

Effect of temperatures, humidity and population density on the spreading of Covid-19 at 70 cities/provinces

¹Jishan Ahmed, ²Md. Hasnat Jaman, ^{3,*}Goutam Saha, ³Pratyya Ghosh

¹Department of Computer Science, Bowling Green State University, USA. Email: jahmed@bgsu.edu

²Department of Geology and Mining, University of Barishal, Bangladesh. Email: mhjaman@bu.ac.bd

³Department of Mathematics, University of Dhaka, Bangladesh.

Email: [*gsahamath@du.ac.bd](mailto:gsahamath@du.ac.bd); pratyaghosh@yahoo.com

*Corresponding author

Abstract

The main goal of this article is to demonstrate the impact of environmental data on the spreading of Covid-19. In this research, data has been collected from 70 cities/provinces that are affected by Covid-19. Here, environmental data refers to temperatures, humidity and population density in each of these cities/provinces. This data has been analyzed using statistical models such as Poisson, Quasi-Poisson and negative Binomial. It is found that a negative Binomial regression model is the best fit for our data. Our results reveal that average high temperature is the vital factor to slow down the spread of Covid-19. In addition, higher population density found to be an important factor for the quick spreading of Covid-19 where it is quite impossible to maintain the social distance and the virus can spread easily.

Keywords: Temperatures, Humidity, Population density, Environment, Covid-19, Generalized linear model.

Introduction

In December 2019, a new RNA virus strain from the family *Coronaviridae* emerged in Wuhan, the capital of Hubei province. (Wu et al., 2020). This novel virus is a betacoronavirus and designated as SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus-2) causing a pneumonia disease called coronavirus disease 2019 (Covid-19) (Gorbalenya et al., 2020). Though SARS-CoV-2 has a low mortality rate (about 2.3%) compared to other coronaviruses like SARS-CoV (about 10%) and MERS-CoV (about 35%), the reproduction number or transmission rate of SARS-CoV-2 virus is very high (2.24 - 3.58) (Ceccarelli et al., 2020; Zhao et al., 2020) causing rapid spreading and becoming a pandemic. Though fever, fatigue and dry cough are the most common symptoms, some patients can develop severe and even fatal complications such as Acute Respiratory Distress Syndrome (ARDS) (D. Wang et al., 2020).

Coronaviruses are enveloped viruses which predominantly deputize through outright contiguity with respiratory droplets of an infected person (generated through coughing and sneezing). By touching the virus contaminated surface afterwards touching one's own face (i.e., eyes, ears, nose and mouth) a distinct person can also be infected. Enveloped viruses can survive for several hours on different surfaces; however, they show sensitivity to heat, detergent and desiccation compared to non-enveloped viruses (Howie et al., 2008). Therefore, environmental factors have a great impact on the transmission of infectious disease by affecting the survival of coronavirus on surface or in air (Casanova et al., 2010). High temperature and high relative humidity environments reduced the transmission of SARS coronavirus (Chan et al., 2011). Ma et al., (2020) found that both 1 unit increase of temperature and absolute humidity were related to the decreasing of Covid-19 death. Some other studies (Hongchao et al., 2020; Oliveiros et al., 2020; Tosepu et al., 2020; J. Wang et

al., 2020) also support that there is a relation between environmental factors and Covid-19 i.e., spreading decreases with increasing temperature. Along with these factors population density and mobility can trigger the spreading of this virus. Although hundreds of scientific articles have been published on Covid-19, there is no documented report found that considers population density and mobility along with environmental factors to know and control the spreading of Covid-19.

The main goal of this research is to provide statistical modeling-based scientific evidence regarding the spreading of the SARS-CoV-2 under the changing circumstances of humidity, temperature and higher population density.

Materials and Methods

In this research, the total number of infected by Covid-19, population density, monthly average humidity and average high and low temperatures data have been collected at 70 cities/provinces around the world from January 18, 2020 to April 24, 2020.

Three different models such as negative Binomial, Poisson, Quasi-Poisson models are considered in this research. These models are assessed using the Akaike Information Criterion (AIC), residual deviance, Pseudo R^2 , and Pearson χ^2 . We have used `glm()` function to fit the Poisson, Quasi-Poisson models and `glm.nb()` function to fit the negative Binomial model from the R software (version 4.0.0) package MASS (R Core Team, 2017). Two-sided statistical tests were considered along with 5% significance level.

Data Observation

Figure 1 shows the distribution of the number of infected people in 70 cities/provinces. Our initial observations suggest that there could be a relationship between the environmental parameters and expansion of Covid-19 across the different geographical locations. Most of the cities/provinces where outbreaks occurred such as Madrid, New York etc., had low temperature and/or low to moderate humidity because coronavirus can survive longer on surfaces or respiratory droplets at this environmental condition. Places with relatively high humidity and high temperature i.e. Banten, Central Luzon etc., showed comparatively less infected people. Another factor, population density and mobility alone can trigger the infection rate logarithmically irrespective of environmental condition. São Paulo, Riyadh etc. cities had high temperature and high humidity but many infected people due to population density and mobility. In cold regions, population density can exacerbate the total Covid-19 infection along with the environment. According to the data analysis, our observation illustrates that there could be a remarkable connection between the environmental parameters and the nature of the Covid-19 virus. In the next section, we will present statistical analysis and try to understand the above mentioned behavior.

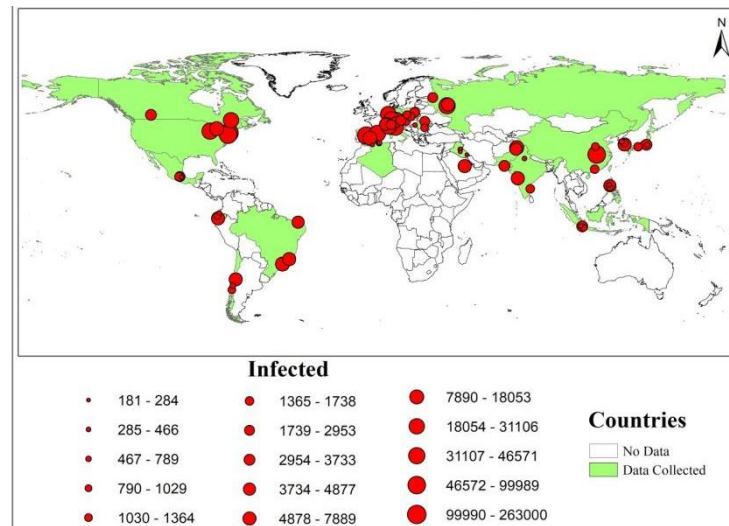


Figure 1: Distribution of Infected population

Statistical analysis

In this paper, a generalized linear model (GLM) framework (Agresti A., 2015) for count data has been deployed to analyze the effects of population density, humidity and average high and low temperatures on the spreading of Covid-19. In the Poisson regression model, it is assumed that the variance and mean of the dependent variable are the same. However, this assumption is not always true, especially while studying the environmental risk to human health due to the fact that the variance is higher than average causing the overdispersion of the data. It is challenging to handle overdispersion in the modeling of count response variables like the number of Covid-19 confirmed cases. In our data, the variance of the infected cases is 1279339997 and mean is 14924.97 - variance is larger than the mean. Also, from Figure 2, we see that our response variable, the count of infected cases is highly skewed. This indicates that our data may be overdispersed. It is convenient to use a negative Binomial model to estimate the parameter due to the presence of overdispersion of the data. Therefore, in this study, we have considered the negative Binomial model and compared our results with the Poisson, quasi-Poisson models as well to detect the overdispersion of our data. The negative Binomial model (Agresti A., 2015) was defined as

$$Y_i \sim NB(\mu_i, \theta)$$

$$\text{Log}(\mu_i) = \alpha + \beta_1 \log(\text{PopDensity}) + \beta_2 \text{Humidity} + \beta_3 \text{AvgHigh} + \beta_4 \text{AvgLow}$$

where i is the index of the city/provinces; Y_i is the observed Infected case on city/province i ; α is the model intercept; PopDensity is the population density in 2020; Humidity is the average humidity; AvgHigh is the average high temperature and AvgLow is the average low temperature within the period respectively.

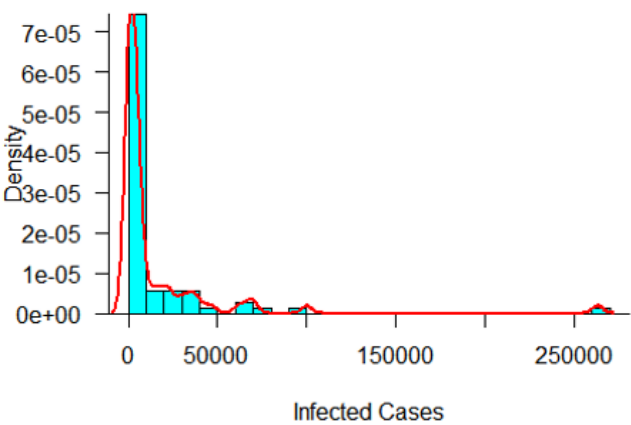


Figure 2: Distribution of Infected case

Results:

Dataset descriptive analysis

In this work, we considered 70 cities/provinces around the world that had the confirmed cases of Covid-19. Figure 3 shows all explanatory variables through normalized heat map representations. The color scale on the right represents the intensity of the variables according to the saturation level of this scale. For example, New York had the highest number of Covid-19 confirmed cases which is shown in this figure with a highly saturated blue color.

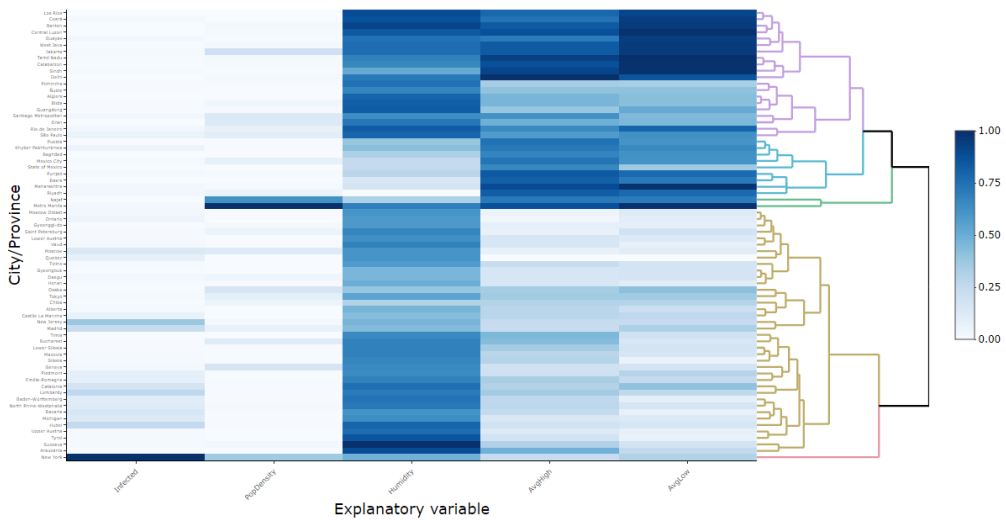


Figure 3: Normalized explanatory variables by cities/provinces

Now, Pearson correlation coefficients are computed as shown in Fig. 4 to determine the possible effects of collinearity. It shows that the Pearson correlation between humidity and temperatures along with the significance measure. There were negative correlations among infected case count and humidity, average high and low temperatures. However, there was a strong positive correlation between average high and low temperatures.

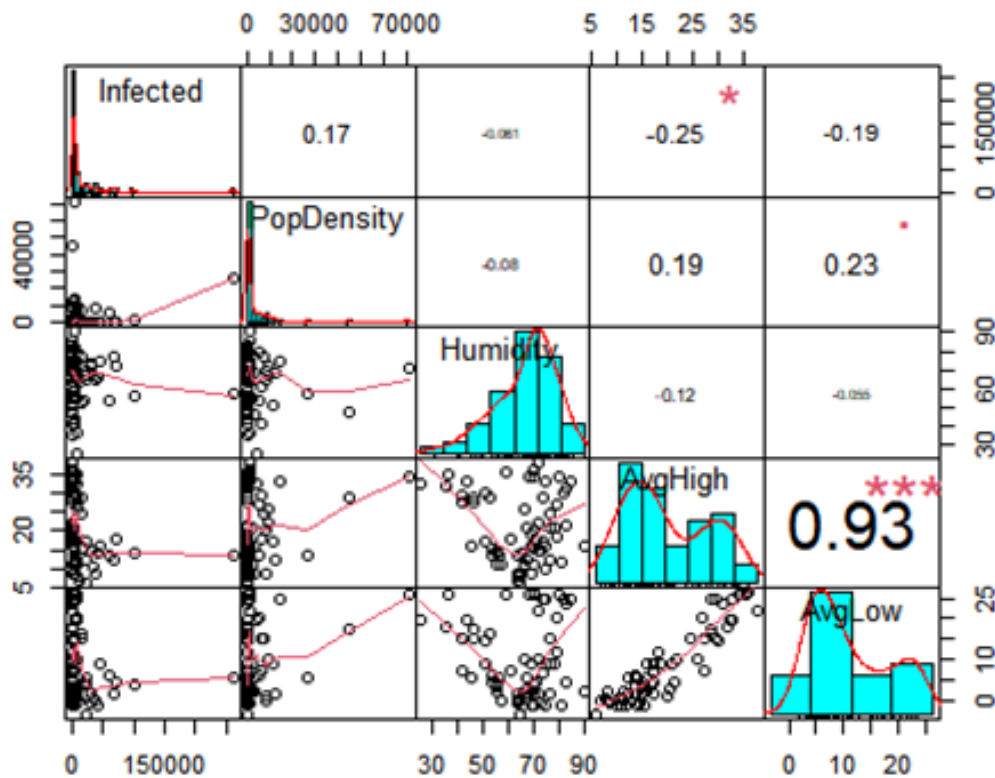


Figure 4: The pairwise plot along with Pearson correlation coefficients of infected cases, population density, humidity and average high and low temperatures.

In the following, we presented the summary statistics for infected cases, population density, humidity and average high and low temperatures as shown in Table 1. It is to be noted that the average number of confirmed infected cases was 14925, the mean value of population density was 4043.4 per km², the mean values of humidity, average high and low temperatures were 65.28%, 20.42°C and 9.41°C.

Table 1: Summary statistics for different parameters

Variables	Mean	SD	Min	Max
Infected cases	14925	35767.86	181	263000
Population Density (/km ²)	4043.4	10563.71	6.0	71263
Humidity (%)	65.28	12.92	26.00	90.00
Average high temperature (°C)	20.42	9.05	6.00	38.09
Average low temperature (°C)	9.41	9.48	-3.230	26.250

Negative Binomial model assessment

Diagnostic plots were used to check the performance of the Poisson, Quasi-Poisson, and the negative Binomial models. In Figures 5-7, the jackknife deviance residuals vs. the fitted values are displayed on the top left panel, and normal QQ plots of the standardized deviance residuals are shown on the top right panel. The dotted line of the normal QQ plots represents the expected line if the standardized residuals are normally distributed (Davison et al., 1991). Cook statistics are shown in the bottom two panels. The bottom left plot shows the Cook statistics vs. the standardized leverages. The horizontal line is drawn at $8/(n-2p)$, and the

vertical line is drawn at $2p/(n-2p)$, where n represents the number of observations and p represents the number of estimated parameters. Points above the horizontal line may be points which have high influence on the model. On the other hand, high leverage points correspond to the right side of the vertical line. We had 70 cities with Covid-19 confirmed cases, and 5 parameters were estimated. For our case, the horizontal line was drawn at 0.13, and the vertical line was drawn at 0.17. The final plot shows us the influential observations using the plot Cook statistic vs. case number. In Figure 5, we clearly see that the Poisson model is inadequate and it does not have any influential points. In contrast, in Figures 5-6, we see that Quasi-Poisson, and the negative Binomial models are far superior to the Poisson model. Both have few influential observations as well as high leverage observations. Therefore, the diagnostic plots (Figures 5-7) clearly suggest that the negative Binomial and Quasi-Poisson models are adequate compared to the Poisson model. However in Table 2, the AICs, Residual deviance, and Pearson χ^2 scores of the three models clearly reveal that the negative Binomial model provided a much better fit to the data than the other models.

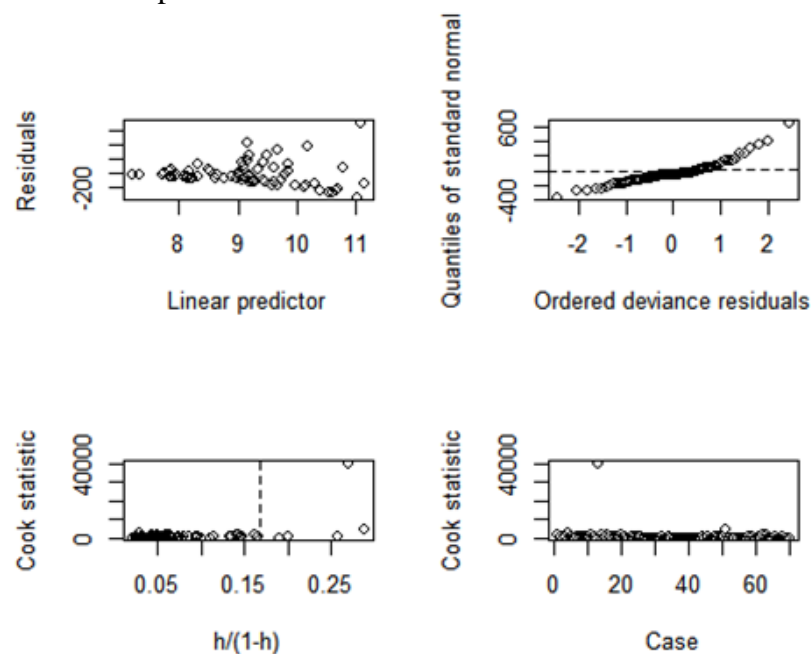


Figure 5: Diagnostic plots for the Poisson model.

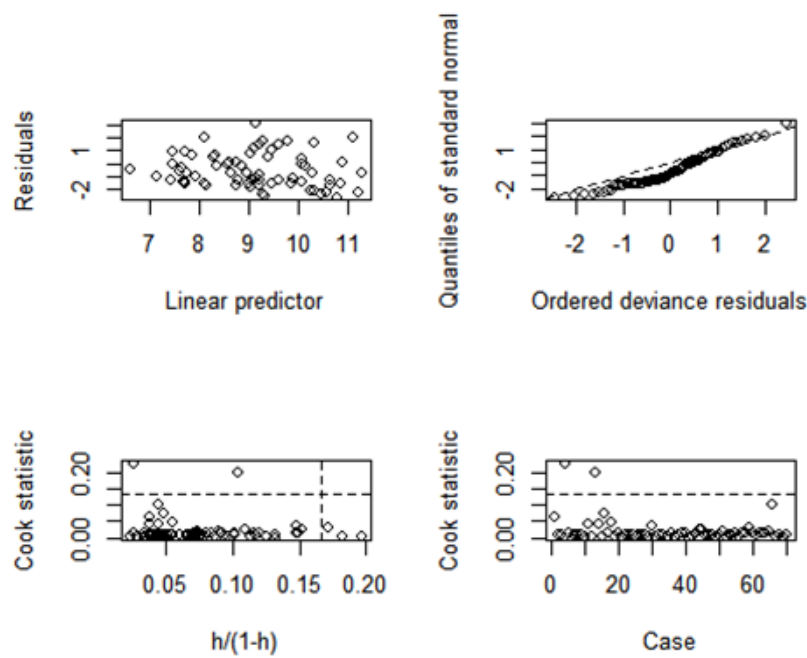


Figure 6: Diagnostic plots for the Quasi-Poisson model.

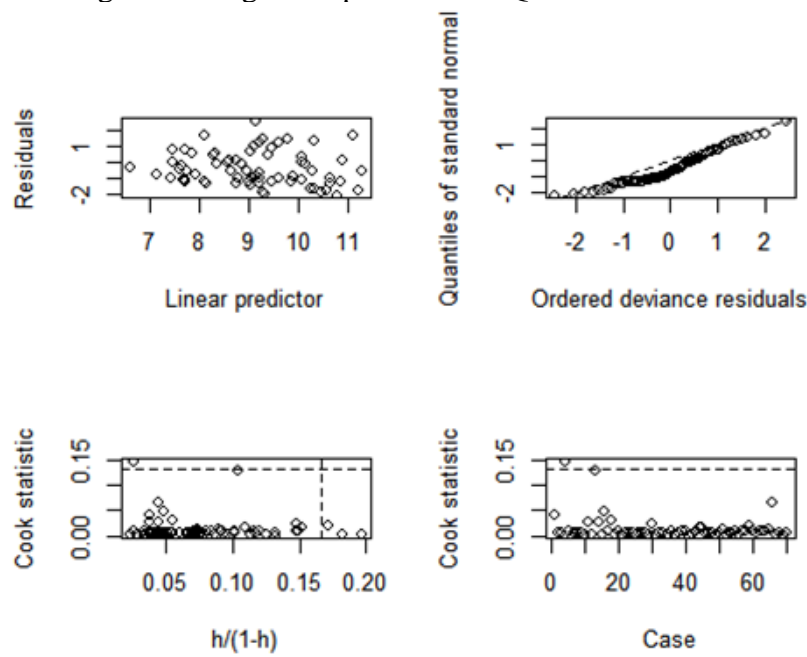


Figure 7: Diagnostic plots for the negative Binomial model.

Table 2: Model evaluation criteria

Models	AIC	Residual deviance	Pseudo R^2	Pearson χ^2
Poisson	1729646	1728935	0.3299394	2172881
Quasi-Poisson	NA	132.28	0.3398827	129.3449
Negative Binomial	1418.9	85.542	0.3399059	83.65256

Table 3 shows the associations of temperature, humidity and population density with Covid-19 infected incidence. The results show that population density (Coefficient estimate: 0.195; 95% CI: 0.0313, 0.363), and average high temperature (Coefficient estimate: -0.195; 95% CI: -0.299, -0.091) were significantly associated with Covid-19. The results also indicate that the “baseline” average Infected case count is 241305.7 (since, $\exp(12.394) = 241305.7$). We can interpret the other exponentiated coefficients multiplicatively. Our results clearly demonstrate that one unit increase in average high temperature decreases the average Infected cases by 0.82 ($\exp(-0.195) = 0.82$) times, whereas one unit increase in average low temperature increases the average Infected cases by 1.08 ($\exp(0.07804) = 1.08$) times. However, low temperature was insignificant.

Table 3: Estimates of the regression coefficients and their associated standard errors and significance after fitting negative Binomial model

Effect	Estimate	Std. Error	95% CI	z value	Pr(> z)
(Intercept)	12.394	1.214	(9.586, 15.286)	10.210	<2e-16 ***
log(PopDensity)	0.195	0.073	(0.0313, 0.363)	2.691	0.00713 **
Humidity	-0.019	0.012	(-0.046, 0.006)	-1.611	0.10720
AvgHigh	-0.195	0.048	(-0.299, -0.091)	-4.102	4.1e-05 ***
AvgLow	0.078	0.046	(-0.017, 0.172)	1.710	0.08720

* Indicates the level of significance as 0: '***'; 0.001: '**'; 0.01: '*'; 0.05: ' . ' .

Discussion:

In this study, we found from Table 3 that the population density and average high were significantly associated with the Covid-19 confirmed infected cases. Since Covid-19 is a highly contagious virus, population density can contribute to the spread of this virus. It is difficult to maintain social distance in densely populated metropolitan cities and countries with tourist attractions. New York, New Jersey, Lombardy, Hubei, Madrid and Catalonia were the epicenter of the Covid-19 due to their dense populations. New York, Lombardy, Madrid and Catalonia are the most popular tourist destinations. Every year millions of tourists visit these cities. Taking the number of tourists into account when modeling the association between population density and Covid-19 could substantially improve the performance of our models. However, we did not consider the number of tourists as an explanatory variable due to the lack of reliable data.

We observed fewer Covid-19 cases in warmer cities like Delhi and Mecca. Seasonal flu epidemics usually occur yearly during the colder months. Covid-19 is primarily spread from person to person through close contact. We can become infected from respiratory droplets when an infected person coughs, sneezes, or talks. Therefore, seasonal flu symptoms such as coughs and sneeze may contribute to the spread of the Covid-19 virus in the colder months. Since Covid-19 vaccines or effective drugs are still under development, identifying the environmental factors that intensify the spread of this virus would be helpful to design a better strategy to lower the spread as well for the future pandemic. Moreover, People from developing countries like Bangladesh may have to wait for two to three years to get the vaccine due to the tremendous demand of the vaccine. Already, Germany and the USA have signed a resolution that frontline health care workers will be vaccinated first. Motivated by this fact, we attempted to find the environmental factors that could intensify the spread of the

Covid-19 virus in our study. It is evident from the negative Binomial regression model results in Table 3 that the population density played a vital role in the spread of this deadly virus. Our results were convincing enough to infer that areas with average high temperatures are less likely to see the surge of massive Covid-19 cases. We can also infer that average low temperature could drive the spread of virus through respiratory droplets. However, our model shows there is very little or no role of humidity for the outbreak of Covid-19.

We would like to acknowledge some limitations of our study.

- The number of infected cases depends on the number of conducted tests. Most developing countries were lagging behind in their testing capacity due to the lack of testing kits.
- Air quality may have some interaction with the temperature. We did not perform any predictions due to these limitations.

In future, we will design a predictive model once we have access to more data along with accurate time series. From this study, we found the role of population density, and average high temperature on the infected cases of Covid-19, which will guide us to design risk predictions through identifying most important environmental factors.

Conclusion

This work is by far the first attempt that selects a model by comparing three statistical models to understand the spreading of Covid-19. We found that the negative Binomial provides the best fit to the data compared to Poisson and Quasi-Poisson models. Our model infers that average high temperature has the most significant role that slows down the spreading of the virus. Population density has played an important role for the spreading of Covid-19 in the 70 cities/provinces. Cities with higher population density pose extreme risk, which provides useful guidelines for policymakers and the public to control the Covid-19 pandemic.

References

- Agresti A., 2015. Foundations of linear and generalized linear models. Hoboken: Wiley.
- Casanova, L.M., Jeon, S., Rutala, W.A., Weber, D.J., & Sobsey, M.D., 2010. Effects of air temperature and relative humidity on coronavirus survival on surfaces. *Appl. Environ. Microbiol.* 76, 2712–2717. doi: 10.1128/AEM.02291-09
- Ceccarelli, M., Berretta, M., Rullo, E.V., Nunnari, G., & Cacopardo, B., 2020. Differences and similarities between Severe Acute Respiratory Syndrome (SARS)-CoronaVirus (CoV) and SARS-CoV-2. Would a rose by another name smell as sweet? *Eur. Rev. Med. Pharmacol. Sci.* 23, 7–9. doi: 10.26355/eurrev
- Chan, K.H., Peiris, J.S.M., Lam, S.Y., Poon, L.L.M., Yuen, K.Y., & Seto, W.H., 2011. The effects of temperature and relative humidity on the viability of the SARS coronavirus. *Adv. Virol.* 2011. doi:10.1155/2011/734690
- Davison, A.C. and Snell, E.J. (1991) Residuals and diagnostics. In *Statistical Theory and Modelling: In Honour of Sir David Cox* D.V. Hinkley, N. Reid, and E.J. Snell (editors), 83–106. Chapman and Hall.
- Gorbalenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, B.W., Penzar, D., Perlman, S., Poon, L.L.M., Samborskiy, D. V., Sidorov, I.A., Sola, I., & Ziebuhr, J., 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544. doi: 10.1038/s41564-020-0695-z
- Hongchao, Q., Shuang, X., Runye, S., P., W.M., Yue, C., Wei, T., Qing, S., Wenge, W., Xinyi, W., & Zhijie, Z., 2020. COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis. *Sci. Total Environ.* 183135. doi: 10.1016/j.scitotenv.2020.138778

- Howie, R., Alfa, M.J., & Coombs, K., 2008. Survival of enveloped and non-enveloped viruses on surfaces compared with other micro-organisms and impact of suboptimal disinfectant exposure. *J. Hosp. Infect.* 69, 368–376. doi: 10.1016/j.jhin.2008.04.024
- Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., & Luo, B., 2020. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Sci. Total Environ.* 724, 138226. doi: /10.1016/j.scitotenv.2020.138226
- Oliveiros, B., Caramelo, L., Ferreira, N.C., & Caramelo, F., 2020. Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases. *medRxiv*.doi: 10.1101/2020.03.05.20031872 .
- R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- Tosepu, R., Gunawan, J., Effendy, D.S., Ahmad, L.O.A.I., Lestari, H., Bahar, H., & Asfian, P., 2020. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Sci. Total Environ.* 725, 138436. doi: /10.1016/j.scitotenv.2020.138436
- Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X., & Peng, Z., 2020. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA - J. Am. Med. Assoc.* 323, 1061–1069. doi: 10.1001/jama.2020.1585
- Wang, J., Tang, K., Feng, K., & Lv, W., 2020. High Temperature and High Humidity Reduce the Transmission of COVID-19. *SSRN Electron. J.* doi:10.2139/ssrn.3551767
- Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C., & Zhang, Y.Z., 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3
- Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., & Wang, M.H., 2020. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *Int. J. Infect. Dis.* 92, 214–217. doi: 10.1016/j.ijid.2020.01.050

Funding Opportunity

Self funded

Author's contributions

All authors contributed equally