Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Exploring the effects of PM$_{2.5}$ and temperature on COVID-19 transmission in Seoul, South Korea

Youngbin Lym$^a$, Ki-Jung Kim$^b$,*

$^a$ Center for Innovation Strategy and Policy, KAIST, South Korea

$^b$ Department of Smart Car Engineering, Doowon Technical University, South Korea

ARTICLE INFO

Keywords:
Hierarchical Bayesian
COVID-19 transmission
Particulate matter 2.5
Temperature
Relative risks
Local municipality

ABSTRACT

With a recent surge of the new severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2, COVID-19) in South Korea, this study attempts to investigate the effects of environmental conditions such as air pollutants (PM$_{2.5}$) and meteorological covariate (Temperature) on COVID-19 transmission in Seoul. To account for unobserved heterogeneity in the daily confirmed cases of COVID-19 across 25 contiguous districts within Seoul, we adopt a full Bayesian hierarchical approach for the generalized linear mixed models. A formal statistical analysis suggests that there exists a positive correlation between a 7-day lagged effect of PM$_{2.5}$ concentration and the number of confirmed COVID-19 cases, which implies an elevated risk of the infectious disease. Conversely, temperature has shown a negative correlation with the number of COVID-19 cases, leading to reduction in relative risks. In addition, we clarify that the random fluctuation in the relative risks of COVID-19 mainly originates from temporal aspects, whereas no significant evidence of variability in relative risks is observed in terms of spatial alignment of the 25 districts. Nevertheless, this study provides empirical evidence using model-based formal assessments regarding COVID-19 infection risks in 25 districts of Seoul from a different perspective.

1. Introduction

Human coronavirus, which causes severe acute respiratory infections in humans, first appeared with the outbreak of severe acute respiratory syndrome (SARS-CoV) in 2002, progressed as Middle East respiratory syndrome (MERS-CoV) in 2012, and recently emerged as the severe acute respiratory syndrome-2 (SARS-CoV-2, COVID-19) (Fung and Liu, 2019). COVID-19 originated in Wuhan, China at the end of December 2019 and later spread worldwide. Eventually, the World Health Organization (WHO) declared the outbreak a pandemic on March 11, 2020, following the Hong Kong flu in 1968 and a novel swine-origin influenza in 2009 (WHO, EURO., 2021). According to the WHO (2021), 127, 349, 248 people were infected with COVID-19 worldwide, with 2,787,593 deaths as of March 31, 2021.

In South Korea, the first novel coronavirus case was confirmed on January 20, 2020 (Ministry of Foreign Affairs, 2020). Since then, the country has faced a series of waves of COVID-19 outbreaks. The first wave was linked with religious gatherings in the southeastern city of Daegu during the early February and March 2020, causing a localized spread. With regard to the second wave of COVID-19 infections, the number of confirmed cases increased sharply in the Seoul metropolitan area (i.e., Seoul capital region consisting of Seoul, Incheon, and Gyeonggi-do), where the surge was closely associated with the Independence Day rally on August 15, 2020 (Ministry of Foreign Affairs, 2020). The third wave began in the Seoul metropolitan area at the end of November 2020 and is still an ongoing process. The daily number of confirmed cases has stayed in triple digits over the past five months amid the third wave (KCDC, 2021).

As described in Fig. 1, an ultrafine dust (PM$_{2.5}$) alert issued in Seoul after the second wave of COVID-19 was controlled (November 15, 2020). After the alert, there appeared to be a new surge of COVID-19 infections, leading to the third wave. This motivated us to test whether there existed a certain association between exposure to air pollutants as well as meteorological factors such as temperature and humidity and COVID-19 spreads (Manisalidis et al., 2020).

Numerous scholars have attempted to understand the COVID-19 infectious disease from various perspectives such as the socio-demography (Sannigrahi et al., 2020), changes in air quality after lockdown (Park et al., 2020; Shen et al., 2021; Xu et al., 2020), environmental conditions (Bashir et al., 2020), air pollution (Wu et al., 2020a), spatial dependency (Briz-Redon and Serrano-Aroca, 2020), and spatiotemporal correlation (Elson et al., 2021; Sartorius et al., 2021), to
name a few. In addition, significant effort has been devoted to investigating various facets of transmission dynamics of COVID-19, including human-to-human diffusion mechanism (Anand, Cabreros, et al., 2021a, 2021b; Bontempi, 2020a), environment-to-human transmission (Bashir et al., 2020; Coccia, 2020a; Rahimi et al., 2021; Sarkodie and Owusu, 2020), and pollution-to-human transmission (Domingo et al., 2020; Maleki et al., 2021). To account for human-to-human diffusion mechanism, some researchers have considered non-pharmaceutical behaviors and policy-based strategy such as restriction measures (Chen et al., 2021; Coccia, 2021c), while others have examined the role of socio-economic factors including commercial exchange (Bontempi, 2020a), GDP (Coccia, 2021d; Islam et al., 2021), and population density (Diao et al., 2020; Coccia, 2021b). Regarding environment-to-human as well as pollution-to-human transmission mechanisms, several researchers have studied the association between COVID-19 transmission and meteorological factors such as temperature (Xie and Zhu, 2020; Islam et al., 2021), humidity (Haque and Rahman, 2020; Diao et al., 2020), and wind speed (Coccia, 2020a, 2020c, 2021b, 2020e), and pollution factors including air pollutants (Barakat et al., 2020; Copat et al., 2020; Hoang et al., 2021) and wastewater (Anand, Cabreros, et al., 2021a, 2021b; Rahimi et al., 2021). It is worth noting that, however, the mechanisms underlying the airborne transport of COVID-19 have not been fully elucidated. According to recent studies (Belosi et al., 2021; Contini and Costabile, 2020), the probability of airborne virus transmission appears to be low in open environments and higher in closed ones, especially in hospitals or quarantine facilities. Therefore, to acquire in-depth understanding of COVID-19 diffusion process, it is required to consider interdisciplinary and multi-dimensional approaches, which entail virus transmission mechanisms, social acceptance of policy measures, socio-demographic features, and economic conditions (Bontempi et al., 2020; Duan et al., 2021).

In this study, we considered previous studies that specifically investigated the correlation between environmental variables (e.g., meteorological covariates and air pollutants) and COVID-19 infections, as summarized in Table 1. Some studies suggest that air pollutants (PM$_{2.5}$, PM$_{10}$) significantly positively correlate with COVID-19 transmission (Chakrabarty et al., 2020; Coccia, 2020b; Copat et al., 2020; Fattorini et al., 2020; Srivastava, 2021; Vasquez et al., 2020; Wu et al., 2020b; Yao et al., 2020; Zhu et al., 2020). Others suggest that air pollutants (PM$_{2.5}$, PM$_{10}$) have a negative association with COVID-19 spread (Bashir et al., 2020). There are also other papers showing that this correlation seems not to be established or remains unclear (Barakat et al., 2020; Belosi et al., 2021; Bontempi, 2020b; Pivato et al., 2021).

When it comes to the influence of meteorological variables on COVID-19 infections, both temperature and humidity appeared to be positively associated with COVID-19. Meanwhile, certain studies have reported mixed outcomes (e.g., Ward et al., 2020; Passerini et al., 2020). Previous studies suggest that the influence of the air pollution and meteorological covariates on the spread of COVID-19 may vary with respect to regions, study periods, and analytic operational frameworks. In consideration of these, the present study specifically attempts to: 1) examine the role of air pollutants and meteorological variables on the daily confirmed COVID-19 transmission cases at 25 local districts in Seoul; 2) test for the lagged influences of PM$_{2.5}$ and other covariates on COVID-19 infection; 3) unveil the latent/random effects stemming from the unexplained variations in COVID-19 spreads across different districts over time after accounting for the influences of covariates or fixed effects. To answer these research questions, we employ a flexible Bayesian hierarchical method along with generalized linear mixed models.

2. Data and methods

2.1. Background of this research

According to the Korea Disease Control and Prevention Agency (KDCA, 2021), a total of 103,088 confirmed cases (198.83 per 100,000 population) of COVID-19 have been reported with 1731 deaths (1.68 % of case fatality rate) as of March 31, 2021. More than half of the cases were confirmed in Seoul (31 %) and other metropolitan areas (32 %), as

![Fig. 1. Distribution of daily confirmed COVID-19 cases in South Korea.](image_url)
the second wave emerged in the Seoul metropolitan area. The incidence rate per 100,000 people in Seoul exceeded 300 during this period. Therefore, we specifically focused on Seoul, and the second and third waves of the pandemic period (August 1–December 31, 2020) to represent the COVID-19 infection trends in South Korea after the first wave.

2.2. Data in details

The capital of South Korea, Seoul, which has the highest population density (24,637.2 people/km²), consists of 25 districts (i.e., gu, an administrative unit). This study considers a district as a unit of analysis for both data collection and analytic operation. For each district, we created a time-series of daily confirmed cases of COVID-19, meteorological variables (i.e., temperature and humidity), and air pollution data (i.e., PM2.5, SO2, CO, NO2, and O3) within our study period (August 1–December 31, 2020). Table 2 presents the corresponding data sources utilized for this research. For example, the number of daily confirmed cases of COVID-19 in each district was obtained from the Seoul City Hall’s Corona 19 Confirmation Board (Seoul Metropolitan City Hall, 2020). Hourly air pollution and daily meteorological data were obtained from Air Korea (Korean Environment Corporation, 2021) and the KMA Weather Data Service (Korea Meteorological Administration, 2021), respectively.

Moreover, as shown in Fig. 2, we conducted a series of data manipulation and validation processes to acquire a clean and balanced dataset for further investigation. The procedure starts with the reconfirmation of the daily confirmed COVID-19 cases of 25 districts so as to ensure the accuracy and validity of the raw data (e.g., handling outliers or observations with missing information). Regarding covariates specific to air pollutants, humidity, and temperature, we consider the daily minimum, average, and maximum values for individual districts. It is worth noting that the temporal dimension of data collection for these covariates ranges from July to December 2020 as we would like to examine the association between the 7-day lagged effects of those variables and COVID-19 infectious risks. We then calculate the number of expected COVID-19 cases for each district based on monthly population and confirmed cases. This not only allows us to adjust the differences in population at risk across 25 districts but also helps compare the relative risks of COVID-19 transmission. Lastly, we add some geospatial information for visualization and statistical analysis.

Regarding our research interest, we present Fig. 3, which shows the distribution of the number of daily confirmed COVID-19 cases per 100,000 people and the average concentration of PM2.5. We observe a new surge of COVID-19 cases in the middle of November, whereas we do not identify any patterns in PM2.5 due to its fluctuation. In that regard, a simple visual inspection does not guarantee the existence of certain association between the number of confirmed cases and PM2.5 concentration in general, requiring further scrutiny. Furthermore, it is necessary to examine the lagged influence of those environmental factors. As indicated by Lauer et al. (2020), the median time of onset of symptoms after exposure to the COVID-19 virus is 5.1 days. Hence, we considered a 7-day lagged effect of environmental covariates.

2.3. Correlation analysis and variable selection

Prior to conducting a model-based formal assessment, we checked whether the collected covariates were correlated to each other. The correlation matrix (Table 3) provides both correlation coefficients and p-values (significance levels for Pearson correlation) of the covariates. We verified that all air pollution variables were highly correlated to each other with strong significance. Similarly, we found a significant correlation between temperature and humidity. Thus, to avoid multicollinearity problems among highly correlated covariates, we considered a sub-selection of variables including PM2.5, temperature,
Environmental Research 203 (2022) 111810

2.4. Analytic operational framework

Considering our research objectives, this study adopted a multi-level modeling framework under a Bayesian perspective (Congdon, 2019; Wikle et al., 2019), which provided more flexibility to collectively incorporate both fixed effects (covariates) and random effects (unobserved heterogeneity). Specifically, our focus relied on the relative risk of COVID-19 transmission across districts over time. Stated formally, we assumed that the number of daily confirmed COVID-19 cases in each district followed a Poisson distribution with the mean structure (i.e., the expected number of occurrences ($μ_{it}$), indexed by district $i$ and time $t$), which is further decomposed into $μ_{it} = E_{it} \times θ_{it}$.

$E_{it}$ refers to the expected number of cases, which is an offset/exposure to adjust the difference in population at risk across districts, while $θ_{it}$ denotes the relative risk of COVID-19 infections. The natural logarithm of the relative risks ($θ_{it}$) can be characterized by the linear combination of several fixed and random effects, which finally yields to the following a full Bayesian hierarchical modeling structure:

$$\log(θ_{it}) = α + \sum_{l=1}^{p} β_{l}X_{i,l,t} + s_{i} + γ_{t} + δ_{it}$$

where $Y_{it}$ indicates the observed daily confirmed cases of district $i$ and day $t$, $α$ is an intercept of the model, $β_{l}$ denotes the fixed effect that the independent variable $X_{i,l,t}$ has on $θ_{it}$. We assumed that the latent influences are characterized by a linear combination of three independent and identically distributed random effects (i.e., district ($s_{i}$), day ($γ_{t}$), their interaction ($δ_{it}$)). Each latent effect is assumed to be following the Gaussian (Normal) distribution with mean 0 and precision, $τ$.

Data: $Y_{it} | θ_{it} \sim \text{Poisson}(E_{it}θ_{it})$ (1)

Process (Level 2): $\log(θ_{it}) = α + \sum_{l=1}^{p} β_{l}X_{i,l,t} + s_{i} + γ_{t} + δ_{it}$

$s_{i} \sim \text{Normal}(0, τ_s)$
$γ_{t} \sim \text{Normal}(0, τ_γ)$
$δ_{it} \sim \text{Normal}(0, τ_δ)$

Parameter (Level 3): $p(\frac{1}{\sqrt{1+k}} > 1) = 0.01$ where $k = s, γ, δ$

$β_{l} \sim \text{Normal}(0, 10^{4})$ for $l = 1, 2, ..., p$ and $p - 1, p$
$i = 1, 2, ..., 24, 25$ districts
$t = 1, 2, ..., 152, 153$ days

where $Y_{it}$ indicates the observed daily confirmed cases of district $i$ and day $t$, $α$ is an intercept of the model, $β_{l}$ denotes the fixed effect that the independent variable $X_{i,l,t}$ has on $θ_{it}$. We assumed that the latent influences are characterized by a linear combination of three independent and identically distributed random effects (i.e., district ($s_{i}$), day ($γ_{t}$), their interaction ($δ_{it}$)). Each latent effect is assumed to be following the Gaussian (Normal) distribution with mean 0 and precision, $τ$.
of variance of the Gaussian distribution \( \tau^{-1} = \sigma^2 \); here, \( \tau_x, \tau_y, \) and \( \tau_t \) denote precisions of 25 districts, 153 days, and 25 \times 153 = 3825 observations, respectively.

Regarding a distribution of precision parameters, we followed Simpson et al. (2017), who facilitated the allocation of weakly informative Penalized Complexity (PC) priors. To define the prior of a precision parameter \( \tau \), we employed a probability statement such that \( p(1/\sqrt{\tau} > U) = \alpha \), where \( U \) and \( \alpha \) denoted a sensible (weakly informative) user-defined upper bound and probability (weight) placed on this event, respectively (Simpson et al., 2017; Moraga, 2019; Lym, 2021). In this study, we select 1 and 0.01 as \( U \) and \( \alpha \), respectively, implying that the probability of the standard deviation (an inverse of the square root of the precision) of each random component exceeding 1 is 0.01, after accounting for the influences of fixed effects by covariates.

Based on the Bayesian hierarchical model described above, we suggest the following six models to explore the association between environmental covariates and the relative risk of COVID-19 transmission:

**Type A: no lagged effect of environmental variables (2)**

**Model 1.** \( \log(\theta_1) = \alpha + \beta_{\text{Popden},1}X_{\text{Popden},t} + \beta_{\text{PM}_{2.5},1}X_{\text{PM}_{2.5},t} + \beta_{\text{Temp},1}X_{\text{Temp},t} + \epsilon_t + \gamma_t + \delta_x \)

**Model 2.** \( \log(\theta_2) = \alpha + \beta_{\text{Popden},1}X_{\text{Popden},t} + \beta_{\text{PM}_{2.5},1}X_{\text{PM}_{2.5},t} + \beta_{\text{Temp},1}X_{\text{Temp},t} + \epsilon_t + \gamma_t \)

**Model 3.** \( \log(\theta_3) = \alpha + \beta_{\text{Popden},1}X_{\text{Popden},t} + \beta_{\text{PM}_{2.5},1}X_{\text{PM}_{2.5},t} + \beta_{\text{Temp},1}X_{\text{Temp},t} + \epsilon_t + \gamma_t + \delta_x \)

**Type B: a 7-day lagged effect of environmental variables (3)**

**Model 1.** \( \log(\theta_4) = \alpha + \beta_{\text{Popden},1}X_{\text{Popden},t} + \beta_{\text{PM}_{2.5},1}X_{\text{PM}_{2.5},t-7} + \beta_{\text{Temp},1}X_{\text{Temp},t-7} + \epsilon_t + \gamma_t + \delta_x \)

**Model 2.** \( \log(\theta_5) = \alpha + \beta_{\text{Popden},1}X_{\text{Popden},t} + \beta_{\text{PM}_{2.5},1}X_{\text{PM}_{2.5},t-7} + \beta_{\text{Temp},1}X_{\text{Temp},t-7} + \epsilon_t + \gamma_t \)

**Model 3.** \( \log(\theta_6) = \alpha + \beta_{\text{Popden},1}X_{\text{Popden},t} + \beta_{\text{PM}_{2.5},1}X_{\text{PM}_{2.5},t-7} + \beta_{\text{Temp},1}X_{\text{Temp},t-7} + \epsilon_t + \gamma_t + \delta_x \)

For each Type, we consider three models constituting only a fixed effect (Model 1), and fixed and random effects (Models 2 and 3). To explore the role of environmental variables, we employed \( \text{PM}_{2.5} \) concentration and temperature, while population density was used as a control variable. In addition, we attempted to test for a 7-day lag effect of covariates on the relative risks of COVID-19, relating to the incubation time of the disease (Briz-Redón and Serrano-Aroca, 2020; Lauer et al., 2020).

### 3. Results and discussion

As indicated, this study employs a full Bayesian hierarchical...
approach for the generalized linear mixed model (GLMM). To implement the formal assessment of selected models, we employed the integrated nested Laplace approximation (INLA) method in lieu of the conventional Markov chain Monte Carlo (MCMC) simulation approach (Rue et al., 2009; Lindgren and Rue, 2015). The analysis was performed using the R-INLA (Rue et al., 2017) package in R programming language.

This study relied on the number of daily confirmed COVID-19 cases of 25 districts in Seoul from August 1 to December 31, 2020. Moreover, we calculated the expected number of cases in each district to account for the differences in population at risk, which enabled us to estimate the relative risk of the infectious disease. Data exploration in Table 3 indicates that there exists multicollinearity among covariates, leading to a sub-selection of variables for further empirical investigation.

We developed three models consisting of fixed effect only (Model 1), and fixed and random effects (Models 2 and 3). To explore the role of environmental variables as well as their lagged influences on COVID-19 transmission, we conducted statistical analysis utilizing Type A and Type B datasets. To adjust for the differences in the scales of these covariates, we standardized them prior to running the models. The outcomes of the selected models are presented in Table 4 and Table 5.

The influence of covariates are compared with the number of confirmed COVID-19 cases along with latent effects in Table 4, while Table 5 presents three models with lagged effects of covariates. When comparing the performance of Bayesian models, DIC (deviance information criterion) and WAIC (widely applicable information criterion) are often used for an optimal model selection. The lower the value of DIC (or WAIC), the more favorable is the model (Spiegelhalter et al., 2002; Wang et al., 2018).

Based on this understanding, we verified that models with latent effects (Models 2 and 3) outperform the fixed effect only model (Model 1), as indicated by a significant drop in DIC and WAIC values. Model 3 (the rightmost column) in Table 4 reveals that the 95% credible interval for PM$_{2.5}$ ranges from $-0.001$ to $0.206$ with a posterior mean of $0.103$. This indicates a 95% probability that the true parameter of PM$_{2.5}$ is within this interval. The estimated relative risks of COVID-19 infections by a unit increase in PM$_{2.5}$ concentration (standardized) is $10.85\% = \exp (0.103) - \exp (0)$, assuming that other factors are fixed. Similar findings can be found in previous works (Coccia, 2020c; 2020d, 2021a, 2021b; Huang et al., 2021; Yao et al., 2020; Zhu et al., 2020).

According to Bontempi (2020b), the pollution can be a source of the virus spread, so-called pollution-to-human transmission mechanism, so that dust and PM$_{2.5}$ can facilitate the transmission of the COVID-19 virus droplets and particles (Domingo et al., 2020; Maleki et al., 2021; Sri-vastava, 2021). A similar statement can be made for temperature while holding other factors constant; a unit increase in temperature would lead to $-45.4\% = \exp (-0.605) - \exp (0)$, indicating a reduction in the relative risks. Several studies confirm that under higher temperature, the spread of COVID-19 tends to be slower (i.e., the virus remains active at low temperature), implying the mitigating effect against the survival and transmission of the virus (Anand, Cabreros, et al., 2020; Rahimi et al., 2021; Sarkodie and Owusu, 2020; Xie and Zhu, 2020). In line with them, our study also verifies the negative influence of temperature on COVID-19 transmission.

Regarding the lagged influence of covariates, we hypothesized that certain associations existed between the 7-day lags of environmental variables and the number of daily confirmed cases. Table 5 suggests that PM$_{2.5}$ is positively correlated with the number of cases, leading to an elevated risk while temperature contributed to lowering the relative risks of COVID-19. Furthermore, we identified that Model 3 outperformed the other two models, implying the importance of incorporating unobserved heterogeneity (latent effects) into the formal modeling framework.

The goodness of fit measures (DIC and WAIC) suggested that Model 3, with lagged environmental covariates, provided the best performance. Based on Model 3 (Type B), Fig. 4 depicts the relative contribution of random effects on the relative risks of COVID-19 transmission. We converted the estimated precision parameters of each random effect (District, Time, and Interaction) to ones with standard deviation. Hence, we verified that the random effect from District was small, indicating a larger precision of the corresponding effect, whereas Time had a larger standard deviation, reflecting a greater contribution to the relative risk variation of COVID-19.

From a different perspective, Fig. 5 shows the posterior mean of the fitted values (the relative risks of COVID-19 transmission) of 25 districts in Seoul seven days after the corresponding high PM$_{2.5}$ levels. We verified the lagged effects of PM$_{2.5}$ on the relative risks of COVID-19 infections, as most districts presented elevated risks (relative risks > 1) at the selected dates. Specifically, when the PM$_{2.5}$ alert was issued during the mid-November 2020 in Seoul, we identified several districts with higher relative risks. Furthermore, a combination of temperature drop and elevated levels of PM$_{2.5}$ on December 11, resulted in a substantial increase in the relative risks of COVID-19 across all districts in Seoul on December 18 (bottom center panel of Fig. 5).

4. Conclusion

This study attempts to elucidate the influence of PM$_{2.5}$ concentration and temperature on the daily transmission of COVID-19 across small municipalities/districts in Seoul during the second and third waves. We specifically relied on these periods and 25 districts within Seoul because a recent surge in COVID-19 cases were concentrated in that area, which also represented the evolution of the COVID-19 pandemic in South Korea. To unveil the association between the daily shifts of COVID-19 and environmental conditions such as air pollution and climate, we initially conducted a simple visual inspection of the data. This allowed us to further examine the lagged effects of environmental variables on the progression of COVID-19 over the study region, along with an in-depth investigation of unexplained variability originating from spatial alignments and temporal influence.

Prior to conducting a model-based formal assessment, we tested for correlation among covariates to cope with potential issues of multicollinearity. A correlation matrix was built based on the Pearson correlation, which enabled us to select a subset of variables including PM$_{2.5}$ concentration, temperature, and population density. In light of these understandings, we employed a full Bayesian hierarchical approach for the generalized linear mixed models, which was flexible enough to cope with both fixed effects by covariates and random fluctuations stemming from unobserved heterogeneity.

Of all the models, the DIC and WAIC measures suggested that Model 3 (i.e., one that considered both fixed and random effects) with a 7-day lagged effect of environmental variables (Type B) outperformed the others. Moreover, there exists a positive correlation between a 7-day lagged effect of PM$_{2.5}$ concentration and the number of confirmed COVID-19 cases, indicating an elevated relative risk of the infection. This is consistent with the findings by Chakrabarty et al. (2020) and Vasquez et al. (2020), to name a few. We also verified that, at high temperatures, the relative risks of COVID-19 infections are likely to be lower, supporting the findings of Ward et al. (2020). In addition, we clarified that the random fluctuation in the relative risks of COVID-19 mainly originated from temporal aspects, whereas no significant evidence of variability in relative risks was observed in terms of spatial alignment of the 25 districts. We assure that a systemic breakdown of random effects allows us to improve our understanding of the daily evolution of COVID-19 transmission across small districts in Seoul amidst the pandemic periods, let alone the lagged impacts of PM$_{2.5}$ concentration and temperature.

---

1 Model 2 incorporates the latent influences from each district (i) and day (t) along with our selected covariates (Model 1), while Model 3 further accounts for an unexplained fluctuation from district × day interaction in addition to Model 2.
In this study, however, there are some limitations which can be further improved in future research endeavors. Firstly, of several transmission dynamics of COVID-19 infectious disease discussed previously, this study partially focuses on environment and pollution-to-human transmission mechanisms, whereas the inclusion of various social factors associated with human-to-human transmission is not fully considered. Some studies show the complexity of pandemic’s diffusion patterns typically caused by a multiplicity of social factors such as commercial exchanges (Bontempi, 2020a), GDP and other factors (Duan et al., 2021), and human-to-human interactions (Bontempi, 2021), requiring a multi-dimensional approach (Bontempi et al., 2020). Since our study utilizes population density as a proxy of human interaction, it could be insufficient to reflect real human-to-human transmission mechanisms. Secondly, this study does not account for the effects of countermeasures including social distancing, mask wearing, and personal sanitation, to name a few. Thirdly, literature suggests that there exists a significant association between wind speed and COVID-19 spread (Coccia, 2020c, 2020d, 2021b, 2020d, 2021b). Due to the unavailability of balanced daily dataset, this study could not include wind speed as a candidate variable for statistical analysis. Lastly, several studies show the impact of different waves of the COVID-19 pandemic on society and suggest policy considerations for future infectious disease (Coccia, 2021c, 2021d, 2021d), which motivate us to further examine the disparity of the impact of COVID-19 during the second and third waves of the pandemic in South Korea in a follow-up study. Nevertheless, we argue that our work contributes to providing empirical evidence of the lagged influence of air pollution and temperature along with the effect of unobserved heterogeneity on COVID-19 transmission within Seoul from a different perspective, complementing others in this research domain.

CRediT author statement

Y.L.: conceptualization, methodology, software, data curation, formal analysis, writing — original draft, visualization, investigation, writing — review and editing, and validation. K.-J.K.: conceptualization, data curation, investigation, validation, writing — original draft, visualization, writing — review and editing, and supervision.

Funding

Ki-Jung Kim was supported by a research grant from Doowon Technical University.

Table 4
Regression outcomes (Type A: no lagged effects of environmental variables).

|          | Mean (S.D) | 95 % C.I                                              |
|----------|------------|-------------------------------------------------------|
| Intercep | 0.427 (0.010) | (0.407, 0.446)                                      |
| PM0.15   | 0.204 (0.006) | (0.191, 0.214)                                      |
| Temperature | -0.690 (0.008)  | (-0.706, -0.677)                                    |
| Population Density | 0.002 (0.008)  | (-0.015, 0.015)                                    |

Random effects

|          | Mean (S.D) | 95 % C.I                                              |
|----------|------------|-------------------------------------------------------|
| τ₀       | 19.729 (6.038) | (10.27, 33.78)                                      |
| τ₁       | 0.997 (0.126) | (0.764, 1.26)                                       |

Goodness of fit

Log-Likelihood

|          | Model 1 | Model 2 | Model 3 |
|----------|---------|---------|---------|
| DIC      | 26762.9 | 17368.06 | 14463.7 |
| WAIC     | 26894.71 | 19003.71 | 14363.7 |
| Log-Likelihood | -13448.3 | -9447.92 | -7823.27 |

Note: 1) Model 1: No random effects (fixed effects only); Model 2: Model 1 + IID District + IID Time; Model 3: Model 2 + IID District × Time Interaction 2) 95 % credible interval (CI) is a Bayesian analog of the frequentist confidence interval, but with different interpretation as it directly copes with the probability of true parameters. 3) The inverse of the variance is precision (i.e., \( r^{-1} = \sigma^2 \)). We consider three independent and identically distributed random effects from district (\( \tau_i \)), time (\( \tau_t \)), and their interaction (\( \tau_{it} \)). 4) Number of observations (N × T): 25 districts × 153 days = 3825.

Table 5
Regression outcomes (Type B: 7-day lagged effects of environmental variables).

|          | Mean (S.D) | 95 % C.I                                              |
|----------|------------|-------------------------------------------------------|
| Intercep | 0.409 (0.010) | (0.389, 0.429)                                      |
| PM0.15   | 0.216 (0.006) | (0.204, 0.228)                                      |
| Temperature | 0.695 (0.008)  | (-0.711, -0.679)                                    |
| Population Density | 0.002 (0.008)  | (-0.014, 0.018)                                    |

Random effects

| τ₀       | 19.70 (5.948) | (10.23, 33.41)                                      |
| τ₁       | 1.17 (0.143)  | (0.91, 1.47)                                       |

Goodness of fit

Log-Likelihood

|          | Model 1 | Model 2 | Model 3 |
|----------|---------|---------|---------|
| DIC      | 26046.21 | 17367.41 | 14666.73 |
| WAIC     | 26157.84 | 18997.93 | 14364.67 |
| Log-Likelihood | -13082.87 | -9435.44 | -7815.65 |

Note: 1) We consider a 7-day lagged effect of the covariates (PM2.5 and temperature). 2) Number of observations (N × T): 25 districts × 153 days = 3825.
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to acknowledge two anonymous reviewers for their perceptive comments which materially improved the quality of this article.

References

Ahmadi, M., Sharifi, A., Dorosti, S., Jafarzadeh Ghouschi, S., Ghanbari, N., 2020. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. Sci. Total Environ. 729, 138705. https://doi.org/10.1016/j.scitotenv.2020.138705.

Anand, U., Adelodun, B., Pivato, A., Suresh, S., Indari, O., Jakhmola, S., Jha, H.C., Jha, P. K., Tripathi, V., di Maria, F., 2021a. A review of the presence of SARS-CoV-2 RNA in wastewater and airborne particulates and its use for virus spreading surveillance. Environ. Res. 196, 110929. https://doi.org/10.1016/j.envres.2021.110929.

Anand, U., Cabreros, C., Mal, J., Ballesteros, F., Sillanpää, M., Tripathi, V., Bontempi, E., 2021b. Novel coronavirus disease 2019 (COVID-19) pandemic: from transmission to control with an interdisciplinary vision. Environ. Res. 197, 111126. https://doi.org/10.1016/j.envres.2021.111126.
Seoul Metropolitan City Hall Corona 19 Confirmation Board, 2020. Available from: https://www.mofa.go.kr/viewer/skin/doc.html?fn=20201021031300238.pdf&rs=viewer/result/202103, 01.15.21.
Seoul open data portal, 2021. Available from: https://data.seoul.go.kr/dataList/dataList.do, 06.15.21.
Shen, L., Zhao, T., Wang, H., Liu, J., Bai, Y., Kong, S., Zheng, H., Zhu, Y., Shu, Z., 2021. Importance of meteorology in air pollution events during the city lockdown for COVID-19 in Hubei Province, Central China. Sci. Total Environ. 754, 142227. https://doi.org/10.1016/j.scitotenv.2020.142227.
Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H., 2017. Penalising model component complexity: a principled, practical approach to constructing priors. Stat. Sci. 32 (1), 1–28. https://doi.org/10.1214/16-sts576.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde, A.V., 2002. Bayesian measures of model complexity and fit. J. Roy. Stat. Soc. B 64 (4), 583–639. https://doi.org/10.1111/1467-9868.00355.
Srivastava, A., 2021. COVID-19 and air pollution and meteorology-an intricate relationship: a review. Chemosphere 263, 128297. https://doi.org/10.1016/j.chemosphere.2020.128297.
Vasquez-Apestegui, V., Parra-Garrido, E., Tapia, V., Paz-Aparicio, V.M., Rojas, J.P., Sánchez-Coyillo, O.R., Gonzales, G.F., 2020. Association between Air Pollution in Lima and the High Incidence of COVID-19: Findings from a Post Hoc Analysis. working paper).
Wang, X., Yue, R.Y., Faraway, J.J., 2018. Bayesian Regression Modeling with INLA (Chapman & Hall/CRC Computer Science & Data Analysis), first ed. Chapman and Hall/CRC.
Ward, M.P., Xiao, S., Zhang, Z., 2020. The role of climate during the COVID-19 epidemic in New South Wales, Australia. Transbound. Emerg. Dis. 67 (6), 2315–2317. https://doi.org/10.1111/tbed.13631.
Who, 2021. World health organization WHO coronavirus disease (COVID-19) dashboard. Available from: https://COVID19.who.int/, 03.31.21.
WHO. EURO, 2021. Available from: https://www.euro.who.int/en/health-topics/communicable-diseases/influenza/pandemic-influenza/past-pandemics, 03.31.21.
Wille, C.K., Mangion, A.Z., Cressie, N.A., 2019. Spatio-temporal Statistics with R. CRC Press, Boca Raton, FL.
Wu, X., Nethery, R.C., Sabath, M.B., Braun, D., Dominici, F., 2020a. Air pollution and COVID-19 mortality in the United States: strengths and limitations of an ecological regression analysis. Sci. Adv. 6 (45), eaad4049. https://doi.org/10.1126/sciadv.aad4049.
Wu, X., Nethery, R.C., Sabath, M.B., Braun, D., Dominici, F., 2020b. Exposure to air pollution and COVID-19 mortality in the United States: a nationwide cross-sectional study. medRxiv 2020. https://doi.org/10.1101/2020.04.05.20054502, 04.05.202054502.
Xie, J., Zhu, Y., 2020. Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci. Total Environ. 724, 138201. https://doi.org/10.1016/j.scitotenv.2020.138201.
Xu, K., Oui, K., Young, L.H., Hsieh, Y.K., Wang, Y.F., Zhang, J., Wan, S., 2020. Impact of the COVID-19 event on air quality in Central China, Aerosol Air Qual. Res. 20 (5), 915–929. https://doi.org/10.4209/aair.2020.04.0150.
Yao, Y., Pan, J., Wang, W., Liu, Z., Kan, H., Qiu, Y., Meng, X., Wang, W., 2020. Association of particulate matter pollution and case fatality rate of COVID-19 in 49 Chinese cities. Sci. Total Environ. 741, 140396. https://doi.org/10.1016/j.scitotenv.2020.140396.
Zhu, Y., Xie, J., Huang, F., Cao, L., 2020. Association between short-term exposure to air pollution and COVID-19 infection: evidence from China. Sci. Total Environ. 727, 138704. https://doi.org/10.1016/j.scitotenv.2020.138704,