Journal of the Royal Statistical Society Series A: Statistics in Society, 2023, **186**, 401–421
https://doi.org/10.1093/jrsssa/qnac006
Advance access publication 27 January 2023
Original Article

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

Network self-exciting point processes to measure health impacts of COVID-19

Paolo Giudici, Paolo Pagnottoni and Alessandro Spelta

Department of Economics and Management, University of Pavia, Pavia, Italy

Address for correspondence: Paolo Pagnottoni, Department of Economics and Management, University of Pavia, Pavia 27100, Italy. Email: paolo.pagnottoni@unipv.it

Abstract

The assessment of the health impacts of the COVID-19 pandemic requires the consideration of mobility networks. To this aim, we propose to augment spatio-temporal point process models with mobility network covariates. We show how the resulting model can be employed to predict contagion patterns and to help in important decisions such as the distribution of vaccines. The application of the proposed methodology to 27 European countries shows that human mobility, along with vaccine doses and government policies, are significant predictors of the number of new COVID-19 reported infections and are therefore key variables for decision-making.

Keywords: COVID-19, epidemic modelling, Hawkes process, interconnectedness, mobility networks

1 Introduction

Interconnectedness is an increasingly relevant concept in many research and practical domains. It features prominently in key issues, such as disease transmission in contact networks, investigation of human behaviour, migration phenomena, and many others. Many types of interconnectedness have been widely studied by means of statistical networks: see, e.g., Khanin and Wit (2006), Tucker et al. (2005), Abegaz and Wit (2013), Wang et al. (2009), Vinciotti et al. (2016), and Giudici and Spelta (2016); along with their implications on society, see, e.g., Ghani and Garnett (1998), Di Zio et al. (2004), Rocha et al. (2017), Barbillon et al. (2017), and Wright et al. (2021). However, there is still a strong need for network models that can leverage interconnectedness for predictive purposes.

The recent coronavirus disease outbreak has ignited further discussions, as human mobility networks play a key role in its diffusion. Differently from the previous acute respiratory syndrome (SARS), which gradually emerged in 2003, the novel coronavirus (SARS-CoV-2) did register a dramatic rate of growth in positive individuals, and early evidence of a high transmission rate (Guan et al., 2020; Wang et al., 2020). This pushed governments worldwide to adopt a wide set of policy measures, such as lockdowns and mobility restrictions, to limit the spread of contagion. Such interventions have been found in many cases to significantly reduce contagion (Hsiang et al., 2020; Xiong et al., 2020), at the expense of strong consequences on mobility patterns (Schlosser et al., 2020; Zhang et al., 2020).

A recent stream of literature exploits mobility networks to examine, monitor, and characterize the diffusion of COVID-19 (Aleta et al., 2020; Chinazzi et al., 2020; Davies et al., 2020; Della Rossa et al., 2020; Flaxman et al., 2020; Kraemer et al., 2020; Maier & Brockmann, 2020; Scala et al., 2020). Within this stream, some studies have employed mobility networks to quantify, by means of a combination of epidemiological and economic models, the economic effects of the

Received: July 20, 2021. **Revised:** October 27, 2022. **Accepted:** November 23, 2022

pandemic (Spelta et al., 2020; Spelta & Pagnottoni, 2021). In these works, the COVID-19 epidemic is viewed through the lens of complex systems, with the aim of understanding the impact of lockdown measures on the virus transmission.

We follow the latter developments, and we leverage mobility networks to understand the impact of policy restrictions on the COVID-19 counts. In particular, we employ self-exciting point processes, a class of birth processes which have been used to model the spread and diffusion of phenomena of various nature: from earthquakes, to email networks and crime forecasting.

Self-exciting point processes have been also designed to model the diffusion of epidemics of different types: Becker (1977) used self-exciting point processes to model the dynamics of the Brazilian smallpox disease; Farrington et al. (2003) employed the methodology to study the effect of vaccinations on the spread of measles in the United States; in Balderama et al. (2012) the authors proposed an extension of the Hawkes process to characterize the diffusion of red banana trees invasive species in a Costa Rican rainforest; Meyer et al. (2012) implemented a Hawkes process to estimate spread rates and describe properties of the human incidence of invasive meningococcal disease. More recently, some studies have dealt with modelling the COVID-19 epidemic with the same methodology, see Zhu et al. (2020), Rambhatla et al. (2020), and Chiang et al. (2020). Alternative time-series approaches to model the COVID-19 epidemic spread can be found, for instance, in Harvey and Kattuman (2020), Giudici et al. (2023) and Celani and Giudici (2022).

In more detail, Chiang et al. (2020) developed models of COVID-19 transmission using Hawkes processes with spatio-temporal covariates at the United States county level, through a regression on Google mobility indices and demographic covariates. They showed how Hawkes process-based models outperform several competing alternative models currently used to monitor the contagion diffusion, such as an ensemble approach and the susceptible-infected-recovered model. Rambhatla et al. (2020) worked on the development of spatio-temporal risk scores to indicate the risk propensity of the disease in different places. They relied on a Hawkes process-based model to assign relatively fine-grain spatial and temporal risk scores by means of high-resolution mobility data based on cell-phone location signals across the United States. Their results showed that spatio-temporal risk scores based on high-resolution mobility data can provide useful insights and ease safe re-openings. Similarly, Zhu et al. (2020) presented a spatio-temporal model with mobility and social demographic covariates to estimate COVID-19 confirmed cases and deaths one-week ahead, at the county level, in the United States. Their retrospective out-of-sample county-level predictions were able to forecast the subsequently observed COVID-19 activity accurately, with a clear identification and quantification of the most important factors which determine the COVID-19 spread.

Against this background, we propose a spatio-temporal network model which takes root from self-exciting point processes and enhances them with spatio-temporal, mobility, and socio-demographic covariates, whose parameters are estimated via maximum likelihood through an expectation maximization (EM) algorithm. The model relies on a large set of relevant predictors, pre-processed through dimensionality reduction techniques such as non-negative matrix factorization (NNMF) and principal component analysis (PCA). We apply our methodology to perform the prediction of the number of new COVID-19 reported infections of 27 European countries over the period ranging from 1 September 2020 to 30 May 2021. After evaluating the model performance, we conduct a policy scenario analysis under the hypothesis of an increased availability of vaccine vials put at disposal of European countries. We statistically assign additional vaccine doses based on a set of re-distribution schemes which progressively allocate a larger (smaller) number of vials depending on the vaccine-to-infected ratio of each country.

We build our model on an heterogeneous, spatial, temporal, and network set of covariates. In particular, we rely on three main data sources of spatio-temporal human mobility patterns: (a) Facebook data, a collection of phone-tracking-based datasets measuring the amount of individuals moving between each country's administrative regions; (b) Google data, a phone-tracking based source of mobility data, which registers daily changes in mobility across location categories, namely retail and recreation, grocery stores and pharmacies, workplaces and transit stations, tracking the trends of people's change in movement throughout the pandemic period; (c) air transport data, and particularly international European air transport data on the number of passengers, collected at airport level, to proxy the number of individuals flowing across countries. We further

consider a set of spatially distributed epidemic covariates: (a) disease risk factors, which comprise heart disease, diabetes, and smoke addiction; (b) social and demographic covariates, which include population density, median age, and life expectancy.

We differ from the literature employing self-exciting point processes for COVID-19 epidemic modelling in four main aspects. First, we exploit techniques borrowed from network theory to build a robust statistical model which is able to take into account for the direct impact of the network of human mobility patterns on the spread of COVID-19. Second, our spatio-temporal self-exciting point process is based on a large and heterogeneous set of predictors of the COVID-19 contagion dynamics, such as mobility data, spatial covariates, and government policy actions, and additionally include a key covariate of interest to policy making—namely, the daily number of vaccines inoculated—which can be tuned to conduct scenario analysis. Third, we build upon the predictions of the illustrated method a number of scenario analyses corresponding to different counterfactual health impacts of several coordinated policy interventions in terms of supply of vaccine vials. Finally, differently from the extant literature which is mainly focused on modelling the evolution of the epidemic in the United States, we focus on 27 European countries.

To the best of our knowledge, there is no study yet on spatio-temporal COVID-19 contagion modelling through Hawkes processes enhanced by mobility networks considering European countries. In addition, our model provides a novel continuous monitoring tool able to dynamically assess the health impact of setups of different policy options, by means of statistically robust, tailored scenario analyses and re-distributive schemes based upon any potentially effective epidemic-tuning covariates of interest.

Our empirical findings show that the proposed model achieves satisfactory performances in terms of predictive accuracy. We demonstrate that a large set of epidemic covariates, such as human mobility patterns, mobility trends, vaccine doses inoculated, government policy stringency, disease risk factors, and demographic factors, are significant predictors of the number of new COVID-19 reported cases in Europe. Our policy scenario analysis shows that an increased number of vaccines put at disposal of European countries would, in any case, cause a slowdown in infection rates, although the effects of different hypothesized re-distribution schemes are crucial to the development of the epidemic. According to our scenario analysis, we discover that the best re-distribution option consists of distributing most resources to countries with lower vaccine-to-infected ratios, though, interestingly, the effects of re-distributing schemes are non-linear.

The remainder of this paper is structured as follows. Section 2 describes the data employed in this study. Section 3 presents the proposed statistical methodology. Section 4 illustrates our empirical results. Section 5 concludes.

2 Data

We consider a large database, standardized for European countries, coming from different sources: (i) COVID-19 diffusion and vaccine data; (ii) Google mobility data; (iii) Facebook mobility data; (iv) airline transport data; (v) spatial country level data; and (vi) government policy data. The considered time period ranges from 1 September 2020 to 30 May 2021, a time span which allows to study multiple waves of COVID-19 infections. We present below a short description of each data source.

2.1 COVID-19 diffusion data

The COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University releases a daily report about COVID-19 cases registered at country level over time. From the dataset, we extract the total number of new confirmed COVID-19 cases in each country. This will be the response variable in our analysis.

2.2 COVID-19 vaccine data

A first relevant explanatory variable of COVID-19 diffusion is the level of vaccination. Data on the total number of vaccines inoculated for each country at the daily level are collected from the COVID-19 Data Repository maintained by the Center for Systems Science and Engineering at Johns Hopkins University.

2.3 Google mobility data

A particular type of mobility data has been made available by Google. This consists of a phone-tracking mobility data which aggregates anonymized information from users who have turned on their location history setting and measures the daily changes in mobility with respect to a baseline value, for each country, across different location categories, including retail and recreation, grocery stores and pharmacies, workplaces, and transit stations. While Facebook and airline transport data measure the flows of people within and between countries, Google data measure the variation in movement trends across different locations. The baseline value represents the median values on a 5-week period from 3 January 2020 to 6 February 2020, a 'normal' business period.

2.4 Facebook mobility data

Facebook has made available the 'Coronavirus Disease Prevention Maps', as a part of its 'Data For Good' programme, a collection of dynamic spatio-temporal datasets illustrating, within each country, population commuting patterns over the COVID-19 period. The maps use anonymized and aggregated data on mobile-phone-based geo-localized movements of people having their geo-positioning option enabled within time intervals of 8 hr, which we aggregate to daily frequencies. We consider mobility flows both between and within the different administrative regions of a country. It is worth mentioning that these data are available at a higher frequency, paving the way for future analyses via multivariate time-series models—see, e.g., Giudici and Pagnottoni (2019)—which could enable to assess impacts in a more timely manner.

2.5 Airline mobility data

To capture flows of individuals between different countries, thereby complementing Facebook data, we consider data from the Air Transport Statistics database provided by EUROSTAT, on a yearly basis. These data are collected at the airport level and describe the number of passengers flying between different airports. We focus on intra-European transportation between countries during 2019 and 2020. The data consist of origin-destination matrices whose element (i, j) represents the number of passengers moving from i to j, where a passenger is defined as any person, excluding on-duty members of the flight and cabin crews, who makes a journey by air between two airports.

2.6 Country data

We also consider country specific demographic covariates which are good candidate predictors of the spread of the COVID-19 disease. We consider the population density, median age, and life expectancy. Moreover, we also consider health risk factors which can worsen the COVID-19 symptoms such as the heart disease propensity, the percentage of diabetics, and the percentage of smokers (male and female). These variables are collected from Our World in Data and are updated yearly.

2.7 Government policy data

The Oxford COVID-19 Government Response Tracker collects publicly available information on a set of indicators of country specific government responses—see Hale and Webster (2020) and Petherick et al. (2020). Eight of the policy indicators (C1–C8) record information on a number of preventive containment and closure policies, such as school closures and mobility restrictions. Four of the indicators (E1–E4) record economic policies, such as income support to citizens or provision of foreign aid. Seven of the indicators (H1–H7) record health system policies such as the COVID-19 testing regime, emergency investments into healthcare and, most recently, vaccination policies.

2.8 Data summary and pre-processing

We first consider the COVID-19 diffusion data, the response variable, whose behaviour in time we aim to explain and predict. In Figure 1, we show the time-series of the total number of COVID-19 reported cases over the period 1 September 2020 to 30 May 2021. From Figure 1, the presence of

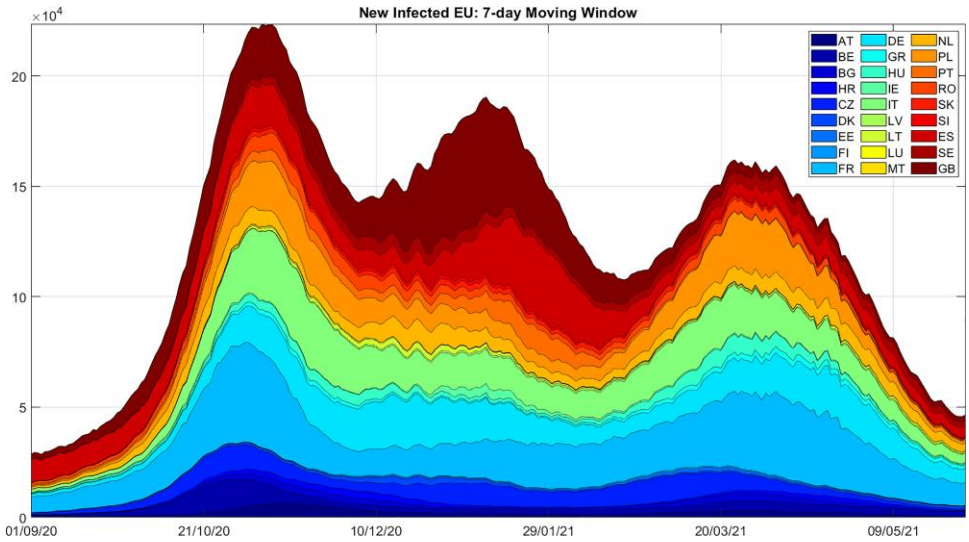

Figure 1. Time evolution of newly infected in European countries, aggregate. The figure shows the dynamics of the 7-day moving average of the total number of new COVID-19 reported cases in the 27 European countries considered, over the period 1 September 2020 to 30 May 2021.

two marked infection peaks in Europe emerges, one in November 2020 and the other one at the beginning of April 2021. There is also evidence of a further peak straddling the two years 2020 and 2021, though this one has affected countries with dissimilar intensities, and with particular strength in UK and Spain. In addition, notice that the infection curves tend to rapidly decrease after the last registered peak of April 2021.

As far as the explanatory variables are concerned, in Figure 2 we show the dynamics of the total number of COVID-19 vaccines inoculated over the period 1 September 2020 to 30 May 2021. Figure 2 shows the rapid increase of vaccines particularly from April 2021, which can be easily correlated to the decrease of cases seen in Figure 1.

The use of mobility data as explanatory variables requires the construction of tailored regressors which we obtain by means of graph theory. In particular, mobility flows consist of connectivity matrices, meaning matrix-valued data containing the number of individuals flowing from administrative region i to administrative region j. Therefore, it is imperative to summarize this information, as it cannot be included into a regression model given the matrix nature of data. The same holds for the airline data, which still consists of connectivity matrices.

Mobility flows can be represented as a graph $\mathcal{G} = (V, E)$, where the vertices in V are regions or countries. The edges in E represent commuting flows from one node to another, i.e., the number of people moving from one region (or country) to another. The mobility graph can then be represented by the weighted adjacency matrix $\mathcal{G} \Leftrightarrow \mathbf{W}(\mathcal{G}) = [w_{od}]$, in which the entry w_{od} is equal to the number of people moving from o to d.

The information contained in mobility graphs can be summarized by a projection method, such as PCA. However, while Google mobility data are defined as deviations from a benchmark and can therefore take on any value into the set of real numbers, Facebook and airline are non-negative, discrete variables. To project them in a lower dimensional space, we need to extend principal components analysis to the more general NNMF—see, e.g., Lee and Seung (1999) and Pecora et al. (2016). NNMF has been widely applied in statistical learning to a number of different areas, such as pattern recognition (Li et al., 2001), multimedia data analysis (Cooper & Foote, 2002), text mining (Pauca et al., 2004), classification (Lee et al., 2009), and clustering (Sha et al., 2002). The method consists of factorizing a matrix $\mathbf{W}$ into two matrices, $\mathbf{B}$ and $\mathbf{L}$, such that both matrices have no negative elements, i.e., $\mathbf{B} \in \mathcal{R}^{K \times F}$ and $\mathbf{L} \in \mathcal{R}^{F \times T}$. To obtain a summary measure, we can set $F = 1$ so that the element B_i corresponds to the variable factor and the element L_i represents the time factor.

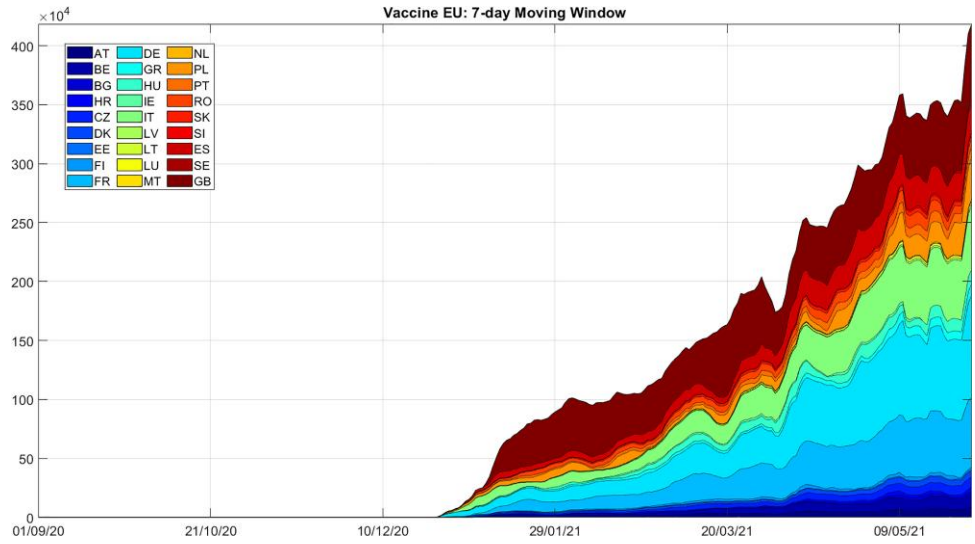

Figure 2. Time evolution of daily number of COVID-19 vaccines inoculated, aggregate. The figure shows the dynamics of the total number of COVID-19 vaccinations in the 27 European countries over the period 1 September 2020 to 30 May 2021.

The application of PCA (Wold et al., 1987) to the Google data and NNMF to the Facebook data leads to the mobility variables represented in Figure 3, which we present for two selected countries. In particular, results are illustrated for Germany (DE) and Italy (IT), given the fact that Germany is the most populated country in the sample, whereas Italy is the country which has suffered most infection counts, especially during the first wave of COVID-19. The representation of Google and Facebook mobility variables for all countries can be found in Online Supplementary Material, Figures S3 and S4 reported in the web-based supplementary information. A similar projection can be obtained for the airline data, which we illustrate in Online Supplementary Material, Figure S2 of the web-based supplementary information as well, together with a joint visualization of the mobility flows between countries (airline mobility data) and within countries (Facebook data), for each analysed European country, in Online Supplementary Material, Figure S1. Such summary measures will be employed as explanatory factors of the COVID-19 infection counts.

Figure 3 shows how the nationwide lockdown measures imposed concomitantly with the epidemic growth has exerted dramatic impacts on different dimensions of human mobility. The upper panels illustrate the geographic distribution of changes in human mobility trends, provided by Google, for groceries, parks, residential areas, retail and recreation sites, groceries and pharmacies, transit station and workplaces. People's movements to retail and recreation sites, given their non-essentiality, were the ones most hardly hit by the imposition of lockdown measures, along with those to transit stations and workplaces. While movements to parks do highly vary over time, those to residential areas and retail stations fluctuate quite steadily. The lower panels show the human mobility patterns, provided by Facebook, within and between administrative regions. It is evident that lockdowns have also tied to the magnitude of commuting flows in a heterogeneous way, depending on the country under consideration. In any case, the mobility within administrative regions was the most relevant part of human mobility over the considered period: that is the reason why the NNMF almost perfectly emulates its behaviour.

We now consider the analysis of the country specific health and demographic explanatory variables which are reported as heat-maps in Online Supplementary Material, Figure S5 of the web-based supplementary information. The demographic and health risk variables can be summarized by two separate projections, obtained applying NNMF, as illustrated in Figure 4. The main reasons behind considering a summarized version of the health and demographic factors, rather than modelling them as single covariates, are: (a) parsimony of the model specification and (b) enhanced predictive performance. As a matter of fact, including all variables as single covariates would

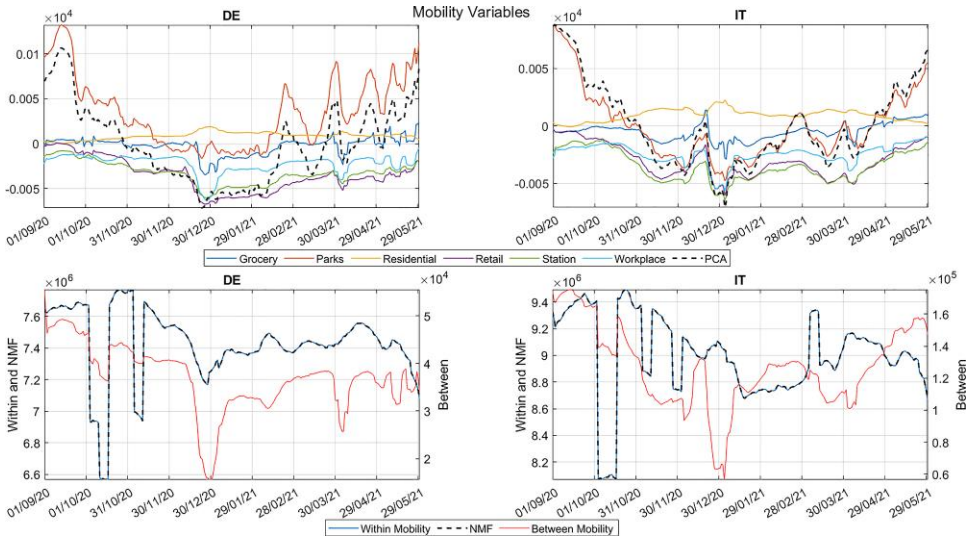

Figure 3. Mobility variables. The figure illustrates the dynamics of the Google (upper panels) and Facebook (lower panels) mobility variables for two representative countries, namely Germany (DE) and Italy (IT). As far as Google variables, they represent movements across different locations: grocery and pharmacies, parks, residential, retail and recreation, stations and workplaces with coloured lines. Each variable quantifies the percentage reduction of mobility with respect to a 5-week baseline corresponding to the period before the outbreak of the pandemic. In the figure, the dashed line represents our considered summary measure: the first principal component summarizes the Google variables. As far as Facebook variables, the figure illustrates the evolution of human mobility patterns provided by the Facebook data, within and between administrative regions. In the figure, the dashed line represents the first non-negative matrix factorization (NNMF) component which summarizes the Facebook variables.

sensibly increase the number of estimated parameters in the Poisson regression of each country, which might even cause problems in the estimation, such as multicollinearity. Moreover, by performing a separate estimation of the model which includes the above-mentioned variables as single covariates, rather than modelling them as factors, we have obtained that reducing the dimensionality of the model improves its predictive accuracy. Figure 4 shows that the risk factor is more important for eastern European countries, such as Poland, Romania, and Hungary and, conversely, the demographic factor matters more for western countries, such as France, Germany, and Italy.

We finally illustrate the behaviour of the policy measures, as summarized by the overall stringency index, as determined by the Oxford Policy Tracker. The index represents the daily level of stringency of policies put in place by national governments. The possible values range on a scale from 0 to 100, increasing with the degree of stringency of the policy mix adopted on each specific date. Figure 5 shows that, for most countries, the highest stringency occurred during the central part of the time period considered.

3 Methods

In this section, we illustrate our proposed methodology. First, we review self-exciting Hawkes processes for epidemics modelling. Second, we introduce our proposal: a self-exciting point process endowed with temporal, mobility, and spatial covariates to explain the epidemic diffusion.

3.1 Background

In the context of branching point processes, the seminal paper by Hawkes (1971) introduces self-exciting point processes, a class of birth processes which admits the possibility that current events might be triggered by prior events, and that current events might trigger further future events.

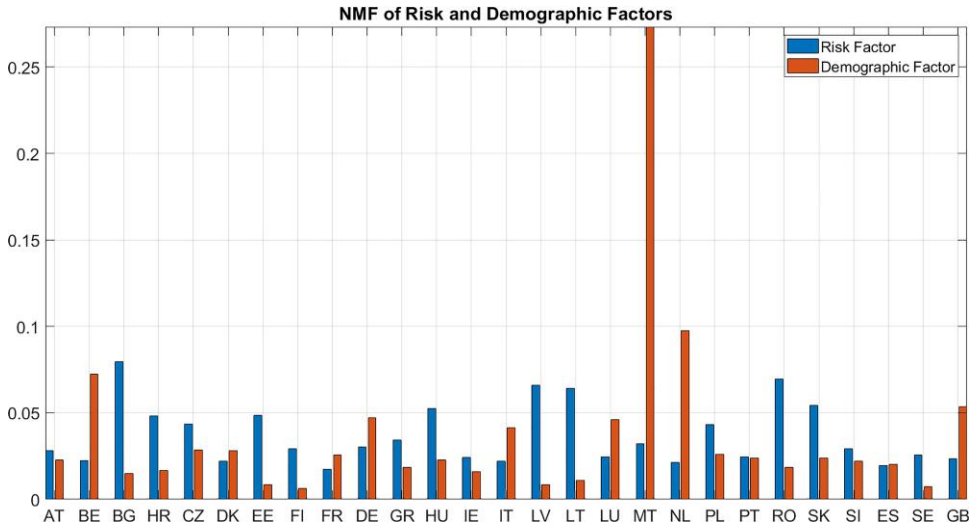

Figure 4. Risk and social factors non-negative matrix factorization (NNMF). The figure shows the NNMF values associated to each country and derived by the social and risk factor variables.

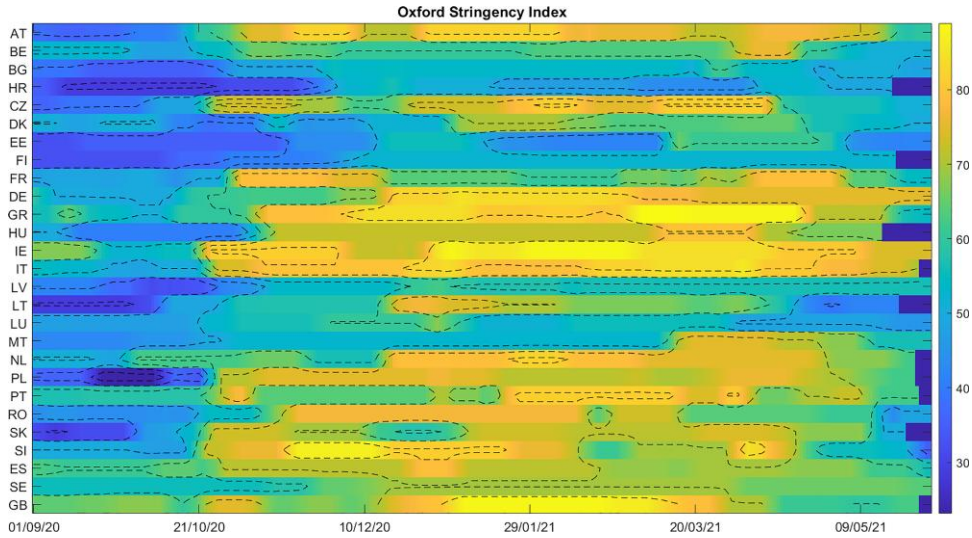

Figure 5. Stringency index. The figure illustrates the time dynamics of the stringency index across the European countries over the study period.

Consider a simple point process of event times $t_i \in [0, T)$, such that $t_i < t_{i+1}$, and a right-continuous counting measure $N(\mathcal{T})$, defined as the number of events occurring at times $t \in \mathcal{T}$. The conditional intensity $\lambda(t \mid \mathcal{H}_t)$ of a point process on the real half-line with respect to a filtration $\mathcal{H}_t$, i.e., the history of all events up to time t, is defined as

$$\lambda(t \mid \mathcal{H}_t) = \lim_{\Delta t \to 0} \frac{E[N(t, t + \Delta t) \mid \mathcal{H}_t]}{\Delta t} \tag{1}$$

with $\lambda(t \mid \mathcal{H}_t)$ representing the infinitesimal rate at which the expected number of points are accumulating at time t, given the history $\mathcal{H}_t$ of all points occurring prior to time t (see Daley & Vere-Jones, 2003).

The Hawkes process (see Hawkes, 1971; Hawkes & Oakes, 1974) represents one of the earliest specifications of $\lambda(t \mid \mathcal{H}_t)$, defined as the conditional intensity of the expected rate at which an event occurs at time t, given all events that occurred previously at times $t > t_j$:

$$\lambda(t \mid \mathcal{H}_t) = \mu + \sum_{t > t_j, t_j \in \mathcal{T}} g(t - t_j) \tag{2}$$

The Hawkes model has been proposed in the field of seismology: see, for instance, Vere-Jones and Ozaki (1982), to design the diffusion process of earthquakes, where the main shocks occur at a constant baseline rate μ over time, and each earthquake at time t_j raises the risk of future earthquakes (aftershocks) through the self-exciting function $g(t - t_j)$, which is often assumed to be power law, exponential, or fat tailed distributed.

Model parameters are generally estimated by means of maximum likelihood. To this aim, the log-likelihood of an observed sequence of events according to an estimated Hawkes model is given by

$$\mathcal{L}(\Theta) = \sum_{i=1}^{n} \log\left(\lambda(t_i; \Theta)\right) - \int_0^T \lambda(t; \Theta)\, \mathrm{d}t \tag{3}$$

where Θ is the set of model parameters and n represents the number of observations.

The literature has extended the self-exciting point process models in several ways over the last few decades. In particular, by means of a multivariate extension of the self-exciting point process, Ogata (1988) proposed the epidemic-type aftershock sequence model to examine the temporal dynamics of aftershock activities of earthquakes. Since then, self-exciting point processes have been applied in different contexts such as crime forecasting (Mohler et al., 2011), events networking (Fox et al., 2016), invasive plant infection (Balderama et al., 2012), war insurgencies (Lewis & Mohler, 2011), finance (Adelfio et al., 2021), and epidemic infection (Meyer et al., 2012), giving rise to a number of methodological improvements.

3.2 Proposal

Within self-exciting processes, the use of additional covariates to model the conditional intensity of a given event has recently become important in the field of crime detection. In particular, Reinhart and Greenhouse (2018) introduce a spatio-temporal self-exciting point process model that incorporates spatial features, near-repeat and retaliation effects, along with triggering to model the dynamics of crime. Mohler et al. (2018) propose a modulated Hawkes process which exploits the incorporation of spatial covariates to construct crime-related social harm indices. Park et al. (2021) analyse gang-related violent crime data using demographic and socio-economic covariates. These studies share the idea of including covariates in the specification of the conditional intensity of the point process.

In this paper, we use an additional set of different types of covariates to model the occurrence of SARS-CoV-2 cases. On the one hand, people commuting flows are key to determine the dynamics of the virus spread, as larger levels of mobility are associated with higher diffusion of the epidemic. On the other hand, spatial covariates such as cluster population density might also exert an effect on the epidemic diffusion process, given that the more people are concentrated per unit of area, the more likely the virus is able to spread across individuals.

Within the context of a self-exciting process, an infection can be attributed to two different components: a background endemic component, coming from the constant rate μ, which is not imputed by the model to any prior infection; and a dynamic epidemic component, in which any infection is generated by a previous one and can in turn generate new ones.

We assume that temporal and spatial covariates can affect only the epidemic component. This assumption is consistent with the fact that the SARS-CoV-2 disease spread is driven mostly by person to person contact rather than being contracted in an exogenous way by means of environmental factors. Our assumption implies, as noted by Reinhart (2018), that any departures from $\mu = 0$

must be interpreted as cases caused by unobserved infections or, in our case, by the inherent tendency of a country to import infections from outside the borders.

More formally, given a set of spatio-temporal covariates $\mathbf{z}_k(t)$ at time t for each country k, where $k = 1, \ldots, K$, we specify a self-exciting process with covariates as follows:

$$\lambda_k(t) = \mu_k + \sum_{t > t_j, t_j \in \mathcal{T}} g(t - t_j; \mathbf{z}_k(t)) \tag{4}$$

In the endemic part of equation (4), we assume that the background infection rate μ_k is constant over time but can vary across countries to capture possible differences in the baseline rate of occurrences across countries. Moreover, we also assume that each event at time t_j increases the risk of future events by means of the modified triggering function $g(t - t_j; \mathbf{z}_k(t))$, which is assumed to be separable into a structural component g_1 (reproduction rate) and a non-structural component g_2 (inter-infection rate) (see Park et al., 2021), so that $g(t - t_j, \mathbf{z}_k(t)) = g_1(\mathbf{z}_k(t))g_2(t - t_j)$.

We then assume that the structural component $G_k^t = g_1(\mathbf{z}_k(t))$, which represents a reproduction rate, is a Poisson random variable, whose expected value is specified by a Poisson regression model as

$$E[G_k^t \mid \mathbf{z}_k(t), \theta] = \mathrm{e}^{(\theta' \mathbf{z}_k(t))} \tag{5}$$

The non-structured component g_2, instead, is assumed to represent an inter-infection time distribution, which, consistently with the literature on epidemics (Cowling et al., 2010; Hellewell et al., 2020; Obadia et al., 2012), can be assumed to follow a Weibull distribution $g_2(t - t_j; \beta, \alpha) = (\beta/(t - t_j))((t - t_j)/\alpha)^{\beta}\, \mathrm{e}^{-((t-t_j)/\alpha)^{\beta}}$, with β and α representing the shape and scale parameters, respectively.

Following the previous assumptions, our model is therefore characterized by the set of parameters $\Theta = [\theta, \alpha, \beta, \mu_k]$, whose log-likelihood can be expressed as

$$\begin{aligned} \mathcal{L}(\Theta) = \sum_{k=1}^{K} \Bigg\{ & \sum_{i=1}^{n_k} p_k(i, i) \log(\mu_k) - \textstyle\int_0^T \mu_k \, \mathrm{d}t \\ & + \sum_{i=2}^{n_k} \Bigg\{ \sum_{j=1}^{i-1} p_k(i, j) \log [g_1(\mathbf{z}_k(t), \theta) g_2(t - t_j \mid \beta, \alpha)] \\ & - \textstyle\int_{t_j}^{T} g_1(\mathbf{z}_k(t), \theta) g_2(t - t_j \mid \beta, \alpha) \, \mathrm{d}t \Bigg\} \Bigg\} \end{aligned} \tag{6}$$

in which p_k represent the branching probabilities for country k and, in particular, $p_k(i, i)$ is the probability that case i was imported, whereas $p_k(i, j)$ represents the probability that case i was caused by case j.

Note that the likelihood in equation (6) corresponds to the likelihood of K separate (non-interacting) self-exciting point processes, for which covariates are aggregated at country level. Maximizing the likelihood is therefore equivalent to separately maximize each of the k country specific marginal likelihoods, with respect to the rate described in equation (4).

Accordingly, parameter estimation can be carried out for each country by means of EM algorithm, which consists of an iterative alternation of an expectation step, in which the branching probabilities p_k are estimated, and a maximization step, during which the model parameters $\Theta = [\theta, \alpha, \beta, \mu_k]$ are updated by maximizing the log-likelihood function. As noted by Lewis and Mohler (2011), the EM algorithm is equivalent to a projected gradient ascent on the likelihood of the Hawkes process.

Specifically, within the expectation step, we estimate the branching probabilities $p_k(i, j)$ for each country. Within the maximization step, the log-likelihood in equation (6) is maximized for each country with respect to the model parameters $\Theta = [\theta, \alpha, \beta, \mu_k]$, conditionally on the estimated branching structure $p_k(i, j)$ derived from the expectation step. The maximization of the

log-likelihood function in equation (6) boils down to three independent optimization problems. The first consists of a Poisson regression of observations $\sum_{i=2}^{n_k} p_k(i, j)$ on the spatial and mobility covariates $\mathbf{z}_k(t)$:

$$\hat{\theta} := \underset{\theta}{\text{argmax}}\ \mathcal{L}_\theta = \underset{\theta}{\text{argmax}} \left\{ \sum_{i=2}^{n_k} \left[\sum_{j=1}^{i-1} p_k(i, j) \right] \theta' \mathbf{z}_k(t) - \mathrm{e}^{(\theta' \mathbf{z}_k(t))} \right\} \tag{7}$$

The second optimisation problem concerns the shape and scale parameters of the inter-infection distribution:

$$\hat{\beta}, \hat{\alpha} := \underset{\beta,\, \alpha}{\text{argmax}}\ \mathcal{L}_{\beta,\alpha} = \underset{\beta,\, \alpha}{\text{argmax}} \left\{ \sum_{i=2}^{n_k} \left\{ \sum_{j=1}^{i-1} p_k(i, j) \log \left[g_2(t_i - t_j \mid \beta, \alpha) \right] \right\} \right\} \tag{8}$$

where $p_k(i, j)$ is the weight of each inter-infection time observation t_i, t_j.

The third optimization problem concerns the background rate μ_k which is obtained from

$$\hat{\mu}_k := \underset{\mu_k}{\text{argmax}}\ \mathcal{L}_{\mu_k} = \underset{\mu_k}{\text{argmax}} \sum_{i=1}^{n_k} p_k(i, i) \log(\mu_k) - \int_0^T \mu_k \, \mathrm{d}t = \sum_{i=1}^{n_k} \frac{p_k(i, i)}{T} \tag{9}$$

We remark that differently from the effect of spatial covariates, that of mobility covariates is time-lagged: the infection time can start before, during, or after the onset of symptoms. In other words, there exists a certain incubation period before which the disease becomes manifest after the contact with an infected individual. To account for this occurrence, we consider the number of new infections triggered by a previous infected case to be dependent upon the level of mobility, but lagged in time. As reported, for example, by the European Centre for Disease Prevention and Control, the infection time duration of the SARS-CoV-2 disease is roughly 14 days (see ECDC, 2020). Thus, we let the delay of mobility covariates vary from 10 to 14 days and choose the best lag order within this range in terms of predictive performance when considering the in-sample estimation phase. We also remark that what presented so far can be extended in a Bayesian framework, for example, following the non-parametric approach of Giudici et al. (2003).

4 Application

4.1 Model estimation

We have applied our proposed model to the described data. From an interpretational viewpoint, the most important aspect of the model is the structural component of the epidemic part given by the Poisson autoregression. The resulting parameter estimates, along with their standard errors, associated t-statistics, and p-values, are contained in Table 1. As in Chiang et al. (2021), standard errors and p-values are calculated from the Poisson regression in the M-step after the EM algorithm reaches convergence.

The results in Table 1 show highly statistically significant coefficients. In particular, we find that the largest negative statistically significant coefficient is that of the stringency index (stringency: −0.049, p-value: <0.001), followed by the vaccine variable (vaccine: −0.0137, p-value: <0.001). These signs are in line with the expectations. The sign of the airline factor coefficient is also negative (airline factor: −0.0038, p-value: <0.001), differently from what would be expected a priori. This may be due to a reverse causal effect: when contagion counts are higher, travel restrictions increase and become more selective, increasing the centrality of the most populated countries. In addition, international contagion might have played role at the very beginning of COVID-19, when travel bans were still not imposed, or during low stringency phases, rather than during nationwide lockdowns. The remaining variables' coefficients are all statistically significant at all conventional significance levels and, as expected, positive. The highest magnitudes are those associated to the Google and Facebook mobility variables, followed by the social factor,

Table 1. Poisson regression

	Estimate	SE	*t*-stat	*p*-value
Intercept	0.0247	0.0003	92.50	<0.001
Vaccine	−0.0137	0.0001	−92.72	<0.001
Stringency	−0.0469	0.0003	−150.72	<0.001
Facebook factor	0.0122	0.0002	64.29	<0.001
Google factor	0.0155	0.0003	55.32	<0.001
Airline factor	−0.0038	0.0002	−20.76	<0.001
Risk factor	0.0014	0.0003	5.27	<0.001
Social factor	0.0146	0.0004	32.87	<0.001

Note: The table reports the Poisson regression coefficients (estimate), standard errors (SEs), test statistics (*t*-stat), and *p*-values (*p*-value) obtained applying the model to the available sample.

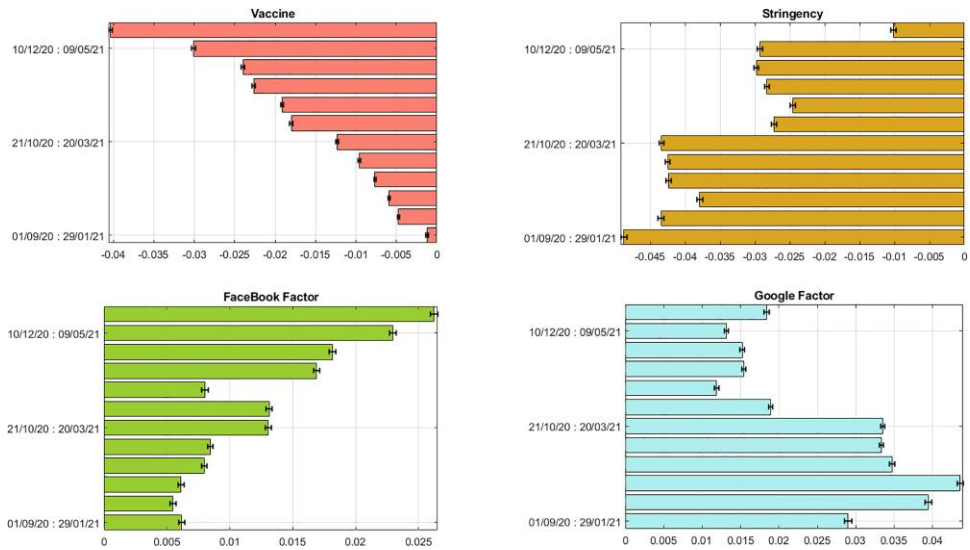

Figure 6. Time evolution of the Poisson regression coefficients. The figure reports the regression coefficients obtained from the dynamic model estimation and associated to the time-series of the number of inoculated vaccines (upper left), Oxford Stringency Index (upper right), Facebook factor (bottom left), and Google factor (bottom right).

which reflects population size and density. The risk factor is also positive, but with a relatively smaller magnitude, according to the model estimates.

To help interpretation, the Poisson regression model can be fit dynamically, producing estimates for each point in time t, based only on data available up to time $t-1$. We set for this purpose an estimation window width of 150 days. Figure 6 reports the resulting dynamic Poisson regression coefficients for the explanatory variables that vary over time: vaccines (upper left), Oxford Stringency Index (upper right), Facebook factor (bottom left), and Google factor (bottom right). Figure 6 shows, in line with the expectations, the increasing importance of vaccines and the decreasing importance of stringency measures, as well as an increasing one associated to the Facebook mobility factor.

For the sake of completeness, we also report the estimation results for the inter-infection and the background parameters of the model. Figure 7 shows the time-series of the estimated dynamic shape (α) and scale (β) parameters of the fitted Weibull distribution, along with the value of the background rate (μ_c). Figure 7 shows that the shape and scale Weibull parameters remain of quite

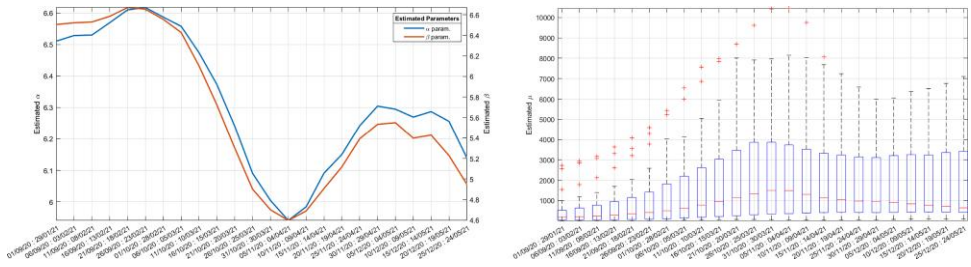

Figure 7. Estimation of the α, β, and μ_c parameters along time. The figure shows the estimates of the shape (α) and scale (β) Weibull parameters obtained from the model estimation (left) and the boxplots associated to the values of the background rate (μ_c) obtained from model estimation (right). For each time window, a background rate for each of the European countries is estimated, and their values are reported in the boxplot within the corresponding period.

comparable size over time, with a sharp decline over the periods ranging approximately from November 2020 to April 2021. In the same period, the country infection background rate estimates increase, almost counterbalancing the decrease in the inter-infection distribution parameters. This indicates a lower performance of the model in capturing the variability of the response variable. Indeed, our results show that, during the same period, the estimated branching probability of an infection to be imported becomes larger, and, at the same time, that of being transmitted lower. This may be a further explanation of the divergence in inter-infection and background parameter estimates, and in the lowering of the model's predictive performance. For the purpose of investigating the robustness of our results, we have replaced the Weibull distribution of inter-infection with a Gamma distribution, as suggested by Obadia et al. (2012) and Chiang et al. (2021). Results remain qualitatively unchanged—see Online Supplementary Material, Figure S7 of the web-based supplementary information.

Before using the model for scenario analysis, a necessary step for policy decision-making is to evaluate its predictive accuracy. The model is back-tested using a dynamic approach. For the in-sample estimation, we have chosen consecutive time windows of 150 days, each one shifted by one week. For each of them, we have produced out-of-sample predictions for the following week. All obtained predicted values are recorded and compared with the observed ones, from February 2021 to the end of the sample. For forecasting purposes, we make use of the branching representation of the Hawkes process. First, we simulate secondary events through the Poisson process based on its background rate; namely, for each event, we simulate a Poisson random variable with mean $e^{(\theta' \mathbf{z}_k(t))}$. Second, we draw the inter-event times of infections from the estimated Weibull distribution. Finally, we simulate 100 realizations of the Hawkes process to estimate its average intensity forecast, as well as its associated confidence intervals.

Figure 8 shows the in-sample and out-of-sample predictions of newly reported COVID-19 cases, aggregated over the 27 European countries, for the period ranging from 1 September 2020 to 30 May 2021. The predicted values are compared with the actually observed ones and are represented along with their associated confidence intervals. Additionally, we report the same predictions for two selected countries, i.e., Italy and Germany. In Online Supplementary material, Figure S6 of the web-based supplementary information, we report the results for each country, which generate the aggregate predictions under analysis. Figure 8 shows that the difference between the one-week ahead predictions and the actual values is relatively small, and this is confirmed by the predictive accuracy at the country level in Online Supplementary Material, Figure S6 of the web-based supplementary information. Furthermore, notice that newly reported cases of Germany seem to be more timely predictable than those of Italy. Additionally, for the purpose of investigating possible under-dispersion in the model residuals, we report in Figure 9 a scatter plot which compares the real infections with the predicted ones. As Figure 9 suggests the point estimates are mostly aligned along the 45° line, thus indicating a limited under-dispersion.

4.2 Policy scenario analysis

In this section, we look into the future management of the pandemic and carry out a scenario analysis to model a hypothetical vaccination strategy to analyse the epidemic outcomes derived from

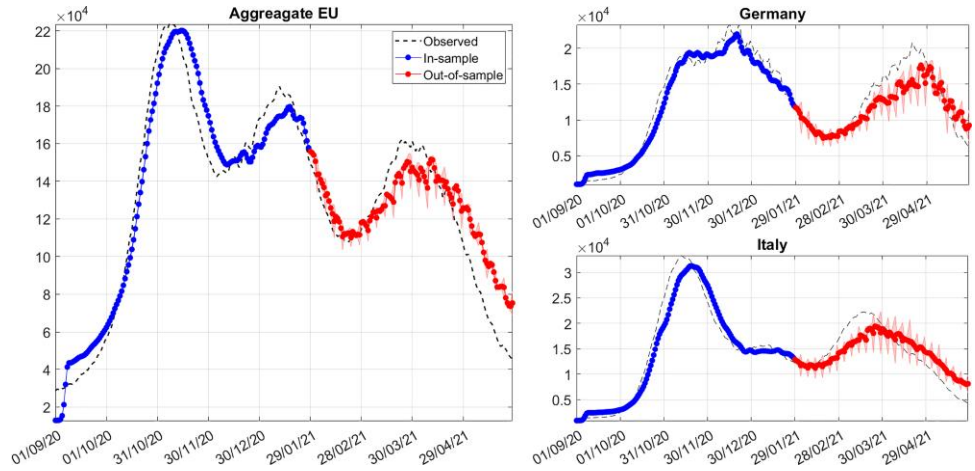

Figure 8. Predictions of newly reported COVID-19 cases. The left panel of the figure shows the aggregate predictions of newly reported COVID-19 cases over the period 1 September 2020 to 30 May 2021. The first part of the curves represents the in-sample estimates over the period from 1 September 2021 to 29 January 2021. The second part of the curves represents the out-of-sample one-week ahead predictions over the period 30 January 2021 to 30 May 2021. The dotted curves represent the actual number of newly reported COVID-19 cases over the same period. The right panel reports the predictions of newly reported COVID-19 case for Germany and Italy over the same period.

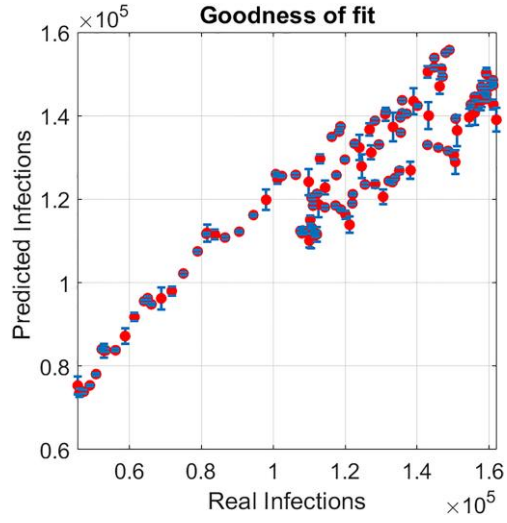

Figure 9. Scatter plot of real infections and out-of-sample forecasts. The figure reports a scatter plot which compares real infections with the predicted ones, along with their associated confidence intervals.

the interplay between (a) an increase of the vaccine doses available in the countries and (b) a co-ordinated re-distribution scheme which allocates the vaccines among the different countries.

More precisely, we evaluate the impact of different vaccination strategies by simulating the number of infected individuals which would result after providing hypothetical additional doses of vaccines. Our modelling strategy is composed by two steps: first, we increase the available quantity of vaccines at each day t by a percentage $q = \{30, \ldots, 100\}$ and, second, we assign these additional doses sorting countries in ascending order according to their vaccine-to-infected ratio and apply a sigmoid function to create a vector of distribution weights.

More formally, let $M_t \in \mathcal{R}^K$ be a sequence of real numbers indicating the vaccine-to-infected ratio observed at day t among the K countries. We define the sequence $sort(M_t) \in \mathcal{R}^K$ to be the

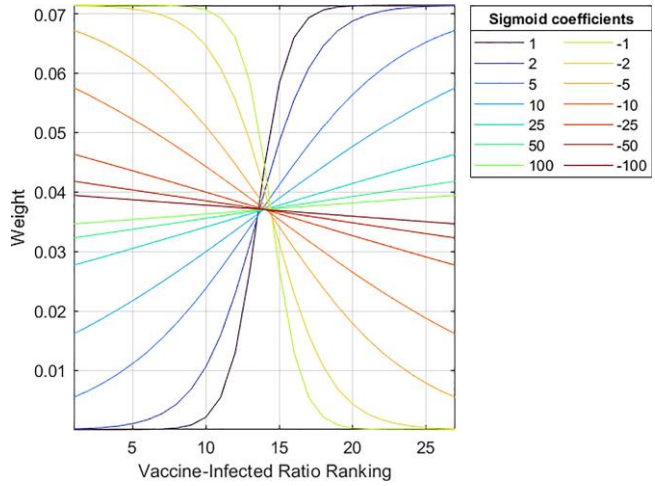

Figure 10. Re-distribution function. The figure shows the sigmoid function used to re-distribute vaccines among European countries. In particular, the sigmoid parameter s generates different re-distribution schemes which assign a weight to each country depending on its vaccine-to-infected ratio. The x-axis defines the ranking of countries according to their vaccine-to-infected ratio in ascending order. The y-axis reports the weights (portions of additional vaccines) assigned by the sigmoid function.

result of sorting M_t, preserving duplicate elements. Let K, $l \in \mathcal{N}$ and for all $1 \leq i \leq K$ let $m_t^i \in \mathcal{R}$ and let $M_t^i = (m_t^i)_{i=1}^K$ be the vaccine-to-infected ratio sequence. Let $B = (b_t^i)_{i=1}^K = sort(M_t)$. We define $M_t(l)$ to be the sequence $M_t(l) = (c_i)_{i=1}^l$ of real numbers satisfying, for all $1 \leq i \leq l$, $c_i = b_{K-l+i}$. Once we have sorted the countries according to the position vector $\mathbf{l}$, we use such vector to assign a weight $\mathbf{w}$ according to a sigmoid function:

$$w(l_i) = \frac{1}{1 + e^{-s(-l_i - \langle l \rangle)}}$$

where $\langle l \rangle$ is the average value of the position vector $\mathbf{l}$ and the parameter set $s = \{1, 10, 25, 50, 100, -1, -2, -5, -10, -25, -50, -100\}$ defines the shapes of the sigmoid function.

Figure 10 shows the sigmoid function used to re-distribute vaccines among European countries in our context. The sigmoid parameter set s generates different re-distribution schemes which assign a weight to each country depending on its vaccine-to-infected ratio. Notice that positive and low values of the parameter s generate a re-distribution scheme which gives less weight to the countries having a low value of the vaccine-to-infected sequence, while negative and low values of s assign more vaccines to the countries with corresponding low vaccine-to-infected ratio. Relatively high values of s in absolute terms tend instead to assign more evenly distributed weights.

The quantity of vaccine $\tilde{v}_t^i$ put at the disposal of country i at time t at each iteration, given the value assumed by parameters q and s, can then be computed as

$$\tilde{v}_t^i = v_t^i + \frac{q}{100} \frac{w(l_i)}{\sum_i w(l_i)} \sum_i v_t^i \tag{10}$$

where v_t^i is the recorded amount of vaccine at disposal of the ith country at time t as in the dataset. We evaluate the model on a grid composed by the different coordinates in the space $q \times s$, recording for each iteration the predicted evolution of the number of infected individuals in each country. In this way, we come up with 144 (namely, 12×12) alternative policy scenarios along with the corresponding health impact in terms of newly reported COVID-19 infections.

Figure 11 shows the percentage of daily avoided COVID-19 infections, with respect to the benchmark case of $q = 0$ (no additional vaccine doses), as the number of total vaccines increases and the re-distribution functions change according to the sigmoid parameter. The colour bar

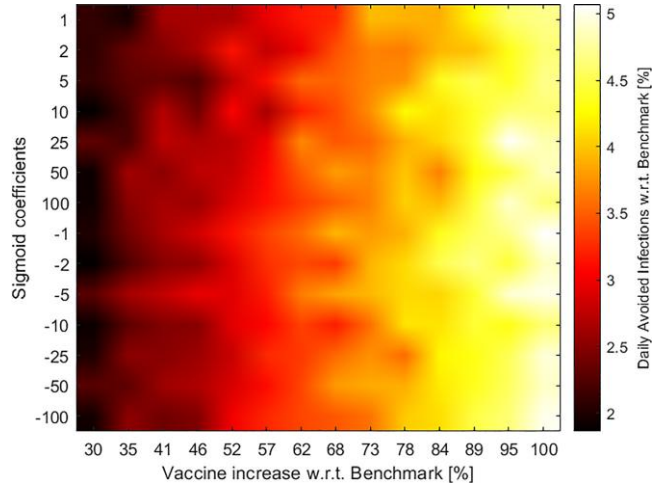

Figure 11. Simulation of infection avoided for European countries in different scenarios. The figure shows the average daily infection avoided as total vaccines increase from 30% to 100% of the actual doses and the re-distribution functions change according to the values assumed by the sigmoid parameter. The colour bar maps colours into infection avoided.

assigns colours based upon the number of infections avoided (ranging from black to white as the infection decreases in intensity). From Figure 11, as expected, increasing values of q, associated to a higher number of vaccine doses available, produce a contraction of the epidemic spread. In contrast, the parameter s, besides affecting the allocation of additional doses, does not seem to exert severe impacts on the number of infected individuals in relative terms.

To shed further light on the effect of policy parameters, we compare the estimated avoided infections by discriminating among the effects produced by changes of q and s. To do so we average, by rows and by columns, the matrix reported in Figure 11, which provides the avoided infection for different pairs of values of q and s. Figure 12 shows, in the left panel, the contribution of the increased vaccine procurement to the total infections avoided. The right panel reports the total infections avoided depending upon different values of the sigmoid parameter s. From Figure 12, notice that, despite the fact that increasing the parameter q substantially reduces the average infections of European countries, the effect of the re-distribution schemes is evident. Indeed, despite the infection reduction effect is not linear as a consequence of changes in s, we observe that an inverse re-distribution scheme (low and negative values of s), which allocates most of the additional doses to countries with a low vaccine-to-infected ratio, is able to reduce infections more than a direct scheme, which rewards countries with a high ratio (low and positive s values). In the middle, a neutral scheme (high s values in absolute terms) more evenly assigns additional doses to countries.

5 Concluding remarks

Human mobility plays a key role in the COVID-19 contagion dynamics. The information contained in mobility networks is crucial to understand the drivers of contagion and to build accurate statistical models for the spread of epidemics, which can be used to make predictions and for decision-making purposes.

Against this background, we propose a network-based spatio-temporal statistical model, based on self-exciting point processes, enhanced with a large set of covariates, which include those derived from mobility patterns. In this way, we can effectively investigate the impact of mobility covariates on the development of the epidemic.

The application of our methodology to the reported COVID-19 cases in European countries shows that our methodology provides satisfactory predictive performances. We discover that human mobility networks, vaccine doses, and government policies, along with disease and demographic risk factors, are all significant predictors of the number of new COVID-19 reported cases.

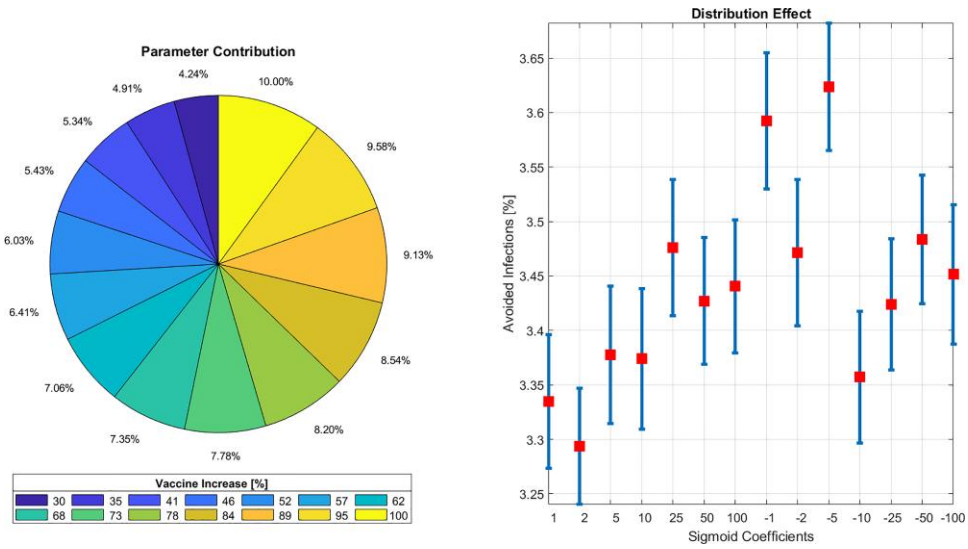

Figure 12. Simulation of avoided infections due to vaccine doses increase (left) and distributional effects (right). The figure shows, in the left panel, the contribution of an increased vaccine procurement to the total infections avoided. The right panel reports the total infections avoided by discriminating across different values of the sigmoid parameter *s*.

In addition, we conduct a policy scenario analysis on the number of vaccine doses put at the disposal of European countries, under different procurement schemes. The policy scenario analysis highlights that an increasing number of vaccines at the disposal of European countries would exert a substantial decline in COVID-19 infection rates, with the distributional scheme adopted being a relevant driver of contagion reduction. Overall, the most effective vaccine distribution policy option would be that of providing a larger part of resources to countries with relatively lower vaccine-to-infected ratios.

We believe that our proposed approach could be employed as a continuous monitoring tool which, based on the available data, is able to evaluate the health impacts associated with different setups of policy measures that aim at preventing disease contagion.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society*.

Conflict of interests: The authors declare no conflicts of interest.

Funding

The research is funded by the European Union, Horizon 2020, research and innovation programme 'PERISCOPE: Pan European Response to the ImpactS of COVID-19 and Future Pandemics and Epidemics', grant agreement no. 101016233, H2020-SC1-PHE-CORONAVIRUS-2020-2-RTD.

Data availability

COVID-19 diffusion data available at https://github.com/CSSEGISandData/COVID-19, COVID-19 vaccine data available at https://ourworldindata.org/coronavirus, Facebook mobility data available at https://dataforgood.fb.com/, airline mobility data available at https://ec.europa.eu/eurostat/data/database, Google mobility data available at https://www.google.com/covid19/mobility/, country data available at https://ourworldindata.org/coronavirus, government policy data available at https://www.bsg.ox.ac.uk/research/research-projects/COVID-19-government-response-tracker

Codes availability

Codes are available upon request.

References

Abegaz F., & Wit E. (2013). Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics*, *14*(3), 586–599. https://doi.org/10.1093/biostatistics/kxt005

Adelfio G., Agosto A., Chiodi M., & Giudici P. (2021). Financial contagion through space-time point processes. *Statistical Methods and Applications*, *30*(2), 665–688. https://doi.org/10.1007/s10260-020-00538-2

Aleta A., Martín-Corral D., Pastore y Piontti A., Ajelli M., Litvinova M., Chinazzi M., Dean N. E., Halloran M. E., Longini Jr I. M., Merler S., Pentland A., Vespignani A., Moro E., & Moreno Y. (2020). Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nature Human Behaviour*, *4*(9), 964–971. https://doi.org/10.1038/s41562-020-0931-9

Balderama E., Schoenberg F. P., Murray E., & Rundel P. W. (2012). Application of branching models in the study of invasive species. *Journal of the American Statistical Association*, *107*(498), 467–476. https://doi.org/10.1080/01621459.2011.641402

Barbillon P., Donnet S., Lazega E., & Bar-Hen A. (2017). Stochastic block models for multiplex networks: An application to a multilevel network of researchers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *180*(1), 295–314. https://doi.org/10.1111/rssa.12193

Becker N. (1977). Estimation for discrete time branching processes with application to epidemics. *Biometrics*, *33*(3), 515–522. https://doi.org/10.2307/2529366

Celani A., & Giudici P. (2022). Endemic–epidemic models to understand COVID-19 spatio-temporal evolution. *Spatial Statistics*, *49*(1), 100528. https://doi.org/10.1016/j.spasta.2021.100528

Chiang W.-H., Liu X., & Mohler G. (2020). Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *medRxiv*.

Chiang W.-H., Liu X., & Mohler G. (2021). Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *International Journal of Forecasting*, *38*(2), 505–520. https://doi.org/10.1016/j.ijforecast.2021.07.001

Chinazzi M., Davis J. T., Ajelli M., Gioannini C., Litvinova M., Merler S., Pastore y Piontti A., Mu K., Rossi L., Sun K., Viboud C., Xiong X., Yu H., Halloran M. E., Longini I. M., & Vespignani A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, *368*(6489), 395–400. https://doi.org/10.1126/science.aba9757

Cooper M., & Foote J. (2002). Summarizing video using non-negative similarity matrix factorization. In *2002 IEEE workshop on multimedia signal processing* (pp. 25–28). IEEE.

Cowling B. J., Lau M. S., Ho L.-M., Chuang S.-K., Tsang T., Liu S.-H., Leung P.-Y., Lo S.-V., & Lau E. H. (2010). The effective reproduction number of pandemic influenza: Prospective estimation. *Epidemiology*, *21*(6), 842–846. https://doi.org/10.1097/EDE.0b013e3181f20977

Daley D., & Vere-Jones D. (2003). Basic properties of the Poisson process. In *An introduction to the theory of point processes: Volume I: Elementary theory and methods* (pp. 19–40). New York, NY: Springer.

Davies N. G., Kucharski A. J., Eggo R. M., Gimma A., John Edmunds W., & Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group (2020). Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: A modelling study. *The Lancet Public Health*, *5*(7), e375–e385. https://doi.org/10.1016/S2468-2667(20)30133-X

Della Rossa F., Salzano D., Di Meglio A., De Lellis F., Coraggio M., Calabrese C., Guarino A., Cardona-Rivera R., De Lellis P., Liuzza D., Lo Iudice F., Russo G., & di Bernardo M. (2020). A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic. *Nature Communications*, *11*(1), 5106. https://doi.org/10.1038/s41467-020-18827-5

Di Zio M., Scanu M., Coppola L., Luzi O., & Ponti A. (2004). Bayesian networks for imputation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *167*(2), 309–322. https://doi.org/10.1046/j.1467-985X.2003.00736.x

ECDC (2020). *Coronavirus disease 2019 (COVID-19) pandemic: Increased transmission in the EU/EEA and the UK—Eighth update*, 8 April 2020. https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-rapid-risk-assessment-coronavirus-disease-2019-eighth-update-8-april-2020.pdf

Farrington C., Kanaan M., & Gay N. (2003). Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics*, *4*(2), 279–295. https://doi.org/10.1093/biostatistics/4.2.279

Flaxman S., Mishra S., Gandy A., Juliette H., Unwin T., Mellan T. A., Coupland H., Whittaker C., Zhu H., Berah T., Eaton J. W., Monod M., Imperial College COVID-19 Response Team, Ghani A. C., Donnelly C. A., Riley S., Vollmer M. A. C., Ferguson N. M., Okell L. C., & Bhatt S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *284*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Fox E. W., Short M. B., Schoenberg F. P., Coronges K. D., & Bertozzi A. L. (2016). Modeling e-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association*, *111*-(514), 564–584. https://doi.org/10.1080/01621459.2015.1135802

Ghani A., & Garnett G. (1998). Measuring sexual partner networks for transmission of sexually transmitted diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *161*(2), 227–238. https://doi.org/10.1111/rssa.1998.161.issue-2

Giudici P., Muliere P., & Mezzetti M. (2003). Mixtures of Dirichlet process priors for variable selection in survival analysis. *Journal of Statistical Planning and Inference*, *111*(1), 101–115. https://doi.org/10.1016/S0378-3758(02)00291-4

Giudici P., & Pagnottoni P. (2019). High frequency price change spillovers in bitcoin markets. *Risks*, *7*(4), 111. https://doi.org/10.3390/risks7040111

Giudici P., & Spelta A. (2016). Graphical network models for international financial flows. *Journal of Business & Economic Statistics*, *34*(1), 128–138. https://doi.org/10.1080/07350015.2015.1017643

Giudici, P., Tarantino, B., & Roy, A. (2023). Bayesian time-varying autoregressive models of COVID-19 epidemics. *Biometrical Journal*, *65*(1), 202200054. https://doi.org/10.1002/bimj.202200054

Guan W.-J., Ni Z.-Y., Hu Y., Liang W.-H., Ou C.-Q., He J.-X., Liu L., Shan H., Lei C.-L., Hui D. S. C., Du B., Li L.-J., Zeng G., Yuen K.-Y., Chen R.-C., Tang C.-L., Wang T., Chen P.-Y., Xiang J., Li S.-Y., *et al.* (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine*, *382*(18), 1708–1720. https://doi.org/10.1056/NEJMoa2002032

Hale T., Webster S., Petherick, A., Phillips, T., & Kira, B. (2020). *Oxford COVID-19 government response tracker (OxCGRT)*. p. 30. https://www.ecb.europa.eu/pub/conferences/shared/pdf/20201016_tracking_the_economy/Thomas_Hale.pdf

Harvey A., & Kattuman P. (2020). Time series models based on growth curves with applications to forecasting coronavirus. *Harvard Data Science Review*, *1*(S1). https://doi.org/10.1162/99608f92.828f40de

Hawkes A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, *33*(3), 438–443. https://doi.org/10.1111/j.2517-6161.1971.tb01530.x

Hawkes A. G., & Oakes D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, *11*(3), 493–503. https://doi.org/10.2307/3212693

Hellewell J., Abbott S., Gimma A., Bosse N. I., Jarvis C. I., Russell T. W., Munday J. D., Kucharski A. J., Edmunds W. J., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Funk S., & Eggo R. M. (2020). Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, *8*(4), e488–e496. https://doi.org/10.1016/S2214-109X(20)30074-7

Hsiang S., Allen D., Annan-Phan S., Bell K., Bolliger I., Chong T., Druckenmiller H., Huang L. Y., Hultgren A., Krasovich E., Lau P., Lee J., Rolf E., Tseng J., & Wu T. (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, *584*(7820), 262–267. https://doi.org/10.1038/s41586-020-2404-8

Khanin R., & Wit E. (2006). How scale-free are biological networks. *Journal of Computational Biology*, *13*(3), 810–818. https://doi.org/10.1089/cmb.2006.13.810

Kraemer M. U. G., Yang C.-H., Gutierrez B., Wu C.-H., Klein B., Pigott D. M., Faria N. R., Li R., Hanage W. P., Brownstein J. S., Layan M., Vespignani A., Tian H., Dye C., Pybus O. G., & Scarpino S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497. https://doi.org/10.1126/science.abb4218

Lee D. D., & Seung H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*-(6755), 788–791. https://doi.org/10.1038/44565

Lee H., Yoo J., & Choi S. (2009). Semi-supervised nonnegative matrix factorization. *IEEE Signal Processing Letters*, *17*(1), 4–7. doi:10.1109/LSP.2009.2027163

Lewis E., & Mohler G. (2011). A nonparametric EM algorithm for multiscale Hawkes processes. In *Proceedings of the 2011 joint statistical meetings*, *vol. 1* (pp. 1–16).

Li S. Z., Hou X. W., Zhang H. J., & Cheng Q. S. (2001). Learning spatially localized, parts-based representation. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001* (*Vol. 1*, pp. I–I). IEEE.

Maier B. F., & Brockmann D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*, *368*(6492), 742–746. https://doi.org/10.1126/science.abb4557

Meyer S., Elias J., & Höhle M. (2012). A space–time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics*, *68*(2), 607–616. https://doi.org/10.1111/biom.2012.68.issue-2

Mohler G., Carter J., & Raje R. (2018). Improving social harm indices with a modulated Hawkes process. *International Journal of Forecasting*, *34*(3), 431–439. https://doi.org/10.1016/j.ijforecast.2018.01.006

Mohler G. O., Short M. B., Brantingham P. J., Schoenberg F. P., & Tita G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, *106*(493), 100–108. https://doi.org/10.1198/jasa.2011.ap09546

Obadia T., Haneef R., & Boëlle P.-Y. (2012). The R0 package: A toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Medical Informatics and Decision Making*, *12*(1), 1–9. https://doi.org/10.1186/1472-6947-12-147

Ogata Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27. https://doi.org/10.1080/01621459.1988.10478560

Park J., Schoenberg F. P., Bertozzi A. L., & Brantingham P. J. (2021). Investigating clustering and violence interruption in gang-related violent crime data using spatial-temporal point processes with covariates. *Journal of the American Statistical Association*, *116*(536), 1674–1687. https://doi.org/10.1080/01621459.2021.1898408

Pauca V. P., Shahnaz F., Berry M. W., & Plemmons R. J. (2004). Text mining using non-negative matrix factorizations. In *Proceedings of the 2004 SIAM international conference on data mining* (pp. 452–456). SIAM.

Pecora A., Kaltwasser P. R., & Spelta A. (2016). Discovering SIFIs in interbank communities. *PLoS ONE*, *11*(12), e0167781. https://doi.org/10.1371/journal.pone.0167781

Petherick A., Kira B., Hale T., Phillips T., Webster S., Cameron-Blake E., Hallas L., Majumdar S., & Tatlow H. (2020). *Variation in government responses to COVID-19* (Blavatnik Centre for Government Working Paper). University of Oxford.

Rambhatla S., Zeighami S., Shahabi K., Shahabi C., & Liu Y. (2020). *Towards accurate spatiotemporal COVID-19 risk scores using high resolution real-world mobility data*, arXiv:2012.07283, preprint: not peer reviewed.

Reinhart A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, *33*(3), 299–318. https://doi.org/10.1214/17-STS629

Reinhart A., & Greenhouse J. (2018). Self-exciting point processes with spatial covariates: Modeling the dynamics of crime. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *67*(5), 1305–1329. https://doi.org/10.1111/rssc.12277

Rocha L. E., Thorson A. E., Lambiotte R., & Liljeros F. (2017). Respondent-driven sampling bias induced by community structure and response rates in social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *180*(1), 99–118. https://doi.org/10.1111/rssa.12180

Scala A., Flori A., Spelta A., Brugnoli E., Cinelli M., Quattrociocchi W., & Pammolli F. (2020). Time, space and social interactions: Exit mechanisms for the COVID-19 epidemics. *Scientific Reports*, *10*(1), 13764. https://doi.org/10.1038/s41598-020-70631-9

Schlosser F., Maier B. F., Jack O., Hinrichs D., Zachariae A., & Brockmann D. (2020). COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences United States of America*, *117*(52), 32883–32890. https://doi.org/10.1073/pnas.2012326117

Sha F., Saul L., & Lee D. (2002). Multiplicative updates for nonnegative quadratic programming in support vector machines. *Advances in Neural Information Processing Systems*, *15*.

Spelta A., Flori A., Pierri F., Bonaccorsi G., & Pammolli F. (2020). After the lockdown: Simulating mobility, public health and economic recovery scenarios. *Scientific Reports*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-73949-6

Spelta A., & Pagnottoni P. (2021). Mobility-based real-time economic monitoring amid the COVID-19 pandemic. *Scientific Reports*, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-92134-x

Tucker A., Vinciotti V., Liu X., & Garway-Heath D. (2005). A spatio-temporal Bayesian network classifier for understanding visual field deterioration. *Artificial Intelligence in Medicine*, *34*(2), 163–177. https://doi.org/10.1016/j.artmed.2004.07.004

Vere-Jones D., & Ozaki T. (1982). Some examples of statistical estimation applied to earthquake data. *Annals of the Institute of Statistical Mathematics*, *34*(1), 189–207. https://doi.org/10.1007/BF02481022

Vinciotti V., Augugliaro L., Abbruzzo A., & Wit E. C. (2016). Model selection for factorial Gaussian graphical models with an application to dynamic regulatory networks. *Statistical Applications in Genetics and Molecular Biology*, *15*(3), 193–212. https://doi.org/10.1515/sagmb-2014-0075

Wang C., Horby P. W., Hayden F. G., & Gao G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, *395*(10223), 470–473. https://doi.org/10.1016/S0140-6736(20)30185-9

Wang Z., Liu X., Liu Y., Liang J., & Vinciotti V. (2009). An extended Kalman filtering approach to modeling nonlinear dynamic gene regulatory networks via short gene expression time series. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *6*(3), 410–419. https://doi.org/10.1109/TCBB.2009.5

Wold S., Esbensen K., & Geladi P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1–3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

Wright M. N., Kusumastuti S., Mortensen L. H., Westendorp R. G., & Gerds T. A. (2021). Personalised need of care in an ageing society: The making of a prediction tool based on register data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *181*(4), 1199–1219. https://doi.org/10.1111/rssa.v184.4

Xiong C., Hu S., Yang M., Luo W., & Zhang L. (2020). Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. *Proceedings of the National Academy of Sciences*, *117*(44), 27087–27089. https://doi.org/10.1073/pnas.2010836117

Zhang J., Litvinova M., Liang Y., Wang Y., Wang W., Zhao S., Wu Q., Merler S., Viboud C., Vespignani A., Ajelli M., & Yu H. (2020). Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*, *368*(6498), 1481–1486. https://doi.org/10.1126/science.abb8001

Zhu S., Bukharin A., Xie L., Santillana M., Yang S., & Xie Y. (2020). *High-resolution spatio-temporal model for county-level COVID-19 activity in the US*, arXiv:2009.07356, preprint: not peer reviewed.