



Centro de Ciências Exatas e Tecnologias - CCET
Programa de Pós-graduação em Engenharia Elétrica

Análise de séries temporais em dados do setor do petróleo e gás natural

Felipe Bezerra Pimentel Araújo

São Luis - MA, 2022

Felipe Bezerra Pimentel Araújo

Análise de séries temporais em dados do setor do petróleo e gás natural

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica do Centro de Ciências Exatas e Tecnologias, da Universidade Federal do Maranhão, como requisito para a obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Allan Kardec Duailibe Barros
Filho

São Luis, Março de 2022

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Araújo, Felipe Bezerra Pimentel.

Análise de séries temporais em dados do setor do
petróleo e gás natural / Felipe Bezerra Pimentel Araújo. -
2022.

61 f.

Orientador(a): Allan Kardec Duailibe Barros Filho.

Dissertação (Mestrado) - Programa de Pós-graduação em
Engenharia Elétrica/ccet, Universidade Federal do
Maranhão, São Luís - MA, 2022.

1. Gás. 2. Modelos estocásticos. 3. Petróleo. 4.
Previsões. 5. Séries temporais. I. Barros Filho, Allan
Kardec Duailibe. II. Título.

CCET - Centro de Ciências Exatas e Tecnologias
Universidade Federal do Maranhão

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica intitulada ***Análise de séries temporais em dados do setor do petróleo e gás natural*** de autoria de Felipe Bezerra Pimentel Araújo, aprovada pela banca examinadora constituída pelos seguintes professores:

Prof. Dr. Allan Kardec Duailibe Barros Filho -
Orientador
Programa de Pós-graduação em Engenharia Elétrica
- UFMA

Prof. Dr. Ewaldo Eder Carvalho Santana -
Membro da banca
Programa de Pós-graduação em Engenharia Elétrica
- UFMA

Prof. Dr. Wendell Ferreira de La Salles-
Membro da banca
Departamento de Tecnologia Química - UFMA

Prof. Dr. Inocêncio Sanches dos Santos Neto -
Membro da banca
Departamento de Química - UEMA

São Luis, 18 de Fevereiro de 2022

AGRADECIMENTOS

Agradeço ao meus pais, Pimentel e Zezé, por proporcionar tudo de melhor nessa vida. E também agradeço eternamente minha irmã Rayssa.

Agradeço a Adriana por todo companheirismo, amor e respeito.

Agradeço ainda aos membros do Laboratório de Processamento da Informação Biológica (PIB - UFMA). Sobretudo ao Luís Fillype pela amizade e Artur Cabral pelos auxílios no decorrer dos compromissos acadêmicos. Agradecimentos especiais ao professor Allan Kardec Barros, pela oportunidade de fazer parte de sua equipe.

RESUMO

A análise de séries temporais é uma ferramenta muito importante para a compreensão do mercado e para a elaboração de planos de ação e estratégias. Através do estudo dos padrões de comportamento de uma variável (como tendência, sazonalidade e ciclo) é possível fazer previsões e, a partir disso, tomar decisões de negócio mais coerentes. O setor de petróleo e gás natural compreende uma quantidade significativa de séries históricas relevantes (preço, produção, capacidade, etc.), de modo que, modelos probabilísticos ou estocásticos podem ser produzidos com o objetivo de fazer previsões para ajudar nas tomadas de decisão. Analisar, identificar e prever os padrões de comportamento são estratégias imprescindíveis para que haja um menor risco financeiro e, consequentemente, menores perdas de investimento e custo operacional.

Palavras-chave: Séries temporais, Previsões, Modelos estocásticos, Petróleo, Gás.

ABSTRACT

Time series analysis is a very important tool for the comprehension of the market and for developing action plans and strategies. By studying the behavior patterns of a variable (such as trend, seasonality, and cycle) it is possible to make predictions and, henceforth, make more effective business decisions. The oil and natural gas sector comprise a significant amount of relevant historical series (price, production, capacity, etc.), therefore, probabilistic or stochastic models can be produced in order to make predictions that will help in decision-making. Analyzing, identifying, and predicting behavior patterns are essential strategies to reduce financial risk and, consequently, reduce capital loss and operating costs.

Keywords: Time series, Forecasts, Stochastic models, Oil, Gas.

LISTA DE FIGURAS

2.1	Preços médios anuais do barril do petróleo entre os períodos de 1861 a 2020.	19
3.1	Gráfico de tempo da cotação diária do barril de petróleo BRENT entre 04/11/2011 até 25/10/2021.	21
3.2	Gráfico de tempo da série produtiva de gás natural no Brasil entre 1970 e 2020.	22
3.3	Série de preços globais anuais de petróleo entre 1861 a 2020 com prováveis ciclos.	23
3.4	Gráfico sazonal mensal da cotação do barril de petróleo BRENT entre 04/11/2011 até 25/10/2021.	24
3.5	Emissões globais de dióxido de carbono entre 1965 e 2020.	25
3.6	Produção de petróleo a nível mundial entre 1965 e 2020.	25
3.7	Gráfico de dispersão entre emissões de dióxido de carbono e produção de petróleo.	26
3.8	Gráfico de dispersão entre capacidade de refino global e produção de petróleo global.	27
4.1	Exportações brasileiras de bens e serviços em termos percentuais do PIB entre 1960-2017	31
4.2	Exportações brasileiras de bens e serviços em termos percentuais do PIB entre 1960-2017 com média móvel de 5 anos.	32
4.3	Decomposição STL para a série produtiva de gás natural no Brasil de 1970 a 2020.	33
5.1	Cotação diária do barril de petróleo BRENT entre 01/02/2019 a 15/04/2019 com previsões dos modelos de média, Naïve, Naïve sazonal e <i>drift</i> e os dados reais do período.	36
5.2	Fluxograma da abordagem Box e Jenkins para ajuste de um modelo preditivo ARIMA.	43
6.1	Fluxograma com metologia de produção de previsões em séries temporais .	44
7.1	Especificação de modelos na série temporal produtiva de gás natural no Brasil de 1970 a 2020.	46
7.2	Validação dos modelos na série temporal produtiva de gás natural no Brasil de 1970 a 2020.	47

7.3	Previsão na série temporal produtiva de gás natural no Brasil com horizonte de 5 anos.	47
7.4	Especificação de modelos na série temporal produtiva de petróleo no Brasil de 1965 a 2020.	48
7.5	Validação dos modelos na série temporal produtiva petróleo no Brasil de 1965 a 2020.	48
7.6	Previsão na série temporal produtiva de petróleo no Brasil com horizonte de 5 anos.	49
7.7	Validação dos modelos na série temporal de consumo de gasolina no Brasil de 1980 a 2020.	49
7.8	Previsão na série temporal de consumo de gasolina no Brasil com horizonte de 5 anos.	50
7.9	Validação dos modelos na série temporal de consumo de diesel no Brasil de 1980 a 2020.	50
7.10	Previsão na série temporal de consumo de diesel no Brasil com horizonte de 5 anos.	51
7.11	Validação dos modelos na série temporal de consumo de bicom bustíveis no Brasil de 1990 a 2020.	51
7.12	Previsão na série temporal de consumo de biocombustíveis no Brasil com horizonte de 5 anos.	52
7.13	Previsão na série temporal de capacidade de refino do Brasil com horizonte de 5 anos.	52
7.14	Previsão na série temporal de rendimento das refinarias do Brasil com horizonte de 5 anos.	53
7.15	Previsão na série temporal de preços do gás natural com índice <i>Henry Hub</i> em um horizonte de 5 anos.	53
7.16	Previsão na série temporal de preços do barril do petróleo em um horizonte de 6 anos.	54
7.17	Previsão na série temporal de capacidade instalada de energia eólica em um horizonte de 5 anos.	55
7.18	Previsão na série temporal de capacidade instalada de energia solar em um horizonte de 5 anos.	55
8.1	Análise SWOT do mercado de petróleo e gás	56

8.2	Matriz de risco dos desafios setor de petróleo e gás	57
-----	--	----

LISTA DE TABELAS

2.1	Densidade de energia em tipos de carvão, óleos brutos e gases naturais. . .	17
5.1	Multiplicadores a serem usados para intervalos de previsão.	37
5.2	Desvio padrão para previsão de várias etapas em quatro métodos de referência	38
5.3	Composição de métodos de suavização exponencial	41
5.4	Classificação de alguns métodos de suavização exponencial	41
7.1	Métricas comparativas entre modelos e dados observados entre 2011 e 2020.	46
7.2	Métricas comparativas entre modelos e dados observados entre 2011 e 2020.	48

Conteúdo

1	INTRODUÇÃO	14
1.1	Objetivo Geral	15
1.2	Objetivos Específicos	15
2	BREVE HISTÓRICO DO SETOR DE PETRÓLEO E GÁS	16
2.1	Carvão	16
2.2	Petróleo e gás natural	18
3	PRINCÍPIOS EM SÉRIES TEMPORAIS	21
3.1	Padrões em séries temporais	22
3.2	Gráficos sazonais	23
3.3	Correlação	24
3.4	Autocorrelação	27
3.5	Ruído branco	28
4	DECOMPOSIÇÃO DE SÉRIES TEMPORAIS	30
4.1	Componentes em séries temporais	30
4.2	Suavização de médias móveis	30
4.3	Desvantagens da decomposição clássica	31
4.4	Decomposição STL	32
5	PREVISÕES COM SÉRIES TEMPORAIS	34
5.1	Modelos preditivos simples	34
5.1.1	Modelo da média	34
5.1.2	Modelo Naïve	34
5.1.3	Modelo Naïve sazonal	34
5.1.4	Modelo do <i>drift</i>	35
5.2	Resíduos e valores ajustados	35
5.3	Intervalos de previsão	36
5.3.1	Intervalos de previsão de várias etapas	38

5.4	Avaliação das previsões	38
5.4.1	Erros de previsão	39
5.4.2	Medidas de acurácia	39
5.5	Métodos de suavização exponencial	40
5.5.1	Estimação de modelos ETS	41
5.6	Modelos ARIMA	42
6	MATERIAL E MÉTODOS	44
6.1	Base de Dados	44
6.2	Pré-processamento	45
6.3	Vizualização	45
6.4	Especificação, Estimação e Verificação	45
7	RESULTADOS	46
7.1	Produção nacional	46
7.1.1	Gás natural	46
7.1.2	Petróleo	47
7.2	Consumo nacional	49
7.2.1	Gasolina	49
7.2.2	Diesel	50
7.2.3	Biocombustíveis	51
7.3	Refino nacional	52
7.4	Preços	53
7.5	Energias renováveis	54
8	DISCUSSÃO	56
8.1	Trabalhos Futuros	57
9	CONCLUSÃO	59
	REFERÊNCIAS	59

1 INTRODUÇÃO

Dados são eventos recolhidos e normalmente armazenados. Informação é o dado investigado e com algum significado. O conhecimento é a informação distinguida, compreendida e empregada a uma finalidade. Esses dados obedecem um ciclo, o ciclo de vida dos dados, composto por: produção, armazenamento, transformação, análise e descarte. A ciência de dados compreende todo o ciclo, ao passo que a estatística, tanto a descritiva como a inferencial, está relacionada, sobretudo, à etapa de análise de dados. Apesar deste trabalho se concentrar em análise de dados é importante ressaltar a importância crescente de uma nova onda industrial que está envolvida com grandes quantidades de dados [1].

A revolução industrial foi uma época da história da humanidade cujos progressos técnicos e metodológicos apresentaram alterações relevantes para a sociedade. Sua primeira etapa data da metade do século XVIII até metade do século XIX e, se embasou no emprego do carvão, ferro e transformações em formas de produção. Na sua segunda fase, compreendida entre 1860 e 1945, destaca-se o início da exploração petrolífera, a utilização da energia elétrica, o emprego do aço e avanços das condições de trabalho. A terceira fase, iniciada após o fim da Segunda Guerra Mundial, foi determinada pelo desenvolvimento da capacidade produtiva com o uso da tecnologia e comunicação [1].

Atualmente o elevado volume de dados caracteriza uma transformação social e cultural, refletindo-se na nova fase da revolução industrial. Mas não é só o volume que representa dados ultimamente, a velocidade, variedade, veracidade e valor são propriedades também consideradas durante o ciclo de vida. Além disso, a redução no custo de armazenamento, miniaturização de componentes e o incremento da capacidade computacional levaram à propagação de equipamentos, dispositivos, e processos capazes de produzir e armazenar uma quantidade cada vez maior de dados. Esse processo de datafication (registro eletrônico de um fenômeno qualquer), associado com eventos de armazenamento e consequentemente reprodução e análise, podem oferecer muitas oportunidades [1].

Para o mundo empresarial estas oportunidades poderão apresentar-se de duas formas: vantagem competitiva ou criação de produtos e/ou serviços orientados a dados. Para o meio acadêmico, a análise de dados de séries temporais do setor de petróleo e gás natural pode ser um instrumento que aplica algum tipo de transformação nos dados em busca de conhecimento. Previsões de séries temporais podem ser importantes em várias situações de tomada de decisão, como por exemplo, decidir a construção de uma nova usina de geração de energia nos próximos cinco anos a partir de previsões de demanda futura; ou a quantidade de suprimentos de um estoque requer previsões de requisitos do próprio estoque [8].

1.1 Objetivo Geral

O objetivo deste trabalho é utilizar essas perspectivas de análise e previsão de séries temporais e transportá-las para o setor de petróleo e gás natural.

1.2 Objetivos Específicos

- Definir modelos que possam auxiliar em previsões e tomadas de decisão;
- descrever o comportamento das principais séries temporais do setor do petróleo e gás.
- verificar a existência de tendências, ciclos, variações sazonais nos dados.

2 BREVE HISTÓRICO DO SETOR DE PETRÓLEO E GÁS

Antes de analisar os dados provenientes do setor do petróleo e gás, se faz necessário uma breve recapitulação do histórico de acontecimentos que marcaram não só a indústria petrolífera, mas também a civilização acerca da busca por fontes de energia. Esse paralelo é importante uma vez que séries temporais são analisadas, sobretudo, com intuito de se fazer previsões sobre o futuro de uma determinada variável. E, para isso, muitos modelos analíticos levam em consideração dados do passado.

O setor de petróleo e gás surge em um contexto de evolução natural da humanidade em busca de fontes energeticamente mais densas. Esse dilema entre energia e sociedade molda a evolução das sociedades humanas, evolução essa, caracterizada por uma crescente complexidade dos arranjos sociais, arranjos produtivos e uma maior qualidade de vida para um número crescente de pessoas. De uma perspectiva biofísica fundamental, o curso da história pode ser visto como a busca por controlar maiores estoques e fluxos de formas mais concentrados e versáteis de energia e como convertê-los, de maneiras mais acessíveis e com custos menores e mais eficientes, em calor, luz e movimento [16].

A dependência humana de fluxos de energia cada vez mais elevados pode ser vista como uma continuação inevitável da evolução orgânica. A combustão controlada em lareiras, fogões e fornalhas transformava a energia química das plantas em energia térmica. Esse calor tem sido usado diretamente em residências, na fundição de metais, na queima de tijolos, no processamento e acabamento de inúmeros produtos. A queima de combustíveis fósseis tornou todos esses usos diretos tradicionais do calor mais difundidos e mais eficientes. Uma série de invenções fundamentais tornou possível converter a energia térmica da queima de combustíveis fósseis em energia mecânica. Isso foi feito primeiro em motores a vapor e de combustão interna, depois em turbinas a gás e foguetes [16].

A combustão de combustíveis fósseis e a geração de eletricidade criaram uma nova forma de civilização de alta energia, cuja expansão agora abrange todo o planeta e cujas fontes de energia primária agora incluem pequenas, mas em rápido crescimento, novas fontes renováveis, especialmente a solar (aproveitada por dispositivos fotovoltaicos ou na concentração de usinas solares) e eólica (convertida por grandes turbinas eólicas). Por sua vez, esses avanços foram baseados em uma concatenação de outros desenvolvimentos [16].

2.1 Carvão

Um desses desenvolvimentos concatenados foi a utilização do carvão em larga escala como fonte de energia. A utilização do carvão como combustível está relacionado intimamente com o conceito de densidade de energia em combustíveis. Esse conceito faz uma relação entre a massa de um corpo com a energia nele armazenada. Ou seja,

quanto mais alta a densidade de energia, mais energia pode ser armazenada ou transportada pela mesma quantidade de massa. Os tipos de carvão, assim como posteriormente, os óleos brutos e gases naturais, apresentam boas quantidades de densidade de energia. Dessa forma, apesar de algumas civilizações antigas terem utilizado o carvão como combustível ocasionalmente, apenas durante as sociedades modernas e o desenvolvimento da Revolução Industrial o carvão passou a ser utilizado em larga escala em cidades e linhas produtivas. A Tabela 2.1 apresenta a densidade de energia de alguns tipos de carvão em comparação com óleos brutos e gases naturais.

Tabela 2.1: Densidade de energia em tipos de carvão, óleos brutos e gases naturais.

Combustível	Densidade de Energia	
	MJ/kg	MJ/m³
Carvões		
Antracitos	31 a 33	
Carvões betuminosos	20 a 29	
Lignites	8 a 20	
Óleos brutos	42 a 44	
Gases naturais		29 a 39

Fonte: Smil, 2017.

O carvão foi muito importante no início das cidades modernas, mas com ele também vieram suas vantagens e desvantagens. A produtividade e a prosperidade das linhas produtivas ganharam um reforço energético importante com a disseminação do carvão na Europa. Fora isso, as invenções dos motores a vapor potencializaram a força de trabalho das grandes fábricas, o que acelerou o processo de urbanização e o transporte de passageiros com locomotivas a carvão. Essa máquina foi o primeiro conversor prático, econômico e confiável da energia química do carvão em energia mecânica, o primeiro motor inanimado energizado por combustível fóssil. No início do século XVIII, os primeiros motores possibilitavam apenas movimento alternativo adequado para bombeamento, mas antes de 1800 surgiram projetos que ofereciam movimento rotativo mais prático. Contribuições de cientistas e engenheiros importantes como Denis Papin's (1647–1712), James Watt (1736–1819) entre outros aumentaram a popularização desses motores pela Europa e posteriormente pelo mundo [16].

Dentre as desvantagens, pode-se destacar a poluição atmosférica das grandes cidades dependentes do carvão, que era um problema constante a saúde das pessoas. Além disso, a densidade de energia dos primeiros tipos de carvão utilizados para aquecimento, cozinha e manufaturas artesanais demandavam a exploração de áreas arborizadas de pelo menos 30 a 50 vezes o tamanho dessas cidades. Grandes cidades pré-industriais localizadas em clima temperado e que dependiam fortemente de carvão vegetal (Pequim na China, por exemplo) teriam exigido uma área arborizada pelo menos de 100 vezes seu

tamanho para garantir um fornecimento contínuo desse combustível [16].

Claramente por questões sustentáveis, por questões de saúde coletiva, pelo desenvolvimento de novas invenções e pela busca de novas fontes energéticas mais densas, o carvão foi sendo transicionado pelo petróleo e gás natural de maneira gradativa, mas ainda atualmente é considerado uma das fontes energéticas fundamentais para o sistema produtivo em que se insere as civilizações contemporâneas.

2.2 Petróleo e gás natural

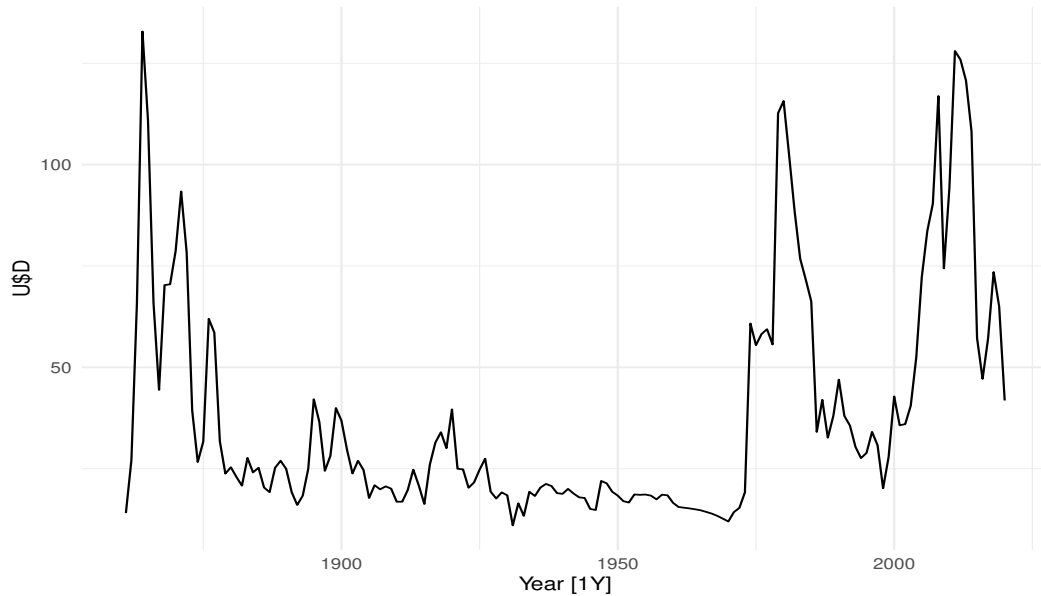
Devido a esses problemas e devido a expansão das sociedades modernas, novos combustíveis fósseis foram sendo reaproveitados por causa de sua maior densidade energética e também em razão de dois fatores principais. O primeiro fator foi a invenção, o desenvolvimento e, eventualmente, a difusão em massa de novas maneiras de converter combustíveis fósseis: introduzindo novos motores principais - começando com motores a vapor e progredindo para motores de combustão interna, turbinas a vapor e turbinas a gás - e criando novos processos para transformar combustíveis brutos, incluindo a produção de coque metalúrgico a partir de carvão, o refino de petróleo bruto para produzir uma ampla gama de líquidos e materiais não combustíveis e o uso de carvão e hidrocarbonetos como matéria-prima em novas sínteses químicas [16].

O segundo fator foi a utilização de combustíveis fósseis para produzir eletricidade, um tipo inteiramente novo, para época, de energia comercial. Qualquer combustível sólido, líquido ou gasoso podia ser queimado, o calor liberado usado para converter água em vapor usado para girar turbogeradores e produzir eletricidade. A tendência de longo prazo era clara: convertia-se uma parcela cada vez maior de combustíveis fósseis em eletricidade e também expandia-se as capacidades de geração de eletricidade primária uma vez que esta fonte de energia era a mais conveniente, mais versátil e, no momento, de seu uso, a forma mais limpa de energia moderna [16].

Logo, em razão desses fatores, a popularização do petróleo e gás natural como fontes de energia primária parece uma evolução natural. O Petróleo, apesar de conhecido desde 200 A.C. pelos chineses, e utilizado em algumas aplicações pontuais de aquecimento, só foi utilizado em grande proporção a partir do século XIX. E, ainda que o ocidente prefira como marco inicial da extração e utilização de petróleo bruto em larga escala a narrativa da corrida do petróleo da Pensilvânia (Pennsylvania oil rush ou Pennsylvanian oil boom), nos Estados Unidos, sabe-se que a primeira fábrica comercial de destilação de petróleo do mundo foi construída pelos russos em 1837, onde hoje se encontra o território do Azerbaijão. Em território norte-americano, porém, o primeiro poço cavado manualmente em 1858 se encontrava no Canadá, na região de Oil Springs. Entretanto, como a busca dos Estados Unidos por petróleo era motivada por um produto substituto do óleo de baleia,

muito caro na época, convencionou-se por motivos de hegemonia que, 27 de agosto de 1859, data em que perfuradores encontraram óleo a uma profundidade de 21 m em Titusville, Pensilvânia, fosse o início da era moderna do petróleo [16]. A Figura 2.1 demonstra bem como marcos históricos importantes no setor de petróleo e gás, influenciaram a volatilidade dos preços médios anuais do barril do petróleo entre os períodos de 1861 a 2020.

Figura 2.1: Preços médios anuais do barril do petróleo entre os períodos de 1861 a 2020.



Fonte: BP Statistical Review of World Energy, 2021.

Nota-se os reflexos da corrida do petróleo na Pensilvânia nos preços médios anuais do barril no início da série temporal, os valores apresentam uma alta significativa (cerca de 130 USD, corrigindo para valores atuais), devido, sobretudo, descobertas de primeiras reservas na América do Norte. Vale destacar outros marcos importantes que afetaram diretamente a oscilação de preços desta série. Durante a década de 1860, apenas três países despontaram como produtores iniciais de petróleo, o Canadá, Os Estados Unidos e a Rússia. O Canadá logo viu sua produtividade baixar sem novas descobertas nas primeiras décadas e, volta a ser um polo importante apenas após a reconstrução pós Segunda Guerra Mundial. Já os Russos, em 1883, abriam exportações de óleo para a Europa após massivos investimentos estrangeiros, fato que auxiliava a expansão de sua nascente indústria petrolífera.

Nos Estados Unidos as descobertas também foram impactantes, sobretudo nos estados da Califórnia, Kansas, Oklahoma e mais importante do campo Spindletop, em 1901, no sul do Texas com capacidade de 100.000 barris/dia. Outras descobertas substanciais de petróleo foram feitas na Indonésia (Sumatra em 1883), no México em 1901 e em 1908 ocorreu a primeira grande descoberta do Oriente Médio no Irã. Além disso, o campo gigante de Mene Grande, na Venezuela, na costa do Lago Maracaibo,

começou a ser entregue em 1914 [16].

O período de reconstrução Pós-Guerra Mundial foi acompanhado também de referências históricas. Cabe destacar as disputas no Irã desde o começo do século XX, quando o governo persa destinava a concessão de exploração de petróleo a recém estabelecida Companhia Petrolífera Anglo-Iraniana (*Anglo-Iranian Oil Co.*), que mais tarde, em partes, tornaria-se a *British Petroleum*. O acordo inicial entre as duas nações era mais vantajoso para os ingleses que detinham 51 % das ações e regalias legais. Isso foi modificando-se a medida que uma postura mais nacionalista do Irã foi se fortalecendo no país, até que em 1932, o xá Reza Pahlevi anulou os privilégios, exigindo maior participação nos lucros.

Com o fim da Segunda Guerra, novas propostas de parcerias foram feitas por outros países, como Holanda, Rússia e Estados Unidos para exploração do petróleo no Irã. O governo Inglês tentou uma nova negociação com o atual primeiro-ministro, mas aconteceu seu assassinato antes da concretização dos novos termos. Dessa forma, um novo primeiro ministro nacionalista chamado Mossadegh foi nomeado, e com ele, a nacionalização da indústria petrolífera em 1951.

A crise de Suez ou guerra do Sinai, em 1956, foi outro fato marcante na oscilação dos preços do petróleo no mercado internacional. Assim como a guerra dos 6 dias e seu contra-ataque, a guerra de Yom Kippur. Ambas culminaram em 1973 no embargo produtivo realizado pela OPEP (Organização dos Países Exportadores de Petróleo), de maioria Árabe, em retaliação à Israel e seus apoiadores. Esse embargo foi chamado de 1ª crise do petróleo e em poucos anos o preço do barril saltou mais de 40 dólares no mercado internacional. Além disso, a Revolução Iraniana em 1979 acrescentou mais instabilidade aos preços.

Por fim, vale mencionar também a invasão do Iraque ao Kuwait em 1990, a crise financeira asiática em 1997, a invasão dos Estados Unidos ao Iraque em 2003 e a Primavera Árabe entre 2010 e 2012. Apesar de alguns eventos aumentarem o nível dos preços e outros os reduzirem, no começo da década de 2010 o saldo positivo nos preços do barril do petróleo era perceptivo, sendo comercializado a cerca de 100 dólares. Dessa forma, constata-se que séries financeiras são altamente voláteis e que os preços praticados, sobretudo na cotação do barril do petróleo, dependem de fatores geopolíticos e históricos relevantes para mudanças rápidas em taxas de variação e no comportamento dos dados.

3 PRINCÍPIOS EM SÉRIES TEMPORAIS

Em análises envolvendo séries temporais é comum que o analista faça um plot inicial dos dados. Essa estratégia trivial mostra-se bastante útil para investigação de ligações entre variáveis, observação de padrões, verificação de taxas de variação no comportamento dos dados e caracterização preliminar de informações aberrantes (*outliers*). Dessa forma, um gráfico notório para séries é o próprio gráfico de tempo. A Figura 3.1 representa o gráfico de tempo da cotação diária do barril de petróleo BRENT entre os períodos de 04/11/2011 até 25/10/2021.

Figura 3.1: Gráfico de tempo da cotação diária do barril de petróleo BRENT entre 04/11/2011 até 25/10/2021.



Fonte: BP Statistical Review of World Energy, 2021.

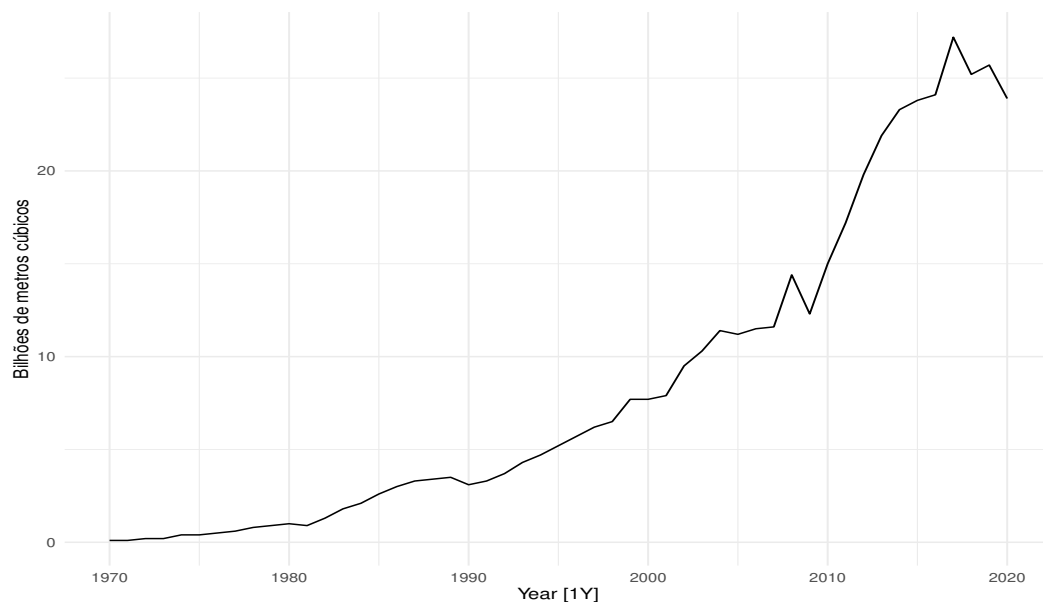
O gráfico de tempo imediatamente revela algumas características interessantes. Existiu um período entre final de 2011 e final de 2014 em que o barril de petróleo se encontrou no patamar acima de 100 USD. Ao final de 2014 e durante 2015 houve uma queda considerável no patamar da cotação, fazendo que o preço do barril alcançasse pontualmente, no começo de 2016, cerca de 30 USD. Isso se deveu ao fato do mercado, na época, ter se assustado com um intenso e inesperado aumento nos estoques de gasolina dos Estados Unidos, que aumentara a sua produção nas áreas do xisto (*shale*). Além disso, os Países da OPEP, se recusavam, em novembro de 2014, a reduzir o teto de produção com objetivo de desestimular a produção de xisto nos EUA, uma vez que a extração era mais cara. Por fim, soma-se a ao cenário tensões geopolíticas envolvendo um anúncio de teste de bomba de hidrogênio pela Coréia do Norte e uma desaleração da economia chinesa. Tudo isso gerou forte influência na queda de preços no período.

Com a regularização da oferta de petróleo no mercado, a cotação do barril apresentou uma recuperação em 75 USD ao final de 2018, se mantendo em 60 USD em 2019. Por fim, é possível ver os efeitos da pandemia de COVID-19 no preço do barril em 2020, que apresentou a mínima do período em março e abril de 2020. Cabe ressaltar, a recuperação dos preços no decorrer de 2021 com as boas notícias de vacinação em massa e controle da situação pandêmica, resultando na recuperação do patamar equivalente ao final de 2018 em outubro de 2021.

3.1 Padrões em séries temporais

Os principais padrões envolvendo séries temporais são tendência (*trend*), sazonalidade (*seasonality*) e ciclo (*cycle*). Uma tendência acontece quando há um incremento ou redução duradouros nos dados. Esse comportamento pode ser linear ou não linear. É comumente relacionada com uma mudança de direção, ou seja, quando há uma transição de tendência crescente para uma tendência decrescente, por exemplo. A Figura 3.2 apresenta a série produtiva de gás natural no Brasil entre 1970 e 2020, nessa série, claramente pode-se observar um tendência crescente nas produções anuais.

Figura 3.2: Gráfico de tempo da série produtiva de gás natural no Brasil entre 1970 e 2020.

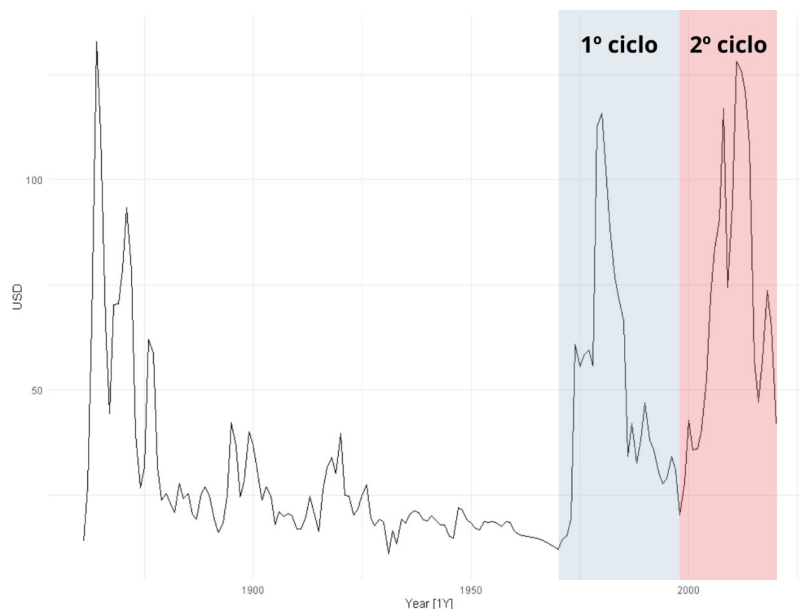


Fonte: BP Statistical Review of World Energy, 2021.

Um padrão sazonal ocorre quando uma série temporal é afetada por fatores sazonais, como a época do ano ou o dia da semana. A sazonalidade é sempre de um período fixo e conhecido. Muitas pessoas confundem comportamento cíclico com comportamento sazonal, mas na verdade são bem diferentes. Se as flutuações não são de frequência fixa,

então elas são cíclicas; se a frequência é imutável e associada a algum aspecto do calendário, então o padrão é sazonal. Em geral, a duração média dos ciclos é maior do que a duração de um padrão sazonal, e as magnitudes dos ciclos tendem a ser mais variáveis do que as magnitudes dos padrões sazonais. Ou seja, Um ciclo ocorre quando os dados exibem subidas e descidas que não são de frequência fixa. Essas flutuações geralmente se devem a condições econômicas e geralmente estão relacionadas ao “ciclo de negócios”. A duração dessas flutuações é geralmente de pelo menos 2 anos [8]. A série de preços globais anuais de petróleo entre 1861 a 2020 apresenta prováveis ciclos entre 1970 e 2020, a Figura 3.5 apresenta os prováveis ciclos nos preços.

Figura 3.3: Série de preços globais anuais de petróleo entre 1861 a 2020 com prováveis ciclos.



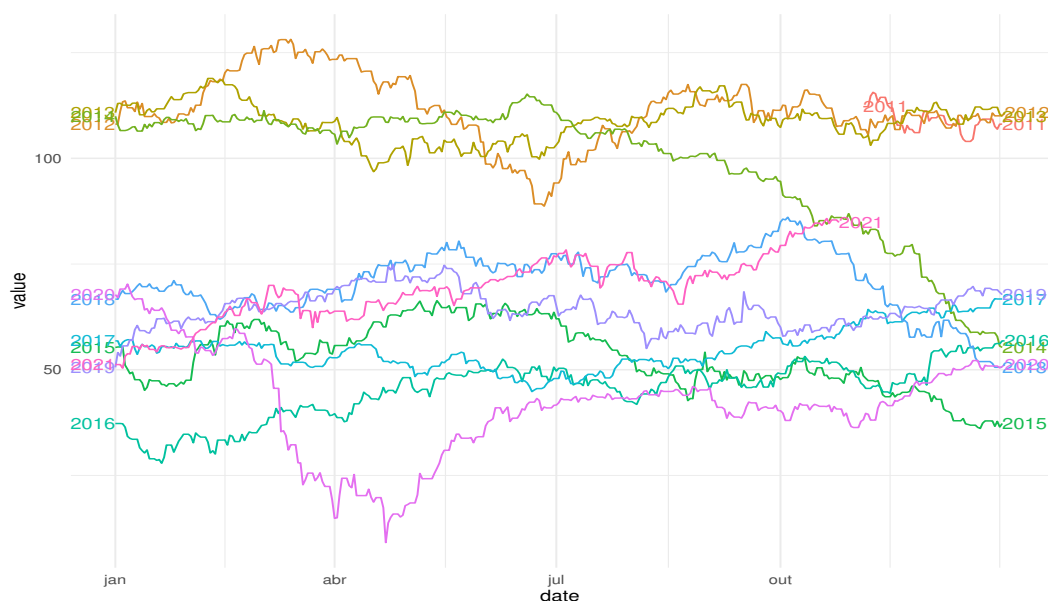
Fonte: Elaborado pelo Autor, 2021.

3.2 Gráficos sazonais

Gráfico sazonal é uma alternativa de visualização de dados que pode informar características interessantes em séries temporais. Um gráfico sazonal é semelhante a um gráfico de tempo, exceto que os dados são demonstrados em relação às sazonalidades individuais nas quais os dados foram observados. Na Figura 3.4 são apresentados os mesmos dados da cotação diária do barril de petróleo BRENT entre os períodos de 04/11/2011 até 25/10/2021, entretanto, agora, os dados estão sobrepostos anualmente em relação aos meses do ano. Um gráfico sazonal permite que o padrão sazonal subjacente seja visto com mais clareza e é especialmente útil na identificação de anos em que o padrão muda.

Nesse caso, fica claro que há uma transição de níveis em direções opostas em 2014 e 2021. Também é possível observar a influência da pandemia de COVID-19 com

Figura 3.4: Gráfico sazonal mensal da cotação do barril de petróleo BRENT entre 04/11/2011 até 25/10/2021.



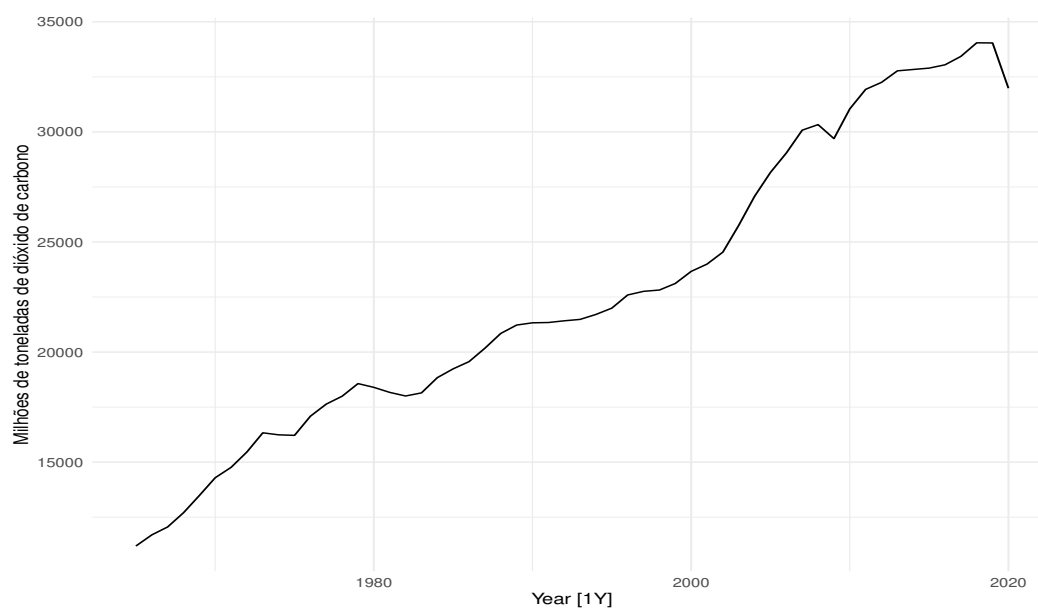
Fonte: Elaborado pelo Autor, 2021.

mais clareza em 2020 e a rápida recuperação do mercado nesse mesmo ano. Fica ainda mais claro as oscilações desse tipo de série temporal financeira no padrão sazonal. Cabe ressaltar um agrupamento de anos em duas faixas de preços principais. A primeira faixa compreende os anos de 2011, 2012, 2013 e o primeiro semestre de 2014 em cerca de 100 dólares o barril, enquanto a segunda faixa compreende os anos de 2015 a 2020 em um intervalo de 40 a 75 dólares. Cabe relevar a queda incomum de 2020 durante a pandemia e uma suposta alta de preços em 2021 devido aos ânimos do mercado com a vacinação global.

3.3 Correlação

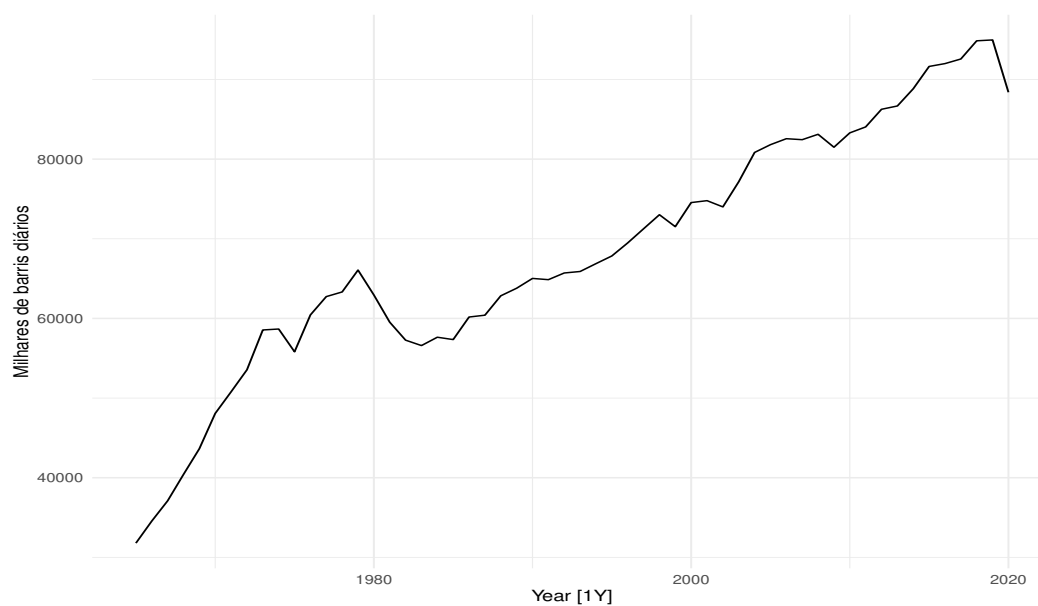
Um outro tipo de visualização importante em análise de séries temporais é o gráfico de dispersão. Esse tipo de visualização de dados é comum quando deseja-se explorar relações entre séries. Por exemplo, as séries de emissões globais de dióxido de carbono e produção de petróleo a nível mundial entre 1965 e 2020 mostradas nas Figuras 3.5 e 3.6 podem estar relacionadas.

Figura 3.5: Emissões globais de dióxido de carbono entre 1965 e 2020.



Fonte: BP Statistical Review of World Energy, 2021.

Figura 3.6: Produção de petróleo a nível mundial entre 1965 e 2020.

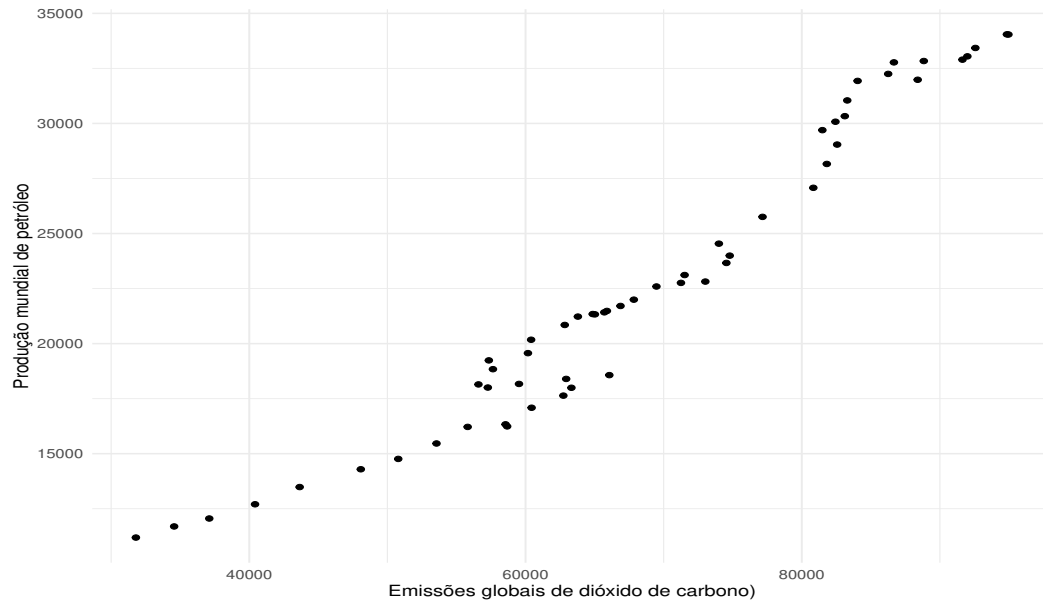


Fonte: BP Statistical Review of World Energy, 2021.

Um gráfico de dispersão pode ser feito relacionando-se produção de petróleo e emissões de dióxido de carbono. A Figura 3.7 demonstra o gráfico de dispersão do forte relacionamento entre as duas séries.

Matematicamente o coeficiente de correlação é a medida que verifica o grau de relação linear entre duas variáveis quaisquer. Esse coeficiente encontra-se em um intervalo de valores compreendidos entre -1 e +1. Quanto mais próximo de +1 ou -1, maior

Figura 3.7: Gráfico de dispersão entre emissões de dióxido de carbono e produção de petróleo.



Fonte: Elaborado pelo Autor, 2021.

é a relação linear diretamente ou inversamente proporcional, respectivamente, entre as variáveis analisadas. Quanto mais próximo de zero, menor é o grau de relacionamento. A Equação 3.1 define o coeficiente de correlação para duas variáveis genéricas x e y .

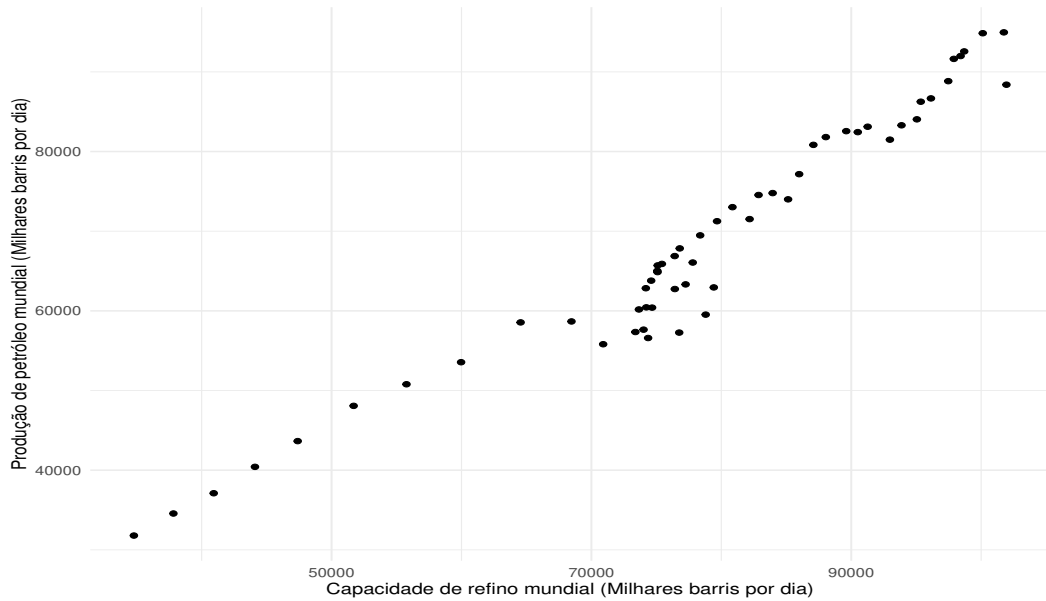
$$r = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum (x_t - \bar{x})^2} \sqrt{\sum (y_t - \bar{y})^2}}. \quad (3.1)$$

O coeficiente de correlação mede apenas a força da relação linear entre duas variáveis e, às vezes, pode ser não vantajoso. Por exemplo, duas variáveis podem não ter uma forte relação linear, mas podem ter uma forte relação não linear e vice-versa [8].

Outra questão importante envolvendo correlação é a não confusão com o conceito de causalidade. Uma variável x pode ser útil para prever uma variável y , mas isso não significa que x está causando y . Até pode ser possível que x esteja causando y , mas pode ser o contrário, que y esteja causando x , ou que a relação entre elas seja mais complicada do que a simples causalidade. Por exemplo, é racional pensar que existe alguma relação de causalidade, além de correlação, entre as séries de capacidade de refino mundial e produção de petróleo mundial. Quanto maior a capacidade de refino, maior será a produção de petróleo mundial? Ou o contrário?. A Figura 3.8 mostra essa relação entre as séries citadas com coeficiente de correlação de 0,968.

Qualquer que seja a confusão que torna difícil determinar quais variáveis estão causando mudanças em outras variáveis, o fato é que não necessariamente torna difícil a previsão. É importante entender que as correlações são úteis para a previsão, mesmo

Figura 3.8: Gráfico de dispersão entre capacidade de refino global e produção de petróleo global.



Fonte: Elaborado pelo Autor, 2021.

quando não há relação causal entre as variáveis, ou quando a causalidade corre na direção oposta ao modelo.

3.4 Autocorrelação

Como correlações entre séries distintas são importantes para modelos preditivos, então parece sensato pensar que as correlações entre os períodos de uma mesma série podem também guardar relações relevantes entre si, esse é o conceito por trás da autocorrelação. A autocorrelação mede a relação linear entre os valores defasados (*lagged values*) de uma série temporal [8].

Segundo [10], antes de se aprofundar nas propriedades da função de autocorrelação é necessário entender o que é um processo estocástico estacionário. Sendo assim, um processo estocástico X_t é uma família de variáveis aleatórias indexadas por $t \in \tau$ e definidas em algum espaço de probabilidade dado. Assim, τ denota um conjunto de índices ordenados que é tipicamente identificado com o tempo. Logo, denomina-se por série temporal a realização ou trajetória (observações ou dados), de um processo estocástico.

Já o conceito de estacionariedade remete ao conhecimento matemático da função de autocovariância do processo estocástico X_t . Como dito, um conhecimento importante na análise de séries temporais é que as realizações em diferentes períodos estão relacionadas entre si. Essa dependência temporal pode ser representada por um modelo explícito ou, de forma descritiva, por covariâncias, e respectivamente, correlações [10].

Dessa forma, segundo [10], "seja X_t um processo estocástico com $\mathbb{V}X_t < \infty$ para todo $t \in \mathbb{Z}$ então a função que atribui a quaisquer dois períodos de tempo t e s , $t, s \in \mathbb{Z}$, a covariância entre X_t e X_s é chamada de função de autocovariância de X_t . A função de autocovariância é denotada por $\gamma_x(t, s)$ ". Esta função é dada por

$$\gamma_x(t, s) = \text{cov}(X_t, X_s) = \mathbb{E}[(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)] = \mathbb{E}X_t X_s - \mathbb{E}X_t \mathbb{E}X_s. \quad (3.2)$$

Dessa forma, um processo estocástico X_t é dito fracamente estacionário se e somente se para todos os inteiros r , s e t as seguintes propriedades são válidas:

- $\mathbb{E}X_t = \mu$ (*constante*);
- $\mathbb{V}X_t < \infty$;
- $\gamma_x(t, s) = \gamma_x(t + r, s + r)$.

No entanto, na prática é mais conveniente observar os coeficientes da autocorrelação em vez das autocovariâncias. A função de autocorrelação (ACF) para processos estacionários é definida como:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} \text{ para todos os inteiros de } h, \quad (3.3)$$

onde h é referido como a ordem. No R é comum a representação da existência de vários coeficientes de autocorrelação, correspondentes a cada painel no gráfico de defasagem. Por exemplo, r_1 mede a relação entre y_t e y_{t-1} , r_2 mede a relação entre y_t e y_{t-2} e assim por diante. No R, o valor de r_k é dado por

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, \quad (3.4)$$

onde T é o comprimento da série temporal.

3.5 Ruído branco

Se o processo estocástico formador da série temporal não apresentar autocorrelação então essa série pode ser caracterizada como um ruído branco (*white noise*). Esse conceito é muito importante em muitas áreas do conhecimento uma vez que o processo de ruído branco é considerado o bloco de construção mais simples, sendo bastante utilizado na construção de outros processos estocásticos.

Em análise de séries temporais o comportamento de ruído branco é importante, por exemplo, no diagnóstico de resíduos de um determinado modelo. Por exemplo, em modelos ARIMA é necessário que o diagnóstico dos resíduos (Os “resíduos” em um modelo de série temporal são o que sobra após o ajuste de um modelo) devem apresentar o comportamento de ruído branco, ou seja, não estarem correlacionados entre si, para que as previsões feitas com esse modelo ajustado não carreguem vieses para o futuro.

Outro papel interessante do ruído branco é na formação de um modelo chamado passeio aleatório (*random walk model*). Esse modelo opera o artifício de utilizar a série diferenciada, que é a subtração entre observações consecutivas na série original, $y'_t = y_t - y_{t-1}$, contudo para quando a série diferenciada na verdade é um ruído branco. Ou seja

$$\varepsilon_t = y_t - y_{t-1} \rightarrow y_t = y_{t-1} + \varepsilon_t. \quad (3.5)$$

Os modelos de passeio aleatório são amplamente utilizados para dados não estacionários, principalmente dados financeiros e econômicos. Isso ocorre porque modelos de passeio aleatório têm longos períodos de tendências aparentes para cima ou para baixo e mudanças repentinas e imprevisíveis na direção, devido sobretudo, ao ruído branco na composição do modelo. Isso é bastante válido uma vez que séries financeiras e econômicas, geralmente, são bem voláteis em relação ao comportamento dos dados [8].

4 DECOMPOSIÇÃO DE SÉRIES TEMPORAIS

A seção 3.1 introduziu alguns conceitos de padrões em séries temporais. Tendência, sazonalidade e ciclos podem ser decompostos, se presentes como padrões em séries temporais.

4.1 Componentes em séries temporais

Segundo [14], um modelo de decomposição consiste em descrever y_t como uma soma de componentes não-observáveis

$$y_t = T_t + S_t + R_t, \quad (4.1)$$

onde T_t é a componente de ciclo-tendência, S_t é a componente sazonal, e R_t é a componente de ruído ou restante (*remainder*), para todo instante de t . Essa é uma abordagem de decomposição aditiva e é bastante utilizada para separação de ciclo-tendência e sazonalidade. Uma variação dessa decomposição é a multiplicativa dada por

$$y_t = T_t \times S_t \times R_t. \quad (4.2)$$

A decomposição aditiva é apropriada se a intensidade das flutuações sazonais, ou a variação em torno do ciclo de tendência, não muda com o nível da série temporal. Quando a variação no padrão sazonal, ou a mudança em torno do ciclo de tendência, parece ser proporcional ao nível da série temporal, então uma decomposição multiplicativa é mais adequada. Decomposições multiplicativas são comuns com séries temporais econômicas [8].

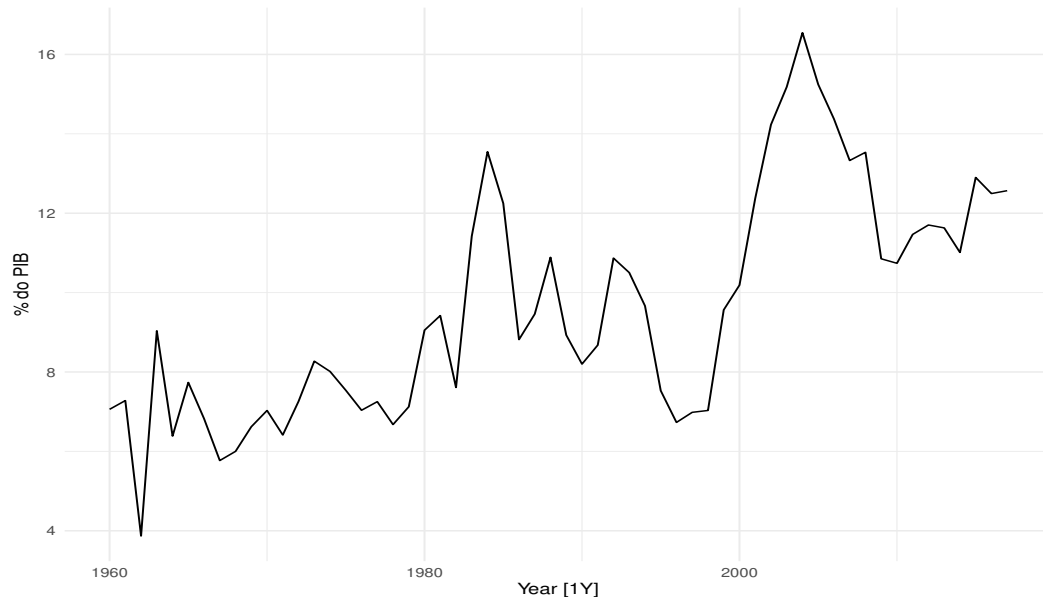
Métodos clássicos de decomposição de séries temporais foram desenvolvidos na década de 1920, mas ainda hoje são utilizados. Logo, é interessante entender como a decomposição clássica de séries temporais funciona e uma primeira etapa para isso é a determinação da componente de ciclo-tendência a partir de suavização de médias móveis.

4.2 Suavização de médias móveis

A suavização de médias móveis é utilizada para determinação da componente de ciclo-tendência porque torna a série temporal muito mais suave e equilibra os movimentos sazonais ou as flutuações de dados aberrantes. A Figura 4.1 mostra exportações brasileiras de bens e serviços em termos percentuais do PIB entre 1960-2017. Exportações de bens e serviços compreendem soja, óleos brutos de petróleo ou de minerais betuminosos, minério

de ferro e seus concentrados, óleos combustíveis de petróleo ou de minerais betuminosos, celulose etc.

Figura 4.1: Exportações brasileiras de bens e serviços em termos percentuais do PIB entre 1960-2017



Fonte: World Bank, 2017.

Um filtro de médias móveis linear de ordem m é dado por

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j}, \quad (4.3)$$

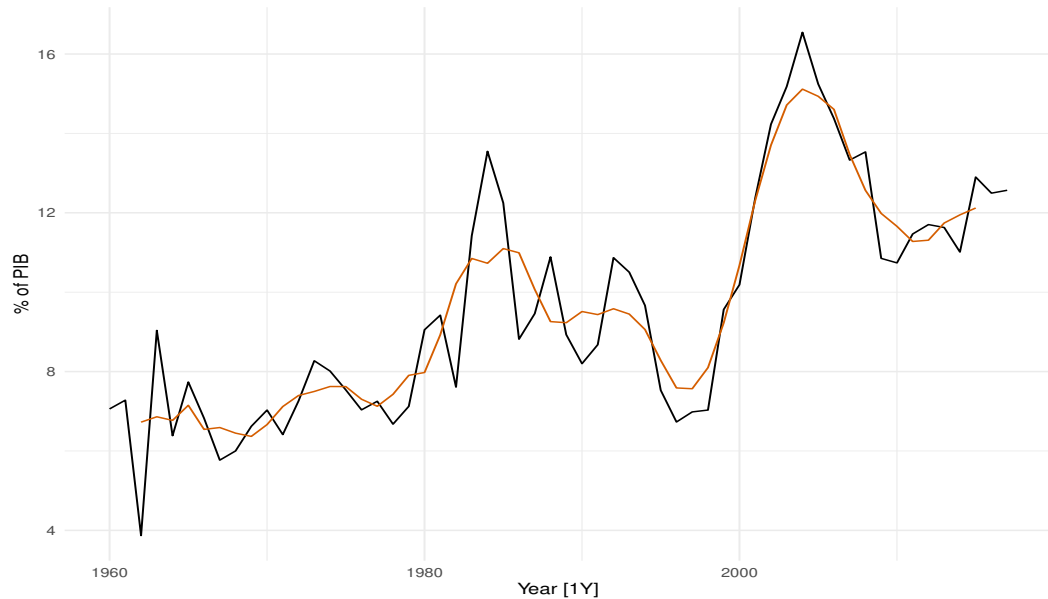
onde $m = 2k + 1$. Isto é, a estimativa da ciclo-tendência no tempo t é obtida pela média dos valores da série temporal dentro de k períodos de t . Observações que estão próximas no tempo também tendem a ter um valor próximo. Portanto, a média elimina parte da aleatoriedade nos dados, deixando um componente de ciclo-tendência suave.

Na Figura 4.2 uma média móvel de ordem 5 é mostrada para as exportações brasileiras de bens e serviços em termos percentuais do PIB entre 1960-2017, fornecendo uma estimativa de ciclo-tendência. O primeiro valor é a média das primeiras cinco observações, 1960-1964; o segundo valor é a média dos valores para 1961-1965; e assim por diante. Ou seja, cada valor é a média das observações na janela de cinco anos centrada no ano correspondente.

4.3 Desvantagens da decomposição clássica

Segundo [8], apesar da decomposição clássica a partir de médias móveis ainda ser amplamente utilizada, isso não é recomendado, uma vez que existem métodos melho-

Figura 4.2: Exportações brasileiras de bens e serviços em termos percentuais do PIB entre 1960-2017 com média móvel de 5 anos.



Fonte: Elaborado pelo autor, 2021.

res. Alguns dos problemas da decomposição clássica são, por exemplo, a estimativa da tendência incompleta nos primeiros e últimos anos da série, já que devido o tamanho da janela da média móvel preestabelecida, não é possível computar valores para as primeiras e últimas observações da série suavizada.

Outra desvantagem é que o método clássico não é robusto para séries que tenham comportamentos não usuais por um determinado período de tempo, por exemplo, um intervalo de tempo sem registros. Além disso, o cálculo da componente sazonal fica restrito a séries em que o padrão sazonal não mude a longo prazo.

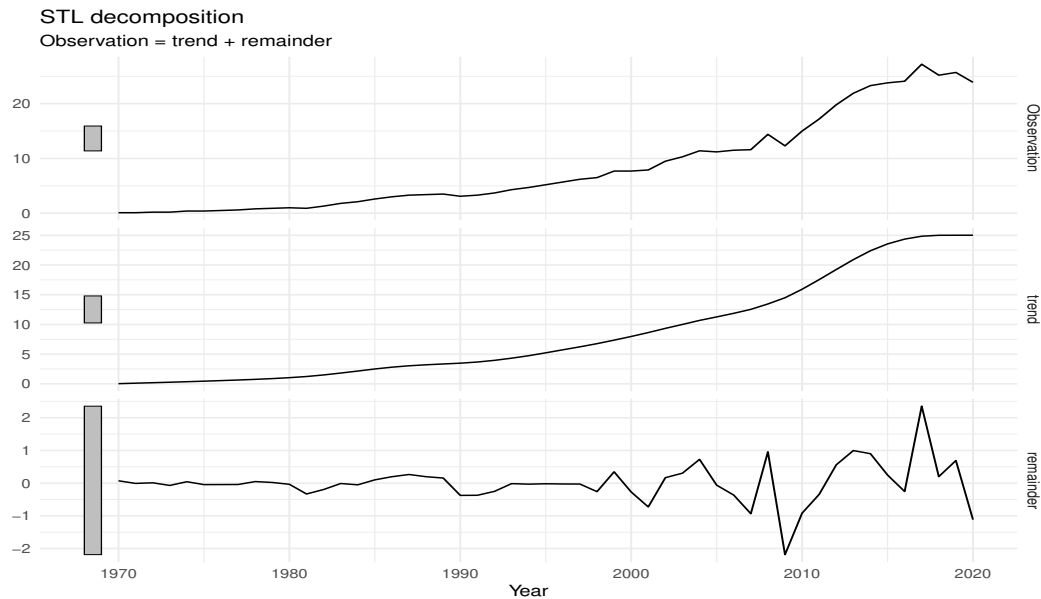
4.4 Decomposição STL

Existem alguns métodos utilizados por Agências oficiais de estatísticas ao redor do mundo para decomposição de séries temporais, cada um com suas vantagens e desvantagens, como por exemplo, os métodos X-11 do *US Census Bureau* do Canadá ou o SEATS do *Bank of Spain*. Contudo, o método de decomposição STL apresenta alguns prós e contras para grande maioria das aplicações.

A técnica de decomposição de séries temporais chamada STL (*Seasonal-Trend Decomposition Procedure Based on Loess*) consiste em uma sequência de operações de suavização, cada uma delas, com uma exceção, emprega a mesma suavização: regressão localmente ponderada ou loess. Essa estratégia de decomposição é robusta e serve para uma ampla gama de dados com vantagens específicas. Em comparação com métodos

clássicos de decomposição, a abordagem STL funciona para qualquer tipo de sazonalidade, pode ser robusta para dados aberrantes e personalizável na escolha das janelas de ciclo-tendência e sazonalidade [4] . A Figura 4.3 representa uma decomposição STL para a série produtiva de gás natural no Brasil de 1970 a 2020.

Figura 4.3: Decomposição STL para a série produtiva de gás natural no Brasil de 1970 a 2020.



Fonte: Elaborado pelo autor, 2021.

Percebe-se que a série analisada não apresenta componente sazonal e que a amplitude da componente de ruído é pequena, ainda que tenha capturado algumas oscilações pontuais, como por exemplo, em 2009 com a crise global e 2020 com a pandemia de COVID-19. A componente de tendência apresenta obviamente o que o gráfico de tempo da série temporal mostra de relevante, um grande acréscimo no nível da série ao longo dos anos, reflexo de uma produção crescente de bilhões de metros cúbicos anuais no Brasil. Fato que parece ter estabilizado a tendência nos anos mais recentes da série, provavelmente um reflexo da contemporânea pandemia no ano de 2020.

5 PREVISÕES COM SÉRIES TEMPORAIS

O processo de produção de previsões em séries temporais obedece algumas etapas. Um fluxograma das etapas de produção de previsões segue as etapas de organização dos dados, visualização, definição de um modelo, estimativa do modelo, avaliação do modelo e cálculo das previsões.

5.1 Modelos preditivos simples

Existem muitos modelos de séries temporais diferentes que podem ser usados para previsão, especificar um modelo apropriado para os dados é essencial para produzir previsões apropriadas. Alguns métodos simples de previsão são inusitadamente eficazes para uma boa gama de comportamentos de dados. Logo, quatro modelos podem se tornar referência para previsões: O método da média, o método Naïve, o método Naïve sazonal e o método com deslocamento (*drift*). Ou seja, quaisquer outros métodos de previsão serão comparados a esses métodos simples para garantir que o novo método seja melhor do que essas alternativas simples.

5.1.1 Modelo da média

A produção de previsões pelo método da média se resume em utilizar a média dos valores históricos como previsões para valores futuros da respectiva série temporal analisada. Isto é, se as observações são definidas como $y_1, \dots, y_T$, então as previsões serão dadas por

$$\hat{y}_{T+h|T} = \bar{y} = \frac{(y_1 + \dots + y_T)}{T}. \quad (5.1)$$

5.1.2 Modelo Naïve

O método ingênuo ou Naïve consiste em apenas repetir a última observação do conjunto de dados formador da série temporal como a próxima estimativa futura. Ou seja, é dado por

$$\hat{y}_{T+h|T} = y_T. \quad (5.2)$$

5.1.3 Modelo Naïve sazonal

Esse método é recomendado para séries que têm altos padrões sazonais. É uma variação do modelo Naïve, mas nesse momento é repetido o último período sazonal da

série analisada, que pode ser o último mês, última semana. como horizonte de previsão do modelo. É dado por

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)}, \quad (5.3)$$

onde $m = \text{período sazonal}$ e k é a parte inteira de $(h - 1)/m$ (ou seja, o número de anos completos no período de previsão antes do tempo $T + h$).

5.1.4 Modelo do *drift*

Outro modelo que é uma variação do método Naïve, mas permitindo que as previsões aumentem ou diminuam ao longo do tempo. Essa quantidade de mudança ao longo do tempo (chamada de *drift*) é definida como a inclinação média vista nos dados históricos. Resumindo, o modelo equivale a traçar uma linha entre a primeira e a última observação da série histórica e extrapolar-la para o futuro. É dado por

$$\hat{y}_{T+h|T} = y_T + h \left(\frac{y_T - y_1}{T - 1} \right). \quad (5.4)$$

A Figura 5.1 representa a série temporal da cotação diária do barril de petróleo BRENT entre 01/02/2019 a 15/04/2019 com previsões para as 10 próximas observações futuras dos modelos de média, Naïve, Naïve sazonal e *drift* e os dados reais do período para critério de comparação de eficiência das estimativas.

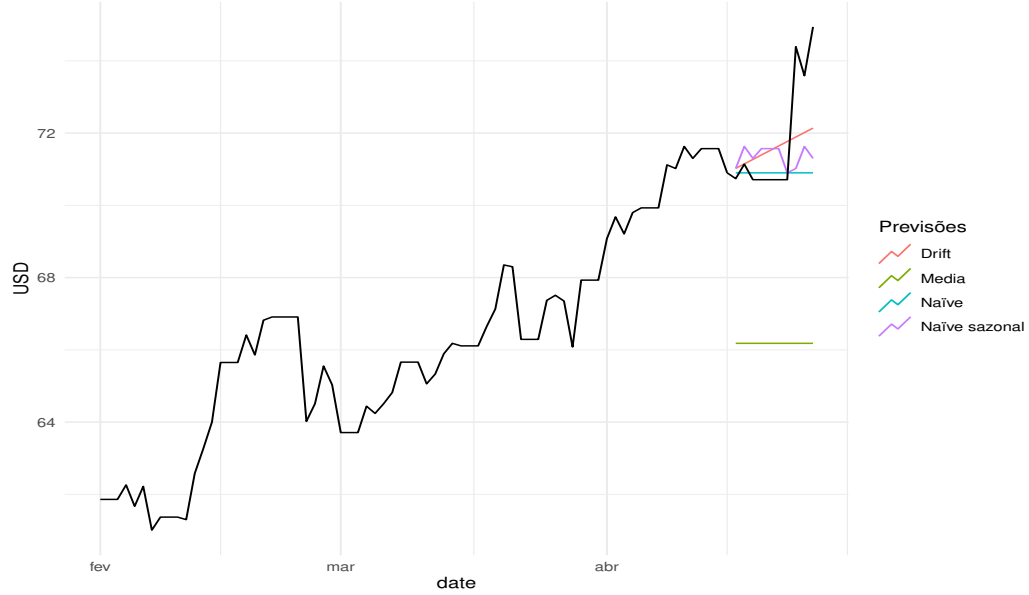
Cabe ressaltar que aparentemente o melhor modelo comparado ao cenário real foi o modelo com *drift* uma vez que capturou a tendência que os dados estavam apresentando até então. Contudo, como a volatilidade dessa série é alta, o ideal é que essa estimativa extrapolada fosse acompanhada de intervalos de confiança para que representasse uma estimativa mais completa. Logo, esses modelos apontam *benchmarks* a serem seguidos, mas nesse caso, modelos preditivos mais robustos podem apresentar melhores previsões.

5.2 Resíduos e valores ajustados

Conceitos muito relevantes quando se trata de especificação, treinamento e validação de um determinado modelo são os de valores ajustados e resíduos do modelo. Por exemplo, quando um modelo é especificado para um determinado conjunto de dados, os valores ajustados são cada observação em uma série temporal podendo ser prevista usando todas as observações anteriores. Isto é, os valores ajustados não são os dados, mas aproximações dos modelos aos dados.

Já os resíduos, como dito, em um modelo de série temporal são o que sobra após o ajuste de um modelo. Ou seja, resíduos são iguais à diferença entre as observações e

Figura 5.1: Cotação diária do barril de petróleo BRENT entre 01/02/2019 a 15/04/2019 com previsões dos modelos de média, Naïve, Naïve sazonal e *drift* e os dados reais do período.



Fonte: Elaborado pelo autor, 2021.

os valores ajustados correspondentes. São úteis para verificar se um modelo capturou adequadamente as informações nos dados.

$$e_t = y_t - \hat{y}_t. \quad (5.5)$$

Bons métodos de previsão possuirão resíduos que não são correlacionados, uma vez que se resíduos apresentarem correlação entre si, significa que há informação não capturada pelo modelo, e dessa forma, as previsões poderão ser enviesadas. Além disso, os resíduos devem ter média zero, caso contrário as previsões também serão tendenciosas. Por isso, busca-se o fato dos resíduos se parecerem com ruído branco em muitos modelos preditivos [8].

Outras características para bons modelos de previsão, mas não obrigatórias, são os resíduos com variância constante (homocedasticidade) e os resíduos serem normalmente distribuídos. Essas duas características estão diretamente relacionadas com os cálculos dos intervalos de confiança.

5.3 Intervalos de previsão

Como previsões são cercadas de incertezas, um artifício para aumentar a eficiência das projeções futuras é baseá-las em uma distribuição de probabilidades.

Segundo Hyndman e Athanasopoulos [8], esse arranjo caracteriza a probabilidade de verificar possíveis valores futuros usando o modelo ajustado. A previsão pontual é a média desta distribuição. A maioria dos modelos de séries temporais produz previsões normalmente distribuídas – ou seja, assume-se que a distribuição de possíveis valores futuros segue uma distribuição normal.

Os intervalos de previsão ou intervalos de confiança fornecem um intervalo dentro do qual espera-se que y_t esteja com uma probabilidade preestabelecida. Por exemplo, supondo que a distribuição de observações futuras seja normal, e com um intervalo de previsão de 80 % para a previsão e horizonte h , o intervalo de previsão é dado por

$$\hat{y}_{T+h|T} \pm 1.28\hat{\sigma}_h, \quad (5.6)$$

onde $\hat{\sigma}_h$ é uma estimativa do desvio padrão da distribuição de previsão de horizonte h . Genéricamente os intervalos de previsão podem ser descritos como

$$\hat{y}_{T+h|T} \pm c\hat{\sigma}_h, \quad (5.7)$$

onde o multiplicador c depende da probabilidade de cobertura (*coverage probability*). É comum quando se trata de previsões os intervalos de 80% e de 95%, mas outras porcentagens podem ser utilizadas. A Tabela 5.1 demonstra outros valores de c , considerando uma distribuição de previsão normal.

Tabela 5.1: Multiplicadores a serem usados para intervalos de previsão.

Porcentagem	Multiplicador
50	0.67
55	0.76
60	0.84
65	0.93
70	1.04
75	1.15
80	1.28
85	1.44
90	1.64
95	1.96
96	2.05
97	2.17
98	2.33
99	2.58

Fonte: Hyndman; Athanasopoulos, 2021.

Segundo [8], o valor dos intervalos de previsão expressam a incerteza nas previsões.

Se apenas previsões pontuais forem produzidas, não há como dizer quão precisas são as previsões. No entanto, se também forem produzidos intervalos de previsão, fica claro quanta incerteza está associada a cada previsão. Por esta razão, as previsões pontuais quase não têm valor sem os intervalos de previsão que as acompanham.

5.3.1 Intervalos de previsão de várias etapas

Como os intervalos de previsão utilizam o desvio padrão da distribuição de previsão ($\hat{\sigma}_h$), então para previsão uma observação à frente ($h = 1$) a estimativa utiliza o desvio padrão dos resíduos, que é dado por

$$\hat{\sigma} = \sqrt{\frac{1}{T-K} \sum_{t=1}^T e_t^2}, \quad (5.8)$$

onde K é o número de parâmetros estimados no método de previsão.

Uma característica comum dos intervalos de previsão é que eles geralmente aumentam em comprimento à medida que o horizonte de previsão aumenta ($h > 1$). Quanto mais à frente se quer prever, mais incerteza estará associada à previsão e, portanto, mais amplos serão os intervalos. Dessa forma, em uma previsão de várias etapas é necessário um método de cálculo mais complicado, que varia com o método de previsão especificado. Esses cálculos assumem que os resíduos não são correlacionados [8]. A Tabela 5.2 traz métodos de cálculo para previsão de várias etapas em quatro métodos de referência.

Tabela 5.2: Desvio padrão para previsão de várias etapas em quatro métodos de referência

Bechmark	desvio padrão de previsão no horizonte h
Média	$\hat{\sigma}_h = \hat{\sigma} \sqrt{1 + 1/T}$
Naïve	$\hat{\sigma}_h = \hat{\sigma} \sqrt{h}$
Naïve sazonal	$\hat{\sigma}_h = \hat{\sigma} \sqrt{k + 1}$
Drift	$\hat{\sigma}_h = \hat{\sigma} \sqrt{h (1 + h/T)}$

Fonte: Hyndman; Athanasopoulos, 2021.

5.4 Avaliação das previsões

Ao escolher modelos, é prática comum separar os dados disponíveis em duas partes, dados de treinamento e dados de teste, onde os dados de treinamento são usados para estimar quaisquer parâmetros de um método de previsão e os dados de teste são usados para avaliar sua precisão. Como os dados de teste não são usados para determinar as pre-

visões, eles devem fornecer uma indicação confiável de quão bem o modelo provavelmente prevê novos dados. Essa subdivisão foi feita na avaliação da série temporal da Figura 5.1.

Alguns pontos são destacáveis quanto ao ajuste de um modelo:

- Um modelo que se ajusta bem aos dados de treinamento não necessariamente fará uma boa previsão;
- Um ajuste perfeito sempre pode ser obtido usando um modelo com parâmetros suficientes, obdecendo o princípio da parcimônia de parâmetros;
- O ajuste excessivo de um modelo aos dados é tão ruim quanto não identificar um padrão sistemático nos dados (problema de *overfitting*).

5.4.1 Erros de previsão

Erro de previsão se refere a imprevisibilidade do acerto das previsões em relação aos valores observados. Ou seja, o erro de previsão pode ser escrito como

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}, \quad (5.9)$$

onde o conjunto de treinamento é composto por $(y_1, \dots, y_T)$ e o conjunto teste por $(y_{T+1}, y_{T+2}, \dots)$. Pode-se medir a precisão das projeções resumindo os erros de previsão em diferentes formas.

5.4.2 Medidas de acurácia

Existem medidas de acurácia que são dependentes das escalas e consequentemente dos erros de previsão e_t . Segundo [8], as duas medidas dependentes de escala mais usadas são baseadas nos erros absolutos ou erros quadrados:

$$\text{Mean Absolute Error} = \text{MAE} = \text{mean}(|e_t|), \quad (5.10)$$

$$\text{Root Mean Squared Error} = \text{RMSE} = \sqrt{\text{mean}(e_t^2)}. \quad (5.11)$$

Ao comparar métodos de previsão aplicados a uma única série temporal ou a várias séries temporais com as mesmas unidades, o MAE é popular por ser fácil de entender e calcular. Um método de previsão que minimize o MAE levará a previsões da mediana, enquanto que minimizar o RMSE levará a previsões da média. Consequentemente, o RMSE também é amplamente utilizado, apesar de ser mais difícil de interpretar.

Existe também métricas que levam em consideração os erros percentuais. O erro percentual é dado por $p_t = 100e_t/y_t$. Logo, uma variação do MAE seria

$$\text{Mean Absolute Percentage Error} = \text{MAPE} = \text{mean}(|p_t|). \quad (5.12)$$

A vantagem desse tipo de métrica é que não apresenta unidades e, por isso, pode ser usada para comparar desempenhos de previsão entre conjuntos de dados.

Por fim, existem métricas que são utilizadas quando diferentes séries temporais têm diferentes unidades de medida. A proposta é escalar os erros de acordo com o MAE de um modelo de previsão simples, nesse caso do método Naïve. O resultado comparativo é

$$q_j = \frac{e_j}{\frac{1}{T-m} \sum_{t=m+1}^T |y_t - y_{t-m}|}, \quad (5.13)$$

com $m = 1$ para série não sazonal e $m > 1$ para série sazonal. Dessa forma uma variação do MAPE é

$$\text{Mean Absolute Scaled Error} = \text{MASE} = \text{mean}(|q_j|). \quad (5.14)$$

5.5 Métodos de suavização exponencial

O método de suavização exponencial visa produzir previsões com base em observações passadas. De acordo com [5] para isso, assume-se que pode haver efeitos sistemáticos de tendência e sazonalidade na geração dos dados. Portanto, é aconselhável atribuir um parâmetro de suavização exponencial ao modelo quando analisar dados históricos. Esse modelo considera o efeito de médias móveis ponderadas exponencialmente decrescente à medida que as observações envelhecem.

O pacote *fable* no R fornece muitos modelos de previsão de suavização exponencial. Esses modelos são organizados de forma que a seleção do melhor método a ser utilizado seja feita de acordo com os principais componentes da série temporal (por exemplo, componentes de tendência e sazonal). Além disso, a maneira como esses componentes são adicionados ao método de suavização (por exemplo, de forma aditiva, multiplicativa ou amortecida) também deve ser considerada. Portanto, cada método é rotulado por duas letras e definem o tipo de tendência e o componente sazonal. Todos os métodos são apresentados na Tabela 5.3.

Alguns desses métodos são conhecidos com outros nomes como pode ser visto na Tabela 5.4.

Tabela 5.3: Composição de métodos de suavização exponencial

Componente de tendência	Componente de sazonalidade		
	N (nenhum)	A (aditivo)	M (Multiplicativo)
N (nenhum)	(N, N)	(N, A)	(N, M)
A (Aditivo)	(A, N)	(A, A)	(A, M)
Ad (Aditivo amortecido)	(Ad, N)	(Ad, A)	(Ad, M)

Fonte: Hyndman; Athanasopoulos, 2021.

Tabela 5.4: Classificação de alguns métodos de suavização exponencial

Abreviatura	Método
(N, N)	Suavização exponencial simples
(A, N)	Método linear de Holt
(Ad, N)	Método de tendência amortecida aditiva
(A, A)	Método de Holt-Winters aditivo
(A, M)	Método de Holt-Winters multiplicativo
(Ad, M)	Método amortecido de Holt-Winters

Fonte: Hyndman; Athanasopoulos, 2021.

Uma terceira letra foi posteriormente adicionada a esta classificação. Esta letra extra refere-se a erros aditivos ou multiplicativos. Isso transformou os métodos da Tabela 5.3 (que gerava apenas previsões pontuais) em modelos estatísticos mais robustos, que podem gerar tanto intervalos de previsão quanto previsões pontuais. Assim, cada modelo de espaço de estados foi rotulado como ETS (Erro; Tendência; Sazonal).

Com essa transformação, um erro e_t com uma certa distribuição de probabilidade é adicionado à equação de suavização. Essa mudança influenciou a equação de observação, a equação de transição e seus respectivos parâmetros de nível, sazonalidade e suavização. A identificação dos intervalos de previsão também muda, dependendo se os erros são aditivos ou multiplicativos. Em outras palavras, essas equações, juntamente com a distribuição estatística dos erros, formam um modelo estatístico completo.

5.5.1 Estimação de modelos ETS

Segundo [14], uma alternativa para estimar os parâmetros minimizando a soma dos erros quadrados é maximizar a verossimilhança (*likelihood*) (para modelos de erro aditivo). A verossimilhança é a probabilidade dos dados surgirem de um modelo especificado. Assim, uma grande verossimilhança está associada a um bom modelo.

Uma grande utilidade de modelos ETS é que os critérios de informação podem ser usados para a seleção do modelo. Esses critérios levam em consideração a estimativa que maximiza a verossimilhança e o número de parâmetros do modelo que é mais apropriado

para uma determinada série temporal.

Para modelos ETS, o Critério de Informação de Akaike (AIC) é definido como

$$AIC = -2\text{Log}(L) + 2k, \quad (5.15)$$

onde L é a verossimilhança do modelo e k é o número total de parâmetros e estados iniciais que foram estimados (incluindo a variância residual). Já o AIC corrigido para viés de amostra pequena (AICc) é definido como

$$AIC_c = AIC + \frac{2k(k+1)}{T-k-1}, \quad (5.16)$$

onde T é o número de observações. O Critério de Informação Bayesiano (BIC) é

$$BIC = AIC + k [\text{Log}(T) - 2]. \quad (5.17)$$

Geralmente na escolha dos melhores modelos, aqueles que tiverem menores valores de AIC_c são os melhores ajustados para previsões [8].

5.6 Modelos ARIMA

Ao contrário dos modelos de suavização exponencial, que usam componentes de tendência e sazonalidade, os modelos ARIMA usam análise de autocorrelação e um teste de raiz unitária para obter o melhor modelo possível para dados de séries temporais. ARIMA significa *Autoregressive Integrated Moving Average* e usa uma abordagem de parametrização baseada na estratégia de Box e Jenkins [2]. O método determina os ordens ARIMA (p, d, q), onde p é a ordem da parte autoregressiva, d é o grau de primeira diferenciação envolvido e q é a ordem da parte da média móvel. A equação para o modelo não sazonal completo é dado por

$$y'_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (5.18)$$

Onde ϕ_p são os pesos autorregressivos, θ_q são os pesos da média móvel e ε_t representa o ruído branco. A constante c (onde μ é o valor médio de y'_t)

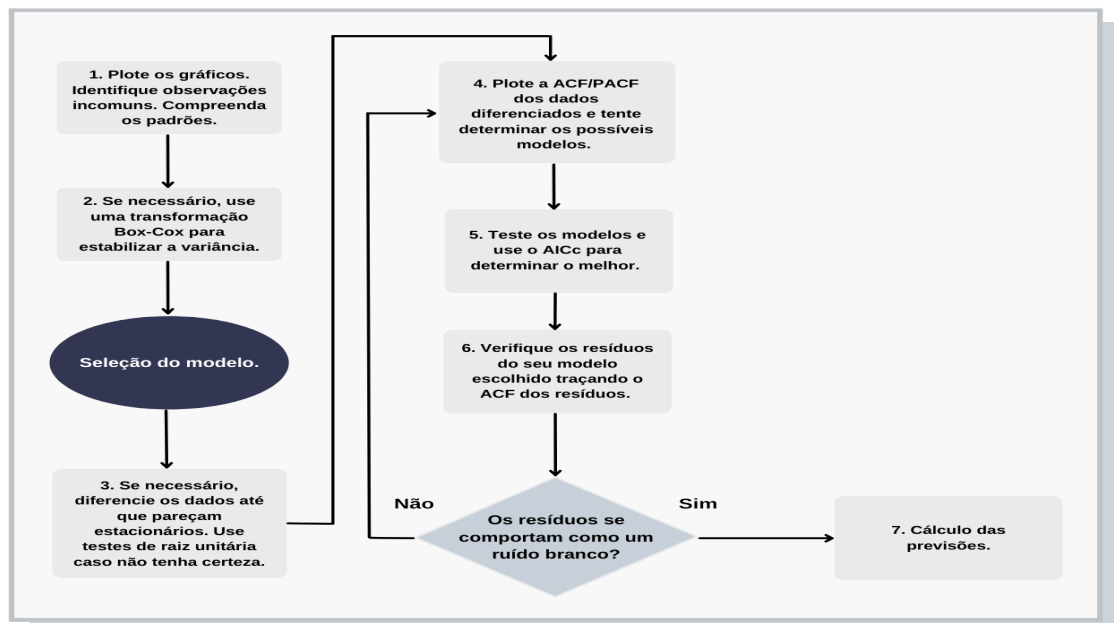
$$c = \mu (1 - \phi_1 - \dots - \phi_p). \quad (5.19)$$

A função `Arima()` do pacote *fable* no R segue a abordagem Box e Jenkins para ajuste de um modelo ARIMA não sazonal. Em primeiro lugar, ele executa um teste

de raiz unitária Kwiatkowski-Phillips-Schmidt-Shin (KPSS) para determinar o valor de d , o grau de primeira diferenciação, que converte uma série temporal não estacionária em estacionária. Após o teste de raiz unitária, as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) são plotadas para que possam ajudar a determinar p e q . Caso sejam insuficientes para identificar p e q , a função Arima () utiliza o critério de informação de Akaike mais baixo e o método de estimação de máxima verossimilhança para calculá-los. Quando os parâmetros são encontrados, é necessário verificar os resíduos antes de poder fazer previsões. Os resíduos do modelo devem se comportar como ruído branco, ou seja, todas as autocorrelações devem estar dentro dos limites. Desta forma, o intervalo de confiança pode ser bem computado, de acordo com [9].

Os resíduos são de grande importância nos modelos ARIMA. Eles devem se comportar como ruído branco porque, segundo [10], “O processo de ruído branco é, portanto, estacionário e temporalmente não correlacionado. Como o ACF não possui estrutura, é impossível fazer inferências de observações passadas para seu desenvolvimento futuro, pelo menos em um cenário de mínimos quadrados com funções de previsão lineares”. Essas características são relevantes, pois, em certo sentido, o processo deve estar em equilíbrio estatístico para produzir intervalos de predição, comportamento enfatizado por [6] .

Figura 5.2: Fluxograma da abordagem Box e Jenkins para ajuste de um modelo preditivo ARIMA.

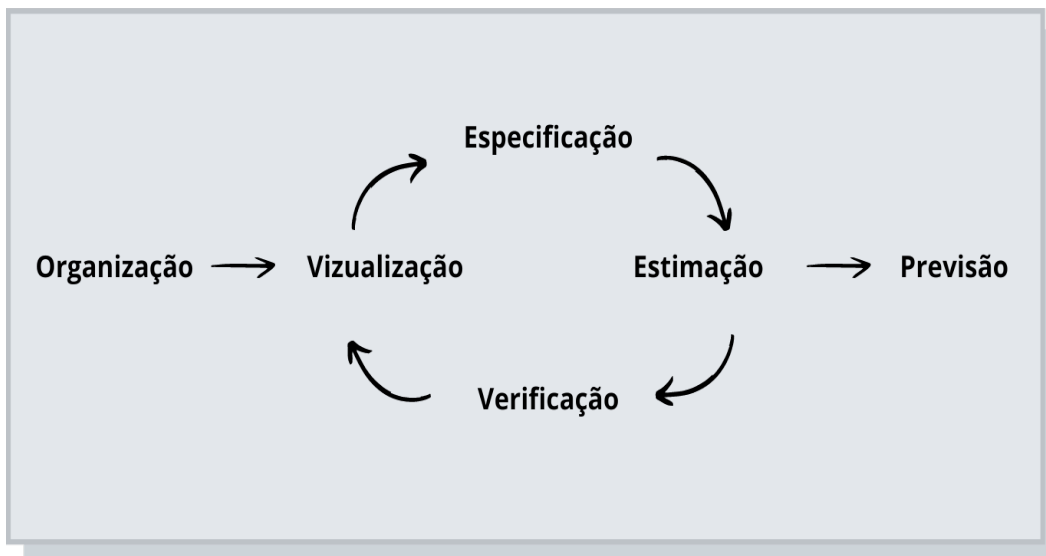


Fonte: Hyndman; Athanasopoulos, 2021.

6 MATERIAL E MÉTODOS

Neste capítulo está descrita toda a metodologia adotada no presente trabalho. A metodologia fundamenta-se sobretudo na produção de previsões através de modelos de séries temporais, conforme o fluxograma da Figura 6.1

Figura 6.1: Fluxograma com metodologia de produção de previsões em séries temporais



Fonte: Hyndman; Athanasopoulos, 2021.

A metodologia baseia-se em análise de dados abertos presentes em anuários estatísticos da *British Petroleum* no seu relatório chamado de *BP Statistical Review of World Energy 2021* [13], utilizado também pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) no seu repositório de dados. Além disso, há pesquisa bibliográfica sobre análise, tendência, sazonalidade, estimação, diagnóstico e previsão com modelos de séries temporais. As simulações computacionais são executadas na linguagem de programação R, que apresenta bons recursos de manipulação de séries de tempo e bibliotecas para múltiplas funcionalidades, inclusive, previsão de dados.

6.1 Base de Dados

Neste trabalho tratamos da base de dados da *British Petroleum* no seu relatório chamado de *BP Statistical Review of World Energy 2021* [13], a qual contém registros de dados de energia primária, petróleo, gás natural, carvão, energia nuclear, hidroeletricidade, energias renováveis, eletricidade e minerais, em alguns casos, há mais de 50 anos. Esses dados, estão reunidos em planilhas com séries temporais diversas, às vezes estratificadas por continentes, países e também de maneira global. Dentre as principais variáveis

de agrupamento cabe destacar consumo, reservas, produção, preços, refino, geração e capacidade.

6.2 Pré-processamento

O processo metodológico deste trabalho se inicia com extração de séries temporais relevantes sobretudo para o contexto em que o Brasil está inserido como produtor de petróleo e gás no cenário mundial. Além disso, abordagem de séries temporais globais de relevância internacional também são analisadas.

Dessa forma o pré-processamento destina-se a filtrar a participação do Brasil em séries produtivas, de consumo, refino, geração e capacidade. Além disso, também é importante a adequação de dados para as análises e previsões, como por exemplo transformação de grandezas, vizualização de dados faltantes etc.

6.3 Vizualização

A intenção é que o diagnóstico de séries temporais possa auxiliar na investigação do procedimento formador da série temporal, a análise pode ainda ser útil para estabelecer previsões de valores futuros, podendo ser de curto prazo ou a longo prazo. O estudo pode ser utilizado para detalhar o desempenho da série, atestando a presença de tendências, ciclos e variações sazonais [14].

6.4 Especificação, Estimação e Verificação

Essas etapas geralmente são agrupadas e são cruciais para boas previsões envolvendo séries temporais. A etapa de especificação é a de definição do modelo (suavização exponencial ou ARIMA) para bons cenário de previsão. A estimação é o treinamento do modelo e a verificação é avaliação do desempenho do modelo a partir de métricas (AIC_c , AIC , BIC , $RMSE$ etc.)

7 RESULTADOS

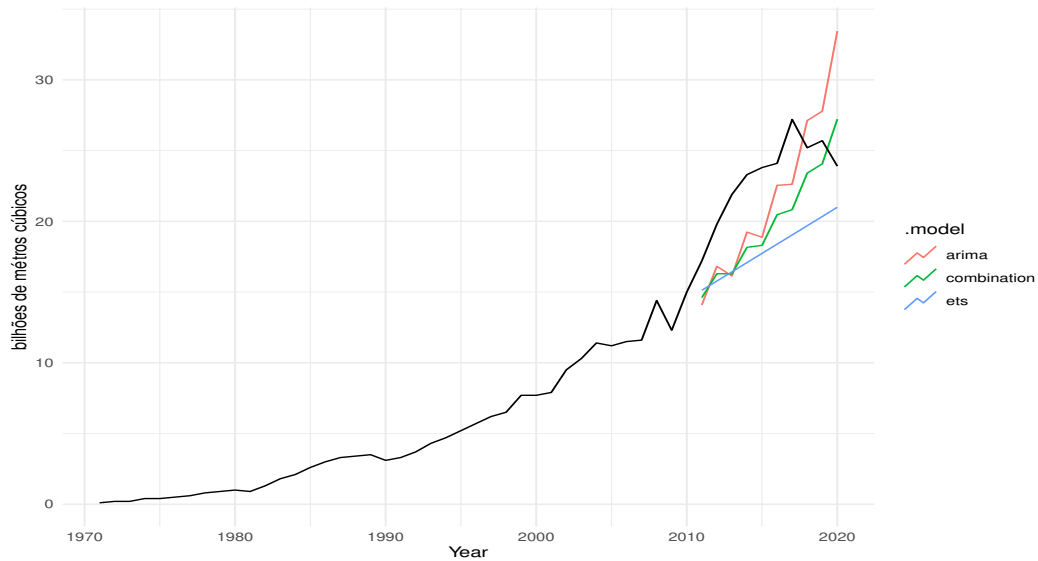
Os resultados serão apresentados de acordo com as variáveis agrupadas no banco de dados: produção, consumo, refino, preços e renováveis.

7.1 Produção nacional

7.1.1 Gás natural

Especificação (Figura 7.1), validação (Figura 7.2) e previsões da série temporal produtiva de gás natural no Brasil de 1970 a 2020 (Figura 7.3). As previsões feitas consideram um horizonte de 5 anos e uma combinação do modelo de suavização exponencial ETS (M,A,N) e ARIMA (1,2,2).

Figura 7.1: Especificação de modelos na série temporal produtiva de gás natural no Brasil de 1970 a 2020.



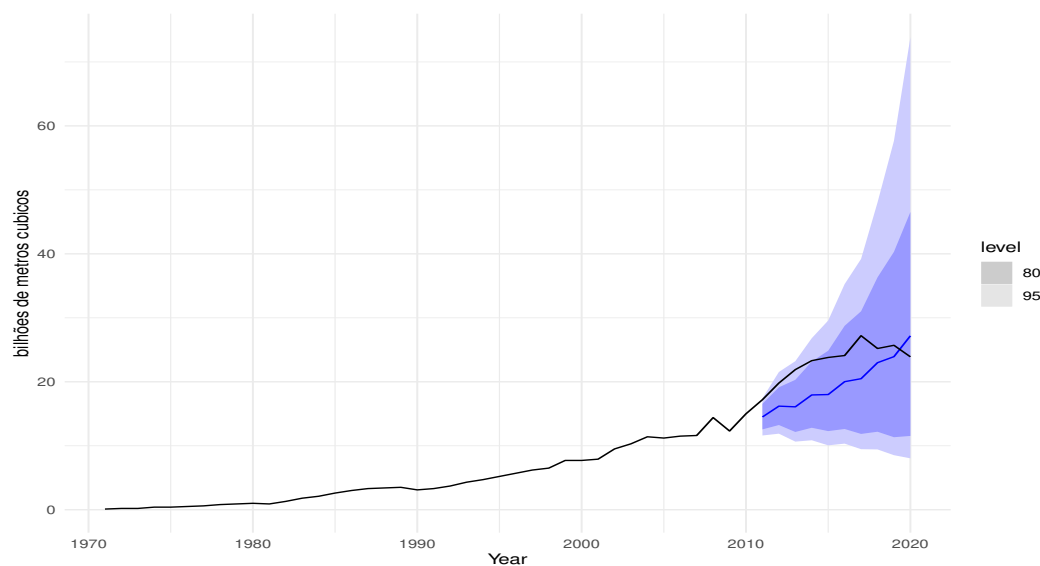
Fonte: Elaborado pelo autor, 2021.

Tabela 7.1: Métricas comparativas entre modelos e dados observados entre 2011 e 2020.

	ME	RMSE	MAE	MPE	MAPE	MASE
Combinação	3.25	4.22	3.91	14.2	16.9	7.63
ARIMA	1.34	4.64	4.05	6.52	17.7	7.91
ETS	5.15	5.41	5.15	21.8	21.8	10

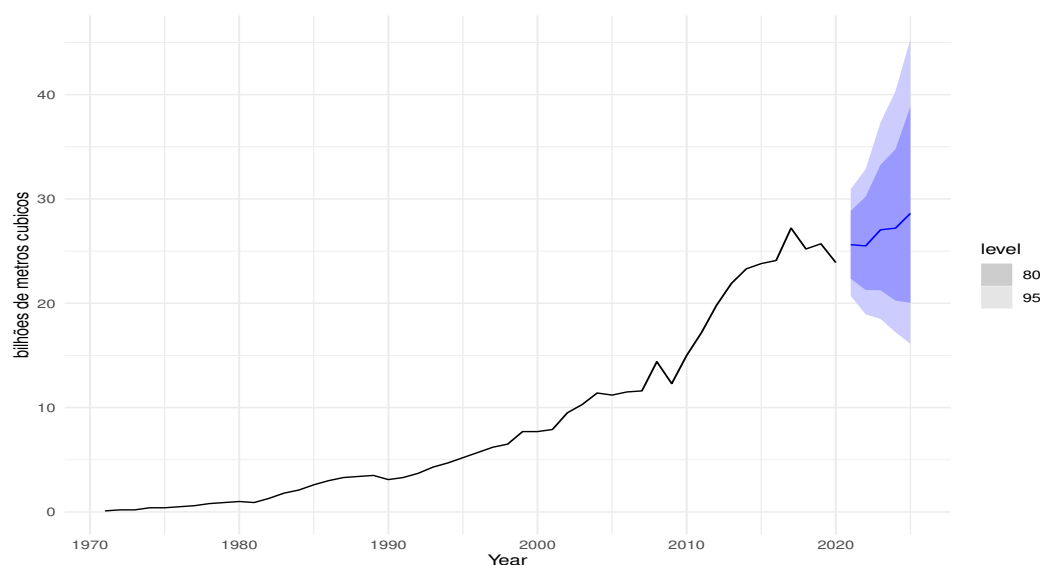
Fonte: Elaborado pelo autor, 2021.

Figura 7.2: Validação dos modelos na série temporal produtiva de gás natural no Brasil de 1970 a 2020.



Fonte: Elaborado pelo autor, 2021.

Figura 7.3: Previsão na série temporal produtiva de gás natural no Brasil com horizonte de 5 anos.

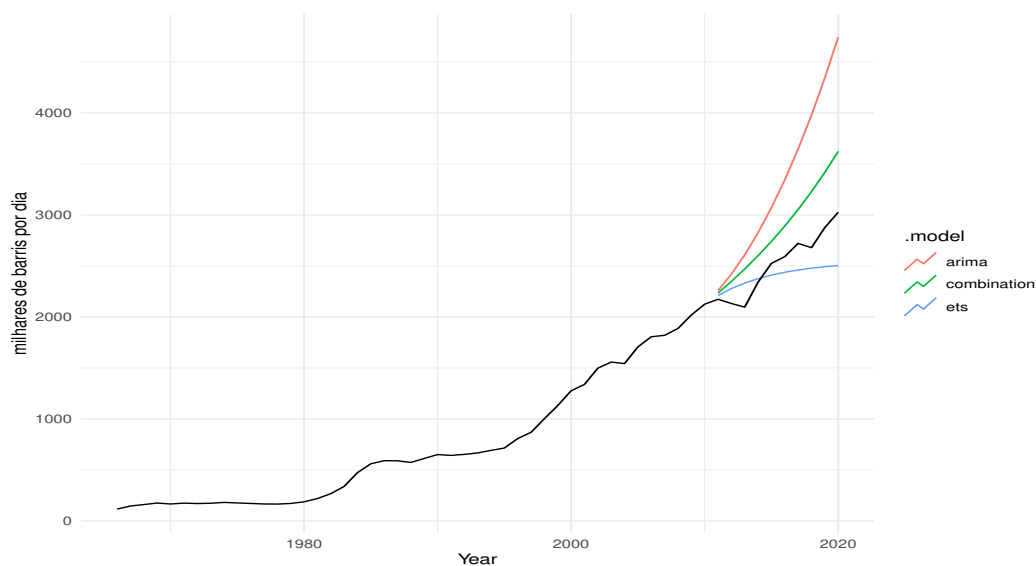


Fonte: Elaborado pelo autor, 2021.

7.1.2 Petróleo

Especificação (Figura 7.4), validação (Figura 7.5) e previsões da série temporal produtiva de petróleo no Brasil de 1965 a 2020 (Figura 7.6). As previsões feitas consideram um horizonte de 5 anos e uma combinação do modelo de suavização exponencial ETS (M,A,N) e ARIMA (1,1,0) com deslocamento.

Figura 7.4: Especificação de modelos na série temporal produtiva de petróleo no Brasil de 1965 a 2020.



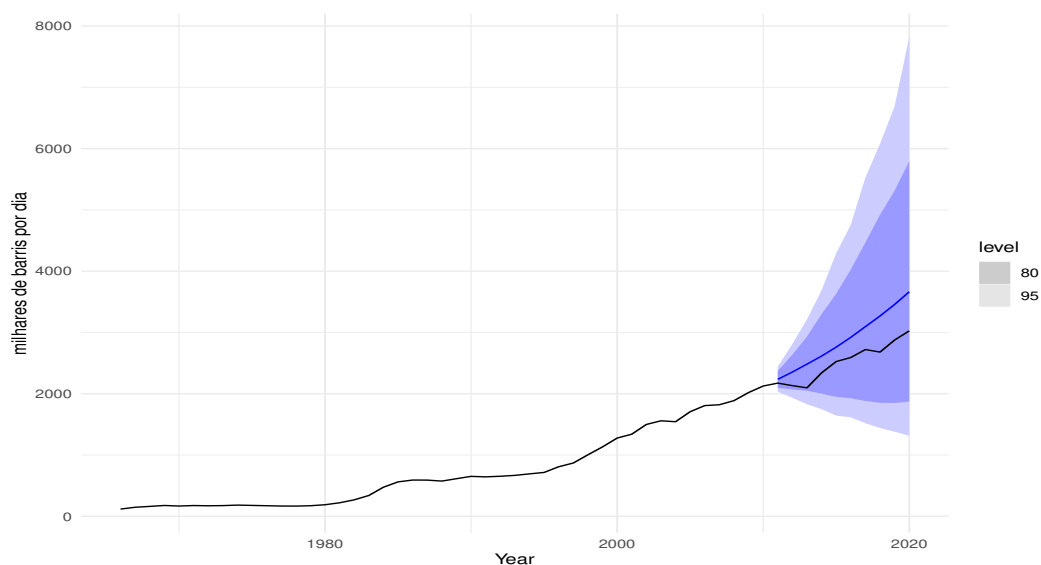
Fonte: Elaborado pelo autor, 2021.

Tabela 7.2: Métricas comparativas entre modelos e dados observados entre 2011 e 2020.

	ME	RMSE	MAE	MPE	MAPE	MASE
ETS	118	254	209	3.69	7.94	4.34
Combinação	-344	381	344	-13.3	13.3	7.15
ARIMA	-807	953	807	-30.3	30.3	16.8

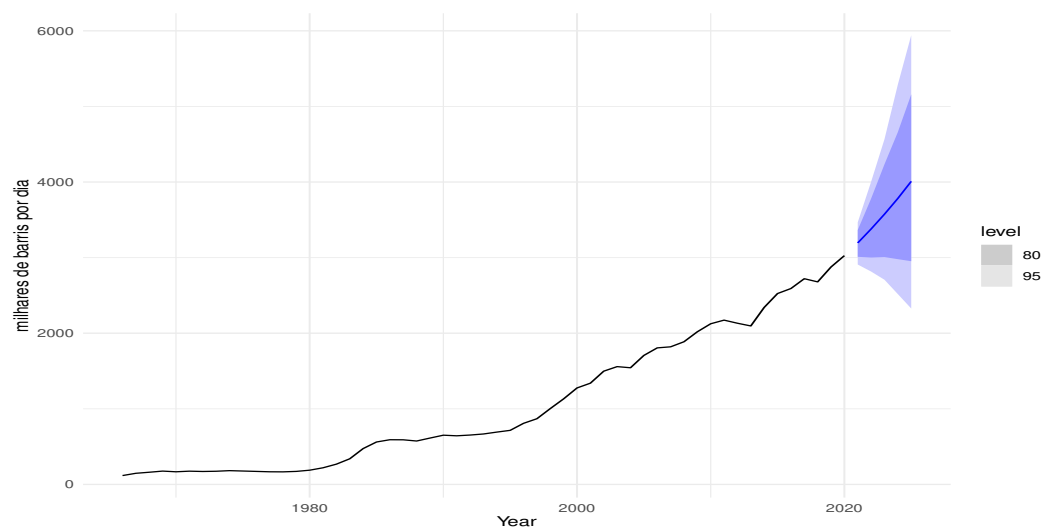
Fonte: Elaborado pelo autor, 2021.

Figura 7.5: Validação dos modelos na série temporal produtiva petróleo no Brasil de 1965 a 2020.



Fonte: Elaborado pelo autor, 2021.

Figura 7.6: Previsão na série temporal produtiva de petróleo no Brasil com horizonte de 5 anos.



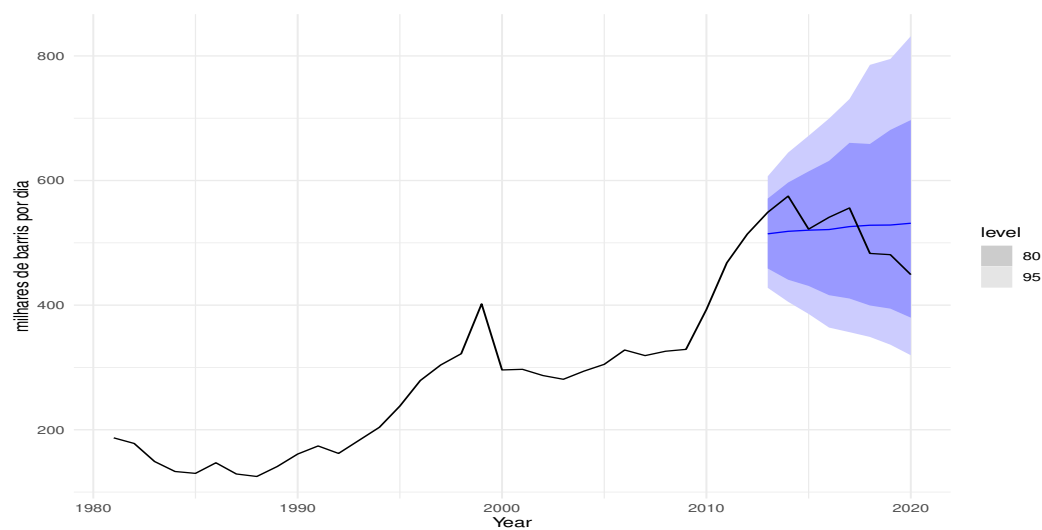
Fonte: Elaborado pelo autor, 2021.

7.2 Consumo nacional

7.2.1 Gasolina

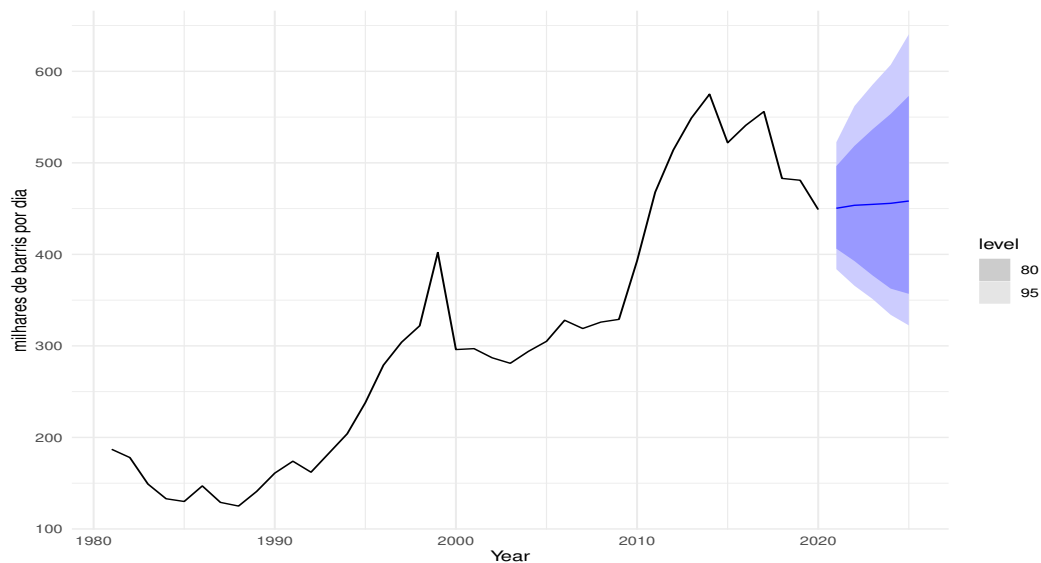
Nessa subseção serão apresentadas a validação (Figura 7.7) e as previsões (Figura 7.8) da série de consumo de gasolina no Brasil de 1980 a 2020. Mais uma vez as previsões calculadas consideram uma combinação média entre modelos de suavização exponencial ETS (M,N,N) e ARIMA (0,1,0) para um horizonte de 5 anos.

Figura 7.7: Validação dos modelos na série temporal de consumo de gasolina no Brasil de 1980 a 2020.



Fonte: Elaborado pelo autor, 2021.

Figura 7.8: Previsão na série temporal de consumo de gasolina no Brasil com horizonte de 5 anos.

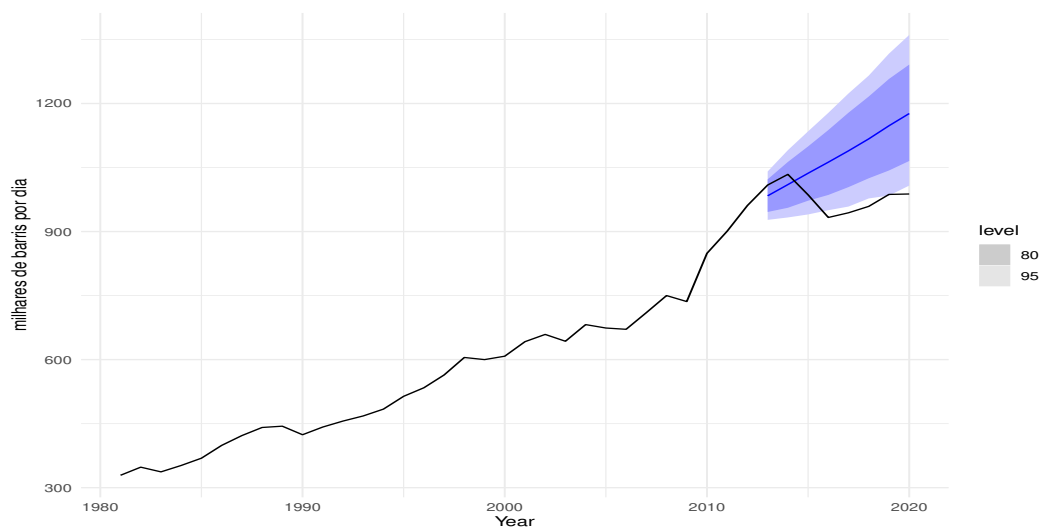


Fonte: Elaborado pelo autor, 2021.

7.2.2 Diesel

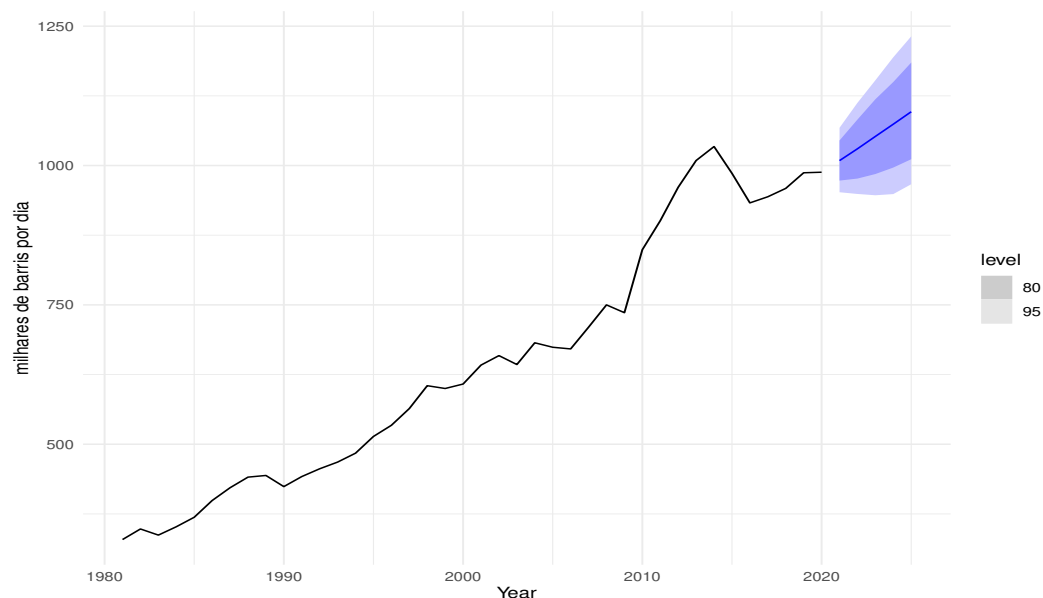
Nessa subseção serão apresentadas a validação (Figura 7.9) e as previsões (Figura 7.10) da série de consumo de diesel no Brasil de 1980 a 2020. As previsões calculadas consideram uma combinação média entre modelos ETS (M,A,N) e ARIMA (0,1,0) com deslocamento.

Figura 7.9: Validação dos modelos na série temporal de consumo de diesel no Brasil de 1980 a 2020.



Fonte: Elaborado pelo autor, 2021.

Figura 7.10: Previsão na série temporal de consumo de diesel no Brasil com horizonte de 5 anos.

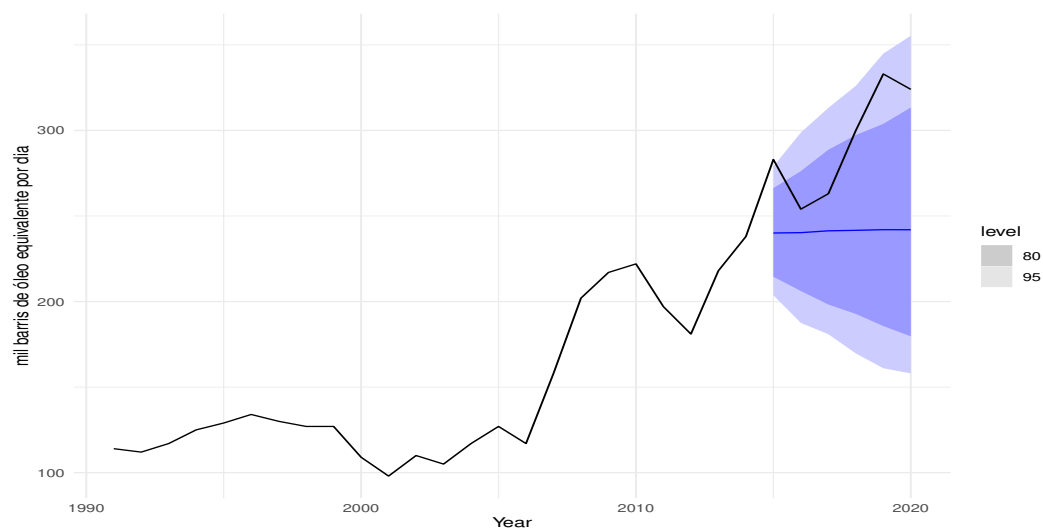


Fonte: Elaborado pelo autor, 2021.

7.2.3 Biocombustíveis

Nessa subseção serão apresentadas a validação (Figura 7.11) e as previsões (Figura 7.12) da série de consumo de biocombustíveis no Brasil de 1990 a 2020. As previsões calculadas consideram uma combinação média entre modelos ETS (M,N,N) e ARIMA (0,1,0) com deslocamento.

Figura 7.11: Validação dos modelos na série temporal de consumo de biocombustíveis no Brasil de 1990 a 2020.



Fonte: Elaborado pelo autor, 2021.

Figura 7.12: Previsão na série temporal de consumo de biocombustíveis no Brasil com horizonte de 5 anos.

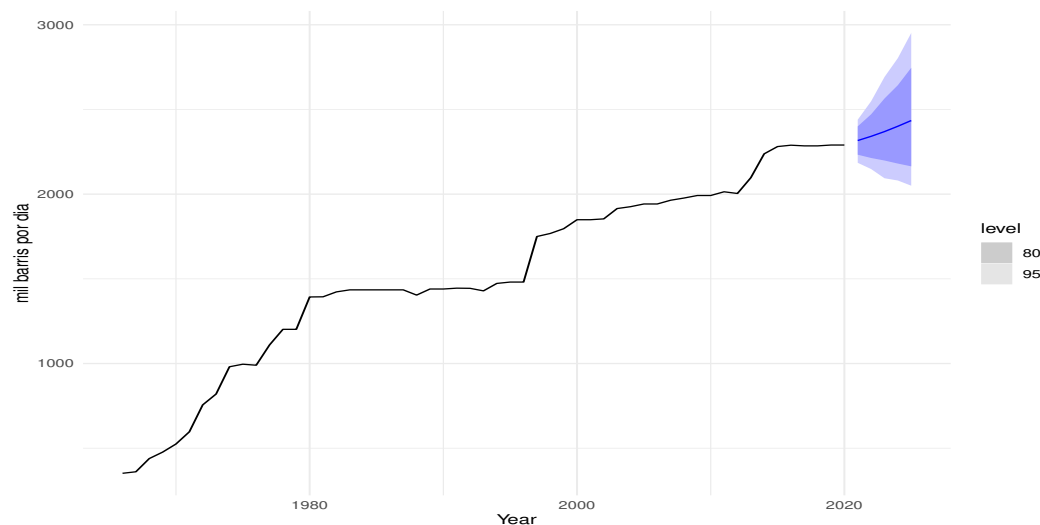


Fonte: Elaborado pelo autor, 2021.

7.3 Refino nacional

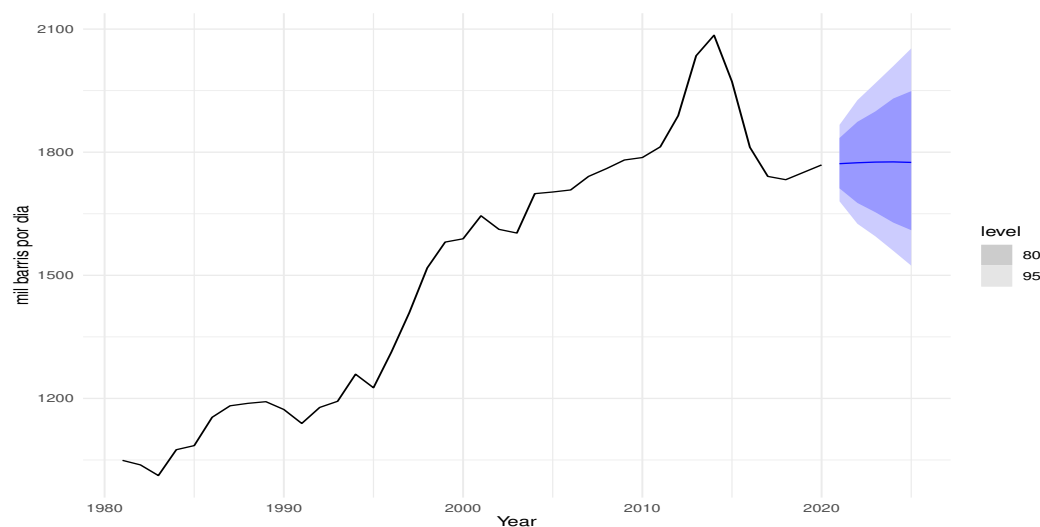
Nessa subseção serão demonstradas as previsões (Figura 7.13) da série de capacidade de refino do Brasil de 1965 a 2020 e o rendimento das refinarias de 1980 a 2020 (Figura 7.14). As previsões calculadas consideram uma combinação média entre modelos ETS (A,A,N) e ARIMA (0,2,1) para capacidade e ETS (M,N,N) e ARIMA (1,1,0) para rendimento.

Figura 7.13: Previsão na série temporal de capacidade de refino do Brasil com horizonte de 5 anos.



Fonte: Elaborado pelo autor, 2021.

Figura 7.14: Previsão na série temporal de rendimento das refinarias do Brasil com horizonte de 5 anos.

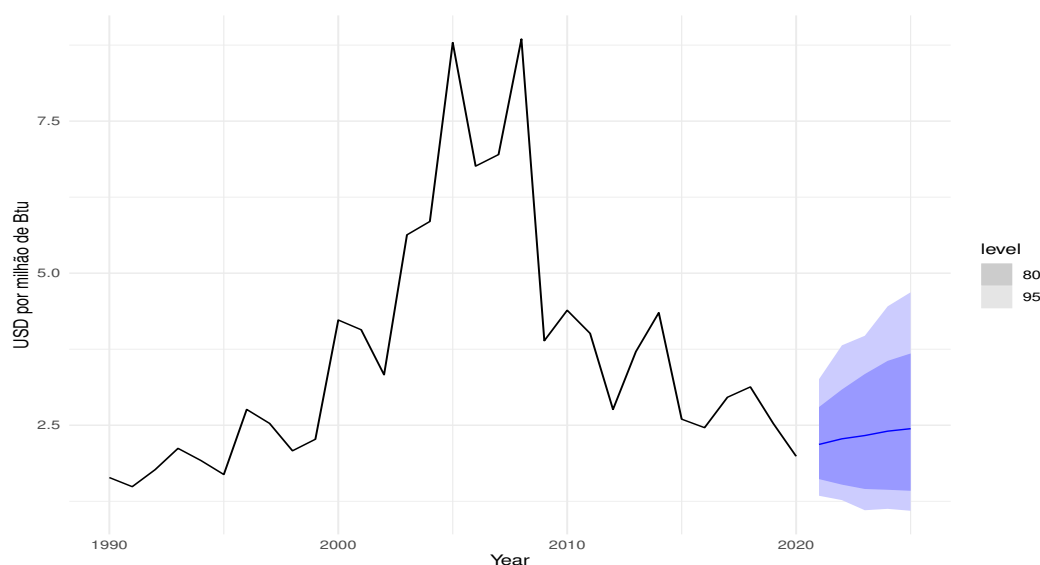


Fonte: Elaborado pelo autor, 2021.

7.4 Preços

Os preços do gás natural são cotados por diferentes índices de acordo com alguns blocos econômicos, como por exemplo o *Heren NBP Index* no Reino Unido ou *Henry Hub* nos Estados Unidos. Será utilizado nas previsões, com horizonte de 5 anos, o índice *Henry Hub*, que tem série temporal de preços em dólares americanos entre 1990 e 2020. Uma combinação de modelos ETS (M,N,N) e ARIMA (1,0,0) serão empregados nas projeções.

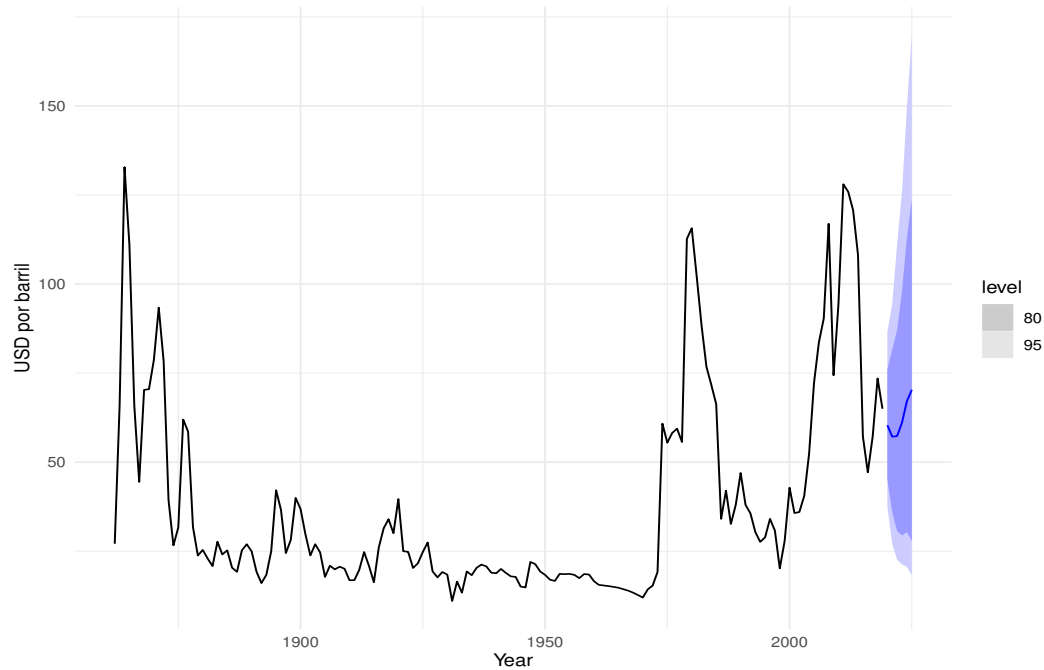
Figura 7.15: Previsão na série temporal de preços do gás natural com índice *Henry Hub* em um horizonte de 5 anos.



Fonte: Elaborado pelo autor, 2021.

Os preços do barril do petróleo são apresentados em uma série longa (desde 1861), deflacionada usando o ano de 2019 como referência ao Índice de Preços ao Consumidor dos EUA. Contudo, considerando a observação atípica de 2020 (devido a pandemia), as previsões serão computadas com observações até 2019. Logo, o horizonte de previsão será de 6 anos e os modelos são ETS (M,A,N) e ARIMA (3,1,2).

Figura 7.16: Previsão na série temporal de preços do barril do petróleo em um horizonte de 6 anos.

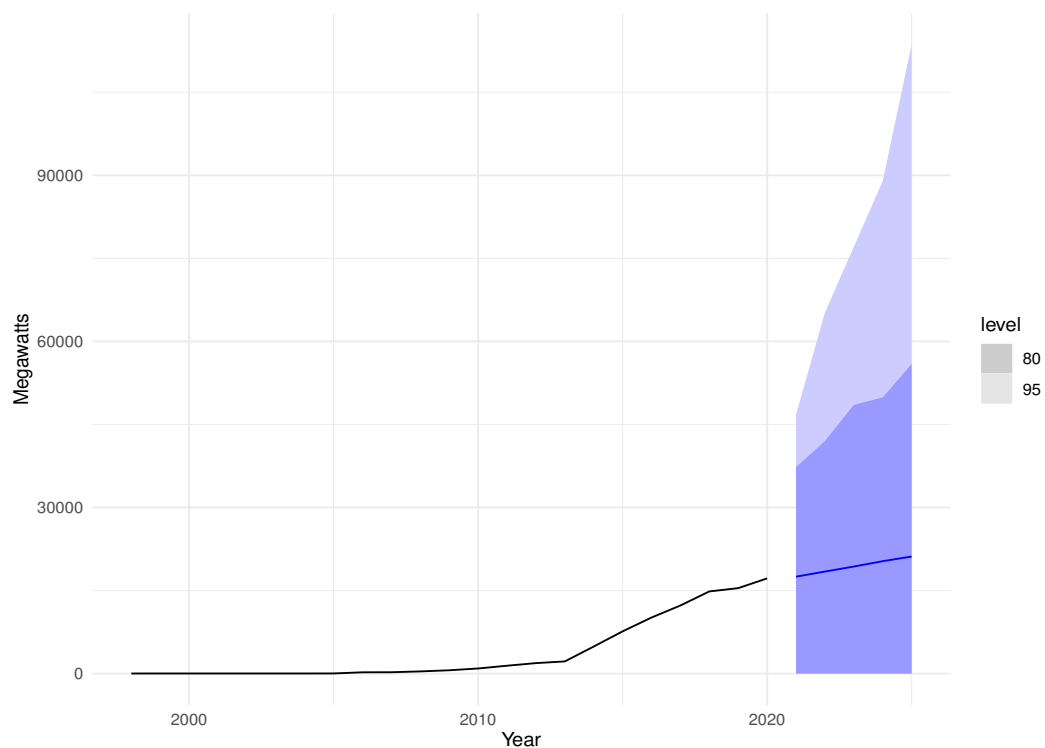


Fonte: Elaborado pelo autor, 2021.

7.5 Energias renováveis

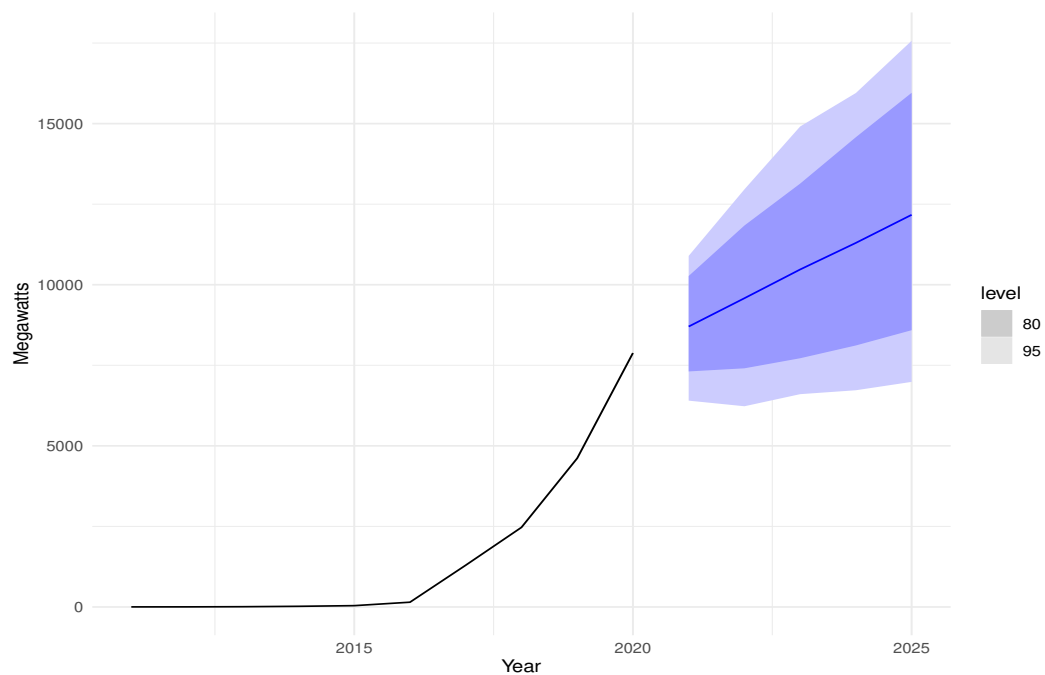
Essa seção abordará as séries temporais da capacidade instalada das energias renováveis eólica e solar. A Figura 7.17 apresenta a previsão com modelo ETS (M,A,N) para um horizonte de 5 anos da capacidade instalada da energia eólica no Brasil. Enquanto que a Figura 7.17 representa as previsões no mesmo horizonte de um modelo de passeio aleatório com *drift*.

Figura 7.17: Previsão na série temporal de capacidade instalada de energia eólica em um horizonte de 5 anos.



Fonte: Elaborado pelo autor, 2021.

Figura 7.18: Previsão na série temporal de capacidade instalada de energia solar em um horizonte de 5 anos.



Fonte: Elaborado pelo autor, 2021.

8 DISCUSSÃO

Através das análises e previsões, é possível notar que a maioria das séries temporais apresentam uma característica de tendência crescente. Isso acontece devido a demanda progressiva por energias primárias no Brasil e no mundo. Cabe ressaltar também que a indústria de petróleo e gás ainda está em aceleração tanto em produção, quanto em consumo e refino. A transição energética para uma economia de baixo carbono é caracterizada também pela participação maior do gás natural como combustível fóssil de transição e pela ampliação também da participação de mais energias renováveis nas matrizes energética e elétrica.

O setor de petróleo e gás têm desafios e oportunidades evidenciados pelos dados. Para analisar os desafios que o setor de petróleo e gás enfrentará em uma perspectiva futura no mercado brasileiro, é necessário investigar os âmbitos do tripé da sustentabilidade dos negócios. Nessa perspectiva utiliza-se uma ferramenta de avaliação estratégica chamada matriz S.W.O.T. (em inglês), ou F.O.F.A. (em português) [11].

Segundo [15], “A análise S.W.O.T. levará você pensar nos aspectos favoráveis e desfavoráveis do negócio, dos seus proprietários e do mercado. É sempre feita em quadrantes e, em cada quadrado são registrados fatores positivos e negativos para a implantação do negócio.”

É com este intuito que a análise S.W.O.T será empregada neste trabalho, como uma ferramenta de avaliação de pontos fortes, pontos fracos, ameaças e oportunidades ao mercado de petróleo e gás. A Figura 8.1 representa a matriz S.W.O.T elaborada com base nos aspectos explorados em seções anteriores, os fatores internos (forças e fraquezas) representam a realidade atual e os fatores externos (oportunidades e ameaças) o futuro potencial.

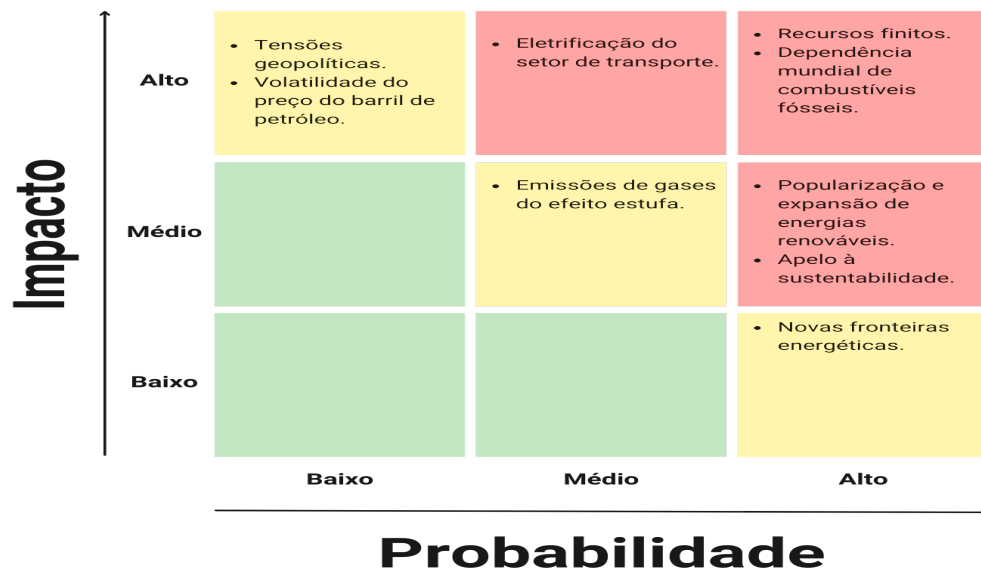
Figura 8.1: Análise SWOT do mercado de petróleo e gás

FORÇAS • Monopólio tecnológico. • Infraestrutura robusta. • Cadeia de suprimentos. • Diversidade da indústria petroquímica.	FRAQUEZAS • Recursos finitos. • Emissões de gases do efeito estufa. • Dependência mundial de combustíveis fósseis.
OPORTUNIDADES • Mercado de crédito de carbono. • Novas fronteiras energéticas. • Incentivo em produção de biocombustíveis. • Utilização de fraturamento hidráulico. • Descarbonização do setor.	AMEAÇAS • Apelo à sustentabilidade. • Volatilidade do preço do barril de petróleo. • Tensões geopolíticas. • Eletrificação do setor de transporte. • Popularização e expansão de energias renováveis. • Novas fronteiras energéticas.

Fonte: Elaborado pelo autor, 2021.

A partir da análise de S.W.O.T. é possível definir estratégias, pesquisas de marketing, objetivos financeiros, mercados-alvo, posicionamento e fatores-chave para tomadas de decisão. Para caracterizar e estratificar esses desafios será utilizado nesse trabalho uma outra ferramenta qualitativa para mensurar os riscos, a matriz de riscos. Essa ferramenta é “definida pela composição das variáveis frequência (probabilidade) e severidade (impacto financeiro), associadas aos eventos de perda (fatores de risco) inerentes ao processo avaliado” [7]. A Figura 8.2 demonstra a matriz de riscos dos desafios setor de petróleo e gás. Essa matriz foi elaborada a partir da análise S.W.O.T, onde seus pontos fracos e ameaças foram utilizados para composição dos riscos.

Figura 8.2: Matriz de risco dos desafios setor de petróleo e gás



Fonte: Elaborado pelo autor, 2021.

8.1 Trabalhos Futuros

Séries financeiras muito voláteis necessitam de atenção especial. Essas séries demandam modelos preditivos que se adaptam melhor as oscilações frequentes de choques exógenos no dados. A sugestão de trabalhos futuros deve envolver modelos preditivos mais robustos e dinâmicos (como em [3]), uma vez que os dados se modificam rapidamente e tensões geopolíticas, notícias especulativas, ameaças a economia global (como pandemias) [17]) são mais difíceis de prever com modelos tradicionais de projeção de valores futuros.

Nessa perspectiva o processamento de linguagem natural pode auxiliar no processo de avaliação de textos jornalísticos em busca de palavras-chave que podem influenciar positivamente ou negativamente os preços de uma série financeira. Logo, essa estratégia pode ser útil na poderação de modelos dinâmicos mais responsivos à volatilidade das observações [12].

Outra sugestão seria concatenar as áreas de análise de séries temporais com pesquisa operacional para um auxílio ainda maior em tomada de decisões mais precisas para o mercado. Ou seja, associar ferramentas de otimização e programação dinâmica com dados de séries temporais.

9 CONCLUSÃO

Neste trabalho foi proposta uma metodologia de previsão de séries temporais a um setor relevante da economia mundial, que apresenta muitos desafios com a transição energética e que está se reinventando com a expansão de novas fronteiras energéticas e novos combustíveis, sobretudo devido ao apelo sustentável e pelo crescimento da demanda energética mundial.

Nota-se que a análise de dados esta cada vez mais presente no cotidiano das pessoas, e que séries temporais são comuns para representar padrões significativos de longo prazo, principalmente se os padrões auxiliam na tomada de decisões mais eficientes, reduzindo custos operacionais e de planejamento.

A indústria de petróleo e gás tem sua relevância e pelas análises expostas ainda deve continuar em franca expansão. De certa forma se modificando com novas pautas, como combustíveis verdes, eletrificação do setor de transportes e geração renovável de energias.

O setor de transporte e de produção de energia elétrica são os principais dependentes de combustíveis fósseis. E ainda que esteja acontecendo uma mudança gradual nas matrizes energéticas e elétricas dos países devido ao apelo sustentável do negócio, as simulações demonstram que essa transição é muito lenta, vide o Brasil que apresenta perspectivas de ampliação de produção, consumo e refino.

Dessa forma, parece razoável investimentos maiores no crescimento produtivo da indústria de petróleo e gás no Brasil. Sobretudo quando cada vez mais países utilizam o fraturamento hidráulico como uma alternativa de sobrevida a poços antes esquecidos e improdutivos. Portanto, fica então evidente o prolongamento da extração em novas fronteiras, como por exemplo, a bacia Pará-Maranhão, que vem apresentando resultados promissores de reservas provadas no arco norte do país.

REFERÊNCIAS

- [1] F. Amaral. *Introdução à Ciência de Dados : mineração de dados e big data*. Alta Books, 2016.
- [2] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [3] H. Chiroma, S. Abdulkareem, and T. Herawan. Evolutionary neural network model for west texas intermediate crude oil price prediction. *Applied Energy*, 142:266–273, 2015.
- [4] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition. *J. Off. Stat*, 6(1):3–73, 1990.
- [5] P. S. Cowpertwait and A. V. Metcalfe. *Introductory time series with R*. Springer Science & Business Media, 2009.
- [6] J. D. Cryer and K.-S. Chan. *Time series analysis: with applications in R*, volume 2. Springer, 2008.
- [7] W. d. L. De Paulo, F. C. Fernandes, L. G. B. Rodrigues, and J. Eidit. Riscos e controles internos: uma metodologia de mensuração dos níveis de controle de riscos empresariais. *Revista Contabilidade & Finanças*, 18:49–60, 2007.
- [8] R. J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice, 3rd edition*. OTexts, 2021.
- [9] R. J. Hyndman and Y. Khandakar. Automatic time series forecasting: the forecast package for r. *Journal of statistical software*, 27(1):1–22, 2008.
- [10] K. N. Klaus Neusser. *Time series econometrics*. Springer, 2016.
- [11] P. Kotler and K. L. Keller. Administração de marketing: a bíblia do marketing. 12^a edição, 2006.
- [12] X. Li, W. Shang, and S. Wang. Text-based crude oil price forecasting: A deep learning approach. *International Journal of Forecasting*, 35(4):1548–1560, 2019.
- [13] B. Looney. Bp statistical review ofworld energy 2021. *BP Statistical Review*, 70th edition, 2021.
- [14] P. A. Morettin and C. Toloí. Análise de séries temporais. In *Análise de séries temporais*, pages 538–538. 2006.
- [15] C. ROSA. *Como Elaborar um Plano de Negócio*. SEBRAE, 2013.

- [16] V. Smil. *Energy and civilization: a history*. MIT Press, 2017.
- [17] J. Wang, G. Athanasopoulos, R. J. Hyndman, and S. Wang. Crude oil price forecasting based on internet concern using an extreme learning machine. *International Journal of Forecasting*, 34(4):665–677, 2018.