

Analise da correlatividade de variáveis meteorológicas com novos casos de Covid-19 nos estados do Brasil

Andressa Marçal
a262878@dac.unicamp.br

Décio Gonçalves
d226072@dac.unicamp.br

Diego Alyson
d230640@dac.unicamp.br

Abstract—Com o impacto mundial que a Covid-19 tem causado, em número de infecções e mortes, se faz necessário investigar quais os fatores que corroboram para a disseminação do vírus desta doença, relacionando fatores climáticos com algumas outras variáveis. Uma das propostas estudadas para tentar modelar as causas da grande quantidade de casos da doença é a relação de influência da temperatura no espalhamento do vírus, seja através do ar ou de superfícies. Também observou-se em outros trabalhos a relação com outras variáveis meteorológicas, sendo elas precipitação total, temperatura do ar, umidade relativa do ar, velocidade do vento, radiação global e pressão atmosférica. Contudo, grande parte destes estudos foram conduzidos em países como China, Irã e Itália, ou de modo global sem se atentar as particularidades de cada local. Dentre os trabalhos investigados, alguns indicam que com o aumento da temperatura é observada uma redução no número de novos casos. Este trabalho tem por objetivo investigar a existência ou não de correlatividade entre novos casos de Covid-19 e variáveis meteorológicas nos estados Brasileiros.

I. INTRODUÇÃO

Historicamente, as condições climáticas já foram descritas como um possível fator que pode ou não influenciar a disseminação de vírus respiratórios, como por exemplo, o vírus da influenza. Nesse cenário atual, a pandemia da doença conhecida mundialmente como Covid-19, causada pelo novo coronavírus, denominado SARS-CoV-2, levou a uma grave crise mundial sanitária, econômica, social e política. O Brasil é um dos países com maior índice de novos casos mundiais, estando atrás somente dos Estados Unidos e Índia. O Covid-19 é uma doença contagiosa e de transmissão viral, acontece da pessoa infectada para uma sadia e, apesar da Covid-19, há uma transmissibilidade superior que a torna mais letal em números absolutos, além de ser um vírus novo para os seres humanos e que, por isso, não confere imunidade prévia, tornando seu grau de infecção ainda maior.

Com números expressivos, 8 milhões de casos da doença, uma média de até mil casos diários, alcançando um número de mortes de 204 mil até janeiro de 2021. A doença espalha-se cada vez mais entre as 27 unidades federativas brasileiras.

Alguns fatores que corroboram para este grande crescimento podem ser considerados, como densidade populacional, extensão territorial, valores socioeconômicos ou o estado meteorológico atual.

Como alguns trabalhos indicam [1]–[3], o estado meteorológico podem influenciar a quantidade de novos

casos, não resumindo-se a temperatura porém a outros efeitos como precipitação total, umidade relativa do ar, velocidade do vento, radiação global e pressão atmosférica.

Em virtude dos climas existentes no Brasil, devido seu grande território, que influenciam os estados meteorológicos de suas cidades, são identificados diversos tipos de climas, sendo os principais: equatorial, tropical, tropical de altitude, tropical úmido, semiárido e subtropical. Esta grande quantidade de possibilidades fornecem ao país uma grande diferença regional e possíveis reações diferentes em relação a doença e seu comportamento esperado.

Algumas perguntas podem ser consideradas, entre elas, a capacidade do tempo atual de uma determinada região influenciar na quantidades de novos casos de uma dada cidade ou estado. Logicamente, outras variáveis devem ser consideradas, como visto anteriormente, valores socioeconômicos, população, entre outros. Porém neste trabalho, será considerado de forma isolada a mudança no tempo e a sua capacidade de influência na quantidade final de infectados.

É importante observar que apesar dos estudos relacionados ao tempo e sua influência, assim como outros estudos quanto ao crescimento da quantidade de casos, não é possível decretar quais variáveis são inteiramente responsáveis por tal comportamento, uma vez que o resultado final é um conjunto de múltiplos fatores e suas influências individuais.

Múltiplas técnicas são propostas para realizar correlações entre variáveis, entre elas, correlações lineares, não lineares e transformadas. Uma possível técnica que fornece bons resultados e que relaciona as variáveis estudadas é o modelo dos mínimos quadrados comuns. Este modelo será aplicado nas variáveis anteriormente mencionadas, mensurando a sua capacidade na mudança da quantidade de novos casos.

Neste trabalho leva-se em consideração a capacidade que o vírus tem de infectar e se espalhar com diferentes valores de temperatura, precipitação total, umidade relativa do ar, velocidade do vento, radiação global e pressão atmosférica, dentro do território brasileiro, através da aplicação de um modelo de Mínimos Quadrados Comuns. Buscamos investigar se estas variáveis têm ou não influencia na quantidade de novos casos registrados no Brasil, ou seja, se existe alguma correlação entre as variáveis meteorológicas com a variação da quantidade de casos de Covid-19 no Brasil.

II. TRABALHOS RELACIONADOS

Em seu trabalho, Tobías e Molina [4] realizaram um estudo mundial e histórico sobre o efeito da temperatura. Este trabalho, considerou diferentes variantes do vírus SARS Cov para tentar correlacionar os conhecimentos anteriores com esta nova variação encontrada e definida como Covid-19. Observou-se neste estudo que o clima pode afetar a disseminação do vírus, onde baixas temperaturas demonstram uma maior quantidade de novos casos confirmados.

Como forma de validar os resultados, uma regressão Quasi-Poisson foi utilizada, que permite observar características específicas para o contexto de doenças infecciosas. Ajustou-se as tendências lineares e quadráticas, finais de semana e o período de *lockdown*. Considerou-se também a autocorrelação residual. Como o período médio de incubação para COVID-19 é de 5 a 6 dias, foi utilizado um modelo de defasagem distribuída de uma semana.

Outros trabalhos consideram esta mesma observação, Sobral *et Al.* [5], analisou as associações entre a transmissão e o número de mortes causadas pelo SARS-CoV-2 e variáveis meteorológicas, temperatura mínima, temperatura máxima e precipitação. Na análise dos dados diários por país, utilizou-se um modelo de dados em painel. O trabalho verificou que um aumento na temperatura média diária a cada 1 grau Fahrenheit reduz o número de casos em aproximadamente 6,4 casos/dia.

Em um estudo feito na China, Peng Shi *et Al.* [1] utilizou uma Regressão localmente ponderada e o gráfico de dispersão de suavização (LOESS), modelos não lineares de *lag* distribuído (DLNMs) e meta-análise de efeitos aleatórios para examinar a relação entre a taxa diária de casos confirmados de Covid-19, e as condições de temperatura. Os resultados indicam que com o aumento da temperatura é observada uma redução no número de novos casos.

Malki *et al.* [6] utilizaram vários modelos de regressão para extrair a relação entre diferentes fatores e a taxa de propagação da Covid-19. Os algoritmos de aprendizado de máquina empregados neste trabalho estimam o impacto de variáveis meteorológicas como temperatura e umidade na transmissão do Covid-19, extraíndo a relação entre o número de casos confirmados e as variáveis meteorológicas em determinadas regiões. Avaliou-se neste trabalho que a temperatura e umidade podem ser características importantes para a previsão da taxa de mortalidade da Covid-19.

Os seguintes modelos de aprendizado de máquina, como modelos lineares (regressão linear, regressão de laço, regressão de crista, rede elástica, regressão de ângulo mínimo, regressão de ângulo mínimo de laço, busca de correspondência ortogonal, cume bayesiano, determinação de relevância automática, regressor agressivo passivo, Random Sample Consensus, TheilSen Regressor, Huber Regressor) foram utilizados.

Além disso, foram utilizado modelos baseados em *ensemble* como *Random Forest*, *Extra Trees Regressor*, *AdaBoost Regressor* e *Gradient Boosting Regressor*. *Extreme Gradient Boosting* (XGBoost), *Light Gradient Boosting Machine*

(*LightGBM*) e *CatBoost Regressor*), *Kernel Ridge*, *Support Vector Machine* (SVM), *K-Nearest Neighbours Regressor* (KNN), *Multi-layers Perceptron* (MLP) e *Decision Tree* foram utilizados para a previsão da propagação do coronavírus.

Biryukov *et al.* [3] apresenta um estudo que correlaciona o aumento da temperatura, umidade relativa e a aceleração da inativação do SARS-CoV-2 em superfícies. Os resultados mostram que o SARS-CoV-2 decaiu mais rapidamente quando a umidade ou a temperatura aumentaram, mas o volume da gota e o tipo de superfície (aço inoxidável, plástico ou luva de nitrila) não afetaram significativamente a taxa de decomposição.

Quando na temperatura ambiente (24°C), a meia-vida do vírus variou de 6,3 a 18,6h dependendo da umidade relativa, mas foi reduzida para 1,0 a 8,9h quando a temperatura foi aumentada para 35°C. Estes resultados sugerem um alto potencial de transmissão, por horas a dias, em ambientes internos. Uma análise de regressão foi realizada para determinar um modelo preditivo de decaimento.

No trabalho de Jahangiri *et Al.* [2] os autores realizaram uma análise de sensibilidade e especificidade da temperatura ambiente e tamanho da população na taxa de transmissão do novo coronavírus (Covid-19) em diferentes províncias do Irã. Neste trabalho observou-se resultados positivos em relação a confirmação que o número de casos de Covid-19 em climas mais quentes é menor do que em climas moderados ou frios. Estas demonstrações foram feitas através de uma curva ROC da taxa de transmissão do coronavírus, mas observou-se que o parâmetro da temperatura ambiente tem uma relação linear com o número de pessoas afetadas com Covid-19.

Iqbal *et Al.* [7] verificou a correlação entre o Covid-19, temperatura e taxa de câmbio na cidade de Wuhan. Os resultados gerais sugerem a insignificância de um aumento na temperatura para conter ou retardar as novas infecções por Covid-19, apresentando resultados contrários a muitos estudos anteriores que sugerem um papel significativo da temperatura em desacelerar a propagação da Covid-19. Para encontrar tais resultados, foram utilizados, Transformada Wavelet Contínua (CWT), Coerência da Transformada Wavelet (WTC), Coerência Wavelet Parcial (PWC) e Coerência Wavelet Múltipla (MWC) para analisar a associação entre a temperatura média diária de Wuhan, número diário de novos casos de Covid-19 na cidade de Wuhan e a taxa de câmbio RMB.

Neto *et Al.* [8] analisa a correlação entre clima, tamanho da população e casos de covid-19 nos estados do Brasil. A análise foi feita desde o primeiro caso de covid-19, no dia 26 de fevereiro até o dia 01 de maio de 2020, o estudo encontrou uma correlação entre as variáveis meteorológicas e o tamanho da população para com a quantidade de novos casos de covid-19.

Naing *et Al.* [9] sugere em seu trabalho que os parâmetros climáticos e as previsões são úteis para reduzir a incerteza na gravidade da transmissão futura de COVID-19. Um destaque é que as previsões serão uma ferramenta útil para a tomada de decisões e as intervenções apropriadas para conter a propagação do vírus na comunidade.

Carleton *et Al.* [10] cita em seu trabalho uma correlação positiva entre a radiação e a diminuição de novos casos de covid-19.

III. PROCESSAMENTO DE DADOS

Nesta seção apresentaremos as bases de dados que serão utilizadas para embasamento e testes desta pesquisa e os processamentos realizados nestes dados. Exploramos dados provenientes de duas bases de dados, para alcançar o objetivo de obter informações de possíveis correlações entre o tempo e a propagação da Covid-19, utilizamos a base de eventos meteorológico do Instituto Nacional de Meteorologia (INMET) e a base de registros de casos e óbitos por covid-19 do portal Brasil.io.

Os dados meteorológicos obtidos foram variação da temperatura, umidade do ar, velocidade do vento, radiação global, precipitação e pressão atmosférica. Escolhemos esses dados para poder responder os questionamentos iniciais do trabalho, onde buscamos entender se há variação ou não de novos casos confirmados com o impacto do tempo. No INMET obteve-se dados em território nacional, por campos de estações meteorológicas do instituto.

Neste trabalho propomos realizar um estudo dos impactos meteorológicos em âmbito nacional e, dando ênfase nos estados Brasileiros, considerando-se as diferenças meteorológicas do Brasil. A base de dados do INMET disponibiliza estes dados por cidade e por histórico temporal.

Para analisar os casos de covid-19 nos estados Brasileiros, utilizamos os dados do portal Brasil.io, que disponibiliza os casos confirmados e óbitos diários, obtidos por meio de boletins das Secretarias Estaduais de Saúde (SES). Os dados são atualizados diariamente, a partir do momento em que um município confirma um caso. A coleta dos dados foi iniciada no dia 25 de Fevereiro de 2020, sendo a nossa última coleta até o dia 31 de Dezembro de 2020. Na base de dados constam 26 estados e o distrito federal e 5.294 cidades do Brasil.

Após a extração dos dados, fizemos uma análise de como esses dados estavam organizados. Na base de dados do Inmet observamos a ausência de algumas cidades do Brasil. O tratamento escolhido para lidar com esse caso, foi a separação dos dados por estados. Já no dataset do Brasil.io, também observou-se essa mesma inconsistência, onde a mesma tática foi utilizada.

Foi aplicada nas duas bases de dados, o procedimento de engenharia de features, para decidir quais as features mais importantes para a nossa pesquisa, onde foi realizado todo o processo de escolha, renomeação, limpeza, balanceamento e criação de novas colunas. Para identificação de *outliers*, foi usada a técnica de cálculo dos z-scores para cada variável.

No dataset do Covid-19, por ser uma base de dados volumosa, optamos por reduzir o escopo utilizando o filtro da data por meses ao invés de dias. E mapeamos uma métrica para filtrar as datas com a quantidade de novos casos confirmados maior ou igual a 1.500 mensais, essa escolha se deu para evitar possíveis *outliers*, tomando como base o início do decreto de pandemia, garantindo uma quantidade mínima de casos.

Após a junção das bases de dados, aplicamos duas novas métricas para analisar o comportamento dos casos confirmados de covid-19, onde foi aplicado um tempo de após 7 dias e em seguida um tempo de após 15 dias, tomamos como base alguns trabalhos citados na literatura, onde apontam que esse é o tempo mínimo de incubação do vírus. A nossa escolha de modelo foi da Regressão Linear OLS e optamos pela correlação de Pearson para interpretar nossos dados.

A. Regressão Linear dos Mínimos Quadrados Ordinários (OLS)

A análise da regressão gera uma equação para descrever a relação estatística entre uma ou mais variáveis preditoras e a variável resposta.

O *p*-value para cada termo testa a hipótese nula de que o coeficiente é igual a zero (sem efeito). Um *p*-value baixo (0,05) indica que pode rejeitar a hipótese nula. Um *p*-value maior (insignificante) sugere que as mudanças na preditora não estão associadas a mudanças na resposta.

O valor R² é a razão entre a variação explicada e a variação total. É um número que mede quão bem a reta da regressão se ajusta aos dados. O valor de R² está sempre contido entre 0 e 1; um valor de R² quase igual a 0 indica um ajuste ruim da reta.

Os coeficientes de regressão representam a mudança média na variável resposta para uma unidade de mudança na variável preditora, mantendo as outras preditoras na constante do modelo. Os coeficientes são frequentemente chamados de coeficientes de inclinação da reta.

B. Correlação de Pearson

Para medir a correlação entre os dados propostos e a sua capacidade de influencia nos resultados observados, optou-se por utilizar a correlação de Pearson.

Este coeficiente é um teste estatístico que mede a relação entre duas variáveis contínuas, resultando em valores entre +1 e -1, onde valores acima de +1 correspondem a uma associação positiva e valores abaixo de 0 indicam uma associação negativa entre as variáveis, indicando assim o quanto esta variável contribui para o aumento ou diminuição do valor da outra.

A correlação de Pearson pode ser calculada da seguinte maneira:

$$r_{xy} = \frac{\sum z_x z_y}{N} \quad (1)$$

Onde, *x* e *y* são as variáveis analisadas, *z_x* e *z_y* os desvios padrões das variáveis e *N* o número de amostras.

Decidiu-se utilizar esta correlação visto que tem-se boas estimativas em relação a correlação estudada quanto existem grandes quantidades de dados envolvidos.

IV. RESULTADOS

O número de casos de Covid-19 voltou a crescer aceleradamente no Brasil na reta final de 2020, o que ficou conhecido como segunda onda. “Segunda onda da pandemia” diz respeito

Table I

TABELA COM RESULTADOS DESCRITIVOS REFERENTES A MÉDIA, DESVIO PADRÃO, VALOR MÍNIMO E MÁXIMO DAS VARIÁVEIS PRECIPITAÇÃO, TEMPERATURA DO AR, UMIDADE DO AR, VELOCIDADE DO VENTO, RADIAÇÃO E PRESSÃO ATMOSFÉRICA DE TODOS OS ESTADOS DO BRASIL.

Estado	precipitacao (mm)	precipitacao (mm)	temperatura_ar (°C)	temperatura_ar (°C)	umidade_ar (%)	umidade_ar (%)	velocidade_vento (m/s)	velocidade_vento (m/s)	radiacao_global (Kj/m²)	radiacao_global (Kj/m²)	pressao_atmosferica (mB)	pressao_atmosferica (mB)
	mean - std	min - max	mean - std	min - max	mean - std	min - max	mean - std	min - max	mean - std	min - max	mean - std	min - max
AC	0.21 - 0.38	0.00 - 3.96	25.61 - 1.48	15.83 - 28.88	83.03 - 5.23	65.53 - 93.93	0.99 - 0.30	0.45 - 3.32	1213.13 - 348.80	235.13 - 2027.08	990.21 - 1.98	983.92 - 998.79
AL	0.14 - 0.26	0.00 - 2.18	26.24 - 1.74	22.61 - 29.68	71.37 - 5.70	56.15 - 86.16	2.03 - 0.58	0.85 - 3.98	1515.25 - 317.53	445.40 - 2102.38	999.26 - 4.15	988.64 - 1007.18
AM	0.24 - 0.29	0.00 - 2.43	27.16 - 0.92	23.88 - 29.96	78.82 - 5.29	64.67 - 89.33	0.65 - 0.13	0.37 - 1.49	756.09 - 140.76	291.49 - 1086.43	1004.69 - 1.42	1000.17 - 1008.38
AP	0.27 - 0.46	0.00 - 3.25	26.82 - 1.05	24.05 - 29.13	80.60 - 5.86	65.75 - 92.17	1.82 - 0.34	0.91 - 2.86	1429.17 - 327.86	416.43 - 1966.50	1006.01 - 1.42	1002.11 - 1010.59
BA	0.13 - 0.17	0.00 - 0.95	24.54 - 1.73	20.51 - 28.62	73.31 - 7.09	53.99 - 89.61	2.09 - 0.41	1.16 - 3.14	1431.12 - 298.99	563.76 - 2010.78	972.70 - 2.56	966.60 - 979.87
CE	0.10 - 0.19	0.00 - 1.10	26.71 - 0.97	24.28 - 28.81	67.91 - 9.96	50.05 - 86.82	2.30 - 0.65	1.01 - 3.61	1309.75 - 216.33	596.52 - 1667.70	982.80 - 3.49	975.21 - 987.91
DF	0.22 - 0.42	0.00 - 2.86	21.59 - 2.16	16.73 - 28.20	67.83 - 16.54	24.51 - 93.57	2.17 - 0.60	1.17 - 4.02	1491.25 - 377.64	384.25 - 2215.68	896.38 - 2.17	890.98 - 902.93
ES	0.16 - 0.31	0.00 - 2.46	23.25 - 2.21	18.20 - 27.87	75.84 - 5.90	61.81 - 93.42	2.13 - 0.43	1.26 - 3.66	1204.08 - 380.67	287.87 - 1999.37	989.75 - 4.23	978.81 - 999.19
GO	0.17 - 0.26	0.00 - 1.58	24.01 - 2.25	17.77 - 30.94	63.15 - 16.09	24.35 - 87.95	1.36 - 0.36	0.78 - 2.77	1426.80 - 261.31	680.95 - 1956.30	931.68 - 3.01	923.64 - 938.20
MA	0.17 - 0.29	0.00 - 2.71	27.15 - 1.08	24.61 - 29.96	70.64 - 10.73	45.22 - 88.86	1.13 - 0.30	0.47 - 2.10	1425.64 - 234.12	565.20 - 2009.97	994.93 - 1.48	991.36 - 999.24
MG	0.16 - 0.24	0.00 - 2.09	21.83 - 2.47	15.98 - 28.70	70.76 - 10.47	38.51 - 90.94	1.71 - 0.35	0.92 - 2.63	1400.57 - 275.87	525.73 - 1994.25	929.46 - 2.50	922.25 - 936.73
MS	0.12 - 0.24	0.00 - 1.55	24.24 - 3.62	10.72 - 31.53	65.47 - 12.81	29.46 - 90.43	1.95 - 0.46	1.03 - 3.49	1375.34 - 347.63	357.89 - 2171.80	967.08 - 3.64	959.52 - 979.25
MT	0.14 - 0.18	0.00 - 1.24	25.96 - 2.01	18.12 - 31.20	67.43 - 14.62	32.42 - 89.73	1.46 - 0.35	0.83 - 3.73	1297.92 - 197.80	540.51 - 1788.42	966.57 - 2.39	958.73 - 972.90
PA	0.26 - 0.28	0.00 - 1.47	26.68 - 0.92	24.17 - 28.88	75.13 - 7.82	58.43 - 89.28	1.05 - 0.16	0.70 - 1.62	1254.42 - 229.51	445.28 - 1819.51	995.90 - 1.36	992.17 - 999.99
PB	0.11 - 0.21	0.00 - 1.68	25.54 - 1.46	22.08 - 28.61	69.43 - 8.14	54.95 - 87.32	2.40 - 0.70	0.62 - 4.11	1409.27 - 230.68	659.54 - 1854.66	968.52 - 3.27	960.09 - 975.35
PE	0.10 - 0.17	0.00 - 1.43	25.24 - 1.74	21.50 - 28.55	64.60 - 7.90	47.22 - 81.36	2.56 - 0.59	1.04 - 3.97	1346.23 - 252.19	595.66 - 1798.00	963.52 - 1.95	957.00 - 968.94
PI	0.13 - 0.24	0.00 - 1.53	27.09 - 1.46	24.40 - 30.92	64.89 - 14.22	34.40 - 87.84	1.87 - 0.57	0.94 - 3.25	1464.41 - 275.57	624.43 - 1937.49	978.55 - 2.06	973.58 - 983.24
PR	0.13 - 0.26	0.00 - 2.61	20.44 - 3.48	8.16 - 28.47	70.96 - 9.22	46.05 - 93.59	1.95 - 0.52	0.93 - 3.90	1303.30 - 410.32	195.43 - 2197.82	941.60 - 4.31	930.56 - 952.49
RJ	0.19 - 0.36	0.00 - 3.00	21.94 - 2.40	15.44 - 7.37	78.81 - 6.17	59.45 - 93.07	1.77 - 0.37	0.98 - 2.74	1209.50 - 412.19	199.73 - 2038.82	980.95 - 4.90	969.67 - 993.45
RN	0.13 - 0.23	0.00 - 1.58	28.26 - 1.24	24.71 - 31.16	67.00 - 10.59	47.43 - 88.32	3.17 - 0.94	1.18 - 4.82	1729.30 - 330.00	673.37 - 2367.93	1004.06 - 1.50	999.68 - 1007.40
RO	0.19 - 0.38	0.00 - 3.37	26.09 - 1.78	14.40 - 30.04	70.06 - 13.23	37.27 - 91.56	1.50 - 0.33	0.94 - 4.09	1356.48 - 290.98	470.20 - 1979.99	972.11 - 5.04	961.84 - 983.37
RR	0.21 - 0.46	0.00 - 3.17	27.57 - 1.60	23.86 - 31.02	68.23 - 11.28	48.17 - 89.96	1.81 - 0.59	0.80 - 3.32	1394.71 - 440.56	388.82 - 2372.70	1002.36 - 1.33	997.76 - 1005.94
RS	0.15 - 0.30	0.00 - 2.27	18.65 - 4.78	5.77 - 27.35	72.86 - 8.60	52.61 - 94.48	2.63 - 0.67	1.33 - 5.00	1325.42 - 484.75	136.55 - 2196.81	975.93 - 4.41	964.02 - 988.56
SC	0.17 - 0.27	0.00 - 1.45	18.20 - 3.68	5.47 - 24.76	77.50 - 7.41	60.64 - 95.27	2.14 - 0.50	0.99 - 4.32	1211.62 - 443.99	162.68 - 2128.22	944.63 - 5.00	932.27 - 956.67
SE	0.13 - 0.25	0.00 - 2.47	25.41 - 1.87	21.70 - 28.81	74.73 - 6.77	59.96 - 90.42	1.40 - 0.67	0.30 - 3.13	1613.96 - 324.61	528.42 - 2309.47	990.85 - 3.16	982.92 - 998.48
SP	0.14 - 0.24	0.00 - 2.13	20.74 - 2.93	10.50 - 28.41	68.19 - 10.36	38.50 - 90.22	1.98 - 0.49	0.96 - 3.69	1309.31 - 320.98	212.49 - 2042.97	943.46 - 3.52	933.53 - 952.49
TO	0.18 - 0.28	0.00 - 1.48	26.39 - 1.65	23.42 - 31.09	69.68 - 14.70	34.88 - 90.76	1.54 - 0.33	0.90 - 2.56	1453.07 - 264.70	744.78 - 1894.95	978.28 - 1.87	974.16 - 983.83

a novos surtos que podem ocorrer após um declínio inicial na taxa de contaminados pelo vírus. Na Figura 1 é possível ver como está esse crescimento de novos casos de Covid em todos os estados brasileiros e distrito federal.

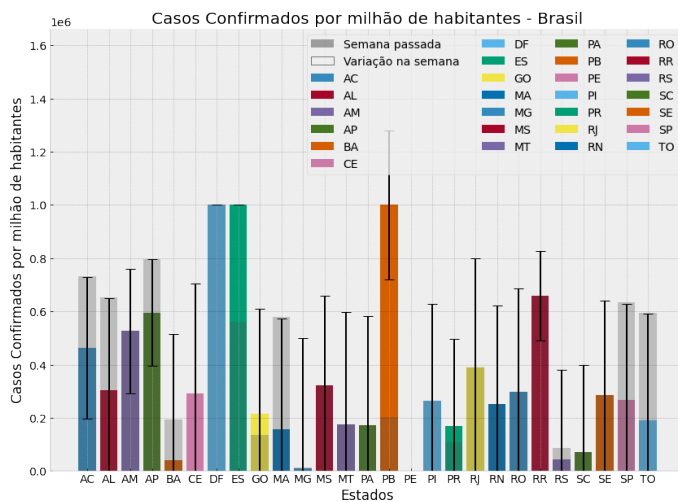


Figure 1. Crescimento de novos casos confirmados de covid-19 nos estados do Brasil.

Foram coletados dados meteorológicos para os 25 estados brasileiros e o distrito federal, até o mês de Dezembro de 2020. Para um melhor entendimento dos dados descritivos, populamos uma tabela com informações de média, desvio padrão, mínimo e máximo das variáveis meteorológicas de precipitação, temperatura do ar, umidade, velocidade do vento, radiação e pressão atmosférica para todos os estados do Brasil e distrito federal. Esse mapeamento nos auxiliou na compreensão dos dados utilizados.

Podemos observar a média de cada variável para avaliar a medida de tendência central que indica o valor onde estão concentrados os dados de um conjunto de valores.

O desvio padrão é uma medida de dispersão, que indica quanto os dados estão afastados da média. Um valor de desvio padrão alto indica que os valores estão mais espalhados, mais longe da média, e um desvio padrão baixo indica que os valores estão mais próximos da média.

O Brasil é regido por um leque de variações, onde podemos observar temperaturas que vão de 5,47 graus à 31,53 graus. Em relação aos dados meteorológicos dos estados analisados, Santa Catarina (SC) foi o local que registrou a menor temperatura média (18,65 C°) e o Rio Grande do Norte (RN) a maior temperatura média (28,26C°).

O maior registro de umidade do ar foi encontrado no estado do Acre (AC), com 83,03%, enquanto o menor foi no estado do Goiás (GO) (63,15%).

Os estados de Rio Grande do Sul (RS) e Sergipe (SE) registraram o maior e a menor velocidade do vento, com 2,33 m/s e 0,30 m/s.

Os estados de Roraima (RO) e Alagoas (AL) registraram a maior e a menor média de pressão atmosférica, chegando a 1002.36mB e 999.26mB.

Os estados do Espírito Santo (ES) e o Amazonas (AM) registraram a maior e a menor taxa de radiação global, sendo Espírito Santo a maior com 1204,08 Kj/m² e o Amazonas com a menor 756,09Kj/m².

E por último, os estados do Amapá (AP) e Pernambuco (PE) registraram a maior e a menor média de precipitação de chuvas, com 0.10mm e 0.27mm. Os dados detalhados são exibidos na Tabela I.

Com o intuito de constatar a correlação entre os dados estudados, aplicou-se a estimativa de Pearson entre as variáveis precipitação total, temperatura do ar, umidade relativa do ar, velocidade do vento, radiação global e pressão atmosférica. O resultado obtido pode ser observado na imagem 2, através da coluna de correlação entre as variáveis e o número de novos casos.

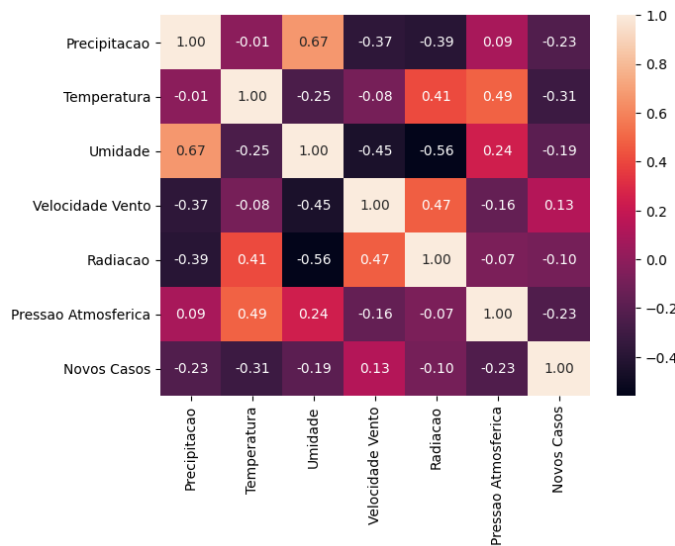


Figure 2. Resultados da Correlação de Pearson para as variáveis meteorológicas e novos casos de Covid-19 nos estados Brasileiros

Observa-se que existe uma associação entre os dados e os novos casos, apesar de serem correlações fracas. Dentre estas, a correlação com maior valor foi encontrada entre os novos casos e a temperatura com valor de 0,31 pontos de correlação. Entre as variáveis consideradas como importantes, através de Pearson para a mudança do número de novos casos é a temperatura, precipitação que é altamente relacionada com umidade e a pressão atmosférica.

O mesmo estudo foi realizado para um período entre sete e quinze dias a partir do dia da coleta, para refletir assim os possíveis dias de incubação do vírus, assim como relatado na literatura médica. Estes valores podem ser observados a partir da imagem 3. Nesta imagem observa-se que alguns valores, após serem ajustados ao período de incubação apresentam uma maior correlação com o número de novos casos, como a temperatura. Outros valores reduziram sua correlação, como a precipitação, demonstrando assim que o período definido como incubação é importante para tal estudo [11].

Outra análise realizada neste trabalho diz respeito a regressão OLS entre os dados observados. Da mesma forma, foram estudados as relações entre os resultados do modelo

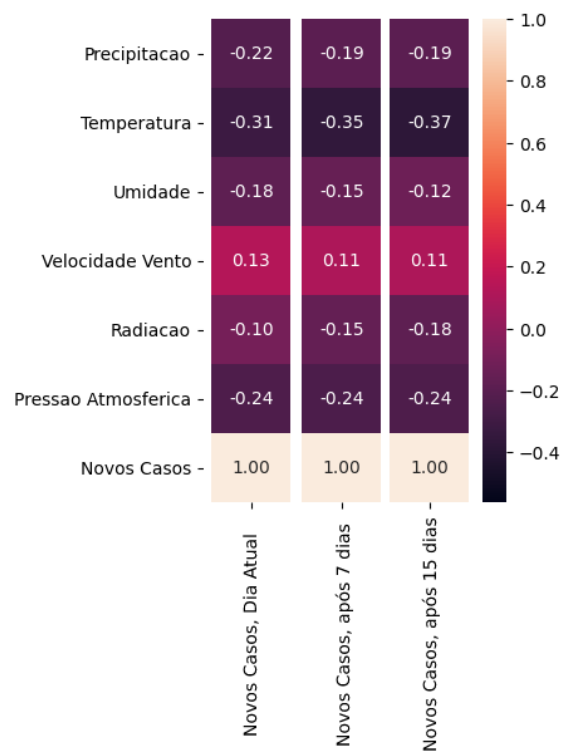


Figure 3. Resultados da Correlação de Pearson para as variáveis meteorológicas e novos casos de Covid-19 nos estados Brasileiros em faixas para o dia atual, sete dias e quinze dias.

para diferentes períodos de tempo, sendo o dia atual igual ao da coleta, após sete dias e após quinze dias. Os resultados obtidos podem ser observados na tabela II.

Para este modelo, múltiplas variáveis podem ser analisadas. Primeiramente, observou-se valores de R^2 acima de 0.66 para todos os tipos de coleta, o que demonstra o qual bem os dados são representados pelo modelo linear produzido. Entre estes resultados, também é observado o p-value de cada uma das variáveis envolvidas. Visto que valores de p-value inferiores a 0,05 representam valores significativos estatisticamente, temos que as variáveis Temperatura, Umidade e Pressão atmosférica respeitam estes limites, alguns ainda sendo reduzidos com a utilização das faixas de tempo de 7 à 15 dias.

Outra métrica observada é o coeficiente da OLS, que repre-

Table II
RESULTADOS DA REGRESSÃO OLS APLICADA AS VARIÁVEIS METEOROLÓGICAS DE COM RELAÇÃO AOS NOVOS CASOS DE COVID-19 NOS ESTADOS DO BRASIL

Parâmetro	Dia = Atual ($R^2 = 0.668$)		Após 7 dias ($R^2 = 0.676$)		Após 15 dias ($R^2 = 0.679$)	
	Coeficiente	p-value	Coeficiente	p-value	Coeficiente	p-value
Precipitação (mm)	-1,73E+07	0,471	-1,04E+07	0,664	-2,11E+07	0. 397
Temperatura (C°)	-4240.3424	0.00	-4707.9808	0. 000	-4645.3082	0. 000
Umidade(C°)	-1048.7666	0,005	-1123.8166	0. 002	-975.4966	0. 008
Velocidade Vento (m/s)	1405.8444	0,721	1091.0245	0. 781	1294.1078	0. 742
Radiação (Kj/m²)	-12.4759	0,275	-17.4795	0. 122	-21.3399	0. 059
Pressão Atmosférica(mB)	244.2443	0.00	267.9899	0. 000	262.2204	0. 000

senta quantas unidades de uma determinada variável influencia em unidades da variável final. Aqui vamos observar apenas as variáveis que têm valores de p -value abaixo de 0.05, devido as observações realizadas acima. Desta forma, o resultado do modelo indica que a cada 1 °C têm-se uma redução de 4707 casos quando analisamos os resultados para faixas de sete dias de diferença. Da mesma forma, para cada unidade de umidade, têm-se uma redução de até 1123 novos casos, enquanto que para valores de pressão em mB, para cada unidade, uma redução de 267 casos.

Com isso, constata-se que, através dos resultados obtidos com o modelo, pode haver uma tendência na diminuição de casos em ambientes com temperaturas mais elevadas, úmidas e com uma menor pressão atmosférica.

V. LIMITAÇÕES/TRABALHOS FUTUROS

Em nosso trabalho encontramos algumas limitações, uma delas foi na dificuldade de encontrar dados referentes a qualidade do ar. Alguns trabalhos na literatura observam a variável da qualidade do ar e correlacionam com a quantidade de novos casos [12], o que pensamos que poderia ser relevante em nossa análise, porém por falta de dados referentes a qualidade do ar no Brasil, essa análise não foi feita. Uma outra lacuna de nosso trabalho se refere aos modelos utilizados, por exemplo em uma análise futura seria interessante o uso da regressão do tipo *Generalized Additive Model* (GAM), esta regressão é bastante difundida na literatura para análises envolvendo variáveis meteorológicas [13]. Com um modelo GAM poderíamos, por exemplo, adicionar informações relativas ao início da pandemia em cada localização, isso nos traria uma modelagem mais realista de como ocorreu o espalhamento em cada lugar analisado. E por fim, uma outra direção para trabalhos futuros seria analisar a possível sazonalidade do COVID-19 durante os períodos de baixa umidade, observando o seu comportamento e variância.

VI. CONCLUSÕES

Com o avanço do vírus tipo SARS-COV-2, e a alta taxa de transmissão e novos casos da doença em múltiplos países, faz-se necessário um estudo profundo quanto as vias e possíveis variáveis que influenciam neste resultado. Muitos trabalhos demonstram que diferentes características podem influenciar a quantidade de novos casos, como quantidade populacional e tamanhos territoriais. Outros estudos tentam correlacionar valores ambientais, como clima, tempo e outros.

Neste estudo demonstramos resultados de diferentes variáveis climáticas e sua importância para o aumento de novos casos mundiais, os resultados obtidos através de uma Regressão Linear dos Mínimos Quadrados Ordinários apontam uma associação linear negativa entre a temperatura, umidade média do ar e número total de casos COVID-19. Enquanto demonstra uma correlação linear positiva entre as variáveis de Pressão Atmosférica e Novos casos. Nenhum outro dado climático analisado mostrou uma correlação significativa.

Os resultados encontrados estão de acordo com algumas expectativas abordadas neste trabalho e em outros trabalhos

relacionados. Os resultados levantados nessa pesquisa, podem auxiliar governos e autoridades sanitárias na tomada de decisões sobre medidas de controle da pandemia. Estudos em outras regiões são importantes para fortalecer os achados além de observações da ação de outras variáveis e seus impactos no resultado final.

REFERENCES

- [1] P. Shi, Y. Dong, H. Yan, C. Zhao, X. Li, W. Liu, M. He, S. Tang, and S. Xi, "Impact of temperature on the dynamics of the covid-19 outbreak in china," *Science of The Total Environment*, p. 138890, 2020.
- [2] M. Jahangiri, M. Jahangiri, and M. Najafgholipour, "The sensitivity and specificity analyses of ambient temperature and population size on the transmission rate of the novel coronavirus (covid-19) in different provinces of iran," *Science of The Total Environment*, p. 138872, 2020.
- [3] J. Biryukov, J. A. Boydston, R. A. Dunning, J. J. Yeager, S. Wood, A. L. Reese, A. Ferris, D. Miller, W. Weaver, N. E. Zeitouni *et al.*, "Increasing temperature and relative humidity accelerates inactivation of sars-cov-2 on surfaces," *MSphere*, vol. 5, no. 4, 2020.
- [4] A. Tobías and T. Molina, "Is temperature reducing the transmission of covid-19?" *Environmental Research*, vol. 186, p. 109553, 2020.
- [5] M. F. F. Sobral, G. B. Duarte, A. I. G. da Penha Sobral, M. L. M. Marinho, and A. de Souza Melo, "Association between climate variables and global transmission of sars-cov-2," *Science of The Total Environment*, vol. 729, p. 138997, 2020.
- [6] Z. Malki, E.-S. Atlam, A. E. Hassanien, G. Dagnew, M. A. Elhosseini, and I. Gad, "Association between weather data and covid-19 pandemic predicting mortality rate: Machine learning approaches," *Chaos, Solitons & Fractals*, vol. 138, p. 110137, 2020.
- [7] N. Iqbal, Z. Fareed, F. Shahzad, X. He, U. Shahzad, and M. Lina, "Nexus between covid-19, temperature and exchange rate in wuhan city: New findings from partial and multiple wavelet coherence," *Science of The Total Environment*, p. 138916, 2020.
- [8] R. A. A. Neto and G. C. Melo, "Correlation between weather, population size and covid-19 pandemic: a study of brazilian capitals," *Journal of Health & Biological Sciences*, vol. 8, no. 1, pp. 1–5, 2020.
- [9] C. Naing, H. Ni, H. H. Aung, E. C. W. Ling, and J. W. Mak, "The influence of climate factors on covid-19 transmission in malaysia: An autoregressive integrated moving average (arima) model," *medRxiv*, 2020.
- [10] T. Carleton, J. Cornet, P. Huybers, K. C. Meng, and J. Proctor, "Global evidence for ultraviolet radiation decreasing covid-19 growth rates," *Proceedings of the National Academy of Sciences*, vol. 118, no. 1, 2021.
- [11] R. G. Garrido and F. d. S. R. G. Garrido, "Covid-19: um panorama com ênfase em medidas restritivas de contato interpessoal," *Interfaces Científicas-Saúde e Ambiente*, vol. 8, no. 2, pp. 127–141, 2020.
- [12] H. Huang, X. Liang, J. Huang, Z. Yuan, H. Ouyang, Y. Wei, and X. Bai, "Correlations between meteorological indicators, air quality and the covid-19 pandemic in 12 cities across china," *Journal of Environmental Health Science and Engineering*, vol. 18, no. 2, pp. 1491–1498, 2020.
- [13] D. N. Prata, W. Rodrigues, and P. H. Bermejo, "Temperature significantly changes covid-19 transmission in (sub) tropical cities of brazil," *Science of the Total Environment*, p. 138862, 2020.