Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics and Data Science

Master's thesis

Bayesian distributed lag non-linear models (DLNM) to describe association between air pollution and COVID-19 in Belgium

Marina Espinasse

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

Prof. dr. Christel FAES

2022
2023

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Bayesian distributed lag non-linear models (DLNM) to describe association between air pollution and COVID-19 in Belgium

Marina Espinasse

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

Prof. dr. Christel FAES

Bayesian distributed lag non-linear models (DLNM) to describe association between air pollution and COVID-19 in Belgium

Marina Espinasse,

Master of Statistics and Data Science, Biostatistics specialization

Hasselt University, 2023

Thesis promotor: professor Christel Faes, Hasselt University

Submission Date: 19th of June 2023

Contents

Abstract

A recent COVID-19 pandemic has been a deleterious phenomenon that plagued a large part of the world. A substantial amount of research was devoted to understand the determinants of COVID-19 dynamics, and air pollution was suggested to be one of them. The connection between air pollution and COVID-19 was established for several countries across the world, but only a few such analyses were done for Belgium.

Here we investigated the effect of the main air pollutants on COVID-19 incidence in Belgium. We applied an advanced statistical approach designed to study non-linear health responses to a predictor across a range of time lags, such as weeks. In addition, our models incorporated spatial and temporal correlations to account for confounding variability in the data. The study revealed a remarkable similarity in COVID-19 incidence between the adjacent municipalities in Belgium, with a larger incidence in the South-East part of the country than in the North-West. The results of the study showed that the maximal levels of black carbon and ozone increase the relative risk of COVID-19 by about 3-4 times, while median concentrations were associated with about 2 times increase of the risk. We have also observed that there was no or a very short lag between exposure to air pollution and health response, and the strongest effect of air pollution on COVID-19 occurred within the first 2 weeks after exposure. Given our findings, we advise policymakers to pay a careful attention to highly air polluted areas while preparing for future pandemics of respiratory diseases.

1 Introduction

Infectious diseases are a recurring problem in a human population (Dobson and Carper, 1996). An extent of the adverse health effects, economic and societal outcomes due to a single plague can be very consequential, as was exemplified by a recent pandemic of the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2, or COVID-19 disease). Therefore, in the wake of the COVID-19, there is a mounting concern for preventing and alleviating major repercussions related to COVID-19 and similar infectious diseases.

Echoing this concern, researchers worldwide have explored the factors associated with the higher risk of COVID-19 disease (Rashedi et al., 2020; Rod et al., 2020; Zheng et al., 2020). The primary determinant accounting for higher case numbers and mortalities of COVID-19 are host-related risk factors, that is, subject's age (Romero Starke et al., 2020), gender (Mukherjee and Pahan, 2021) and accompanying diseases such as diabetes (Abdi et al., 2020), hypertension (Peng et al., 2021), or cancer (Saini et al., 2020). Besides, there are also environmental and occupational risk factors that pertain to the intensity of social interactions or the conditions that stimulate virus spread, for instance, the lack of personal protective items (masks and gloves) and insufficient ventilation, especially in the enclosed spaces (nursery homes, dormitories, prisons) (Rashedi et al., 2020; Abdelzaher et al., 2020). Change in COVID-19 transmission and therefore, risk, was also found to be affected by climate: regions with steady warm and wet conditions were less strongly affected by COVID-19 (Mecenas et al., 2020; Liu et al., 2021) and accordingly, warm seasons of a year in temperate climates were less favorable for the virus spread (Sajadi et al., 2020).

Although the common risks of COVID-19 are relatively well explored, it is still uncertain why some of the counties and regions had higher incidences and moralities related to COVID-19 that others. Research studies suggested that the discrepancies in the observed COVID-19 cases can be attributed to the additional factors affecting risks and fatalities of respiratory diseases such as COVID-19, proposing air pollution as the plausible candidate (Pansini and Fornacca, 2020; Comunian et al., 2020).

To reflect the importance of air pollution in epidemiological research, we will introduce the common air pollutants, their origin, limit concentrations, and their known effects on human health. Air pollution studies usually consider the

following polluting substances, most abundant and often related to negative health outcomes: black carbon (BC), nitrogen dioxide (NO_2), ozone (O_3), and particulate matter smaller than 10 µm (PM_{10}) and smaller than 2.5 µm ($PM_{2.5}$). Due to potentially harmful effects of exposure to these pollutants (Lelieveld et al., 2015; Bhalla et al., 2014; Anenberg et al., 2012), EU Directive 2008/50/EC has established limit concentrations for some of the substances. For example, annual mean value of the NO_2 in Europe is set to be 40 µg/m^3, while annual means of particulate matter PM_{10} and $PM_{2.5}$ should not exceed 50 and 25 µg/m^3, respectively. No annual limit value is yet established for O_3, but a proposed target value is calculated to be at most 120 µg/m^3 over the consecutive 8 hours (Fierens et al., 2015). No limit values are yet suggested for BC pollution.

Regarding the origin of considered air pollutants, particulate organic matter (PM_{10}, $PM_{2.5}$) pollution emerges from a range of human activities, such as transport, various industries, agriculture, and heating of buildings (Fierens et al., 2015). BC is also a particulate matter, but it originates from combustion processes, and enter the atmosphere as, for instance, diesel soot. NO_2 is originated primarily from the road transport, and to a lesser extent, from energy production, industry and building heating. O_3 on the other hand, is not produced by any human-related activities, but is a sub-product of Volatile Organic Compounds (VOCs) or nitrogen oxides NO_x (Zhang et al., 2019). For example, during hot summer days, NO_2 may break down into NO and a free radical of oxygen ($O-$), which will further react with oxygen and form O_3.

There is compelling evidence of the detrimental health effects of NO_2, O_3, and particulate matter-related pollution. Prolonged exposure to particulate matter pollution induces inflammation processes in the lungs, asthma, and other chronic pulmonary diseases (Bai et al., 2007; Becker and Soukup, 1999; Kim et al., 2018), with O_3 pollution causing similar respiratory dysfunctions (Zhang et al., 2019). Exposure to NO_2 may also induce asthma but in addition, can increase the risk of cardiovascular diseases (Sunyer et al., 2002; Barnett et al., 2006).

Due to a direct effect on lung functions and thus, on the susceptibility of lungs to viral attacks, air pollutants play an important role for the risk of respiratory infections (Comunian et al., 2020). For instance, it was shown that PM_{10} is

an important determinant of influenza prevalence (Morales et al., 2017). The risk of another respiratory disease, SARS, was nearly two times larger in the highly polluted areas of China (high Air Pollution Index), compared to less polluted regions (Cui et al., 2003). Moreover, in addition to increasing humans' frailty to respiratory viruses, air pollution particles may serve as a carrier for viruses, augmenting their viability and transportability in the air (Setti et al.; Comunian et al., 2020). Hence, a higher risk of respiratory infections can be observed in the regions with heavy pollution.

To test the hypothesis that there is a connection between air pollution and higher risks and fatalities of COVID-19, multiple research studies have analyzed pollution levels and concurrent COVID-19 statistics. For example, a study by Zhang et al.(2020) showed that an increase in air pollution by particulate matter and NO_2 (by 10 units of Air Quality index) in China corresponded to 5-7% increase in confirmed COVID-19 cases per day. Another study compared pollution levels by NO_2 and $PM_{2.5}$ with mortality rates from COVID-19 in eight countries. In most of the cases (USA, Italy, Iran, France, UK), the authors found a positive correlation between air pollution and COVID-19 mortality (Pansini and Fornacca, 2020). These findings agreed with other similar studies that tested for air pollution effect on COVID-19 fatalities: in UK (Travaglio et al., 2021), Italy (Ogen, 2020; Coker et al., 2020), and USA (Bashir et al., 2020). In addition to the mentioned examples, there are many more studies from USA, European countries, and China, that corroborated a contribution of air pollution to a larger number of infections and fatalities from COVID-19 (Ali and Islam, 2020; Marquès and Domingo, 2022).

Despite a large number of studies connecting air pollution and COVID-19, only a minor part of published literature featured a similar analysis for Belgium, to the best of our knowledge. Belgium is a country with generally low levels of atmospheric pollution, but limit annual mean concentrations of the common pollutants were sometimes surpassed, particularly, in the urban areas (Fierens et al., 2015). In this context, a possible role of air pollution in the spread of COVID-19 in Belgium should also be evaluated for explaining the observed number of cases and fatalities during the pandemic. Therefore, in this study we aim to examine the potential association between air pollution in the 581 Belgian municipalities and corresponding cases of COVID-19 in year 2021.

In contrast to the previously mentioned studies, many of which used correlation analyses or regression models to relate COVID-19 cases to air pollution (Pansini and Fornacca, 2020; Wu et al., 2020; Bashir et al., 2020; Hendryx and Luo, 2020), we employ an alternative statistical approach to answer the same question. Because earlier research demonstrated that health outcomes (e.g., infection) may be delayed by weeks or months (Copat et al., 2020; Lowe et al., 2021) in response to environmental antecedents such as air pollution or temperature, we assumed that COVID-19 risks may as well be delayed after an exposure to air pollution. Then, to fully account for delayed responses, we utilize a statistical tool specifically designed for such lagged relationships–the Distributed lag non-linear models (DLNM) (Gasparrini et al., 2010), which allow to describe the non-linear responses to exposure and the lagged effects of exposure over a range of selected lags.

By applying this statistical method for COVID-19 cases and air pollution data in Belgium, we intend to answer two specific questions: 1) is air pollution associated with the increased risk of COVID-19 in Belgium and 2) how is the relative risk of COVID-19 affected by air pollution over a range of time lags since exposure?

We expect that our study will shed new light on the involvement of air pollution in the COVID-19 consequences in Belgium and will provide the knowledge base for identifying areas most at risk of COVID-19 or of similar infectious diseases. Thus, the study results may also facilitate a better preparedness for the outbreaks of respiratory infectious diseases in the future.

2 Methods

2.1 Data structure and preparation

Data on COVID-19 cases in Belgium was retrieved from the Infectious Diseases Data Exploration and Visualizations (EPISTAT) repository. The data set on COVID-19 contained information on the total daily cases of COVID-19 in each municipality in Belgium, during 2020 through the first months of 2023. The data set provided no information on the age or any other specific characteristics about people who contracted COVID-19. For the purposes of the present project, we filtered out COVID-19 cases recorded in year 2021. To

aggregate COVID-19 cases per week, we have calculated the sum of COVID-19 cases over the consecutive 7 days for each municipality.

During the preparation of COVID-19 cases data set, we have observed that about 55% of observations were recorded as "< 5" . To be able to include these data in the analyses, we have replaced them with a random draw from the uniform distribution with limits 1 and 4. Being aware that this approach my influence the results of our analyses, we have verified that summary statistics of COVID-19 in Belgium do not change considerably with repeated replacement of "< 5" records. Indeed, the 5 replicates of COVID-19 data sets with replaced "< 5" records differed only in decimal values of their COVID-19 summary statistics. In addition, we also presumed that low values of COVID-19 cases (in the range of 1 to 4 cases) are unlikely to be associated with pollution levels, hence, may not strongly affect the results. Therefore, we did not consider other approaches to replace "< 5" records.

The air pollution data was accessed from the Belgian Interregional Environment Agency data repository (IRCEL-CELINE). The data contained population weighted daily mean concentrations (in µg/m^3) of the following five pollutants : black carbon (BC), Nitrogen Dioxide (NO_2), ozone (O_3), particulate matter (PM_{10} and $PM_{2.5}$). The estimation of all pollutants' concentrations was done by the IRCEL-CELINE Agency using RIO 4x4 km^2 interpolation technique (Janssen et al., 2008). The verified air pollution data was available up to year 2021 only. The air pollution values were also aggregated per week by calculating the average air pollution over the consecutive 7 days for each municipality.

Although, our primary interest was on the effects of pollution of COVID-19, we have also included the vaccination rate data in our analyses, because the vaccination against COVID-19 could be a confounding factor affecting the number of COVID-19 cases. The vaccination data for year 2021 was accessed from the same source as the COVID-19 data set, and contained cumulative number of people per age group and municipality, who received a COVID-19 vaccine of type A (first dose), B(second dose), C(vaccine requiring only 1 dose), E (booster dose), E2 (second booster dose), or E3 (third booster dose). For the current analysis, we did not differentiate between the age groups or vaccine types, instead, we summed the total number of people vaccinated per week and municipality, regardless of age and vaccine type. However, 15% of vaccination

counts were recorded as "< 10", and to make such observations quantitative, we have replaced them with a random draw from a uniform distribution with limits 1 and 9.

Data on the population of each municipality was downloaded from the STATBEL website, and was used for the calculation of the incidence of COVID-19 per 1000 population.

2.2 Distributed lag non-linear models (DLNM)

The epidemiological research that focuses on the health-related reactions to environmental agents, like anomalous temperatures or pollution, is often confronted with a delay in the responses that is hard to estimate by standard statistical methods. However, a relatively unconventional approach, distributed lag models (DLM), allows to address this methodological difficulty.

DLM were originally developed in the field of econometrics to predict industrial capital expenditures from present and past appropriations (Almon, 1965). However, it was later shown that a similar approach can be successfully adopted for epidemiological research, for example, to measure delayed mortality cases due to air pollution (Schwartz, 2000). Subsequent research also demonstrated the usefulness of distributed lag models for relating mortality to temperature fluctuations or humidity (Braga et al., 2001; Armstrong, 2006; Zanobetti et al., 2000), while later studies extended the application of distributed lag models to non-mortality outcomes such as dengue risk (Lowe et al., 2018, 2021) and malaria (Laguna et al., 2017).

Here we will take a look at the methodology of the distributed lag models, and their extension to non-linear distributed lag models (DLNM) (Gasparrini et al., 2010). As a first step, we can recall the formulation of a general time-series model yet ignoring the presence of delayed responses:

$$g(\mu_t) = \alpha + \sum_{j=1}^{J} s_j(x_{tj}; \boldsymbol{\beta_j}) + \sum_{k=1}^{K} \gamma_k u_{tk} \tag{1}$$

where g is a monotonic link function with a corresponding parameter vector $\boldsymbol{\beta_j}$, t is an indicator of the number of time points, s_j is a smooth function of the variable x_j, and u_k and γ_k are other linear predictors and their coefficients,

respectively.

In a distributed lag model, s_j is defined by the use of basis functions, which allow to describe a high degree of complexity of the relationship between the stressor (such as heat) and outcome (mortality). The basis functions are a set of transformations of the original variable x, such as polynomial functions, splines or other transformations producing smoothed curves.

In a general form, the smooth function can be written as follows

$$s(x_t; \boldsymbol{\beta}) = \boldsymbol{z_t^T \beta} \tag{2}$$

where $\boldsymbol{z_t}$ forms the tth row of the basis matrix $\boldsymbol{Z}$ with dimensions $n \times v_x$, which was obtained by applying basis functions to a vector of exposures $\boldsymbol{x}$. In the parametric definition of the basis matrix, usually used for distributed lag models, the dimension v_x is equal to the degrees of freedom, or the degree of flexibility of the function which defines the relationship between exposure and response (Gasparrini, 2011).

In a next step, to account for a delayed effect (outcome in terms of past exposures), we derive L lags of the ordered exposure, that yields a matrix $\boldsymbol{Q}$ with dimensions $n \times (L+1)$, where n is the number of observations and $\boldsymbol{q_t} = [x_t, ..., x_{t-\ell}..., x_{t-L}]]^T$ is a vector of lagged exposures at time t, corresponding to the columns of the matrix $\boldsymbol{Q}$. In addition, to avoid collinearity in the model due to positive correlations between the consecutive exposures, a model allows to constrain distributed lagged functions to have similar effect within lag intervals or to have a continuous polynomial or a spline form (Zanobetti et al., 2000).

Next, basis functions can be applied to the previously defined lags to yield matrix $\boldsymbol{C}$ with dimensions $(L+1) \times v_\ell$. For example, if 10 lags were estimated for a vector of exposures $\boldsymbol{x}$ and then a 3-degree polynomial basis functions were applied to each lag, the resultant matrix $\boldsymbol{C}$ will have 11 rows and 3 columns, corresponding to the number of lags and the number of basis function transformations, respectively. In other cases, basis matrix $\boldsymbol{C}$ can be a vector of ones, for a moving average model, or an identity matrix, for an unconstrained distributed lag model (Gasparrini et al., 2010).

The general form of a smooth function (Equation 2) can now be rewritten as a general definition of the distributed lag models

$$s(x_t; \boldsymbol{\eta}) = \boldsymbol{q_t^T C \eta} \tag{3}$$

where the rows of the basis matrix are replaced by the rows $\boldsymbol{q_t}$ of the lags matrix, and the parameter vector is represented by the matrix $\boldsymbol{C}$ and a vector of unknown parameters $\boldsymbol{\eta}$.

Then the matrix $\boldsymbol{W} = \boldsymbol{QC}$ of transformed variables v_ℓ is included in the design matrix of the model for the estimation of parameters $\boldsymbol{\eta}$. The estimated parameters $\hat{\boldsymbol{\eta}}$ are subsequently used for the interpretation of the linear effects $\boldsymbol{\beta}$ at each lag:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{C\hat{\eta}} \tag{4}$$

And the variance of $\hat{\boldsymbol{\beta}}$ is estimated as:

$$V(\hat{\boldsymbol{\beta}}) = \boldsymbol{CV(\hat{\eta})C^T} \tag{5}$$

The illustrations above are relevant for delayed responses that can be described by relatively simple linear relationships with exposure, but a further extension that relaxes the assumption of linearity can be made through distributed lag non-linear models (DLNM). The concept of DLNM is based on the *cross-basis* functions, which are also basis functions as exemplified earlier but they are bi-dimensional, capturing the function of exposure-response relationship across the covariate x (e.g., temperature) and its lagged effects.

The initial steps in the estimating cross-basis function are similar to the steps of the linear distributed lag model. First, we apply basis function transformation to x (strata, natural cubic splines, linear threshold or polynomial) and create a matrix of transformed covariates $\boldsymbol{Z}$. Next, we define lags of the previously derived basis functions in the matrix $\boldsymbol{Z}$, to yield an array $\boldsymbol{R}$ of dimension $n \times v_x \times (L+1)$, where n is the number of observations, v_x is the number of the basis function transformations, and L is the maximal lag applied. Then, we also obtain a matrix $\boldsymbol{C}$ of basis function transformations of the lag vector ℓ, and finally define a DLNM as follows:

$$s(x_t; \boldsymbol{\eta}) = \sum_{j=1}^{v_x} \sum_{k=1}^{v_\ell} \boldsymbol{r_{tj.}^T c_{.k}} \eta_{jk} = \boldsymbol{w_{t.}^T \eta} \tag{6}$$

which represents the cross-basis function of the DLNM (Gasparrini et al., 2010). The selection of the basis function transformation for exposure-response and lagged response depends on the analysis objectives, but simpler and more interpretable models are produced by lag-stratified basis function transformations than by natural cubic splines transformations (Armstrong, 2006).

The estimation of effects from DLNM requires a technically heavy description, but the main attributes are that the effects of exposure at each lag can be estimated, as well as the effect of a single level of exposure across all lags. In other words, we can estimate the mortality by temperature at lag 1, 5, or 10 days, and similarly, we can estimate the mortality by lag, when temperature was 10, 15 or 25°C (Armstrong, 2006).

Since DLNM allows flexibility in choosing basis function transformation and associated parameters (e.g., the number of knots for natural cubic spline transformation), the justification of selected parameters is warranted. Gasparini et al (2010) propose the use of a modified AIC and BIC, tailored for quasi-likelihood models, when overdispersed observations are assumed. In other cases, standard AIC can be used (Armstrong, 2006) or a DIC when Bayesian techniques are applied for parameter estimations (Lowe et al., 2021).

In this study, we apply DLNM to describe the delayed health responses (COVID-19 infection) to air pollution. The exact model formulation, the choice of basis functions and model comparison criteria will be described in detail in the Section 2.4.

2.3 Integrated Nested Laplace Approximation

Bayesian methods have received a great popularity in statistical research, covering the fields of ecology, environment, and medical sciences (Congdon, 2019; Kéry and Royle, 2020; Qian et al., 2022). Formerly, Bayesian approaches were implemented only using Markov-Chain Monte Carlo (MCMC) techniques involving one or another type of sampling from the posterior (e.g., Gibbs sampling, random walk Metropolis-Hastings). Although MCMC-based methods are flexible and functional, they are not without complications. First, significant computing power is required for most of the MCMC procedures, which makes them relatively slow, especially for multi-parameter hierarchical models. Secondly, convergence to a true posterior can be difficult to achieve and evaluate objectively

(Lesaffre and Lawson, 2012).

An alternative approach for Bayesian estimation was proposed by Rue et al. (2009; 2017). This alternative method featured the use of Integrated Nested Laplace Approximation (INLA) to marginal posterior distributions. The INLA procedure was designed to be used for a large class of latent Gaussian models (LGM) and was shown to provide more accurate posterior estimations for these models than MCMC methods. LGM type of models comprises a range of widely applicable models such as linear and generalized linear modes, spatial and spatial-temporal models, models for disease mapping and many more (Wang et al., 2018), allowing the use of INLA for various statistical problems.

INLA makes several important assumptions about the model parameters. First, observations $\boldsymbol{y}$ are assumed to be conditionally independent given the latent field $\boldsymbol{\eta}$ (vector of model covariates together with random effects if present) and its vector of hyperparameters $\boldsymbol{\theta_1}$ (Equation 7). Second, the latent field $\boldsymbol{\eta}$ of the model must be a latent Gaussian random field (GRMF, Rue and Held (2005)), which follows a Normal distribution with a sparse precision matrix $\boldsymbol{Q^{-1}}$ (Rue et al., 2009) (Equation 8). Finally, the vector of observations' hyperparameters $\boldsymbol{\theta_1}$ and the vector of the latent field's hyperparameters $\boldsymbol{\theta_2}$ follow a joint prior distribution $\pi(\boldsymbol{\theta})$ (Equation 9), where $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$.

$$L(\boldsymbol{\eta}, \boldsymbol{\theta_1}|\boldsymbol{y}) = \pi(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta_1}) = \prod_{i \in I} \pi(y_i|\eta_i, \boldsymbol{\theta_1}) \tag{7}$$

$$\boldsymbol{\eta}|\boldsymbol{\theta_2} \sim \mathcal{N}(0, \boldsymbol{Q^{-1}(\theta_2)}) \tag{8}$$

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) \tag{9}$$

Next, the joint posterior distribution of the latent field and associated hyperparameters of the model is formulated as follows:

$$p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\eta}|\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta}) \tag{10}$$

The interest lies in estimating posterior marginal distributions of individual latent effects $p(\eta_i|\boldsymbol{y}), i = 1....n$ and marginal distributions of hyperparameters $p(\theta_j|\boldsymbol{y}), j = 1, 2$. To obtain these marginals, INLA uses a series of approximations that will be described below.

As a first step, INLA proposes an approximation $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y})$ to $p(\boldsymbol{\theta}|\boldsymbol{y})$ as:

$$\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{p(\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{p}(\boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{y}))}\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*(\boldsymbol{\theta})} \tag{11}$$

This is a Laplace approximation of a joint posterior of hyperparameters. Here, $\tilde{p}(\boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{y})$ is a Gaussian approximation to the full conditional of $\boldsymbol{\eta}$, and $\boldsymbol{\eta}^*$ is the mode of this full conditional given the vector $\boldsymbol{\theta}$. Taylor series expansion is used for this approximation, and usually, a log transformation of the hyperparameters $\boldsymbol{\theta}$ is applied to facilitate approximation.

In a second step, INLA approximates marginal distributions of latent effects $p(\eta_i|\boldsymbol{\theta}, \boldsymbol{y})$. Rue et al. (2009) describes three approaches to approximate $p(\eta_i|\boldsymbol{\theta}, \boldsymbol{y})$. The first one relies on GMRF property of the latent filed and uses Gaussian approximation to estimate marginal mean and variance. Although Gaussian approximation is the fastest, it is often inaccurate (Rue and Martino, 2007). Therefore, the second approach corrects for inaccuracy of the Gaussian approximation by again using Laplace approximation, as a result providing accurate and exact estimations. On the other hand, the second method is computationally extensive, while a third method–simplified Laplace approximation–benefits from accuracy of the Laplace approximation but requires less computation time. Hence, simplified Laplace approximation is a default option in INLA, and in most cases it yields sufficiently accurate results.

In a third step, INLA's algorithm explores the $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y})$ to use it in the numerical integration of $p(\eta_i|\boldsymbol{y})$. This step requires reparametrization of $\boldsymbol{\theta}$-space, which involves the computation of the negative Hessian matrix $\boldsymbol{H^{-1}}$ at the mode of $\boldsymbol{\theta}$ and eigenvalue decomposition of $\boldsymbol{H^{-1}}$ as $\boldsymbol{H^{-1}} = \boldsymbol{V\Lambda V'}$. Decomposition $\boldsymbol{H^{-1}}$ allows to obtain a new variable $\boldsymbol{z} = (\boldsymbol{V\Lambda^{1/2}})^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)$. This z-parameterization is used to explore the space of $\boldsymbol{\theta}$ either using a regular grid centered at the mode ($z = 0$) or using a set of strategically placed points (a central composite design method, or CCD). The CCD is more efficient than the regular grid approach when the number of hyperparameters is larger than 2. But both methods can be computationally intensive, therefore, there is an alternative exploration method—empirical Bayes, where modal configuration is used for the integration over $p(\boldsymbol{\theta}|\boldsymbol{y})$. The empirical Bayes approach to explore $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y})$ performs well when variability of the hyperparameters is not too large.

Finally, as a result of exploration of $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y})$ the selected set of integration

points $\boldsymbol{\theta}^{(k)}$ and their respective weights $\Delta\boldsymbol{\theta}^{(k)}$ are used for integrating the hyperparameters and marginalizing over the latent effects (Equation 12).

$$\pi(\eta_i|\boldsymbol{y}) \simeq \sum_{k=1}^{K} \tilde{\pi}(\eta_i|\boldsymbol{\theta}^{(k)}, \boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta}^{(k)}|\boldsymbol{y})\Delta\boldsymbol{\theta}^{(k)} \tag{12}$$

The marginals of hyperparameters are obtained by numerical integration of $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y})$, where $\boldsymbol{\theta_{-j}}$ is the vector of $\boldsymbol{\theta}$ excluding the j^{th} element (Equation 13). Alternatively, Laplace approximation is applied to estimate the same marginal densities.

$$\tilde{p}(\boldsymbol{\theta_j}|\boldsymbol{y}) = \int \tilde{p}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta_{-j}} \tag{13}$$

The details on the Bayesian model parameterization in INLA and the choice of prior and likelihood are described in detail in the Section 2.4.

2.4 The formulation of Bayesian DLNM with INLA

Hierarchical structure of the models

We began model formulation by choosing the fundamental structure and the appropriate likelihood for the model. As the data likelihood we used Negative Binomial distribution, to account for a likely overdispersion of COVID-19 cases in our data. We have also tested the Poisson likelihood but ruled it out early in the model selection process.

Designing a hierarchical structure for our models, we decided to include both spatial and temporal random effects. The temporal random effect was considered for a base model because we have data spanning from the beginning to the end of the year 2021. Hence, we expected a temporal dependency in the COVID-19 cases from one month to another and from one week to the next within a month. The temporal random effect was specified as the random walk model of order 2 (RW2), which is a common choice for the analysis of spatial-temporal epidemiology data (Blangiardo et al., 2013; Moraga, 2019). RW2 model represents a nonparametric temporal model that assumes independence of the second-order increments.

$$\gamma_t|\gamma_{t-1}, \gamma_{t-2} \sim \mathcal{N}(2\gamma t - 1 + \gamma_{t-2}, \sigma^2) \tag{14}$$

Next, as the COVID-19 data was provided per municipality, we presumed that ignoring a possible similarity in COVID-19 incidences between the neighboring municipalities would be incorrect. This is a common concern in the spatial health data, such as ours, and the conventional approach to accommodate spatial dependency is to use Conditional Autoregressive type of random-effects models, also called Besag models (Besag et al., 1991).

The standard Besag model is based on the correlated spatial random effect u_i, which has a conditional normal distribution:

$$b_i|\mathbf{b_i}, \tau_b \sim \mathcal{N}(\frac{1}{n_{\delta_i}} \sum_{j \in \delta_i} b_j, \frac{1}{n_{\delta_i} \tau_b}) \tag{15}$$

where b_i is a single spatial unit (e.g., a municipality), n is the number of neighbors surrounding spatial unit b_i, τ_b is the total spatial variance among all the spatial units, and δ_i is a set of neighbors of a spatial unit b_i.

The advantage of this model is that interpretation of the spatial random effect is intuitive: the mean of the conditional distribution for a spatial unit b_i is equal to the average spatial component value over its neighbors, while the variance is proportional to the total variance divided by the number of neighbors. In the other words, the larger the number of neighbors, the lower will be the variance (and higher the precision) of a spatial unit b_i and the more information can be obtained about it. The number of neighbors for each spatial unit is usually defined using a common boundary point approach (Qeen's case), where the two municipalities that share one or several points of their borders, are considered as neighbors.

Although the Besag model is convenient and relatively easy to apply, it is limited to spatially correlated random effects. Often, however, spatial data features also an uncorrelated, a stochastic spatial variation, which cannot be described by u_i (Riebler et al., 2016). To avoid underestimation of the spatial random effects, the Besag York and Mollie (BYM) model was proposed (Besag et al., 1991). In this model, spatial effect $\mathbf{b}$ is decomposed into the sum of the two components: a spatially structured $\mathbf{u}$ with a CAR prior distribution as in the Besag model, and an unstructured, iid, component $\mathbf{v}$.

$$\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \tau_u^{-1}\mathbf{Q}^-) \tag{16}$$
$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \tau_v^{-1}\mathbf{I}) \tag{17}$$

where $\boldsymbol{Q}$ is the precision matrix of $\mathbf{u}$.

BYM model though also has a small flaw–the lack of scaling of a precision prior, due to which application of the same prior to different spatial structures results in different marginal variances (Riebler et al., 2016). This shortcoming was corrected in a BYM2 model proposed by Simpson (2017). Unlike the BYM original model, BYM2 model contains a scaled spatial component $\boldsymbol{u_*}$:

$$\mathbf{b} = \frac{1}{\sqrt{\tau_b}}(\sqrt{1-\phi}\mathbf{v} + \sqrt{\phi}\mathbf{u}_*) \tag{18}$$

where τ_b is a common variance for both spatial components and ϕ is a mixing parameter.

While in the BYM model $\mathbf{u}$ and $\mathbf{v}$ are difficult to interpret separately, in the BYM2 model the mixing parameter ϕ simplifies this interpretation. The value of ϕ corresponds to the proportion of a total spatial variance explained by the $\mathbf{u}$ and $\mathbf{v}$. For example, when ϕ is close to 1, most of the spatial variation in the data is explained by the scaled structured component $\mathbf{u}_*$, while very small values of ϕ indicate that spatial variation is mainly a stochastic spatial noise. In addition, BYM2 model facilitates the use of penalized complexity (PC) priors for the precision of random effects. PC priors are preferable for hierarchical Bayesian models, because they are robust, stable under reparameterizations, and favor more parsimonious models (Simpson et al., 2017). Therefore, for our analysis, we preferred BYM2 spatial model for the spatial random component.

We used PC priors both for the hyperparameters of the spatial random effect (ϕ and precision τ_b), and also for the precision of the RW2 temporal model. The PC prior for τ_b was defined conventionally, as $p(\frac{1}{\sqrt{\tau_b}}) > 0.5 = 0.01$ (Moraga, 2019; Simpson et al., 2017). And the same prior was used for the precision of the RW2 random effect. A default PC prior was used for the mixing parameters ϕ, defined as $P(\phi < 0.5) = 0.5$.

Finally, the adjacency structure for our BYM2 model was defined using `poly2nb` function in R. The resultant spatial structure is shown in the Supplementary Figure 6.

Cross-basis function definition

For the main effects (5 pollutants and a vaccination rate) we calculated lagged effects, up to 6 weeks lags for pollution and up to 2 weeks lag for vaccination. Based on these lagged variables we defined a cross-basis function for each covariate using function `crossbasis` in the `dlnm` package in R. Cross-basis function applies a desired smooth function on the main effect of a covariate and on the lagged vectors, then combining both effects in a design matrix $\mathbf{w_{t.}^T}$ mentioned in the Equation 6.

For our data, we chose natural spline smooth function for both exposure-response (no lag) effect and for lag-response effect. We set 2 equally spaced knots for the smooth function of exposure-response, resulting in 3 basis functions (2 knots + 1), and 1 knot was specified for lag-response, also resulting in 3 basis functions (1 knots + 1 + intercept) (Gasparrini, 2011). The final cross-basis functions for each covariate had $3 * 3 = 9$ basis functions.

For the main effects of pollution and vaccination we did not manipulate the priors, using default INLA priors (Normally distributed priors with zero mean and large variance).

DLNM general formula

The base DLNM was formulated in INLA as follows:

$$y_{ij}|\mu_{ij}, \kappa \sim NegBin(\mu_{ij}, \kappa) \tag{19}$$
$$log(\mu_{ij}) = log(p_{ij}) + log(pop_{ij}) \tag{20}$$
$$log(p_{ij}) = u_i + v_i + \alpha_j + f.w(x, l) \tag{21}$$

here, μ_{ij} is the Negative Binomial distribution mean; κ is the overdispersion parameter of the Negative Binomial distribution; $log(p_{ij})$ is the COVID-19 incidence in a municipality i and year j; $log(pop_{ij})$ is the population of a municipality divided by 1000 (an offset); u_i is the spatially structured random effect and v_i is a stochastic spatial random effect; α_j is the RW2 random effect of year; $f.w(x, l)$ is the exposure-lag-response (cross-basis) function, or a combination between $f(x)$, the function describing non-linear response to a predictor (e.g., pollutant) and $w(l)$, the lag-response function of delayed responses to a predictor over the maximal number of lags l.

Model selection process

In a step 1 of a model selection process, we constructed the base model that included only the spatial and temporal random effects and then added one covariate to this model. We then compared the relative contribution of each covariate to the model fit. In the second step, we added a second covariate to the model selected in the first step and again compared the goodness of fit of the candidate models.

For comparing model fit we used Deviance Information Criterion (Spiegelhalter et al., 2002), where smaller value indicates a better fit. In addition, we calculated the Conditional Predictive Ordinate (Congdon, 2019). CPO values indicate a conditional probability of a single value y_i predicted by the model that was fit without value y_i. CPO is large for y_i values well predicted by the model, but for the ease of model comparisons, we calculated the log score, which is the minus mean of the log of CPO values, and the lower log score complies with a better model fit.

All analyses were performed in the R software version 4.2.2 using packages `r-inla` (version 23.01.12) and `dlnm` (version 2.4.7).

3 Results

3.1 Exploratory analysis

As a part of the exploratory analysis, we estimated monthly incidence (per 1000 citizens) of COVID19 in Belgian municipalities in 2021 (Figure 1). A cursory look at the Figure 1 reveals that during 2021 there was a considerable temporal variation of COVID-19 incidence: most of the country's municipalities had highest incidence in November and December, while June and July had overall lowest incidences. This observation suggested a strong seasonality of COVID-19, also indicating that a temporal component of COVID-19 cases must be accounted for in the statistical models. On the other hand, within a single month, we could observe large differences in COVID-19 incidence among the municipalities. For example in September 2021, higher incidences were observed in the eastern municipalities (Liege province, Weismes, Plombières, Trois-Ponts) and in the South (Luxembourg province, Rouvroy and Meix-devant-Vitron).

On the contrary, northern and westerns part of the country (West Flanders, East Flanders and Antwerpen provinces) had generally lower COVID-19 incidences in September, with no cases recorded in municipalities Vleteren, Horbeke, and Maarkedal. Spatial dissimilarity of COVID-19 incidence could also be noticed in June through July, while less remarkable spatial differences were observed in winter. The presence of a spatial heterogeneity of COVID-19 cases across Belgium in 2021 could not be ignored in the statistical analyses, therefore it was further translated into a random spatial component of the statistical models, as described in the Methods section 2.4.

Pollution levels also varied spatially across the country. The extreme records of BC pollution in 2021 (above 1.9 µg/m^3) were observed in municipalities Gent, Bredene, Antwerpen, and Sint-Gillis in the western and northern parts of the country (West Flanders and Antwerpen provinces). Most NO_2 polluted (40 µg/m^3 or above) municipalities in 2021 were Sint-Gillis and Sint-Jans-Molenbeek in the Brussels province. PM_{10} pollution was most severe (above 45 µg/m^3) in Oudsbergen and Houthalen-Helchteren in the North-East of the country (Limburg province), while $PM_{2.5}$ was highest (above 28 µg/m^3) in Kortrijk and Deerlijk (West Flanders province) . Among the municipalities with the highest O_3 pollution (above 82 µg/m^3) in 2021 were Habay, Meix-devant-Virton, Attert, and Musson, all located in the Luxembourg province in the South-East.

During exploratory data analysis, we have checked for collinearity between pollutants, as their concentrations may be mutually dependent (Fierens et al., 2015). A strong positive correlation of 0.7-0.9 was detected between $PM_{2.5}$ and PM_{10} pollutants, and between NO_2 and BC pollutants (Supplementary Figure 5). Therefore during the model selection process we kept in mind that these pairs of pollutants could not be included in the same model.

3.2 Model selection

The results of the first step of a model selection process are shown in the Table 1. We compared the set of models including only 1 covariate (e.g., only BC or only O_3). Among these models, a model with BC was the best (Table 1). Therefore in the next step, we added another covariate to the model already containing BC. We observed that adding O_3 to this model improved the fit (had lower DIC and log score), while adding vaccination rate did not change

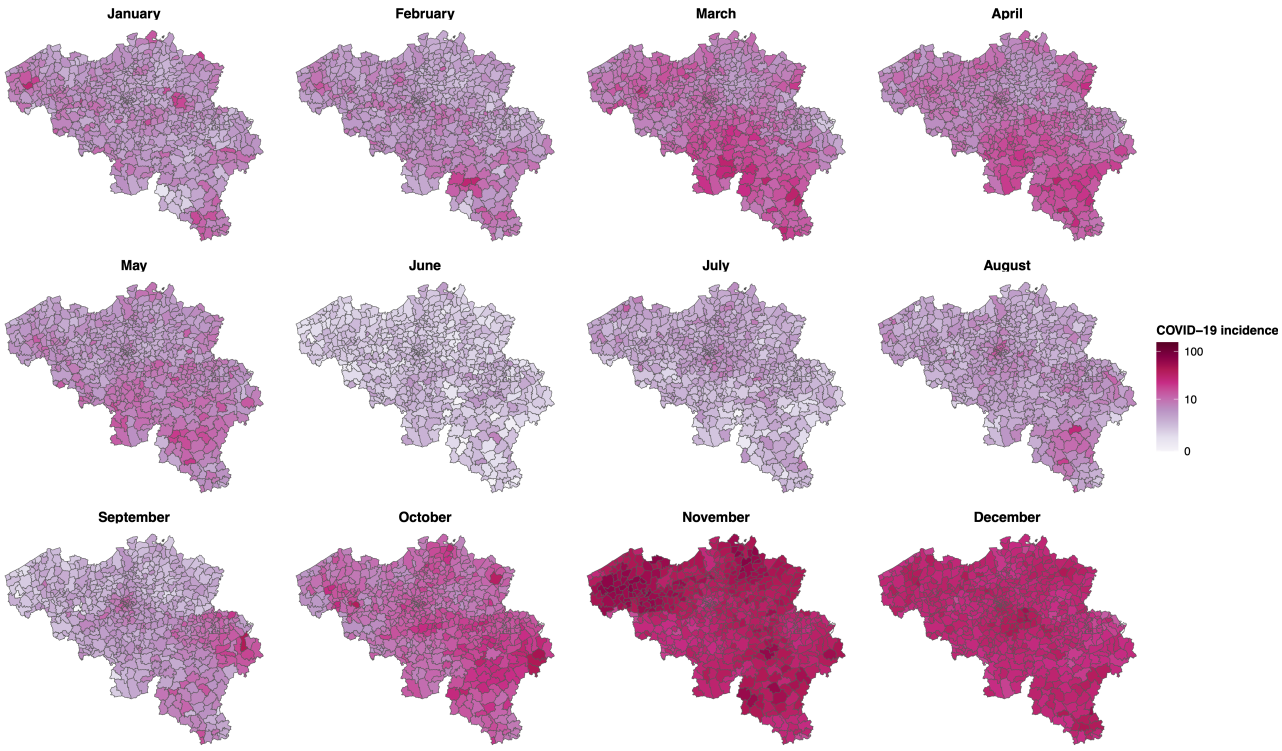

Figure 1: Incidence rate of COVID-19 infection per 1000 population, in Belgium in 2021

the model fit (Table 2). We have considered also the models with NO_2 and PM_{10} or $PM_{2.5}$ as a second covariate in addition to BC, but because all three pollutants were strongly positively correlated with BC, we did not include them in the final model to avoid multicollinearity. Thus, we have selected a model with BC and O_3 as the final.

3.3 Selected DLNM results

Our selected DLNM contained two types of random effects: a spatial random effect BYM2 and a temporal trend RW2. For the spatial random effect, the mixing parameter ϕ was equal to 0.88, indicating that a large part of the spatial variation was attributed to a structured, spatially correlated random effect. In other words, it means that spatial similarity between the adjacent municipalities was relatively large in our COVID-19 data. This result is also reflected by the

Table 1: The results of the first step of the model selection process: the base model compared with the models including one additional covariate. DIC stands for Deviance Information Criterion; log score is the minus mean of the log Conditional Predictive Ordinate. The mathematical formulation of the models is the same as shown in Equation 21.

Model components	**DIC**	**Log score**
Base	229739	3.8033
Base + O_3	229371	3.7972
Base + BC	228729	3.7868
Base + NO_2	229214	3.7946
Base + PM_{10}	229402	3.7977
Base + $PM_{2.5}$	226531	3.7998
Base + Vaccination	229734	3.8032

Table 2: The results of the second step of the model selection process: the model model including BC is compared with the models additionally including O_3 and vaccination. DIC stands for Deviance Information Criterion; log score is the minus mean of the log Conditional Predictive Ordinate. The mathematical formulation of the models is the same as shown in Equation 21.

Model components	**DIC**	**Log score**
Base + BC	228729	3.7868
Base + BC + O_3	228016	3.7750
Base + BC + Vaccination	228721	3.7866

map of spatial random effects, where neighboring municipalities tend to have similar values of the random effect (Figure 2A).

A prominent separation can be observed between the municipalities with a positive spatial component, that is, on average, larger COVID-19 cases, and a negative spatial random effect. Therefore, nearly all the municipalities located in the North and North-East of the country (the Flanders Region), had lower cases of COVID-19 than on average in the country. On the contrary, most of the municipalities in the South-West (the Walloon Region) had on average more cases of COVID-19. But the amplitude of differences in COVID-19 counts implied by spatial random effects was not large: from at about 3 times lower to 3 times larger number of COVID-19 cases, while for the majority of municipalities the difference in the number of cases was within 50%.

The temporal random effect RW2 demonstrated well the seasonality of COVID-19 (Figure 2B). The two extreme deviations from the average COVID-19 incidence were observed in week 25 (June 21, 2021) and in week 51 (November 25, 2021). The summer drop in COVID-19 lasted for about 2 weeks after which it stabilized around the average in August-October until an abrupt increase in

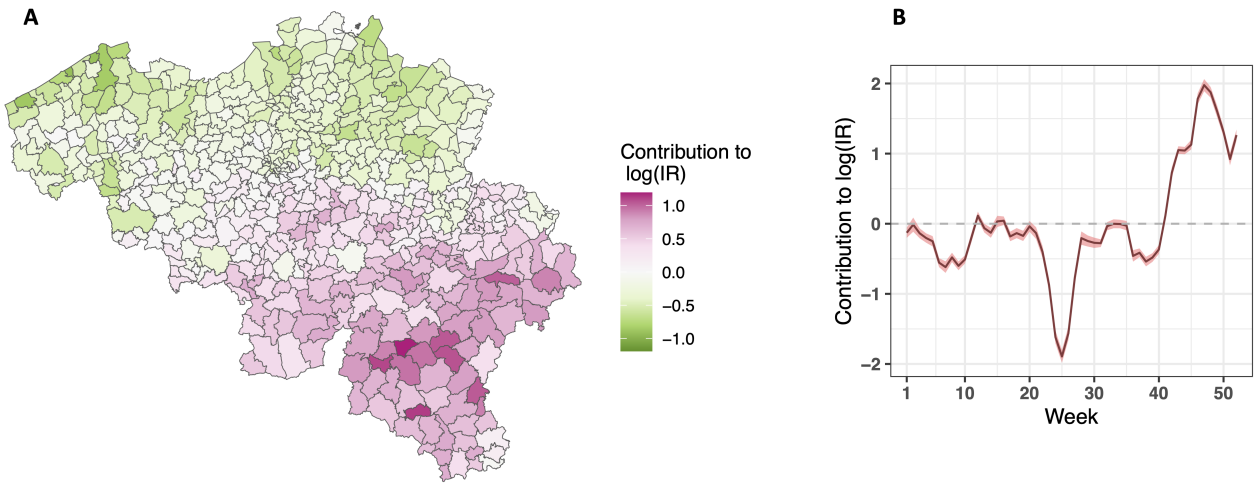

Figure 2: (A) spatial random effects of the selected Bayesian DLNM. This spatial effect denotes municipality-specific deviations from the average COVID-19 incidence; (B) temporal random effect type RW2. The temporal effect denotes week-specific deviations from the average COVID-19 incidence.

November, a month with the largest COVID-19 incidence in Belgium in 2021 (Figure 1) . We noticed also that a temporal random effect was not very smooth across the study timeline. This characteristic of the temporal trend can be due to inherent variability of our data and to our choice of default priors, which allowed a relatively large marginal variance of the temporal component.

Both temporal and spatial random effects, therefore, demonstrated that there was a strong temporal and spatial dependency in the COVID-19 data, as we correctly presumed during the analysis design.

The plots of the association between the risk of COVID-19 and *BC* concentrations across the lag of 6 weeks, are shown on the Figure 3. The strongest effect of *BC* (more than 4 times increase in the relative risk) was observed for the highest concentrations, above 2 µg/m^3. The *BC* concentrations between 0.5 to 1.5 µg/m^3 were also associated with a large increase in the relative risk: from about 30% to 2.5 times increase during the first 2 weeks after exposure (Figure 3A). Median concentrations of *BC* were associated with less than 2 times increase in the relative risk of COVID-19 (Figure 3B), and the relative risk decreased to 1 at the lag of 5 weeks. *BC* concentrations below 0.5 µg/m^3 were not predicted to have a measurable increase in the RR.

Both plots indicated that the effect of *BC* on the relative risk of COVID-19 can be expected during the first few weeks of exposure, so there was no delay

or only a short delay in the health response to this pollutant, but the effect remained for 5 weeks after exposure.

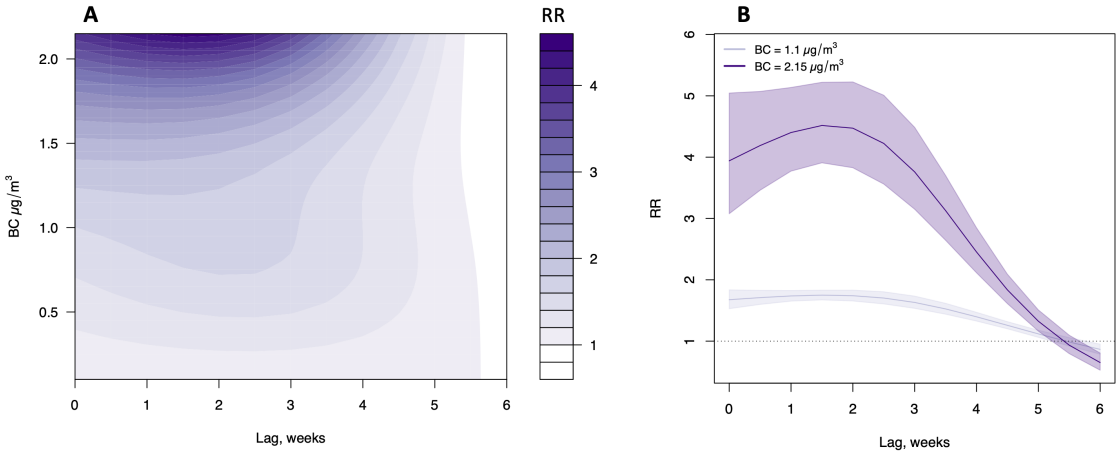

Figure 3: Lag-response of COVID-19 to Black Carbon (BC) pollution. (A) contour plot of the association between the relative risk of COVID-19 (RR) and the level of BC pollution, relative to the minimum observed pollution. (B) lag-response association between the risk of COVID-19 and two levels of BC pollution: the median level of pollution (light purple) and the highest level of pollution (dark purple).

The contour and lag-response plots for O_3 (Figure 4) showed generally similar effects on COVID-19 as BC. The largest concentrations of O_3 were associated with 3.6 times increase in the relative risk of COVID-19 during the first 2 weeks, but more realistic, median concentrations of O_3 increased the risk by at most 2 times during the first 2 weeks after exposure (Figure 4B). For O_3, there were also no delay in the response, but the effect diminished at about week 5, like the effect of BC (Figure 4A).

In sum, evaluating DLNM results for both pollutants, we may underline that COVID-19 risk induced by air pollution was increased without a long delay, and the strongest effect was predicted for the first 2 weeks after exposure with virtually no effect at a lag of 6 weeks .

The cumulative effect of both air pollutants is shown in the Supplementary Figure 7. At most 16 times increase in the risk of COVID-19 was predicted given cumulative exposure to O_3 . However, for BC, only small cumulative effects are shown for low concentrations of the pollutant, but an abrupt increase in the cumulative relative risk is predicted at the concentrations above 1.7 µg/m^3. This abrupt change maybe due to instability of the prediction, as the BC concentrations larger than 2 µg/m^3 were observed only a few times in the data

set used for the analysis.

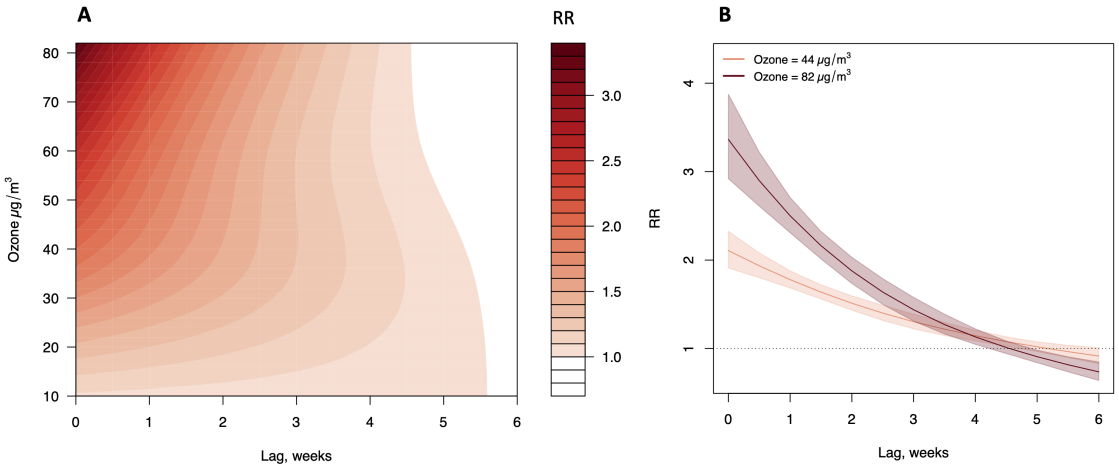

Figure 4: Lag-response of COVID-19 to ozone (O_3) pollution. (A) contour plot of the association between the relative risk of COVID-19 (RR) and the level of O_3 pollution, relative to the minimum observed pollution. (B) lag-response association between the risk of COVID-19 and two levels of O_3 pollution: the median level of pollution (light red) and the highest level of pollution (dark red).

4 Discussion

Our study attempted to test the association between air pollution and COVID-19 incidence in Belgium and to identify how such associations change across time lags. Although we did not answer these questions for each of the air pollutants prevailing in the country, we discovered a strong effect of BC and O_3 on COVID-19 risk during 1-2 weeks after exposure, while only a modest effect was found at 3-5 weeks lag after exposure. In addition, our study revealed a strong spatial correlation between the adjacent municipalities in terms of COVID-19 incidence, which can be attributed to air pollution as well as additional factors not measured in this study.

4.1 The spatial structure of COVID-19 cases in Belgium

Our spatial random effect demonstrated a clear spatial correlation between the COVID-19 incidence per municipality (Figure 2). Similarities were observed within the clusters of municipalities on a small scale, but also within the two Belgian regions–the Flanders and the Walloon (Wallonia) (Supplementary

Figure 8). The generally smaller spatial random effect for municipalities in Flanders indicated on average lower count of cases in this region in 2021 compared to the Walloon.

There can be multiple reasons why COVID-19 affected Belgian municipalities differently. Among the likely factors proposed by the preceding research are differences in socio-economic and demographic factors (Verwimp et al., 2020; Gadeyne et al., 2021; Natalia et al., 2022). For example, studies by Meurisse et al. (2022) and Verwimp (2020) connected higher COVID-19 incidence to municipalities with a lower socio-economic status. One of the socio-economic factors is income, and in relation to COVID-19, lower income households may have less health literacy or can be exposed to higher COVID-19 risk at work because their, usually low-skilled, jobs do not permit distance working. Hence, because the municipalities of Walloon have comparatively lower income than the municipalities of Flanders (STATBEL), the Walloon region might have been more vulnerable to COVID-19. Moreover, on top of the socio-economic determinants, we believe that limited connectivity between the two regions related to work, study or leisure travels, have added to the large differences in the observed COVID-19 counts.

Another factor potentially responsible for strong spatial separation is population density, where, contrary to common sense, less densely populated areas had higher incidence of COVID-19. This discrepancy was due to inhabitants of such municipalities taking fewer precautions under a false impression that COVID-19 transmission is less likely outside cities (Meurisse et al., 2022). Indeed, population density is slightly lower in Walloon (STATBEL), which may have contributed to a larger incidence of COVID-19 in this region in 2021.

Last, a geographical aspect may have also been instrumental in increasing COVID-19 incidence in Walloon. The Wallon region share borders with both Germany and France, the two countries that were often high on COVID-19 charts in 2021 (Our World in Data). Because national borders of the European Union countries are open for the Union citizens, cross-border mobility, in our case between Belgium and Germany or Belgium and France, could have been an auxiliary transmission vector of COVID-19 (Wilder-Smith, 2021).

In the context of our analysis, we have also considered pollution as a potential factor driving COVID-19 variability across Belgium. Interestingly, the

IRCEL-CELINE air pollution report (Fierens et al., 2015) indicates generally higher 8-hour levels of O_3 pollution in the North-Eastern area of Flanders and in the nearly whole area of Walloon. Similarly, the highest average weekly O_3 levels (above 80 µg/m^3) were observed mainly in the Walloon region based on 2021 data. The BC pollution, however, was not strikingly different between the two regions (Fierens et al., 2015), and was the largest in some of the municipalities in Flanders. Thus, we think that the O_3 pollution was not the sole driver of COVID-19 spatial variability, but to some extent, it was associated with higher COVID-19 incidence in the South-East of Belgium.

4.2 The effects of air pollution on COVID-19 in Belgium

Our study put forward BC pollution in Belgium as one of the co-factors of COVID-19 incidence. Connecting our findings to the available literature, we discovered that there are only a small number of studies that explicitly analyzed BC contribution to COVID-19 incidence or fatalities. For example, a study by Rathod et al. (2021) based in India, analyzed the BC pollution data and COVID-19 dynamics in 2020. They revealed that a season of intensive BC production in Delhi area, due to residual crop combustion, has played a significant role in the surge of COVID-19 cases in October-November 2020. Moreover, they also indicated that the increased BC pollution have boosted COVID-19 mortality rates with a lag of 10 days, implying that the strongest effect of BC occurred within the 2 weeks after exposure, as our study also demonstrated. Another research, presented at the International Society for Environmental Epidemiology (ISEE) in September 2022, showed that long-term exposure to BC in Denmark was associated with about 10% increase in the hazard of COVID-19-related death, and a 10% increase in the hazard of hospitalization (Zhang et al., 2022). In addition, we have identified a research work from Belgium leaded by Vos et al. (2022), also presented at the ISEE. The Belgian researchers postulated that an increase in average exposure to BC prolongs hospital stay of COVID-19 patients by almost 3 days and that it also increases the odds of admission to ICU by 20%. Although all these studies did not use the same methodology as we did in the present analysis, our work reiterated their qualitative conclusions regarding BC effects on COVID-19.

Our study has also revealed a strong effect of O_3 pollution on the relative risk

of COVID-19 in Belgium. This finding concurred with the conclusions of earlier research that air pollution by O_3 is an important co-factor driving COVID-19 incidence (Lim et al., 2021; Gujral and Sinha, 2021; Stufano et al., 2021; Zoran et al., 2020). Our study also agreed with the finding that the strongest effect of O_3 is likely to occur during the first weeks of exposure (Zhu et al., 2020). Therefore, results of the present analysis have validated and expanded the earlier evidence of positive association between O_3 pollution and COVID-19, showing that also in Belgium this association was important.

Regarding the lagged effects of air pollution, we found that there was no or very small delay in responses to O_3 and BC, and the strongest effects were concentrated within 1-2 weeks following exposure (Figures 3 and 4). To elucidate why the delay in pronounced effects was very short or absent, we refer to the mechanisms of air pollution effects on the respiratory organs of humans.

When humans are exposed to high concentrations of air pollutants, such as particulate matter or O_3, they become predisposed to respiratory tract infections. In detail, air pollution causes inflammatory processes in human lungs and oxidative stress (excess of free radicals in the body) (Lai et al., 2021). In addition, exposure to air pollution alters body's natural virus-fighting mechanisms: it decreases self-clearing processes in lungs (mucociliary clearance), decreases virus uptake by macrophages (white blood cells that destroy harmful microorganisms), and increases permeability of lungs‘ epithelial tissue, favoring viral spread (Wang et al., 2020; Woodby et al., 2021).

We do not have exact estimation how soon inflammatory and related processes make the person susceptible to respiratory viruses, but we assume that there is a less than a week delay, which is supported by our data and earlier research (Schwartz, 2000; Clarke et al., 1999). For the comparison, we can recall other studies that analyzed delayed health-related responses to environmental stressors (e.g., precipitation). For example, a study on the dengue risk in Brazil showed that following extremely dry conditions, the risk of dengue increases after about 3-5 months (Lowe et al., 2021). Such a long lag before the start of a disease was explained by changes in human behavior in response to extreme water shortages. In such circumstances, people utilize water storage outside homes, that later serve as breeding habitats for the *Aedes aegypti* larvae (a larvae of mosquito spreading dengue). The time that it takes people to realize

the need for water reserves, and the time necessary for *A. aegypti* to develop to a mosquito stage, comprises a lag of 3-5 months until an outbreak of dengue. However, the cause-effect mechanisms between dry climate conditions and dengue are likely very different from the association between air pollution and respiratory infection risks, which supposedly requires only a short lag.

While a greater part of the published literature establishes a link between air pollution and COVID-19, some studies with the divergent conclusions should also be mentioned. An already cited work from Belgium by Vos et al. (2022) did not find any effect of O_3 on COVID-19 dynamics, although our results agreed with respect to BC effects. Another study by Hoang et al. (2021) applied DLNM to estimate the effect of various air pollutants on COVID-19 cases in two cities and surrounding provinces of South Korea. Surprisingly, they reported no effect of O_3 on increasing COVID-19 cases in any of the locations.

In addition, there are studies that suggested a lower number of COVID-19 cases under a higher air pollution concentration. For instance, despite the evidence of NO_2 detrimental effects on human health and COVID-19 susceptibility (Pansini and Fornacca, 2020; Ogen, 2020), the research by Zoran et al. (Zoran et al., 2020) found a negative correlation between NO_2 concentration and COVID-19 cases in Italy. This finding was likely observed due to a negative correlation between NO_2 and O_3, which in turn had a negative and a positive correlation with COVID-19, respectively. A negative correlation was also found between SO_2 air pollutant and counts of confirmed COVID-19 events in China (Zhou et al., 2021). In the latter work, the authors commented that the mechanisms of SO_2 and COVID-19 connection are unclear and need to be further explored.

The contradictions between the studies relating air pollution to COVID-19 may point to an important issue of pertinent data analyses. The studies cited in this project used various approaches to define study areas and to statistically analyze the data, which can be one of the reasons for disparate study outcomes. For example, correlation analysis may not reveal the same findings as regression-type models, which both can diverge from DLNM approach used in our study. Thus, to achieve more solid and reproducible results, we believe that similar future studies can be improved by the corrections of data analyses. While we cannot suggest corrections for already published research,

we would like to underline limitations of our own work and propose strategies for rectification.

5 Limitations of the study and ideas for correction

The major limitation of this study is that we tried to include several pollutants in the same statistical model. On the one hand, such analysis was appealing, because it allowed to disentangle the differences in each pollutant's effect on COVID-19. On the other hand, it caused collinearity problem in our analyses and in most of the previous research. An alternative approach would be to use a single covariate serving as a proxy for average pollution levels. For instance, an Air Quality Index (AQI), which is based on the scaled concentrations of the main air pollution substances.

Next, as a spatial unit of our analysis we used Belgian municipality. This approach was justified by the spatial resolution of the data (pollution and COVID-19 were both measured on a municipality level) but was also suitable to account for municipality-specific responses to COVID-19. However, air pollution generated within the borders of a single municipality may redistribute across municipalities' borders following local air masses circulation. We assume that a possible better approach would be to define "air pollution domains", or areas with distinct air masses circulation and therefore, distinct air pollution levels. This analysis would also require data from neighbor countries, because air pollution may span across national borders.

Here we have listed the two critical limitations of our study and corresponding remedies for upcoming research. But we have more recommendations for improvement, and we refer an interested reader to the Supplementary Section 8.2.

6 Ethical thinking, societal relevance and stakeholder awareness

As our study illustrates, there was a link between air pollution and COVID-19 incidence in Belgium in 2021. Although our study was only exploratory, we believe that the effect of air pollution on human health and on the likelihood of

respiratory diseases, is real. Therefore, we would like to offer two recommendations for public health authorities in Belgium.

First, we encourage to continue developing policies to combat air pollution in the country for minimizing the risk of respiratory dysfunctions and related fatalities. Second, we would like to point stakeholder's attention to the most air polluted areas of the country. In case there would be an outbreak of a similar disease as COVID-19 in the future, the residents of these regions, particularly the elderly, would be most at risk. Hence, in such a situation, they should be prioritized by the pandemic preparedness campaigns.

Regarding the ethical aspect of our study, we have adhered to the ethical principles of using public health and other national data in an academic work. We have revealed no individual subject's data by our analyses, and we have carefully cited all the sources that have provided data access for us.

7 Conclusions

Concluding our study, we would like to emphasize that our results agreed with the evidence of the detrimental effect of air pollution on human health. Therefore, the salient message of our work is that Belgium should strive to keep air pollution levels low.

Finally, we would like to remark on the methodological approach of our study. DLNM proved to be appropriate for identifying air pollution effects on COVID-19, but they have a much wider range of applications particularly in combination with a Bayesian framework. DLNM are suitable for scientific endeavors looking into lagged health responses to environmental or anthropogenic factors on a large spatial and temporal scale. We thus inspire epidemiologists to explore and apply DLNM to their research problems, for which other statistical methods can be deficient.

References

H. Abdelzaher, B. M. Saleh, H. A. Ismail, M. Hafiz, M. A. Gabal, M. Mahmoud, S. Hashish, R. M. A. Gawad, R. Y. Gharieb, and A. Abdelnaser. Covid-19 genetic and environmental risk factors: a look at the evidence. *Frontiers in pharmacology*, 11:579415, 2020.

A. Abdi, M. Jalilian, P. A. Sarbarzeh, and Z. Vlaisavljevic. Diabetes and covid-19: A systematic review on the current evidences. *Diabetes Research and Clinical Practice*, 166:108347, 2020. ISSN 0168-8227.

N. Ali and F. Islam. The effects of air pollution on covid-19 infection and mortality—a review on recent evidence. *Frontiers in public health*, 8:580057, 2020.

S. Almon. The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1):178–196, 1965. ISSN 00129682, 14680262.

S. C. Anenberg, J. Schwartz, D. Shindell, M. Amann, G. Faluvegi, Z. Klimont, G. Janssens-Maenhout, L. Pozzoli, R. Van Dingenen, E. Vignati, et al. Global air quality and health co-benefits of mitigating near-term climate change through methane and black carbon emission controls. *Environmental health perspectives*, 120(6):831–839, 2012.

B. Armstrong. Models for the relationship between ambient temperature and daily mortality. *Epidemiology*, pages 624–631, 2006.

N. Bai, M. Khazaei, S. F. van Eeden, and I. Laher. The pharmacology of particulate matter air pollution-induced cardiovascular dysfunction. *Pharmacology & therapeutics*, 113(1):16–29, 2007.

A. G. Barnett, G. M. Williams, J. Schwartz, T. L. Best, A. H. Neller, A. L. Petroeschevsky, and R. W. Simpson. The effects of air pollution on hospitalizations for cardiovascular disease in elderly people in australian and new zealand cities. *Environmental health perspectives*, 114(7):1018–1023, 2006.

M. F. Bashir, B. Jiang, B. Komal, M. A. Bashir, T. H. Farooq, N. Iqbal, M. Bashir, et al. Correlation between environmental pollution indicators and covid-19 pandemic: a brief study in californian context. *Environmental research*, 187:109652, 2020.

S. Becker and J. M. Soukup. Exposure to urban air particulates alters the macrophage-mediated inflammatory response to respiratory viral infection. *Journal of Toxicology and Environmental Health Part A*, 57(7):445–457, 1999.

J. Besag, J. York, and A. Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43:1–20, 1991.

K. Bhalla, M. Shotten, A. Cohen, M. Brauer, S. Shahraz, R. Burnett, K. Leach-Kemon, G. Freedman, and C. J. Murray. *Transport for health: the global burden of disease from motorized road transport*. 2014.

M. Blangiardo, M. Cameletti, G. Baio, and H. Rue. Spatial and spatio-temporal models with r-inla. *Spatial and spatio-temporal epidemiology*, 4:33–49, 2013.

A. L. F. Braga, A. Zanobetti, and J. Schwartz. The time course of weather-related deaths. *Epidemiology*, 12(6):662–667, 2001.

R. Clarke, P. Catalano, P. Koutrakis, G. K. Murthy, C. Sioutas, J. Paulauskis, B. Coull, S. Ferguson, and J. Godleski. Urban air particulate inhalation alters pulmonary function and induces pulmonary inflammation in a rodent model of chronic bronchitis. *Inhalation toxicology*, 11(8):637–656, 1999.

E. S. Coker, L. Cavalli, E. Fabrizi, G. Guastella, E. Lippo, M. L. Parisi, N. Pontarollo, M. Rizzati, A. Varacca, and S. Vergalli. The effects of air pollution on covid-19 related mortality in northern italy. *Environmental and Resource Economics*, 76:611–634, 2020.

S. Comunian, D. Dongo, C. Milani, and P. Palestini. Air pollution and covid-19: the role of particulate matter in the spread and increase of covid-19's morbidity and mortality. *International journal of environmental research and public health*, 17(12):4487, 2020.

P. D. Congdon. *Bayesian hierarchical models: with applications using R*. Chapman and Hall/CRC, 2019.

C. Copat, A. Cristaldi, M. Fiore, A. Grasso, P. Zuccarello, G. O. Conti, S. Santo Signorelli, and M. Ferrante. A first review to explore the association of air pollution (pm and no2) on severe acute respiratory syndrome coronavirus (sars-cov-2). *Preprints*, 2020.

Y. Cui, Z.-F. Zhang, J. Froines, J. Zhao, H. Wang, S.-Z. Yu, and R. Detels. Air pollution and case fatality of sars in the people's republic of china: an ecologic study. *Environmental health*, 2(1):1–5, 2003.

A. P. Dobson and E. R. Carper. Infectious diseases and human population history. *Bioscience*, 46(2):115–126, 1996.

F. Fierens, C. Vanpoucke, E. Trimpeneers, O. Peeters, S. Quidé, T. de Vos, P. Maetz, and V. Hutsemékers. Annual report air quality in belgium 2015. *Belgian Interregional Environment Agency: Brussels, Belgium*, 2015.

S. Gadeyne, L. Rodriguez-Loureiro, J. Surkyn, W. Van Hemelrijck, W. Nusselder, P. Lusyne, and K. Vanthomme. Are we really all in this together? the social patterning of mortality during the first wave of the covid-19 pandemic in belgium. *International journal for equity in health*, 20 (1):1–17, 2021.

A. Gasparrini. Distributed lag linear and non-linear models in r: the package dlnm. *Journal of statistical software*, 43(8):1, 2011.

A. Gasparrini, B. Armstrong, and M. G. Kenward. Distributed lag non-linear models. *Statistics in medicine*, 29(21):2224–2234, 2010.

H. Gujral and A. Sinha. Association between exposure to airborne pollutants and covid-19 in los angeles, united states with ensemble-based dynamic emission model. *Environmental research*, 194:110704, 2021.

M. Hendryx and J. Luo. Covid-19 prevalence and fatality rates in association with air pollution emission concentrations and emission sources. *Environmental Pollution*, 265:115126, 2020.

T. Hoang, T. Q. Nguyen, and T. T. A. Tran. Short-term exposure to ambient air pollution in association with covid-19 of two clusters in south korea. *Tropical Medicine & International Health*, 26(4):478–491, 2021.

S. Janssen, G. Dumont, F. Fierens, and C. Mensink. Spatial interpolation of air pollution measurements using corine land cover data. *Atmospheric Environment*, 42(20):4884–4903, 2008. ISSN 1352-2310.

M. Kéry and J. A. Royle. *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models*. Academic Press, 2020.

D. Kim, Z. Chen, L.-F. Zhou, and S.-X. Huang. Air pollutants and early origins of respiratory diseases. *Chronic diseases and translational medicine*, 4(2): 75–94, 2018.

F. Laguna, M. E. Grillet, J. R. León, and C. Ludeña. Modelling malaria incidence by an autoregressive distributed lag model with spatial component. *Spatial and Spatio-temporal Epidemiology*, 22:27–37, 2017. ISSN 1877-5845.

A. Lai, M. L. Chang, R. P. O'Donnell, C. Zhou, J. A. Sumner, and T. K. Hsiai. Association of covid-19 transmission with high levels of ambient pollutants: Initiation and impact of the inflammatory response on cardiopulmonary disease. *Science of the Total Environment*, 779:146464, 2021.

J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525(7569):367–371, 2015.

E. Lesaffre and A. B. Lawson. *Bayesian biostatistics*. John Wiley & Sons, 2012.

Y. K. Lim, O. J. Kweon, H. R. Kim, T.-H. Kim, and M.-K. Lee. The impact of environmental variables on the spread of covid-19 in the republic of korea. *Scientific reports*, 11(1):5977, 2021.

Z. Liu, L. Zhu, Y. Wang, Z. Zhou, and Y. Guo. The correlation between covid-19 activities and climate factors in different climate types areas. *Journal of Occupational and Environmental Medicine*, 63(8):e533, 2021.

R. Lowe, A. Gasparrini, C. J. Van Meerbeeck, C. A. Lippi, R. Mahon, A. R. Trotman, L. Rollock, A. Q. Hinds, S. J. Ryan, and A. M. Stewart-Ibarra. Nonlinear and delayed impacts of climate on dengue risk in barbados: A modelling study. *PLoS medicine*, 15(7):e1002613, 2018.

R. Lowe, S. A. Lee, K. M. O'Reilly, O. J. Brady, L. Bastos, G. Carrasco-Escobar, R. de Castro Catão, F. J. Colón-González, C. Barcellos, M. S. Carvalho, et al. Combined effects of hydrometeorological hazards and urbanisation on dengue

risk in brazil: a spatiotemporal modelling study. *The Lancet Planetary Health*, 5(4):e209–e219, 2021.

M. Marquès and J. L. Domingo. Positive association between outdoor air pollution and the incidence and severity of covid-19. a review of the recent scientific evidences. *Environmental Research*, 203:111930, 2022.

P. Mecenas, R. T. d. R. M. Bastos, A. C. R. Vallinoto, and D. Normando. Effects of temperature and humidity on the spread of covid-19: A systematic review. *PLoS one*, 15(9):e0238339, 2020.

M. Meurisse, A. Lajot, B. Devleesschauwer, D. Van Cauteren, H. Van Oyen, L. Van den Borre, and R. Brondeel. The association between area deprivation and covid-19 incidence: a municipality-level spatio-temporal study in belgium, 2020–2021. *Archives of Public Health*, 80(1):1–10, 2022.

P. Moraga. *Geospatial health data: Modeling and visualization with R-INLA and shiny*. CRC Press, 2019.

K. F. Morales, J. Paget, and P. Spreeuwenberg. Possible explanations for why some countries were harder hit by the pandemic influenza virus in 2009–a global mortality impact modeling study. *BMC infectious diseases*, 17:1–12, 2017.

S. Mukherjee and K. Pahan. Is covid-19 gender-sensitive? *Journal of Neuroimmune Pharmacology*, 16:38–47, 2021.

Y. A. Natalia, C. Faes, T. Neyens, and G. Molenberghs. The covid-19 wave in belgium during the fall of 2020 and its association with higher education. *Plos one*, 17(2):e0264516, 2022.

Y. Ogen. Assessing nitrogen dioxide (no2) levels as a contributing factor to coronavirus (covid-19) fatality. *Science of the Total Environment*, 726:138605, 2020.

R. Pansini and D. Fornacca. Higher virulence of covid-19 in the air-polluted regions of eight severely affected countries. *medRxiv*, pages 2020–04, 2020.

M. Peng, J. He, Y. Xue, X. Yang, S. Liu, and Z. Gong. Role of hypertension on the severity of covid-19: A review. *Journal of cardiovascular pharmacology*, 78(5):e648, 2021.

S. S. Qian, M. R. DuFour, and I. Alameddine. *Bayesian Applications in Environmental and Ecological Studies with R and Stan*. CRC Press, 2022.

J. Rashedi, B. Mahdavi Poor, V. Asgharzadeh, M. Pourostadi, H. Samadi Kafil, A. Vegari, H. Tayebi-Khosroshahi, and M. Asgharzadeh. Risk factors for covid-19. *Infez Med*, 28(4):469–474, 2020.

A. Rathod and G. Beig. Impact of biomass induced black carbon particles in cascading covid-19. *Urban Climate*, 38:100913, 2021.

A. Riebler, S. H. Sørbye, D. Simpson, and H. Rue. An intuitive bayesian spatial model for disease mapping that accounts for scaling. *Statistical methods in medical research*, 25(4):1145–1165, 2016.

J. Rod, O. Oviedo-Trespalacios, and J. Cortes-Ramirez. A brief-review of the risk factors for covid-19 severity. *Revista de saude publica*, 54, 2020.

K. Romero Starke, G. Petereit-Haack, M. Schubert, D. Kämpf, A. Schliebner, J. Hegewald, and A. Seidler. The age-related risk of severe outcomes due to covid-19 infection: a rapid review, meta-analysis, and meta-regression. *International journal of environmental research and public health*, 17(16):5974, 2020.

D. B. Rubin. Multiple imputation after 18+ years. *Journal of the American statistical Association*, 91(434):473–489, 1996.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, 2005.

H. Rue and S. Martino. Approximate bayesian inference for hierarchical gaussian markov random field models. *Journal of statistical planning and inference*, 137(10):3177–3192, 2007.

H. Rue, S. Martino, and N. Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392, 2009.

H. Rue, A. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren. Bayesian computing with inla: a review. *Annual Review of Statistics and Its Application*, 4:395–421, 2017.

K. S. Saini, M. Tagliamento, M. Lambertini, R. McNally, M. Romano, M. Leone, G. Curigliano, and E. de Azambuja. Mortality in patients with cancer and coronavirus disease 2019: A systematic review and pooled analysis of 52 studies. *European Journal of Cancer*, 139:43–50, 2020. ISSN 0959-8049.

M. M. Sajadi, P. Habibzadeh, A. Vintzileos, S. Shokouhi, F. Miralles-Wilhelm, and A. Amoroso. Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (covid-19). *JAMA network open*, 3(6):e2011834–e2011834, 2020.

J. Schwartz. The distributed lag between air pollution and daily deaths. *Epidemiology*, 11(3):320–326, 2000.

L. Setti et al. Evaluation of the potential relationship between particulate matter (pm) pollution and covid-19 infection spread in italy. sima, università di bologna, università degli studi di bari aldo moro. 2020 apr.

D. Simpson, H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. Penalising model component complexity: A principled, practical approach to constructing priors. 2017.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639, 2002.

A. Stufano, S. Lisco, N. Bartolomeo, A. Marsico, G. Lucchese, H. Jahantigh, L. Soleo, M. Moretti, P. Trerotoli, G. De Palma, et al. Covid19 outbreak in lombardy, italy: An analysis on the short-term relationship between air pollution, climatic factors and the susceptibility to sars-cov-2 infection. *Environmental Research*, 198:111197, 2021.

J. Sunyer, X. Basagana, J. Belmonte, and J. Anto. Effect of nitrogen dioxide and ozone on the risk of dying in patients with severe asthma. *Thorax*, 57(8): 687–693, 2002.

M. Travaglio, Y. Yu, R. Popovic, L. Selley, N. S. Leal, and L. M. Martins. Links between air pollution and covid-19 in england. *Environmental pollution*, 268: 115859, 2021.

P. Verwimp et al. *The spread of COVID-19 in Belgium: a municipality-level analysis*. ECARES, 2020.

S. Vos, E. De Waele, P. Goeminne, E. Bijnens, E. Bongaerts, A. De Weerdt, P. Jorens, and T. Nawrot. Covid-19 disease severity in hospitalized patients is associated with short-and long-term exposure to air pollutants. In *ISEE Conference Abstracts*, volume 2022, 2022.

B. Wang, H. Chen, Y. L. Chan, and B. G. Oliver. Is there an association between the level of ambient air pollution and covid-19? *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 319(3):L416–L421, 2020.

X. Wang, Y. Yue, and J. J. Faraway. *Bayesian regression modeling with INLA*. Chapman and Hall/CRC, 2018.

A. Wilder-Smith. Covid-19 in comparison with other emerging viral diseases: risk of geographic spread via travel. *Tropical Diseases, Travel Medicine and Vaccines*, 7:1–11, 2021.

B. Woodby, M. M. Arnold, and G. Valacchi. Sars-cov-2 infection, covid-19 pathogenesis, and exposure to air pollution: What is the connection? *Annals of the new York Academy of Sciences*, 1486(1):15–38, 2021.

X. Wu, R. C. Nethery, M. B. Sabath, D. Braun, and F. Dominici. Air pollution and covid-19 mortality in the united states: Strengths and limitations of an ecological regression analysis. *Science advances*, 6(45):eabd4049, 2020.

A. Zanobetti, M. P. Wand, J. Schwartz, and L. M. Ryan. Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics*, 1 (3):279–292, 2000.

J. Zhang, Y. Wei, and Z. Fang. Ozone pollution: a major health hazard worldwide. *Frontiers in immunology*, 10:2518, 2019.

J. Zhang, Y.-H. Lim, T. Lange, S. Loft, L. H. Mortensen, G. M. Napolitano, R. G. Westendorp, J. Brandt, C.-H. Tom, and Z. J. Andersen. Long-term exposure to air pollution and covid-19 mortality and morbidity in denmark (aircoden). In *ISEE Conference Abstracts*, volume 2022, 2022.

Z. Zhang, T. Xue, and X. Jin. Effects of meteorological conditions and air pollution on covid-19 transmission: Evidence from 219 chinese cities. *Science of the total environment*, 741:140244, 2020.

Z. Zheng, F. Peng, B. Xu, J. Zhao, H. Liu, J. Peng, Q. Li, C. Jiang, Y. Zhou, S. Liu, et al. Risk factors of critical & mortal covid-19 cases: A systematic literature review and meta-analysis. *Journal of infection*, 81(2):e16–e25, 2020.

J. Zhou, L. Qin, X. Meng, and N. Liu. The interactive effects of ambient air pollutants-meteorological factors on confirmed cases of covid-19 in 120 chinese cities. *Environmental Science and Pollution Research*, 28:27056–27066, 2021.

Y. Zhu, J. Xie, F. Huang, and L. Cao. Association between short-term exposure to air pollution and covid-19 infection: Evidence from china. *Science of the total environment*, 727:138704, 2020.

M. A. Zoran, R. S. Savastru, D. M. Savastru, and M. N. Tautan. Assessing the relationship between ground levels of ozone (o3) and nitrogen dioxide (no2) with coronavirus (covid-19) in milan, italy. *Science of The Total Environment*, 740:140005, 2020.

8 Appendix

8.1 Supplementary figures

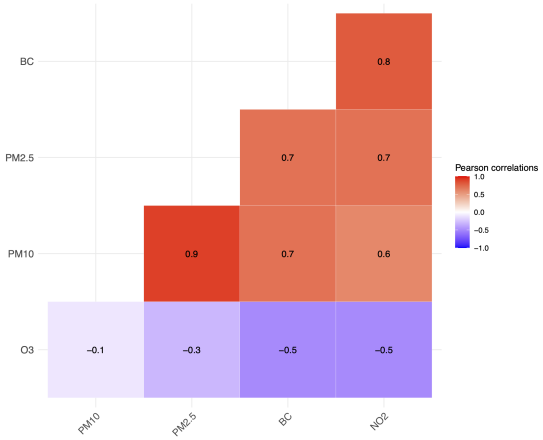

Figure 5: Pearson correlations between the 4 air pollutants from the IRCEL-CELINE data: black carbon (BC), Nitrogen Dioxide (NO_2), ozone (O_3), and particulate matter PM_{10} and $PM_{2.5}$.

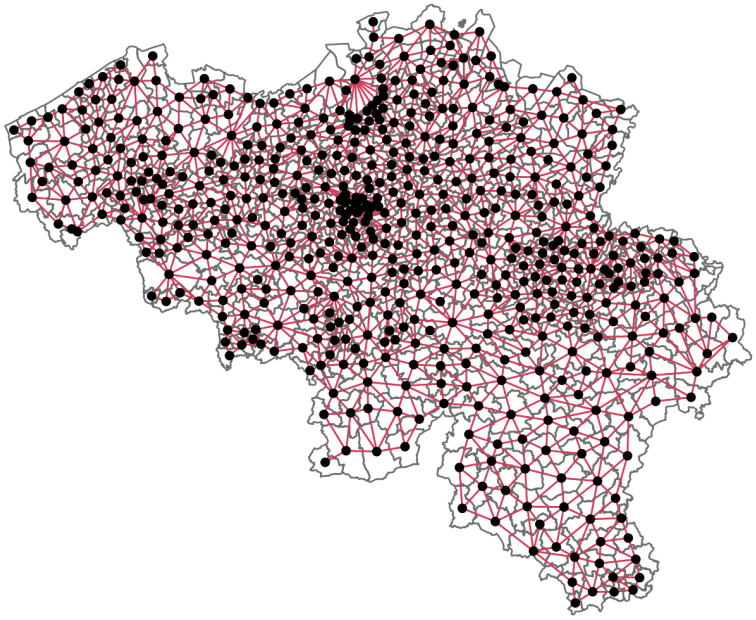

Figure 6: Adjacency matrix used for defining spatial structure of the Bayesian DLNM. The black dots denote centroids of municipalities' polygons and red lines connect the municipalities considered as neighbors.

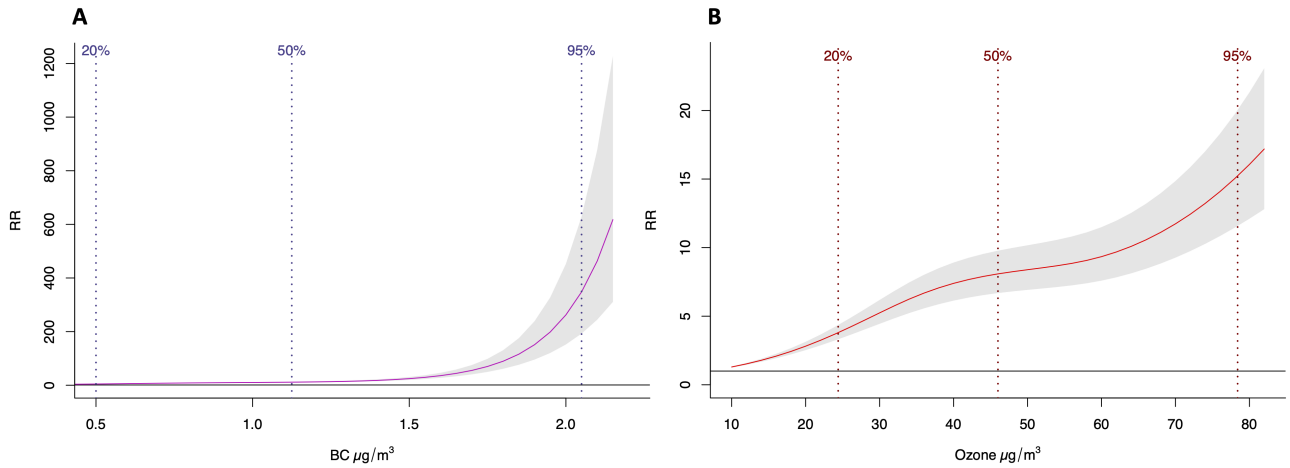

Figure 7: Cumulative effects of exposure to BC (A) and O_3 (B) on the relative risk of COVID-19. Dotted lines denote precentiles of pollutants concetrations: 20%, 50%, and 95%.

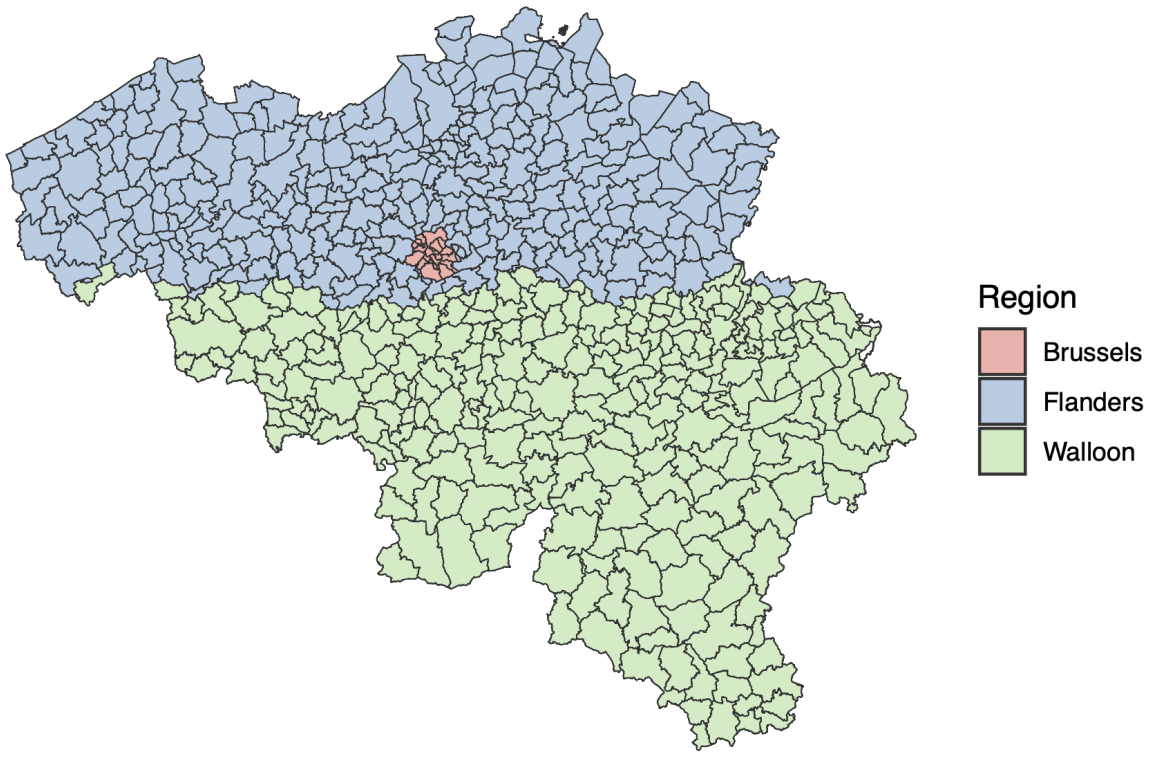

Figure 8: A map of the Belgian regions: Brussels (red), Flanders (blue) and Walloon (green). The municipalities borders are shown by grey lines.

8.2 Supplementary tables

Table 3: Additional limitations of our study with suggestions of alternative procedures by future studies.

Limitation	**Solutions employed in our study**	**Strategies for improvement**
COVID-19 cases data had about 50% of observations recorded as "< 5" counts	"< 5" values were replaced by a random draw from the uniform distribution. A simple sensitivity analysis did not show large changes in the data, if replacements were repeated, but study results may have been affected	Multiple imputation techniques might have been a better approach for our study. More precise estimation of pollution effects might be obtained based on the analyses of 10 imputed data sets (Rubin, 1996). However, this analysis is currently difficult to combine with a Bayesian framework
Our data had a spatial and a temporal structure, hence spatial-temporal interaction (e.g., municipality-specific temporal trends) could have been relevant	We have included a spatial and a temporal random effect but opted against an interaction effect for two reasons. Spatial-temporal interaction in some cases results in an overfitted model, when an interaction effect absorbs most of the variation in the data. This obstacle has been reported by similar studies (Natalia et al., 2022). Moreover, spatial-temporal interaction may conceal the effect of air pollution, that also varied spatially and temporally	Including interaction effect is more appropriate for the studies based on longer data than ours. For example, a similar analysis of air pollution effects on respiratory diseases risk, comprising 10 years of data could have a spatial effect of a municipality, a temporal effect of a week within each year, and a spatial-temporal interaction between a municipality and a year
A conventional choice of parameters was made for cross-basis functions of DLNM	We selected natural splines with 2 knots for exposure-response and 1 knot for lag-response basis functions. Functions of more complex shapes can also be selected but may also lead to overspecified models	For a larger number of knots, a penalization of smooth function may need to be applied, but at the time of writing (June 2023), this approach is not yet implemented in INLA. But future research teams should not hesitate to contact INLA core team (R-INLA) for advise and instructions

8.3 R code

All the the R codes used for the analysis in this study are accessible from the github repository via the following link: COVID-19 DLNM.