

# Analysing spatiotemporal patterns of Covid-19 confirmed deaths at the NUTS-2 regional level

**Andrea Bucci**

(corresponding author)

Università degli Studi

“G. d’Annunzio” Chieti-Pescara,

Italy

E-mail: andrea.bucci@unich.it

**Luigi Ippoliti**

Università degli Studi

“G. d’Annunzio” Chieti-Pescara,

Italy

**Pasquale Valentini**

Università degli Studi

“G. d’Annunzio” Chieti-Pescara,

Italy

**Keywords:**

Covid-19,

spatial clustering,

B-splines,

poisson log-normal regression,

conditional autoregressive processes

During the ongoing Covid-19 pandemic, understanding the spatiotemporal patterns of the virus is crucial for policymakers to intervene promptly. The relevance of spatial proximity in the spread of the pandemic necessitates adequate tools, and noisy data must be properly treated. This study proposes obtaining clusters of European regions using smoothed curves of daily deaths from March 2020–March 2022. A functional representation of the curves was implemented to extract the features used in a clustering algorithm that allows spatial proximity. In a spatial regression model, the authors also investigated the role of clusters and pre-existing conditions on cumulative deaths, and observed that air pollution, health conditions, and population age structure are significantly associated with Covid-19 confirmed deaths.

## Introduction

Since its outbreak in Wuhan (China) in December 2019, the coronavirus disease 2019 (Covid-19) has spread rapidly and uncontrollably throughout the world (Kincses–Tóth 2020). At the start of this pandemic, the most affected countries, such as Italy, France, Germany, and Spain, implemented a series of large-scale interventions to control the epidemic. However, an animated debate has arisen about whether and to what extent these interventions have been effective in slowing down the pandemic. The question remains unanswered, as related literature (see, for example, Alfano–Ercolano 2020, Islam et al. 2020, May 2020, Melnick–Ioannidis 2020, Kosfeld et al. 2021, Bourdin et al. 2021, Nyikos et al. 2021, Antalóczy et al. 2022, Bourdin et al. 2022 and references therein) has revealed considerable differences in the efficacy of the implemented public health interventions in Europe.

In this context, several studies have shown that the Covid-19 disease patterns exhibit both spatial and temporal variability (Guliyev 2020, D'Urso et al. 2022, Bucci et al. 2022, Desmet–Wacziarg 2022). Indeed, epidemic curves vary significantly between and within countries, indicating a substantial regional dimension of the pandemic that affects both public management and the implementation of preventive public health measures. The identification of a data-generating process for these patterns is complicated because of the pandemic's complexity. However, statistical analysis of this phenomenon can provide governments with useful insights for implementing public health strategies. For these reasons, in this study, we identify groups of regions that share common Covid-19 death trajectories and describe possible associations of pre-existing conditions with the burden of the pandemic using a regression model. Identifying clusters of regions that exhibit similar spatial and temporal patterns in terms of the impact of Covid-19 on health can help policymakers understand the effectiveness of control and response strategies against past outbreaks. It can also encourage the implementation of more targeted actions by governments, and the development of synergistic actions by neighbouring countries for future waves.

Previous studies have focused on the clustering of Covid-19 confirmed death time series at different territorial levels (Shariati et al. 2020, Zhang et al. 2020, Andersen et al. 2021). For example, D'Urso et al. (2022) analysed the spatiotemporal patterns of the disease in Italian regions and identified that spatial contiguity is crucial for understanding the pandemic. Bucci et al. (2022) identified the same result in their study of the Nomenclature of Statistical Territorial Units (Nomenclature des Unités Territoriales Statistiques) NUTS-2 regions in Europe. Despite the existence of several studies on the identification of common patterns in territorial units, most limit the analysis to specific single countries or detect groups at the national level. However, in most countries, public health management is decentralised, and regional units should be given more attention. Consequently, in contrast to others, we focus on regional data (NUTS-2) from several European countries. In particular, we examined the results of clustering time series of Covid-19 confirmed deaths from 142 European NUTS-2 regions in France, Germany, Italy, Spain, Switzerland, and the United Kingdom.

Motivated by the spatiotemporal nature of the studied phenomenon, we used a hierarchical clustering algorithm in which the objective function to be optimised is a weighted mixture of attributes (*i.e.* time series features) of NUTS-2 regions and geographical/locational dissimilarity matrices. Early formulations of the concept behind the weighted optimisation of geographical features and attributes can be observed in Webster–Burrough (1972), Wise et al. (1997), and Haining et al. (2000). In this study, we used the Ward-like hierarchical clustering algorithm proposed by Chavent et al. (2018) which is equipped with a freely available and easy-to-implement R package. Depending on the number of groups selected, the algorithm requires the

tuning of a hyperparameter,  $\alpha$ , which regulates the weight to be assigned to the geographic and attribute components. Therefore, the coordinate variables are treated separately from the “regular” attributes, resulting in a compromise between the two types of dissimilarities. By regulating the effect of spatial contiguity in the clustering procedure, this study not only identifies groups of regions sharing common temporal patterns but also provides insights into how various countries in different clusters dealt with the pandemic in the given context and the consequences they faced. As a result, local authorities and European leaders can implement similar data-driven policies in regions sharing the same cluster.

In this context, we used a spatial regression model to investigate the phenomenon’s association with a number of potential triggering factors to further describe its geographical distribution. Several studies (see, for example, Alves et al. 2021, Amdaoud et al. 2021, Ehlert 2021, and Aritenang 2022) have already shown that socio-economic and demographic conditions can help explain the outcomes of the pandemic. However, most of them remain inconclusive owing to the ever-changing scenario of the pandemic and dynamic understanding of the disease. For this reason, we further investigated the relationship between cumulative deaths because of Covid-19 and a set of explanatory variables, considering the information regarding pandemic patterns through clusters. Our aim was to understand if an association between pre-existing conditions and death counts exists, and if this remains valid after controlling the patterns of the time series of deaths. In fact, the health outcomes of a pandemic would inevitably depend on restrictive measures, the behaviour of citizens in response to these measures, and the broader determinants of health, such as social, economic, physical, and environmental variables (Egri 2017). For example, income inequality, absence of job security and adequate pay often translate into a higher number of non-specialized workers whose health conditions may be generally low, and because of poor access to care (Perucca et al. 2019, Kapás 2022). Health infrastructure also influences the number of deaths from a disease as it determines the possibility for a patient to be properly treated (Ngepah 2021, Pál et al. 2021). In addition, air pollution can exacerbate the effects of the pandemic and subsequently cause a higher number of deaths in the most polluted regions, as stated by Comunian et al. (2020), Lolli et al. (2020), and Bourdrel et al. (2021). Consequently, understanding if some explanatory variables affect the total number of confirmed deaths would allow the government to protect the most vulnerable regions in terms of the joint effect of the pandemic and already existing conditions. It is worth noting that this study only used ecological data. This indicates that, to avoid the so-called ecological fallacy, we do not draw direct conclusions at the individual level. However, our findings can be used to generate hypotheses regarding the possible determinants of the spread and effects of Covid-19 before individual-level studies become available.

The remainder of this paper is organised as follows. Preliminary, we provide a simple exploratory analysis of the data and describe how to obtain features (attributes) of

interest from the observed time series. This section also describes the proposed clustering algorithm. In following section we present the results of the clustering and spatial model estimation of the total number of deaths. Finally, the discussion and conclusions close the study.

## Materials and methods

### Data

To identify patterns of interest in time and space, we considered daily confirmed death counts of Covid-19 for 142 NUTS-2 regions in France, Germany, Italy, Spain, Switzerland, and the United Kingdom<sup>1</sup> for 744 days (March 11, 2020–March 31, 2022). Data are available from the national repositories [1–6]. The choice of these countries was primarily motivated by their relevance in the European context (these are the European countries with the highest gross domestic product (GDP) at market prices in 2019) and by the evolution of the pandemic that first hit this part of Europe. All countries included in our analysis adopted a universalistic National Health Service (NHS). However, regional autonomy, in terms of health expenditure and strategies, is significantly different between the countries considered. In Italy, Spain, and Germany, different health plans without national endorsement can be implemented. This has led to a variety of health strategies and outcomes throughout each country which may have influenced the outcomes of the Covid-19 pandemic (Garattini et al. 2022). In Switzerland, the central government can act only in certain areas, such as the financing of the health system and the assessment of the quality and safety of pharmaceuticals and medical devices, while the cantons (NUTS-3) are responsible for securing healthcare provision for the population (De Pietro 2015). Meanwhile, France is characterised by a low level of decentralisation. The government directly finances and organises the delivery of health and social services, whereas regional agencies coordinate prevention, health, and supportive care. Similarly, in the United Kingdom, the NHS is responsible for the health of the population, each country has its own healthcare system (Flynn et al. 2020), and regional teams are only capable of supporting the commissioning of healthcare for different regions of the country.

It is well known that, for various reasons, deciding how to define a Covid-19 case or a Covid-19 death is not a straightforward task because the data available from public sources are inherently noisy. However, the time series of confirmed Covid-19 deaths represents a clear and simple indicator of the epidemic's state of evolution, and is also used by authorities as one of the parameters for emergency management. To reduce possible problems associated with the presence of measurement errors, we

<sup>1</sup> Excluding Wales, the Scottish Highlands, and outermost regions. Wales and the Scottish Highlands have been excluded owing to lack of daily data. The outermost regions are extremely distant from the continental territory and the evolution of the pandemic in these regions would not reflect what happened in the rest of the country.



considered herein. Furthermore, to enhance comparability among regions and produce more objective clusters, we work with the same scale so that each transformed series, henceforth denoted as  $\hat{y}_i$ ,  $i = 1, \dots, n$ , assumes values in the interval  $[0,1]$ . Finally, for clustering purposes, each  $\hat{y}_i$  is perfectly fitted by a regression B-spline function such that the vector of spline coefficients,  $\hat{\mathbf{y}}_i$ , now used as time series “regular” attributes, is estimated as follows:

$$\hat{\mathbf{y}}_i := \operatorname{argmin} (\hat{\mathbf{y}}_i - \mathbf{Z}\boldsymbol{\gamma}_i)'(\hat{\mathbf{y}}_i - \mathbf{Z}\boldsymbol{\gamma}_i), \quad i = 1, \dots, n.$$

As a result, the set of normalised time series  $\{\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_n\}$  is summarised by a set of spline coefficient vectors of  $\mathbb{R}^L$ ,  $\{\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_n\}$ .

### Descriptive analysis

The set of normalised and smoothed data,  $\hat{\mathbf{y}}_i$ , stratified by country is shown in Figure A1 in the Appendix. As can be observed, the temporal evolution of the pandemic denotes several phases in all the countries. The first peak is present in almost all the time series from March–May 2020, while the second wave of the pandemic presents a single or multiple peaks from October 2020–April 2021, and a third wave is observable from December 2021–February 2022. Not surprisingly, time series from regions belonging to the same country show similar patterns, although the series shows some variability, particularly in the second and third phases of the pandemic. This is particularly true for the United Kingdom, Italy, and France. Variability within a country has a two-fold origin. On the one hand, it may be because of regional differences in citizens’ compliance with preventive measures (Daoust et al. 2022). On the other hand, this may imply that local interventions, particularly those used in the second/third phase of the pandemic, can produce different effects on health status. This is evident from the plot of the two countries that have managed restrictions in different ways, that is, Italy and Germany. In fact, the effects of the pandemic in Italy were extremely variable in the second phase of the pandemic, partly motivated by the implementation of local restrictions by the Italian government at the regional level. Conversely, in Germany, the restrictions were similar to those applied in the first phase of the pandemic, and were directed to the whole population. This decision led to fewer differences between the patterns of confirmed deaths in the German regions. However, an inspection of the correlations computed for pairs of time series suggests the presence of supranational patterns across regions. In particular, although the correlations tend to decay with the spatial distance of the region centroids, with values ranging between 0.5 and 0.8 on average within 1,000 Km, significantly large correlations ( $> 0.5$ ) can still be observed at larger distances. As a result, the clustering procedure follows a strategy that accounts for spatial proximity but does not impose any hard contiguity constraints.

### Spatial clustering of time series

In this section, we briefly describe the hierarchical clustering algorithm proposed by Chavent et al. (2018). Assume that  $D_0 = [d_{0,ij}]$  is the  $(n \times n)$  dissimilarity matrix describing the pairwise distances between the  $n$  vectors of spline coefficients  $\hat{\mathbf{y}}_i$ , for  $i = 1, \dots, n$ . Assume that  $D_1 = [d_{1,ij}]$  is the  $(n \times n)$  dissimilarity matrix describing the distances in terms of geographical coordinates (*i.e.*, centroids  $\mathbf{s}_i, \mathbf{s}_j$ ) for pairs of regions  $\mathcal{A}_i$  and  $\mathcal{A}_j$ . Then, the algorithm defines the mixed pseudo-inertia of cluster  $\mathcal{C}_k^\alpha$  as follows:

$$J_\alpha(\mathcal{C}_k^\alpha) = (1 - \alpha) \sum_{i \in \mathcal{C}_k^\alpha} \sum_{j \in \mathcal{C}_k^\alpha} d_{0,ij}^2 + \alpha \sum_{i \in \mathcal{C}_k^\alpha} \sum_{j \in \mathcal{C}_k^\alpha} d_{1,ij}^2,$$

where the contributions of the geographic and attribute components are balanced by mixing parameter  $\alpha$ . In practice, geographical coordinates are used to constrain the original solution towards a point where all clusters consist of contiguous observations. A large weight for the centroid coordinates, in essence, forces a contiguous solution, similar to what would happen if only the coordinates were considered. This extreme case would assist in obtaining contiguity but would not provide a relevant result in terms of attribute similarity. The mixed pseudo-within-cluster inertia of partition  $\mathcal{P}_K^\alpha = (\mathcal{C}_1^\alpha, \dots, \mathcal{C}_K^\alpha)$  is

$$W_\alpha(\mathcal{P}_K^\alpha) = \sum_{k=1}^K J_\alpha(\mathcal{C}_k^\alpha)$$

which is the sum of the mixed inertia of each cluster. To obtain a new partition  $\mathcal{P}_K^\alpha$  in  $K$  clusters from a given partition  $\mathcal{P}_{K+1}^\alpha$  in  $K + 1$  clusters, the concept is to aggregate two clusters of  $\mathcal{P}_{K+1}^\alpha$ , such that the new partition has minimum within-cluster inertia. The optimisation problem is as follows:

$$\arg \min_{\mathcal{A}, \mathcal{B} \in \mathcal{P}_{K+1}^\alpha} W_\alpha(\mathcal{P}_{K+1}^\alpha) - W_\alpha(\mathcal{P}_K^\alpha)$$

which is equivalent to the following optimisation problem (see, Chavent et al. 2018 for more details)

$$\arg \min_{\mathcal{A}, \mathcal{B} \in \mathcal{P}_{K+1}^\alpha} J_\alpha(\mathcal{A} \cup \mathcal{B}) - J_\alpha(\mathcal{A}) - J_\alpha(\mathcal{B}).$$

### Choice of $K$ and $\alpha$

Determining the optimal number of clusters in a dataset is crucial for all clustering algorithms.

Generally, the optimal number of clusters is defined subjectively, and the choice is strongly related to the method used for measuring similarities. In hierarchical clustering, a practical and subjective solution consists of inspecting the dendrogram for determining whether a clear number of clusters emerges. Accordingly, to obtain more objective solutions, we consider the results from both statistical testing methods and other criteria based on internal validation measures which include information

on within-cluster scattering and between-cluster separation. Examples of these methods include the *gap statistic* of Tibshirani et al. (2001) and *average silhouette* approach of Kaufman–Rousseeuw (1990). While the former compares the total intra-cluster variation for different values of  $K$  with their expected values under the null reference distribution of the data, the latter measures the quality of clustering by determining how close each object is to the points of its neighbouring clusters.

The other key point of the algorithm is the choice of a suitable value for the mixing parameter  $\alpha \in [0,1]$ . This parameter logically depends on the number of clusters  $K$  and this dependence is an issue when determining an optimal value for parameter  $\alpha$ . Optimising the algorithm with respect to  $K$  and  $\alpha$  simultaneously appears to be a difficult task. Therefore, in this study, we follow the practical approach of Chavent et al. (2018), which is conditional on the number of groups  $K$  and allows the selection of the value of  $\alpha$  that offers the best compromise between the loss in terms of attributes and the loss of geographical homogeneity. For a given grid of  $M$  values for  $\alpha$ , this is achieved by measuring the proportion of *explained pseudo-inertias* (Chavent et al. 2018)

$$Q_0(\mathcal{P}_K^{\alpha_m}) = 1 - \frac{W_0(\mathcal{P}_K^{\alpha_m})}{W_0(\mathcal{P}_1^{\alpha_m})} \quad \text{and} \quad Q_1(\mathcal{P}_K^{\alpha_m}) = 1 - \frac{W_1(\mathcal{P}_K^{\alpha_m})}{W_1(\mathcal{P}_1^{\alpha_m})} \quad \text{for } m = 1, \dots, M$$

that, for partition  $\mathcal{P}_K^{\alpha_m}$ , describe the loss of attribute homogeneity and the loss of geographical homogeneity, respectively. As a result, by examining the patterns of  $Q_0(\mathcal{P}_K^{\alpha_j})$  and  $Q_1(\mathcal{P}_K^{\alpha_j})$  as functions of  $\alpha_m$ , the algorithm attempts to find a contiguous spatial solution with the smallest weight  $\alpha_m$  provided to the centroid coordinates. This approach appears to be consistent with the hints provided by the exploratory data analysis and is thus pursued in the clustering procedure.

## Results

### Choice of the number of clusters and mixing parameter

Following the approach introduced in the previous section, we computed  $D_1$  as a proximity matrix equal to the distances (in kilometres) between the centroids of the aerial units. In this manner, we can detect the proximity effects of the pandemic, such as local restrictive measures and sharing of resources and account for the relevance of spatial proximity in the spread of the virus during the first wave. We also computed  $D_0$  as the matrix of the Euclidean distances between the coefficients of the splines. This approach then foresees the choice of the number of clusters and mixing parameter  $\alpha$  which denotes the relevance of geographical distances in the definition of clusters.

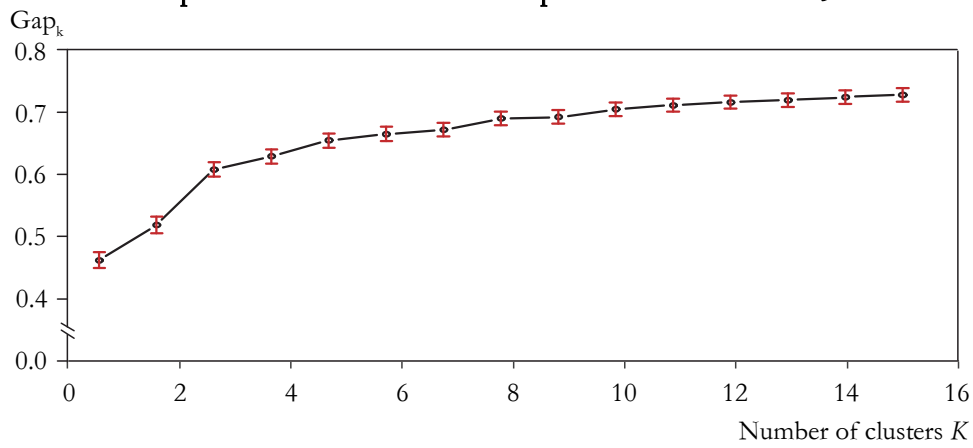
As further discussed in previous section, we used both the *gap statistic* by Tibshirani et al. (2001) and *average silhouette width* by Kaufman–Rousseeuw (1990) to select the optimal number of clusters. Using only the information from  $D_0$ , we calculated the *gap statistic* for  $K$  ranging from 1 to 15 and sampling with  $M = 200$  replicates. By observing the values of this measure in Figure 1, we can deduce that the highest *gap*



*statistic* is observed for 15 clusters. However, it can also be noticed that the standard error bars overlap for values of  $K$  from 11 to 15. Simultaneously, the silhouette plot in Figure 2 shows that the highest value of average silhouette width can be observed for  $K = 11, 12, 13, 14, 15$ . Following Tibshirani et al. (2001), we selected the smallest number of  $K$  such that the standard error bars of the *gap statistic* over the  $M$  replicates overlap. Therefore, the number of clusters was set to  $K = 11$ .

Figure 1

### Gap statistic for values of $K$ comprised between 1 and 15



*Notes:* In red we also report the standard error bars calculated as in Tibshirani et al. (2001) on the bootstrapped sample with  $M = 200$ .

Figure 2

### Average silhouette width calculated as the mean of the silhouette width among several clusters comprised between 1 and 15

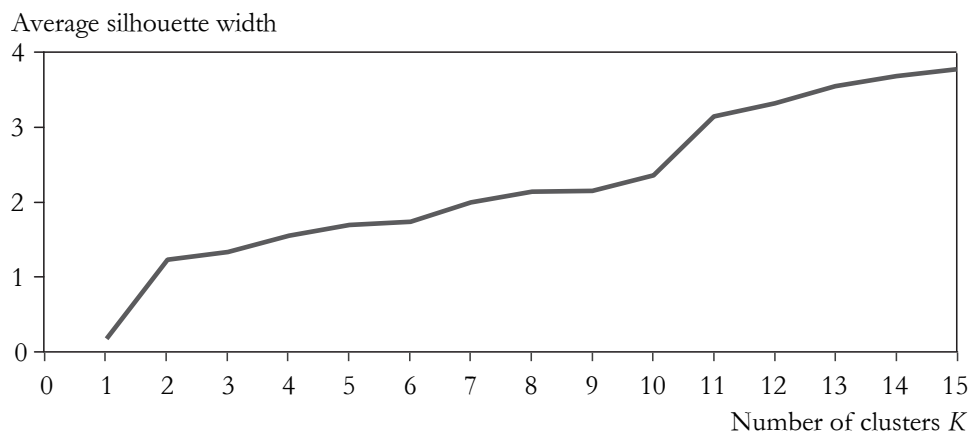
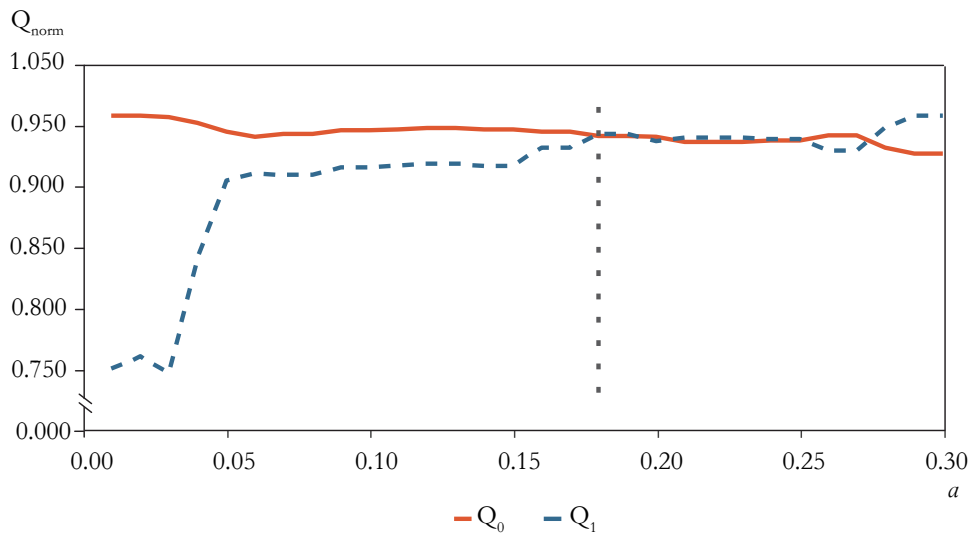


Figure 3

Plot of  $Q_0$  and  $Q_1$  for the selection of the mixing parameter  $\alpha$  for a partition of  $K = 11$  clusters



Notes: Normalised proportion of explained pseudo-inertias  $Q_0(\mathcal{P}_K^\alpha)$  versus  $\alpha$  (in red) and  $Q_1(\mathcal{P}_K^\alpha)$  versus  $\alpha$  (in dotted light blue). The vertical line denotes  $\alpha = 0.18$ .

The choice of  $\alpha$  relies on the values of the *explained pseudo-inertias* as suggested by Chavent et al. (2018). In this case, we allowed  $\alpha$  to vary between 0.01 and 0.30. We analysed only values of  $\alpha$  below 0.30; because for larger weights to the dissimilarity matrix  $D_1$ , the clusters would entirely coincide with the countries, and no interesting information could be inferred from the analysis. By observing the values of normalised pseudo-inertias calculated with  $D_0$  and  $D_1$  in Figure 3, we can deduce that the optimal choice for  $\alpha$ , conditioned on  $K = 11$  clusters, is 0.18, which corresponds to a loss of only 1.3% of feature homogeneity and a 19.3% increase in spatial homogeneity.

### Clustering results

Figure A2 in the Appendix shows the clusters on a map. It is interesting to note that Northern Germany belongs to a single cluster, that is, cluster 1. The same occurred for most Spanish regions. Simultaneously, Southern France is entirely grouped in cluster 9, while the North–South division of Italy seems to be totally reflected by the clusters. This indicates that spatial proximity can still be relevant in enhancing the information using time series patterns to obtain clusters of regions. In addition, some of the most polluted regions in Europe (EEA 2020) belong to the same cluster, that is, cluster 8. In such regions, the behaviour of the pandemic in terms of death counts

is notably different from that in other Italian regions. It is worth noting that this group includes the Lombardy region, where the European spread of the pandemic began, and Catalonia, which is grouped with Lombardy as in Bucci et al. (2022). Figure A2 also highlights that the United Kingdom can be split into two sub-regions: the South-Eastern territory belongs to cluster 5, while the rest of the country is grouped in cluster 2. Interestingly, most of the groups divide the countries in terms of latitude, in particular for Italy, the United Kingdom, and Germany. This is consistent with the theory that the evolution of Covid-19 is strongly linked to the average temperature of the region (Wu et al. 2020).

Looking at the time series per cluster in Figure A3 in the Appendix, it is evident that German regions (in clusters 1 and 7) are characterised by a peak in the second phase of the pandemic. This peak is clearly defined in cluster 1, which includes most German regions. Conversely, regions from clusters 2, 4, and 8 were mostly affected in the first phase of Covid-19 spread. The French regions belonging to cluster 6 showed a long peaking phase from October 2020–March 2021. Furthermore, the regions from Spain in cluster 4 exhibited a high peak in the first months of the pandemic and lower peaks in the second phase. The plot of the time series in cluster 3 shows that the Ile-de-France region and its neighbours were heavily hit in the first phase, but for a short period, and the second phase of the pandemic has no relevant peaks. Finally, the Italian regions of clusters 10 and 11 suffered a dramatic second wave of deaths between October 2020 and April 2021. The regions of clusters 9 and 11 were also the most affected in the third phase of the pandemic, with a peak in February 2022.

The descriptive analysis of our clusters shows that variability in terms of the time series is rather limited. This is similar to what was identified by Bucci et al. (2022), where the clusters in Germany, the United Kingdom, and Italy were particularly similar to what was identified in this study. A cluster of high-risk regions in Northern Italy and in Eastern France up to Ile-de-France was also found by Amdaoud et al. (2021). D'Urso et al. (2022) shared the concept of three main clusters in Italy, identifying an almost overlapping cluster for the central part of the Italian territory.

### **Analysis of the effects of socioeconomic, demographic, and environmental variables on total death counts**

The spatial aspect of the pandemic, as highlighted in the previous section, is crucial to understand the local patterns of the consequences of the virus and act accordingly. We also note that this aspect is not the only one characterising the spread of the pandemic, because most of the regions have experienced the effects of Covid-19 sooner or later. A relevant question at this point may be the role of initial conditions in the spread of the virus and the exacerbation of its effects (Ngepah 2021).

To understand whether some variables are associated with the total number of deaths, after controlling the presence of the regions in a cluster, we fit a spatial

generalised linear mixed (GLM) model to regions, where the response variable is the total number of deaths in each region, and the random effects are modelled by a conditional autoregressive (CAR) prior specified as in Leroux et al. (2000). The purpose of this specification is mainly to understand whether being grouped in a specific cluster has led to different outcomes in terms of the total number of deaths, and if explanatory variables influence death counts after controlling the patterns of the series through clusters. It would be of further interest to analyse the interaction between clusters and the variables considered; however, the number of regions in each cluster was not sufficient to obtain robust estimates of coefficients.

Let  $Y = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))'$  denote the  $n$ -dimensional vector of cumulative confirmed Covid-19 deaths, and let  $X(\mathbf{s}_i) = (1, x_{i1}, \dots, x_{ip})'$  be the vector of  $p$  covariates available for each region  $\mathcal{A}_i$  with geographical coordinates  $\mathbf{s}_i$ . Then, the following Bayesian hierarchical Poisson-LogNormal model (Banerjee et al. 2014) was used for our count data:

$$Y(\mathbf{s}_i) | \mu_i = E_i \lambda_i \sim \text{Poisson}(E_i \lambda_i) \quad \text{for } i = 1, \dots, n \quad (2)$$

$$\log(\lambda_i) = \mathbf{X}(\mathbf{s}_i)' \boldsymbol{\beta} + O_i + \psi_i \quad (3)$$

where  $E_i$  denotes the expected cases,  $\lambda_i$  the relative risk for the  $i$ -th region,  $O_i$  represents the offset computed as the logarithm of the population, and  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$  represents the vector of the covariate effects. The random effects  $\boldsymbol{\psi} = (\psi_1, \dots, \psi_n)$  are intended to capture any residual spatial autocorrelation present in the death count data and are assigned with a conditional autoregressive prior specified as in Leroux et al. (2000). Specifically, the distribution of  $\psi_i$  conditional on all other effects  $\psi_j$ , for  $i \neq j$ , is expressed as follows:

$$\psi_i | \psi_j, i \neq j, \tau^2, H \sim N \left( \frac{\rho \sum_j \psi_j \eta_{ij}}{\rho \sum_j \eta_{ij} + 1 - \rho}, \frac{\tau^2}{\rho \sum_j \eta_{ij} + 1 - \rho} \right) \quad (4)$$

where  $\rho$  represents a spatial autocorrelation parameter such that  $\rho = 0$  implies independence, while  $\rho = 1$  corresponds to the intrinsic CAR model,  $\eta_{ij}$  is the  $ij$ -th element of the symmetric  $(n \times n)$  adjacency matrix  $H$ , equal to one if  $(s_i, s_j)$  share a common border, zero otherwise, and with  $\eta_{ii} = 0$ . A multivariate Gaussian prior is assumed for  $\boldsymbol{\beta}$ , with zero mean and a diagonal covariance matrix with a value of 100,000 on the diagonal elements, whereas the hyperpriors are  $\tau_u^2 \sim \text{InverseGamma}(1, 0.01)$  and  $\rho \sim \text{Uniform}(0, 1)$ . The inference for this model is based on 10,000 Markov chain Monte Carlo (MCMC) samples obtained by running the chain for 15,000 samples and discarding 5,000 samples as the burn-in period. The estimated parameters were obtained through the `S.CARleroux()` function in the `CARBayes` R package by Lee (2013).

In our model,  $\mathbf{X}(\mathbf{s}_i)$  contains a set of covariates and factor variable for the clusters. Many social, economic, and environmental factors were attributed as potential determinants of the outcomes of the Covid-19 pandemic. Some researchers relate

these outcomes to the aging population (Gardner et al. 2020) or inadequate healthcare infrastructure (Le et al. 2021, Alfonso et al. 2021), while others analyse the role of the natural environment (Wu et al. 2020) or social conditions in the region (Kim–Bostwick 2020, Wildman 2021). In most cases, understanding the mechanism underlying the association between death counts and pre-existing conditions is not straightforward. For instance, a higher economic status in a region not only permits better care, but also increases networking, thus facilitating the spread of the virus (Ehlert 2021). For this reason, this analysis is intended to provide insights into hypothesis-making rather than assuming causal effects.

In selecting the covariates, we were guided by existing ecological studies on respiratory diseases (Horwitz et al. 2020) and the availability of data for the administrative units we considered. Among the variables used as covariates, to proxy for the level of healthcare infrastructure (Ngepah 2021, Ehlert 2021), we included the number of hospital beds per hundred thousand inhabitants because health resource shortages have worsened the effects of the virus during the earlier phases of its spread (Remuzzi–Remuzzi 2020). The influence of income and wellness on Covid-19 deaths were discussed in several studies (Plümper–Neumayer 2020, Tan et al. 2021, Welsh et al. 2022). Accordingly, we used the GDP of the region as a wellness measure and the percentage of persons at risk of poverty or social exclusion as a proxy for social inequalities in the European regions (Mari–Dell’Olmo et al. 2021). Clinical data and surveillance reported two important aspects of the pandemic: the elderly population and patients with comorbidities, such as diabetes (Barron et al. 2020), hypertension, and cardiovascular and respiratory diseases (Ssentongo et al. 2020) were at a greater risk of dying from Covid-19. For this reason, our set of covariates comprehends the death rates of diabetes, respiratory, and cardiovascular diseases, and the rate of the population aged above 75 years (Daoust 2020). The levels of particulate matter (PM 2.5) (computed as the average of the annual levels in the period 2012–2016) were included in the set of characteristics of the regions because they are often related to the spread of the pandemic (Comunian et al. 2020, Lolli et al. 2020) and provided that the population which is already exposed to air pollution may experience stronger effects of the virus. Finally, to control any geographic-related features, such as temperature and humidity (Wu et al. 2020), the set of covariates also included the latitude and longitude of the region centroids. We then included cluster-related dummies to capture differences in terms of the total burden of the pandemic across clusters. To avoid singularity in the design matrix, the dummy for the first cluster was excluded and considered as a reference for the interpretation of the relative risks. Table 1 provides a more detailed description of the explanatory variables.

Table 1

**Description of socio-economic, demographic, and  
environmental explanatory variables**

Variable	Description	Year	Source
Beds	Number of hospital beds per hundred thousand inhabitants	2018	Eurostat
GDP	Per capita GDP in thousands Euro	2018	Eurostat
At risk	Percentage of total population at risk of poverty or social exclusion	2018	Eurostat
Diabetes	Crude death rate (per 1 000 persons) by diabetes, expressed as the ratio of the number of deaths during the year to the average population in that year	2017	Eurostat
Respiratory	Crude death rate (per 1 000 persons) by respiratory diseases, expressed as the ratio of the number of deaths during the year to the average population in that year	2017	Eurostat
Cardio	Crude death rate (per 1 000 persons) by cardiovascular diseases, expressed as the ratio of the number of deaths during the year to the average population in that year	2017	Eurostat
Over 75	Percentage of population with an age $\geq$ 75 years	2020	Eurostat
PM 2.5	5-years average of annual concentrations (micrograms per cubic meter) of ground-level fine particulate matter (PM 2.5)	2012–2016	NASA Socio-economic Data and Applications Center
Latitude	Latitude of the centroid in decimal degrees		
Longitude	Longitude of the centroid in decimal degrees		

The results of the GLM model with spatial effects are listed in Table 2, where all coefficient results are presented as an incidence rate ratio (relative risk) of an increase of one unit in each covariate. This study relied only on ecological-level data. This indicates that we do not make any generalization about individuals or specific population groups or infer any causality between the covariates and dependent variable.

The estimated relative risks show that the GDP, percentage of people at risk of poverty, and total number of hospital beds are not statistically relevant in the modelling of the number of deaths, while the other covariates and factor variables related to clusters have credible intervals for the relative risks that do not include one. Except for the death rate because of cardiovascular diseases which has no significant effect on cumulative death counts, the covariates used to proxy the health status of the population have a positive association with the rate of deaths by Covid-19. This is not surprising given that the Covid-19 pandemic may have exacerbated the precarious health conditions of the population. For example, for an additional unit of the percentage of the population with an age equal to or greater than 75 years, the

rate of deaths increases by 7.4%. This is also true for PM which is associated with a 4.2% increase in the death rate, highlighting a possible link between air pollution and disease exacerbation. It is worth noting that this effect may reflect the association of the dependent variable with other variables such as the level of industrialisation or population density. The latitude has a negative association with the number of deaths, meaning that regions at higher latitudes have experienced a lower death rate. Instead, among the regions considered, the Eastern regions experienced higher death rates. The estimates for most of the ten levels of the factor variables, compared to the baseline category of the first cluster, were significant. This indicates that by fixing the other explanatory variables, clusters may also help to discriminate among the total number of deaths. Finally, inferential results for the estimated spatial autoregressive parameter,  $\rho$ , suggest that the fit of the data improves by including spatial correlation effects in the model.

Table 2

### Estimated relative risks with 95% credible intervals

Variable	Relative risk	Lower Bound	Upper Bound
Beds	1.000	1.000	1.000
GDP	1.000	1.000	1.000
At risk	1.004	0.995	1.010
Diabetes	1.002*	1.001	1.005
Respiratory	1.007*	1.006	1.008
Cardio	0.998	0.998	1.002
Over 75	1.074*	1.052	1.088
PM 2.5	1.042*	1.034	1.048
Latitude	0.988*	0.983	0.999
Longitude	1.040*	1.034	1.049
Cluster 2	1.774*	1.665	1.913
Cluster 3	1.708*	1.566	1.862
Cluster 4	1.560*	1.502	1.824
Cluster 5	2.195*	2.020	2.378
Cluster 6	1.289*	1.101	1.879
Cluster 7	0.847	0.755	0.935
Cluster 8	1.303*	1.108	1.395
Cluster 9	0.989	0.907	1.035
Cluster 10	0.840*	0.768	0.903
Cluster 11	0.981	0.807	1.139
$\tau^2$	0.222	0.159	0.304
$\rho$	0.520	0.283	0.773
DIC	1696.782		
LMPL	-892.45		

*Notes:* Estimated median relative risks are presented with credible intervals of 95%. The table includes the deviance information criterion (DIC) and log-pseudo-marginal likelihood (LMPL). \* denotes a credible interval of 95% for relative risk that does not overlap unity.  $\rho$  and  $\tau^2$  were not transformed.

## Discussion and conclusions

The results of the clustering analysis in this manuscript suggest that there are 11 groups of regions in Europe that exhibit similar Covid-19 patterns in the time series of daily confirmed cases. These clusters often coincide with an entire territory within a country, for example, Northern Italy in cluster 8 and Northern Germany in cluster 1.

Although the task of providing a useful interpretation for clusters and attempting to generalise the characteristics of the Covid-19 pandemic is not straightforward, some interesting findings can be deduced from our analysis. First, the analysis of the time series suggests that the clusters are primarily characterised by different phases of the pandemic. Moreover, all regions experienced the effects of the disease sooner or later without being able to completely limit its spread. This could suggest that the implementation of persistent restrictive measures, such as in the first phase of the pandemic in Germany and Italy, could have helped reduce the outcomes of the virus. Second, based on the size of the mixing parameter in the clustering procedure and estimated spatial correlation from the GLM model, spatial proximity appears to be relevant in the modelling of Covid-related variables of interest. This is not surprising, because most countries have implemented restrictive measures aimed at the entire population, particularly in the first phase of the pandemic or during certain periods, such as Christmas holidays. Our study further underlines the relevance of local differences in deaths during the spread of Covid-19. Regional conditions vary across Europe and even within a single country, with clusters of patterns showing a spatial distribution. Thus, an understanding of the pattern of deaths by Covid-19 across countries can assist decision-makers take restrictive measures in well-defined areas or move resources where most required. Regional choices in terms of public health policies are crucial to disentangle the spatial aspects of the pandemic, even in countries where health is managed at a central level for a two-fold reason: Covid-19 appears to spread in a non-uniform manner across the regions but tends to remain clustered; the low availability of intensive care units requires accurate management of the resources, and acting at a local level can help to detect criticality in the health system and fix it. Consequently, for countries where the pandemic has a different impact within the territory, acting locally can help prevent the effects of the pandemic in further possible waves, for example in regions belonging to the same cluster in previous phases but not yet overwhelmed by the pandemic. Finally, what emerges from our study is that to avoid working with noisy data, a common protocol in defining cases and deaths because of Covid-19 is necessary. In this study, we solved the possible discrepancies between the real and measured counts by smoothing the time series; however, more robust data would inevitably result in a more robust analysis.



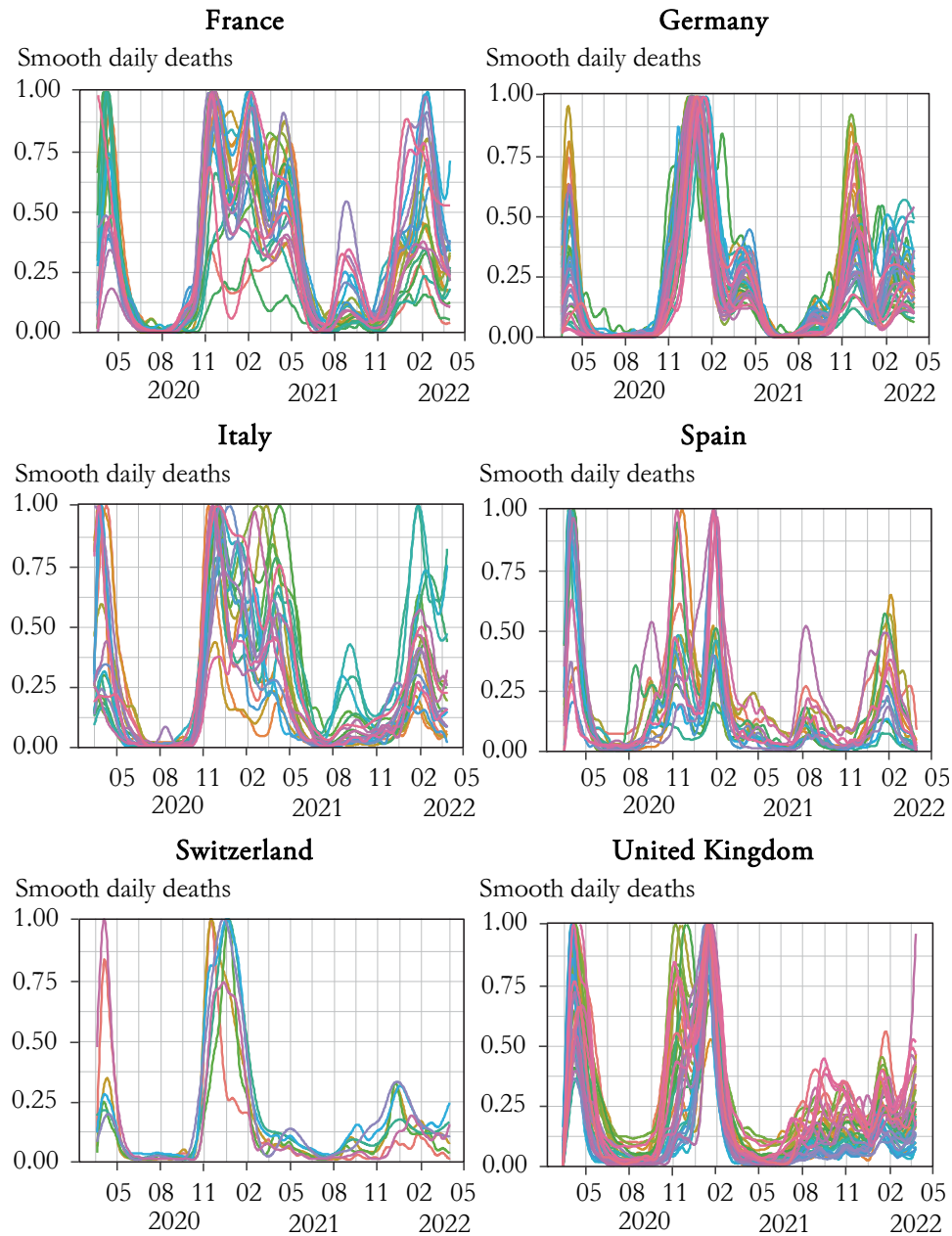
In our study, we also analysed the impact of socio-economic, demographic, and environmental variables, and the effect of being in a cluster on the total number of deaths in each region. Among the considered variables, GDP, the percentage of persons at risk of poverty or social exclusion, the death rate because of cardiovascular diseases, and hospital beds are the only variables not significantly associated with the dependent variable. For what concerns the GDP, this result reflects the small differences in terms of GDP among the regions considered in this analysis. It is likely that GDP would have had a more relevant effect if the analysis also covered the poorest European countries. The same reflection can be made regarding the risks of social exclusion. The variability of this indicator in the countries considered in this study is limited, and this could have led to a non-significant effect on the outcome of the pandemic. The number of hospital beds, on the contrary, has changed a lot during the pandemic, and the lack of significance of the association with the number of deaths because of Covid-19 may be owing to outdated data. Conversely, all other covariates had a statistically significant association with the death rate by Covid-19. Therefore, pre-existing conditions appear to be somehow linked to the evolution of the pandemic and may have helped limit or exacerbate the effects of the spread of Covid-19. This implies that restrictive measures should be complemented by preventive measures in different fields. Improving health infrastructure can help curb mortality and avoid overloading in hospitals, while promoting healthy lifestyles and treatment of communicable diseases can be crucial to alleviate the burden of a pandemic as well as reduce respiratory damage because of air pollution.

Future studies could rely on the approach presented in this manuscript to predict the evolution of the pandemic at an aggregated level. In fact, grouping regions with similar temporal evolution of the pandemic can help forecast the phenomenon while maintaining the setting of the forecast in a low-dimensional space. Finally, the availability of data for a larger number of regions would allow the evaluation of the interaction between clusters and pre-existing conditions.

**Appendix**

Figure A1

**Normalized time series of smoothed confirmed deaths for each country**



*Note:* The bold lines indicate the average of the time series per country.

Figure A2

Map of the  $K = 11$  clusters obtained through hierarchical clustering  
by mixing parameter  $\alpha = 0.18$

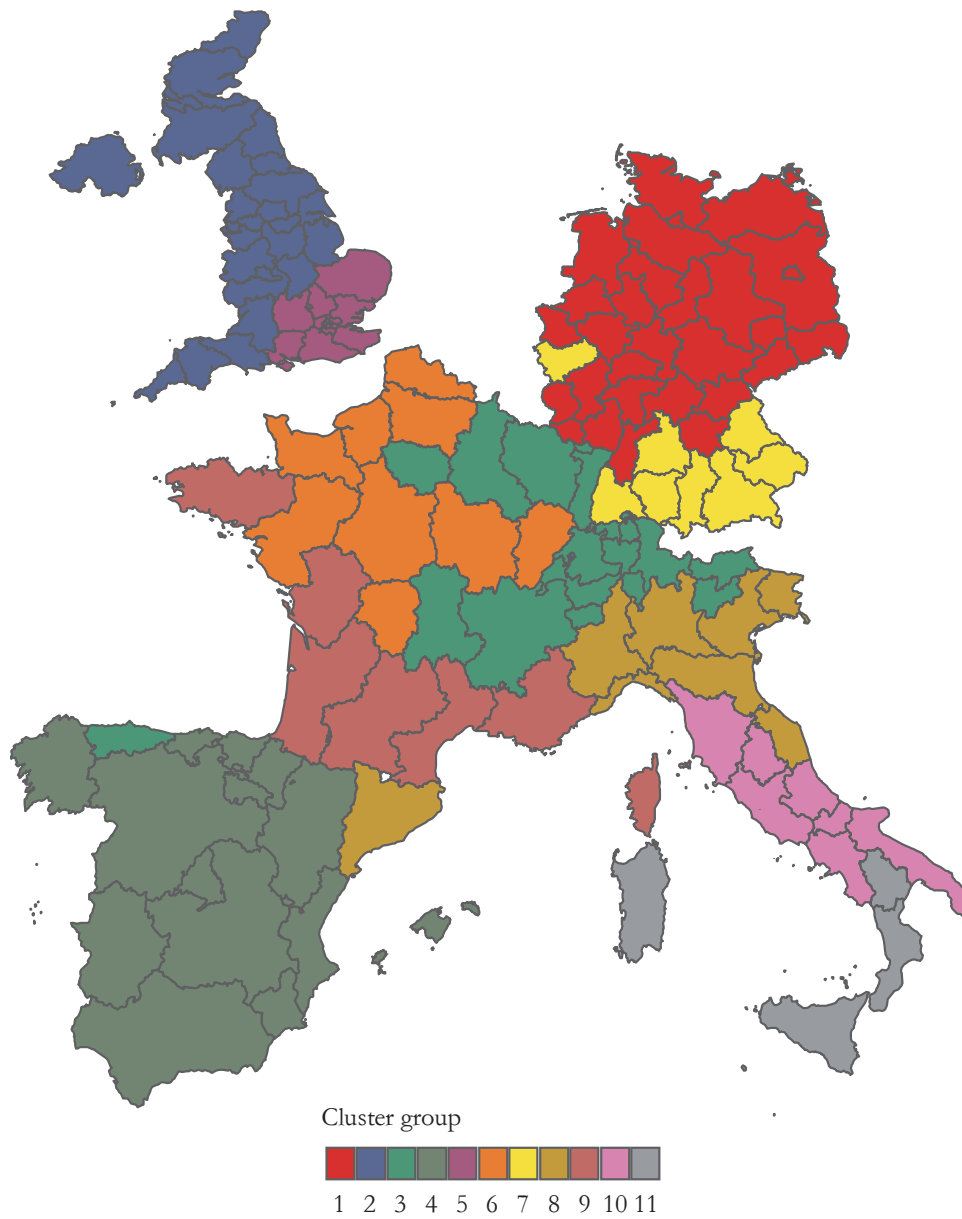
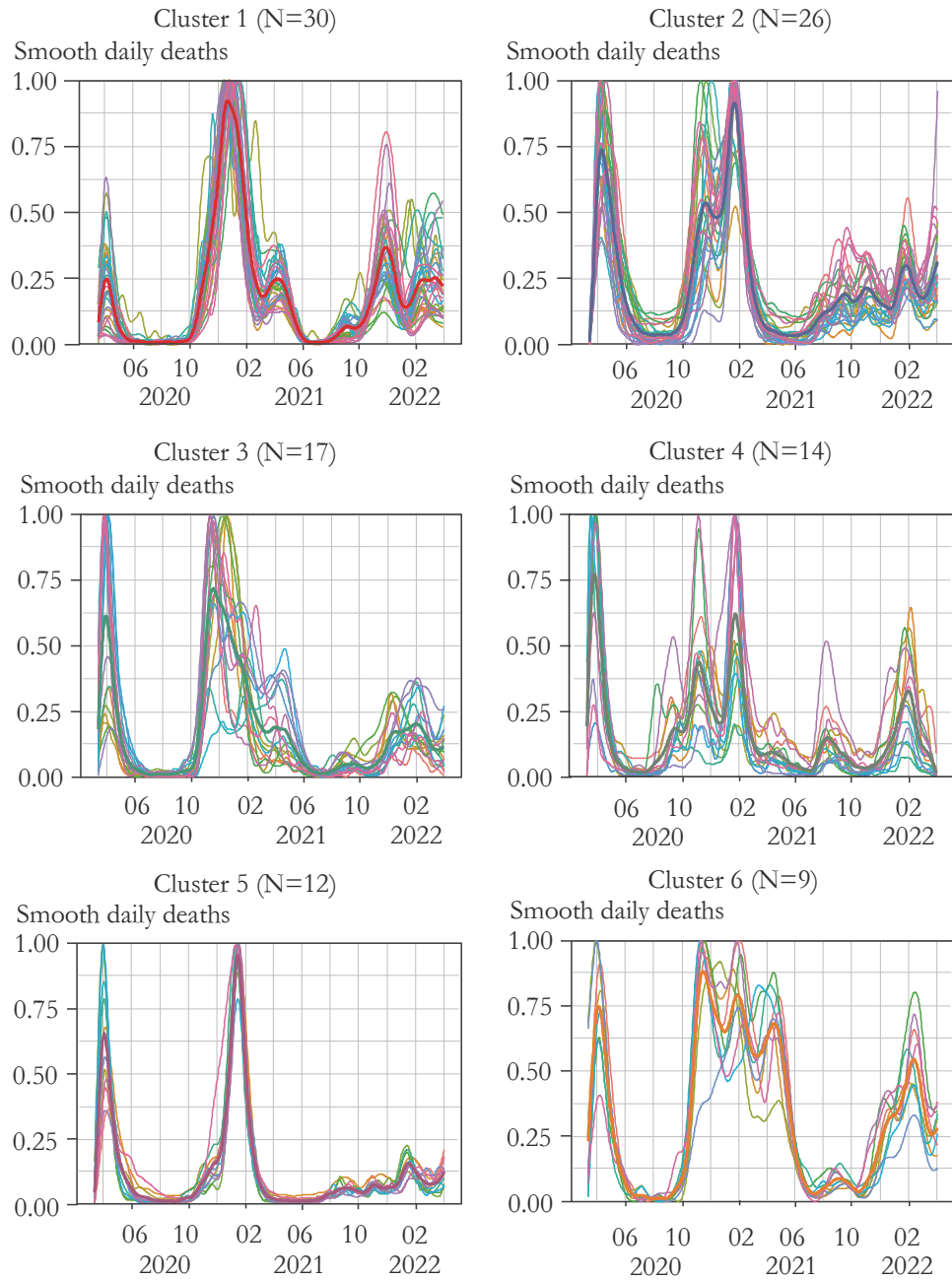
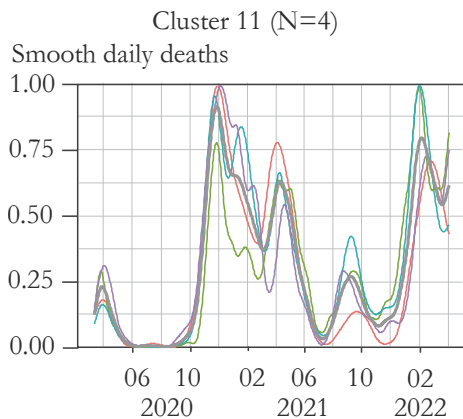
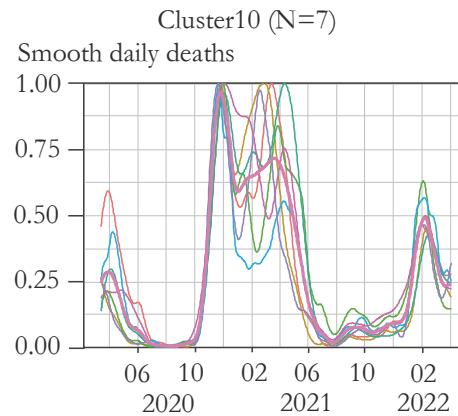
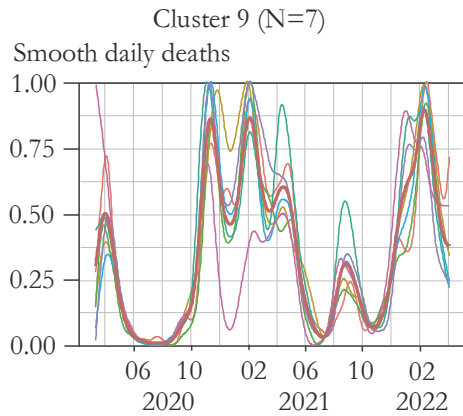
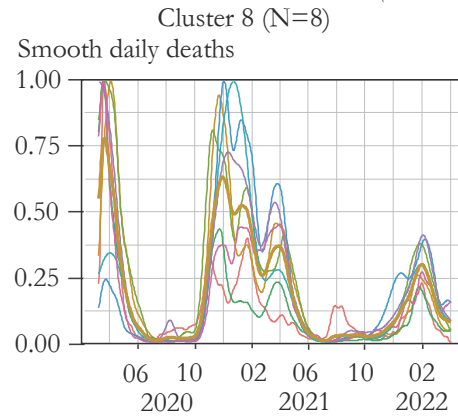
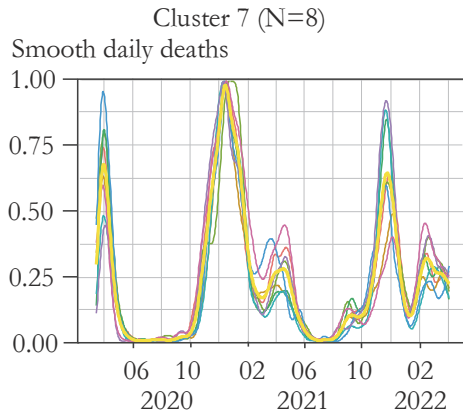


Figure A3

**Normalised time series of smoothed confirmed deaths for each cluster**



*(Figure continues on the next page.)*

*(Continued.)*

*Notes:* The bold lines indicate the average of the time series per cluster. The values between brackets denote the number of regions in each cluster.

### Acknowledgments

Andrea Bucci acknowledges the grant received by the project PON “Ricerca e Innovazione” 2014-2020 – Azione I.2 “Mobilità dei Ricercatori” – Avviso di cui al D.D. n. 407 del 27 febbraio 2018 e D.D. n.1621 del 12/08/2019 “Attraction and International Mobility” - Linea 1 (Mobilità dei Ricercatori) (CUP n. D24I19001680001 – Id. proposta: AIM1894803 Linea 1: attività 3 Area di specializzazione SNSI: Smart, Secure and Inclusive Communities).

### REFERENCES

- ALFANO, V.–ERCOLANO, S. (2020): The efficacy of lockdown against Covid-19: A cross-country panel analysis *Applied Health Economics and Health Policy* 18: 509–517. <https://doi.org/10.1007/s40258-020-00596-3>
- ALFONSO, Y. N.–LEIDER, J. P.–RESNICK, B.–MCCULLOUGH, J. M.–BISHAI, D. (2021): US public health neglected: Flat or declining spending left states III equipped to respond to Covid-19 *Health Affairs* 40 (4): 664-671. <https://doi.org/10.1377/hlthaff.2020.01084>
- ALVES, J.–SOARES, P.–ROCHA, J. V.–SANTANA, R.–NUNES, C. (2021): Evolution of inequalities in the coronavirus pandemics in Portugal: An ecological study *European Journal of Public Health* 31 (5): 1069–1075. <https://doi.org/10.1093/eurpub/ckab036>
- AMDAOUD, M.–ARCURI, G.–LEVRATTO, N. (2021): Are regions equal in adversity? A spatial analysis of spread and dynamics of Covid-19 in Europe *The European Journal of Health Economics* 22: 629–642. <https://doi.org/10.1007/s10198-021-01280-6>
- ANDERSEN, L. M.–HARDEN, S. R.–SUGG, M. M.–RUNKLE, J. D.–LUNDQUIST, T. E. (2021): Analyzing the spatial determinants of local Covid-19 transmission in the United States *Science of The Total Environment* 754: 142396. <https://doi.org/10.1016/j.scitotenv.2020.142396>
- ANTALÓCZY, K.–BIRIZDÓ, I.–SASS, M. (2022): Local investment promotion in a Hungarian medium-sized town and the implications of the Covid pandemic *Regional Statistics* 12 (1): 27–50. <https://doi.org/10.15196/RS120104>
- ARITENANG, A. F. (2022): The impact of urban characteristics on the spread of Covid-19 in 2020: The case of Java Island cities, Indonesia *Regional Statistics* 12 (3): 3–17. <https://doi.org/10.15196/RS120301>
- BANERJEE, S.–CARLIN, B. P.–GELFAND, A. E. (2014): *Hierarchical modeling and analysis for spatial data* Chapman and Hall/CRC.
- BARRON, E.–BAKHAI, C.–KAR, P.–WEAVER, A.–BRADLEY, D.–ISMAIL, H.–KNIGHTON, P.–HOLMAN, N.–KHUNTI, K.–SATTAR, N.–WAREHAM, N. J.–YOUNG, B.–VALABHJI, J. (2020): Associations of type 1 and type 2 diabetes with Covid-19-related mortality in England: A whole-population study *The Lancet Diabetes & Endocrinology* 8 (10): 813–822. [https://doi.org/10.1016/S2213-8587\(20\)30272-2](https://doi.org/10.1016/S2213-8587(20)30272-2)
- BOURDIN, S.–BEN MILED, S.–SALHI, J. (2022): The drivers of policies to limit the spread of Covid-19 in Europe *Journal of Risk and Financial Management* 15 (2): 67. <https://doi.org/10.3390/jrfm15020067>

- BOURDIN, S.–JEANNE, L.–NADOU, F.–NOIRET, G. (2021): Does lockdown work? A spatial analysis of the spread and concentration of Covid-19 in Italy *Regional Studies* 55 (7): 1182–1193. <https://doi.org/10.1080/00343404.2021.1887471>
- BOURDREL, T.–ANNESI-MAESANO, I.–ALAHMAD, B.–MAESANO, C. N.–BIND, M. A. (2021): The impact of outdoor air pollution on Covid-19: a review of evidence from in vitro, animal and human studies *European Respiratory Review* 30 (159): 200242. <https://doi.org/10.1183/16000617.0242-2020>
- BUCCI, A.–IPPOLITI, L.–VALENTINI, P.–FONTANELLA, S. (2022): Clustering spatio-temporal series of confirmed Covid-19 deaths in Europe *Spatial Statistics* 49: 100543. <https://doi.org/10.1016/j.spasta.2021.100543>
- CHAVENT, M.–KUENTZ-SIMONET, V.–LABENNE, A.–SARACCO, J. (2018): ClustGeo: An R package for hierarchical clustering with spatial constraints *Computational Statistics* 33: 1799–1822. <https://doi.org/10.1007/s00180-018-0791-1>
- COMUNIAN, S.–DONGO, D.–MILANI, C.–PALESTINI, P. (2020): 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. <https://doi.org/10.3390/ijerph17124487>
- DAOUST, J. F. (2020): Elderly people and responses to Covid-19 in 27 countries *PLOS ONE* 15 (7): e0235590. <https://doi.org/10.1371/journal.pone.0235590>
- DAOUST, J. F. (2022): Is the unequal Covid-19 burden in Canada due to unequal levels of citizen discipline across provinces? *Canadian Public Policy* 48 (1): 124–143. <https://doi.org/10.3138/cpp.2021-060>
- DE PIETRO, C.–CAMENZIND, P.–STURNY, I.–CRIVELLI, L.–EDWARDS-GARAVOGLIA, S.–SPRANGER, A.–WITTENBECHER, F.–QUENTIN, W. (2015): Switzerland: Health system review *Health Systems in Transition* 17 (4): 1–288.
- DESMET, K.–WACZIARG, R. (2022): JUE Insight: Understanding spatial variation in Covid-19 across the United States *Journal of Urban Economics* 127: 103332. <https://doi.org/10.1016/j.jue.2021.103332>
- D'URSO, P.–DE GIOVANNI, L.–VITALE, V. (2022): Spatial robust fuzzy clustering of Covid 19 time series based on B-splines *Spatial Statistics* 49:100518. <https://doi.org/10.1016/j.spasta.2021.100518>
- EGRI, Z. (2017): Regional health inequalities in the European microregion from the East Central European perspective *Regional Statistics* 7 (1): 197–224. <https://doi.org/10.15196/RS0711>
- EHLERT, A. (2021): The socio-economic determinants of Covid-19: A spatial analysis of German county level data *Socio-Economic Planning Sciences* 78: 101083. <https://doi.org/10.1016/j.seps.2021.101083>
- EILERS, P. H. C.–MARX, B. D. (1996): Flexible smoothing with B-splines and Penalties *Statistical Science* 11 (2): 89–102. <https://doi.org/10.1214/ss/1038425655>
- FLYNN, D.–MOLONEY, E.–BHATTARAI, N.–SCOTT, J.–BRECKONS, M.–AVERY, L.–MOY, N. (2020): Covid-19 pandemic in the United Kingdom *Health Policy and Technology* 9 (4): 673–691. <https://doi.org/10.1016/j.hlpt.2020.08.003>
- GARATTINI, L.–BADINELLA MARTINI, M.–ZANETTI, M. (2022): The Italian NHS at regional level: same in theory, different in practice *European Journal of Health* 23: 1–5. <https://doi.org/10.1007/s10198-021-01322-z>

- GARDNER, W.–STATES, D.–BAGLEY, N. (2020): The coronavirus and the risks to the elderly in long-term care *Journal of Aging & Social Policy* 32 (4–5): 310–315.  
<https://doi.org/10.1080/08959420.2020.1750543>
- GULIYEV, H. (2020): Determining the spatial effects of Covid-19 using the spatial panel data model *Spatial Statistics* 38: 100443. <https://doi.org/10.1016/j.spasta.2020.100443>
- HAINING, R. F.–WISE, S.–MA, J. (2000): Designing and implementing software for spatial statistical analysis in a GIS Environment *Journal of Geographical Systems* 2 (3): 257–286. <https://doi.org/10.1007/PL00011458>
- HORWITZ, L. I.–CHANG, C.–ARCILLA, H. N.–KNICKMAN, J. R. (2020): Quantifying health systems' investment in social determinants of health, by sector, 2017–19. *Health Affairs* 39 (2): 192–198. <https://doi.org/10.1377/hlthaff.2019.01246>
- KAPÁS, J. (2022): Has Covid-19 caused a change in the dynamics of the unemployment rate? The case of North America and continental Europe *Regional Statistics* 12 (1): 3–26.  
<https://doi.org/10.15196/RS120107>
- KAUFMAN, L.–ROUSSEEUW, P. J. (1990): *Finding groups in data: An introduction to cluster analysis* John Wiley, Hoboken, N. J. <https://doi.org/10.1002/9780470316801>
- KIM, S. J.–BOSTWICK, W. (2020): Social vulnerability and racial inequality in Covid-19 deaths in Chicago *Health Education & Behavior* 47 (4): 509–513.  
<https://doi.org/10.1177/1090198120929677>
- KINCSES, Á.–TÓTH, G. (2020): How coronavirus spread in Europe over time: National probabilities based on migration networks *Regional Statistics* 10 (2): 228–231.  
<https://doi.org/10.15196/RS100210>
- KOSFELD, R.–MITZE, T.–RODE, J.–WÄLDE, K. (2021): The Covid-19 containment effects of public health measures: A spatial difference-in-differences approach *Journal of Regional Science* 61: 799–825. <https://doi.org/10.1111/jors.12536>
- LE, N.–LE, A. V.–PARIKH, J.–BROOKS, J. P.–GARDELLINI, T.–IZURIETA, R. (2021): Ecological and health infrastructure factors affecting the transmission and mortality of Covid-19 *Journal of Advances in Virology Research* 1 (1): 1–7.  
<https://doi.org/10.15744/javr.2021.1.101>
- LEE, D. (2013): CARBayes: An R package for Bayesian spatial modeling with conditional autoregressive priors *Journal of Statistical Software* 55 (13): 1–24.  
<https://doi.org/10.18637/jss.v055.i13>
- LEROUX, B. G.–LEI, X.–BRESLOW, N. (2000): Estimation of disease rates in small areas: A new mixed model for spatial dependence. In: HALLORAN, M. E.–BERRY, D. (eds.): *Statistical Models in Epidemiology, the Environment, and Clinical Trials* pp. 179–191., Springer New York, New York, NY.
- LOLLI, S.–CHEN, Y. C.–WANG, S. H.–VIVONE, G. (2020): Impact of meteorological conditions and air pollution on Covid-19 pandemic transmission in Italy *Scientific Reports* 10 (1): 16213. <https://doi.org/10.1038/s41598-020-73197-8>
- MARÍ-DELL'OLMO, M.–GOTSSENS, M.–PASARÍN, M. I.–RODRÍGUEZ-SANZ, M.–ARTAZCOZ, L.–GARCIA DE OLALLA, P.–RIUS, C.–BORRELL, C. (2021): Socioeconomic inequalities in Covid-19 in a European urban area: Two waves, two patterns. *International Journal of Environmental Research and Public Health* 18 (3): 1256.  
<https://doi.org/10.3390/ijerph18031256>
- MAY, T. (2020): Lockdown-type measures look effective against Covid-19 *BMJ* 2020: 370.  
<https://doi.org/10.1136/bmj.m2809>



- MELNICK, E. R.–IOANNIDIS, J. P. A. (2020): Should governments continue lockdown to slow the spread of Covid-19? *BMJ* 2020: 369. <https://doi.org/10.1136/bmj.m1924>
- NGEPAH, N. (2021): Socio-economic determinants of global Covid-19 mortalities: Policy lessons for current and future pandemics *Health Policy and Planning* 36 (4): 418–434. <https://doi.org/10.1093/heapol/czaa161>
- NYIKOS, GY.–SOHA, B.–BÉRES, A. (2021): Entrepreneurial resilience and firm performance during the Covid-19 crisis – Evidence from Hungary *Regional Statistics* 11 (3): 29–59. <https://doi.org/10.15196/RS110307>
- PÁL, V.–LADOS, G.–MAKRA, Z. I.–BOROS, L.–UZZOLI, A.–FABULA, S. (2021): Concentration and inequality in the geographic distribution of physicians in the European Union, 2006–2018 *Regional Statistics* 11 (3): 3–28. <https://doi.org/10.15196/RS110308>
- PERUCCA, G.–PIACENZA, M.–TURATI, G. (2019): Spatial inequality in access to healthcare: Evidence from an Italian Alpine region *Regional Studies* 53 (4): 478–489. <https://doi.org/10.1080/00343404.2018.1462481>
- PLÜMPER, T.–NEUMAYER, E. (2020): The pandemic predominantly hits poor neighbourhoods? SARS-CoV-2 infections and Covid-19 fatalities in German districts *European Journal of Public Health* 30 (6): 1176–1180. <https://doi.org/10.1093/eurpub/ckaa168>
- REMUZZI, A.–REMUZZI, G. (2020) Covid-19 and Italy: What next? *The Lancet* 395 (10231): 1225–1228. [https://doi.org/10.1016/S0140-6736\(20\)30627-9](https://doi.org/10.1016/S0140-6736(20)30627-9)
- SHARIATI, M.–MESGARI, T.–KASRAEE, M.–JAHANGIRI-RAD, M. (2020): Spatiotemporal analysis and hotspots detection of Covid-19 using geographic information system (March and April, 2020) *Journal of Environmental Health Science and Engineering* 18 (2): 1499–1507. <https://doi.org/10.1007/s40201-020-00565-x>
- SSENTONGO, P.–SSENTONGO, A. E.–HEILBRUNN, E. S.–BA, D. M.–CHINCHILLI, V. M. (2020): Association of cardiovascular disease and 10 other pre-existing comorbidities with Covid-19 mortality: A systematic review and meta-analysis *PLOS ONE* 15 (8): e0238215. <https://doi.org/10.1371/journal.pone.0238215>
- TAN, A. X.–HINMAN, J. A.–ABDEL MAGID, H. S.–NELSON, L. M.–ODDEN, M. C. (2021): Association between income inequality and county-level Covid-19 cases and deaths in the US *JAMA Network Open* 4 (5): e218799–e218799. <https://doi.org/10.1001/jamanetworkopen.2021.8799>
- TIBSHIRANI, R.–WALTHER, G.–HASTIE, T. (2001): Estimating the number of clusters in a data set via the gap statistic *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2): 411–423. <https://doi.org/10.1111/1467-9868.00293>
- WEBSTER, R.–BURROUGH, P. (1972): Computer-based soil mapping of small areas from sample data II classification smoothing *Journal of Soil Science* 23 (2): 222–234. <https://doi.org/10.1111/j.1365-2389.1972.tb01655.x>
- WELSH, C. E.–SINCLAIR, D. R.–MATTHEWS, F. E. (2022): Static socio-ecological Covid-19 vulnerability index and vaccine hesitancy index for England *The Lancet Regional Health – Europe* 14: 100296. <https://doi.org/10.1016/j.lanepe.2021.100296>
- WILDMAN, J. (2021): Covid-19 and income inequality in OECD countries *The European Journal of Health Economics* 22: 455–462. <https://doi.org/10.1007/s10198-021-01266-4>
- WISE, S.–HAINING, R.–MA, J. (1997): Regionalisation tools for exploratory spatial analysis of health data. In: FISCHER, M. M.–GETIS, A. (eds.): *Recent Developments in Spatial*

*Analysis: Spatial Statistics, Behavioural Modelling, and Computational Intelligence*  
pp. 83–100., Springer, New York.

WU, X.–NETHERY, R. C.–SABATH, M. B.–BRAUN, D.–DOMINICI, F. (2020): Air pollution and Covid-19 mortality in the United States: Strengths and limitations of an ecological regression analysis *Science Advances* 6 (45): eabd4049.

<https://doi.org/10.1126/sciadv.abd4049>. 33148655

ZHANG, X.–RAO, H.–WU, Y.–HUANG, Y.–DAI, H. (2020): Comparison of spatiotemporal characteristics of the Covid-19 and SARS outbreaks in mainland China *BMC Infectious Diseases* 20 (1): 805. <https://doi.org/10.1186/s12879-020-05537-y>

### INTERNET SOURCES

EEA (2020): *European Environment Agency – Air quality in Europe – 2020* Technical Report. EEA Report No 9/2020.

[https://www.eea.europa.eu/publications/air-quality-in-europe-2020-report/at\\_download/file](https://www.eea.europa.eu/publications/air-quality-in-europe-2020-report/at_download/file) (downloaded: April 2022).

ISLAM, S. N.–CHENG, H. W. J.–HELGASON, K.–HUNT, N.–KAWAMURA, H.–LAFLEUR, M. (2020): Variations in Covid strategies: Determinants and lessons *United Nations, Department of Economic and Social Affairs*.

[https://digitallibrary.un.org/record/3893518/files/wp172\\_2020.pdf](https://digitallibrary.un.org/record/3893518/files/wp172_2020.pdf)  
(downloaded: April 2022)

### WEBSITES/DATABASES

[1] Sante publique France <https://geodes.santepubliquefrance.fr> (downloaded: May 2022)

[2] NPGeo CoronaHub 2020 (Germany)

<https://npgeo-corona-npgeo-de.hub.arcgis.com> (downloaded: May 2022)

[3] Civil Protection Department (Italy) <https://github.com/pcm-dpc/Covid-19>

(downloaded: May 2022)

[4] Escovid19data (Spain) <https://github.com/montera34/escovid19data>

(downloaded: May 2022)

[5] Open Government Data Canton of Zurich (Switzerland)

[https://github.com/openZH/covid\\_19](https://github.com/openZH/covid_19) (downloaded: May 2022)

[6] UK Government (United Kingdom) <https://coronavirus.data.gov.uk/>

(downloaded: May 2022)