Correlation between GDP and PM2.5 Change During Covid-19

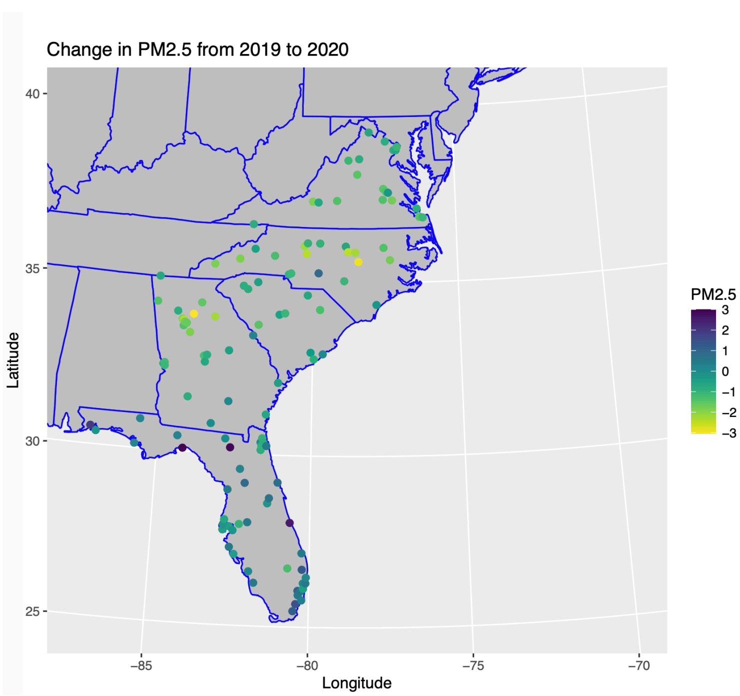
Hongjian Yang

1. Introduction

Covid-19 has impacted everyone's life. Students start to take classes online, employees start to work from home, and people refuse to leave their home. As a result, traffic is much less than normal, and it is expected that our air quality will improve slightly since car exhaust has been a source of air contamination. PM2.5, a pollutant in the air, is emitted mostly from car exhaust. When people drive less, PM2.5 concentration is supposed to decrease.

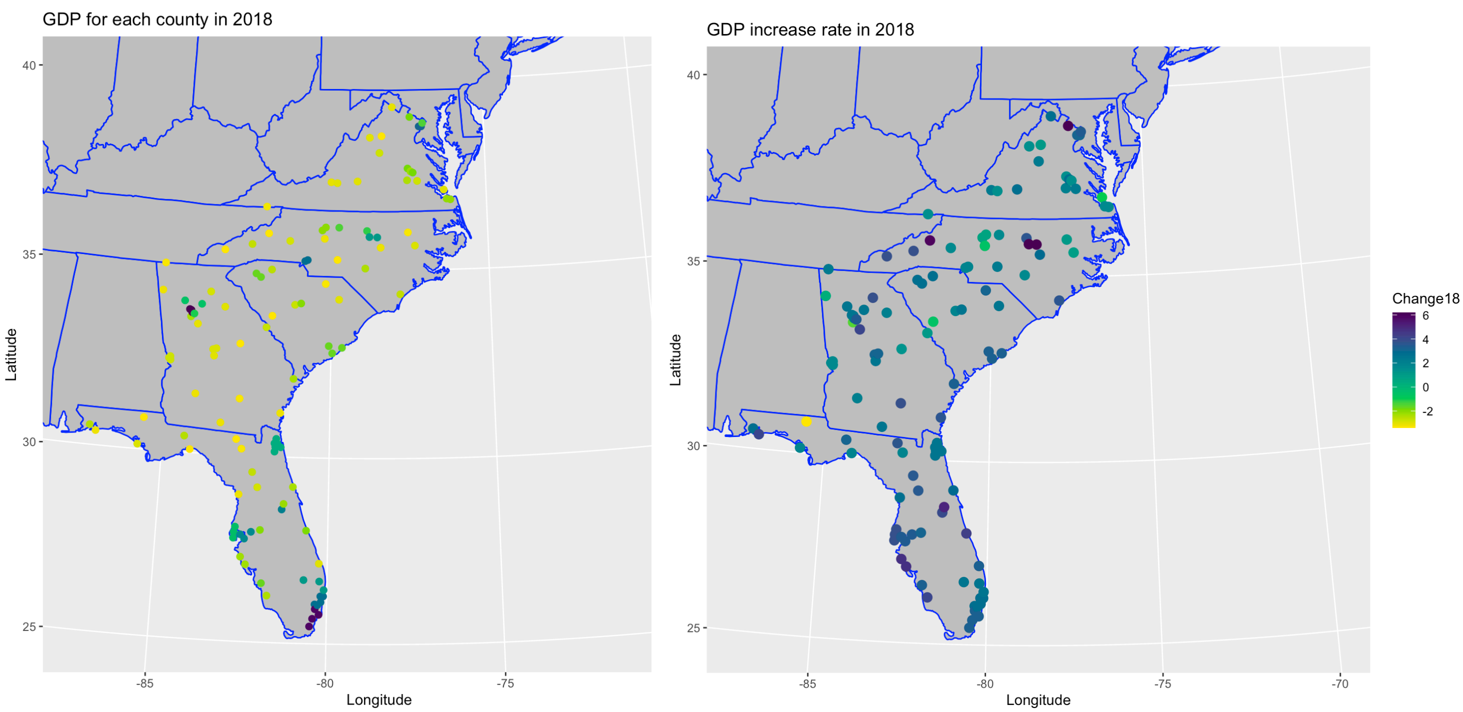
However, PM2.5 concentration is not likely to decrease by the same level in all areas. Major cities, like New York where GDP is usually very high, might see a higher degree of decrease since human activity was more intense during normal time. Small towns, like Apex where GDP is often low, might experience roughly the same level of PM2.5 concentration. This report will examine the correlation between GDP and amount of change in PM2.5 concentration to better understand how Covid-19 has impacted the air quality.

2. Data Description



The air quality data, as shown in the figure to the left, come from EPA website at **www.epa.gov/outdoor-air-quality-data/download-daily-data**. This report uses data from North Carolina, South Carolina, Virginia, Georgia, and Florida from April to June in 2019 and 2020, respectively. To eliminate anomaly observations, this report only focuses on observations sites where at least 10 observations are recorded from April to June in both years.

The county-level GDP data comes from Bureau of Economic Analysis at **www.bea.gov/system/files/2019-12/lagdp1219.pdf**. Only 2018 GDP information is used since 2019 GDP has not been published yet. This report uses GDP and GDP change rate in the southeastern five states described above. Plots of GDP and GDP change rate distributions are shown below. The GDP has been normalized with a mean of 3 to prepare for model fitting.



3. Methods

Since air quality data often has a spatial pattern, the proposed models include longitude and latitude. The model will also include GDP and GDP change rate for each site as two covariates.

The basic model proposed in this report is exponential model with nugget effect:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + Z_i + \epsilon_i$$

where β_0 is the intercept, X_1, X_2 are longitude and latitude, and X_3, X_4 are GDP and GDP change rate. For the exponential model, the covariance function is defined as:

$$Cov(Z_i, Z_j) = \sigma^2 \rho(Z_i, Z_j)$$

and
$$\rho(Z_i, Z_j) = exp(-d/\phi)$$

This report will compare the proposed model with three other models: exponential model without nugget, Matern model with nugget, and Matern model without nugget, where Matern covariance function is defined as:

$$\rho(Z_i, Z_j) = \frac{2^(1-\nu)}{\Gamma(\nu)}(\sqrt{2\nu}\frac{d}{\phi})^{\nu} K_{\nu}(\sqrt{2\nu}\frac{d}{\nu})$$

and d is the distance between location Z_i and Z_j.

To compare the performance of different models, three metrics, AIC, BIC, and MSE will be reported, where

AIC = $-2log\{L(\hat{\beta}, L(\hat{\theta})\} + 2k$

BIC = $-2log\{L(\hat{\beta}, L(\hat{\theta})\} + log(n)k$

and k is the number of parameters in (β, θ).

4. Model Comparison

Here are the summary of the four models described above:

	MSE	MAD	AIC	BIC
Exponential With Nugget	**0.8**	**0.59**	**228.6**	**248.32**
Exponential Without Nugget	**0.73**	**0.63**	**228.3**	**245.56**
Matern With Nugget	**0.77**	**0.64**	**230.54**	**252.73**

	MSE	MAD	AIC	BIC
Matern Without Nugget	**0.77**	**0.64**	**228.54**	**248.27**

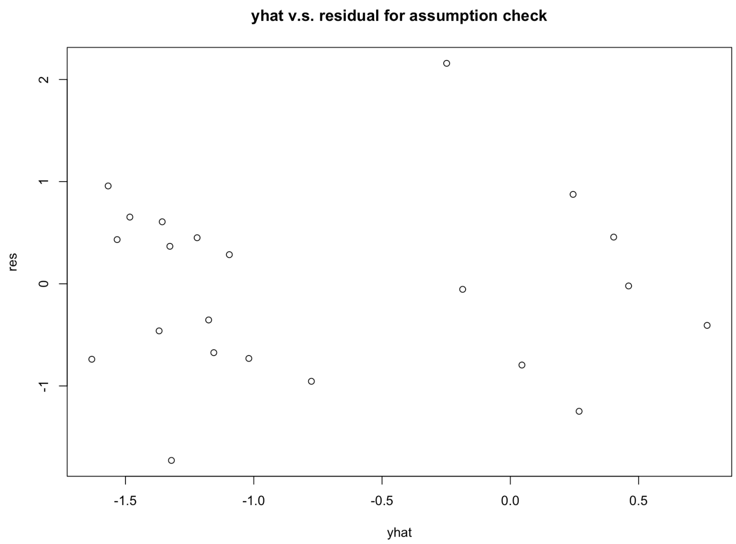
This comparison table tells us that for both the exponential and Matern models, the performance is better when there is no nugget effect term. Compare all four models, exponential without nugget model achieves the best performance with lowest MSE and lowest AIC and BIC.

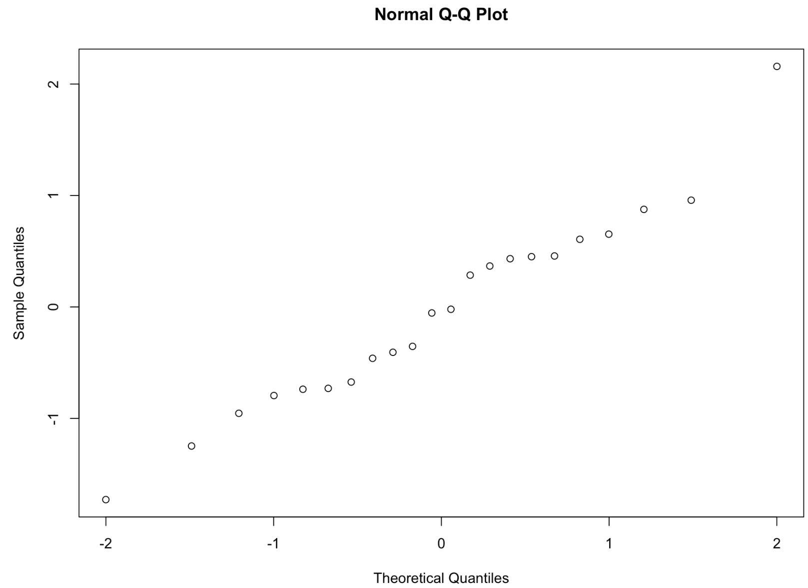
The final model predicts the PM2.5 concentration change pretty well, with a mean squared error of 0.73.

5. Model Checking

Since the final model is exponential without a nugget effect, the main assumption is that the spatial dependence is isotropic. This assumption is generally met by looking at the plot: No special pattern is recognized. Secondly, there is no nugget effect in the model. It means almost no localized random error under this model.

Three other assumptions are also checked for linear regression. Firstly, the expected value of residuals is around 0. After calculation, the expected value is -0.042; Secondly, the variance of error is constant which is met by plotting the residuals with respect to the predicted values, as shown below. Lastly, the residuals is roughly normal which is checked by plotting the quantile-quantile plot as shown below.





6. Summarize the Final Model

The parameters from the final model are as follows:

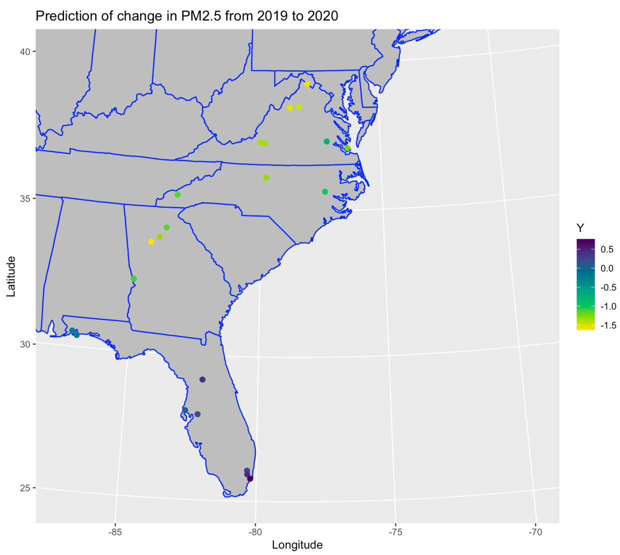
$$\beta_0 = 10.69, \beta_1 = 0.042, \beta_(2) = -0.218, \beta_3 = -0.159, \beta_4 = -0.003$$

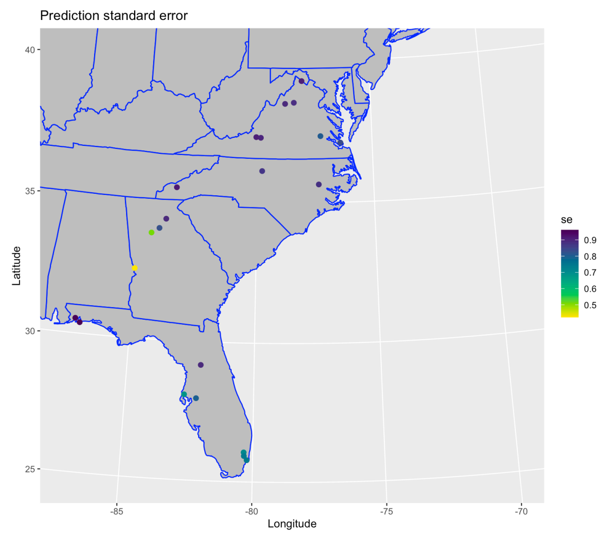
$$\tau^2 = 0, \sigma^2 = 0.83, \phi = 1.01$$

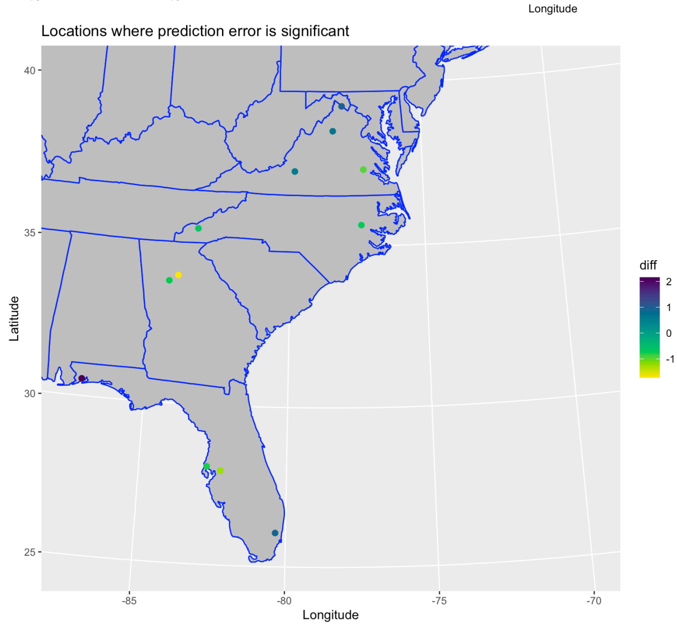
At location (0,0) and a GDP of 0 (a relatively poor area), the change of PM2.5 concentration will increase by 10.69 from 2019 to 2020. The change in PM2.5 increase by 0.042 unit when we move from west cost to east coast on the same latitude, while it decreases by 0.218 unit when we move from south to north on the same longitude. If GDP increases by 1 unit, PM2.5 decrease by 0.159 unit. If GDP change rate increase by 1 unit, PM2.5 decreases by 0.003 unit. The total variation is about 0.83 unit and beyond a distance of 1.01 unit, the data is no longer correlated.

7. Spatial Prediction

Because GDP data is not available at every location, I am only able to use about 100 data points for training and prediction. Plots below are prediction, standard error of prediction, and locations where different is significant.







6. Conclusion

From the result of this paper, we can conclude that Covid-19 has more significantly impacted areas where economical activity is high and GDP is high. For high-GDP areas, the decrease in PM2.5 is higher. It is the same for area with high GDP growth rate.