



Article

Association between the New COVID-19 Cases and Air Pollution with Meteorological Elements in Nine Counties of New York State

Carlos Díaz-Avalos ¹, Pablo Juan ², Somnath Chaudhuri ³, Marc Sáez ^{4,5} and Laura Serra ^{4,5,*}

¹ Department of Probability and Statistics, IIMAS, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico; zhangkalo@gmail.com

² Department of Mathematics and IMAC, Universitat Jaume I, Castellón, 12006 Castellón, Spain; juan@mat.uji.es

³ Department of Mathematics, Universitat Jaume I, 12006 Castellón, Spain; al383341@uji.es

⁴ Research Group on Statistics, Econometrics and Health (GRECS), University of Girona, 17003 Girona, Spain; marc.saez@udg.edu

⁵ CIBER of Epidemiology and Public Health (CIBERESP), 28029 Madrid, Spain

* Correspondence: laura.serra@udg.edu

Received: 27 October 2020; Accepted: 2 December 2020; Published: 4 December 2020



Abstract: The principal objective of this article is to assess the possible association between the number of COVID-19 infected cases and the concentrations of fine particulate matter (PM_{2.5}) and ozone (O₃), atmospheric pollutants related to people's mobility in urban areas, taking also into account the effect of meteorological conditions. We fit a generalized linear mixed model which includes spatial and temporal terms in order to detect the effect of the meteorological elements and COVID-19 infected cases on the pollutant concentrations. We consider nine counties of the state of New York which registered the highest number of COVID-19 infected cases. We implemented a Bayesian method using integrated nested Laplace approximation (INLA) with a stochastic partial differential equation (SPDE). The results emphasize that all the components used in designing the model contribute to improving the predicted values and can be included in designing similar real-world data (RWD) models. We found only a weak association between PM_{2.5} and ozone concentrations with COVID-19 infected cases. Records of COVID-19 infected cases and other covariates data from March to May 2020 were collected from electronic health records (EHRs) and standard RWD sources.

Keywords: COVID-19; INLA; RWD; PM_{2.5}; O₃; New York

1. Introduction

In recent months, a series of studies related to the pandemic caused by the new SARS-CoV-2 pathogen have been conducted around the world. Most of those studies have focused on epidemiological and molecular biology aspects aiming to learn as fast as possible about the characteristics of the virus and its epidemic spread. One aspect of the pandemic that has been mentioned in the press and other media is the effect that epidemic control measures like nationwide lockdown and the associated reduction in human activities had on environmental issues such as air pollution. There are several published studies regarding the effect that reduction of human activities due to the current pandemic has had on air pollution levels. A recent study conducted during two distinct phases of lockdown over the Yangtze River Delta Region applied Weather Research and Forecasting and Comprehensive Air Quality Model with Extensions (WRF-CAMx) modeling system with monitoring data to investigate the impact of human activity pattern changes on air quality [1]. The result shows a notable reduction in the

concentrations of PM_{2.5}, NO₂, and SO₂ during Level I and Level II lockdown periods compared with 2019. However, ozone did not show any indicative reduction. Similar research work with air quality data (2017–2020) in 22 cities of India analyzes the effect of restricted human mobility and reports 43%, 31%, 10%, and 18% decreases in the concentration of PM_{2.5}, PM₁₀, CO, and NO₂ in the study cities during lockdown period compared to previous years [2]. In the same note, Zambrano-Monserrate et al. show significant association between contingency measures and improvement in air quality revolve around the reduction of PM_{2.5} and NO₂ concentrations in China, France, Germany, Spain, and Italy during the lockdown phases in the study countries [3]. Another interesting study analyzed recent data released by National Aeronautics and Space Administration (NASA) and European Space Agency (ESA) and indicated that concentrations of NO₂ in some of the epicenters of COVID-19 such as Wuhan, Italy, Spain, and USA has reduced up to 30% during the different stages of nationwide lockdowns [4].

As air pollution is related to climatic and human factors, modeling the dynamics of their concentration in space and in time is a complex task. However, with a significant reduction in human activities in many countries, the presence of the COVID-19 pandemic provides a unique opportunity to analyze, at least partially, the effect of human activities on air pollution levels. Studies focusing on the effect of restricted human activities due to the COVID-19 pandemic on air quality show comparisons of air pollution levels for a series of days in 2019 and compare with pollution levels for the same days in 2020 [5]. Selvam et al. in a recent study [6] compared pollution levels based on time trends along 2020 before and after the lockdown. In the same context, Adams in his study [7] compared pollution levels during the weeks of lockdown with pollution levels observed in the same time period in 2019. Goldberg [8] showed the existence of drops in NO₂ levels in several US cities using transport and satellite data. Krecl et al. in their recent study [9] have reported similar findings for the city of Sao Paulo, and identical patterns were found by Wang [10] in six sites in China. Some studies focused on the spatiotemporal dynamics of the disease including environmental effects, socioeconomic factors, health service resources, and demographic conditions that vary from different counties in the United States [11].

The social behavior and working habits of people show spatial variation although they are difficult to measure. To assess adequately the changes in pollution levels due to the COVID-19 pandemic in the study region, it is essential to have access to reliable data sources. This leads us to rely on real-world data (RWD) that comes from diverse sources other than traditional randomized controlled trials (RCTs) [12]. In a recent report by the U.S. Food and Drug Administration (FDA) on Real-World Evidence (RWE) program under Section 505F(b) FD&C Act, RWD is defined as the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources (Cornell Law School web (<https://www.law.cornell.edu/uscode/text/21/355>)). After the 21st-Century Cures Act, passed in 2016, the health care community with the guidelines from the FDA is using RWD to develop guidelines and decision support tools for use in clinical practice and healthcare issues. RWD can be obtained from observational studies or from existing digital sources such as healthcare databases including electronic health records, pharmacies, and health insurance databases, etc. Data gathered from product and disease registries in mobile devices, patient-generated data including in home-use settings or, internet-based websites and applications that enable users to create and share in social networking mediums also considered as RWD and facilitates modern RWE research [13]. In the current study, we emphasized on the EHRs that contribute as an important source of RWD. We used the number of confirmed COVID-19 cases reported in nine counties in the state of New York, USA. Details of the RWD data source and concerned stakeholders are reported in Section 2.1.

According to a current COVID-19 pandemic report by the Centers for Disease Control and Prevention (CDC), USA, New York is the worst affected state in the United States of America [14]. Out of 62 counties of New York, counties having major cities with a high population are being affected the most [15]. In connection to this, the present study was conducted in the southern-most nine counties of New York namely, Bronx, Kings, Nassau, New York, Queens, Richmond, Rockland, Suffolk, and Westchester as shown in Figure 1. These counties are all in the vicinity of the New York city and

have reported the highest number of deaths and confirmed cases of the virus during the COVID-19 pandemic [16].

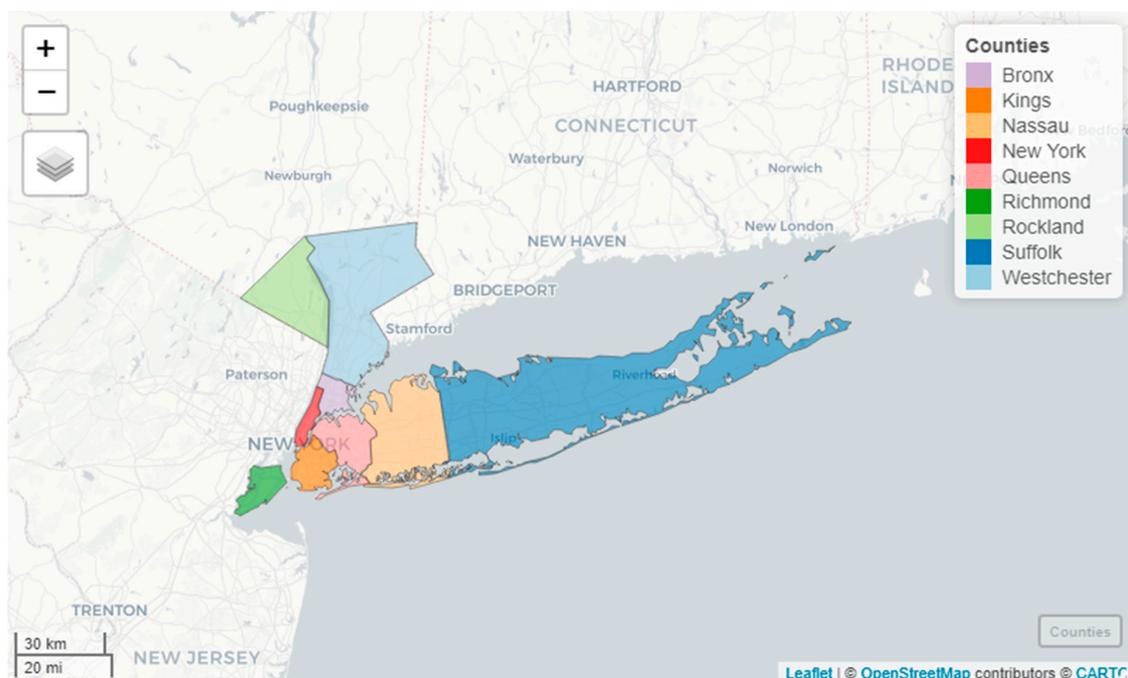


Figure 1. Study area (nine counties in the southernmost region of New York State).

Air pollution levels are suspected to be associated with the number of COVID-19 infected cases. Some studies claimed that prolonged exposure to air pollution may increase the vulnerability and mortality rates due to COVID-19, although the relative role of air pollution and aerosols in the spread of the virus and mortality rates is still under debate [17,18]. On the other hand, in most countries, once the number of cases has passed a given threshold value, social changes such as nationwide lockdown, work from home have resulted in a significant decrease in the traffic and transport needs for people, thus reducing automotive traffic as well as industrial pollution. The inclusion of time in the analysis has also been considered due to the time correlation in the data used in our analysis [19].

In this paper, we present the results of the analysis of the association between the number of COVID-19 cases with some variables related to meteorology and air pollution. Our analysis is based on fitting generalized linear mixed models (GLMM) using the air pollution variables ($PM_{2.5}$ and O_3) where the number of COVID-19 infected cases and meteorological variables are used as covariates for nine counties in the state of New York currently affected by the SARS-CoV-2 pandemic. Like other studies of this nature, we used the integrated nested Laplace approximation (INLA) methodology [20,21] and compared the results before and after the COVID-19 pandemic [22,23]. Most of the studies related to COVID-19 incidence and air pollution are based on comparisons between pollution levels the weeks before lockdown and during the lockdown. Others compare the air pollution levels during the COVID-19 lockdown with pollution levels the same days of lockdown but in previous years. Although such studies reported drops in pollution levels during the COVID-19 pandemic, their results cannot tell whether such reductions are the result of a long time trend due to pollution control measures or if they are the result of local climatic conditions during the COVID-19 lockdown in the city under study. In this context, it is noteworthy to mention about a recent study [24] analyzing the effects of the lockdown on air pollution in Athens, Greece during pre-, post-, and lockdown periods, as well as with respect to previous years in conjunction with meteorological parameters. Another interesting research work studied the ambiguous phenomena of severe haze pollution that occurred in northern China during the lockdown despite the abrupt decreases in gaseous emissions caused by record-low

anthropogenic activities. The study shows in context to surface meteorology, that the severe air pollution episode over the study area coincided with the abnormally low planetary boundary layer (PBL) height, which triggered strong aerosol–PBL interactions [25].

In the current study, we present the results of the analysis of the association between air pollution levels as a response variable, with COVID-19 incidence and several meteorological factors as explanatory covariates. We fit a generalized linear mixed model (GLMM) which includes a spatial and a temporal term. The structure of this model allows to filter out the effect of meteorological variables as well as spatial variation and local time trends from the association between air pollution and COVID-19 incidence, providing a statistically more powerful test to the hypothesis of no association between the response and the covariates in our study.

Our goal is to assess the effect of changes in traffic density and industrial activities on air pollution measurements after filtering the effect of meteorological variables. We use the number of COVID-19 cases as a proxy for a reduction in automotive traffic and a decrease in industrial activities. Some recent studies [26,27] have shown a strong association ($R^2 > 0.89$) between COVID-19 incidence and mobility. These authors showed that the higher the mobility the higher the number of contagions. Thus, due to the unavailability of mobility data for our study, we found it reasonable to use the number of new COVID-19 cases as a proxy for human mobility. Our analysis includes geographic space as a random effect [28].

2. Methods

2.1. Data Settings

The meteorological data for the study site were retrieved from the online Applied Climate Information System (ACIS) developed and maintained by the National Oceanic and Atmospheric Administration (NOAA) Regional Climate Centers (RCCs), USA [29]. ACIS is a joint project of the RCCs, the National Centers for Environmental Information (NCEI), and the National Weather Service. The accessed data set is based on the average (median) of several stations in the counties. It provides average daily records of five meteorological components like dew point (in degree Fahrenheit), humidity (in percent), precipitation (in inches), temperature (in degree Fahrenheit), and wind speed (in miles per hour).

The meteorological variables show variations per day as expected, and except for the temperature, there is no apparent trend in the time window of this study, which suggests that they may be assumed to be second order stationary. Figures 2 and 3 show the temporal variation for the dew point and daily average temperature. Similar plots for all the current study counties are reported in the Appendix A, Figures A1–A3. The air quality dataset for the study area was collected from the open-portal Air Quality System (AQS) of the Environmental Protection Agency (EPA) of the United States of America [30]. Both air quality and meteorological data set are from 2 March 2020 to 2 May 2020. Except for Nassau County, the other counties have daily average records of two major air quality components such as particulate matter, fine particles with an aerodynamic diameter of 2.5 μm or less ($\text{PM}_{2.5}$) in $\mu\text{g}/\text{m}^3$, and ozone (O_3) in mg/m^3 . O_3 records were not available for Nassau County. Figures 4 and 5 depict the temporal variation of the two air quality parameters expressed in logarithmic values in each county of the study region.

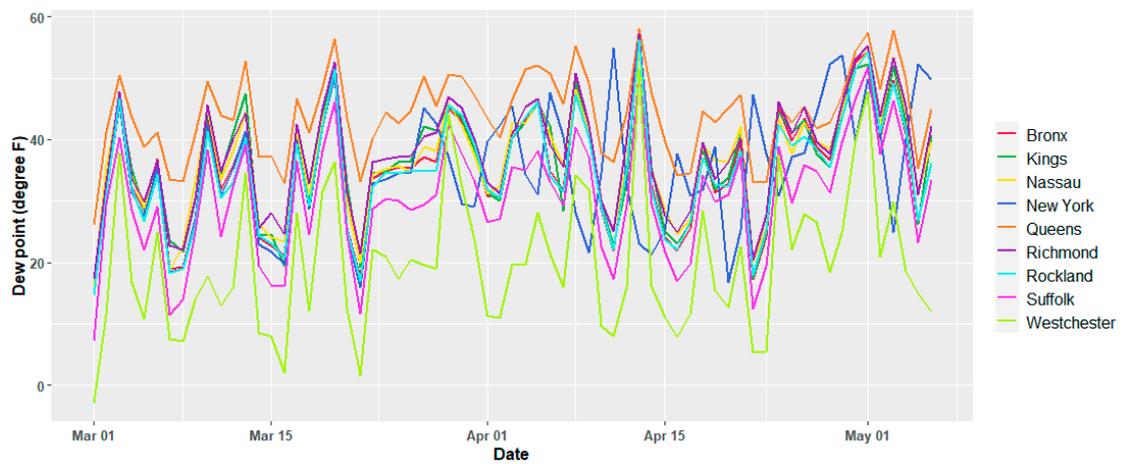


Figure 2. Variation of daily average dew point for individual counties.

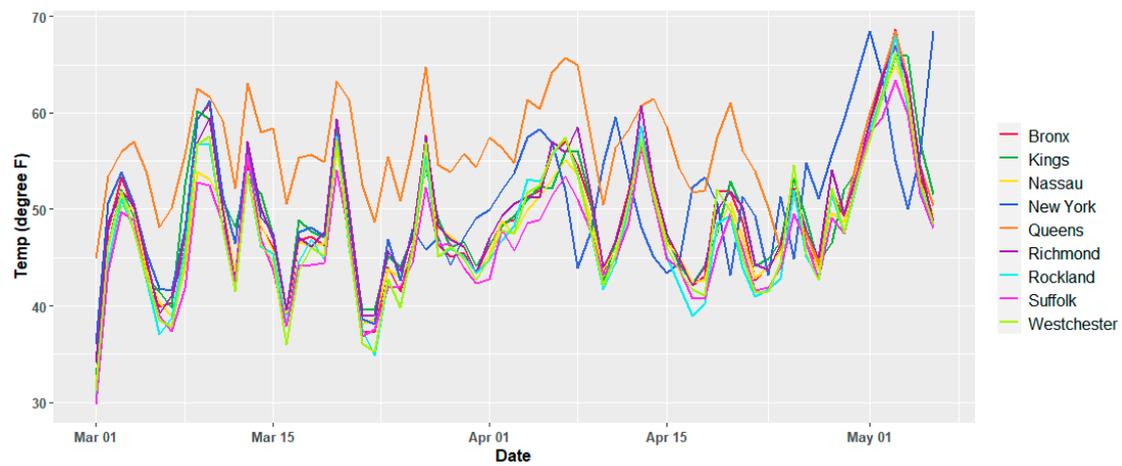


Figure 3. Variation of daily average temperature for individual counties.

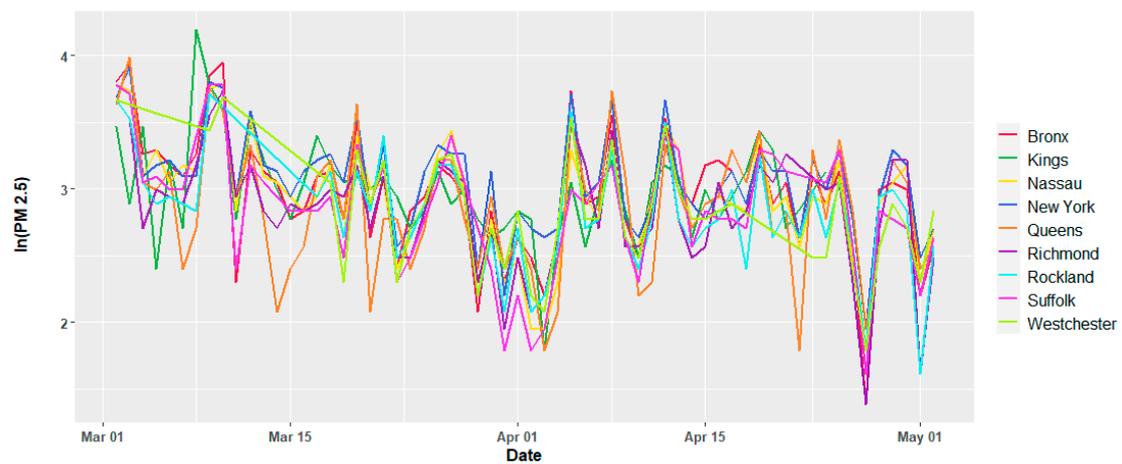


Figure 4. Variation of the daily average concentration of particulate matter $PM_{2.5}$ (in $\mu g/m^3$) in the natural logarithmic scale for individual counties.

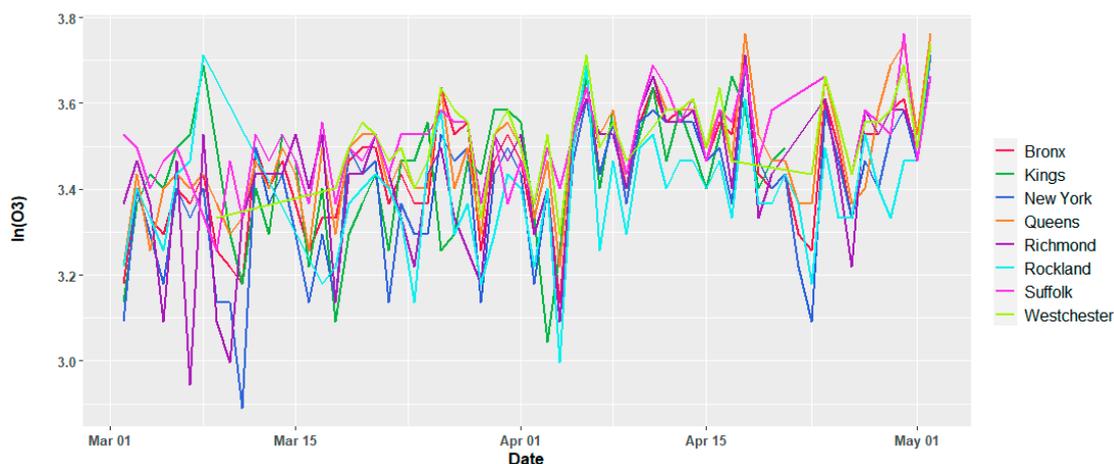


Figure 5. Variation of the daily average concentration of ozone (in mg/m^3) in natural logarithmic scale for individual counties.

In context to the EHRs data source for RWD, the number of COVID-19 daily infected records were retrieved from the official website of the Department of Health, New York state [31], and verified from the open data repository of CDC [32]. Data for infected cases for all the counties were available from 2 March 2020 to 2 May 2020 as reported in Figure 6 for individual counties. All datasets used in the current study were collected from sources without restrictions and that have open access.

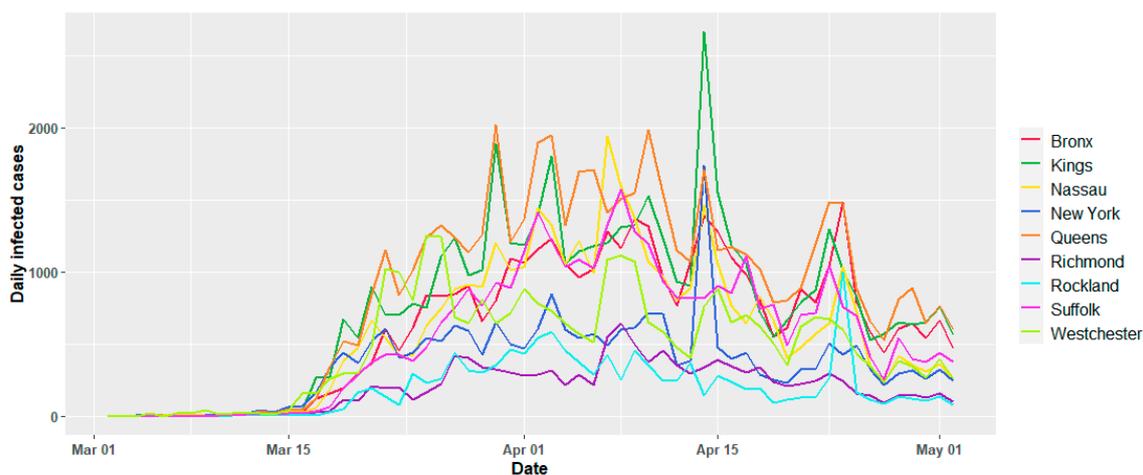


Figure 6. Variation of daily confirmed COVID-19 infected cases for individual counties.

2.2. Statistical Methodology

Hierarchical Mixed Linear Model for COVID-19 Mortality

For Bayesian analysis, the association between $\text{PM}_{2.5}$ and ozone with the number of new cases of COVID-19, we considered a set of different explanatory variables, besides the number of cases of COVID-19, namely temperature, dew point, humidity, wind speed, and precipitation. The environmental covariates are well known to affect air pollutant concentrations as well as human activities since for example, blizzards tend to make people stay at home and reduce automotive traffic. Our goal is not to compare the pollutant concentrations before and during the COVID-19 pandemic but to evaluate whether or not the observed changes in $\text{PM}_{2.5}$ and ozone are directly linked to mobility reductions in human mobility in the study area, using the number of new COVID-19 cases as a proxy for human mobility.

In our context, spatiotemporal data can be idealized as realizations of a stochastic process indexed by a space and a time dimension.

$$Y(s, t) \equiv \{y(s, t) | (s, t) \in D \times T \in R^2 \times R\} \quad (1)$$

where D is a (fixed) subset of R^2 and T is a subset of R . The data can then be represented by a collection of observations $y = \{y(s_1, t_1), \dots, y(s_n, t_n)\}$, where the set (s_1, \dots, s_n) indicates the spatial locations where the measurements are recorded, and (t_1, \dots, t_n) are the corresponding time points. In this study the data came aggregated on a daily time basis, so t represents days. Our aim is to study the association between air pollution components (PM_{2.5}, O₃) as the response variables with the number of cases of COVID-19 and some meteorological factors as covariates, using hierarchical Bayesian analysis. We assumed that $Y(st)$ follows a log-normal distribution, this is, $\ln[Y(s, t)] \sim N(\mu_{s,t}, \sigma^2)$. For the hierarchical modeling process, we use the canonical link function,

$$\eta_{s,t} = g(\mu_{s,t}) = \ln[Y(s, t)] \quad (2)$$

We assume a linear association of the response variable with the covariates of the form,

$$\eta_{st} = \beta_0 + \sum_{m=1}^M \beta_m z_{m,st} + \sum_{l=1}^L f_l(v_{l,st}) \quad (3)$$

where $\beta = (\beta_0, \beta_1, \dots, \beta_M)$ are the coefficients that quantify the effect of covariates $z_j = (z_{1j}, \dots, z_{Mj})$ on the response, and $f = \{f_1(\cdot), \dots, f_L(\cdot)\}$ is a collection of functions defined in terms of a set of covariates $v = (v_1, \dots, v_L)$. From this definition, varying the form of the functions $f_l(\cdot)$ we can estimate different kinds of models, from standard and hierarchical regression, to spatial and spatiotemporal models (Rue et al. 2009). The vector of parameters is represented by $\theta = \{\beta_0, \beta, f_l\}$.

In our particular case we propose the model,

$$\eta_{st} = \ln(\mu_{st}) = \beta_0 + \sum_{m=1}^M \beta_m z_{m,st} + U_s + \gamma_t \quad (4)$$

where β_j represents the heterogeneity, in segment j and time t , $z_{\alpha,it}$ are the covariates, β_α the coefficients associated with covariates, U_s is a two dimensional random field capturing the spatial dependence effect, and γ_t is a one dimensional random process to capture the temporal effect. The term $\sum_{m=1}^M \beta_m z_{m,st}$ represents the part of the overall trend in the response explained by the covariates. γ_t is assumed to follow a random walk structure of order 1, thus accounting for local variation in time. This structure has the advantage of being able to capture trends and seasonality not accounted for by the covariates [33].

Using the covariates available, the full model takes the form

$$\eta_{st} = \beta_0 + \beta_1 z_{temp,st} + \beta_2 z_{wind,st} + \beta_3 z_{st} + \beta_4 z_{rh,st} + \beta_5 z_{dew,st} + \beta_6 z_{COVID19,st} + U_s + \gamma_t \quad (5)$$

We assumed Gaussian priors for the intercept and the coefficients for fixed effects with zero means and precisions equal to 0.001. The assumed prior for the precision $\tau = (1/\sigma^2)$ of the spatial and temporal terms were Gamma with parameters 1 and 0.00005.

For the spatial covariance structure, we used the Matérn family, which specifies the covariance function as $\Sigma_{ij} = Cov(\theta_{it}, \theta_{ju}) = \sigma_C^2 M(s_i, s_j \vee v, \kappa)$ where $\sigma_C^2 > 0$ is the variance component and

$$M(v, \kappa) = \frac{2^{1-v}}{\Gamma(v)} (\kappa \|h\|)^v K_v(\kappa \|h\|) \quad (6)$$

Controls the spatial correlation at distance $\|h\| = \|s_i - s_j\|$. Here, K_v is a modified Bessel function of the second kind, $\kappa > 0$ is a spatial scale parameter and the smoothness parameter is $v > 0$. When $v + \frac{d}{2}$ is an integer, a computationally efficient piecewise linear representation can be constructed by using a

different representation of the Matérn field $S(j)$, namely as the stationary solution to the stochastic partial differential equation (SPDE) [34],

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} s(j) = W(s) \tag{7}$$

where $\alpha = \nu + d/2$ is an integer, $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator and $W(s)$ is spatial white noise.

The main idea of the SPDE approach consists in defining the continuously indexed Matérn GF $X(s)$ as a discrete indexed GMRF by means of a basis function representation defined on a triangulation of the domain D ,

$$S(j) = \sum_{l=1}^n \varphi_l(s) \omega_l \tag{8}$$

where n is the total number of vertices in the triangulation, $\{\varphi_l(s)\}$ is the set of basis function and $\{\omega_l\}$ are zero-mean Gaussian distributed weights. Figure 7 shows two ways to build the triangulation needed to approximate a continuous Gaussian random field using the INLA-SPDE method [35]. The mesh on the left shows the triangulation when we ignore the boundary of the polygon of the study area D and instead use the convex hull defined by the coordinates of the data points. The plot on the right side of Figure 7 shows the triangulation obtained when we use the polygon defining the boundary of D .

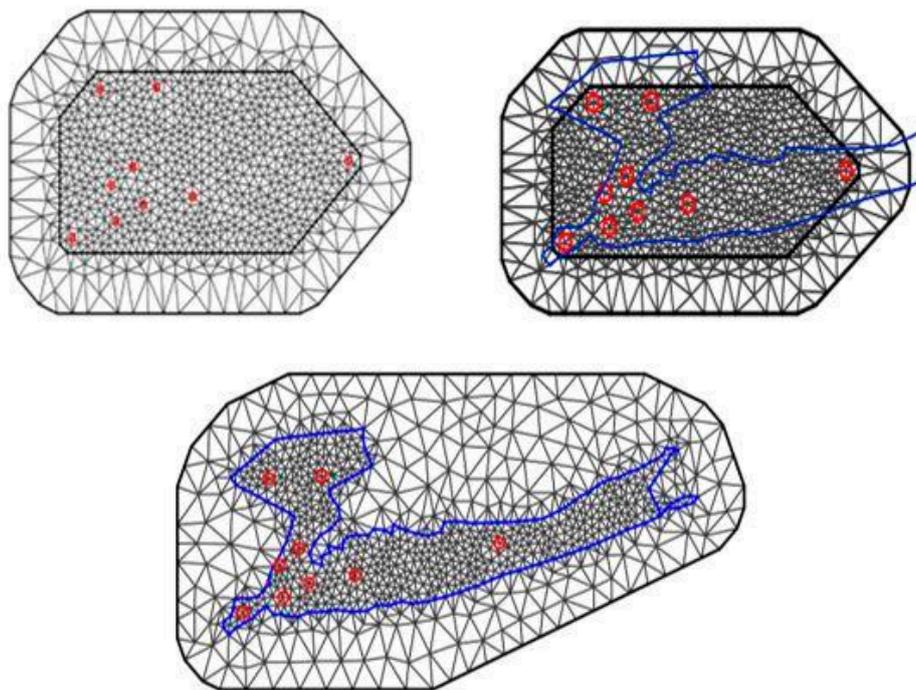


Figure 7. Stochastic partial differential equation (SPDE) mesh without boundary (top) and with the boundary of the study region (bottom). Locations of the weather stations in the study counties are highlighted as red circles. Blue line depicts the boundary region for the entire study area.

The basis functions $\varphi_l(s)$ are not random, but rather were chosen to be piecewise linear on each triangle,

$$\varphi_l(s) = \{1 \text{ at vertex } l \text{ and } 0 \text{ elsewhere}\} \tag{9}$$

The key step is to calculate $\{\omega_l\}$, which reports on the value of the spatial field at each vertex of the triangle. The values inside the triangle will be determined by linear interpolation [34].

Thus, expression (6) is an explicit link between the Gaussian field $S(j)$ and the Gaussian Markov random field and defined by the Gaussian weights $\{\omega_l\}$ that can be given by a Markovian structure.

The temporal dependence (on t) is assumed smoothed functions, in particular random walks of order 1 [36]. Thus, the random walk model of order 1 (RW1) for the Gaussian vector $\gamma = (\gamma_1, \dots, \gamma_t)$ is constructed assuming independent increments:

$$\gamma_t \sim \gamma_{t-1} N(\gamma_{t-1}, \tau_t^{-1}) \quad (10)$$

For the spatial effect in models is used SPDE. In Figure 7, some of the possibilities are presented.

Once the models have been obtained, they can be compared using the deviance information criterion (DIC) [37], which is a Bayesian model comparison criterion given by

$$DIC = \text{“goodness of fit”} + \text{“complexity”} = D(\underline{\theta}) + 2p_D \quad (11)$$

where $D(\underline{\theta})$ is the deviance evaluated at the posterior mean of the parameters and p_D denotes the effective number of parameters, which measures the complexity of the model. When the model is true, $D(\underline{\theta})$ should be approximately equal to the effective degrees of freedom, $n - p_D$, and DIC may under-penalise complex models with many random effects.

On the other hand, the conditional predictive ordinate (CPO) [38] is also analyzed. This expresses the posterior probability of observing the value (or set of values) when the model is fitted to all the data except y_i

$$CPO = \pi(y_i^{obs} | y_{-i}) \quad (12)$$

Here, y_{-i} denotes the observations y with the i th component removed. This facilitates computation of the cross-validated logscore [39] for model choice $(-\text{mean}(\ln(CPO)))$.

Therefore, the lowest values of DIC and $(-\text{mean}(\ln(CPO)))$ suggest the model with the best fit. A large number of parameters means more complexity. The best models are those with a high level of complexity and a high goodness-of-fit.

3. Results

Figure 8 shows the scatter plots of the number of daily COVID-19 new cases against pollutant concentrations for two selected counties (Bronx and Suffolk). Similar plots for all the current study counties are reported in the Appendix A, Figures A4 and A5. For Bronx only, daily data for PM_{2.5} and ozone were available. The selected plots suggest the existence of a slight association of the PM_{2.5} and ozone concentrations with the appearance of new cases of COVID-19. It appears to be a positive association between ozone and COVID-19 new cases and a decreasing association with PM_{2.5} and CO. Pearson correlation between pollutant concentration and the number of new cases of COVID-19 in the different counties was also low, corroborating the presence of a weak association. A state of emergency for the State of New York was declared on 7 March 2020. In compliance with this declaration, all nonessential activities were shut down. This included the closing of businesses, offices, and schools at all levels, and banned the gathering of public events and meetings with more than 50 people. It also included a stay at home order to the population, but it was not mandatory nor enforced by law officers. People thus had some mobility to go to grocery stores, visit hospitals and doctors' offices, walk their dogs, and go to laundromats. People could even go to enjoy the outdoors and do other activities in open spaces (<https://www.theguardian.com/us-news/2020/mar/20/new-york-90-day-stay-home-order-what-it-means>). Although traffic and transport were reduced, the “nonmandatory” stay at home order makes it difficult to predict if the variation of pollutant levels during the pre- and lockdown phases were the result of traffic reduction associated with the lockdown.

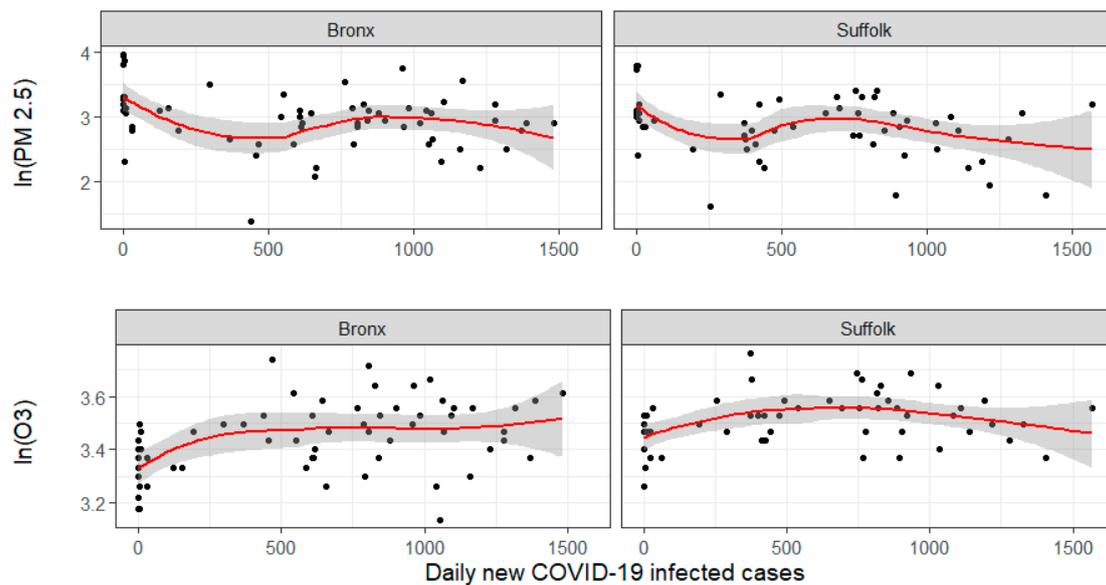


Figure 8. Scatter plots for the number of daily COVID-19 new infected cases vs. logarithmic value of $PM_{2.5}$ concentrations (measured in $\mu g/m^3$) (upper panel) and logarithmic value of ozone concentrations (measured in mg/m^3) (lower panel) for two selected counties Bronx and Suffolk. The red line depicts a Locally weighted scatterplot smoothing (Lowess) for the scattered data plot. Similar plots for all the current study counties are reported in the Appendix A, Figures A4 and A5.

The results for the hierarchical models fitted for each pollutant by county are reported in Tables 1 and 2. In total, 11 models were fitted, where the rows correspond to the fitted model number, and the columns to the covariates available for inclusion in the different models. The white color cells depict the covariates actually included in each model as well as the significant coefficient estimates obtained in boldface. Some models included only the COVID-19 new cases, the number of confirmatory tests performed in the study area, and space and time variation terms, other models included only the meteorological covariates, and the rest included some combination of both groups of covariates. This obeyed the presence of collinearity in the covariates, which resulted in changes in the regression coefficient signs depending on the covariates included in the different models. As the number of covariates was relatively high and because there are no automatic variable selection algorithms implemented in the INLA method, we only tested some combinations of covariates that in our perception were of interest. Tables 3 and 4 show the test statistics for the models fitted for $PM_{2.5}$ and ozone, respectively. The tables show the DIC and CPO values which are the most commonly used diagnostics for model quality [40]. From the values in Table 3 we see that the best model is model 11, which includes meteorological covariates, affected, plus spatial and time dependence of $PM_{2.5}$ concentrations. The effect of daily new COVID-19 positive cases is significant, but the coefficient for this covariate is substantially small. For the rest of the models fitted to $PM_{2.5}$, some covariates were statistically different from zero, but these are not the same effect in model 11 in terms of the diagnostic statistics.

Table 3 presents the estimates of the models fitted to the ozone concentrations in the study area. In this case, the model 11 is the best of all models under the DIC and CPO criteria and includes all the covariates.

For model 9, the coefficient for the effect of COVID-19 new cases on the ozone concentration was significant albeit it has a small value ($\beta_{COVID} = 0.005$). The inclusion of the meteorological covariates in the model for this pollutant reduced the value for the β_{COVID} coefficient to a point that, albeit significant, its effect on the response variable is negligible.

Table 1. Significant variable coefficients of 11 fitted models (for PM_{2.5}).

Model	Meteorological Factors					Affected		
	T	Dew	H	Ws	Prec	New Pos	Space	Time
1								
2						0.003		
3								
4						0.004		
5	0.068	-0.052	0.022	0.010	-0.086			
6						0.003		
7	0.068	-0.052	0.022	0.010	-0.086			
8						0.004		
9								
10	0.061	-0.047	0.023	0.050	-0.046			
11	0.060	-0.048	0.023	0.053	-0.039	0.000		

Bold if value is significant and grey not in model.

Table 2. Significant variable coefficients of 11 fitted models (for O₃).

Model	Meteorological Factors					Affected		
	T	Dew	H	Ws	Prec	New Pos	Space	Time
1								
2						0.004		
3								
4						0.005		
5	0.077	-0.074	0.035	0.015	-0.005			
6						0.004		
7	0.077	-0.074	0.035	0.015	-0.005			
8								
9						0.005		
10	0.077	-0.073	0.034	0.016	-0.021			
11	0.076	-0.073	0.035	0.018	-0.005	0.000		

Bold if value is significant and grey not in model.

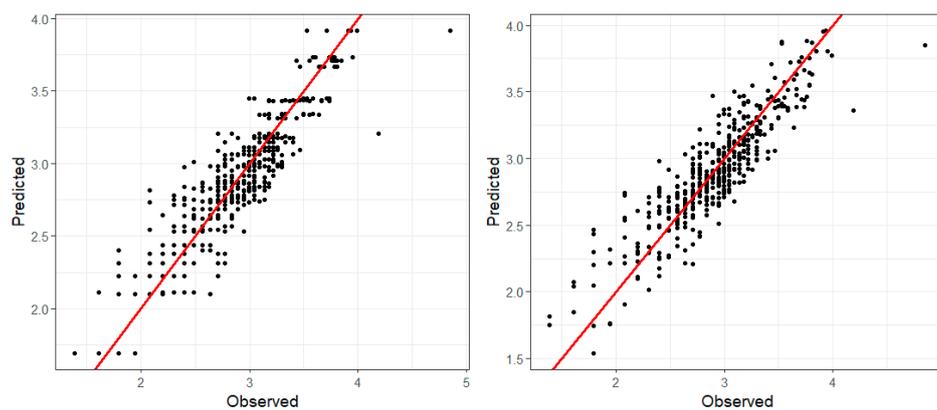
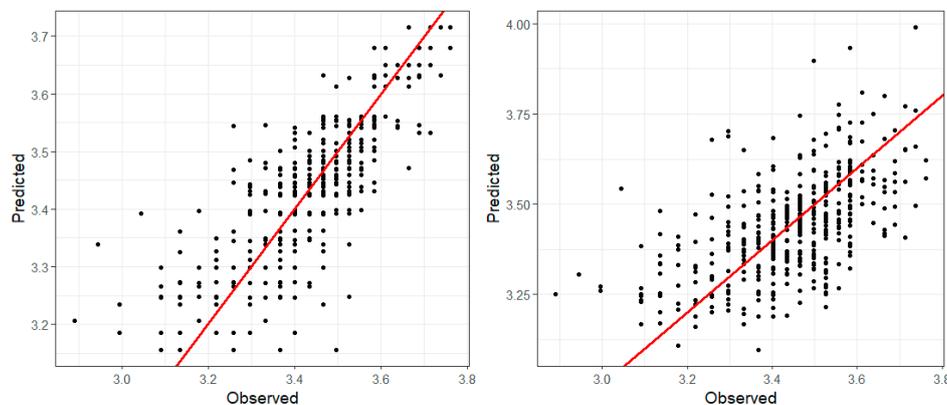
Table 3. Deviance information criterion (DIC) and conditional predictive ordinate (CPO) of models (for PM_{2.5}).

MODEL	DIC	CPO
1	2591.32	2.50
2	2232.89	2.16
3	2591.43	2.50
4	1983.99	1.92
5	747.14	0.72
6	2232.81	2.15
7	747.00	0.72
8	1983.81	1.92
9	2591.30	2.50
10	261.63	0.27
11	257.43	0.27

Table 4. Deviance information criterion (DIC) and conditional predictive ordinate (CPO) of models (for O₃).

MODEL	DIC	CPO
1	2371.41	4.88
2	2011.63	2.25
3	1869.86	2.09
4	191.40	0.26
5	2011.63	2.19
6	191.24	0.24
7	2371.29	2.66
8	1869.69	2.09
9	174.75	0.22
10	261.63	0.27
11	165.53	0.22

Figure 9 shows the plots of observed vs. fitted PM_{2.5} values obtained from models 9 (left plot) and 11 (right plot). Figure 10 depicts the plots of observed vs. fitted O₃ values (in mg/m³) for models 9 (left plot) and 11 (right plot). Besides the DIC and CPO values, the plots also support the conclusion that the models provide an acceptable fit to the data and that despite the observed patterns in Figure 8, the spatial and temporal variation observed in PM_{2.5} is not strongly associated with the appearance of new cases of COVID-19. This lack of association may be the result of the absence of strong measures against the COVID-19 pandemic given that the study area had not a law enforced lockout. Another factor might be that sources of PM_{2.5} remained at normal activity levels, although we have no information to confirm this hypothesis.

**Figure 9.** Plots of observed vs. fitted PM_{2.5} values (in µg/m³) for models 9 (left) and 11 (right). The red line depicts the linear regression between the observed and the predicted values.**Figure 10.** Plots of observed vs. fitted O₃ values (in mg/m³) for models 9 (left) and 11 (right). The red line depicts the linear regression between the observed and the predicted values.

4. Discussion

After the outbreak of the COVID-19 scientists all over the world started an unprecedented effort to understand different aspects of the pandemic. The lockdown decrees in many countries provided a unique opportunity to compare the levels of ozone, and particulate matter in urban areas under a reduced influence of human activities. Some of such studies have based their comparisons on point locations [5,9,41] whilst others have used aerial data to make comparisons in pollutant concentrations [8,10]. Ozone spatial and temporal distribution is closely related to NO₂ concentrations, although its dynamics are more difficult to model. NO₂ concentrations were not used as a response variable due to the lack of this information in our data. Thus, we cannot have a deep discussion about reductions of this pollutant in our study area and its association with human mobility in the nine counties considered. However, our goal was not to model the spatiotemporal dynamics of ozone in the study area, but, unlike other analyses where comparisons of pollutant concentrations during the COVID-19 pandemic have been reported, our aim was to measure through a GLMM test the association of PM_{2.5} and ozone with COVID-19 new daily cases after filtering out the effect of meteorological variables. The quantification of such association is represented by the corresponding regression coefficient estimates in the GLMM fitted. When only the COVID-19 new cases are included as a covariate along with the temporal and spatial components, the regression coefficient in most of the cases are 0.003 and 0.004 (Table 1), it is thus expected an increase of 0.3–0.4% in PM_{2.5} concentration for each new COVID-19 case. Note that we are not implying that COVID-19 cases bring an increase in air pollution, but that given the strong correlation between new COVID-19 and mobility, the result implies that an increase in mobility results in an increase in PM_{2.5} concentration.

We used a generalized linear model with a spatial and a temporal component to account for the effect of spatial and temporal association in the variance of the estimates, which reduces the number of false rejections of significant results [42]. Modeling air pollution is a complex task due to the multiple factors that affect pollutant concentrations at different scales. Albeit their limitations, the models we presented here are useful to gain insight on the possible effect that the reduction of human activities may have on PM_{2.5} and ozone concentrations in an urbanized area. For PM_{2.5} and O₃, all the inclusion of covariates in the model, improved the fit as measured by the DIC and CPO, the usual statistics to compare model quality in a Bayesian setting. However, the sparse spatial distribution of the covariate data used in the models explains in part the slight association between air pollutants and the number of new cases reported of COVID-19 we found. However, the correlation between the observed and predicted data from the models is indicative that model quality is acceptable but that models could be improved. Such improvements may come from obtaining the covariate and response information from a finer spatial scale or by posting a better functional relationship between the pollutant concentration and the covariates.

Linear models are a useful tool to measure the association between a response variable and a set of covariates. The GLMM model we used allows measuring the association between the concentrations of PM_{2.5} and ozone with the number of new cases of COVID-19 after filtering out the effects of meteorological variables that explain part of the observed variability of PM_{2.5} and ozone in the study area. By filtering such effects, we are able to tell that although there was a reduction in the concentrations of these two pollutants, only a small part of such reduction is explained by the number of COVID-19 new cases, and hence due to a reduction of mobility. The term “linear” refers to linearity in the model parameters and not in the covariates [43]. Linear models are sensitive to misspecification of the linearity between the response and the covariates. However, the exploratory data analysis made before the modeling process did not show evidence of nonlinearity between PM_{2.5} or ozone with the covariates. The inclusion of spatial and temporal terms in the models we fitted has helped to filter out the effect of spatial and temporal variability of the covariates as well, allowing us to obtain the proper power for statistical hypothesis tests, a key property of spatiotemporal statistical modeling [44]. The model results have shown the existence of a reduction in concentrations of ozone in the study area but after filtering out the meteorological covariates such reductions have only a slight

association to COVID-19 new cases. Ozone is a by-product of complex interactions between heat sources and atmospheric nitrogen in different forms, mainly nitrogen oxides, and is highly associated with automotive traffic in urban areas [45]. PM_{2.5} on the other hand, is related to industrial activity and to carbon emissions by diesel engines that power trucks and buses used in public transportation. Although the lockdown decree in the New York City area reduced significantly human mobility, our model results show that the reduction in PM_{2.5} and ozone observed in our study area is only slightly associated with changes in human mobility, as measured through its association with COVID-19 new cases. However, the current study is not comparing the values of pollutant data during pre-lockdown or in previous years, and can be incorporated in future studies in this domain. However, similar results are observed with the findings of Zangari et al. [46], who found that reductions in NO₂ and PM_{2.5} during the COVID lockdown in the State of New York have little association with reductions in mobility and point out that such reductions are more likely associated to a constant reduction of those pollutants in the atmosphere due to actions taken by authorities in previous years and to weather variables. These authors also mention the spatial variability in the NO₂ and PM_{2.5} reductions across the state. The lockdown imposed in many cities around the world resulted in notorious reductions in human mobility [47]. The New York City transportation system reported reductions in human mobility up to 60% from March to June 2020 [48]. Unfortunately, the daily data were not available to us for the nine counties considered in our study, and this key information was not included in our models.

In recent years, RWD from routine healthcare settings, like EHR powered by statistical analysis and machine learning techniques, produces RWE [49].

It provides insight adding potential to identify new indications in a real-world perspective to make strategic healthcare decisions beyond traditional clinical trials [50]. In context to this, the results of our study have contributed to rationalize the conjecture that a decrease in human-related air pollution results in a positive impact in nature, using RWD and providing a contribution to RWE for future decisions in the pollution-related health care issues and control measures.

5. Conclusions

In this work, we found that the number of new COVID-19 cases in nine counties in New York State has had little influence on the spatiotemporal distribution of both PM_{2.5} and O₃ concentrations, once the observed confounders (meteorological variables), as well as spatial and temporal dependency, were controlled. Despite the declaration of strict lockout for the population in the New York area and the concomitant reduction of human mobility, the observed association was not as strong as those reported by other scientists. The main difference between our results and others reported in the literature is that other authors have done comparisons of pollutant concentrations during the pandemic with previous time periods. We undertook a longitudinal analysis of PM_{2.5} and ozone atmospheric concentrations with COVID-19 new cases and meteorological factors. Our results are valid at local time and spatial scales. The increase in the number of new COVID-19 cases observed after the reopening of economic activities results in an increase in human mobility. This poses a big opportunity to test if pollution levels increase again.

The models we proposed in this work are useful to test the new observed trends. As COVID-19 contagion is influenced by people's social activities and mobility, the authorities should consider tightening confinement. Finally, the popular claim that the increase in COVID-19 new cases brings, as a consequence, a reduction in air pollution levels associated with reduced human mobility is reflected in our study area although this reduction is not as strong as expected.

Author Contributions: Conceptualization, L.S., P.J. and C.D.-A.; methodology, P.J., M.S.; software, S.C.; validation, L.S., P.J. and S.C.; formal analysis, L.S. and M.S.; resources, S.C.; data curation, L.S., P.J. and S.C.; writing—original draft preparation, L.S., C.D.-A., P.J. and M.S.; writing—review and editing, L.S., P.J., S.C. and C.D.-A.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This work is part of the ‘Cohort-Real World Data’ subprogram of the CIBER of Epidemiology and Public Health (CIBERESP). Carlos Diaz-Avalos was partially supported by grant PAPIIT-UNAM IG100221.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

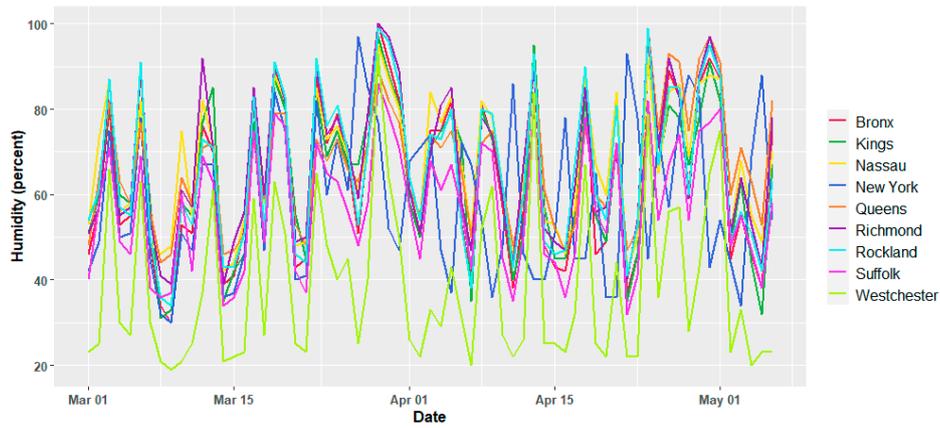


Figure A1. Variation of daily average humidity for individual counties.

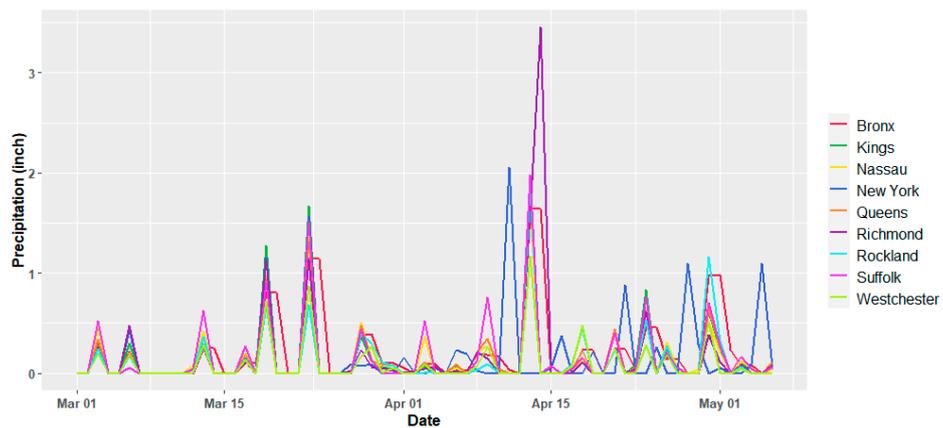


Figure A2. Variation of daily average precipitation for individual counties.

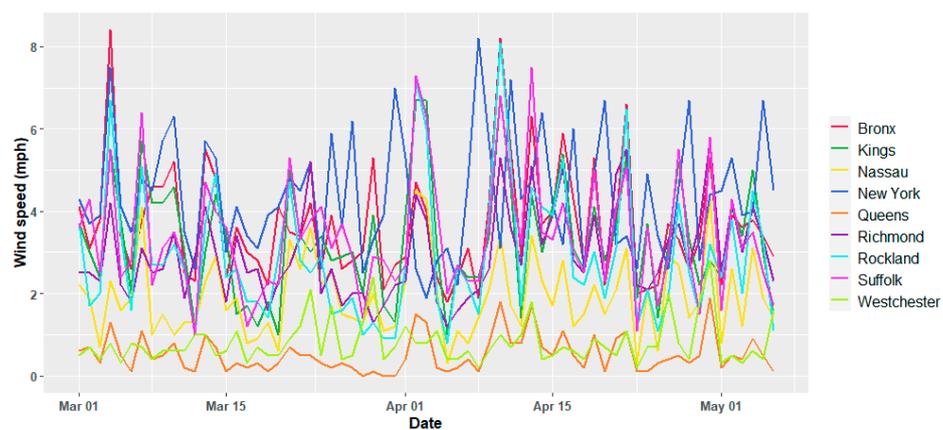


Figure A3. Variation of daily average wind speed for individual counties.

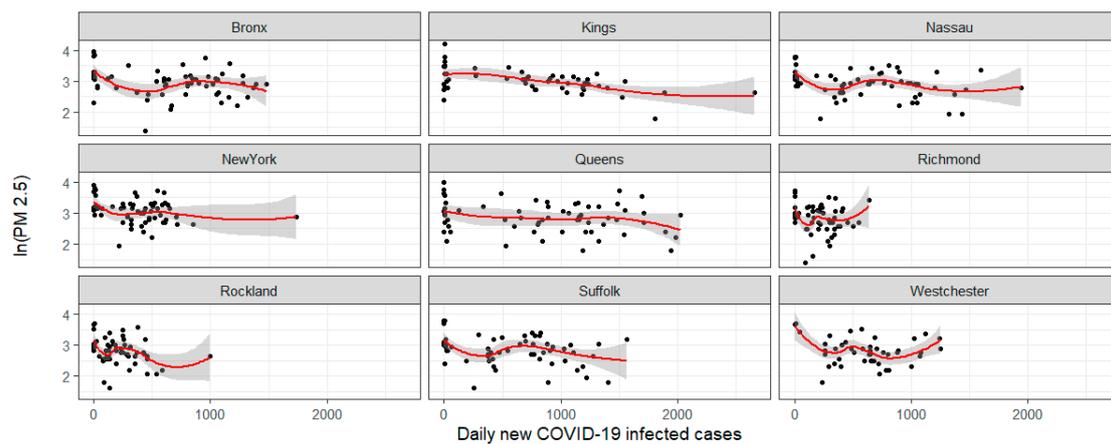


Figure A4. Scatter plots for the number of daily COVID-19 new cases vs. $PM_{2.5}$ (in $\mu g/m^3$) represented in natural logarithmic values for individual counties.

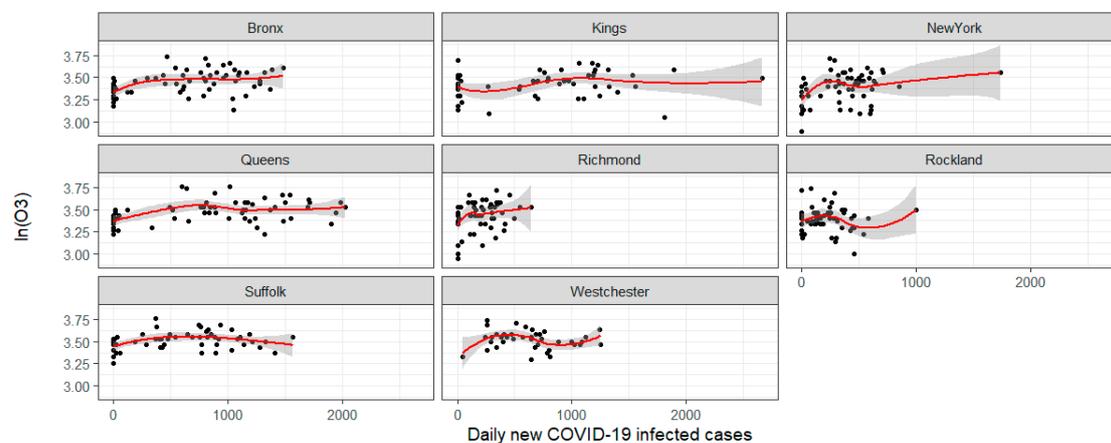


Figure A5. Scatter plots for the number of daily COVID-19 new cases vs. ozone (in mg/m^3) represented in natural logarithmic values for individual counties.

References

- Li, L.; Li, Q.; Huang, L.; Wang, Q.; Zhu, A.; Xu, J.; Liu, Z.; Li, H.; Shi, L.; Li, R.; et al. Air quality changes during the COVID-19 lockdown over the Yangtze River Delta Region: An insight into the impact of human activity pattern changes on air pollution variation. *Sci. Total Environ.* **2020**, *732*, 139282. [[CrossRef](#)]
- Sharma, S.; Zhang, M.; Gao, J.; Zhang, H.; Kota, S.H. Effect of restricted emissions during COVID-19 on air quality in India. *Sci. Total Environ.* **2020**, *728*, 138878. [[CrossRef](#)]
- Zambrano-Monserrate, M.A.; Ruano, M.A.; Sanchez-Alcalde, L. Indirect effects of COVID-19 on the environment. *Sci. Total Environ.* **2020**, *728*, 138813. [[CrossRef](#)]
- Muhammad, S.; Long, X.; Salman, M. COVID-19 pandemic and environmental pollution: A blessing in disguise? *Sci. Total Environ.* **2020**, *728*, 138820. [[CrossRef](#)] [[PubMed](#)]
- Dutheil, F.; Baker, J.S.; Navel, V. COVID-19 as a factor influencing air pollution? *Environ. Pollut. (Barking Essex 1987)* **2020**, *263 Pt A*, 114466. [[CrossRef](#)]
- Selvam, S.; Muthukumar, P.; Venkatramanan, S.; Roy, P.D.; Manikanda Bharath, K.; Jesuraja, K. SARS-CoV-2 pandemic lockdown: Effects on air quality in the industrialized Gujarat state of India. *Sci. Total Environ.* **2020**, *737*, 140391. [[CrossRef](#)] [[PubMed](#)]
- Adams, M.D. Air pollution in Ontario, Canada during the COVID-19 State of Emergency. *Sci. Total Environ.* **2020**, *742*, 140516. [[CrossRef](#)] [[PubMed](#)]
- Goldberg, D.L.; Anenberg, S.C.; Griffin, D.; McLinden, C.A.; Lu, Z.; Streets, D.G. Disentangling the impact of the COVID-19 lockdowns on urban NO_2 from natural variability. *Geophys. Res. Lett.* **2020**, *47*, e2020GL089269. [[CrossRef](#)] [[PubMed](#)]

9. Krecl, P.; Targino, A.C.; Oukawa, G.Y.; Cassino Junior, R.P. Drop in urban air pollution from COVID-19 pandemic: Policy implications for the megacity of São Paulo. *Environ. Pollut. (Barking Essex 1987)* **2020**, *265 Pt B*, 114883. [CrossRef]
10. Wang, Y.; Wen, Y.; Wang, Y.; Zhang, S.; Zhang, K.M.; Zheng, H.; Xing, J.; Wu, Y.; Hao, J. Four-Month Changes in Air Quality during and after the COVID-19 Lockdown in Six Megacities in China. *Environ. Sci. Technol. Lett.* **2020**. [CrossRef]
11. Wang, M.; Jiang, A.; Gong, L.; Luo, L.; Guo, W.; Li, C.; Zheng, J.; Li, C.; Yang, B.; Zeng, J.; et al. Temperature significant change COVID-19 Transmission in 429 cities. *medRxiv* **2020**. [CrossRef]
12. Makady, A.; de Boer, A.; Hillege, H.; Klungel, O.; Goettsch, W. What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews. *Value Health* **2020**, *20*, 858–865. [CrossRef] [PubMed]
13. RWE Navigator. 2020. Available online: <https://rwe-navigator.eu/use-real-world-evidence/sources-of-real-world-data/> (accessed on 15 September 2020).
14. Centers for Disease Control and Prevention, CDC. Cases, Data, and Surveillance. 2020. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/> (accessed on 2 June 2020).
15. New York State, NYC. Information on Novel Corona, NYC. 2020. Available online: <https://coronavirus.health.ny.gov/latest-news> (accessed on 21 May 2020).
16. Johns Hopkins University. COVID-19 United States Cases by County—Johns Hopkins Coronavirus Resource Center. 2020. Available online: <https://coronavirus.jhu.edu/us-map> (accessed on 2 June 2020).
17. Conticini, E.; Frediani, B.; Caro, D. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environ. Pollut.* **2020**, *261*, 114465. [CrossRef] [PubMed]
18. Contini, D.; Costabile, F. Does Air Pollution Influence COVID-19 Outbreaks? *Atmosphere* **2020**, *11*, 377. [CrossRef]
19. Ceylan, Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci. Total Environ.* **2020**, *729*, 138817. [CrossRef] [PubMed]
20. Briz-Redón, Á.; Serrano-Aroca, Á. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. *Sci. Total Environ.* **2020**, *728*, 138811. [CrossRef] [PubMed]
21. Saez, M.; Tobias, A.; Varga, D.; Barceló, M.A. Effectiveness of the measures to flatten the epidemic curve of COVID-19. The case of Spain. *Sci. Total Environ.* **2020**, *727*, 138761. [CrossRef]
22. Tobías, A.; Carnerero, C.; Reche, C.; Massagué, J.; Via, M.; Minguillón, M.C.; Alastuey, A.; Querol, X. Changes in air quality during the lockdown in Barcelona (Spain) one month into the SARS-CoV-2 epidemic. *Sci. Total Environ.* **2020**, *726*, 138540. [CrossRef]
23. Núñez-Delgado, A. What do we know about the SARS-CoV-2 coronavirus in the environment? *Sci. Total Environ.* **2020**, *727*, 138647. [CrossRef]
24. Grivas, G.; Athanopoulou, E.; Kakouri, A.; Bailey, J.; Liakakou, E.; Stavroulas, I.; Kalkavouras, P.; Bougiatioti, A.; Kaskaoutis, D.; Ramonet, M.; et al. Integrating in situ Measurements and City Scale Modelling to Assess the COVID-19 Lockdown Effects on Emissions and Air Quality in Athens, Greece. *Atmosphere* **2020**, *11*, 1174. [CrossRef]
25. Su, T.; Li, Z.; Zheng, Y.; Luan, Q.; Guo, J. Abnormally shallow boundary layer associated with severe air pollution during the COVID-19 lockdown in China. *Geophys. Res. Lett.* **2020**. [CrossRef] [PubMed]
26. Iacus, S.M.; Santamaria, C.; Sermi, F.; Spyrtatos, S.; Tarchi, D.; Vespe, M. Human mobility and COVID-19 initial dynamics. *Nonlinear Dyn.* **2020**, *101*, 1901–1919. [CrossRef] [PubMed]
27. Carteni, A.; Francesco, L.D.; Martino, M. How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. *Sci. Total Environ.* **2020**, *741*, 140489. [CrossRef]
28. Lindgren, F.; Rue, H. Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **2015**, *63*, 1–25. [CrossRef]
29. National Weather Service Forecast Office, New York. NOWData. 2020. Available online: <https://w2.weather.gov/climate/xmacis.php?wfo=okx> (accessed on 2 June 2020).
30. United States Environmental Protection Agency, US EPA. Outdoor Air Quality Data. 2020. Available online: <https://www.epa.gov/outdoor-air-quality-data/download-daily-data> (accessed on 2 June 2020).
31. New York State, NYC. New York State COVID-19 Data. 2020. Available online: <https://data.ny.gov/browse?tags=covid-19> (accessed on 2 June 2020).
32. Centers for Disease Control and Prevention, CDC. Cases Deaths by County|CDC. 2020. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/county-map.html> (accessed on 2 June 2020).

33. Cameletti, M.; Blangiardo, M. *Spatial and Spatio-Temporal Bayesian Models with R—INLA*; John Wiley Sons, Inc.: Hoboken, NJ, USA, 2015; pp. 235–258. ISBN 9781118950203.
34. Lindgren, F.; Rue, H.; Lindstrom, J. An explicit link between Gaussian fields and Gaussian Markov random fields the SPDE approach. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2011**, *73*, 423–498. [[CrossRef](#)]
35. Illian, J.B.; Sørbye, S.H.; Rue, H. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Ann. Appl. Stat.* **2012**, *6*, 1499–1530. [[CrossRef](#)]
36. R-INLA Project. 2020. Available online: <http://www.r-inla.org/home>. (accessed on 17 April 2020).
37. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P. Bayesian measures of model complexity and fit. *J. R. Stat. Soc.* **2002**, *64*, 583–639. [[CrossRef](#)]
38. Pettit, A.L.I. The conditional predictive ordinate for the normal distribution. *J. R. Stat. Soc. Ser. B* **1990**, *52*, 175–184. [[CrossRef](#)]
39. Gneiting, T.; Raftery, A.E. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [[CrossRef](#)]
40. Rue, H.; Martino, S.; Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2009**, *71*, 319–392. [[CrossRef](#)]
41. Singh, V.; Singh, S.; Biswal, A.; Kesarkar, A.P.; Mor, S.; Ravindra, K. Diurnal and temporal changes in air pollution during COVID-19 strict lockdown over different regions of India. *Environ. Pollut. (Barking Essex 1987)* **2020**, *266 Pt 3*, 115368. [[CrossRef](#)]
42. Cressie, N. *Statistics for Spatial Data*; John Wiley Sons, Inc.: Hoboken, NJ, USA, 1993; pp. 277–379. ISBN 9781119115151.
43. Seber, G.A.F.; Lee, A.A.F. *Linear Regression Analysis*; John Wiley Sons, Inc.: Hoboken, NJ, USA, 2003; pp. 4–5. ISBN 10: 0471415405.
44. Cressie, N.; Wikle, C.K. *Statistics for Spatio-Temporal Data*; John Wiley Sons, Inc.: Hoboken, NJ, USA, 2013; pp. 297–356. ISBN 978-0-471-69274-4.
45. Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and Health Impacts of Air Pollution: A Review. *Front. Public Health* **2020**, *8*. [[CrossRef](#)] [[PubMed](#)]
46. Zangari, S.; Hill, D.T.; Charette, A.T.; Mirowsky, J.E. Air Quality changes in New York City during the COVID-19 pandemic. *Sci. Total Environ.* **2020**, *742*. [[CrossRef](#)]
47. Warren, M.S.; Skillman, S.W. Mobility changes in response to COVID-19. *arXiv* **2020**, arXiv:2003.14228.
48. University Transportation Research Center, The City College of New York. Mobility Trends in New York City During COVID-19 Pandemic: Analyses of Transportation Modes Throughout 13 June 2020. Available online: <https://www.nyctaxinews.com/Mobility-Report-May2020.pdf> (accessed on 14 June 2020).
49. Zou, K.H.; Li, J.Z.; Imperato, J.; Potkar, C.N.; Sethi, N.; Edwards, J.; Ray, A. Harnessing Real-World Data for Regulatory Use and Applying Innovative Applications. *J. Multidiscip. Healthc.* **2020**, *13*, 671–679. [[CrossRef](#)] [[PubMed](#)]
50. Makady, A.; Ham, R.T.; de Boer, A.; Hillege, H.; Klungel, O.; Goettsch, W. GetReal Work Package Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies. *Value Health* **2017**, *20*, 520–532. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).