



UNIVERSIDADE FEDERAL DO RIO
GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
ENG07053 - TRABALHO DE DIPLOMAÇÃO EM
ENGENHARIA QUÍMICA



Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

Autor: Thomaz Ochoa Favarini

Orientador: Jorge O. Trierweiler

Porto Alegre, dezembro de 18

Agradecimentos

Registro meus agradecimentos aos meus pais, Paulo Eduardo Reus Favarini e Mariza Ochoa Favarini por terem me transmitido seus valores e me concedido as oportunidades necessárias para a conclusão do curso de Engenharia Química na Universidade Federal do Rio Grande do Sul

Agradeço à UFRGS, que me acolheu e abriu as portas para que eu pudesse ter uma educação de excelência.

Registro meu reconhecimento ao orientador Prof. Dr. Jorge O. Trierweiler pelo auxílio na busca de meus objetivos e por me guiar durante a elaboração deste trabalho.

À Alina da Cunha El Jundi por seu companheirismo e compreensão ao longo da jornada acadêmica e profissional.

Por fim, agradeço às minhas avós Claire e Jacira, por todos os exemplos fornecidos ao longo dos anos e pela confiança em mim depositada.

Resumo

A grande disponibilidade de dados disponíveis permite aos mais diversos setores da economia obter maior volume de informação sobre os seus processos. Dentro deste contexto, a análise de séries temporais possibilita aos mesmos uma maneira de aproveitar esta disponibilidade de informação de maneira mais eficiente, criando modelos que permitem maximizar o entendimento em relação às séries temporais e ainda estabelecer relações que permitam inferir o comportamento futuro acerca de determinado conjunto de dados. Este trabalho, abordou metodologias que utilizam somente variáveis endógenas para representar as características determinísticas de uma série temporal e desta forma obter melhor conhecimento acerca do comportamento dos dados e realizar simulações preditivas. Para tanto foram utilizadas as metodologias SARIMA, Decomposição da série temporal, e Suavização exponencial. Estas técnicas foram aplicadas a três séries temporais: Evolução da concentração de CO₂ na atmosfera; Evolução da variação de pressão em um adsorvedor de umidade de poço de petróleo *offshore*; Evolução da cotação da ação *BRKM3*. Através de comparação entre as previsões realizadas por cada um dos modelos pode-se verificar que a natureza dos processos geradores do conjunto de dados tem grande influência na capacidade preditiva dos modelos, sendo a série da evolução da concentração de CO₂ na atmosfera a que obteve melhor predição do comportamento dos dados originais, apresentando $R^2 = 0,98$. Quanto às metodologias de previsão, aquela que possuiu melhor desempenho para representar os comportamentos das três séries temporais estudadas foram os modelos SARIMA.

Palavras-chave: Séries temporais. SARIMA. Suavização exponencial.

Lista de Ilustrações

Figura 1- Metodologia utilizada para elaborar as previsões das séries temporais	12
Figura 2 - Método B, aplicado na realização de forecast via decomposição das séries temporais	13
Figura 3 – Método A, aplicado na realização de forecast via decomposição das séries temporais	13
Figura 4 - Evolução da concentração de CO_2 em ppm de 1974 a 1985.....	21
Figura 5 - Autocorrelação e autocorrelação parcial da evolução da concentração de CO_2	21
Figura 6 - Comparação da concentração de CO_2 Medida (linha azul) com a concentração prevista com o modelo obtido através da metodologia A (linha laranja).	22
Figura 7 - Autocorrelação e autocorrelação parcial dos resíduos gerados pelo método A.	23
Figura 8 - Comparação da concentração de CO_2 Medida em ppm (linha azul) com a concentração prevista com o modelo SARIMA(2,1,0)(0,1,1,12) (linha laranja).	24
Figura 9 - Autocorrelação e autocorrelação parcial dos resíduos do modelo gerado pelo modelo SARIMA(2,1,0)(0,1,1,12).	24
Figura 10 - Comparação da concentração em ppm de CO_2 Medida (linha azul) com a concentração prevista pelo modelo de suavização exponencial (linha laranja).	25
Figura 11 - Autocorrelação e autocorrelação parcial dos resíduos do modelo gerado pelo modelo de suavização exponencial tripla	26
Figura 12 - Evolução da diferença de pressão em adsorvedor com amostras de intervalo de 10 minutos.	27
Figura 13 - Autocorrelação e autocorrelação parcial do delta P até o lag 50..	28

Figura 14 - Comportamento sazonal da série temporal identificado pela Transformada Rápida de Fourier utilizando 10 harmônicas.....	29
Figura 15 - Comparação da diferença de pressão medida (linha azul) do conjunto de teste com a diferença de pressão prevista pela utilização do método B de decomposição (linha laranja).....	29
Figura 16 - Autocorrelação e autocorrelação parcial do delta P dos resíduos até o lag 50.	29
Figura 17 - Histograma dos resíduos indicando que estes não seguem distribuição normal em torno de 0.	30
Figura 18 - Comparação da diferença de pressão medida (linha azul) do conjunto de teste com a diferença de pressão prevista pela utilização do modelo SARIMA (2,0,1)(0,1,0,25) (linha laranja).	31
Figura 19 - Autocorrelação e autocorrelação parcial dos resíduos gerado pelo modelo de SARIMA (2,0,1)(0,1,0,25) na base de treino.....	31
Figura 20 – Figura evidenciando a variação dos coeficientes AR1, AR2 e MA1 nas 60 simulações realizadas.	32
Figura 21 - Evolução da cotação da BRKM3 em intervalo diário minutos, conjunto de treino.....	33
Figura 22 -Autocorrelação e autocorrelação parcial do preço da ação até o lag 50.	33
Figura 23 - Comparação da cotação medida (linha azul) do conjunto de treino com a cotação prevista pelo modelo ARIMA (2,1,3) (linha laranja).....	35
Figura 24 - Comparação da cotação medida (linha azul) do conjunto de treino com a cotação prevista pelo modelo ARIMA (29,1,9) (linha laranja).....	35
Figura 25 - Autocorrelação e autocorrelação parcial dos resíduos do modelo até o lag 50.	36
Figura 26 - Comparação de 7 dias da cotação medida do conjunto de teste (linha azul) com a cotação prevista pelo modelo ARIMA (29,1,9) (linha laranja) .	37
Figura 27 – Comparação entre a primeira derivada do modelo preditivo ARIMA(29,1,9) e a primeira derivada da ação do conjunto de teste.	37

Figura 28 - Comparação de 53 dias da cotação medida do conjunto de teste (linha azul) com a cotação prevista pelo modelo ARIMA (29,1,9) (linha laranja). 38

Figura 29 - Comparação da cotação do conjunto de teste (linha azul) com a cotação prevista pelo modelo de suavização exponencial dupla (linha laranja). . 39

Lista de Tabelas

Tabela 1: Valores médios determinados para cada unidade de tempo do ciclo de 12 meses.....	22
Tabela 2 - Valores determinados dos coeficientes do modelo SARIMA(2,1,0)(0,1,1,12).....	23
Tabela 3 - Coeficientes determinados para o método da exponenciação tripla	25
Tabela 4 - Representação do <i>range</i> utilizado na otimização do AIC	30
Tabela 5 - Valores determinados das constantes do modelo SARIMA(2,0,1)(0,1,0,25).....	31
Tabela 6 - Representação do <i>range</i> utilizado na otimização do AIC.	34
Tabela 7 - Valores determinados dos coeficientes do modelo ARIMA (29,1,9).	36
Tabela 8 - Coeficientes determinados pelo método da exponenciação dupla.38	

Lista de Símbolos

Lista de Símbolos séries temporais

x_t : Valor observado na série temporal no instante t

T_t : Estimador da tendência da série temporal no instante t

S_t : Estimador da sazonalidade da série temporal no instante t

C_t : Estimador do componente cíclico da série temporal no instante t

E_t : Estimador dos resíduos da série temporal no instante t

B_m : Parâmetro do ajuste polinomial de ordem m

$\hat{T}_t^{(g)}$: Estimador médio de tendência de ordem g no instante t

X_{t-k} : Valor observado na série temporal no instante $t - k$

W_{t-k} : Parâmetro ponderador do instante $t - k$

x_{ij} : Valor observado na série temporal no ciclo i e posição j

T_{ij} : Estimador de tendência da série temporal no ciclo i e posição j

S_{ij} : Estimador de sazonalidade da série temporal no ciclo i e posição j

E_{ij} : Estimador dos resíduos da série temporal no ciclo i e posição j

m : É o total de observações em um ciclo completo, é o período da sazonalidade

a_j : É o parâmetro obtido pelo método de regressão para a posição j

d_{ij} : Conjunto de variáveis periódicas (Seno, cosseno)

k : Frequência do sinal periódico, onde: $0 < k < N - 1$

N : Total de observações na amostra

Lista de símbolos Processos Estacionários

h : número de defasagens da série temporal

$\hat{\rho}_h$: Autocorrelação defasada em h instantes, onde: $-1 \leq \hat{\rho}_h \leq 1$

$\hat{\mu}$: Média da amostra

a_h : autocorrelação parcial defasada h instantes, onde: $-1 \leq a_h \leq 1$

σ^2 : Variância característica de um ruído branco

x'_t : Primeira diferenciação da série temporal

R^2 : Coeficiente de determinação

x_{t-m} : Variável x no instante $t - m$

Lista de Símbolos Decomposição de séries temporais

I : Total de ciclos que uma posição j participou

Lista de Símbolos Suavização exponencial

α : Fator de suavização

F_t : Previsão da série temporal para o instante t

\hat{x}_{t+1} : Previsão da série temporal para o instante $t + 1$

β : Fator de suavização da tendência

γ : Fator de suavização da sazonalidade

Lista de Símbolos Modelos ARIMA e SARIMA

ϕ_p : Coeficiente autoregressivo de ordem p

θ_q : Coeficiente de médias móveis de ordem q

X_t : Previsão da série temporal para o instante t

α : Intercepção média

$\Delta_d x$: Série temporal diferenciada d vezes

Δ_D^s : Série temporal diferenciada D vezes sazonalmente

$\hat{\theta}_Q$: Coeficiente de médias móveis sazonal de ordem P

$\hat{\phi}_P$: Coeficiente autorregressivo sazonal de ordem Q

$\hat{\sigma}^2$: Variância residual

\hat{y}_i : Dados originais

\bar{y}_i : Valor médio do conjunto de dados originais

y_i : Valores gerados pelo modelo

k : Número total de parâmetros a serem estimados

x Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

T: Número total de parâmetros a serem estimados

Lista de Abreviaturas e Siglas

ADF – Augmented Dickey-Fuller

AIC – Akaike's Information Criterion

AR – Autoregressive

ARIMA – Autoregressive Integrated Moving Average

ARMA – Autoregressive Moving Average

Corr - Correlação

DFT – Discrete Fourier Transform

FFT – Fast Fourier Transform

I – Integrated

LSE – Least Squares Estimation

MA – Moving Average

MIN – Minimize function

SARIMA – Seasonal Autoregressive Integrated Moving Average

SSE – Sum Squared Error

Sumário

Agradecimentos	ii
Resumo	iii
Lista de Ilustrações	iv
Lista de Tabelas	vii
Lista de Símbolos	viii
Lista de Abreviaturas e Siglas	xi
Sumário	xii
1 Introdução	1
2 Revisão Bibliográfica	3
2.1 Tendência	3
2.1.1 Tendência polinomial	4
2.1.2 Suavização	4
2.1.2.1 Média móvel (MA)	4
2.1.2.2 Média móvel ponderada (WMA).	5
2.2 Sazonalidade	6
2.2.1 Método de regressão	6
2.2.2 Transformada rápida de Fourier	7
2.3 Processos estacionários	7
2.3.1 Outros métodos de estacionarização	9
2.3.1.1 Diferenciação	9
2.3.1.2 Diferenciação sazonal	9
3 Metodologia	11
3.1 Pré-tratamento das séries temporais	11
3.2 Visualizar as séries temporais	12
3.3 Definir e aplicar modelos de previsão	12

3.3.1	Decomposição de séries temporais	12
3.3.1.1	Método A	13
3.3.1.2	Método B	14
3.3.2	Suavização exponencial	14
3.3.2.1	Suavização exponencial simples.	15
3.3.2.2	Suavização exponencial de Holt	15
3.3.2.3	Suavização exponencial tripla	16
3.3.3	Modelos ARIMA e SARIMA	16
3.3.3.1	Modelos autoregressivos	16
3.3.3.2	Modelos de média móvel	17
3.3.3.3	Modelos autoregressivos e de média móvel ARMA	17
3.3.3.4	Modelos ARIMA	17
3.3.3.5	Modelos SARIMA	18
3.4	Critérios de avaliação dos resultados obtidos	18
4	Estudo de Caso	20
4.1	Evolução da concentração de CO ₂	20
4.1.1	Pré-tratamento da série temporal	20
4.1.2	Visualizar a série temporal	21
4.1.3	Definir e aplicar modelos de previsão	21
4.1.3.1	Decomposição da série temporal, Método A	22
4.1.3.2	SARIMA	23
4.1.3.3	Suavização exponencial tripla	25
4.1.4	Avaliar resultados obtidos	26
4.1.5	Previsões de longo prazo	26
4.2	Diferença de pressão em um adsorvedor	27
4.2.1	Pré-tratamento da série temporal	27
4.2.2	Visualizar a série temporal	27
4.2.3	Definir e aplicar modelos de previsão	28
4.2.3.1	Decomposição da série temporal, Método B	28

4.2.3.2	SARIMA	30
4.2.4	Avaliar resultados obtidos	32
4.2.5	Avaliação da dinâmica da série temporal	32
4.3	Evolução da cotação da ação BRKM3	32
4.3.1	Pré-tratamento da série temporal	32
4.3.2	Visualizar a série temporal	33
4.3.3	Definir e aplicar modelos de previsão	34
4.3.3.1	SARIMA	34
4.3.3.2	Suavização exponencial dupla	38
4.3.4	Avaliar resultados obtidos	39
5	Conclusões e Sugestões para Trabalhos Futuros	40
	Referências	42
	Apêndice A	44

1 Introdução

Uma série temporal pode ser definida como a sequência de observações sucessivas tomadas ao longo do tempo (BOX; JENKINS; REINSEL, 2008) estas séries são encontradas nas mais variadas áreas como na engenharia, medicina, negócios, economia e ciências sociais, alguns exemplos de observações no tempo são: medição da pressão encontrada em poços de petróleo *offshore* tomada de hora em hora, apurações mensais das temperaturas médias de Porto Alegre, vendas mensais de petróleo, índices diários da bolsa de valores e assim sucessivamente.

A análise das séries temporais, por sua vez, também possui importantes aplicações que se estendem em todos os setores onde há a ocorrência das séries temporais. Alguns dos objetivos destas análises podem ser: Investigar o mecanismo gerador da série temporal; obter melhor entendimento sobre o comportamento da série; estimar funções de transferência (relacionando entradas e saídas), analisar os efeitos de eventos não usuais em um sistema; controle de processos; e realizar previsões dos valores futuros das séries temporais (BOX; JENKINS; REINSEL, 2008), que é o foco do presente trabalho.

Em um cenário altamente competitivo a busca pelo aumento de eficiência e produtividade é fundamental para a sobrevivência das organizações. Um dos meios para atuar neste quesito é entender de maneira mais clara quais serão os eventos futuros aos quais a empresa ou os setores estarão submetidos e, de posse deste conhecimento, agir de maneira antecipada para obter o melhor aproveitamento. Isto é válido para inúmeras situações, por exemplo: decidir o melhor período para realizar a substituição/otimização de um equipamento; aumentar ou diminuir a utilização da capacidade produtiva fabril; otimizar estoques; decidir qual o melhor momento para adquirir ou vender um ativo; obter melhor entendimento dos fenômenos, interpretando situações que diferem da prevista; entre outras.

A realização de previsões em relação a eventos futuros das séries temporais utiliza a característica de dependência entre as sucessivas observações ao longo do tempo (BOX; JENKINS; REINSEL, 2008). Deste modo, o conhecimento da estrutura internas das séries temporais e o estabelecimento de técnicas para capturar e entender esta dependência são fundamentais.

Este trabalho tem como objetivo comparar três métodos de previsão de séries temporais, utilizando somente variáveis endógenas, em três conjuntos de dados distintos. A avaliação da qualidade do modelo será feita visualmente pela comparação gráfica dos valores previstos pelos realizados, quantitativamente através do indicador do coeficiente de determinação e realizando avaliação dos resíduos, por meio da utilização do teste Aumented Dickey Fuler e das funções de autocorrelação e autocorrelação parcial para verificar se os resíduos seguem modelo estacionário.

A estrutura do texto é dividida em seis capítulos: o Capítulo 1 é a introdução, que apresenta a definição de série temporal, as motivações e os objetivos do trabalho; O Capítulo 2 aborda revisão bibliográfica, o tema central é a representação e identificação dos componentes de uma série temporal; o Capítulo 3 aborda as metodologias utilizadas para realizar a previsão de séries temporais; no Capítulo 4 são apresentadas e analisadas as séries temporais e o resultado dos modelos é avaliado. No Capítulo 5 são sintetizadas as principais conclusões do trabalho e são feitas recomendações para a realização de trabalhos futuros.

2 Revisão Bibliográfica

Uma série temporal é composta por uma ou mais das seguintes estruturas internas:

$$x_t = T_t + S_t + C_t + E_t$$

- Tendência (T_t): é o movimento de aumento ou diminuição da variável observada ao longo do tempo.
- Sazonalidade (S_t): são movimentos periódicos e repetitivos presentes em uma série temporal.
- Componente cíclico (C_t): são movimentos periódicos menos frequentes que a sazonalidade e podem não apresentar um período de variação fixo (PAL; PRAKASH, 2017).
- Resíduos (E_t): Este componente não é dependente do tempo e representa a porção irreduzível da função que o modelo não consegue contabilizar (PAL; PRAKASH, 2017) segundo BROCKWELL e DAVIS (1996), este resíduo deve se comportar conforme um ruído branco.

Onde x_t representa a variável x no instante de tempo t .

2.1 Tendência

Tendência é o movimento de incremento ou decréscimo do valor da função ao longo do tempo. Esta característica das séries temporais pode ser percebida analisando visualmente os dados através de um gráfico da variável observada em função do tempo (PAL; PRAKASH, 2017). Como uma definição quantitativa, pode-se definir a tendência como sendo a mudança de longo prazo na média da função temporal (CHATFIELD, 2000).

Há vários métodos para realizar a estimação da função de tendência T_t (MORETTIN; TOLOI, 2006) e estes se dividem em dois grupos principais:

- **Regressão paramétrica:** Ajustar uma série temporal utilizando uma função descrita utilizando parâmetros finitos. Como uma função polinomial, exponencial modelos ARMA, ARIMA e SARIMA.
- **Regressão não paramétrica:** Neste tipo de regressão não é necessário realizar estimativa de parâmetros para realizar a regressão. São exemplos dessa classe a média móvel, Regressão kernel, regressão polinomial local, regressão splines, wavelet, entre outras.

Neste trabalho a tendência será estimada utilizando os seguintes métodos:

- Polinomial;

- Média móvel;
- Média móvel ponderada.

2.1.1 Tendência polinomial

Ao realizar uma estimação por tendência polinomial, o objetivo é modelar a série temporal utilizando um polinômio que terá seus coeficientes determinados através do ajuste por mínimos quadrados (*LSE*). Desta forma, é identificado o polinômio que melhor se adequa aos dados que estão sendo utilizados (BROCKWELL; DAVIS, 1996).

$$LSE = \text{MIN} \sum_{t=1}^n (x_t - T_t)^2$$

$$T_t = B_0 + B_1t + \dots + B_mt^m$$

Onde B_m é o parâmetro de ordem m determinado por *LSE*.

2.1.2 Suavização

Ao representar a tendência por um polinômio são utilizadas todas as observações para elaborar a função que representará a tendência durante todo o intervalo de tempo. Uma outra alternativa consiste na utilização apenas de um conjunto de dados próximos, dentro de uma janela menor de dados. A ideia é que em um instante t a tendência será estimada usando-se as observações Z_s ao redor de t (MORETTIN; TOLOI, 2006). Enquadram-se nessa categoria os métodos de Média Móvel (“moving average”, MA) e Média Móvel Ponderada (“weighted moving average”, WMA), os quais serão apresentados nas próximas subseções.

2.1.2.1 Média móvel (MA)

A realização da média móvel tem como efeito suavizar a série temporal, reduzindo o seu ruído.

$$\hat{T}^{(g)}_t = \frac{X_{t-k} + X_{t-k+1} + \dots + X_t + X_{t+k-1} + X_{t+k}}{2k + 1}$$

Onde a tendência média é representada por $\hat{T}_t^{(g)}$, o intervalo considerado é $t \pm k$ e k equivale ao alcance da média. A ordem da média móvel é calculada como $g = 2k + 1$.

A ordem da média móvel pode ser ímpar (simétrica) ou par (assimétrica), para tornar a série simétrica é necessário realizar a média duas vezes.

$$\hat{T}_t^{(2)} = \frac{X_{t-1} + X_t}{2}$$

$$2 \times \hat{T}_t^{(2)} = \frac{\hat{T}_t^{(2)} + \hat{T}_{t+1}^{(2)}}{2} = \frac{1}{2} \left(\frac{X_{t-1} + 2X_t + X_{t+1}}{2} \right)$$

É importante apontar que a repetição da média móvel causa uma ponderação, aumentando o peso dos valores mais recentes (X_t) (PAL; PRAKASH, 2017). Em séries onde a sazonalidade é presente para eliminá-la basta considerar o período da sazonalidade na ordem e repetição da média móvel, ou seja, para dados com sazonalidade anual a remoção da sazonalidade é feita realizando a seguinte média móvel $2x\hat{T}_t^{(12)}$, em dados com sazonalidade semanal a remoção dessa sazonalidade é realizada com a média $\hat{T}_t^{(7)}$.

2.1.2.2 Média móvel ponderada (WMA).

Segundo PAL e PRAKASH (2017) pode-se optar por utilizar a média móvel ponderada uma vez que esta suaviza a série temporal considerando maior peso para os dados mais próximos do valor calculado.

Representação matemática da média móvel ponderada:

$$\hat{T} = X_{t-k}W_{t-k} + X_{t-k+1}W_{t-k+1} + \dots + X_tW_t + X_{t+k-1}W_{t+k-1} + X_{t+k}W_{t+k}$$

Onde:

$$W_{t-k} + W_{t-k+1} + \dots + W_t + W_{t+k-1} + W_{t+k} = 1$$

Apesar da média móvel e a média móvel ponderada serem boas ferramentas para visualizar e identificar a tendência de maneira mais clara, segundo BROCKWELL e

DAVIS (1996) elas não são úteis para elaboração de modelos preditivos uma vez que elas levam em consideração tantos valores passados quanto futuros.

2.2 Sazonalidade

Sazonalidade refere-se a movimentos periódicos repetitivos presentes em séries temporais e que podem ser modelados como a soma de ondas senoidais de período conhecido. Uma característica das séries sazonais é a alta autocorrelação que estas apresentam em *lags* sazonais, isto é, em *lags* múltiplos do período m . Um bom procedimento de captura da sazonalidade deve ser capaz de remover grande parte desta correlação (MORETTIN; TOLOI, 2006).

Uma série temporal pode ser representada da seguinte forma:

$$x_{ij} = T_{ij} + S_{ij} + E_{ij}$$

Onde os índices i e j representam a quantidade de ciclos e a posição dentro de um ciclo respectivamente. Se o padrão sazonal não variar muito ao longo dos anos, a equação é simplificada para:

$$x_{ij} = T_{ij} + S_j + E_{ij}$$

Dentre os métodos utilizados para estimar a sazonalidade, deve-se citar o método de regressão, o método das médias móveis e o da transformada rápida de Fourier.

Neste trabalho será abordado o método de regressão e o da transformada rápida de Fourier.

2.2.1 Método de regressão

Estes modelos são utilizados quando a periodicidade da sazonalidade é conhecida, ou seja, nas séries às quais devido ao seu comportamento constante da sazonalidade, permitem prever a sazonalidade com base nos meses anteriores (MORETTIN; TOLOI, 2006).

Considerando que a série não apresente componente de tendência:

$$S_t = \sum_{j=1}^m a_j d_{jt}$$

Onde d_{jt} são as variáveis periódicas (senos, cossenos ou variáveis sazonais). Como foi considerada sazonalidade constante, a_j independe de t . Desta maneira,

temos que, para t correspondente a uma posição específica j dentro do ciclo, d_{jt} é igual a 1 e 0 caso contrário.

Dessa maneira, utilizando o método dos mínimos quadrados é possível determinar o vetor dos m valores de a (MORETTIN; TOLOI, 2006), onde m representa o total de observações dentro de um ciclo completo.

2.2.2 Transformada rápida de Fourier

Esta metodologia é baseada na possibilidade do componente sazonal poder ser representado como um sinal periódico em um tempo discreto e com a utilização de uma série de Fourier converter em uma combinação linear de harmônicos de uma dada frequência fundamental (MARTELOTTE, 2014).

Para calcular a transformada discreta de Fourier é utilizada a seguinte equação:

$$DFT(x)[k + N] = \sum_{n=0}^{N-1} x[n] * \exp\left(-\frac{2\pi ink}{N}\right)$$

Onde k representa a frequência de 0 a $N - 1$, n é o índice de tempo, N representa o número de observações na amostra $DFT(x)$ é a transformada discreta de Fourier (DOWNEY, 2016).

Neste trabalho, a transformada de Fourier será calculada através do algoritmo da Transformada Rápida de Fourier, o qual reduz significativamente o tempo para computar conversões do domínio do tempo para da frequência (STREMLER, 1990, *apud* LEWIS; HERBERT; BELL, 2003).

2.3 Processos estacionários

Se uma série temporal pode ser prevista exatamente pelos valores passados, esta é classificada como determinística. No entanto, na maioria dos casos, a maior parte das séries temporais apresentam uma parcela significativa com características estocásticas, e nestas somente parte do futuro é determinado pelos valores passados (CHATFIELD, 2000).

Ao representar uma série temporal pela adição das parcelas determinísticas, através das funções dos componentes internos (tendência, sazonalidade, componente cíclico), e subtrair estes componentes das observações medidas, é obtido resíduo não nulo devido às características estocásticas da série temporal.

Deste modo, um bom modelo de previsão deve ser capaz de gerar um resíduo que se comporte como uma sequência de variáveis aleatórias não correlacionadas, com média zero e variância σ^2 . Este tipo de sequência temporal é conhecido como ruído branco (BROCKWELL; DAVIS, 1996), veremos no próximo parágrafo que o ruído branco devido à sua média e variância é também classificado como estacionário (BROCKWELL; DAVIS, 1996), característica importante que indica boa modelagem da parte determinística da série temporal.

Neste trabalho, foi considerada estacionária toda função que possui média e variância constantes e finitas (GUIDOLIN; PEDIO, 2018).

Devido à dificuldade de identificar se estas variáveis são constantes ao longo da série temporal, são utilizadas outras ferramentas para analisar se o conjunto de dados é ou não estacionário. Três ferramentas comumente utilizadas são as funções de autocorrelação, autocorrelação parcial e o teste *Augmented Dickey-Fuller*- ADF (MACKINNON, 2010).

A função autocorrelação reflete o grau de dependência linear de uma série temporal em diferentes tempos, ou ainda, indica a correlação da série temporal com ela mesma defasada em h instantes de tempo. Dada uma amostra de N observações da variável x a autocorrelação ($\hat{\rho}_h$) pode ser calculada utilizando a seguinte equação (GUIDOLIN; PEDIO, 2018):

$$\hat{\rho}_h = \frac{\sum_{t=h+1}^N (x_t - \hat{\mu})(x_{t-h} - \hat{\mu})}{\sum_{t=h+1}^N (x_t - \hat{\mu})^2}$$

Na qual a média, $\hat{\mu}$, da amostra é representada por:

$$\hat{\mu} = N^{-1} \sum_{t=1}^N x_t$$

A função autocorrelação parcial entre um instante x_t e x_{t-h} é a autocorrelação entre estas variáveis condicionado às interações $x_{t-1}, x_{t-2}, \dots, x_{t-h+1}$, isto é, a autocorrelação depois de removida a relação entre a variabilidade já explicada entre os instantes x_{t-1} e x_{t-h+1} . Formalmente a autocorrelação parcial é descrita pela seguinte equação (GUIDOLIN; PEDIO, 2018):

$$a_h = \text{Corr}(x_t, x_{t-h} | x_{t-1}, \dots, x_{t-h+1})$$

De acordo com teste ADF na presença de autocorrelação, a diferenciação de primeira ordem de uma série temporal x'_t pode ser representada como uma regressão linear de instantes de tempo anteriores e dos diferenciais de primeira ordem até um atraso (lag) de índice m . A regressão pode ser descrita pela seguinte equação (PAL; PRAKASH, 2017):

$$x'_t = \gamma x_{t-1} + \beta_1 x'_{t-1} + \dots + \beta_h x'_{t-h} + E_t$$

$$\gamma = (-1 + \sum_{i=1}^h \phi_i)$$

Se a série for não estacionária ela apresentará raiz unitária ($\gamma = (-1 + \sum_{i=1}^m \phi_i) = 0$) e a hipótese nula será aceita. Caso contrário a série é dita estacionária e a hipótese nula é rejeitada (PAL; PRAKASH, 2017).

Como será exemplificado no decorrer do trabalho na seção 3.3.1, ϕ são os coeficientes de regressão do modelo.

2.3.1 Outros métodos de estacionarização

Quando a remoção de tendência e da sazonalidade não são suficientes para transformar a série temporal em estacionária dois métodos são comumente aplicados:

2.3.1.1 Diferenciação

A ideia fundamental da diferenciação é realizar a diferença entre dois pontos sucessivos de uma série temporal de maneira que a diferença tenha média e variância constantes e, desta forma, a série temporal poder ser tratada como estacionária.

$$x'_t = x_t - x_{t-1} = c + E_t$$

Nos casos em que a primeira diferenciação não tornar a série estacionária, podem ser realizadas sucessivas diferenciações (PAL; PRAKASH, 2017).

2.3.1.2 Diferenciação sazonal

A diferenciação sazonal é utilizada quando a série temporal apresenta caráter sazonal e o período, m , da sazonalidade é conhecido. Desta forma o conjunto de observações pode ser convertido em estacionário realizando a diferenciação entre x_t e x_{t-m} .

10 Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

$$x_t^m = x_t - x_{t-m} = E_t$$

Onde E_t representa o erro e x_t^m , a diferenciação sazonal, que deve comportar-se como um ruído branco (PAL; PRAKASH, 2017).

3 Metodologia

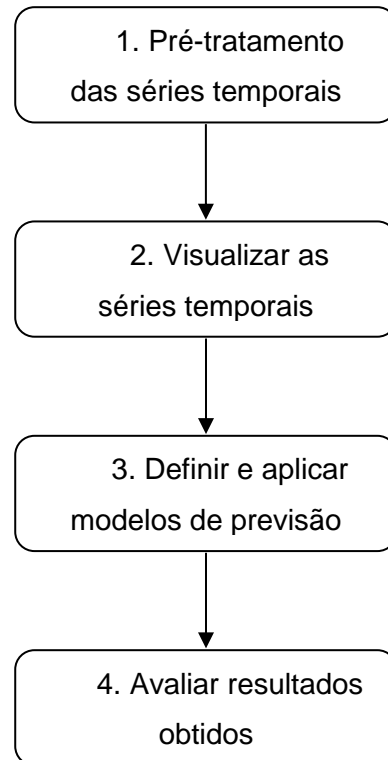
A metodologia utilizada na elaboração do trabalho está ilustrada na Figura 1. As simulações foram realizadas utilizando os pacotes Pandas, Numpy e Statsmodels da programação em Python.

3.1 Pré-tratamento das séries temporais

Em um primeiro momento, é preciso obter as bases de dados que serão analisadas. Como na maioria dos casos são utilizados dados que foram gerados com outros objetivos, cinco etapas se tornam necessárias:

- **Definir tamanho da amostra:** Nas análises de séries temporais deste trabalho são feitos estudos sobre intervalos discretos. Portanto, a correta amostragem é fundamental para que possa ser realizada uma boa simulação dos dados. Quando a base de dados obtida possui tamanho superior ao desejado, comumente se agrupa a série temporal em valores médios de um conjunto de observações (visando a remoção de oscilações que influenciariam como ruídos) ou se remove uma quantidade de pontos da amostra em intervalos de tempo constante, ou seja, se efetua uma re-amostragem.
- **Remover vazios:** A remoção dos dados nulos (ou *not-a-number*) também é fundamental, uma vez que grande parte dos modelos não simulam de maneira adequada quando estes estão presentes.
- **Remover outliers:** Para que a visualização e as previsões não sejam prejudicadas é importante realizar a remoção de *outlier*. Para tanto, podem ser utilizados diferentes métodos como filtros por frequência das observações ou por filtros em percentil elevado. Devido ao escopo estes métodos não serão focados ao longo deste trabalho.
- **Definir coluna de tempo:** Para realizar as simulações é importante que a coluna dos dados cronológicos que são utilizados no eixo das abscissas seja de fácil entendimento sendo frequentemente necessário converter ao formato desejado.
- **Segmentar os dados em conjuntos de treino e teste:** Para construir o modelo é importante separar a base que será usada para a elaboração do modelo (base de treino) e a que será utilizada para verificar se a predição obteve bons resultados (base de teste).

Figura 1- Metodologia utilizada para elaborar as previsões das séries temporais



3.2 Visualizar as séries temporais

Nesta etapa os componentes internos (tendência, sazonalidade, componente cíclico) das séries temporais são identificados e também são obtidos os valores das funções de autocorrelação e autocorrelação parcial. Este procedimento é fundamental para definição dos modelos de previsão que serão utilizados.

3.3 Definir e aplicar modelos de previsão

Neste trabalho foram abordados os seguintes modelos visando a estimação de valores futuros para as séries temporais:

- Decomposição das séries temporais;
- Suavização exponencial;
- Modelos ARIMA e SARIMA.

3.3.1 Decomposição de séries temporais

Estes modelos têm como principal característica decompor as séries temporais em seus principais componentes: tendência, sazonalidade e erro e a partir das funções determinadas construir um modelo de previsão.

Neste trabalho foram construídos modelos de previsão por decomposição das séries temporais realizando os seguintes procedimentos em dois métodos distintos, conforme ilustrado na Figura 2 e na Figura 3.

Figura 3– Método A, aplicado na realização de forecast via decomposição das séries temporais

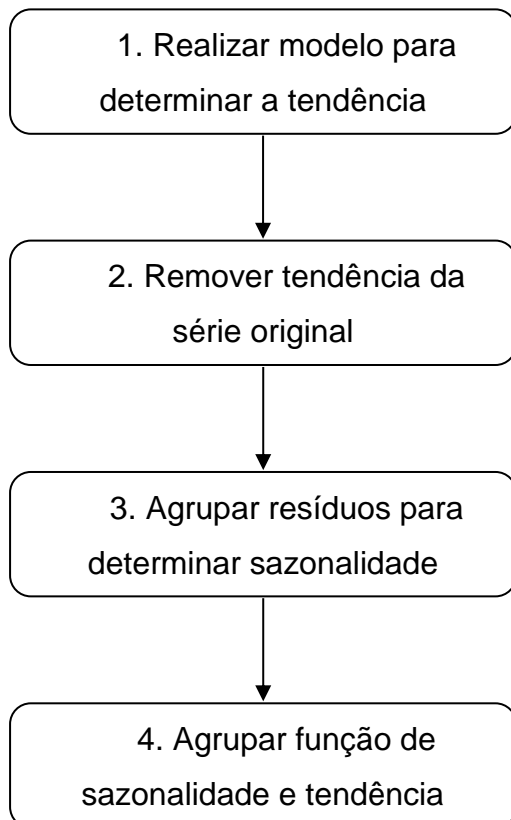
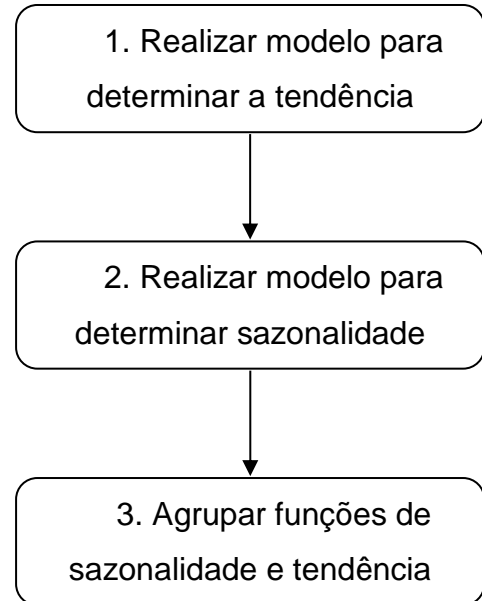


Figura 2 - Método B, aplicado na realização de forecast via decomposição das séries temporais



3.3.1.1 Método A

Considerando que na série temporal não haja componente cíclica, a sua equação pode ser representada por:

$$x_t = T_t + S_t + E_t$$

A tendência T_t é determinada utilizando um dos procedimentos apresentados no capítulo 2.1.

Em seguida, é realizada a remoção da tendência conhecida da série original.

$$x_t - T_t = S_t + E_t$$

Os resíduos (sazonalidade e erro) são agrupados com base no período m da sazonalidade. Desta maneira terá um valor característico para cada uma das observações dependendo com a posição dele referente à curva de sazonalidade.

Ao utilizar este método, assume-se que a sazonalidade é constante e não tem sua amplitude alterada ao longo dos anos.

Deste modo a sazonalidade característica de cada um dos instantes dentro do período m é representada pela seguinte equação:

$$S_j = I^{-1} \sum_{i=1}^I x_i$$

Onde $j = 1, 2, \dots, m$ em que m representa o período da sazonalidade, I representa o total de ciclos em que um determinado j participou e i representa um ciclo completo.

Por fim o modelo é então definido como:

$$x_{ij} = S_j + T_{ij}$$

3.3.1.2 Método B

Considerando que na série temporal não exista componente cíclica, a sua equação pode ser representada por:

$$x_t = T_t + S_t + E_t$$

A tendência T_t e a sazonalidade S_t são determinadas utilizando um dos procedimentos apresentados no Capítulo 2.1 e 2.2 respectivamente.

Por fim ambos as equações são adicionadas e a previsão é realizada utilizando a seguinte equação:

$$x_t = T_t + S_t$$

3.3.2 Suavização exponencial

Nestes modelos os dados passados de uma série temporal são utilizados de maneira ponderada a fim da realização de previsões. Segundo MORETTIN E TOLOI (2006), técnicas específicas deste tipo assumem que os valores extremos da série representam aleatoriedade e que através de suavizações desses extremos é possível identificar um padrão básico.

3.3.2.1 Suavização exponencial simples.

O modelo atua de maneira similar à média móvel ponderada, porém a ponderação é feita de maneira exponencial e somente os dados do passado são utilizados, permitindo a elaboração de modelos preditivos (BROCKWELL; DAVIS, 1996):

$$x_{t+1} = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \dots$$

Onde α é o fator de suavização e varia de 0 a 1. Para a realização da previsão é utilizada a seguinte equação.

$$F_1 = x_1$$

$$F_2 = \alpha x_1 + (1 - \alpha)F_1$$

...

$$F_{t+1} = \alpha x_t + (1 - \alpha)F_{t-1}$$

Nesta equação, F_t representa o valor da previsão. Quando $\alpha = 1$, $F_{t+1} = x_t$ e quando $\alpha = 0$, $F_{t+1} = F_{t-1}$. Ou seja, quanto maior alfa menor é o peso para os dados mais antigos e maior é para os mais recentes (PAL; PRAKASH, 2017).

3.3.2.2 Suavização exponencial de Holt

Quando a série temporal apresenta tendência, a suavização exponencial simples não representará a série de maneira adequada. Para contornar este problema, é utilizada a suavização de Holt na qual é adicionada uma segunda constante que tem como funcionalidade representar a tendência da série temporal (MORETTIN; TOLOI, 2006).

Neste modelo a previsão da série temporal é representada como a soma do nível e da tendência (PAL; PRAKASH, 2017).

$$\hat{x}_{t+1} = F_t + T_t$$

Onde F_t representa o nível e T_t a tendência:

$$F_t = \alpha x_t + (1 - \alpha)(F_{t-1} + T_{t-1})$$

$$T_t = \beta(F_t - F_{t-1}) + (1 - \beta)(T_{t-1})$$

Nesta equação β representa o fator de suavização da tendência.

3.3.2.3 Suavização exponencial tripla

A suavização exponencial tripla permite à simulação incorporar o nível, a tendência e a sazonalidade (PAL; PRAKASH, 2017), para tanto, a seguinte equação é utilizada:

$$F_t = \alpha(x_t - S_{t-L}) + (1 - \alpha)(F_{t-1} + T_{t-1})$$

S_t representa a sazonalidade e é calculada utilizando a seguinte equação (HYNDMAN; ATHANASOPOULOS, 2018):

$$S_t = \gamma(x_t - F_t) + (1 - \gamma)(S_{t-m})$$

Onde γ representa o fator de suavização da sazonalidade.

3.3.3 Modelos ARIMA e SARIMA

Em essência, estes modelos utilizam de regressões entre as observações passadas das séries temporais para elaborar modelos preditivos (PAL; PRAKASH, 2017).

3.3.3.1 Modelos autoregressivos

Segundo Chatfield (2000), se uma série temporal é dita autoregressiva de ordem p então ela é uma soma linear ponderada dos p valores passados adicionados a um termo de característica aleatória.

Nestes modelos a regressão da série é realizada com o *lag* da série temporal. Este modelo autoregressivo é definido como AR(p) onde o p refere-se à ordem do componente AR (PAL; PRAKASH, 2017).

Um modelo AR(1) é descrito pela seguinte função (HYNDMAN; ATHANASOPOULOS, 2018):

$$X_t = \alpha + \phi X_{t-1} + E_t$$

Um modelo de ordem p AR(p) é descrito pela seguinte função:

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + E_t$$

Neste modelo E_t representa o termo de característica aleatória, que se a modelagem for perfeita, deverá ser equivalente a um ruído branco, em um instante t , e ϕ é o coeficiente de regressão do modelo.

3.3.3.2 Modelos de média móvel

Os modelos com média móvel utilizam a dependência temporal entre os erros residuais para prever valores futuros. Este modelo auxilia no ajuste de séries onde ocorreram eventos inesperados como a ruptura de um equipamento.

Um modelo MA(1) é descrito pela seguinte função:

$$X_t = \alpha + \theta E_{t-1} + E_t$$

Um modelo de ordem q AR(q) é descrito pela seguinte função (HYNDMAN; ATHANASOPOULOS, 2018):

$$X_t = \alpha + \theta_1 E_{t-1} + \theta_2 E_{t-2} + \dots + \theta_q E_{t-q} + E_t$$

Netas equações q representa a ordem do modelo e α representa a intercepção média.

3.3.3.3 Modelos autoregressivos e de média móvel ARMA

Estes modelos capturam a característica de regressão da própria série temporal dos modelos AR e, também, incorporam o modelo de média móvel dos erros dos modelos MA (PAL; PRAKASH, 2017).

Um modelo ARMA(1,1) é descrito pela seguinte função:

$$X_t = \alpha + \theta_1 E_{t-1} + \phi_1 X_{t-1} + E_t$$

Um modelo ARMA(p,q) de ordem p,q é descrito pela seguinte função (DURBIN; KOOPMAN, 2001):

$$X_t = \alpha + \theta_1 E_{t-1} + \phi_1 X_{t-1} + \dots + \theta_q E_{t-q} + \phi_p X_{t-p} + E_t$$

3.3.3.4 Modelos ARIMA

É uma generalização dos modelos ARMA incluindo um componente de integração que tem como função converter a série em estacionária. O componente de integração é responsável por diferenciar a série uma ou mais vezes, diminuindo caráter não estacionário da mesma (HYNDMAN; ATHANASOPOULOS, 2018).

Para diferentes valores de integração (d) a resposta muda da seguinte forma:

$$d = 0 : x_t = x_t$$

$$d = 1 : x_t = x_t - x_{t-1}$$

$$d = 2 : x_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2}$$

Desta maneira um modelo de ordem autoregressiva p , de integração d e de média móvel q , ARIMA(p,d,q) pode ser representado da seguinte maneira:

$$\Delta_d x_t = \alpha + \phi_p \Delta_d x_{t-p} + \phi_1 \Delta_d x_{t-1} + \dots + \theta_1 E_{t-1} + \theta_q E_{t-q} + E_t$$

3.3.3.5 Modelos SARIMA

Estes modelos consideram um componente sazonal para a autorregressão, para a média móvel e para a diferenciação (HYNDMAN; ATHANASOPOULOS, 2018), com o objetivo de realizar previsão mais acertada quando a série temporal apresentar sazonalidade característica.

Um modelo SARIMA(p,d,q) (P,D,Q,s) pode ser representado pela seguinte equação, onde as variáveis P,D,Q e s representam a ordem da autorregressão sazonal, da diferenciação sazonal, da média móvel sazonal e do tamanho do ciclo, respectivamente (SEABOLD; PERKTOLD, 2010).

$$\Delta_d \Delta_D^s x_t = \alpha + \hat{\phi}_p \phi_p \Delta_d \Delta_D^s x_{t-p} + \hat{\phi}_1 \phi_1 \Delta_d \Delta_D^s x_{t-1} + \dots + \hat{\theta}_1 \theta_1 E_{t-1} + \hat{\theta}_q \theta_q E_{t-q} + E_t$$

3.4 Critérios de avaliação dos resultados obtidos

O primeiro ponto avaliado ao verificar a qualidade de uma previsão de série temporal é justamente a capacidade do modelo em representar o comportamento dos dados de maneira visual.

Além disso, neste trabalho, o critério de AIC (*Akaike's Information Criterion*) é utilizado para auxiliar na escolha do melhor modelo SARIMA, o critério de Akaike é calculado utilizando a seguinte equação:

$$AIC = \ln(\hat{\sigma}^2) + \frac{2k}{T}$$

Onde T é o tamanho da amostra, $\hat{\sigma}$ é a variância residual do modelo e k é o número total de parâmetros a serem estimados (GUIDOLIN; PEDIO, 2018).

Para a comparação de diferentes metodologias de estimação foram utilizados os indicadores que também auxiliam no ajuste dos dados e que são usados para determinar a qualidade de um modelo (FACCHIN, 2005, *apud* MOTKE, 2018, p. 16). O principal destes é o da soma quadrática do erro (SSE), que mede os desvios dos dados previstos pelo modelo e os dados originais da série temporal, aliado a variância dos dados (S_{yy}), obtém-se a expressão final do cálculo do coeficiente de determinação, R^2 . As equações abaixo demonstram o cálculo destes índices:

$$SSE = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

$$S_{yy} = \sum_{i=1}^N (y_i - \bar{y})^2$$

$$R^2 = 1 - \frac{SSE}{S_{yy}}$$

Onde y_i são os valores gerados pelo modelo, \hat{y}_i os dados originais utilizados para elaboração do modelo e \bar{y} o valor médio do conjunto de dados originais.

Ainda, foram realizados testes para verificar se os modelos implementados tornaram os resíduos em estacionários. Para esta análise, foram utilizadas as funções autocorrelação e autocorrelação parcial e o teste *Augmented Dickey-Fuller* (ADF).

Neste trabalho os gráficos de autocorrelação e autocorrelação parcial foram plotados com intervalo de confiança de 95%. Se uma função for estacionária, então, 95 % da amostra da autocorrelação deveria estar dentro do intervalo de confiança (BROCKWELL; DAVIS, 1996).

4 Estudo de Caso

Neste trabalho são abordadas três séries temporais distintas de maneira a possibilitar a análise de quais modelos possuem melhor desempenho em cada uma das aplicações. Para tanto, foram selecionados os seguintes dados:

- Medida de concentração em partes por milhão (ppm) mensal de CO_2 no ar de 1974 a 1987.
- Evolução da cotação diária da ação da Braskem, BRKM3, de fevereiro de 2017 a 2018.
- Medida a cada minuto da diferença de pressão de um leito de adsorção de umidade de uma unidade desumidificadora em uma plataforma de extração de petróleo durante 30 mil minutos.

A escolha das séries temporais reside na natureza de cada uma das mesmas. A concentração de CO_2 apesar de apresentar tendência possui oscilações mais demoradas e padronizadas, desta forma, deve apresentar maior facilidade de predição.

A série da cotação da Braskem é bastante volátil devido à natureza “nervosa” do mercado de ações. Espera-se que previsibilidade para este conjunto de dados seja inferior aos demais.

Os dados da unidade de adsorção são industriais pertencendo a um sistema mais fechado que os demais, no entanto, as variações abruptas de diferença de pressão e o caráter de variância não estacionária devem aumentar a dificuldade na predição dessas séries.

4.1 Evolução da concentração de CO_2

Os dados disponíveis são médias mensais da concentração de CO_2 na atmosfera em ppm (TANS; KEELING, 2018) entre o período de maio de 1974 a setembro de 1987.

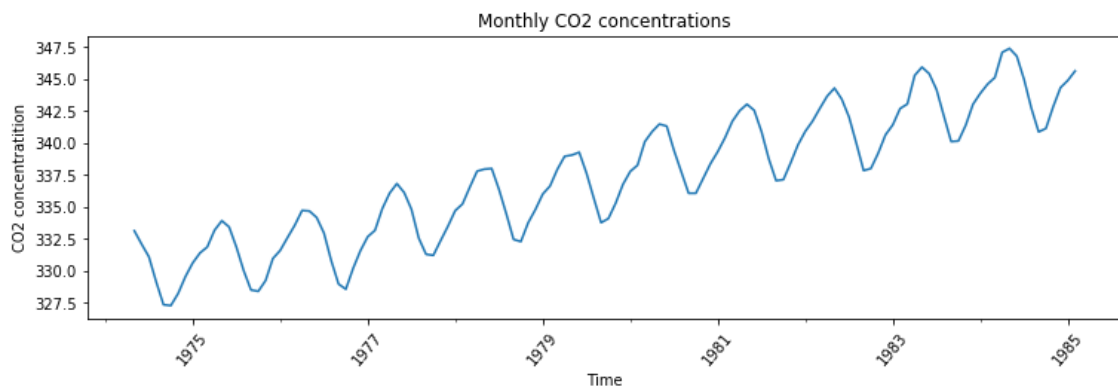
4.1.1 Pré-tratamento da série temporal

Para a elaboração das previsões foram utilizados os dados até fevereiro de 1985 (conjunto de treino) e a partir daí se realizou a previsão comparando com o restante dos dados disponíveis (conjunto de teste).

4.1.2 Visualizar a série temporal

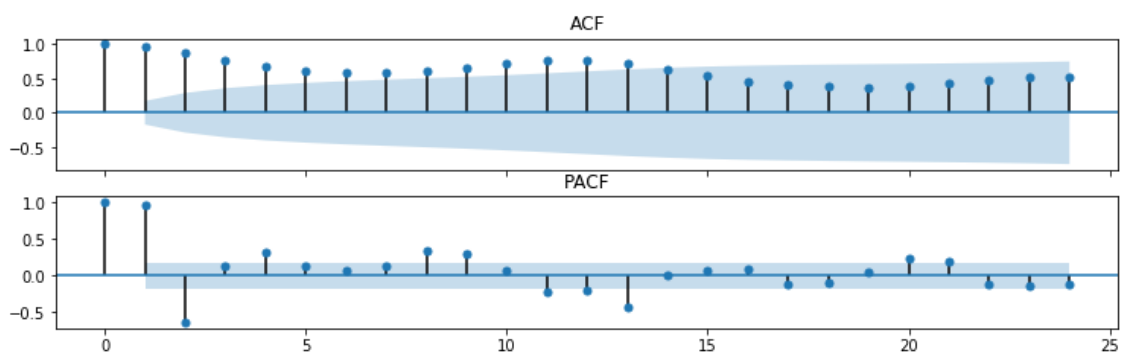
Seguindo a metodologia, é plotada a série temporal, as funções de autocorrelação e autocorrelação parcial, para obter melhor entendimento acerca dos dados e identificar a presença e as características dos componentes internos dos dados, conforme indicado na Figura 4.

Figura 4 - Evolução da concentração de CO_2 em ppm de 1974 a 1985



Fonte: TANS; KEELING, 2018.

Figura 5 - Autocorrelação e autocorrelação parcial da evolução da concentração de CO_2 .



Fonte: Elaborado pelo autor.

Desta maneira é perceptível a presença de tendência e, a partir da análise dos picos do gráfico de autocorrelação, se permite estimar a sazonalidade em 12 meses.

Nota-se a existência de forte autocorrelação entre os dados da série temporal e elevada autocorrelação parcial até o *lag* 2.

4.1.3 Definir e aplicar modelos de previsão

Devido à presença de tendência e sazonalidade bem definidas e constantes serão utilizados os seguintes modelos:

- Decomposição da série temporal, Método A;

- Suavização exponencial tripla;
- SARIMA.

4.1.3.1 Decomposição da série temporal, Método A

A partir da regressão linear realizada, foi identificada a seguinte equação para a tendência:

$$y_t = 329,419 + 0,122x_t$$

Foi removida a tendência da série original a fim de determinar a sazonalidade. O valor determinado para as variações sazonais são os seguintes:

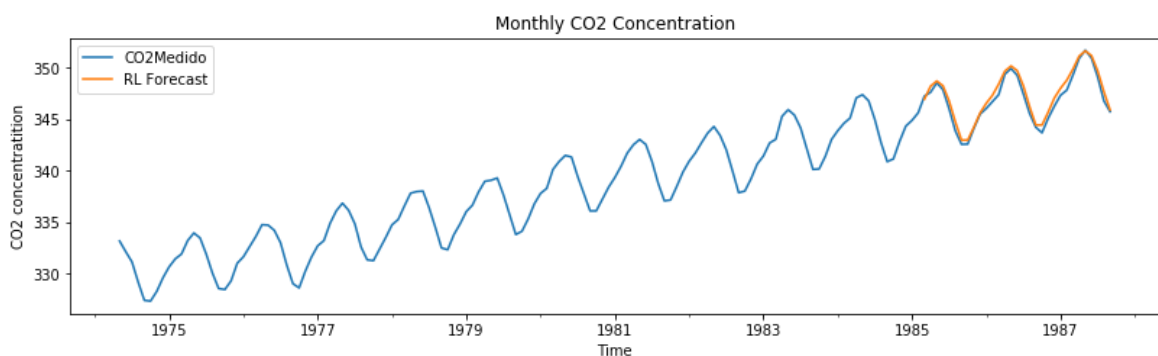
Tabela 1: Valores médios determinados para cada unidade de tempo do ciclo de 12 meses.

Mês	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
Média	-0,07	0,58	1,54	2,69	3,07	2,48	0,87	-1,27	-3,17	-3,29	-2,15	-0,90

Fonte: Elaborado pelo autor.

O modelo desenvolvido foi extrapolado para o restante de 1985, para os anos de 1986 e 1987 e comparado com os valores medidos no mesmo período.

Figura 6 - Comparação da concentração de CO₂ Medida (linha azul) com a concentração prevista com o modelo obtido através da metodologia A (linha laranja).

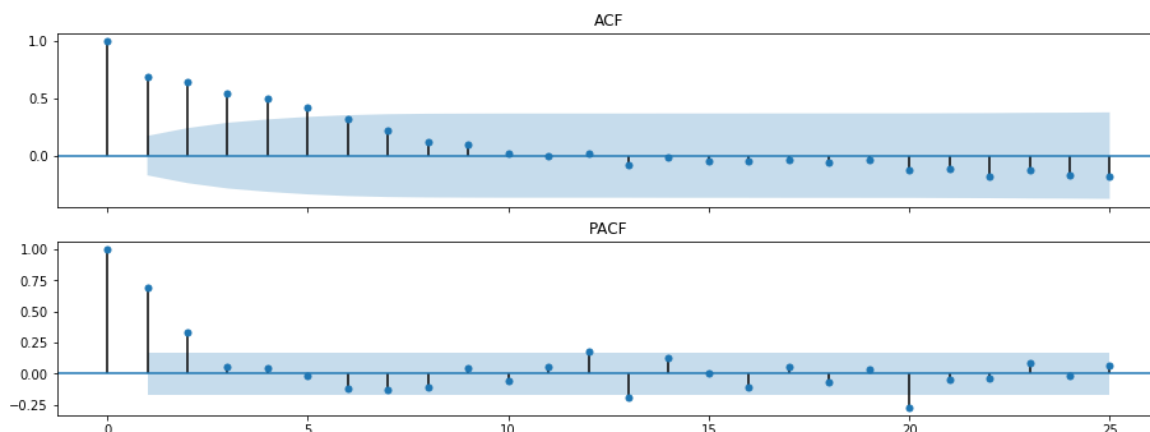


Fonte: Elaborado pelo autor.

As funções de autocorrelação e autocorrelação parcial foram aplicadas aos resíduos para verificar se ainda há correlação entre os dados remanescentes. Conforme indicado pela Figura 7, verifica-se que apesar da autocorrelação para lags

maiores terem reduzido, ainda há correlação significativa entre os instantes mais próximos, indicando dependência entre os resíduos.

Figura 7 - Autocorrelação e autocorrelação parcial dos resíduos gerados pelo método A.



Fonte: Elaborado pelo autor.

Apesar disto o valor de R^2 obtido foi de 0,969 e o teste *Augmented Dickey-Fuller* (ADF) apresentou valor de 0,0236, indicando um bom modelo de previsão e que os resíduos são estacionários.

4.1.3.2 SARIMA

Para identificar a ordem do modelo, aliado aos gráficos de PACF e ACF e análise visual, foram realizadas otimização simulando todas as constantes (p,d,q,P,D,Q) de 0 a 3 de modo a reduzir o valor de AIC. Este valor foi obtido para um modelo SARIMA (2,1,0) (0,1,1,12).

As constantes obtidas com o modelo são apresentadas na Tabela 2. ϕ_1 é a constante autorregressiva no *lag* 1, ϕ_2 é a constante autorregressiva no *lag* 2, $\hat{\theta}_1$ é a constante de média móvel sazonal no *lag* 12 (um ciclo completo).

Tabela 2 - Valores determinados dos coeficientes do modelo SARIMA(2,1,0)(0,1,1,12).

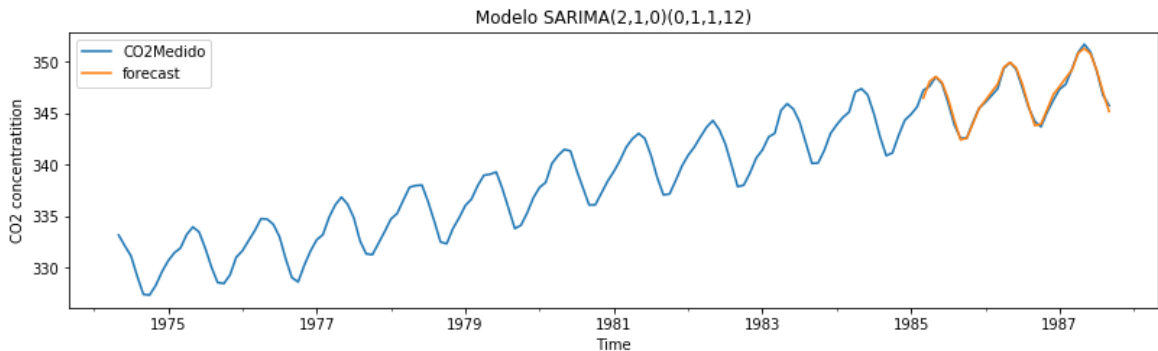
Coeficiente	Valor
ϕ_1	-0.4513
ϕ_2	-0.0982
$\hat{\theta}_1$	-0.6065

Fonte: Elaborado pelo autor.

24 Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

A Figura 8 representa a comparação da concentração de CO_2 medido contra os dados previstos pelo modelo SARIMA.

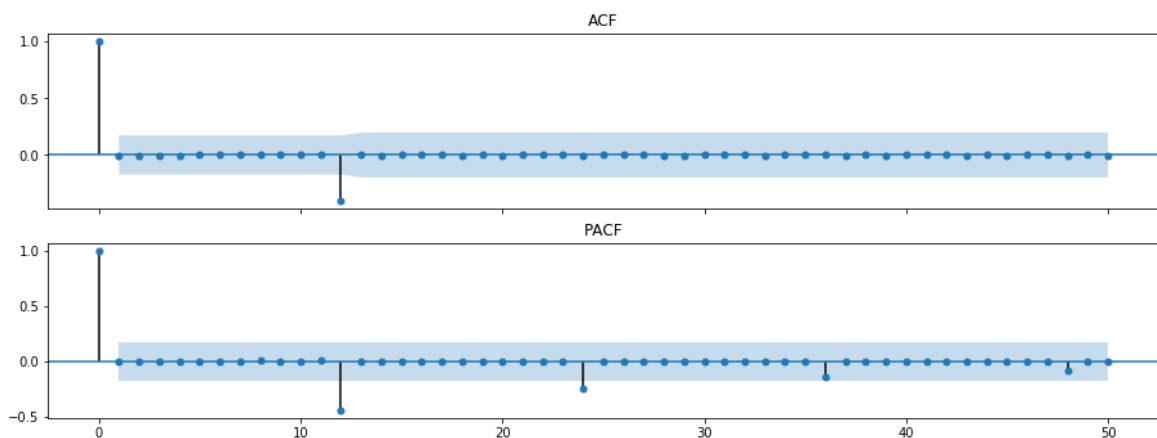
Figura 8 - Comparação da concentração de CO_2 Medida em ppm (linha azul) com a concentração prevista com o modelo SARIMA(2,1,0)(0,1,1,12) (linha laranja).



Fonte: Elaborado pelo autor.

As funções de autocorrelação e autocorrelação parcial foram aplicadas aos resíduos para verificar se ainda há correlação entre os dados remanescentes. Conforme indicado pela Figura 9, verifica-se que a autocorrelação foi bem capturada pelo modelo, indicando baixa dependência temporal entre os resíduos.

Figura 9 - Autocorrelação e autocorrelação parcial dos resíduos do modelo gerado pelo modelo SARIMA(2,1,0)(0,1,1,12).



Fonte: Elaborado pelo autor.

Os valores de R^2 e do teste ADF são respectivamente 0,980 e 0,000, indicando um bom modelo e confirmando que os resíduos são estacionários.

4.1.3.3 Suavização exponencial tripla

Devido à presença de tendência e de sazonalidade foi realizada exponenciação sazonal tripla. Os parâmetros foram obtidos de modo a minimizar o erro da predição e estão representados na Tabela 3.

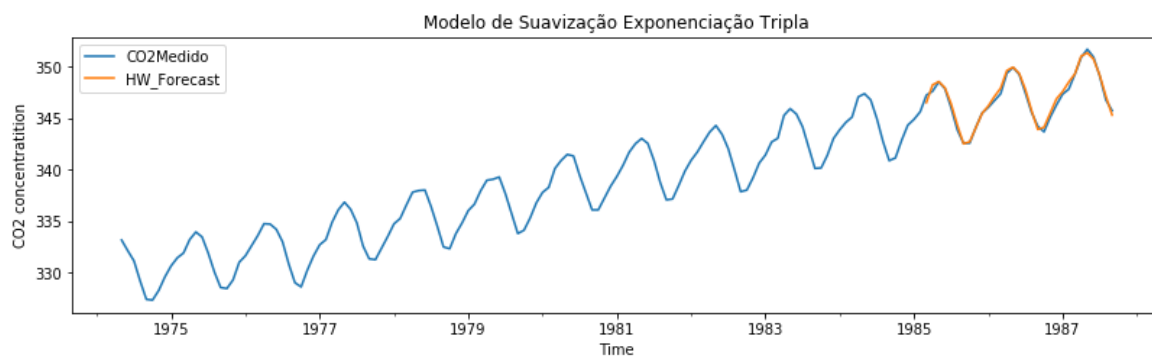
Tabela 3 - Coeficientes determinados para o método da exponenciação tripla.

Coeficiente	Valor
α	0,56
β	$1,54 \cdot 10^{-8}$
γ	0,31

Fonte: Elaborado pelo autor.

A figura 10 representa a comparação da concentração de CO₂ medido contra os dados previstos pelo modelo de exponenciação tripla.

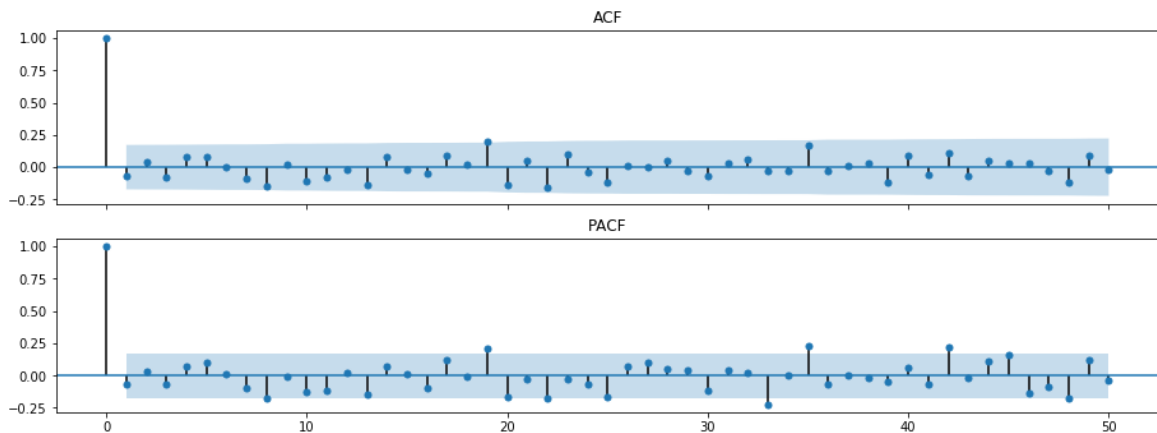
Figura 10 - Comparação da concentração em ppm de CO₂ Medida (linha azul) com a concentração prevista pelo modelo de suavização exponencial (linha laranja).



Fonte: Elaborado pelo autor.

As funções de autocorrelação e autocorrelação parcial foram aplicadas aos resíduos para verificar se ainda há correlação entre os dados remanescentes. Conforme indicado pela Figura 11, verifica-se que a autocorrelação foi bem capturada pelo modelo, indicando baixa dependência temporal entre os resíduos.

Figura 11 - Autocorrelação e autocorrelação parcial dos resíduos do modelo gerado pelo modelo de suavização exponencial tripla.



Fonte: Elaborado pelo autor.

Os valores de R^2 e do teste ADF são respectivamente 0,978 e 0,000, indicando um bom modelo e confirmando que os resíduos são estacionários.

4.1.4 Avaliar resultados obtidos

Os modelos que apresentaram melhores resultados, devido às características estacionárias dos resíduos, ao coeficiente de determinação e à capacidade de representar visualmente a série temporal original, foram os de suavização exponencial tripla e o modelo SARIMA(2,1,0)(0,1,1,12).

4.1.5 Previsões de longo prazo

Os dois modelos que apresentaram os maiores R^2 foram estendidos para localizar em qual período a concentração média mensal alcançaria 400 ppm pela primeira vez, está é uma importante posição visto que o incremento de CO_2 na atmosfera gera diversos impactos na flora e fauna.

Segundo o modelo SARIMA e o de Suavização exponencial tripla esta marca seria atingida somente em abril de 2023, no entanto, em abril de 2014 foi registrada média mensal de 401 ppm, demonstrando que a característica de tendência da série temporal, no decorrer do tempo, incrementou seu carácter exponencial (TANS; KEELING,2018).

4.2 Diferença de pressão em um adsorvedor

Os dados utilizados foram medidos em intervalo de um minuto totalizando 32.700 minutos de operação de um adsorvedor operando com intuito de reduzir a umidade de água a 1 ppm em uma corrente de gás de uma plataforma de petróleo.

4.2.1 Pré-tratamento da série temporal

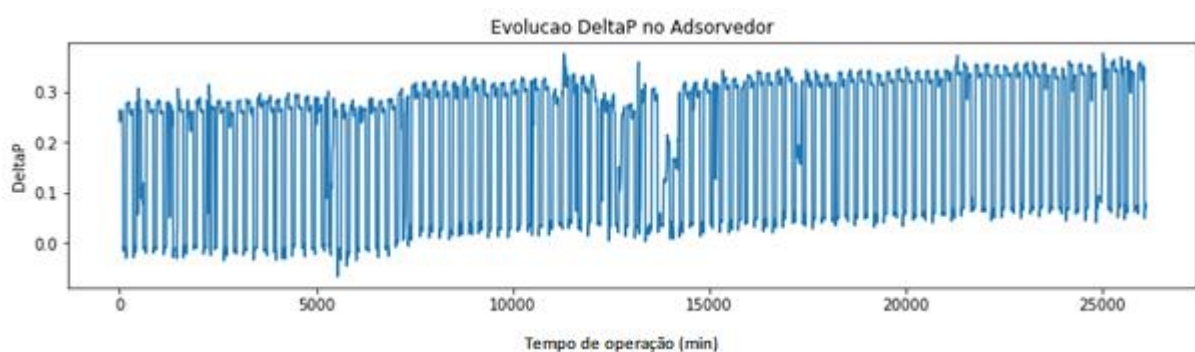
Devido ao elevado número de pontos, optou-se por se utilizar uma amostragem de 10 minutos, ou seja se pegou 1 ponta de cada 10 amostragens operacionais, tendo-se o cuidado de se pegar o início do ciclo para evitar defasagens decorrentes do processo de re-amostragem.

Para o conjunto de treino, foi utilizada a amostra do intervalo de 0 a 26.100 minutos de operação. O conjunto de teste foi realizado comparando a previsão desenvolvida com as amostras dos minutos 26.100 ao 32.700.

4.2.2 Visualizar a série temporal

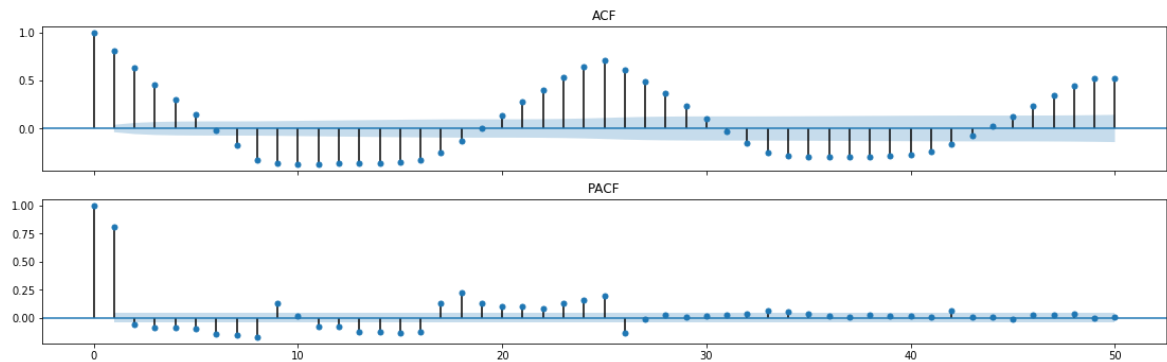
Para obter melhor entendimento acerca dos dados e identificar a presença e as características dos componentes internos da série temporal, foi plotado o conjunto de treino, Figura 12, e calculadas as funções de autocorrelação e autocorrelação parcial, Figura 13.

Figura 12 - Evolução da diferença de pressão em adsorvedor com amostras de intervalo de 10 minutos.



Fonte: Dados de adsorvedor de umidade em planta plataforma de extração de petróleo.

Figura 13 - Autocorrelação e autocorrelação parcial do delta P até o lag 50.



Fonte: Elaborado pelo autor.

Nota-se a presença de tendência, a partir da análise dos picos do gráfico de autocorrelação permite-se estimar a sazonalidade em 25 amostras ou 250 minutos. Ainda, destaca-se forte autocorrelação entre os dados e elevada autocorrelação parcial até o lag 2.

4.2.3 Definir e aplicar modelos de previsão

Devido às características observadas na etapa de visualização da série temporal, foram utilizados os seguintes modelos:

- Decomposição da série temporal, Método B;
- SARIMA.

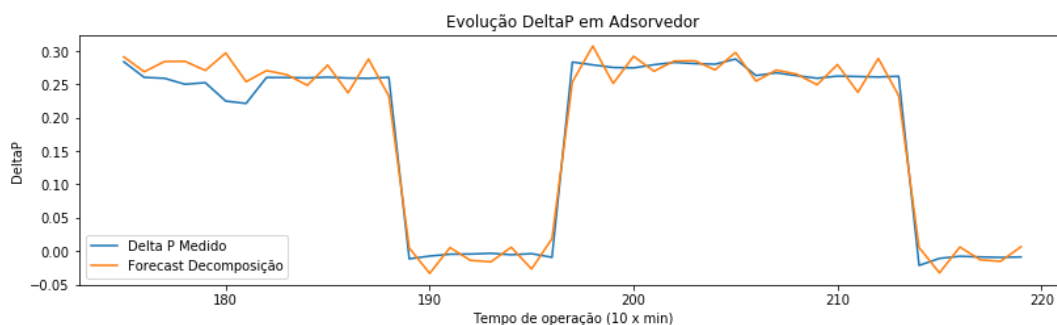
4.2.3.1 Decomposição da série temporal, Método B

A tendência foi capturada realizando regressão polinomial, a equação identificada está representada abaixo.

$$y_t = 0,163 + 3,27x_t$$

Foi então removida a tendência da série original a fim de determinar a sazonalidade através da utilização do método da FFT. Para tanto, foi selecionado um ciclo de adsorção que foi considerado representativo, sem ocorrências operacionais espúrias, e o resultado (Figura14) identificado foi replicado aos demais períodos.

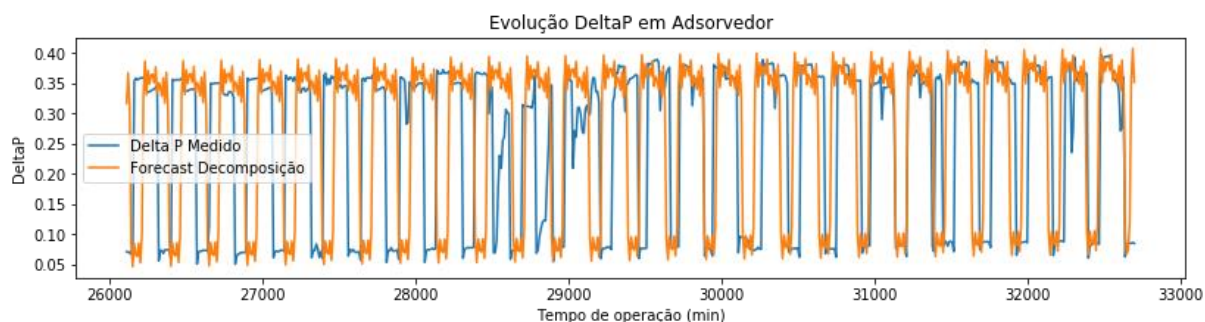
Figura 14 - Comportamento sazonal da série temporal identificado pela Transformada Rápida de Fourier utilizando 10 harmônicas.



Fonte: Elaborado pelo autor.

Os componentes sazonais e de tendência foram adicionados e o modelo foi extrapolado para realizar a comparação com o conjunto de teste:

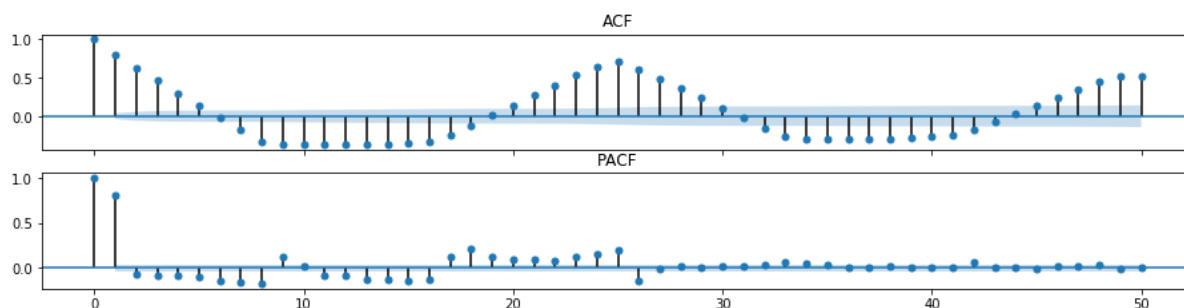
Figura 15 - Comparação da diferença de pressão medida (linha azul) do conjunto de teste com a diferença de pressão prevista pela utilização do método B de decomposição (linha laranja).



Fonte: Elaborado pelo autor.

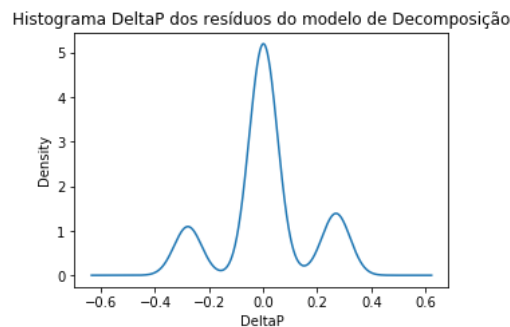
Os valores de R^2 e do teste ADF são respectivamente 0,782 e 0,000, indicando modelo razoável e que os resíduos são estacionários. No entanto, as funções de ACF, PACF e a distribuição dos resíduos apontam que há comportamentos determinísticos que não foram capturados pelo modelo.

Figura 16 - Autocorrelação e autocorrelação parcial do delta P dos resíduos até o lag 50.



Fonte: Elaborado pelo autor.

Figura 17 - Histograma dos resíduos indicando que estes não seguem distribuição normal em torno de 0.



Fonte: Elaborado pelo autor.

4.2.3.2 SARIMA

Para identificar a ordem do modelo foram realizadas otimização simulando as constantes (p, d, q, P, D, Q) conforme indicado na Tabela 4 de modo a identificar quais modelos obtêm menor AIC.

Tabela 4 - Representação do *range* utilizado na otimização do AIC

	p	d	q	P	D	Q
Início	0	0	0	0	0	0
Fim	13	1	5	3	2	3

Fonte: Elaborado pelo autor.

Uma vez identificados os coeficientes que minimizam o valor de AIC, eles foram utilizados na realização de simulações e foram avaliadas visualmente a qualidade das previsões produzidas por estes modelos. O modelo escolhido foi o SARIMA (2,0,1)(0,1,0,25)

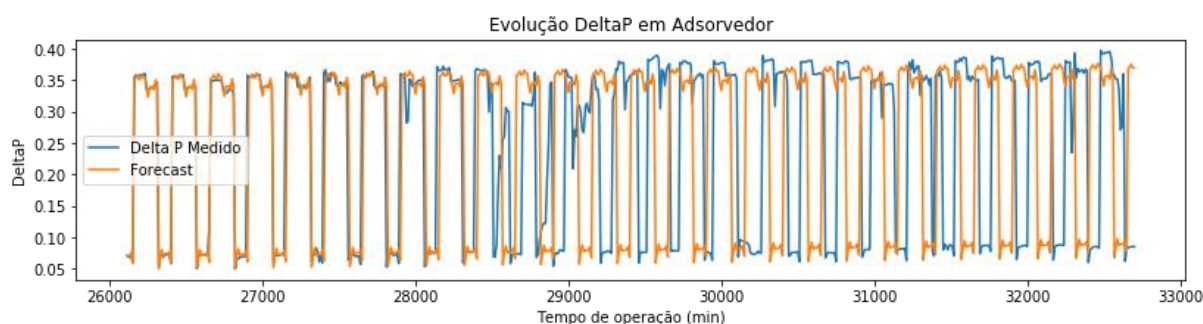
As constantes obtidas com o modelo são apresentadas na Tabela 5. ϕ_1 é a constante autorregressiva no *lag* 1, ϕ_2 é a constante autorregressiva no *lag* 2, θ_1 é a constante de média móvel sazonal no *lag* 1 e α é a intercepção média no eixo das ordenadas.

Tabela 5 - Valores determinados das constantes do modelo SARIMA(2,0,1)(0,1,0,25).

	ϕ_1	ϕ_1	θ_1	α
Valor	0,4308	0,1376	0,1295	0,0003

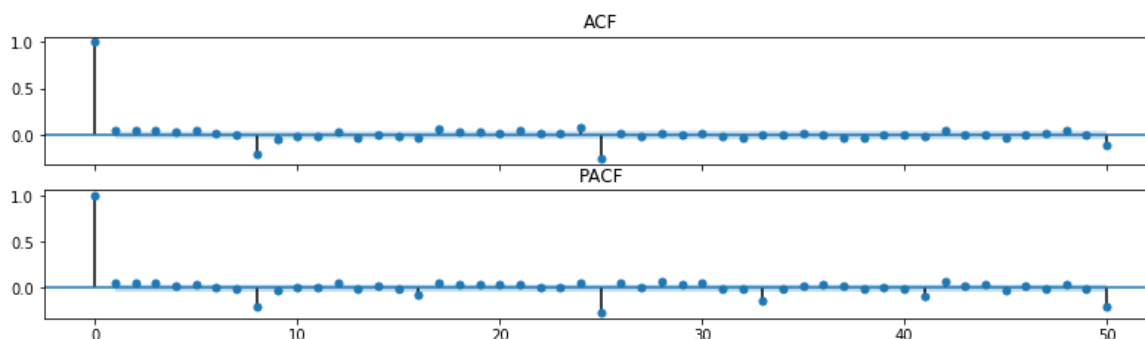
Fonte: Elaborado pelo autor.

A Figura 18 representa a comparação da diferença de pressão medida no conjunto de teste contra os dados previstos pelo modelo de SARIMA(2,0,1)(0,1,0,25).

Figura 18 - Comparação da diferença de pressão medida (linha azul) do conjunto de teste com a diferença de pressão prevista pela utilização do modelo SARIMA (2,0,1)(0,1,0,25) (linha laranja).

Fonte: Elaborado pelo autor.

As funções de autocorrelação e autocorrelação parcial foram aplicadas aos resíduos para verificar se ainda há correlação entre os dados remanescentes. Conforme indicado pela Figura 19, verifica-se que a autocorrelação foi bem capturada pelo modelo, indicando baixa dependência temporal entre os resíduos.

Figura 19 - Autocorrelação e autocorrelação parcial dos resíduos gerado pelo modelo de SARIMA (2,0,1)(0,1,0,25) na base de treino.

Fonte: Elaborado pelo autor.

Os valores de R^2 e do teste ADF são respectivamente 0,762 e 0,000, indicando modelagem razoável e confirmando que os resíduos são estacionários.

4.2.4 Avaliar resultados obtidos

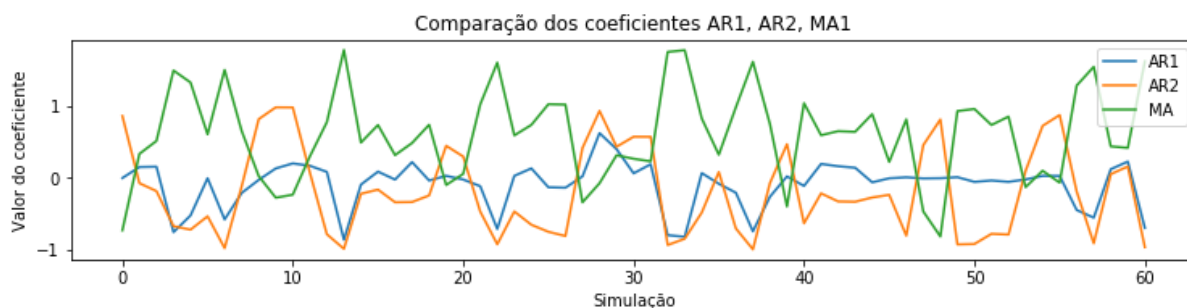
O modelo que apresentou melhor resultado, devido às características estacionárias dos resíduos, ao coeficiente de determinação e à capacidade de representar visualmente a série temporal original, foi o SARIMA (2,0,1)(0,1,0,25).

4.2.5 Avaliação da dinâmica da série temporal

Para obter melhor entendimento em relação à dinâmica da operação do adsorvedor, foram realizadas 60 simulações do modelo SARIMA(2,0,1)(0,1,0,25) com intervalos constantes de 75 observações, porém, variando o início de 0 a 2585 com passo de 25 observações e o fim do intervalo de 75 a 2610, também com passo de 25.

Uma vez formada a matriz dos intervalos com os coeficientes p_1, p_2 e q_1 foi avaliada a existência de algum padrão ou tendência para os resultados determinados. No entanto, as simulações se mostraram inconclusivas devido à característica aleatória dos coeficientes conforme indica a Figura 20.

Figura 20 – Figura evidenciando a variação dos coeficientes AR1, AR2 e MA1 nas 60 simulações realizadas.



Fonte: Elaborado pelo autor.

4.3 Evolução da cotação da ação BRKM3

Os dados disponíveis são de fevereiro de 2017 a novembro de 2018 e representam a cotação diária das ações da Braskem, BRKM3, na bolsa de valores do Brasil.

4.3.1 Pré-tratamento da série temporal

Para a elaboração e avaliação dos modelos as bases foram divididas em dois períodos. Para o conjunto de treino foram empregados os dados de fevereiro de 2017 a 13 de agosto de 2018. Com a base treino foram definidos os modelos que melhor

desempenham na previsão da série temporal, além de realizar a avaliação dos resíduos.

A base teste, necessária para realizar a comparação com os dados de previsão gerados pelo modelo, inicia no dia 14 de agosto e termina em 01 de novembro.

4.3.2 Visualizar a série temporal

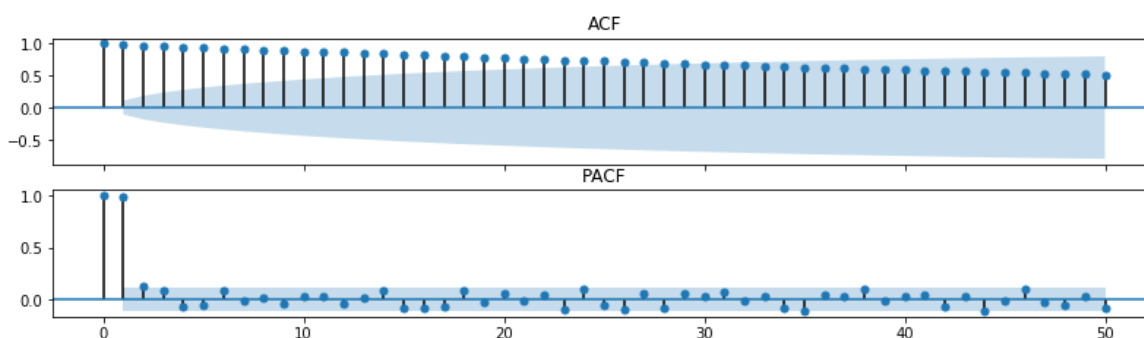
Foi plotada a série temporal para obter melhor entendimento acerca dos dados e identificar a presença e as características dos componentes internos da série temporal.

Figura 21 - Evolução da cotação da BRKM3 em intervalo diário minutos, conjunto de treino.



Fonte: Bolsa de valores do Brasil (B3, 2018), dados do fechamento da ação BRKM3.

Figura 22 - Autocorrelação e autocorrelação parcial do preço da ação até o lag 50.



Fonte: Elaborado pelo autor.

A visualização da série temporal e das funções de ACF e PACF indicam que apesar da grande volatilidade dos dados, há presença de tendência para a cotação da Braskem apesar da ausência de sazonalidade.

Ainda, Nota-se elevada de autocorrelação parcial até o lag 2, e elevada autocorrelação parcial até o lag 27.

4.3.3 Definir e aplicar modelos de previsão

Devido à presença de tendência, à tendência irregular, à inexistência de sazonalidade e à elevada autocorrelação serão utilizados os seguintes modelos:

- ARIMA;
- Suavização exponencial dupla.

4.3.3.1 SARIMA

Para identificar a ordem do modelo foram realizadas otimização simulando as constantes (p,d,q) conforme indicado na Tabela 6, de modo a identificar quais modelos obtêm menor AIC.

Tabela 6 - Representação do *range* utilizado na otimização do AIC.

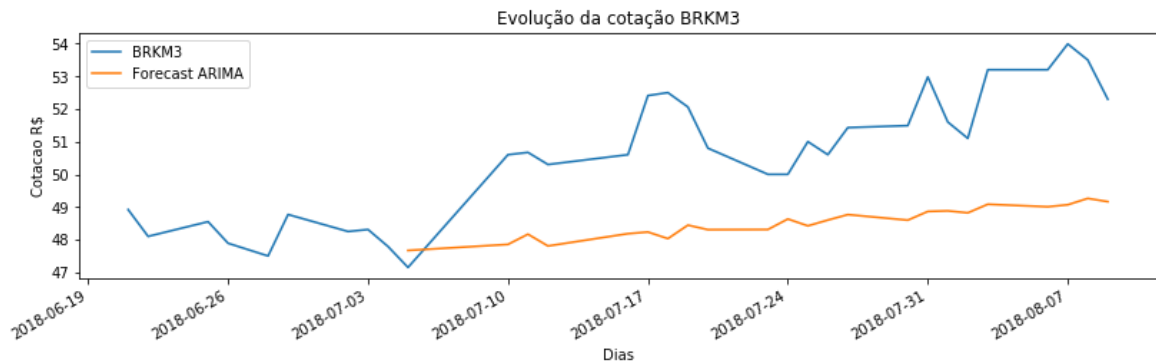
	p	d	q
Início	0	1	0
Fim	29	2	10

Fonte: Elaborado pelo autor.

Os valores que minimizavam o AIC foram utilizados na realização de simulações e estas foram avaliadas visualmente em relação à qualidade das previsões produzidas. Apesar dos modelos de ordem mais baixa produzirem resultados de AIC similares aos de elevada ordem, a simulação gráfica indica melhor capacidade de representar o comportamento da série temporal para os coeficientes de maior ordem. Esta avaliação gráfica é representada nas Figuras 23 e 24.

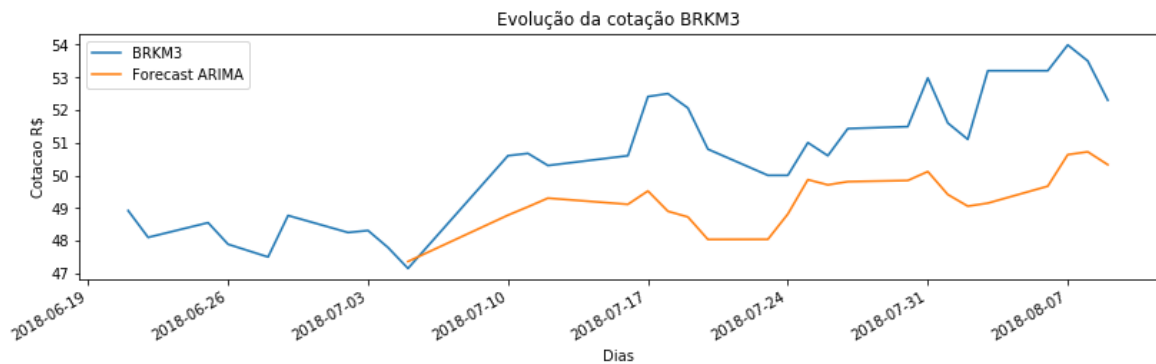
O melhor valor de AIC obtido na otimização foi de 1.024 para o modelo ARIMA (2,1,3), a comparação será feita com o modelo ARIMA (29,1,9) que demonstrou melhor aderência aos dados originais mesmo tendo apresentado AIC de 1.042.

Figura 23 - Comparação da cotação medida (linha azul) do conjunto de treino com a cotação prevista pelo modelo ARIMA (2,1,3) (linha laranja).



Fonte: Elaborado pelo autor.

Figura 24 - Comparação da cotação medida (linha azul) do conjunto de treino com a cotação prevista pelo modelo ARIMA (29,1,9) (linha laranja).



Fonte: Elaborado pelo autor.

Os coeficientes obtidos para o modelo ARIMA (29,1,9) estão representados na Tabela 7.

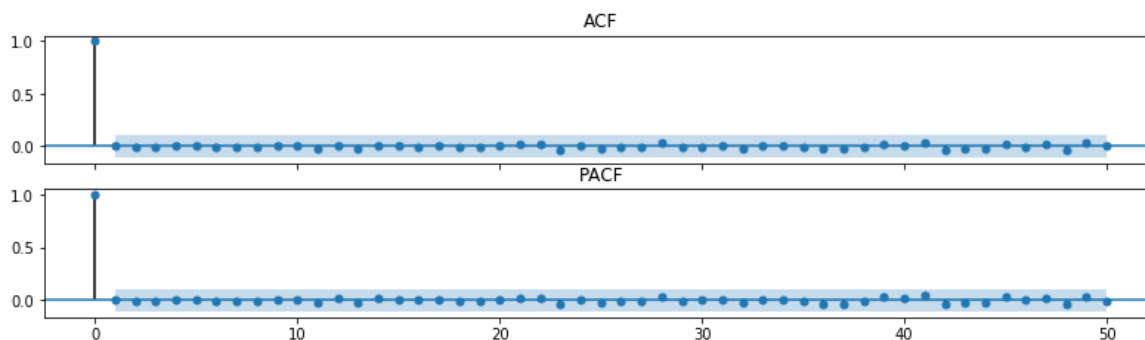
Tabela 7 - Valores determinados dos coeficientes do modelo ARIMA (29,1,9).

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}
Valor	0,2711	-0,3961	0,0888	-0,1163	-0,0886	0,0493	0,2520	-0,3384	0,7471	0,0292	0,1621	-0,0936
	ϕ_{13}	ϕ_{14}	ϕ_{15}	ϕ_{16}	ϕ_{17}	ϕ_{18}	ϕ_{19}	ϕ_{20}	ϕ_{21}	ϕ_{22}	ϕ_{23}	ϕ_{24}
Valor	-0,1411	0,2561	-0,0271	0,0480	-0,0896	0,0545	-0,0207	-0,0391	-0,1349	0,1535	-0,1054	0,0872
	ϕ_{25}	ϕ_{26}	ϕ_{27}	ϕ_{28}	ϕ_{29}	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
Valor	-0,1145	0,1165	0,0220	0,0409	-0,0758	-0,4122	0,3464	-0,0164	0,0872	-0,0227	0,0889	-0,3829
	θ_8	θ_9	α									
Valor	0,4327	-0,8796	0,0318									

Fonte: Elaborado pelo autor.

A avaliação dos resíduos é realizada através das funções de autocorrelação parcial, autocorrelação, Figura 25, e do teste ADF.

Figura 25 - Autocorrelação e autocorrelação parcial dos resíduos do modelo até o lag 50.

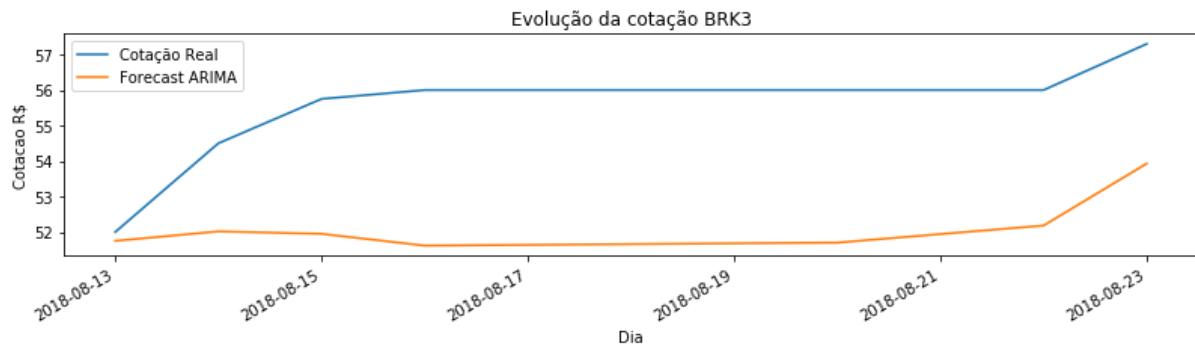


Fonte: Elaborado pelo autor.

O valor apresentado para o teste ADF foi 0, isto alinhado às funções ACF e PACF indicam que os resíduos são estacionários.

A Figura 26 representa a comparação da cotação da BRKM3 do conjunto de teste contra os dados previstos pelo modelo ARIMA(29,1,9).

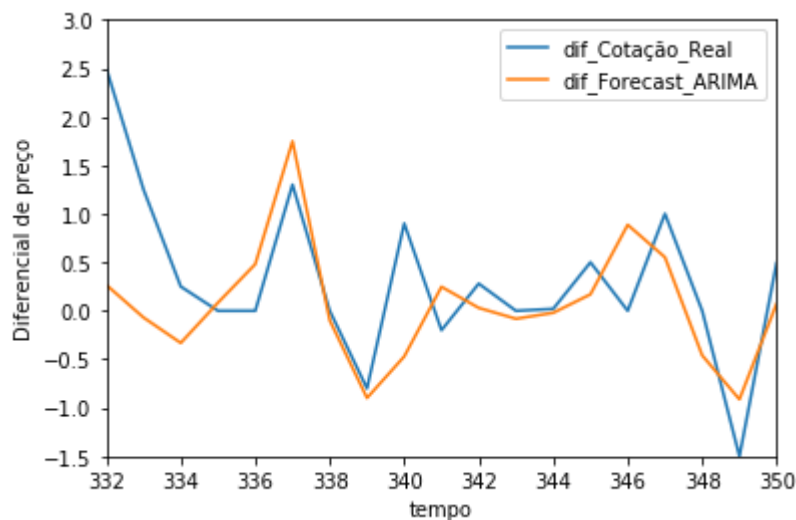
Figura 26 - Comparação de 7 dias da cotação medida do conjunto de teste (linha azul) com a cotação prevista pelo modelo ARIMA (29,1,9) (linha laranja).



Fonte: Elaborado pelo autor.

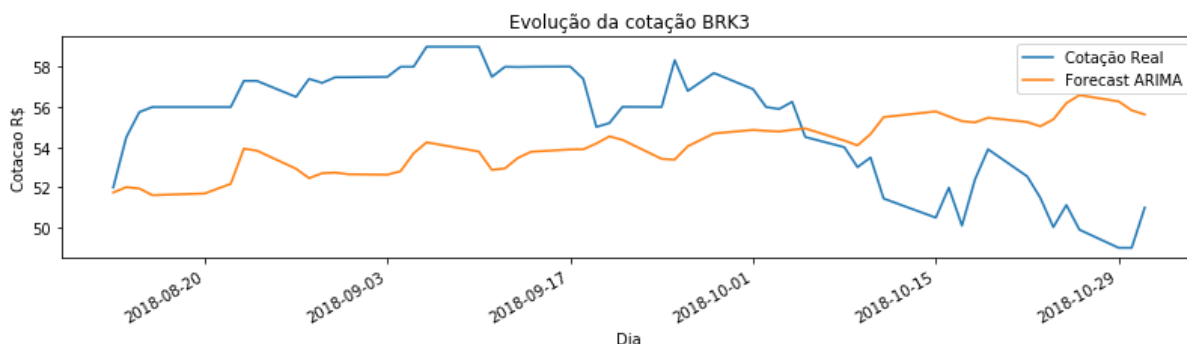
Pode-se inferir que apesar dos valores previstos não corresponderem aos da cotação realizada o modelo conseguiu capturar parcialmente o comportamento da série temporal em um intervalo de tempo não muito distante. Capturando a tendência de alta e baixa das ações, conforme representado na Figura 27.

Figura 27 – Comparação entre a primeira derivada do modelo preditivo ARIMA(29,1,9) e a primeira derivada da ação do conjunto de teste.



Fonte: Elaborado pelo autor.

Figura 28 - Comparação de 53 dias da cotação medida do conjunto de teste (linha azul) com a cotação prevista pelo modelo ARIMA (29,1,9) (linha laranja).



Fonte: Elaborado pelo autor.

A Figura 28 mostra que conforme se estende o período da previsão, a qualidade da representação do comportamento da série temporal pelo modelo decai significativamente.

Em virtude do distanciamento entre o valor previsto e o valor realizado, o coeficiente de determinação não foi quantificado como um dos métodos de avaliação da qualidade do modelo para esta série temporal.

4.3.3.2 Suavização exponencial dupla

Devido à presença de tendência foi realizada exponenciação sazonal dupla. Os parâmetros foram obtidos de modo a minimizar o erro da predição e estão representados na Tabela 8.

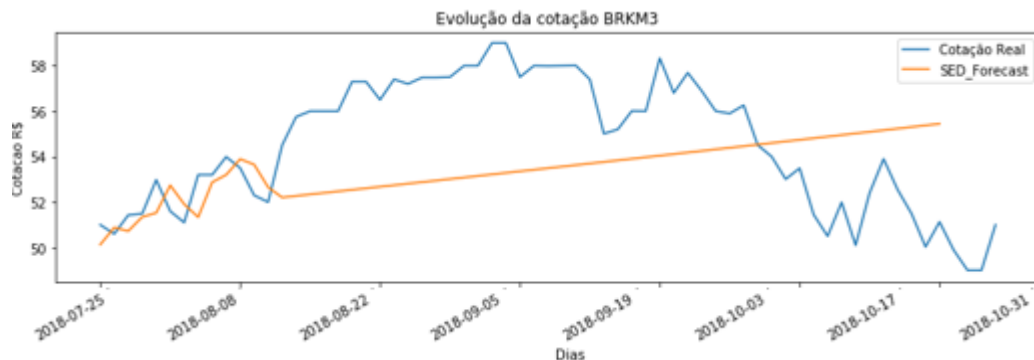
Tabela 8 - Coeficientes determinados pelo método da exponenciação dupla.

	α	β
Valor	0,79	0

Fonte: Elaborado pelo autor.

A Figura 29 representa a comparação da cotação do conjunto teste contra os dados previstos pelo modelo de suavização exponencial dupla.

Figura 29 - Comparação da cotação do conjunto de teste (linha azul) com a cotação prevista pelo modelo de suavização exponencial dupla (linha laranja).



Fonte: Elaborado pelo autor.

Devido à característica “nervosa” da série temporal e tendência variável, o modelo de suavização exponencial, que utiliza parâmetros para ponderar os valores passados da série, não consegue representar a previsão da cotação da ação de maneira adequada.

4.3.4 Avaliar resultados obtidos

O modelo que melhor conseguiu representar os dados da cotação da Braskem, BRKM3, foi o ARIMA (29,1,9). Apesar de não ser exato na previsão, este modelo auxiliou na identificação do comportamento dos valores futuros desta série temporal conforme representado na Figura 26 e na Figura 27.

O aumento do período da previsão piora significativamente a previsão devido à característica “nervosa” dos dados, conforme ilustrado na Figura 28.

5 Conclusões e Sugestões para Trabalhos Futuros

Neste trabalho, os comportamentos de três conjuntos de dados temporais de natureza diferente foram analisados e através de modelos preditivos foram realizadas previsões para os próximos pontos destas séries temporais. Para tanto, foram abordados três conjuntos de técnicas de previsão: As de decomposição dos componentes internos, as de suavização exponencial e as técnicas do grupo SARIMA.

A decisão de quais modelos seriam aplicados em cada uma das séries temporais ocorreu após a análise e entendimento do comportamento dos conjuntos de dados e de modo a aplicar modelos que conseguiriam representar da maneira mais adequada cada uma das três séries temporais.

Desta forma, observou-se que para os dados de CO_2 os três modelos testados conseguiram simular de maneira satisfatória o comportamento da série com destaque ao SARIMA (2,1,0) (0,1,1,12) e ao de Suavização exponencial tripla. Verificou-se também que a tendência de aumento do CO_2 na atmosfera foi intensificada após o período estudado neste trabalho.

Para representar o comportamento da unidade adsorvedora de umidade nos ciclos de adsorção e dessorção optou-se pela utilização dos modelos SARIMA e de regressão linear com a utilização de séries de Fourier para representar a sazonalidade. A representação gráfica dos valores futuros e os resíduos indicaram que o modelo SARIMA (2,0,1) (0,1,0,25) capturou a parte determinística de maneira superior ao método de decomposição. Ainda, foram realizadas simulações dos coeficientes do modelo SARIMA em intervalos variados da série temporal para tentar identificar algum padrão ou alteração na série temporal, porém, este estudo foi inconclusivo.

O modelo que melhor conseguiu representar os dados da cotação da Braskem, BRKM3, foi o ARIMA (29,1,9). Apesar de não ser exato na previsão, este modelo auxiliou na identificação do comportamento dos valores futuros desta série temporal. Também foi avaliada a dificuldade de se ter um valor preciso para os dados da BRKM3 conforme o aumento da distância da previsão, devido à característica “nervosa” dos dados.

Por fim, respondendo ao segundo questionamento, o modelo que melhor conseguiu representar o comportamento futuro das séries temporais em geral foi o

SARIMA. Este fato explica-se pela capacidade dos modelos SARIMA de inserir em suas equações padrões de comportamento passados e utilizá-los para realizar as previsões.

Para futuros trabalhos, afim de melhorar o entendimento acerca do comportamento do adsorvedor, sugere-se que a avaliação do mesmo seja realizada utilizando a razão entre a raiz quadrada da diferença de pressão e a vazão de entrada no equipamento. Ainda, recomenda-se que as simulações dos coeficientes do modelo SARIMA em intervalos variados da série temporal sejam realizados utilizando não valores de periodicidade constantes, de início e término das simulações, mas com a utilização de funções que indiquem a posição de início e término exato de cada ciclo.

Em relação às metodologias de previsão, poderiam ser adotados modelos que levem em consideração não somente variáveis endógenas, mas também as exógenas como os modelos SARIMAX, permitindo utilizar maior volume de informação sobre o fenômeno que se deseja realizar previsão.

Referências

- BOX, George E. P.; JENKINS, Gwilym M.; REINSEL, Gregory C. **Time Series Analysis**. Forecasting Control. 4. ed. Hoboken: John Wiley & Sons, 2008.
- BROCKWELL, Peter J.; DAVIS, Richard A. **Introduction to Time Series and Forecasting**. 3. ed. New York: Springer, 1996.
- CHATFIELD, Christopher. **Time-series forecasting**. Boca Raton: CRC Press, 2000. ISBN 1-58488-063-5.
- DOWNEY, Allen B. **Think DSP: Digital Signal Processing in Python**. 1. ed. Sebastopol (CA): O'Reilly Media, 2016.
- DURBIN, J.; KOOPMAN, S. J. **Time series analysis by state space methods**. New York: Oxford University Press, 2001.
- GUIDOLIN, Massimo; PEDIO, Manuela. **Essentials of time series for financial applications**. Academic Press/Elsevier, 2018.
- HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting: Principles and Practice**. 2. ed. OTexts, 2018. Disponível em: <<https://otexts.org/fpp2/>>. Acesso em: 18 nov. 2018
- LEWIS, Bruce G.; HERBERT, Ric D.; BELL, Rod D. The Application of Fourier Analysis to Forecasting the Inbound Call Time Series of a Call Centre. **Proceedings of the International Congress on Modeling and Simulation MODSIM03**, Townsville, pp. 1281–1286, 2003. Disponível em: <https://www.mssanz.org.au/MODSIM03/Volume_03/B10/06_Lewis.pdf>. Acesso em: 22 ago. 2018.
- MACKINNON, James G. Critical values for cointegration tests. **Queen's Economics Department Working Paper**, n. 1227, 2010. Disponível em: <<https://www.econstor.eu/bitstream/10419/67744/1/616664753.pdf>>. Acesso em: 20 set. 2018
- MARTELOTTE, Marcela Cohen. **Projeto de filtros para ajuste sazonal robustos a variações na sazonalidade**. 2014. 298 f. Tese (Doutorado em Engenharia Elétrica). Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. 2014. Disponível em: <<http://www02.smt.ufrj.br/~eduardo/teses/marcela-cohen-doutorado.pdf>>. Acesso em: 03 nov. 2018
- MORETTIN, Pedro A.; TOLOI, Clélia M. C. **Análise de séries temporais**. 2. ed. São Paulo: EgargBlucher, 2006.
- MOTKE, Mateus Berno. **Predição de pressão de fundo em poços de petróleo via redes neurais**. 2018. 32 f. Trabalho de conclusão (Graduação). Universidade Federal do Rio Grande do Sul. Escola de Engenharia. Curso de Engenharia Química, Porto Alegre, BR-RS, 2018. Disponível em:

<<http://www.bibliotecadigital.ufrgs.br/da.php?nrb=001074436&loc=2018&l=d2f335d2100a49ec>>. Acesso em: 17 nov. 2018

PAL, Avishek; PRAKASH, PKS. **Practical Time Series Analysis**: Master Time Series Data Processing, Visualization, and Modeling using Python. Birmingham: Packt Publishing, 2017.

SEABOLD, Skipper; PERKTOLD, Josef. Statsmodels: Econometric and Statistical Modeling with Python. **Proceedings of the 9th Python in Science Conference**, Austin, v. 57. pp 57-61, 2010.

TANS, Pieter; KEELING, Ralph. **Scripps Institution of Oceanography - Trends in Atmospheric Carbon Dioxide**. Disponível em: <<https://www.esrl.noaa.gov/gmd/ccgg/trends/data.html>> Acesso em: 12 nov. 2018.

B3, 2018 Disponível em: <http://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/cotacoes/> Acesso em: 20 nov. 2018

Apêndice A

Simulação da série temporal da ação BRKM3

```
In [1]: import os
import pandas as pd
%matplotlib inline
from matplotlib import pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from __future__ import print_function
import numpy as np
import seaborn as sns
from statsmodels.tsa.arima_model import ARIMA
import statsmodels.api as sm
import statsmodels.tsa.api as smtsa
from statsmodels.tsa import stattools
from statsmodels.tsa.holtwinters import ExponentialSmoothing
```

```
In [2]: ##Importando os dados
data = pd.read_excel('C:/Users/Thomaz/Documents/ENQ/2018_02/TCC/Dados/Bovespa Historical Data4_train.xlsx', sheet_name='p2')
```

```
In [3]: # SOMENTE PARA AJUSTAR A BD
```

```
In [4]: data.tail(12)
```

Out[4]:

	Data	Bov	BRKM3	anomes
321	2018-07-26	79405.34	50.60	2018/07
322	2018-07-27	79866.10	51.43	2018/07
323	2018-07-30	80275.59	51.49	2018/07
324	2018-07-31	79220.43	52.98	2018/07
325	2018-08-01	79301.65	51.60	2018/08
326	2018-08-02	79636.69	51.10	2018/08
327	2018-08-03	81434.98	53.20	2018/08
328	2018-08-06	81050.76	53.20	2018/08
329	2018-08-07	80346.52	53.99	2018/08
330	2018-08-08	79151.70	53.50	2018/08
331	2018-08-09	78767.99	52.30	2018/08
332	2018-08-13	77496.45	52.00	2018/08

```
In [5]: data.index = data.Data
```

```
In [6]: data.head()
```

Out[6]:

	Data	Bov	BRKM3	anomes
Data				
2017-02-17	2017-02-17	67748.42	31.00	2017/02
2017-02-20	2017-02-20	68532.86	31.48	2017/02
2017-02-21	2017-02-21	69052.03	31.40	2017/02
2017-02-22	2017-02-22	68589.54	31.00	2017/02
2017-02-23	2017-02-23	67461.39	31.00	2017/02

VISUALIZAR A SÉRIE

```
In [9]: #Plotar gráfico, pode ser necessário alterar os eixos.
plt.figure(figsize=(13.5, 3.5))
data['BRKM3'].plot(color='b')
plt.title('Evolução diária do preço de fechamento da ação BRKM3')
plt.xlabel('Tempo')
plt.ylabel('R$')
plt.xticks(rotation=50)
```

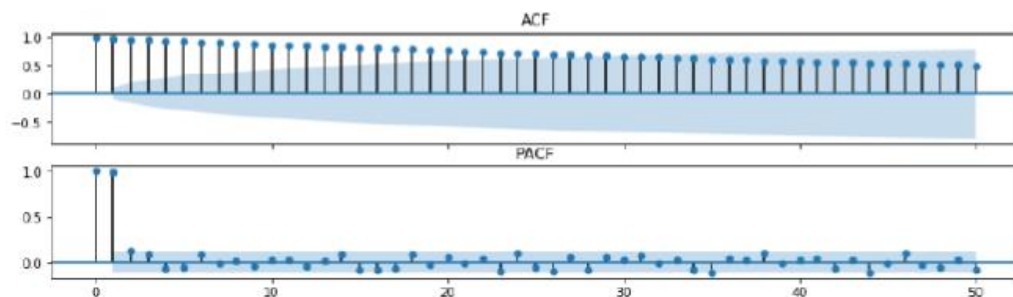
```
Out[9]: (array([736361., 736420., 736481., 736542., 736603., 736664., 736726.,
736785., 736846., 736907.]), <a list of 10 Text xticklabel objects>)
```



```
In [10]: fig, axarr = plt.subplots(2, sharex=True)
fig.set_size_inches(13.5, 3.5)

plot_pacf(data['BRKM3'], ax=axarr[1], lags=50)
axarr[1].set_title('PACF')
plot_acf(data['BRKM3'], ax=axarr[0], lags=50)
axarr[0].set_title('ACF')
```

```
Out[10]: Text(0.5,1,'ACF')
```



ARIMA

Mantenho como texto devido à demora para realizar a otimização.

```
aicVal=[] for d in range(1,2): for ari in range(0,29): for maj in range(0,10): try: arima_obj = sm.tsa.SARIMAX(data[brk'].tolist(),
order=(ari,d,maj),dynamic=True) arima_obj_fit=arima_obj.fit() aicVal.append([ari, d, maj, arima_obj_fit.aic]) except ValueError:
pass
```

46 Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

```
In [13]: ##aicVal

In [14]: data.info()

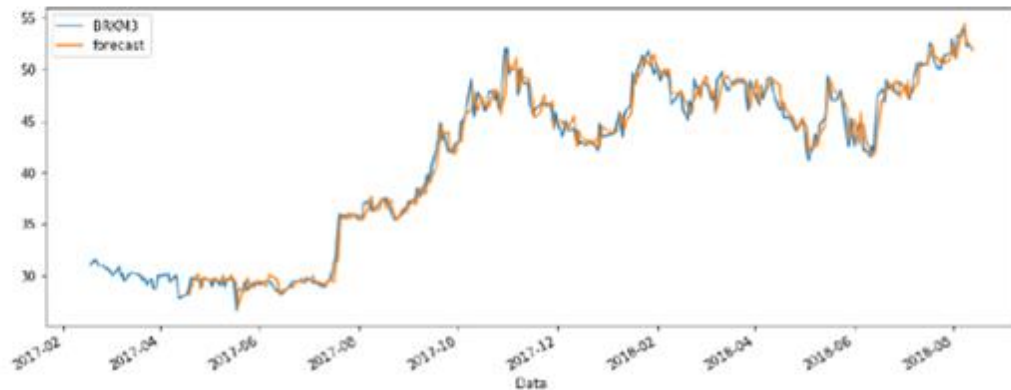
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 333 entries, 2017-02-17 to 2018-08-13
Data columns (total 4 columns):
Data      333 non-null datetime64[ns]
Bov       333 non-null float64
BRKM3     333 non-null float64
anomes    333 non-null object
dtypes: datetime64[ns](1), float64(2), object(1)
memory usage: 13.0+ KB

In [15]: model = sm.tsa.SARIMAX(data['BRKM3'], order=(2, 1, 9), trend='c')
results = model.fit()
print(results.summary())
```

Simulação um passo à frente

```
In [16]: data['forecast'] = results.predict(start = 39, end= 332)
data[['BRKM3', 'forecast']][0:400].plot(figsize=(13.5, 5))
```

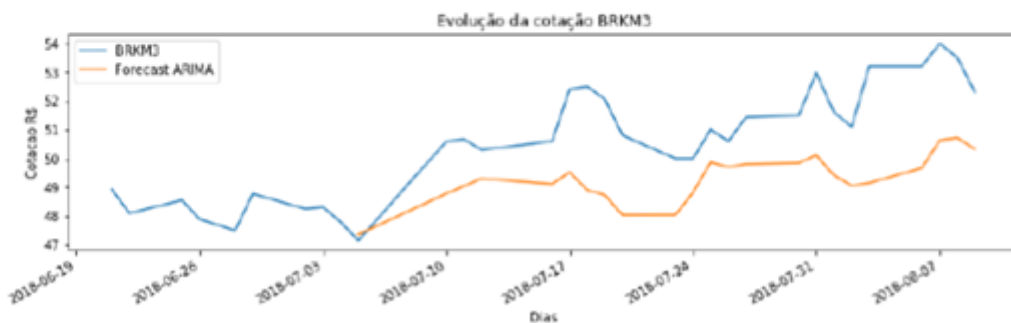
```
Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x273a135d2e8>
```



Simulação dinâmica

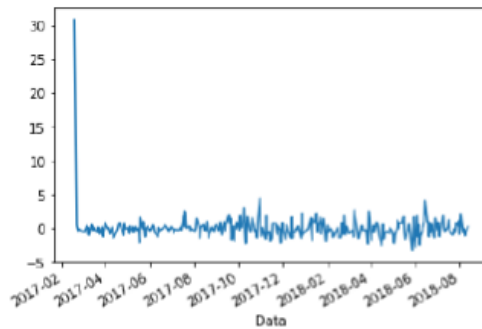
```
In [17]: data['Forecast ARIMA'] = results.predict(start = 309, end= 332, dynamic=True)
data[['BRKM3', 'Forecast ARIMA']][300:332].plot(figsize=(13.5, 3.5), title='Evolução da cotação BRKM3')
plt.xlabel('Dias')
plt.ylabel('Cotacao R$')
```

```
Out[17]: Text(0, 0.5, 'Cotacao R$')
```



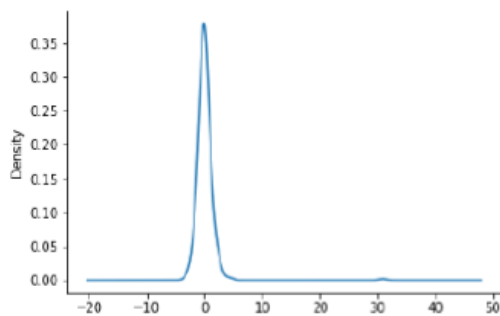
```
In [19]: results.resid.plot()
```

```
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x2739fbaf240>
```



```
In [20]: results.resid.plot(kind='kde')
```

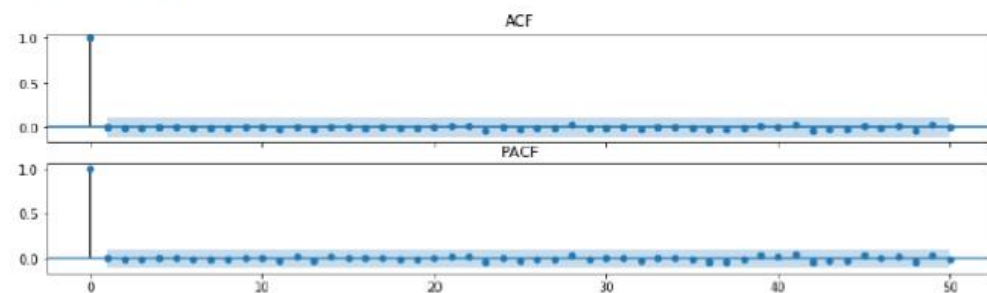
```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x273a14779b0>
```



```
In [21]: fig, axarr = plt.subplots(2, sharex=True)
fig.set_size_inches(13.5, 3.5)
```

```
plot_pacf(results.resid, ax=axarr[1],lags=50)
axarr[1].set_title('PACF')
plot_acf(results.resid, ax=axarr[0],lags=50)
axarr[0].set_title('ACF')
```

```
Out[21]: Text(0.5, 1, 'ACF')
```



```
In [22]: #Run ADF test on the irregular variations
adf_result = stattools.adfuller(results.resid, autolag='AIC')
print('p-val of the ADF test on irregular variations in employment data:', adf_result[1])
```

```
p-val of the ADF test on irregular variations in employment data: 0.0
```

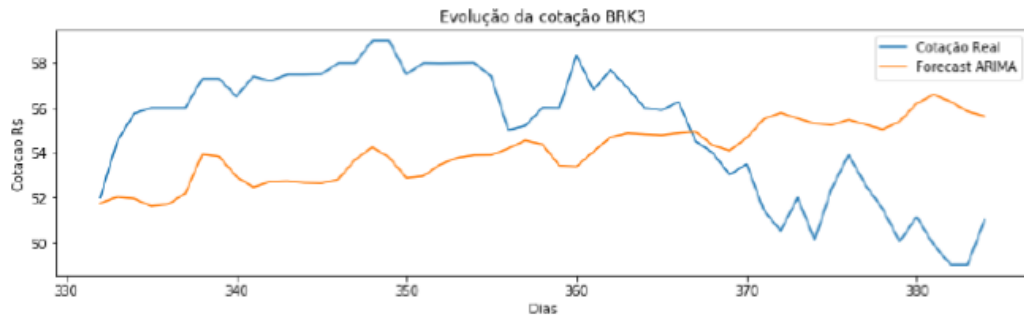
```
teste = pd.read_excel('C:/Users/Thomaz/Documents/ENQ/2018_02/TCC/Dados/brk2_teste.xlsx')
```

48 Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

```
In [29]: #Verificar qualidade do modelo
teste['Forecast ARIMA'] = results.predict(start=332, end=385, dynamic=True)
teste[['Cotação Real', 'Forecast ARIMA']][332:385].plot(figsize=(13.5,3.5), title='Evolução da cotação
o BRK3')
plt.xlabel('Dias')
plt.ylabel('Cotacao R$')
```

```
C:\Users\Thomaz\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:531: ValueWarning:
No supported index is available. Prediction results will be given with an integer index beginning
at `start`.
ValueWarning)
```

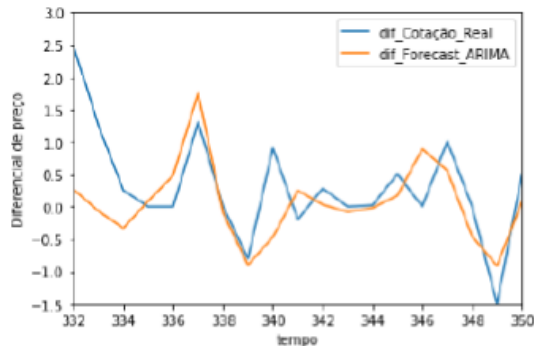
Out[29]: Text(0,0.5,'Cotacao R\$')



```
In [45]: dif_Forecast_ARIMA = np.diff(teste['Forecast ARIMA'])
dif_Cotação_Real = np.diff(teste['Cotação Real'])
```

```
In [70]: ax = plt.subplot(111)
plt.plot(dif_Cotação_Real, label='dif_Cotação_Real')
plt.plot(dif_Forecast_ARIMA, label='dif_Forecast_ARIMA')
plt.xlim(332,350)
plt.ylim(-1.5,3)
plt.ylabel('Diferencial de preço')
plt.xlabel('tempo')
ax.legend()
```

Out[70]: <matplotlib.legend.Legend at 0x273a629fdd8>



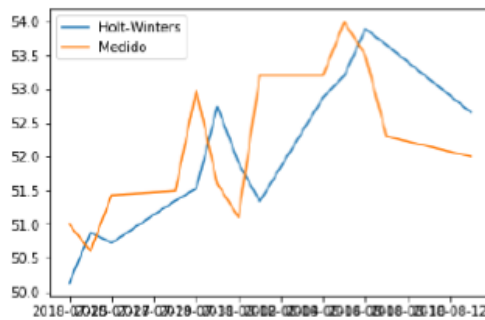
SUAVIZACAO EXPONENCIAL

```
In [103]: ## Criar modelo utilizando a base de treino
modelH = smtsa.ExponentialSmoothing(data['BRKM3'], trend='mul').fit()
pred = modelH.predict(start=320, end=332)
```

```
C:\Users\Thomaz\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:225: ValueWarning:
A date index has been provided, but it has no associated frequency information and so will be ignored when e.g. forecasting.
  ' ignored when e.g. forecasting.', ValueWarning)
```

```
In [104]: plt.plot(pred.index, pred, label='Holt-Winters')
plt.plot(pred.index, data['BRKM3'][320:333], label='Medido')
plt.legend(loc='best')
```

```
Out[104]: <matplotlib.legend.Legend at 0x1b740e1a860>
```

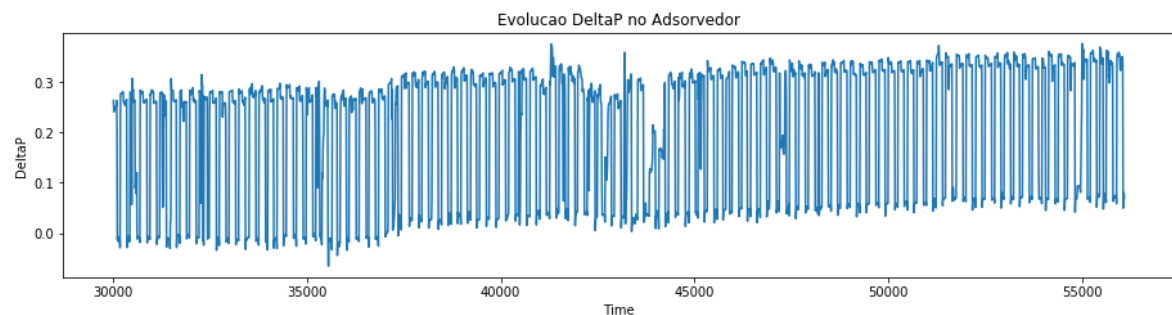


```
In [105]: ## Exibir parâmetros
modelH.params
```

```
Out[105]: {'damping_slope': nan,
'initial_level': 31.04375457944968,
'initial_seasons': array([], dtype=float64),
'initial_slope': 1.0012824161866694,
'lamba': None,
'remove_bias': False,
'smoothing_level': 0.7885396833992049,
'smoothing_seasonal': nan,
'smoothing_slope': 0.0,
'use_boxcox': False}
```

Simulação da série temporal do adsorvedor

```
plt.figure(figsize=(15.5, 3.5))
plt.plot(df2['df'])
plt.xlabel('Time')
plt.ylabel('DeltaP')
plt.title('Evolucao DeltaP no Adsorvedor')
plt.show()
```

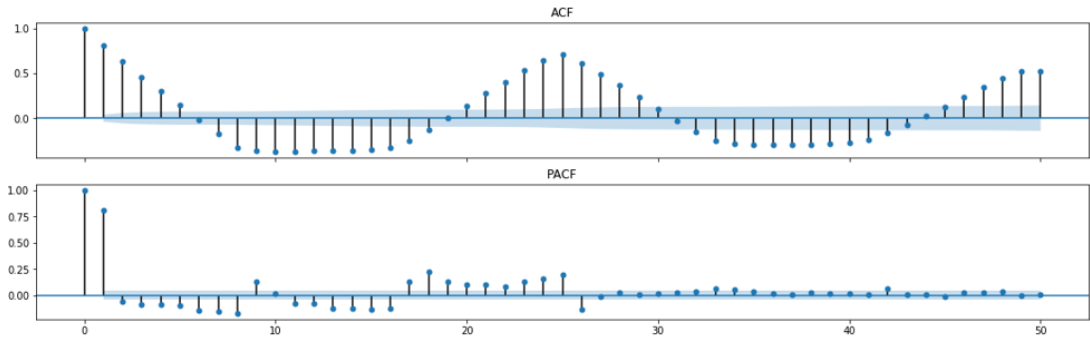


50 Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

```
In [7]: fig, axarr = plt.subplots(2, sharex=True)
fig.set_size_inches(19, 5.5)

plot_pacf(df['df'], ax=axarr[1],lags=50)
axarr[1].set_title('PACF')
plot_acf(df['df'], ax=axarr[0],lags=50)
axarr[0].set_title('ACF')
```

Out[7]: Text(0.5,1,'ACF')



MODELO SARIMA

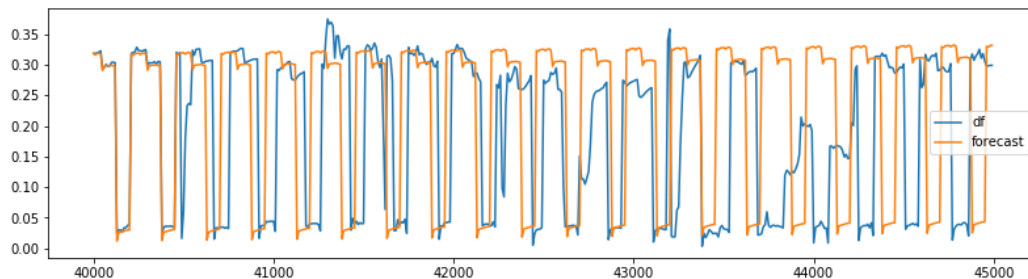
Realizar otimização do AIC

```
aicVal=[] for d in range(0,1): for ari in range(0, 13): for maj in range(0,5): for ARI in range (0,3): for D in range (0,2): for MAJ in range (0,3): try: arima_obj = sm.tsa.SARIMAX(df['df'].tolist(), order=(ari,d,maj),seasonal_order=(ARI,D,MAJ,25)) arima_obj_fit=arima_obj.fit() aicVal.append([ari, d, maj, ARI, D, MAJ, arima_obj_fit.aic]) except ValueError: pass
```

```
#: # Modelar a função
model = sm.tsa.SARIMAX(df['df'],order=(2,0,1),seasonal_order=(0,1,0,25),trend= 'c', dynamic=True)
results = model.fit()
print(results.summary())
```

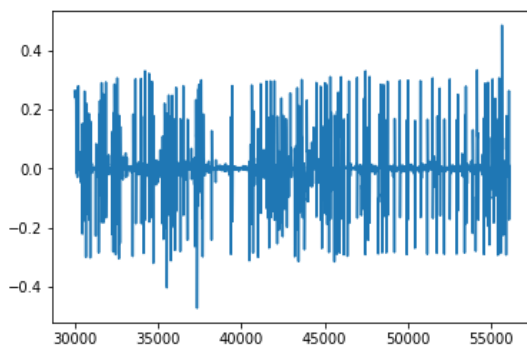
```
#Verificar qualidade do modelo
df['forecast'] = results.predict(start =-1000, end= 2610, dynamic=True)
df[['df','forecast']][1000:1500].plot(figsize=(13.5,3.5))
```

<matplotlib.axes._subplots.AxesSubplot at 0x12cdbc84978>



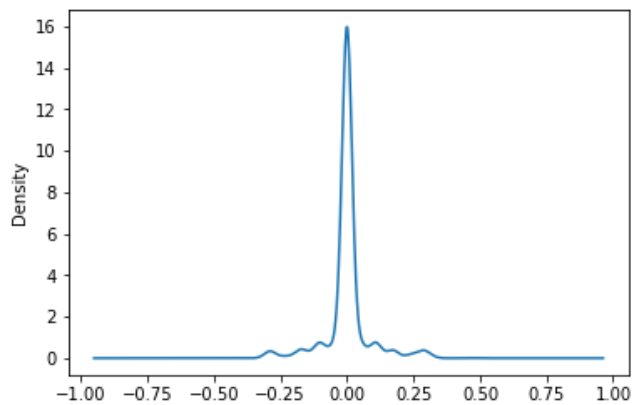
```
results.resid.plot()
```

<matplotlib.axes._subplots.AxesSubplot at 0x12cdb5d2a20>




```
results.resid.plot(kind='kde')
```

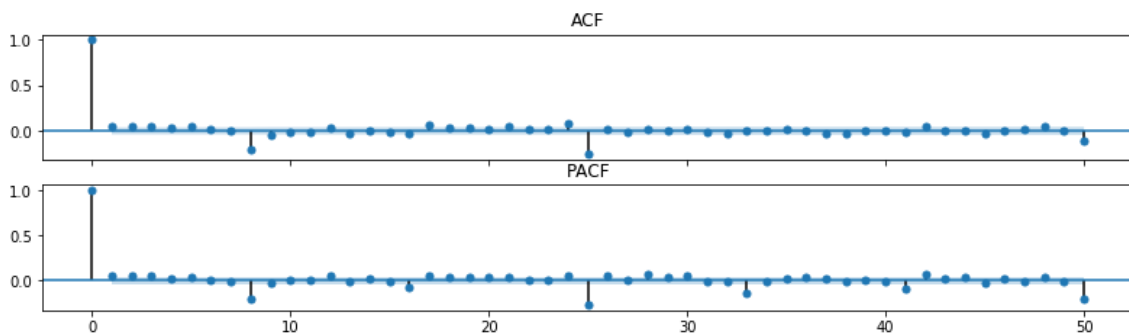
```
<matplotlib.axes._subplots.AxesSubplot at 0x12cddb5e160>
```



```
fig, axarr = plt.subplots(2, sharex=True)
fig.set_size_inches(13.5, 3.5)

plot_pacf(results.resid, ax=axarr[1],lags=50)
axarr[1].set_title('PACF')
plot_acf(results.resid, ax=axarr[0],lags=50)
axarr[0].set_title('ACF')
```

```
Text(0.5,1,'ACF')
```



```
from statsmodels.tsa import stattools
```

```
#Run ADF test on the irregular variations
adf_result = stattools.adfuller(results.resid, autolag='AIC')
print('p-val of the ADF test on irregular variations in employment data:', adf_result[1])
```

```
p-val of the ADF test on irregular variations in employment data: 3.1828347674678043e-24
```

Ler Arquivos gerados

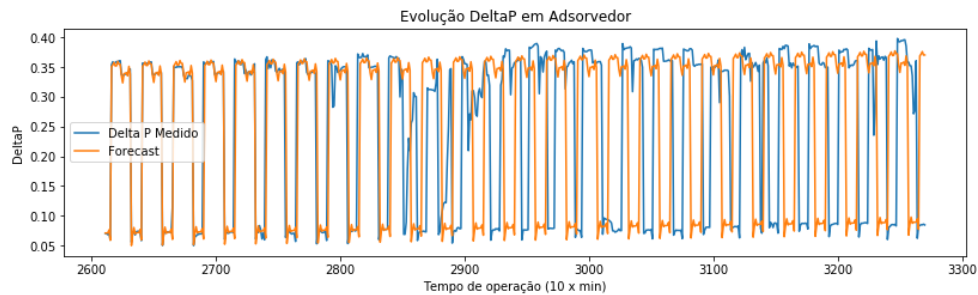
```
: teste = pd.read_excel('C:/Users/Thomaz/Documents/ENQ/2018_02/TCC/Dados/book.xlsx')
```

52 Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

```
#Verificar qualidade do modelo
teste['Forecast'] = results.predict(start =2611, end= 3271, dynamic=True)
teste[['Delta P Medido', 'Forecast']] [2611:3271].plot(figsize=(13.5,3.5),title='Evolução DeltaP em Adsorvedor')
plt.xlabel('Tempo de operação (10 x min)')
plt.ylabel('DeltaP')
```

C:\Users\Thomaz\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:531: ValueWarning: No supported index is available. Prediction results will be given with an integer index beginning at `start`.
ValueWarning)

Text(0,0.5,'DeltaP')



Aplicando o Método B

Regressão Linear - capturar tendência ¶

```
: from sklearn.linear_model import LinearRegression
```

```
: trend_model = LinearRegression(normalize=True, fit_intercept=True)
trend_model.fit(np.arange(df.shape[0]).reshape((-1,1)),df['df'])
```

```
: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=True)
```

```
: print('Trend model coefficient={} and intercept={}'.format(trend_model.coef_[0],
trend_model.intercept_
))
```

Trend model coefficient=3.2658430951974774e-05 and intercept=0.16345646911062195

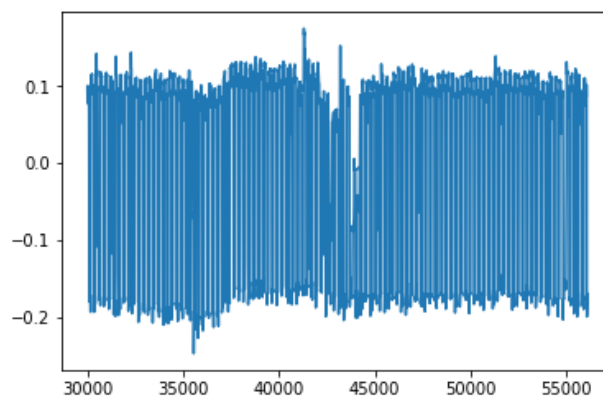
```
tm = trend_model.predict(np.arange(df.shape[0]).reshape(-1,1))
```

```
df['tm'] = tm
```

```
df['dt'] = df['df'] - df['tm']
```

```
In [319]: df['dt'].plot()
```

```
Out[319]: <matplotlib.axes._subplots.AxesSubplot at 0x1b31f11acc0>
```



```
def fourierExtrapolation(x, n_predict):
    n = x.size
    n_harm = 10
    t = np.arange(0, n)
    p = np.polyfit(t, x, 1)
    x_notrend = x
    x_freqdom = fft.fft(x_notrend)
    f = fft.fftfreq(n)
    indexes.sort(key = lambda i: np.absolute(f[i]))

    t = np.arange(0, n + n_predict)
    restored_sig = np.zeros(t.size)
    for i in indexes[:1 + n_harm * 2]:
        ampli = np.absolute(x_freqdom[i]) / n
        phase = np.angle(x_freqdom[i])
        restored_sig += ampli * np.cos(2 * np.pi * f[i] * t + phase)
    return restored_sig
```

```
x = np.array(df['dt'][195:220].values)
n_predict = 3246
extrapolation = fourierExtrapolation(x, n_predict)
fig, axarr = plt.subplots(1, sharex=True)
fig.set_size_inches(15.5, 5.5)
plt.plot(np.arange(0, 105), df['dt'][195:300], 'b', label = 'x', linewidth = 3)
plt.plot(np.arange(0, extrapolation.size), extrapolation, 'r', label = 'extrapolation')
plt.legend()
```

```
ex =pd.DataFrame(extrapolation)
```

```
teste['Sazonalidade']=ex
```

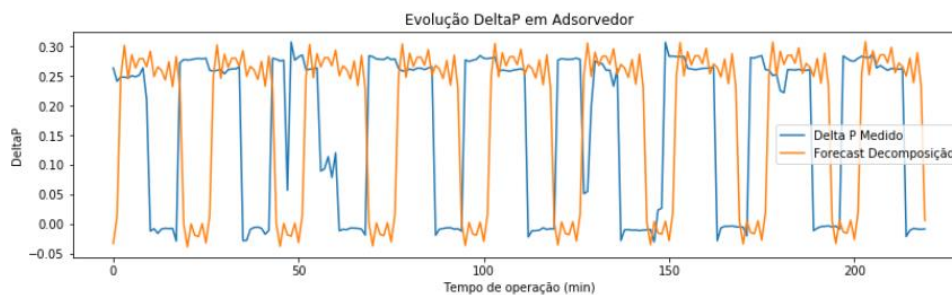
```
tm = trend_model.predict(np.arange(teste.shape[0]).reshape(-1,1))
```

```
teste['tm'] = tm
```

```
teste['Forecast Decomposição'] = teste['tm'] + teste['Sazonalidade']
```

```
: teste[['Delta P Medido', 'Forecast Decomposição']][0:220].plot(figsize=(13.5,3.5),title='Evolução DeltaP em Adsorvedor')
plt.xlabel('Tempo de operação (min)')
plt.ylabel('DeltaP')
```

```
: Text(0,0.5,'DeltaP')
```



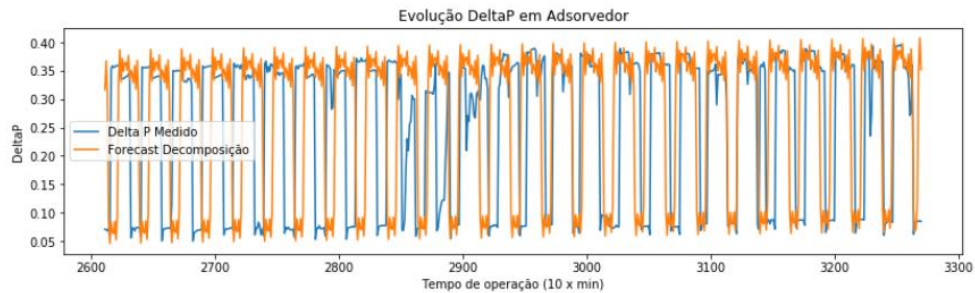
54 Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

```
teste.to_excel('C:/Users/Thomaz/Documents/ENQ/2018_02/TCC/Dados/fourier.xlsx')
```

```
fft = pd.read_excel('C:/Users/Thomaz/Documents/ENQ/2018_02/TCC/Dados/fourier_cc.xlsx')
```

```
fft[['Delta P Medido', 'Forecast Decomposição']][2611:3271].plot(figsize=(13.5,3.5),title='Evolução DeltaP em Adsorvedor')  
plt.xlabel('Tempo de operação (10 x min)')  
plt.ylabel('DeltaP')
```

```
Text(0,0.5,'DeltaP')
```



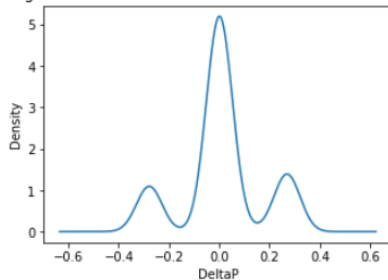
```
#Run ADF test on the irregular variations  
adf_result = stattools.adfuller(fft['Delta P Medido'][0:2610], autolag='AIC')  
print('p-val of the ADF test on irregular variations in employment data:', adf_result[1])
```

p-val of the ADF test on irregular variations in employment data: 6.215998337866845e-07

```
fft['Resíduos']][2200:2610].plot('kde', title='Histograma DeltaP dos resíduos do modelo de Decomposição',figsize=(5,3.5))  
plt.xlabel('DeltaP')
```

```
Text(0.5,0,'DeltaP')
```

Histograma DeltaP dos resíduos do modelo de Decomposição



Simulação da série temporal do CO2

```
##Importando os dados  
datac = pd.read_excel('C:/Users/Thomaz/Documents/ENQ/2018_02/TCC/Dados/Monthly_CO2_Concentrations.xlsx')
```

```
datac.index = datac.Date
```

CRIAR BASE PARA DESENVOLVER O MODELO

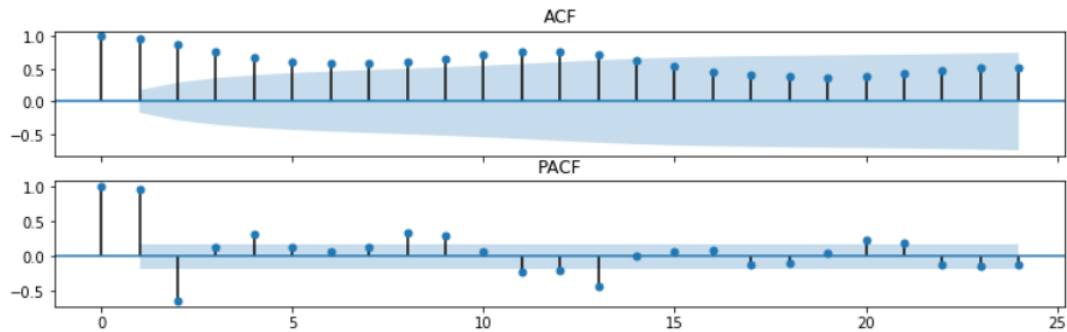
```
## Base para treinar o modelo  
data = datac[:130]
```

```
## Base para testar o modelo  
teste = datac[130:]
```

```
fig, axarr = plt.subplots(2, sharex=True)
fig.set_size_inches(12.5, 3.5)

plot_pacf(data['CO2'], ax=axarr[1],lags=24)
axarr[1].set_title('PACF')
plot_acf(data['CO2'], ax=axarr[0],lags=24)
axarr[0].set_title('ACF')
```

```
Text(0.5,1,'ACF')
```



PRIMEIRO MODELO - REGRESSÃO LINEAR ¶

```
from sklearn.linear_model import LinearRegression
```

```
trend_model = LinearRegression(normalize=True, fit_intercept=True)
trend_model.fit(np.arange(data.shape[0]).reshape((-1,1)), data['CO2'])
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=True)
```

```
print('Trend model coefficient={} and intercept={}'.format(trend_model.coef_[0],
                                                         trend_model.intercept_)
      )
```

```
Trend model coefficient=0.12158633874557882 and intercept=329.4191426893717
```

```
residuals = np.array(data['CO2']) - trend_model.predict(np.arange(data.shape[0]).reshape((-1,1)))
```

```
tm = trend_model.predict(np.arange(data.shape[0]).reshape(-1,1))
```

```
tm = trend_model.predict(np.arange(data.shape[0]).reshape(-1,1))
```

```
data['tm'] = tm
```

```
residuals = data['CO2'] - tm
data['Residuals'] = residuals
monthwise_avg = data.groupby(by=['Month'])['Residuals'].aggregate(['mean'])
```

```
data['media'] = data.Month.map(monthwise_avg['mean'])
```

```
irr_var = data['CO2'] - data['tm'] - data['media']
```

```
from statsmodels.tsa import stattools
```

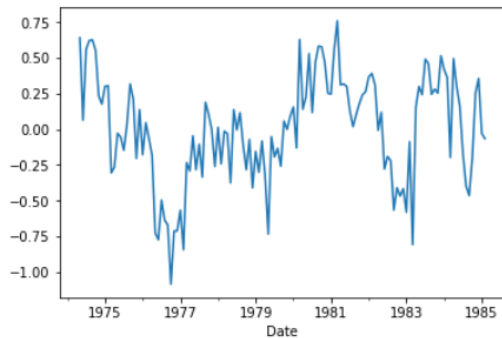
```
data['F_RL'] = data['media'] + data['tm']
```

```
data['iv'] = irr_var
```

56 Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

```
data['iv'].plot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x238182310f0>
```



```
adf_result = stattools.adfuller(data['iv'], autolag='AIC')
```

```
print('p-val of the ADF test:', adf_result[1]) #Unidade do indicador
```

```
p-val of the ADF test: 0.023578883820030202
```

MODELO SARIMA

```
aicVal=[]
for d in range(1,2):
    for ari in range(0, 3):
        for maj in range(0,3):
            for ARI in range (0,3):
                for D in range (0,2):
                    for MAJ in range (0,3):
                        try:
                            arima_obj = sm.tsa.SARIMAX(data['CO2'].tolist(), order=(ari,d,maj),seasonal_order=(ARI,D,MAJ,12))
                            arima_obj_fit=arima_obj.fit()
                            aicVal.append([ari, d, maj, ARI, D, MAJ, arima_obj_fit.aic])
                        except ValueError:
                            pass
```

```
model = sm.tsa.SARIMAX(data['CO2'],order=(2,1,0),seasonal_order=(0,1,1,12))
results = model.fit()
print(results.summary())
```

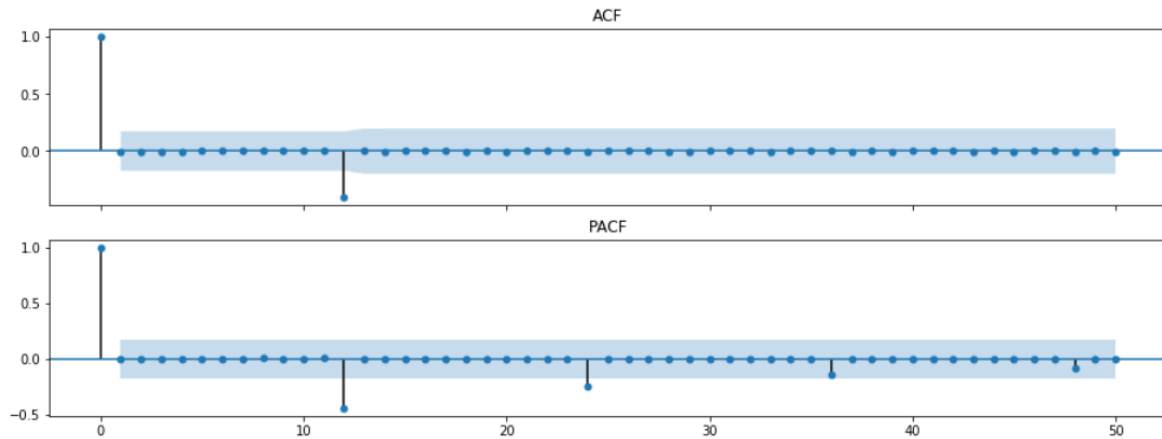
```
data['forecast'] = results.predict(start = 50, end= 134)
data[['CO2', 'forecast']].plot(figsize=(13.5,3.5))
```

```
results.resid.plot(kind='kde', xlim=(-50,50))
```

```
fig, axarr = plt.subplots(2, sharex=True)
fig.set_size_inches(15.5, 5.5)

plot_pacf(results.resid, ax=axarr[1],lags=50)
axarr[1].set_title('PACF')
plot_acf(results.resid, ax=axarr[0],lags=50)
axarr[0].set_title('ACF')
```

```
Text(0.5,1,'ACF')
```



```
#Run ADF test on the irregular variations
adf_result = stattools.adfuller(results.resid, autolag='AIC')
print('p-val of the ADF test on irregular variations in employment data:', adf_result[1])
```

p-val of the ADF test on irregular variations in employment data: 0.0

COMPARAR SARIMA vs BASE ORIGINAL

```
from pandas.tseries.offsets import DateOffset
```

```
future_dates = [data.index[-1] + DateOffset(months=x) for x in range(0,620)]
```

```
future_dates_df = pd.DataFrame(index=future_dates[1:],columns=data.columns)
```

```
future_df = pd.concat([data,future_dates_df])
```

```
X = np.array(data['CO2'])
```

```
future_df['CO2Medido'] = X
```

```
F = future_df['CO2'].dropna()
```

```
future_df['forecast'] = results.predict(start = 130, end = 900, dynamic= True)
```

HOLT WINTERS MODEL

```
## Criar modelo utilizando a base de treino
modelH = smtsa.ExponentialSmoothing(data['CO2'], seasonal='add', trend='add', seasonal_periods=12).fit()
pred = modelH.predict(start=data['CO2'].index[0], end=data['CO2'].index[-1])
plt.plot(pred.index, pred, label='Holt-Winters')
plt.plot(pred.index, data['CO2'], label='Medido')
plt.legend(loc='best')
```

```
## Exibir parâmetros
modelH.params
```

```
future_df['HW_Forecast'] = modelH.predict(start=130, end=900)
```

58 Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso

```
# PACF e ACF dos resíduos
fig, axarr = plt.subplots(2, sharex=True)
fig.set_size_inches(15.5, 5.5)

plot_pacf(modelH.resid, ax=axarr[1],lags=50)
axarr[1].set_title('PACF')
plot_acf(modelH.resid, ax=axarr[0],lags=50)
axarr[0].set_title('ACF')
```