Concentração de Ozônio na região de Bangu, Rio de Janeiro

Uma abordagem em séries temporais

Lucas Machado Moschen[1]*

Resumo
A poluição do ar é um dos temas mais imporantes da contemporaneidade e, este trabalho, pretende recortar esse problema na região de Bangu, no Rio de Janeiro. A abordagem foi consttruída com o ferramental de séries temporais, utilizando conceitos como regressão linear, autocorrelação, estacionariedade e tempo, obviamente. Os resultados não foram tão bons quanto os esperados, mas oo processo para alcance foi bem estruturado, com testes de hipóteses e Cross-Validation. A maior parte do teoria do trabalho foi contruída com o auxílio do livro Forecasting, devidamente refrenciado. Gráficos foram gerados para elucidar a questão e trazer esse problema tão próximo à nossam mesa de discussão.

Palavras-chave
Poluição do ar — Ozônio — Predição

[1] *Fundação Getulio Vargas - Escola de Matemática Aplicada*
*__Corresponding author__: lucas.machadomoschen@gmail.com

Conteúdo

Introdução

A cidade do Rio de Janeiro, localizada no estado do Rio de Janeiro, no Brasil, é uma das 50 cidades mais lindas do mundo, segundo sites de viagem como Conde Nast Traveler [1], além de ser senso comum entre as brasileiras e os brasileiros. Isso traz muita visibilidade para a capital e com isso, os turistas. Apesar disso, existem regiões menos visitadas pela população em geral, dentre essas, está Bangu, um bairro localizado na zona oeste do Rio. É um dos distritos mais populosos do Rio de Janeiro, com mais de 200 mil habitantes, segundo o Armazenzinho do Rio de Janeiro [2]. A região de Bangu é uma das mais quentes da cidade, com as maiores temperaturas máximas da cidade, em média, (nas estações registradas), segundo o Instituto Nacional de Metereologia (INMET) [3].

Nessa região, o quimíco ozônio apresenta concentração muito alta e é considerada, algumas vezes, inadequada pelos índices de qualidade do ar. Na maioria dos anos, o índice foi maior do que todas as outras estações, segundo os dados de monitoramento da qualidade do ar do Data.Rio, da prefeitura da cidade [4], como apresentado na Figura 1. Exploraremos esses dados ao longo do texto. O motivo de ter capturado as máximas, ao invés das mínimas ou médias é a importância que elas no caso de dados faltantes, já que quando se faz a média, ao não ter a medição em determinado período com

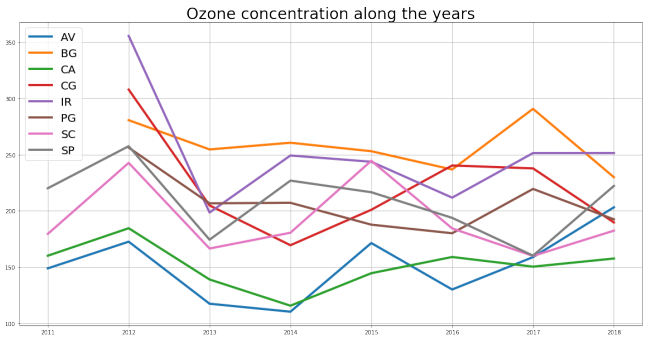

Figura 1. Um gráfico com as máximas do gás ozônio entre os anos de 2012 a 2018 nas 8 estações registradas. Bangu é representada pela linha laranja (BG).

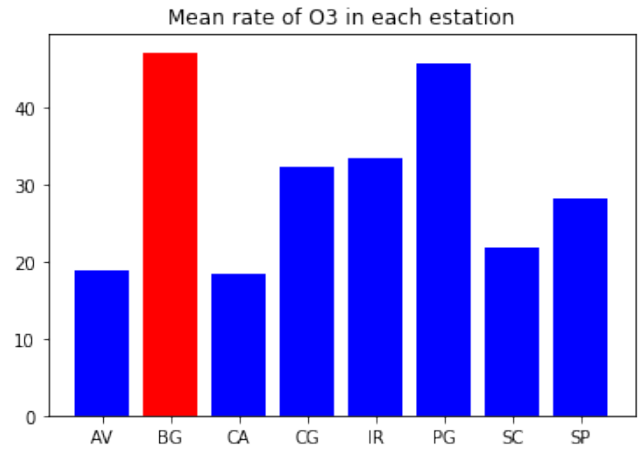

Figura 2. Gráfico com a média de gás ozônio no período de 2012 a 2018. Bangu está sinalizado em vermelho e apresenta a maior das médias. Os rótulos são: {AV: Copacabana, BG: Bangu, CA: Centro, CG: Campo Grande, IR: Irajá, PG: Pedra da Guaratiba, SC: São Cristóvão, SP: Tijuca}

relativas baixas da quantidade, a média torna-se mais alta do que realmente é, e vice-versa.

Mas também pensando nisso, utilizando o mesmo banco de dados, é possível ver que a quantidade média de gás ozônio em todo o período de estudo (2012 - 2018), é maior na região de Bangu, como é possível ver na Figura 2. Outra região que também assusta com sua média é Pedra de Guaratiba, porém, vamos nos atentar ao bairro de Bangu.

A partir dessas visualizações, surge grande interesse no entendimento do ozônio nessa região, para posterior análise do impacto causado.

Neste trabalho, dividirei minhas atenções entre o estudo de séries temporais, em particular sobre a concentração de ozônio ao longo dos anos de 2012 a 2018 e análise de alguns métodos na literatura de previsão. Esses estudos pretendem seguir o livro Forecasting, de Rob Hyndman and George Athanasopoulos [5]. No outro sentido, este livro pretende realizar uma crítica à qualidade do ar naquela região e introduzir possíveis análises das consequências, lançando mão de inferência causal para potenciais futuros trabalhos.

Diferente da análise de amostras aleatórias de observações, discutidas em outros contextos estatísticos, a série temporal é baseada em valores sucessivos que representam medidas tomadas em intervalos de espaço igualmente espaçados. Isto é, trata-se de uma observação ao longo do tempo. A concentração de ozônio vai de encontro com essa definição, pois sua medida é captada a cada hora. A partir desses dados, podemos inferir algumas análises.

O banco de dados é conduzido pelo MonitorAr-Rio, que conta com oito estações espalhadas pela cidade e uma unidade móvel. Esse órgão emite um Boletim Diário, onde descreve condições metereológicas das últimas 24 horas e da qualidade do ar. Também propõe uma tendência para o próximo dia. Mais informações sobre esse boletim podem ser encontradas em [6]. Para esse trabalho, coletei os dados sobre informações horárias de diversos índices, dentre eles Temperatura, Velocidade do Vento e Concentração de vários poluentes. São essas informações que me interessam. Na imagem 3 é mostrado um boletim do dia 6 de novembro de 2019. Nesse dia, Bangu registrou um nível de ozônio tão alto que a qualidade do ar foi considerada como má. São com essas premissas que inicio meu trabalho.

1. Metodologia

Este trabalho busca um referencial bibliográfico a fim de compreender o conhecimento adquirido na área de séries temporais, através de livros e artigos devidamente citados.

As análises numéricas foram realizadas na linguagem Python através da plataforma do Jupyter Notebook, devido à fácil visualização dos gráficos. Esses gráficos que foram gerados pelas Bibliotecas *Pandas* e *Matplotlib.pyplot*. Veja Apêndice D.

O banco de dados é disponibilizado pelo Data.Rio em formato *CSV* e apresenta os dados numéricos. Entretanto, ele apresenta muitos dados faltantes, devido a manutenções realizadas nas estações. O tratamento para a maioria dos cálculos aqui apresentados desconsidera-os, porém, para futuros trabalhos, pretende-se estimar esses valores.

2. Poluição do Ar

Poluição do ar é uma mistura de partículas e gases, muitas vezes não visíveis aos olhos humanos. As formas visíveis são amplamente conhecidas, como fumaça, fuligem e mofo. Um índice utilizado para medir os níves de poluição do ar é o índice de Qualidade do Ar (IQA), que é disponibilizado pelo boletim diário, como visto na imagem 3.

2.1 Gases Poluentes

A atmosfera da Terra consiste em um sistema dinâmico e complexo de gases naturais que são necessários para a vida, de acordo com [7]. O planeta apresenta um mecanismo de defesa que absorve parte desses gases, fazendo com que um ciclo seja formado. Entretanto, altos níveis de concentração

MonitorAR Rio
Programa de Monitoramento da Qualidade do Ar

PREFEITURA DA CIDADE DO RIO DE JANEIRO
SECRETARIA MUNICIPAL DE MEIO AMBIENTE
BOLETIM DE QUALIDADE DO AR

06/11/2019 16:00H
Quarta-feira

Exibir Boletim Anterior:
Data: Exibir

Estação	Concentração Máxima Poluentes Monitorados					Índice de Qualidade do Ar (IQA)	Classificação
	Dióxido de Enxofre (SO_2) [µg/m³](3)	Monóxido de Carbono (CO) [ppm](2)	Material Particulado (MP_{10}) [µg/m³](3)	Ozônio (O_3) [µg/m³](2)	Dióxido de Nitrogênio (NO_2) [µg/m³](1)		
Centro	NM	0,4	24,4	177,7	NM	144	Inadequada
Copacabana	1,4	0,1	72,0	78,4	NM	61	Regular
São Cristóvão	6,7	0,1	25,5	146,8	NM	92	Regular
Tijuca	3,1	0,4	32,8	130,4	61,3	81	Regular
Irajá	4,4	0,5	33,2	144,9	49,3	91	Regular
Bangu	4,2	0,5	52,3	208,7	ND	201	Má
Campo Grande	3,5	0,3	25,5	109,3	31,0	68	Regular
Pedra de Guaratiba	NM	NM	25,2	80,8	NM	50	Boa
Unidade Móvel Fiocruz	Temporariamente desativada para reposicionamento						

Condições Meteorológicas observadas no período:

A passagem de uma de uma frente fria pelo oceano deixou o tempo instável na cidade do Rio de Janeiro nesta quarta-feira. O dia teve predomínio de céu nublado e houve registro de chuva fraca em alguns pontos do Município, resultando no suave declínio das concentrações dos poluentes. Assim, a qualidade do ar ficou classificada como REGULAR na maior parte dos locais monitorados; BOA na estação Pedra de Guaratiba; INADEQUADA na estação Centro e, na estação Bangu, MÁ.

Tendência da Qualidade do Ar para as Próximas 24h:

O tempo seguirá instável devido ao posterior transporte de umidade após a passagem da frente fria. O céu irá variar entre nublado a parcialmente nublado, com previsão de chuva fraca a moderada ao longo do dia. Assim, espera-se o ligeiro declínio das concentrações dos poluentes, o que deverá deixar a qualidade do ar nas próximas 24h classificada entre BOA a INADEQUADA nas localidades monitoradas.

Figura 3. Boletim registrado no dia 06/11/2019

de gases podem causar diversos efeitos nos organismos vivos, variando de gás para gás.

Dentre as principais fontes dos poluentes, estão a queima de combustíveis fósseis, queima de mata e áreas de agricultura. Os gases poluentes incluem:

- **Óxidos de Carbono:** O monóxido de carbono (CO) é oriundo da combustão incompleta e não apresenta cheiro ou cor. Já o dióxido de carbono é um gás que contribui para o efeito estufa e, em excesso na atmosfera, devido à queima de combustíveis fósseis, pode causas sérios danos.
- **Óxidos de Nitrogênio:** Também emitidos por veículos e tem uma aparência marrom. O dióxido de nitrogênio é um dos gases mais perigososos para a poluição do ar, e sua toxidade é facilmente identificável.
- **Óxidos de Enxofre:** Causa primária da chuva ácida, muito comum na Europa. É natural após erupções vulcânicas. É uma forte causa de problemas respiratórios.

Na próxima subsessão, trataremos do gás ozônio, o ator principal da nossa análise, segundo [8].

2.2 O gás Ozônio

O gás ozônio (O_3) contém três átomos de oxigênio. Até pequenas concentrações desse gás são consideradas tóxicas e explosivas. Ele ocorre naturalmente na atmosfera, porém em pequenas quantidades, quando absorve radiação ultravioleta. Em condições especiais, óxidos de nitrogênio e hidrocarbonos podem produzir ozônio em concontração alta o suficiente para causar irritação nos olhos e na mucosa.

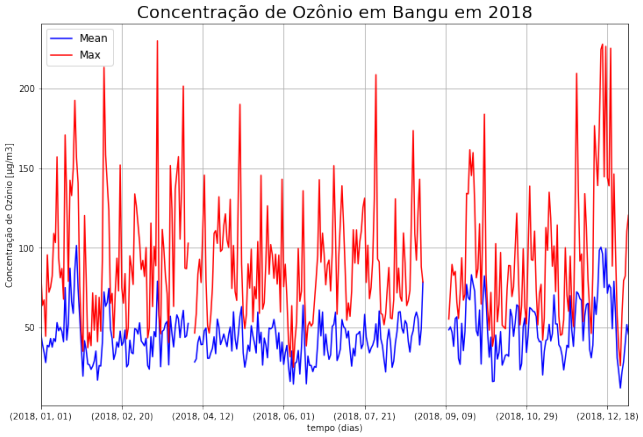

Figura 4. Concentração de Ozônio em 2018.

3. Análise da Série Temporal

Uma série temporal é uma coleção de observações x_t registrada no tempo t. Esse tempo pode ser discreto ou contínuo. Na nossa análise, o tempo é contínuo, entretanto, interpretamos ele como discreto, já que as medições são guardadas a cada hora. O objetivo de analisar a série temporal é tentar compactar a informação disponibilizada pela prefeitura para interpretação a posteriori, estudar a relação com outras variáveis medidas e prever futuros valores usando algum modelo. Nesse caso, mostraremos mais de um. Observe a série temporal de Ozônio no ano de 2018 na Figura 4. Grande parte desse texto também é contido no livro Econometria de Séries Temporais [9].

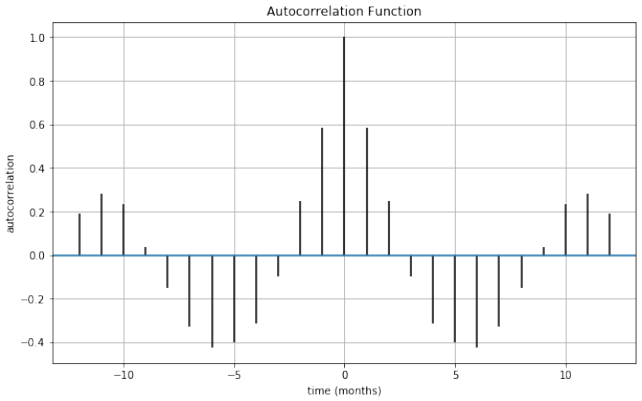

Figura 5. Função de autocorrelação em relação aos meses. As linhas azuis representam os limites.

3.1 Autocorrelação

Precisamos encontrar padrões na série temporal, a fim de encontrar possíveis tendências, não necessariamente lineares, sazonalidades ou mudanças cíclicas. Para isso, utilizo uma medida de relação linear entre valores com atraso da série. Definimos r_k como essa medida entre os valores y_t e y_{t-k}, para todos os valores de y_t capturados. Esse método é derivado da correlação de Pearson.

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

A partir disso, é possível gerar a função de autocorrelação. Neste trabalho, explorei essa função em relação aos meses e às horas. Em relação aos meses, meu k varia entre os valores de 0 a 12. De fato $r_0 = 1$. Espera-se que a autocorrelação seja zero se o tempo não tenha influência sobre os dados. Nesse caso, chamamos a série de ruído branco. Claro que como o conjunto de dados tem tamanho finito, a autocorrelação dificilmente será 0. Desta maneira, para esses casos, é esperado que os valores estejam entre os limites $\pm\frac{2}{\sqrt{T}}$ com probabilidade 95%, onde T é o tamanho da amostra. A figura 5 representa a função. Note o quão insignificantes se tornam os limites.

Valores altos para pequenos atrasos indicam tendência na série, já que existe uma correlação alta entre os valores de um mês com os valores do mês anterior. Entretanto, como o decréscimo não é suave, a tendência não é tão observada. Também é interessante observar o efeito da sazonalidade em períodos de 6 meses. Isso pode estar relacionado a período mais quente e frio, e a alternância, nesses casos, da concentração de ozônio.

No caso das horas, também foi observada alta correlação para valores de atraso pequenos, mas o resultado foi pouco elucidativo.

3.2 Estacionariedade

Uma série é dita estacionária quando ao passar do tempo, seus valores mantem-se ao redor da média e variância constante. Quando uma série apresenta tendência, ela não é estacionária. A importância de uma série estacionária é para a realização do modelo, já que vários deles são descritos sobre séries estacionárias. Para testar se nossa série temporal é estacionária, consideremos o teste Dickey-Fuller Aumentado, um tipo de teste de uma raíz, uma causa para a não estacionariedade.

$$H_0 : \textit{there is unit root } (\textit{non stationary})$$

$$H_1 : \textit{there is no unit root } (\textit{stationary})$$

Considero o nível de significância de 5%. O p-valor do teste esteve na ordem de 10^{-30}, e a hipótese nula foi rejeitada. O teste foi realizado em Python, através da função *adfuller* do módulo para análide de séries temporais *statsmodels.tsa*. Para esse teste em Python, conferir [10]. Outro fator importante de se analisar é a ergodicidade, porém, nesse ensaio, assumirei a série como engórdica.

4. Modelando Séries Temporais

Para modelar uma série temporal, a fim de fazer futuras predições, existem diversos métodos apresentados pela literatura. Os dados consideram as máxima média a cada 8h, que são consideradas para o cálculo do IQA no banco de dados. Para os valores faltantes, preenchi com o valor anterior e deixo para futuros trabalhos estudar métodos mais eficazes. Os dados não são afetados de forma significativa por inflação, mudança na população ou calendário, pois o intervalo temporal em anos não é muito grande.

Para realizar os seguintes métodos, desenvolvo um processo para a análise do resultado com o diagnóstico dos dos resíduos através do teste para autocorrelação de *Portmanteau*. Nesse teste, testamos se as primeiras h autocorrelações são significantemente diferentes da esperada em ruído branco. Existem várias estatísticas para esse teste, como, por exemplo, o teste *Box-Pierce* e, mais preciso, o teste *Ljung-Box*, conforme apêndice A.

Para avaliar a precisão de nossas previsões, utilizo o método de Cross-Validation para dividir os dados, mais de uma vez, em treino e teste. Ao fim, para cada divisão temporal, calculo a precisão do modelo através do método da Média Absoluta de Erros de Escala (Veja apêndice B) e, depois, faço a média desse cálculo de previsão.

Métodos Estudados nesse trabalho:

1. Método da Média;
2. Regressão Linear;
3. Suavização Exponencial;

4.1 Método da Média

Nesse método de previsão, o modelo simplesmente afirma que a próxima observação é a média das observações anteriores. Ele tem sua importância para teste de sanidade e verificação dos métodos de análise de resultado.

4.2 Regressão Linear

A regressão linear admite que exista uma relação linear entre duas ou mais séries temporais. Assim,

$$y_t = \beta_0 + \beta_1 x_{1,t} + ... + \beta_n x_{n,t} + \varepsilon_t$$

A variável ε captura tudo que as variáveis escolhidas não capturam. Observe que os parâmetros indicam o quanto as variáveis são relacionadas e como elas se relacionam (positivamente ou negativamente). Assumimos algumas coisas sobre os erros:

- A média 0;
- São não autocorrelacionados. Caso não fossem, existiria informação adicional que não foi extraída dos dados;
- São não relacionados com as variáveis preditoras, que representam as variáveis $x_{i,t}$.
- É interessante, mas não necessário, que os erros sejam normalmente distribuídos.

O método que utilizo para escolher os parâmetros é a estimação de mínimos quadrados. Também, para esse modelo, lanço mão do coefieciente de determinação, o R^2 (Veja Apêncice C).

As variáveis que utilizarei para esse modelo são Temperatura, Radiação Solar, Velocidade do Vento e Umidade Relativa. Essas quatro variáveis apresentam os maiores valores absolutos de correlação e tem relação direta com a formação do ozônio. A radiação solar, é um exemplo já construído na literatura e tem relação direta com o ozônio [11].

Além disso, eu crio 11 variáveis indicadoras para os meses, a fim de capturar a sazonalidade ou tendências em certos períodos. É importante dizer que não é necessária mais uma variável, pois ela estará incluída no parâmetro de interceptação, que não acompanha variáveis independentes. Ao colocar essa variável, podemos criar uma *variável indicadora armadilha*. Essas variáveis indicam 1 para o mês correspondente e 0 caso contrário.

4.3 Suavização Exponencial

Proposta no final dos anos de 1950, por Brown, Holt e Winters, a suavização exponenicial tem motivado diversos sucessos em previsões. Basicamente, esse método faz uma média ponderada das observações passadas, só que esses pesos decaem exponencialmente com o tempo. Especificamente, os pesos decrescem com uma razão geométrica. Nesse método, existem diversas variações. Chamamos de Suavização Exponencial Simples a seguinte equação:

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha) l_0,$$

onde, os parâmetros α e l_0 devem ser estimados. l_0 a primeira estimação, enquanto α é um valor entre 0 e 1 que indica o quando de importância se dá aos eventos passados.

Entretanto, nesse trabalho, desenvolvo um modelo para capturar a sazonalidade dos dados, onde existe a equação de previsão e três equações de suavização: o componente de tendência, o componente sazonal e o nível da suavização. Nesse caso, existem mais dois parâmetros a serem estimados, além de um parâmetro de sazonalidade que indica o número de períodos do ano. Esse modelo foi desenvolvido por Holt e Winters.

A fim de encontrar os parâmetros, também utilizo o método de minimização de quadrados dos erros.

5. Resultados e Discussões

Nessa sessão apresento os principais resultados referentes aos métodos que apliquei aos dados de previsão.

5.1 Analisando a Média

Para esse método, o teste de Ljung-Box indicou que existe correlação temporal, i.e, autocorrelação. O valor do p-valor ficou extremamente próximo de 0, o que indica que a hipótese nula é rejeitada a um nível 5%. Além disso, o valor do nosso erro MASE ficou maior que 1, sendo equivalente a 1.16, o que significa que ele não melhora a estimativa da média de todos os dados e estimar baseado nessa média.

De fato, esse não é um bom preditor para esses dados, visto que eles carregam muita variação. Entretanto, ele estabelece um limite mínimo de qualidade para os próximos métodos para esse conjunto de dados. Confira a Figura 6.

5.2 Analisando a Regressão Linear

O método de Regressão Linear também não foi capaz de capturar a variabilidade dos nossos dados. Nesse caso, fiz duas predições, uma que incluísse as variáveis indicadoras dos meses, e outra que não, apenas as variáveis já citadas anteriormente. Nos dois casos, os resultados não foram tão bons. Nos dois casos, o teste de Ljung-Box evidenciou que os resíduos possuiam autocorrelação, o que significa que o modelo não foi capaz de capturar toda a informação temporal contida nos dados. Acredito que isso se deva ao fato da grande variabilidade já citada. Alguma transformação necessitaria ter sido feita nesses dados a fim de resolver esse problema. Entretanto, deixo esse fato para futuros trabalhos.

Sobre os resultados, quando considerei os meses como variáveis do meu problema, o erro calculado foi de 1.01. Esse resultado já foi melhor do que a média, mas ainda não capturou dados suficientes sobre os dados. Já o coeficiente de determinação foi de 0.5, relativamente baixo, o que evidencia que essas variáveis não foram boas para estimar minha série temporal. Acredito que isso se deva à falta de algum parâmetro e os valores faltantes em cada variável, que somados podem ter influenciado negativamente os dados.

Quando retirei as variáveis sobre os meses, obtive um resultado do erro de 0.96, meu primeiro valor menor do que um nesse sentido. Entretando, o coeficiente de determinação foi de 0.42, bem inferior ao anterior. Isso é explicado pelo seguinte fato. Adicionar variáveis não reduz o coeficiente de

determinação, pois se reduzisse, bastaria colocar o parâmetro correspondente como 0. Nesse sentido, não acredito que o resultado anterior foi melhor. Confira o resultado desse teste, sem as variáveis indicadoras, na Figura 7.

5.3 Analisando a Suavização Exponencial

Esse método também não foi feliz para modelar o problema. Nesse caso, o problema é possivelmente na quantidade de informações, que dificulta a estimação. O método teve o pior desempenho no cálculo do erro, apesar de ter reduzido os valores das estatísticas de Ljung-Box, apesar de que ainda o teste considerou os resíduos autocorrelacionados. Na imagem 8, coloquei todos os valores estimados pelo modelo, e os valores estimados para o conjunto de teste. Os resíduos apresentaram o melhor resultado de encaixe dos dados. Uma possível maneira de corrigir esse problema seria considerar os valores de concentração por mês.

O valor do erro foi de 1.28.

5.4 Discussões sobre os Resultados

Os resultados não foram bons como esperado antes do problema ser enunciado, entretanto, acredito que o processo tenha sido sólido o suficiente para a extensão para melhorar as simulações numéricas e a teoria sobre outros métodos seja enunciada de maneira natural.

Os métodos apresentados por séries temporais, aqui neste trabalho, se mostraram não suficientes para explicar o comportamento da concentração de ozônio, apenas com esse banco de dados. Acredito que outros métodos devam ser aplicados em ordem de obter resultados melhores.

Entretanto, um resultado ficou claro. Podemos observar, principalmente no método de Regressão Linear, que muitos dos dias apresentam quantidade de ozônio maior do que 80. Esse valor chave é definido pelo Boletim de Qualidade do Ar, como máximo de ozônio para que o ar seja considerado bom. Entretanto, nenhum valor superou o valor de 160, atingindo a qualidade de ar inadequada. Dessa forma, as políticas esperadas de atuação não precisam ser intensas, apenas controles bem localizados.

5.5 Futuros Trabalhos

1. Melhorar a forma de inserir dados novos no lugar de dados faltantes. Para isso, propor métodos de regressão linear ou inferência causal;

2. Melhorar a regressão linear, fazendo diferentes testes com variáveis fora do banco de dados, como, por exemplo, produção em indústrias locais;

3. Aplicar transformações na série temporal de estudo, afim de obter melhores resultados;

4. Aplicar testes da importância dos *outliers* nos modelos e retirá-los, quando evidente.

5. Fazer os mesmos estudos, só que no sentido mais amplo, mensal.

6. Apêndice

A Estatística do Teste Ljung-Box

Considere a estatística:

$$Q^* = T(T+2)\sum_{k=1}^{h}(T-k)^{-1}r_k^2,$$

onde, T é o número de observações, h é o maior atraso de autocorrelação. Esse teste sugere valores de h não superiores a $T/5$. Observe que se os valores de r_k forem próximos a 0, o valor de Q^* será pequeno. Se Q^* veio de um ruído branco, Q^* espera-se que tenha uma distribuição χ^2 com $h-K$ graus de liberdade, onde K é o número de parâmetros no modelo.

B Mean Absolute Scaled Error

Erros de escala foram propostos por Hyndman e Koehler, em 2006, para ser uma alternativa aos métodos antereiores, como os erros percentuais. Uma forma de definir foi, a partir de métodos de previsão simples:

$$q_j = \frac{e_j}{\frac{1}{T-m}\sum_{t=m+1}^{T}|y_t - y_{t-m}|}$$

No nosso caso, foi utilizado $m = 1$. Note que q_j é independente da escala dos dados.

A partir disso, podemos definir $MASE = media(|q_j|)$.

C Coeficiente de determinação

Ele sumariza quão bem uma modelo linear foi encaixado nos dados. Ele é calculado como:

$$R^2 = \frac{\sum(\hat{y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2}$$

Ele reflete a proporção de variação na variável prevista. Se a proporção é próxima de 1, significa que os preditores são próximos dos valores reais.

Entretanto esse índice não pode ser usado de forma equivocada. A adição de novos parêmetros não reduz o seu valor e o valor dele por si só diz pouca coisa sobre o modelo, já que existem outras formas de analisarmos os dados.

D Analise de Séries Temporais com Python

Para a análise temporal, algumas biliotecas são de extrema importância, como, por exemplo, *pandas* para lidar com banco de dados, *numpy* para lidar com a matemática do problema e armazenamento de memória, *sklearn* para obter diversos modelos para séries temporais e Machine Learning, *statsmodels* com todo o seu ferramental estatístico de testes e preditores já implementados e *matplotlib.pyplot* para montar os gráficos.

Sugiro como refêrencia: [12] e [13].

Agradecimentos

Agradeço ao professor Cláudio pela indicação do Livro em questão e agradeço à prefeitura do Rio por organizar dados tão importantes sobre a cidade e disponibilizá-los de maneira simples para o entendimento.

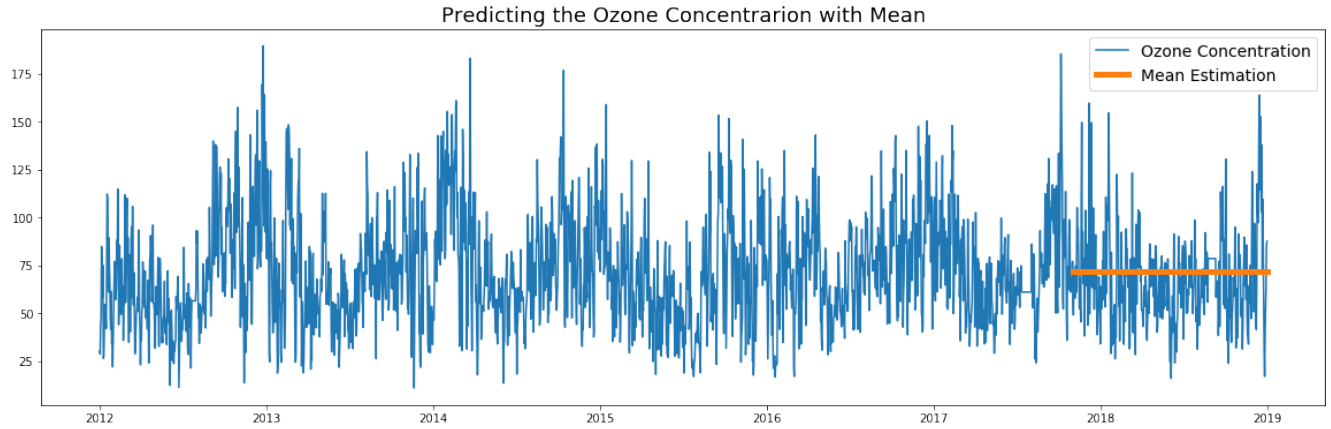

Figura 6. Previsão utilizando o método da Média

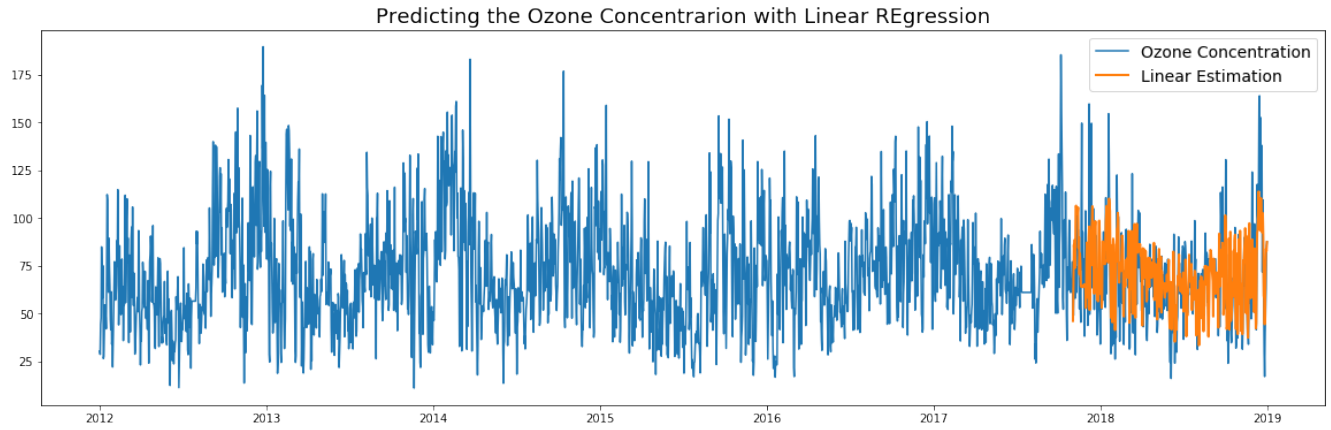

Figura 7. Previsão utilizando o método de Regressão Linear

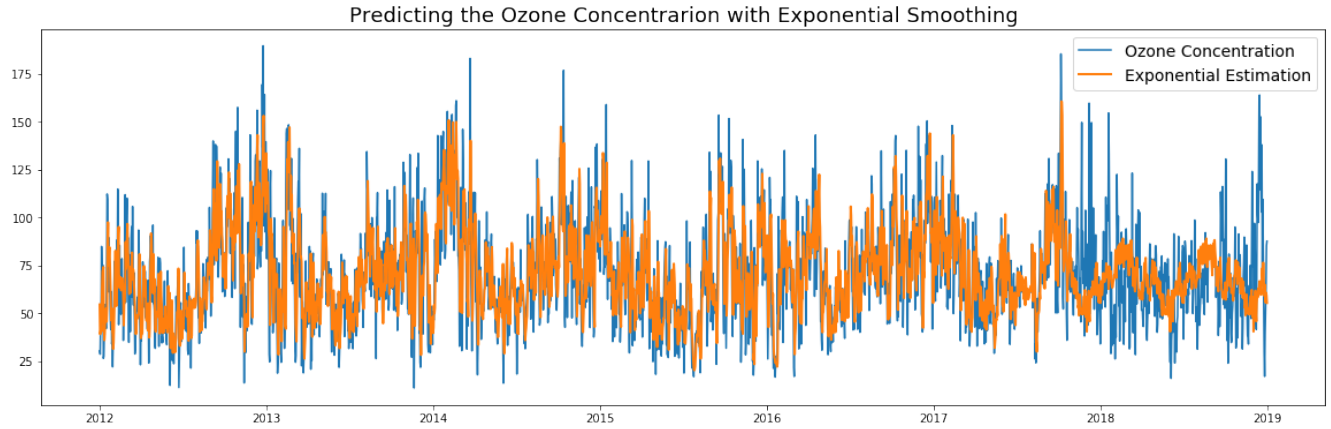

Figura 8. Previsão utilizando o método da Suavização Exponencial

Referências

[1] Caitlin Morton. The 50 most beautiful cities in the world. https://www.cntraveler.com/galleries/2016-01-08/the-50-most-beautiful-cities-in-the-world, 4 de abril 2019.

[2] Armazenzinho do Rio. Dados do rio. http://pcrj.maps.arcgis.com/apps/MapJournal/index.html?appid=9843cc37b0544b55bd5625e96411b0ee, 2019.

[3] INMET. Gráficos climatológicos. http://www.inmet.gov.br/portal/index.php?r=clima/graficosClimaticos, 2019.

[4] Prefeitura do Rio de Janeiro. Dados horários do monitoramento da qualidade do ar - monitorar. http://www.data.rio/datasets/dados-hor%C3%A1rios-do-monitoramento-da-qualidade-do-ar-monitorar, 2019.

[5] Rob J Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*, volume 4. OTexts: Melbourne, Australia, https://otexts.com/fpp2/, 2 edition, 2018.

[6] MonitorAr-Rio. Boletim de qualidade do ar. http://jeap.rio.rj.gov.br/je-metinfosmac/institucional/docs/saibamais_boletim.pdf, 2019.

[7] Melissa Sherrard. Gases that cause air pollution. https://sciencing.com/gases-cause-air-pollution-7445467.html, 2018.

[8] Encyclopedia britannica. https://www.britannica.com, 2019.

[9] Rodrigo de Losso da Silveira Bueno. *Econometria de Séries Temporais*. Cengage Learning, 2 edition, 2012.

[10] InsightsBot. Augmented dickey-fuller test in python. http://www.insightsbot.com/augmented-dickey-fuller-test-in-python/, 2019.

[11] Thomas K. Van Heuklon. Estimating atmospheric ozone for solar radiation models. *Solar Energy*, 22(1):63–68, 1979.

[12] Marco Peixeiro. The complete guide to time series analysis and forecasting. https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775, 2019.

[13] Davide Burba. An overview of time series forecasting models. https://towardsdatascience.com/an-overview-of-time-series-forecasting-models-a2fa7a358fcb, 2019.