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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Clusteringspatio-temporal series of confirmed COVID-19 deaths in Europe

A. Bucci a, L. Ippoliti a,*, P. Valentini a, S. Fontanella b

a Department of Economics, University G. d’Annunzio, Chieti-Pescara, Italy
b National Heart and Lung Institute, Imperial College London, UK

ARTICLE INFO
Article history:
Received 6 May 2021
Received in revised form 22 September 2021
Accepted 23 September 2021
Available online 6 October 2021

Keywords:
COVID-19
Bayes nonparametrics
Dynamic linear models
Model-based clustering
Spatio-temporal analysis

ABSTRACT
The impact of the COVID-19 pandemic varied significantly across different countries, with important consequences in the definition of control and response strategies. In this work, to investigate the heterogeneity of this crisis, we analyse the spatial patterns of deaths attributed to COVID-19 in several European countries. To this end, we propose a Bayesian nonparametric approach, based on mixture of Gaussian processes coupled with Dirichlet process, to group the COVID-19 mortality curves. The model provides a flexible framework for the analysis of time series data, allowing the inclusion in the clustering procedure of different features of the series, such as spatial correlations, time varying parameters and measurement errors. We evaluate the proposed methodology on the death counts recorded at NUTS-2 regional level for several European countries in the period from March 2020 to February 2021.

1. Introduction
Since the officially reported outbreak in China of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), best known as coronavirus disease 2019 (COVID-19), the world has been facing an international public health emergency. The rapid and global spread of the disease prompted an extensive range of government interventions across the whole world in an attempt to contain the spread of further infection and prevent health system strain. These measures, such as the detection and isolation of infected individuals, contact-tracing, quarantine measures, social
distancing, and closure of non-essential businesses, have since become significant components of public health policies. While Hsiang et al. (2020) have shown that these interventions helped prevent or delay more than 500 million infections across several countries, data on their efficacy are conflicting and many questions remain open. In fact, despite all the efforts, the human cost of coronavirus continues to mount, with about 4.7 million people known to have died at the time of this publication. Like other pandemics, COVID-19 is characterized by a critical spatial dimension such that the impact of this health crisis differs significantly not only across countries, but also across regions and municipalities within countries, both in terms of infections and related deaths. Hence, understanding the differences in the evolution of mortality by COVID-19 between countries constitutes a key factor to discover the dynamics of the disease and the relative effectiveness of preventive interventions on its spread.

To help decision-makers to monitor the efficacy of control and response strategies, appropriate statistical models that allow for the representation and the understanding of spatial patterns and trajectories of both infections and deaths attributed to COVID-19 are required. Recently, several studies have been conducted to analyse the evolution of the COVID-19 pandemic from different perspectives. However, despite the growing literature, there seems to be a gap in the investigation of the different ways in which the pandemic evolves over time and spreads across countries.

Apart from merely providing descriptive statistics, only a few studies rely on proper temporal and spatial analyses of COVID-19, demonstrating the impact of morbidity, mortality and global geographical dissemination of the disease across countries. For example, Shang and Xu (2020) and Kumar et al. (2022) analysed the evolving patterns of the COVID-19 pandemic in time and space, modelling the trajectory of confirmed cases and deaths adopting a change-point perspective.

Regional short-term predictions of COVID-19 intensive care unit beds in Italy were provided by Farcomeni et al. (2020), and a spatial panel data model was proposed by Guliiev (2020) to determine the spatial effects of COVID-19. Some Bayesian spatio-temporal models were also developed by Chen et al. (2020), Feng (2022), Sahu and Böhning (2022), Lee et al. (2022) and Bartolucci and Farcomeni (2022), to assess and predict the distribution of the pandemic at its different stages.

Using a different perspective, Zhang et al. (2020) proposed a space–time scan statistical analysis (Kulldorff et al., 2005) to detect spatio-temporal clusters associated with changes in local containment strategies in China. Spatial clustering techniques were also proposed by Shariati et al. (2020) to study the cumulative incidence and cumulative mortality rates over two time periods of interest, and by D’Urso et al. (2022) to group the 20 Italian regions by taking into account the spatial and temporal features of several variables related to the pandemic.

In this paper, to understand the dynamics of the pandemic and how they differ across different countries and regions, we also propose a clustering approach to group 145 European NUTS-2 Regions according to the temporal similarity of their time series of COVID-19 confirmed deaths. In particular, we seek to evaluate: (a) the way the dynamics of the time series differ across the areas, (b) the existence of possible timing effects resulting in an earlier or later evolution of the phenomenon and, finally, (c) the shape effects resulting in more or less persistent patterns in some regions and specific intervals of time. While the first question aims at quantifying the spatial heterogeneity in terms of temporal trends, the other two refer to time shift and shape effects which arise when the time location of a specific phenomenon and its temporal profile change across the areas. Answers to these questions can offer key information about the impact of the implemented control measures, how these have elicited similar effects across regions and on the necessity of further local interventions in case of persistent patterns. In particular, the identification of clusters of high incidence areas allows future health resources to be targeted appropriately at regions greatly affected in terms of deaths.

There is an extensive literature for clustering time series. Popular approaches include hierarchical and non-hierarchical methods, fuzzy clustering methods, machine learning methods and model-based methods — see, for example, Maharaj et al. (2019) and references therein. Here, we adopt a Bayesian nonparametric approach where time series patterns are modelled as a mixture of Gaussian processes (GP), with a Dirichlet process (DP) prior (Ferguson, 1973) over mixture components. This procedure is appealing on several grounds as, for example, the number of clusters does not need to be determined in advance, but is automatically selected during the clustering procedure within
**Table 1**

| Countries     | Source                                      | Link                                                      |
|---------------|---------------------------------------------|-----------------------------------------------------------|
| France        | Santé publique France                       | https://geodes.santepubliquefrance.fr                     |
| Germany       | NPGeo Corona Hub 2020                       | https://npgeo-corona-npgeo-de.hub.arcgis.com              |
| Italy         | Civil Protection Department                 | https://github.com/pcm-dpc/COVID-19                       |
| Spain         | Escovid19data                               | https://github.com/montera34/escovid19data                |
| Switzerland   | Open Government Data Canton of Zurich       | https://github.com/openZH/covid_19                        |
| United Kingdom| UK Government                               | https://coronavirus.data.gov.uk/                         |

a suitable Markov chain Monte Carlo (MCMC) algorithm. Furthermore, the Gaussian mixture is defined through the observation equation of a state-space model (West and Harrison, 1997), which, not only allows for time varying parameters, but also accommodates for temporal trends, spatial correlations and measurement noise.

The remainder of the paper is organized as follows. In Section 2, we offer a simple exploratory analysis of the data. The methodological contribution is outlined in Section 3 and results of the study are presented in Section 4. Finally, Section 5 concludes the paper with a brief discussion.

2. The data

To investigate the dynamics of the pandemic, we consider data for a period of 52 weeks (18 March 2020–18 February 2021) on COVID-19 confirmed deaths recorded for 145 NUTS-2 Regions in France, Germany, Italy, Spain, Switzerland and United Kingdom. Data were collected from National repositories as reported in Table 1.

Weekly counts of deaths are a key indicator of overall epidemic impact and trajectory and may help improving data quality respect to daily data. Thereby, in our analysis, daily data have been aggregated with a weekly frequency to account for differences in the recording mechanism among countries and to filter out weekend effects.

The raw data, stratified by country, are depicted in Fig. 1. The temporal evolution of the pandemic highlights a clear trend across countries, characterized by two or three peaks in the periods March – May 2020 and October 2020 – March 2021 and by low death counts for all regions during the summer. As expected, the time series within the same country present similar patterns; however, high variability can be observed in all the country-level plot, especially for the second and third waves. This suggests that the impact of the pandemic has been highly asymmetric within countries and that there might exist a supranational pattern across regions. This seems to be confirmed to some extent by Fig. 2 which shows the time series correlations as a function of the spatial distance computed by means of the region centroids. As it can be noticed, although the correlations tend to decay with the spatial distance, high correlation values can still be observed at larger distances. This feature suggests that discriminative methods based on a pairwise spatial similarity metric are not suitable in this context. For this reason, in the following, we adopt a modelling strategy that, while explicitly accounting for the spatial correlation in the model specification, does not directly consider the spatial effect in the definition of the clustering structure. Furthermore, to enhance comparability among the regions, we examine standardized time series (i.e., for each time series we subtract its mean and divide by its standard deviation) maintaining the shape of their patterns. This allows comparable ranges focusing on the structural similarities of the regions rather than just on the amplitude levels of their time series.

3. The statistical model

Following Section 2, we consider population level summaries of confirmed COVID-19 deaths collected over time over a fixed study area, $D$. The complete set of information is denoted as

---

1 The Outermost Regions, i.e. Martinique, Mayotte, Guadeloupe, French Guiana, Réunion, Saint-Martin for France, Canary Island and Cueva for Spain have not been included in the analysis.
Fig. 1. Weekly time series of deaths (per million inhabitants) by COVID for each Country.

Fig. 2. Time series correlations represented as a function of the spatial distance (in kilometres). The red line represents the empirical LOESS fit.

\[
\{ Y_t(s_i), s_i \in D \}, \quad i = 1, \ldots, n; \quad t = 1, \ldots, T, \quad \text{where } n \text{ is the number of areal units, } s_i \in D, \text{ and } T \text{ is the length of the time series. We also assume that at the first level of the hierarchy, the model} \]
for the (standardized) data is given by
\[ Y_t(s_i) = \eta_t(s_i) + \epsilon_t(s_i) \]
where \( \eta_t(s_i) \) is the true (latent) process of interest and \( \epsilon_t(s_i) \) is a spatially and temporally uncorrelated Gaussian error term, assumed to have zero mean and time varying variance, i.e. \( \epsilon_t(s_i) \sim N(0, \sigma^2_t) \). This error term corresponds to measurement error and/or representativeness error associated with the true process \( \eta_t(s_i) \). In particular, the fact that there may be deaths misattributed to COVID-19 and, conversely, others not attributed to COVID-19, some uncertainty is expected surrounding the true number of COVID-19 deaths. Because of this, it is desired to predict the smooth process \( \eta_t(s_i) \) rather than the observed noisy process \( Y_t(s_i) \).

With the aim of recognizing similarities and differences in the shape of the patterns of the weekly time series, a finite mixture is used as flexible model. The central assumption here is that the \( n \) time series arise from \( M \ll n \) hidden classes and, within each class, all time series can be characterized by a common data generating mechanism which is defined in terms of a probability distribution for the entire time series, depending on unknown class-specific parameters \( \phi_m \). Accordingly, denoting with \( \eta(s_i) = (\eta_1(s_i), \eta_2(s_i), \ldots, \eta_T(s_i))^\top \), the \((T \times 1)\) vector of time series of the latent process \( \eta \) at areal unit \( s_i \), the density of the mixture is written as
\[ f(\eta(s_i)) = \sum_{m=1}^{M} \pi_m f(\eta(s_i) | \phi_m, \psi), \quad \text{for} \quad i = 1, \ldots, n \]
where we adopt a standard truncated Dirichlet process model to define the prior over the mixing probabilities, \( \pi_m \), based on some (large) upper bound \( M \) see, Ishwaran and James (2002). A class assignment index \( Z_i \) taking a value in the set \{1, \ldots, M\} is then introduced for each time series, \( \eta(s_i) \), to indicate which class the time series belongs to:
\[ f(\eta(s_i) | Z_i, \pi, \phi_1, \ldots, \phi_M, \psi) = f(\eta(s_i) | \phi_{Z_i}, \psi) = \begin{cases} f(\eta(s_i) | \phi_1, \psi), & \text{if} \ Z_i = 1; \\ f(\eta(s_i) | \phi_2, \psi), & \text{if} \ Z_i = 2; \\ \vdots \\ f(\eta(s_i) | \phi_M, \psi), & \text{if} \ Z_i = M. \end{cases} \]
The class assignment indices \( Z = \{1, \ldots, M\} \) are random and independently distributed apriori, with prior class assignment distribution being the same for all sites, such that, \( P(Z_i = m | \pi) = \pi_m \), \( m = 1, \ldots, M \). The weights in \( \pi = (\pi_1, \ldots, \pi_M) \), with \( 0 \leq \pi_m \leq 1 \) and \( \sum \pi_m = 1 \), are assumed to be unknown model parameters estimated along with the data. An important aspect of the model is that we do not assume to know a priori which time series belongs to which group. For each time series, the group indicator variable \( Z_i \) is estimated along with the group-specific parameters \( \phi = (\phi_1, \ldots, \phi_M) \) from the data. Finally, \( \psi \) summarizes unknown parameters in the probabilistic distribution that are not related with the group indicators.

Assume that the \( \phi_m \) are independently drawn from an uncertain prior distribution \( G \), where the uncertainty about \( G \) is expressed through a Dirichlet process model. Then, the mixture formulation gives the following standard hierarchical model (Escobar and West, 1995)
\[ \eta(s_i) \mid \phi_m \sim N(\mu_m, \Sigma_m), \]
where the means \( \mu_m \) and variances \( \Sigma_m \) determine the parameters \( \phi_m = (\mu_m, \Sigma_m), \) and
\[ (\phi_m \mid G) \sim G, \quad (G \mid \alpha, G_0) \sim DP(\alpha, G_0). \]
The Dirichlet process is thus defined by a distribution function \( G_0 \) which is the prior expectation of \( G \), so that \( E(G(\phi)) = G_0(\phi) \), and \( \alpha \), a precision parameter determining the concentration of the prior for \( G \) about \( G_0 \). From the Pólya Urn Scheme it also follows that
\[ \phi_m \mid \phi_1, \phi_2, \ldots, \phi_{m-1} \sim \frac{\alpha}{m - 1 + \alpha} G_0(\phi_m) + \frac{1}{m - 1 + \alpha} \sum_{k=1}^{m-1} \delta_{\phi_k}(\phi_m) \] (3)
where $\delta_\phi(\phi_m)$ denotes a unit point mass distribution at $\phi = \phi_m$ — see, for example, Escobar and West (1995) and Manolopoulou et al. (2010).

The Dirichlet process prior allows to construct the weights for model classes following a stick-breaking representation (Sethuraman, 1994), where the $\pi_1, \pi_2, \ldots$ are obtained iteratively as follows

$$\pi_1 = V_1, \quad \pi_m = V_m \times \prod_{k=1}^{m-1} (1 - V_k), \quad \text{for } m > 1.$$  \hspace{1cm} (4)

The variables $V_1, V_2, \ldots$ represent a sequence of independent Beta$(1, \alpha)$ random variables for which we have $k < M$ and $V_M = 1$ to ensure that $\sum \pi_m = 1$. The coefficient $\alpha$ tunes the number of clusters in a direct manner, with larger values implying a larger number of clusters a priori. In particular, as $\alpha$ tends to zero, most of the samples $\phi_m$ share the same value, whereas when $\alpha$ tends to infinity, the $\phi_m$ are almost i.i.d. samples from $G_0$. Placing a prior on $\alpha$ thus allows us to draw inferences about the number of mixture components; a typical choice (Ishwaran and James, 2002) is to assume a Gamma prior, i.e. $\alpha \sim \Gamma(\xi_1, \xi_2)$, where $E(\alpha) = \xi_1/\xi_2$. Choosing a large value for $\xi_2$ is particularly relevant, because it encourages clustering; Escobar and West (1995) suggest $\alpha \sim \Gamma(2, 4)$, with $E(\alpha) = 0.5$.

The prior specification for each component $m$, completes with an appropriate choice for $G_0(\phi_m)$. In the following, since we aim to group areas together that exhibit similar temporal trends, we are only interested in modelling the mean with a mixture distribution. This is easily accommodated within the framework here by setting $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_M$ to equal some parameter $\Sigma$. This parametrisation has been studied widely in literature and references can be found, for example, in West (1992), West and Cao (1992) and Ishwaran and Zarepour (2000). To specify the form of the mean we consider the following state space model

$$\eta_i(s_i) = h(s_i) \mu_t + v_i(s_i), \quad i = 1, \ldots, n.$$ \hspace{1cm} (5)

$$\mu_{t+1} = \mu_t + u_t.$$ \hspace{1cm} (6)

where $\mu_t = (\mu_{t,1}, \ldots, \mu_{t,M})'$ is a $(M \times 1)$ state vector, $u_t$ is a zero mean temporally uncorrelated error with $(M \times M)$ covariance matrix $\Sigma_u = \text{diag}(\sigma_{u,1}^2, \ldots, \sigma_{u,M}^2)$ and $h(s_i)$ is an $(M \times 1)$ incidence vector where its $m$th element is equal to 1 if $Z_i = m$ and zero otherwise. Hence, conditionally on the class assignment variables $Z_i$, the structure of $h(s_i)$ implies that

$$\eta_i(s_i) =\mu_{t,m} + v_i(s_i), \quad m = 1, \ldots, M.$$ \hspace{1cm} (7)

where $\mu_{t,m}$ is the $m$th element of $\mu_t$ and the error terms $v_i(s_i)$ are uncorrelated in time but correlated in space. In particular, the $(n \times 1)$ vector $v_t = (v_1(s_1), \ldots, v_n(s_n))'$, denotes a zero mean spatial Conditional Autoregressive (CAR) process (Cressie, 1993) with precision matrix $Q_t = \tilde{D}_t (I - C_t)$, where $\tilde{D}_t = \text{diag}(\sigma_{v_{1,t}}^{-2}, \ldots, \sigma_{v_{n,t}}^{-2})$, $\sigma_{v_{i,t}}^{-2}$ are conditional variances and $C_t$ is a matrix which captures the spatial dependence at time $t$, with elements $c_{ij}$ satisfying the condition $c_{ij} \sigma_{v_{i,t}}^{-2} = c_{ij} \sigma_{v_{i,t}}^{-2} - \text{for a discussion on the symmetry conditions see, for example, Ippoliti et al. (2018).}$ A possible choice for $C_t$ requires first the specification of $W$, a symmetric regional neighbour-incidence matrix (with elements $w_{ij}$ equal to 1 if regions $i$ and $j$ are adjacent and 0 otherwise) and then the definition of $D_t$ as $\tilde{D}_t = \sigma_{v_{i,t}}^{-2} D$, where $D$ is a diagonal matrix whose $d_i$ element is the $i$th row sum of $W$, i.e. $d_i = w_{i,\cdot}$. This specification implies that $\sigma_{v_{i,t}}^{-2} = \sigma_{v_{i,t}}^{-2}/w_{i,\cdot}$ and that $C_t = \rho_t D^{-1} W$ where $\rho_t$ is a spatial dependence parameter. For simplicity, it is assumed here that each region has at least one neighbour and that, for isolated regions or islands, the set of neighbours is determined by the regions within a distance of 150 kilometres. The use of disconnected graphs (i.e., a graph containing a singleton node/region with no neighbours or a graph split in different sub-graphs) for the CAR specification is not straightforward (Freni-Sterrantino et al., 2018) and it is not considered here. Also, we note that, although we consider the time series to be spatially correlated, using our approach it is possible to cluster into a single group multiple time series that have similar temporal patterns, but that are located far away from each other. This is in contrast with other approaches that, by
imposing Markov random field (MRF) priors on the mixture parameters (Blekas et al., 2007) or on the cluster assignment variables (Jiang and Serban, 2012), force the spatial clustering of the series.

The state space model described above represents a local level model where the mean of \( \eta_t(s_i) \) is group specific and varies over time. If \( \mu_m = (\mu_{1.m}, \ldots, \mu_{T,m})' \) is considered as a vector of parameters then, Eq. (6) can be interpreted as defining a hierarchical prior for \( \mu_m \), \( m = 1, \ldots, M \). In particular, if

\[
R = \begin{bmatrix}
1 & 0 & 0 & 0 & \cdots & \cdots & 0 \\
-1 & 1 & 0 & 0 & \cdots & \cdots & 0 \\
0 & -1 & 1 & 0 & \cdots & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & 0 & 0 & -1 & 1
\end{bmatrix}
\]

is a square first difference matrix, and \( u_m = (u_{1,m}, \ldots, u_{T,m})' \), we can write

\[
R \mu_m = u_m, \quad m = 1, \ldots, M
\]
such that \( G_0(\mu_m) = N(0, V_0) \) with \( V_0 = \sigma^2 u_m R^{-1} R^{-1} \) and \( \mu_{0,m} \sim N(\mu_m, \sigma^2 m) \) for the initial conditions. In words, the stochastic trend term has a variance which is increasing with time (and thus can wander over an increasing wide range), but \( \mu_{t,m} \) changes only gradually over time being consistent with the assumption that \( \mu_{t,m} \) and \( \mu_{t-1,m} \) tend to be close to one another. It is thus worth noting that the state equation can be interpreted as already providing us with a prior for \( \mu_{t,m} \) implying that \( \mu_{t+1,m} \mid \mu_{t,m} \sim N(\mu_{t,m}, \sigma^2 m) \) with \( f(\mu_m | \sigma^2 m) = \prod_{t=1}^T f(\mu_{t+1,m} | \mu_{t,m}, \sigma^2 m) \).

This is an example of a hierarchical prior, since the prior for \( \mu_m \) depends on \( \sigma^2 m \) which, in turn, requires its own prior.

Finally, for estimation purposes, it is also worth considering that, by isolating each variable \( \nu_i(s_i) \) in turn, and considering the n individual conditional distributions \( \nu_i(s_i) | \nu_i(-s_i), i = 1, \ldots, n \), at each site, where \( \nu_i(-s_i) \) denotes all other values than \( \nu_i(s_i) \), the general CAR specification described above for \( \nu_i \) implies that

\[
\eta_i(s_i) | \eta_i(s_j), i \neq j, \sim N \left( \mu_i + \sum_{j=1}^n c_{ij} (\eta_i(s_j) - h(s_j)' \mu_j), \sigma^2_{\eta, i} / d_i \right)
\]

for spatial coefficients \( c_{ij} = \rho_i w_{ij} / w_{i,+} \) and \( d_i = w_{i,+} \). A special case is the Intrinsic Conditional Autoregressive (ICAR) model which uses \( d_i = w_{i,+} \) and \( \rho_i = 1 \) (Cressie, 1993). Incidentally we note that Eq. (8), by Brook’s Lemma (Brook, 1964), gives rise to a pseudo-likelihood approximation of the joint distribution of \( \eta_i(s_i) \), which is efficient for simple Gaussian fields see, Besag (1975, 1977).

3.1. Posterior inference

Assume that \( \theta \) is a vector collecting all model parameters. Estimation of \( \theta \) is based on the posterior distribution \( f(\theta | Y) \) of \( \theta \) given the data \( Y \). The posterior distribution \( f(\theta | Y) \) is defined through Bayes’ theorem as the product of the observed-data likelihood function \( f(Y | \theta) \) and the prior \( f(\theta) \). Under this framework, we use Markov chain Monte Carlo (MCMC) methods and posterior inference is obtained by implementing a Gibbs sampler. Computationally, we only need to calculate the full conditionals of each parameter given all other parameters, which is usually not hard. In the following, we provide details for the relevant conditional distributions. First, it is important to recall (see, for example, Sahu and Mardia, 2005) that

\[
Y_t(s_i) | \eta(s_i) \sim N(\eta(s_i), \sigma^2_{\eta, t})
\]

and that

\[
\eta = (\eta(s_1), \ldots, \eta(s_n))' \sim N(H\mu, Q^{-1})
\]
where $H$ is a $(n \times M)$ matrix with $j$th row equal to $h(s_i)'$. It thus follows that the full conditional distribution of $\eta$ is the multivariate Normal distribution $N(\Sigma \eta, \Sigma_\eta)$ where

$$\Sigma_\eta^{-1} = \sigma^{-1}_e \mathbf{I} + Q, \quad \xi = \sigma^{-1}_e y_t + Q \mu_t.$$

Then, using Eq. (8), we can thus write

$$f(\eta | \theta) \propto \prod_{i=1}^n \left( \prod_{m=1}^M \prod_{t=1}^T f(\eta_i(s) | Z_i, \eta_i(s), i \neq j) \right).$$

Conditional on $\eta_i(s)$, the state equation $\mu_s$ of model (5)–(6) can be sampled by means of the well-known forward filtering–backward sampling algorithm of Carter and Kohn (1994) or Durbin and Koopman (2002).

A further important part of the estimation refers to the sampling of $\alpha$ whose posterior only depends on the data through the variables $V_k$. Accordingly, $\alpha$ has the usual posterior distribution (Ishwaran and James, 2002)

$$\alpha \sim \text{Ga}(\varsigma_1 + M - 1, \varsigma_2 - \sum_{k=1}^{M-1} \log(1 - V_k)).$$

The posterior for the class assignment indices $Z$ is multinomial with probabilities

$$f(Z_t = m \mid \eta(s), \phi) \propto \pi_m N(\eta(s) \mid \mu_m, \Sigma).$$

Furthermore, since the precision matrix of the CAR process can be time dependent, we specify an additional state-space model. In particular, for the conditional variance $\sigma_{\epsilon,t}^2$, we introduce the following representation

$$\begin{align*}
\eta_{\epsilon,t}^m &= \mathbf{1} g_{v,t} + e_{v,t} \\
g_{v,t+1} &= g_{v,t} + v_{v,t}
\end{align*}$$

where the $j$th element of $\eta_{\epsilon,t}^m$ is $\eta_{\epsilon,t}^m(s_i) = \log[\eta_t(s_i)^2 + c]$, $\eta_t(s_i)$ denotes the $j$th element of $\eta_{\epsilon,t}^m = \mathbf{D}^{1/2}(1 - C_t)^{1/2}(\eta - \mu)$, $g_{v,t} = 2 \log(\sigma_{v,t})$, $e_{v,t}$ is a $(n \times 1)$ vector of independent errors, $v_{v,t}$ is $N(0, \sigma_v^2)$ and is independent of $v_t$ and $u_t$. The constant $c$ is an offset that we set at 0.001.

The system in this form has a linear, but non-Gaussian state space form, because the innovations in the measurement equations are distributed as a log $\chi^2(1)$. As described in Kim et al. (1998), in order to further transform the system in a Gaussian one, a mixture of seven Normals is used as an approximation of the log $\chi^2$. Let $S_{ij} \in \{1, 2, \ldots, 7\}$ denote the indicator variable of the $i$th Normal from which $e_{v,t}(s_i)$ is drawn from, and let $S_i = (S_{i1}, \ldots, S_{i7})'$ and $S = (S_1, \ldots, S_n)'$ be component indicators for all the elements of $e_{v,t}$. Conditional on $S$ (and the other parameters), Eqs. (9) and (10) define a Gaussian state-space model and, hence, the algorithm by Durbin and Koopman (2002) can be used to draw $g_{v,t}$. In particular, let $q_j$, be the component probability of the $j$th Normal of the mixture with mean, $m_j$, and variance $V_j^2$, for $j = 1, \ldots, 7$ - see, Kim et al., 1998. Then

$$\Pr(S_{ij} = j \mid \eta, g_{v,t}) \propto q_j N(\eta_{\epsilon,t}^m(s_i) \mid g_{v,t} + m_j - 1.2704, V_j^2),$$

for $j = 1, \ldots, 7$, $i = 1, \ldots, n$, and $t = 1, \ldots, T$.

Regarding the evolution of the spatial correlation parameter, $\rho$, we specify the following state equation

$$\rho_{t+1} = \rho_t + \psi_t$$

where $\psi_t \sim N(0, \sigma_\psi^2)$, is independent of $u_t$, $v_t$ and $v_t$. Eq. (12) is combined with the measurement equation defined in (8), so that the algorithm of Durbin and Koopman (2002) can be used to draw the states.
Finally, posterior simulation of $\sigma_{\epsilon,t}^2$ follows closely that of $\sigma_{\nu,t}^2$. Let $Y^*_t(s_i) = Y_t(s_i) - \eta_t(s_i)$, $Y^{**}_t(s_i) = \log \left( (Y^*_t(s_i))^2 + c \right)$ and $\tau_{\epsilon,t} = 2\log(\sigma_{\epsilon,t})$. Then, the state space model with measurement and state equations is specified by

$$
Y^*_t(s_i) = \eta_t(s_i) + \epsilon_{\epsilon,t}(s_i) \\
\tau_{\epsilon,t+1} = \tau_{\epsilon,t} + \nu_{\epsilon,t}
$$

(13)

As for $g_{\nu,t}$, the procedures described by Kim et al. (1998) and Durbin and Koopman (2002) are used to draw $\tau_{\epsilon,t}$.

4. Clustering time series of confirmed COVID-19 deaths

Here, we complete the statistical analysis of the data introduced in Section 2 by applying the methodology discussed in Section 3 to group the 145 regions into distinct communities based on their temporal patterns.

In particular, we fit several competing models by considering $\sigma_{\epsilon,t}^2$, $\sigma_{\nu,t}^2$ and $\rho_t$ either as time-varying parameters or as constant through time. Overall, we evaluate $2^{36}$ different model parametrizations.

For all the fitted models, the MCMC algorithm is run for 60,000 iterations and posterior inference is based on the last 10,000 draws using every 5th member of the chain to avoid autocorrelation within the sampled values. Convergence of the chains of the model is monitored visually through trace plots as well as using the R-statistic of Gelman (1996).

When convergence is attained, to avoid label switching (Stephens, 2000), we follow Nieto-Barajas and Contreras-Cristan (2014) and Dahal (2006) and choose the representative clustering structure which minimizes the deviation from a pairwise clustering matrix which, taking into account all MCMC iterations, provides an estimate of the probability that the two time series belong to the same cluster. Then, in order to compare among cluster structures resulting from different model specifications, we use the following heterogeneity measure

$$
HM(G_1, \ldots, G_M) = \sum_{m=1}^{M} \frac{2}{n_m - 1} \sum_{(i<j) \in G_m} \sum_{t=1}^{T} (\eta_t(s_i) - \eta_t(s_j))^2
$$

where $\{G_1, \ldots, G_M\}$ denotes the set of indices for a structure of $M$ clusters with size-groups $n_1, \ldots, n_M$. This measure of cluster validity assessment is expected to be as small as possible. Additionally, for model comparison, we also evaluate a measure of goodness of fit. In particular, we consider the logarithm of the pseudo marginal likelihood (LPML)

$$
LPML = \sum_{i=1}^{n} \log(\widehat{CP}_i)
$$

which requires the computation of the Conditional Predictive Ordinate (CPO) statistics (Geisser and Eddy, 1979; Mukhopadhyay and Gelfand, 1997)

$$
\widehat{CP}_i = \left( \frac{1}{L} \sum_{l=1}^{L} \frac{1}{f(y(s_i) | \eta^{(l)}(s_i), \Sigma_{\epsilon}^{(l)})} \right)^{-1},
$$

where $\Sigma_{\epsilon}^{(l)} = \text{diag}(\sigma_{\epsilon,1}^{(l)}, \sigma_{\epsilon,2}^{(l)}, \ldots, \sigma_{\epsilon,T}^{(l)})$. Given posterior samples of $\eta^{(l)}(s_i)$ and $\Sigma_{\epsilon}^{(l)}$, $l = 1, \ldots, L$, the LPML is a predictive measure of model performance with larger values indicating a better model fit.

By considering the model selection criteria described above, Table 2 suggests that models $M_5$ and $M_7$ do not provide the best fit of the data, but produce the most homogeneous clusters. On the other hand, a better fit can be obtained by using models $M_2$ and $M_6$ which also give the lowest number of clusters. Accordingly, a good balance between heterogeneity and goodness of fit seems to be offered by model $M_2$, which is characterized by the highest LPML and a good value of the
Table 2
LPML and HM statistics for different model parametrizations. The last column reports the number of estimated groups.
| Model                  | LPML    | HM      | M    |
|------------------------|---------|---------|------|
| $M_1(\sigma^2_{\nu,t}, \rho_t, \sigma^2_{\epsilon,t})$ | 1864.60 | 2181.20 | 19   |
| $M_2(\sigma^2_{\nu,t}, \rho_t, \sigma^2_{\epsilon})$  | 9085.03 | 2289.01 | 12   |
| $M_3(\sigma^2_{\nu,t}, \rho, \sigma^2_{\epsilon,t})$  | 1329.00 | 2193.60 | 16   |
| $M_4(\sigma^2_{\nu,t}, \rho_t, \sigma^2_{\epsilon,t})$ | −115.92 | 2363.10 | 23   |
| $M_5(\sigma^2_{\nu}, \rho, \sigma^2_{\epsilon})$     | 544.45  | 1790.10 | 19   |
| $M_6(\sigma^2_{\nu,t}, \rho, \sigma^2_{\epsilon})$     | 8557.31 | 2588.42 | 12   |
| $M_7(\sigma^2_{\nu,t}, \rho, \sigma^2_{\epsilon})$     | 862.44  | 1809.09 | 19   |
| $M_8(\sigma^2_{\nu}, \rho_t, \sigma^2_{\epsilon})$     | −124.84 | 2922.09 | 24   |

heterogeneity measure HM. For this model, the regions are grouped in $M = 12$ clusters, both the conditional variance and the spatial parameter of the CAR are time varying parameters, and the variance of the measurement error is constant. The dynamics of the time-varying parameters, with 95% credible intervals, are shown in Figs. 3 and 4. The pattern of the conditional variance $\sigma^2_{\nu,t}$, coherently with the data, highlights three peaks overall the study period. There are also nearly zero values between the end of May and the end of September, denoting small individual (regional) effects in this period. There is the hint that this might be due to the implementation of national and subnational strict measures during the first phase and favourable climate conditions, though this is difficult to prove. The dynamic of $\rho_t$ shows that, on average, the spatial dependence parameter varies between 0.85 and 0.92 reaching its minimum value in the second week of May. Although the range of values appears limited, its plot suggests that the spatial correlation increases with the resumption of social and economic activities reaching two peaks of different amplitude around the second week of November and the third of January. As it can be noticed from Figs. 1 and 3, these peaks are aligned both with the last two of $\sigma^2_{\nu,t}$ and with those of several time series.

A spatial map of the regions grouped by using model $M_2$ is shown in Fig. 5, where each cluster is represented by a specific colour. Fig. 6 also offers a graphic visualization of the clusters where the posterior mean group functions are overlaid with the time series of each region. As expected, the curves in the same cluster appear similar in shape, whereas present different patterns or trends across the clusters. The regions that are geographically close are likely to be in the same cluster and this applies in particular to several regions in Italy, France, and Germany. However, as expected from the exploratory data analysis (Section 2), the clusters are not always dominated by the spatial proximity of the regions. An example is provided by the second cluster, which mainly includes regions from the South of Germany and the West-Southern part of Italy.

The temporal pattern of the mean groups represented in Fig. 6 show that, after peaking in early April, the pandemic appeared to be well controlled until the beginning of August. As a result of the introduction of several containment measures and with increasing temperatures, all the regions have experienced relatively low counts of deaths, with the mean curves resulting flat for most part of the summer period. Focusing on the first phase of the pandemic, the subplots also suggest that the rate of change of the early counts of deaths in a few clusters were higher than all of the other clusters.

This is confirmed in Fig. 7 by the dynamics of the first derivative of the group mean functions, which help to understand their rate of change as well as the persistence of the phenomenon. As it can be noticed, the regions in clusters 5 and 7 show the highest rate of change (i.e., a higher velocity) during March and April. From the exploratory analysis of the data, it also results that cluster 5 includes some of the regions with the highest mortality rate as, for example, the Lombardy region (Italy) with the provinces of Bergamo and Brescia, which have been at the centre of Italy’s coronavirus outbreak started in February 2020.

Cluster 7 includes a few regions from Spain and one from England. This cluster appears highly heterogeneous in the period October 2020–February 2021. Furthermore, its mean function
Fig. 3. The temporal dynamics of the posterior mean conditional variance $\sigma^2_{\nu,t}$. 95% credible intervals are represented by shaded areas.

Highlights a slightly delayed onset of the pandemic compared to cluster 5, which peaked earlier than cluster 7 in the first phase. Cluster 8 is similar in shape to cluster 7; however, the latter appears more heterogeneous, with a few time series showing different timing effects, especially in January and February 2021. In contrast with these patterns, cluster 1, mainly composed by regions from the Central and North-Eastern part of Germany and most of the Swiss Confederation, shows the smallest rate of change and slow dynamics of the phenomenon in the first phase. All other clusters do not seem to differ substantially in terms of variations during the first COVID-19 wave.

Further differences among clusters can be appreciated by studying the dynamics of the series in the period August 2020–February 2021. In general, all the series show that the number of deaths has surged after growth slowed in summer. Almost all the series, in fact, have been trending upward during the autumn. Exceptions to these trends are represented by the pattern in cluster 5 where the regions do not seem to show any important tendency of rebound after the first phase.

Understanding the potential reasons behind this trend may be crucial to identify key factors in the definition of effective preventive policies. Changes in the healthcare system, in the exposure of the vulnerable population as well as restrictive measurements, such as local lockdowns, might all have played an important role in flattening the mortality curves. For example, many regions have seen a marked improvement in the healthcare system, both in terms of intensive care units (ICU) capacity as well as access to novel effective treatments, that have guaranteed a timely and effective assistance to the most severe cases with a consequent reduction of the mortality rate. A further explanation for the reduced mortality in cluster 5, during the second wave, could be found in the harvesting effect. According to this theory, the most vulnerable population, made of elderly and those with health conditions, might have died during the first COVID-19 wave, especially in regions with a high infection rate. By contrast, the regions that were spared from the first phase have then shown an increase of the mortality in the second phase. Clusters 1, 2, 3, 6, 9 and 10 show similar bimodal mean shapes in the last five months, with some differences appearing in the rate...
of change and in the persistence of the phenomenon. In particular, the number of deaths in clusters 1 and 2 rises steeply than in other clusters from late September, while in clusters 9 and 10 the resurgence appears slower and delayed to reach a plateau until the beginning of February.

These temporal patterns also differ from the dynamics of clusters 4 and 11 for which only one peak is observed near the end of the third week of January, denoting a significant reappearance of the phenomenon in UK. For the regions in cluster 4, this resurgence could be linked to the English Coronavirus variant, which has rapidly outcompeted pre-existing variants in the Southeast region of the country (Davies et al., 2021). Finally, the dynamics of cluster 12, which includes most of the regions with the lowest population density, highlight a significant rate of change (see, Fig. 7) of mortality between October and November 2020, with regions reaching their peak around the second week of November. In contrast with many other clusters, for this group of regions, the mortality has steadily decreased from the third week of November to attain low-stable levels in 2021.

5. Conclusions

In this paper, we have analysed the regional patterns of spatio-temporal data on confirmed deaths by COVID-19 in NUTS-2 regions from six European countries between March 2020 and February 2021. To uncover homogeneous patterns of COVID-19 progression, we have implemented a Bayesian nonparametric approach that models (standardized) time series patterns as a mixture of Gaussian processes, specifying a Dirichlet process prior over mixture components. Although the statistical analysis was carried out under Gaussianity assumptions, our model could be easily cast within the more general framework of generalized linear models (GLM), where the data are assumed to be from an exponential family with distribution $F$. For example, considering the death counts, one obvious choice would be to assume a Poisson regression model with a log link function. In
order to keep computations for the posterior distributions simple, one may then apply the data augmentation technique proposed by Tanner and Wong (1987).

The proposed methodology has allowed grouping the 145 regions into 12 clusters providing evidence that the effects of the pandemic varied a lot both temporally and across countries and regions. Some of the areas that have been hit the hardest in the first phase (for example, the Northern region of Lombardy in Italy) have seen a steep decline in the number of deaths soon after implementing control strategies, with contagion and deaths far less widespread in autumn and winter 2020. In contrast, other areas, which have not been affected in the first wave, have seen a sharp rise in deaths during the second and third waves. An example is given by the southern regions of Italy, which became a new hotspot.

One advantage of our analysis concerns the possibility of defining community mitigation strategies not only at countries levels, but also at the level of clusters, where their members (administrative areas) appear homogeneous in terms of temporal patterns. This clustering effect calls for a strong inter-governmental coordination and for a territorial approach in order to define joint
solutions and enhance acceptance of measures at all levels. The aim is thus to facilitate a regional cooperation to support recovery strategies by ensuring coherent mitigation guidelines, pooling resources, and strengthening investment opportunities. In particular, promoting sub-national governments could help to evaluate the degree of asymmetry and differentiate aid schemes to align with the differentiated impact of COVID-19.

In policy-making, trying to understand the underlying causes of this heterogeneity represents a crucial step to evaluate the efficacy of the different mitigation strategies adopted by central and local governments. However, assessing the effect of the individual intervention policies is a very challenging task due to the inherent multidimensional nature of the phenomenon under investigation. From the statistical analysis, in fact, there is a clear hint that differences in population density, age structure, urbanization and socio-economic factors contribute to differentiate the impact of COVID-19. Indeed, even if coronavirus infects exposed subjects indiscriminately, the severity of disease and social and economic effects are not being felt equally throughout the European territory, partly due to existing differences in health between and within countries, mostly related to socio-economic inequities. To fully understand the impact of containment policies on the pandemic dynamic, it would be necessary to explicitly relate these factors to the death counts within the proposed methodological framework. Albeit the proposed model could easily accommodate the introduction of these factors, in practice, we have found it infeasible as most of the covariates arising from multiple sources are often spatially misaligned or aggregated over different geographical units and, more importantly, they are only spatially referenced, making it impossible to evaluate their temporal effect on disease spreading. Accordingly, here, we have not attempted to establish the
causality of the observed differences between clusters, but we have simply provided a tool to evaluate the evolution of the pandemic across European regions. The inclusion of these covariates and the other risks factors will be the objects of future studies.

Acknowledgments

Andrea Bucci acknowledges that the research is part of 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. D24119001680001 – Id. proposta: AIM1894803 Linea 1: attività 3 Area di specializzazione SNSI: Smart, Secure and Inclusive Communities).

References

Bartolucci, F., Farcomeni, A., 2022. A spatio-temporal model based on discrete latent variables for the analysis of COVID-19 incidence. Spatial Stat. 49, 100504. http://dx.doi.org/10.1016/j.spasta.2021.100504.
Besag, J., 1975. Statistical analysis of non-lattice data. The Statistician 24, 179–195.
Besag, J., 1977. Efficiency of pseudolikelihood estimation for simple Gaussian fields. Biometrika 64, 616–618.
Blekas, K., Nikou, C., Galatsanos, N., Tsekos, N.V., 2007. Curve clustering with spatial constraints for analysis of spatiotemporal data. In: Tools with Artificial Intelligence. Vol 1, ICTAI 2007. 19th IEEE International Conference, pp. 529–535.
Brook, D., 1964. On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. Biometrika 51, 481–483.
Carter, C., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.
