



**ANÁLISE DE SÉRIES DE TEMPO APLICADA À DIFUSÃO DA
ENERGIA FOTOVOLTAICA NO BRASIL**

ARTHUR DE BRITO DIAS

**TRABALHO DE CONCLUSÃO DE CURSO EM ENGENHARIA ELÉTRICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**FACULDADE DE TECNOLOGIA
UNIVERSIDADE DE BRASÍLIA**

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**TIME SERIES ANALYSIS APPLIED TO PHOTOVOLTAIC POWER
DIFFUSION IN BRAZIL**

**ANÁLISE DE SÉRIES DE TEMPO APLICADA À DIFUSÃO DA
ENERGIA FOTOVOLTAICA NO BRASIL**

ARTHUR DE BRITO DIAS

**ORIENTADOR: DR. ANÉSIO DE LELES FERREIRA FILHO
COORIENTADOR: DR. FERNANDO CARDOSO MELO**

**TRABALHO DE CONCLUSÃO DE CURSO EM
ENGENHARIA ELÉTRICA**

PUBLICAÇÃO: PPGEA.TD-001/11

BRASÍLIA/DF: NOVEMBRO - 2020

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**ANÁLISE DE SÉRIES DE TEMPO APLICADA À DIFUSÃO DA
ENERGIA FOTOVOLTAICA NO BRASIL**

ARTHUR DE BRITO DIAS

**TRABALHO DE CONCLUSÃO DE CURSO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA
ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA COMO
PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL.**

APROVADA POR:

**Prof. Dr. Anésio de Leles Ferreira Filho – ENE/Universidade de Brasília
Orientador**

**Prof. Dr. Fernando Cardoso Melo – ENE/Universidade de Brasília
Coorientador**

**Prof. Dr. Rafael Amaral Shayani – ENE/Universidade de Brasília
Membro Interno**

**Dra. Livia Maria de Rezende Raggi – ANEEL/SRD
Membro Externo**

BRASÍLIA, 30 DE NOVEMBRO DE 2020.

FICHA CATALOGRÁFICA

DIAS, ARTHUR DE BRITO;

Análise de Séries de Tempo aplicada à Difusão da Energia Fotovoltaica no Brasil [Distrito Federal] 2020.

xi, 66p., 210 x 297 mm (ENE/FT/UnB, Bacharel, Engenharia Elétrica, 2020).

Trabalho de Conclusão de Curso – Universidade de Brasília, Faculdade de Tecnologia.

Departamento de Engenharia Elétrica

1. Análise de Séries de Tempo

3. Difusão de Produtos

I. ENE/FT/UnB

2. Mercado Fotovoltaico

4. Previsão Estatística

II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

DIAS, A.B. (2020). Análise de Séries de Tempo aplicada à Difusão da Energia Fotovoltaica no Brasil . Trabalho de Conclusão de Curso em Engenharia Elétrica, Publicação PPGEA.TD-001/11, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 66p.

CESSÃO DE DIREITOS

AUTOR: Arthur de Brito Dias

TÍTULO: Análise de Séries de Tempo aplicada à Difusão da Energia Fotovoltaica no Brasil .

GRAU: Bacharel ANO: 2020

É concedida à Universidade de Brasília permissão para reproduzir cópias desta trabalho de conclusão de curso e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte dessa trabalho de conclusão de curso pode ser reproduzida sem autorização por escrito do autor.

Arthur de Brito Dias

Departamento de Engenharia Elétrica (ENE) - FT

Universidade de Brasília (UnB)

Campus Darcy Ribeiro

CEP 70919-970 - Brasília - DF - Brasil

*A tranquilidade é trágica. Nunca
seja tranquilo!*

Antônio Abujamra

DEDICATÓRIA

À minha mãe, Maria Helena, a mulher que me ensinou a correr atrás de tudo que significar vitória e felicidade na minha vida, me agarrando ao amor e à disciplina quando as angústias porventura aparecerem. Ao meu pai, João, o homem que me mostrou que a vida é mais bonita quando contemplada de forma calma e inteligente, sempre orientada a aprender algo novo dia após dia, e sempre abraçar o próximo. Vocês são meu maior símbolo de força, perseverança e amor; "tudo de bom em mim começou com vocês".

A todos os meus familiares, que sempre me acolheram e me acompanharam nos momentos felizes da vida. Especialmente ao meu primo Vinícius, que sempre foi e sempre será um referencial de sabedoria e excelência profissional, e ao meu primo Fabrício, por estar a quase duas décadas compartilhando muitos dos melhores momentos de companheirismo e amizade da minha vida.

Aos docentes do Departamento de Engenharia Elétrica da Universidade de Brasília, especialmente aos professores Anésio, Fernando e Eduardo Peixoto. São mestres que enriqueceram enormemente minha vivência como discente desta Faculdade, por meio de todos os ensinamentos transmitidos pelas disciplinas do curso e pelas orientações acadêmicas, com carinho e cuidado proporcionais à qualidade e à exigência.

Aos colegas do 1º/2015: Pedro Ferreira, André, Letícia, Felipe, Marcelo, Thiago, Rodrigo Fischer, Hiandra, Gabriel Lins, Allan, Pedro Caio, Rodrigo Borba, Edson, Caio, Gustavo, Vinícius, Osmar, Victor, Gabriel Soutinho, Pedro Nunes e Luiz Filomeno. Foi um privilégio compartilhar estes anos de graduação com estas pessoas de corações e mentes gigantes, e graças a vocês, ter cursado Engenharia Elétrica é sinônimo de lembrar de muitas vitórias, perdas, aprendizados, risadas, festas e conversas maravilhosas.

Ao corpo técnico da Superintendência de Regulação dos Serviços de Distribuição da ANEEL, em especial aos mestres Carlos, Daniel, Hugo, Marília, Henrique, Nara, Larissa, Lívia, Clélia, Maria Lúcia, Renato, Pedro e Luciano. Estes excelentes profissionais me acolheram durante meu período de estágio e me proporcionaram conhecimentos importantes sobre o mercado elétrico brasileiro e os papéis da Engenharia na esfera pública.

A todos os amigos-professores da dança de salão, em especial Wesley, Carolina, João Luis, Luiz Felipe, Matheus Barbosa, Matheus Henrique, Sabrina, Emily, Bruna, Izabella, Sanderson, Matheus Rinco, Luiz Liberal, Letícia, Christian, Izabella, Juliana, Tatiana, Flávio, Tiago, Ludmila, Hércules, Roberta, Reynaldo, Maria Clara, Leandro, Joana, Paulo e Thaynan. Compartilhar de tantas danças, bailes e aulas com vocês é realizar um sonho que eu não sabia que eu tinha dentro de mim.

RESUMO

Título: Análise de Séries de Tempo Aplicada à Difusão da Energia Fotovoltaica no Brasil

Autor: Arthur de Brito Dias

Orientador: Dr. Anésio de Leles Ferreira Filho

Coorientador: Dr. Fernando Cardoso Melo

Programa de Graduação em Engenharia Elétrica

Brasília, 30 de novembro de 2020

A capacidade de se prever e projetar tendências do mercado fotovoltaico é uma atividade importante para a definição de diretrizes de planejamento energético, estruturação tarifária e simulações de impactos na rede devidos a diferentes níveis de penetração fotovoltaica. Com vistas à contribuição na área de projeções de potência fotovoltaica, este trabalho apresenta algumas metodologias já consagradas na área de análise de séries de tempo, propondo a modelagem das séries históricas brasileiras de microgeração fotovoltaica residencial e microgeração fotovoltaica comercial no período entre agosto de 2013 e dezembro de 2019. Foram obtidos três categorias de modelos estatísticos(suavização exponencial, ARIMA e modelos estruturais) para cada série apresentada, os quais apresentaram uma faixa de erro percentual de previsão(MAPE) de 0,7% a 3,2%.

Palavras-chave: Energia Fotovoltaica, Estatística, Previsão, Séries de Tempo

ABSTRACT

Title: Time Series Analysis Applied to Photovoltaic Power Diffusion in Brazil

Author: Arthur de Brito Dias

Supervisor: Dr. Anésio de Leles Ferreira Filho

Co-Supervisor: Dr. Fernando Cardoso Melo

Graduate Program in Electrical Engineering

Brasília, November 30th, 2020

The ability to predict and project trends in the photovoltaic market is an important activity for the definition of energy planning guidelines, tariff structuring and simulations of impacts on the network due to different levels of photovoltaic penetration. Aiming contribution in the area of photovoltaic power forecasting, this work presents some methodologies already established in the area of time series analysis, proposing the modelling of Brazilian historical series of residential photovoltaic microgeneration and commercial photovoltaic microgeneration in the period between August 2013 and December 2019. Three categories of statistical models (exponential smoothing, ARIMA and structural models) were obtained for each series presented, which presented a range of percentage error of prediction (MAPE) of 0.7 % to 3.2 %.

Keywords: Photovoltaics, Statistics, Forecasting, Time Series

SUMÁRIO

1	INTRODUÇÃO	1
1.1	CONTEXTO	1
1.2	MOTIVAÇÃO.....	1
1.3	OBJETIVOS	3
2	FUNDAMENTAÇÃO TEÓRICA E METODOLOGIA.....	4
2.1	ANÁLISE DE SÉRIES TEMPORAIS E PROCESSOS ESTOCÁSTICOS	4
2.1.1	AUTOCORRELAÇÃO E AUTOCORRELAÇÃO PARCIAL	5
2.1.2	ESTACIONARIEDADE DE SÉRIES TEMPORAIS	7
2.1.3	TESTES DE HIPÓTESE PARA ESTACIONARIEDADE	8
2.1.3.1	TESTE KPSS	8
2.1.3.2	TESTE DE DICKER-FULLEY AUMENTADO.....	9
2.2	MODELOS DE REGRESSÃO NÃO-LINEAR	10
2.2.1	ESTIMAÇÃO POR MÍNIMOS QUADRADOS NÃO-LINEARES	11
2.2.2	MODELO DE DIFUSÃO DE BASS.....	11
2.3	MODELOS DE SUAVIZAÇÃO EXPONENCIAL.....	12
2.4	MODELOS AUTORREGRESSIVOS	14
2.5	MODELOS DE MÉDIA MÓVEL	15
2.6	MODELOS MISTOS	15
2.6.1	DIFERENCIAÇÃO	16
2.6.2	MODELOS ARIMA	16
2.7	MODELOS DE ESPAÇO DE ESTADOS	17
2.8	MODELOS ESTRUTURAIS E O FILTRO DE KALMAN.....	17
2.8.1	MÉTRICAS DE QUALIDADE DE AJUSTE.....	19
2.8.1.1	CRITÉRIO DE INFORMAÇÃO DE AKAIKE	19
2.8.1.2	TESTE DE LJUNG-BOX	19
2.9	CONSIDERAÇÕES FINAIS	20
3	REVISÃO BIBLIOGRÁFICA	21
3.1	SIMULAÇÃO DO MERCADO BRASILEIRO VIA MODELO DE BASS.....	21
3.2	ANÁLISE ECONOMETRICA DA DIFUSÃO FOTOVOLTAICA	23
3.3	DIFUSÃO ESPACIAL E TEMPORAL FOTOVOLTAICA	25
3.4	SIMULAÇÕES DE DIFUSÃO BASEADAS EM ABM	28
3.5	CONSIDERAÇÕES FINAIS	29
4	METODOLOGIA	30

4.1 PESQUISA METODOLÓGICA INICIAL	30
4.2 ESCOLHA E TRATAMENTO DE DADOS	31
4.3 METODOLOGIA ESTATÍSTICA	32
4.3.1 OBTENÇÃO DOS MODELOS DE SUAVIZAÇÃO EXPONENCIAL	33
4.3.2 OBTENÇÃO DOS MODELOS ARIMA	33
4.3.3 OBTENÇÃO DOS MODELOS DE ESPAÇO DE ESTADOS	35
4.4 FERRAMENTAS COMPUTACIONAIS	35
4.5 CONSIDERAÇÕES FINAIS	36
5 RESULTADOS	37
5.1 SÉRIE DE MICROGERAÇÃO LOCAL RESIDENCIAL	37
5.1.1 VISUALIZAÇÃO GRÁFICA E TRANSFORMAÇÕES	37
5.1.2 AJUSTES ARIMA	40
5.1.3 AJUSTES DE SUAVIZAÇÃO EXPONENCIAL	41
5.1.3.1 AJUSTES DE MODELOS DE ESPAÇO DE ESTADOS	41
5.2 QUALIDADE PREDITIVA DOS MODELOS SELECIONADOS	42
5.3 SÉRIE DE MICROGERAÇÃO COMERCIAL LOCAL	44
5.3.1 VISUALIZAÇÃO GRÁFICA E TRANSFORMAÇÕES	44
5.3.2 AJUSTES ARIMA	46
5.3.3 AJUSTES DE SUAVIZAÇÃO EXPONENCIAL	47
5.3.3.1 AJUSTES DE MODELOS DE ESPAÇO DE ESTADOS	47
5.4 QUALIDADE PREDITIVA DOS MODELOS SELECIONADOS	48
5.5 CONSIDERAÇÕES FINAIS	49
6 CONCLUSÃO E TRABALHOS FUTUROS	51
REFERÊNCIAS BIBLIOGRÁFICAS	51
A TABELAS	56

LISTA DE FIGURAS

2.1	Componentes de Séries de Tempo, Moreira <i>et. al.</i>	5
2.2	Demanda mensal de energia na Austrália(1980-1995), Hyndman e Athasanopoulos	7
2.3	Autocorrelograma para a série temporal acima, Hyndman e Athasanopoulos. Nota-se a junção dos efeitos de sazonalidade e tendência crescente.	7
3.1	Fluxograma da metodologia de simulação de mercado.	23
3.2	Representação de <i>cluster</i> individual.	25
3.3	Representações gráficas de espalhamento de <i>clusters</i>	26
3.4	Representação de atributos por célula espacial, Zhao <i>et al.</i>	27
5.1	Potência Instalada - Microgeração Residencial Local	37
5.2	Potência Instalada - Microgeração Residencial Local(Log)	38
5.3	Potência Instalada - Microgeração Residencial Local(Diferenciada)	39
5.4	Autocorrelograma - Microgeração Residencial Local	40
5.5	Autocorrelograma Parcial - Microgeração Residencial Local	40
5.6	Previsão ARIMA(0,1,2) - Microgeração Residencial Local	42
5.7	Previsão de SE(sem amortecimento) - Microgeração Residencial Local	43
5.8	Previsão de EE com Tendência Estocástica - Microgeração Residencial Local	43
5.9	Potência Instalada - Microgeração Comercial Local	44
5.10	Potência Instalada - Microgeração Comercial Local(Log)	44
5.11	Potência Instalada - Microgeração Comercial Local(Diferenciada)	45
5.12	Autocorrelograma - Microgeração Comercial Local	46
5.13	Autocorrelograma Parcial - Microgeração Comercial Local.....	46
5.14	Previsão ARIMA(0,1,2) - Microgeração Comercial Local	48
5.15	Previsão de SE(sem amortecimento) - Microgeração Comercial Local	49
5.16	Previsão de EE com Tendência Estocástica - Microgeração Comercial Local .	49

LISTA DE TABELAS

2.1	Valores críticos de cauda superior para o teste KPSS, Kwiatkowski <i>et al.</i>	9
2.2	Tabela de valores críticos para o Teste ADF.	10
3.1	Erros médios quadráticos e parâmetros finais ajustados.	24
3.2	Previsões de energia instalada.	24
5.1	Ajustes ARIMA - Microgeração Residencial Local	41
5.2	Ajustes SE - Microgeração Residencial Local	41
5.3	Ajustes EE - Microgeração Residencial Local	41
5.4	Erros de Previsão(ln) - Microgeração Residencial Local	42
5.5	Ajustes ARIMA - Microgeração Comercial Local	47
5.6	Ajustes SE - Microgeração Comercial Local	47
5.7	Ajustes EE - Microgeração Comercial Local	47
5.8	Erros de Previsão(ln) - Microgeração Comercial Local	48
A.1	Série de Tempo: Microgeração Residencial Local	56
A.2	Série de Tempo: Microgeração Comercial Local	58
A.3	Previsão ARIMA(0,1,2) - Microgeração Residencial Local.	61
A.4	Previsão de Suavização Exponencial - Microgeração Residencial Local	62
A.5	Previsão de Espaço de Estados - Microgeração Residencial Local	63
A.6	Previsão ARIMA(0,1,3) - Microgeração Comercial Local	64
A.7	Previsão de Suavização Exponencial - Microgeração Comercial Local	65
A.8	Previsão de Espaço de Estados - Microgeração Comercial Local	66

1

INTRODUÇÃO

1.1 CONTEXTO

A crescente depleção de combustíveis fósseis e perspectivas futuras de escassez dessas matrizes provoca, principalmente, desde a segunda metade do século XX, a prospecção por fontes de energias alternativas. A revolução na composição da matriz energética mundial a partir da consolidação energias renováveis é um passo importante para o desenvolvimento das sociedades globais para as próximas décadas. Este processo tem como motivação uma série de demandas socioambientais, como a redução da geração de subprodutos nocivos advindos da indústria energética(e principalmente de matrizes baseadas em combustíveis fósseis) e a redução dos níveis globais de emissão de gases de efeito estufa, e a melhora das condições da vida em ambientes urbanos, com a melhora nas condições respiratórias e a consequente diminuição da pressão sobre custos financeiros e estruturais nas redes de saúdes desses locais. Dentro dessa busca por novas alternativas energéticas, a energia fotovoltaica se destaca como uma das fontes renováveis que exhibe crescimento mais expressivo na matriz energética mundial. Ela exibiu um crescimento exponencial entre o período de 1977 a 2018, sendo que a potência fotovoltaica total instalada no mundo atingiu, no final daquele período, um total de 633 GW [1], dentre os quais 276,6 GW (54%) foram instalados entre 2015 e 2018. A tendência de custo energético no mercado fotovoltaico mostra um decréscimo bastante acelerado, seguindo o princípio chamado Lei de Swanson: o preço do módulo de silício cristalino cai 20% a cada vez que a potência total instalada solar mundial dobra.[2].

1.2 MOTIVAÇÃO

Há muita complexidade em realizar previsões a respeito deste ritmo de crescimento, e consequentemente, projetar cenários futuros de penetração fotovoltaica. Neste sentido, a abordagem da maioria dos órgãos de pesquisa e órgãos reguladores ao redor do mundo é realizar regressões estatísticas não-lineares sobre os conjuntos de séries históricas de crescimento fotovoltaico. No entanto, esses tipos de metodologias falham em captar nuances mais detalhadas das curvas de crescimento dessas séries históricas, visto que o mercado fotovoltaico é um sistema com diversas variáveis influenciadoras e uma grande quantidade de agentes, cada um com muitas possibilidades de decisão. Isso resulta em curvas de expansão com discordâncias consideráveis entre valores previstos e valores reais, observáveis em resultados de órgãos como a WEO[3] e a IEA[4, 5].

Com vistas ao melhoramento das discrepâncias estatísticas para o mercado fotovoltaico brasileiro, o trabalho proposto buscou referências aos métodos de estado da arte dentro do campo de regressões estatísticas e modelos de previsão. Entre estes métodos, Modelagens Baseadas em Agentes (MBA) e métodos de Machine Learning são empregados com bastante sucesso no campo das previsões estatísticas. No entanto, o primeiro método requer uma quantidade e distinção de variáveis de entrada tais que a obtenção de todas elas não pode ser feita de forma completa no escopo deste trabalho; já o segundo método apenas apresenta resultados confiáveis para séries de tempo com mais pontos de observação do que as séries anteriores.

Além destes dois métodos, também existem metodologias mais consolidadas, como os estudos de regressões estatísticas lineares baseadas em análise de séries de tempo. Tal campo engloba uma variedade de modelos como métodos de espaço de estados, suavização exponencial e métodos autorregressivos, que possuem uma extensa literatura de referência e podem ser implementados de forma relativamente simples por meio de linguagens de programação como Python e R. Portanto, este trabalho tem como objetivo analisar e classificar as séries de tempo colhidas à luz da teoria de análise de séries temporais, para então construir modelos de ajuste e previsão destas séries.

Esses tipos de modelos oferecem resultados com importantes aplicações práticas. Em âmbito técnico, por exemplo, a inserção de novas unidades fotovoltaicas apresenta risco de perdas técnicas e sobrecargas. Sendo elementos de geração de energia, o posicionamento dessas unidades de geração ao longo da rede de distribuição insere diversos fluxos de potência em vários nós da malha, resultando em vários impactos de nível de tensão, corrente e perdas técnicas, que podem inclusive contribuir para efeitos nocivos ao sistema elétrico se não forem previamente simulados e dimensionados de forma apropriada. Nesse contexto, uma previsão acertada da tecnologia fotovoltaica e, conseqüentemente, dos níveis de penetração fotovoltaica em um sistema elétrico, permite modelar cenários futuros mais seguros para fins de planejamento técnico.

Projeções robustas da penetração da tecnologia fotovoltaica também possuem importância estratégica a respeito de políticas tarifárias futuras. Como exemplo prático, pode-se citar as recentes discussões a respeito da revisão da Resolução 482/2012 -ANEEL, que introduz a compensação de crédito de energia vinda da geração distribuída por meio do sistema de *net metering*. Uma das principais discussões a respeito da validade da REN 482/2012 é sobre o fato da conta de energia recair de forma integral sobre todas as componentes do cálculo tarifário, o que, segundo projeções de potência instalada da ANEEL, resultaria em um ônus de utilização de rede de até 68 bilhões de reais a longo prazo[6].

Tal projeção fez a ANEEL propor novos cenários de compensação de crédito que abatessem menos componentes tarifárias, cenários que resultam em medidas de payback menos atrativas para prováveis adotantes de energia fotovoltaica. Dada a variedade dos agentes de

mercado, podem-se aplicar estratégias de análise e previsão de parâmetros desse mercado, como número de unidades consumidoras adotantes de geração distribuída e potência instalada total, em uma determinada região. Tais procedimentos são importantes para avaliar cenários futuros de ônus tarifário sobre consumidores de um determinado sistema elétrico.

1.3 OBJETIVOS

Assim, o trabalho teve como objetivo a elaboração de análises das séries de tempo relacionadas ao nível de crescimento da geração fotovoltaica em âmbito nacional, para estudar os modelos pesquisados e gerar a descrição e classificação das séries de tempo. A partir desses dados preliminares, o código é capaz de gerar os modelos econométricos mais adequados para o ajuste das séries de tempo, dentro dos modelos teóricos apresentados anteriormente.

Os modelos gerados são, então, utilizados para realizar ajustes das séries de tempo, e consequentemente, obtém-se curvas de previsão dos níveis de crescimento fotovoltaico para cada categoria de instalação fotovoltaica desejada. Cada curva possui seus respectivos intervalos de confiança, para o horizonte de 1 ano a partir da data da última observação (a partir de junho de 2018) onde o modelo econométrico com os melhores parâmetros de erro estatístico tem suas previsões escolhidas como resultado definitivo do trabalho comparativo.

2 FUNDAMENTAÇÃO TEÓRICA E METODOLOGIA

2.1 ANÁLISE DE SÉRIES TEMPORAIS E PROCESSOS ESTOCÁSTICOS

Processos estocásticos[7] são processos definidos por um conjunto de variáveis aleatórias tomadas ao longo do tempo, e que evoluem tanto em função da passagem do tempo, como do intervalo do espaço amostral definido para o processo estacionário. Matematicamente, define-se um determinado processo estocástico P como:

$$P = \{X(t, \xi) | t \in [a; b]\} \quad (2.1)$$

Onde $X(t, \xi)$ é chamada de variável aleatória do processo estocástico, e pode ser entendida como alguma grandeza que será amostrada em diferentes instantes de t.

De acordo com Shmueli e Lichtendall Jr. [8], séries de tempo podem ser decompostas em quatro componentes principais:

- Componente de Nível;
- Componente de Tendência;
- Componente de Sazonalidade;
- Componente de Erro Aleatório ou Ruído;

A componente de nível diz respeito ao valor médio de uma determinada observação. A componente de tendência diz respeito à variação de valor entre duas observações diferentes, e dita tendências mais gerais e persistentes de crescimento ou decrescimento da série; a componente de sazonalidade também reflete padrões de crescimento e decrescimento, mas de caráter cíclico ou periódico, e costuma representar um padrão de curto prazo. Por fim, a componente de erro aleatório representa a incerteza associada à medida de cada observação, e é inerente a qualquer processo estocástico. Tipicamente, esta componente de ruído é tomada como seguindo uma distribuição Gaussiana.

Quantificando as componentes em cada observação da série de tempo, costuma-se definir a composição das séries de tempo de duas maneiras diferentes. Seja uma série de tempo com-

posta por observações y_t , e μ_t , C_{cyclic_t} , C_{trend_t} e ε_t as componentes nível, cíclica, tendência e ruído, respectivamente. Uma série de tempo de composição aditiva pode ser definida como:

$$y_t = \mu_t + C_{cyclic_t} + C_{trend_t} + \varepsilon_t \quad (2.2)$$

Em situações específicas, pode-se definir uma série de tempo multiplicativa por:

$$y_t = \mu_t * C_{cyclic_t} * C_{trend_t} * \varepsilon_t \quad (2.3)$$

Nem todas as séries de tempo possuem componentes sazonais ou de tendência, mas em sistemas reais, todas as séries exibem componente de nível e de erro aleatório. Dito isso, as técnicas de previsão de séries de tempo têm o objetivo, de uma forma ou de outra, de identificar a presença ou ausência dessas componentes em uma determinada série, e feito isso, identificar a valoração de cada uma dos componentes no sistema. Alternativamente de acordo com Moreira *et. al*[9], alguns autores separam a sazonalidade em componentes de ciclicidade e sazonalidade, para oscilações com períodos superiores a um ano e inferiores a um ano, respectivamente.

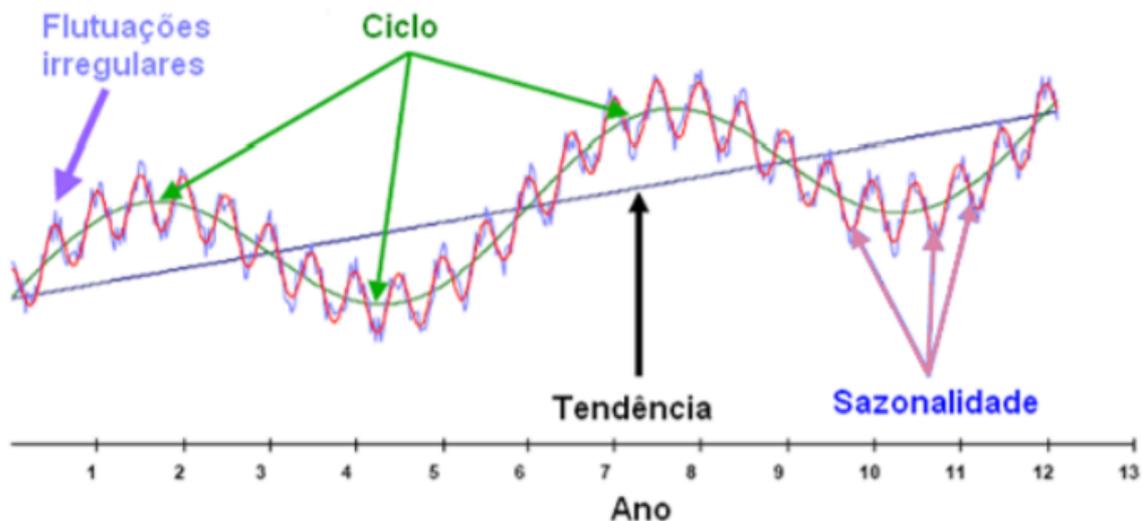


Figura 2.1 – Componentes de Séries de Tempo, Moreira *et. al*.

2.1.1 Autocorrelação e Autocorrelação Parcial

Análoga à definição de correlação entre duas variáveis aleatórias distintas, a definição de autocorrelação para séries de tempo é dada, segundo Hyndman e Athanasopoulos [10], por:

$$ACF(Y_t, k) = Corr(Y_t, Y_{t-k}) = \frac{\sum_{n=k+1}^T (y_n - \bar{y})(y_{n-k} - \bar{y})}{\sum_{n=1}^T (y_n - \bar{y})^2} \quad (2.4)$$

Onde y_t é uma observação da série de tempo no instante t , T é o comprimento da série de tempo em número de observações \bar{y} é o valor médio das observações da série de tempo, k indica o índice de defasagem, ou *lag*. Conceitualmente, a autocorrelação indica o grau de correlação e dependência linear da série de tempo com uma versão deslocada dela mesma no tempo, em k períodos de amostragem.

Para a modelagem de séries de tempo, é importante também a obtenção do grau de correlação entre uma observação num tempo t e um tempo $t + k$, mas excluindo a influência das observações intermediárias $y_{t+1}, y_{t+2} + \dots + y_{t+k-1}$; a esta extensão da autocorrelação dá-se o nome de autocorrelação parcial:

$$PACF(Y_t, k) = Corr(Y_t, Y_{t-k} | y_{t+1}, y_{t+2}, y_{t+3}, \dots, y_{t+k-1}) \quad (2.5)$$

Estas duas funções podem ser posteriormente exploradas em um gráfico chamado autocorrelograma, que indica o grau de correlação (variando entre -1 e 1) da série temporal, em função da defasagem k . Os padrões encontrados em autocorrelogramas fornecem indícios importantes a respeito de quais componentes estão presentes na formação da série de tempo[10]. De forma geral, temos que:

- Séries temporais com uma forte componente de tendência exibem valores semelhantes para pontos de amostra adjacentes ou próximos entre si. Isso se traduz em autocorrelogramas com valores altos de correlação para cada defasagem k , que diminuem gradativamente à medida que k aumenta.
- Séries com forte componente de sazonalidade exibem um perfil cíclico no autocorrelograma: as covariâncias exibem um valor máximo nos valores de k referentes ao período de oscilação da sazonalidade, diminuem até um mínimo e voltam a aumentar, atingindo o valor máximo no próximo k múltiplo do período, analogamente a um perfil senoidal.

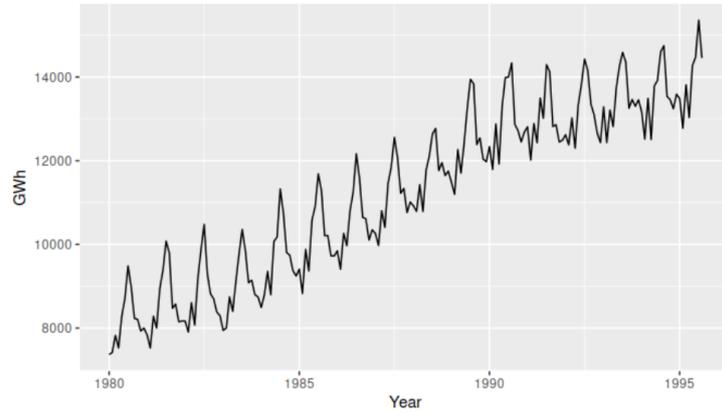


Figura 2.2 – Demanda mensal de energia na Austrália(1980-1995), Hyndman e Athasano-
poulos

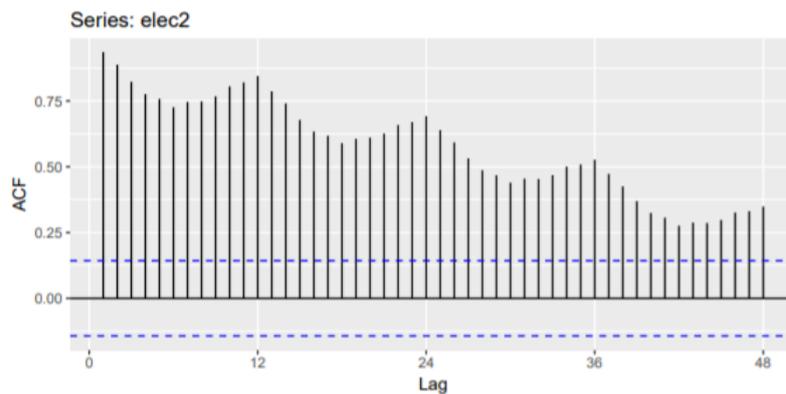


Figura 2.3 – Autocorrelograma para a série temporal acima, Hyndman e Athasanopoulos.
Nota-se a junção dos efeitos de sazonalidade e tendência crescente.

2.1.2 Estacionariedade de Séries Temporais

De acordo com Morettin e Tolo [11], consideram-se classicamente um conjunto de hipóteses simplificadoras no tratamento de séries de tempo como processos estocásticos, que resultam na classificação dos processos estocásticos em três categorias:

- Processos estacionários e não-estacionários;
- Processos Gaussianos ou não-Gaussianos;
- Processos Markovianos ou não-Markovianos;

Concernando a primeira categoria, definem-se como processos estacionários como os processos estocásticos $\{Z(t)\}$ cujas distribuições estatísticas conjuntas(e com isso, suas

características estatísticas não se alteram para alguma realização futura $\{Z(t+\tau)\}$, deslocada para algum instante τ no futuro.

Dentro dos processos estacionários, definem-se processos fracamente estacionários como os processos cujas funções de autocovariância r_k e valor médio $E[Z_t]$ não se alteram ao longo do tempo [12]. Posteriormente neste trabalho, discutem-se estas duas últimas condições serem pressupostos iniciais para análise de séries de tempo.

2.1.3 Testes de Hipótese para Estacionariedade

A estacionariedade da série de tempo é uma condição importante para a construção dos modelos estudados neste trabalho, visto que a maioria destes modelos prevê a definição de um processo estacionário nas suas equações. Aqui serão apresentados dois testes de hipótese nula frequentemente utilizados em Econometria para averiguar as probabilidades de estacionariedade: o teste de Dicker-Fulley aumentado e o teste KPSS.

2.1.3.1 Teste KPSS

Para a realização do teste KPSS, Kwiatkowski *et.al* [13] propõem um teste de hipótese tal que a hipótese nula é a estacionariedade da série (ou uma estacionariedade de tendência), com a não-estacionariedade da série como hipótese alternativa.

A estatística de teste assume que a série de tempo pode ser decomposta em uma equação linear dada por:

$$y_t = \xi t + r_t + \epsilon_t \quad (2.6)$$

Onde ξt é uma componente de tendência determinística, dependente do instante t ; ϵ_t é um erro aleatório, e r_t é um termo de passeio aleatório, dado por:

$$r_t = r_{t-1} + u_t \quad (2.7)$$

Onde $u_t (0; \sigma_u)$. A estatística utilizada no método é chamada LM, e é dada por:

$$LM = \sum_{t=1}^T \frac{S_t^2}{\sigma^2} \quad (2.8)$$

Na equação acima, σ_ϵ é a variância do erro aleatório ϵ obtida pela regressão da série de tempo para o ajuste 2.6, e S_t é a soma de resíduos parcial relativa a essa mesma regressão,

definida por:

$$S_t = \sum_{i=1}^T e_i \quad (2.9)$$

A estatística calculada é utilizada como entrada em duas distribuições probabilísticas específicas para cálculo dos valores críticos para avaliação das hipóteses: η_μ e η_τ . A primeira é utilizada para uma avaliação de estacionariedade ao redor de um nível médio da série de tempo, enquanto a segunda é utilizada para uma avaliação ao redor de uma tendência linear. Ambas distribuições são derivadas, resumidamente, do produto da variância dos erros σ por pontes brownianas.

Nível de Significância (α)	10%	5%	2,5%	1%
Valor crítico (η_μ)	0,347	0,463	0,574	0,739
Valor crítico (η_τ)	0,119	0,146	0,176	0,216

Tabela 2.1 – Valores críticos de cauda superior para o teste KPSS, Kwiatkowski *et al.*

2.1.3.2 Teste de Dicker-Fulley Aumentado

Diferentemente do teste KPSS, o teste de Dicker-Fulley[14] aumentado utiliza como hipótese nula a não estacionariedade da série de tempo, enquanto usa como hipóteses alternativas a estacionariedade ou a estacionariedade de tendência. Este teste utiliza uma regressão semelhante à do teste KPSS, que é definida por:

$$\Delta y_t = \alpha + \beta t + \gamma_{t-1} + \sum_{i=1}^{p-1} \sigma_i \Delta y_{t-i} + \epsilon_t \quad (2.10)$$

Partindo do pressuposto de que a hipótese nula prevê $\gamma = 0$, a estatística de teste utilizada é calculada com o parâmetro linear estimado γ_{t-1} e o erro padrão $SE(\hat{\gamma})$ associado ao parâmetro:

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (2.11)$$

Essa regressão permite três modos dependendo das definições pré-estabelecidas: o modo $\alpha \neq 0; \beta \neq 0$ modela uma série de tempo com valores médios(ou interceptos) e componentes de tendência; o modo $\alpha = 0$ realiza uma regressão linear sem um componente de valor médio, e o modo $\beta = 0$ realiza uma regressão sem componente de tendência.

Além disso, é necessário também determinar a ordem de $\sum_{i=1}^{p-1} \sigma_i \Delta y_{t-i}$, definindo-se previamente o valor p . Uma abordagem possível para essa determinação, proposta por Ng e Perron[15], é supor um valor p máximo para iniciar o processo de teste. Se o valor da esta-

tística de teste tiver módulo maior que 1,6, prossegue-se com o cálculo. Em caso negativo, diminui-se o valor de p e repete-se a avaliação. O valor da estatística leva à não rejeição da hipótese nula se for maior do que algum valor crítico alvo.

Tabela 2.2 – Tabela de valores críticos para o Teste ADF.

Tamanho da Amostra	Sem Tendência		Com Tendência	
	1%	5%	1%	5%
25	-3,75	-3,00	-4,38	-3,60
50	-3,58	-2,93	-4,15	-3,50
100	-3,51	-2,89	-4,04	-3,45
250	-3,46	-2,88	-3,99	-3,43
500	-3,44	-2,87	-3,98	-3,42
Infinitas Amostras	-3,43	-2,86	-3,96	-3,41

Fonte: Dickey e Fulley [14].

Uma equação para a seleção do primeiro valor p é fornecida por Schwert [16]:

$$p = 12 \left(\frac{T}{100} \right)^{0,25} \quad (2.12)$$

Onde T é o número de amostras da série de tempo.

2.2 MODELOS DE REGRESSÃO NÃO-LINEAR

Em muitos problemas estatísticos, o objetivo principal é encontrar alguma relação entre conjuntos diferentes de variáveis aleatórias. De acordo com Seber e Wild[17], em alguns casos específicos é possível definir uma das variáveis aleatórias como um variável dependente y , e um outro conjunto de variáveis aleatórias $x_1, x_2, x_3, x_4, \dots$, chamadas variáveis independentes, que podem ajudar a explicar o comportamento de y em função das variáveis dependentes:

$$y \approx f(x_1, x_2, x_3, x_4, \dots) \quad (2.13)$$

Dependendo da natureza do sistema, a relação acima é conhecida analiticamente, e trata-se de alguma equação que expressa a variável y como uma combinação não-linear das variáveis independentes; tais equações geralmente são provenientes de definições de diversas áreas de estudo[18], e, portanto, costumam envolver alguma quantidade de parâmetros constantes. Para uma série de observações de variáveis independentes, esse conjunto de parâme-

tros é inicialmente desconhecido, e pode ser representado por um vetor θ :

$$y \approx f(x_1, x_2, x_3, x_4, \dots; \theta) \quad (2.14)$$

Define-se como análise de regressão não-linear o processo de relacionar as variáveis independentes e dependentes observadas na série de tempo, ajustando esta série à curva característica desta equação. Assim, com base na hipótese de que o sistema seja regido em certo nível pela equação analítica conhecida, o problema reduz-se a encontrar o vetor desconhecido θ .

2.2.1 Estimação por Mínimos Quadrados Não-Lineares

O método de estimação por mínimos quadrados não-lineares parte do ponto de que o conjunto de parâmetros θ que melhor satisfaz o ajuste de curva é aquele que minimiza a soma de quadrados do erro $S(\theta)$:

$$S(\theta) = \sum_{n=1}^n (y_i - f(x_i; \theta))^2 \quad (2.15)$$

Se considerarmos $f(x_1, x_2, x_3, x_4, \dots, \theta)$ como diferenciáveis em respeito a θ , tem-se a condição para minimização:

$$\left. \frac{\partial \sum_{n=1}^n (y_i - f(x_i; \theta))^2}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0 \quad (2.16)$$

Com as duas equações acima pode-se concluir:

$$\sum_{n=1}^n \{y_i - f(x_i; \theta)\} \left. \frac{\partial (f(x_i; \theta))}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0 \quad (2.17)$$

A equação 2.17 geralmente não possui soluções analíticas para $\hat{\theta}$, e precisa ser resolvida por métodos iterativos.

2.2.2 Modelo de Difusão de Bass

O modelo atualmente utilizado pela ANEEL para projeção de crescimento fotovoltaico foi proposto por Frank Bass em 1969[19], e se baseia na teoria de difusão de produtos de Rogers[20]. Consolidado como um dos trabalhos mais referenciados dentro da área de projeções de mercado, já foi objeto de estudo de uma ampla gama de artigos de pesquisa. Primeiramente, assume-se que a difusão de um determinado produto ou bem de consumo apresenta diferentes tipos de compradores ao longo do tempo:

- Inovadores: iniciam o processo de difusão adotando o produto sem pressões externas iniciais. Geralmente são considerados indivíduos com afinidade ao risco e à incerteza;
- Imitadores: permeiam o processo de difusão da tecnologia desde momentos iniciais até os momentos de consolidação final da tecnologia. Diferentemente dos inovadores, são indivíduos que sofrem influência externa, e tendem a comprar o produto se mais pessoas ao redor também resolveram comprá-lo anteriormente;

Com base nessa teoria, propôs-se uma equação diferencial que descrevesse, de forma matemática, tal cenário de difusão:

$$\frac{f(t)}{1 - F(t)} = p + \frac{q}{m} F(t) \quad (2.18)$$

A partir de uma difusão iniciada em $t=0$, dado em anos, $F(t)$ é número total de adoções até um instante t , e $f(t)$ é a taxa de crescimento das novas adoções, ou seja, $\frac{d}{dt}F(t)$. p, q e m são parâmetros constantes, e são denominados respectivamente fator de inovação, fator de imitação, e mercado potencial final.

Manipulando-se a equação acima e definindo a condição de contorno $F(0) = 0$, podemos chegar às soluções analíticas tanto para a taxa de novas adoções $s(t) = mf(t)$ quanto para o número cumulativo de adoções $S(t) = mF(t)$:

$$s(t) = mf(t) = m \frac{d}{dt}F(t) = m \frac{(p+q)^2}{q} \frac{1 - e^{-(p+q)t}}{\left(1 + \frac{q}{p}e^{-(p+q)t}\right)^2} \quad (2.19)$$

$$S(t) = mF(t) = m \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} \quad (2.20)$$

Para uma série histórica de vendas de um determinado produto, os parâmetros m, p e q podem ser obtidos por meio de métodos iterativos de análise de regressão, como o método dos mínimos quadrados ordinários(OLS) e o método dos mínimos quadrados não-linear(NLS); ambos os métodos fornecem o conjunto de parâmetros que minimiza o somatório de mínimos quadrados do modelo, ou seja, a soma dos quadrados dos resíduos entre as observações reais e os pontos fornecidos pela curva com os parâmetros ajustados.

2.3 MODELOS DE SUAVIZAÇÃO EXPONENCIAL

Seja uma série de tempo qualquer, com $t - 1$ observações. Uma previsão do valor da série no instante t pode ser expressa como \hat{y}_t . Segundo Hyndman *et al.*[21](apud Brown,

1959), um modelo de suavização exponencial supõe que uma previsão do valor da série no instante $t + 1$ pode ser dada pela soma entre a previsão anterior e o erro da previsão anterior:

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t) \quad (2.21)$$

Onde α é uma constante pré-determinada, $0 < \alpha < 1$.

Manipulando e expandindo esta expressão, conclui-se também que a previsão para o instante $t + 1$ é igual a $\hat{y}_{t+1} = \alpha[y_t + (1-\alpha)y_{t-1} + (1-\alpha)^2y_{t-2} + (1-\alpha)^3y_{t-3} + \dots + (1-\alpha)^t\hat{y}_1]$. Observando a equação, percebe-se que este modelo é uma média móvel que pondera cada observação passada com um peso que decresce exponencialmente. Dessa forma, quanto mais distante no tempo está a observação, menos ela influencia no futuro. Outra forma de definir o método de suavização exponencial seria:

$$l_t = \begin{cases} x_0, & t = 0 \\ \alpha x_t + (1 - \alpha)l_{t-1}, & t = 1, 2, 3, \dots \end{cases} \quad (2.22)$$

Onde l_t é um valor da série suavizada, x_t é uma observação da série de tempo original no instante t . No que diz respeito à decomposição de séries de tempo, a série suavizada corresponde a um ajuste da componente de nível.

Hyndman *et al.*[21](apud Holt,1957) também introduz a componente de tendência para previsão por suavização exponencial:

$$l_t = \alpha x_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (2.23)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (2.24)$$

$$\hat{y}_{t+h} = l_t + bh \quad (2.25)$$

β também é uma constante pré-determinada, $0 < \beta < 1$.

Gardner, Jr. e McKenzie[22] propuseram também o método amortecido, que projeta uma curva que diminui gradualmente sua declividade ao longo do tempo, o que pode aumentar a qualidade preditiva no caso de horizontes de previsão mais extensos:

$$l_t = \alpha x_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1}) \quad (2.26)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)\phi b_{t-1} \quad (2.27)$$

$$\hat{y}_{t+h} = l_t + \left(\sum_{i=1}^h \phi^i \right) h \quad (2.28)$$

ϕ deve ser uma constante tal que $0 < \phi \leq 1$, tanto para evitar a inclusão de termos negativos no ajuste, quanto para não introduzir uma componente de crescimento exponencial na equação.

2.4 MODELOS AUTORREGRESSIVOS

Um processo estocástico puramente autorregressivo é aquele que pode ser descrito como uma combinação linear de um número de observações anteriores do processo, acrescidas de um erro aleatório ϵ_t :

$$z_t = \mu + \sum_{i=1}^p \varphi_i z_{t-i} + \epsilon_t \quad (2.29)$$

Os índices φ_i são parâmetros constantes, μ é uma constante de valor médio do processo, e o número p indica quantas observações imediatamente anteriores serão utilizadas na soma ponderada, identificando o grau $AR(p)$ do processo. Uma notação bastante utilizada nos estudos de Econometria em séries de tempo é o operador de defasagem:

$$B^n z_t = z_{t-n} \quad (2.30)$$

De forma que o processo também pode ser definido como:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_k B^p) \bar{z}_t = \epsilon_t \quad (2.31)$$

Onde $\bar{z}_t = z_t - \mu$. Definindo $(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_k B^p) = \phi(B)$:

$$\phi(B) \bar{z}_t = \epsilon_t \quad (2.32)$$

2.5 MODELOS DE MÉDIA MÓVEL

Modelos puros de média móvel, por sua vez, são aqueles definidos como uma combinação linear de erros aleatórios (ou ruídos) provenientes de observações anteriores:

$$z_t = \mu + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t \quad (2.33)$$

De forma semelhante, podemos expressar esse tipo de modelo como:

$$\bar{z}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_k B^q) \epsilon_t \quad (2.34)$$

Fazendo $(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_k B^q) = \theta(B)$:

$$\bar{z}_t = \theta(B) \epsilon_t \quad (2.35)$$

2.6 MODELOS MISTOS

Certos fenômenos estocásticos, se modelados mediante somente um dos dois modelos acima, resultam em uma descrição com um excesso de parâmetros de ajuste, o que fere o princípio de Parcimônia, que prevê a utilização de uma quantidade mínima de parâmetros para a descrição de um sistema. Tal cenário pode acabar comprometendo a qualidade de previsão do modelo com uma soma de quadrados do erro, como demonstrado por Ledolter e Abraham [23]. Assim, propõe-se a classe de modelos mistos autogressivos de média móvel (ARMA) [24] para minimizar a complexidade de modelagem de certos processos aleatórios.

Os modelos ARMA são definidos pela junção direta de modelos autorregressivos e modelos de média móvel:

$$z_t = \mu + \sum_{i=1}^p \varphi_i z_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t \quad (2.36)$$

Os modelos ARMA são definidos pelos seus parâmetros p e q , sendo identificados pelos índices $ARMA(p, q)$. Utilizando a notação compacta:

$$\phi(B) \bar{z}_t = \theta(B) \epsilon_t \quad (2.37)$$

Por questões de convergência do processo, estes últimos processos lineares geralmente satisfazem as condições de estacionariedade se os polinômios $\phi(B)$ e $\theta(B)$ não possuírem raízes localizadas fora de um círculo unitário no plano complexo. Em muitas situações, estes polinômios podem apresentar raízes exatamente sobre o círculo unitário, onde dizemos que o sistema possui uma ou mais raízes unitárias, e é não-estacionário.

2.6.1 Diferenciação

Uma operação comumente realizada em séries de tempo para tentar remover a não-estacionariedade de uma série de tempo é a diferenciação. Uma diferenciação de primeira ordem é a obtenção de uma série de tempo y_t a partir da subtração entre elementos adjacentes de uma série de tempo:

$$y'_t = y_t - y_{t-1} = (1 - B)y_t \quad (2.38)$$

Uma diferenciação de segunda ordem é dada por:

$$y''_t = y'_t - y'_{t-1} = y_t - y_{t-1} - y_{t-1} + y_{t-2} = (1 - B)^2 y_t \quad (2.39)$$

Por fim, uma diferenciação de ordem d é denotada pelo operador:

$$\nabla^d y_t = (1 - B)^d y_t \quad (2.40)$$

Como nota-se pelas equações, uma série diferenciada d vezes possuirá $t - d$ observações, em relação à série original.

2.6.2 Modelos ARIMA

Uma classe mais geral de modelos estocásticos lineares são os modelos autoregressivos integrados de média móvel. Eles são derivados ao se diferenciar d vezes um processo não-estacionário z_t até que ele se estabilize; a série diferenciada pode ser ajustada como se fosse um modelo $ARMA(p, q)$ padrão, e então o ajuste final é obtido pela reversão das operações de diferenciação sobre este modelo $ARMA$. Assim, temos um modelo $ARIMA(p, q, d)$, denotado por:

$$\phi(B)\nabla^d \bar{z}_t = \theta(B)\epsilon_t \quad (2.41)$$

2.7 MODELOS DE ESPAÇO DE ESTADOS

Modelos de espaço de estado univariados são um formato mais geral de sistemas estocásticos lineares, definidos por Commandeur *et.al* [25] pelo sistema linear:

$$y_t = \mathbf{A}\alpha_t + \epsilon_t \quad (2.42)$$

$$\alpha_{t+1} = \mathbf{B}\alpha_t + \zeta_t \quad (2.43)$$

A equação (3.42) é chamada equação de observação, onde a observação y_t do sistema no instante é função do nível α_t e do ruído de observação $\epsilon_t \sim N(0, \sigma_\epsilon)$. A equação(3.43), chamada equação de estado, define um estado futuro de α_{t+1} como uma combinação linear do estado anterior α_t e do ruído de estado $\zeta_t \sim N(0, \sigma_\zeta)$. Dessa forma, a equação (3.43) informa a natureza do processo oculto que governa o sistema estocástico.

No sistema acima, \mathbf{A}, \mathbf{B} , são parâmetros constantes ou variáveis, geralmente pré-determinados de acordo com a categoria de sistemas a serem modelados. Esse tipo de modelo é capaz de generalizar os outros modelos discutidos aqui, permitindo uma representação de uma ampla gama de sistemas estocásticos lineares, incluindo os métodos de suavização exponencial, como demonstrado por Hyndman *et.al* e Box e Jenkins[24].

2.8 MODELOS ESTRUTURAIS E O FILTRO DE KALMAN

O Filtro de Kalman é um estimador recursivo para a obtenção de estimativas à respeito do nível de um sistema em um determinado instante, bem como de outros parâmetros do sistema. Seja um modelo de espaço de estados definido por:

$$x_{t+1} = x_t + \nu_t \quad (2.44)$$

$$z_t = x_t + \omega_t \quad (2.45)$$

A essa classe mais simples de modelo de espaço de estados, com as constantes de sistema iguais a 1, dá-se o nome de modelos estruturais. Os parâmetros ν_t e ω_t são ruídos gaussianos com variâncias respectivas Q e R . Neste algoritmo, são interpretadas respectivamente como as incertezas na medição da grandeza e na previsão da grandeza para momentos futuros. Desses parâmetros, obtêm-se a equação de atualização da matriz de covariância do sistema,

dada por:

$$P_t = P_{t-1} + Q \quad (2.46)$$

Com essas condições iniciais, uma primeira etapa de previsão fez previsões *a priori*(-) do nível do sistema e da covariância do estado, com as equações recursivas:

$$\hat{x}_k^- = \hat{x}_{k-1}^+ \quad (2.47)$$

$$P_k^- = \hat{P}_{k-1}^+ + Q \quad (2.48)$$

O sobrescrito (-) diz respeito a um valor obtido *a priori*, e (+) diz respeito a um valor obtido *a posteriori*.

Após a previsão *a priori*, a próxima etapa de atualizações começa com a obtenção do erro de medição $\tilde{y}_k = z_k - \hat{x}_k^-$. Em seguida, calcula-se o ganho de Kalman, dado por:

$$K_k = \frac{P_k^-}{P_k^- + R} \quad (2.49)$$

Para a obtenção da nova estimativa(*a posteriori*) de \hat{x}_k^+ , realiza-se a operação:

$$\hat{x}_k^+ = \hat{x}_k^- - K_k \tilde{y}_k \quad (2.50)$$

Assim, o valor atualizado do estado estimado \hat{x}_k é dado por uma acréscimo da diferença de erro ponderado pelo ganho de Kalman. Por fim, é atualizado o valor da estimativa da covariância do estado:

$$\hat{P}_k^+ = (1 - K_k) \hat{P}_k^- \quad (2.51)$$

Assim, os valores calculados de \hat{x}_k e \hat{P}_k *a posteriori* (+) nesta iteração k são a entrada *a priori*(-) da etapa de previsão iteração $k + 1$, reiniciando o algoritmo. A estimativa do nível \hat{x}_k pode ser iterada até uma margem de erro desejada. Para modelos de espaço de estados mais complexos, o algoritmo é generalizado com a inclusão das possíveis constantes **A**, **B**, como exemplificado por Thacker e Lacey[26].

2.8.1 Métricas de Qualidade de Ajuste

2.8.1.1 Critério de Informação de Akaike

O critério de informação de Akaike (AIC) é uma métrica que indica a perda relativa de informação à respeito do sistema ao se ajustar um determinado modelo. O critério é dado pela equação:

$$AIC = 2k - 2\ln\hat{\mathbf{L}} \quad (2.52)$$

Na equação acima, k é o número de parâmetros que o modelo ajustam e $\ln\hat{\mathbf{L}}$ a função de log verossimilhança maximizada. Do ponto de vista qualitativo, o AIC ilustra a qualidade do modelo como uma diferença entre o número de parâmetros do sistema (representando o risco de sobreajuste) e a verossimilhança maximizada do modelo (representando o risco de subajuste).

O valor fornecido pelo AIC não diz respeito a um grau de qualidade absoluto do modelo em questão. Ele deve sempre ser utilizado em uma análise comparativa entre um conjunto de modelos, ainda com a condição necessária de que todos os parâmetros de cada modelo do conjunto tenham sido obtido exatamente dos mesmos pontos da série de tempo. Dessa forma, dentro de um certo conjunto de prováveis modelos de uma mesma série de tempo, o modelo que oferece o menor valor de AIC é o modelo que mais tende a reter informação útil a respeito do sistema estocástico.

2.8.1.2 Teste de Ljung-Box

Quando um modelo estatístico é ajustado a uma série de tempo, ele fornece um valor estimado \hat{y}_t para um ponto real y_t . A diferença $\hat{y}_t - y_t$ é chamada de resíduo. De forma geral, os modelos referenciados neste trabalho estão bem ajustados à série de tempo quando a série de resíduos produzida pelo modelo tem autocorrelação nula, ou seja, não pode ser distinguida de um ruído branco.

Para averiguar se uma série de resíduos possui características de ruído branco, utilizamos o teste de Ljung-Box, cuja estatística é dada por:

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{r}_k}{(n-k)} \quad (2.53)$$

Onde n é o número da amostra, \hat{r}_k a autocorrelação dos resíduos, e h o número de lags utilizados na função de autocovariância. Para $Q > \chi_{1-\alpha, h}^2$, ou seja, Q acima do percentil $1-\alpha$ de uma distribuição Chi-quadrada com h graus de liberdade, a hipótese nula de resíduos não

autocorrelacionados é recusada. Para o escopo deste trabalho, é rejeitado qualquer p-valor menor que 0,05, e são considerados 20 *lags* para a função de autocorrelação.

2.9 CONSIDERAÇÕES FINAIS

Este capítulo apresentou o ferramental teórico relativo aos métodos estatísticos fundamentais discutidos na metodologia e nos resultados deste trabalho, abrangendo principalmente os fundamentos da teoria de regressão estatística, modelagens de sistemas estocásticos lineares e métricas de teste estatístico, áreas que foram importantes na construção da metodologia de previsão proposta.

3

REVISÃO BIBLIOGRÁFICA

Este capítulo expõe os trabalhos mais relevantes no campo da previsão de espalhamento de tecnologias fotovoltaicas, além de trabalhos de econometria de séries de tempo em contextos fora do mercado fotovoltaico, mas que agregam para a construção da metodologia deste trabalho.

3.1 SIMULAÇÃO DO MERCADO BRASILEIRO VIA MODELO DE BASS

O trabalho que motivou o início destes estudos foi a análise de Konzen[27], no qual realizou-se uma análise dos principais fatores socioeconômicos que influenciam no processo de difusão fotovoltaica residencial em nível nacional, e a partir desta análise, quantificar essas influências e apresentar uma simulação de difusão fotovoltaica por meio do método de Bass, acrescido de parametrizações adicionais que refletem a influência daqueles fatores. O intervalo de análise e projeção do mercado foi definido a partir de 2013, decorridos 240 dias da aprovação da REN N° 482/2012, e se estende até 2024.

Assim, Konzen subdivide o território brasileiro em suas áreas geoeletricas de concessão, com base no sistema georeferenciado da ANEEL. Com a utilização de ferramentas do Arc-GIS, dados socioeconômicos do Censo 2010 do IBGE foram cruzados com essas divisões geográficas e associados às regiões das respectivas distribuidoras de energia; os principais fatores quantificados foram:

- Renda Familiar Mensal;
- Quantidade de residências do tipo casa própria;
- Taxa de crescimento do estado;

Com estes três fatores, elaborou-se um fluxograma para calcular o mercado potencial inicial m da curva de difusão de Bass, acrescido de um fator de aptidão igual a 0,8 para representar os efeitos de formatos de telhado na adequabilidade das residências à instalação de módulos fotovoltaicos.

Além destes fatores socioeconômicos, o trabalho também levou em conta a influência do *payback* do investimento na tecnologia fotovoltaica; com base nos estudos de Beck, o

mercado potencial inicial foi multiplicado por uma função de influência do *payback* fmm , de acordo com a expressão:

$$m_{final} = fmm * m = e^{-SPB*TPB(t)} * m \quad (3.1)$$

Onde:

- m é o mercado potencial inicial;
- SPB é o fator de sensibilidade ao *payback*, definido como 0,3
- $TPB(t)$ é o *payback* de investimento fotovoltaico na área de atuação da distribuidora para um determinado ano t .

Por sua vez, o *payback* anual em cada distribuidora foi calculado com base em uma planilha de fluxo de caixa de elaboração própria do trabalho, que levou em conta os seguintes dados:

- As bases tarifárias de cada distribuidora,
- Parâmetros de custo de investimento em módulos solares, por consumidor;
- Parâmetros de produtividade e degradação média de módulos fotovoltaicos;
- Diferentes condições de subsídios fiscais e políticas de incentivo;

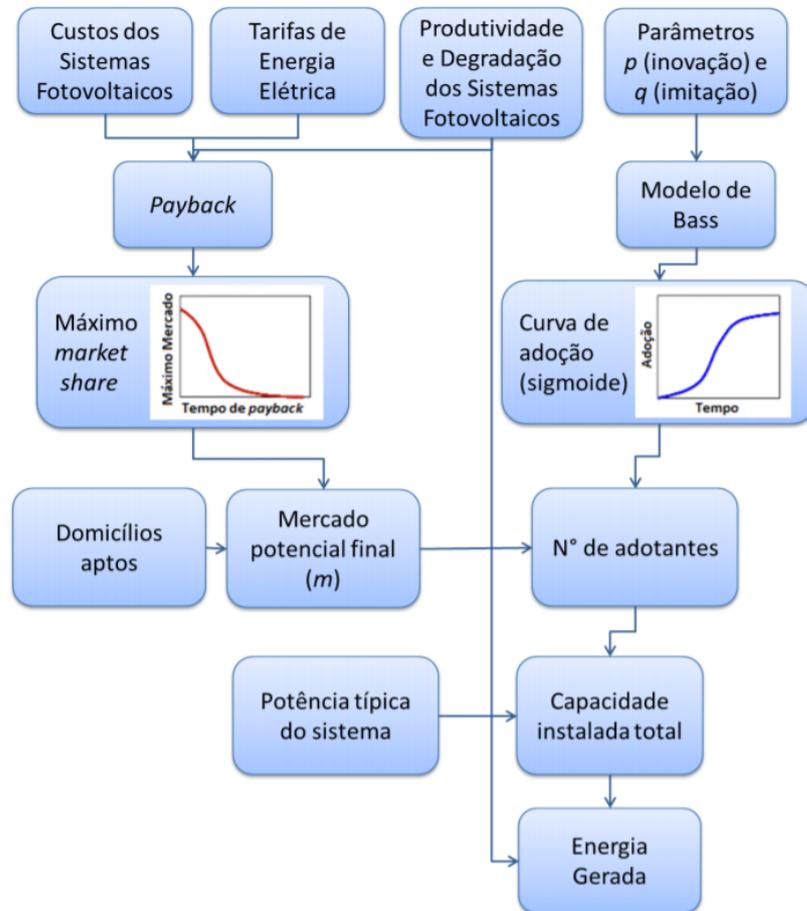
Para a determinação dos cenários tarifários, foram definidas quatro propostas de incentivos:

- Cenário I: Negócio como usual;
- Cenário II: Desconto no Imposto de Renda;
- Cenário III: Isenção fiscal na compra de módulos fotovoltaicos;
- Cenário IV: Convênio ICMS 6(decisão do CONFAZ de cobrança de ICMS sobre geração de energia) não será aplicado pelos estados;
- Cenário V: Junção dos cenários II, III e IV;

O modelo de Bass é finalmente definido pelos fatores p e q , escolhidos por Konzen com base em médias históricas desses parâmetros no mercado norteamericano(no ano do trabalho, a série histórica brasileira não tinha observações o suficiente para captura desses valores).

Assim, foram definidos valores de $p = 0,0015$ e $q = 0,3$ para um modelo de difusão *business as usual*, e do valor de $p = 0,002$ no caso de existência de políticas de desconto no imposto de renda e compensação virtual, (visto que a parcela inovadora dos adotantes se sentiria mais estimulada) bem como o valor de $q = 0,4$ na presença de políticas de isenção fiscal dos módulos fotovoltaicos e implementação do *virtual net metering*, acelerando o processo de adoção.

Figura 3.1 – Fluxograma da metodologia de simulação de mercado.



Fonte: Konzen(2015)

3.2 ANÁLISE ECONOMETRICA DA DIFUSÃO FOTOVOLTAICA

A consolidação de tecnologias econométricas pode ser observada nos resultados obtidos por Nguyen *et al.*[28], em que a modelagem por métodos autorregressivos integrados com média móvel(ARIMA) é aplicado em conjunto com uma nova metodologia baseada em ajustes de série de Taylor. Neste estudo, para modelar cenários de penetração fotovoltaica no território dos Estados Unidos, foi analisada uma série de tempo com 146 observações, que se

estende de janeiro de 2000 a fevereiro de 2012, relativas à quantidade de novas instalações fotovoltaicas em cada ponto observado.

Após a etapa de análise econométrica, a série é ajustada para um determinado modelo ARIMA com componente sazonal, com parâmetros SARIMA(p,q,d)(P,Q,D). Uma vez que todos os parâmetros são obtidos, os resíduos do modelo são avaliados de acordo com estatísticas de Ljung-Box. No entanto, uma nova ferramenta é adicionada ao conjunto de métodos de análise: a sequência de resíduos encontrados no modelo é reajustada dentro de uma série de Fourier. Para tanto, seja o conjunto de resíduos $\{\epsilon_n\} = \epsilon(1), \epsilon(2), \epsilon(3), \dots, \epsilon(k), \dots, \epsilon(n)$, onde $k = \overline{1, n}$ é sempre um número inteiro. Propõe-se um estimador $\hat{\epsilon}(k)$, definido por:

$$\hat{\epsilon}(k) = \frac{1}{2}a_0 + \sum_{i=1}^D \left[a_i \cos\left(\frac{2\pi i}{n-1}k\right) + b_i \text{sen}\left(\frac{2\pi i}{n-1}k\right) \right] \quad (3.2)$$

Onde, $D = n - 1$ é denominado frequência de distribuição mínima da série de Fourier. Os parâmetros $a_0, a_1, a_2, a_3, \dots, a_D, b_1, b_2, b_3, \dots, b_D$, são obtidos por meio da equação acima, pelo método dos mínimos quadrados ordinários. Estes parâmetros são utilizados para a definição da nova série de resíduos, que são reintegrados diretamente ao modelo SARIMA inicial para formar o modelo final FARIMA(p,q,d)(P,Q,D). A qualidade das previsões foi avaliada de acordo com o critério do erro médio absoluto. Os resultados foram comparados aos erros de previsão obtidos se a parametrização residual em série de Fourier não fosse utilizada.

Tabela 3.1 – Erros médios quadráticos e parâmetros finais ajustados.

Modelo	MAPE	S1	S2	C	P	ρ
ARIMA(1,1,1)(1,1,1)	0,0193	2,5526	0,1980	0,0776	1,000	0,9807
FARIMA(1,1,1)(1,1,1)	0,0008	2,5526	0,0131	0,0051	1,000	0,9992

Fonte: Nguyen *et al.*, tradução própria.

Finalmente, o modelo FARIMA é utilizado para prever o montante de potência fotovoltaica nos Estados Unidos nos meses de janeiro e fevereiro de 2012, obtendo um erro médio absoluto inferior a 2%, considerado satisfatório pelas restrições do estudo.

Tabela 3.2 – Previsões de energia instalada.

Mês	Real(Trilhões de BTU)	Previsto(Trilhões de BTU)	MAPE
Jan/2012	15,048	14,857	0,013
Feb/2012	14,596	14,351	0,017

Fonte: Nguyen *et al.*, tradução própria.

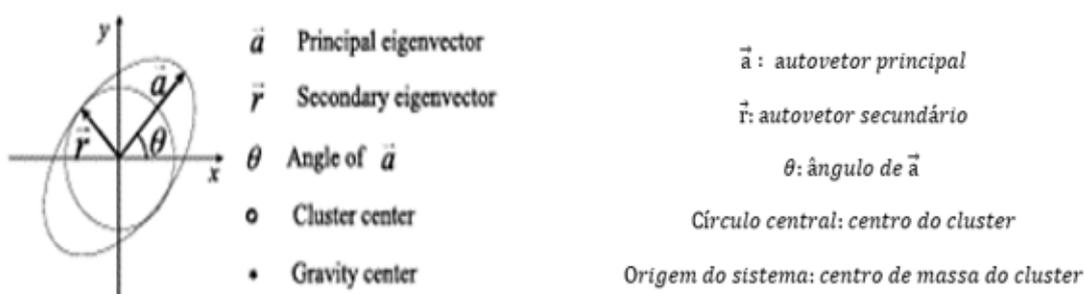
3.3 DIFUSÃO ESPACIAL E TEMPORAL FOTOVOLTAICA

Os trabalhos de Zhao *et al*[29]. propõem uma abordagem de projeção fotovoltaica baseada não somente em uma análise econométrica da difusão de geração distribuída, mas também no estudo de aspectos microscópicos e macroscópicos de difusão espacial de tecnologias. O local escolhido para a análise de difusão fotovoltaica espacial é o distrito de Pudong, pertencente à província de Shanghai.

Os dados temporais do trabalho consistem de uma série histórica de capacidade instalada em base trimestral, que abrange desde o quarto trimestre de 2013 até o terceiro trimestre de 2014. Esta informação temporal é integrada ao conjunto de dados espaciais, que reúne um grande número de informações como distribuições de relevo e tipo de uso da terra; a integração desses dois tipos de dados permitem a geração de um banco de dados espaço-temporal, que associa às distribuições da geração fotovoltaica a características geográficas territoriais, gerando diversas camadas de dados com ênfase em diferentes aspectos da distribuição fotovoltaica. Tal base de dados é arranjada por células, subdivisões da malha do mapa territorial, cada uma trabalhada como um objeto individual contendo seus próprios atributos.

Como unidades básicas de difusão fotovoltaica, o estudo concebe agrupamentos de gerações fotovoltaicas, definidos pelos seus autovetores primário e secundário, que por sua vez, descrevem regiões que indicam tanto o espalhamento das células quanto as prováveis direções de difusão. Os agrupamentos de gerações solares crescem ao longo das células da malha espacial, de forma que as células, ao longo do tempo, podem receber ou não a condição de portadora de geração solar como um atributo, de forma estocástica.

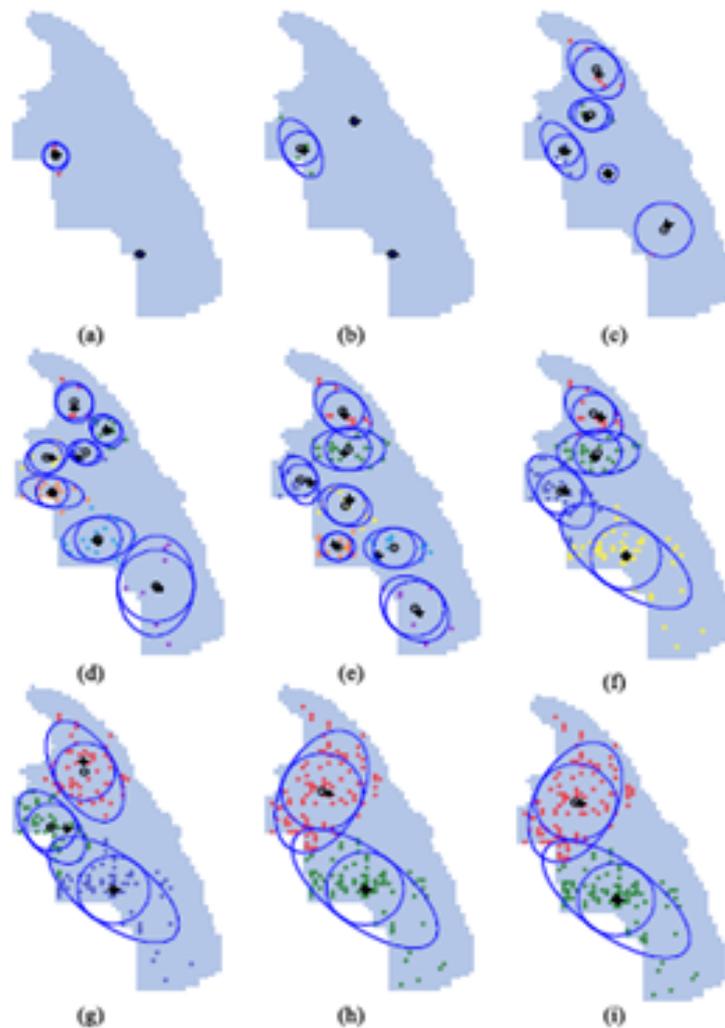
Figura 3.2 – Representação de *cluster* individual.



Fonte: Zhao *et al.*, tradução própria.

Como indicado pela figura acima, os agrupamentos de sistemas fotovoltaicas possuem a capacidade de se mesclarem e formarem agrupamentos maiores, com tendências conjuntas de crescimento numérico e difusão espacial.

Figura 3.3 – Representações gráficas de espalhamento de *clusters*.



Fonte: Zhao *et al.*, tradução própria.

Também são integrados ao banco de dados espacial os fatores internos e externos que influenciam o processo da difusão da tecnologia. Para cada célula espacial, são atribuídas variáveis de influência externa, como existência de políticas governamentais e influências de vizinhança, bem como fatores de influência interna, como tipo de fornecimento energético da célula e tipo de telhado associado à célula espacial.

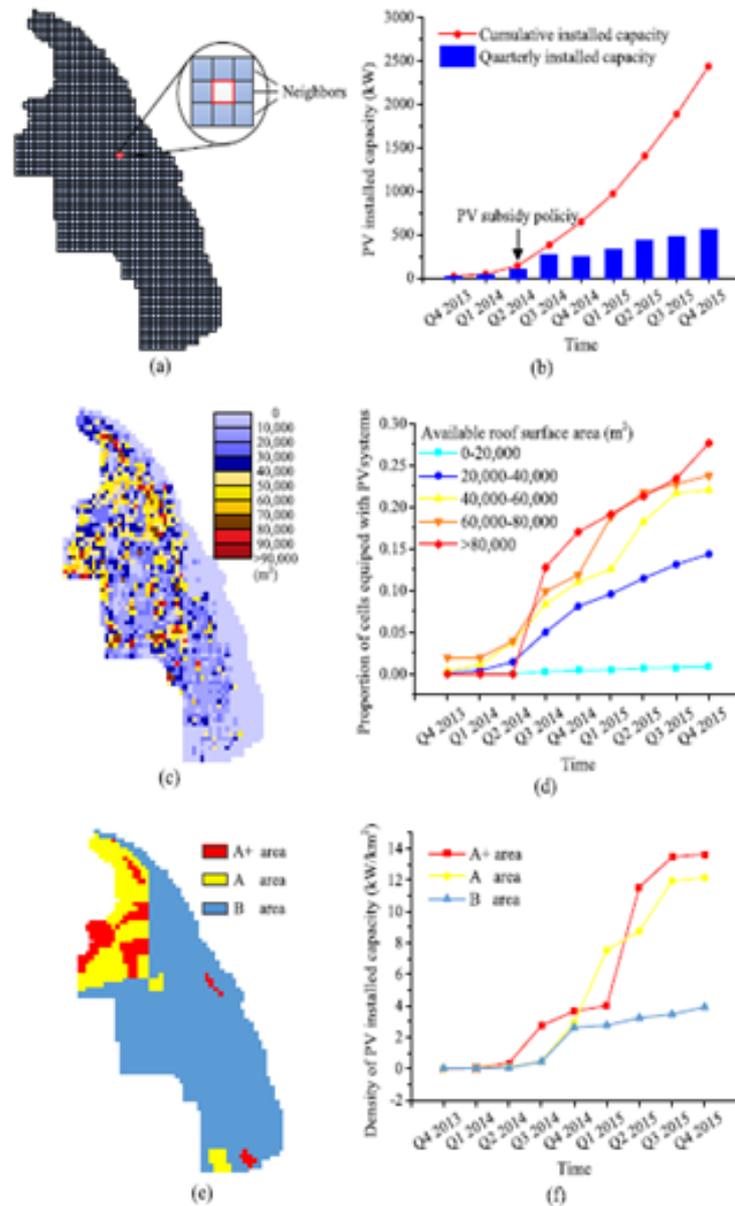


Figura 3.4 – Representação de atributos por célula espacial, Zhao *et al.*

A equação de regressão utilizada para o crescimento da curva é dada por:

$$Y(t) = \frac{e^{r_g t}}{\frac{1}{Y_0} - \frac{1+e^{r_g t}}{Y_{GM}}} \quad (3.3)$$

Onde Y_0 é a capacidade instalada em algum período inicial t_0 , Y_{GM} é a capacidade máxima ajustada pela regressão da curva, e r_g é a taxa de crescimento trimestral da curva.

Após esta projeção temporal inicial, uma certa capacidade de geração solar é projetada para um intervalo futuro definido. Em seguida, utiliza-se uma máquina de vetores de suporte (SVM, *State Vector Machine*) para, a partir do t_0 inicial, estimar os estados futuros das

células espaciais, de forma que as células com mais fatores favoráveis ao mercado fotovoltaico terão uma probabilidade de assumir o estado de adotantes da tecnologia, seguindo a orientação dos autovetores de aglomerados de geração solar.

Por fim, a carga de geração distribuída projetada é associada à previsão dos estados futuros das células espaciais para definir uma densidade de probabilidade das novas instalações fotovoltaicas. A alocação espacial das futuras gerações fotovoltaicas parte de condições de contorno deste denso probabilístico, e é definida com base na otimização das distâncias entre os novos pontos de geração utilizando o método de ponderação por entropia, que por sua vez utiliza os atributos de influência de cada célula como indicadores.

3.4 SIMULAÇÕES DE DIFUSÃO BASEADAS EM ABM

Outra metodologia, mais complexa, mas com grandes potenciais no campo da difusão fotovoltaica, são os modelos baseados em agentes (ABM, *Agent-Based Models*), uma classe de modelos computacionais que busca simular e projetar o comportamento estocástico de sistemas dinâmicos. Nestes sistemas, objetos fundamentais, denominados agentes, são espalhados em uma dada estrutura espacial e temporal e são capazes de interagir entre e si e tomarem decisões, influenciados por atributos intrínsecos a cada agente e pelas interações interagentes, assumindo estados diferentes a cada iteração.

Nesse contexto, Schiera et.al[30], propuseram aplicação de um ABM integrado a uma base de dados GIS, para análise de projeção de tecnologia fotovoltaica em Turim, na Itália. O esquema de simulação foi estruturado em camadas principais, subdivididas por sua vez em diversas subcamadas que operam como microaplicações independentes, capazes de estabelecer um fluxo realimentado de informações com as outras camadas do ambiente de simulação. Nessa estrutura, diversos fatores como dados climáticos, renda, pressões sociais, formato do telhado da casa são integrados como atributos aos agentes consumidores, que num determinado momento podem exibir maior ou menor probabilidade de adotarem a tecnologia fotovoltaica.

As influências interagentes são modeladas de acordo com o modelo de Redes de Mundo Pequeno (SWN), onde os agentes são espacialmente relacionados por um grafo circular, e relações podem ser traçadas entre pontos vizinhos, ou por pontos mais distanciados, mas ainda sim com poucas relações necessárias para a interação. Nesse tipo de grafo, a distância entre dois indivíduos é proporcional ao logaritmo do número de indivíduos que pertencem à rede.

Cada agente, por sua vez, tem seu comportamento modelado pela Teoria do Comportamento Planejado, que supõe que cada decisão tomada pelo indivíduo depende basicamente

de três fatores: atitude frente ao comportamento, normas subjetivas e controle comportamental, sendo que a decisão final será uma combinação linear dessas componentes. De forma simplificada, cada agente possui uma probabilidade de comportamento b dada por:

$$b = W_{bi} * bi + W_{pbc} * pbc \quad (3.4)$$

O valor bi representa a intenção de comportamento, definida pelas convicções individuais e crenças normativas que influenciam o agente a realizar ou não alguma ação. Já o valor pbc representa o nível de controle que o agente supõe ter sobre a decisão, com dificuldades e facilidades percebidas. W_{bi} e W_{pbc} são pesos de ponderação respectivos.

3.5 CONSIDERAÇÕES FINAIS

Este capítulo apresentou, de forma resumida, as metodologias mais elaboradas no contexto de previsão de processos de difusão fotovoltaica. Entre uma variedade considerável de metodologias teóricas e computacionais, o ponto de partida da pesquisa foi, de fato, os trabalhos de Konzen[27], por sua contextualização atrelada ao cenário fotovoltaico brasileiro e suas implicações nas projeções de mercado. Quanto às outras metodologias, o trabalho de Nguyen *et al.* foi uma influência importante, por oferecer um método mais simples e antigo, mas com forte fundamentação teórica e comprovação empírica.

4

METODOLOGIA

4.1 PESQUISA METODOLÓGICA INICIAL

A primeira parte do trabalho de pesquisa foi destinado à compreensão e revisão dos materiais que fornecem informações sobre como são realizadas as projeções dos níveis da matriz fotovoltaica em território nacional. Dessa forma, os estudos começaram na análise das notas técnicas 028/2018 da EPE[31] e 0056/2017 da ANEEL [32]. Ambas as notas técnicas explicitam a aplicação do método de difusão de Bass na projeção a longo prazo da penetração da energia fotovoltaica no Brasil, e têm como referência metodológica principal os trabalhos de Konzen[27], previamente mencionados.

Com o esforço de pesquisa já centrado nesta última dissertação, procurou-se compreender os aspectos teóricos e os pressupostos socioeconômicos utilizados para a aplicação da equação de Bass, bem como para a calibração dos parâmetros do método para ajustá-lo ao mercado fotovoltaico brasileiro. Acompanharam essa etapa de pesquisa algumas implementações simplificadas do modelo de Bass em ambiente MATLAB, para fins de fixação do conteúdo estudado.

Finda a ambientação inicial aos conceitos teóricos, procurou-se diretamente os especialistas em regulação que compõem o grupo de Qualidade da Energia, pertencente à Superintendência de Regulação dos Serviços de Distribuição - ANEEL, na qual o autor deste trabalho atua como estagiário. Os diálogos com os técnicos permitiram uma compreensão mais aprofundada a respeito de como o método de Bass é aplicado especificamente dentro da superintendência, principalmente em termos computacionais.

Ademais, o corpo técnico forneceu informações importantes sobre as desvantagens e limitações do método de Bass. As simulações do modelo apresentaram, ao longo de 2019, discrepâncias significativas em relação aos valores reais, posteriormente atualizados. A taxa de crescimento mensal da energia fotovoltaica superava, de forma sustentada, os valores previamente projetados na planilha de cálculo, mesmo mediante sucessivas alterações dos parâmetros p e q da curva, e mediante atualizações do mercado potencial m , provenientes das atualizações de bases tarifárias das distribuidoras.

Após discussões posteriores com o corpo técnico e com o orientador do trabalho, iniciou-se uma segunda parte do trabalho de pesquisa, agora focada em procurar métodos alternativos que pudessem substituir o modelo de Bass e oferecer previsões com um maior grau de qualidade estatística. As alternativas encontradas foram os métodos de Econometria de séries de tempo: análises fundamentadas no grau de correlação estatística e dependência linear entre várias observações ao longo de um período de tempo, e ao mesmo tempo, independentes de

um modelo explanatório para oferecer previsões acuradas.

As novas propostas foram discutidas com o corpo técnico da SRD-ANEEL, que forneceu a base de dados do sistema SAS(*Statistical Analysis System*), que contém informações constantemente atualizadas sobre toda a potência fotovoltaica instalada em âmbito nacional. Uma visualização prévia das tendências de crescimento da energia solar permitiu a escolha final dos métodos de análise de tempo aplicados neste trabalho, que serão explicitados a seguir.

4.2 ESCOLHA E TRATAMENTO DE DADOS

Seguindo as recomendações próprias do corpo técnico da ANEEL, inicialmente filtraram-se quatro séries históricas do sistema SAS:

- Microgeração Local Residencial;
- Microgeração Local Comercial;
- Minigeração Local Total;
- Minigeração Remota Total;

As quatro séries históricas abrangem o período de janeiro de 2013 a dezembro de 2019, amostradas em base mensal. O início da série histórica foi escolhido mediante a determinação da REN nº 482, que previu um prazo de 240 dias, a partir de sua aprovação, para que as distribuidoras se adequassem completamente à norma.

As séries históricas foram submetidas a uma inspeção basicamente visual, para identificação simplificada de alguns fenômenos como tendências de crescimento e choques estruturais. Esta primeira análise revelou que as séries históricas de minigeração, por serem baseadas em implantações bastante esparsas, configuram um padrão de vendas intermitente adjunto de uma leve componente de tendência de crescimento, um padrão destoante das séries de microgeração, mais comportadas do ponto de vista teórico.

Os padrões de crescimento intermitente suscitam a aplicação de certos métodos econométricos que dificultam a comparação, em termos de qualidade de previsão, com as outras duas séries temporais mais comportadas. Assim, decidiu-se remover a análise das séries de microgeração do escopo do trabalho.

A pesquisa a respeito dos métodos econométricos passíveis de serem utilizados no contexto das séries de tempo apresentadas levou à escolha de três modelos de ajuste e previsão de séries de tempo, cuja implementação será detalhada a seguir.

4.3 METODOLOGIA ESTATÍSTICA

Os modelos de análise estatística escolhidos foram os modelos ARIMA, modelos de suavização amortecida, e modelos estruturais(formas mais simples dos modelos de espaço de estados). Modelagens com componentes sazonais foram desconsideradas por conta da ausência de evidências, tanto no formato da série de tempo, quanto nos fundamentos do mercado fotovoltaico, que este mercado seja marcado por padrões sazonais.

O modelo ARIMA foi escolhido por ser um método mais antigo e consolidado, fortemente embasado por teorias de análise espectral e análise de sistemas lineares, e apresenta uma vasta gama de referências acadêmicas e validações empíricas, apesar de ser mais indicado para janelas de previsão de prazo mais curto e necessitarem de novas análises de parâmetros e calibração manual destes, em caso de novas entradas na série temporal.

O modelo de suavização exponencial com tendência foi escolhido por representar uma versão modificada de um dos modelos de previsão mais fundamentais, que é o método de média móvel. Além disso, Gardner *et.al*[22] apresenta este método como uma solução que já comprovou sua robustez em um grande número de séries de tempo de competições M3. Por fim, sua implementação é simples e flexível, permitindo uma aplicação para uma variedade de configurações de parâmetros, com um ajuste automático à medida que novas entradas são adicionadas nas séries de tempo sob análise.

Por fim, os modelos estruturais constituem um tipo de abordagem teórica mais recente, mas que se destaca por possuir potenciais vantagens em relação aos outros métodos; Weiss[33] afirma que uma grande variedade de modelos de séries de tempo podem ser escritas no formato de espaço de estados(como por exemplo, os próprios modelos ARIMA). Além disso, modelos de espaço de estados são não estacionários, e não precisam de transformações como transformada logarítmica e diferenciação. Por fim, são modelos que conseguem lidar melhor com volatilidades e choques estruturais, e permitem a introdução de variáveis explanatórias no caso de modelos mais complexos.

A partir da escolha dos modelos a serem trabalhados, o processo de análise começou na coleta e visualização das séries de tempo a serem analisadas. Nessa etapa, componentes explicitamente presentes nos dados, como padrões de sazonalidade, componentes de tendência e choques estruturais. Dependendo da natureza destas componentes, a visualização permite a realização de transformações que coloquem a série num formato mais apropriado para o ajuste dos modelos.

Após a análise comparativa entre modelos das mesmas categorias, uma avaliação de qualidade preditiva será realizada entre os melhores ajustes obtidos das três metodologias; as métricas de erro de previsão utilizadas foram o erro médio absoluto percentual(MAPE), o erro médio absoluto(MAE), e a raiz quadrada do erro médio(RMSE). Estas métricas foram

analisadas e obtidas avaliando as métricas médias de erro ao se ajustarem os modelos com 80% das observações de cada série de tempo e testando os modelos para prever os 20% finais da série.

4.3.1 Obtenção dos Modelos de Suavização Exponencial

A presença de uma componente de tendência bem definida na série log-transformada sugere a utilização de ajuste de suavização exponencial modificada para modelar uma componente aditiva de tendência. Além disso, foi ajustado também um modelo com tendência amortecida para comparação, visto que Hyndman e Athanasopoulos[10] afirmam, empiricamente, que a adição de tendências não amortecidas pode resultar em valores de previsão sobreestimados. Por fim, os parâmetros α , β e ϕ não foram pré-definidos, e foram ajustados por maximização da função de verossimilhança e do filtro de Kalman aplicado para o modelo de espaço de estados dos modelos de suavização exponencial.

4.3.2 Obtenção dos Modelos ARIMA

Para os modelos ARIMA, o fluxo de trabalho é semelhante aos indicados por Box e Jenkins [24] e Hyndman e Athanasopoulos [10]; em um ramo da metodologia de escolha, aplicou-se a estrutura padrão de escolha de prováveis modelos de Box-Jenkins com a determinação dos índices $ARIMA(p,d,q)$ com base em análises de autocorrelação e estacionariedade da série de tempo. Em outro ramo, um algoritmo compara diversos modelos ARIMA pré selecionados e escolhe mais outro modelo que minimiza o Critério de Informação de Akaike.

A estrutura de Box - Jenkins é um algoritmo empírico para a seleção de prováveis modelos de ajuste para uma determinada série de tempo, generalizada pelos passos a seguir:

1. Inicialmente, testes de hipótese são realizados para se saber as condições de estacionariedade da série, que exigem que alguma operação de diferenciação seja realizada (com diferenciações prévias tendo sido realizadas ou não). O número de diferenciações necessária para estabilizar a série é o índice d .
2. Uma vez que os testes e as diferenciações dão segurança o suficiente para considerar a série estacionária, extraem-se a autocorrelação e a autocorrelação da série. Os parâmetros p e q são escolhidos com base na observação destes gráficos e da aplicação das regras gerais[34]:

	ACF	PACF
Modelo $AR(p)$	Decaimento geral	Corte abrupto em $k = p$
Modelo $MA(q)$	Corte abrupto em $k = q$	Decaimento geral
Modelo $ARIMA(p, d, q)$	Decai com corte em $k = q$	Decai com corte em $k = p$

3. Definidos um conjunto de modelos prováveis $ARIMA(p_1, d, q_1)$, $ARIMA(p_2, d, q_2)$, $ARIMA(p_3, d, q_3)$,..., são realizados os ajustes para cada modelo acima da série de tempo. Efetua-se um diagnóstico dos resíduos de cada modelo, e computam-se os critérios de informação de cada ajuste. Por fim, são obtidas as previsões dentro da janela temporal da série de testes, e são computados os erros de previsão de cada modelo. Em ambas metodologias, os parâmetros dos modelos são estimados por máxima verossimilhança.

No outro ramo, a série de tempo estacionária é submetida ao algoritmo de minimização do critério de Akaike proposto por Hyndman e Khandakhar[35], que realiza uma série de modelos de forma automática, selecionando o que obtiver o menor AIC. De forma resumida, o algoritmo é composto das etapas:

1. Um teste KPSS é automaticamente realizado para avaliação do índice d ;
2. Parâmetros p e q são testados na sequência:
 - (a) $ARIMA(0, d, 0)$;
 - (b) $ARIMA(2, d, 2)$;
 - (c) $ARIMA(1, d, 0)$;
 - (d) $ARIMA(0, d, 1)$;
3. O modelo com o menor AIC é pré-selecionado como o mais bem ajustado;
4. Variações de ± 1 são feitas nos índices p e q do modelo pré-selecionado, e o valor médio μ é adicionado ou retirado. Este passo é repetido até o AIC não poder ser mais minimizado, e por fim o modelo final ARIMA é escolhido;

No final, todos os modelos obtidos por ambos os métodos são submetidos a uma análise de resíduos, para atestar a normalidade destes. Essa análise inclui a elaboração de um autocorrelograma residual, de um histograma de resíduos e da realização de testes de Ljung-Box para verificação das hipóteses de independência dos resíduos.

4.3.3 Obtenção dos Modelos de Espaço de Estados

Os modelos de espaço de estados escolhidos para a análise deste trabalho são os modelos estruturais indicados por [33]. Os ajustes foram realizados com dois modelos. O primeiro é chamado modelo linear local:

$$y_t = \alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon); \quad (4.1)$$

$$\alpha_{t+1} = \alpha_t + \xi_t, \quad \xi_t \sim N(0, \sigma_\xi); \quad (4.2)$$

Onde y_t é alguma observação do sistema num instante t , α_t é o chamado nível do processo, ϵ_t é o erro aleatório associado à observação, e ξ_t o erro aleatório relativo ao nível do processo.

O segundo modelo a ser ajustado é um modelo linear local com uma componente de tendência estocástica ν_t , dado por:

$$y_t = \alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon); \quad (4.3)$$

$$\alpha_{t+1} = \alpha_t + \nu_t + \xi_t, \quad \xi_t \sim N(0, \sigma_\xi); \quad (4.4)$$

$$\nu_{t+1} = \nu_t + \zeta_t, \quad \zeta_t \sim N(0, \sigma_\zeta); \quad (4.5)$$

Por se tratarem de modelos não-estacionários, ambos são aplicados diretamente sobre as séries não linearizadas. Os parâmetros σ_ϵ , σ_ξ , e σ_ζ , assim como os níveis de estado α_t , foram estimados por meio do filtro de Kalman, que também estima as previsões dos modelos. Os ajustes também passaram por uma análise de Ljung-Box igual à dos modelos anteriores, e também foram comparados em termos do critério AIC.

4.4 FERRAMENTAS COMPUTACIONAIS

A linguagem R possui sintaxe relativamente simples, e contém uma gama robusta de pacotes estatísticos que proporcionam, de maneira simples e completa, implementações de modelos, realização de testes estatísticos e confecção de gráficos. Em particular, o ambiente de desenvolvimento escolhido para a construção do código foi o RStudio, uma IDE cuja interface simples e limpa permite a escrita do código paralelamente à visualização rápida de

resultados gráficos e a listagens de atributos de modelos, tornando o processo de desenvolvimento mais dinâmico.

Para todos os tipos de modelos escolhidos, os parâmetros do sistema foram definidos por meio de algoritmos computacionais de maximização de função de verossimilhança (*Maximum Likelihood Estimation*). No caso das funções escolhidas, a série de tempo é ajustada de acordo com o modelo de espaço de estados que represente um determinado modelo; a partir daí, estimativas de níveis são obtidas para cada modelo por meio de filtragem de Kalman. Finalmente, os valores médios desses níveis entram nas equações de log-verossimilhança de cada modelo, cuja maximização fornece os valores médios dos parâmetros de um determinado modelo.

Os pacotes estatísticos, bem como as funções utilizadas neste trabalho, foram:

- **tseries**, para análise geral de séries de tempo. Foram utilizadas as funções *read.ts*, que recebeu os dados históricos em formato *.dat* e os converteu em um objeto do tipo **TS**, um tipo de lista adequada para posterior manipulação da série com outras funções. O pacote também contém a função *adf.test()*, que recebe uma série de tempo e retorna as métricas do teste Dicker-Fulley Aumentado;
- **ggplot2**, para a plotagem e manipulação de resultados gráficos;
- **urca**, para a realização do teste KPSS com a função *ur.kpss*;
- **forecast**, com as funções de modelagem *auto.arima()* e *Arima()*, a função de modelagem para suavização exponencial *ets()*, a função de modelagem de modelos estruturais *StructTS()*, a função de previsão geral de modelos *forecast()* e a função de visualização de resíduos *check.residuals()*;

4.5 CONSIDERAÇÕES FINAIS

Este capítulo apresentou a aplicação metodológica dos princípios estatísticos fundamentais discutidos na teoria deste trabalho, abarcando principalmente os fundamentos da teoria de regressão estatística, abrindo espaço para obtenção e discussão de resultados posteriores.

5 RESULTADOS

5.1 SÉRIE DE MICROGERAÇÃO LOCAL RESIDENCIAL

5.1.1 Visualização Gráfica e Transformações

Neste primeiro momento, as séries de tempo foram basicamente expostas em formato gráfico para visualização:

Figura 5.1 – Potência Instalada - Microgeração Residencial Local

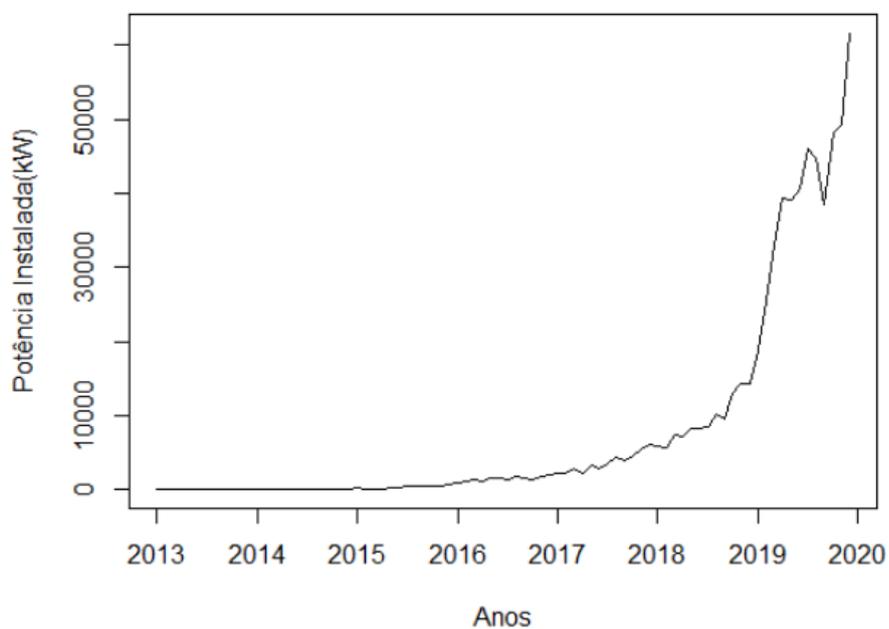
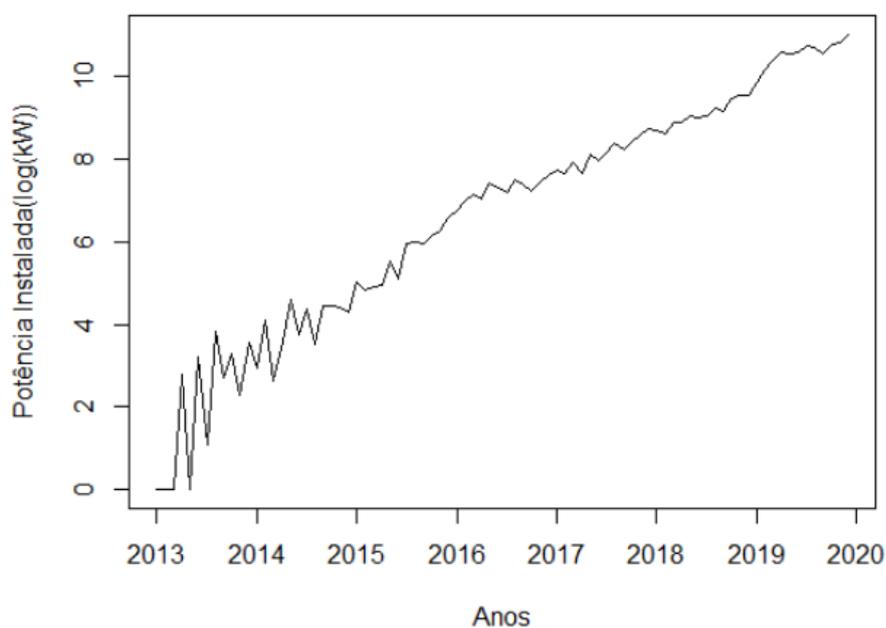


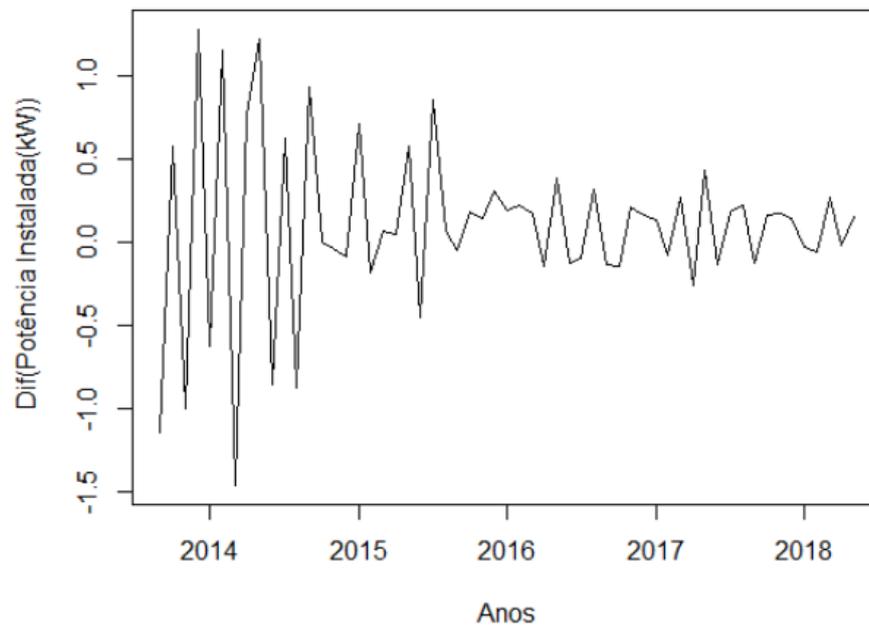
Figura 5.2 – Potência Instalada - Microgeração Residencial Local(Log)



A série original (tabela A.1 do Apêndice) apresenta uma tendência semelhante a um crescimento exponencial, onde o ano de 2019 exibe uma grande aceleração no ritmo de novas instalações, embora apresente um vale de instalações no mês de setembro. O crescimento exponencial sugere uma transformação log sobre o conjunto de dados, o que facilita o ajuste da série mediante os modelos lineares anteriormente referidos.

A série temporal log transformada (tabela A.2 do Apêndice) apresenta uma tendência de crescimento aproximadamente linear, como previsto. A transformação logarítmica atenuou os efeitos de vale no final da série de tempo, mas a presença de alguns valores nulos no início da série introduziu oscilações que se estabilizam após um tempo. Após a retirada destes pontos(até o mês de agosto), aplicou-se então uma diferenciação de primeira ordem na série, conforme sugerido pela função **ndiffs**:

Figura 5.3 – Potência Instalada - Microgeração Residencial Local(Diferenciada)



As estatísticas de teste KPSS, para a série diferenciada, foram $\eta_{\mu} = 0,1052$ e $\eta_{\tau} = 0,1024$, o que leva à não rejeição da estacionaridade para todas as faixas de valores críticos. Por fim, o teste ADF forneceu um valor $DF_{\tau} = -7,0798$, o que rejeita fortemente a hipótese nula de não estacionariedade. Dessa forma, podemos assumir com um bom grau de segurança que a série diferenciada é estacionária.

5.1.2 Ajustes ARIMA

Figura 5.4 – Autocorrelograma - Microgeração Residencial Local

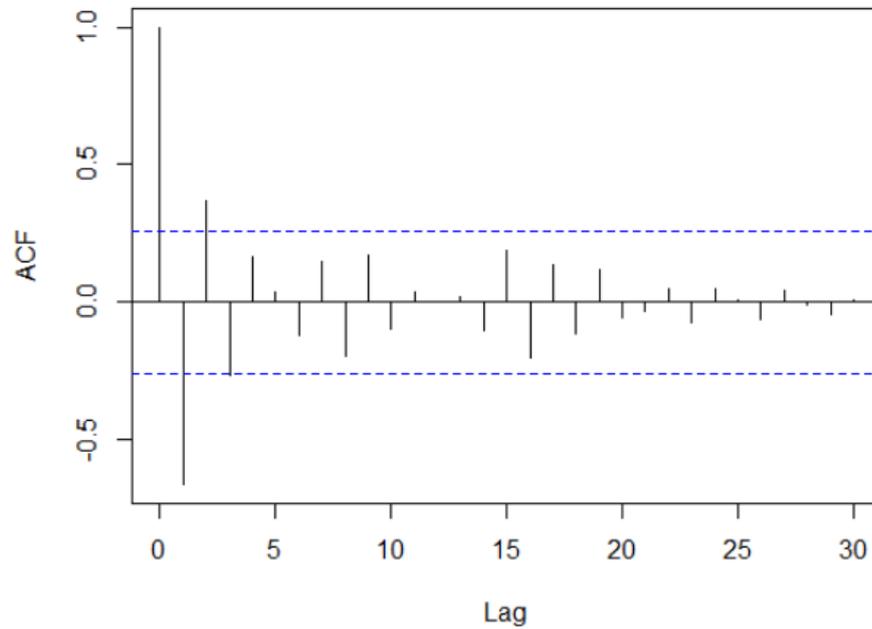
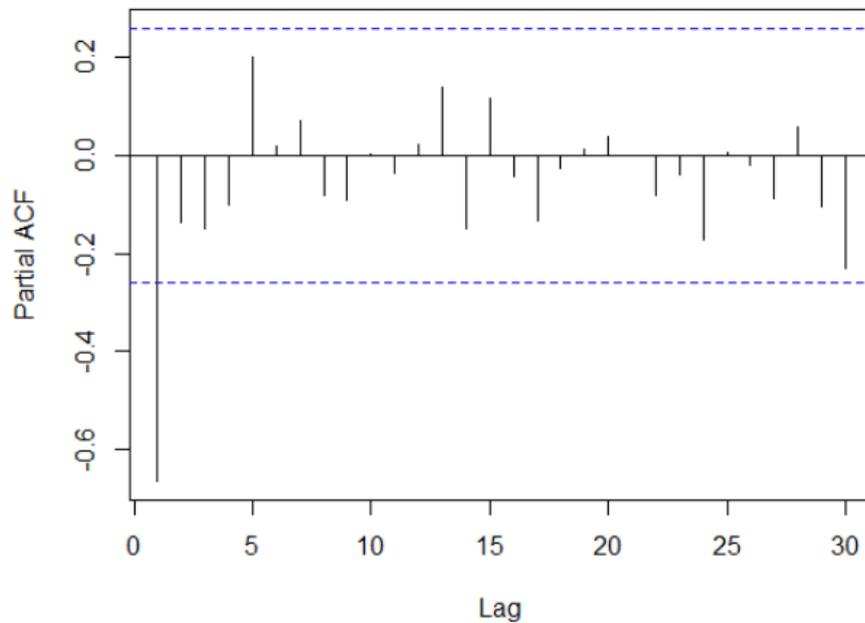


Figura 5.5 – Autocorrelograma Parcial - Microgeração Residencial Local



O último pico significativo da ACF está no *lag* 2, enquanto a PACF possui somente um único pico significativo no *lag* 1, padrão que, segundo a estrutura de Box-Jenkins, sugere um modelo $ARIMA(1, 1, 2)$. O algoritmo de minimização de Hyndman-Khandakar sugere um modelo $ARIMA(0, 1, 2)$. Os modelos encontrados foram:

Tabela 5.1 – Ajustes ARIMA - Microgeração Residencial Local

Modelo	ϕ_1	θ_1	θ_2	μ	AIC
ARIMA(1,1,2)	0,1044	-1,2301	0,6033	0,1017	51,76
ARIMA(0,1,2)	0,0000	-1,1596	0,5460	0,1026	49,94

Os critérios de informação têm um valor muito parecido entre si, o que sugere que a perda de informação dos ajustes é semelhante. Os testes de Ljung-Box forneceram p-valores $p_{(1,1,2)} = 0,9804$ e $p_{(0,1,2)} = 0,9759$, o que indica que os resíduos de ambos os modelos provavelmente configuram como ruído. Assim, assume-se que os dois modelos ajustaram a série de tempo com qualidade satisfatória, embora $ARIMA(0, 1, 2)$ tenha obtido um critério de informação superior.

5.1.3 Ajustes de Suavização Exponencial

O algoritmo automático ajustou dois modelos de suavização exponencial com componente de tendência, obtendo os seguintes parâmetros:

Tabela 5.2 – Ajustes SE - Microgeração Residencial Local

Modelo	α	β	ϕ	AIC
S.E sem amortecimento	0,2197	0,0604	-	133,8879
S.E com amortecimento	0,1535	0,1535	0,9217	134,3470

Os testes de Ljung-Box forneceram p-valores $SE_{semamortecimento} = 0,24$ e $SE_{comamortecimento} = 0,17$, o que indica que os resíduos de ambos os modelos provavelmente configuram como ruído. No entanto, o modelo sem amortecimento possui critérios de informação e p-valor ligeiramente melhores que o modelo amortecido, o que leva a crer que trata-se do modelo mais bem ajustado.

5.1.3.1 Ajustes de Modelos de Espaço de Estados

Os ajustes de modelos estados obtidos foram:

Tabela 5.3 – Ajustes EE - Microgeração Residencial Local

Modelo	σ_ϵ	σ_ξ	σ_ζ	AIC
Nível Local	0,0921	0,0596	-	87,1694
Tendência Local	0,1050	0,000	0,002	90,1430

Os testes de Ljung-Box forneceram p-valores $p_{nivellocal} = 0,0944$ e $p_{tendencialocal} =$

0,3215, o que indica que os resíduos de ambos os modelos provavelmente configuraram como ruído. Apesar do modelo com nível local possuir critérios de informação ligeiramente melhores que o modelo de tendência estocástica, O modelo com tendência local parece ser melhor ajustado pelo p-valor maior e pela tendência de crescimento ter sido totalmente capturada, a julgar pela variância nula do nível médio.

5.2 QUALIDADE PREDITIVA DOS MODELOS SELECIONADOS

Após a escolha dos métodos mais adequadamente ajustados à série, foram computadas as previsões (tabelas A.3, A.4 e A.5) para 19 períodos à frente, dos modelos ARIMA(0,1,2), suavização exponencial não amortecido, e estrutural com tendência estocástica:

Tabela 5.4 – Erros de Previsão(ln) - Microgeração Residencial Local

Modelo	RMSE(ln(KW))	MAE(ln(kW))	MAPE(%)
ARIMA(0,1,2)	0,219	0,171	1,679
SE(sem amortecimento)	0,268	0,220	2,138
EE(tendência local)	0,385	0,316	3,029

Figura 5.6 – Previsão ARIMA(0,1,2) - Microgeração Residencial Local

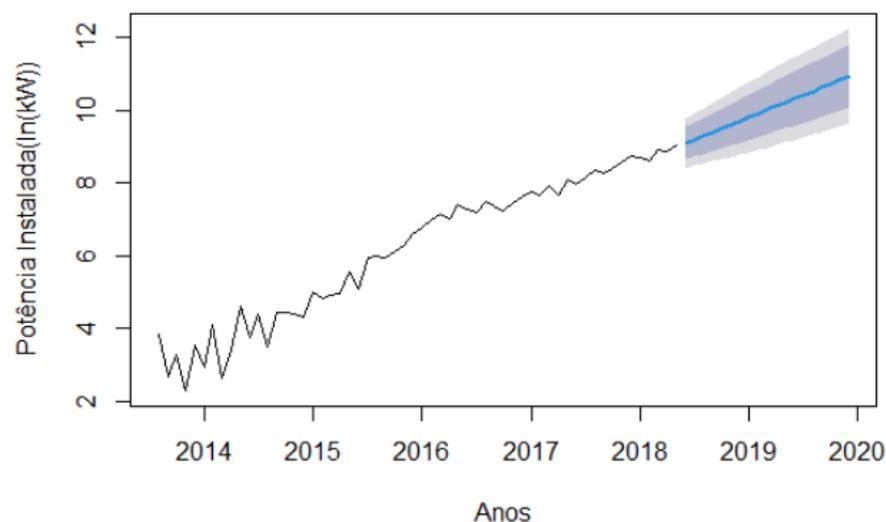


Figura 5.7 – Previsão de SE(sem amortecimento) - Microgeração Residencial Local

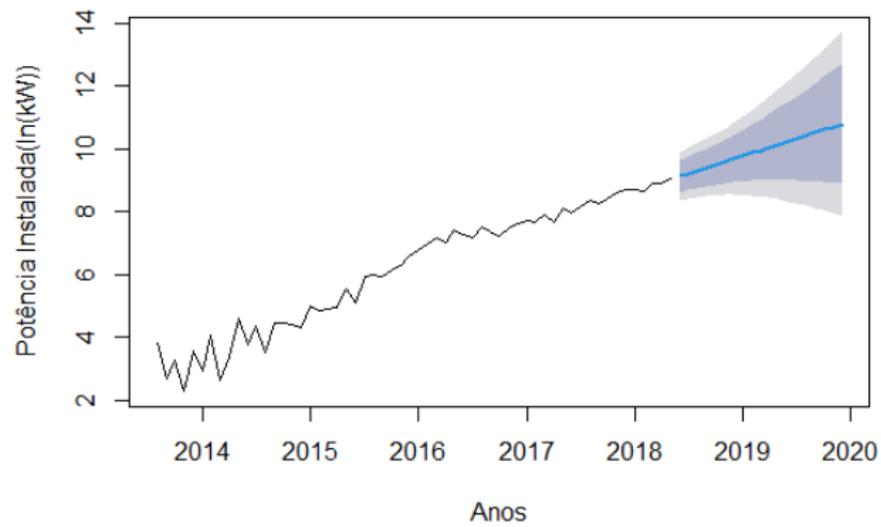
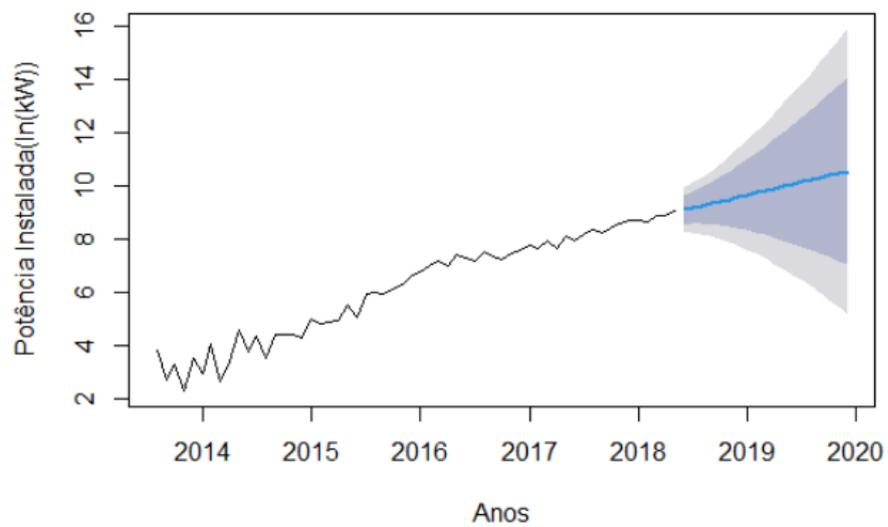


Figura 5.8 – Previsão de EE com Tendência Estocástica - Microgeração Residencial Local



A observação das métricas de erro indica que o modelo de melhor qualidade preditiva para essa série é o modelo ARIMA(0,1,2), para quaisquer métricas de erro.

5.3 SÉRIE DE MICROGERAÇÃO COMERCIAL LOCAL

5.3.1 Visualização Gráfica e Transformações

Figura 5.9 – Potência Instalada - Microgeração Comercial Local

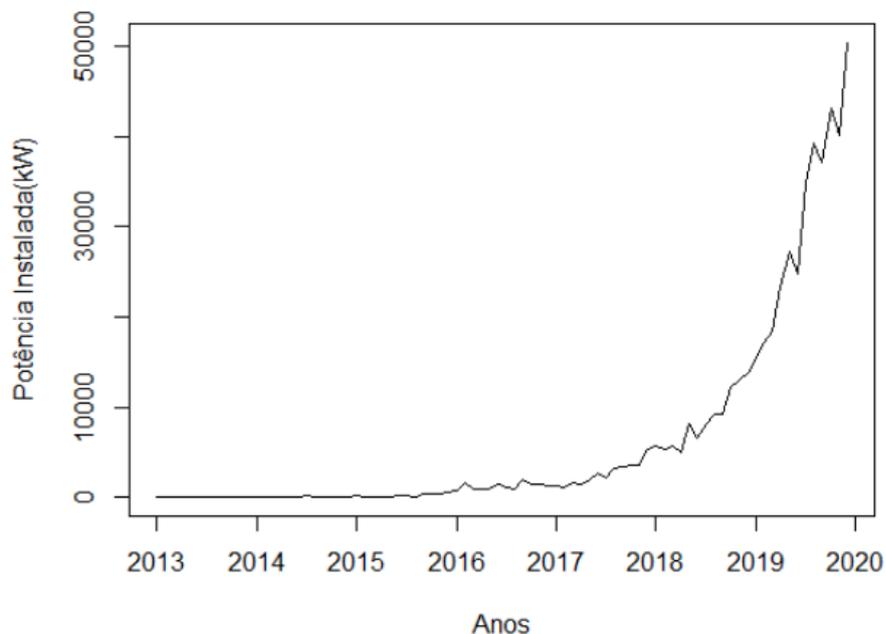
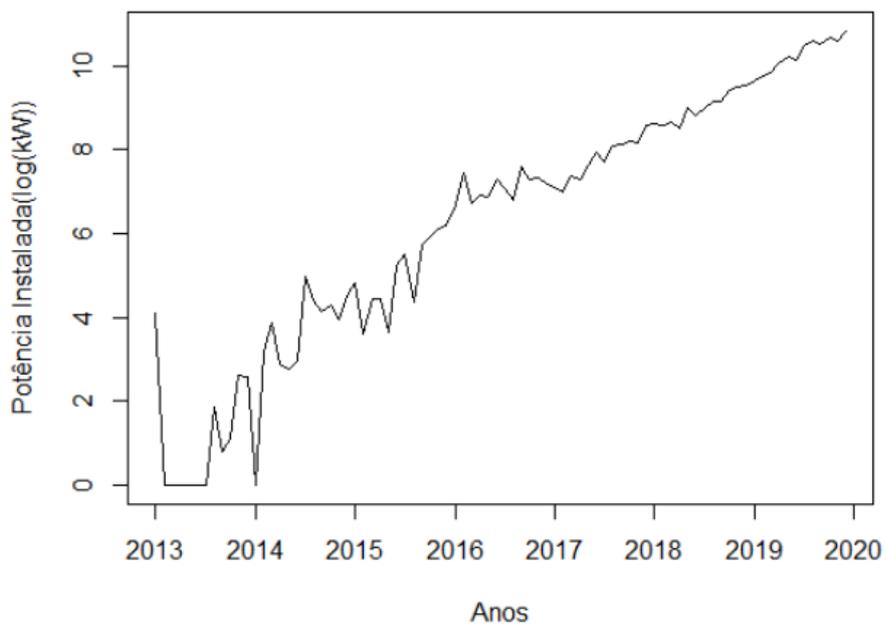


Figura 5.10 – Potência Instalada - Microgeração Comercial Local(Log)

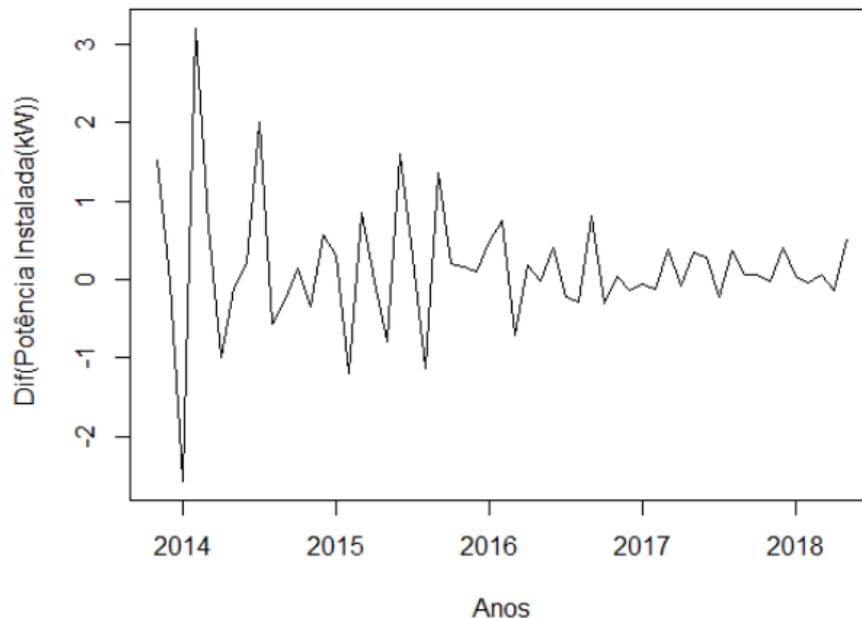


A série original (tabelaA.1 do Apêndice) apresenta uma tendência semelhante a um crescimento exponencial, mas com um número maior de oscilações, se comparada à série re-

sidencial. O crescimento exponencial também sugere aqui uma transformação log sobre o conjunto de dados, o que facilita o ajuste da série mediante os modelos lineares anteriormente referidos.

A série temporal log transformada (tabelaA.2 do Apêndice) apresenta uma tendência de crescimento aproximadamente linear, como previsto. A transformação logarítmica atenuou os efeitos de vale no final da série de tempo da mesma forma que na primeira série, mas a presença de mais valores nulos no início da série aumentou um pouco a instabilidade da série transformada. Após a retirada destes pontos da série log-transformada(até o mês de outubro de 2013), aplicou-se então uma diferenciação de primeira ordem na série, novamente conforme sugerido pela função **ndiffs**:

Figura 5.11 – Potência Instalada - Microgeração Comercial Local(Diferenciada)



As estatísticas de teste KPSS, para a série diferenciada, foram $\eta_{\mu} = 0,1434$ e $\eta_{\tau} = 0,0562$, o que leva à não rejeição da estacionaridade para todas as faixas de valores críticos. Por fim, o teste ADF forneceu um valor $DF_{\tau} = -6,200$, o que rejeita fortemente a hipótese nula de não estacionariedade. Dessa forma, podemos assumir com um bom grau de segurança que a série diferenciada é estacionária.

5.3.2 Ajustes ARIMA

Figura 5.12 – Autocorrelograma - Microgeração Comercial Local

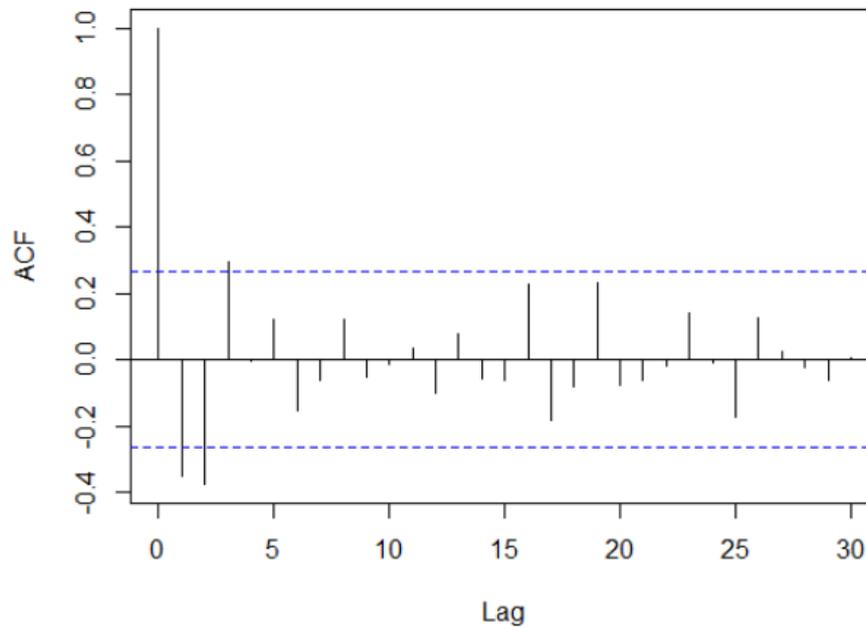
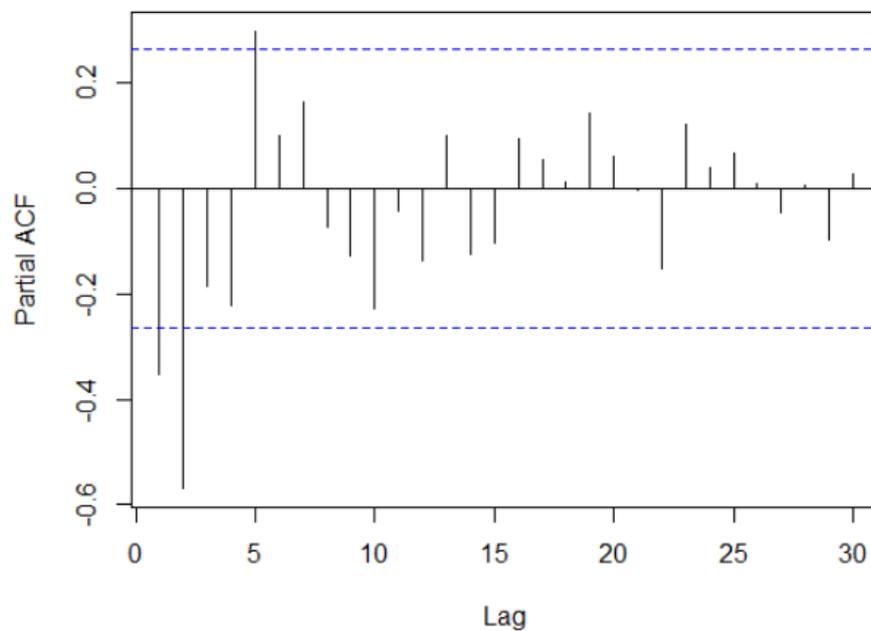


Figura 5.13 – Autocorrelograma Parcial - Microgeração Comercial Local



O último pico significativo da ACF está no *lag* 2, enquanto a PACF possui somente um único pico significativo no *lag* 2, padrão que, segundo a estrutura de Box-Jenkins, sugere um modelo $ARIMA(2, 1, 2)$. O algoritmo de minimização de Hyndman-Khandakar sugere um modelo $ARIMA(0, 1, 3)$. Os modelos encontrados foram:

Tabela 5.5 – Ajustes ARIMA - Microgeração Comercial Local

Modelo	ϕ_1	ϕ_2	θ_1	θ_2	θ_3	μ	AIC
ARIMA(2,1,2)	0,1959	-0,5924	-1,0104	0,6885	0,000	0,1340	107,45
ARIMA(0,1,3)	0,0000	0,0000	0,1166	0,1188	0,1084	0,0509	101,95

Os critérios de informação têm um valor muito parecido entre si, o que sugere que a perda de informação dos ajustes é semelhante. Os testes de Ljung-Box forneceram p-valores $p_{(2,1,2)} = 0,6204$ e $p_{(0,1,3)} = 0,9617$, o que indica que os resíduos de ambos os modelos provavelmente configuram como ruído. Assim, assume-se que os dois modelos ajustaram a série de tempo com qualidade satisfatória, embora *ARIMA*(0, 1, 3) tenha obtido um critério de informação superior.

5.3.3 Ajustes de Suavização Exponencial

O algoritmo automático ajustou dois modelos de suavização exponencial com componente de tendência, obtendo os seguintes parâmetros:

Tabela 5.6 – Ajustes SE - Microgeração Comercial Local

Modelo	α	β	ϕ	AIC
S.E sem amortecimento	0,2609	0,00004	-	187,4497
S.E com amortecimento	0,00004	0,00004	0,9783	180,6013

Os testes de Ljung-Box forneceram p-valores $SE_{semamortecimento} = 0,1352$ e $SE_{comamortecimento} = 0,0039$, o que indica que o modelo amortecido falhou em capturar toda a informação útil a respeito do sistema estocástico. Dada a pequena diferença entre os critérios de informação, o modelo SE sem amortecimento parece ser o mais adequadamente ajustado.

5.3.3.1 Ajustes de Modelos de Espaço de Estados

Os ajustes de modelos estados obtidos foram:

Tabela 5.7 – Ajustes EE - Microgeração Comercial Local

Modelo	σ_ϵ	σ_ξ	σ_ζ	AIC
Nível Local	0,2710	0,1319	-	138,76
Tendência Local	0,3066	0,0309	0,00002	132,98

Os testes de Ljung-Box forneceram p-valores $p_{nívellocal} = 0,0631$ e $p_{tendencialocal} = 0,09255$, o que indica que os resíduos de ambos os modelos provavelmente configuram

como ruído. No entanto, o modelo com tendência local possui critérios de de informação ligeiramente melhores que o modelo de nível local, o que leva a crer que trata-se do modelo mais bem ajustado.

5.4 QUALIDADE PREDITIVA DOS MODELOS SELECIONADOS

Após a escolha dos métodos mais adequadamente ajustados à série, foram computadas as previsões (tabelas A.6, A.7 e A.8), para 19 períodos à frente, dos modelos ARIMA(0,1,3), suavização exponencial não amortecido, e estrutural com tendência estocástica:

Tabela 5.8 – Erros de Previsão(ln) - Microgeração Comercial Local

Modelo	RMSE(ln(kW))	MAE(ln(kW))	MAPE(%)
ARIMA(0,1,3)	0,319	0,287	2,865
SE(sem amortecimento)	0,310	0,303	3,182
EE(tendência local)	0,089	0,074	0,794

Figura 5.14 – Previsão ARIMA(0,1,2) - Microgeração Comercial Local

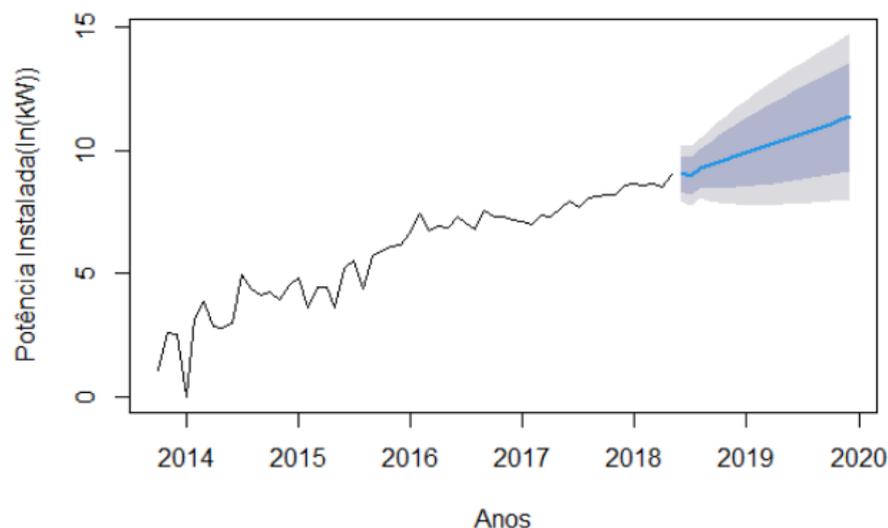


Figura 5.15 – Previsão de SE(sem amortecimento) - Microgeração Comercial Local

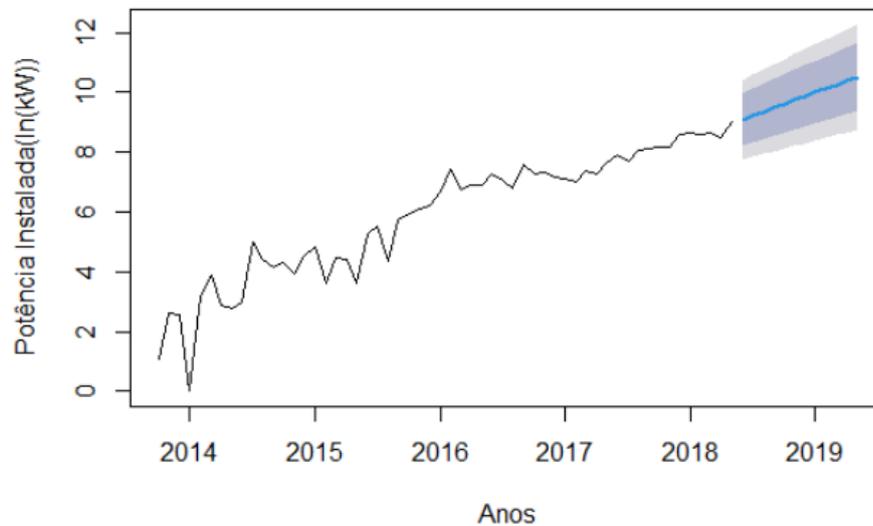
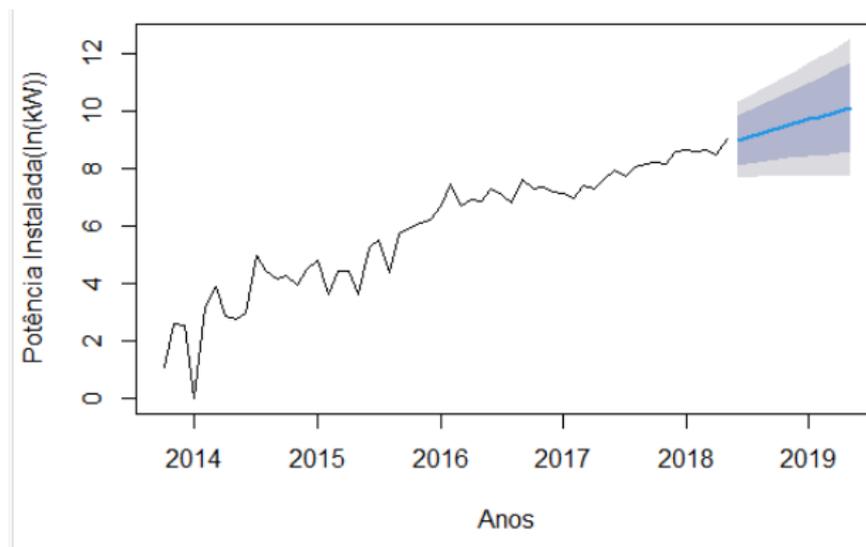


Figura 5.16 – Previsão de EE com Tendência Estocástica - Microgeração Comercial Local



Para o caso da série de microgeração residencial local, o modelo estrutural de tendência local foi o que obteve as melhores métricas de erro de previsão, por uma margem considerável.

5.5 CONSIDERAÇÕES FINAIS

Este capítulo explicita os resultados obtidos pelas metodologias propostas e discutidas no capítulo anterior. Os resultados refletem um processo de análise *a priori* das séries de tempo em si, e passa por análise comparativas entre métodos de ajuste de séries de tempo, tendo sido selecionados os modelos com melhor performance estatística. A partir destes modelos

escolhidos, foram obtidas previsões de níveis de potência fotovoltaica instalada para cada série de tempo.

6

CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho evidencia a importância da multidisciplinaridade e interlocução das diversas áreas das ciências exatas na resolução de problemas da engenharia. com fundamentos de Estatística e Econometria de séries de tempo, produziram-se resultados de previsão e projeção de potência instalada com potencial aplicação na construção de cenários e na proposição de novas políticas públicas para auxiliar os rumos futuros do mercado fotovoltaico no cenário nacional.

Os diálogos e trocas de conhecimentos entre as esferas acadêmica e pública enriqueceram muito a realização deste trabalho. Isto permitiu o fornecimento de dados diretamente relacionados às tendências nacionais da energia solar, o que permite uma comparação e contextualização dos resultados obtidos de tal forma que a proposição de novas metodologias para a projeção de cenários de difusão da energia solar se torna mais imediata e coesa.

A complexidade deste mercado define tanto as limitações deste trabalho, como suas potencialidades futuras. Os modelos de série de tempo apresentados e desenvolvidos aqui não possuem utilidade explanatória, ou seja, não são capazes, por si, de quantificar a influência de todos os fatores que contribuem para as tendências de mercado. No entanto, a maior qualidade estatística destes métodos permitem a construção de cenários de forma mais segura que métodos como a difusão de Bass, que, embora explanatório, possui uma configuração de parâmetros que não permite uma atualização da curva no ritmo de mudança das variáveis do sistema.

Assim, o caminho para os trabalhos futuros residem na aplicação destes resultados de análise econométrica como uma entrada em sistemas que sejam capazes de alocar a potência projetada de forma não explanatória em metodologias capazes de agregar e quantificar variáveis socioeconômicas de forma mais precisa e abrangente, o que requer uma coleta de dados cujo detalhamento e extensão estão além do escopo deste trabalho.

Este processo pode ser aliado à aplicação de metodologias mais complexas como modelos econométricos mistos, modelos de simulação de sistemas baseados em agentes(MBA) e modelos de difusão espacial de inovações(associando a captura de dados socioeconômicos à distribuição georreferenciada destes em um determinado território), o que pode resultar no desenvolvimento de sistemas que possam simular o processo de crescimento e espalhamento da tecnologia fotovoltaica de uma forma mais próxima das interações entre os diversos agentes de mercado na vida real.

REFERÊNCIAS BIBLIOGRÁFICAS

- 1 IEA-PVPS-TCD. *Trends in Photovoltaic Applications 2019*. [S.l.], 2019.
- 2 PARTAIN, L. et al. ‘swanson’s law’ plan to mitigate global climate change. In: . [S.l.: s.n.], 2016. p. 3335–3340.
- 3 METAYER, M.; FELL, H.-J. The projections for the future and quality in the past of the world energy outlook for solar pv and other renewable energy technologies. In: . [S.l.: s.n.], 2015.
- 4 OSMUNDSEN, T. How the IEA exaggerates the costs and underestimates the growth of solar power. 2014. Disponível em: <<https://energypost.eu/iea-exaggerates-costs-underestimates-growth-solar-power/>>.
- 5 WHITMORE, A. Why have IEA renewables growth projections been so much lower than the out-turn? 2013. Disponível em: <<https://energycentral.com/c/ec/why-have-iea-renewables-growth-projections-been-so-much-lower-out-turn>>.
- 6 SCG/SMA/SRD-ANEEL. *Revisão das regras aplicáveis à micro e minigeração distribuída – Resolução Normativa nº 482/2012 Relatório de Análise de Impacto Regulatório nº 0004/2018*. [S.l.], 2018.
- 7 LATHI, B. P. *Modern Digital and Analog Communication Systems*. 2nd. ed. USA: Oxford University Press, Inc., 1995. ISBN 0030284074.
- 8 SHMUELI, G.; JR, K. C. L. *Practical time series forecasting with r: A hands-on guide*. [S.l.]: Axelrod Schnall Publishers, 2016.
- 9 MOREIRA, P. et al. *Aplicação de Séries Temporais na Previsão da Média Mensal da Taxa de Câmbio do Real para o Dólar Comercial usando o Modelo de Holt*. 2010.
- 10 HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. [S.l.]: OTexts, 2018.
- 11 MORETTIN, P.; TOLOI, C. *Análise de séries temporais: modelos lineares univariados*. Blucher, 2018. ISBN 9788521213529. Disponível em: <<https://books.google.com.br/books?id=UwC5DwAAQBAJ>>.
- 12 FLORESCU, I. *Probability and Stochastic Processes*. Wiley, 2014. ISBN 9781118593134. Disponível em: <<https://books.google.com.br/books?id=Kdq6BQAAQBAJ>>.
- 13 KWIATKOWSKI, D. et al. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of econometrics*, v. 54, n. 1-3, p. 159–178, 1992.
- 14 FULLER, W. A. *Introduction to statistical time series*. [S.l.]: John Wiley & Sons, 2009. v. 428.

- 15 NG, S.; PERRON, P. Unit root tests in arma models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, Taylor Francis, v. 90, n. 429, p. 268–281, 1995. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476510>>.
- 16 SCHWERT, G. W. Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 20, n. 1, p. 5–17, 2002.
- 17 SEBER, G.; WILD, C. *Nonlinear Regression*. [S.l.]: John Wiley Sons, 2003. ISBN 0-471-47135-6.
- 18 ASUERO, A. G.; GONZÁLEZ, G. Fitting straight lines with replicated observations by linear regression. iii. weighting data. *Critical Reviews in Analytical Chemistry - CRIT REV ANAL CHEM*, v. 37, p. 143–172, 08 2007.
- 19 BASS, F. M. A new product growth for model consumer durables. *Management Science*, v. 15, n. 5, p. 215–227, 1969. Disponível em: <<https://doi.org/10.1287/mnsc.15.5.215>>.
- 20 ROGERS, E. *Diffusion of Innovations, 4th Edition*. Free Press, 2010. ISBN 9781451602470. Disponível em: <<https://books.google.com.br/books?id=v1ii4QsB7jIC>>.
- 21 HYNDMAN, R. et al. *Forecasting with exponential smoothing: the state space approach*. [S.l.]: Springer Science & Business Media, 2008.
- 22 GARDNER, E. S.; MCKENZIE, E. Why the damped trend works. *Journal of the Operational Research Society*, Springer, v. 62, n. 6, p. 1177–1180, 2011.
- 23 LEDOLTER, J.; ABRAHAM, B. Parsimony and its importance in time series forecasting. *Technometrics*, v. 23, p. 411–414, 11 1981.
- 24 BOX, G.; JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. [S.l.]: Holden-Day, 1976.
- 25 COMMANDEUR, J. J.; KOOPMAN, S. J. *An introduction to state space time series analysis*. [S.l.]: Oxford University Press, 2007.
- 26 THACKER, N.; LACEY, T. The kalman filter. 1997.
- 27 KONZEN, G. *Difusão de Sistemas Fotovoltaicos Residenciais Conectados à Rede no Brasil: Uma simulação via modelo de Bass*. Dissertação (Mestrado) — Universidade de São Paulo, São Paulo, 2014.
- 28 Nguyen, T. et al. Forecasting model for the solar photovoltaics consumption in united states of america. In: *2012 10th International Power Energy Conference (IPEC)*. [S.l.: s.n.], 2012. p. 284–288.
- 29 Zhao, T. et al. Spatio-temporal analysis and forecasting of distributed pv systems diffusion: A case study of shanghai using a data-driven approach. *IEEE Access*, v. 5, p. 5135–5148, 2017.
- 30 Schiera, D. S. et al. Analysis of rooftop photovoltaics diffusion in energy community buildings by a novel gis- and agent-based modeling co-simulation platform. *IEEE Access*, v. 7, p. 93404–93432, 2019.

- 31 EPE. *Modelo de Mercado da Micro e Minigeração Distribuída(4MD): Metodologia - Versão PDE 2027*. [S.l.], 2018.
- 32 SCG/SMA/SRD-ANEEL. *Nota Técnica 0056/2017:Atualização das projeções de consumidores residenciais e comerciais com microgeração solar fotovoltaicos no horizonte 2017-2024*. [S.l.], 2017.
- 33 WEISS, C. E. The use of time-series methods for diffusion modelling: An evaluation. *PhD First-Year Report (June 2014), Judge Business School, University of Cambridge*, 2014.
- 34 PELGRIN, F. *Lecture 5: Box-Jenkins Methodology*. University of Lausanne, Department of Mathematics(IMEA-NICE), 2011. Disponível em: <https://math.unice.fr/~frapetti/CorsoP/Chapitre_5_IMEA_1.pdf>.
- 35 HYNDMAN, R. J.; KHANDAKAR, Y. et al. *Automatic time series for forecasting: the forecast package for R*. [S.l.]: Monash University, Department of Econometrics and Business Statistics . . . , 2007.

APÊNDICE

A TABELAS

Tabela A.1 – Série de Tempo: Microgeração Residencial Local

Mês	Residencial - Micro Local(kW)	Log Transformada(log(kW))
jan/13	0,00	0,00
fev/13	0,00	0,00
mar/13	0,00	0,00
abr/13	15,00	2,77
mai/13	0,00	0,00
jun/13	24,00	3,22
jul/13	2,00	1,10
ago/13	45,50	3,84
set/13	13,85	2,70
out/13	25,68	3,28
nov/13	8,79	2,28
dez/13	34,20	3,56
jan/14	17,83	2,94
fev/14	58,87	4,09
mar/14	12,88	2,63
abr/14	28,70	3,39
mai/14	99,69	4,61
jun/14	41,92	3,76
jul/14	79,22	4,38
ago/14	32,57	3,51
set/14	83,77	4,44
out/14	83,37	4,44
nov/14	80,77	4,40
dez/14	73,15	4,31
jan/15	149,90	5,02
fev/15	125,44	4,84
mar/15	135,29	4,91
abr/15	141,26	4,96
mai/15	253,45	5,54
jun/15	160,61	5,09
jul/15	379,05	5,94
ago/15	407,89	6,01
set/15	386,51	5,96

Tabela A.1 continua da página anterior

Mês	Residencial - Micro Local(kW)	Log Transformada(ln(kW))
out/15	462,74	6,14
nov/15	533,01	6,28
dez/15	729,75	6,59
jan/16	875,52	6,78
fev/16	1091,82	7,00
mar/16	1297,14	7,17
abr/16	1123,49	7,03
mai/16	1649,64	7,41
jun/16	1469,47	7,29
jul/16	1329,58	7,19
ago/16	1821,09	7,51
set/16	1609,98	7,38
out/16	1386,27	7,24
nov/16	1721,76	7,45
dez/16	2023,11	7,61
jan/17	2292,58	7,74
fev/17	2128,59	7,66
mar/17	2789,66	7,93
abr/17	2140,24	7,67
mai/17	3306,46	8,10
jun/17	2900,85	7,97
jul/17	3446,20	8,15
ago/17	4323,77	8,37
set/17	3825,59	8,25
out/17	4499,12	8,41
nov/17	5329,31	8,58
dez/17	6093,89	8,72
jan/18	5948,79	8,69
fev/18	5594,40	8,63
mar/18	7301,62	8,90
abr/18	7166,24	8,88
mai/18	8339,56	9,03
jun/18	8194,68	9,01
jul/18	8487,84	9,05
ago/18	10272,65	9,24
set/18	9452,44	9,15
out/18	13010,92	9,47

Tabela A.1 continua da página anterior

Mês	Residencial - Micro Local(kW)	Log Transformada(ln(kW))
nov/18	14225,04	9,56
dez/18	14209,30	9,56
jan/19	18755,61	9,84
fev/19	25579,13	10,15
mar/19	33124,96	10,41
abr/19	39387,60	10,58
mai/19	39137,25	10,57
jun/19	40581,66	10,61
jul/19	45942,81	10,74
ago/19	44635,46	10,71
set/19	38373,56	10,56
out/19	48078,94	10,78
nov/19	49285,32	10,81
dez/19	61651,60	11,03

Tabela A.2 – Série de Tempo: Microgeração Comercial Local

Mês	Comercial - Micro Local(kW)	Log Transformada(ln(kW))
jan/13	60,48	4,12
fev/13	0,00	0,00
mar/13	0,00	0,00
abr/13	0,00	0,00
mai/13	0,00	0,00
jun/13	0,00	0,00
jul/13	0,00	0,00
ago/13	5,41	1,86
set/13	1,20	0,79
out/13	2,00	1,10
nov/13	12,70	2,62
dez/13	12,12	2,57
jan/14	0,00	0,00
fev/14	23,62	3,20
mar/14	47,60	3,88
abr/14	16,92	2,89
mai/14	15,04	2,78
jun/14	18,72	2,98
jul/14	145,52	4,99

Table A.2 continua da página anterior

Mês	Comercial - Micro Local(kW)	Log Transformada(ln(kW))
ago/14	82,33	4,42
set/14	62,51	4,15
out/14	72,22	4,29
nov/14	50,56	3,94
dez/14	90,00	4,51
jan/15	123,44	4,82
fev/15	36,17	3,62
mar/15	85,42	4,46
abr/15	83,16	4,43
mai/15	37,08	3,64
jun/15	186,61	5,23
jul/15	246,18	5,51
ago/15	77,87	4,37
set/15	304,21	5,72
out/15	375,35	5,93
nov/15	444,94	6,10
dez/15	490,20	6,20
jan/16	796,98	6,68
fev/16	1676,95	7,43
mar/16	829,27	6,72
abr/16	992,97	6,90
mai/16	973,45	6,88
jun/16	1462,17	7,29
jul/16	1173,65	7,07
ago/16	882,79	6,78
set/16	1979,68	7,59
out/16	1456,69	7,28
nov/16	1514,80	7,32
dez/16	1306,84	7,18
jan/17	1225,02	7,11
fev/17	1080,74	6,99
mar/17	1580,03	7,37
abr/17	1470,97	7,29
mai/17	2055,00	7,63
jun/17	2755,90	7,92
jul/17	2208,64	7,70
ago/17	3192,45	8,07

Table A.2 continua da página anterior

Mês	Comercial - Micro Local(kW)	Log Transformada(ln(kW))
set/17	3401,31	8,13
out/17	3602,03	8,19
nov/17	3528,44	8,17
dez/17	5312,89	8,58
jan/18	5601,48	8,63
fev/18	5325,32	8,58
mar/18	5681,00	8,65
abr/18	4920,64	8,50
mai/18	8202,85	9,01
jun/18	6560,08	8,79
jul/18	8148,57	9,01
ago/18	9186,45	9,13
set/18	9226,10	9,13
out/18	12256,06	9,41
nov/18	12902,12	9,47
dez/18	13851,89	9,54
jan/19	15397,15	9,64
fev/19	17223,12	9,75
mar/19	18391,89	9,82
abr/19	23308,33	10,06
mai/19	27292,78	10,21
jun/19	24815,86	10,12
jul/19	34762,65	10,46
ago/19	39242,94	10,58
set/19	37069,35	10,52
out/19	43117,87	10,67
nov/19	40123,43	10,60
dez/19	50439,32	10,83

Tabela A.3 – Previsão ARIMA(0,1,2) - Microgeração Residencial Local.

Mês	Mínima(kW)	Média(kW)	Máxima(kW)	Real(kW)
jun/18	5663,6	8918,5	14043,7	8194,7
jul/18	6172,4	9700,2	15366,3	8487,8
ago/18	6620,0	10839,0	17728,8	10272,7
set/18	7121,4	12003,1	20250,8	9452,4
out/18	7668,4	13305,4	23085,4	13010,9
nov/18	8265,8	14749,0	26290,5	14225,0
dez/18	8927,5	16334,6	29910,6	14209,3
jan/19	9642,1	18105,0	33961,1	18755,6
fev/19	10434,8	20049,3	38560,1	25579,1
mar/19	11292,7	22224,6	43738,2	33125,0
abr/19	12245,6	24586,7	49562,0	39387,6
mai/19	13278,8	27281,5	56105,1	39137,3
jun/19	14399,2	30241,4	63512,0	40581,7
jul/19	15629,8	33859,4	71824,9	45942,8
ago/19	16982,5	37122,3	81144,6	44635,5
set/19	18452,3	41149,9	91765,3	38373,6
out/19	20069,4	45614,4	103568,7	48078,9
nov/19	21828,1	50512,7	116890,3	49285,3
dez/19	23622,6	55993,0	131925,5	61651,6

Fonte: elaboração própria.

Tabela A.4 – Previsão de Suavização Exponencial - Microgeração Residencial Local

Mês	Mínima(kW)	Média(kW)	Máxima(kW)	Real(kW)
jun/18	5522,4	9171,8	15237,5	8194,7
jul/18	5941,1	10055,8	17045,7	8487,8
ago/18	6340,1	11035,9	19223,0	10272,7
set/18	6708,5	12110,4	21864,2	9452,4
out/18	7037,8	13278,8	25081,9	13010,9
nov/18	7322,5	14573,1	29011,7	14225,0
dez/18	7559,8	15989,8	33818,6	14209,3
jan/19	7749,0	17499,8	39705,9	18755,6
fev/19	7890,8	19224,6	46926,5	25579,1
mar/19	7986,7	21098,4	55796,4	33125,0
abr/19	8038,9	23154,8	66712,2	39387,6
mai/19	8050,2	25386,2	80173,1	39137,3
jun/19	8023,3	27860,5	96809,0	40581,7
jul/19	7961,1	30545,3	117416,2	45942,8
ago/19	7866,6	33522,4	143003,6	44635,5
set/19	7742,7	36789,7	174852,4	38373,6
out/19	7592,5	40335,0	214591,7	48078,9
nov/19	7418,8	44266,2	264298,7	49285,3
dez/19	7224,5	48532,0	326626,6	61651,6

Fonte: elaboração própria.

Tabela A.5 – Previsão de Espaço de Estados - Microgeração Residencial Local

Mês	Mínima(kW)	Média(kW)	Máxima(kW)	Real(kW)
jun/18	5242,0	8865,2	15219,6	8194,7
jul/18	5218,7	9613,2	17658,9	8487,8
ago/18	5167,4	10434,8	20877,9	9452,4
set/18	5109,1	11315,3	25100,8	10272,7
out/18	4929,2	12270,1	30622,4	13010,9
nov/18	4695,2	13329,4	37839,4	14209,3
dez/18	4422,1	14457,0	47288,0	14225,0
jan/19	4122,9	15676,8	59696,8	18755,6
fev/19	3808,8	17016,5	76057,8	25579,1
mar/19	3488,9	18452,3	97727,0	33125,0
abr/19	3170,8	20029,3	126563,3	38373,6
mai/19	2860,4	21589,3	165126,8	39137,3
jun/19	2562,3	23387,5	216956,1	39387,6
jul/19	2280,0	25564,5	286962,7	40581,7
ago/19	2015,9	27749,2	381992,2	44635,5
set/19	1771,4	30090,6	511624,4	45942,8
out/19	1547,4	32662,1	689321,1	48078,9
nov/19	1344,0	35241,2	934074,9	49285,3
dez/19	1160,8	38444,6	1272790,4	61651,6

Fonte: elaboração própria.

Tabela A.6 – Previsão ARIMA(0,1,3) - Microgeração Comercial Local

Mês	Mínima(kW)	Média(kW)	Máxima(kW)	Real(kW)
jun/18	4154,0	8517,5	17717,3	8194,7
jul/18	3556,1	7784,4	17101,7	8487,8
ago/18	4769,0	10403,6	22939,3	9452,4
set/18	4666,8	11848,0	30435,8	10272,7
out/18	4665,9	13493,0	39524,3	13010,9
nov/18	4732,3	15366,3	50595,6	14209,3
dez/18	4850,2	18032,7	64094,5	14225,0
jan/19	5011,3	20089,5	80542,8	18755,6
fev/19	5211,3	22878,6	100560,5	25579,1
mar/19	5448,2	26081,0	124886,0	33125,0
abr/19	5721,6	29701,9	154400,6	38373,6
mai/19	6031,8	33859,4	190155,7	39137,3
jun/19	6380,3	38560,1	233406,2	39387,6
jul/19	6768,8	43957,4	285648,6	40581,7
ago/19	7200,0	50060,1	348666,1	44635,5
set/19	7676,7	55546,8	424584,2	45942,8
out/19	8202,4	64859,9	515931,4	48078,9
nov/19	8781,0	73864,4	625717,0	49285,3
dez/19	9417,2	84456,2	757519,6	61651,6

Fonte: elaboração própria.

Tabela A.7 – Previsão de Suavização Exponencial - Microgeração Comercial Local

Mês	Mínima(kW)	Média(kW)	Máxima(kW)	Real(kW)
jun/18	3714,9	8777,0	21042,8	8194,7
jul/18	4107,7	9995,6	24658,7	8487,8
ago/18	4546,1	11383,4	28870,1	9452,4
set/18	5035,3	13028,9	33773,4	10272,7
out/18	5581,2	14837,8	39480,0	13010,9
nov/18	6190,7	16880,9	46119,2	14209,3
dez/18	6870,9	19224,6	53840,8	14225,0
jan/19	7630,4	21806,3	62818,2	18755,6
fev/19	8478,5	24833,8	73252,5	25579,1
mar/19	9425,7	28281,5	85376,2	33125,0
abr/19	10483,6	32272,4	99458,7	38373,6
mai/19	11665,6	36752,9	115812,1	39137,3
jun/19	12986,3	41771,8	134797,2	39387,6
jul/19	14462,3	47571,0	156831,7	40581,7
ago/19	16112,1	54175,4	182399,1	44635,5
set/19	17956,6	61696,6	212058,5	45942,8
out/19	20019,0	69562,8	246456,8	48078,9
nov/19	22325,4	79220,3	286342,4	49285,3
dez/19	24905,3	90218,4	332579,8	61651,6

Fonte: elaboração própria.

Tabela A.8 – Previsão de Espaço de Estados - Microgeração Comercial Local

Mês	Mínima(kW)	Média(kW)	Máxima(kW)	Real(kW)
jun/18	3326,6	7862,6	18787,5	8194,7
jul/18	3525,3	8777,0	21861,4	8487,8
ago/18	3699,7	9709,9	25510,4	9452,4
set/18	3870,4	10741,9	29862,0	10272,7
out/18	4037,1	11895,5	35059,6	13010,9
nov/18	4199,1	13159,8	41277,4	14209,3
dez/18	4356,1	14558,5	48727,1	14225,0
jan/19	4507,5	16122,0	57666,4	18755,6
fev/19	4653,1	17835,5	68409,5	25579,1
mar/19	4792,3	19731,1	81339,7	33125,0
abr/19	4925,0	21828,1	96925,5	38373,6
mai/19	5050,7	24179,4	115739,7	39137,3
jun/19	5169,3	26741,3	138484,4	39387,6
jul/19	5280,3	29583,3	166020,1	40581,7
ago/19	5383,7	32760,2	199403,0	44635,5
set/19	5479,2	36260,3	239931,8	45942,8
out/19	5566,6	40093,7	289204,0	48078,9
nov/19	5645,9	44354,9	349187,8	49285,3
dez/19	5716,8	49117,9	422309,5	61651,6

Fonte: elaboração própria.