e Paulo Machado Pereira, e namorada, Juliana Moreira Vieira, pelo carinho.

Aos meus orientadores: Nival Nunes e Maria Luiza por aceitarem me orientar nesta tese. Agradeço também ao estímulo, parceria e paciência.

Aos meus amigos, por todo apoio.

Aos colegas de trabalho, por terem feito minhas noites de escrituração da tese mais divertidas.

“You must fall in love with your data, but not always with your model”

G. M. Jenkins, (1979)

RESUMO

PEREIRA, Cesar Machado. *Modelagem Fuzzy para previsão de uma série temporal de energia elétrica*. 2015. 101f. Dissertação (Mestrado em Engenharia Eletrônica) – Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2015.

Esta dissertação testa e compara dois tipos de modelagem para previsão de uma mesma série temporal. Foi observada uma série temporal de distribuição de energia elétrica e, como estudo de caso, optou-se pela região metropolitana do Estado da Bahia. Foram testadas as combinações de três variáveis exógenas em cada modelo: a quantidade de clientes ligados na rede de distribuição de energia elétrica, a temperatura ambiente e a precipitação de chuvas. O modelo linear de previsão de séries temporais utilizado foi um SARIMAX. A modelagem de inteligência computacional utilizada para a previsão da série temporal foi um sistema de Inferência Fuzzy. Na busca de um melhor desempenho, foram feitos testes de quais variáveis exógenas melhor influenciam no comportamento da energia distribuída em cada modelo. Segundo a avaliação dos testes, o sistema Fuzzy de previsão foi o que obteve o menor erro. Porém dentre os menores erros, os resultados dos testes também indicaram diferentes variáveis exógenas para cada modelo de previsão.

Palavras-Chave: Previsão; Séries Temporais; Energia Elétrica; SARIMAX; Sistemas Fuzzy.

ABSTRACT

PEREIRA, Cesar Machado. *Fuzzy modeling to forecast a time series electric power*. 2015. 101f. Dissertação (Mestrado em Engenharia Eletrônica) – Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2015.

This dissertation tests and compares two types of predicting models to the same time series. A time series of electricity distribution was observed and, as a case study, were opted for the metropolitan region of Bahia State. Three exogenous variables were tested in each model: the number of customers connected to the electricity distribution network, the temperature and the precipitation of rain. The linear model time series forecasting used was a SARIMAX. The modelling of computational intelligence used to predict the time series was a Fuzzy Inference System. For better performance, in each model was tested all the exogenous variables to fit the influence in the energy distributed. According to the evaluation of the tests, the Fuzzy forecasting system presented the lowest error. But among the smallest errors, the results of the tests also indicated different exogenous variables for each forecast model.

Keywords: Forecast; Time Series; Electric Power; SARIMAX; Fuzzy Inference System.

LISTA DE ILUSTRAÇÕES

Figura 1 — Série temporal de energia distribuída original da Coelba.....	22
Figura 2 — Proporção de clientes por região.....	23
Figura 3 — Série de energia distribuída das 5 regiões da Coelba.....	23
Figura 4 — Série de energia distribuída da região Metropolitana.....	25
Figura 5 — Proposta da tese de dissertação.....	26
Figura 6 — Série não estacionária.....	32
Figura 7 — Série não estacionária após a primeira diferenciação.....	33
Figura 8 — Série estacionária após a segunda diferenciação.....	33
Figura 9 — Ciclo iterativo de Box-Jenkins.....	37
Figura 10 — Exemplo do comportamento do correlograma de séries não estacionárias.....	39
Figura 11 — Função de pertinencia sigmoidal.....	49
Figura 12 — Funções de pertinencia.....	49
Figura 13 — Modelo Geral de um Sistema de Inferência Fuzzy.....	52
Figura 14 — Série de energia distribuída da região Metropolitana sem finais de semana.....	55
Figura 15 — Precipitação de chuvas.....	57
Figura 16 — Temperatura ambiente.....	58
Figura 17 — Número de Clientes.....	59
Figura 18 — Mudanças dos coeficientes.....	60
Figura 19 — Metodologia aplicada no SARIMAX.....	61
Figura 20 — Distribuição amostral da série.....	62
Figura 21 — Teste de estacionariedade da série temporal da região Metropolitana.....	63
Figura 22 — FAC da região Metropolitana.....	64
Figura 23 — Transformação da série temporal da região Metropolitana.....	64
Figura 24 — Distribuição amostral da série transformada.....	65
Figura 25 — Teste de homogeneidade da transformada da região Metropolitana.....	66
Figura 26 — Autocorrelograma da série da região Metropolitana.....	67
Figura 27 — Autocorrelograma parcial sérieda região Metropolitana.....	67
Figura 28 — Teste de estacionariedade dos resíduos do teste 12 (1).....	69
Figura 29 — Teste de estacionariedade dos resíduos do teste 13 (1).....	70
Figura 30 — Teste de estacionariedade dos resíduos do teste 23 (1).....	71
Figura 31 — Teste de estacionariedade dos resíduos do teste 123 (1).....	72

Figura 32 — Teste SARIMAX 12 (1).....	78
Figura 33 — Resíduos do teste SARIMAX 12 (1).....	79
Figura 34 — Teste SARIMAX 13 (1).....	80
Figura 35 — Resíduos do teste SARIMAX 13 (1).....	80
Figura 36 — Teste SARIMAX 23 (1).....	81
Figura 37 — Resíduos do teste SARIMAX 23 (1).....	82
Figura 38 — Teste SARIMAX 123 (1).....	83
Figura 39 — Resíduos do teste SARIMAX 123 (1).....	83
Figura 40 — Teste 12 (1) do Sistema Fuzzy.....	85
Figura 41 — Resíduos do teste do Sistema Fuzzy 12 (1).....	85
Figura 42 — Teste 13 (1) do Sistema Fuzzy.....	86
Figura 43 — Resíduos do teste do Sistema Fuzzy 13 (1).....	87
Figura 44 — Teste 23 (1) do Sistema Fuzzy.....	88
Figura 45 — Resíduos do teste do Sistema Fuzzy 23 (1).....	88
Figura 46 — Teste 123 (1) do Sistema Fuzzy.....	89
Figura 47 — Resíduos do teste do Sistema Fuzzy 123 (1).....	90

LISTA DE TABELAS

Tabela 1— Avaliações dos modelos de regressão.....	60
Tabela 2— Ordem do modelo SARIMAX.....	68
Tabela 3— Parâmetros do teste 12 (1) do modelo SARIMAX.....	68
Tabela 4— Parâmetros do teste 13 (1) do modelo SARIMAX.....	69
Tabela 5— Parâmetros do teste 23 (1) do modelo SARIMAX.....	70
Tabela 6— Parâmetros do teste 123 (1) do modelo SARIMAX.....	71
Tabela 7— Teste de sobrefixação do teste 12 (1)..	72
Tabela 8— Teste de sobrefixação do teste 13 (1)..	72
Tabela 9— Teste de sobrefixação do teste 23 (1)..	73
Tabela 10— Teste de sobrefixação do teste 123 (1)..	73
Tabela 11— Teste da função de pertinência do teste 12 (1).	74
Tabela 12— Teste da função de pertinência do teste 13 (1).	75
Tabela 13— Teste da função de pertinência do teste 23 (1).	75
Tabela 14— Teste da função de pertinência do teste 123 (1).	75
Tabela 15— Parâmetros do Sistema de Inferência Fuzzy.....	75
Tabela 16— Testes de variáveis do modelo SARIMAX.	77
Tabela 17— Parâmetros das variáveis do teste SARIMAX 12 (1).	79
Tabela 18— Erros percentuais do teste SARIMAX 12 (1).	79
Tabela 19— Parâmetros das variáveis do teste SARIMAX 13 (1).	81
Tabela 20— Erros percentuais do teste SARIMAX 13 (1).	81
Tabela 21— Parâmetros das variáveis do teste SARIMAX 23 (1).	82
Tabela 22— Erros percentuais do teste SARIMAX 23 (1).	82
Tabela 23— Parâmetros das variáveis do teste SARIMAX 123 (1).	84
Tabela 24— Erros percentuais do teste SARIMAX 123 (1).	84
Tabela 25— Testes de variáveis do Sistema de Inferência Fuzzy.	84
Tabela 26— Parâmetros das variáveis do teste 12 (1) do Sistema Fuzzy.....	86
Tabela 27— Erros percentuais do teste 12 (1) do Sistema Fuzzy.....	86
Tabela 28— Parâmetros das variáveis do teste 13 (1) do Sistema Fuzzy.....	87
Tabela 29— Erros percentuais do teste 13 (1) do Sistema Fuzzy.....	87
Tabela 30— Parâmetros das variáveis do teste 23 (1) do Sistema Fuzzy.....	89
Tabela 31— Erros percentuais do teste 23 (1) do Sistema Fuzzy.....	89

Tabela 32— Parâmetros das variáveis do teste 123 (1) do Sistema Fuzzy	90
Tabela 33— Erros percentuais do teste 123 (1) do Sistema Fuzzy	90
Tabela 34— MAPE dos testes de variáveis do Sistema Fuzzy e SARIMAX.....	91
Tabela 35— DP dos testes de variáveis do Sistema Fuzzy e do SARIMAX.....	91

LISTA DE ABREVIACOES

ABCE	<i>Associao Brasileira de Concessionrias de Energia Eltrica</i>
ABRADE	<i>Associao Brasileira de Distribuidores de Energia Eltrica</i>
ABRAGE	<i>Associao Brasileira das Empresas Geradoras de Energia Eltrica</i>
ABRATE	<i>Associao Brasileira das Grandes Empresas de Transmisso de Energia Eltrica</i>
ACL	<i>Ambiente de Contratao Livre</i>
ACR	<i>Ambiente de Contratao Regulada</i>
AIC	<i>Critrios Akaike</i>
AICc	<i>Critrios Akaike corrigido</i>
ANEEL	<i>Agncia Nacional de Energia Eltrica</i>
ANFIS	<i>Sistema de Inferncia Neuro-Fuzzy Adaptativo</i>
APINE	<i>Associao Brasileira dos Produtores Independentes de Energia Eltrica</i>
AR	<i>Auto-regressive</i>
ARMA	<i>Auto-regressive Moving Average</i>
ARIMA	<i>Auto-regressive Integrated Moving Average</i>
AST	<i>Anlise de Sries Temporais</i>
CCEE	<i>Cmara de Comercializao de Energia Eltrica</i>
CNPE	<i>Conselho Nacional de Poltica Energtica</i>
Coelba	<i>Companhia de Eletricidade do Estado da Bahia</i>
DP	<i>Desvio Padro</i>
FAC	<i>Funo de Autocorrelao</i>

FACP	<i>Função de Autocorrelação Parcial</i>
FBP	<i>Fuzzy Bayes Predictor</i>
FMP	<i>Fuzzy Markov Predictor</i>
FIS	<i>Fuzzy Inference System</i>
GWh	<i>Gigawatts-hora</i>
HMM	<i>Hidden Markov Models</i>
INMET	<i>Instituto Nacional de Meteorologia</i>
MA	<i>Moving Average</i>
MAPE	<i>Mean absolute percentage error</i>
MME	<i>Ministério de Minas e Energia</i>
MWh	<i>Megawatts-hora</i>
OLS	<i>Ordinary least squares</i>
ONS	<i>Operador Nacional do Sistema Elétrico</i>
PLD	<i>Preço de Liquidação das Diferenças</i>
PIB	<i>Produto Interno Bruto</i>
Procel	<i>Programa Nacional de Conservação de Energia Elétrica</i>
Proinfa	<i>Programa de Incentivo às Fontes Alternativas de Energia Elétrica</i>
SARIMA	<i>Seasonal Auto-regressive Integrated Moving Average</i>
SARIMAX	<i>Seasonal Auto-regressive Integrated Moving Average with exogenous regressors</i>
SFBR	<i>Sistema Fuzzy Baseado em Regras</i>

SUMÁRIO

INTRODUÇÃO	17
1 O SISTEMA DE DISTRIBUIÇÃO DE ENERGIA	19
1.1 O sistema de distribuição da Bahia.....	21
1.2 A série temporal objeto de estudo	24
1.3 Objetivo	25
1.4 Trabalhos Relacionados	26
1.5 Estruturação da dissertação	28
2 DESCRIÇÃO DAS TÉCNICAS DE MODELAGEM	30
2.1 Séries Temporais	30
2.2 Estacionariedade das Séries Temporais	31
2.3 Modelo Autoregressivo (AR)	34
2.4 Modelo de Médias Móveis (MA)	34
2.5 Modelo Autoregressivo e de Média Móvel Multiplicativo (ARMA).....	35
2.6 Modelo Autoregressivo Integrado de Média Móvel (ARIMA)	36
2.7 Identificação dos modelos ARIMA (p, d e q)	38
2.7.1 <u>Função de Autocorrelação (FAC)</u>	38
2.7.2 <u>Função de Autocorrelação Parcial (FACP)</u>	39
2.7.3 <u>Identificação dos modelos</u>	40
2.7.4 <u>Estimativas dos parâmetros do modelo</u>	41
2.7.5 <u>Crítérios Akaike (AIC)</u>	41
2.8 Modelo Multiplicativo Autoregressivo Integrado de Média Móvel (SARIMA)	42
2.9 Modelo Multiplicativo Autoregressivo Integrado de Média Móvel com Variáveis Exógenas (SARIMAX)	43
2.10 Modelos de Regressão para as Variáveis Explicativas no SARIMAX	44
2.10.1 <u>Modelo de Regressão Linear (OLS)</u>	45
2.10.2 <u>Modelo Cochrane-Orcutt (CO-LS)</u>	45
3 SISTEMAS DE INFERÊNCIA FUZZY	47
3.1 Lógica Fuzzy	47
3.1.1 <u>Conjuntos Fuzzy</u>	47
3.1.2 <u>Funções de pertinência</u>	48
3.1.3 <u>Variáveis Linguísticas</u>	50

3.1.4	<u>Operações Lógicas</u>	51
3.2	O sistema Fuzzy	51
4	EXPERIMENTOS E MODELOS OBTIDOS	54
4.1	Descrição dos Dados	54
4.1.1	<u>Série Temporal de Energia Distribuída</u>	54
4.1.2	<u>Número de Clientes</u>	56
4.1.3	<u>Precipitação de Chuvas</u>	56
4.1.4	<u>Temperatura Ambiente</u>	57
4.2	Análise exploratória dos dados	58
4.3	Modelo de previsão SARIMAX	60
4.3.1	<u>Teste de Estacionariedade</u>	61
4.3.2	<u>Transformação da Série Temporal</u>	63
4.3.3	<u>Teste de Estacionariedade da Série Transformada</u>	65
4.3.4	<u>Análise Descritiva</u>	66
4.3.5	<u>Testes dos Parâmetros do Modelo SARIMAX</u>	68
4.4	Modelo de previsão do Sistema de Inferência Fuzzy	73
4.5	Avaliação	76
4.5.1	<u>Erro Médio Percentual</u>	76
4.5.2	<u>Desvio Padrão</u>	76
5	RESULTADOS OBTIDOS POR MEIO DOS MODELOS	77
5.1	Resultados do modelo SARIMAX	77
5.1.1	<u>Resultados do modelo SARIMAX no teste 12 (1)</u>	78
5.1.2	<u>Resultados do modelo SARIMAX no teste 13 (1)</u>	79
5.1.3	<u>Resultados do modelo SARIMAX no teste 23 (1)</u>	81
5.1.4	<u>Resultados do modelo SARIMAX no teste 123 (1)</u>	82
5.2	Resultados do Sistema de Inferência Fuzzy	84
5.2.1	<u>Resultados do Sistema de Inferência Fuzzy no teste 12 (1)</u>	85
5.2.2	<u>Resultados do Sistema de Inferência Fuzzy no teste 13 (1)</u>	86
5.2.3	<u>Resultados do Sistema de Inferência Fuzzy no teste 23 (1)</u>	87
5.2.4	<u>Resultados do Sistema de Inferência Fuzzy no teste 123 (1)</u>	89
5.3	Comparativo entre os modelos	91

CONCLUSÃO.....	92
REFERÊNCIAS	95
ANEXO A.....	100

INTRODUÇÃO

Segundo o noticiário do Canal Energia¹, o Brasil possui 132,6 mil megawatts de capacidade instalada de geração de eletricidade, e segundo indicadores do Instituto Brasileiro de Geografia e Estatística² (IBGE), uma população de 203,6 milhões de habitantes, que são os consumidores atuais e os potenciais das empresas que distribuem energia.

De acordo com o Operador Nacional do Sistema Elétrico³ (ONS) e a Câmara de Comercialização de Energia Elétrica⁴ (CCEE), o setor elétrico brasileiro está estruturado para garantir a segurança do suprimento de energia elétrica, promover a inserção social, por meio de programas de universalização do atendimento, e também a modicidade tarifária e de preços.

Em linhas gerais, o sistema de elétrico brasileiro é controlado e coordenado pela ONS, fiscalizado e regulado pela Agência Nacional de Energia Elétrica (ANEEL). Além disso, a CCEE é a instituição responsável pela contabilização e por realizar a liquidação financeira das diferenças entre os montantes contratados e os montantes efetivamente consumidos.

Entretanto há inúmeros outros órgãos envolvidos com o setor elétrico: Conselho Nacional de Política Energética (CNPE) e o Ministério de Minas e Energia (MME). E as associações: Associação Brasileira de Concessionárias de Energia Elétrica (ABCE), Associação Brasileira de Distribuidores de Energia Elétrica (ABRADEE), Associação Brasileira das Empresas Geradoras de Energia Elétrica (ABRAGE), Associação Brasileira das Grandes Empresas de Transmissão de Energia Elétrica (ABRATE), Associação Brasileira dos Produtores Independentes de Energia Elétrica (APINE).

Em uma visão geral da ABRADEE, a indústria de energia elétrica é basicamente composta por geradores espalhados pelo país e pelas linhas de transmissão e de distribuição de energia, que compõem a chamada “indústria de rede”. Todo o sistema é eletricamente conectado, exigindo o balanço constante e instantâneo entre tudo o que é produzido e consumido.

Entre 2003 e 2004 o governo federal lançou as bases do modelo atual do Setor Elétrico Brasileiro, que é sustentado pelas Leis nº 10.847 e 10.848, de 15 de março de 2004, e pelo Decreto nº 5.163, de 30 de julho de 2004. Assim, a comercialização de energia elétrica passou

¹Disponível em: <<http://www.canalenergia.com.br/zpublisher/materias/Noticiario.asp?id=104620>>. Acesso em: 06 jan. 2015.

²Disponível em: <<http://www.ibge.gov.br/>>. Acesso em: 06 jan. 2015.

³Disponível em: <www.ons.org.br/>. Acesso em: 06 jan. 2015.

⁴Disponível em: <www.ccee.org.br/>. Acesso em: 06 jan. 2015.

a contar com dois ambientes: o Ambiente de Contratação Regulada (ACR), com agentes de geração e de distribuição de energia; e o Ambiente de Contratação Livre (ACL), com geradores, distribuidores, comercializadores, importadores e exportadores, além dos consumidores livres e especiais.

Sob o novo modelo, o governo federal definiu um conjunto de agentes institucionais com competências e atribuições bem definidas, objetivando garantir o bom funcionamento setorial e alcançar a modicidade tarifária, a segurança no suprimento e universalização do acesso (TOLMASQUIM, 2011 apud SOUZA et al., 2014).

A relação entre as duas pontas da indústria de rede, ou seja, entre as distribuidoras e as geradoras se deve através dos leilões de energia elétrica realizados pela CCEE. Os leilões de compra de energia elétrica realizados pela CCEE, por delegação da ANEEL, ocupam papel essencial no Ambiente de Contratação Regulada (ACR). Os compradores e vendedores de energia participantes dos leilões formalizam suas relações comerciais por meio de contratos registrados no âmbito do ACR. Os contratos do ACR têm regulação específica para aspectos como o preço da energia, submercado de registro do contrato e vigência de suprimento.

Como ainda há diferenças entre a energia produzida e a consumida pelo mercado há um mercado de curto prazo operado pela CCEE que equaciona este problema (SOUZA et al., 2014).

No Ambiente de Contratação Livre (ACL) o acordo se deve a ambas as partes. Os produtores independentes, comercializadores e os consumidores livres têm liberdade para negociar a compra de energia. Nele se estabelece o volume, o preço e o prazo de suprimento. As operações do ACL são feitas por Contratos de Compra de Energia.

Todos os contratos de compra e venda de energia realizados no mercado ACR e no ACL devem ser registrados na CCEE. Neste sentido, a CCEE realiza a medição dos montantes efetivamente produzidos e consumidos por cada agente. As diferenças apuradas, positivas ou negativas, são contabilizadas para posterior liquidação financeira no Mercado de Curto Prazo e valoradas ao Preço de Liquidação das Diferenças (PLD).

Ademais, o Mercado de Curto Prazo é definido como o segmento da CCEE onde são contabilizadas as diferenças entre os montantes de energia elétrica contratados de geração e de consumo. Assim, no Mercado de Curto Prazo não existem contratos, ocorrendo a contratação multilateral, conforme as Regras de Comercialização.

O setor elétrico em todo o mundo vem sofrendo um processo de transformação quer no campo político, quer no econômico. A privatização e a desregulamentação do sistema

elétrico foram os grandes responsáveis pelas mudanças de cenário deste sistema (SERRÃO, 2003).

Seguindo o tema do setor elétrico, o capítulo a seguir se destinará a uma parte do setor que energia que sofreu com as mudanças e é peça fundamental para em relacionamento com o consumidor final.

1 O SISTEMA DE DISTRIBUIÇÃO DE ENERGIA

O segmento de distribuição recebe grande quantidade de energia do sistema de transmissão e a distribui para os consumidores finais, residências e pequenos comércios e indústrias. Segundo a ABRADDEE⁵, no Brasil, esse segmento é composto por 63 concessionárias, as quais são responsáveis pela administração e operação de linhas de transmissão de menor tensão (abaixo de 230 mil Volts), como exemplo, as instaladas nas ruas e avenidas das cidades.

De acordo com a ANEEL⁶ o segmento de distribuição de energia se caracteriza como a parte do setor elétrico dedicado à entrega de energia elétrica para um usuário final. O sistema de distribuição pode ser considerado como o conjunto de instalações e equipamentos elétricos que operam, geralmente, em tensões inferiores a 230 kV, incluindo os sistemas de baixa tensão.

Segundo a ABRADDEE⁷: a conexão, o atendimento e a entrega efetiva de energia elétrica ao consumidor do ambiente regulado ocorrem por parte das distribuidoras de energia. A energia distribuída, portanto, é a energia efetivamente entregue aos consumidores conectados à rede elétrica de uma determinada empresa de distribuição.

As distribuidoras instalam no final da cadeia de energética um pequeno aparelho que consegue medir a quantidade de energia por eles utilizada por cada local de consumo. A medição é feita por hora e chamamos de horário de pico o momento em que uma localidade utiliza maior quantidade de energia elétrica (FERREIRA, 2012).

Os valores máximos de carga são geralmente os mais cruciais para as empresas do setor elétrico conhecer antecipadamente, considerando que podem ocasionar quedas no sistema (ALMEIDA, 2008).

⁵Disponível em: <<http://www.abradee.com.br/setor-eletrico/visao-geral-do-setor>>. Acesso em: 06 jan. 2015.

⁶Disponível em: <www.aneel.gov.br>. Acesso em: 06 jan. 2015.

⁷Disponível em: <<http://www.abradee.com.br/setor-de-distribuicao/a-distribuicao-de-energia>>. Acesso em: 06 jan. 2015.

De acordo com a Companhia Paranaense de Energia⁸ (COPEL), o horário de pico é o período definido e composto por três horas diárias consecutivas, durante o qual o consumo de energia elétrica tende a ser maior. Até o ano de 2013 no Brasil, nos dias de semana o horário é das 18 horas às 21 horas, em horário de verão das 19 horas às 22 horas. O pico de consumo de energia, principalmente quando as pessoas chegam do trabalho acendem as luzes, ligam os condicionadores de ar e a televisão e tomando banho com a água aquecida por chuveiros elétricos.

O fornecimento de energia é um dos serviços mais desafiantes da sociedade moderna. Para que o consumidor disponha de energia no momento que aciona um interruptor ou conecta um aparelho elétrico na tomada em sua residência é preciso de todo o sistema em cadeia esteja apto a operar de forma coordenada, com centrais geradoras, linhas de transmissão, subestações, linhas e transformadores de distribuição.

Dentre os diversos programas do governo que influenciam a distribuição de energia elétrica há o programa “Luz para Todos”⁹ (LPT), que completou 10 anos em 2014, que serve como incentivo para investimento no setor elétrico. O programa luz para todos é uma política de responsabilidade do Estado e tem os objetivos de universalizar o serviço de eletricidade e promover a inclusão social.

Como não se armazena energia elétrica, é preciso sincronizar a produção de energia com o consumo. Portanto, a operação de geração de energia elétrica no início da cadeia precisa ajustar-se continuamente às oscilações no consumo de energia.

Segundo o Instituto Acende Brasil, sistemas elétricos devem ser concebidos e operados de forma a prover a confiabilidade desejada pela população. Isso requer que o sistema seja dimensionado adequadamente para atender à carga prevista. Além do zelo pela continuidade do serviço, há também a questão da conformidade da corrente elétrica. Oscilações no fornecimento de energia elétrica podem prejudicar a operação de aparelhos elétricos ou mesmo danificá-los.

A energia elétrica é essencial no dia-a-dia da sociedade, seja nas residências ou nos diversos segmentos da economia. Atividades rotineiras como assistir televisão ou o uso do ar condicionado só são possíveis através da distribuição da energia elétrica. Fábricas, supermercados, shoppings e outros lugares precisam de energia elétrica para funcionar.

Portanto, a estimação da demanda futura de energia elétrica é de extrema importância para o planejamento do sistema de geração de eletricidade, pois pode indicar a necessidade de

⁸Disponível em: <<http://www.copel.com>>. Acesso em: 06 jan. 2015.

⁹O programa LPT foi antecedido pelo programa “Luz no Campo”, que perdurou entre por 3 anos.

expansão da capacidade de oferta (OLIVEIRA; MORITA, 2005). De acordo com a previsão de energia elétrica distribuída também se pode desenvolver estratégias para maximizar os lucros e minimizar riscos (ZHOU et al., 2006). A previsão de energia elétrica também se torna importante, principalmente para o setor em questão, devido aos elevados custos de erros de previsão (CAMPOS; JESUS; MENDES, 2007).

Uma rede de energia elétrica confiável, segura e estável é uma necessidade básica para o desenvolvimento de um país. Portanto a rede elétrica brasileira é um tema de bastante interesse (ARRUDA, 2013).

Estudo divulgado em janeiro de 2014 na nona edição do Programa Energia Transparente pelo Instituto Acende Brasil¹⁰ mostra descompasso entre a vida real e as projeções no sistema elétrico nacional. Na avaliação da entidade, a raiz de todo esse cenário está na modelagem que é utilizada tanto para o planejamento quanto a operação.

Como uma previsão adequada pode dar suporte a uma decisão minimizadora de risco por parte dos tomadores de decisão, ela é essencial para o planejamento organizacional de qualquer setor da economia.

Atualmente, a redução dos preços das tarifas energéticas vem agravando uma crise no setor de energia elétrica. Pois a intervenção proporciona um problema financeiro às empresas distribuidoras de energia, devido à falta de oferta de energia barata, ao alto valor da energia no mercado a vista (*spot*), ocasionado pela falta de chuvas.

A diferença de preços que a distribuidora contrata a energia e a distribui se torna uma problemática nesse sistema, portanto, é bastante relevante realizar formas mais sofisticadas de prever a energia elétrica distribuída.

Além da diferença de preço entre os mercados, há o problema da contratação de energia. Pois antes de se saber o mercado a empresa distribuidora de energia faz um contrato de compra com as geradoras. Assim a distribuidora tem que comprar energia antes mesmo de saber a quantidade total de energia que consumirão.

1.1 O sistema de distribuição da Bahia

Neste contexto, o Estado da Bahia tem como empresa responsável a Companhia de Eletricidade do Estado da Bahia (Coelba) para a distribuição de energia elétrica. Segundo

¹⁰Disponível em: <www.acendebrasil.com.br>. Acesso em: 06 jan. 2015.

dados informados no *site* da própria empresa¹¹, a Coelba é a terceira maior distribuidora de energia elétrica do país em número de clientes e a sexta em volume de energia fornecida.

A criação da Coelba foi em 1960 e atualmente ela está presente em 415 dos 417 municípios da Bahia, atendendo a mais de 14 milhões de habitantes em uma área de concessão de 563 mil quilômetros quadrados. Segundo informações fornecidas pela própria empresa, vide documento em anexo, em 2010 ela possuía 4.900.783 clientes, sendo 88% destes, clientes residenciais, com um total de 15.329 gigawatts-hora¹² (GWh) de energia distribuída. Atualmente a empresa conta com 2,5 mil colaboradores, 12 mil parceiros, 43 agências de atendimento e 8 agências móveis além da Rede Coelba Serviços.

A energia elétrica distribuída¹³ da Coelba pode ser ilustrada na série temporal a seguir em escala de megawatts-hora (MWh). Os dados foram fornecidos pela própria empresa em estudo, vide documento em anexo. A série temporal apresentada tem periodicidade diária ao longo de janeiro de 2010 até outubro de 2013, compondo assim 1.377 pontos de amostra. No eixo das abscissas está os dias e no eixo das ordenadas o MWh.

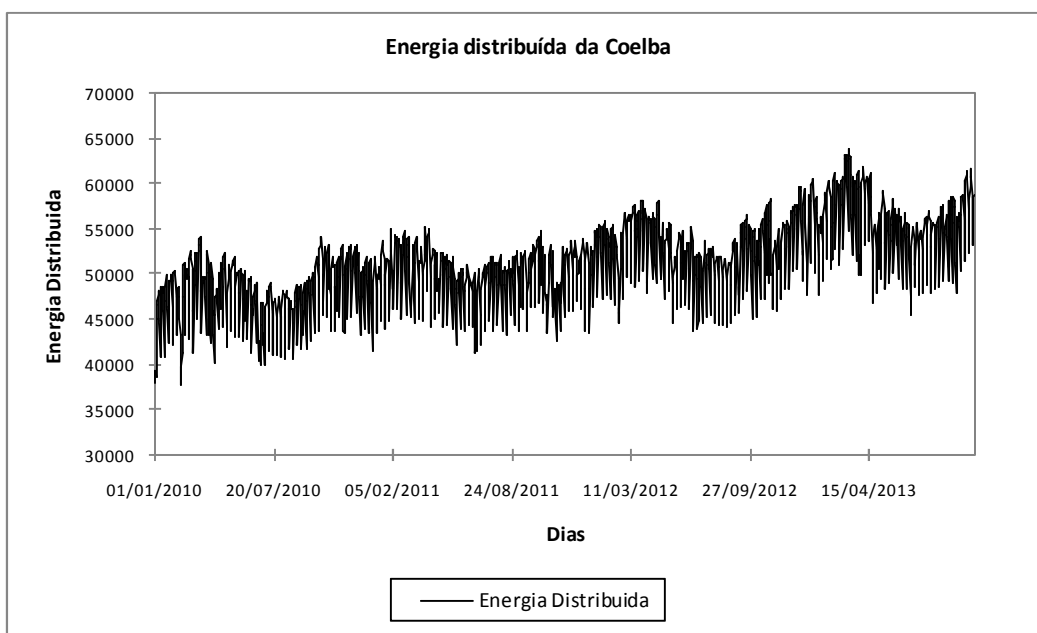


Figura 1 — Série temporal de energia distribuída original da Coelba.

A série temporal de energia distribuída em MWh apresentada acima tem um padrão ao longo do período bem definido e uma tendência de alta. Podem ser observáveis ciclos sazonais.

¹¹Disponível em: <<http://www.coelba.com.br/>>. Acesso em: 06 jan. 2015.

¹²Para cada forma de energia foi estabelecida uma medida. No caso da energia elétrica estabeleceu-se o Watt-hora (Wh).

¹³A energia distribuída é usualmente chamada de carga. No caso em questão é a carga diária.

Como uma classificação interna da própria Coelba, essa série temporal pode ser segmentada em cinco regiões do Estado da Bahia: Metropolitana, Norte, Centro, Sul-Sudoeste e Oeste. A proporção da quantidade de clientes em cada uma dessas cinco regiões é ilustrada na figura abaixo.

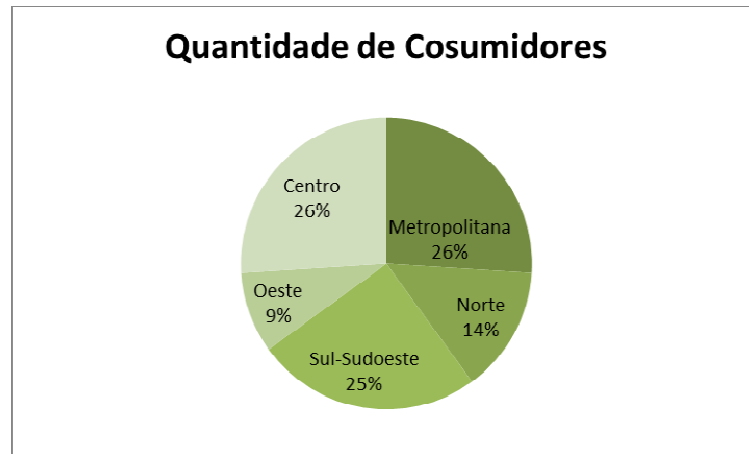


Figura 2 — Proporção de clientes por região.

A figura 3 apresentada a seguir ilustra as cinco regiões destacadas acima em séries temporais de distribuição de energia elétrica. No eixo das abscissas está os dias e no eixo das ordenadas o MWh com os 1.377 pontos de amostra. A região Sul-Sudoeste está representada em azul claro, a Oeste em azul escuro, a Norte em verde, a Centro em roxo e Metropolitana em vermelho.

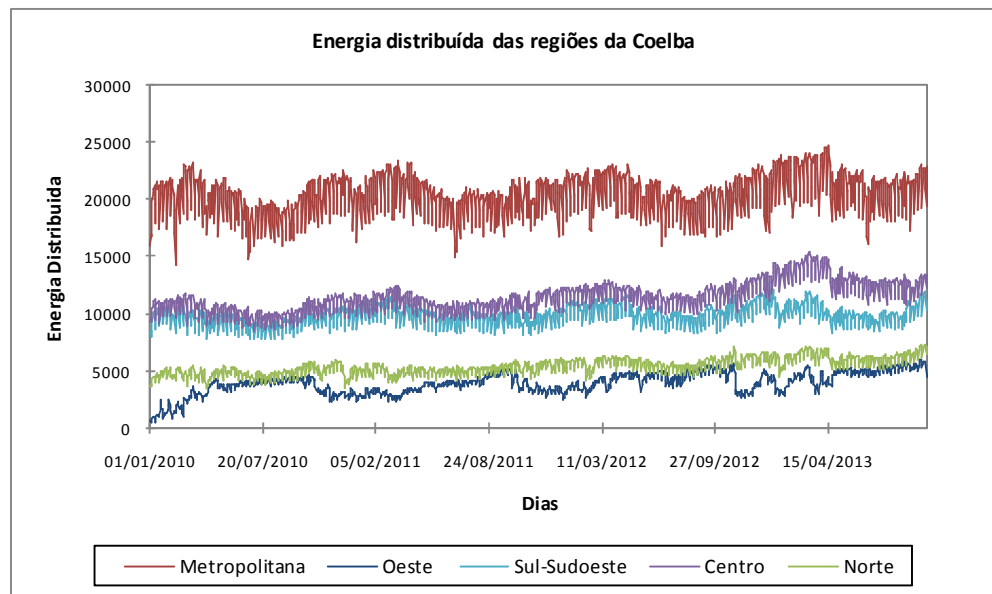


Figura 3 — Série de energia distribuída das 5 regiões da Coelba.

No exemplo da figura acima é observável uma diferença significativa entre as 5 séries apresentadas. Pois cada região representada possui um tipo de consumidor diferente e um padrão de consumo específico.

1.2 A série temporal objeto de estudo

A série temporal objeto de estudo desta dissertação é a da região Metropolitana da Coelba. Esta escolha tem por fundamento devido as variáveis exógenas escolhidas como entrada do sistema. Como a região Metropolitana é uma região urbana e mais concentrada geograficamente, os dados medidos em uma determinada localização são mais representativos para toda a região metropolitana. Por exemplo, há níveis de temperatura diferentes entre dois pontos distintos do Estado da Bahia. Mas com a delimitação da região metropolitana do Estado da Bahia, há um número mais característico, que melhor represente o nível de precipitação e de temperatura em toda essa região.

Na região Metropolitana, a quantidade de energia distribuída nesta região é bastante dependente do uso de eletrodomésticos, principalmente o ar condicionado. Exemplificando o padrão de consumo e das variáveis exógenas entre as regiões apresentadas neste capítulo 1 é o consumo das geladeiras velhas na população de baixa renda da região Metropolitana. Assim foi criado o programa Nova Geladeira¹⁴, pois as geladeiras velhas representavam 70% do consumo de energia nessas comunidades. Segundo informação do próprio *site* da empresa no programa Nova Geladeira, o gasto por residência chegou a cair 40% com a troca para novas geladeiras com o selo de eficiência energética. Esta iniciativa foi incentivada pela própria Coelba.

Os dados apresentados a seguir foram fornecidos pela própria empresa em estudo. São séries temporais que têm periodicidade diária ao longo de janeiro de 2010 até outubro de 2013, compondo assim 1.377 pontos de amostra. A figura abaixo ilustra a energia distribuída em megawatts-hora (MWh) da região Metropolitana.

¹⁴Disponível em: <<http://www.coelba.com.br/>>. Acesso em: 06 jan. 2015.

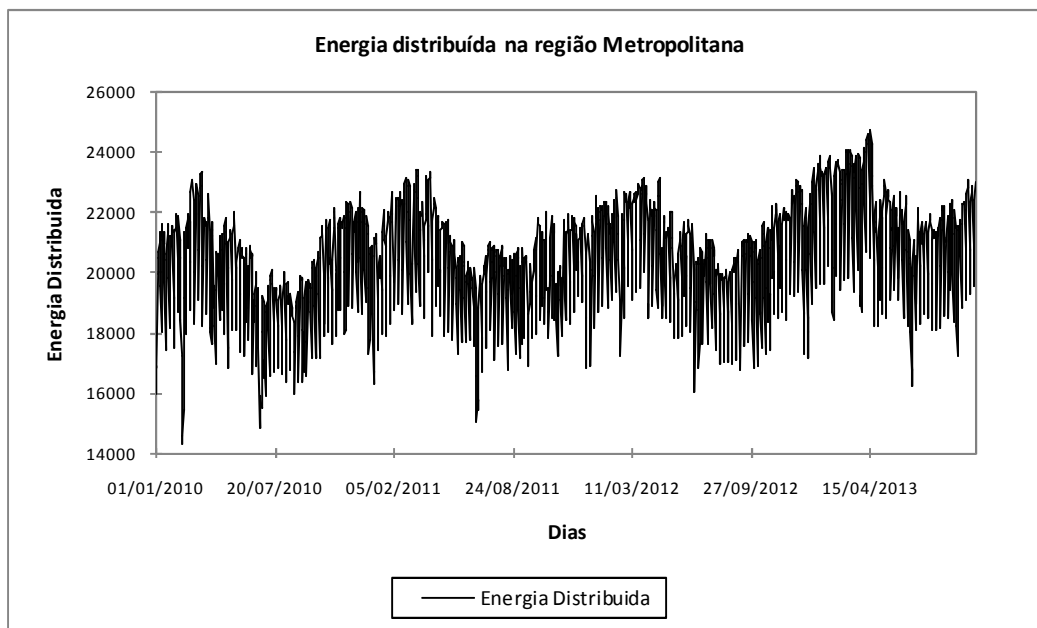


Figura 4 — Série de energia distribuída da região Metropolitana.

1.3 Objetivo

Esta dissertação tem por objetivo comparar modelagens¹⁵ para previsão de séries temporais. As modelagens comparadas são: uma linear SARIMAX e uma de Sistema de Inferência Fuzzy. Ambos os modelos são aplicados para a previsão de uma série temporal de distribuição de energia elétrica da região Metropolitana do Estado da Bahia.

Para fins de comparação, se propõe assim realizar testes em ambos os modelos com o uso de diferentes combinações de três variáveis exógenas¹⁶. As variáveis exógenas indicadas para ambos os modelos foram: a temperatura ambiente, a precipitação de chuvas e o número de clientes.

Com o uso das variáveis exógenas, os modelos buscam a previsão de um passo a frente, no dia anterior.

A figura e os tópicos abaixo ilustram e destacam a proposta desta tese:

- Modelagem de um algoritmo SARIMAX.
- Modelagem de um Sistema de Inferência Fuzzy.

¹⁵ Um modelo matemático é uma construção abstrata de um fenômeno real, que deveria comportar-se em concordância com a situação real em todos os seus aspectos relevantes (SOUZA; CAMARGO, 1996).

¹⁶ Variáveis exógenas são dados que se aceitam para estudar o desempenho da variável explicada.

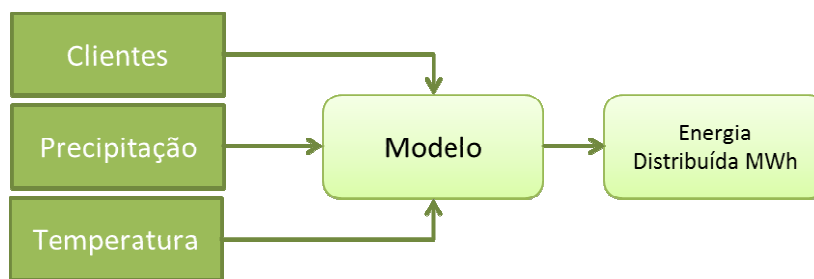


Figura 5 — Proposta da tese de dissertação.

Em um primeiro momento, foi apresentado o mercado de energia elétrica no Brasil e na Bahia.

Agora, esta tese se propõe a apresentar algumas abordagens existentes para a aplicação da previsão de energia elétrica, descrevendo os modelos utilizados com relativa frequência na literatura científica referente ao assunto e apresentando os conceitos e equações envolvidas dos modelos utilizados na presente tese.

Pretende-se também através da presente dissertação realizar uma pesquisa buscando verificar os aspectos ligados ao consumo de energia elétrica de um setor residencial, apresentando algumas variáveis exógenas para o mercado de energia elétrica no Brasil e no Estado da Bahia.

Em fim, são comparados os modelos de previsão de série temporal SARIMAX e o Sistema de Inferência Fuzzy. Com a comparação entre os dois modelos é possível destacar e apresentar considerações sobre os resultados obtidos nos testes.

1.4 Trabalhos Relacionados

Diversos modelos são aplicados para a previsão de energia elétrica. Os modelos matemáticos ARMA, ARIMA e SARIMAX e os sistemas de inferência Fuzzy, do inglês *Fuzzy Inference System* (FIS) foram encontrados nos trabalhos acadêmicos citados ao longo desta seção.

O modelo estatístico mais utilizado é o modelo autoregressivo integrado de média móvel (ARIMA). Na área de energia esse modelo está sendo utilizado por desde 1978. O autor Uri (1978) o utilizou para prever a ocorrência mensal de picos de energia mediante o comportamento de alterações nas variáveis econômicas e climáticas.

Em 1999 Chavez, Bernat e Coalla (1999) utilizaram o modelo de análise de séries temporais Box-Jenkins para previsão da produção e do consumo de energia elétrica. Os países utilizados foram: Áustria e Espanha.

Os autores Saab, Badr e Nasr (2001) investigaram diferentes metodologias para se realizar a previsão de energia elétrica em Lebanon. Um de seus modelos foram o ARIMA. Ediger e Tatildil (2002) utilizaram a técnica semiestatística para formular o modelo para prever a demanda de energia primária na Turquia e análise de padrões cíclicos.

Na previsão do preço da geração e do consumo de energia elétrica na Espanha os autores Cuaresma et al. (2004) e Conejo et al. (2005) também utilizaram o modelo ARIMA.

Os estudos de Serrão (2003) e Pandian et al. (2006) usaram técnicas de modelagem matemática para prever o comportamento de consumo de energia através de condições meteorológicas, jornadas de trabalho, dias da semana e a hora do dia.

O Sistema de Inferência Fuzzy (FIS em inglês) vem sendo utilizada em modelos de previsão de séries temporais do setor elétrico no curto prazo, séries de meia hora, horárias e diárias. Os trabalhos de Gooijer, Jan e Hyndman (2006) e Pandian et al. (2006) são alguns desses exemplos.

Os autores Texeira et al. (2001) aplicaram um sistema híbrido chamado de *Fuzzy Markov Predictor* (FMP). O sistema híbrido se utiliza da lógica Fuzzy e das Cadeias Ocultas de Markov, o *Hidden Markov Models*. O sistema faz a previsão de energia elétrica do instante imediatamente posterior sobre uma série mensal. Durante o trabalho os autores comparam o sistema proposto com dois métodos tradicionais Box-Jenkins e uma *Winter Exponential Smoothing*.

Diversas variáveis também são utilizadas nos modelos de previsão. Kadoshin, Nishiyama e Tosshihide (2000) utilizaram em seus modelos de previsão o crescimento da população e o desenvolvimento da economia.

Já Chow (2001) discutiu o consumo de energia em Hong Kong para o período 1984-1997, com especial ênfase para o setor doméstico. Reddy e Balachandra (2003) buscaram vários fatores que influenciam a demanda de energia na Índia no ano de 2001. Isto foi feito através do desenvolvimento de um modelo matemático integrado incorporando inúmeros fatores, como o PIB e o crescimento da população.

Já os trabalhos de Conejo (2005) e Zhou et al., (2006) fazem a previsão do preço da energia elétrica. Ambos se utilizam do modelo ARIMA como base do trabalho.

Pelas pesquisas de Pandian et al. (2006) e Serrão (2003) comprovou-se que o comportamento de consumo de energia, em uma perspectiva de curto prazo, é sensível às condições meteorológicas, a parte da jornada de trabalho, os dia da semana e as horas do dia. Por este motivo, dispor de um bom sistema de previsão de carga, usando essas variáveis climáticas associadas a carga é de suma importância.

Os autores Song e Chissom propuseram o uso de sistemas Fuzzy em séries temporais nos trabalhos (SONG; CHISSOM, 1993), (SONG; CHISSOM, 1993) e (SONG; CHISSOM, 1994). E os autores Cheng e Li propuseram uma série temporal Fuzzy, onde a caracterizam como uma sequência linguística ordenada em termos de tempo (LI; CHENG, 2012).

O trabalho proposto pelos autores Azadeh, Pazhoheshfar e Saberlic (2010) utiliza um sistema Fuzzy, através do ANFIS, para a previsão de energia anual sobre países industrializados como Estados Unidos da América, Reino Unido, Japão, França e Itália. As variáveis de entrada do sistema são o produto interno bruto (PIB) e a população.

1.5 Estruturação da dissertação

Esta dissertação está organizada em 5 capítulos e mais uma parte final de conclusão.

Nos dois capítulos seguintes é apresentada uma revisão bibliográfica das técnicas de modelagem utilizadas em séries temporais e de inteligência computacional. No capítulo 2, haverá uma abordagem sobre os modelos clássicos de previsão de séries temporais. Neles estão inclusos o: AR, MA, ARMA, ARIMA, SARIMA e SARIMAX. Nas seções do capítulo 2 também é apresentada a forma de identificação dos parâmetros do modelo, junto aos métodos utilizados para a sua aplicação. No capítulo 3 a abordagem será sobre a técnica de inteligência computacional de Sistema de Inferência Fuzzy.

O capítulo 4, de experimentos, detalha a base de dados utilizada e as variáveis exógenas que alimentam os sistemas apresentados. As variáveis exógenas de quantidade de clientes, precipitação de chuvas e temperatura ambiente são apresentadas nas subseções da seção 4.1. A seção 4.2 apresenta a análise exploratória dos dados com o pré-processamento. A Análise de Séries Temporais e os testes para a escolha dos parâmetros do modelo estatístico SARIMAX são apresentados na seção 4.3. Assim como na seção 4.3, a seção 4.4 detalha a aplicação do modelo ANFIS e as escolhas dos parâmetros a serem testados. A última seção apresenta a forma de avaliar os testes de cada modelo. O erro médio percentual (MAPE) e o Desvio Padrão dos resíduos são detalhados nas subseções da seção 4.5.

A partir dos parâmetros definidos na seção anterior, o capítulo 5 apresenta os resultados obtidos através dos modelos anteriormente apresentados. A primeira seção apresenta os resultados do modelo SARIMAX. Em três subseções são destacados com gráficos e tabelas a melhor combinação de duas e três variáveis exógenas do modelo SARIMAX. A segunda seção apresenta os resultados do Sistema de Inferência Fuzzy. Nas

três subseções destacadas, são apresentadas as mesmas combinações de duas e três variáveis exógenas da seção 5.1. Nessa subseção também são apresentados gráficos e tabelas de cada teste. Na terceira seção são apresentados os comparativos entre os dois diferentes modelos.

Por fim, também são apresentadas as conclusões dos testes realizados ao longo desta dissertação. Também são descritas as considerações finais das variáveis exógenas utilizadas ao longo deste estudo junto a propostas para trabalhos futuros.

2 DESCRIÇÃO DAS TÉCNICAS DE MODELAGEM

Diversas técnicas de modelagem matemática são aplicadas em séries temporais para previsão. Dentre as ferramentas de inteligência computacional existentes, será abordada neste trabalho a descrita ao longo desta seção¹⁷.

Existem várias abordagens diferentes para modelagem de séries temporais. O modelo mais usual de séries temporais é o Box-Jenkins. O trabalho de Box-Jenkins denominado “*Time Series Analysis: Forecasting and Control*” foi criado em 1976 e tem sido utilizado em inúmeros trabalhos acadêmicos.

A seção a seguir introduz o que é uma série temporal, explicitando definições e sua relação no mundo real. Como é imprescindível a estacionariedade de uma série temporal para aplicação dos modelos clássicos, a seção de estacionariedade das séries temporais precede as seções dos modelos autoregressivos, de médias móveis e autoregressivos e de médias móveis.

A seção 2.6 descreve o modelo ARIMA e é sucedida da seção 2.7 que descreve as formas de identificação e validade de seus componentes.

Por fim, as seções 2.8 e 2.9 descrevem o modelo SARIMA e SARIMAX, respectivamente. E a seção 2.10 descreve o método OLS utilizado para as variáveis exógenas desta dissertação.

2.1 Séries Temporais

Uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. Os autores Souza e Camargo definem uma série temporal como um conjunto de observações de uma dada variável, ordenadas segundo o parâmetro tempo (SOUZA; CAMARGO, 1996). Para tanto, há um conjunto de técnicas estatísticas disponíveis que dependem do modelo definido, bem como do tipo de série analisada e do objetivo do trabalho.

Os autores Latorre e Cardoso definem uma série temporal como uma sequência de dados obtidos em intervalos regulares de tempo durante um período específico (LATORRE; CARDOSO, 2001). É uma coleção de observações feitas sequencialmente ao longo do tempo. Para tanto, há um conjunto de técnicas estatísticas disponíveis que dependem do modelo

¹⁷Ressalta-se que nenhuma técnica de previsão é perfeita, sempre haverá erros envolvidos neste processo, pois são vários os fatores que influenciam o ambiente de negócios (CHASE; JACOBS; AQUILANO, 2006).

definido (ou estimado para a série), bem como do tipo de série analisada e do objetivo do trabalho (LATORRE; CARDOSO, 2001).

As séries temporais são medidas de determinadas variáveis (vendas e receitas de um determinado produto, quantidade de matéria-prima, vazão de um determinado sistema, etc.) tomadas a intervalos regulares de tempo (MOORE et al., 2006).

O interesse no estudo de séries temporais reside no fato de que sua análise trata com observações dependentes, dependência esta em que estamos particularmente interessados (SOUZA; CAMARGO, 1996).

Séries temporais ocorrem com muita frequência na prática e no dia-a-dia, surgindo nas mais variadas áreas de aplicação. As técnicas associadas a elas nos auxiliam a prever um resultado futuro baseados em dados passados e presentes (RODRIGUES; SILVA; LINDEN, 2007).

Uma previsão é uma estimativa quantitativa (ou conjunto de estimativas) acerca da verossimilhança de eventos futuros, baseados na informação atual e passada (SOUZA; CAMARGO, 1996).

A análise de series temporal é uma importante área da previsão de demanda onde dados históricos de algumas variáveis são coletados visando obter um modelo que descreva a relação subjacente destas variáveis (ZHANG, 2003). A motivação na previsão de séries temporais é explorar a possível relação funcional e estatística em certo período de tempo. Sendo assim, útil para a tomada de decisão (LI e CHIANG, 2013).

Na análise de uma série temporal, primeiramente deseja-se modelar o fenômeno estudado para, a partir daí, descrever o comportamento da série, fazer estimativas e, por último, avaliar quais os fatores que influenciaram o comportamento da série, buscando definir relações de causa e efeito entre duas ou mais séries (LATORRE; CARDOSO, 2001).

Ao se comentar série temporal, se assumirá uma série temporal como uma realização de um processo estocástico com as propriedades de estacionariedade e ergodicidade. Processos estocásticos são sistemas que evoluem no tempo e/ou no espaço de acordo com leis probabilísticas (SOUZA; CAMARGO, 1996).

2.2 Estacionariedade das Séries Temporais

Como dito no início deste capítulo, a estacionariedade de uma série é um requisito para a aplicação dos modelos autoregressivos, de médias móveis e autoregressivos e de

médias móveis em séries temporais. Para a aplicação do modelo de Box-Jenkins, é necessário que a série temporal seja estacionária ou não apresente tendência, ou sazonalidade (SÁFADI, 2004).

Como requisito do Box-Jenkins o processo estocástico precisa ser estacionário e ergódico. Um processo é estacionário de segunda ordem quando ele oscila em torno de uma média com uma variância constante, de modo que a auto-correlação dependa somente do *lag*, ou seja, independente do tempo (MORETTIN, 2006). Um processo estacionário é dito ergódico se apenas uma realização do mesmo é o suficiente para se obter todas as estatísticas do mesmo (SOUZA; CAMARGO, 1996).

Na ilustração a seguir, Z_t é um conjunto de observações passadas, uma série histórica qualquer. Onde t é representa um tempo qualquer. Fonte: SOUZA; CAMARGO (1996).

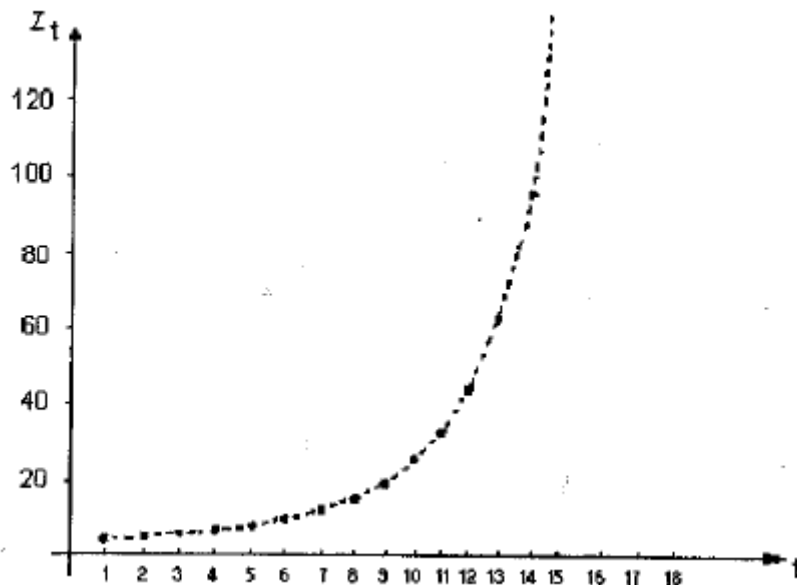


Figura 6 — Série não estacionária.

Segundo Fava (2000) e Pindyck e Rubinfeld (2004), séries não estacionárias podem ser transformadas em séries estacionárias quando suas observações são diferenciadas uma ou mais vezes. Para a transformação em uma série estacionária é necessário aplicar a diferenciação. Aplicando a primeira diferenciação na série original $\{P_t\} = \{P_1, P_2, \dots, P_n\}$ temos como resultado: $\{\Delta P_t\} = \{P_2 - P_1, P_3 - P_2, \dots, P_n - P_{n-1}\}$.

A figura abaixo ilustra a primeira diferenciação da série temporal Z_t apresentada na figura anterior. Fonte: SOUZA; CAMARGO (1996).

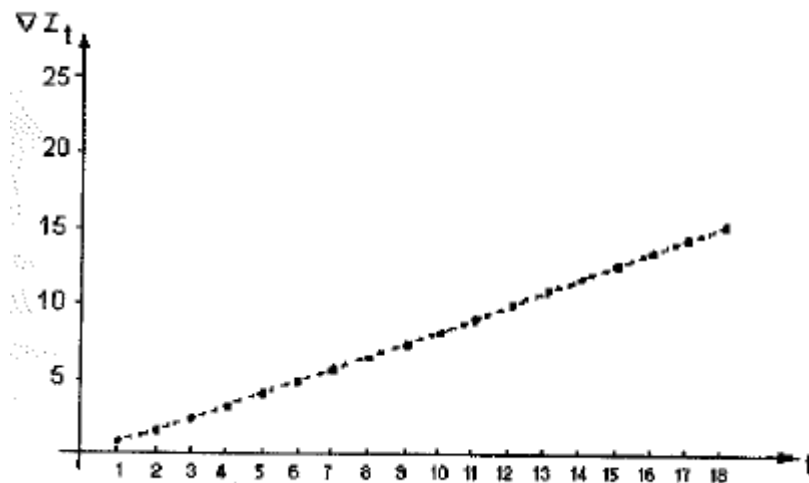


Figura 7 — Série não estacionária após a primeira diferenciação.

É observável que a série temporal ilustrada acima ainda possui tendência de alta. Segundo Zhou et al., a diferenciação elimina a tendência de uma série (ZHOU et al., 2006). Como a série ainda não se tornou estacionária após a primeira diferenciação, é necessário aplicar mais uma diferenciação na série para ela se tornar estacionária.

A figura a seguir ilustra a série temporal após a segunda diferenciação de Z_t . Pode-se observar que a série se tornou estacionária. Fonte: SOUZA; CAMARGO (1996):

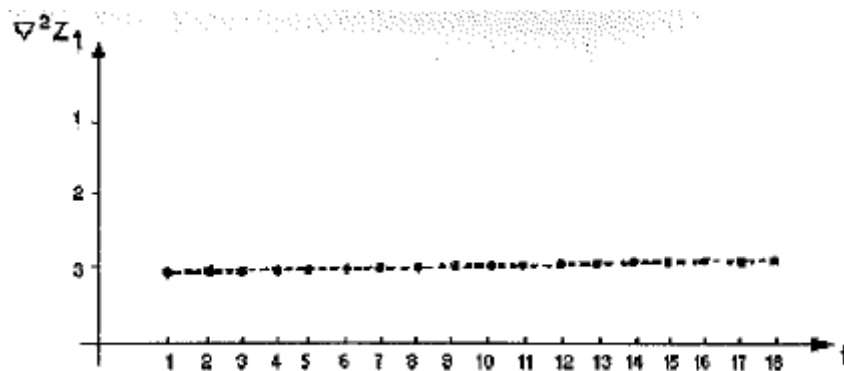


Figura 8 — Série estacionária após a segunda diferenciação.

Os autores Souza e Camargo destacam que a importância da estacionariedade de uma série “... está no fato de que, quando se trabalha com uma série estacionaria, se está em presença de uma função amostral do processo que tem a mesma forma em todos os instantes do tempo $T \in N$, o que acarreta possibilidades de obtenção de estimativa das características do processo de forma bastante simples, em caso contrário, não seria tarefa fácil” (SOUZA; CAMARGO, 1996).

2.3 Modelo Autoregressivo (AR)

O modelo autoregressivo (AR) expressa o número de observações passadas de uma série temporal que intervém na equação. Ou seja, os dados de determinada série temporal, em determinados pontos no tempo, estão fortemente correlacionados com os valores que os antecedem e os que o sucedem (LEVINE et al., 2005).

Os coeficientes autoregressivos (ϕ_p) são parâmetros que descrevem como um valor corrente relaciona-se com seus valores passados e podem ser estimados através do método de mínimos quadrados (PELLEGRINI, 2000; MATOS, 2000).

A especificação de um modelo autoregressivo de primeira ordem, AR(1) é dada pela equação abaixo (WHEELWRIGTH, 1984):

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t \quad (1)$$

A especificação para um modelo de ordem p, AR(p) é dada pela equação abaixo:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (2)$$

Onde:

- p é a ordem do modelo AR.
- Y_t corresponde à observação da série temporal no tempo t .
- ϕ_i corresponde aos parâmetro do modelo AR de ordem p , onde $i = 1, \dots, p$.
- ε representa o erro de eventos aleatórios que não podem ser explicados pelo modelo.

De acordo com Box-Jenkins o modelo AR(p) pode ser escrito da seguinte forma (SOUZA; CAMARGO, 1996):

$$Y_t = a_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_n Y_{t-n} \quad (3)$$

Onde:

- Y_t corresponde à observação da série temporal no tempo t .
- ϕ é o grau do polinômio p .
- a_t ruído branco.

O nome autoregressivo se deve ao fato de que W_t no instante t é função dos W 's nos instantes anteriores à t . A soma dos parâmetros ϕ_i deve ser menor que 1 para que o processo possa ser considerado estacionário.

2.4 Modelo de Médias Móveis (MA)

Os modelos de Médias Móveis, do inglês *Moving Average* (MA), são gerados por meio de uma média ponderada das perturbações aleatórias ou erros aleatórios de defasagem q (PINDYCK; RUBINFELD, 1981) e geram previsões médias com menor variabilidade que os dados originais. Os modelos de MA possuem como características a simplicidade e o baixo custo computacional (BRESEGHELLO, 2005).

Os modelos MA são semelhantes aos modelos AR, onde, ao invés dos valores passados da série, neste modelo, os valores considerados são os valores correntes dos erros aleatórios e seus q valores passados. (FAVA, 2000).

Um modelo de MA de ordem q é definido segunda a equação abaixo (WHEELWRIGTH, 1984).

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (4)$$

Onde:

- q é a ordem do modelo MA.
- Y_t corresponde à observação da série temporal no tempo t .
- θ_i corresponde ao parâmetro do modelo MA de ordem q , onde $i = 1, \dots, q$.
- ε representa o erro de eventos aleatórios que não podem ser explicados pelo modelo.
- μ é a expectativa de Y_t , geralmente assumida como 0.

De acordo com Box-Jenkins o modelo MA(q) pode ser escrito da seguinte forma (SOUZA; CAMARGO, 1996):

$$Y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_n B^n) a_t \quad (5)$$

Onde:

- Y_t corresponde à observação da série temporal no tempo t .
- B é o operador de atraso.
- θ é o grau do polinômio q .
- a_t ruído branco.

2.5 Modelo Autoregressivo e de Média Móvel Multiplicativo (ARMA)

Alguns processos aleatórios estacionários possuem qualidades tanto autoregressivas como de médias móveis e podem ser modelados através de um processo misto autoregressivo e de médias móveis ARMA(p, q). Esse processo pode ser identificado pela equação abaixo (WHEELWRIGTH, 1984):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (6)$$

Onde:

- p é a ordem do modelo AR.
- q é a ordem do modelo MA.
- Y_t corresponde à observação da série temporal no tempo t .
- ϕ_i corresponde aos parâmetros do modelo AR de ordem p , onde $i = 1, \dots, p$.
- ε representa o erro de eventos aleatórios que não podem ser explicados pelo modelo.
- θ_q corresponde ao parâmetro do modelo MA de ordem q , onde $i = 1, \dots, q$.
- μ é a expectativa de Y_t , geralmente assumida como 0.

De acordo com Box-Jenkins o modelo ARMA(p,q) pode ser escrito da seguinte forma (SOUZA; CAMARGO, 1996):

$$Y_t = \theta(B) \cdot \phi^{-1}(B) a_t \quad (7)$$

Onde:

- Y_t corresponde à observação da série temporal no tempo t .
- B é o operador de atraso e seu efeito sobre uma variável discreta é: $B^k Z_t = Z_{t-k}$.
- θ é o grau do polinômio q .
- ϕ é o grau do polinômio p .
- a_t ruído branco.

2.6 Modelo Autoregressivo Integrado de Média Móvel (ARIMA)

Até agora, obteve-se um processo estacionário a partir de um processo ruído branco. Como no mundo real, poucas séries temporais são estacionárias (SOUZA; CAMARGO, 1996). Deve-se buscar ajustar o modelo para que a estrutura residual forme um ruído branco¹⁸, ou seja, o resíduo se torna uma variável aleatória independente e distribuída (SÁFADI, 2004).

A partir do operador diferença definido anteriormente pode-se construir um modelo mais abrangente, composto de uma parte autoregressiva variável, explicado pelo comportamento passado da própria variável, uma parte de médias móveis, explicado pelos distúrbios de períodos anteriores, e outra definida como integração, onde o número de

¹⁸ O termo ruído branco é uma analogia com a luz branca, pois nela todas as frequências estão presentes em quantidades iguais.

defasagens (d) transforma a série em uma série estacionária, como foi descrito na seção estacionariedade das séries temporais.

A figura a seguir ilustra o fluxograma que descreve o ciclo iterativo de Box-Jenkins (SOUZA; CAMARGO, 1996).

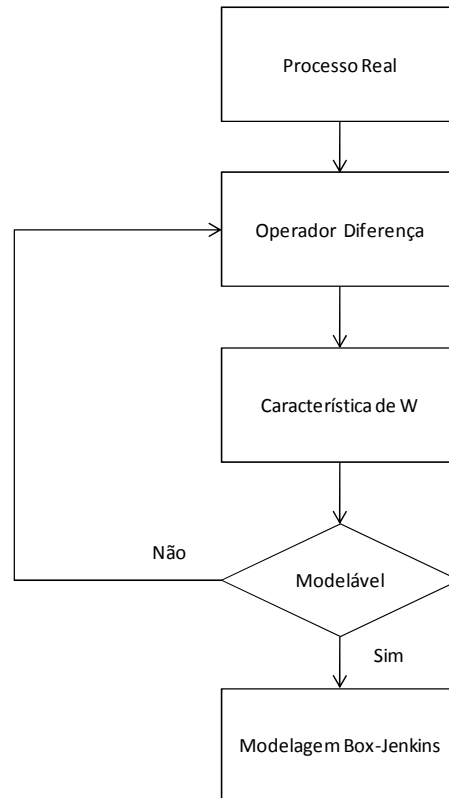


Figura 9 — Ciclo iterativo de Box-Jenkins.

A modelagem ARIMA (do inglês *Auto-Regressive Integrated Moving Average*) de séries temporais foi proposta por Box-Jenkins em 1970 (BOX; JENKINS, 1970). Seguindo o modelo de Souza e Camargo, o modelo ARIMA(p, d, q) pode ser escrito da seguinte forma (SOUZA; CAMARGO, 1996):

$$\nabla Z_t = (1 - B)Z_t = w_t \text{ e } \xi(B) = \phi(B)\nabla^d = \phi(B)(1 - B)^d \quad (8)$$

Onde:

- ∇Z_t é o operador diferença.
- Z_t é uma integral de w_t .
- ∇ é um polinômio do tipo AR.
- w_t é a diferença de Z_t .
- $\phi(B)$ é um operador autoregressivo não-estacionário de ordem $p + d$.
- B é o operador de atraso.
- μ é a expectativa de Y_t , geralmente assumida como 0.

Na metodologia Box-Jenkins do ARIMA, parte do princípio que cada valor da série temporal pode ser explicado por seus valores prévios, em virtude da utilização de uma estrutura de correlação temporal, que é algo que geralmente há entre os valores da série (GUJARATI, 2000; FAVA, 2000;).

2.7 Identificação dos modelos ARIMA (p, d e q)

Tendo definido o modelo ARIMA, é importante a busca dos termos p, d e q que compõem o modelo ARIMA. Segundo Souza e Camargo, a identificação dos modelos ARIMA consiste teoricamente de dois estágios (SOUZA; CAMARGO, 1996):

- Escolha adequada do grau de diferenciação, d.
- Escolha adequada dos graus do polinômio do modelo ARMA, p e q.

Diversos autores afirmam que para determinar a ordem dos termos p, d e q, as principais ferramentas a serem utilizadas são a FAC, a FACP e seus respectivos correlogramas (GUJARATI, 2000), (MORETTIN, 2008) e (LIU; HUDAK, 1994).

Conforme descrito acima, para o grau de diferenciação aplicam-se diferenciações na própria série temporal.

Há um “*trade-off*”¹⁹ para a identificação do número de parâmetros a serem inclusos no modelo. Embora um aumento no número de parâmetros do modelo reduza o erro de previsão, pode haver diminuição do grau de liberdade da série.

2.7.1 Função de Autocorrelação (FAC)

A estatística de autocorrelação, ou função de autocorrelação (FAC), de ordem k é uma ferramenta amplamente utilizada nos modelos de Box-Jenkins (ENDERS, 2004). Um coeficiente de autocorrelação p_1 mede a correlação entre a variável no período t em relação ao período defasado $t - 1$ e é chamado de coeficiente de autocorrelação de lag 1, assim como um coeficiente de auto correlação p_k é dito de coeficiente de autocorrelação de lag k (PELLEGRINI, 2000).

A FAC serve para medir a extensão para qual o valor tomado no tempo t, depende daquele tomado no tempo $t - k$ (SOUZA; CAMARGO, 1996).

¹⁹*Trade-off* é utilizado para designar situações de escolha entre opções conflitantes.

Segundo Greene (2003) a função autocovariância de um processo y_t pode ser expresso por:

$$\lambda_k = Cov[y_t, y_{t-k}] \quad (9)$$

A partir desta última pode-se estimar a FAC pela divisão da função autocovariância pela variância λ_0 . Assim temos:

$$\rho_k = \frac{\lambda_k}{\lambda_0}, -1 \leq \rho_k \leq 1 \quad (10)$$

Assim o número ótimo de defasagens no modelo pode ser determinado com base nos valores da FAC (MATOS, 2000).

A FAC também auxilia na análise de séries temporais, pois é uma importante ferramenta para a investigação das propriedades empíricas da série temporal, como: identificação da condição de estacionariedade e identificação de modelos potenciais a serem utilizados na modelagem e previsão da série temporal (GUJARATI, 2000; MORETTIN, 2008).

A característica de uma série temporal não estacionária é que estas apresentam fortes correlações seriais, fazendo com que os valores da função de autocorrelação declinem muito lentamente à medida que k aumenta (*lags*²⁰) (FAVA, 2000). A imagem a seguir extraída de Junior (JUNIOR, 2005) ilustra esse comportamento.

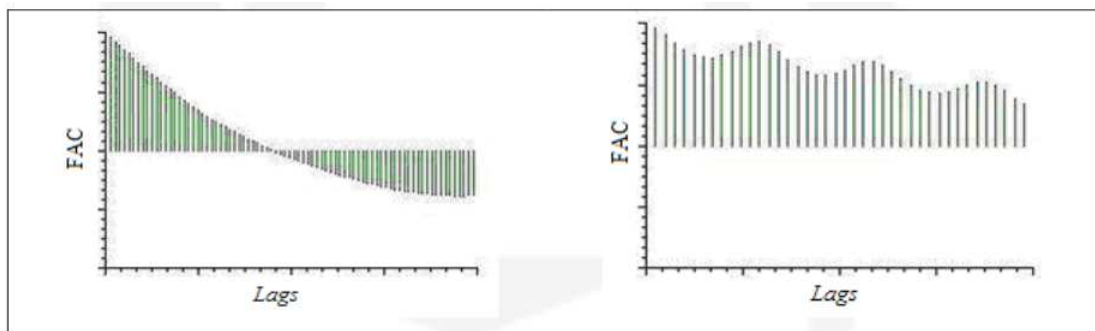


Figura 10 — Exemplo do comportamento do correlograma de séries não estacionárias.

Em ambas FACs se observa um declínio do histograma, caracterizando assim uma série temporal não estacionária.

2.7.2 Função de Autocorrelação Parcial (FACP)

Assim como a FAC, a função de autocorrelação parcial (FACP) também é utilizada para a identificação dos modelos a serem utilizados para a modelagem e previsão da série

²⁰Valores atrasados (FAVA, 2000).

temporal. A FACP também pode auxiliar na identificação da condição de estacionariedade da série temporal (MATOS, 2000; GUJARATI, 2000; MORETTIN, 2008).

Segundo Greene a correlação parcial entre duas variáveis y_t e y_{t-k} é a correlação simples existente entre elas menos aquela parte explicada linearmente pelas defasagens intermediárias (GREENE, 2003). Algebricamente:

$$\rho_k = \text{corr}[y_t - E\left(\frac{y_t}{y_{t-1}}, \dots, y_{t-k+1}\right), y_{t-k}] \quad (11)$$

A função $E(\cdot)$ poderia ser a regressão linear se a média condicional fosse linear, mas não pode. O ótimo preditor linear é a regressão linear, porém, assim o que nós temos é:

$$\rho_k = \text{corr}[y_t - \beta_1 y_{t-1} - \beta_2 y_{t-2} - \dots - \beta_k y_{t-k+1}, y_{t-k}] \quad (12)$$

Onde:

$$\beta = [\beta_1, \dots, \beta_{k-1}] = \{\text{Var}[y_{t-1}, \dots, y_{t-k+1}]\}^{-1} * \text{Cov}[y_t(y_{t-1} - \dots - y_{t-k+1})]' \quad (13)$$

Esta equação será reconhecida como um vetor de coeficientes da regressão. Como tal, o que nós estamos computando é a correlação entre um vetor de resíduos e y_{t-k} .

2.7.3 Identificação dos modelos

Uma das questões-chave na análise de séries temporais é determinar se o valor que observamos no tempo t depende do que foi observado no passado ou não. A visualização do FAC ou da FACP ajuda a identificar os modelos adequados para explicar as observações passadas e fazer previsões. A teoria mostra que a função PACF de um AR (p) - um processo de auto-regressivo de ordem p - é zero para defasamentos maiores do que p .

A dificuldade de identificação do método proposto por Box-Jenkins é pelo fato de se trabalhar com estimadores, que nunca seguem exatamente o comportamento das funções teóricas (SOUZA; CAMARGO, 1996).

Um gráfico de FAC e FACP contém os histogramas e mais duas linhas tracejadas. As linhas tracejadas são limites de significância para as autocorrelações. Os limites de significância utilizados nesta tese são de 95%. Para um intervalo de confiança de 95%, o valor crítico é 1,96. O desvio padrão é a raiz quadrada de $1/n$, onde n é o comprimento da série. A fórmula abaixo destaca esse cálculo:

$$LS = \frac{1,96}{\sqrt{n}} \quad (14)$$

$$LI = \frac{-1,96}{\sqrt{n}} \quad (15)$$

Onde o LS é o Limite Superior e o LI é o Limite Inferior.

Os *lags* cujos valores da autocorrelação ultrapassam esses limites são ditos como significantes (GOODRICH; STELLWAGEM, 1999). Como exemplo: uma autocorrelação é significativa de *lag* 1 se, a partir do *lag* 2, as autocorrelações estão todas dentro do intervalo, indicando um “corte” no *lag* 1.

Os autores Souza e Camargo resumem o comportamento das FAC e FACP para cada modelo através dos quadros abaixo (SOUZA; CAMARGO, 1996). Como ilustração, pode-se destacar as características destas funções no que tange a alguns processos:

- AR(p): ACF infinita decrescente e PACF finita (nula para *lags* maiores que p).
- MA(q): ACF finita (nula para *lags* maiores que q) e PACF infinita decrescente.
- ARMA(p,q): ACF infinita (para *lags* maiores que $p - q$) e PACF infinita (para *lags* maiores que $p - q$).

De forma geral, para se identificar a ordem p observa-se se a FAC decresce e se a FACP apresenta um corte. Se isto acontecer, o *lag* onde este corte ocorre nos fornece a ordem p . Analogamente para o modelo MA as FAC e FACP devem apresentar comportamento inverso para se encontrar a ordem q .

Torna-se importante salientar que a inspeção gráfica da série é uma ferramenta poderosa para se fazer inferência e para uma melhor prospecção do modelo. Nestes termos, imputou-se ao modelo inicial um teste de sobrefixação, cujo objetivo é verificar se o número de parâmetros é correto.

2.7.4 Estimativas dos parâmetros do modelo

Para a estimativa dos parâmetros do modelo utiliza-se um teste para os resíduos e um teste de sobrefixação. No teste dos resíduos procura-se constatar que o resíduo gerado pelo modelo é um ruído branco. Assim, o erro não apresentaria nenhuma estrutura de correlação.

O teste de sobrefixação consiste em gerar modelos de ordem superior ao identificado, de forma que se possa reforçar a pertinência deste.

2.7.5 Crítérios Akaike (AIC)

Esta seção descreve as estatísticas de ajuste e de análise dos resíduos

Os critérios Akaike (AIC) mede a verossimilhança do modelo. Este critério é calculado através da seguinte equação:

$$AIC = 2k - 2\ln(L)$$

(16)

Onde k é o número do parâmetro do modelo e L é o maior valor da função do modelo.

$$AIC = \ln(MQR) + 2\frac{NP}{NR} \quad (17)$$

Onde MQR é a média dos quadrados dos resíduos, NP é número de parâmetros do modelo e NR é o número de resíduos.

O critério AICc é o AIC com uma correção para tamanhos de mostra finita. Este critério é calculado através da seguinte equação:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad (18)$$

Onde n indica o tamanho da amostra.

Em resumo, escolhe-se aquele que apresenta o menor AIC.

2.8 Modelo Multiplicativo Autoregressivo Integrado de Média Móvel (SARIMA)

Esta seção apresenta o modelo de Box-Jenkins com uma componente sazonal²¹. Podem-se citar como exemplos de séries temporais sazonais as séries de consumo e geração de energia elétrica.

Box-Jenkins (1970) generalizaram o modelo ARIMA para lidar com sazonalidade e definiu-se um modelo ARIMA multiplicativo, denominado SARIMA. Em sua fórmula o modelo SARIMA contém parte sazonal na primeira parte da fórmula, e parte não sazonal dos parâmetros na segunda parte da fórmula (ESPINOSA; PRADO; GHELLER, 2010).

Como extensão natural, pode-se definir o modelo SARIMA como (SOUZA; CAMARGO, 1996):

$$\phi(B) \cdot \phi(B^S) \cdot \nabla_s^D \cdot \nabla^d Z_t = \theta(B) \cdot \theta(B^S) \cdot a_t \quad (19)$$

Onde:

- $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ é um operador não sazonal autoregressivo.
- $\nabla^d = (1 - B)^d$ é um operador diferença não sazonal de ordem d .
- $\phi(B^S) = (1 - \phi_1 B^S - \phi_2 B^{2S} - \dots - \phi_p B^{pS})$ é um operador sazonal autoregressivo.
- $\nabla_s^D = (1 - B^S)^D$ é um operador sazonal de ordem D .
- $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ é um operador não sazonal de médias móveis.

²¹ Sazonalidade é a tendência do processo em repetir certo tipo de comportamento dentro de um determinado período.

- $\theta(B^s) = (1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs})$ é um operador sazonal de médias móveis.
- $\nabla_S^D \cdot \nabla^d Z_t = w_t$ é um filtro não linear aplicado a série original Z_t que produz um processo estacionário w_t .
- a_t é um ruído branco.

O modelo de previsão autoregressivo integrado de média móvel multiplicativo sazonal (SARIMA) é formulado com as componentes apresentadas a seguir (WANGDI et al., 2010):

- p e P - o autoregressivo e o sazonal autoregressivo, respectivamente;
- d e D - as diferenças não sazonais e sazonais, respectivamente;
- q e Q - os parâmetros de média móvel e o sazonal, respectivamente.
- s , representando o comprimento do período sazonal.

A análise para a construção dos modelos SARIMA $(p, d, q)(P, D, Q)_s$ é, de certa forma, semelhante aos modelos ARIMA (p, d, q) , pois um modelo é uma extensão do outro (FAVA, 2000).

2.9 Modelo Multiplicativo Autoregressivo Integrado de Média Móvel com Variáveis Exógenas (SARIMAX)

A fim de melhorar o desempenho de previsão de valores futuros, outra série de tempo pode ser incorporada nos modelos de séries temporais (CHRISTO; SOUZA, 2006).

A adição de uma entrada externa para um modelo é chamado usando uma variável exógena. Por exemplo, o seguinte é um modelo ARMA ordinária com uma única variável exógena, ARMAX.

$$Y_t = \sum_{i=1}^p (\phi_i Y_{t-i}) + \sum_{j=1}^q (\theta_j a_{t-j}) + \sum_{h=1}^b \beta_h X_{t-h} + \omega_t \quad (20)$$

Onde:

- X_t são as variáveis exógenas.

Assim, a composição de um modelo SARIMAX é pelas estruturas: autoregressiva, Média Móvel, integrador, uma componente sazonal e variáveis exógenas. Em resumo, a componente autoregressiva, ou AR, é uma autoregressão que é uma regressão que tenta explicar os valores usando seus valores anteriores. A componente de Média Móvel, ou MA. A parte “I” (do inglês *integrated*), que permite que a série seja não estacionária. A parte sazonal, S, adiciona periodicidade a ARIMA. Por fim, adicionou-se a variável exógena, X, que

basicamente permite o uso de variáveis externas a serem consideradas no modelo, como por exemplo, temperatura.

O modelo SARIMAX estima a variável dependente com base em variáveis exógenas e modela o resíduo por meio de um modelo SARIMA. O modelo SARIMAX estima a variável dependente por meio de uma regressão linear múltipla em conjunto com o ajuste de um modelo multiplicativo autoregressivo integrado de médias móveis. O modelo explica a variável dependente por meio de: variáveis exógenas; defasagens das variáveis exógenas; defasagens da variável dependente (MOURA; MONTINI; CASTRO, 2011).

Portanto, faz-se necessidade de apresentar os modelos de regressão a seguir.

2.10 Modelos de Regressão para as Variáveis Explicativas no SARIMAX

Para possibilitar a adição de variáveis exógenas no SARIMA usam-se modelos de regressão. Os modelos de regressão dinâmica combinam a dinâmica de séries temporais e o efeito de variáveis explicativas. A palavra “dinâmica” significa um método de regressão no qual se inclui a estrutura de dependência de uma série temporal.

Eles devem ser usados quando existe uma estrutura de dependência entre a variável de interesse e possíveis variáveis causais e, ao mesmo tempo, quando a estrutura de correlação da série dependente indicar que não se pode supor a independência dos erros.

Nos modelos de regressão dinâmica, a variável dependente é explicada por seus valores defasados e pelos valores atuais e passados de variáveis causais ou exógenas. (ZANINI, 2000)

Na definição do modelo adequado, é necessário levar em conta a significância dos parâmetros.

Os modelos de regressão dinâmica podem ser descritos pela equação a seguir:

$$\Phi(B)Y_t = \beta x_t + \varepsilon_t \quad (21)$$

Onde:

- Y_t é a variável dependente (endógena) no instante t .
- β é o vetor de coeficientes das variáveis causais, que será estimado por mínimos quadrados.
- x_t é o vetor de variáveis causais (exógenas) no instante t .
- ε_t é o ruído aleatório associado ao modelo, onde supomos que os ε_t são independentes e identicamente distribuídos com densidade $N(0, \sigma^2)$.

- $\Phi(B)$ é o polinômio autoregressivo de ordem p .

Esta seção apresenta de forma resumida dois modelos de regressão. A seguir é apresentado o Modelo de Regressão Linear, que será utilizado nos experimentos desta dissertação. Em seguida, é apresentado o modelo Cochrane-Orcutt (1949).

2.10.1 Modelo de Regressão Linear (OLS)

O modelo utilizado nas variáveis explicativas do SARIMAX deste trabalho é o *Ordinary Least Square* (OLS). O modelo de regressão linear OLS ajusta as variáveis exógenas utilizando o modelo de regressão linear clássico, onde os resíduos são modelados utilizando um modelo SARIMA. Ele se encaixa aos dados para um modelo especificado na tentativa de minimizar a soma total dos quadrados dos erros.

Sem dúvida o OLS é uma das técnicas mais antiga e mais amplamente aplicada para estimativa de esforço de software é a de regressão linear simples.

Esta técnica ajusta a função de regressão linear a um conjunto de dados contendo variáveis independentes e dependentes, este tipo de regressão também é comumente referido como regressão múltipla (DEJAER et al., 2012).

A regressão OLS assume o seguinte modelo linear dos dados (DEJAER et al., 2012):

$$e_i = x_i' \beta + b_0 + \epsilon_i \quad (22)$$

Onde:

- x_i' representa o vetor contendo os valores de i th observações.
- β é o vetor que contém os parâmetros de inclinação que são estimados pela regressão.
- b_0 é um escalar de interceptação.
- ϵ_i é o erro associado a cada observação.

O modelo OLS corresponde a minimizar a soma dos quadrados diferenças entre os valores observados e preditos. O erro ϵ_i é minimizado através de $\min \sum_{i=1}^N \epsilon_i^2$.

2.10.2 Modelo Cochrane-Orcutt (CO-LS)

O modelo de regressão generalizado de Cochrane-Orcutt é dado por (COCHRANE; ORCUTT, 1949):

$$\Phi(B)Y_t = \beta x_t + W_t \quad (23)$$

$$R(B)W_t = \epsilon_t \quad (24)$$

Onde $R(B)$ é um polinômio autoregressivo.

As duas fórmulas apresentadas acima podem ser escritas na forma abaixo.

$$W_t = \varphi(B)Y_t - \beta X_t \quad (25)$$

Desta última expressão nota-se que o modelo de regressão generalizado de CO-LS introduz defasagens tanto na variável dependente (Y_t) quanto nas causais. A fórmula também indica também uma relação de entre Y_t e X_t não é afetada pela introdução do polinômio autoregressivo $R(B)$.

3 SISTEMAS DE INFERÊNCIA FUZZY

Este capítulo tem por fundamento apresentar ao leitor uma revisão bibliográfica do modelo de inteligência computacional aplicado ao longo do trabalho. Neste capítulo 3 está contida a lógica nebulosa, conhecida como Lógica Fuzzy, junto a conceitos de conjuntos Fuzzy, funções de pertinência, as variáveis linguísticas e as operações lógicas. A última subseção destina-se ao Sistema Fuzzy a proposto.

3.1 Lógica Fuzzy

A lógica Fuzzy, conhecida como lógica nebulosa, ou difusa, é uma técnica de inteligência computacional que tem como objetivo modelar o modo aproximado de raciocínio do ser humano (CAMPOS; JESUS; MENDES, 2007). Assim é possível tomar decisões em um ambiente de incerteza e imprecisão (ZADEH, 1965), o que não ocorre com a lógica discreta.

Como a lógica Fuzzy é formada a partir da teoria de conjuntos Fuzzy. As subseções abaixo descrevem esse entendimento por completo.

3.1.1 Conjuntos Fuzzy

Formalmente, um conjunto nebuloso do universo de discurso é definido por uma função de pertinência, $\mu_A: U \rightarrow [0,1]$. Essa função associa a cada elemento de um conjunto um grau. A função de pertinência indica o grau de compatibilidade (SANDRI e CORREA, 1999).

Um conjunto Fuzzy é um conjunto sem uma fronteira nítida²², ou claramente definida. Na forma clássica a associação de um elemento em relação a um conjunto é assumida com termos binários. Ou seja, o elemento pode somente pertencer ou não pertencer a um conjunto, 0 ou 1.

Diferentemente da forma clássica, na forma nebulosa um elemento pode pertencer a mais de um conjunto Fuzzy com diferentes graus de pertinência (TANSCHKEIT, 2007). Segundo Kuok, Fu e Wong o conceito de conjuntos Fuzzy é melhor do que o clássico porque

²²Já os conjuntos CRISP possuem uma fronteira claramente definida.

os conjuntos Fuzzy permitem uma suave transição entre conjuntos (KUOK; FU; WONG, 1998).

Pode-se exemplificar essa diferença entre Fuzzy e a forma clássica através do conjunto de dias de finais de semana. O conjunto de dias de finais de semana contém: sábado e domingo. Mas sexta-feira é considerada por todos como uma parte do fim de semana.

Conjuntos clássicos não trabalham este tipo de classificação parcial. Ou algo é considerado contido ou não.

Portanto com a teoria de conjuntos Fuzzy, trabalha-se com a forma como as pessoas realmente percebem o conceito de fim de semana. Em outras palavras, a sua grande vantagem é a capacidade de responder a uma questão de sim ou não com um não-muito-sim-ou-não houve resposta.

Na prática, a sexta-feira pode ser considerada como pertencente a 0,74 do conjunto final de semanal.

3.1.2 Funções de pertinência

A função de pertinência, em inglês, *Membership Function* (MF) é uma curva que define a forma como cada ponto no espaço de entrada é mapeado para um valor de pertinência, que no exemplo anterior era de 0,74. A função de pertinência deve estar entre 0 e 1.

É muito importante a escolha do tipo de função de pertinência, pois ela influencia diretamente nos resultados do sistema Fuzzy. Portanto a sua escolha depende do contexto do problema estudado e da experiência do usuário.

Retornando ao exemplo do final de semana, uma função de pertinência em forma de degrau não se adequa ao problema. Como se definiu sexta-feira como 0,74 pertencente ao conjunto final de semana, uma função com crescimento *sigmoidal* definiria melhor essa transição entre os dias.

A função de pertinência *sigmoidal* é apresentada a seguir (JANG; GULLEY, 1995).

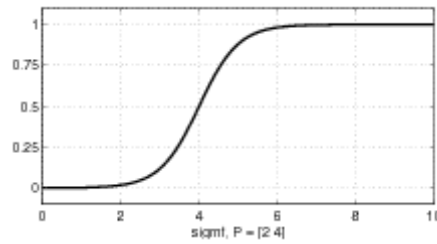


Figura 11 — Função de pertinência sigmoide.

As funções de pertinência mais usuais são as parametrizáveis. Dentre elas: a triangular (*trimf*) e a trapezoidal (*trapmf*). Mas também há: a Gaussiana (*gaussmf* e *gaussmf2*), a *generalizedbell* (*gbellmf*) e a *sigmoide* (*sigmf*).

As figuras apresentadas a seguir ilustram as funções utilizadas no Sistema de Inferência Fuzzy desta dissertação (JANG; GULLEY, 1995).

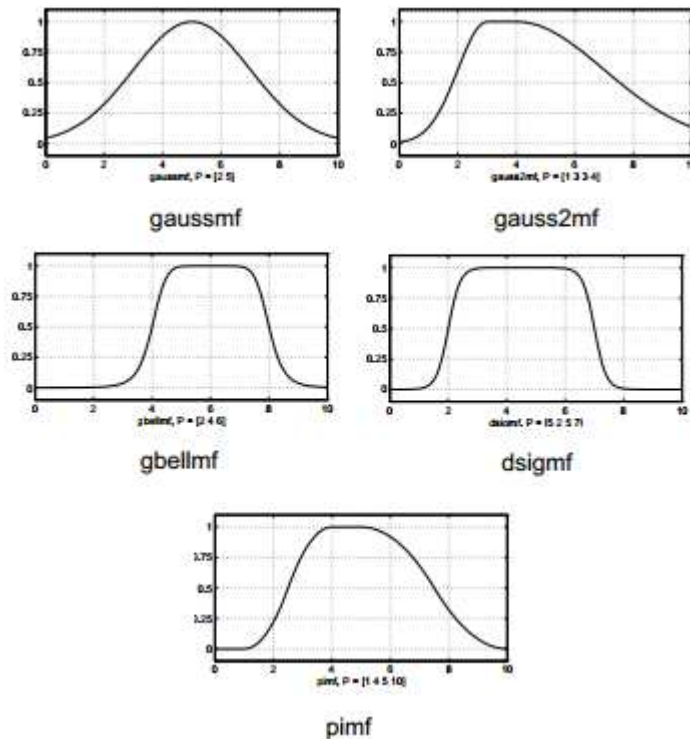


Figura 12 — Funções de pertinência.

A seguir são apresentadas as fórmulas das respectivas funções de pertinência apresentadas acima. Sendo a, b, c, d os parâmetros numéricos dos respectivos modelos. Esses parâmetros ajustam a forma da função em questão.

As funções de pertinência são caracterizadas por um conjunto de valores, Dos quais podem ser denominados de a, b, c e d. Onde a e a última letra de cada função de pertinência determinam o intervalo dentro do qual a função de pertinência assume valores diferentes de zero.

$$f(x; \sigma; c) = e^{-\frac{(x-c)^2}{2\sigma^2}} \quad (26)$$

Os parâmetros σ e c listados representam a função *gaussmf*. A função de pertinência *gauss2mf* é uma combinação de dois destes dois parâmetros. A primeira função, especificado por *sig1* e *c1*, determina a forma da curva mais à esquerda. A segunda função especificada por *sig2* e *c2* determina a forma da curva mais à direita. Sempre que $c1 < c2$, a função *gauss2mf* atinge um valor máximo de 1. Caso contrário, o valor máximo é menor do que um.

A função de *gbellmf* é apresentada abaixo:

$$f(x; a; b; c) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}} \quad (27)$$

Onde o parâmetro b é normalmente positivo. O parâmetro c localiza o centro da curva.

A função de pertinência *pimf* é avaliada nos pontos determinados pelo vetor x . Os parâmetros a e d determinam os "pés" da curva, enquanto b e c determinam seus "ombros". A função de pertinência *pimf* é dada por:

$$f(x; a, b, c, d) = \begin{cases} 0, x \leq a \\ 2 \left(\frac{x-a}{b-a} \right)^2, \frac{a+b}{2} \leq x \leq b \\ 1 - 2 \left(\frac{x-b}{b-a} \right)^2, \frac{a+b}{2} \leq x \leq b \\ 1, b \leq x \leq c \\ 1 - 2 \left(\frac{x-c}{d-c} \right)^2, c \leq x \leq \frac{c+d}{2} \\ 2 \left(\frac{x-d}{d-c} \right)^2, \frac{c+d}{2} \leq x \leq d \\ 0, x \geq d \end{cases} \quad (28)$$

Por fim, a função *dsigmf* é a diferença entre duas funções *sigmoidals*.

$$f(x; a, c) = \frac{1}{1 + e^{-a(x-c)}} \quad (29)$$

3.1.3 Variáveis Linguísticas

Uma variável linguística é aquela cujos valores são palavras em lugar de números (TANSCHHEIT, 2007). Na lógica Fuzzy, uma variável linguística é uma variável cujos valores são nomes de conjuntos Fuzzy. A principal função das variáveis linguísticas é fornecer uma maneira sistemática para uma caracterização aproximada de fenômenos através da descrição linguística. Isto permite um melhor entendimento do problema que está sendo analisado, permitindo o tratamento de sistemas que são muito complexos para serem analisados através de mecanismos matemáticos convencionais (GOMIDE; GUDWIN; TANSCHHEIT, 1995).

3.1.4 Operações Lógicas

A partir do entendimento da inferência Fuzzy, pode-se apresentar as operações lógicas.

Da mesma forma que na teoria dos conjuntos clássica, também é possível realizar operações de interseção, união e negação com conjuntos Fuzzy. Os operadores de interseção e união podem ser definidos através das normas e conormas triangulares (t-normas e s-normas, respectivamente). Seguem-se as respectivas definições (PEDRYCZ, 1989). Uma t-norma é uma função binária $t: [0,1] \times [0,1] \rightarrow [0,1]$, que satisfaz às seguintes condições, $\forall x, y, z, w \in [0,1]$:

$$\text{Monotonicidade:} \quad t(x, w) \leq t(y, z), \text{ para } x \leq y \text{ e } w \leq z \quad (1)$$

$$\text{Comutatividade:} \quad t(x, y) = t(y, x) \quad (2)$$

$$\text{Associatividade:} \quad t(t(x, y), z) = t(x, t(y, z)) \quad (3)$$

$$\text{Condições limites:} \quad t(x, 0) = 0 \text{ e } t(x, 1) = x \quad (4)$$

3.2 O sistema Fuzzy

A partir das informações apresentadas sobre lógica Fuzzy pode-se conceituar um sistema Fuzzy baseado em regras (SFBR). Os Sistemas Fuzzy representam uma das áreas de aplicação mais importantes da Teoria dos Conjuntos Fuzzy, sendo que o modelo mais utilizado é aquele baseado em regras. Eles são uma extensão dos sistemas clássicos baseados em regras, porque utilizam regras do tipo “se-então”, onde os antecedentes e consequentes são proposições Fuzzy (HERRERA, 2008; JANG; GULLEY, 1995).

Existem dois tipos de sistemas de inferência que variam conforme o tipo de saída determinado: Mamdani e Sugeno. O Mamdani é usado através da teoria dos conjuntos Fuzzy. O Sugeno, ou Takagi-Sugeno-Kang foi introduzido em 1985. Ele é semelhante ao Mamdani em muitos aspectos. A principal diferença entre Mamdani e o Sugeno é que as funções de pertinência de saída Sugeno são ou linear ou constante (SUGENO, 1985).

As regras Fuzzy “se-então”, “ou”, “senão” e “e” são regras cuja saída pode ser número ou linguístico. Exemplificando, “se x é A então y é B ” como um ponto Fuzzy $A \times B$ (ZADEH, 1965).

A estrutura mais usual de um processo controlado por um controlador Fuzzy é mostrada na figura abaixo e a sua descrição também segue no texto mais abaixo:

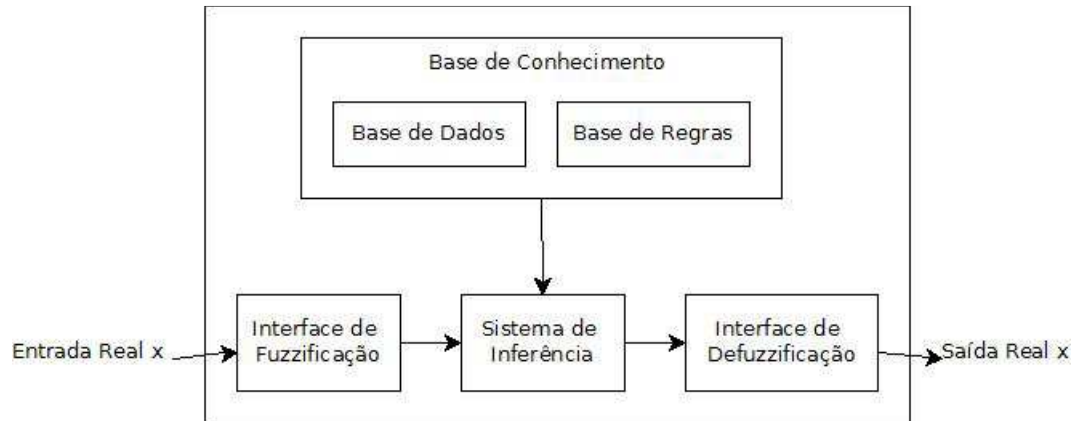


Figura 13 — Modelo Geral de um Sistema de Inferência Fuzzy.

Ele é composto por uma interface de fuzzificação, uma unidade de inferência, uma interface de defuzzificação e por uma base de conhecimento, que por sua vez divide-se em Base de dados e a Base de Regras.

Fuzzificação: a principal função de um fuzzificador é converter os valores reais de entrada (escalar ou vetorial) em um grau de pertinência a conjuntos Fuzzy para que sejam tratados pela máquina de inferência. Segundo Wang (1997), dentre os fuzzificadores mais utilizados: *Singleton*, Gaussiano e o Triangular.

Inferência: Nesta etapa ocorrem as operações com conjuntos Fuzzy. Os operadores mais usuais são “e” e “ou”.

Regras: As regras podem ser formadas a partir de um especialista, através da linguística. Mas como a extração de regras do especialista é uma tarefa muito difícil, há alguns métodos de extração de regras de dados numéricos.

Dados: Os dados são objeto de estudo. Nele estão armazenadas as informações necessárias para todo o Sistema de Inferência Fuzzy.

Defuzzificação: é definido como um mapeamento de um conjunto fuzzy, produzido pela máquina de inferência, em um valor real. Isto é, especificar um ponto na saída que melhor represente o conjunto Fuzzy. Na escolha de um defuzzificador, os critérios de: plausibilidade (o valor de saída é intuitivo), simplicidade computacional, e de continuidade, devem ser considerados. Segundo Wang (1997), dentre os defuzzificadores mais utilizados: centro de gravidade, centro ponderado e o máximo.

Os autores Gomide, Gudwin e Tanscheit explicam a interface de fuzzificação através dos valores das variáveis de entrada para condicionar universos de discurso, transformando números em conjuntos fuzzy, de modo que possam se tornar instâncias de variáveis linguísticas. A base de conhecimento consiste de uma base de regras, caracterizando a estratégia de controle e suas metas. A base de dados armazena as definições necessárias sobre

discretizações e normalizações dos universos de discurso, as partições fuzzy dos espaços de entrada e saída e as definições das funções de pertinência. O procedimento de inferência processa os dados Fuzzy de entrada, junto com as regras, de modo a inferir as ações de controle Fuzzy, aplicando o operador de implicação Fuzzy e as regras de inferência da lógica Fuzzy. A interface de defuzzificação transforma as ações de controle Fuzzy inferidas em ações de controle não-fuzzy. Em seguida, efetua um escalamento, de modo a compatibilizar os valores normalizados vindos do passo anterior com os valores dos universos de discurso reais das variáveis (GOMIDE, GUDWIN e TANSCHKEIT, 2000).

Os sistemas de inferência Fuzzy são utilizados para formar modelos em séries temporais. Esses sistemas têm sido amplamente aplicados devido à sua interpretabilidade, capacidade de lidar com dados não lineares e por não são necessárias premissas rígidas para os dados.

4 EXPERIMENTOS E MODELOS OBTIDOS

Este capítulo está dividido em quatro seções, cada seção é uma etapa da aplicação da metodologia nesta dissertação: a primeira e segunda seção descrevem a base de dados utilizada no trabalho, as variáveis explicativas e o pré-processamento dos dados. A terceira e quarta seção explicitam os modelos de previsão propostos: SARIMAX e o Sistema de Inferência Fuzzy, respectivamente. A quinta e última seção descreve como serão avaliados os modelos apresentados.

4.1 Descrição dos Dados

Apesar de não haver um consenso sobre as melhores variáveis que descrevem o funcionamento do serviço de distribuição de energia (PINHEIRO, 2012), optou-se pelo uso de três variáveis exógenas: a quantidade de clientes, a precipitação de chuvas e a temperatura ambiente foram utilizadas nos modelos.

A escolha dessas três variáveis exógenas para este trabalho se deve ao seu largo uso nos trabalhos relacionados descritos no item 1.4. Essas variáveis exógenas também foram escolhidas por serem acompanhadas em relatórios gerenciais da empresa em questão.

Os dados disponíveis de normais climatológicas são do Instituto Nacional de Meteorologia²³ (INMET). Os outros dados medidos foram fornecidos pela própria empresa em estudo, vide documento em anexo.

Os dados do INMET e os fornecidos pela empresa foram medidos no município de Salvador devido a sua maior densidade demográfica na região Metropolitana do Estado da Bahia.

4.1.1 Série Temporal de Energia Distribuída

Como apresentado no capítulo introdutório: a base de dados utilizada para este trabalho é de distribuição de energia elétrica da empresa distribuidora de energia Coelba.

²³ Os dados foram extraídos do site do INMET. Disponível em: <<http://www.inmet.gov.br/portal/>>. Acesso em: 17dez. 2014.

A série temporal objeto de estudo tem periodicidade diária. Para inferir os modelos em um mesmo padrão de distribuição de energia elétrica, foram utilizados somente os dias de semana, descartando assim os finais de semana (sábados e domingos) da série temporal. Os feriados foram mantidos na série temporal, exceto os que são nos finais de semana.

A exclusão dos finais de semana se deve a um diferente padrão de consumo entre os dias de semana e os finais de semana, como por exemplo, o tempo de acionamento de aparelhos eletrônicos como televisão, ar condicionado e máquinas de lavar roupa. Além disso, o horário de pico dos dias de semana geralmente é diferente do horário de pico dos fins de semana. Assim a relação entre a energia distribuída e as variáveis exógenas²⁴ durante os dias de semana é diferente da relação entre ambas nos finais de semana. Portanto, a série temporal de energia elétrica com início em janeiro de 2010 e fim em outubro de 2013 é composta por 983 pontos.

Portanto, a série temporal de dias de semana de energia elétrica distribuída com início em janeiro de 2010 e fim em outubro de 2013 é composta por 983 pontos. A figura abaixo ilustra essa nova composição.

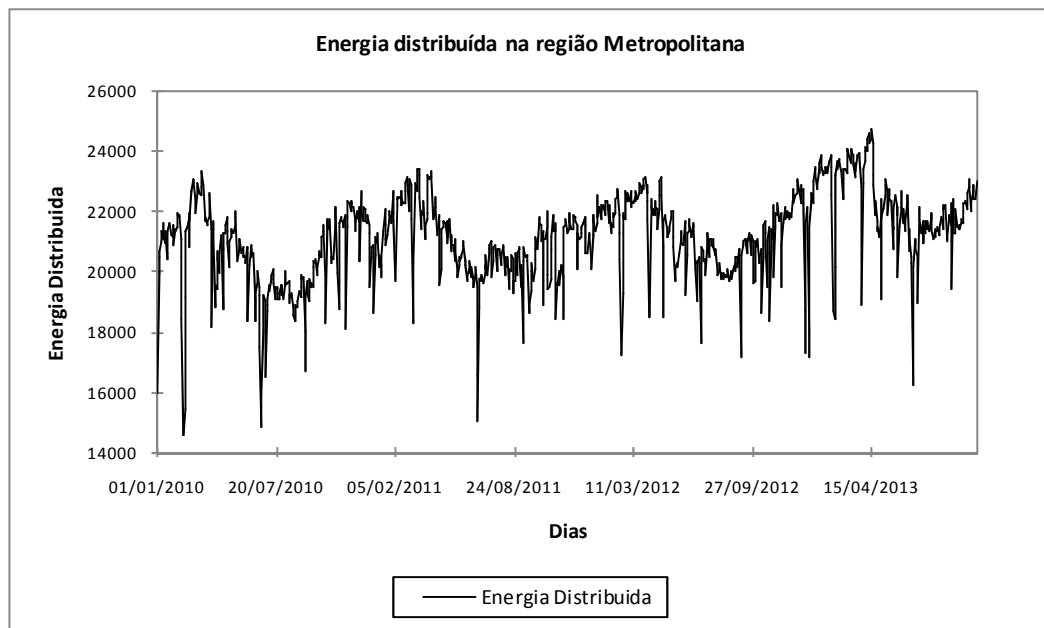


Figura 14 — Série de energia distribuída da região Metropolitana sem finais de semana.

Pode-se observar que a série temporal acima possui sazonalidade e tendência. As sazonalidades presentes são semanais, de estação de ano e anual.

²⁴As variáveis exógenas de temperatura, precipitação e número de clientes serão utilizadas neste trabalho.

4.1.2 Número de Clientes

A variável “clientes” é a quantidade de unidades comerciais ou residenciais ligadas à rede de energia da distribuidora. Segundo a nota técnica DEA 16/11 do MME o número de ligações à rede elétrica depende de variáveis demográficas, como população, número de domicílios e o número de habitantes por domicílio. A variável “clientes” representa o crescimento do mercado consumidor em relação ao intervalo de dados utilizados.

Definiu-se como entrada do sistema a variação percentual da quantidade de consumidores dentre dois dias consecutivos. O dado representa o crescimento do mercado consumidor em relação ao intervalo de dados utilizados. A fórmula abaixo apresenta essa representação.

$$x_1 = \frac{clientes_t}{clientes_{t-1}} - 1 \quad (30)$$

Onde $clientes_t$ é no instante em questão e $clientes_{t-1}$ é no instante imediatamente anterior.

A variável quantidade de clientes consumidores representa variações de forma horizontal e de forma vertical. O crescimento horizontal representa o atendimento a novas cidades e novos bairros. O crescimento vertical representa ao aumento de carga em lugares já segmentados como mercados. Esses lugares são novas construções em prédios e casas residenciais, por exemplo.

4.1.3 Precipitação de Chuvas

A precipitação de chuvas é o nível de chuva medido em milímetro (mm).

Para a formação da variável, optou-se pela diferença do valor da precipitação diária em relação às normais climatológicas. Os dados de normal climatológica foram extraídos do site do INMET e é uma medição da precipitação entre os anos de 1961 e 1990.

A figura a seguir ilustra todos os dados de precipitação utilizados. Em formato de linha, o dado de precipitação medido em mm. Em formato de área, o dado de precipitação da normal climatológica apresentada acima.

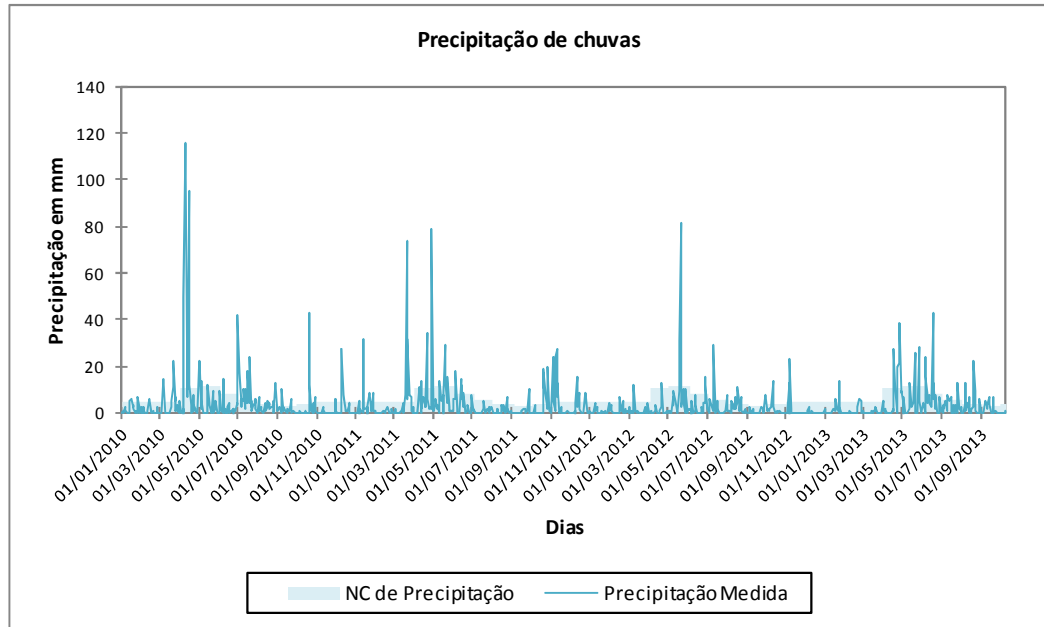


Figura 15 — Precipitação de chuvas.

Para uso nos modelos: fez-se o cálculo da diferença de chuva medida com a informada pela normal climatológica. A fórmula abaixo apresenta essa representação.

$$x_2 = C_{medida} - C_{NC} \quad (31)$$

Onde C_{medida} é a precipitação medida no dia em questão e C_{NC} é a normal climatológica de precipitação do mês em questão de Salvador.

4.1.4 Temperatura Ambiente

A temperatura ambiente em graus Celsius é a terceira variável explicativa utilizada.

As temperaturas utilizadas foram as temperaturas máximas e mínimas diárias. Para a formação da variável, foi usada a diferença do valor da temperatura diária em relação às normais climatológicas.

Assim como a precipitação, os dados da normal climatológica foram extraídos do site do INMET e é uma medição das temperaturas máximas e mínimas diárias entre os anos de 1961 e 1990.

A figura a seguir ilustra todos os dados de temperatura utilizados. Em formato de linha, os dados de temperatura máxima e mínima medidos em graus Celsius. Em formato de área, os dados de temperatura máxima e mínima das normais climatológicas apresentados acima.

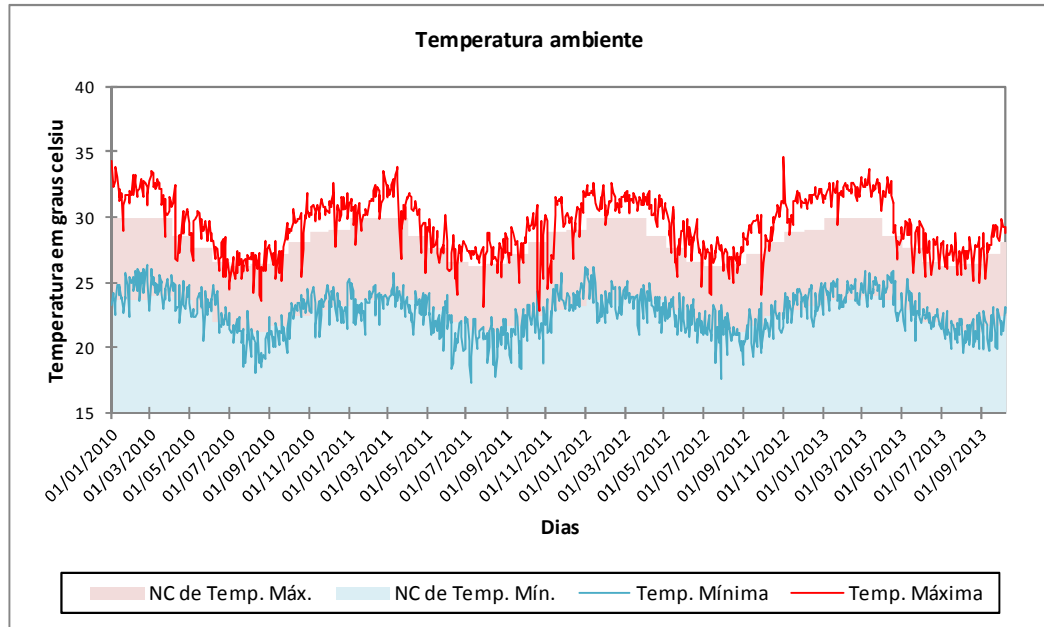


Figura 16 — Temperatura ambiente.

Para uso nos modelos: fez-se o cálculo de diferença entre a temperatura máxima diária medida e a normal climatológica máxima e a diferença entre a temperatura mínima diária medida e a normal climatológica mínima em cada dia. Pois assim, com a soma entra as diferenças das temperaturas máximas e das temperaturas mínimas obteve-se um único valor de sensibilização da temperatura. A fórmula abaixo apresenta essa representação.

$$x_3 = (T_{máx} - NC_{máx}) + (T_{mín} - NC_{mín}) \quad (32)$$

Onde $T_{máx}$ e $T_{mín}$ são as temperaturas máxima e mínima medida no dia em questão e $NC_{máx}$ e $NC_{mín}$ são as normais climatológicas de temperatura máxima e mínima do mês em questão de Salvador.

4.2 Análise exploratória dos dados

A identificação de observações que afetam as estimativas dos parâmetros (observações influentes) é parte importante para validação de modelos. A presença de tais observações pode tornar inadequadas as conclusões com base num ajuste que não incorpore o caráter atípico destas.

Estes tipos de observações podem ser de dois tipos. No primeiro estes podem ser resultantes de erros de medidas ou de registro dos dados. Neste caso, estas são somente dados ruins que prejudicam a estimação dos modelos. Por outro lado, eles podem refletir de maneira verídica a distribuição do processo, exibindo dispersões, *heteroscedasticity*, assimetria,

leptokurtosis no modelo. Portanto, tais observações podem conter informações anormais que é, no entanto, essencial para a estimação exata do modelo.

Entretanto, para se determinar o tipo de observação influente é difícil quando se olha somente para os dados.

Na figura 17 é possível observar que a quantidade de clientes varia de dia para dia. Portanto é justificável utilizar a variação entre dois dias consecutivos (fórmula 27) como variável exógena para os modelos de previsão.

Na figura a seguir também é observável que a quantidade de clientes apresenta um comportamento peculiar junto à observação 578. Embora a energia distribuída, mostrada na figura 14, tenha tido um aumento consistente, a série clientes apresenta uma perda de mais de 300.000 clientes no dia, representado pela amostra 578. Esta observação pode sugerir erros de registro.

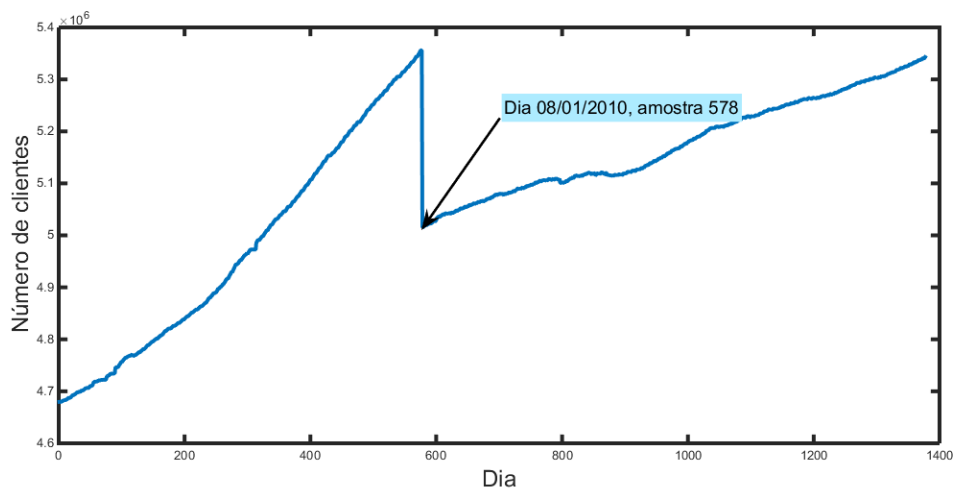


Figura 17 — Número de Clientes.

Uma abordagem para identificação de observações influentes é avaliar o efeito da exclusão de uma observação na estimativa dos parâmetros do modelo. Foram realizados testes diagnósticos baseados em sequências onde em cada instância uma linha dos valores, tanto dos preditores quanto da resposta, foi deletada. As estatísticas, para medir a influência de observações individuais, foram geradas através do cômputo de uma regressão para cada conjunto de dados com uma amostra deletada e foram comparadas com as estatísticas do conjunto de dados completo.

Mudanças significativas nas estimativas dos coeficientes depois de deletarmos uma observação constituem a preocupação principal. Estas diferenças foram avaliadas através das variâncias dos coeficientes individuais (para cada conjunto com uma observação deletada).

A Figura a seguir mostra as mudanças nas estimativas dos coeficientes β .

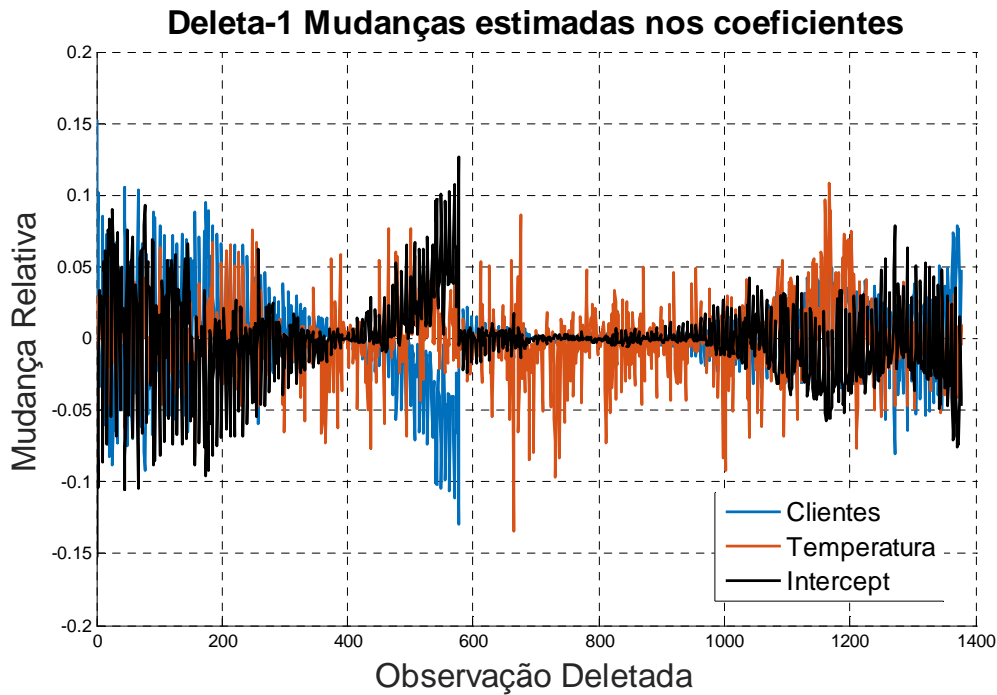


Figura 18 — Mudanças dos coeficientes.

Pode-se observar que as variações antes da observação 578 são consideravelmente altas. A tabela a seguir, com avaliações dos melhores modelos de regressão para os dois casos, mostra que o corte feito nas observações não compromete os resultados.

Tabela 1— Avaliações dos modelos de regressão.

[Redacted Table Content]

Por opção deste trabalho foi realizado um tratamento desta amostra. Substituindo essa amostra pela média dos 5 valores anteriores.

4.3 Modelo de previsão SARIMAX

Esta seção descreve o modelo de previsão SARIMAX com a combinação das variáveis exógenas que melhor representaram o comportamento da série temporal de energia distribuída. Para este fim, foi feita uma análise de série temporal (AST), formação de funções de autocorrelações e autocorrelações parciais, além de testes para a escolha das variáveis que melhor modelaram a série de distribuição de energia. Na AST foi realizada a análise descritiva da série temporal, a transformação da série em uma série estacionária, a análise descritiva para a escolha de cada um dos parâmetros do modelo SARIMAX.

Optou-se pelo método OLS para ser utilizado no SARIMAX. O seu uso se justifica pela análise dos resíduos, pois ele se comprovou suficiente. Para os experimentos propostos, foram utilizadas até três variáveis regressoras.

A execução do algoritmo SARIMAX foi realizada no *software* XLSTAT²⁵. O *software* XLSTAT é próprio para uso de modelos matemáticos em séries temporais.

A figura abaixo ilustra o passo a passo da aplicação do modelo estatístico SARIMAX. Primeiro foi aplicado o teste de homogeneidade para saber se a série é estacionária ou não. Como a série não é estacionária foi necessário transformá-la em uma série estacionária. Com a transformação é possível realizar uma nova análise descritiva para encontrar os parâmetros do modelo. Por fim, cabe realizar os testes com os parâmetros escolhidos para se verificar se é necessário reinterpretá-los.



Figura 19 — Metodologia aplicada no SARIMAX.

Para AST do modelo SARIMAX foram utilizados todos os 983 dados da série temporal. Esses mesmos 983 dados também foram utilizados para os testes.

As seções a seguir descrevem detalhadamente como o modelo foi aplicado até a formação dos parâmetros.

4.3.1 Teste de Estacionariedade.

Primeiramente, é apresentado o gráfico de uma normal que serve como comparação com a curva de energia distribuída.

²⁵O XLSTAT é um software de análise estatística compatível com a plataforma Excel. O XLSTAT é um produto de Addinsoft SARL.

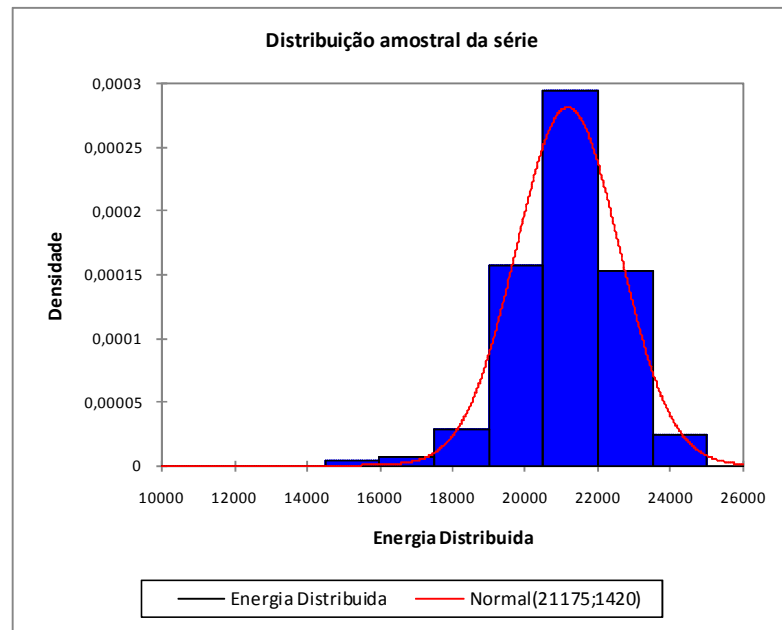


Figura 20 — Distribuição amostral da série.

O histograma da série original revela:

- Uma assimetria da série, pois se observa uma amostra mais populosa em valores baixos.
- A amostra é mais populosa nos valores centrais.
- Acima de certo mínimo, praticamente não se registra a ocorrência de valores. Ou seja, uma capacidade máxima de energia distribuída é observada.

Para a análise descritiva da série proposta indica se uma série é estacionária. Como destacado acima, o teste abaixo confirma que a série temporal em questão possui tendência e sazonalidade, não sendo assim, estacionária.

O teste de Pettitt²⁶ indicou que a série é não estacionária, a ilustração abaixo exhibe o resultado do teste para a série temporal em questão (SALARIJAZI et al., 2012).

²⁶O teste de Pettitt é um teste não paramétrico que não requer nenhuma suposição sobre a distribuição de dados (SALARIJAZI et al., 2012).

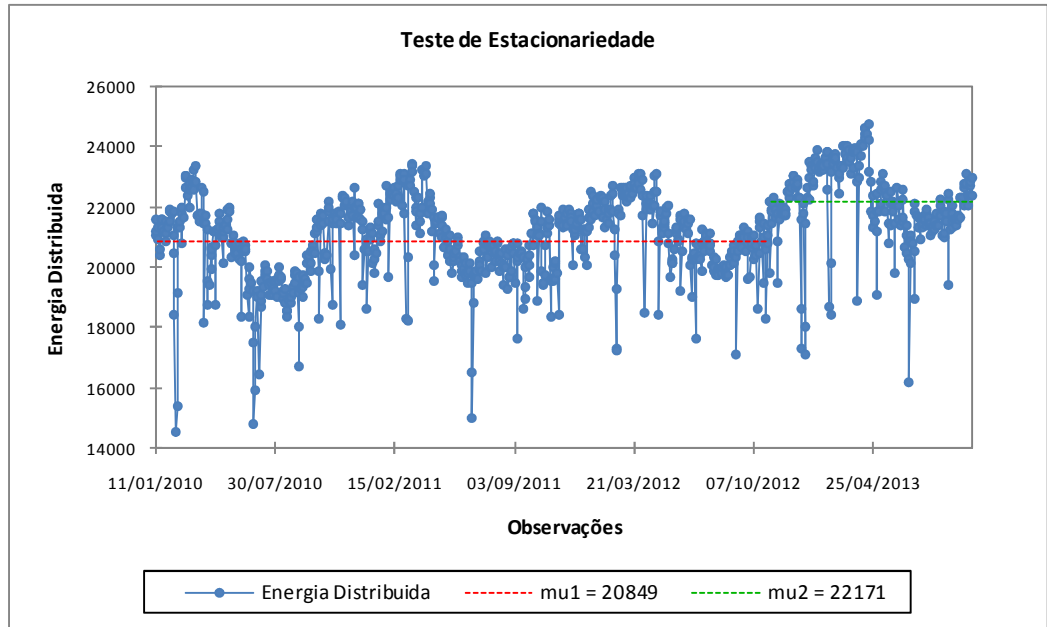


Figura 21 — Teste de estacionariedade da série temporal da região Metropolitana.

Portanto, é necessário aplicar o operador de diferença para transformá-la em estacionária.

4.3.2 Transformação da Série Temporal.

Como foi indicado na seção anterior que a série temporal não é estacionária, é necessário transformá-la para torná-la estacionária. Portanto foi necessário realizar a diferença da série proposta.

Portanto foi feita a primeira diferença da série.

Como na imagem apresentada abaixo a FAC indica uma sazonalidade de 5 períodos, foi feito também a primeira diferença em 5 períodos. Na FAC a seguir foi assumida a hipótese de ruído branco para a série temporal de energia distribuída da região metropolitana da Coelba com o limite inferior e o limite superior de 95%, as duas linhas tracejadas apresentada na imagem a seguir.

É observável que a FAC decresce lentamente, corroborando a constatação de não estacionariedade da série. Portanto é necessário sim diferenciá-la.

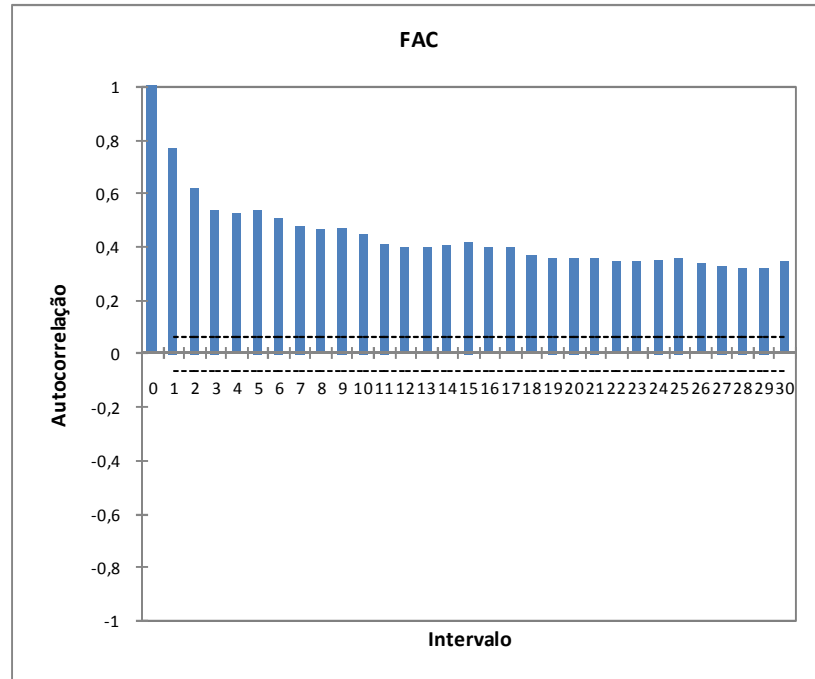


Figura 22 — FAC da região Metropolitana.

A imagem abaixo ilustra a transformação da série temporal em questão aplicando a primeira diferença e a diferença sazonal de 5 períodos. Ou seja, os parâmetros descritos no capítulo 2 são: $d=1$, $D=1$ e $s=5$:

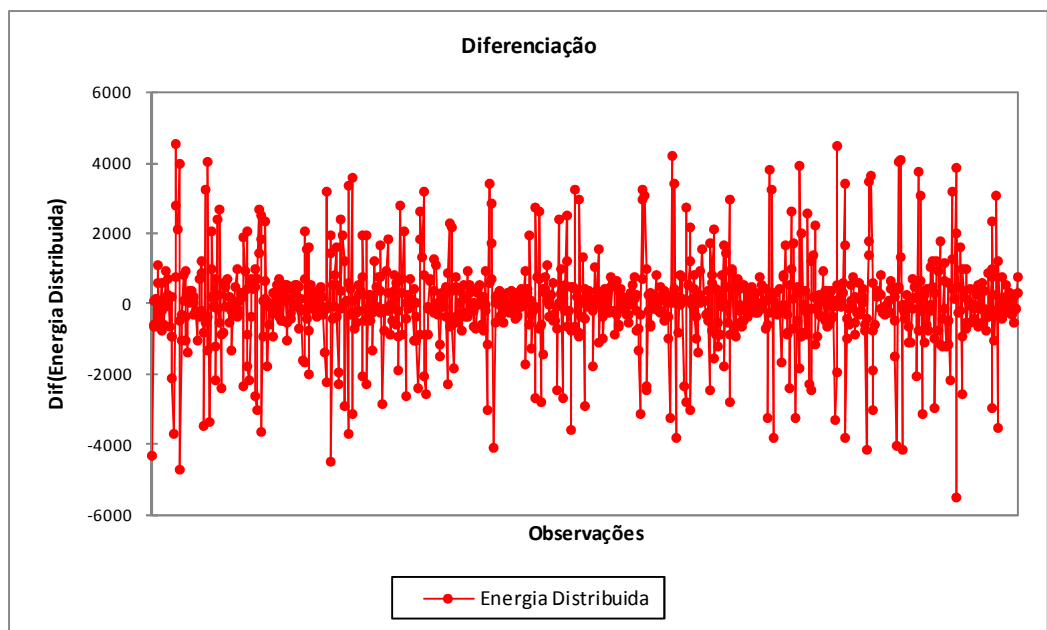


Figura 23 — Transformação da série temporal da região Metropolitana.

Agora se faz necessário aplicar um novo teste para verificar se as séries se tornaram estacionárias.

4.3.3 Teste de Estacionariedade da Série Transformada.

Após a transformação é esperado corrigir a assimetria observada anteriormente. Portanto, é apresentado o gráfico de uma normal com a curva de energia distribuída transformada.

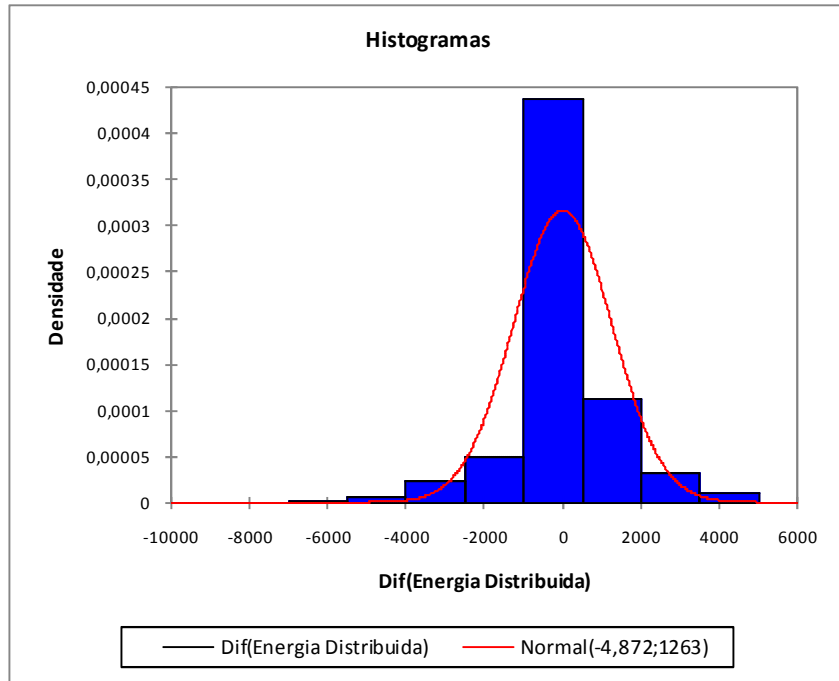


Figura 24 — Distribuição amostral da série transformada.

Através da imagem acima, pode-se considerar uma melhor simetria na série.

O teste de estacionariedade da série temporal transformada indica se com a transformação realizada na seção anterior a série se tornou estacionária.

A ilustração abaixo exibe o resultado do teste série temporal da região Metropolitana.

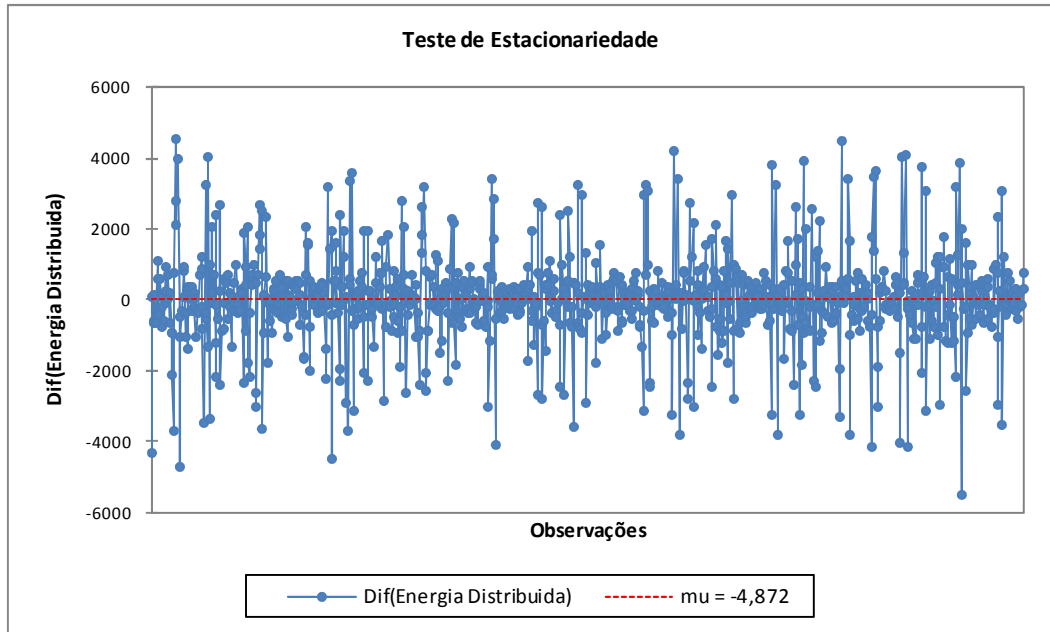


Figura 25 — Teste de homogeneidade da transformada da região Metropolitana.

Pode-se observar que o teste indicou que a série temporal se tornou estacionária. Portanto, foi retirada da série temporal a sazonalidade e a tendência.

Neste momento é possível a análise da série para a busca dos parâmetros do modelo SARIMAX.

4.3.4 Análise Descritiva.

Como descrito no capítulo 2, através do autocorrelograma (FAC) e do autocorrelograma parcial (FACP) os parâmetros descritos no capítulo 2 (p , P , q e Q) são interpretados. As duas figuras apresentadas a seguir ilustram o FAC e o FACP da série temporal de energia distribuída da região metropolitana da Coelba.

Na FAC e na FACP apresentadas a seguir foi assumida a hipótese de ruído branco para a série temporal de energia distribuída da região metropolitana da Coelba com o limite inferior e o limite superior de 95%, as duas linhas tracejadas apresentada na imagem a seguir.

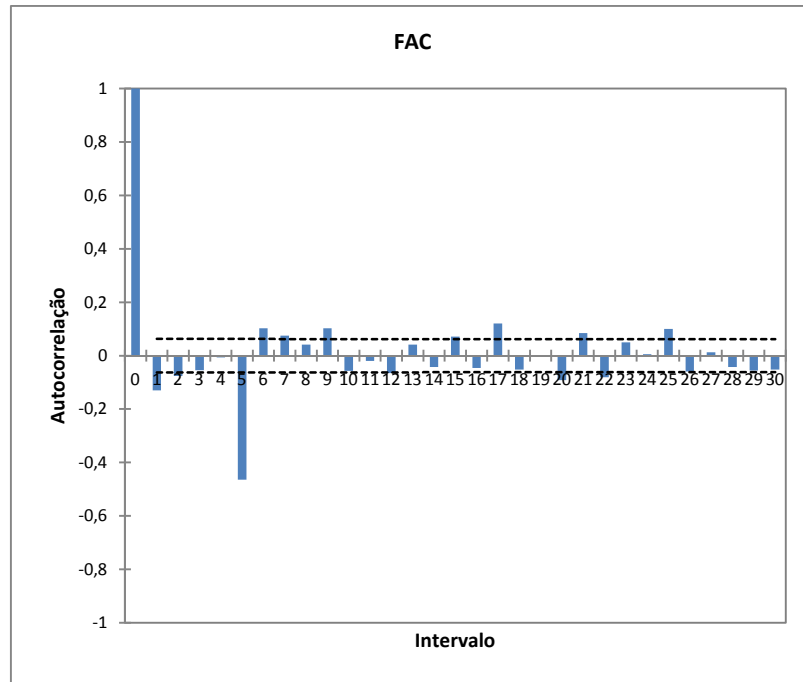


Figura 26 — Autocorrelograma da série da região Metropolitana.

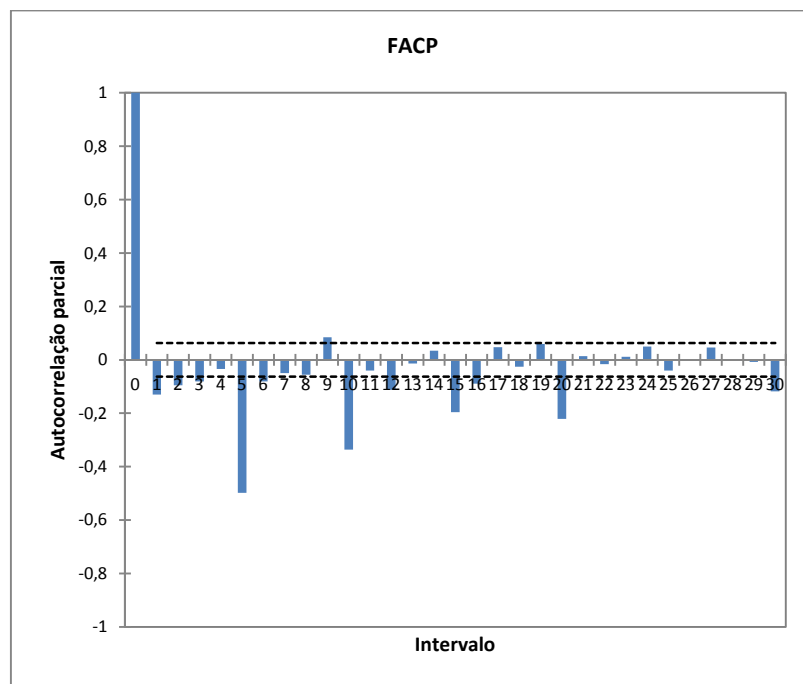


Figura 27 — Autocorrelograma parcial série da região Metropolitana.

Seguindo a descrição da interpretabilidade da FAC e da FACP do capítulo 2. O FAC e o FACP indicam os parâmetros p , P , q e Q .

A figura de autocorrelograma indica claramente uma sazonalidade de 5 períodos.

Observa-se que a FAC apresenta um corte no *lag* 1 e a FACP também. Identificando a ordem $q = 1$ e $p = 1$, ou seja, um MA(1) e um AR(1).

Os *lags* sazonais (5, 10, 15, ...) da FACP são significantes, apresentando um corte significativo no *lag* 1. Em conjunto com os cortes também significantes nos *lags* sazonais da FAC. O que nos dá na parte sazonal um AR(1) e um MA(1).

Portanto, o modelo SARIMAX é apresentado resumidamente na tabela a seguir:

Tabela 2— Ordem do modelo SARIMAX.



É importante interpretar esta formulação matemática: a energia distribuída no dia de hoje é influenciada pela energia distribuída no dia anterior, pela energia distribuída de 5 dias atrás e pela energia de 6 dias atrás.

4.3.5 Testes dos Parâmetros do Modelo SARIMAX.

Esta etapa tem por fundamento verificar se os parâmetros do modelo SARIMAX apresentado na tabela acima são adequados a série ou não. Para a verificação, são analisados os resíduos e os AIC.

O teste 12 (1) apresentou um AIC de 16.238. Os valores dos parâmetros do teste 12 (1) são apresentados na tabela a seguir.

Tabela 3— Parâmetros do teste 12 (1) do modelo SARIMAX.

Parâmetro	Valor	Erro
AR(1)	0,054	0,032
SAR(1)	0,022	0,034
MA(1)	-0,480	0,033
SMA(1)	-0,999	0,001

A imagem abaixo ilustra que o teste de estacionariedade dos resíduos também apresentou resultado satisfatório.

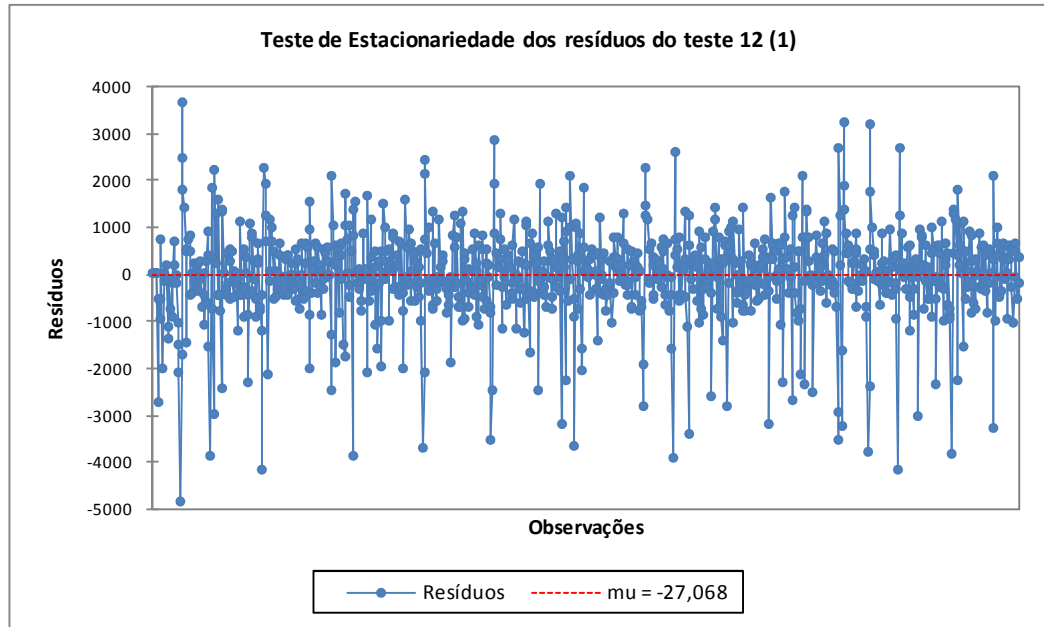


Figura 28 — Teste de estacionariedade dos resíduos do teste 12 (1).

No eixo das ordenadas estão os resíduos e no eixo das abscissas os dias entre os anos de 2010 e 2013.

O teste 13 (1) apresentou um AIC de 16.034. Os valores dos parâmetros do teste 13 (1) são apresentados na tabela a seguir.

Tabela 4— Parâmetros do teste 13 (1) do modelo SARIMAX.

[Conteúdo da tabela oculto]

A imagem abaixo ilustra que o teste de estacionariedade dos resíduos também apresentou resultado satisfatório, ou seja, os resíduos formam um ruído branco.

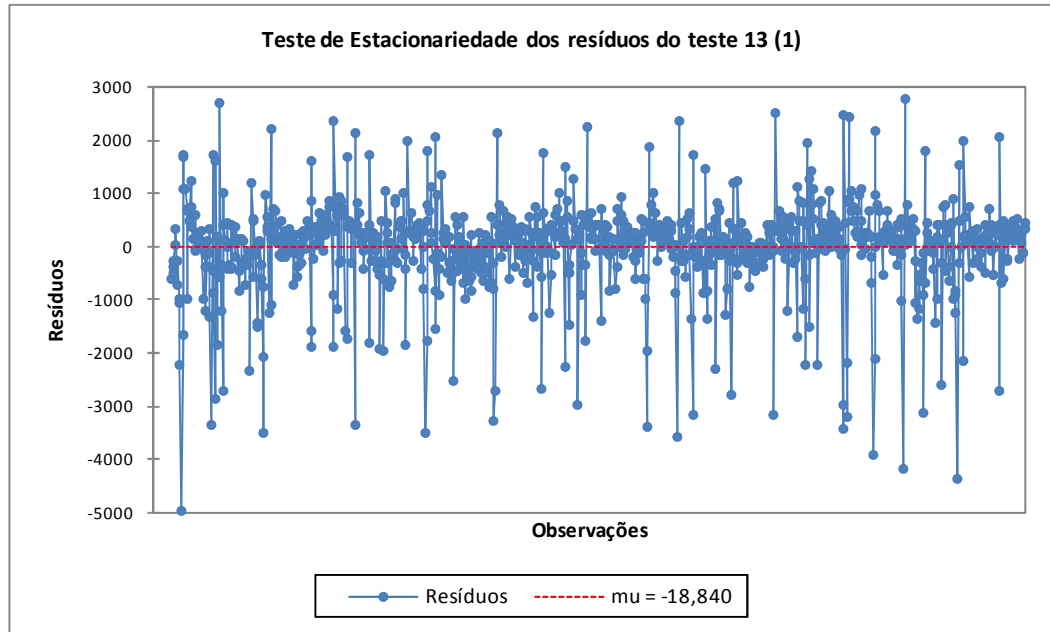


Figura 29 — Teste de estacionariedade dos resíduos do teste 13 (1).

No eixo das ordenadas estão os resíduos e no eixo das abscissas os dias entre os anos de 2010 e 2013.

O teste 23 (1) apresentou um AIC de 16.148. Os valores dos parâmetros do teste 23 (1) são apresentados na tabela a seguir.

Tabela 5— Parâmetros do teste 23 (1) do modelo SARIMAX.

--	--

A imagem abaixo ilustra que o teste de estacionariedade dos resíduos também apresentou resultado satisfatório, ou seja, os resíduos formam um ruído branco.

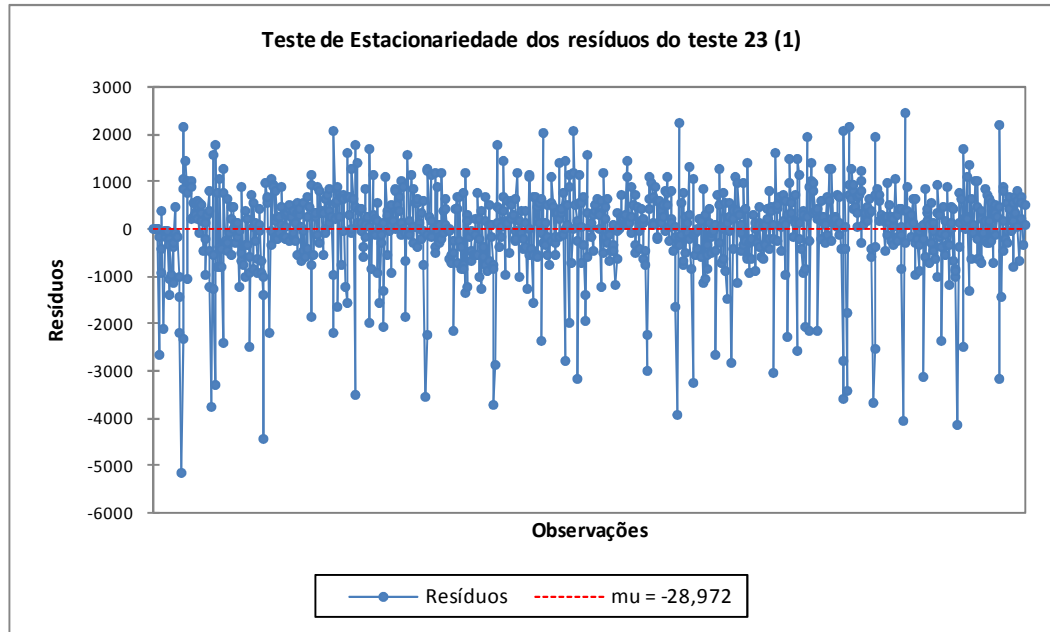


Figura 30 — Teste de estacionariedade dos resíduos do teste 23 (1).

No eixo das ordenadas estão os resíduos e no eixo das abscissas os dias entre os anos de 2010 e 2013.

O teste 123 (1) apresentou um AIC de 16.150. Os valores dos parâmetros do teste 123 (1) são apresentados na tabela a seguir.

Tabela 6— Parâmetros do teste 123 (1) do modelo SARIMAX.

Parâmetro	Valor	Erro
AR(1)	0,491	0,034
SAR(1)	0,081	0,032
MA(1)	-0,937	0,014
SMA(1)	-1,000	0,000

A imagem abaixo ilustra que o teste de estacionariedade dos resíduos também apresentou resultado satisfatório, ou seja, os resíduos formam um ruído branco.

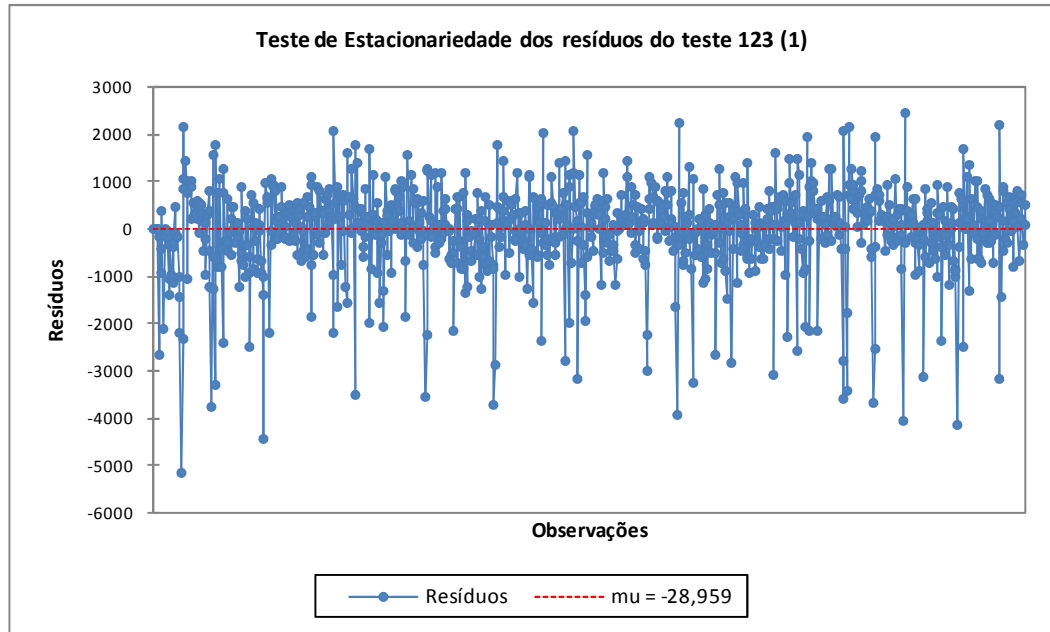


Figura 31 — Teste de estacionariedade dos resíduos do teste 123 (1).

No eixo das ordenadas estão os resíduos e no eixo das abscissas os dias entre os anos de 2010 e 2013.

As tabelas a seguir apresentam novos testes variando os valores do modelo SARIMAX. Esses novos testes asseguram a escolha do modelo SARIMAX.

Tabela 7— Teste de sobrefixação do teste 12 (1).

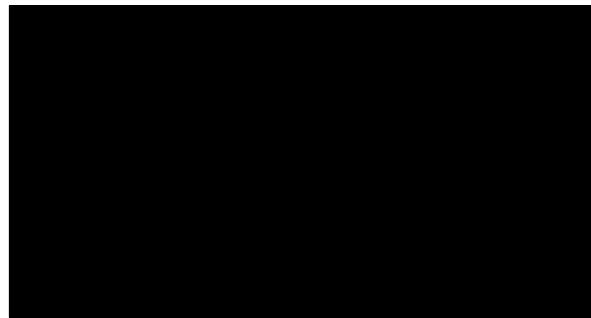
[Conteúdo da tabela não legível]		
----------------------------------	--	--

Tabela 8— Teste de sobrefixação do teste 13 (1).

Teste 13 (1)	MAPE	AIC
(1, 1, 1)(1, 1, 1) 5	2,69%	16.034
(0, 1, 1)(1, 1, 1) 5	2,82%	16.142
(1, 1, 0)(1, 1, 1) 5	2,83%	16.167
(1, 1, 1)(0, 1, 1) 5	2,72%	16.043
(1, 1, 1)(1, 1, 0) 5	3,38%	16.043
(0, 1, 1)(1, 1, 0) 5	3,74%	16.505
(1, 1, 0)(0, 1, 1) 5	2,82%	16.168

Tabela 9— Teste de sobrefixação do teste 23 (1).

Teste 23 (1)	MAPE	AIC
(1, 1, 1)(1, 1, 1) 5	3,07%	16.148
(0, 1, 1)(1, 1, 1) 5	3,20%	16.243
(1, 1, 0)(1, 1, 1) 5	3,31%	16.297
(1, 1, 1)(0, 1, 1) 5	3,08%	16.152
(1, 1, 1)(1, 1, 0) 5	3,74%	16.152
(0, 1, 1)(1, 1, 0) 5	4,14%	16.603
(1, 1, 0)(0, 1, 1) 5	3,32%	16.298

Tabela 10— Teste de sobrefixação do teste 123 (1).


Pode-se comprovar que o menor MAPE e o menor AIC foram dos modelos definidos anteriormente. Todos esses modelos estão marcados de negrito na primeira linha das tabelas apresentadas.

Enfim, a seção destinada aos experimentos do modelo SARIMAX atingiu um dos objetivos primordiais desta tese. O próximo passo da metodologia consiste em realizar os testes no modelo de previsão do Sistema de Inferência Fuzzy.

4.4 Modelo de previsão do Sistema de Inferência Fuzzy

Esta seção descreve o modelo de previsão do Sistema de Inferência Fuzzy. O modelo proposto é um Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS). Este modelo proposto foi executado no *software* MATLAB com o uso do *toolbox* ANFIS Editor para a formação *Sugeno Fuzzy System*. O Sistema proposto utiliza o método SUGENO com saída constante.

O ANFIS mapeia os dados de entrada por meio de funções de pertinências e seus parâmetros associados, e, em seguida, através das funções de pertinências de saída e seus parâmetros, é possível obter a interpretabilidade da entrada com a saída do sistema.

Os parâmetros associados às funções de pertinência mudam através do processo de ajuste que é composto por um determinado número de épocas e um método para minimizar os

erros. Utilizou-se 30 como a quantidade de épocas para o ajuste do modelo, que se ajusta através de um método híbrido. O método híbrido em questão é uma combinação do método dos mínimos quadrados e o método de gradiente descendente *backpropagation*.

Para o ajuste do ANFIS foram utilizados como *input* do modelo todos os 983 dados da série temporal. Assim como no modelo SARIMAX, esses mesmos 983 dados também foram utilizados para os testes. Pois no ANFIS é opcional utilizar a validação do modelo.

Definiram-se quatro como o número de funções de pertinência para cada entrada do sistema. Todas as quatro funções de pertinência devem ser iguais por ser uma limitação do *ANFISeditor* do MATLAB.

Para fins de comparação, como foi informado no SARIMAX, foi definido como entrada do Sistema de Inferência Fuzzy a energia distribuída do dia anterior e a de cinco dias anteriores. No modelo SARIMAX, a energia distribuída do dia anterior foi informada pela primeira diferenciação e a energia distribuída de cinco dias anteriores foi informada pela sazonalidade de 5 dias, uma semana. Portanto, além das variáveis exógenas foi utilizada como entrada no modelo essas duas séries de energia distribuída.

Para a escolha dos parâmetros do Sistema de Inferência Fuzzy foi utilizado testes com diferentes combinações dos tipos e de quantidade de funções de pertinência. Ao testar as combinações buscou-se um menor erro. Como só é possível o uso das funções de pertinência fornecidas, foi realizado testes com as funções de pertinência: *gbellmf*, *gaussmf*, *gauss2mf*, *pimf* e *dsigmf*.

As tabelas a seguir ilustram os testes para a escolha da função de pertinência. Destacado em negrito está a função de pertinência que apresentou o menor erro e, portanto a que foi elegida como parâmetro do modelo.

Tabela 11— Teste da função de pertinência do teste 12 (1).

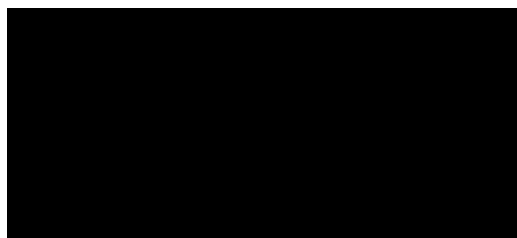


Tabela 12— Teste da função de pertinência do teste 13 (1).



Tabela 13— Teste da função de pertinência do teste 23 (1).

Teste 23 (1)	MAPE	DP
gbellmf	2,33%	745
gaussmf	2,23%	727
gauss2mf	2,86%	852
pimf	3,11%	912
dsigmf	2,51%	819

Tabela 14— Teste da função de pertinência do teste 123 (1).

Teste 123 (1)	MAPE	DP
gbellmf	2,07%	681
gaussmf	1,99%	662
gauss2mf	2,63%	790
pimf	2,94%	867
dsigmf	2,32%	730

A tabela abaixo resume as informações especificadas no Sistema de Inferência Fuzzy.

Tabela 15— Parâmetros do Sistema de Inferência Fuzzy.

Entrada:	Tipo:	Funções:	Épocas:
2	gaussmf	4	30
3	gaussmf	4	30

A primeira coluna, “Entrada”, descreve somente a quantidade de variáveis exógenas utilizadas nos testes. A segunda coluna, o tipo de função de pertinência utilizada em cada respectivo teste, no caso a *gaussmf*. A terceira coluna, “Funções”, demonstra a quantidade de funções de pertinência para cada entrada no respectivo teste, escolhida como 4 desde o início. A última coluna destaca a quantidade de épocas utilizada nos testes. Em todos os testes utilizou-se a quantidade de 30 épocas para o treinamento do algoritmo.

Para fins de comparação com o modelo SARIMAX, foram utilizadas as mesmas combinações de variáveis exógenas para a modelagem dos testes do Sistema Fuzzy. Portanto, o Sistema de Inferência Fuzzy atingiu outro objetivo primordial desta tese. Neste momento, cabe a este capítulo apresentar a forma de avaliação ambos os modelos.

4.5 Avaliação

Para avaliar os dois algoritmos apresentados na seção 4.3 e 4.4 foram utilizados o erro percentual médio e o desvio padrão entre as séries originais e as séries previstas por cada modelo estatístico. Ambos avaliadores são calculados somente com uma casa decimal de precisão.

As duas seções a seguir, 4.5.1 e 4.5.2 apresenta cada medida de erro utilizada.

4.5.1 Erro Médio Percentual

O erro percentual médio, mais usualmente em inglês *Mean Absolute Percentage Error* (MAPE), é um parâmetro largamente utilizado em outros trabalhos acadêmicos e mede, em termos absolutos, o desvio da série original com a série saída do modelo. Nele é possível medir a consistência de um modelo.

A fórmula do MAPE é destacada abaixo:

$$M = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - Y_t}{A_t} \right| \quad (33)$$

Onde A_t é o valor original e Y_t é o valor previsto por cada modelo.

4.5.2 Desvio Padrão.

O desvio padrão (DP) medido é sobre a diferença entre as séries originais e as séries previstas pelo modelo. Ele pode ser escrito através da fórmula abaixo:

$$DP = \sqrt{\frac{\sum (A_t - \bar{Y}_t)^2}{(n-1)}} \quad (34)$$

Onde \bar{Y}_t é a média da amostra e n é o tamanho da amostra, no caso o valor de n é de 983.

A partir de ambos os avaliadores apresentados nesta seção é possível definir o modelo mais preciso para a previsão da série temporal de energia elétrica da região Metropolitana do Estado da Bahia.

5 RESULTADOS OBTIDOS POR MEIO DOS MODELOS

Cabe a este capítulo apresentar os resultados e as diferenças entre os testes obtidos utilizando as diferentes ferramentas descritas no capítulo 4. Ao destacar as diferenças entre as performances há considerações descritas sobre os resultados obtidos.

Este capítulo inicia com a apresentação dos resultados do modelo SARIMAX seguido da apresentação dos resultados do Sistema Fuzzy. A última seção destina-se ao comparativo entre ambos os modelos. Convém registrar mais uma vez, que os dois modelos de previsão se propõem a realizar a previsão de um passo a frente.

Nas duas seções a seguir são apresentados separadamente cada teste com a combinação de duas variáveis exógenas e um teste com a combinação de três variáveis exógenas. Em seus detalhamentos são apresentados gráficos e tabelas.

5.1 Resultados do modelo SARIMAX

Esta seção apresenta os resultados obtidos com o modelo SARIMAX. Inicialmente são apresentados em uma mesma tabela os resultados de previsão do modelo com diferentes combinações das variáveis exógenas.

Os testes indicados na tabela abaixo apresentam o MAPE e o desvio padrão dos resíduos (DP) para cada combinação de variáveis indicadas nas seções anteriores. Há destacado em negrito o teste que apresentou o melhor resultado.

Tabela 16— Testes de variáveis do modelo SARIMAX.

Código dos Testes:	Variáveis Exógenas			SARIMAX	
	Clientes	Temp	Prec	MAPE	DP
teste 12 (1)	X	X		3,20%	964
teste 13 (1)	X		X	2,69%	867
teste 23 (1)		X	X	3,07%	919
teste 123 (1)	X	X	X	3,08%	919

O resultado do teste 13 (1) indicado em negrito na tabela acima destaca a melhor combinação dos testes. Ou seja, a combinação que obteve o menor MAPE e o menor DP. O teste 13 (1) é feito a partir da combinação da variável exógena quantidade de clientes e de precipitação de chuvas. O MAPE do teste 13 (1) é de 2,69% e o DP é de 867.

Todos os quatro testes destacados na tabela acima são detalhados nas seções a seguir.

5.1.1 Resultados do modelo SARIMAX no teste 12 (1)

Esta seção descreve e ilustra os resultados do teste 12 (1) mais detalhadamente. A figura abaixo ilustra a série temporal original, em preto, e a série temporal prevista pelo modelo SARIMAX, em vermelho.

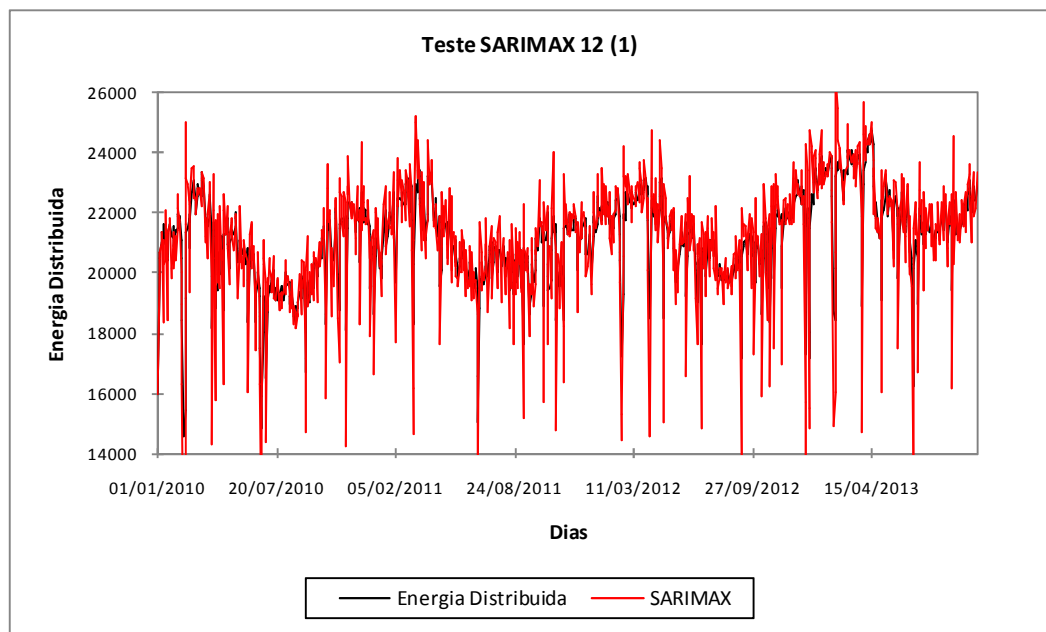


Figura 32 — Teste SARIMAX 12 (1).

A figura abaixo ilustra o resíduo do teste 12 (1). O resíduo consiste na diferença entre o valor previsto no modelo e o valor a ser previsto pelo modelo, ou seja, da série temporal da região Metropolitana da Coelba.

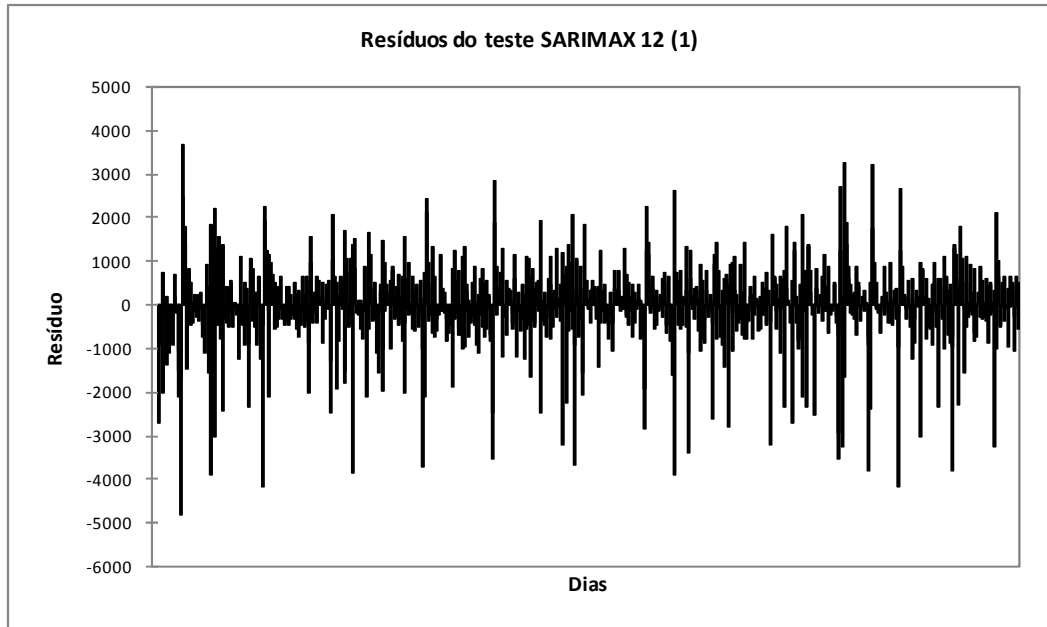


Figura 33 — Resíduos do teste SARIMAX 12 (1).

A tabela apresentada a seguir destaca os parâmetros das variáveis exógenas do modelo SARIMAX no teste 12 (1).

Tabela 17— Parâmetros das variáveis do teste SARIMAX 12 (1).

[Conteúdo da tabela redigido]	
-------------------------------	--

Como destacado acima, a tabela abaixo destaca os erros percentuais do modelo SARIMAX na série original, sendo os erros: médio percentual e o desvio padrão (DP) dos resíduos.

Tabela 18— Erros percentuais do teste SARIMAX 12 (1).

	MAPE	DP
Teste 12 (1)	3,20%	964

O modelo estatístico SARIMAX obteve um erro médio percentual de 3,20%. E o desvio padrão dos resíduos de 964.

5.1.2 Resultados do modelo SARIMAX no teste 13 (1)

Esta seção descreve e ilustra os resultados do teste 13 (1) mais detalhadamente. A figura abaixo ilustra a série temporal original, em preto, e a série temporal prevista pelo modelo SARIMAX, em vermelho.

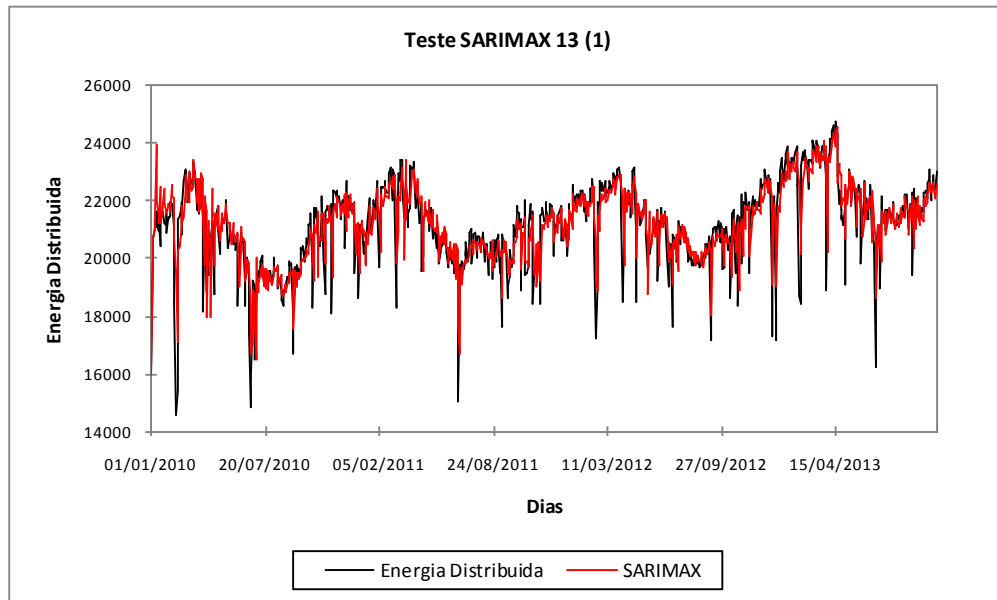


Figura 34 — Teste SARIMAX 13 (1).

A figura abaixo ilustra o resíduo do teste 13 (1). O resíduo consiste na diferença entre o valor previsto no modelo e o valor a ser previsto pelo modelo, ou seja, da série temporal da região Metropolitana da Coelba.

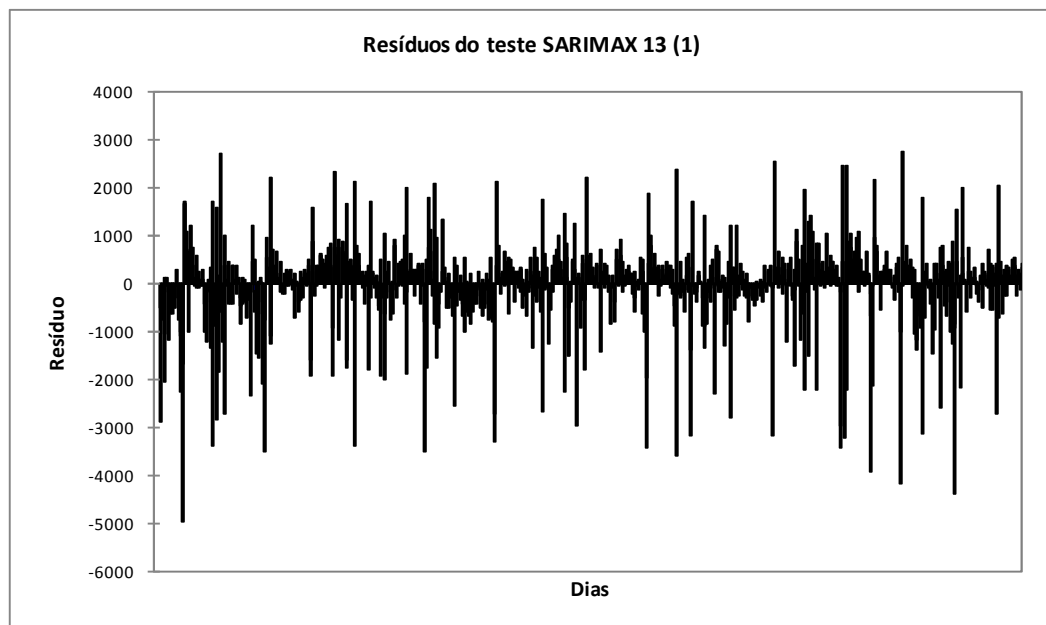


Figura 35 — Resíduos do teste SARIMAX 13 (1).

A tabela apresentada a seguir destaca os parâmetros das variáveis exógenas do modelo SARIMAX no teste 13 (1).

Tabela 19— Parâmetros das variáveis do teste SARIMAX 13 (1).

Parâmetro	Valor	Erro
Constante	21.093,33	
Clientes (t-1)	3278,408	0,000
Precip (t-1)	28,057	3,747

Como destacado acima, a tabela abaixo destaca os erros percentuais do modelo SARIMAX na série original, sendo os erros: médio percentual e o desvio padrão (DP) dos resíduos.

Tabela 20— Erros percentuais do teste SARIMAX 13 (1).

	MAPE	DP
Teste 13 (1)	2,69%	864

O modelo estatístico SARIMAX obteve um erro médio percentual de 2,69%. E o desvio padrão dos resíduos de 864.

5.1.3 Resultados do modelo SARIMAX no teste 23 (1)

Esta seção descreve e ilustra os resultados do teste 23 (1) mais detalhadamente. A figura abaixo ilustra a série temporal original, em preto, e a série temporal prevista pelo modelo SARIMAX, em vermelho.

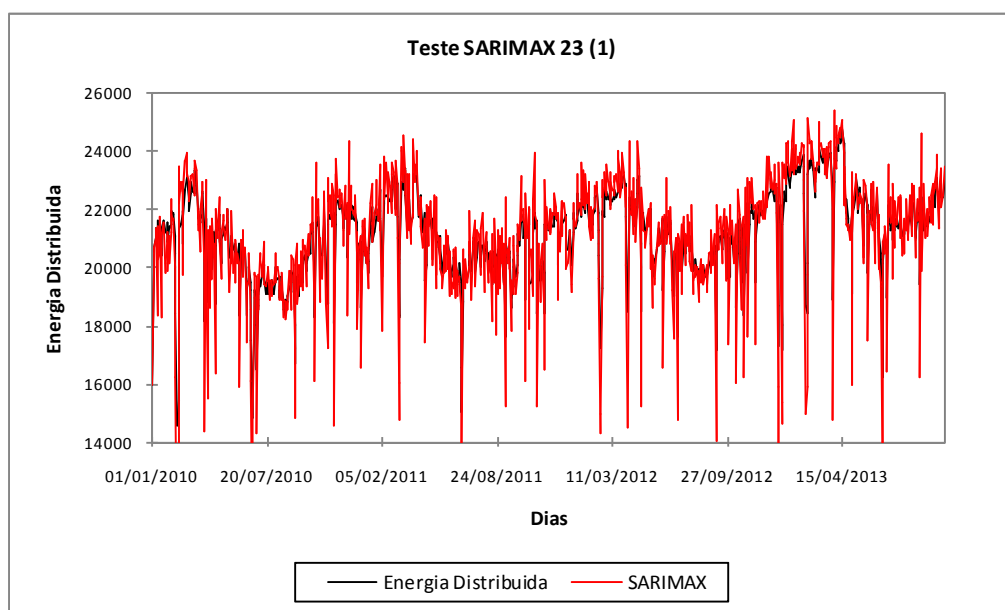


Figura 36 — Teste SARIMAX 23 (1).

A figura abaixo ilustra o resíduo do teste 23 (1). O resíduo consiste na diferença entre o valor previsto no modelo e o valor a ser previsto pelo modelo, ou seja, da série temporal da região Metropolitana da Coelba.

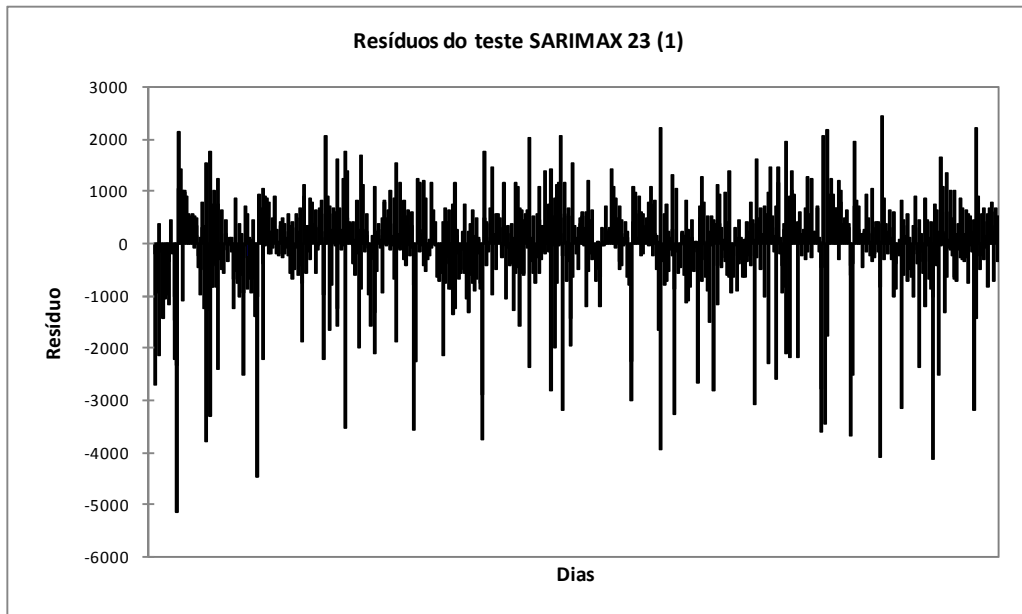


Figura 37 — Resíduos do teste SARIMAX 23 (1).

A tabela apresentada a seguir destaca os parâmetros das variáveis exógenas do modelo SARIMAX no teste 23 (1).

Tabela 21— Parâmetros das variáveis do teste SARIMAX 23 (1).

Parâmetro	Valor	Erro
Constante	20.840,38	
Temperatura (t-1)	287,472	28,513
Precip (t-1)	-3,616	5,822

Como destacado acima, a tabela abaixo destaca os erros percentuais do modelo SARIMAX na série original, sendo os erros: médio percentual e o desvio padrão (DP) dos resíduos.

Tabela 22— Erros percentuais do teste SARIMAX 23 (1).

	MAPE	DP
Teste 23 (1)	3,07%	919

O modelo estatístico SARIMAX obteve um erro médio percentual de 3,07%. E o desvio padrão dos resíduos de 919.

5.1.4 Resultados do modelo SARIMAX no teste 123 (1)

Esta seção descreve e ilustra os resultados do teste 123 (1) mais detalhadamente. A figura abaixo ilustra a série temporal original, em preto, e a série temporal prevista pelo modelo SARIMAX, em vermelho.

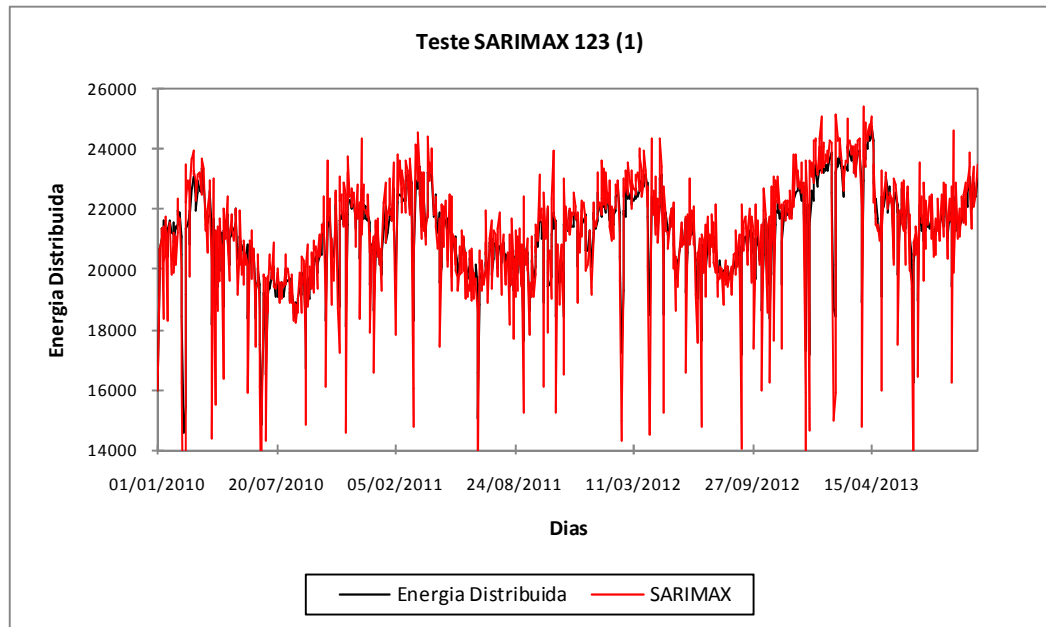


Figura 38 — Teste SARIMAX 123 (1).

A figura abaixo ilustra o resíduo do teste 123 (1). O resíduo consiste na diferença entre o valor previsto no modelo e o valor a ser previsto pelo modelo, ou seja, da série temporal da região Metropolitana da Coelba.

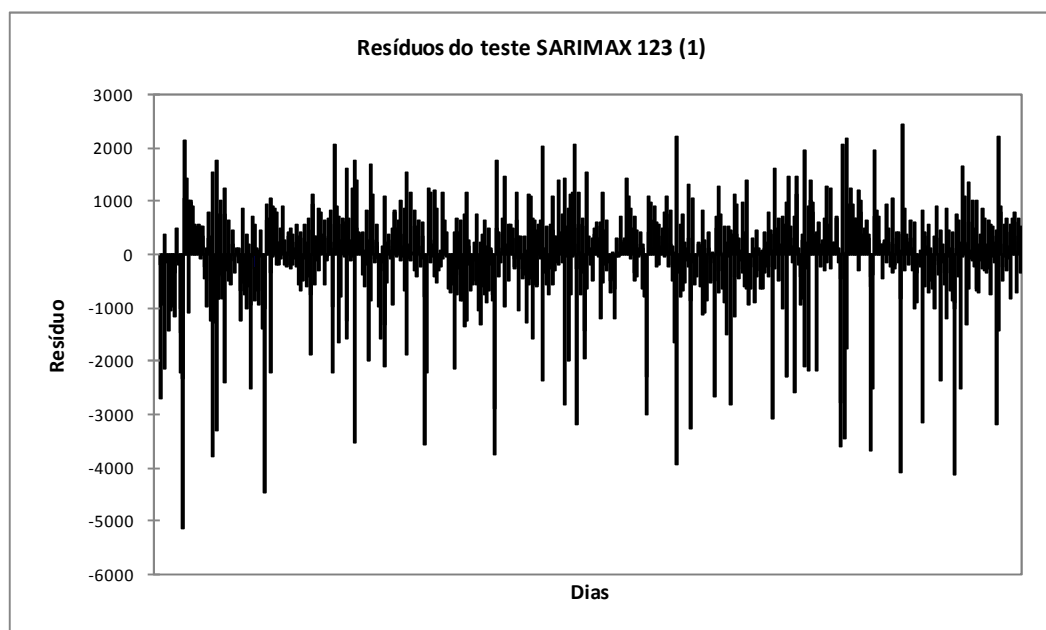


Figura 39 — Resíduos do teste SARIMAX 123 (1).

A tabela apresentada a seguir destaca os parâmetros das variáveis exógenas do modelo SARIMAX no teste 123 (1).

Tabela 23— Parâmetros das variáveis do teste SARIMAX 123 (1).

Parâmetro	Valor	Erro
Constante	20.840,01	
Clientes (t-1)	5.697,09	0,00
Temperatura (t-1)	287,533	28,514
Precip (t-1)	-3,631	5,822

Como destacado acima, a tabela abaixo destaca os erros percentuais do modelo SARIMAX na série original, sendo os erros: médio percentual e o desvio padrão (DP) dos resíduos.

Tabela 24— Erros percentuais do teste SARIMAX 123 (1).

	MAPE	DP
Teste 123 (1)	3,08%	919

O modelo estatístico SARIMAX obteve um erro médio percentual de 3,08%. E o desvio padrão dos resíduos de 919.

5.2 Resultados do Sistema de Inferência Fuzzy

Esta seção apresenta os resultados obtidos com o Sistema Fuzzy. Assim como na seção 5.1, são apresentados em uma mesma tabela os resultados de previsão do modelo com diferentes combinações das variáveis exógenas.

Os testes indicados na tabela abaixo apresentam o MAPE e o desvio padrão dos resíduos (DP) para cada combinação de variáveis indicadas nas seções anteriores. Há destacado em negrito o teste que apresentou o melhor resultado.

Tabela 25— Testes de variáveis do Sistema de Inferência Fuzzy.

Código dos Testes:	Variáveis Exógenas			Sistema de Inferência Fuzzy	
	Clientes	Temp.	Prec.	MAPE	DP
teste 12 (1)	X	X		2,18%	705
teste 13 (1)	X		X	2,31%	762
teste 23 (1)		X	X	2,23%	727
teste 123 (1)	X	X	X	1,99%	662

O resultado do teste 123 (1) indicado em negrito na tabela acima destaca a melhor combinação dos testes. Ou seja, a combinação que obteve o menor MAPE e o menor DP. O teste 123 (1) é feito a partir da combinação de todas as três variáveis exógenas: quantidade de clientes, temperatura ambiente e precipitação de chuvas. O MAPE do teste 123 (1) é de 1,99% e o DP é de 662.

Todos os quatro testes destacados na tabela acima são detalhados nas seções a seguir.

5.2.1 Resultados do Sistema de Inferência Fuzzy no teste 12 (1).

Esta seção descreve e ilustra os resultados do teste 12 (1) mais detalhadamente. A figura a seguir ilustra a série temporal original, em preto, e a série temporal prevista pelo Sistema de Inferência Fuzzy, em vermelho.

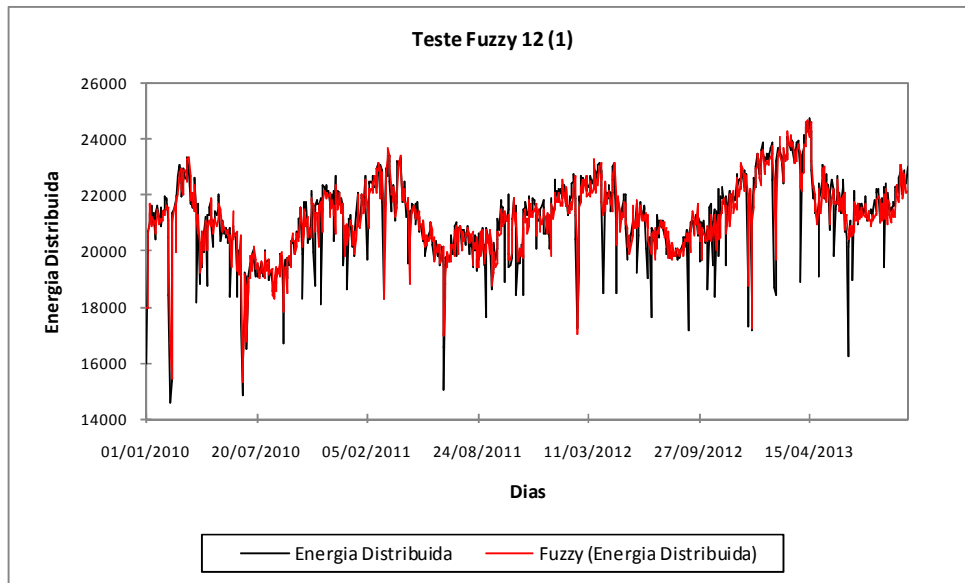


Figura 40 — Teste 12 (1) do Sistema Fuzzy.

A figura abaixo ilustra o resíduo do teste 12 (1). O resíduo consiste na diferença entre o valor previsto no modelo e o valor a ser previsto pelo modelo, ou seja, da série temporal da região Metropolitana da Coelba.

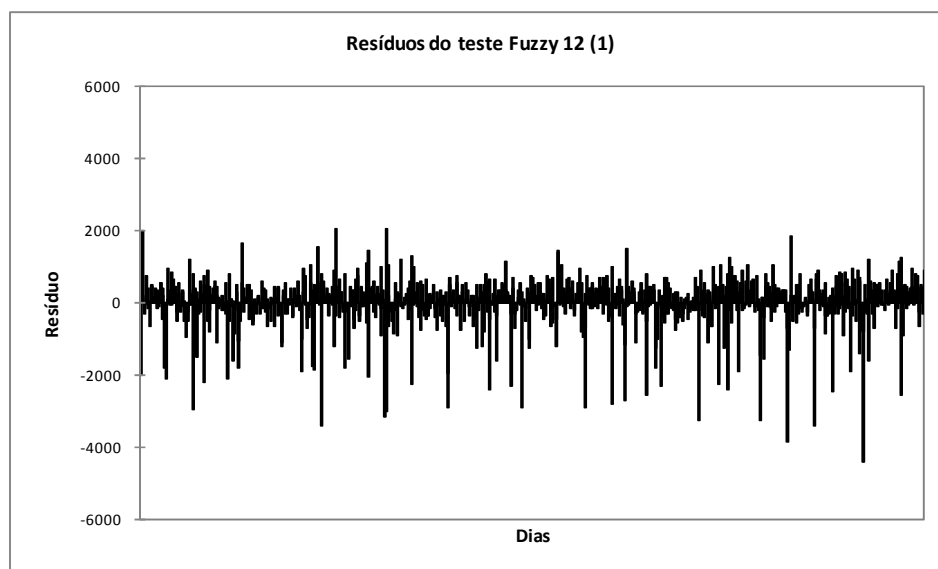


Figura 41 — Resíduos do teste do Sistema Fuzzy 12 (1).

A tabela apresentada a seguir destaca os parâmetros das variáveis exógenas do Sistema de Inferência Fuzzy no teste 12 (1).

Tabela 26— Parâmetros das variáveis do teste 12 (1) do Sistema Fuzzy



Como destacado na seção 5.2, a tabela abaixo destaca os erros percentuais do modelo Sistema de Inferência na série original, sendo os erros: médio percentual e o desvio padrão (DP) dos resíduos.

Tabela 27— Erros percentuais do teste 12 (1) do Sistema Fuzzy

	MAPE	DP
Teste 12 (1)	2,18%	705

O Sistema de Inferência Fuzzy obteve um erro médio percentual de 2,18%. E o desvio padrão dos resíduos de 705.

5.2.2 Resultados do Sistema de Inferência Fuzzy no teste 13 (1).

Esta seção descreve e ilustra os resultados do teste 13 (1) mais detalhadamente. A figura a seguir ilustra a série temporal original, em preto, e a série temporal prevista pelo Sistema de Inferência Fuzzy, em vermelho.

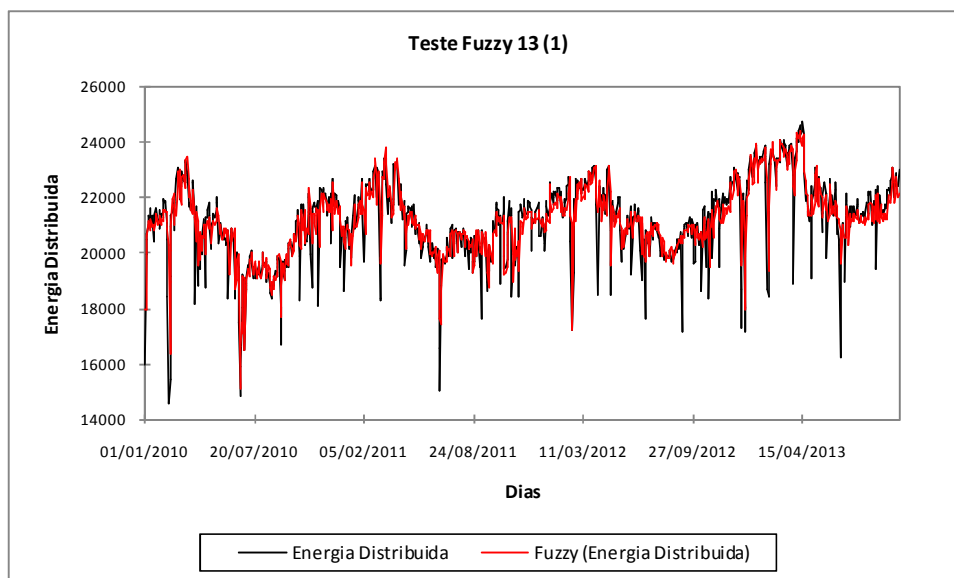


Figura 42 — Teste 13 (1) do Sistema Fuzzy.

A figura abaixo ilustra o resíduo do teste 13 (1). O resíduo consiste na diferença entre o valor previsto no modelo e o valor a ser previsto pelo modelo, ou seja, da série temporal da região Metropolitana da Coelba.

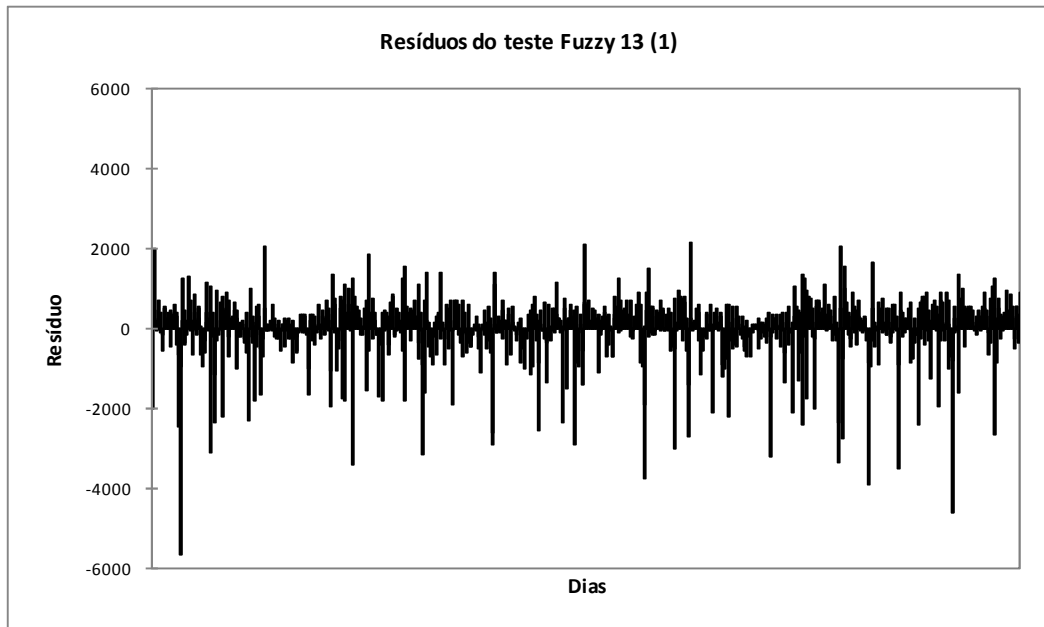


Figura 43 — Resíduos do teste do Sistema Fuzzy 13 (1).

A tabela apresentada a seguir destaca os parâmetros das variáveis exógenas do Sistema de Inferência Fuzzy no teste 13 (1).

Tabela 28— Parâmetros das variáveis do teste 13 (1) do Sistema Fuzzy

[Tabela 28 oculta]	
--------------------	--

Como destacado na seção 5.2, a tabela abaixo destaca os erros percentuais do Sistema de Inferência Fuzzy na série original, sendo os erros: médio percentual e o desvio padrão (DP) dos resíduos.

Tabela 29— Erros percentuais do teste 13 (1) do Sistema Fuzzy

	MAPE	DP
Teste 13 (1)	2,31%	762

O Sistema de Inferência Fuzzy obteve um erro médio percentual de 2,31%. E o desvio padrão dos resíduos de 762.

5.2.3 Resultados do Sistema de Inferência Fuzzy no teste 23 (1).

Esta seção descreve e ilustra os resultados do teste 23 (1) mais detalhadamente. A figura abaixo ilustra a diferença entre a série temporal original, em preto, com a prevista pelo Sistema de Inferência Fuzzy, em vermelho.

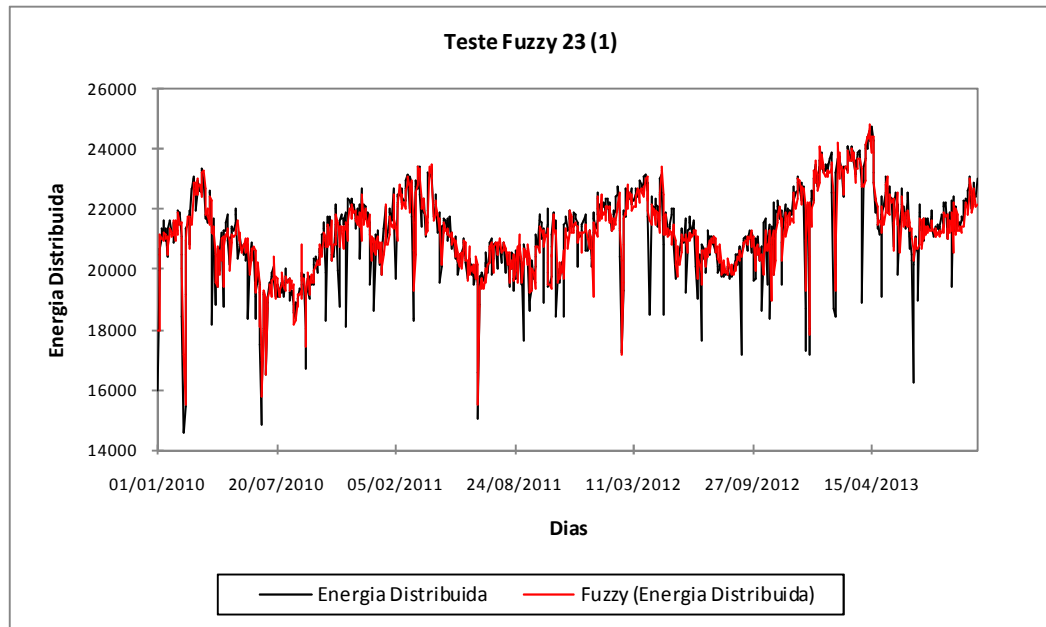


Figura 44 — Teste 23 (1) do Sistema Fuzzy.

A figura abaixo ilustra o resíduo do teste 23 (1). O resíduo consiste na diferença entre o valor previsto no modelo e o valor a ser previsto pelo modelo, ou seja, da série temporal da região Metropolitana da Coelba.

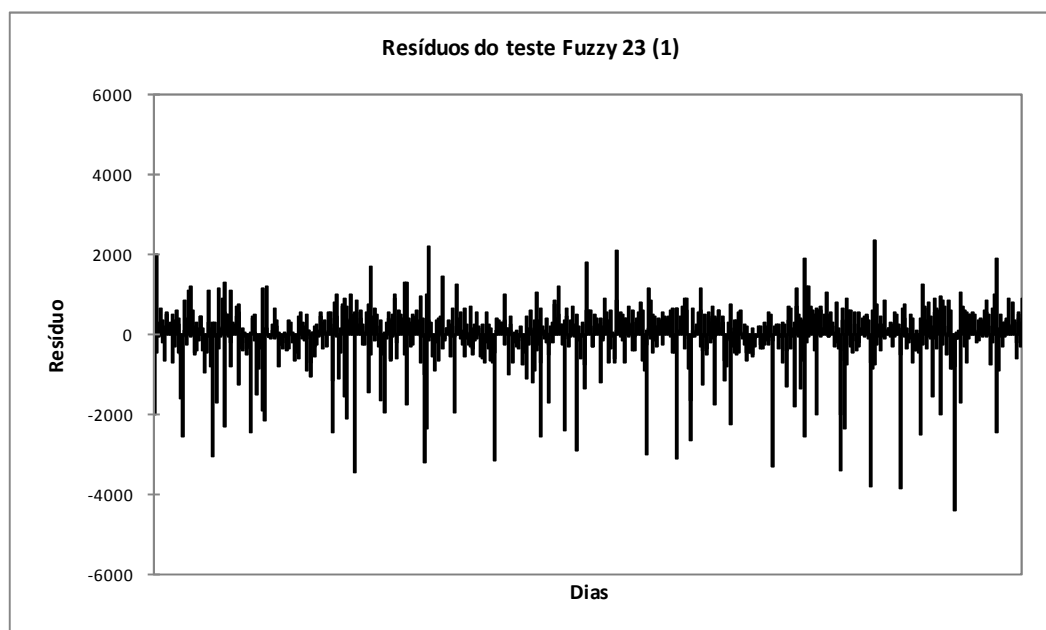


Figura 45 — Resíduos do teste do Sistema Fuzzy 23 (1).

A tabela apresentada a seguir destaca os parâmetros das variáveis exógenas do Sistema de Inferência Fuzzy no teste 23 (1).

Tabela 30— Parâmetros das variáveis do teste 23 (1) do Sistema Fuzzy

Teste 23 (1):	Entrada:
Tipo	gausmf
Quantidade	4
Épocas	30
Regras	256

Como destacado na seção 5.2, a tabela abaixo destaca os erros percentuais do Sistema de Inferência Fuzzy na série original, sendo os erros: médio percentual e o desvio padrão (DP) dos resíduos.

Tabela 31— Erros percentuais do teste 23 (1) do Sistema Fuzzy

	MAPE	DP
Teste 23 (1)	2,23%	727

O Sistema de Inferência Fuzzy obteve um erro médio percentual de 2,23%. E o desvio padrão dos resíduos de 727.

5.2.4 Resultados do Sistema de Inferência Fuzzy no teste 123 (1).

Esta seção descreve e ilustra os resultados do teste 123 (1) mais detalhadamente. A figura a seguir ilustra a série temporal original, em preto, e a série temporal prevista pelo Sistema de Inferência Fuzzy, em vermelho.

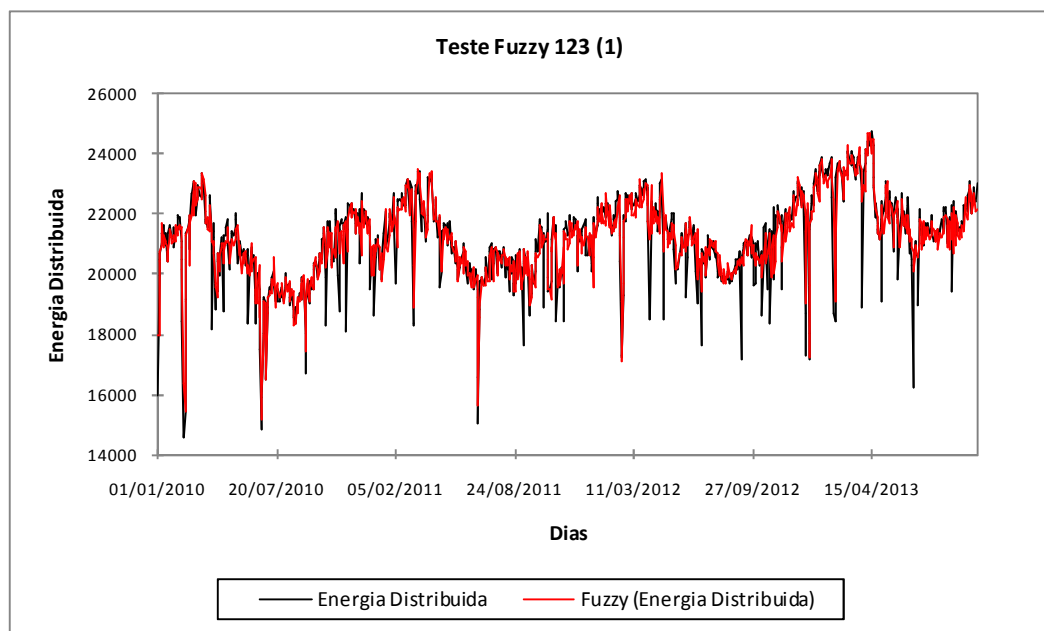


Figura 46 — Teste 123 (1) do Sistema Fuzzy.

A figura abaixo ilustra o resíduo do teste 123 (1). O resíduo consiste na diferença entre o valor previsto no modelo e o valor a ser previsto pelo modelo, ou seja, da série temporal da região Metropolitana da Coelba.

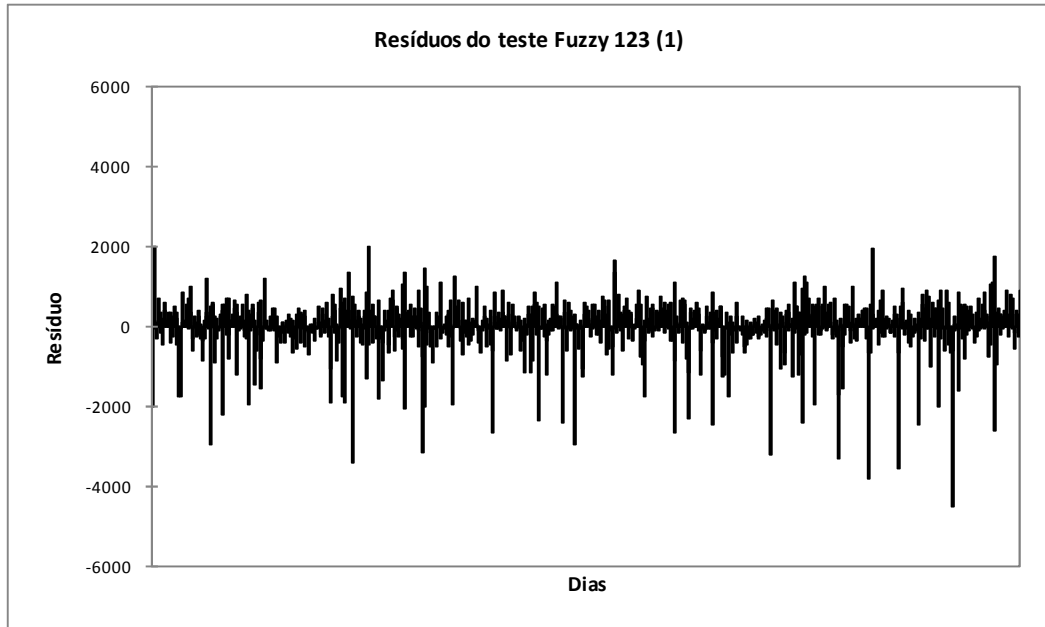


Figura 47 — Resíduos do teste do Sistema Fuzzy 123 (1).

A tabela apresentada a seguir destaca os parâmetros das variáveis exógenas do Sistema de Inferência Fuzzy no teste 123 (1).

Tabela 32— Parâmetros das variáveis do teste 123 (1) do Sistema Fuzzy

Teste 123 (1):	Entrada:
Tipo	gaussmf
Quantidade	4
Épocas	30
Regras	1.024

Como destacado na seção 5.2, a tabela abaixo destaca os erros percentuais do Sistema de Inferência Fuzzy na série original, sendo os erros: médio percentual e o desvio padrão (DP) dos resíduos.

Tabela 33— Erros percentuais do teste 123 (1) do Sistema Fuzzy

[Conteúdo oculto]

O Sistema de Inferência Fuzzy obteve um erro médio percentual de 1,99%. E o desvio padrão dos resíduos de 662.

5.3 Comparativo entre os modelos

Cabe a esta seção destacar as diferenças entre o modelo SARIMAX e o Sistema de Inferência Fuzzy. Assim como nas seções 5.1 e 5.2, a seção 5.3 indicada na tabela abaixo os testes descritos nas seções anteriores.

As tabelas abaixo apresentam os resultados do modelo SARIMAX (seção 5.1), do Sistema de Inferência Fuzzy (seção 5.2) e a diferença entre ambos.

Tabela 34— MAPE dos testes de variáveis do Sistema Fuzzy e SARIMAX.

MAPE	SARIMAX	Fuzzy	SARIMAX - Fuzzy
teste 12 (1)	3,20%	2,18%	1,02%
teste 13 (1)	2,69%	2,31%	0,38%
teste 23 (1)	3,07%	2,23%	0,84%
teste 123 (1)	3,08%	1,99%	1,08%

Tabela 35— DP dos testes de variáveis do Sistema Fuzzy e do SARIMAX.

DP	SARIMAX	Fuzzy	SARIMAX - Fuzzy
teste 12 (1)	964	705	259
teste 13 (1)	867	762	105
teste 23 (1)	919	727	192
teste 123 (1)	919	662	257

Os resultados dos testes com o Sistema de Inferência Fuzzy apresentaram os melhores resultados em todas as combinações de variáveis exógenas. O teste 123 (1) do Sistema de Inferência Fuzzy apresentou um MAPE 1,08% menor e um DP 257 menor que o do SARIMAX.

O resultado do teste 13 (1) do SARIMAX destaca um melhor desempenho com a combinação de 2 variáveis exógenas ao modelo. Diferentemente do Sistema de Inferência Fuzzy que além de possuir como melhor combinação de 2 variáveis o teste 12 (1) tem o teste 123 (1) com o menor erro dentre os testes realizados.

CONCLUSÃO

As primeiras seções apresentaram o mercado de energia elétrica no Brasil e no Estado da Bahia. Foi apresentada a empresa distribuidora de energia elétrica do Estado da Bahia, a Coelba. Também foi destacada a importância e a grandeza do problema de se prever a série temporal diária de distribuição de energia elétrica.

O setor elétrico se insere em uma conjuntura de toda uma população, portanto, o estudo do consumo residencial de energia é particularmente interessante porque nele são refletidas as enormes disparidades econômicas e sociais do país e, qualquer política de mudança efetiva desta realidade repercutirá sobre a estrutura do consumo referente a este setor.

Através da metodologia Box-Jenkins foi aplicado um modelo clássico de previsão de séries temporais. Para a aplicação da previsão da série temporal pelo modelo SARIMAX foi realizado uma AST. Assim foram feitos testes de estacionariedade, FAC, FACP, AIC até a de previsão da série temporal em si. Mas a escolha dos parâmetros não foi feita somente pela interpretabilidade. Além da interpretabilidade foram feitos inúmeros testes para comprovar a escolha dos parâmetros, esses testes estão destacadas na seção 4.3.4. Com a escolha dos parâmetros, também foram realizados testes para diferentes variáveis exógenas. Cabe destacar mais uma vez que o teste 13 (1) obteve o menor erro entre os testes do SARIMAX, com um MAPE de 2,69% e um DP de e 867.

Para o Sistema de Inferência Fuzzy, assim como no modelo SARIMAX, foram realizados testes com diferentes tipos de funções de pertinência. Na comparação entre os dois modelos, o Sistema de Inferência Fuzzy mostrou-se ser o mais preciso na previsão da energia distribuída com o uso de todas as três variáveis: de clientes, temperatura ambiente e precipitação de chuvas, como destacado nas tabelas 33 e 34.

As tabelas 33 e 34 apresentam que a melhor combinação de variáveis para o modelo SARIMAX não foi a melhor combinação de variáveis para o Sistema Fuzzy. Ou seja, O Sistema de Inferência Fuzzy apresentou o menor erro no teste 123 (1) e o SARIMAX no teste 13 (1).

Concluiu-se que, de uma forma geral, neste trabalho, com o uso da série temporal da região Metropolitana da Coelba nos dois modelos propostos, não há “melhores” variáveis exógenas para prever a quantidade diária de energia distribuída na região Metropolitana da Bahia. Pode-se afirmar que para o modelo SARIMAX aplicado a “melhor” combinação de variáveis exógenas foi a de quantidade de clientes e a precipitação de chuvas. E para o

Sistema de Inferência Fuzzy aplicado a “melhor” combinação foi de todas as três variáveis exógenas.

É possível afirmar que no modelo SARIMAX aplicado a inserção de mais “informação” ao sistema, ou seja, de mais uma variável de entrada no sistema não resultou em um menor erro na previsão. Diferentemente do Sistema de Inferência Fuzzy.

O Sistema de Inferência Fuzzy perdeu a sua interpretabilidade, pois gerou uma quantidade de regras muito grande. Sendo geradas 256 regras no teste 12 (1), teste 13 (1) e no teste 23 (1) e no teste 123 (1) gerado 1.024 regras. Para recuperar a interpretabilidade é necessário outro sistema para reduzir essa quantidade de regras.

Cabe ressaltar que a partir da escolha da série temporal de energia distribuída da região Metropolitana da Coelba, foi possível o uso das variáveis: quantidade de clientes, temperatura ambiente e precipitação de chuvas. Através dos testes, comprovou-se que as variáveis exógenas influenciam de forma diferente em cada modelo no comportamento da energia distribuída.

Esta dissertação não se objetivou somente na busca do melhor modelo de previsão proposto. Pois a proposta apresentada não foi limitada em encontrar menores erros percentuais, mas sim em entender o problema em questão e modelá-lo. O conhecimento dos dados modelados e sua aplicabilidade é um diferencial desta dissertação, justificado pela grande quantidade de testes realizados com diferentes combinações de parâmetros e variáveis.

Portanto, “... o estudo dos processos de previsão que aqui se analisam envolve um conjunto de técnicas, que busca estabelecer uma base teórica que norteia o estudo de previsões de séries...” (SOUZA; CAMARGO, 1995).

Por fim, cabe sugerir trabalhos futuros. As sugestões apresentadas abaixo servem para o refinamento de modelos de previsão de energia elétrica. Como por exemplo, a formulação de um sistema híbrido sendo composto tanto pelo modelo SARIMAX como pelo Sistema Fuzzy. Assim seria possível a combinação de modelos com características diferentes. Uma combinação linear, combinando as previsões dos dois modelos, também se justifica na busca de um menor erro de previsão.

Em relação ao modelo SARIMAX, propõe-se ajustar uma função de transferência de Box-Jenkins para as séries temporais utilizadas. Branqueando as séries temporais, ou seja, ajustar um modelo ARMA ou SARIMA para as séries temporais de entrada e de saída. Para correlacionar os seus respectivos ruídos para ver se há dependência.

Na modelagem proposta foi desprezada a parte destinada a classe industrial da energia distribuída, devido à sua pequena representatividade em relação aos consumidores

residenciais que são 88% dos clientes da empresa em estudo. Como a classe industrial foi ignorada nesta tese, é importante também buscar variáveis para modelar o comportamento da classe industrial. Pois a classe industrial também tem a sua participação na energia distribuída. O PIB e a quantidade de venda de produtos seriam algumas sugestões de variáveis para modelizar.

Além da classe industrial é importante se ater as regiões que se localizam mais no interior do Estado. Muitas dessas regiões são formadas por populações rurais, onde a quantidade de energia distribuída é menos sensível a variações de temperatura. Pois, em geral, os consumidores não possuem um ar condicionado ou outros aparelhos eletrodomésticos. Programas governamentais como do empreendimento Minha Casa Minha Vida e o LPT seriam algumas dessas variáveis que devem ser modelizadas.

O autor Serrão afirma que para cada classe de consumidores é possível se observar algumas características particulares (SERRÃO, 2003), como:

- Residencial: carga com crescimento constante, sensível às condições meteorológicas, as mudanças econômicas e tarifárias, etc.
- Comercial: carga sensível à sazonalidade resultante de eventos como datas comemorativas, condições meteorológicas e mudanças econômicas.
- Industrial: carga bem específica, definida pelo número de turnos de operação.

Para maiores periodicidades que a diária pode-se alimentar o modelo com a relação de consumo do consumidor com o valor da tarifa de energia. Pois é a lei da oferta e da demanda, de uma forma geral, mantendo todas as outras variáveis: Quanto mais cara a tarifa, menor vai ser o consumo e quanto mais barata a tarifa maior será o consumo.

REFERÊNCIAS

- ALMEIDA, E. de S. **Previsão de carga a curto prazo: uma comparação de modelos**. Campos dos Goytacazes. RJ. UENF, 2008.
- ARRUDA, M. V. N. **Implementação de projetos Smart Grid no Brasil. Instituto Federal de educação, ciência e tecnologia**. Mato Grosso, Cuiabá, 2013.
- AZADEH, V. N. A.; PAZHOHESHFAR, P.; SABERIC, M. **An Adaptive-Network-Based Fuzzy Inference System for Long-Term Electric Consumption Forecasting (2008-2015): A Case Study of the Group of Seven (G7) Industrialized Nations: U.S.A, Canada, Germany, United Kingdom, Japan, France and Italy**. IEEE, 2010.
- BOX, G. E. P.; JENKIS, G. M. **Time Series Analysis: Forecasting and Control**. San Francisco. Holden Day, 1970.
- BRASIL. Conversão da MPv nº 145, de 2003. Lei nº 10.847, de 15 de março de 2004. Autoriza a criação da Empresa de Pesquisa Energética – EPE e dá outras providências. D.O.U. DE 16/03/2004, P. 1.
- BRASIL. Conversão da MPv nº 144, de 2003. Lei nº 10.848, de 15 de março de 2004. Dispõe sobre a comercialização de energia elétrica, altera as Leis nos 5.655, de 20 de maio de 1971, 8.631, de 4 de março de 1993, 9.074, de 7 de julho de 1995, 9.427, de 26 de dezembro de 1996, 9.478, de 6 de agosto de 1997, 9.648, de 27 de maio de 1998, 9.991, de 24 de julho de 2000, 10.438, de 26 de abril de 2002, e dá outras providências. D.O.U. DE 16/03/2004, P. 2.
- BRASIL. Decreto nº 5.163, de 30 de julho de 2004. Regulamenta a comercialização de energia elétrica, o processo de outorga de concessões e de autorizações de geração de energia elétrica, e dá outras providências. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 30 jul. 2004. Seção 1, p.1.
- BRESEGHELLO, F. N. **Estudo comparativo de métodos de previsão de demanda: uma aplicação ao caso dos aeroportos com tráfego aéreo regular administrados pelo DAESP**, 2005. Tese de mestrado – Instituto Tecnológico de Aeronáutica, São José dos Campos.
- CAMPOS, R. J.; JESUS, T. A.; MENDES, E. M. A. M. **Uma abordagem fuzzy para a previsão de curto-prazo do consumo de energia elétrica**. In XXX Congresso Nacional de Matemática Aplicada - CNMAC, Brasil, Florianópolis, Setembro, 2007.
- CHASE, R. B.; JACOBS, F. R.; AQUILANO, N. J. **Administração da produção para a vantagem competitiva**. 10 ed. Porto Alegre/RS: Ed. Bookman, 724 p. ISBN 0-07-250636-9, 2006.
- CHAVEZ, G. S.; BERNAT, J. X.; COALLA, H. L. **Forecasting of energy production and consumption in Asturias (Northern Spain)**. Energy, 24, 183–98, 1999.
- CHOW, L. C. H. **A study of sectoral energy consumption in Hong Kong (1984–97) with special emphasis on the household sector**. Energy Policy, 29, 1099–110, 2001.

CHRISTO, E. S.; SOUZA, R. C. **Uma abordagem estatística para a previsão de potência reativa em sistemas elétricos**. Pesquisa Oper. vol. 26. número 2. Rio de Janeiro. Maio-Agosto. 2006.

COCHRANE, D., ORCUTT, G. H. (1949). **Application of least squares regression to relationships containing auto-correlated error terms**. Journal of the American Statistical Association, 44(245), 32-61. 1949.

CONEJO, A. J. **Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models**. IEEE Transactions on Power Systems, Vol. 20, No. 2. Maio, 2005.

CONEJO, A. J.; CONTRERAS, J.; ESPÍNOLA, R.; PLAZAS, M. A. **Forecasting electricity prices for a day-ahead pool-based electric energy market**. International Journal of Forecasting, Vol. 21, pp. 435-65, 2005.

CUARESMA, J. C.; HLOPUSKOVA, J.; KOSSMEIER, S.; OBERSTEINER, M. **Forecasting electricity spot-prices using linear univariate time series models**. Applied Energy, Vol. 77, pp. 87-106, 2004.

DEJAER, K.; VERBEKE, W.; MARTENS, D.; BAESSENS B. **Data Mining Techniques for Software Effort Estimation: A Comparative Study**. IEEE Transactions on Software Engineering, v.38 n.2, p.375-397, Março, 2012.

EDIGER, V. S.; TATLLDIL, H. **Forecasting the primary energy demand in Turkey and analysis of cyclic patterns**. Energy Conversion and Management, 43, 473-87, 2002.

ENDERS, W. **Applied Econometric Time Series**. 2 ed. United States of America: Ed. Wiley Series In Probability And Statistics, 466 p. ISBN 0-471-23065-0, 2004.

ESPINOSA, M. M.; PRADO, S. M.; GHELLERE, M. **Uso do modelo SARIMA na previsão do número de focos de calor para os meses de junho a outubro no Estado de Mato Grosso**. Departamento de Estatística/ICET. Universidade Federal de Mato Grosso, Cuiaba, MT, 2010.

FAVA, V. L. **Metodologia de Box-Jenkins para modelos univariados**. In: Vasconcellos, M. A. S.; Alves, D. Manual de econometria: nível intermediário. São Paulo: Atlas, 2000.

FERREIRA, F. B. **A evolução automobilística e aplicação de smartgrid para utilização de energia renovável**. Universidade São Francisco. Itatiba, 2012.

GUJARATI, D. N.; **Econometria básica**. 3 ed. São Paulo: Makron Books, 2000.

GOMIDE, F.; GUDWIN, R.; TANSCHKEIT, R. **Conceitos fundamentais da teoria de conjuntos fuzzy, lógica fuzzy e aplicações**. In: Porc. 6th IFSA Congress-Tutorials, 1995.

GOODRICH, R.L.; STELLWAGEN, E. A. **Forecast Pro for Windows**. Business Forecast Systems, 1999.

GOOIJER, de; JAN, G.; HYDMAN, R. J. **25 years of time series forecasting**. International Journal of Forecasting, 22(3):443-473, 2006.

GREENE, Willian H. **Econometric analysis**. 5. ed. New Jersey: Prentice Hall, 2003.

HERRERA, F. Genetic fuzzy systems: taxonomy, current research trends and prospects. **Evolutionary Intelligence**, v. 1, n. 1, p. 27–46, 1 mar. 2008.

JANG, J. S. R.; GULLEY, N. **The Fuzzy Logic Toolbox for use with MATLAB**. The MathWorks, Inc., Natick, Massachusetts, 1995.

JUNIOR, F. J. M. **Proposta de um método para o controle estatístico de processo para observações autocorrelacionadas**. 2005. 138 p. Dissertação (Mestrado em engenharia de produção) – Programa de Pós-Graduação em Engenharia de Produção (PPGEP), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.

KADOSHIN, S.; NISHIYAMA T.; TOSSHIHIDE, I. **The trend in current and near future of energy consumption from a statistical perspective**. *Applied Energy*, 67, 407–17. 2000.

KOUK, C. M.; FU, A.; WONG, M. H. **Mining Fuzzy Association Rules in Databases**, SIGMOD Record, pp. 41-46, Vol. 27, No. 1, 1998.

LATORRE, M. R. D. O.; CARDOSO M. R. A. **Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos**. *Revista Bras. Epidemiol*, 4:145-52, 2001.

LEVINE, D. M.; STEPHAN, D.; KREHBIEL, T. C.; BERENSON, M. L. **Estatística – Teoria e aplicações usando o Microsoft Excel**. 3ª ed. Rio de Janeiro/RJ: 819 p. Ed. LTC, 2005.

LI, S.T.; CHENG, Y. C. **A Stochastic HMM-Based Forecasting Model for Fuzzy Time Series**. *IEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetic*, Vol. 40, No. 5. Outubro, 2010.

LI C.; CHIANG, T. W. **Complex Neurofuzzy ARIMA Forecasting – A new Approach Using Complex Fuzzy Sets**. *IEEE Transactions on Fuzzy Systems*. Vol. 21. No. 3. Junho, 2013.

LIU, L; HUDAK, G. P. **Forecasting and Time Series Analysis Using the SCA Statistical System**. Scientific Computing Associated, 1994.

MATOS, O. C. **Econometria básica: teoria e aplicações**. São Paulo: Atlas, 2000.

MOORE, D.S.; McCABE, G. P.; DUCKWORTH, W. M.; SCLOVE, S. L.. **A prática da estatística empresarial: como usar dados para tomar decisões**. 1ª Ed. Rio de Janeiro/RJ: Ed. LTC, 952 p. ISBN 85-216-1500-0, 2006.

MORETTIN, Pedro A. **Econometria Financeira: um curso de séries temporais financeiras**. ABE: São Paulo, 2006.

MOURA, Fernando Alves de; MONTINI, Alessandra de Ávila; CASTRO, João Bosco Barroso de. **Modelagem do Consumo de Energia Elétrica Residencial no Brasil Através de Modelos ARMAX**. Seminario de Administração. XIV SemeAd. Outubro, 2011.

OLIVEIRA, L. E. S.; MORITA, M. E. **Introdução aos Modelos Escondidos de Markov (HMM)**. Pontifícia Universidade Católica do Paraná – PUC-Pr, 2005.

PANDIAN, S. C.; DURAISWAMY, K.; RAJAN, C. C. A.; KANAGARAJ N. **Fuzzy approach for short term load forecasting**. Electric Power Systems Research, 76: pp. 541-548, 2006.

PEDRYCZ, W. **Fuzzy control and fuzzy systems**. [s.l.] Research Studies Press, 1989.

PELLEGRINI, F.R. **Metodologia para implementação de sistemas de previsão de demanda** – Dissertação de Mestrado em Engenharia de Produção. Porto Alegre: UFRGS, 2000.

PINDYCK, R. S.; RUBINFELD, D. L. **Econometria: modelos e previsões**. ed. 4. São Paulo: Campus-Elsevier, 2004.

PINDYCK, R. S.; RUBINFELD, D. L. **Econometric models & economic forecasts**. 2 ed. United States of America: Ed. McGraw - Hill Book Company, 1981, 630 p. ISBN 0-07-050096-7.

PINHEIRO, T. M. M. **Distribuidoras de Energia Elétrica no Brasil**. Dissertação de mestrado. Universidade de Brasília, 2012. Disponível em: http://www.aneel.gov.br/biblioteca/trabalhos/trabalhos/Dissertacao_Thelma_Pinheiro.pdf.

REDDY, B. S.; BALACHANDRA, P. **Integrated energy–environment-policy analysis: a case study of India**. Utility Policy, 11(2), 59–73. 2003.

RODRIGUES, L. C.; SILVA, P. P. C.; LINDEN, R. **Séries Temporais no Consumo de Energia Elétrica no Estado do Rio de Janeiro**. Revista Visões 1. 2007.

SAAB, S.; BADR, E.; NASR, G. **Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon**. Energy, 26, 1–14. 2001.

SÁFADI, T. **Uso de séries temporais na análise de vazão de água na represa de Furnas**. Ciência Agrotécnica, Lavras, v.28,n.1 ,p. 142-148, jan/fev, 2004.

SALARIJAZI M., ALI-MOHAMMAD A., ARASH A. and ALIREZA D. **Trend and change-point detection for the annual stream-flow series of the Karun River at the Ahvaz hydrometric station**. African Journal of Agricultural Research. Vol. 7(32), pp: 4540-4552, 2012.

SANDRI, S.; CORREA, C. **Lógica nebulosa**. 1999.

SERRÃO, F. C. C. **Modelo de previsão de carga de curto-prazo utilizando redes neurais e lógica fuzzy**. Dissertação de Mestrado. Pontifícia Universidade Católica, Brasil, Rio de Janeiro, 2003.

SONG, Q.; CHISSOM, B. S. **Forecasting enrollments with fuzzy time series - Part I**, Fuzzy Sets Syst., vol. 54. pp. 1-9, 1993.

SONG, Q.; CHISSOM, B. S. **Fuzzy time series and its models**, Fuzzy Sets Syst., vol. 54. pp. 269-277, 1993.

SONG, Q.; CHISSOM, B. S. **Forecasting enrollments with fuzzy time series - Part II**, Fuzzy Sets Syst., vol. 62. pp. 1-8, 1994.

SOUZA, R. C.; CAMARGO, M. E. **Análise e previsão de séries temporais: os modelos ARIMA**. SEDIGRAF, 1996.

SOUZA, R. C.; OLIVEIRA, F. C.; FERREIRA, P. C.; MARCATO, A. L. M.; DIAS, B. H.; BRANDI, R. S.; RAMOS, T. P. **Planejamento da operação de sistemas hidrotérmicos no Brasil**. Geração de cenários e otimização. Editora PUC-Rio, 2014.

SUGENO, M. **Industrial applications of fuzzy control**. Elsevier Science Pub. Co., 1985.

TANSCHKEIT, R. **Sistemas fuzzy**, In: **Inteligência computacional aplicada a administração, economia e engenharia em Matlab**, pp. 229-264, São Paulo, Thomson Learning, 2007.

TEIXEIRA, M. A.; ZAVERUCHA, G.; SILVA, V. N. A. L.; RIBEIRO, G. F. **Fuzzy Bayes Predictor in Electric Load Forecasting**, **International Joint Conference on Neural Networks**, Washington DC, vol. 4, pp. 2339-234, 2001.

URI, N. D. **Forecasting peak system load using a combined time series and econometric model**. *Applied Energy*, Vol. 4, pp. 219-227, 1978.

XIE, M.; SANDELS, C.; ZHU, K.; NORDSTROM, L. **A seasonal ARIMA model with exogenous variables for elspot electricity prices in Sweden**. *European Energy Market (EEM)*, 2013 10th International Conference on the, vol., no., pp.1,4, 27-31. Maio de 2013.

WANG, Li-Xin. **A course in fuzzy systems and control**, 1997.

WANGDI, K; SINGHASIVANON, P; SILAWAN, T; LAWPOOLSRI, S; WHITE, N J.; KAEWKUNGWAL, J. **Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: A case study in endemic districts of Bhutan**. *Malaria Journal*, 2010.

WHEELWRIGHT, S. C. **Manufacturing Strategy: Defining the Missing Link**. *Strategic Management Journal*. vol. 5, p.77-91, 1984.

ZADEH, L. A., **Fuzzy sets**. *Fuzzy sets, Information and Control*, 8:338-353, 1965.

ZANINI, Alexandre. **Redes Neurais e regressão dinâmica: um modelo híbrido para previsão de curto prazo da demanda de gasolina automotiva no Brasil**. Dissertação de Mestrado em Engenharia Elétrica: Teoria de Controle e Estatística. Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro, 2000.

ZHANG, G. P. **Time series forecasting using a hybrid ARIMA and neural network model**. *Neurocomputing-Elsevier, USA*, n. 50, p. 159-175, 2003.

ZHOU, M.; YAN, Z.; NI, Y. X.; LI, G.; NIE, Y. **Electricity price forecasting with confidence-interval estimation through an extended ARIMA approach**. *IEEE Proc-Gener. Transm. Distrib*, Vol 153, No 2. Março de 2006.

ANEXO A

Os dados de energia distribuída ao longo da tese foram fornecidos e autorizados pela empresa Iberdrola. A equipe de planejamento da empresa Iberdrola indicou as variáveis exógenas para teste nos modelos.

DECLARAÇÃO DE USO DE INFORMAÇÕES

Declaro que o aluno Cesar Machado Pereira do Programa de Pós-Graduação em Engenharia Eletrônica de matrícula ME 1220075 utilizou os dados e as informações da empresa Coelba para fins acadêmicos do projeto de pesquisa.

Por ser verdade, firmo a presente.

Rio de Janeiro, 25 de fevereiro de 2015.


Eduardo Capelastegui Saiz
Diretor de Controle e Gestão
Iberdrola Brasil S.A.

Eduardo Capelastegui Saiz