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.
Modeling infectious disease dynamics: Integrating contact tracing-based stochastic compartment and spatio-temporal risk models
Mateen Mahmood a, André Victor Ribeiro Amaral a,∗, Jorge Mateu b, Paula Moraga a

Computer, Electrical and Mathematical Sciences and Engineering Division. King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
Department of Mathematics, Universitat Jaume I, Spain

ARTICLE INFO

Article history:
Received 13 January 2022
Received in revised form 15 June 2022
Accepted 15 July 2022
Available online 9 August 2022

Keywords:
Compartment modeling
Contact tracing
Infectious diseases
Pedestrian mobility
Spatio-temporal modeling

ABSTRACT

Major infectious diseases such as COVID-19 have a significant impact on population lives and put enormous pressure on healthcare systems globally. Strong interventions, such as lockdowns and social distancing measures, imposed to prevent these diseases from spreading, may also negatively impact society, leading to jobs losses, mental health problems, and increased inequalities, making crucial the prioritization of riskier areas when applying these protocols. The modeling of mobility data derived from contact-tracing data can be used to forecast infectious trajectories and help design strategies for prevention and control. In this work, we propose a new spatial-stochastic model that allows us to characterize the temporally varying spatial risk better than existing methods. We demonstrate the use of the proposed model by simulating an epidemic in the city of Valencia, Spain, and comparing it with a contact tracing-based stochastic compartment reference model. The results show that, by accounting for the spatial risk values in the model, the peak of infected individuals, as well as the overall number of infected cases, are reduced. Therefore, adding a spatial risk component into compartment models may give finer control over the epidemic dynamics, which might help the people in charge to make better decisions.

© 2022 Elsevier B.V. All rights reserved.

∗ Corresponding author.
E-mail addresses: muhammad.mahmood.1@kaust.edu.sa (M. Mahmood), andre.ribeiroamaral@kaust.edu.sa (A.V.R. Amaral), mateu@uji.es (J. Mateu), paula.moraga@kaust.edu.sa (P. Moraga).

https://doi.org/10.1016/j.spasta.2022.100691
2211-6753/© 2022 Elsevier B.V. All rights reserved.
1. Introduction

The current situation of coronavirus disease (COVID-19) and its ongoing waves and emerging variants not only endanger public health but also pose a significant threat to society's general stability (Moraga et al., 2020; Al-Salem et al., 2021). This long existence of a pandemic scenario cannot be confronted with permanent lockdowns and social distancing restrictions as it generates secondary problems such as unemployment, economic decline, and overall harm to social function (Kang et al., 2020). To tackle these problems, policy-makers ease restrictions, which allow the public to return to their normal routine. However, there is a critical need to monitor populations' interactions because only through strict surveillance of such human behavior, contacts can be traced and prevention measures can be implemented in an efficient way to avoid recurring disease outbreaks while minimizing disruption of normal life.

Detection and control of these population interactions are in need of Digital Contact Tracing (DCT) (Anglemyer et al., 2020), which can quickly identify prior contacts of a detected infected individual to identify the exposed ones. Although tracking interactions of these exposed individuals is demanding, their identification can effectively restrict the pandemic growth (Müller et al., 2020). DCT depends on information of fine-scale individual-level mobility, and collected data can be modeled to simulate a disease outbreak scenario and identify how an infection propagates in space and time.

Despite immense progress of research done in stochastic models for disease spread, the inclusion of the spatial component in such models is generally new. This latest introduction of spatial aspect is made available due to the recent advancement of location-based services by providing geolocation of individual-level mobility and their associated interactions to a very fine-scale (accuracy of less than a meter) (Zheng, 2015). In spite of this improvement in geospatial technologies and development in computing procedures, a major bottleneck is the highly invasive nature of these continuous recordings of individual-level mobility, which makes these data seldom publicly available (Reichert et al., 2020). Hence, infectious diseases modeling approaches that include a spatial component for an individual level are still difficult to be employed.

This scarcity of spatial capability in terms of infectious disease modeling is exploited by long-standing pandemics, such as COVID-19. Such a situation requires extraordinary attention of stakeholders, especially with the surface and aerosol stability of infectious diseases (Van Doremalen et al., 2020; Simmerman et al., 2010). Preparing for an extreme scenario like this requires proactive measures in terms of having a readily available solution (e.g., accounting for spatial information) that can be used as required (Mahmood et al., 2021). At the same time, when the entire humanity is at stake, the requirement of saving lives becomes more critical than ensuring privacy, and the use of these continuous recordings of individual-level mobility is acceptable (Benreguia et al., 2020). Similarly, Prodanov (2021) suggests that innovations in epidemiological tools are of significant utility to simulate epidemic scenarios and different control strategies.

One such novel tool for Contextual Contact Tracing (CCT) was developed by Mahmood et al. (2021). CCT focuses on the consideration of spatial risk of a contact's geographical location to influence the overall outcome of a contact; the idea being that some areas are more disease inducing than others. Mahmood et al. (2021) implemented a SIR-based compartment model (Kermack and McKendrick, 1927) with two additional compartments related to quarantine (Hernández-Orallo et al., 2020). The work was concerned with the tracking of infectious trajectories (pedestrian mobility) and the investigation of the stakeholders' requirements of identification of (i) high-risk areas, and (ii) individuals to be quarantined.

The CCT tool integrated the SIR-based compartment model and the spatial risk, where risk was defined based on (i) mobility trajectories (susceptible and infected individuals) and (ii) location of contacts. This spatial risk was then introduced into the SIR model as an intensity of a contact which directly affected the transmission rate.

In this study, we extend the previous CCT work in three aspects to address some of their limitations. Firstly, through the use of a simulated, yet more realistic data set, with a higher count of individuals. Secondly, with an improved frequency of computing contacts and spatial risk. And lastly, with a statistically comprehensive and supervised process for identification of spatial risk,
rather than through an unsupervised method as Self Organizing Maps (SOM) (Hulle and Marc, 2012), followed in previous works. With our supervised method we have direct control, through our statistical model underlying the spatial risk, of the way the risk is behaving, its variability and statistical distribution—aspects that were not considered in the previous unsupervised method. We demonstrate the functionality of this newly proposed CCT extension by simulating different scenarios of disease spreading in the city of Valencia, Spain. To do this, we use spatial risk values, estimated using a Poisson process for the number of infected-susceptible contacts, and update the contacts counting accordingly. We compare the results obtained with the newly proposed model and a simpler and previously introduced method (Hernández-Orallo et al., 2020), and show that the introduction of a spatial risk component imposes finer control over the epidemic parameters, which may help policy-makers to make better decisions. In that way, this new extended method may contribute to the capability of empowering the analyst to associate spatial risk into individual-level SIR-based compartment modeling.

The remainder of this paper is as follows. Section 2 describes (i) the simple SIR-based compartment model, (ii) the previously implemented spatial-SIR model, and (iii) the newly proposed enhancement in the form of a spatio-temporal-SIR model. Section 3 discusses the simulation procedure along with an introduction of the used data sets. This section also presents the results of mobility simulation as well as of the different disease outbreak scenarios. Section 4 stresses our approach limitations and presents conclusions from this work with some recommendations for the future.

2. Methodology

2.1. Base and spatial-SIR modeling

In order to extend previous approaches on how to model the dynamics of an epidemic, we first review how the non-spatial stochastic SIR models work, as well as the stochastic SIR model with spatial components.

2.1.1. Base-SIR modeling

Here, we consider the base-SIR model as the one discussed in Hernández-Orallo et al. (2020). These authors work with an event-based stochastic approach (Gillespie’s Method) (Gillespie, 1977) for the SIR model, originally presented by Kermack and McKendrick (1927), in which they also incorporate the idea of contact tracing. An exhaustive description of this approach is presented by Mahmood et al. (2021).

In this base-SIR model, assuming a closed environment (i.e., there are no births, deaths or migrations), there are five different infectious states; that is, compartments an individual can be during an epidemic, which represent the states of Susceptible (S), Infected (I), Recovered (R), Quarantine Susceptible (Qₐ), and Quarantine Infected (Qᵢ). Also, an individual can be transferred from one compartment to another in seven different ways, namely: \( S \to I \), \( S \to Qₐ \), \( S \to Qᵢ \), \( I \to Qᵢ \), \( Qₐ \to S \), \( I \to R \), and \( Qᵢ \to R \). Fig. 1 summarizes how each compartment relates to others. At the beginning of a pandemic, it is assumed that a (typically small) portion of the population is infected, so the dynamics of the pandemic may evolve as people move across compartments.

In this event-based approach, the key task is to determine the rate of each event in which an individual is transferred from one compartment to another. These rates mainly rely on three types of parameters: (i) epidemic-specific, (ii) contacts-based, and (iii) simulation-related.

Epidemic parameters include the transmission rate \( \beta \) (which depends on the probability of transmission \( p \)) and the recovery rate \( \gamma \). Also, the basic reproductive ratio \( R_0 \), which can be defined as \( \beta / \gamma \), represents the expected count of cases directly affected by a single case. Another disease specific parameter is the time in quarantine \( t_Q \), which is the duration an individual will stay in the quarantine compartment after being quarantined. The rates are also influenced by the contact tracing information collected in form of identified contacts among individuals. Regarding the contacts-based parameters, average degree \( \kappa \), infectious contacts \( K_i \) and prior contacts of an individual...
$C_i$ are all considered to compute the rate of each event for each individual in the population. Lastly, simulation related parameters include the rate of detection $\delta$ and the tracing efficiency $q'$.  

More specifically, the interactions in the population may be represented by a network graph in which the nodes represent the individuals and the edges represent their contacts (Enright and Kao, 2018). However, we can also choose to see it as a matrix, say $G$. In our case, for $N$ individuals, $G$ will be a $(N \times N)$ matrix in which $G_{ij} = 1$, if there exists a contact between individuals $i$ and $j$, and $G_{ij} = 0$, otherwise. In that way, note that $G$ is symmetric, that is, $G_{ij} = G_{ji}$, for every pair $(i, j)$. Also, $G_{ii} = 0$, for all $i \in \{1, \ldots, N\}$. Similarly, $G(t)$ represents the existing contacts in a time-window $t$.

Within this framework, and as described by Mahmood et al. (2021), the average degree $\kappa$, computed for some time period $T$, is given by

$$\kappa = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{j=1}^{T} K_i(t_j) \right),$$

where $K_i(t)$ represents the count of contacts of a person $i$ with other individuals in $G(t)$. However, if we want to have a measure of only infectious contacts, we can define the infectious contacts $K_i$ for some time-window $t$ as

$$K_i(t) = \sum_{j=1}^{N} G_{ij}(t) I_j(t),$$

such that $I_j(t)$ is an indicator function for the event in which the individual $j$ can infect others. And as a way to track back the contacts, we can define, for a time-window $t$ and time period $\Delta$, the prior contacts of an individual $C_i$ as

$$C_i(t, \Delta) = \begin{cases} 1 \text{ if } \sum_{j=1}^{N} (\max_{\tau \in [t-\Delta, t]} G_{ij}(\tau)) D_j(t) > 0 \text{ \ otherwise,} \end{cases}$$

such that $D_j(t)$ is an indicator function for the event in which the individual $j$ is infected and traced. Also, for $q$ corresponding to the fraction of traced individuals being quarantined, we can define the tracing efficiency as $q' = q \cdot \tau_r$, where $\tau_r^{-1}$ corresponds to the average tracing time.

Therefore, as in Table 1, the aforementioned parameters influence the rate of each event for each specific person, and then a stochastic element determines the next event. Here, stochastic elements include (i) the type of next event and the person to whom it will happen, and (ii) the time of next event. This process is followed for each specific person in each round of the modeling process until the end of epidemic (or simulation) period.
Table 1
Rate equations for events in the base-SIR model, as in Mahmood et al. (2021).

| Event      | Rate equation                                                                 |
|------------|--------------------------------------------------------------------------------|
| $S \rightarrow I$ | $(1 - C_i(t, \triangle)) \cdot b \cdot \kappa_i(t)$                           |
| $S \rightarrow Q_s$ | $q' \cdot C_i(t, \triangle) \cdot (1 - (b \cdot \kappa_i(t)))$               |
| $S \rightarrow Q_i$ | $q' \cdot C_i(t, \triangle) \cdot b \cdot \kappa_i(t)$                      |
| $I \rightarrow Q_s$ | $\delta$                                                                      |
| $I \rightarrow R$ | $\gamma$                                                                     |
| $Q_i \rightarrow R$ | $\tau_0$                                                                    |

2.1.2. Spatial-SIR modeling

Extending the base-SIR model, Mahmood et al. (2021) incorporate the idea of temporally varying spatial risk. This approach consists in computing a risk value for each piece of the total considered area, which will play an important role in how the parameters from the model described in Section 2.1.1 are updated in each round of this iterative process. Following the procedure described by Mahmood et al. (2021), the risk values are determined for each cell from a regular lattice placed over the desired area, and the model parameters are daily updated based on this standardized computed risk. Also, this spatial risk mainly relies on how infected and susceptible people move across the cells; in particular, the authors consider the spent time by an infected person in a specific cell, the number of people in that sub-region, and the number of contacts (whether or not involving an infected individual) in the analyzed rectangle. These different sources of information are summarized in a unique grid by applying an unsupervised clustering technique named Self Organizing Map (SOM) (Hulle and Marc, 2012).

Once all quantities in a time window ($t - 1$) are determined, and based on the computed risk values, the intensity of a contact is updated based on the contacts locations in the time window $t$. In this case, the intensity can be seen as a representation of probable risk for the susceptible individual involved in the infectious contact. These quantities are computed for each contact (involving an infectious individual) and are to be used while computing rates for the next event. This revision only alters the rates of events associated to the susceptible individuals. Then, this updated parameter is used to compute the new rates of the base-SIR model, which will dictate how the pandemic evolves based on the identification (which may lead to public policies preventing people from going to those regions) of risky areas.

2.2. New spatio-temporal-SIR model

In this paper, we focus on improving the previous work of a spatial-SIR model by (i) proposing a new way to simulate pedestrian mobility data, (ii) increasing the modeling update frequency of the epidemic’s dynamic, and (iii) improving the way the spatial risk is computed by implementing a supervised process in the form of a spatial stochastic model.

2.2.1. Simulation of pedestrian mobility

Due to the difficulty in obtaining real-world mobility data, we opted for simulating pedestrians movement across a selected region. This simulation process is performed using the software named Simulation of Urban Mobility (SUMO) (Lopez et al., 2018). For a given region, we are able to generate map-matched trajectory data for people walking on streets during some specific time-windows.

From the generated trajectory data, we just have to extract the quantities of interest, such as the number of contacts or duration of each contact. And based on these processed data sets, we finally include the epidemic dynamics, setting the initial number of infected people, computing the SIR model parameters, etc.

To obtain the infrastructure data about the selected region, we used another tool named OpenStreetMap (OSM) (Haklay and Weber, 2008). Using an Application Programming Interface
(API) that connects to the OSM platform, we can download an .osm file with data about the streets (and other information, such as buildings) of the selected region that could be read by SUMO and used to simulate the pedestrian movement, as described.

This is a novelty compared to Mahmood et al. (2021), since they used a data set extracted from Tsai and Chan (2015), in which the movements of 115 students on the campus of National Chengchi University were recorded during a 15-day window. As these data sets are difficult to obtain, the ability to simulate (in a reasonably realistic way) such trajectory data is a huge advantage in fitting contact tracing models.

2.2.2. Spatial risk update frequency

Another main difference from Mahmood et al. (2021) is how often we are now updating the risk values and, therefore, the epidemic dynamics. Previously, contacts and the spatial risk were computed on a daily basis, as it is common to have a temporal frequency of a day in epidemiological studies (Hernández-Orallo et al., 2020).

From a contact tracing perspective, this is typically acceptable, however, for the spatial risk computations, per-day spatial risk tends to ignore how temporally varying spatial risk evolves during the day. Hence, we integrated here the concept of temporal windows, which refers to a four-hour window as the frequency of computations.

If we allow people to move around from 8 am to 8 pm, we will have now three windows of four hours each. These time windows could have been shorter, though. However, this limitation is not only concerned with the computational power, but also with the data volume in each time window. Note that if the windows are too short, then it may be more difficult to identify a contact. In our case, a contact happens only if two individuals have been less than one meter apart for at least 60 s, as in Hernández-Orallo et al. (2020).

2.2.3. Spatio-temporal modeling of the spatial risk

As introduced in Section 2.1.2, recall that we are interested in studying the risk across a given region; however, instead of considering a spatially continuous domain, and as a way to discretize the studied area, we divide the region of interest into cells determined by a non-overlapping regular grid. The cells cannot be too large so that we do not lose granularity in the results, but also cannot be too small so that we observe sufficiently many contacts in each cell. Therefore, in Section 3, as a compromise between granularity and the number of observed contacts, we divide the studied region into a regular lattice with each cell size of approximately 90 m × 100 m, which is the same as partitioning the total study area of 2,000,000 square meters into a total of 230 cells.

The Relative Risk (RR) for a cell \( j \) will be used to update the \( G \) matrix in a way that the entries of \( G \) will be replaced by the RR, value, if the two corresponding individuals (for a \( N \times N \) matrix, where \( N \) is the number of individuals, we are referring to the two individuals indexed by the row and column of \( G \)) have contacted each other in that particular cell. This new \( G \) matrix will also be used to update the infectious contacts \( K_i \) and the prior contacts of an individual \( C_i \), which finally update the first three rate equations from Table 1. In a nutshell, information about contact and other relevant covariates will be aggregated and modeled accordingly aiming to estimate the RR, as represented in Fig. 2.

Therefore, in order to determine how risky a particular cell is, one could compute the Standardized Contact Ratio (SCR), that is, the ratio of the observed counts of an event of interest (an infected-susceptible contact) and the expected counts of the same event (Moraga, 2019). In particular, for each cell \( j \),

\[
SCR_j = \frac{Y_j}{E_j},
\]

where \( Y_j \) represents the number of contacts between one infected and one susceptible person inside cell \( j \), and \( E_j \) represents the number of contacts of the same type that one would expect to observe in \( j \), if population from this cell behaved in the same way as the standard population. More specifically, \( Y_j = c_j^A \) and \( E_j = (\text{pop}_j \cdot \theta_{all}) \), such that

\[
\theta_{all} = \frac{\sum_{all} Y_j}{\sum_{all} \text{pop}_j},
\]
Fig. 2. Relative Risk (RR) modeling strategy for aggregated contact-tracing data and important covariates.

where \( \text{pop}_j = (c_{ij}^{is} + c_{ij}^{ii} + c_{ij}^{ss}) \), \( c_{ij}^{is} \) is the number of contacts between an infected person and a susceptible one, \( c_{ij}^{ii} \) the number of contacts between two infected people, and \( c_{ij}^{ss} \) the number of contacts between two susceptible people.

However, if the number of reported occurrences of the event of interest in a particular cell, say \( j \), is not sufficiently large, then \( E_j \) may become less reliable (Moraga, 2018). To overcome this issue, one may choose to model \( Y_j \) according to a process such that \( Y_j \) follows a Poisson distribution with mean \( (E_j \cdot \theta_j) \), where \( \theta_j \) represents the Relative Risk in cell \( j \), \( \forall j \). Thus,

\[
Y_j \sim \text{Poisson}(E_j \cdot \theta_j), \quad \text{for } j = 1, \ldots, n
\]  

(1)

where \( \log(\theta_j) = \beta_0 + \beta_1 \cdot x_1 + \cdots + \beta_m \cdot x_m + u_j \), such that \( x_1, \ldots, x_m \) are covariates and \( u_j \) is a random effect. In particular, we can let \( u_j \) represent the unstructured exchangeable random effect that models uncorrelated noise; i.e., \( u \sim \text{Normal}(0, \sigma_u^2 I_n) \), where \( I_n \) is the \( n \)-by-\( n \) identity matrix. Hence, the estimated \( \hat{\theta}_j \), \( \forall j \), will be given by \( \hat{\theta}_j \).

However, for Model (1), notice that different structures for \( \log(\theta_j) \) can be proposed; for instance, one can add a random effect \( v_j \) that models the spatial dependence between relative risks. This is usually denoted as Besag–York–Mollié (BYM) model (Besag et al., 1991), and in this case, \( v_j \) follows an Intrinsic Conditional Autoregressive (ICAR) distribution that smooths the data according to a certain neighborhood structure (Moraga, 2019). The BYM model presents some difficulties, though, as discussed in Banerjee et al. (2003); for instance, the non-identifiability of \( v \), and the problem in setting sensible priors on the precision of the random effects (Morris et al., 2019). In this regard, a model with different parameterization, namely BYM2 (Riebler et al., 2016), can be employed. In the BYM2 model, and regarding the two random effects, a single precision parameter is assigned to the combined components, and a mixing parameter determines the amount of spatial and non-spatial variation (Morris et al., 2019).

Finally, as we may not observe contacts in many different cells (imagine a real-world scenario in which people barely meet each other in peripheral regions, whilst in central or touristic places they are constantly contacting other individuals), there is no reason to believe that these regions are more or less risky than the overall considered area. Therefore, if \( \text{pop}_j = 0 \) for some \( j \), then we set \( \text{RR}_j \) to 1. This is also the reason why we decided not to account for the spatially structured random effects \( v_j \) in Model (1), that is, since we do not have enough data for most cells, the fitting procedure for a model with dependent random effects might result into numerical instabilities.

Once we have \( \text{RR}_j, \forall j \), estimated, and as described in Section 2.1.1, the intensity of a contact, which affects how the susceptible-related parameters in the next time-window are changed, is updated. This procedure carries the spatial information from a time-window to the next one and updates the epidemic dynamic accordingly.

Also, although Model (1) can be seen as a Generalized Linear Mixed Model (GLMM) and, therefore, can be fitted using many different computational tools, the addition of other random
effects, as in the BYM2 model, may impose additional challenges if one decides not to take advantage of some already-implemented computational solutions. For instance, when fitting the BYM2 model, one needs to scale the model so that the geometric mean of the marginal variances is one (Riebler et al., 2016). Hence, we chose to estimate RR using the Integrated Nested Laplace Approximation (INLA) method (Rue et al., 2009), which can deal with spatio-temporal models and large data sets. In R, INLA can be easily performed by using the R-INLA package (Rue et al., 2009). In Section 4, we will discuss some drawbacks (e.g., low coverage in out-of-sample predictions) of using R-INLA and briefly compare it with other solutions.

3. Simulation

In this section, we demonstrate how mobility data for pedestrians can be simulated in a reasonably realistic way, and show how the base-SIR model and the newly proposed model, described in Section 2.2, can be used to simulate an epidemic. The results for the two approaches will be compared, and we stress the differences obtained when accounting for the estimated Relative Risk into our model.

3.1. Simulated data

As described in Section 2.2.1, data was simulated using SUMO (and OSM). For this paper, aiming to obtain inferences in a real setting, we have simulated the movement of 1000 pedestrians across Valencia city, Spain, for 15 days. The number of days used in the mobility simulation could have been larger, but then the pedestrians’ movement data generation would have become unnecessarily computationally expensive. Fig. 3 shows the map of the studied region with the overlapped cells. Besides, as a period of 15 days is too short for an epidemic to occur, we self-concatenated these 15 days data set multiple times to have a data set for a longer duration (150 days). Also, we defined a contact as the situation in which two individuals have been less than 1 m apart for at least 60 s (Hernández-Orallo et al., 2020).
Simulation was performed with the specifications that (i) each individual has a single trajectory in a day for a duration of two hours (which can be in between 8am to 8pm) and (ii) starting and ending points of each trajectory were chosen at random within the study area (but preferably in the outer periphery—fringe factor in SUMO). Furthermore, the region was divided into 230 cells (as a consequence of the total area size) and, aiming at increasing the number of contacts, we decided to set 10% of the cells, i.e., 23 cells (the same as the cells with the highest building densities) as high priority; that is, the chance of people visiting these cells along their day of walking is very high. Fig. 4 shows the maps for the selected 23 cells (left) and the one for the cells where at least one contact has occurred (right).

As we were expecting, from Fig. 4, one can see that the cells with higher buildings density are similar to the cells where people contacted each other. This is important since we want to guarantee that a sufficient number of contacts occurs during each temporal-window.

3.2. Implementation

With the goal of simulating the evolution of an epidemic in a particular city, in our case Valencia, we adopted the procedure described in Section 2.2.3 to estimate the Relative Risks for each cell $j$, $j = 1, 2, \ldots, 230$, in each step of the previously explained iterative process. In particular,

$$Y_j \sim \text{Poisson}(E_j \cdot \theta_j), \text{ for } j = 1, \ldots, 230,$$

where $\log(\theta_j) = \beta_0 + \beta_1 \cdot x_{ij} + u_j$, such that $x_{ij}$ represents the number of buildings in cell $j$ and $u_j \sim \text{Normal}(0, \sigma_u^2)$.

Also, as the model was fitted using the R-INLA package, we used the default prior distributions for the parameters and hyperparameters; that is, $\beta_0 \sim \text{Normal}(0, (1/\tau_0))$, for $\tau_0 = 0$, $\beta_1 \sim \text{Normal}(0, (1/\tau_1))$, for $\tau_1 = 0.001$, and $\log(1/\sigma_u^2) \sim \text{LogGamma}(1.5 \times 10^{-5})$. Then, the estimated Relative Risks were obtained based on the computed mean of the posterior sample from the corresponding parameters.

Additionally, following Hernández-Orallo et al. (2020), we set the initial number of infected people to 150 (out of the 1000 pedestrians), transmission rate to 1/5, recovery rate to 1/15 (and therefore, $R_0 = 3$), time in quarantine to 14 days, rate of detection to 1/20, and tracing efficiency to 1/10. Some of these parameters are disease-specific and, therefore, may vary depending on the study focus. Other parameters will be further investigated in the sensitivity analysis. Also, in order to have multiple realizations of the simulated epidemic, we repeated this procedure 10 times, so that we can summarize the results by averaging over all repetitions. However, as these process realizations are not guaranteed to end in the same time point, we will accommodate for these differences by extrapolating the values up to the longest series before computing their mean. The corresponding results are shown and discussed in the next subsection.
3.3. Results

In order to compare how the contact tracing and temporal components, detailed in Section 2.2, affect the way a pandemic evolves, we fit, for the same mobility data set, the base-SIR and spatio-temporal-SIR models. That is, for the approaches described in Sections 2.1.1 and 2.2, we fit the models considering the implementation details, if applicable, presented in Section 3.2. Also, regarding the mobility data, Fig. 5 shows the trajectory of 10 randomly chosen people in “day 1” of simulation.

3.3.1. Results for the base-SIR model

Considering the model introduced in Section 2.1.1, we can plot the different realizations of the simulated epidemic. In particular, we follow the number of individuals in each of the five compartments (S, I, R, Q_S, and Q_I) over the days. In this way, we can track how the number of infected cases increases (or decreases) in the different stages of an epidemic, described according to a model that does not account for a spatio-temporal component. Fig. 6 shows the results of the evolution of the epidemic for the simulated data, as well as a summary (computed by averaging over all the realizations) of the corresponding individual evolution.

From Fig. 6, one can see that the number of infected individuals, set to 150 at the beginning of the simulation, as described in Section 3.2, rapidly increases in the very first days of the epidemic. At the same time, as contacts occur, pedestrians (both susceptible and infected) started to be quarantined. Also, the number of recovered cases increased at a fast rate in the first third of the studied time period. However, as people started to be moved to other compartments, e.g., Q_S and Q_I, the number of infected people decreased over the days. As a consequence, from the second third on, fewer infected people were moved to quarantine and the number of recovered cases increased up to its maximum (around 900 individuals). Finally, with no more new cases of infectious, the remaining quarantined people were assigned to the susceptible or recovered compartments, and the epidemic ended.

3.3.2. Results for the spatio-temporal-SIR model

Following what has been presented in Section 3.3.1, we now simulate the dynamic of an epidemic using the ideas introduced in Section 2.2. However, differently from before, we also have to estimate the Relative Risk for each cell and for each window. Fig. 7 shows the estimated RR for
Fig. 6. Number of individuals in each compartment (S, I, R, Q₅, and Q₄) over the days using the base-SIR model. Light lines represent the 10 realizations of the simulated epidemic, and the bold line represents the average number of individual, as described in Section 3.2.

Fig. 7. Estimated Relative Risks for all 230 cells in the second window of days 1, 2, 3, and 4.

the first window in days 1, 2, 3, and 4, for the first realization of the simulated process. Recall that, if \( pop_j = 0 \), for some cell \( j \), then we set \( RR_j \) to 1. Also, depending on the RR values computed for a time window \( (t - 1) \), the parameters that control the epidemic dynamics will be updated accordingly, which will change the conditions under which the spread of the disease in the time window \( t \) is simulated.

Besides, we can also analyze how the pandemic evolves over time, i.e., we can plot the number of people in each of the five compartments for each time window. Here, recall that differently from the analysis that we have done in Section 3.3.1, changes take place in a finer scale, which plays an important role in the implementation of more efficient preventive measures by the policy-makers. Also, as before, the obtained realizations from the simulated process are summarized through their average. Fig. 8 shows such plots.
Fig. 8. Number of individuals in each compartment (S, I, R, Q_S, and Q_I) over the days using the spatio-temporal-SIR model. Light lines represent the 10 realizations of the simulated epidemic, and bold line represents the average number of individual, as described in Section 3.2.

From Fig. 8, we first observe that the peak of infected people is lower (approximately 250 individuals) in comparison with the results shown in Section 3.3.1. This may be explained by the fact that, mainly due to the computed Relative Risks, people are being quarantined (both Q_S and Q_I) more often; that is, if a person has visited a particular cell in (t − 1) and we have observed infected-susceptible contacts on that cell during the same time-window, then this person (infected or susceptible) has a higher chance to be quarantined in t. Also, since the occurrence of newly infected cases was distributed over a longer period, the total number of infected individuals, which can be obtained by analyzing the Infected (I), Recovered (R), and Quarantine Infected (Q_I) compartment in the last simulated day, was also reduced. Therefore, accounting for the estimated Relative Risk has an impact on the rate of individuals being quarantined, which also impacts the total number of infected cases.

3.3.3. Sensitivity analysis

Aiming to verify if the obtained results hold for different scenarios, we will repeat the procedure described in Sections 3.3.1 and 3.3.2 for different sets of parameters. In particular, we will vary the initial number of infected individuals (150 and 250) and the tracing efficiency (0.1, 0.5, and 0.9). As a remark, we also ran the epidemics varying the number of days in the mobility data generation and the number of seconds that define a contact. For the former, the simulated S, I, R, Q_S, and Q_I did not change much (recall that the number of simulated days for mobility data is being self-concatenated, and as long as the mobility dynamics are being reasonably modified over time, the results are expected to be very similar), and for the latter, the overall number of infected individuals increase as the number of seconds that defines a contact decreases; however, the difference is almost negligible for similar time intervals (e.g., 30, 45, and 60 s). Based on our simulated mobility data, most interactions that last more than 30 s also last more than 60 s. Of course, if we set this threshold to much higher values, for instance, 600 s, the epidemic dynamics changes; nonetheless, defining a contact for only interactions that last that long might be unreasonable. To see why, just imagine classifying a contact as possibly infectious only after the tenth minute of close-distance interaction.

From Table 2, recall that scenario 01 is the same as the one discussed in Sections 3.3.1 and 3.3.2. Also, Figs. 9, 10, 11, 12, and 13 in Appendix show the results for scenarios 02, 03, 04, 05, and 06, respectively. From these plots, we can see that when increasing the initial number of infected individuals, the peak in the infectious curve also increases; however, the overall epidemic dynamics for each of the base-SIR and spatio-temporal-SIR models remain the same. Regarding the changing in the tracing efficiency, the results are substantially different for the two models. As we increase q′, that is, the tracing efficiency, the number of individuals in the Q_S compartment gets drastically larger, and the peak in the infected individuals decreases by a small percentage. This is due to the fact that, by setting the tracing efficiency to larger values, we are now able to better identify
Table 2
Different simulated scenarios based on the initial number of infected individuals (2nd column) and the tracing efficiency (3rd column).

| Scenario | Initial # of infected ind. | Tracing efficiency |
|----------|-----------------------------|--------------------|
| 01       | 150                         | 0.1                |
| 02       | 150                         | 0.5                |
| 03       | 150                         | 0.9                |
| 04       | 250                         | 0.1                |
| 05       | 250                         | 0.5                |
| 06       | 250                         | 0.9                |

infectious contacts (i.e., contacts between a susceptible and an infected individual) and send the corresponding person to the quarantine.

In summary, although the specific scenarios change when we use different set of parameters for the simulation procedure, the main difference between the base-SIR and the spatio-temporal-SIR models is kept the same: that is, by incorporating the Relative Risk into our modeling approach, we can decrease the overall number of infected individuals by quarantining people who have been into an infectious contact.

4. Conclusion

In this paper, we have proposed a new spatio-temporal-SIR procedure that models the Relative Risk in each subarea of the study region and updates the epidemic dynamics based on the observed infectious contacts. We did this by extending the work from Mahmood et al. (2021) in three major aspects: (i) describing how to simulate mobility data using SUMO, (ii) introducing the idea of temporal-window, which allows to have finer control over the epidemic dynamics, and (iii) proposing a supervised procedure for the computation and update of the spatial risk. In this regard, as contact tracing data might be difficult to obtain, we have demonstrated the use of this new method by means of simulation. We think this method may be further tested as contact tracing apps evolve and data are made available—although privacy may always be an issue regarding collecting individuals’ movements.

These changes improve previous approaches since (i) allow researchers to obtain reasonably realistic mobility data, (ii) empower decision-makers with more often updated information about risky areas, and (iii) establish a methodology to aggregate information about contact tracing data and transform them into useful insights.

This type of work is important as it provides easy-to-interpret data for policy-makers. Having close to real-time knowledge about risky areas in, say a city, allows politicians to propose policies that will prevent individuals from walking around dangerous spots during a specific time-window. This is crucial since one way to slow down the spreading of a disease is to guide individuals to avoid in-person interactions. However, as these restrictions also impact other areas of the citizens’ lives, e.g., economy, work, leisure, etc., it is necessary to have frequently updated information, so that policies are not unnecessarily restrictive and are more efficient in focusing on areas that matter.

The obtained results, presented along Section 3.3, show that the introduction of the estimated Relative Risk values into the model may contribute to the reduction of the epidemic peak (when we compare the approach described in Section 2.2 with the one described in 2.1.1). Also, as people are more likely to be quarantined, the overall number of infected cases is expected to be lower. Of course, quarantining more individuals may cause other consequences to society, but an important detail comes from the fact that we are not doing this by random; instead, we are focusing on pedestrians who visited risky areas. Thus, by preventing individuals from going to those places, we can also reduce the number of infected cases without overload the quarantine compartment.

Still, although the results seem positive, our proposed framework (and implementation procedure) has limitations. Firstly, about Model (1), recall that we could have added a structured random effect $v_j$ that accounts for the spatial dependence among the cells; however, when we do not have enough data for most locations—as may happen in scenarios where either the studied region...
is too large and people are not evenly distributed across the observed area or the population is too small (and, therefore, there are less contacts)—the estimation procedure may not be doable in practice. In that case, we only accounted for the unstructured random effects that models the uncorrelated noise. Secondly, as discussed by Sahu (2022), R–INLA gives a low coverage in out-of-sample predictions. As an alternative, CARBayes (Lee, 2013) can be used when fitting, among others, a BYM or a Leroux (Leroux et al., 2000) model. Also, we can fit the BYM2 model using RStan (Stan Development Team, 2021; Morris et al., 2019). Both CARBayes and RStan are based on Markov chain Monte Carlo (MCMC) methods.

In summary, contact tracing data can be used as an important tool in health surveillance. Keeping track of individuals movements (and interactions) may help decision-makers understand the population contact patterns and, based on this acquired knowledge, propose policies that may prevent individuals from visiting risky areas and contribute to the spread of the analyzed infectious disease. Also, constantly updated information about risk in a given region plays an important role in preventing people from visiting dangerous spots without jeopardizing other aspects of their lives since proposed policies may be less restrictive, only forbidding access to certain areas during some parts of the day.

**Appendix**

Simulation results for the scenarios 02, 03, 04, 05, and 06 (Figs. 9, 10, 11, 12, and 13, respectively), as defined in Table 2, for both base-SIR and spatio-temporal-SIR models.
Fig. 10. Number of individuals in each compartment (S, I, R, Q_S, and Q_I) over the days using the base-SIR (top) and the spatio-temporal-SIR (bottom) models. Light lines represent the 10 realizations of the simulated epidemic, and the bold line represents the average number of individual. Simulation corresponds to scenario 03 from Table 2.
Fig. 11. Number of individuals in each compartment (S, I, R, Q_s, and Q_i) over the days using the base-SIR (top) and the spatio-temporal-SIR (bottom) models. Light lines represent the 10 realizations of the simulated epidemic, and the bold line represents the average number of individual. Simulation corresponds to scenario 04 from Table 2.
Fig. 12. Number of individuals in each compartment (S, I, R, Qₜ, and Qᵢ) over the days using the base-SIR (top) and the spatio-temporal-SIR (bottom) models. Light lines represent the 10 realizations of the simulated epidemic, and the bold line represents the average number of individual. Simulation corresponds to scenario 05 from Table 2.
Fig. 13. Number of individuals in each compartment (S, I, R, Qₜ, and Qᵢ) over the days using the base-SIR (top) and the spatio-temporal-SIR (bottom) models. Light lines represent the 10 realizations of the simulated epidemic, and the bold line represents the average number of individual. Simulation corresponds to scenario 06 from Table 2.
References

Al-Salem, Waleed, Moraga, Paula, Ghazi, Hani, Madad, Syra, Hotez, Peter J. 2021. The emergence and transmission of COVID-19 in European countries, 2019–2020: a comprehensive review of timelines, cases and containment. Int. Health.

Anglemyer, Andrew, Moore, Therese HM, Parker, Lisa, Chambers, Timothly, Brady, Alice, Chiu, Kellia, Parry, Matthew, Wilczynska, Magdalena, Flemyng, Ella, Bero, Lisa, 2020. Digital contact tracing technologies in epidemics: a rapid review. Cochrane Database Syst. Rev. (8).

Banerjee, Sudipto, Carlin, Bradley P., Gelfand, Alan E., 2003. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC.

Benreguia, Badreddine, Moumen, Hamouma, Merzoug, Mohammed Amine, 2020. Tracking covid-19 by tracking infectious trajectories. IEEE Access 8, 145242–145255.

Besag, Julian, York, Jeremy, Mollié, Annie, 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Statist. Math. 43 (1), 1–20.

Enright, Jessica, Kao, Rowland Raymond, 2018. Epidemics on dynamic networks. Epidemics 24, 88–97.

Gillespie, Daniel T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81 (25), 2340–2361.

Haklay, Mordechai, Weber, Patrick, 2008. Openstreetmap: User-generated street maps. IEEE Pervasive Comput. 7 (4), 12–18.

Hernández-Orallo, Enrique, Manzoni, Pietro, Calafate, Carlos Tavares, Cano, Juan-Carlos, 2020. Evaluating how smartphone contact tracing technology can reduce the spread of infectious diseases: the case of COVID-19. IEEE Access 8, 99083–99097.

Hulle, Van, Marc M., 2012. Self-organizing maps.

Kahle, David, Wickham, Hadley, 2013. ggm: spatial visualization with ggplot2. R J. 5 (1), 144–161. URL https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

Kermack, William Ogilvy, McKendrick, Anderson G., 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. Ser. A Math. Phys. Sci. 115 (772), 700–721.

Lee, Duncan, 2013. CARBayes: an R package for Bayesian spatial modeling with conditional autoregressive priors. J. Stat. Softw. 55 (13), 1–24.

Leroux, Brian G., Lei, Xingye, Breslow, Norman, 2000. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Statistical Models in Epidemiology, the Environment, and Clinical Trials. Springer, pp. 179–191.

Muller, Markus, Derlet, Peter M, Mudry, Christopher, Aepli, Gabriel, 2020. Testing of asymptomatic individuals for fast feedback-control of COVID-19 pandemic. Phys. Biol. 17 (6), 060057.

Prodanov, Dimiter, 2021. Asymptotic analysis of the SIR model. Applications to COVID-19 modelling.

Riebler, Andrea, Sørbye, Sigrunn H, Simpson, Daniel, Rue, Håvard, 2016. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. Stat. Methods Med. Res. 25 (4), 1145–1165.

Riebler, Andrea, Sørbye, Sigrunn H, Simpson, Daniel, Rue, Håvard, 2016. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. Stat. Methods Med. Res. 25 (4), 1145–1165.

Rue, Håvard, Martino, Sara, Chopin, Nicolas, 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2), 319–392.

Sahu, Sujit K., 2022. Bayesian Modeling of Spatio-Temporal Data with R. Chapman and Hall/CRC.

Simmerman, James Mark, Suntarattiwong, Piyarat, Levy, Jens, Gibson, Robert V., Cruz, Christina, Shaney, Jeffrey, Jarman, Richard G., Chopitayasunondh, Taweey, 2010. Influenza a virus contamination of common household surfaces during the 2009 influenza A (H1N1) pandemic in Bangkok, Thailand: Implications for contact transmission. Clin. Infect. Dis. (ISSN: 1058-4838) 51 (9), 1053–1061. http://dx.doi.org/10.1086/656581.

Stan Development Team, 2021. RStan: the R interface to Stan. URL https://mc-stan.org/.

Tsai, Tzu-Chieh, Chan, Ho-Hsiang, 2015. NCCU trace: Social-network-aware mobility trace. IEEE Commun. Mag. 53 (10), 144–149.

Van Doremalen, Neeltje, Bushmaker, Trenton, Morris, Dylan H, Holbrook, Myndi G, Gamble, Amandine, Williamson, Brandi N, Tamin, Azaiib, Harcourt, Jennifer L, Thornburg, Natalie J, Gerber, Susan I, et al., 2020. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. N. Engl. J. Med. 382 (16), 1564–1567.

Zheng, Yu, 2015. Trajectory data mining: an overview. ACM Trans. Intell. Syst. Technol. (TIST) 6 (3), 1–41.