A Spatiotemporal Analytical Outlook of the Exposure to Air Pollution and Covid-19 Mortality in the USA

Sounak Chakraborty, Tanujit Dey, Yoonbae Jun,
Chae Young Lim, Anish Mukherjee, Francesca Dominici

Francisca Adobea Ofori

December 2023

1 Objectives of the paper

The study seeks to understand the link between exposure to fine particulate matter (PM2.5) and COVID-19 mortality. Using a spatiotemporal model with Bayesian regression and random effects, the research accounts for social and environmental factors. Weekly data from four diverse U.S. locations are analyzed to explore how PM2.5 exposure relates to COVID-19 deaths while considering variations in time and space.

2 Methods Used for Analysis

2.1 Spatial Negative Binomial Model (SNB)

The paper introduces a Spatial Negative Binomial Model (SNB) to analyze weekly COVID-19 deaths at the county level. It uses a negative binomial distribution, considers covariates like PM2.5 exposure and population density, and includes county-specific spatial random intercepts for spatial dependence. The model is compared with a spatiotemporal counterpart in the analysis.

2.2 Spatiotemporal Negative Binomial Model (STNB)

The paper introduces a Spatiotemporal Negative Binomial Model (STNB) for analyzing COVID-19 death counts over time and space. This model, addressing over-dispersion, uses a negative binomial distribution and logistic link function. It incorporates fixed-effect covariates, nonlinear time-fixed effects, and spatial bivariate random effects for intercept and slope, capturing potential temporal heterogeneity across counties. The model includes a bivariate Intrinsic Conditional Autoregressive (ICAR) prior for spatial dependence.

2.3 Spatiotemporal Zero-Inflated Negative Binomial Model (STZINB)

It also uses a Spatiotemporal Zero-Inflated Negative Binomial Model (STZINB) to address zero inflation in COVID-19 death counts across counties and time. It separates the probability of zero counts from actual counts and models them using logistic regression with fixed-effect covariates, nonlinear time-fixed effects, and spatial random effects. Two model variations are considered: STZINB-NLT with a nonlinear fixed-time trend and STZINB-LT with a simpler linear fixed-time trend for the zero count component. The paper explores these models to assess the need for a nonlinear time trend in zero inflation.

2.4 Bayesian Inference

2.4.1 Prior Specification and MCMC Settings

The paper outlines the prior and hyper-parameter specifications for the Spatiotemporal Zero-Inflated Negative Binomial (STZINB) model. The probability for latent at-risk indicators is expressed as $exp(\theta_{1ij})/[1 + exp(\theta_{1ij}]$. Priors for β_1 and β_2 are assumed to be normal, and a uniform prior is used for the dispersion parameter (r). Time-basis functions are constructed by standardizing time points. Model comparison is performed using Deviance Information Criterion (DIC). Three MCMC chains with 11,000 iterations (1,000 burn-in) are run for each model, and convergence is assessed using standard MCMC diagnostics.

2.4.2 Conditional Posterior Distribution and Model Fitting

The paper employs a Markov Chain Monte Carlo (MCMC) approach for updating parameters in the Spatiotemporal Zero-Inflated Negative Binomial (STZINB) model. It details steps for updating at-risk indicators, coefficients for binary and count model components, dispersion parameter, and spatial random effects. The process involves drawing from various distributions, including Bernoulli and Pólya-Gamma, with adjustments for numerical stability. The authors use a Metropolis–Hastings method for updating certain parameters and provide conditional prior distributions for the spatial random effects. The MCMC chains involve 11,000 iterations with a 1,000 iteration burn-in, and convergence is assessed using standard diagnostics.

3 Data

The study analyzed daily COVID-19 death counts in U.S. counties from March 23, 2020, to August 31, 2020, using Johns Hopkins University data. Four regions were examined, and spatiotemporal models considered weekly death counts with population size as an offset. Spatial models analyzed cumulative counts. Data from Wu et al. (2020) on air pollution and twelve key variables (socio-economic, demographic, healthcare, and lifestyle factors) were included. Sensitivity analysis explored different counting approaches. The study aimed to understand COVID-19 dynamics, considering various factors, with data spanning spatial but not temporal variations.

4 Results

The study compares Bayesian spatiotemporal models for COVID-19 mortality, finding positive associations between long-term PM2.5 exposure and deaths across regions. Socioeconomic factors show mixed effects. Spatial analyses reveal regional variations and distinct temporal patterns in COVID-19 deaths. Comparisons with a spatial negative binomial model highlight consistent PM2.5 associations but with differing effect sizes. Sensitivity to aggregation periods and potential overfitting are noted for spatiotemporal models.

5 Conclusions

The study explores the relationship between long-term exposure to PM2.5 and county-level COVID-19 weekly death counts using various spatiotemporal negative binomial models. The analysis considers different regions, accounting for sociocultural and healthcare system variations. The selected models show consistent positive associations between PM2.5 and COVID-19 death counts, with varying effect sizes across regions. The complexity of the models, incorporating nonlinear time effects and spatial random effects, provides flexibility but may lead to increased estimation variability. The findings highlight the importance of considering regional heterogeneity and spatiotemporal dynamics in studying the impact of air pollution on COVID-19 mortality. The study emphasizes the need for patient-level data and ongoing research to strengthen the understanding of this association.

6 Discussions

The study reveals a strong and statistically significant link between long-term PM2.5 exposure and COVID-19 death counts in several U.S. locations. Despite regional differences, the consistency of findings across spatiotemporal models emphasizes the robustness of the observed association. Because of the potential ecological fallacy, caution is suggested in interpretation, emphasizing the importance of future study with granular patient-level data. The study's methodological rigor, as proven by extensive simulations, increases confidence in its contribution to understanding the intricate relationship between air pollution and COVID-19 outcomes.