Name: Kristine Dinh
STAT 673 - Journal Article Assignment
Date: December 9, 2020

Time-Series Analysis for the Association of COVID-19 and Climate in Mainland China

Executive Summary: COVID-19 is a deathly illness that caused by a coronavirus started in late 2019. Many doctors and scientists want to find the variables that effect the infection of this virus. Knowing the variables that effect the infection of coronavirus, everyone in the world can be more aware and avoid getting infected. One of the most obvious factors is climate change over time. An analysis by Qi et al. (2020) was analyzing a time series relationship between COVID-19 and two meteorological factors like average temperature and relative humidity percentage in Mainland China. The authors included a study of COVID-19 cases in 31 Chinese provinces over the period of two months between December 2019 and February 2020. They found a negative relationship between meteorological factors and daily COVID-19 cases using a generalized additive model. In other words, if the temperature is hotter, on average, the number of COVID-19 cases will decrease. In addition, if the percentage of relative humidity are higher, then the number of COVID-19 cases will also decrease. Some significant finding that Qi et al. mentioned were a significant interaction between average temperature and average relative humidity percentage have a 95% confident interval of 0.004 to 0.07 in Hubei, one of China's provinces. This indicates that, on average, as average temperature in this province increases by 1 degree Celsius, the daily confirmed COVID-19 cases will decrease by 36% to 57% while average relative humidity is between 67% to 85.5%. In addition, on average, as average relative humidity increases by 1%, the daily COVID-19 confirmed cases will decrease by 11% to 22% while average temperature is between 5.04 and 8.2 degree Celsius. Qi et al. also mentioned that the number of changes is vary based on each province in Mainland China.

Introduction: According to the World Health Organization, COVID-19 has become a pandemic since the late 2019 and beginning of 2020. The number of daily confirmed cases has been rising exponentially. Most countries stop operating indoor businesses, large group gathering, and unnecessary traveling. Since this is such a significant issue in the world right now that might cause the economy to be in regression, scientists and researchers would ask many questions in favor of this issue. One of the most obvious questions would be examining the relationship between COVID-19 and meteorological factors including temperature and relative humidity over time. With the answers to this question, health care professionals would expect to have more patients in certain seasons compared to others. This way, they can prepare accordingly at peak season to have enough supplies for the patients. In addition, normal citizens can have a better understanding of which time of the year they would higher risker of getting infected with coronavirus. This way, they can stay home to stop the spread of coronavirus. The purpose of Qi et al.'s analysis was to prove the association of climatological factors and COVID-19, scientifically. With the historical data of daily confirmed cases in Mainland China, the authors have signified the correlation between the frequency of daily confirmed cases and meteorological factors including temperature and humidity.

Methodology: All methodologies and results are from the study by Qi et al. called *COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis*. The COVID-19 confirmed cases data were collected from the National Health Commission of People's Republic of China from December 1, 2019 to February 11, 2020. The date of the data ranged different depending on each province in Mainland China. The climate change data were from the Weather Underground. This dataset included meteorological variables including the average temperature and average relative humidity of each province in China. Using the two data sets given above, a generalized additive model (GAM) was used for the model development process. Qi et al. considered a model with 14-day exponential moving average function with lag effects of the interaction between average temperature and average relative humidity. The daily count of confirmed COVID-19 cases was also transformed using log-transformation to have a more stationary time series process and smaller values to work with. After the model development process, the authors also validated the model using the same time series model from different provinces including Wuhan and Hubei. All analysis in their paper were done in R with version 3.5.3.

Results/Discussion: Before building the model, it is important to assess the relationship between the response variable and each of the predictors. Climate changing over time could be a huge factor in changing the count of COVID-19 confirmed cases. It was found that the daily confirmed cases increased dramatically on January 20, 2020 in Hubei province when the temperature was slightly falling from December 2019 to January 2020. In addition, the percentage of average relative humidity was between 42.17% to 96.92%. Other provinces had an average humidity rate of larger than 88% on January 23, 2020 and February 8, 2020. In other words, the number of cases generally increased as the average temperature or average relative humidity decreased. The daily count of COVID-19 confirmed cases by each climate factor are visualized in Figure 1. As one can see in Figure 1a, the log-transformation of the daily confirmed cases is decreasing as the relative humidity percentage from the 25th, 50th and 75th percentile increase. In addition, the number log-transformed daily confirmed cases also decrease as the average temperature in the 25th, 50th, and 75th percentile increase. This suggests that these two variables can contribute significant impact to the fluctuation of daily COVID-19 confirmed cases.

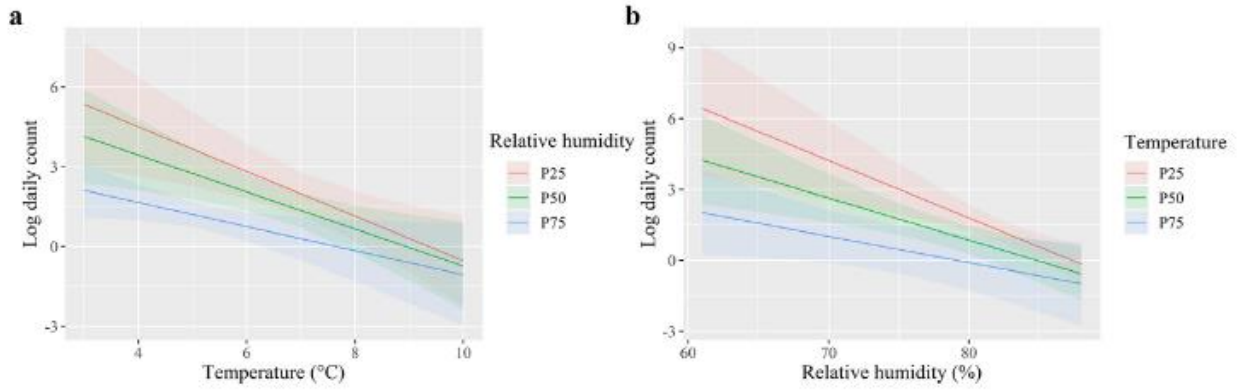

Figure 1. The correlation between log-transformed daily confirmed cases vs meteorological factors including average temperature and average relative humidity by percentiles.

Now that we have explored the variables and found two significant variables that could impact the daily confirmed cases, we can proceed to the model development process. Qi et al. used a GAM to demonstrate the relationship between the target variable (daily confirmed cases) and predictors (average temperature and average relative humidity). Table 1 shows the summary of the regression of the GAM. This table depicts that all variables, except β_5, in the model are significant with small p-values of less than 0.05. In addition, the interaction between average temperature and average relative humidity percentage can also be classified as significant with a p-value of 0.029. Furthermore, the confident interval for all coefficient estimates are clearly saying that the coefficient estimates are significant since there is no zero included in the intervals, except one coefficient for time series component, β_5.

Table 1. *Summary regression of the GAM model for Hubei province. The columns represent parameter of the features, 95% confident interval of coefficient estimates, Z test statistics, and significant p-values, respectively from left to right.*

Parameter	Estimate (95% CI)	Z-value	p-Value
Intercept, β_0	38.14 (14.54–61.74)	3.17	0.002
AT, β_1	−3.61 (−6.46 to −0.75)	−2.47	0.013
ARH, β_2	−0.43 (−0.7 to −0.16)	−3.16	0.002
AT ×ARH, β_3	0.04 (0.004–0.07)	2.18	0.029
Baidu index, β_4	3.51×10^{-5} (1.96×10^{-5}, 5.06×10^{-5})	4.43	<0.001
Time			
β_5	3.42 (−0.9, 7.74)	1.55	0.120
β_6	4.96 (2.87–7.04)	4.66	<0.001

Now that the model has been developed, some of the coefficients for the GAM will be interpreted. The correlation between daily COVID-19 confirmed cases and climate factor have been found to be moderately negative. Looking at Table 1, the interaction between average temperature and average relative humidity percentage have a coefficient of 0.04 and a 95% confident interval of 0.004 and 0.07. This indicates that if there is an increase in average temperature, the effect of average relative humidity on the daily confirmed cases will decrease at Hubei province. Intuitively, Qi et al. believed that temperature changes could make people's nasal mucosa prone to small ruptures. Therefore, this could create opportunities for the coronavirus to enter the body easier.

In terms of average temperature, Table 1 shows the relationship between temperature and daily confirmed COVID-19 cases to be negative with an estimated coefficient of -3.61 and a 95% confident interval between -6.46 to -0.75. This indicated that as temperature increases by 1 degree Celsius, the number of confirmed cases will decrease on an average of 4 log-transformed cases. In addition, average relative humidity percentage also found to have a negative relationship with daily confirmed COVID-19 cases. This indicates that as the percentage of relative humidity increases by 1%, the number of confirmed cases will decrease, on average, by 1 log-transformed case. Likewise, the temporal components of this model are also significant. As seen in Table 1, β_5 and β_6 are the regression coefficients of natural splines of time with two degrees of freedom. These two coefficient estimates have a positive relationship with confirmed cases. Therefore, this signifies that as time passed by, the number of daily COVID-19 cases would increase.

After the model was finished building, we need to validate the model. In the article by Qi et al., the model was validated by providing the predictors for two provinces including Wuhan and Hubei in China using the same GAM. The authors intended to see if the model could predict the number of confirmed cases over time for different provinces from different periods of time. However, the predictions were biased based on time series. Since this model was only built using data from two-month periods from December 2020 to February 2020. However, the results came out to be as expected. Predicted number of COVID-19 cases in province Wuhan and Hubei seem to be decreasing as average temperature and average relative humidity percentage increase. Therefore, this created a negative relationship between the target variable and the meteorological factors, while the interaction between average temperature and average humidity percentage is positive.

Despite the biased predictions of the model while outside the spatial and temporal ranges, the model was still considered decent to represent the descriptive relationship of daily confirmed COVID-19 cases and meteorological factors including average temperature and average relative humidity percentage. The final model was presented as the following equation:

$$\log(\mu_t) = \beta_0 + \beta_1 matemp_t \times \beta_2 mahmd_t + \beta_3 matemp_t \times \beta_4 mabindex_t + \beta_5 \times ns_1 + \beta_6 \times ns_2,$$

where μ_t is the expected values of daily confirmed COVID-19 cases, β_0 is the intercept, β_1 is the effect of moving average of average temperature, β_2 is the effect of moving average of average relative humidity percentage, β_3 is the interaction between average temperature and average relative humidity percentage, β_4 is the effect of moving average of Baidu index from the weather data, and β_5 and β_6 are the coefficient of the splines of time. The estimated values of beta coefficients can be found in Table 1.

Conclusions

Since COVID-19 became the center topic in 2020 and a pandemic in the world, scientists were seeking for solution to help decrease the daily confirmed cases. Two automatic factors that came into mind were meteorological and time series factors that could affect the daily confirmed cases. Qi et al. had proved in their analysis that there is a negative correlation between the number of confirmed COVID-19 cases and meteorological factors including average daily temperature and average daily relative humidity percentage of each province, especially Hubei. This indicated that we should expect less cases in winter seasons and more COVID-19 cases in summer seasons. With this information, health care professionals would have more idea of when to expect more patients at the hospital to prepare enough supplies to treat all patients. In addition, normal citizens would also have a better idea of their risk of getting infected with coronavirus depending on climate. They would make choices of staying home when the summer comes to avoid getting infected. On the other hand, the interaction between average temperature and average relative humidity percentage was found to be positive. This means that the effect of average relative humidity on expected cases would decreases as the average temperature increases. In addition, time also had an important effect in this model. It was proven that time have a positive effect on the number of cases. Therefore, we would expect the daily confirmed COVID-19 cases to increase as time passes by.

There are limitations to this analysis. Qi et al. mentioned in their paper that there might be more risk variables that would contribute to rise of confirmed COVID-19 cases including social economic status. For example, a low-income family who are living in a crowded apartment would have higher risk of getting infected. Qi et al. also discussed how the analysis was conducted in a short period of time and the weather data was only collected from the capital city of each province. This led to an inaccuracy for other cities in the province since each province have a large area in China. In addition, this study was only been done on the months of December 2019 to February 2020 and specifically in Hubei province. If we want to project the model to predict daily confirmed cases from different time or different province, the information might be unstable. To further study this project, the authors would study the meteorological factors on COVID-19 more in depth. In addition, Qi et al. also mentioned that they would study the spatial and temporal component further.

Bibliography

Qi, Hongchao, et al. “COVID-19 Transmission in Mainland China Is Associated with Temperature and Humidity: A Time-Series Analysis.” *Science of The Total Environment*, vol. 728, 2020, p. 138778., doi:10.1016/j.scitotenv.2020.138778.

WHO 2020. WHO characterizes COVID-19 as a pandemic.