

# The role of mobility in the dynamics of the COVID-19 epidemic in Andalusia

Z. Rapti<sup>a</sup>

*Department of Mathematics and Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign*

J. Cuevas-Maraver

*Grupo de Física No Lineal, Departamento de Física Aplicada I,  
Universidad de Sevilla. Escuela Politécnica Superior, C/ Virgen de Africa, 7, 41011-Sevilla, Spain and  
Instituto de Matemáticas de la Universidad de Sevilla (IMUS). Edificio Celestino Mutis. Avda. Reina Mercedes s/n, 41012-Sevilla, Spain*

E. Kontou and S. Liu

*Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign*

Y. Drossinos

*Thermal Hydraulics & Multiphase Flow Laboratory, Institute of Nuclear & Radiological Sciences and Technology, Energy & Safety,  
N.C.S.R. "Demokritos", GR-15341 Agia Paraskevi, Greece*

P. G. Kevrekidis, M. Barmann, and Q.-Y. Chen

*Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01003-4515, USA*

G.A. Kevrekidis

*Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218, USA*

(Dated: March 28, 2023)

Metapopulation models have been a popular tool for the study of epidemic spread over a network of highly populated nodes (cities, provinces, countries) and have been extensively used in the context of the ongoing COVID-19 pandemic. In the present work, we revisit such a model, bearing a particular case example in mind, namely that of the region of Andalusia in Spain during the period of the summer-fall of 2020 (i.e., between the first and second pandemic waves). Our aim is to consider the possibility of incorporation of mobility across the province nodes focusing on mobile-phone time dependent data, but also discussing the comparison for our case example with a gravity model, as well as with the dynamics in the absence of mobility. Our main finding is that mobility is key towards a quantitative understanding of the emergence of the second wave of the pandemic and that the most accurate way to capture it involves dynamic (rather than static) inclusion of time-dependent mobility matrices based on cell-phone data. Alternatives bearing no mobility are unable to capture the trends revealed by the data in the context of the metapopulation model considered herein.

Keywords: metapopulation; human mobility; COVID-19 epidemic; gravity law

---

<sup>a</sup> Corresponding author: [zrapti@illinois.edu](mailto:zrapti@illinois.edu)

## I. INTRODUCTION

Human mobility has played an indisputable role in COVID-19 dynamics [9, 28] with as many as 86% of global cases having been imported from Wuhan, the original location of the pandemic. Studies of the epidemic in China have shown that in the early stages thereof, the probability of an outbreak was correlated with the frequency of imported cases from Wuhan [28]. The trajectory of the epidemic over a similar time period was also studied in [29] using an SEIR (Susceptible-Exposed-Infected-Recovered) stochastic metapopulation model, where it was determined that undocumented infections played a crucial role in the rapid spread of the epidemic. These early and high profile studies render it clear that a systematic consideration of such mobility aspects of the pandemic and of theoretical models thereof is a key ingredient towards appreciating its potential for spreading across countries and regions.

Continuing along this vein, in a subsequent study [44], the probability of case importations to countries having airports with direct flights to and from mainland China was estimated. It was assumed that the probability of importation is proportional to the number of airports in the country with direct connections to mainland China. With the implementation of Wuhan's travel ban and the subsequent international travel restrictions, [9] analyzed the effect of quarantine measures on local, national, and international pandemic spread. Even though the spread of the virus could only be delayed in the Chinese mainland, the mitigation of the transmission would be notable around the world. Modeling the course of the epidemic in other countries such as England and Wales, [13] also incorporated daily commuting as an important factor in the spread of the disease. It was assumed that infectious hosts may infect others both at home during the night and away during the day in the span of a day's cycle. A study evaluating confinement and other mitigation measures in Spain [2] used workforce mobility as a proxy for confinement. For Brazil, commuter and airline data were used to calibrate a stochastic epidemic model [11]. The model was used to investigate the spatial spread of the disease at various geographical scales (ranging from municipalities to states). It should be clear that these are only some select studies within a continuously expanding large volume of literature, which has now also been reviewed, e.g., in [7] (see also earlier reviews such as [8, 31]).

Of course, such models have a time-honored history in earlier instances of disease-spread modeling. For example, the Global Epidemic and Mobility (GLEAM) team integrates real-world pandemic transmission models with mobility data, including airline transportation network flows, ground mobility flows, and sociodemographic features, to capture spatiotemporal connections between mobility and an epidemic's spread [4, 9]. A model for influenza in the US [33] accounted for both daily commuting and random travels between states. One of the main findings there was that the metapopulation model more accurately predicts the onset, peak timing and intensity than models only accounting for specific locations. A study of long-term influenza patterns in the US [42] used mortality data and the gravity model, whereby population flows between nodes of the metapopulation network are determined by considerations akin to Newton's law of gravity, to study the spreading of influenza across states. References [4, 49] also showed correlation between infection spread and human movements. Theoretical metapopulation studies, where the travel rates are given by the gravity model, also exist [5]. Other works use the rates at which hosts leave and return to their permanent locations to infer the coupling strengths in their ODE model [25]. In earlier work, the bubonic plague epidemic was modeled using a similar approach [24], with adjacent metapopulations on a lattice coupled to rates chosen to fit historical data.

Studies looking at human mobility under lenses that go beyond gravity models also exist. One such example is the radiation model [37], which is based on the assumption that population density dictates employment opportunities, so when density is low, commuters need to travel longer distances. Hence, the predicted flux depends on the origin and destination populations and on the population of the region surrounding the origin location. More recently, a new mobility law [36] has been proposed showing that the number of visitors to any location is proportional to the inverse square of the product of the frequency of visits and distance traveled. This law has been applied in the context of urban mobility (within-city mobility), where it has shown a remarkable agreement with data.

When traffic data are available, they may be leveraged using entropy maximization techniques [22, 41] aiming to reconstruct origin-destination matrices [46] describing human mobility among various locations. However, in more recent considerations where mobile phone data are available, these have been found to more accurately represent the actual movements of people [39, 45]. During the first and second COVID-19 pandemic waves, in the US [3, 21], Japan [48], and in China [9], among others, mobile phone location data were utilized to explore the effects of mobility on the reported cases reduction.

We should also note in passing that other approaches to examining the spatial spread of COVID-19 have also been deployed, including, e.g., models based on partial differential equations [27, 30, 43]. These modeling efforts take into account local population density by modifying the transmission coefficients accordingly [27] (compared to an ordinary differential equations model), emphasize the importance of inflows from neighboring regions [43], and utilize time-varying diffusion coefficients to account for the effect of mitigation measures [27, 30].

In the present work, we wish to explore some of the practical challenges of applying a metapopulation model to a concrete region during the COVID-19 pandemic, and also when attempting to systematically compare model results with existing data. In line with our earlier studies [12], we bring to bear an epidemic model that accounts for both symptomatic and asymptomatic infections and includes appropriate recovered compartments, as well as a compartment for the fatalities, since the latter appears to be the most accurate dataset [23]. However, since we have examined already aspects of the identifiability of such models, as well as their usefulness in the context of age-structured populations [12], we do not focus on such aspects herein. Instead,

our emphasis is on the availability of different approaches to couple the nodes of such a model into a network pattern for a metapopulation description of a region of interest. In that vein, we compare and contrast the findings of an implementation neglecting the mobility between provinces, with one incorporating it. When incorporating such mobility traits, we comment on our attempts to do so, based on “standard” techniques such as those stemming from gravity models or transportation-based origin-destination matrices.

Our case example of interest is the region of Andalusia in Spain for numerous reasons, including the familiarity of our group with the region (aiding an understanding of the observed mobility patterns and, e.g., their seasonal variation). A significant feature facilitating and enabling our study is a large-scale data analysis of the Transportation Ministry of the Spanish Government [32] that provides time-resolved mobility data across the provinces within this region and hence a dynamic incorporation of the relevant patterns based on an “as accurate as possible” characterization of the mobility within the area of interest. We calibrate the model using fatality data from Andalusia [15], focusing on the summer and early fall period of 2020 (i.e., from around the end of the first and the beginning of the second pandemic wave). During this period, mitigation measures were relatively relaxed and mobility among provinces was high due to summer vacations and later due to higher education-related relocation. We find that we are unable to obtain a quantitative match with the observed data in each province (and hence Andalusia as whole) without mobility—or with static patterns of mobility produced by some of the above mentioned “standard” techniques—. Instead, our most accurate quantitative description of the observations stems from the incorporation of the above described “dynamic mobility” as obtained from the time dependent mobile-phone data in [32].

Our presentation is structured as follows. In Section 2, we present the model, including the relevant metapopulation network considerations. We also show how mobility matrices, an input to the metapopulation model, that are obtained from different data sets compare. In Section 3, we present our results, including the parameter fitting approach used and the comparison with the existing data for COVID-19 fatalities in each Andalusian province. Finally, in Section 4, we present our conclusions and a discussion towards future steps within these classes of models. The Appendix contains information on the determination of the Origin-Destination matrix using either the gravity method or mobile-phone data.

## II. MODELING FRAMEWORK

### A. Epidemic model for each node

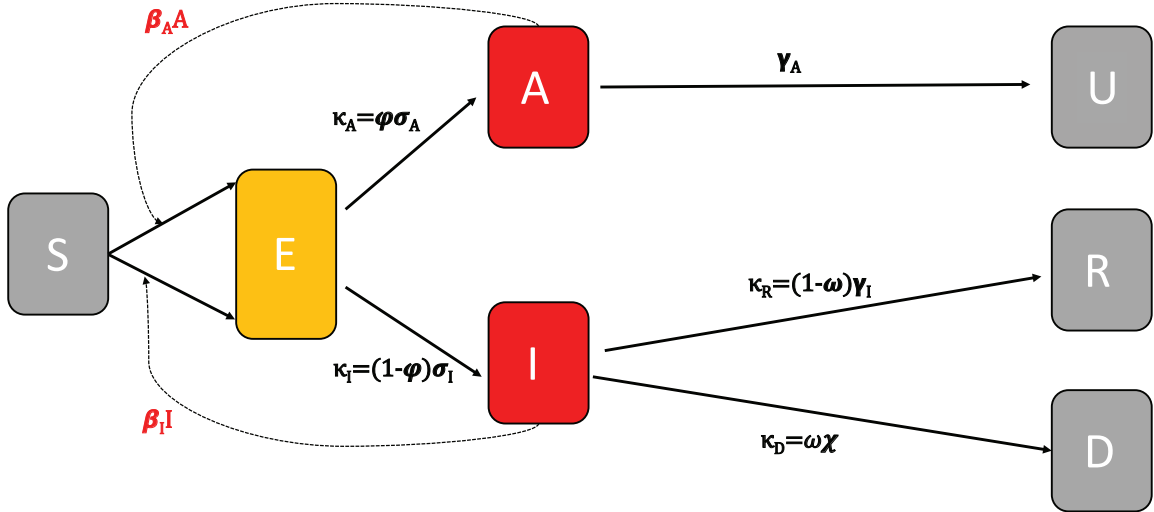


FIG. 1. Schematic diagram of the Susceptible-Exposed-Asymptomatic-Infected-Recovered (SEAIR) model for each metapopulation node.

In the ordinary differential equation (ODE) model that we put forth (a slight variant of the ones previously considered, e.g., in [12, 27]), there are seven compartments for each node. Susceptible individuals,  $S$ , become exposed (latently infected,

not infectious yet),  $E$ , after contact with either asymptotically infectious hosts,  $A$ , or symptomatically infectious hosts,  $I$ . Recall that the importance of asymptotically induced transmission, especially in the context of COVID-19 has been argued in numerous studies [7, 34]. We assume standard incidence  $\beta_{AS,IS}/(N - D)$ , where  $N - D = S + E + A + I + U + R$  is the total living population. However, as the number of individuals in the deceased class  $D$  is quite small in most cases, from now on it will be ignored compared to  $N$ , namely we will set the incidence to  $\beta_{AS,IS}/N$ , where the transmission coefficient  $\beta_{AS,IS}$  can be assumed constant over the considered periods of time. We have selected the time interval under consideration as one involving high mobility without changes in mitigation measures, so as to reflect more clearly the genuine role of transportation effects in the model results.

Once in the exposed class  $E$ , a fraction of hosts  $\phi$  never develop symptoms and moves into the asymptotically infectious class  $A$  at a rate  $\sigma_A$ . Asymptomatic hosts are assumed to recover at an average rate  $\gamma_A$  and move into the recovered compartment  $U$ . The remaining exposed host fraction  $1 - \phi$  develops symptoms at a rate  $\sigma_I$  and these individuals move into the symptomatically infectious class  $I$ . A fraction  $\omega$  of symptomatic hosts die at an average rate  $\chi$  (moving into the compartment  $D$ ) and the remaining fraction  $1 - \omega$  recovers at a rate  $\gamma_I$  and moves into the recovered class  $R$ . A schematic diagram of the above description is shown in Figure 1. The relevant equations governing the spreading of the epidemic read:

$$S' = -\beta_{AS}S\frac{A}{N} - \beta_{IS}S\frac{I}{N} \quad (1)$$

$$E' = \beta_{AS}S\frac{A}{N} + \beta_{IS}S\frac{I}{N} - (\kappa_A + \kappa_I)E \quad (2)$$

$$A' = \kappa_A E - \gamma_A A \quad (3)$$

$$I' = \kappa_I E - (\kappa_R + \kappa_D)I \quad (4)$$

$$U' = \gamma_A A \quad (5)$$

$$R' = \kappa_R I \quad (6)$$

$$D' = \kappa_D I, \quad (7)$$

where we set

$$\kappa_A = \phi\sigma_A, \quad \kappa_I = (1 - \phi)\sigma_I, \quad \kappa_R = (1 - \omega)\gamma_I, \quad \kappa_D = \omega\chi. \quad (8)$$

In what follows, in order to reduce parameter redundancy in the model, we fit the following seven parameters and parameter combinations

$$\beta_{AS}, \beta_{IS}, \kappa_A = \phi\sigma_A, \kappa_I = (1 - \phi)\sigma_I, \kappa_R = (1 - \omega)\gamma_I, \kappa_D = \omega\chi, \gamma_A \quad (9)$$

This version of the model will be used when considering the fatalities within Andalusia's provinces but without any (mobility-induced) coupling between them and when considering the entire Andalusia (no metapopulation). It is straightforward to observe that for system (1-7), the total population  $N = S + E + A + I + R + U + D$  is conserved. Moreover, the subset  $\{S \geq 0, E \geq 0, A \geq 0, I \geq 0, U \geq 0, R \geq 0, D \geq 0\}$  of  $\mathbb{R}^7$  is positively invariant for the system. Hence, the system is well-posed for any initial condition.

All variables and model parameters are defined in Table I.

## B. Metapopulation model

We are implementing a coupling between the different provinces in line with [5]. Namely, we assume that individuals are indistinguishable and travel from node  $i$  to node  $j$  at some rate given by human mobility data, without assigning any base location to them. Hence, individuals in node  $i$  are instantaneously assigned to node  $j$  upon arrival, regardless of their prior node (no memory). The same individual may change multiple nodes, in principle, within the model. Connections between the nodes depend on the mobility flow of susceptible  $S$ , exposed  $E$ , and infectious hosts,  $A$  and  $I$ . To avoid a highly complicated model, we do not incorporate terms such as  $S_i A_j$ ,  $S_i I_j$  in the equations., i.e., we assume that the primary source of infection is through interactions of susceptible with infectious individuals within each node (no direct long-range transmission).

The metapopulation model assumes the following form (with  $i = 1, \dots, i_{\max}$  where  $i_{\max} = 8$  since Andalusia has eight

TABLE I. Variables and parameters

Variable	
$S$	susceptible hosts
$E$	exposed hosts
$A$	asymptomatically infectious hosts
$I$	symptomatically infectious hosts
$U$	recovered hosts (transitioning from $A$ )
$R$	recovered hosts (transitioning from $I$ )
$D$	deceased hosts
$t$	time (days)
Parameter	
$\beta_{IS}$	transmission rate between $I$ and $S$ (1/day)
$\beta_{AS}$	transmission rate between $A$ and $S$ (1/day)
$\kappa_A$	transition rate from $E$ to $A$ (1/day)
$\kappa_I$	transition rate from $E$ to $I$ (1/day)
$\gamma_A$	transition rate from $A$ to $U$ (1/day)
$\kappa_R$	transition rate from $I$ to $R$ (1/day)
$\kappa_D$	transition rate from $I$ to $D$ (1/day)

provinces):

$$S'_i = -\beta_{AS}S_i \frac{A_i}{N_i} - \beta_{IS}S_i \frac{I_i}{N_i} + \theta \left( \sum_j M_{ij} \frac{S_j}{N_j} - \sum_j M_{ji} \frac{S_i}{N_i} \right) \quad (10)$$

$$E'_i = \beta_{AS}S_i \frac{A_i}{N_i} + \beta_{IS}S_i \frac{I_i}{N_i} - (\kappa_A + \kappa_I)E_i + \theta \left( \sum_j M_{ij} \frac{E_j}{N_j} - \sum_j M_{ji} \frac{E_i}{N_i} \right) \quad (11)$$

$$A'_i = \kappa_A E_i - \gamma_A A_i + \theta \left( \sum_j M_{ij} \frac{A_j}{N_j} - \sum_j M_{ji} \frac{A_i}{N_i} \right) \quad (12)$$

$$I'_i = \kappa_I E_i - (\kappa_R + \kappa_D)I_i + \theta \left( \sum_j M_{ij} \frac{I_j}{N_j} - \sum_j M_{ji} \frac{I_i}{N_i} \right) \quad (13)$$

$$U'_i = \gamma_A A_i \quad (14)$$

$$R'_i = \kappa_R I_i \quad (15)$$

$$D'_i = \kappa_D I_i \quad (16)$$

$$N'_i = \theta \left( \sum_j M_{ij} - \sum_j M_{ji} \right) \quad (17)$$

The last equation shows how the population of node  $i$  is updated over time. Our model is along the lines of [29, 33]. If mobility is ignored by setting  $\theta = 0$ , the total population within each node  $N_i$  is conserved. Otherwise, when  $\theta = 1$ , solely the total population over all provinces is conserved. We note that  $\theta$  is a binary parameter, assuming the value 1 (0) when human mobility is considered (not considered).  $M_{ij}$  is the daily rate of people traveling from  $j$  to  $i$ . Then, one multiplies this rate with the proportion of  $S, E, A, I$  in the total node population  $N_j$ . This can be interpreted as the probability of an individual from these four classes traveling if we choose randomly from  $N_j$ . Symptomatically infectious individuals  $I$  are assumed to be able to move, but not  $U$  or  $R$ . In any event, the latter two do not affect further dynamics in the network as they are terminal classes of the model. Allowing  $U$  and  $R$  to move (since no re-infection is considered on the time-scales used in the present work) only has the effect of redistributing the recovered population among the network nodes: the infection dynamics are not expected to be directly affected. However, movement of individuals in these compartments may still change the population size of a given location, which could slightly affect incidence (frequency-dependent transmission with  $1/N_i, 1/N_j$  terms). It is relevant to also note that, over the time scale considered, these individuals are assumed to have immunity (upon recovery) and, hence, it is

not considered to be a possibility for the population in  $U$  or  $R$  to re-enter the susceptible population, over the time frame of interest. Therefore, we only allow the susceptible class,  $S$ , which consists the majority of the population, and the exposed,  $E$ , and infectious,  $A$  and  $I$ , classes to move among the network nodes. Furthermore, since there were no mobility restrictions at the time, we assume that the movement of exposed, asymptomatic, and infected individuals is the same as the movement of susceptible individuals. While quarantine and isolation were required at the time for infectious and exposed individuals, we consider that they all travel and at the same rate since it is not straightforward to estimate compliance. Should, however, such compliance data become available, it would be an easy fix to multiply  $M_{ij}$  with the appropriate compliance rate.

One may easily observe that as long as  $N_i > 0$ ,  $i = 1, \dots, 8$  the metapopulation model is well-posed for any initial condition. This follows since the region

$$\{S_i \geq 0, E_i \geq 0, A_i \geq 0, I_i \geq 0, U_i \geq 0, R_i \geq 0, D_i \geq 0\} \subset \mathbb{R}^{56}$$

is positively invariant and for the time period studied, while  $N_i$  fluctuate, they stay positive for all  $i = 1, \dots, 8$ .

We note that based on the mobility data [32], the network of the eight Andalusian provinces is a complete graph and the population flows  $M_{ij}$  are time-dependent. In the following subsection we discuss how we determined the daily movement rates, i.e., the population flows, between two network nodes, and alternative ways to determine them if mobile-phone records are not available.

### C. Human mobility estimation

Mobility flows are commonly estimated from mobile-phone records. In this work the flows we analyzed are based on a study performed by the Ministry of Transportation, Mobility and Urban Agency (Ministerio de Transportes, Movilidad y Agenda Urbana-MITMA) [32] that included all of Spain for the period beginning on March 14, 2020. The main data source was anonymized mobile phone data for more than 13 million mobile lines as well as locations of communication towers and antenna orientations. Population data as well as information about the transportation network (airport locations, railways) were leveraged. Figure 2 shows the time-dependent population flows for each province, i.e., each network node, of Andalusia as determined by the mobile-phone data. They are shown for the time duration of our study, starting on July 10, 2020 till October 29, 2020 (112 days). Note the significant time dependence of the inter-province population flows.

Daily mobility data among locations, such as those provided by mobile phones, may not be always readily and publicly available. When this is the case, other types of data and alternative models are used to determine the population flows. One avenue is to rely on census surveys and base the coupling of the epidemiological model on daily commuting data [13, 14]. In this approach, workdays and weekend, as well as commuters and non-commuters, should be distinguished using additional travel surveys. Failure to consider non-work related trips may lead to an erroneous slowing down of the epidemic [14]. This fine tuning is not required when using time-varying mobility matrices as in the present work.

Another avenue is to utilize commonly used trip-distribution modeling techniques, like the gravity model to construct the origin-destination (O-D) matrix for the metapopulation network. The gravity law is used extensively in the literature to model travel demand between O-D pairs (e.g., [16, 20]). We assume a region where  $n$  denotes the nodes (or centroids) of the cities in the regional transportation network and  $m$  their highway links. A trip matrix element (number of trips per day) is denoted by  $w_{ij}$ , where  $i$  and  $j$  are the origin and destination nodes of the considered trip, respectively. Given the population of these cities and their distances, the O-D matrix elements are computed by

$$w_{ij} = C \frac{N_i^\alpha N_j^\gamma}{e^{\beta_G \text{dis}_{ij}}} \quad (18)$$

where  $C$  is a constant,  $\text{dis}_{ij}$  is the distance between the O-D pair  $(ij)$ ,  $\alpha$  and  $\gamma$  are parameters associated with the populations  $N_i$  and  $N_j$  of the pair  $(ij)$ , and  $\beta_G$  is a constant parameter whose value—measured in units of inverse distance, indeed in our case of 1/Km—depends on the distance between the network nodes, as explained in Appendix A. Once the elements of the O-D matrix have been estimated, the force of infection on susceptible hosts  $S_j$  in location  $j$  in the metapopulation model, which reads  $\beta_{AS} A_j / N_j$ ,  $\beta_{IS} I_j / N_j$  in Eqs. (10, 11) is modified as [47]

$$\beta_{AS} \frac{1}{N_j} \left( A_j + C N_j^\gamma \sum_{i \neq j} \frac{A_i^\alpha}{e^{\beta_G \text{dis}_{ij}}} \right), \quad \beta_{IS} \frac{1}{N_j} \left( I_j + C N_j^\gamma \sum_{i \neq j} \frac{I_i^\alpha}{e^{\beta_G \text{dis}_{ij}}} \right). \quad (19)$$

A notable difference between models implementing (19) and the metapopulation model (10-17) is the following. Whereas the model defined by (10-17) introduces mobility via changes directly in the rates of change of the  $S$ ,  $E$ ,  $A$ , and  $I$  populations (for  $\theta = 1$ ), models using (19) implement mobility through modification of the transmission terms.

It is relevant to note that the accuracy of gravity-like models has received considerable recent criticism [36, 37]. In the present work, we will not embark on a detailed comparison of a metapopulation model based on the gravity law and our own approach

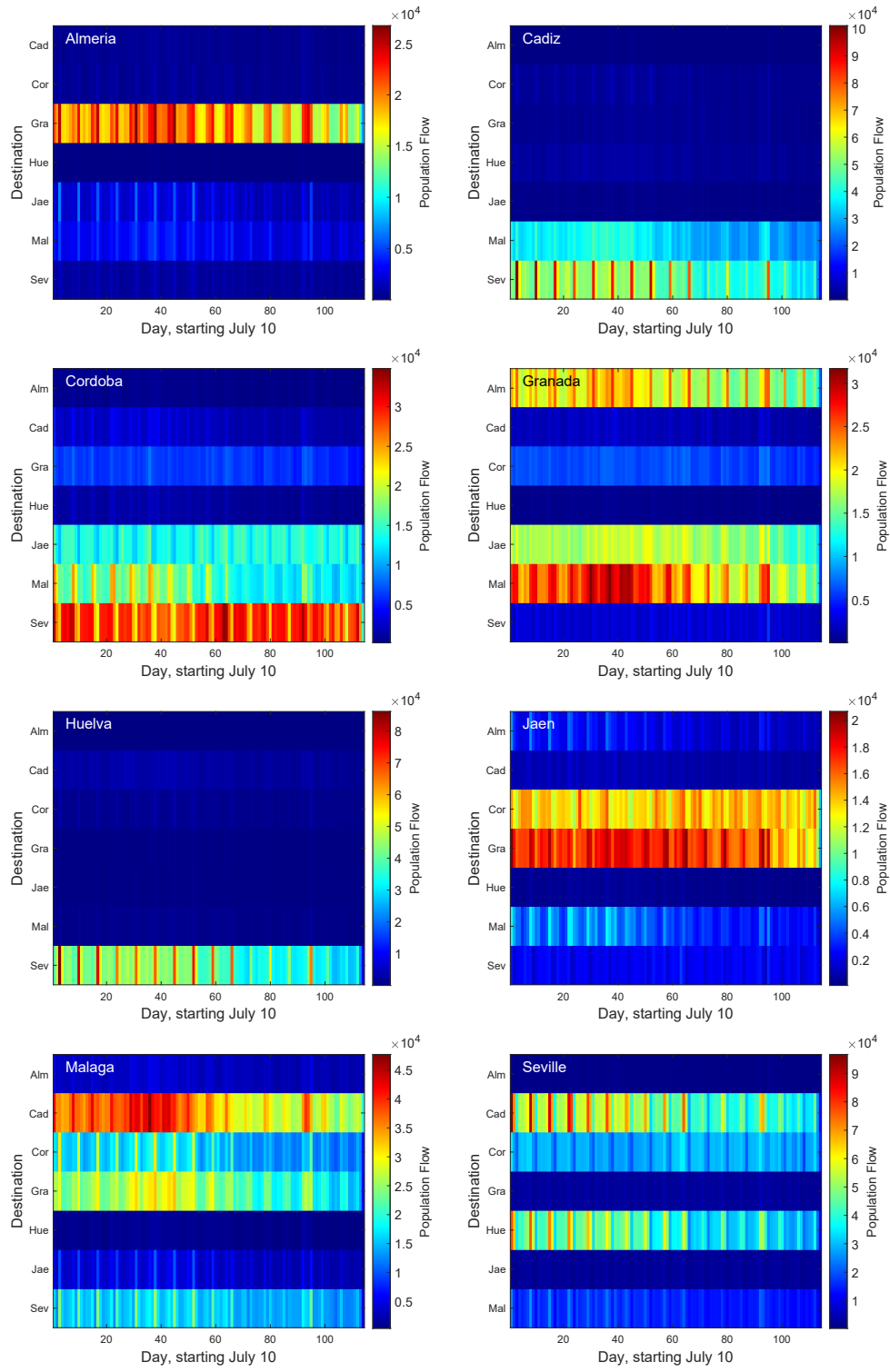


FIG. 2. Time-varying daily population flows (the daily rate of individuals traveling  $M_{ij}$ ) for each Andalusian province as determined from mobile-phone data [32]. The destinations are shown on the vertical axis. The code name for the eight provinces in Andalusia is: Alm: Almeria, Cad: Cadiz, Cor: Cordoba, Gra: Granada, Hue: Huelva, Jae: Jaen, Mal: Malaga, Sev: Seville.

(based on time-dependent mobile-phone records). Nevertheless, for completeness, we would like to illustrate that in the absence of alternative and possibly quite superior data sets, the method can be used to capture some principal features of mobility flows

in workdays (Friday and Tuesday; no mobility restrictions in place); see, in particular, Figure 3. More concretely, due to the scarcity of reported traffic count data –they are averaged over a year – only a static O-D matrix can be obtained. Also, the traffic count data available to us were from 2019-2020, namely prior to the pandemic. The O-D matrices in Figure 3 show that the gravity-law method roughly captures the main mobility trends. For instance, there is substantial support within the matrix between the rows 2-4 and columns 6-8 (and vice-versa), as well as e.g. between Seville and Huelva or Malaga etc. Therefore, in the absence of more detailed and accurate mobility information, it can be used as an alternative. The O-D matrix presented in Fig. 3, left panel, reproduces the gravity-law data shown in Table VI in the Appendix. It should be noted however, that the gravity-law O-D is in terms of vehicle trips per day, whereas the mobility flows from the Spanish government reported in [32] are expressed in terms of people traveling per day. To convert one into the other one would need to know on average the number of people traveling in vehicles: as the comparison is qualitative we opted not to convert population flows to trips per day.

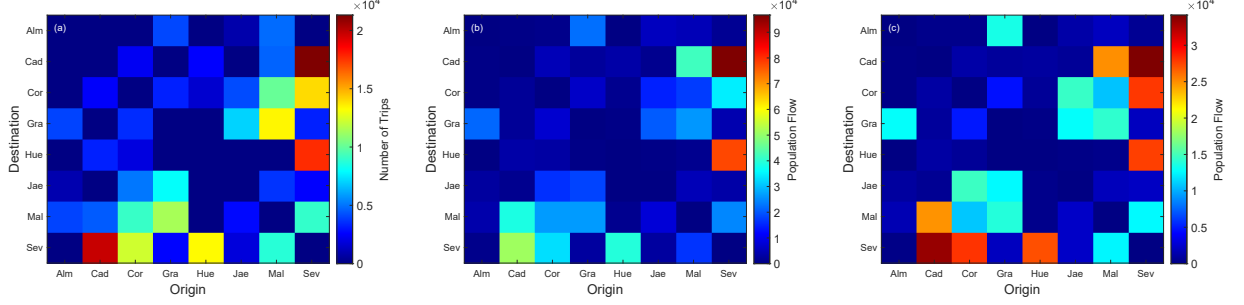


FIG. 3. Origin-Destination matrices based on the gravity-law method (vehicle trips per day) for a pre-pandemic day (panel (a), left) and from the mobility data based on mobile-phone records (people traveling per day) for Friday, July 10, 2020 (panel (b), middle) and Tuesday, October 27, 2020 (panel (c), right).

### III. RESULTS

#### A. Period of study and rationale

The time period considered begins on July 10, 2020 and ends on October 29, 2020 (112 days). We use the first 84 days from July 10 till October 1, 2020 as the fitting period, and the remaining 28 days from October 2 to October 29, 2020 as the prediction interval. Since the goal of the present study is to investigate of the role of mobility on the spread of an epidemic, the period of study was chosen to satisfy the following two conditions. First, that there would be no imposed mobility restrictions except at the end. In fact, on October 29, the regional government imposed a curfew at nights and closed the border with the rest of Spain and limited mobility between the provinces. That is the reason we chose to perform our analysis up to the end of October 2020, and not longer, as afterwards the mobility patterns were modified due to the imposed restrictions on travel. Second, the period should include the initial exponential growth of the epidemic peak.

#### B. Parameter fitting and model predictions

We first use the model of Eqs. (1)-(7) for the entire region of Andalusia. We use the norm

$$\mathcal{N} = \sum_{i=1}^{84} \left( \frac{D_{\text{num}}(t_i)}{D_{\text{obs}}(t_i)} - 1 \right)^2, \quad (20)$$

as the objective function. We minimized it to fit the fatality data  $D_{\text{obs}}(t_i)$ , where  $t_i$  stands for day  $i$ , since our start point of July 10, 2020, and  $D_{\text{num}}(t_i)$ , denotes the fatality estimate for the same day, obtained from the model. It is worth noting again that we are not attempting to fit to case data, since these are believed to be significantly less reliable than fatality data, due to under-reporting as has been the case in other countries as well [12, 26]. Indeed, when trying to fit both case and fatality data, we have obtained results that are considerably less satisfactory than the ones presented below.

In addition to the seven parameters shown in (9), we also obtain estimates for the initial parameters  $I_0, A_0, E_0$  when the entire autonomous community of Andalusia is considered. We performed 500 optimizations with an initial guess for each parameter uniformly sampled within a pre-specified range. The upper and lower limits of the variation ranges were used as boundaries in



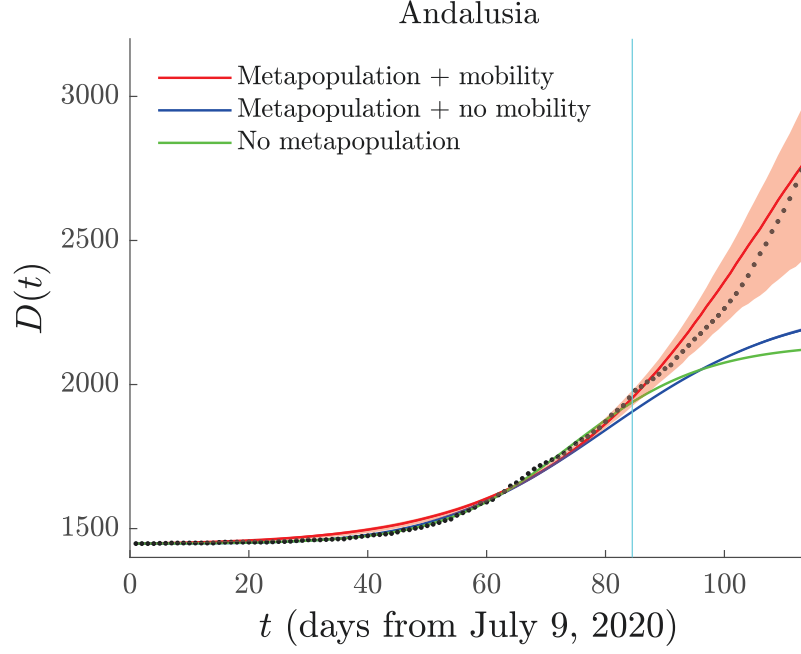


FIG. 4. Model fit and prediction of the fatalities time series for the entire region of Andalusia. Data points are shown as black dots, the output of the metapopulation model with mobility ( $\theta = 1$ ) is shown as a red curve, the output of the metapopulation model with mobility turned off ( $\theta = 0$ ) is shown as a blue curve, and the fit to the ODE model (Eqs. 1-7) is shown as a green curve. The light blue vertical line corresponds to the date when fitting stops (day 84) and prediction begins. The interquartile range is highlighted in red.

the constrained minimization algorithm (implemented in Matlab via the `fmincon` function). The outcome of the fitting, for values taken from July 10 to October 1, 2020 (84 days), allowed us to retrieve an approximation for the initial values for the  $I$ ,  $E$  and  $A$  compartments. Their median (they had a very small dispersion) was used as initial condition for the metapopulation model (10-17), weighted by  $\omega_j = C_j/C$ , with  $C_j$  being the number of cases in the  $j$ -th province in the period from July 4 to July 10 and  $C$  the total number of cases in Andalusia in the whole period. We minimized the norm

$$\mathcal{N} = \sum_{j=1}^8 \mathcal{N}_j \quad (21)$$

with

$$\mathcal{N}_j = \sum_i \left( \frac{D_{j,\text{num}}(t_i)}{D_{j,\text{obs}}(t_i)} - 1 \right)^2 \quad i = 1, \dots, 84, \quad (22)$$

and  $D_{j,\text{num}}(t_i)$  and  $D_{j,\text{obs}}(t_i)$  being, respectively, the fatality estimate and data for the day  $t_i$  at province  $j$ . In the metapopulation model, we focused on two values of  $\theta$ ,  $\theta = 1$  and  $\theta = 0$ , which will be denoted as the metapopulation model with and without mobility, respectively. As mentioned previously, the coupling matrices  $M_{ij}$  were obtained from mobile-phone data [32].

Figure 4 shows the fit of the SEAIR model of Eqs. (1)-(7), no metapopulation, together with the metapopulation model (10-17) with ( $\theta = 1$ ) and without ( $\theta = 0$ ) mobility for the case of only Andalusia. Part of the data (the first 84 days, from July 10 to October 1) is used for parameter fitting, and the remaining is used for prediction (till day 112, from October 2 to October 29). We observe that while all three curves are close to each other and trail the data points with a satisfactory level of accuracy during the fitting period (since we are fitting them to the data), they diverge afterwards. Only the metapopulation model with mobility follows the same trend as the data in the prediction interval. One possible reason is that during summer the fatality curves in all provinces behave similarly, i.e., they are quite homogeneous, but later on they follow different trends, and they become heterogeneous. Hence, the overall fatality curve (black dots in the figure), the one corresponding to the entire autonomous community of Andalusia, diverges from the homogeneous curve.

Another explanation is that it is possible to fit different models to the same data set, but not all models will be able to make accurate predictions. This is especially true when fitting to epidemic data in the period before the inflection point of the epidemic peak has been reached [35].

Further insight on the dynamical evolution of the fatalities in each of the provinces is provided in Figure 5. During the months considered in the present study, due to relaxation or complete absence of mitigation measures, different nodes of the network

exhibit different characteristics. This can be attributed to some nodes being touristic destinations (Malaga, Huelva), others being close to country borders (Cadiz with Gibraltar and Huelva with Portugal), while yet others undergoing annual exodus over the summer months (Seville). This is evident in Figures 2 and 6, where we show the variation in mobility flows and population, respectively, for the eight provinces forming our network.

It is relevant to make the following observations in connection with the results. With the exception of Cordoba (featuring a systematic underestimation within the prediction interval for which we do not have a definitive explanation) and Huelva (with a corresponding overestimation in the prediction interval), data points typically follow qualitative trends consonant with the interquartile range. Huelva is a major vacation hub, both in the summer period and during weekends, mainly from residents in Seville (who also commonly spend their holidays in the province of Cadiz). If people return to their permanent residence to receive treatment and quarantine (or are anyway logged as cases within these regions), this may explain the disparity between the observed and predicted fatalities. It is worthwhile to note an apparently similar overestimation trend within the prediction interval for Cadiz; however, in this case, the situation is somewhat less clear, due to an opposite trend within the fitting interval. Also, high population density over the summer could partially explain the overestimation: the model is trained with more people residing there, who subsequently depart to return to their regular residence. Also, compared to other provinces, fatalities are relatively small in number, which makes it prone to stochastic effects [1, 6], as is also evident in the trends of the data.

Figure 6 shows the population in each province during the period of our study. Two major trends emerge. First, there is a weekly oscillation, due to increased mobility during the weekends. This is due to residents traveling from their primary residence to vacation destinations, such as the ones we described before between Seville and Huelva or Cadiz; similar patterns are found between other pairwise transitions: e.g., in the case of Cordoba, such movements happen to and from Malaga, Seville and Jaen. In any event, the real-time data used in this work provide a clear picture of the dynamics across the network and the key interactions across its nodes. Second, there is a significant variation in the population of most provinces, ranging from mild (0.99-1.06 in Jaen, 0.93-1 in Cordoba) to extreme (0.75-1.2 in Huelva). Others, experience a peak in late summer (Almeria, Cadiz, Malaga) before their population drops again in October. Granada and Seville exhibit a reverse behavior, where their population increases in the fall, when people resume living in their permanent residence. This is the seasonal trend that is superposed to the weekly trend. A similar observation may be made by considering Figure 2 where the time-dependent population flows between any two provinces are shown. In line with our above observations, some clear signatures are obvious, such as weekly periodicity, overall increased mobility in the summer months and other trends, such as the consistent mobility between specific pairs of provinces, as discussed above.

Figure 7 depicts the time evolution of the fatalities in the form of a heat map. We can observe how the model predicts the spreading of the epidemic from Almeria to neighboring provinces. Note, however, that in the reported data there was a spot in Malaga, probably caused by people traveling from other places in the world (a process which is not included in the current work). We also observe that Seville and Malaga are the provinces that eventually exhibit the highest number of fatalities, an observation correlated to their higher population. The maps also show that the model without mobility predicts a very much smaller number of fatalities than the model with mobility. Although at an early stage of the prediction both models are fairly comparable, later on, within the prediction interval, the model with mobility is significantly more accurate towards predicting the spread of the epidemic within the metapopulation network than the model without it. Both the detailed (individual province, cf. Fig. 5) quantitative findings, and this overarching figure are convincing, in our view, of the relevance at such regional levels of the consideration of metapopulation approaches. Additionally, the concrete trends that our mobility data reveal illustrate the relevance of the dynamic consideration of the coupling matrices  $M_{ij}$ .

The best-fit parameters are shown in Table II, whereas Table III presents the initial conditions for each province in the metapopulation model. For the initial condition of the population ( $N_0$ ) we took the census data for January 1, 2020 [17]. We must note that the data we compared with correspond to the day when events (such as deaths) actually occurred, and were extracted from the data available in the National Epidemiological Center of Spain [15], as well as that each fatality is assigned to the residence province.

Table IV shows the residuals for the fitting and predictions. The former is found by computing (20) at each simulation and taking the median and quartiles of all these values; for getting the latter, the same procedure is followed but extending the summation in (20) to 114

### C. Effective reproduction number

Given the importance of the reproduction number during the initial stages of an epidemic wave, we use the Next Generation Matrix approach [18] to evaluate the effective reproductive number  $R_t$ . In doing so, we treat this epidemic wave as a “new epidemic” assuming that most of the population is still susceptible. This assumption practically renders  $R_t = R_0$ , namely the effective reproductive number is equal to the basic reproductive number  $R_0$ . For  $t > 0$  it holds  $R_t = R_0 S(t)/N$ , [10]. Hence, the calculated effective reproduction number refers to the first day of our simulations, July 10th, 2020. The reproduction number

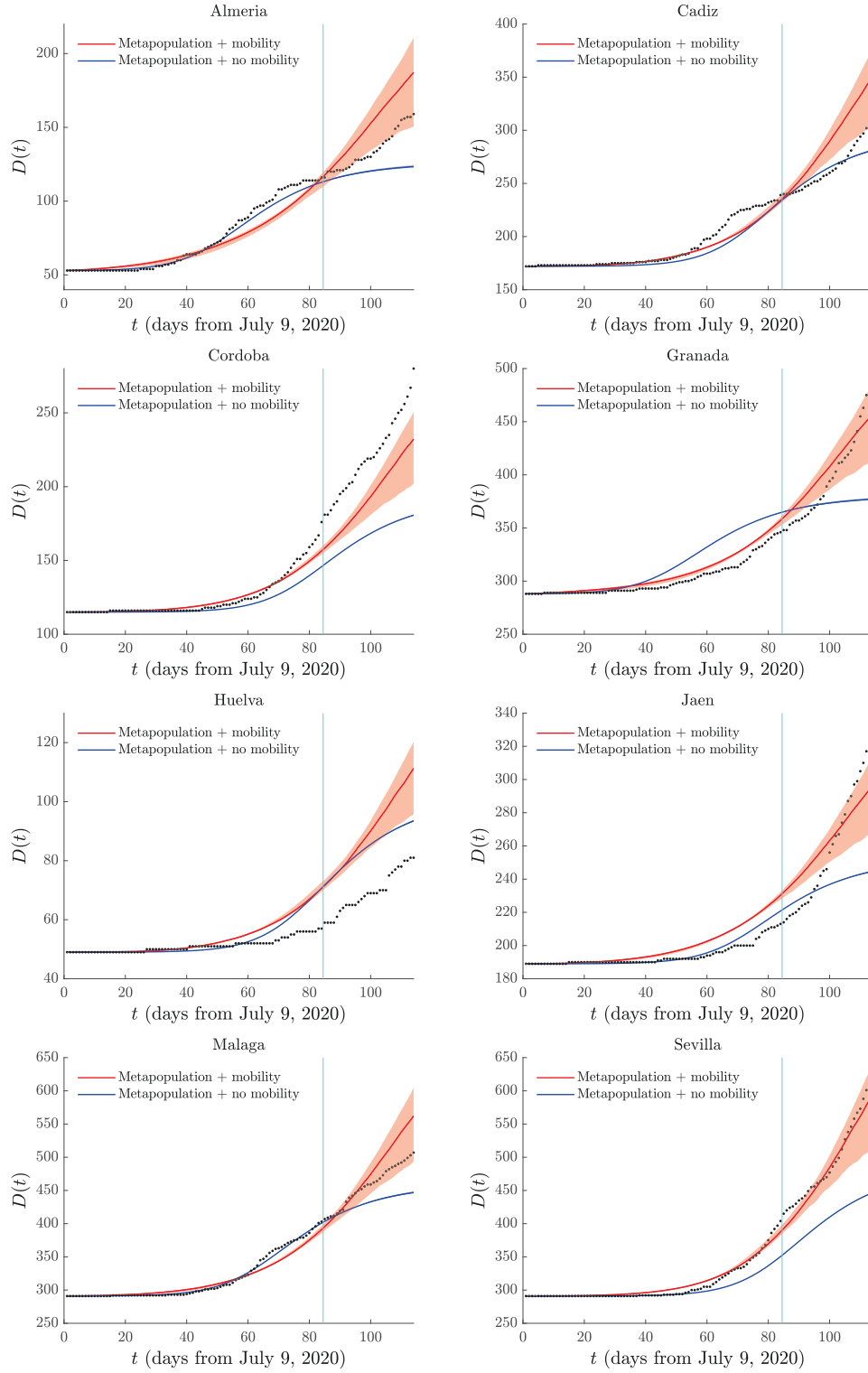


FIG. 5. Fatality data (black dots) and model fit for each province of Andalusia. The output of the metapopulation model, Eqs. (10)-(17), with mobility ( $\theta = 1$ ) is shown as a red curve and the output of the metapopulation model with mobility turned off ( $\theta = 0$ ) is shown as a blue curve. The light blue vertical line corresponds to the date when fitting stops (day 84) and prediction begins. The interquartile range is highlighted in light-red for red curve, while for the blue curve, the interquartile range is so narrow that it is not visible in the plot. Note the different y-axis scales.

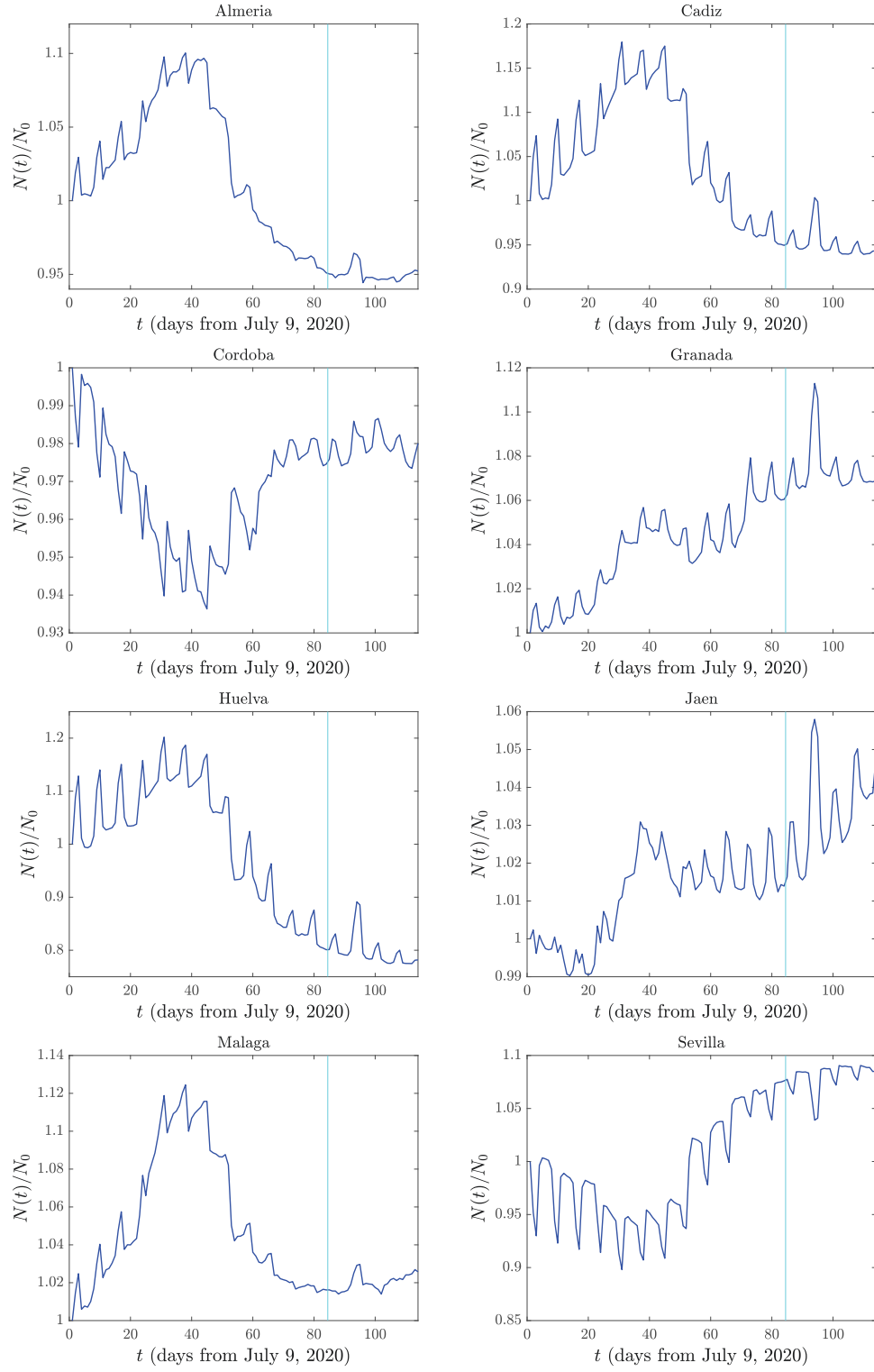


FIG. 6. The ratio of the population of each province  $N_i(t)$  to the the initial population  $N_0$ , the latter based on data from (17). The light red vertical line corresponds to the day when fitting stops (day 84) and validation begins. Note the different y-axis scales.

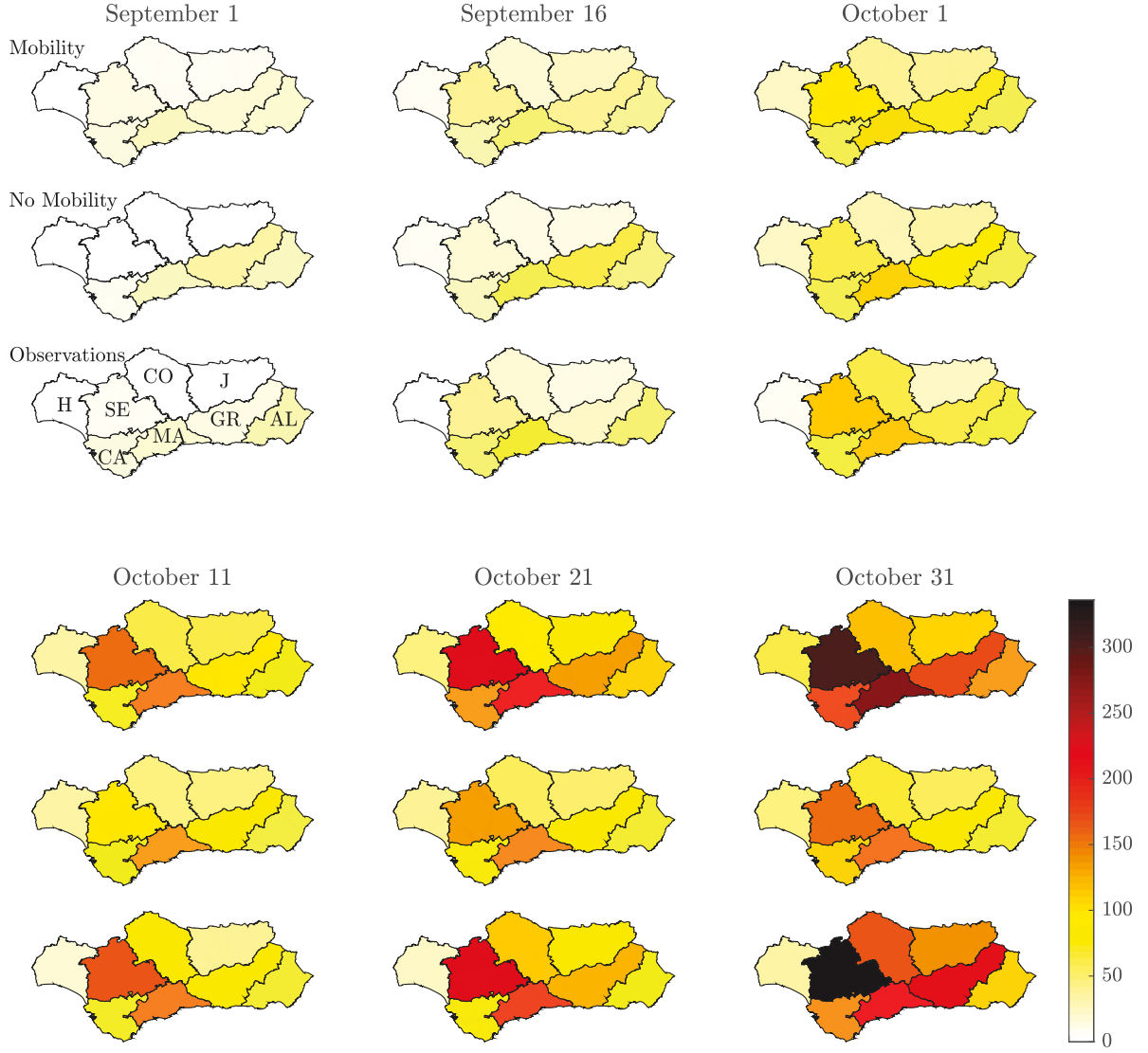


FIG. 7. Snapshots of the evolution of the number of fatalities occurring from (and including) July 10 at each province at different days from September to October. Top and middle row maps correspond to the numerical fit/prediction of the metapopulation model with and without mobility, whereas bottom row maps represent the observed number of fatalities. Bottom map in the first snapshot includes the code for the name of each province (AL: Almeria, CA: Cadiz, CO: Cordoba, GR: Granada, H: Huelva, J: Jaen, MA: Malaga, SE: Seville).

for the one-node model (1-7) (with either  $N - D$  in the denominator or the simplified version  $N - D \approx N$ ) is

$$R_t = \frac{\kappa_A}{\kappa_A + \kappa_I} \frac{\beta_{AS}}{\gamma_A} + \frac{\kappa_I}{\kappa_A + \kappa_I} \frac{\beta_{IS}}{\kappa_D + \kappa_R}. \quad (23)$$

The first term is the contribution to  $R_t$  from asymptomatic hosts  $A$  while the second is the contribution from the symptomatically infectious hosts  $I$ . Each term represents the fraction of asymptomatic  $\kappa_A/(\kappa_A + \kappa_I)$  or symptomatically infected  $\kappa_I/(\kappa_A + \kappa_I)$  hosts generated in the lifespan of an exposed host  $E$ , or equivalently the fraction of individuals reaching  $A$  or  $I$  after going through state  $E$ , multiplied by the number of new infected hosts generated in the lifespan of the corresponding infectious host,  $\beta_{AS}/\gamma_A$ ,  $\beta_{IS}/(\kappa_D + \kappa_R)$ , respectively.

Using the estimated parameters shown in the first column of Table II, the value of  $R_t$  is (the interquartile range in parenthesis)

$$R_t = 1.4848 \text{ (1.4798 – 1.4903)}.$$

TABLE II. Best fit parameter values

Parameter	Median (interquartile range) (No metapopulation)	Median (interquartile range) (Metapopulation w/o mobility)	Median (interquartile range) (Metapopulation w/ mobility)
$\beta_{IS}$	0.475749 (0.439256–0.497102)	0.428599 (0.422961–0.446900)	0.394443 (0.274636–0.511767)
$\beta_{AS}$	0.483056 (0.469496–0.506530)	0.490886 (0.469169–0.492347)	0.412055 (0.287619–0.513972)
$\kappa_A$	0.338381 (0.333010–0.347806)	0.493038 (0.469970–0.494282)	0.313193 (0.254518–0.372569)
$\kappa_I$	0.280513 (0.270371–0.287488)	0.426593 (0.377388–0.453029)	0.278822 (0.227327–0.331485)
$\gamma_A$	0.317751 (0.308228–0.324681)	0.461193 (0.421694–0.474062)	0.318094 (0.262533–0.384494)
$\kappa_R$	0.334552 (0.326885–0.345297)	0.192424 (0.191560–0.201661)	0.320300 (0.242333–0.417852)
$\kappa_D$	0.000106 (0.000102–0.000113)	0.000056 (0.000054–0.000060)	0.000455 (0.000287–0.000558)
$E_0/I_0$	1.59 (1.57–1.61)	—	—
$A_0/I_0$	1.60 (1.58–1.63)	—	—
$I_0$	587.05 (582.71–591.87)	—	—

TABLE III. Initial conditions

Province	$I_0$	$E_0$	$A_0$	$N_0 = S_0$
Almeria	175	278	281	727945
Cadiz	26	41	42	1244049
Cordoba	9	15	15	781451
Granada	246	391	395	919168
Huelva	7	11	11	524278
Jaen	14	23	23	631381
Malaga	95	150	152	1685920
Seville	14	23	23	1950219

TABLE IV. Residuals

Province	No metapopulation		Metapopulation without mobility		Metapopulation with mobility	
	Fitting	Prediction	Fitting	Prediction	Fitting	Prediction
Almeria	—	—	0.074 (0.074–0.074)	0.855 (0.852–0.859)	0.497 (0.444–0.776)	1.620 (1.121–2.653)
Cadiz	—	—	0.169 (0.169–0.169)	0.232 (0.232–0.235)	0.095 (0.076–0.111)	0.535 (0.314–0.852)
Cordoba	—	—	0.227 (0.227–0.228)	2.097 (2.095–2.102)	0.059 (0.043–0.071)	0.512 (0.334–1.081)
Granada	—	—	0.266 (0.266–0.267)	0.686 (0.685–0.691)	0.041 (0.026–0.047)	0.147 (0.113–0.333)
Huelva	—	—	0.406 (0.403–0.406)	1.821 (1.813–1.824)	0.584 (0.513–0.829)	3.595 (2.610–4.605)
Jaen	—	—	0.019 (0.019–0.019)	0.532 (0.531–0.532)	0.126 (0.116–0.131)	0.304 (0.260–0.496)
Malaga	—	—	0.005 (0.005–0.005)	0.216 (0.216–0.218)	0.052 (0.040–0.061)	0.254 (0.154–0.509)
Sevilla	—	—	0.168 (0.168–0.169)	1.194 (1.193–1.199)	0.030 (0.024–0.047)	0.110 (0.063–0.436)
Andalusia	0.00069 (0.00068–0.00083)	8.752 (8.603–8.942)	0.006 (0.006–0.006)	0.434 (0.433–0.435)	0.012 (0.007–0.018)	0.129 (0.060–0.412)

We calculated  $R_t$  based on the 500 sets of parameter values, see the discussion following Eq. (20). The interquartile range was calculated as follows. First, using the 500 accepted sets of the model parameters, we calculate  $R_t$ . From those values, we then obtain the lower and upper quartiles and the median. When the metapopulation model is used without mobility ( $\theta = 0$ ), then the same expression, Eq. (23), applies with the parameters of the second column yielding

$$R_t = 1.5972 (1.5945 - 1.5983).$$

Finally, for the entire metapopulation network,  $R_t$  is calculated as follows. We define the relevant vectors, focusing on the

infectious/infected compartments ( $E_i, A_i, I_i$ ) and ignoring the rest ( $S_i, U_i, R_i, D_i$ ):

$$\mathcal{F} = \begin{pmatrix} \frac{\beta_{AS}}{N_1} S_1 A_1 + \frac{\beta_{IS}}{N_1} S_1 I_1 \\ 0 \\ 0 \\ \vdots \\ \frac{\beta_{AS}}{N_8} S_8 A_8 + \frac{\beta_{IS}}{N_8} S_8 I_8 \\ 0 \\ 0 \end{pmatrix}, \quad \mathcal{V} = \begin{pmatrix} (\kappa_A + \kappa_I) E_1 - \theta \left( \sum_j M_{1j} \frac{E_j}{N_j} - \sum_j M_{j1} \frac{E_1}{N_1} \right) \\ -\kappa_A E_1 + \gamma_A A_1 - \theta \left( \sum_j M_{1j} \frac{A_j}{N_j} - \sum_j M_{j1} \frac{A_1}{N_1} \right) \\ -\kappa_I E_1 + (\kappa_R + \kappa_D) I_1 - \theta \left( \sum_j M_{1j} \frac{I_j}{N_j} - \sum_j M_{j1} \frac{I_1}{N_1} \right) \\ \vdots \\ (\kappa_A + \kappa_I) E_8 - \theta \left( \sum_j M_{8j} \frac{E_j}{N_j} - \sum_j M_{j8} \frac{E_8}{N_8} \right) \\ -\kappa_A E_8 + \gamma_A A_8 - \theta \left( \sum_j M_{8j} \frac{A_j}{N_j} - \sum_j M_{j8} \frac{A_8}{N_8} \right) \\ -\kappa_I E_8 + (\kappa_R + \kappa_D) I_8 - \theta \left( \sum_j M_{8j} \frac{I_j}{N_j} - \sum_j M_{j8} \frac{I_8}{N_8} \right) \end{pmatrix}$$

We then find the Jacobian matrices of  $\mathcal{F}, \mathcal{V}$  with respect to  $E_i, A_i, I_i$  in the order in which they appear. This yields two  $24 \times 24$  matrices of the form:

$$F = \begin{pmatrix} F_{11} & O_{3 \times 3} & O_{3 \times 3} & \dots & O_{3 \times 3} \\ O_{3 \times 3} & F_{22} & O_{3 \times 3} & O_{3 \times 3} & O_{3 \times 3} \\ O_{3 \times 3} & O_{3 \times 3} & F_{33} & O_{3 \times 3} & O_{3 \times 3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ O_{3 \times 3} & O_{3 \times 3} & O_{3 \times 3} & \dots & F_{88} \end{pmatrix}, \quad F_{ii} = \begin{pmatrix} 0 & \beta_{AS} \frac{S_i}{N_i} & \beta_{IS} \frac{S_i}{N_i} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad O_{3 \times 3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$V = \begin{pmatrix} V_{11} & V_{12} & \dots & V_{18} \\ V_{21} & V_{22} & \dots & V_{28} \\ \vdots & \vdots & \ddots & \vdots \\ V_{81} & V_{82} & \dots & V_{88} \end{pmatrix}, \quad V_{ii} = \begin{pmatrix} \kappa_A + \kappa_I + \theta \sum_j \frac{M_{ji}}{N_i} & 0 & 0 \\ -\kappa_A & \gamma_A + \theta \sum_j \frac{M_{ji}}{N_i} & 0 \\ -\kappa_I & 0 & \kappa_R + \kappa_D + \theta \sum_j \frac{M_{ji}}{N_i} \end{pmatrix}$$

$$V_{ij} = \begin{pmatrix} -\theta \frac{M_{ij}}{N_j} & 0 & 0 \\ 0 & -\theta \frac{M_{ij}}{N_j} & 0 \\ 0 & 0 & -\theta \frac{M_{ij}}{N_j} \end{pmatrix}$$

We note that for the calculation at  $t = 0$ , we set  $\frac{S_i(0)}{N_i(0)}$  in the  $F_{ii}$  matrices and  $\frac{M_{ji}(0)}{N_i(0)}, \frac{M_{ij}(0)}{N_j(0)}$  in  $V_{ii}, V_{ij}$ , respectively.

The reproduction number is the spectral radius of  $FV^{-1}$  which in our case has the value

$$R_t = 1.3349 (1.2806 - 1.4581).$$

This is exactly the same value one obtains when using Eq. (23) if one completely disregards the mobility terms in matrix  $V$ , i.e., with the parameter values of the third column in Table II. Hence, the change in  $R_t$  is due to the different values in third column of Table II, and not due to the terms containing  $M_{ij}$  in matrix  $V$ . In other words, the effect of mobility is to change the estimated parameters; while they do not alter the  $R_t$  (which, as mentioned earlier is effectively evaluated at the first day of our simulations), they have a large effect on the dynamics later. When mobility is included in the model, the interquartile intervals of each parameter value are significantly wider, which is also reflected in the corresponding interval for  $R_t$ .

#### IV. DISCUSSION-CONCLUSION

In the present work, we revisited the formulation of metapopulation models, motivated by the interest towards describing a “relatively small” region (the autonomous community of Andalusia within Spain) with well-defined and available in a time-resolved manner data regarding the mobility across provinces. It is also a region without an extensive influx (or outflux) of populations, e.g., through major international airport hubs. This appears to render this case a fertile ground for the application of metapopulation models.

In that vein, in addition to a prototypical model for each node, involving susceptibles, exposed, asymptomatic and symptomatically infected, as well as recovered from each of these categories and fatalities, we considered different possibilities on how to incorporate human mobility across the nodes. We explored the model for the entire autonomous community of Andalusia (without sub-nodes), the model where the nodes do not feature mobility between them (independent nodes) and the canonical case proposed where mobility is incorporated. One of the main findings of the present work is that in the absence of mobility

among nodes the model is unable to predict the wave of infections that took place in the fall of 2020. It has long been known that human mobility is crucial at the beginning stages of an epidemic, when the infection is seeded in various locations [9, 28, 45]. It has also been noted that mobility may also affect contact rates which in turn affect disease transmission [45]. The present study suggests that population flows are critically important in periods during an epidemic when there are no restrictions on mobility. Moreover, while there are numerous ways of incorporating mobility, for example via static origin-destination matrices as calculated via gravity models, we believe that at present the optimal inclusion should be time-resolved. Dynamical information stemming from mobile-phone data seamlessly incorporates aspects such as the weekly or seasonal variations of human mobility; hence it more accurately captures the resulting increases or decreases in the probability of formation of an epidemic wave of infection. However, when this is not possible, we also offer details on how origin-destination matrices obtained by the gravity-law can be calculated to be used in a metapopulation model.

Nevertheless, we certainly refrain from assigning full responsibility to human mobility for the wave of infections in the fall of 2020, or indeed more generally during the second wave of the pandemic. It is clear that there exist numerous factors that may have contributed to the relevant features, including, e.g., seasonality [13] and humidity [19]. It would be interesting to further explore these factors and their interplay with mobility both in the context of the second wave (as here) in other regions, but also as concerns subsequent waves of the pandemic, where other key factors, such as the existence and the role of vaccinations [40] need to be taken into consideration. Such studies will be deferred to future publications.

#### ACKNOWLEDGMENTS

ZR, EK, SL, PGK, MB and GAK acknowledge support through the C3.ai Inc. and Microsoft Corporation. ZR also acknowledges support from the NSF through grant DMS-1815764. JCM acknowledges support from EU (FEDER program 2014-2020) through both Consejería de Economía, Conocimiento, Empresas y Universidad de la Junta de Andalucía (under the projects P18-RT-3480 and US-1380977), and MCIN/AEI/10.13039/501100011033 (under the project PID2020-112620GB-I00).

#### DATA AVAILABILITY STATEMENT

The datasets and codes generated and/or analyzed in this study are available at <https://gitlab.com/peacogr/metapopulation>.



## Appendix A: Origin-Destination matrices

We describe the methodology used to compute origin-destination (O-D) matrices for light-duty vehicle highway travels compatible with empirical data of coarse grained flows in highway transportation networks. We obtained the O-D matrices either via the gravity model or from the mobile-phone data reported by the Ministry of Transportation, Mobility and Urban Agency (Ministerio de Transportes, Movilidad y Agenda Urbana-MITMA) [32] (see, also the discussion in Section II C in the main text). Specifically, we constructed the O-D matrix as follows: given population data from the Instituto Nacional de Estadística [17] and distances of city pairs we applied the gravity law to calculate the vehicle number of trips per day between any two nodes. The gravity-model methodology is presented in Section A 1. Section A 2 describes briefly how we translated the mobile-phone data into population flows, people traveling per day between provinces. Figure 3 in the main text compares the static, pre-pandemic, gravity-model calculated O-D matrix with time-dependent O-D matrices obtained from mobile-phone data for two characteristic working days: Friday, July 10, 2020 and Tuesday, October 27, 2020.

### 1. Methodology: Gravity law and traffic counts

Given the population of provinces and distances between the capital city of province pairs, the gravity model as described by Eq. (18) in the main text can be leveraged with the parameters presented in Table V. When the distance between an O-D pair ( $i, j$ ) is larger than 300 km,  $\beta_G$  is set to be N/A, which denotes that the denominator will be approaching 0, as well as  $w_{ij}$  would be approaching 0.

TABLE V. Parameters of gravity law [4, 38]

$d(\text{km})$	Parameter	Estimate
$\leq 300$	$\alpha$	0.46
	$\gamma$	0.64
	$\beta_G$	0.0122
$> 300$	$\alpha$	0.35
	$\gamma$	0.37
	$\beta_G$	N/A
	$C$	0.04289

The Origin-Destination matrix for the eight Andalusian provinces calculated from the gravity law with the parameters reported in Table V is presented in Table VI, which summarizes the daily number of vehicle trips between O-D pairs ( $i, j$ ). As mentioned in the main text this O-D matrix is static, and the data we used referred to 2019-2020, i.e., before the pandemic started.

TABLE VI. O-D matrix (daily number of vehicle trips) based on the gravity law.

	Almeria	Cadiz	Cordoba	Granada	Huelva	Jaen	Malaga	Sevillea
Almeria	0	0	0	4139	0	873	4863	0
Cadiz	0	0	2391	0	2645	0	4704	21312
Cordoba	0	2536	0	3323	1658	4211	10110	14025
Granada	4018	0	3531	0	0	7051	13425	3253
Huelva	0	3248	1920	0	0	0	0	17683
Jaen	964	0	5092	8022	0	0	3728	2612
Malaga	4015	4509	9138	11417	0	2787	0	9059
Seville	0	19750	12257	2675	13346	1888	8758	0

## 2. Methodology: Mobile-phone data

All mobility data used in this work are available in the website referenced in [32]. The database aggregates more than 13 million anonymized mobile lines to provide data for the number of persons traveling per day, i.e., population flows. The procedure we followed to determine  $M_{ij}$ , the daily rate of people traveling from province  $i$  to province  $j$ , was to download the spreadsheet, and then to select the origin and destination of a trip according to following codes: 04 for Almería, 11 for Cádiz, 14 for Córdoba, 18 for Granada, 21 for Huelva, 23 for Jaén, 29 for Málaga, and 41 for Sevilla. The trip date is under the column “fecha” and the trips, one per individual, under the column “viajes”. This information is extracted and assembled into the  $M_{ij}(t)$  matrices for day  $t$  that provide the desired population flows as persons traveling per day.

- 
- [1] S. Ando, Y. Matsuzawa, H. Tsurui, T. Mizutani, D. Hall, and Y. Kuroda. Stochastic modelling of the effects of human-mobility restriction and viral infection characteristics on the spread of covid-19. *Scientific Reports*, 11(1):6856, 2021.
  - [2] A. Arenas, W. Cota, J. Gomez-Gardeñes, S. Gomez, S. Granell, J. T. Matamalas, D. Soriano-Panos, and B. Steinegger. Modeling the spatiotemporal epidemic spreading of COVID-19 and the impact of mobility and social distancing interventions. *Physical Review X*, 10:041055, 2020.
  - [3] H.S. Badr, H. Du, M. Marshall, E. Dong, M.M. Squire, and L.M. Gardner. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(11):1247–1254, 2020.
  - [4] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the USA*, 106:21484–21489, 2009.
  - [5] V. Belik, T. Geisel, and D. Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X*, 1:011001, 2011.
  - [6] F. Calleri, G. Nastasi, and V. Romano. Continuous-time stochastic processes for the spread of covid-19 disease simulated via a monte carlo approach and comparison with deterministic models. *Journal of Mathematical Biology*, 83(4):1–26, 2021.
  - [7] D. Calvetti, A.P. Hoover, J. Rose, and E. Somersalo. Metapopulation network models for understanding, predicting, and managing the coronavirus disease COVID-19. *Frontiers in Physics*, 8:261, 2020.
  - [8] D. Chen. Modeling the spread of infectious diseases: A review. in: *Analyzing and modeling spatial and temporal dynamics of infectious diseases*, pages 19–42, 2014.
  - [9] M. Chinazzi, J.T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M.E. Halloran, I.M. Longini, and A. Vespignani. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368:395–400, 2020.
  - [10] A. Cintron-Arias, C. Castillo-Chavez, L. M. A. Bettencourt, A. L. Lloyd and H. T. Banks. The estimation of the effective reproductive number from disease outbreak data. *Mathematical Biosciences and Engineering*, 6:261–282, 2009.
  - [11] G. S. Costa, W. Cota, and S. C. Ferreira. Outbreak diversity in epidemic waves propagating through distinct geographical scales. *Physical Review Research*, 2:043306, 2020.
  - [12] J. Cuevas-Maraver, P.G. Kevrekidis, Q.Y. Chen, G.A. Kevrekidis, Z. Rapti, and Y. Drossinos. Lockdown measures and their impact on single- and two-age-structured epidemic model for the COVID-19 outbreak in Mexico. *Mathematical Biosciences*, 336:108590, 2021.
  - [13] L. Danon, E. Brooks-Pollock, M. Bailey, and M. Keeling. A spatial model of COVID-19 transmission in England and Wales: early spread, peak timing and the impact of seasonality. *Philosophical Transaction of the Royal Society B*, 376:20200272, 2021.
  - [14] L. Danon, T. House, and M.J. Keeling. The role of routine versus random movements on the spread of disease in Great Britain. *Epidemics*, 1:250–258, 2009.
  - [15] Ministerio de Ciencia e Innovacion. COVID-19 en Espana. <https://cnecovid.isciii.es/covid19/>.
  - [16] J. Ortúzar and L.G. Willumsen. *Modelling transport*. John Wiley & Sons, 2011.
  - [17] Instituto Nacional de Estadística. <https://www.ine.es/en/>.
  - [18] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28:365–382, 1990.
  - [19] Y. Drossinos, J. P. Reid, W. Hugentobler, and N. I. Stilianakis. Challenges of integrating aerosol dynamics into SARS-CoV-2 transmission models. *Aerosol Science and Technology*, 56:777–784, 2022.
  - [20] S. Erlander and N.F. Stewart. *The gravity model in transportation analysis: theory and extensions*, volume 3. Vsp, 1990.
  - [21] E. L. Glaeser, G. Gorbach, and S. J. Redding. Jue Insight: How much does COVID-19 increase with mobility? Evidence from New York and four other U.S. cities. *Journal of Urban Economics*, 127:103292, 2022.
  - [22] S. Gomez, A. Fernandez, S. Meloni, and A. Arenas. Impact of origin-destination information in epidemic spreading. *Scientific Reports*, 9:2315, 2019.
  - [23] I. Holmdahl and C. Buckee. Wrong but useful- what COVID-19 epidemiological models can and cannot tell us. *The New England Journal of Medicine*, 383(4):303–305, 2020.
  - [24] M. J. Keeling and C. A. Gilligan. Metapopulation dynamics of bubonic plague. *Nature*, 407:903–906, 2000.
  - [25] M. J. Keeling and P. Rohani. Estimating spatial coupling in epidemiological systems: a mechanistic approach. *Ecology Letters*, 5:20–29, 2002.
  - [26] G. A. Kevrekidis, Z. Rapti, Y. Drossinos, P.G. Kevrekidis, M. A. Barmann, Q. Y. Chen, and J. Cuevas-Maraver. Backcasting COVID-19: A physics-informed estimate for early case incidence. <https://arxiv.org/pdf/2202.00507.pdf>, 2022.

- [27] P. G. Kevrekidis, J. Cuevas-Maraver, Y. Drossinos, Z. Rapti, and G. A. Kevrekidis. Reaction-diffusion spatial modeling of COVID-19: Greece and Andalusia as case examples. *Physical Review E*, 104:024412, 2021.
- [28] M.U.G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D.M. Pigott, L. du Plessis, N.R. Faria, R. Li, W.P. Hanage, J.S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O.G. Pybus, and S.V. Scarpino. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368:493–497, 2020.
- [29] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman. Substantial undocumented infections facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368:489–493, 2020.
- [30] Y. Mammeri. A reaction-diffusion system to better comprehend the unlockdown: Application of SEIR-type model with diffusion to the spatial spread of COVID-19 in France. *Computational and Mathematical Biophysics*, 8:102–113, 2020.
- [31] H. McCallum, N. Barlow, and J. Hone. How should pathogen transmission be modelled? *Trends in ecology & evolution*, 16(6):295–300, 2001.
- [32] Movilidad y Agenda Urbana Ministerio de Transportes. Estudio de movilidad con big data. <https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data>.
- [33] S. Pei, S. Kandula, W. Yang, , and J. Shaman. Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences of the USA*, 115:2752–2757, 2018.
- [34] M. Peirlinck, K. Linka, F. Sahli Costabal, J. Bhattacharya, E. Bendavid, J. A. P. Ioannidis, and E. Kuhl. Visualizing the invisible: The effect of asymptomatic transmission on the outbreak dynamics of COVID-19. *Computer Methods in Applied Mechanics and Engineering*, 2020.
- [35] B. Prasse, M. A. Achterberg, and P. V. Mieghem. Accuracy of predicting epidemic outbreaks. *Physical Review E*, 105:014302, 2022.
- [36] M. Schlapfer, L. Dong, K. O’Keeffe, P. Santi, M. Szell, H. Salat, S. Anklesaria, M. Vazifeh, C. Ratti, and G. B. West. The universal visitation law of human mobility. *Nature*, 593:522–540, 2021.
- [37] F. Simini, M. C. Gonzalez, , A. Maritan, and A.-L. Barabasi. A universal model for mobility and migration patterns. *Nature*, 484:96–100, 2012.
- [38] M. Stefanouli and S. Polyzos. Gravity vs radiation model: two approaches on commuting in Greece. *Transportation research procedia*, 24:65–72, 2017.
- [39] M. Tizzoni, P. Bajardi, Decuyper A., G. K. Kam King, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. Gonzalez, and V. Colizza. On the human mobility proxies for modeling epidemics. *PLoS Computational Biology*, 10:e1003716, 2014.
- [40] T. Usherwood, Z. LaJoie, and V. Srivastava. A model and predictions for COVID-19 considering population behavior and vaccination. *Scientific Reports*, 11:12051, 2021.
- [41] H.J. Van Zuylen and L.G. Willumsen. The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological*, 14(3):281–293, 1980.
- [42] C. Viboud, O. N. Bjornstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312:447–451, 2006.
- [43] A. Viguerie, G. Lorenzo, F. Auricchio, D. Baroli, T. J. R. Hughes, A. Patton, A. Reali, T. E. Yankeelov, and A. Veneziani. Simulating the spread of COVID-19 via a spatially-resolved susceptible–exposed–infected–recovered–deceased (SEIRD) model with heterogeneous diffusion. *Applied Mathematics Letters*, 111:106617, 2021.
- [44] C. R. Wells, P. Sah, S. M. Moghadas, A. Pandey, A. Shoukat, Y. Wang, Z. Wang, L. A. Meyers, B. H. Singer, and A. P. Galvani. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proceedings of the National Academy of Sciences of the USA*, 117:7504–7509, 2020.
- [45] A. Wesolowski, C. O. Buckee, K. Engo-Mongen, and C. J. E. Metcalf. Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *The Journal of Infectious Diseases*, 214:S414–420, 2016.
- [46] L. G. Willumsen. Simplified transport models based on traffic counts. *Transportation*, 10:257–278, 1981.
- [47] Y. Xia, O. N. Bjornstad, and B. T. Grenfell. Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *The American Naturalist*, 164(2):267–281, 2004.
- [48] T. Yabe, K. Tsubouchi, N. Fujiwara, T. Wada, Y. Sekimoto, and S.V. Ukkusuri. Non-compulsory measures sufficiently reduced human mobility in tokyo during the covid-19 epidemic. *Scientific reports*, 10(1):1–9, 2020.
- [49] G. K. Zipf. The P1 P2/D hypothesis: on the intercity movement of persons. *American Sociological Review*, 11:677–686, 1946.