Human mobility and epidemics

Giulia Pullano

THESE DE DOCTORAT DE
SORBONNE UNIVERSITE

Spécialité

Biomathématiques

L'Ecole Doctorale Pierre Louis de Santé Publique : Epidémiologie et Sciences de l'Information Biomédicale

Présentée par

Mme. Giulia Pullano

Pour obtenir le grade de

DOCTEUR de SORBONNE UNIVERSITE

Sujet de la thèse :

Mobilité humaine et propagation des épidémies

soutenue le 28/09/2021

devant le jury composé de :

Mme. Vittoria COLIZZA (Inserm, Sorbonne Université)	Directeur de thèse
Mme. Stefania RUBRICHI (Orange S.A.)	Co-encadrant
Mme. Shweta BANSAL (Georgetown University)	Rapporteurs
M. Marco FIORE (IMDEA Networks Institute)	Rapporteurs
M. Alain BARRAT (CNRS, Aix-Marseille Université)	Examinateurs

Sorbonne Université
Bureau d'accueil, inscription des doctorants et base de données
Esc G, 2ème étage
15 rue de l'école de médecine
75270-PARIS CEDEX 06

Tél. Secrétariat : 01 42 34 68 35
Fax : 01 42 34 68 40
Tél. pour les étudiants de A à EL : 01 42 34 69 54
Tél. pour les étudiants de EM à MON : 01 42 34 68 41
Tél. pour les étudiants de MOO à Z : 01 42 34 68 51
E-mail : scolarite.doctorat@upmc.fr

Contents

Contents iii

1 Introduction 1

2 The role of human mobility in infectious disease transmission 5
2.1 International travels and the global spread . . . 5
2.2 Recurrent mobility and national spread . . . 12
2.3 Social mixing and local spread . . . 18
2.4 Conclusions . . . 21

3 Mobile phone data and epidemic spread 23
3.1 Mobile phone data for epidemiological purposes . . . 23
3.2 Mobile phone data to fight emergencies in real-time . . . 25
3.3 Type and sources of mobile phone data . . . 26
3.4 Data access and privacy issues . . . 28
3.5 Extraction of mobility indicators from mobile phone data . . . 29
3.6 Conclusions . . . 32

4 Integrating mobility into epidemic models 33
4.1 Basic and age-structured compartmental models . . . 34
4.2 Compartmental models specific for COVID-19 . . . 35
4.2.1 Parametrization of contact matrices . . . 38
School attendance . . . 38
Presence at work . . . 39
Adoption of physical distancing . . . 40
4.3 Metapopulation models . . . 41
4.4 Modeling risk of introduction . . . 46
4.5 Mobility definition and resolution needed to inform metapopulation models . . . 47
4.5.1 Statistical comparison of coupling matrices . . . 48
4.5.2 The aggregation approach affects the invasion dynamics . . . 53
Improving coupling matrices . . . 57
4.5.3 Discussion . . . 59
4.6 Conclusions . . . 60

5 Travel bans and global epidemic importations of COVID-19 61
5.1 Early stage COVID-19 pandemic . . . 61
5.2 Article #1: Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020 . . . 62
5.3 Conclusions . . . 68

6 Parametrizing social mixing from mobility data to model the first wave of COVID-19 in France **71**
6.1 Social mixing reductions due to control strategies 71
6.2 Article #3: Underdetection of COVID-19 cases in France threatens epidemic control 72
6.3 Conclusions 90

7 Conclusions and Perspectives **93**

Bibliography **97**

List of Figures

2.1 Illustrative maps of multiscale structure of human mobility. 6
2.2 Epidemic acceleration. 8
2.3 The increase of global connectivity and air travel. 9
2.4 Properties of the worldwide air traffic. 10
2.5 The spread of the Black Death and of the 2009 A/H1N1 pandemic. . . 11
2.6 Individual mobility patterns. 13
2.7 Early-stage COVID-19 transmission in China. 14
2.8 Comparison of GLEAM predictions at the global and regional level obtained with and without commuting flows. 15
2.9 Comparison of GLEaM predictions at the local level obtained with and without commuting. 16
2.10 Contact Matrices. 19

3.1 Mobile phone penetration. 23
3.2 Individual trajectories inferred by mobile phone data. 27
3.3 The distance decays of taxicab trips and mobile phone displacements. 30

4.1 Compartmental model scheme. 35
4.2 Percentage of students going to school. 39
4.3 Percentage of individuals not going to workplace locations. 40
4.4 Percentage of individuals avoiding physical contact 41
4.5 Schematic illustration of metapopulation scheme. 42
4.6 The aggregation of the individual trajectories. 49
4.7 Coupling forces. 51
4.8 Differences between coupling forces. 52
4.9 Differences on modelled epidemics. 54
4.10 Epidemic invasion trees. 56
4.11 Sensitivity. 58

List of Tables

4.1 Parameters and values to define the COVID-19 age-structured transmission model. 37

Chapter 1

Introduction

What is the role of human mobility in infectious disease transmission? Human mobility affect the mixing among populations and thus crucially alter the probability of coming in contact with infected individuals and the likelihood of disease propagation [1, 2]. Infected individual who travels may also introduce pathogens and generate local clusters in not-affected locations, or may lead to re-seeding events in places with low or no epidemic activity.

The availability in the last twenty years of mobility data ranging from international air traffic data, census commuting data to daily individual trajectory data has widely contributed to a better understanding of human mobility patterns and their impact on the epidemic spread. Targeted models spanning several levels of complexity - from homogenous mixing populations to explicit contact and mobility structures - have been developed and informed by mobility data [3, 4, 5, 6, 7, 8, 9]. Researchers found that mobility traffic spans several orders of magnitude in intensity and spatio-temporal scales. Individual travels indeed are mainly characterized by recurrent short movements to few locations, and few international long-range travels [5, 7, 10]. The collective effects of the sum of the individual trajectories generate coupling forces among locations, which lead to a preferential path of transmission [5, 11]. It was also observed how the heterogeneity of the coupling among locations has a crucial influence on the speed of disease spatial transmission [12]. Epidemiological studies aimed to understand the role of human mobility are mostly based on mobility fluxes defined from theoretical models, static mobility data such as commuting census data or mobility data collected during peacetime.

However, patterns of mobility may be altered during an ongoing epidemic. Individuals may change their behaviours due to: i) mobility restrictions put in place by governments to mitigate the epidemic activity; or ii) individual adaptive behaviours to the epidemic, like risk aversion. In recent years, researchers have had the intuition that mobile phone data could be used to monitor the population's behaviours in response to emergencies [13, 14]. For instance, they used mobile phone traces to highlight the role of mass gatherings on the spread of waterborne diseases like cholera [15], or quantified seasonal population fluxes in Kenya to assess how these drive rubella transmission dynamics [16]. Moreover, researchers highlighted the potential of mobile phones to track individual behaviors to assess the impact of mobility restrictions enforced to mitigate the Ebola Virus epidemic in Sierra Leone [17].

Dealing with such sensitive data led to many issues on the privacy, resulting in complications in legal agreements which often delay data accessibility [18, 19]. The reluctance of data sharing is also increased with the raised awareness of the limits of data anonymization procedures [19, 20].

The COVID-19 health global crisis has underlined the necessity for rapid access to mobility data in order to help mitigate the viral diffusion [21]. Network operators and companies across the world made thus huge efforts to quickly pre-process and

share their data through legally and ethically compliant agreements. Mobile phone data have been shared at national level, and they have been largely used to track individual behaviours within countries [11, 13, 14, 15, 22, 23, 24, 25, 26, 27, 28, 29], while air traffic data have been mainly used to monitor international mobility fluxes over time and global importations [4, 5, 10, 30].

This fast data availability on a large scale and high-resolution allowed researchers to quantify mobility changes since the early stage of the COVID-19 pandemic. For instance, during the first national lockdown in France, trips were reduced by 65% [31]. Similar results were found also in Belgium [32], Spain [33], and Italy [34]. In addition, due to risk aversion individuals living in regions with higher incidence showed a higher reduction of mobility during mobility restrictions [17, 31], generating mobility changes heterogeneous in space. Spatial heterogeneities also depend on socio-economic constraints as the standard of living, labor structure, household crowding and retail stores availability [31, 35].

The massive and detailed information of mobility in space and time we collected during COVID-19 has thus opened new challenges on i) quantifying mobility restrictions in terms of mobility reductions; ii) integrating real-time mobility data into models, in order to increase their predictive power by accounting for mobility changes.

My doctoral thesis deal with theoretical and applied research work aimed at integrating mobility data on different spatial scales into mathematical models for public heath applications. It is an interdisciplinary research work of mathematical and computational modeling, data science, and epidemiology. I developed my doctoral work at Institut Pierre Louis d'Epidémiologie et de Santé Publique, INSERM, Sorbonne Université, Paris, France, at EPIc Lab within the team of the Surveillance et Modélisation des maladies transmissibles (SUMO), supervised by Dr. Vittoria Colizza. My PhD thesis was promoted and funded by Orange SA, within the framework of French CIFRE (Convention Industrielle de Formation par la Recherche) fellowships, under the supervision of Dr. Stefania Rubrichi. Thanks to this collaboration between Orange and Inserm, I had the possibility to access and analyse mobile phone traces. Since January 2020, I have dedicated my research exclusively to the COVID-19 epidemic. In this context, we collaborated with Santé publique France [36] and Réseau Sentinelles group [37] in the assessment of the COVID-19 epidemic in France. Our group helped inform public health and political authorities on the situation and efficacy of interventions.

The work is organized as follows. In Chapter 2 I discuss how on large, medium, and short scales mobility affects epidemic spread, leading to the high predictability of the transmission of diseases. I explain, how international travel drives the invasion of pathogens at a global level, while recurrent daily mobility affects epidemic peak timing and incidence at local and national level.

In Chapter 3 I discuss the use of mobile phone data to quantify mobility fluxes. Particular attention is devoted to the use of such data in fighting emergencies in real-time, as in the context of COVID-19 pandemic. The different types and sources of these data are described, and I discuss the privacy issues in dealing with such sensitive data. In the last Section of the Chapter, I review the aggregation process for computing mobility indicators from individual trajectories extracted from mobile phone data.

Chapter 4 is dedicated to describing the theoretical frameworks to integrate mobility into epidemic models. First, I review the cornerstones of disease transmission modelling, introducing basic and age-structured compartmental models and I

present the age-structured transmission model we designed and developed specifically to fight the COVID-19 epidemic in France. Particular attention is devoted to the innovative method we proposed to parametrize social mixing through mobile phone data in order to account for the evolving individual behaviours due to mobility restrictions and adaptive behaviours to the epidemic.

Then, I describe the metapopulation frameworks largely used by researchers to model explicitly spatial structures and mobility among populations. The attention is focused on the transmission model we designed and developed to assess how to integrate mobile phone data at a national level into metapopulation models. Then, I present how we quantified the risk of the introduction of a new pathogen carried by international travellers using air traffic data. To conclude, I present an original work issued from my thesis on assessing the appropriate definition and resolution of mobility at national level to inform metapopulation models from mobile phone data.

In Chapters 5 and 6, I present the two main research works that I performed during the COVID-19 pandemic, considering the impact of human mobility across different scales. Our work on the COVID-19 pandemic resulted in 8 published papers, 1 as first author and 2 as co-first author. In my thesis I selected two of them for presentation: Pullano, Giulia, et al. "Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020.", *Eurosurveillance* (2020) and Pullano, Giulia et al. "Underdetection of cases of COVID-19 in France threatens epidemic control", *Nature* (2021). In the first one, we computed the risk of global epidemic importations of COVID-19 into Europe from China due to international travellers. In the second one, we estimated the underdetection of COVID-19 cases after the first wave, when a new surveillance system was put in place to systematically detect all cases.

Chapter 2

The role of human mobility in infectious disease transmission

In the last decades there are several examples of emerging or endemic diseases globally and locally transmitted by mobility fluxes including the ongoing COVID-19 [38, 39, 40, 41], influenza [10, 42, 43, 44, 45], Ebola virus [17, 46], Middle East Respiratory Syndrome (MERS) [47], Severe Acute Respiratory Syndrome (SARS) [6, 48], cholera [15, 49], dengue [50] and malaria [51, 52]. Mobility fluxes affect the infectious disease transmission by altering the time of arrival of the infection in previously not-affected locations, their incidence and the frequency of the epidemics [5, 7, 10, 11, 22, 24, 53, 54, 55]. In this Chapter, I will explain in detail the role of mobility in shaping epidemics in space and time, providing examples on past and ongoing epidemics.

Mobility traffic spans several orders of magnitude in intensity and spatio-temporal scales ranging from short-range recurrent patterns such as commuting fluxes at national level [22, 24, 55], to international long-range trips. (Fig 2.1) [5, 7, 10], and such patterns play a different role in infectious disease transmission. The Chapter is therefore divided into three different sections. In Section 2.1, I will explain the role of international travelers in spreading infectious diseases globally. In Section 2.2, I will describe the role of national internal mobility in spreading epidemics within the country once first cases have emerged. To conclude, in Section 2.3, I will focus on how social mixing among individuals generated by individual movements affects spread at local level.

2.1 International travels and the global spread

Human travels and migrations are the major drivers of carrying infections worldwide. Given the improvements in transportation infrastructures and means, nowadays big cities across the world are strongly connected among themselves and to the neighbouring locations, allowing travellers to move from one location on earth to any other point on the planet in only 1/2 days. This leads to a continuous social mixing among subpopulations geographically far from each other, thus creating a faster epidemic spatial transmission among countries.

International travel may contribute to localized or global epidemics

Once a new pathogen is introduced into a location by travellers, infections may lead to global epidemics such as COVID-19 or H1N1 influenza, causing localized epidemics such as Schistosomiasis in Corse or not generate epidemic clusters, such Ebola when it was introduced into some countries in Europe. In this Section, I will present the most relevant recent human diseases transmitted by vectors, or by

human-to-human, and I will discuss how they spread across countries driven by mobility.

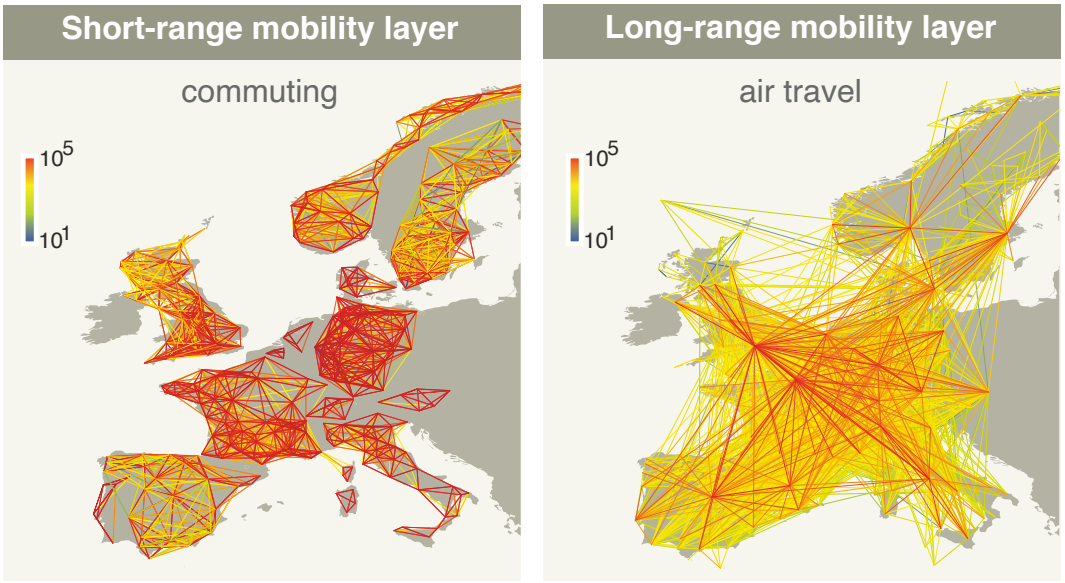


FIGURE 2.1: Illustrative maps of multiscale structure of human mobility in Europe. The short-range mobility layer covers commuting fluxes among municipalities and the long-range mobility layer covers the air travel flows between worldwide airport pairs connected. The colour bar represents mobility traffic i.e. the number of people moving among the connected locations. Figure from [56].

Vectorborne diseases. Vectorborne diseases generate predominantly localized epidemics as they are transmitted by the bite of infected species and may generate epidemics only in environments that have suitable conditions for vectors to survive and reproduce. The two vectorborne diseases with the highest prevalence in the world are Malaria and Schistosomiasis [8]. In the context of Malaria, the vectors are the mosquitoes, and they live in tropical and temperate countries, especially in African countries [57]. Researchers in [58] showed the top ten air travel risk routes for malaria-carrying mosquitoes by air traffic. By combining air traffic volumes and climatic similarity between origin-destination airports, they found that all routes fly from endemic African countries to European destinations in specific climate conditions in July, August or September. Schistosomiasis, on the other hand, is transmitted by direct contact with fresh water contaminated by specific snails, and it is most commonly found in Africa, Asia, and South America [59]. Cases of schistosomiasis introduced by air travel from Africa were discovered recently in Corsica as there were local snails to host agents of schistosomiasis [8].

Human-to-human transmission diseases. In recent years, air travel has contributed to several epidemics of global health significance including 2002–2004 Severe Acute Respiratory Syndrome (SARS) outbreak, the 2009 H1N1 Pandemic, the 2012 Middle East Respiratory Syndrome (MERS) outbreak, the 2014-2016 Ebola Outbreak in West Africa, and the 2019 coronavirus disease (COVID-19) pandemic [8].

The 2014 Ebola virus disease (EVD) epidemic began in Sierra Leone in 2014 crossing the borders also to Guinea and Liberia resulting in fatality rates ranging from 30% to 80% [8]. Since the active spread in West-Africa, cases began appearing in several countries via international air travel such as Nigeria, the USA, Italy, United Kingdom, Spain. Nigeria experienced the largest cluster due to air travellers with

19 confirmed cases in two cities of which seven died [60]. In spite of the fact that the travel ban reduced air traffic by 60% between the West African regions most affected by Ebola and the rest of the world, such a reduction was insufficient to prevent the exportation of Ebola cases. Travel restrictions indeed only led to delaying the risk of case importation per country from a few days to a few weeks [30]. Since the severe symptoms generated by the virus are easily recognizable, imported cases however were promptly detected and isolated, leading at most to small localized clusters.

The greatest threats for Global Heath are, however, diseases with an airborne transmission (i.e. transmitted through the air over time and distance) as they have the highest risk of generating global epidemics not requiring physical contacts to be transmitted. There are many examples of airborne viruses indeed that have generated simultaneous epidemics in non contiguous geographical areas across the world.

For instance, the SARS outbreak caused by a SARS-CoV in 2002/2003 spread from southern China to more than two dozen countries in North America, South America, Europe, and Asia. The global epidemic lasted about six months, with over 8,000 infected people. The World Health Organization (WHO) estimated a fatality rate of 14-15% [61]. The MERS outbreak caused by a MERS-CoV, which originated in Saudi Arabia in 2012 reached 27 countries in Europe, North America and Asia and in the Middle East. This generated more then 2,000 confirmed cases and around 800 deaths, leading to a fatality rate of around 30% [62]. Analyses of traffic data from/to Saudi Arabia show this latter covers a central role in connecting different regions of the world resulting in large traffic fluxes towards the continents of Asia, Europe and Africa [47]. South Korea experienced the largest epidemic outside of Saudi Arabia, which was introduced by an infected traveller returning from the affected country.

At the beginning of the 20th century, the 1918 influenza pandemic highlighted the potential of an airborne disease to spread globally thanks to the new interconnected world. In fact, the virus infected about 500 million individuals, then one-third of the population of the world. Later on, the 2009 Influenza H1N1 pandemic demonstrated the acceleration of global spread due to the increase in international mobility fluxes. In fact, it emerged in Mexico and it reached the United States in less then one month and it quickly spread across the world (See Figure 2.2), reaching the 21% of the global population [63]. Following the international alert, an air travel ban from/to Mexico was put in place, and air traffic was reduced by about 40%. However, this was not sufficient to contain the global emergence, producing only a delay of around 2 days in the arrival of infection out of Mexico. Bajardi et al. found that even with a 90% travel reduction, the resulting delay in the exportation worldwide would have been less than 2 weeks [10].

Lastly, the most recent emerging airborne virus was the severe acute respiratory syndrome coronavirus 2 (SARS CoV 2). This virus is the pathogen responsible for COVID-19. The ongoing COVID-19 pandemic has shown very clearly how interconnected the modern world has become. In fact, the rapid increase of COVID-19 infections all over the world generated a pandemic by reaching nearly every country in the World in around 70 days. The speed of spatial transmission exceeded all the other pandemics (See Figure 2.2) and resulted in more than 100 million cases and millions of deaths globally to date [64]. The global distribution of COVID-19 cases in the first wave occurred particularly in Asia, where there are most of the flight routes from China. Lau et al. [65] in found a positive correlation among passenger volume in any continent and international COVID-19 cases ($r^2 = 0.98$, $p < 0.1$). In the early stage of the COVID-19 pandemic, China implemented a travel ban from/to Wuhan (the epicentre of the epidemic), 59 airline companies suspended or limited flights to Mainland China and several countries including USA, Russia, Australia,

and Italy imposed travel restrictions. In this case also, however, air travel restrictions were not sufficient to contain the global emergency. Air travel restrictions may work, only when they are put in place in combination with other social distancing measures in the affected areas [38, 66]. Otherwise, new infections will quickly reach neighbouring locations, and these latter in turn, will spread the disease globally. In fact, after the travel ban in Wuhan, the other big cities in China including Shanghai, Beijing, and Shenzhen continued exporting infected COVID-19 cases all over the world as they were already previously infected [67].

The most alarming factor of the global spread is the speed of worldwide diffusion. The figure 2.2 shows the cumulative number of infected countries over time for a set of pandemics over the past century. The speed of global invasion has dramatically increased. Such an acceleration phenomenon is related to the increase in volume and connections resulting in an even higher heterogeneity in the distribution of the connections between locations given the increased urban concentralisation. In the next Section, I will explain the structure of air traffic network and how its affect global spread.

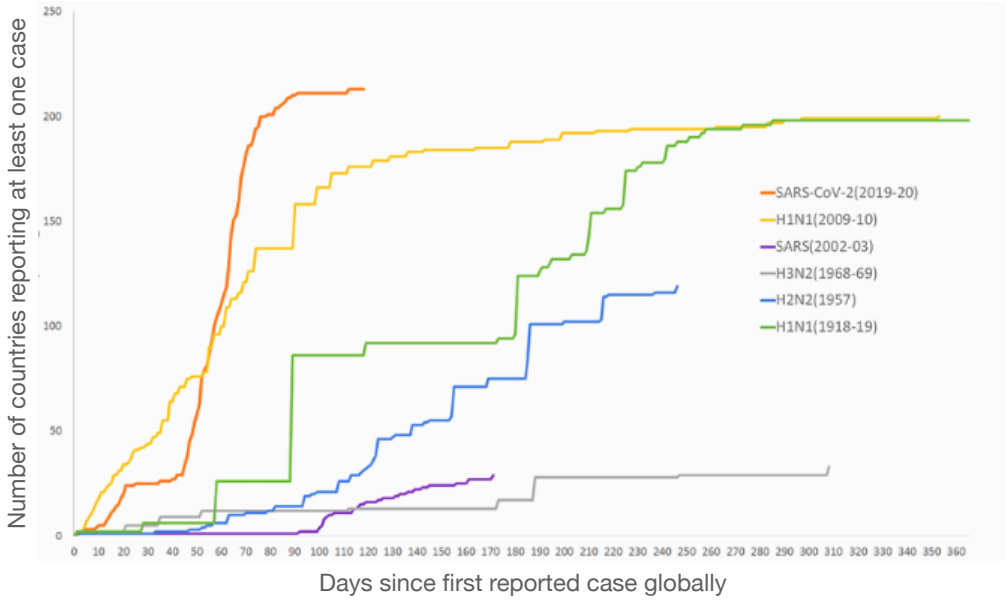


FIGURE 2.2: Plot of cumulative number of countries reporting first imported cases by day (day= 1 is when first case is reported in the epicentre). Coloured curves shows six disease pandemics over the past century. Figure from [68].

The architecture of air traffic flows

In the past decades, we have experienced the increase of global connectivity and mobility through air travel as Figure 2.4 shows. The number of connections of any airport and the number of passengers (air traffic) have become highly heterogeneous [4, 7, 69] and such intrinsic feature of air traffic flows play a crucial role in shaping epidemics [7, 69, 70].

Barrat et al. in [4] analysed the list of any pair of airports connected by direct flights in 2002 and the number of available seats on any given connection and showed that the traffic handled by each airport T with the corresponding number of connections k follows the non-linear form $T \approx k^{\beta}$ with $\beta \cong 1.5$ (See Figure 2.3). This

means that mobility connections between airports span several orders of magnitude: there are few hubs showing a huge number of connections, essentially big cities, while, the majority of the airports show few connections. Figure 2.4 also shows a very different behaviour for the real data network and its randomized version. The air traffic between connections belonging to highly connected airports tends to have a higher value than the one corresponding to the randomized weighted network. This means there is a strong correlation between the traffic among airports and the topology of the air traffic network, where the larger an airport is in terms of number of connections, the more traffic it can handle. Such airports with numerous connections and high volumes of passengers are called airport hubs. In the following section, I will explain how they drive the global epidemic diffusion.

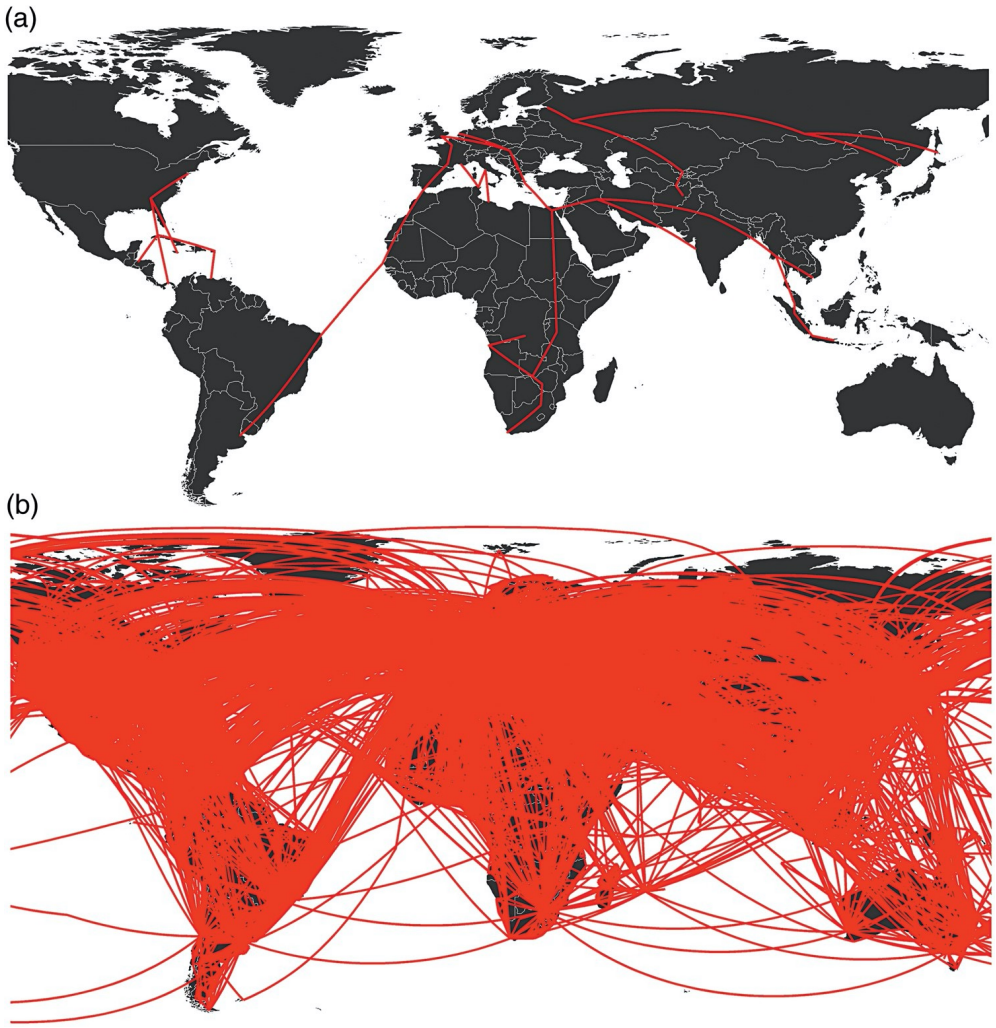


FIGURE 2.3: The increase of global connectivity and air travel. The international commercial air network in (a) 1933 and (b) 2010. Figure from [71].

From radial to heterogeneous diffusion

In the pre-industrial era, only a few means of transport were available and trips on the average timescale of one day were limited to short distances, resulting in contagions between individuals only in adjacent locations. A historical example is

given by the bubonic plague, called "the Black Death" which spread from China to Western Europe in the 14th century and took around 15 years with a velocity in the order of 300–700 km per year and whose spatio-temporal spread was dominated by radial diffusion [72] as shown in Figure 2.5 a).

At present given the high heterogeneity of air traffic fluxes previously discussed, global spread is not only dominated by radial diffusion as we have witnessed with the emergence and spread of novel pathogens to distant countries in a few days or weeks, such as the severe acute respiratory syndrome (SARS) in 2002, the 2009 A/H1N1 influenza pandemic, or the COVID-19 pandemic [68]. In Figure 2.5 b the 2009 pandemic that originated in Mexico and spread to more than 100 countries in a few months is reported. In this new perspective, the global spread presents different spatial transmission patterns, moving from a local diffusion (e.g. Black Death) to a diffusion that presents even long-range contagions. These "jumps" across countries are due to international travellers carrying pathogens across the planet and they are the reason why we are experiencing such epidemic acceleration.

Researchers indeed found that the increase of heterogeneity in the air traffic network reduces the the global invasion threshold i.e. the minimum number of travellers between locations in order to have a macroscopic number of infected subpopulations [70]. Therefore, the higher the heterogeneity is among airport connections, the lower the number of passengers that is needed to generate global epidemics. This means that the topology of the mobility network is a key driver to shaping infectious disease transmission.

Considering the same air traffic network presented in the previous Section, Colizza et al. in [5] found a direct relationship between the air traffic structure and the epidemic transmission.

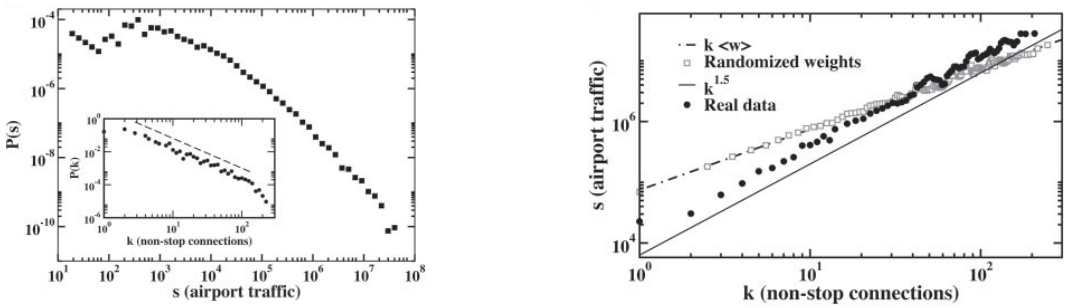


FIGURE 2.4: Properties of the worldwide air traffic. Left. The distribution $P(k)$ follows a power-law behavior with exponent 1.8 ± 0.2 where k is the number of connections of any airport. The total traffic handled by the airport as function of the number of connections of the airport. Figure from [4]

In fact, researchers compared the level of heterogeneity of the simulated global epidemics generated by real air traffic, with simulated epidemics generated in a synthetic scenario in which air traffic between airports had been homogeneously distributed. The latter case displays a homogeneous evolution of the simulated epidemics during a long time window. On the other hand, in the real scenario, cities are affected differently. In the real scenario, they then measured the similarity between two different outbreak stochastic realizations to understand if differences among cities' prevalence depended on the stochastic nature of the epidemic transmission

or to the nature of the air traffic network. They found a high level of predictability which is explained by the presence of preferential routes for spatial transmission that are weakly affected by the stochastic nature of the epidemic transmission.

In fact, the heterogeneity of traffic flows introduces dominant connections between airports with a higher exchange number of passengers (hub airports). These connections select preferential pathways, increasing the epidemic predictability [5].

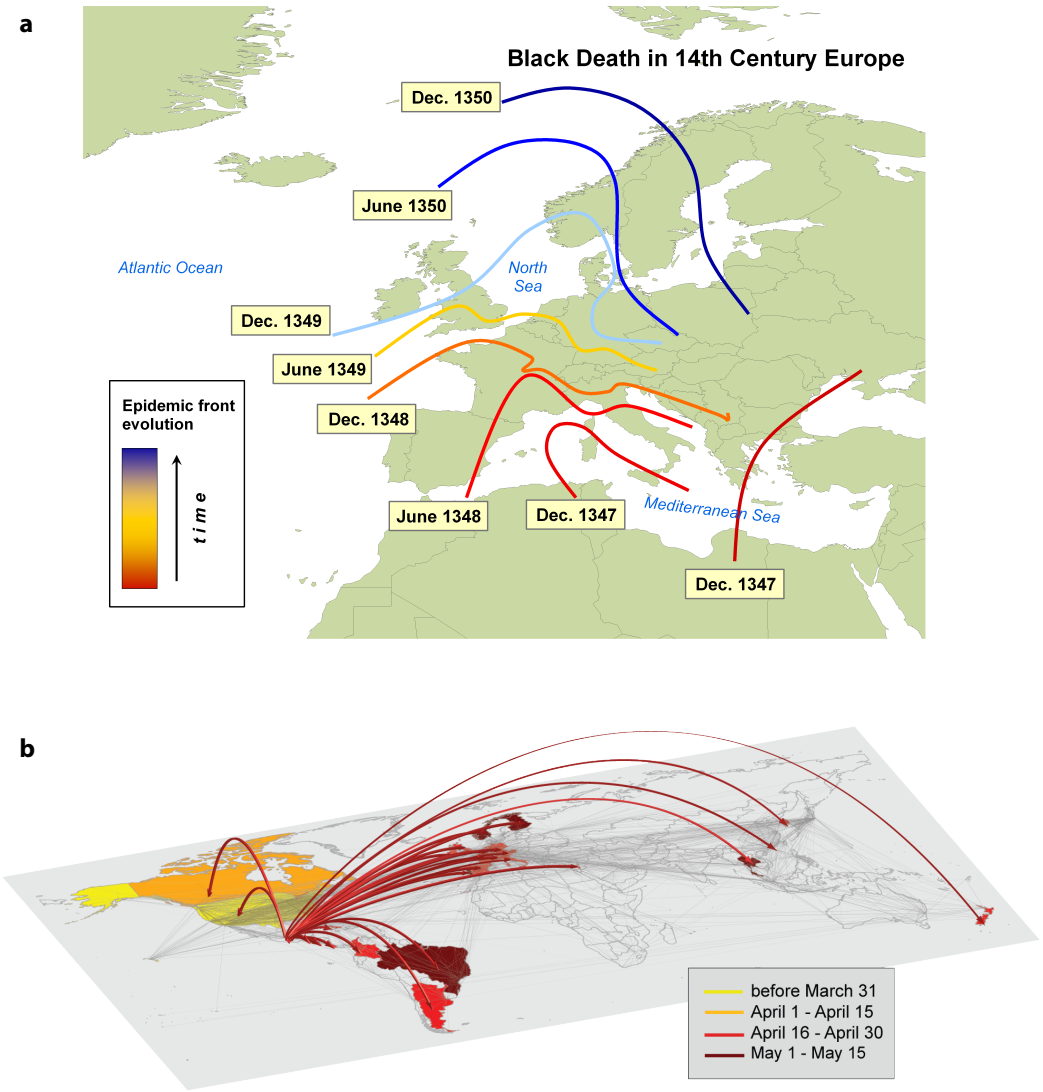


FIGURE 2.5: The spread of the Black Death and of the 2009 A/H1N1 pandemic. (a) The spread of the Black Death in the 14th Century in Europe had a radial diffusion, with an epidemic front wave crossing the continent from South to North. (b) The 2009 A/H1N1 pandemic originated in Mexico and reached all continents in a few months through air travel. Figure b) from [10].

Since the high predictability of spatial invasion related to mobility fluxes, it is possible to extract the global mobility patterns relevant to the spatial transmission and to design spatial epidemic models ad hoc for controlling and containing the emerging disease [38, 48, 73, 74, 75, 76, 77]. This especially allows to compute the

risk of introduction due to air travelers in the early phase of an epidemic which is the most relevant phase for epidemic surveillance [38, 47, 78]. In Section 5.2, I will present my research work on assessing the risk of importation of COVID-19 cases in Europe in the early stage of COVID-19 pandemic.

2.2 Recurrent mobility and national spread

Once local clusters of a disease emerge in a country, new infections may rapidly spread across the entire country due to internal mobility fluxes. In this Section, I will discuss the role of daily individual displacements in shaping national spread. First, I will discuss relevant mobility patterns of individuals, and then I will explain how such patterns affect epidemics.

Mobility recurrent patterns

Individuals have a recurrent travel distance and mainly return every day to the few locations that they frequent quite often [22, 55]. Researchers found that the distribution of displacements over all users and of their radius of gyrations (i.e. a metric defining the characteristic distance travelled by individuals) are well approximated by a truncated power-law (See Figure 2.6 c, d). To do that, they analyzed the individual trajectories of 6 million European mobile phone users (the use of mobile phone data to quantify human mobility is described in Chapter 3). Most individuals travel only over short-range distances, but a few regularly move over hundreds of kilometers. For example, in Figure 2.6 b, the user visits a total of 12 different locations, but he spends 88% of his time in two locations. Other studies proved the average number of frequently visited locations by individuals is only 2.14, and we may suppose that these locations are the home and the workplace [24]. Daily individual mobility is thus mainly dominated by commuting fluxes. Commuting flows appear to connect mainly neighboring subpopulations and few long-range high-populated cities, and they are on average an order of magnitude larger than the average airline traffic flow (See Figure 2.1). Moreover, commuting flows are recurrent as they refer to round trip processes with a characteristic time of the order of $1/3$ day (average duration of a workday). From a spatial diffusion perspective, this strong coupling involves a rapid diffusion among neighboring areas and long-range connected cities giving rise to the heterogeneous spatial spread commonly experienced in national spread.

The most recent example of a national rapid and heterogeneous diffusion due to individual movements was experienced during the early stage of the COVID-19 pandemic. The virus emerged in Wuhan, the ninth-most populous Chinese city. As the city was strongly connected by mobility fluxes (domestic flights, railways, highways) to the rest of the country, most of the early detected cases reported outside the epicentre came from Wuhan and were distributed across China (57%) [66]. Exported cases from Wuhan contributed to initiating local clusters, both in neighboring provinces (e.g., Henan) and in more distant provinces (e.g., Guangdong and Zhejiang). The earliest evidence of COVID-19 outside of Wuhan, however, was detected in big cities (e.g., Beijing, Shanghai, Guangzhou and Shenzhen) and these highly connected cities contributed to the further spread of COVID-19 to places less directly connected to Wuhan. A strict lockdown was put in place in Wuhan city on 23 January 2020 and subsequently in 14 other cities across Hubei province.

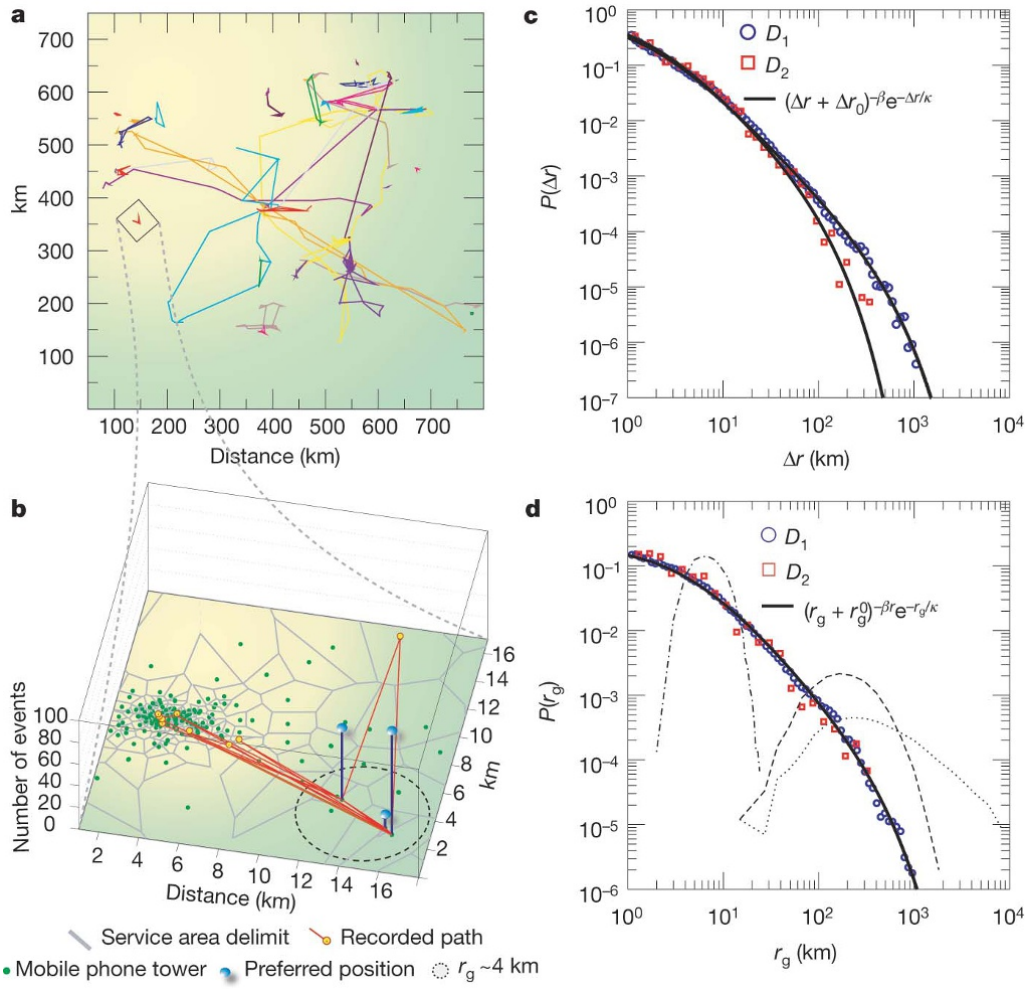


FIGURE 2.6: Individual mobility patterns. a) Week-long trajectory of 40 mobile phone users. Colors represent different users. b) The detailed trajectory of a single user. The different phone towers are shown as green dots, and the spatial partitioning (Voronoi cells) in gray marks the approximate reception area in which each tower handles the signal. The frequency of visits for each location is shown as a vertical bar and the circle represents the radius of gyration centered in the trajectory's mass center. c) Probability density function $P(\Delta r)$ of travel distances for two mobile phone data sets $D1$ and $D2$. The solid line is a fitted power low $((\Delta r + \Delta r_0)^{-}\beta exp(-\Delta r/k))$with exponent $\beta = 1.75$ and $\Delta r_0 = 1.5$ km and cutoff values k are $400km$ and $80km$ for $D1$ and $D2$ respectively. d) The distribution $P(r_g)$ of the radius of gyration measured for the users after 6 months of mobility patterns' observation. The solid lines are a fitted power law $((r_g + r_g^0)^{-}\beta_r exp(-r_g/k))$ with exponent $\beta = 1.65$, $r_g^0 = 5.8km$ and $k = 350$ km. The dotted, dashed and dot-dashed curves show $P(r_g)$ for a modelled scenario in which agents follow a random walk (RW), Lévy flight (LF) or truncated Lévy flight (TLF). Figure from [22].

By analysing daily individual trajectories, in [66] researchers found that the total number of cases (until 10 February 2020) outside of Wuhan was very well predicted by the mobility fluxes out of Wuhan alone ($R^2 = 0.89$ from a log-linear regression using cumulative cases) and in less than six days localized epidemics were established in most Chinese provinces (See Figure 2.7). After importations, similar patterns of

invasion at national level were observed in the other countries in the world including in US and Europe [39, 79, 80, 81, 82].

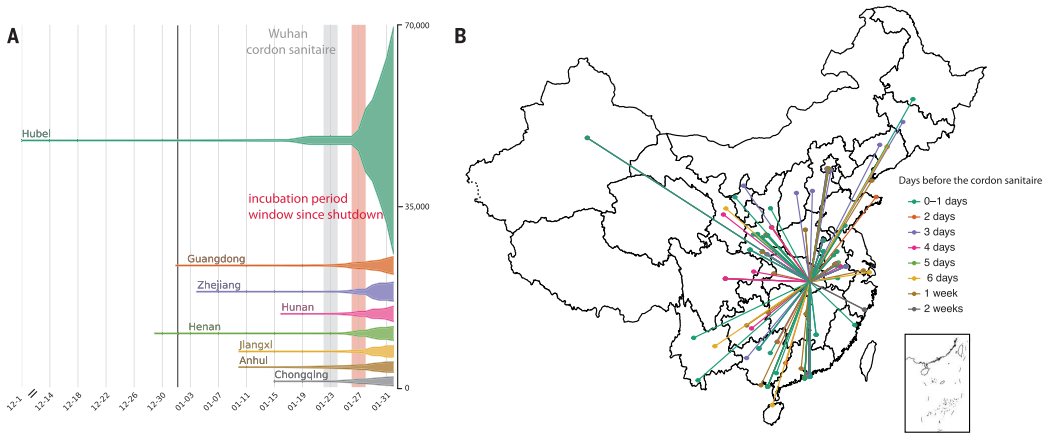


FIGURE 2.7: Early-stage COVID-19 transmission in China. A) Epidemic curve of the COVID-19 outbreak in the most affected provinces in China. The width of each horizontal tube represents the number of reported cases in each province. The gray shaded area represents the implementation of travel ban from/to Whuan and lockdown in Wuhan. The red shaded area show the end of the first incubation period after the implementation of restrictions. The black line represents the closure of the Wuhan seafood market where the virus was emerged. B) Map of COVID-19 confirmed cases with ravel history from Wuhan before travel restrictions. Colors indicate the travel date in terms of number of day after travel restrictions. Figure from [66].

The impact of recurrent patterns on synchronizing epidemics

Researchers have largely studied the impact of mobility at shaping epidemics within countries. For instance in [7], they simulated global epidemics through the GLEAM epidemic model [56] and analysed the epidemic outputs of different scenarios. They found that at a global level, the effect of commuting flows on prevalence profiles is relevant in the tail of the epidemic wave (Figure 2.8 A top): Commuting fluxes produce a faster decay of the new infections compared with a scenario in which locations are not connected by commuting fluxes. As shown in Figure 2.8 B), epidemic waves in many regions of the world last longer in the absence of commuting, showing that the commuting fluxes accelerate the time of arrival and may synchronize local epidemics.

They did not find significant changes in the epidemic profiles of locations with airport hubs as these are rapidly reached by infected individuals via air traffic, whereas the time in neighboring locations with limited airline connections crucially depends on commuting flows (See Figure 2.9). In fact, once hubs are infected, the epidemic may quickly spread out to the neighboring locations via a radial diffusion. This means that incidence profiles in the neighboring locations and in the hub overlap and fade out at the same time, generating the peak of importations occurring at the same time as the peak of the local epidemic wave [83]. On the other hand, when the volume of mobility between locations is not enough to generate synchronization effects, the epidemic emerges in these places with a certain delay. The non-synchronized epidemics (e.g. in one place infected cases decrease over time and in the other places new cases show an exponential growth) create rescue effects in the

place with lower epidemic activity. This mechanism is called a re-seeding event, and it may imply disease persistence, leading to a new re-growth of cases in the locations presenting few cases.

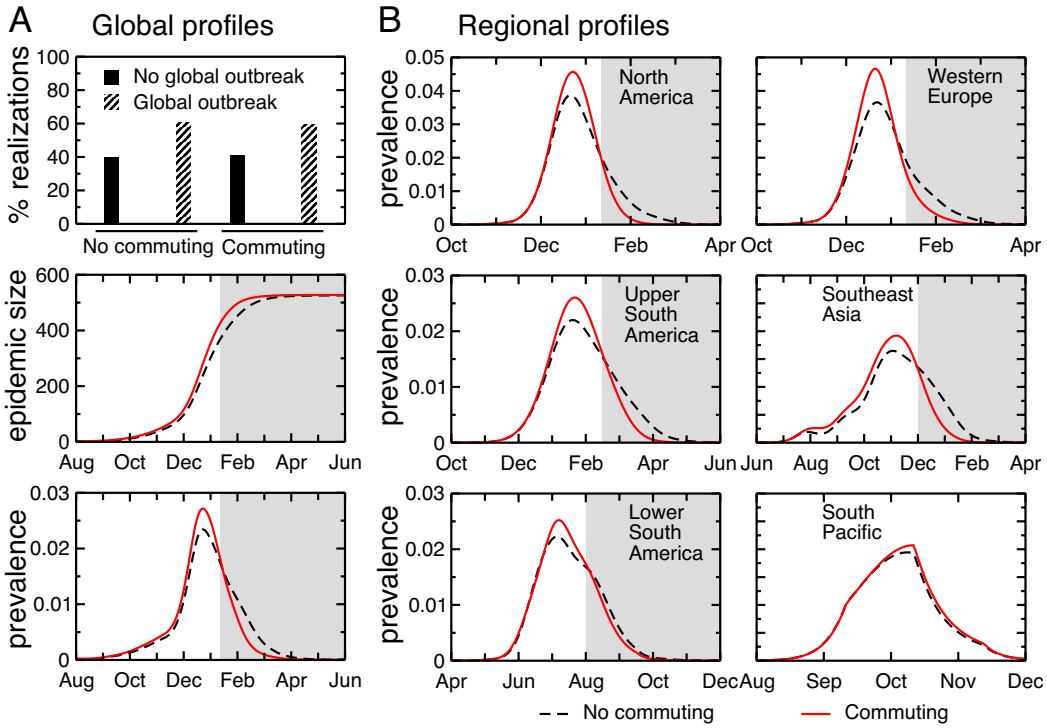


FIGURE 2.8: Comparison of GLEAM predictions [56] at the global and regional level obtained with and without commuting flows. Results refer to a pandemic influenza scenario starting in Hanoi. A) Top. Probability of outbreak. About 40% of the realizations lead to extinction at the source, whereas the remaining 60% causes a pandemic reaching more than 100 countries. A) Middle and Bottom. Global profiles for the epidemic size (number of cases per 1,000) and the prevalence. B) Prevalence at regional level. Figure from [7].

By studying influenza in the United States over the past 30 years, Viboud et al. in [44] found a synchrony in the time of arrival of epidemics across states which decays with geographical distance and is positive correlated with state population sizes. The most populous locations strongly connected by commuting fluxes indeed exhibit synchronized epidemics. To understand the relationship between disease spread and severity, distance, and population size, they compared the time of arrival and incidence to relevant measures of human movement, and they found that the commuting flow compared to long-distance trips and air travel is a higher predictor of influenza spread.

In order to better understand synchronized effects on seasonal influenza, Crepy et al. also studied seasonal influenza outbreaks and they analysed epidemiological data over 30 years in the US, and over 20 years in France [54]. For all pairs of locations (regions in France and states in US), they computed the Pearson correlation coefficient for incidence over each epidemic. Indeed, by carrying pathogens from an infected area to a non-infected one, they observed a difference in the times of arrival and peak times.

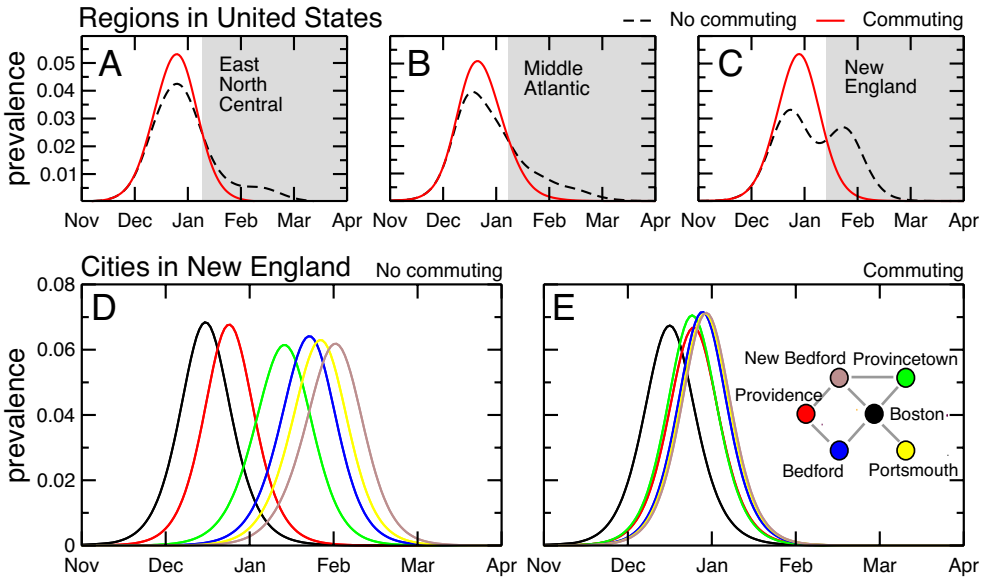


FIGURE 2.9: Comparison of GLEaM [56] predictions at the local level obtained with and without commuting. (A–C) Prevalence profiles in East North Central, Middle Atlantic and New England. The grey shaded area highlights the effect of commuting which it is in the faster decay and absence of multiple peaks. (D and E) Prevalence profiles for Boston area and the surrounding cities with no commuting (D) and with commuting (E). in (E) is also shown an illustrative representation of the short-range mobility connections. Commuting fluxes increase the synchronization effect, by reducing of around 1 month the delay between peaks among cities. Figure from [7].

In the case of US, for instance, considering the correlation between California and other states, some states present high fluctuations in the correlation e.g. for Wisconsin with a 95% CI ranging from 0.46 to 0.92, while others, like Arizona , display more robustness (0.84-0.94). This robustness means that recurrent pattern exists. Correlations between states are found for neighboring states (<600km), as well as distance ones (>3000km). Moreover, in US they found a positive correlation (linear fit with a coefficient of 0.74) among domestic air traffic between two states and their robustness in the correlations. Finally, they found that domestic air traffic makes a greater contribution to such robustness than distance or temperature (estimates were 0.407, -0.096, and 0.220, respectively). By considering inter-regional road and railways traffic volume, the same results have been found for France. This means that spatial spread has a preferential spatial path driven by individual mobility, leading to synchronization effects in locations strongly connected by both short and long-range mobility fluxes .

Therefore, high mobility inside countries tends to synchronize epidemics in neighbouring locations and between highly populated cities [67], while international travel as I described in the previous section tend to synchronize national epidemics, hastening global spread [84].

Changes in mobility in specific settings and for targeted age groups as in the case of holidays might affect spatial transmission. Holidays indeed alter mobility and social interactions as children and adults travel for the holidays and spend time with their families, rather than to commute to work and schools. In this context in [85], researchers found that the changes in seasonal influenza incidence patterns were

not synchronous across age groups during holidays. In fact, every year both children and adults experience a temporary decrease and rebound in incidence around Christmas, but adults only experience a reduction after the holiday while children start to decrease before then. In the next Section, I will discuss in detail the role of social mixing between age groups and its influence in shaping epidemics.

Source and sinks of imported cases

I previously discussed the invasion dynamics of an emerging disease and the resulting synchronized epidemics. More in general, in the context of ongoing epidemics or endemic diseases, people living in a given place are exposed to an infection which can be carried by (i) people living there and not moving around (ii) infected individuals living in other places and visiting by traveling (visitors), and (iii) residents who have been infected in other locations and have then returned back (returning residents) (See Section 4.3).

For example, the highest risk of acquiring HIV in Namibia depends on localized transmission from people not moving (60%), 25% of the risk depends on returning residents infected in other communities (returning residents), and 15% of the risk depends on infected visitors (visitors) [86].

Regarding mobility fluxes, therefore, some cities have a high percentage of visitors which results in a high risk of importing new infections (sink). While, other locations where residents move a lot may result in a high risk of exporting new cases (source). Clearly, big cities may play both roles. For instance, [52], researchers identified the sources and sinks of imported infections of malaria due to human travel in Kenya and the high-risk sites of parasite importation e.g. they found that the capital Nairobi imports the largest fraction of infections of malaria by residents infected during trips to the coast in the Lake Victoria region. Source and sink of malaria have been implemented also by Chang et al. [87] in Bangladesh. They found that the highest proportion of imported cases come from the south-western forested area of the CHT region (source) to more populated areas (sink). Similar results have also been found in [88] by Floyd et al. that proved that malaria parasites are more often carried from rural to urban districts (2.986% of movements) than vice versa (0.009% of movements).

Detecting sources and sinks of imported infections due to human mobility thus provide a starting point to improve targeted implementation of surveillance programs e.g. improving prevention strategies in sink areas and medical treatments in source ones. For instance, by quantifying source and sinks, Wesolowski et al. [26] estimated seasonal dengue virus importation and computed dynamic risk maps at high spatial resolution with direct application to dengue containment and epidemic preparedness.

Particular situations, however, (e.g. pilgrimage, traditional festivals, holidays) may generate new source and sink locations by leading many individuals to move to specific places at the same moment. For instance, it is well known how a mass gathering due to a pilgrimage was the factor responsible for the cholera outbreak in Senegal in 2005 [15]. The epidemic was originally localized in the region of Diourbel and its surroundings with hundreds of cases per week, but after the religious pilgrimage it suddenly spread to around a dozen of regions of the country, with over 27,000 reported cases. Researchers found a correlation between the mass gathering and the sudden peak of cases in Diourbel and the national spread by pilgrims returning to their homes (coefficient of determination between modelled and reported weekly cases $R^2 = 0.78$). In fact, during the pilgrimage, the number of travellers

away from home almost doubles with respect to an average day. People travelled to the holy city of Touba from all over the country, resulting in an estimated number nearly 6 times the usual population of people present in the city. This underlines the importance of monitoring real-time human behaviour to assess dynamically the source and sink of potential emerging epidemics.

Mobility fluxes, therefore, affect epidemics by mixing individuals and thus altering the probability and frequency of encounters among them. In the next Section, I will explain in detail what occurs at local level within a mixed population and how social mixing favours new infections.

2.3 Social mixing and local spread

While local transportation, commuting, or global travels allow the disease to reach different places by mixing individuals, at local level such mixing patterns i.e. who meets whom (where, when, how long, and how often) may actually lead to new infections [76, 84, 89, 90, 91].

Co-location for social mixing

One of the largest and earliest efforts to understand social mixing driven by co-location and its role in the epidemic transmission was carried out in 2008 by the POLYMOD project. They conducted a large scale population-based survey in eight European countries asking participants to provide their daily contacts. They translated survey data into contact matrices whose elements report the rate of face-to-face contacts among individuals of different age groups in different settings: home, work, transports, doing leisure activities, and doing other not specified activities [89]. Contacts are also broken down into physical contacts and non-physical ones. Such matrices cannot capture all the complex details of any face-to-face interactions, however, they provide a useful combination of the social interactions of a population level which reveals key factors relevant for epidemic transmission [89, 91].

The analysis of this survey shows that the reported contacts present a regular pattern of contact frequency by age: the number of contacts rises with age in children; contacts then fall to a lower plateau in adults until the age of 50 and decrease after that age. Contacts also change also over time. On weekdays individuals make 30%–40% more contacts than on Sundays. Moreover, individuals interact differently in different locations. Most of the contacts at home (75%) and at school (50%) and during leisure activities (50%) were physical, whereas in other settings most of the contacts are non-physical. The same results were found in all countries (Figure 2.10). The dominant characteristic is that individuals in all age groups tend to mix in an assortative manner (i.e., preferentially with others of similar age). In the Figure 2.10, the assortative pattern is shown by the yellow diagonal. This pattern is most evident in young people aged 5–24, and least pronounced in older people aged 55–69. Another characteristic is the two parallel secondary diagonals representing children mixing with adults (mainly at home). (Figure 2.10).

Similar studies have been successively conducted in various other countries [92, 93] including France in 2015 [94] and they found similar contact patterns. This is a crucial result as it underlines how there are common patterns at population level in the way individuals mix among age groups. Considering the difficulty of designing a survey and the similar contact patterns found in most countries, researchers developed estimation methods to compute contact matrices. Researchers indeed mixed

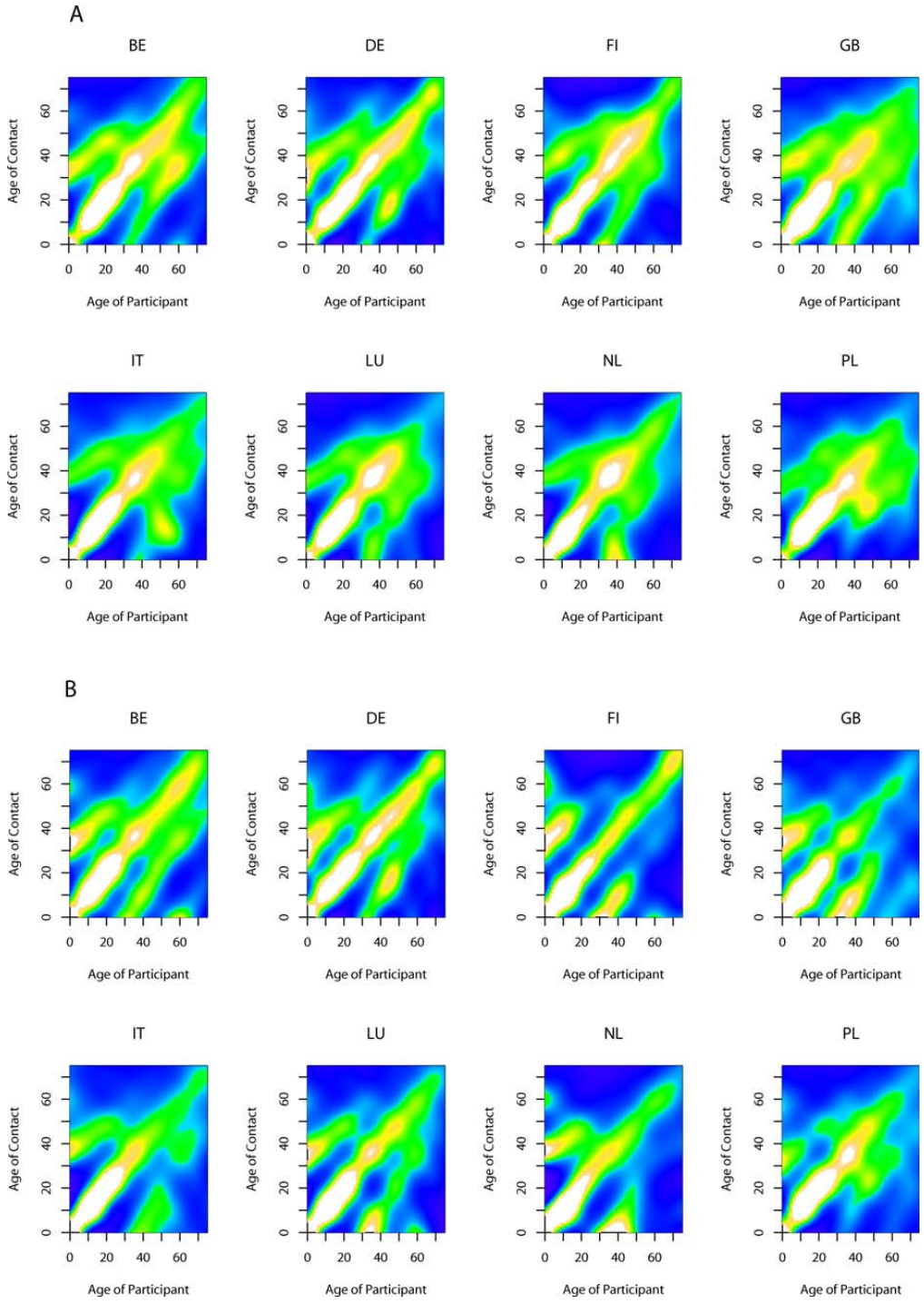


FIGURE 2.10: Smoothed Contact Matrices for many European countries on A) All Reported Contacts and B) Physical Contacts. White indicates high contact rates, green intermediate contact rates, and blue low contact rates. Rates are relative to the country-specific contact intensity. Figure from [89].

the data from POLYMOD project and local census to compute contact matrices in 144 other countries [95].

To summarize, we discussed how contact patterns are different according to location, with most contacts made at school and at home and fewer contacts during transport and in public places. Contact patterns were also found to be assortative with age in all locations, and mixing among different ages groups occurred mainly at home. At the same time, social mixing has a seasonality effect as daily interactions of individuals may vary over time e.g. during the workday, people tend to spend most of the day at work or school interacting in pairs, while, at the weekend or during school holidays people are more likely to interact with their family [93, 96].

Contact matrices, however, are static as they are collected in a given time slot, therefore do not capture any potential change in social mixing e.g. during an outbreak. In this respect, online questionnaires and behavioral data collections app have been developed to monitor dynamical social mixing [93].

Social mixing affects local epidemic profiles

Static and dynamical contact matrices or explicit contact networks among individuals have been largely embedded into epidemic models to study the role of social mixing and its changes in shaping epidemics [76, 81, 84, 89, 90, 91, 92, 97, 98, 99, 100, 101].

Since interactions between individuals vary according to age and the place where they occur, indicators as household size, the fraction of workers and students in the population crucially affect epidemic profiles [91]. Indeed, results show how countries characterized by large household groups and by a large fraction of students and workers in the population like Ireland face more severe epidemics of influenza than countries like Germany and Bulgaria [84].

Moreover, the reduction of social mixing at weekends and holidays strongly affects pathogen circulation as individuals do not go to school or work and spend time mainly with friends and family. As children make more contacts overall and have several close contacts at school, this increases their risk of being infected, making schools an important source of transmission to households [97, 99, 102, 103].

In order to get a better understanding on this, several research works have been done to understand the role of school closure in shaping epidemics. During the 2009 H1N1 pandemic, a correlation was observed between the changes in contact patterns due to the opening of schools and the onset of widespread transmission of H1N1 in the US [101].

Moreover, in the context of seasonal influenza in the US, Ewing et al. found in [103] that the reduction of contacts at school due to Christmas holidays decreases the risk of epidemic activity among children, increasing it instead in adults. They also found that winter holidays reduce the viral circulation of influenza overall across all metropolitan areas in US, delay the peak timing, and thus increase the synchrony of local epidemics. Since during school holidays, social contacts are reduced but travelling increases, they also analysed the role of holiday-related trips in shaping epidemics, finding that travel has minimal effects compered to social mix changes.

Corroborating, De Luca et al. found that changes in social mixing due to weekends and holidays result in a delay of the peak time incidence of seasonal influenza in Belgium [97]. By simulating seasonal influenza waves, they analyzed both a scenario with only workdays and, a more realistic one, considering changes in social patterns due to weekends and holidays. Comparing the simulated outcomes, they

found that the realistic model delays the peak time (median among Belgian municipalities has 3.7 weeks of delay). They also found that the realistic model reduces peak time incidence (33% median relative change) and the total epidemic size (11% median relative change).

Since results suggest that most infections occur in specific places such as workplaces and schools, monitoring social mixing is thus crucial to detecting locations favouring the spreading and applying targeted interventions of social distancing. Research studies on this topic largely emerged during the COVID-19 pandemic with the ongoing emergency and the availability of empirical data capturing behavioral change induced by the diseases and strict interventions.

Indeed during the COVID-19 pandemic governments had to put in place strict interventions in place as the health-care systems were overwhelmed in several countries due to the high viral circulation. Since early 2020, most countries in the world experimented national lockdowns (i.e. a restriction policy where people do not move except for basic needs) to promptly bring down transmission below the level needed to sustain the epidemic [38, 39, 66, 81, 92, 98, 104, 105]. In this context, several works have been done to assess the impact of the interventions put in place and the expected impact of potential new interventions both to mitigate new infections during the ascending phase of the epidemic, and to avoid the new increase of cases during the descending phase.

In fact, as I discussed in the previous Sections, travel restrictions only delay importations in non-affected locations, while social distancing interventions are the only means to reduce the basic reproductive number R_0 when viral circulation is high and testing strategies are not efficient enough to detect and isolate cases (See Section 4.1 to the definition of R_0) [66]. For instance in Italy results suggest that the sequence of restrictions posed to mobility and human-to-human interactions succeeded in reducing transmission by 45% [39]. In France, we estimated the basic reproductive number R_0 before the implementation of interventions to be at 3.18 [3.09, 3.24] (95% confidence interval) and thanks to a reduction of the average number of contacts of 81%, R_0 this decreased to 0.68 [0.66, 0.69] during lockdown [81]. Researchers also estimated that in a synthetic scenario of an uncontrolled epidemic in Wuhan and Shanghai (without intervention measures, travel restrictions, and spontaneous behavioural responses of the population), the mean infection attack rate had to be ranging from 64% to 92% after a year of disease circulation with R_0 in the range of 2 to 3 [92]. On the other hand, they estimated that in a scenario where social distancing measures were implemented as the new virus emerged, $R0$ would remain under the threshold, and thus the epidemic would not take off in other locations.

2.4 Conclusions

In this chapter, I presented the role of human mobility on epidemic transmission. In Section 2.1, I explained the role of international fluxes mainly composed of air travellers at rapidly spread disease globally.

In Section 2.2, I focused on epidemic spread at national level. I focused principally on how individual mobility and commuting fluxes affect epidemics resulting in contagion among neighboring locations and highly populated long-range locations. In Section 2.3, I explained how social mixing caused by individual mobility affects local epidemics. In the next Chapter 3, I will present how to quantify mobility patterns from mobile phone data and how to translate individual behaviours

to population patterns relevant to infection transmission, more specifically I will describe the most common aggregation method used in literature. Then, In Chapter 4, I will describe the integration of mobility data in a comprehensive theoretical framework, in order to model epidemic spread. Finally, in Chapter 5 and 6, I will present my research work where I will quantify mobility fluxes at different scales and from a different source and I will show how to integrate mobility into theoretical frameworks introduced in Chapter 4 to help authorities to control spatial transmission and to better implement surveillance strategies.

Chapter 3

Mobile phone data and epidemic spread

The availability of human mobility data on a large scale and high resolution has impacted several research fields such as urban planning, social sciences and in particular the spatial epidemiology of infectious diseases [18]. As human movements are a key driver of spatial transmission of infectious diseases, tracking and analyzing individual daily trajectories help understand geographical conditions that lead to epidemic diffusion. In the previous Chapter, I showed how human mobility is responsible for the spatial spread of infectious diseases. In this chapter, I will describe how to quantify human mobility from mobile phone data for epidemiological purposes.

3.1 Mobile phone data for epidemiological purposes

Mobile phones are one of the most pervasive technologies. In 2019, there were already around 5 billion unique mobile phone subscribers globally, which means a penetration rate of 67% of the global population, as shown in Fig 3.1. Even in the less-developed countries, such as Sub-Saharan Africa, the penetration rate of mobile phone subscriptions has reached 45% of the population. Projections for 2025 show an increasing trend in every country reaching an expected penetration rate of 70.5% globally.

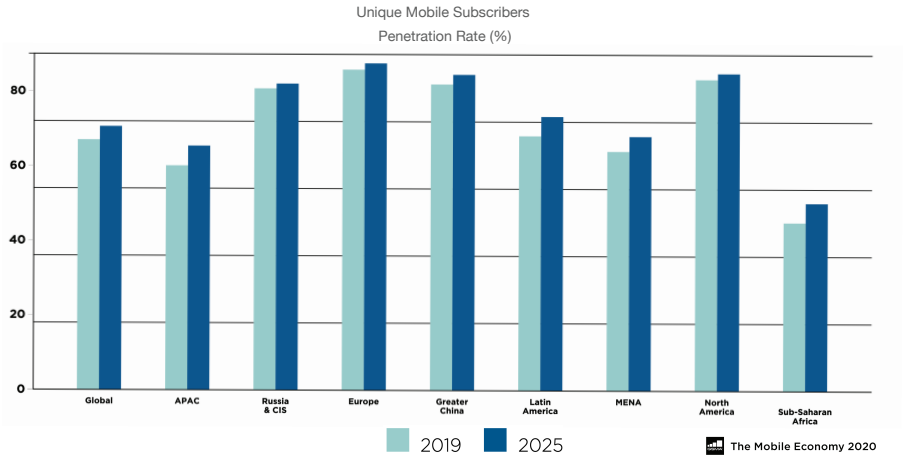


FIGURE 3.1: The figure shows the penetration rate (%) of unique mobile phone subscriptions in 2019 and the projections for 2025.

As mobile phones are now an integral part of modern life, mobile phone data has emerged as a deeply valuable source to understand human mobility and pathogen spread across spatial and temporal scales [11, 14, 15, 25, 26, 27, 28, 29]. Mobile-derived human mobility has been used for epidemiological studies in the context of COVID-19 [31, 34, 39, 66, 81, 104, 106, 107, 108, 109, 110, 111, 112], malaria [52], dengue [26], cholera [113], Ebola [17], and human immunodeficiency virus (HIV) infection [114]. Traditionally, human mobility has been measured by data from censuses, travel history surveys or traffic surveys. Although survey and census data have the advantage of having a very refined spatial resolution which is often the census block, in data-poor countries this type of data is only updated every 5/10 years. In the last decades, other sources have also been largely used – e.g. global air traffic and infrastructure data. Global air traffic data has been analyzed to measure national and international mobility and its impact on the global spread of pathogens and vectors at city or airport level [5, 10, 30, 38]. While, infrastructure data - e.g. ground transportation data - have been used to define the connectivity within countries, for example to assess the health accessibility [115] or to detect patterns in the national spread of epidemics [54]. The crucial power of mobile phone data over other data sources is that it allows for a better and more specific understanding of individual mobility behavior. Furthermore, thanks to its high penetration in the entire population, the information is on a large scale and long term. Although mobile phone data have been largely used to study national mobility [14], the analysis of international mobility fluxes however, is recent and very limited [116]. At the moment, the most largely used source of data to study international mobility is air traffic data.

Since early 2008, the access to this data allowed researchers to understand human mobility dynamics in terms of quantity [22, 23]. Researchers in [22] have found that individual trajectories have a certain regularity in time and space and are highly influenced by their historical behavior. Individuals have a recurrent travel routine and return to the few locations that they frequent a lot almost every day. For example, analyzing mobile phone data from Portugal, researchers discovered that the average number of frequently
is only 2.14 and that 95% of the users frequently visit less than 4 locations [24].

Since this data enables routine monitoring of population activities, it also has the potential to detect atypical behaviors. Candia et al. [23], first, introduced a method to study the statistical fluctuations in space and time of individual user behaviors, with respect to their average behavior. Given this background,in 2010 Lu et al. had the intuition that mobile phone data might be used to monitor population's behaviors in response to emergencies. As a result, they asked the biggest network operator in Haiti, to share mobile phone data from millions of users from before and after the earthquake. Analyzing this data, they found that the population of the capital of Haiti, Port-au-Prince, decreased by almost one-quarter soon after the quake, and slowly increased in the following 11 months [13]. These results have been validated by data coming from intensive surveys conducted by the United Nations.

This work was the first clear demonstration of the reliability of this data. Scientists started collaborating with network operators or companies to analyse mobile phone data from tens of millions of phone owners in several countries all over the world. Combining mobile phone data with epidemiological information, researchers had thus the possibility to better study the impact of human mobility patterns on the spread of infectious disease. One of the earliest studies was [52]. Researchers used 1 year of mobile phone data from 15 million individuals in Kenya. They quantified connectivity in terms of mobility network among 692 sites that have cell towers within 10 km of their boundaries, and they detected the high-risk sites of

parasite importation of malaria [16]. With the same dataset, researchers quantified seasonal population fluxes in Kenya between districts, and they embedded them into an epidemic model to assess how these drive rubella transmission dynamics.

In 2012 Orange.S.A., the French multinational telecommunications corporation, organized two open innovation challenges within a project called Data for Development (D4D) [117] in which Orange Telecom offered researchers its aggregated mobile phone data of 5 million Orange customers in Ivory Coast and of 9 million customers in Senegal at department level. The challenge marked the first time a large database of mobile network data was opened to the international scientific community for use in research. Thanks to D4D datasets, researchers have highlighted the role of mass gatherings on the spreading of waterborne diseases like cholera [15] in Senegal. They found that a mass gathering, which took place during the initial phase of the Cholera outbreak, influenced the course of the epidemic. This study underlined the potential of such a source of data to perform analysis that would not be feasible by static sources of human mobility like census and survey data.

The D4D challenge also helped to better characterize the interplay between mobility patterns and epidemic activity. In fact, by using the dataset from Ivory Cost, researchers extracted 224 different features to describe the number and the average duration of users' calls and the locations visited by users [118]. They also extracted migration features by tracking the changes in the home locations of the users. They defined home locations as the places where individuals make the largest number of calls during nighttime and the migration process as the moment when people do not sleep at home for more than 3 nights. Thanks to the high resolution of the data in time and in space, they were able to map each indicator at different temporal and spatial scales at the same time. Given all these features, they then built regression models and evaluated their performance when predicting prevalence rate, finding that mobility fluxes during nighttime, the spatial areas covered by users, and the frequent migrations are strongly linked to HIV transmission. Most recently, by using mobile phone data from Namibia, researchers have proven that the majority of the risk of acquiring HIV was due to localized transmission from people not moving (%60), 25% of the risk was due to returning infected residents from other communities (returning residents), and 15% of the risk was due to infected visitors (visitors).

3.2 Mobile phone data to fight emergencies in real-time

Mobile phone data have a crucial role in the assessment of the emerging and ongoing epidemics as elements for strategic planning, outbreak preparedness, and response include knowing the patterns of movements, the interplay of movements at different time and spatial scales, the presence of high-risk locations in highly dense areas. Furthermore, most important is how these aspects evolve in time as individuals change their behavior in response to the epidemic.

By aiming to implement control strategies in order to contain the rapid spread of the virus within Guinea, Sierra Leone, and Liberia, epidemiologists called for its use during the 2014 West Africa Ebola virus epidemic [46]. In the absence of mobile phone data from the currently affected countries, they revolved then around the use of mobility models fitted on data provided by D4D challenge in Ivory Cost [119]. Not only were the data not specific to the affected countries, but it was also collected in peacetime - i.e. in the absence of an ongoing outbreak, inevitably leading to disruption of regular patterns. More recently, researchers had access to mobile phone data from Sierra Leone and they had the possibility to do a retrospective study

to quantify the population mobility reductions associated with travel restrictions [17].

Researchers therefore started to underline the need to develop protocols for the sharing of operator data to have rapid access to it during public health emergencies [46]. In May 2019, at least 20 mobile-phone companies donated their proprietary information to such efforts [14], including operators who back an initiative called Big Data for Social Good, sponsored by the GSMA. The urgent need for real-time mobility data has then been recently experienced during the COVID-19 pandemic [21]. The European Commission called European mobile network operators to share their data. A call for action was also developed by the World Bank, the International Telecommunication Union (ITU), GSMA (Global System for Mobile Communications) and the World Economic Forum with the aim of using the power of mobile big data to fight COVID-19 [120]. As a consequence, operators and digital companies from all around the world started sharing aggregated mobile phone information on human travels and positioning. For example, Google and Apple started publishing periodically community mobility reports by providing changes in human activity behaviors compared to pre-pandemic scenarios [121]. Facebook also made online maps of population movement for 169 countries available [122].

The rapid and pervasive sharing of this data allowed a better estimate of human behavioral adaptations to COVID-19 pandemic and the effectiveness of ongoing control measures in many countries – e.g. Austria, Belgium, Chile, China, Germany, France, Italy, Spain, United Kingdom, and the United States [34, 66, 104, 107, 108, 109, 110, 111, 112]. Real-time data matched with epidemiological information were also used to integrate epidemic models to quantify the impact of the interventions put in place [39, 81, 106]. In addition, to find the optimal future restrictions needed to avoid new waves - e.g. closure of non-essential activities, partial attendance at school, curfew [81, 123, 124, 125]. For example in [106], in order to inform reopening strategies in 10 metropolitan areas in the US after a lockdown, they mapped the hourly movements of 98 million people from census blocks to points of interest such as restaurants and religious establishments, connecting 56,945 census block groups to 552,758 points of interest with 5.4 billion hourly connections. Integrating a transmission epidemic model with this information, they found that few points of interest account for a large majority of the infections. They suggested that the optimal strategy in this context would be restricting the maximum occupancy in these targeted points of interest, instead of implementing homogeneous travel restrictions.

COVID-19 pandemic showed how mobile phone data can be an essential source of information for decision-makers and authorities during the ongoing epidemics. By collaborating with researchers and health authorities, operators may share their data as quickly as possible in an ethical and privacy-preserving manner to help fight ongoing epidemics and for studying invasion dynamics in both emerging and endemic diseases.

3.3 Type and sources of mobile phone data

Mobile phone data includes mobile network data and location history data from mobile apps. Mobile network data is collected by network operators and include call detail records (CDRs), extended data records (XDRs) and signaling data. A mobile phone, switched on, regularly notifies its position in terms of the closest tower cell where it is currently located. The information on the mobile phone positioning either by active communication events - calls, text messages, or Internet usage - or by

signaling between the network and the device. A CDRs and XDRs are generated for billing purposes every time people use their phones to call or message. Each record contains several attributes: the caller and callee IDs, the time-stamp, the duration of the activity, the type of communication - national, international, call outgoing, call incoming -, and the identifier of the antenna that handled the activity. This data allows the extraction of the individual trajectories by interpolating every user's displacements based on two consecutive calls/SMS. It means that CDRs can measure individual displacements only if individuals are using the mobile phone (See Figure 3.2). Individual trajectories from CDRs, however, are interpolated on any two consecutive calls and this is strongly dependent on users' calling behavior [126, 127].

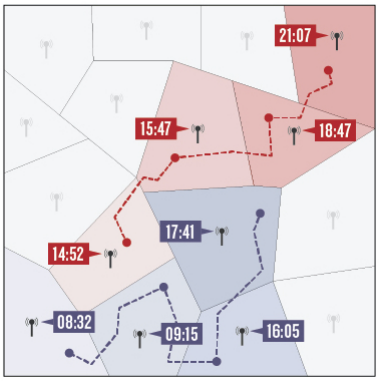


FIGURE 3.2: Individual trajectories of two mobile phone users inferred by CDRs. The color code represents the different mobile phone users. The time label is the time when users do an activity - e.g. calling or sending a message. For any activity, the signal is handled by the closest cell tower and the call detail record stored. Figure from [14]

.

The second type of mobile phone data collected by operators is signaling data. Signaling data overcomes the bias of calling frequency as they are driven by updates of the network. In fact, signaling data provides the geolocation of all the events that are generated by the network itself as Location Area updates. A Location Area is a set of tower cell stations that are grouped together to optimize signaling. Location updates can occur for i) periodic update, which occurs to provide information on which cell towers the phone is connected to; ii) handover, which occurs when a user involved in a call moves between two cell towers; iii) mobility location update, which occurs when a user moves between two Location Areas even if he is not using the mobile. Despite the fact that signaling data is network-driven data, it has some other biases. In fact, generally speaking, network data can not measure spatial movements better than cell tower-level spatial resolution.

More recently, given the extensive use of smartphones, another source of mobile phone data has emerged. Various mobile phone applications record the position of users with high spatial precision. This data is called location history data, and it can be extracted from the usage of mobile-based social media - e.g. Tweets, Facebook, or other applications. In this case, the location is identified by the phone's internal GPS [21], the individual trajectories of the users are thus spatially refined as GPS tracker data. This is collected by location intelligence companies [128], specific apps or companies that develop mobile phone operating systems (e.g. Google, Apple) [129]. The high resolution of mobile-based location history data means they are

one of the most powerful sources of information for better understanding human activities and travel routes across long periods and countries. However, smartphone penetration is still very low in low-income countries. In fact, in 2019, the penetration rate was globally 49% ranging from 76% in US and Europe to 25% in Sub-Saharan Africa.

However, mobile phone data may have ownership biases as they do not capture the entire population. Since individuals may own multiple SIM cards and multiple individuals may use the same SIM cards [130]. However, despite the bias, several studies have been done to validate their reliability showing that aggregated network data can accurately replicate population counts and migration patterns from censuses [131, 132, 133]. They have been also validated as a good proxy of commuting fluxes [11]. In [11], Tizzoni et al. proved that commuting fluxes extracted from mobile phone data present same patterns as census commuting data. They also introduced a transmission model with both mobility fluxes and they found similar epidemic outcomes.

The state of the art presented in the previous section is mainly based on CDRs datasets as operators in the past preferentially shared this kind of data. While location history data is an emerging source that companies have recently started sharing.

3.4 Data access and privacy issues

From the discussion in previous chapters emerges the important role of mobile phone data to help fight epidemics. However, there are also several risks associated with sharing and dealing with such sensitive data. In fact, the high level of detail of individuals' behaviours extracted from mobile phone data might threaten the security of citizens, more especially of minorities e.g. racial, religious.

In most democratic countries, such as in Europe, both national and regional legal regulations limit the use of personal data. In this regard, the main pillar of data protection legislation in the European countries is the General Data Protection Regulation (GDPR) [134] and the 'Privacy and electronic communications' directive (e-privacy) with its national transposition [135].

However, most companies, including mobile network operators, still tend to be reluctant to share their data, even when aggregated and anonymized. The main concerns are due to i) the limits of data anonymization, as personal information could be still retrieved after the process [136], and sometimes to ii) the lack of clear and harmonized legal frameworks for sharing the data.

To date, external access to private data has proceeded under specific bilateral agreements, mostly between research institutions and data providers, or in the context of data challenges like D4D (Data 4 Development). During the COVID-19 emergency, companies, operators, and researchers underlined the need for standardized sharing models, legally compliant and ethically acceptable so to reduce concerns about sharing data [21, 46]. In this regard, researchers also proposed creating an ethics and privacy advisory committee to supervise research projects dealing with data [21].

Awarding contracts between companies and researchers/health authorities and computational efforts to aggregate and anonymize data requirement time, which is not always affordable with the speed of the spread of an emerging disease. In fact, the emergency of the COVID-19 pandemic showed the importance of implementing a health emergency sharing protocol early. Both for the use of digital data and for algorithms to pre-process such datasets.

3.5 Extraction of mobility indicators from mobile phone data

Mobile phone individual trajectories give us mobility indicators about different spatial and temporal resolutions, which may be used to study mobility fluxes and to provide information for epidemic models. Here I present the most common mobility indicators used to quantify coupling between locations due to mobility fluxes. Given the high-resolution of such data, we may consider different levels of details of the individual trajectories in the aggregation process to compute coupling forces.

Translating individual trajectories into coupling forces between locations means defining: (i) the mobility process that is relevant, and (ii) the resolution in space and time that is needed. Namely, to define when two locations i and j need to be coupled ($p_{ij} \neq 0$) and to quantify the probability of coupling p_{ij} between the connected locations.

By selecting the level of details, aggregated individual trajectories have been largely used to study many diseases through different methodological approaches. The aggregation method at the highest resolution is when the entire resolution of the individual trajectories is preserved. This approach was used by researchers in [25, 27] to assess the impact of non-pharmaceutical interventions in simulated synthetic epidemics. They used coupling matrices to inform a mechanistic transmission model where human mobility among locations is explicitly modelled through the movements of the individuals (See Section 4.3). In this work, they constructed the coupling matrix from CDRs accounting for the number of displacements between each two consecutive calls made by mobile phone users. I refer to this matrix as displacement-based coupling matrix D and the mathematical formulation is the following:

Displacements based coupling matrix D

$$p_{ij}^{D} = \frac{\sum_u D_{i,j}^u}{\sum_k \sum_u D_{i,k}^u} \tag{3.1}$$

where $D_{i,j}^u$ is the number of times user u moves from location i to j. This method is at high resolution as it keeps the temporal sequence of any displacement of the users into account. The coupling probability between two locations is defined as the probability of moving from one place to another. In this approach, the time spent by users in any place is not considered. In fact, each displacement has the same weight in the total sum regardless of the time elapsed between two consecutive displacements. In a further work [137], Lima et al. introduced another formulation where $D_{i,j}^{'u}$ is re-defined as the sum of the time elapsed between any displacement between i and j done by the user u.

Though the former method covers each individual's movement, it is highly dependent on the place, time and frequency of mobile phone activities made by them. By using signalling data, in [138] Kang et al. suggested an improvement of the definition of displacements. They divided the space in grids - 500 m by 500 m - and defined a trip when users make two consecutive network connections in two different grids. This way, they reduced short-distance trips within an identical grid and the bias due to the frequency of mobile phone activities made by users. Comparing these aggregated mobility fluxes with taxicab trips, they found that taxicab trips within 1 km are much less than trips derived by mobile phone data. While taxicab

trips longer than 3 km are much more than the others. They also found that the distance decay effect in trips extracted from mobile phone data is stronger than that of taxicab trips as shown in Figure 3.3.

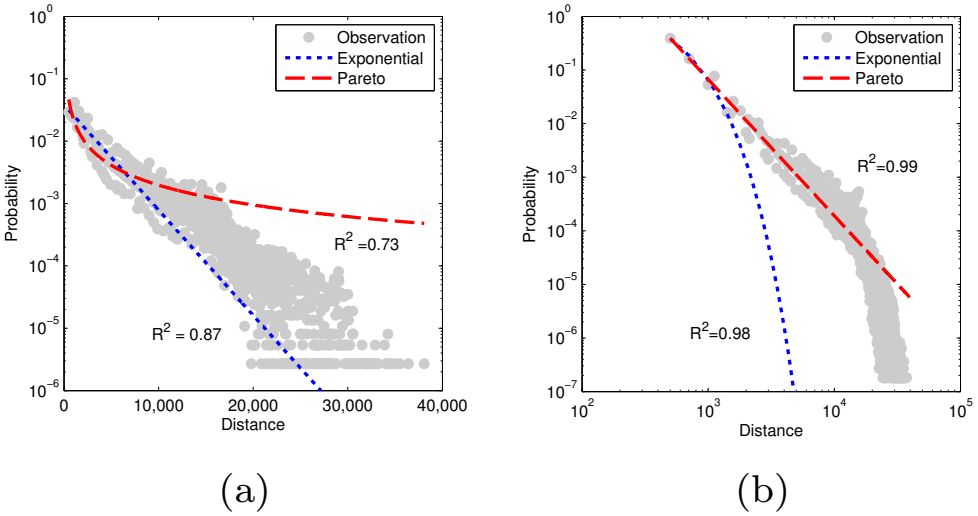


FIGURE 3.3: The distance decays of taxicab trips (a) and mobile phone displacements (b). in (a) it is shown as taxicab trips follow the exponential distribution with an exponent about 2.9, while in (b) it is shown as mobile phone movements follow the power-law distribution with an exponent about 2.5. Figure from [138]

.

These differences are partly due to the fact that taxicabs are mainly used by people in specific circumstances and they better indicate medium and long-distance travels. On the other hand, the high resolution of individual displacements extracted by mobile phones makes deriving a proper definition of an individual trip hard. This work underlined how in the Displacements based aggregation, individuals' trip may be fragmented into several displacements leading to an underestimation of the geographical distance of the real journey.

By reducing the level of details of the trajectories even more, Wesolowski et al. in [16, 26] aggregated displacements at time slots of one day. They evaluated the users' most visited location every day and computed the daily displacements between locations obtaining a time series of daily trips. The most visited locations were computed as the place where the users made the most number of calls each day. With this approach, they quantified i) the impact of human mobility in the emergence of dengue epidemics in Pakistan in non endemic areas [26] and ii) the impact of seasonal travel patterns on driving rubella transmission dynamics in Kenya [16]. In the first work, they estimated the number of introduced cases in j from i by sampling from a binomial distribution $B(n_j, p_{ij})$ with p_{ij} the coupling probability and n_i the number of infected cases in a given day. While in the second one, they analyzed the time series of daily trips, climatic changes and school holidays as observable in a linear regression with seasonal transmission as the response variable.

Location based coupling matrix L

A totally different approach to aggregate CDRs trajectories at medium resolution was suggested by researchers in [15]. They connected the home location of each

user to all the other visited locations, with a coupling force that was proportional to the number of calls made in each location. They used the resulting coupling matrix to inform a metapopulation model with an effective spatial transmission where individuals in each location do not move but feel a force of infection depending on coupling matrices (See Section 4.3). They used this approach to study the role of mass gatherings in the spreading of cholera outbreaks in Senegal. The same model and aggregation procedure was also used to study the spatial spread of schistosomiasis within a network of connected villages in the endemic region of Saint-Louise, in Senegal [29]. The aggregation process they used is radically different from the displacement-based one. In this case the trajectory is lost in favour of a coupling home - visited location. I refer to this method as location based coupling matrix L and the mathematical formulation is the following:

$$p_{ij}^{L} = \frac{\sum_{u_i} L_{i,j}^{u_i}}{\sum_k \sum_{u_i} L_{i,k}^{u}} \tag{3.2}$$

where $L_{i,j}^{u_i}$ is the number of calls made in j by user u_i living in i. Users living in i are detected by extracting individuals that do the most number of calls in i during nighttime - from 7 pm to 7 am. This is a well-established method used in literature to define the home location of mobile phone users [18]. In this method, the information of the temporal sequence of any individual displacement is lost, and the coupling is proportional to the number of calls made by users in each location. By assuming that the number of calls made by users in a place is proportional to the time that they spend in it, this method measures the coupling between any two locations i-j as the probability of being in j living in i. In this approach, all the locations users visits are divided into places where they spend most of their time and places where they just spend few minutes. However, by accounting for the number of calls, in this case as in the displacement-based one, there is a bias due to the heterogeneity of the activity of mobile phone users.

A similar aggregation method was used in [106] to inform the reopening after the first lockdown due to COVID-19 epidemic in ten of the largest US metropolitan areas. They used already aggregated location history data from SafeGraph, a company which aggregates anonymized location data from mobile applications. Researchers quantified coupling matrices in terms of the number of movements of mobile phone users from their home location to specific points of interest, which are non-residential locations that people visit such as work places, restaurants, grocery stores and religious establishments. Given the GPS positioning of the location history data, it was possible to extract these types of places. Thus, selecting the points of interest of the users by excluding all the ones where people were only passing through.

Commuting based coupling matrix C

An even lower resolution approach than the latter is to only select the most visited locations by excluding all the places where people only pass through, considering them secondary points of interest. This approach was proposed by Wesolowski et al. in [52] in order to identify the main importation routes that contribute to the spreading of malaria in Kenya through a spatially-explicit probabilistic model. First, they assigned the home location for each user. Then, they computed the destination of each daily trip made by users as the location where they spent the majority of their

time during the day. With this approach, the number of connections between locations are a subset of the ones detected with the location based method. Therefore, I refer to it as commuting based coupling matrix C and the mathematical formulation is the following:

$$p_{ij}^{C} = \frac{\sum_{u_i} C_{i,j}^{u_i}}{\sum_k \sum_{u_i} C_{i,k}^{u_i}} \tag{3.3}$$

$C_{i,j}^{u_i}$ is the amount of time user u_i spends at his most visited location daily in j. Therefore, this method takes only locations where users spend most of their time into account. Any other less visited location is neglected. Since the most visited location often coincides with the work location of an individual, this aggregation process may be considered an extraction of the commuting fluxes. In [11], Tizzoni et al. found similar patterns by using a similar definition computed in commuting fluxes from mobile phone data in three European countries. Followed by comparing them with mobility fluxes from census. The results showed that commuting fluxes from mobile phone data capture more than 87% of the total census fluxes.

However, commuting fluxes aside, each presented coupling matrix has been integrated into disparate epidemic models without paying attention to the accuracy of the process of aggregation. Aggregating individual trajectories properly means assessing how many details of the trajectory are needed, which details are negligible, and which ones produce a bias when embedded into models. Considering the privacy issues (See Section 3.4), the biggest operators and companies must aggregate individual trajectories before sharing data with researchers. In this context, knowing the proper aggregation methods may be very important in order for operators and companies to develop standard algorithms to process their data for epidemiological purposes.

3.6 Conclusions

To conclude, in this Section I presented how mobile phone data has been largely used to quantify mobility. Then I integrated this into epidemic models to study its role in the epidemic diffusion in many countries around the world. More specifically, in Section 3.2 I explained how such data has been used to fight epidemics in real-time. Then, in Section 3.3, I described how mobile phone data is collected, its different types and the information we can extract from it. Finally, in Section 3.5 I presented the most common methods to aggregate individual trajectories extracted from mobile phone data in order to define mobility indicators for epidemiological means. In the next Section, I will present how to integrate mobility into epidemic models. Particular emphasis will be dedicated to my research work aimed at assessing the impact on informing transmission models with different aggregation approaches.

Chapter 4

Integrating mobility into epidemic models

Epidemic models are used to assess the mechanisms behind the transmission of infectious disease in order to quantitatively understand its prevalence and distribution in the general population, as well as the key factors which determine its incidence, spread in time and space, and persistence. The first epidemic model with a mathematical approach was developed by Daniel Bernoulli in 1766 to study the expected impact of vaccinations to help eradicate smallpox [139]. Almost two centuries later, in 1906 Hamer theorized that the spread of an epidemic depends on social contacts between infectious and susceptible individuals [140] and in 1927, Kermack and McKendrick introduced a theoretical formulation of the compartmental models to describe the dynamics of infectious diseases [141]. Nowadays, mathematical and computational epidemiology are based on this powerful framework. Moreover, the improvement of the computational capability in recent decades has allowed the development of computational models not analytically solvable and to integrate large sources of data [2]. The effort of researchers has created disparate models, ranging on very different levels of complexity and targeted for every specific epidemiological settings. Epidemic models indeed span different levels of realism ranging from the simplest models in which individuals have all the same probability to interact among each other to more realistic models in which spatial, social and mobility structures are integrated. The higher the resolution of the model, the more data is needed to inform it. In the next Chapter, I will introduce a few of these models and I will pay particular attention to how design models account for the role of human mobility in shaping new infections. More specifically, in Section 4.1 I will explain how to describe the evolution of the disease by compartmental models and how to embed social structures into models. Next, in Section 4.2, I will explain in detail the age-structured model specific for the COVID-19 disease I developed during my PhD thesis. In Section 4.3, I will present spatially explicit epidemic models which account explicitly for mobility fluxes between locations using a metapopulation approach. Particular attention will be devoted to the model I designed and developed to asses the use of mobile phone data to inform transmission models. In Section 4.4, I will describe the method to model invasion dynamic and the risk of importation into any country. I will present in detail, methods I used to compute the risk of COVID-19 imported cases into Europe from China in the early-stage of the COVID-19 pandemic. Finally, in Section 4.5, I will present the theoretical work I performed during my PhD on assessing the mobility definition and resolution needed to inform metapopulation models using mobile phone data.

4.1 Basic and age-structured compartmental models

Compartmental models describe the interaction between host and pathogen by classifying any host with respect to its health statuses. The transitions between one compartment (or status) and another are formalized as a set of coupled differential equations. Such transitions can be event-driven like the infection that depends on the event to make contact with infected individuals, or spontaneous like the time needed to recover if you are ill. The number and type of compartments depend on the dynamic of the disease. However, all compartmental models are a variation of two basic archetypes: SIR or SIS [142].

S stands for susceptible, and it is when an individual is in good health and he is susceptible to an infection. After coming in contact with an infected individual, the susceptible person could be infected and become infectious (I). Finally, in SIR hosts recover (R) having an immunity that allows them to avoid further infection, while in SIS they do not have immunity and become susceptible again.

In both models this transition of the reaction process is $SI \xrightarrow{\beta} 2I$ in which β is the transmission rate. Instead, the transitions from I to R in SIR and from I to S in SIS are spontaneous and can be written as $I \xrightarrow{\mu} R$ and $I \xrightarrow{\mu} S$, respectively, where μ is the recovery rate.

I have reported here the equations in the case of SIR model:

$$\frac{ds(t)}{dt} = -\beta\langle k\rangle i(t)s(t) = \beta i(t)s(t) = -\lambda s(t) \tag{4.1}$$

$$\frac{di(t)}{dt} = \lambda s(t) - \mu i(t) \tag{4.2}$$

$$\frac{dr(t)}{dt} = \mu i(t) \tag{4.3}$$

whit N the population size, $N = S + I + R$ and $s = S/N$, $i = I/N$ and $r = R/N$ respectively. The factor $\langle k\rangle$ represents the average number of contacts of a single individual. Here, we make a homogeneous mixing assumption, assuming that $\langle k\rangle$ is a constant value for every person in the population. This means that all individuals can interact randomly with each other.

Social contact structures may be then added to the model to describe key epidemiological behaviours in the case of the disease affecting sub-populations heterogeneously (e.g. per age groups) or to answer specific questions such assessing the implementation of targeted intervention strategies in particular social settings. Social interactions commonly appear with various structural heterogeneity among individuals [90], by changing patterns depending on locations where they occur and age of the individuals. Even if several works have been done to quantify the precise social network, collecting data about them on a large scale for entire populations is extremely difficult. Contact matrices per age group are thus a middle way approach to modelling social mixing between homogeneously mixed populations and effective social contacts [81, 92, 94, 95, 97, 143, 144]. Contact matrices are often embedded into age-structured models to capture the impact of social mixing patterns and its changes on the epidemic transmission [97, 144, 145, 146]. In age-structured epidemic models, the host population is divided into n age groups. For any age class i, the force of infection $\lambda(i,t)$ at the time t is driven by social interactions among age groups as follows [97]:

$$\lambda(i,t) = \beta \sum_j C_{ij} \frac{I^j(t)}{N} \tag{4.4}$$

with $j \in \{1, 2, .., n\}$.

C is the contact matrix. The C_{ij} element of the matrix represents the per capita number of contacts that an individual of age i has with an individual of age j. The force of infection is proportional to the sum of the probabilities of encountering an infectious individual of any age class j. Considering only one age bracket, we find the equation 4.1 that is the force of infection in the case where age classes are not considered.

Contact matrices estimate the number and intensity of face-to-face contacts people have in a day per age groups. Matrices are computed in several locations and for physical and non-physical contacts and successively aggregated into one matrix to generate the effective descriptions of human mixing patterns relevant to the spreading of a specific disease (See Section 2.3).

In my research work, using real-time individual behaviours (e.g. the estimate of the percentage of people going to school, at work, or the adaptation of social distancing measures), I parametrized changes in social mixing compared to the static contact matrices estimated in a regular period in France. In the following section 4.2, I will describe the age-structured transmission model specific per COVID-19 we designed, and I will explain how we parametrized contact matrices during and in exit of the first national lockdown.

4.2 Compartmental models specific for COVID-19

SIR and SEIR-like archetypes may be used to design specific disease models. Ad hoc models have been designed to study several diseases including Ebola [30], COVID-19 [142], MERS [47], SARS [147]. In this section, I will present the age-structured SEIR-like model specific for COVID-19 epidemic in France that I used in my research work in Section 6.2. The model was adapted from the model already designed in [81].

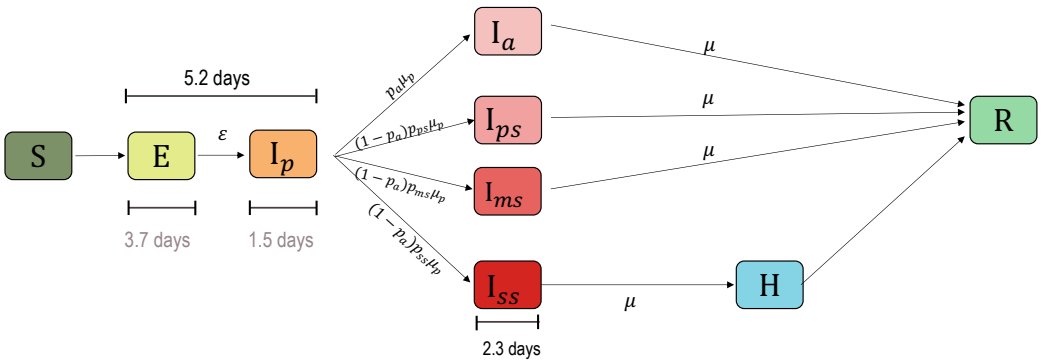


FIGURE 4.1: The figure shows the compartmental model specific for COVID-19 epidemic. S=Susceptible, E=Exposed, I_p= Infectious in the prodromic phase, I_a=Asymptomatic Infectious, I_{ps}=Paucisymptomatic Infectious, I_{ms}=Symptomatic Infectious with mild symptoms, I_{ss}=Symptomatic Infectious with severe symptoms, H=severe case admitted to the hospital, R=Recovered.

In order to model the COVID-19 epidemic, we considered a compartmental model where transmission dynamics evolve according to the scheme reported in Figure 4.1. Specifically, we consider four age classes. Younger children (yc): [0-10] years old; teenagers (t): [11,18] years old; adults (a): [19-64] years old and seniors (s): 65+ years old. The first two age groups refer to students, [0-10] for pre-school and primary school, [11-18] for middle school and high school. In the context of COVID-19, social structures are deeply important because the disease does not affect age groups homogeneously e.g. there is a very small percentage (<≤5 %) of children in COVID-19 confirmed cases worldwide [148].

The dynamic of the disease is a generalization of the classic $SEIR$ model as it is defined as follows. Once susceptible individuals get the infections, they will first enter in E. Later on, individuals will move to the prodromic status I_p and become able to transmit the disease. At the end of the prodromic phase, with a spontaneous transition individuals may either not show symptoms and thus to be asymptomatic (I_a) or develop symptoms. The proportion of asymptomatic individuals in our model is 40%, as estimated by [149]. By doing a sensitivity analysis, we also examined a probability of being asymptomatic equal to $p_a = 50\%$ and $p_a = 30\%$. Concerning individuals developing symptoms, we considered different degrees of severity: i) paucisymptomatic individuals (I_{ps}), ii) individual with mild symptoms (I_{ms}) or iii) individual with severe symptoms (I_{ss}). Probability rates to move from I_p status to one of the four infectious ones are all informed by previous studies [149, 150, 151]. Current evidence from household studies, contact tracing investigations, and modeling works suggested that young people (younger children and adolescents) have a reduced susceptibility and are more likely to become either asymptomatic or paucisymptomatic [92, 152, 153, 154, 155, 156]. Further evidence from serological investigations in high schools and primary schools suggested a different role of children compared to adolescents [152], with the latter more likely to act as asymptomatic adults, whereas children are likely less infectious. Based on this knowledge, we considered that: young people have a 50% relative susceptibility compared to adults and become either asymptomatic or paucisymptomatic when infected; adolescents have the same reduction in transmissibility as adults in absence of symptoms ($r_\beta = 0.55$) [157], whereas younger children have a smaller transmissibility, $r_\beta[0,11) = 0.25$. Individuals showing severe symptoms may move into the hospitalized compartment (H). Finally, all compartments converge in status R in which there are people who have recovered or died. Values used to parameterize the model are reported in 4.1. The model was developed for every region in France by using demographics and age profile data provide by INSEE (National Institute of Statistics and Economic Studies) [158].

The theoretical formulation of the compartmental model for a given age class i is based on the following set of differential equations:

$$\frac{dS^i}{dt} = -\lambda_i S^i \tag{4.5}$$

$$\frac{dE^i}{dt} = \lambda_i S^i - \epsilon E^i \tag{4.6}$$

$$\frac{dI_p^i}{dt} = \epsilon E^i - \mu_p I_p^i p_a^i - \mu_p I_p^i (1 - p_a^i)[p_{ps}^i + p_{ms}^i + p_{ss}^i] \tag{4.7}$$

$$\frac{dI_a^i}{dt} = p_a^i \mu_p I_p^i - \mu I_a^i \tag{4.8}$$

Variable	Description	Value
θ^{-1}	Incubation period	5.2d
μ_p^{-1}	Duration of prodromal phase	1.5d, computed as the fraction of pre-symptomatic transmission events out of pre-symptomatic plus symptomatic transmission events.
ϵ^{-1}	Latency period	$\theta^{-1} - \mu_p^{-1}$
p_a	Probability of being asymptomatic	0.4
p_{ps}	If symptomatic, probability of being paucisymptomatic	1 for children 0.2 for adults, seniors
p_{ms}	If symptomatic, probability of developing mild symptoms	0 for children 0.7 for adults 0.6 for seniors
p_{ss}	If symptomatic, probability of developing severe symptoms	0 for children 0.1 for adults 0.2 for seniors
g	Generation time	6.6d
μ^{-1}	Infectious period	2.3d, chosen accordingly to generation time distribution
r_β	Relative infectiousness of I_p, I_a, I_{ps}	0.25 for younger children 0.55 for adolescents, adults, seniors
s	Relative susceptibility	0.5 for younger children, adolescents 1 for adults, seniors

TABLE 4.1: Parameters and values to define the COVID-19 age-structured transmission model.

$$\frac{dI_{ps}^i}{dt} = (1 - p_a^i) p_{ps}^i \mu_p I_p^i - \mu I_{ps}^i \tag{4.9}$$

$$\frac{dI_{ms}^i}{dt} = (1 - p_a^i) p_{ms}^i \mu_p I_p^i - \mu I_{ms}^i \tag{4.10}$$

$$\frac{dI_{ss}^i}{dt} = (1 - p_a^i) p_{ss}^i \mu_p I_p^i - \mu I_{ss}^i \tag{4.11}$$

$$\frac{dH^i}{dt} = \mu I_{ss}^i - \lambda_{H,R} H^i \tag{4.12}$$

$$\frac{dR^i}{dt} = \lambda_{H,R} H^i + \mu I_{ps}^i + \mu I_{ms}^i + \mu I_{ss}^i \tag{4.13}$$

with

$$\lambda_i = \frac{\beta}{N} \left[\sum_j C_{ij}(t) s r_\beta I_p^j + \sum_j C_{ij}^{I_a}(t) s r_\beta I_a^j + \sum_j C_{ij}^{I_{ps,ms}}(t) s r_\beta I_{ps}^j + \sum_j C_{ij}^{I_{ps,ms}}(t) s I_{ms}^j + \sum_j C_{ij}^{I_{ss}}(t) s I_{ss}^j \right] \tag{4.14}$$

and $i, j \in \{yc, t, a, s\}$ and N is the population size.

The contact matrix $C(t)$ provides the average contact rates made by healthy individuals between different age classes over time t. We also considered the individual

behavioural changes due to the onset of symptoms and self-isolation of individuals testing positive for COVID-19.

From $C(t)$, we thus defined $C^{I_a}(t)$, $C^{I_{ps,ms}}(t)$ and $C^{I_{ss}}(t)$ as follows:

- $C^{I_a}(t) = x_a \cdot C(t) \cdot 0.1 + (1 - x_a) \cdot C(t)$
- $C^{I_{ps,ms}}(t) = x_s \cdot C(t) \cdot 0.1 + (1 - x_s) \cdot C(t)$
- $C^{I_{ss}}(t) = x_s \cdot C(t) \cdot 0.1 + 0.25 \cdot (1 - x_s) \cdot C(t)$

where x_a and x_s are fitted parameters representing the percentage of individuals testing positive for COVID-19 without and with symptoms, respectively. We assumed a reduction of 90% of contacts for individuals with a positive test. An additional reduction of 75% of contacts was considered for individuals not tested and showing severe symptoms, because we supposed that they self-isolate in any case as they are sick.

Social mixing on a regular weekday between healthy individuals in France was estimated by a national survey in 2015 [94]. The survey provides contact matrices in different settings: home, school, work, transport, leisure, and other locations not specified with the possibility to break down contact matrices for physical and non-physical contacts. Therefore, the contact matrix in a regular condition for a a pre-pandemic scenario $C(t)$ is the sum over all the empirical location-based matrices. To take into account the changes of social mixing due to interventions put in place by the French government and adaptive behaviors of individuals to the epidemic (e.g.risk aversion), we parameterized contact matrices over time. The parametrization is explained in the following Section.

4.2.1 Parametrization of contact matrices

Anomalous changes in patterns are not detected from contact matrices provided by the 2015 Survey in France [94]. They only provide seasonal behavior changes during weekends or holidays. Here, I present how we parameterized contact matrices from behavioral data to better describe real-time evolving social mixing during first wave of COVID-19 epidemic in France. We thus had to infer changes in human behavior depending on the implementation put in place by French government.

Contact matrices are thus mechanistically parameterized by region to reduce contacts in specific social settings for targeted age classes. We obtained synthetic contact matrices reproducing social distancing interventions during the first national lockdown and exit phase interventions in the 7 weeks after lockdown. Details on the timeline of the first wave of COVID-19 in France have been discussed in Section 6.1.

To parametrize the matrices, we reduced social mixing by considering multi source data on the percentage of students going to school, the percentage of workers going to the workplace and on the percentage of people compliant to preventive measures of social distancing.

School attendance

Schools during first lockdown were closed, therefore we considered contacts at school equal to zero. While, contacts at school after lockdown was parameterized by accounting for the percentage of reported attendance at pre-school, primary, middle and high school provided by the Ministry of Education [159] per region (See Figure 4.2). For instance, the attendance of 14.5% in Île-de-France region in pre-schools and

primary schools, corresponds to a reduction of 85.5% in the number of contacts made at school by individuals in the same age class. As students did not go to school, we accordingly reduced contacts made in the means of transport.

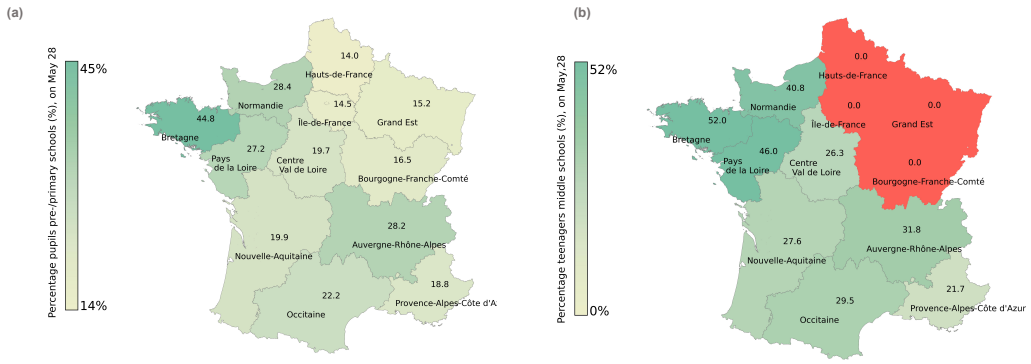


FIGURE 4.2: Percentage of students going to school, as reported by the Ministry of Education on May 28, 2020 [159]. (a) Students of pre- and primary schools (b) Students of middle schools. Red areas refer to regions where schools remaining closed at that time as they presented an high viral circulation.

Presence at work

Given the closure of non-essential activities, factories and companies during the lockdown and the recommendations on remote working and activities that were not yet reopened in the exit of lockdown, we also reduced contacts at work. To do that, we used the estimated variation of presence at workplaces per week based on mobile phone location data provided by Google Mobility Trends [121] (See Figure 4.3). The Figure shows the 7 weeks after the first national lockdown, but we also analyzed variation of presence at workplaces during lockdown. Contacts at work and for different means of transport (as we did in the case of the school attendance) were therefore modified according to this percentage. As individuals (students and workers) did not commute, we also increased contacts in households by considering the changes in contact patterns during weekend versus weekday and the proportion of adults working during the weekend.

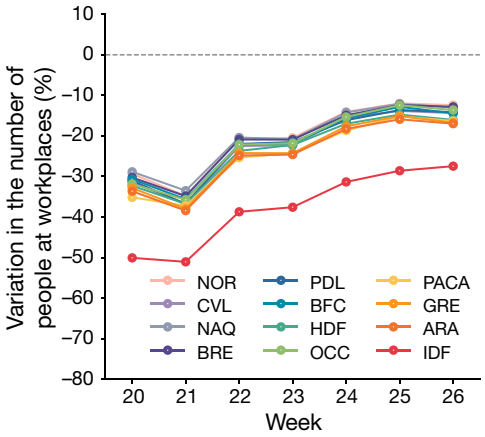


FIGURE 4.3: Percentage of individuals not going to workplace locations. The plot shows the estimated change in presence at workplace locations in the 7 weeks after the first lockdown and by region based on Google location history data [121]. Region acronyms are listed in the paper reported in Section 6.2.

Adoption of physical distancing

Since during lockdown, it was compulsory to stay at home, going out only for basic needs, we considered compliance of 100% of physical distancing by accounting for physical contacts only in households according to the results we found in [81] and in agreement with data collected afterwards [160]. Then, to account for individual adoption of preventive behaviour after lockdown, we used the percentage of population avoiding physical contacts estimated from a large-scale survey conducted by Santé publique France [160]. We fit the data of the survey with a linear regression to infer the weekly percentage of individuals avoiding physical contacts (Figure 4.4 (a)). We therefore reduced contact matrices week by week, according to the fitted percentage of individuals avoiding physical contacts per week. Santé publique France data also show how older individuals present a higher risk aversion compared to the other age classes. In fact, on average, seniors respected physical distancing 28% more than the other age classes as shown in Figure 4.4 (b). For this reason, we considered a further reduction of 30% in contacts for older individuals in the exit phase.

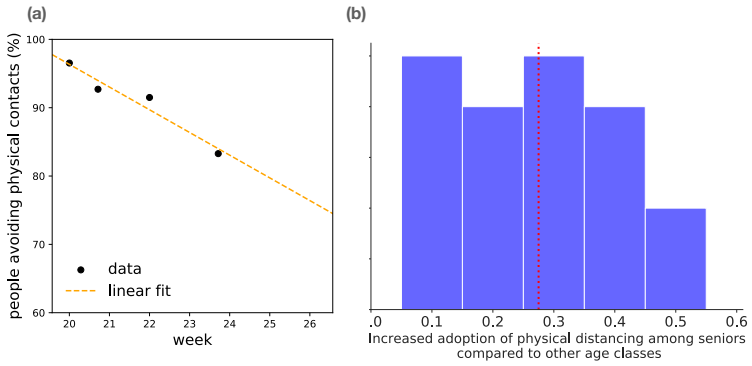


FIGURE 4.4: Percentage of individuals avoiding physical contact estimated from a large-scale survey conducted by Santé publique France [160]. a) Fitted percentage of individuals avoiding physical contact with respect to lockdown over time. Black dots are the estimation by Santé publique France. b) Probability distribution of the increased adoption of physical distancing adopted by seniors compared to other classes (n=20, with median 0.275 and 95% CI [0.074, 0.516]). The red dotted line indicates the median value.

Information on the progressive reopening of activities indicates that leisure and other activities were only partially open in the study period. Data, however, are not fine-grained enough to parameterize our model, so we assume a 50% opening of these activities and explore variations in the sensitivity analysis.

4.3 Metapopulation models

Modern epidemic models recognize the importance of the role of mobility fluxes and how these can substantially alter the probability of encounters, patterns of exposure, and the likelihood of disease propagation [2, 5, 47, 56, 66, 69, 161]. In the previous sections, I described the reaction-diffusion process used to model the disease dynamics in homogeneous mixing and age-structured models. In age-structured models, the mobility is implicitly modelled accounting for the number of contacts among age classes in any setting. However, the natural framework to explore the interplay between mobility, spatial structure and the epidemic transmission is the metapopulation approach. In this framework, compartmental models are integrated into a structured system, in which individual movements and space are explicitly considered.

In recent decades, metapopulation models have been largely developed and informed by demographic and mobility data for analyzing the spread of infectious diseases across time and spatial scales [5, 48, 69, 162, 163, 164]. The assumption behind the model is that the system is characterized by a spatial fragmentation of the population in sub-populations also called patches. Each patch represents a single well-defined social or ecological entity such as a geographical area, social community in human-related disease; or a pack, a herd, a habitat, in animal-related disease.

The metapopulation model has a multiscale structure as shown in Figure 4.5. At the microscopic level, each patch i is composed of a population of individuals. The health status of the individuals in any patch is modelled by a compartmental model. The compartment dynamics accounts for the possibility that individuals in the same location may make contact with and change their health status according to the infection dynamics. In the illustrative Figure 4.5 an example of a metapopulation

model is considered where a SIR dynamics in homogeneous mixing is modelled in each patch.

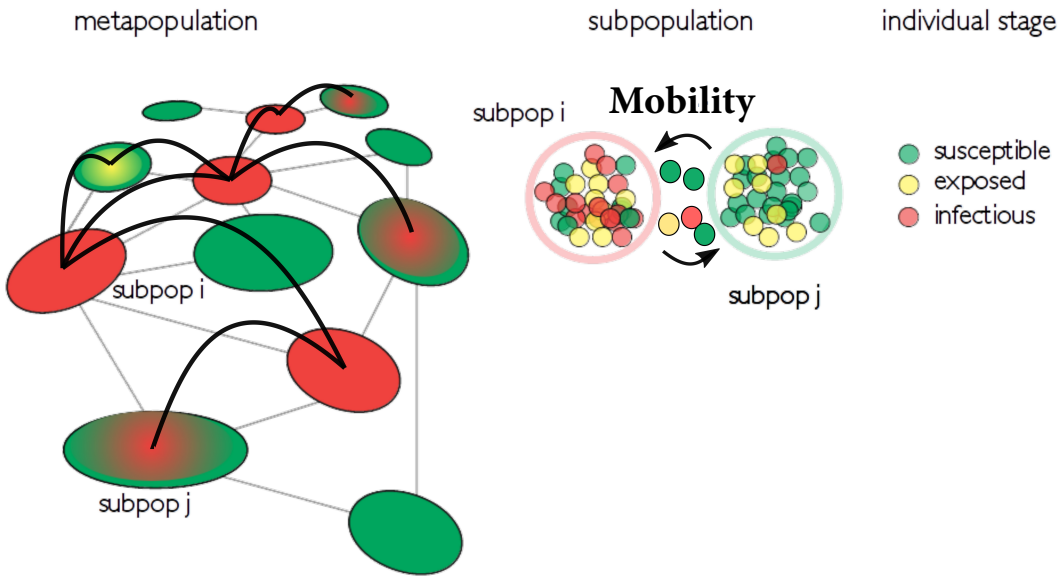


FIGURE 4.5: At the macroscopic level the system is composed of a set of coupled subpopulations. Coupling is defined from mobility connections. At the microscopic level, each subpopulation is a population of individuals. The infection dynamics in Figure is described by a compartmental model with only three compartment: S=Susceptible, E=Exposed, I= Infectious individuals.

At the macroscopic level, the metapopulation model has a network structure where each subpopulation is a node and each node i is connected to another node according to the movement of individuals between the two nodes. The mobility of individuals between subpopulations can be described using two different approaches. The first one is called **explicit metapopulation approach** as mobility fluxes between any pair of patches are defined by explicit mechanistic approaches [165, 166]. Individuals move from one patch to another with a given rate of movement and, once people have moved to a new patch, they feel the force of infection λ depending on the patch they have moved to. The second approach is called **effective metapopulation**. In this case, mobility patterns can be accounted for by effective couplings expressed as a force of infection generated by the infectious individuals in patch i on the individuals in patch j [167, 168, 169, 170, 171]. Both approaches have been widely adopted in recent years and informed by mobility data (e.g. mobile phone data, commuting data, or air traffic data) [5, 7, 29, 30, 38, 97, 106, 142, 145]. In Section 4.5, I presented my theoretical research work on how to aggregate individual trajectories from mobile phone data to inform such transmission models.

Explicit metapopulation model

The mechanistic approach used in explicit metapopulation models has a Markovian assumption as it does not imply any memory of the individual movements. The Markovian character is in the fact that individuals are not labelled according to their original patch and at each time step, the same probability of moving is applied to all individuals in the subpopulation without having memory of their origin. At

any time step, individuals thus may move according to a given probability p_{ij}. p_{ij} expresses the probability of an individual to travel from the patch i to the patch j and is defined by the mobility fluxes w_{ij} which are the number of travellers moving from the subpopulations i to j. w_{ij} may be inferred from a mobility process based on transportation, air traffic and commuting data [5, 10, 30, 38, 54]. More recent models have also used mobility fluxes from mobile phone data [11, 14, 15, 25, 26, 27, 28, 29].

Here, I reported the differential equations for an explicit metapopulation model in the simple case of a $SEIR$ compartmental model:

$$\frac{dS_i}{dt} = -\beta_i S_i \frac{I_i}{N_i} + \Omega_i^S \tag{4.15}$$

$$\frac{dE_i}{dt} = \beta S_i \frac{I_i}{N_i} - \epsilon E_i + \Omega_i^E \tag{4.16}$$

$$\frac{dI_i}{dt} = \epsilon E_i - \mu I_i + \Omega_i^I \tag{4.17}$$

$$\frac{dR_i}{dt} = \mu I_i + \Omega_i^R \tag{4.18}$$

with $\Omega_i^X = \sum_j (p_{ji} X_j - p_{ij} X_i)$ and

$$p_{ij} = \frac{w_{ij}}{N_i}. \tag{4.19}$$

where i defines all the parameters and variables referring to subpopulation i and S_i, E_i, I_i, R_i, denoting the number of susceptible, exposed, infected and recovered individuals at time t in a patch i.

In general, accounting for fact that the duration of trips delays the time of introduction of a pathogen, regardless of the pathogen, the fluctuations on the path of the spatial spread increase proportionally with pathogen-generation time[172]. By labelling the home location of individuals, model improvements have been then made in this direction to design a non-Markovian model which accounts for returning residents and the length of stay in any location [173, 174, 175]. All these complicated mechanistic patterns, however, can be accounted for more easily by using an effective metapopulation approach.

Effective metapopulation model

The effective metapopulation approach models mobility as an associated force of infection among subpopulations. The approach is also called kernel transmission method, as the individual movements are effectively translated into a spatial transmission kernel between subpopulations [176]. Using this approach, individuals not explicitly moving from one patch to another, but living in i felt a force of infection proportionally with the coupling between subpopulations. This approach is Non-Markovian as it has it has memory of the home location of individuals [44, 177, 178]. The advantage of this approach is that it does not require exceptional computational efforts to be computed as opposed to the explicit metapopulation approach. Moreover, another advantage is that the length of stay does not have to be inferred explicitly from mobility data, but is implicitly factored into the definition of the coupling forces (See Section 4.5). I will now report the differential equations for the effective metapopulation approach in the simple case of a $SEIR$ compartmental model:

$$\frac{dS_i}{dt} = -\lambda_i S_i \tag{4.20}$$

$$\frac{dE_i}{dt} = \lambda_i S_i - \epsilon E_i \tag{4.21}$$

$$\frac{dI_i}{dt} = \epsilon E_i - \mu I_i \tag{4.22}$$

$$\frac{dR_i}{dt} = \mu I_i \tag{4.23}$$

λ_i is the force of infection felt by susceptible people i.e. people in compartment S living in i and is defined as follows:

$$\lambda_i = \beta \sum_j f(M) I_j \tag{4.24}$$

$f(M)$ takes the role of the mobility into account and it may be defined depending on the mobility process and the disease considered.

I will now present the novel implicit SEIR metapopulation model I designed and developed. The model is stochastic, discrete, and Non-Markovian. The force of infection for a susceptible in patch i is given by three contributions: (i) the transmission of the infection by infectious individuals resident in i and not moving; (ii) the transmission of the infection by infectious individuals resident in j and spending time in i (visitors); (iii) the transmission of the infection by infectious individuals resident in i who spent time in j, got infected there and came back to their residency (returning residents).

Here I report the differential equations of the model I designed:

$$\frac{dS_i}{dt} = -\lambda_i S_i \tag{4.25}$$

$$\frac{dE_i}{dt} = -\epsilon E_i + \lambda_i S_i \tag{4.26}$$

$$\frac{dI_i}{dt} = -\mu I_i + \epsilon E_i \tag{4.27}$$

$$\frac{dR_i}{dt} = \mu I_i(t) \tag{4.28}$$

Taking into account the three terms of the mobility, the force of infection in patch i is calculated as:

$$\lambda_i = \lambda_{ii} + \sum_{i \neq j} \lambda_{ji}^v + \sum_{i \neq j} \lambda_{ij}^r. \tag{4.29}$$

with

$$\lambda_{ii} = \beta p_{ii}^2 \frac{I_i}{\hat{N}_i} \tag{4.30}$$

$$\lambda_{ji}^v = \beta p_{ii} p_{ji} \frac{I_j}{\hat{N}_i} \tag{4.31}$$

$$\lambda_{ij}^r = \beta p_{ij} \frac{\hat{I}_j}{\hat{N}_j} \tag{4.32}$$

where p_{ij} is the coupling probability between patches i and j,

$$\hat{N}_i = p_{ii}N_i + \sum_{j \neq i} p_{ji}N_j \tag{4.33}$$

$\hat{N}_i$ is the effective population in i considering the coupling among patches and

$$\hat{I}_i = p_{ii}I_i + \sum_{j \neq i} p_{ji}I_j \tag{4.34}$$

$\hat{I}_i$ is the effective number of infected individuals in i.

p_{ij} is called **coupling probability** as it is not simply the probability of moving but, as previously discussed in Section 4.5 represents the probability that in the subpopulation j a resident in j interacts with an individual living in i. The processes to define coupling forces among locations is not trivial, and it is discussed in detail in Section 4.5. In fact, in my research work, I focused my attention on how to use mobile phone data for translating high-resolution individual trajectories into coupling forces between locations for effective metapopulation models.

Invasion trees

With the aim of modelling the invasion dynamics, it is possible to compute the epidemic invasion trees [7, 11]. An invasion tree is the most probable spatial transmission path of the infection, and it is computed as follows. Considering a disease-free location i, as soon as $I_i \neq 0$, a direct link - between i and the source location from which the infection came - is tracked. Computing several runs, the invasion path is computed by summing for any pair of locations i, j the number of runs in which this contagion link exists. The invasion path is thus a network in which the nodes are the patches of the metapopulation and the weights of the links are the percentage of the number of runs in which this contagion link exists. Once the network of the invasion path is obtained, the invasion tree may be computed by measuring the direct maximum spanning tree [179]. A spanning tree is computed from a network by extracting the subgraph containing no cycles and the maximum spanning is the one that has a larger sum of weights on its links.

We compared the invasion trees through two metrics: the betweenness centrality distance [180] and the invasion distance. The betweenness centrality distance between coupling matrices X and Y is defined as follows:

$$d_{centrality}(X, Y) = \sqrt{\sum_{k=1}^{n} (b_i^X - b_i^Y)^2}. \tag{4.35}$$

where the betweenness centrality of a node i is given by the expression:

$$b(i) = \sum_{s \neq j \neq k} \frac{\sigma_{st}(i)}{\sigma_{jk}} \tag{4.36}$$

σ_{jk} is the total number of the shortest paths from node j to node k $\sigma_{jk}(i)$ is the number of those paths that pass through i. The shortest path in a network is a path between two nodes in which the sum of the weights of its links is minimized. Instead, the invasion distance in any invasion tree is measured for each node, and it is the number of edges connecting the considered node with the root (epidemic seed).

Considering every node, we obtain the distribution probability of the invasion distance and the betweenness centrality distance. We thus compering different invasion trees by comparing such probability distributions.

In addition to the epidemic trees, we also defined the invasion probability $p_{inv}(t)$ for a patch i as the probability that the epidemic has arrived in i in a given time t. We evaluated it at $t_{5\%}$, $t_{10\%}$, $t_{20\%}$ when respectively 5%, 10%, 20% of the patches were infected. For every patch we counted the number of runs where the epidemic arrived before $t_{5\%}$, $t_{10\%}$, $t_{20\%}$, respectively.

4.4 Modeling risk of introduction

Since the spatial transmission due to human movements, it is possible to assess the risk of importation from an affected (origin) area to a non-affected (destination) area. To compute the risk of invasion, two key factors have to be considered: prevalence in the location of origin and the mobility fluxes between considered locations. Importations may occur at different spatial scales ranging from local level carried by short-range trips to worldwide carried by long-range travel. Different sources of mobility data may be used depending on the context of interest. Within countries, mobile phone data have been largely used to assess the risk of introduction into new locations [16, 26, 52]. On the other hand, since the data available on mobile phones does not yet allow for the global tracking of international movements, the risk of international introduction has so far been commonly calculated from air traffic flows. [38, 47, 75, 78].

In my research work presented in Section 5.2, I computed the risk of importation of COVID-19 cases into Europe from China.

Computation of the probability of importing at least 1 case in Europe

First of all, we computed the probability of importing at least 1 case into Europe by considering a given infected area in China acting as seed of exportation. P_{EU} is the probability provided by Epirisk [181] that an individual travels from that area to Europe. If x cases are exported, the risk of importation of at least 1 case to Europe is computed as

$$1 - (1 - P_{EU})^x \tag{4.37}$$

Computation of the probability of importing 1 case in Europe except France, conditioned to observing

In a second moment, we computed the probability of importing 1 case into Europe except France, conditioned to observing 3 imported cases in France on January 27, 2020. Given the travel ban in Wuhan at that time, in this case, we considered 13 cities that are highly connected to Wuhan as possible seeds of case exportation out of China.

The general formula is defined as $P(y|\vec{x})$ where y is the number of cases potentially imported to Europe except the m countries in Europe with detected cases. Then, we decomposed $P(y|\vec{x})$ in:

$$P(y|\vec{x}) = \frac{P(y, \vec{x})}{P(\vec{x})} \tag{4.38}$$

$P(y, \vec{x})$ and $P(\vec{x})$ come from (m + 1)-dimensional and (m+ 2)-dimensional multinomial distributions, respectively. $P(\vec{x})$ is (m + 1)-dimensional as describe the probability of importing to one of the m countries, and probability to import somewhere else. $P(y, \vec{x})$ is (m+ 2)-dimensional as describe the probability of importing to one of the m countries, to import somewhere in Europe except for the m countries, and the probability to import somewhere else. We computed separately the numerator and the denominator:

$$P(\vec{x}) = n!(\prod_{j=1}^{m} \frac{P_j^{x_j}}{x_j!})\frac{(1-g)^{n-c}}{(n-c)!} \tag{4.39}$$

$$P(y, \vec{x}) = n!(\prod_{j=1}^{m} \frac{P_j^{x_j}}{x_j!})\frac{q^y}{y!}\frac{(1-g-q)^{n-c-y}}{(n-c-y)!} \tag{4.40}$$

Therefore,

$$P(y|\vec{x}) = \frac{(n-c)!}{y!(n-c-y)!}\frac{q^y(1-g-q)^{n-c-y}}{(1-g)^{n-c}} = \binom{n-c}{y}(1-\frac{q}{1-g})^{n-c}(\frac{q}{1-g-q})^y \tag{4.41}$$

with:
n: number of cases exported from China;
$\vec{x}$: m-dimensional vector encoding the number of cases in each European country with detected cases;
$c = \sum_{j=1}^{m} x_j$: number of detected cases imported to Europe;
p=m-dimensional vector encoding the importation probabilities in each European country with detected cases;
$g = \sum_{j=1}^{m} p_j$: probability of importing to any of the m countries with detected cases;
q: probability of importing to Europe except for the m countries;

Setting $y = 0$, we found the probability of having at least one case in Europe conditioned at the number of the imported cases (c) in Europe as:

$$P(y > 0, \vec{x}) = 1 - (1 - \frac{q}{1-g})^{n-c} \tag{4.42}$$

4.5 Mobility definition and resolution needed to inform metapopulation models

As I discussed in Section 3.5, there are different aggregation approaches to compute coupling forces among locations. By integrating coupling forces into metapopulation models, a crucial point is to accurately define the mobility process that is relevant to the epidemic spread, and the resolution that is needed to describe the invasion dynamics. As little attention has been devoted to this in literature, I focused on identifying the meaningful mobility fluxes extracted from mobile phone data needed to estimate the coupling probability between locations for epidemic transmission models.

To answer this question, I selected the three principal aggregation methods presented in the previous section: the Displacement-based coupling matrix D, the Location-based coupling matrix L and the Commuting-based coupling matrix C, at high,

medium and low resolution, respectively. I then assessed the impact of their different levels of aggregation on the modelled epidemic diffusion. Comparing D and L, the aim was to consider two possible approaches which take into account all the displacements of the users: the coupling between any two consecutive visited locations in D and the coupling between home location and any visited location in L. Instead, comparing L with C, the focus was to understand if commuting fluxes extracted in C could be enough for epidemiological purposes or instead, all visited locations should be considered.

To implement D,L and C, I used the pseudo-anonymised mobile phone individual trajectories collected by Orange. The dataset I used consists of a set of Call Detail Records (CDRs) of phone calls and text exchanges between Orange's customers in Senegal. I had access to this data as I was an Orange PhD student and I thus was effectively an Orange employee, otherwise non-aggregated data cannot be accessed by external researchers. As I already discussed, to comply with privacy regulations it is very common for mobile phone data providers to share already aggregated data with external researchers. Having the opportunity to have access to such high-resolution dataset, I found it extremely important to assess the impact of different aggregation procedures. By knowing the impact of different aggregation, researchers may have a better understanding on which aggregation to use or request of mobile data providers.

I analyzed a set of 15,859,942,126 records of 9,569,425 million mobile phone users (80 % of the Senegalese population) from January to December 2013 handled by 15999 Orange antennas. I only considered users having more than 30 days of mobile phone usage and having less than 1000 activities (calls/SMS) per week. Since shared phones is a very common phenomenon in Africa [130] the threshold on the maximum number of activities per week allows the avoidance of bias due to multi-ownership of the phone. Moreover, the cell towers are heterogeneously distributed over the whole territory of Senegal covering all 46 urban municipalities and 357 out of 437 rural ones. Besides, fluctuations in time of the number of active antennas means that the number of covered municipalities is time-varying over the year. So, considering only the covered municipalities throughout 2013, I selected 394 ones (46 urban and 348 rural). The 46 urban areas are located in the following cities: Dakar, Guediawaye, Pikine, Rufisque, Thies. After this preprocessing, I computed the coupling matrices D,L and C at municipality level per any given month. The illustrative Figure **??** shows the resulting three coupling matrices obtained from the individual trajectory of only one user.

To assess the role of the mobility definition on disease diffusion, I made a statistical comparison of the estimated coupling matrices and I then integrated the coupling matrices in a novel metapopulation model.

4.5.1 Statistical comparison of coupling matrices

Extracting the three coupling matrices D, L, C per month, I obtained 12 matrices for each method. Connections between municipalities depend on the aggregation process as described in Figure 4.6). To understand similarities and differences between each pair of matrices, I measured the Pearson coefficient of their elements, and I found that all matrices are highly correlated ($R_{C,L} = 1$,$R_{D,L} = 0.97$,$R_{D,C} = 0.99$). Even though all three matrices are correlated to each other, elements in D differ from the other two methods of an order of magnitude. In fact, while C and L have a quite similar probability distribution of the coupling probabilities, in D the median of the distribution, is around 1 order of magnitude lower (Figure 4.7 a). The Largest

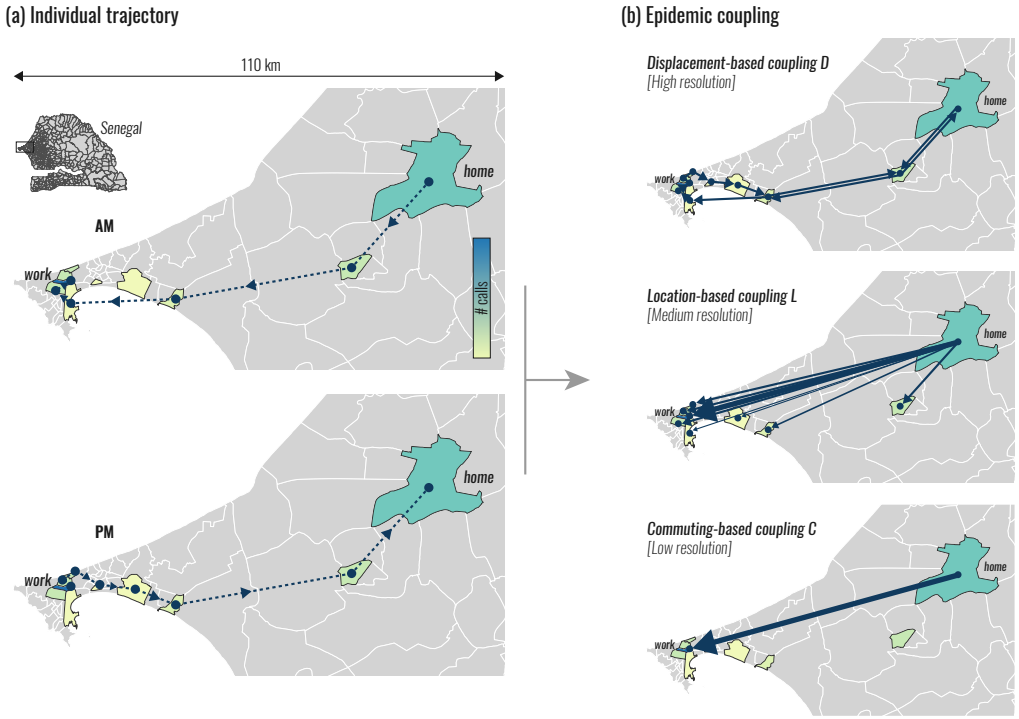


FIGURE 4.6: The aggregation of the individual trajectories. *D*) High-resolution method [27] *L*) Medium-resolution method [15] *C*) Lower-resolution method [52].

differences are on links connecting Urban/Rural municipalities (Figure 4.7 b). Also, the median of the geographical distance distribution is lower in D compared with the other two methods (Figure 4.7 c). Another key difference between D and L and C is on the outgoing probability in any location. Considering the coupling matrix, the outgoing probability is defined as follows:

$$p_{out} = \sum_{i \neq j} p_{ij} = 1 - p_{ii}. \tag{4.43}$$

Such differences increase between municipalities far from the urban areas (Figure 4.7 d,e,f) in which outgoing probability in D is around two orders of magnitude smaller compared with L and C.

I focused the comparison of the matrices on i) common connected locations and ii) links detected in one method and not in another one. In Figure 4.8 a) and b) I reported the distribution of relative variation of the common links between any pair of methods. Considering subsets of links that have relative differences higher than a certain cut-off, I found that the biggest differences are between D and L for links with the highest weight in L and the lowest in D. The probability of coupling in D can be up to 1000 times lower compared with the L (Figure 4.8 c, e). I found that these differences increase with the increase of the geographical distance between the coupled municipalities (Figure 4.8 c). Assuming the earth as a spherical body, I evaluated the geographical distance using the Haversine Formula between the centroids of any two municipalities. The same results were found in the comparison between D and C. Instead, C and L are quite similar. We found that links in C are no more than 10 times bigger than in L and their discrepancies are quite stable on the

geographical distance. (Figure 4.8 d, f).

Concerning (ii), I found (Figure 4.8 g) that the 37% of the links in L (27% in C) are not detected in D: 25% are links that have a weight that is higher than the median (26% in C) (Figure 4.8 g), 71% (73% in C) are links between municipalities at long-range distance (Figure 4.8 g). Otherwise, the fraction of links, existing in D and not in the other methods, represent people who pass from one municipality to another one not living in the location of origin. Moreover, the 12% of links detected in L do not exist in C. These links connect home locations to destinations in which people do secondary activities and thus as is shown in Figure 4.8 all these links have coupling probability lower than the median.

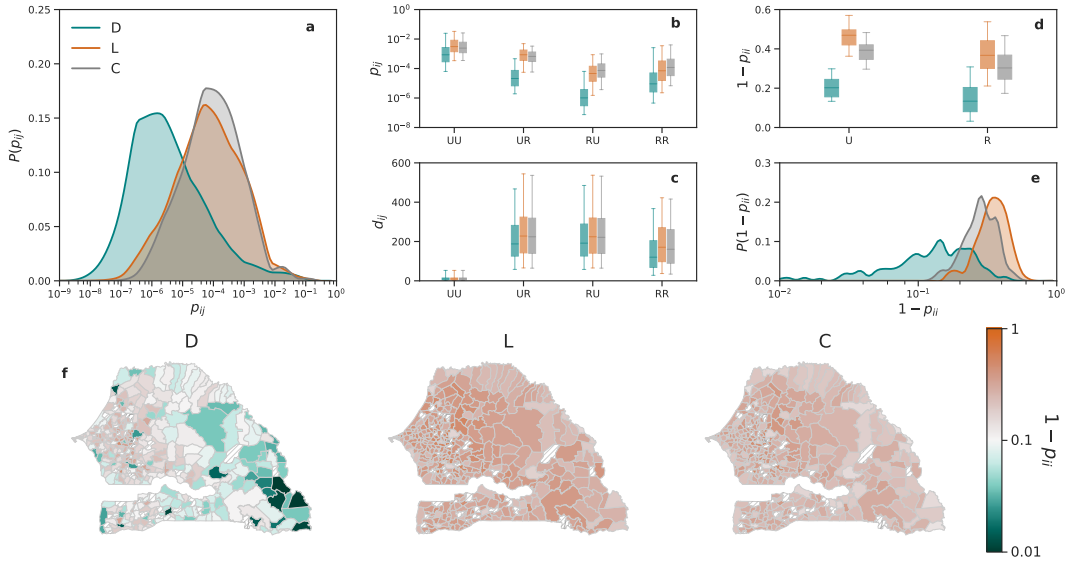


FIGURE 4.7: Coupling forces. a) Coupling probability distribution in *D*, *L*, *C*. Curves represent the average of the coupling probabilities among the 12 matrices. b),c),d) Box plots indicate the 95% reference range in January of the coupling probability, the outgoing probability distribution and the geographical distance, respectively. These represent a subset of links extracted by breaking down municipalities into Urban (U) and Rural (R). e) Outgoing probability distribution in *D*, *L*, *C*. f) Map of the outgoing probability in any municipality for *D*, *L* and *C*. Same distribution patterns have been found throughout thee years.

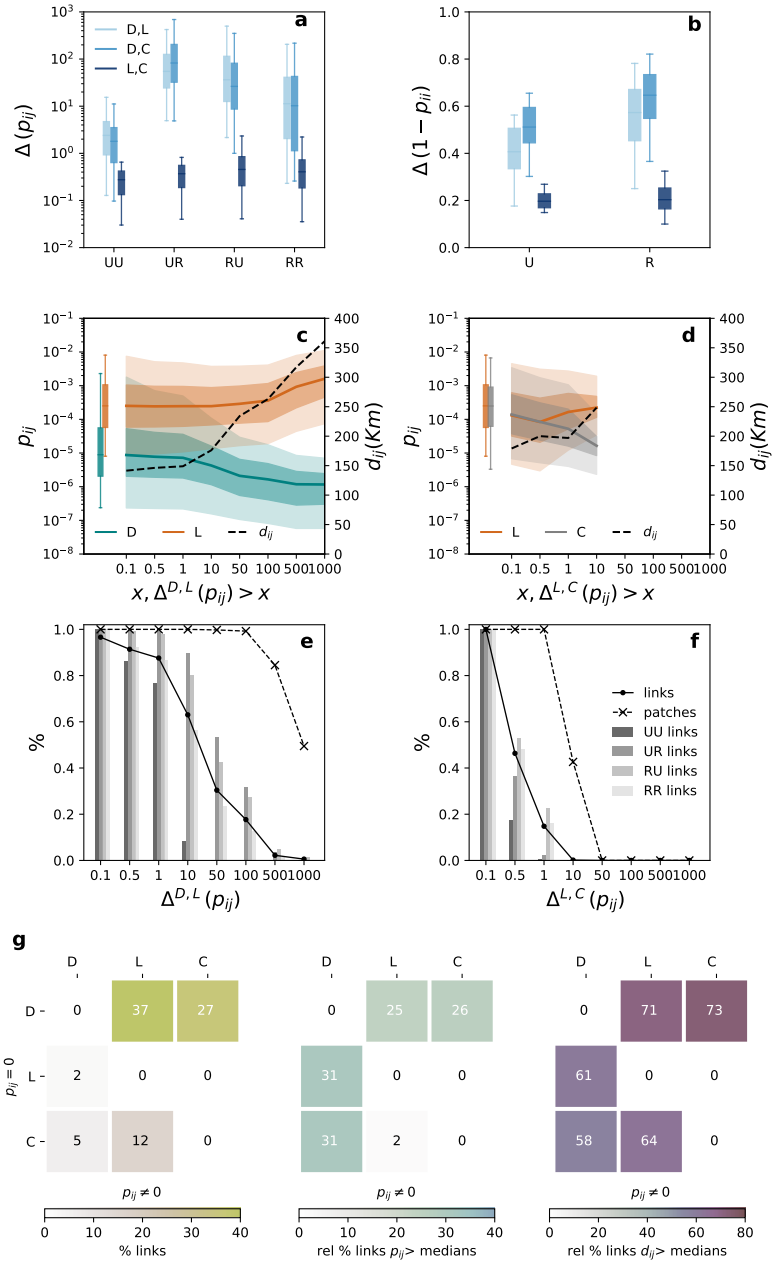


FIGURE 4.8: Differences between D, L and C. Box plots indicate the 95% reference range of the distribution of the relative difference of coupling probabilities of the common links among D, L, C and of the relative difference of outgoing probability of the common links among D, L, C in b). In c) and d) The curves display the coupling probability distribution for a subset of common links. Subsets of links are extracted a different cut-off of the relative variation between D, L and L, C on the coupling probability. The dotted lines show the median of the geographical distance distribution for the same subset of links. e) and f) show the percentage of links and patches present in any subset and the relative percentage of links extracted by breaking down municipalities into Urban (U) and Rural (R). g) Not-common links analysis. In the three matrices are shown the percentage of links detected by the aggregation method in the column and not detected by the aggregation method in the row (yellow matrix), the relative percentage of such not-common links with coupling probability higher than the median (green matrix) and the relative percentage of such not-common links with geographical distance higher than the median (violet matrix).

4.5.2 The aggregation approach affects the invasion dynamics

To assess the impact of the different coupling definitions on the epidemic diffusion, I used them to inform a spatially-explicit epidemic model at the level of municipalities. To design the model, I partitioned the Senegalese population into subpopulations at municipality level basing them on 2013 Census provided by the Senegalese national institute of statistics ANSD [182]. I thus considered 13092348 inhabitants distributed among these municipalities. The force of infection for a susceptible individual in a municipality i is given by three contributions: (i) the transmission of the infection by infectious individuals resident in i and not moving; (ii) the transmission of the infection by infectious individuals resident in j and spending time in i (visitors); (iii) the transmission of the infection by infectious individuals resident in i who spent time in j, got infected there, and came back to their residency (returning residents). All these components are weighted with the coupling matrices (See Section 4.3).

I considered three different epidemic scenarios: i) the top value in the confidential interval of Ebola Virus, $R_0 = 3$ (high transmissibility) ii) the top value in the confidential interval of Influenza, $R_0 = 1.5$ (medium transmissibility) iii) a scenario of an epidemic with control measures in action, $R_0 = 1.1$ (Low transmissibility). I then evaluated how the simulated epidemic behaviour depends on the underlying spatial and time aggregation scheme, by investigating the time to the first infection in each location and the invasion epidemic dynamics.

For each scenario, I compared the output of the simulated epidemic in D, L, C equally initialized. I initialized the epidemics with 10 infected people and I explored 92 potential epidemic seeds i.e. municipalities in which the epidemic began (46 urban areas and 46 rural ones). Epidemic seeds are selected by considering all the urban areas and the top 10% of rural ones which have the highest variation on the coupling probabilities among different approaches. I investigated a set of global observables to characterize the simulated epidemic in time and space. I then based the comparison of the matrices on two main observables. i) the arrival time of the infection in each location and ii) the epidemic invasion tree from the epidemic seed to the rest of the country. The arrival time of the epidemic in a given municipality is the time when the first case arrives. Instead, the invasion epidemic tree is the more likely route of the spatial invasion from the epidemic seed to the rest of the municipalities. Methods on how invasion epidemic paths are implemented are reported in Section 4.4.

Overall, the lower coupling probability measured in D compared to the other methods results in delayed arrival times. The arrival time in D ranges from few weeks to almost 410 days, while in L and C it is quite a bit lower, not exceeding 250 days. Comparing L and C, the median of the relative variation on the arrival times is 0 regardless of the transmissibility of the disease (Figure 4.9 a). Instead, comparing L and C with D, the relative variation is relatively higher (around 100%).

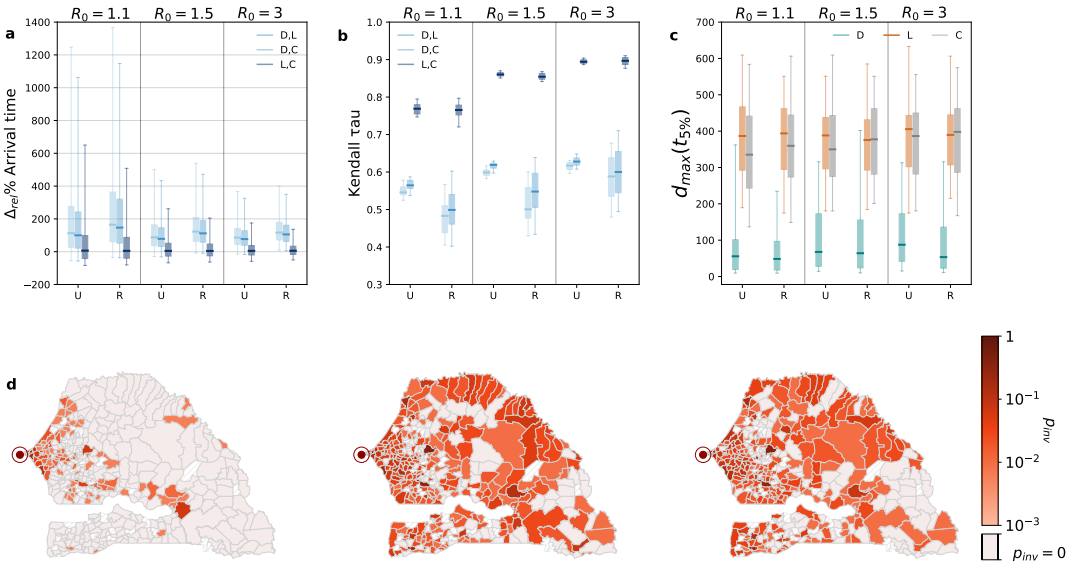


FIGURE 4.9: Differences on mode led epidemics. Box plots indicate the 95% reference range of a) The relative variation of the arrival times in any municipality between any pair of the three methods b) The Kendall tau probability by comparing D, L, D, C and L, C c) The maximum geographical distance achieved from the seed at $t_{5\%}$. $t_{5\%}$ is the time when the 5% of the municipalities have been infected ($t_{5\%}$). Analysis of urban (U) and rural (R) epidemic seeds were done separately. d) Visualization on the map of the invasion probability p_{inv} at $t_{5\%}$ in the three methods with a municipality of the capital Dakar as seed and $R_0 = 1.1$. The location with the red dot is the epidemic seed.

In addition, L and C present a high correlation in the ranking of the arrival times in each location (Kendall tau coefficient ranging from 0.65 to 0.90). Instead, it is not so high between D and the other two methods as shown in Figure 4.9 b). It means that L and C not only reproduce a quite similar distribution of the arrival time but also simulate the arrival of the epidemic in each place with the same order. Furthermore, the maximum distance achieved when the 5% of the municipalities have been infected is much smaller in D compared with L and C (Figure 4.9 d). By analysing the invasion trees, I observed a relation between the arrival time in D and the geographical distance and this explains the differences with C and L. As the map shows in Figure 4.9 e for $R_0 = 1.1$, the first 5% of municipalities infected in D are clustered close to the capital Dakar, instead, in L and C the epidemic invasion is more heterogeneous in space. Similar results are found with $R_0 = 1.5$, $R_0 = 3$.

Focusing on the spatial invasion, we found L and C reproduce similar paths of invasion of the epidemic in the country. To implement epidemic invasion trees, I selected only 2 seeds: i) the most populated urban area in Dakar and ii) a rural area in the department of Saraya. I selected this rural area as it is the farthest municipality from Dakar, and it is in one of the municipalities with the highest variation in the coupling probabilities between D and L. I analysed the similarity between epidemic trees by using a metrics based on the betweenness centrality (See Section 4.4). I found that the betweenness distance on the trees between C and L is lower compared with D,L and D,C as shown in Figure 4.10 a (D,L <0.05, D,C and D,L ranges from 0.05 to 0.25). Invasion trees are shown in Figure 4.10 d. I also found that in D the spatial transmission is mainly fragmented into short paths (Figure 4.10 b, c). To compare D, L and C, I thus quantified the epidemic invasion distance (d_{inv}) for any municipality i as the number of links in the invasion trees i.e. the number of municipalities reached from the seed before the infection arrives in i. Focusing on the first layer of infection ($d_{inv} = 1$), as is shown in Figure 4.10 d, e, in D, the infected locations are clustered close to the seed, while in L and C, these are more heterogeneously distributed in space. It means that C and L reproduce more realistic epidemic patterns integrating long-range transmission with local-range dispersal. A key difference in the invasion trees in D compared to C and L is played by one municipality: Touba, the second most populated Senegalese city after the capital Dakar. In C and L around the 15% of the municipalities are infected by Touba, instead, in D less than 5%. In Figure 4.9e it is also shown that the invasion probability at $t_{5\%}$ in Touba is around one order of magnitude higher in L and C than in D (epidemic seed is at Dakar). I defined the invasion probability $p_{inv}(t)$ for a patch i as the probability that the epidemic arrives in i in a given time t. Considering the strong daily connections for commuting and commercial exchanges between Dakar and Touba [183], this suggests that L and C perform better the modelled epidemics by detecting Touba as a relevant epidemic hotspot. In any case, the differences between the simulated epidemic outcomes decrease, as expected, with high values of the basic reproduction number.

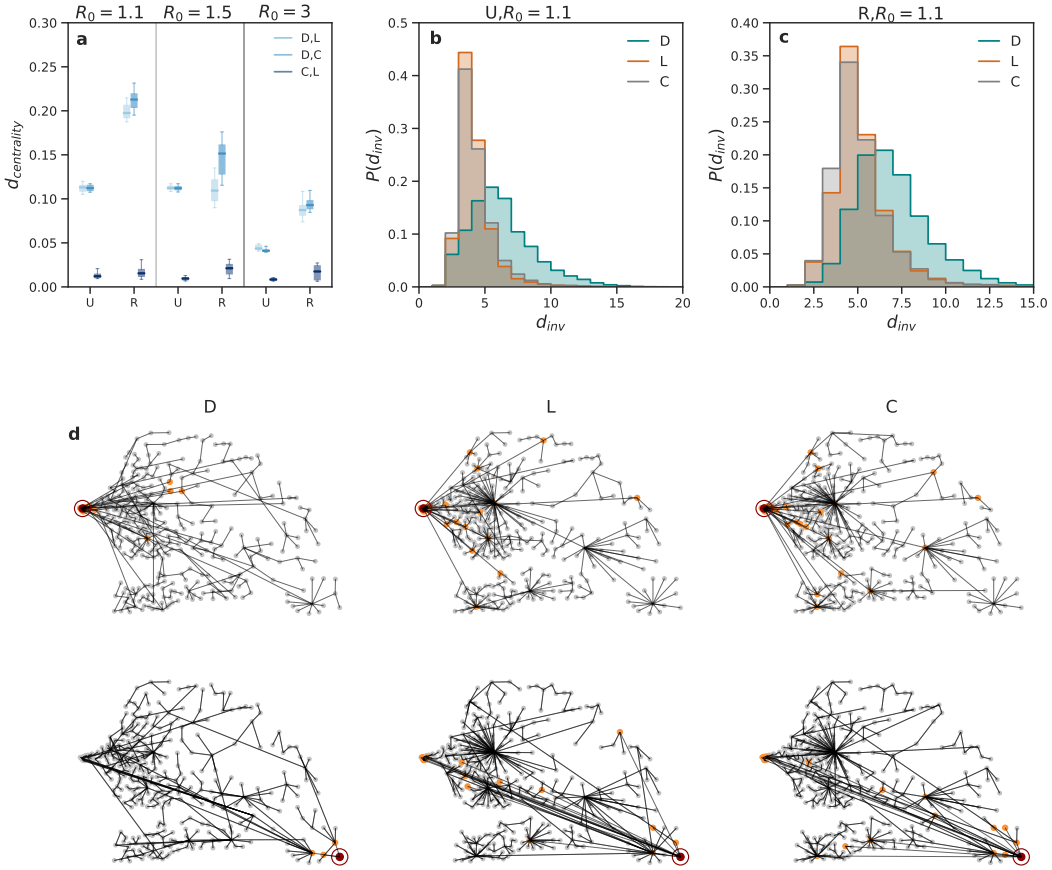


FIGURE 4.10: Epidemic invasion trees. Epidemic invasion trees have been computed for $R_0 = 1.1, 1.5, 3$ considering both an urban (U) and a rural seed (R). a) Box plots indicate the 95% reference range of the betweenness centrality distance index measured between the epidemic infection trees of any pair of methods. b) and c) The distribution of the invasion distance in the three methods when $R_0 = 1.1$. d) and e) The invasion trees when $R_0 = 1.1$. Epidemic seeds are in red. Locations directly infected by the seed are in orange, while locations infected by other nodes are in gray.

Improving coupling matrices

Finally, I introduced the matrices D', L', C' to improve the previous matrices D, L, C and I analyzed the differences in the relative epidemic outcomes. In L', I considered the time spent in a place by interpolating exactly the time spent in any location instead of using the number of calls as a proxy of time as in L. In C', I evaluated the most visited location over 12h (7am-7pm) instead of over all day as in C. In D', I considered the time lapse between each two displacements, instead of the number of displacements as in D.

Few differences exist between L, L' and C, C', while relevant ones have been found between D, D'. I found that in L' the outgoing probability decreases compared with L. It is probably a bias of the overestimation of the outgoing probability in L due to the fact that at night users make fewer calls. Relevant differences between D and D', instead, could mean that users make a lot of consecutive calls in a short period in the same place involving an underestimation of the outgoing probability in D as any two consecutive calls account for a displacement. Then, I found that the outgoing probability in C' is higher compared with C (Figure 4.11 a). Intuitively, it means that the most visited location, computed only over the 12 daily hours, more likely does not match with the home location compared with the one computed over the 24 hours.

Our findings indicate that preserving the full resolution of the observed trajectory of individual movements (D, D') may bias the spatio-temporal diffusion of the simulated epidemic in both the timing and pattern of invasion. Instead, aggregating on visited locations (L, L'), even if the sequence of the trajectories is not considered, reproduces realistic simulated patterns. The high similarity between L, L' and the commuting-based method C, C' suggests that secondary visited locations have no significant impact on the spread, and commuting-like mobility is the main driver of disease diffusion. D and D' are thus not well-defined enough to be embedded into transmission models. They could be a good option in the case where it is important to know the actual path of the individuals e.g. to study migration phenomena. As individuals change their home location, in this case, it would be incorrect to couple home-visited locations as in L or home-work as in C.

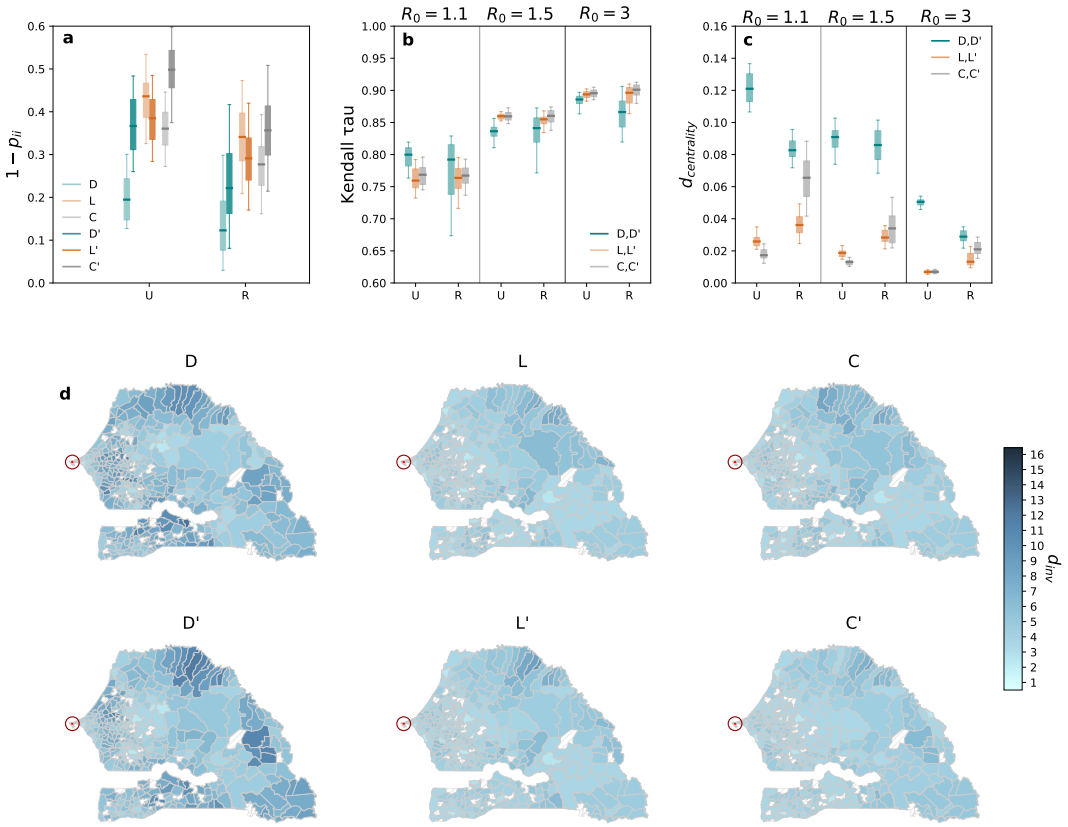


FIGURE 4.11: Sensitivity. a) The outgoing probability in January in D, L, C and D', L', C' by breaking down municipalities into Urban (U) and Rural (R) seeds, b) Kendall tau probability by comparing D, D', L, L' and C, C' by breaking down into urban and rural seeds. c) Betweenness centrality distance index measured between the epidemic invasion trees? of any pair of methods. Box plots indicate the 95% reference range. Betweenness centrality is measured on the invasion trees by selecting only an urban (U) and a rural seed (R). d) Visualization on map of the epidemic invasion distance when the capital Dakar (red circle) is the epidemic seed and $R_0 = 1.1$.

4.5.3 Discussion

Adequately aggregating human movements becomes particularly relevant for improving the reliability of infection disease models and predictive capacity. In my research work, I showed that it is necessary to aggregate mobile phone data properly, as the level of details of the individual trajectories in the aggregation process significantly impacts the epidemic outputs of the models.

I found two key results: i) preserving the maximum resolution of the individual trajectories may involve a bias in the simulated invasion dynamics; ii) commuting fluxes are enough to reproduce the spatial transmission. Keeping the full resolution of the individual trajectories, Displacement-based method D does not couple the actual origin and destination locations of the individuals. This implies an error when such a matrix is embedded into metapopulation models, as the modelled spatial transmission is delayed and do not present long-range contagions. In fact, D reproduces a process of invasion that is thus radial. Finally, p_{ij} in D is defined, accounting for users' displacements at different time scales by depending on the time elapsed between two consecutive calls. This may involve a bias when the matrix is integrated into discrete epidemic models which have a given timescale (e.g. weekly, daily, hourly). All these considerations suggest that D is not properly indicated to be integrated into transmission models aimed at reproducing the invasion dynamics. However, it may be embedded into probabilistic epidemic models and be used for epidemiological studies in which the temporal sequence of the individual trajectories is important, e.g. the impact of migration flows in the spatial transmission.

Conversely, C and L reproduce more realistic spatial and temporal epidemic patterns by simulating both short than long-range contagions. The high similarity between L and C means that all leisure activities considered in computing L can be neglected, and the information on daily commuting is already sufficient to define the coupling forces well. The commuting method may thus be considered a valuable aggregation approach in terms of accuracy and reliability for metapopulation models.

The work has, however, some limitations. Concerning the model, I developed a SEIR model with a hypothesis of homogeneous mixing in each municipality. Substructures of the population based on social and mobility behaviours may be considered. However, my aim was not to reproduce a real epidemic but rather to get a better understanding of the use of mobile phone data and the details required for the design of metapopulation models. Then, the model is not validated on epidemiological data. The realism of L and C is assumed on the basis that they reproduce heterogeneous patterns in the spatial spread, as it is well established in literature that epidemics spread through complex spatio-temporal patterns combining both short-range and long-range seeding events. This has been experienced both in epidemics with a medium transmissibility rate like Influenza [184] and in faster epidemics like Ebola [185].

The study was performed in Senegal in a peacetime period when ongoing epidemics were not present. Mobility patterns reproduce a regular situation presenting seasonal patterns, and they are thus reliable to describe commuting fluxes. We studied the aggregation processes at municipality level as it is the lowest administrative level. There is not a correct level of spatial fragmentation for studying infectious disease dynamics, it really depends on the target question and the spatial resolution of the epidemiological data available. Further research work should be done on comparing aggregation approaches at different spatial scales. Reducing the spatial scale, however, means averaging all the heterogeneity, therefore we expect the differences

between the aggregation methods to decrease drastically. We also expect that results we found will not change in other countries having a similar cultural, social, and economic situation.

While addressing a specific methodological problem of reducing data resolution, our findings also have important implications for the identification of the most relevant locations to be targeted for disease prevention and control and for the data and resolution needs in case of an emerging epidemic.

4.6 Conclusions

In this Chapter, I presented the mathematical and computational modelling background of my research and I discussed my theoretical work on assessing ways to embed mobility into transmission models. I introduced some basics of mathematical models used for studying infectious transmission. More specifically, in Section 4.1 I introduced compartmental models to describe the dynamics of disease in a homogeneous mixed population and I explained how to integrate social structures. I then focused on the age-structured compartmental model specific for describing COVID-19 epidemic in Section 4.2. In Section 4.2.1, I explained in detail the parametrization of the contact matrices that we used in the model to define social mixing during the first wave of COVID-19 in France. In Section 4.3, I introduced ways to integrate the mobility dimension in transmission models with a metapopulation approach, with a main focus on the stochastic metapopulation models we designed and developed. Given the modelled epidemics, I also explained how to compute the most likely spatial path of invasion, so-called invasion tree. Then, in Section 4.4 I explained mathematical tools to model the risk of introduction of a pathogen in a previously non-affected area. To conclude, in Section 4.5, I presented how to aggregate mobile phone individual trajectories to extract specific mobility indicator for metapopulation models. In the next Chapters, I will present my two main research works and I will show how we accounted for the mobility dimension at different scales and in different phases of the epidemic to study COVID-19 pandemic.

Chapter 5

Travel bans and global epidemic importations of COVID-19

In this chapter, I present my research work on the assessment of the risk of the introduction of COVID-19 infected cases from China to Europe by air travellers in the early stage of the COVID-19 pandemic. By spotting locations with the highest risk of importation, the aim was to allow European countries to be prepared to face the emerging epidemic in order to implement targeted surveillance in places served by high-risk airports.

In Section 5.1, I will describe the early stage of Covid-19 pandemic. Then, in Section 5.2, I will present how we integrated air traffic data into a mathematical framework to compute the risk of importing at least one case of COVID-19 from China to Europe prior to the travel ban in Hubei province, and after the travel ban on January 27 2020, when little information on the route of transmission of the disease and its effective viral circulation in China was available. The mathematical framework is extensively described in Section 4.4. The presented research work was published on January 30, 2020, in the scientific journal Eurosurveillance.

5.1 Early stage COVID-19 pandemic

On December 31, 2019 Chinese authorities detected a cluster of cases of pneumonia of unknown aetiology in Wuhan City, Hubei Province of China [79]. A novel coronavirus disease (2019-nCoV) was identified, on February 11, 2020, and the WHO announced the official name COVID-19. The situation evolved so fast that on January 23, 2020, 571 confirmed cases of 2019-nCoV were reported in several provinces in China and Chinese authorities enforced a travel ban on international flights in the province of Hubei. On January 27, 2020, local clusters were detected in most of the provinces in China, and forty-one imported cases were confirmed in the rest of the world including three cases in Europe, all in France. There was also a confirmed case in Germany not coming from China. He got infected by a Chinese guest visiting his company in Germany [186].

In order to prevent importation and secondary transmissions in not-affected areas, on January 25, 2020, the WHO Director for Europe issued a public statement underlining the importance of being ready to detect international travellers from locations in China where there were ongoing transmissions, at the local and national levels, . Even if WHO announced that the COVID-19 disease had human-to-human transmission, however, the route of transmission and epidemiological features of the disease were not clear at that stage.

During previous outbreaks due to other coronaviruses (MERS and SARS), human-to-human transmission occurred through droplets, contact, and fomites, suggesting that the route of transmission of COVID-19 could be similar. Starting from this knowledge, WHO published travel advice for international traffic on 27 January, advising exit screening for travellers coming from Chinese provinces and isolation for confirmed cases. Exit screening included detecting symptomatic individuals e.g. via temperature scanner and interviewing passengers with potential exposure to high-risk contacts, who had left the affected areas.

In order to help national health authorities to optimize surveillance efforts, we quantified the potential risk of importation of infected cases by air travellers to Europe from China with the little information available at that moment. We computed the importation risk in each country within Europe, and for each county, we assessed the risk of every airport. In the following Section, I will present the published article about this work.

5.2 Article #1: Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020

Rapid communication

Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020

Giulia Pullano[1] , Francesco Pinotti[1] , Eugenio Valdano[2] , Pierre-Yves Boëlle[1] , Chiara Poletto[1] , Vittoria Colizza[1]
1. INSERM, Sorbonne Université, Institut Pierre Louis d'Epidémiologie et de Santé Publique, IPLESP, Paris, France
2. Center for Biomedical Modeling, The Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, United States

Correspondence: Vittoria Colizza (vittoria.colizza@inserm.fr)

Citation style for this article:
Pullano Giulia , Pinotti Francesco , Valdano Eugenio , Boëlle Pierre-Yves , Poletto Chiara , Colizza Vittoria . Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020. Euro Surveill. 2020;25(4):pii=2000057. https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000057

Article submitted on 23 Jan 2020 / accepted on 30 Jan 2020 / published on 30 Jan 2020

As at 27 January 2020, 42 novel coronavirus (2019-nCoV) cases were confirmed outside China. We estimate the risk of case importation to Europe from affected areas in China via air travel. We consider travel restrictions in place, three reported cases in France, one in Germany. Estimated risk in Europe remains high. The United Kingdom, Germany and France are at highest risk. Importation from Beijing and Shanghai would lead to higher and widespread risk for Europe.

Starting December 2019, cases of pneumonia of unknown aetiology were reported in the city of Wuhan, in the province of Hubei in China [1]. The infective pathogen was later identified as a novel coronavirus, called 2019-nCoV [2]. As at 26 January 2020, a total of 1,988 confirmed cases have been reported from China [3]. The main affected area is in the province of Hubei, but as at 27 January 2020, confirmed cases have also been reported in 32 other provinces [4].

Forty-one travel-related cases were confirmed as at 27 January 2020, all coming from China. Twenty-seven cases were imported to Asia, six to North America, five to Oceania, and three to Europe [3,5-7]. Thirty of them were exported from Wuhan. In Europe, all three cases were imported to France. They were confirmed on 24 January 2020, with travel dates on 18 January 2020 (2 cases) and 22 January 2020 (1 case). One case was confirmed in Germany on 27 January 2020 with no history of travel to China but contact with a Chinese guest visiting their company [8]. In an effort to contain the spread of the virus, Chinese authorities enforced a travel ban in the province of Hubei starting on 23 January 2020 (3 a.m. Central European Time). This includes a complete ban on international flights [9].

Here we estimate the risk of importation of 2019-nCoV cases to Europe from infected areas in China by air travel. We compare the risk prior to the travel ban in Hubei province, with the risk updated to the outbreak situation of 27 January 2020, accounting for three cases imported to France and one case confirmed in Germany.

Modelling risk of importation

For this study, Europe is defined according to the Wikipedia contemporary geographical definition but with exclusion of transcontinental countries (Azerbaijan, Georgia, Kazakhstan, Russia and Turkey) [10]. The risk of importation to Europe is estimated as the probability that at least one case is imported from China to Europe. It is based on estimates from the platform EpiRisk [11] and accounts for origin-destination air travel flows of January 2019 from the Official Airline Guide (OAG) database of the GLEAM Project [11-13]. Details of the computation are provided in the Supplementary Material.

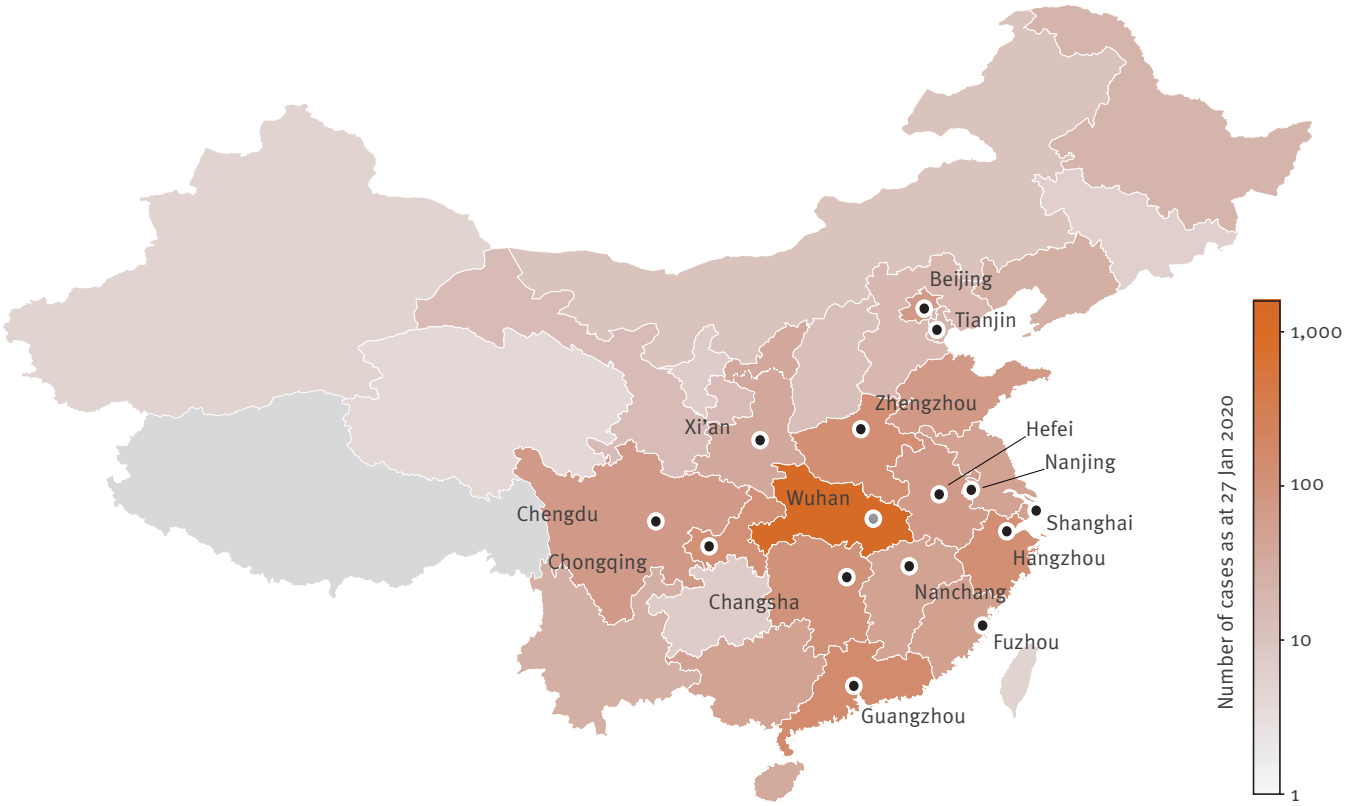
To estimate the risk in Europe prior to the travel ban in the Hubei province, we consider Wuhan as the only seed of the international spread [3,5-7]. We then provide a colour-coded map of Europe to report for each country the probability that a case imported to the continent arrives there, when coming from Wuhan only. For sensitivity, we tested whether the risk changes considering air travel flows of the month of February 2019.

To estimate the risk in Europe following the travel ban, we consider as possible seeds of case exportation out of China the cities that are highly connected to Wuhan based on de-identified and aggregated domestic population movement data (2013–2015) derived from Baidu Location-Based Services [14]. These cities are depicted in Figure 1. They were also found to be highly correlated with those reporting a high number of cases in the corresponding provinces [14].

To account for the current situation, including the three cases in France and one in Germany, we estimate the risk of importation to Europe except France

Figure 1

Map of Chinese provinces colour coded according to the number of cases of 2019-nCoV [4] as at 27 January 2020

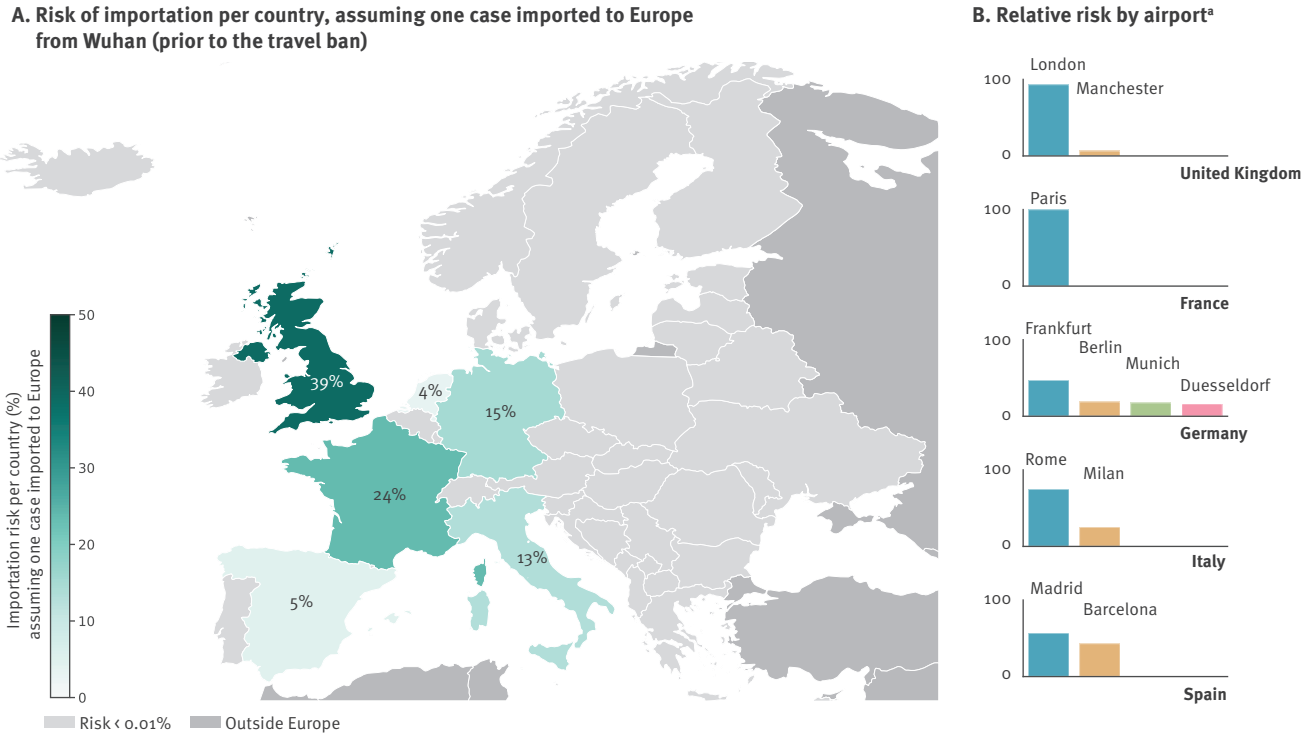


2019-nCoV: Novel coronavirus.

The 14 cities selected for the multi-source seeding [14] are shown with black dots: Beijing (in the province of Beijing), Changsha (Hunan), Chengdu (Sichuan), Chongqing (Chongqing), Fuzhou (Fujian), Guangzhou (Guangdong), Hangzhou (Zhejiang), Hefei (Anhui), Nanchang (Jiangxi), Nanjing (Jiangsu), Shanghai (Shanghai), Tianjin (Tianjin), Xi'an (Shaanxi), Zhengzhou (Henan). Wuhan (grey dot) is subject to the current travel ban.

Figure 2

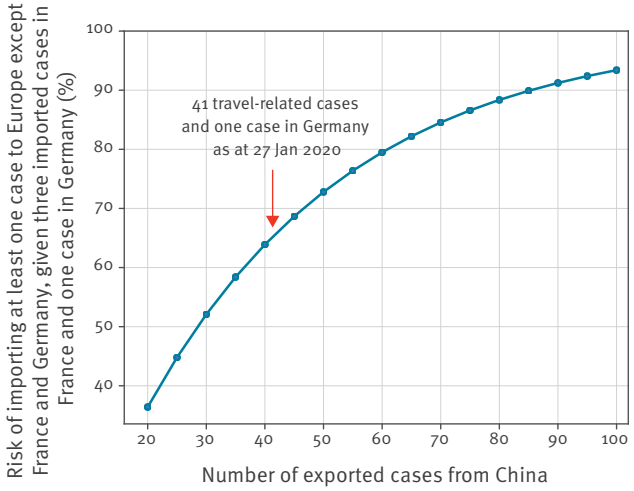
(A) Country-specific risk of importation assuming one case imported to Europe from Wuhan before the travel ban, and (B) relative risk by airport[a], January 2020



[a] When a city is served by several airports, these airports are considered as one entity.

Figure 3

Risk, as a function of the cumulative number of exported cases from China, of importing at least one case to Europe except France and Germany, given three imported cases reported in France and one case confirmed in Germany, January 2020



and Germany as the probability that Europe (France and Germany excluded) imports at least one travel-related case from China, conditioned to the observation of three cases imported to France and one case in Germany. Details of the computation are provided in the Supplementary Material. We estimate the risk for a varying number of exported cases from China, cumulative in time, to account for likely detection delays or under-detection of travel-related cases. As before, we then provide a colour-coded map of Europe to report for each country the probability that a case imported to the continent arrives there, when coming from cities depicted in Figure 1, except Wuhan. For sensitivity, we also tested whether the risk changes due to the additional inclusion of Wuhan in the multi-source seeding.

Estimated importation risk from Wuhan before the travel ban in Hubei province

The exportation of 30 cases from Wuhan before the travel ban, as reported so far, was estimated to put Europe at 61% risk of importing at least one case. The risk was localised in Western European countries, with the highest risk estimated for the United Kingdom (UK; 39%), followed by France (24%), and Germany (15%) (Figure 2). In some countries, importations are likely to occur at multiple airports (e.g. Germany, Italy, Spain), whereas in others the risk is mostly concentrated in airports serving the capital city (e.g. London in the UK, and Paris in France).

Estimated importation risk from considered areas of China following the travel ban in Hubei province

The probability that at least one case is imported to Europe except France and Germany, given the three imported cases reported in France and one case confirmed in Germany, is high (Figure 3). It is estimated to be more than 64% for the number of travel-related exportations from China reported so far (41 travel-related and one confirmed case in Germany). The probability becomes larger than 80% if 60 cases are exported from China.

In the event that one travel-related case is imported to Europe, the risk of importation is highest in the UK (25%) (Figure 4). Germany and France, which already have confirmed cases, rank second and third with a probability of 16% and 13% to receive another case, respectively. Italy (11%) and Spain (9.5%) rank as fourth and fifth in terms of risk. The risk is in general higher in more populated countries (Supplementary Figure S1). Also Eastern Europe and Northern Europe would be at risk of importing cases.

In the UK and France, the airports serving the capital cities continue to contribute the largest likelihood of importing cases (London contributes to 83% of the risk, Paris contributes to 94% of the risk, respectively). The estimates account for the travel ban imposed in the province of Hubei. Including travel flows from Wuhan, to account for cases who may have flown before the travel ban and are not yet detected, does not alter the estimations (data not shown).

Discussion and conclusions

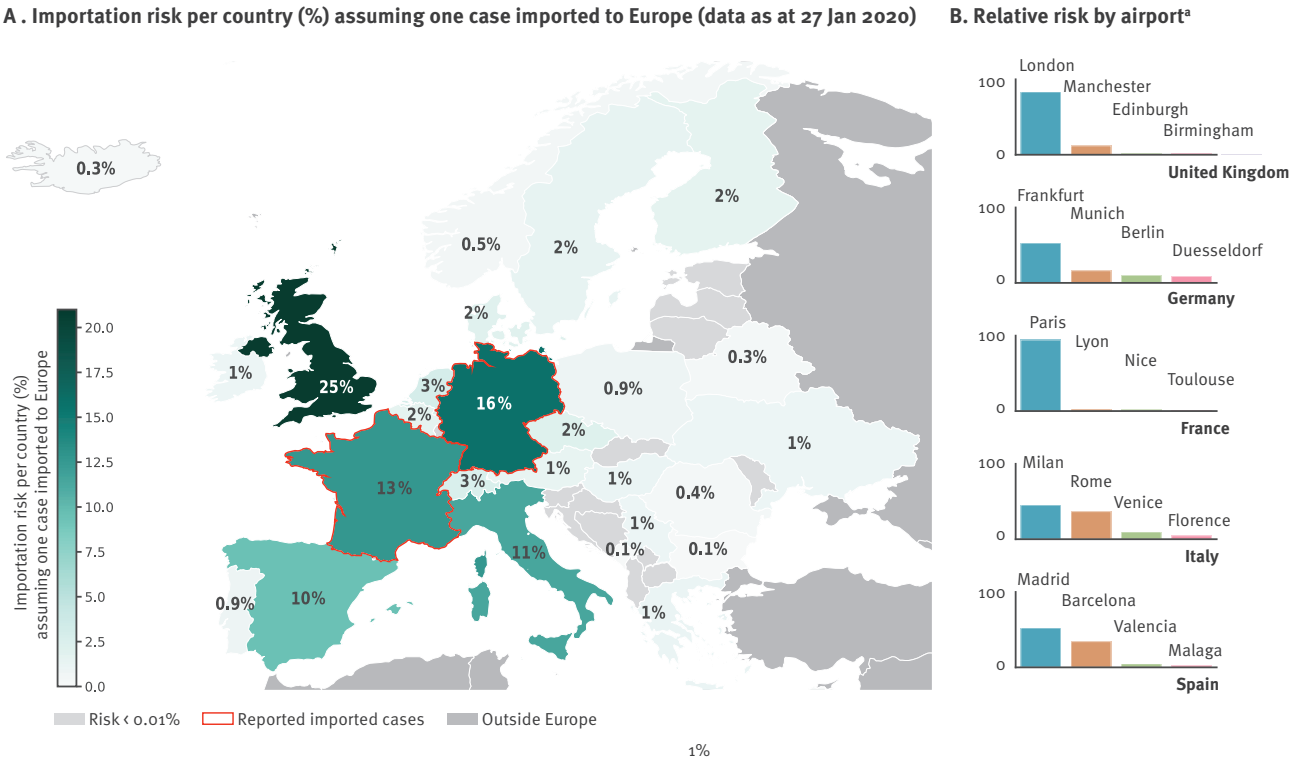
France reported on 24 January 2020 the importation of three 2019-nCoV confirmed cases from China, and Germany confirmed its first case on 27 January 2020 with no history of travel to China. They are still the first and only imported cases confirmed in Europe, at the time of writing. We estimate that the risk of importation of at least one case to Europe except France and Germany is high. It is larger than 80% if 60 travel-related cases are exported from China. The three countries at highest risk are the UK, Germany, and France (confirming estimates reported by other studies [12,14,15]), with the latter two countries already reporting cases. Delays are expected from date of importation to date of identification that may bias observations at the time of writing. All three cases imported to France were confirmed on 24 January 2020, with two travelling on 18 January 2020 (6 days delay) and one on 22 January 2020 (2 days delay).

The risk pattern of 2019-nCoV importation estimated for Europe varies considerably depending on the geographical extent of the affected areas in China. In particular, a larger area acting as seed of exportation that includes Shanghai and Beijing (two cities with larger number of travellers to more widespread areas in Europe) would likely result in a higher and more widespread risk for Europe.

Our results are based on available data and estimates of the affected provinces in China and account for origin-destination travel fluxes from these provinces, as well as the travel ban enforced in the Hubei province.

Figure 4

(A) Country-specific risk of importation assuming one case imported to Europe from the multi-source seeding of Figure 1 and (B) relative risk by airporta, January 2020



[a] When a city is served by several airports, these airports are considered as one entity. For each country, only the four most important cities in terms of agglomeration of airports and passenger traffic are represented.

Updated to data as at 27 January 2020.

However, estimates are sensitive to different health-seeking behaviours that infected travellers may have, and to the active surveillance practices put in place in European countries. We did not provide estimates of the expected number of imported cases per country, as this depends on the number of travel-related exported cases from China, a variable that is still hard to assess at this early stage.

Risk maps will need to be rapidly updated as the outbreak situation evolves.

Acknowledgements

We thank the GLEAM Project for the use of EpiRisk, Simon Cauchemez and REACTing (https://reacting.inserm.fr/) for useful discussions.

Funding statement: This study was partially supported by the ANR project DATAREDUX (ANR-19-CE46-0008-03) to VC; the H2020 MOOD project to VC, CP, PYB.

Conflict of interest

None declared.

Authors' contributions

GP, FP, EV performed the analysis.

EV, PYB, CP, VC conceived and designed the analysis.

VC wrote the manuscript.

References

1. World Health Organization (WHO). Pneumonia of unknown cause – China. Geneva: WHO; 2020.
2. World Health Organization (WHO). Novel Coronavirus – China. Geneva: WHO; 2020.
3. European Centre for Disease Prevention and Control (ECDC). Outbreak of acute respiratory syndrome associated with a novel coronavirus, China; First cases imported in the EU/EEA; second update – 26 January 2020. Stockholm: ECDC; 2020. Available from: https://www.ecdc.europa.eu/sites/default/files/documents/Risk-assessment-pneumonia-Wuhan-China-26-Jan-2020_0.pdf
4. Mobs Lab. Situation report Mainland China dashboard. [Accessed 27 Jan 2020]. Available from: https://docs.google.com/spreadsheets/d/e/2PACX-1vQUoSIALScXx8VXDX7yKNKWWPKE1YjFlWc6VTEVSN45CklWWf-uWmprQIyLtoPDA18tX9cFDr-aQ9S6/pubhtml
5. Johns Hopkins University Center for Systems Science and Engineering. Wuhan Coronavirus (2019-nCoV) Global Cases, January 27, 2020, 9am EST. [Accessed 27 Jan 2020]. Available from: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6
6. Australian Government, Department of Health. Update on novel coronavirus, Jan 28, 2020. [Accessed 30 Jan

2020]. Available from: https://www.health.gov.au/news/update-on-novel-coronavirus

7. Korea Centers for Disease Control & Prevention. The fourth imported case of 2019 novel coronavirus(2019-nCoV) has been confirmed in Korea. 27 Jan 2020. [Accessed 27 Jan 2020]. Available from: https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030
8. European Centre for Disease Prevention and Control (ECDC). ECDC statement following reported confirmed case of 2019-nCoV in Germany. 28 Jan 2020. Stockholm: ECDC. [Accessed 30 Jan 2020]. Available from: https://www.ecdc.europa.eu/en/news-events/ecdc-statement-following-reported-confirmed-case-2019-ncov-germany
9. China halts flights and trains out of Wuhan as WHO extends talks. Channel News Asia. 26 Jan 2020. [Accessed 26 Jan 2020]. Available from: https://www.channelnewsasia.com/news/asia/wuhan-virus-quarantine-city-flights-trains-china-12306684?cid=h3_referral_inarticlelinks_24082018_cna
10. Wikipedia. the free encyclopedia. St. Petersburg (FL): Wikimedia Foundation, Inc. 2001 – Europe; contemporary definition. [Modified 28 Jan 2020; Accessed 28 Jan 2020]. Available from: https://en.wikipedia.org/wiki/Europe#Definition
11. EpiRisk. Available from: https://epirisk.net/
12. Chinazzi M, Davis JT, Gioannini C, Litvinova M, Pastore y Piontti A, Rossi L, et al. Preliminary assessment of the International Spreading Risk Associated with the 2019 novel Coronavirus (2019-nCoV) outbreak in Wuhan City. Available from: https://www.mobs-lab.org/uploads/6/7/8/7/6787877/wuhan_novel_coronavirus__6_.pdf
13. GLEAM Project. Available from: https://www.gleamproject.org/
14. Lai S, Bogoch II, Watts A, Khan K, Li Z, Tatem A. Preliminary risk analysis of 2019 novel coronavirus spread within and beyond China. Available from: https://www.worldpop.org/resources/docs/china/WorldPop-coronavirus-spread-risk-analysis-v1-25Jan.pdf
15. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of Unknown Etiology in Wuhan, China: Potential for International Spread Via Commercial Air Travel. J Travel Med. 2020;taaa008. https://doi.org/10.1093/jtm/taaa008 PMID: 31943059

License, supplementary material and copyright

Any supplementary material referenced in the article can be found in the online version.

5.3 Conclusions

In the article, we integrated air traffic data into a mathematical model to compute the risk of importation of COVID-19 cases from Chinese provinces to Europe at the end of January 2020, before and after the travel ban in the province of Hubei.

We estimated the probability of importing at least one case to Europe, excluding France and Germany, given that the already imported cases were more than 64%, and considering the forty-one travel-related confirmed cases on January 27, 2020. Moreover, we found that the countries at highest risk on 27th January 2020 were UK (25%), Germany (16%), France (13%), Italy (11%), and Spain (9.5%), and that London, Frankfort, Paris, Milan and Madrid, were the airports at highest relative risk in each country, respectively, suggesting that targeted surveillance efforts in such places be improved.

Our results were quite in accord with what we then experienced, showing how most of the detected cases arrived in hub airports strongly connected with Chinese provinces, for example, Paris, Milan and London. The first cluster of COVID-19 with local transmission was in fact confirmed in Bavaria, Germany on January 29, 2020, and afterward, new imported cases from China were detected; two in France (Paris airport), two in Italy (Milan airport), two in UK (York and London airports), one in Finland and one in Sweden [187, 188]. Local spread in northern Italy was then detected on February 22, 2020 [186]. In a short time, cases also started to spread among European counties and soon COVID-19 reached all of Europe. On March 13, 2020, as the number of new cases became greater than those in China, the World Health Organization (WHO) began to consider Europe as the active centre of the COVID-19 pandemic [186].

At the time when this research work was done, there was no knowledge on the underreporting of imported cases, we were thus unable to accurately estimate the importation risk, even when taking undetected imported cases into account. To generalize our analysis, we estimated the relationship between the risk of importation into Europe and the number of imported cases from China. As this empirical low shows, the risk of importation becomes higher than 80% if only 60 travel-related cases are exported from China. Subsequent studies have estimated that less than 40% of exported cases were detected by surveillance systems [187, 188]. This means that imported cases should have been more than 100 at that time, corresponding to a risk of importation into Europe of over 90%.

Our results, however, were not predictions, but were simply aimed at detecting airports with the highest risk of importation. As local transmission was already self-sustainable in China when we performed the work, we should have integrated air traffic into transmission models, estimating the expected number of imported cases due to the actual situation of incidence. However, at that time we did not know the stage reached by the epidemic seeing that the first period of the spread was silent due to the high percentage of infected individuals passing undetected with mild or no symptoms. Moreover, we did not have any information on the disease dynamics and on the delay between infection and case report to calibrate an ad hoc predictive model. Later on, having a broader understanding of the disease, works were done in this direction [38, 189].

Given the lack of information on the disease in January 2020 and the fast epidemic transmission, the efforts of the surveillance systems were not sufficient to contain the COVID-19 global diffusion. In fact, travel restrictions in the early stage of the pandemic were too few and too late.

First, the local transmission in most of the Chinese provinces explains the negligible impact of travel restrictions put in place only in Hubei province. In fact, on 23 January 2020, the epidemic was seeded in several locations across mainland China [38]. On [41], researchers inferred that the epidemic was already growing exponentially in multiple major cities of China with a lag time behind the Wuhan outbreak of about 1–2 weeks. This meant that big cities outside Wuhan become the new potential origin of importation. In fact, before the travel ban, around 86% of the internationally imported cases came from Wuhan. After the implementation, around 86% of the internationally imported cases came from Shanghai (28.1%), Beijing (14%), and Shenzhen (12.8%) as reported on [38].

Secondly, the travel restrictions were implemented too late, as other countries already had local transmission. There were more cases of coronavirus than officially reported, around two out of three cases passed undetected across international borders [187]. Before the travel ban, 30 exported cases from Wuhan were confirmed. Considering underreporting, at least 50 undetected but infectious travellers were already on the move when the travel bans went into effect. Given that situation, and other public health non-pharmaceutical interventions, such as increasing awareness, isolating sick people, or at worst, social distancing measures like lockdown became necessary to try to mitigate the ongoing spread.

Chapter 6

Parametrizing social mixing from mobility data to model the first wave of COVID-19 in France

In this Chapter, I present my research work on estimating the underreporting of symptomatic cases of COVID-19 in France in May-June 2020 (on emerging from the first national lockdown) through epidemic transmission models specific for COVID-19. During an epidemic, adaptive behaviours such as risk aversion and social and mobility restrictions put in place by governments involve changes in social mixing and thus reduce contacts. This implies many challenges in designing proper models which capture such changes. In my research work, in order to model the first wave of COVID-19 in France, I parametrized social contacts over time using mobility and behavioural data to properly describe the local transmission at that time. In the following Section, I will present the intervention put in place by the French government to mitigate the first wave of Covid-19 and how I embedded them into the transmission models. The presented research work was published in Nature Journal on December 21, 2020.

6.1 Social mixing reductions due to control strategies

In order to mitigate the first wave of the COVID-19 epidemic in France, a national lockdown was implemented (March 17 - May 10, 2020) in which schools, non-essential businesses, companies and factories were closed. Lockdown was efficient in reducing incidence, for example, in Île-de-France the reproductive number decreased from around 3 prior to the lockdown to 0.68 during lockdown thanks to a reduction of 81% in the average number of contacts [81]. The French government thus decided to reopen non-essential businesses, companies and factories and to partially reopen schools. In order to avoid a resurgence of infections, however, teleworking and social distancing were highly recommended. At the same time, an aggressive testing strategy was planned in order to promptly identify infectious individuals and isolate them with the aim of keeping the epidemic activity low. By aiming to help health authorities to improve tracking of infected cases on emerging from the first lockdown, we thus decided to quantify the rate of detection of symptomatic cases in each region in France. Improving the surveillance system is crucially important as it allows us to reduce interventions.

To this end, we designed and developed age-structured regional transmission models specific for COVID-19 (see Section 4.2) accounting for the incubation period, pre-symptomatic and asymptomatic transmission, age-dependent susceptibility and infectiousness, different degrees of symptom severity and hospitalization.

In the age-dependent models, we consider the structure of social relationships i.e. social mixing per age bracket. Social mixing within the population is informed in the model through contact matrices that quantify the average daily number of contacts between age groups performed by individuals in several settings: at work, at home, at school, during leisure time, during transport, or in other unspecified contexts. Contact matrices in France for a regular situation were estimated from a survey launched in 2015 [94]. When control strategies were put in place, mobility and social interactions in each of these locations changed depending on policies implemented such as school closure, work from home, closure of places of mass gathering (e.g. gym, theatre, cinema). In my research work, I thus parametrized changes in mobility and thus in social contacts in France during the first national lockdown (March 17 - May 10, 2020) and in the 7 weeks after lockdown (May 11 - June 28, 2020). We parametrized such matrices through mobility and behavioural data. In particular, we reduced contact consideration % of attendance at school [159], adoption of physical distancing over time [190] and percentage of teleworkers informed by mobile phone data [121]. Other studies also proved that seniors have a higher risk aversion behaviour compared to other age groups leading to an average additional 30% reduction of their physical contacts [190] . After parametrizing contact matrices, we integrated them into the regional transmission models and we quantified the underdetection of COVID-19 cases in France at emergence from lockdown.

In the following section, I will present the resulting article about my research work.

6.2 Article #3: Underdetection of COVID-19 cases in France threatens epidemic control

Article

Underdetection of cases of COVID-19 in France threatens epidemic control

https://doi.org/10.1038/s41586-020-03095-6

Received: 11 August 2020

Accepted: 8 December 2020

Published online: 21 December 2020

Check for updates

Giulia Pullano[1,2,7], Laura Di Domenico[1,7], Chiara E. Sabbatini[1], Eugenio Valdano[1], Clément Turbelin[1], Marion Debin[1], Caroline Guerrisi[1], Charly Kengne-Kuetche[1], Cécile Souty[1], Thomas Hanslik[1,3,4], Thierry Blanchon[1], Pierre-Yves Boëlle[1], Julie Figoni[5], Sophie Vaux[5], Christine Campèse[5], Sibylle Bernard-Stoecklin[5] & Vittoria Colizza[1,6] ✉

As countries in Europe gradually relaxed lockdown restrictions after the first wave, test–trace–isolate strategies became critical to maintain the incidence of coronavirus disease 2019 (COVID-19) at low levels[1,2]. Reviewing their shortcomings can provide elements to consider in light of the second wave that is currently underway in Europe. Here we estimate the rate of detection of symptomatic cases of COVID-19 in France after lockdown through the use of virological[3] and participatory syndromic[4] surveillance data coupled with mathematical transmission models calibrated to regional hospitalizations[2]. Our findings indicate that around 90,000 symptomatic infections, corresponding to 9 out 10 cases, were not ascertained by the surveillance system in the first 7 weeks after lockdown from 11 May to 28 June 2020, although the test positivity rate did not exceed the 5% recommendation of the World Health Organization (WHO)[5]. The median detection rate increased from 7% (95% confidence interval, 6–8%) to 38% (35–44%) over time, with large regional variations, owing to a strengthening of the system as well as a decrease in epidemic activity. According to participatory surveillance data, only 31% of individuals with COVID-19-like symptoms consulted a doctor in the study period. This suggests that large numbers of symptomatic cases of COVID-19 did not seek medical advice despite recommendations, as confirmed by serological studies[6,7]. Encouraging awareness and same-day healthcare-seeking behaviour of suspected cases of COVID-19 is critical to improve detection. However, the capacity of the system remained insufficient even at the low epidemic activity achieved after lockdown, and was predicted to deteriorate rapidly with increasing incidence of COVID-19 cases. Substantially more aggressive, targeted and efficient testing with easier access is required to act as a tool to control the COVID-19 pandemic. The testing strategy will be critical to enable partial lifting of the current restrictive measures in Europe and to avoid a third wave.

Surveillance and detection aim to rapidly identify and isolate cases to prevent onward transmission of SARS-CoV-2 in the community and to avoid a substantial resurgence of cases of COVID-19. After an initial period—during which, because of a limited capacity, testing for SARS-CoV-2 infections mainly focused on severely ill patients—a new testing policy was implemented in France to systematically screen for potential infections with SARS-CoV-2 and enable lifting of the lockdown restrictions on 11 May 2020[8].

The specific characteristics of COVID-19, however, hinder the identification of cases[9–11]. Large proportions of asymptomatic infectious individuals[12], and the presence of mild or paucisymptomatic infections that easily go unobserved[9,11], present serious challenges to the detection and control of SARS-CoV-2[9,10,13]. Missing a substantial portion of infectious individuals compromises the control effort, enabling the virus to silently spread[10–12]. Synthesizing evidence from virological[3] and participatory syndromic surveillance[4] with mathematical models[2,14] that account for behavioural data[15–18], we assessed the performance of the new testing policy in France and identified its main limitations for actionable improvements.

COVID-19 surveillance

Management of the COVID-19 pandemic in France after lockdown in spring (May–June) 2020 involved the generation of a centralized database that collected all data on virological testing (SI-DEP[3], the information system for testing). All individuals with symptoms that were compatible with COVID-19[19] were invited to consult their general

[1]INSERM, Sorbonne Université, Institut Pierre Louis d'Epidémiologie et de Santé Publique, IPLESP, Paris, France. [2]Orange Labs, Sociology and Economics of Network and Services (SENSE), Chatillon, France. [3]UFR des Sciences de la Santé Simone-Veil, Université Versailles-Saint-Quentin-en-Yvelines, Versailles, France. [4]AP-HP, Service de Médecine Interne, Hôpital Ambroise Paré, Boulogne Billancourt, France. [5]Santé publique France, Direction des maladies infectieuses, Saint-Maurice, France. [6]Tokyo Tech World Research Hub Initiative, Institute of Innovative Research, Tokyo Institute of Technology, Tokyo, Japan. [7]These authors contributed equally: Giulia Pullano, Laura Di Domenico. ✉e-mail: vittoria.colizza@inserm.fr

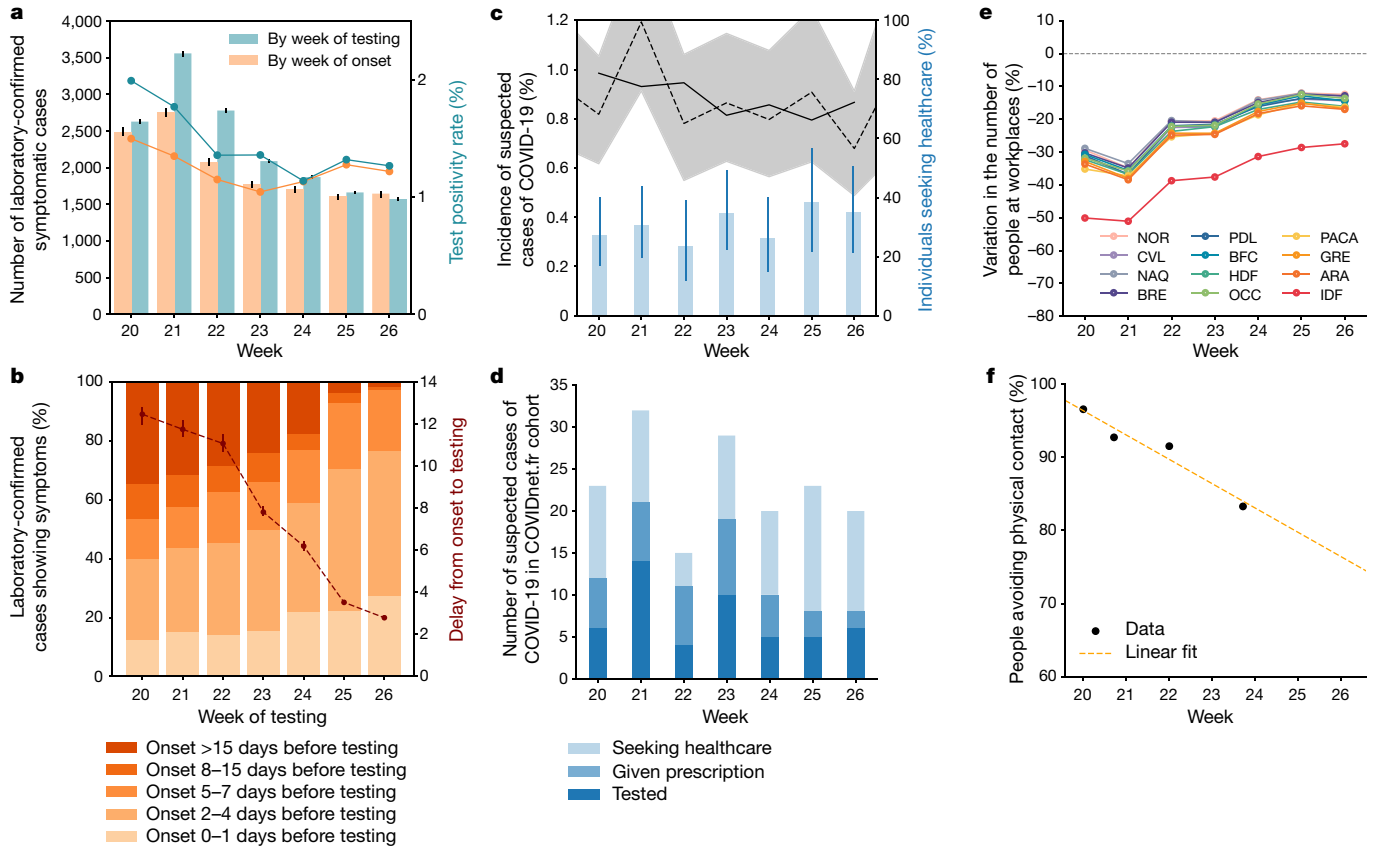


Fig. 1 | Virological surveillance, participatory syndromic surveillance and behavioural data for model parameterization. a, Estimated number of virologically confirmed symptomatic cases in mainland France by week of testing and week of onset (bar graphs), and test positivity rate (line graphs). Estimates are based on the imputation of individuals without symptoms who tested positive at the time of testing into asymptomatic or presymptomatic; imputation of missing data on clinical status at the time of testing into asymptomatic, presymptomatic or symptomatic; and imputation of the date of onset of symptoms for presymptomatic and symptomatic cases (Methods). Imputations were performed $n = 100$ times. Uncertainties (black bars) correspond to the 95% confidence intervals. Test positivity rates were computed for cases with complete information. Data for weeks 20–26 were consolidated in week 30. **b**, Breakdown of virologically confirmed cases with symptoms and complete information in the SI-DEP database by week of testing according to the declared onset of symptoms (left *y* axis; $n = 5{,}514$). The estimated time from onset to testing is also shown (right *y* axis; median and 95% confidence interval obtained from $n = 100$ imputations of the onset date). **c**, Weekly incidence of suspected cases of COVID-19 (median (dashed line), 95% confidence interval (shaded area) and 3-week moving average (solid line)), and percentage of individuals seeking healthcare (median and 95% confidence interval), estimated from the participatory surveillance system, COVIDnet.fr (average weekly $n = 7{,}481$). **d**, The number of suspected cases of COVID-19 in the participatory cohort who sought healthcare, and among those individuals, the number of individuals who received a prescription and performed a virological test when given the prescription. **e**, Estimated change in presence at workplace locations over time and by region based on Google location history data[17]. Region acronyms are listed in Table 1. **f**, Percentage of individuals avoiding physical contact with respect to lockdown, estimated from a large-scale survey conducted by Santé publique France[18].

practitioner and obtain a prescription for a virological test[8]. Contacts of confirmed cases were traced and tested. A total of 20,777 virologically confirmed cases were notified from 13 May (week 20) to 28 June (week 26) in mainland France. These cases included individuals with or without symptoms at the time of testing who tested positive for SARS-CoV-2 or individuals who tested positive for SARS-CoV-2 for whom information on clinical status at the time of testing was missing (Extended Data Fig. 1). Accounting for presymptomatic individuals among those presenting with no symptoms at the time of testing and after imputation of missing data (Methods), an estimated 16,165 (95% confidence interval, 16,101–16,261) symptomatic cases were tested in the study period (Fig. 1a). The average delay from symptom onset to testing decreased from 12.5 days in week 20 to 2.8 days in week 26 (Fig. 1b and Extended Data Fig. 1). Accounting for this delay (Methods and Extended Data Fig. 2), we estimated that 14,061 (13,972–14,156) virologically confirmed symptomatic cases had an onset of symptoms in the study period, showing a decreasing trend over time (2,493 in week 20, 1,647 in week 26). The test positivity rate decreased in the first weeks and stabilized at around 1.2% (mean over weeks 24–26).

A digital participatory system was additionally considered for COVID-19 syndromic surveillance in the general population[20], including those who did not consult a doctor. Called COVIDnet.fr, it was adapted from the platform GrippeNet.fr (which is dedicated to the surveillance of influenza-like illnesses[4]) to respond to the COVID-19 health crisis in early 2020. It is based on a set of volunteers who weekly self-declare their symptoms, along with sociodemographic information. On the basis of symptoms declared by an average of 7,500 participants each week, the estimated incidence of suspected cases of COVID-19[19] decreased from about 1% to 0.8% over time (Fig. 1c). Of 524 suspected cases, 162 (31%) consulted a doctor in the study period. Among them, 89 (55%) received a prescription for a test, resulting in the screening of 50 individuals (56% of those given the prescription) (Fig. 1d).

COVID-19 pandemic trajectories and detection rates

We used stochastic discrete age-stratified epidemic models[2,14] based on demography, age profile[21] and social contact data[15] of the 12 regions of mainland France to account for age-specific contact activity and role in

Table 1 | Population, confirmed and projected symptomatic cases, estimated detection rate and trends

Region	Acronym	Population (millions)	Number of laboratory-confirmed symptomatic cases by week of onset		Number of projected symptomatic cases by week of onset		Estimated detection rate (%) for symptomatic cases		Trend in detection rate
			Week 20	Week 26	Week 20	Week 26	Week 20	Week 26	
Île-de-France	IDF	12.3	737	574	12,427 (8,104–14,136)	1,704 (1,258–2,004)	6 (5–9)	34 (29–46)	+146%
Grand Est	GRE	5.5	323	135	4,868 (2,992–5,848)	756 (568–914)	7 (6–11)	18 (15–24)	+99%
Hauts de France	HDF	6.0	308	225	4,476 (2,381–6,648)	396 (219–538)	7 (5–13)	57 (42–100)	+186%
Auvergne-Rhône-Alpes	ARA	8.0	204	181	3,552 (2,017–5,283)	312 (173–451)	6 (4–10)	58 (40–100)	+244%
Occitanie	OCC	5.9	166	106	851 (397–1,400)	128 (57–235)	19 (12–42)	83 (45–100)	+165%
Provence–Alpes–Côte d'Azur	PACA	5.1	164	73	3,040 (1,665–4,625)	157 (83–239)	5 (4–10)	46 (31–88)	+289%
Pays de la Loire	PDL	3.8	127	96	1,158 (463–1,846)	255 (103–423)	11 (7–27)	38 (23–93)	+45%
Bourgogne–Franche–Comté	BFC	2.8	118	36	1,591 (854–2,379)	154 (88–235)	7 (5–14)	23 (15–40)	+95%
Nouvelle Aquitaine	NAQ	6.0	115	43	1,040 (482–1,691)	94 (38–166)	11 (7–24)	46 (26–100)	+54%
Centre-Val de Loire	CVL	2.6	94	44	1,706 (812–2,511)	79 (34–142)	6 (4–12)	56 (31–100)	+187%
Brittany	BRE	3.3	80	23	672 (294–1,155)	113 (51–206)	12 (7–27)	20 (11–45)	−28%
Normandy	NOR	3.3	55	112	725 (322–1,194)	153 (63–258)	8 (5–17)	73 (43–100)	+342%
France[a]		64.6	2,493	1,647	35,704 (30,290–40,748)	4,319 (3,773–4,760)	7 (6–8)	38 (35–44)	+142%

Regions are ranked by decreasing number of confirmed cases in week 20. The trend is computed comparing the average of the estimated detection rate in the weeks of June (weeks 23–26) with the average in the weeks of May (weeks 20–22). For the number of projected symptomatic cases and the estimated detection rate, data are medians and 95% confidence intervals obtained from $n = 500$ independent stochastic runs.

[a]Data are for mainland France; Corsica and overseas territories were excluded.

COVID-19 transmission. Disease progression is specific to COVID-19[2,14] and parameterized using the current knowledge to include presymptomatic transmission[22], and asymptomatic[12] and symptomatic infections with different degrees of severity[9,11,23,24]. The model was shown to capture the transmission dynamics of the pandemic in Île-de-France in the first wave and was used to assess the effect of lockdown and exit strategies[2,14]. Full details are reported in the Methods.

Intervention measures were modelled as mechanistic modifications of the contact matrices, accounting for a reduction in the number of contacts engaged in specific settings, and were informed from empirical data. Lockdown data were obtained from previously published studies[2,14]. The exit phase was modelled considering region-specific data of school attendance based on the data from the Ministry of Education[16], partial presence at workplaces based on estimates from location history data of mobile phones[17] (Fig. 1e), a reduction in the adoption of physical distancing over time and the increased risk aversion of older individuals based on survey data[18] (Fig. 1f), and the partial reopening of activities. A sensitivity analysis was performed on the reopening of activities, as data were missing for an accurate parameterization of associated contacts. Testing and isolation of detected cases were implemented by considering a 90% reduction in contacts for the virologically confirmed cases of COVID-19[2,14]. Region-specific models were fitted to regional hospital admission data (Fig. 2) using a maximum likelihood approach. Further details are reported in the Methods and Supplementary Information.

The projected number of cases decreased over time in all regions, in agreement with the decreasing tendency reported in hospital admissions during the study period (Fig. 2 and Extended Data Fig. 3). Overall, 103,907 (95% confidence interval, 90,216–116,377) new symptomatic infections were predicted in mainland France in weeks 20–26 (from 35,704 (30,290–40,748) in week 20 to 4,319 (3,773–4,760) in week 26). Île-de-France was the region with the largest predicted number of symptomatic cases (from 12,427 (8,104–14,136) to 1,704 (1,258–2,004) from week 20 to week 26), followed by Grand Est and Hauts-de-France (Table 1 and Extended Data Table 1).

Projections were substantially higher than the number of virologically confirmed cases (Figs. 2, 3). The estimated detection rate for symptomatic infections in mainland France in the period of weeks 20–26 was 14% (12–16%), suggesting that about 9 out of 10 new cases with symptoms were not identified by the surveillance system. A lower detection rate was found for asymptomatic infections (Extended Data Fig. 5). The estimated detection rate increased over time (7% (6–8%) in week 20, 38% (35–44%) in week 26) (Table 1). By the end of June, five regions had a median detection rate above 50%, and six regions had a detection rate within the confidence interval of model projections (Fig. 3b–d). All regions except Brittany displayed average increasing trends in the estimated detection rate in June compared with May. We did not find any significant associations between the detection rate and the number of detected cases, or the test positivity rate (Extended Data Fig. 4). However, the detection rate was negatively associated with model-predicted incidence (Spearman correlation, $r = -0.75$, $P < 10^{-15}$) (Fig. 3f). In addition, the data followed a power-law function, $\pi = 66 \times i^{-0.51}$, where π is the weekly detection rate of symptomatic cases (expressed as a percentage) and i the projected weekly incidence (number of cases per 100,000). This function quantifies the relationship between the detection capacity of the test–trace–isolate system and the circulation of the virus in the population. It clearly shows that the detection capacity rapidly decreases as the incidence of COVID-19 increases.

Validation of the model was performed in two ways. First, we compared our model projections of the percentage of the population infected with the results of three independent seroprevalence studies performed after the first wave in France[7,25,26] (Methods). Modelling results are in agreement with serological estimates at the national and regional level (Fig. 3e and Extended Data Fig. 6). Second, we compared the projected incidence of symptomatic cases of COVID-19 in week 26 (6.69 (5.84–7.37) cases per 100,000) with the value obtained from the number of virologically confirmed cases (2.55 (2.48–2.61) cases per 100,000) and two estimates based on COVIDnet.fr data (Fig. 3g). The first estimate applies the measured test positivity rate to the incidence of self-reported suspected cases of COVID-19 (estimate 1, which yielded 8.6 (95% confidence interval, 6.2–11.5) cases per 100,000); the second additionally assumes that only 55% would be confirmed as a suspected case by a physician and prescribed a test (according to COVIDnet.fr

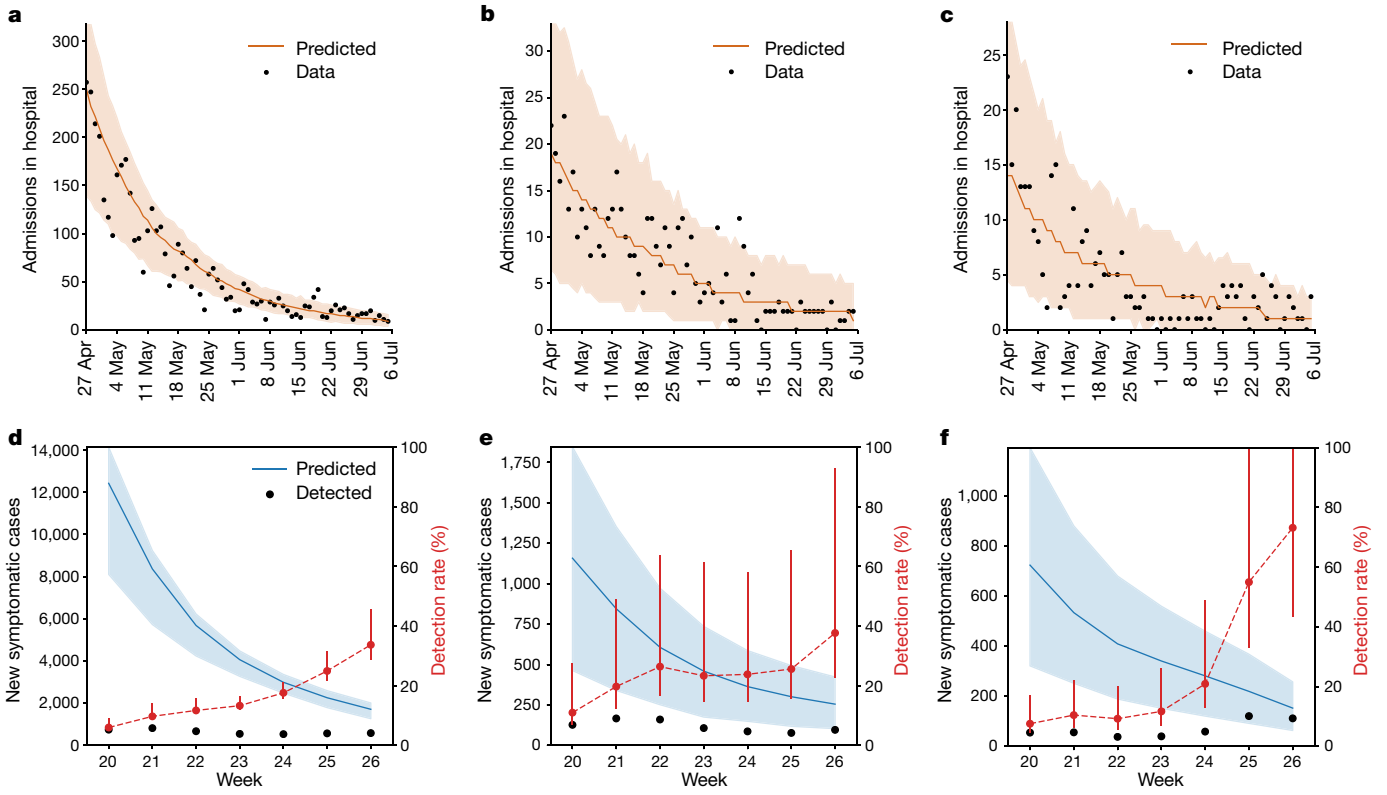


Fig. 2 | Hospital admissions and number of new symptomatic cases. **a**–**c**, Hospital admissions over time; data (points) and simulations (median and 95% confidence intervals) for Île-de-France (**a**), Pays de la Loire (**b**) and Normandy (**c**). Hospital admission data up to week 27 (consolidated in week 28) were used to infer parameter values. **d**–**f**, Projected number of new symptomatic cases over time (median and 95% confidence interval) and estimated number of virologically confirmed symptomatic cases by week of onset (points), for the same regions (Île-de-France (**d**), Pays de la Loire (**e**) and Normandy (**f**)) (left *y* axis). The estimated detection probability of symptomatic cases (%) is also shown (red points, median and 95% confidence interval; right *y* axis). In all panels, 95% confidence intervals were obtained from $n = 500$ independent stochastic runs. Plots for the remaining regions are shown in Extended Data Fig. 3.

data; estimate 2, which yielded 4.7 (3.4–6.3) cases per 100,000). Our projections are in line with plausible estimates from COVIDnet.fr, and suggest that, on average, at least 80% of suspected cases should be tested to reach the predicted incidence.

Sensitivity analysis showed that the findings were robust to elements of the contact matrices that could not be informed by empirical data (Supplementary Figs. 8, 9). Furthermore, a model selection analysis showed that changes in contact patterns over time due to restrictions and the activities of individuals of different age classes after lockdown (for example, partial attendance at school and remote working) are needed to accurately capture the transmission dynamics (Supplementary Table 2 and Supplementary Fig. 5).

Discussion

Despite a test positivity rate in mainland France well below the recommendations (5%) of the WHO[5], a substantial proportion of symptomatic cases (9 out of 10) remained undetected in the first 7 weeks after lockdown.

Low detection rates in mid-May were in line with estimates for the same period from a seroprevalence study in Switzerland[27]. Surveillance improved substantially over time, leading to half of the French regions reporting numbers of cases that were compatible with model projections. The framework progressively strengthened with increasing resources over time, as shown by a more-rapid detection of cases (78% reduction in the average delay from symptom onset to testing from May to June). At the same time, the system benefited from a substantial and concurrent decrease in epidemic activity in all regions.

Despite this positive trend, our findings highlight structural limitations and a critical need for improvement. Some areas remained with limited diagnostic exhaustiveness. This is particularly concerning in those regions that were predicted to have large numbers of weekly infections (Île-de-France, in which only one out of three symptomatic cases was detected by the end of June, and Grand Est, in which one out of five was detected). Almost all patients (92%) who were clinically diagnosed by sentinel general practitioners as suspected cases of COVID-19 were prescribed a test[20]. However, only 31% of individuals with COVID-19-like symptoms consulted a doctor according to participatory surveillance data. Overall, these figures suggest that a large number of symptomatic cases of COVID-19 were not screened because they did not seek medical advice despite the recommendations. This was confirmed by serological studies. In France, only 48% of symptomatic participants with antibodies against SARS-CoV-2 reported consulting a general practitioner[7]; in Spain, between 16% and 20% of individuals with antibodies against SARS-CoV-2 reported a previous virological screening[6]. By combining estimates from virological and participatory surveillance data, we extrapolated an incidence rate from crowdsourced data that is compatible with model projections, under the hypothesis that the large majority of suspected cases would get tested (>80%). This finding further supports testing of all suspected cases of COVID-19. Large-scale communication campaigns should reinforce recommendations to raise awareness in the population and strongly encourage healthcare-seeking behaviour especially in patients with mild symptoms. At the same time, investigations to identify reasons for not consulting a doctor could be quickly performed through the participatory surveillance system.

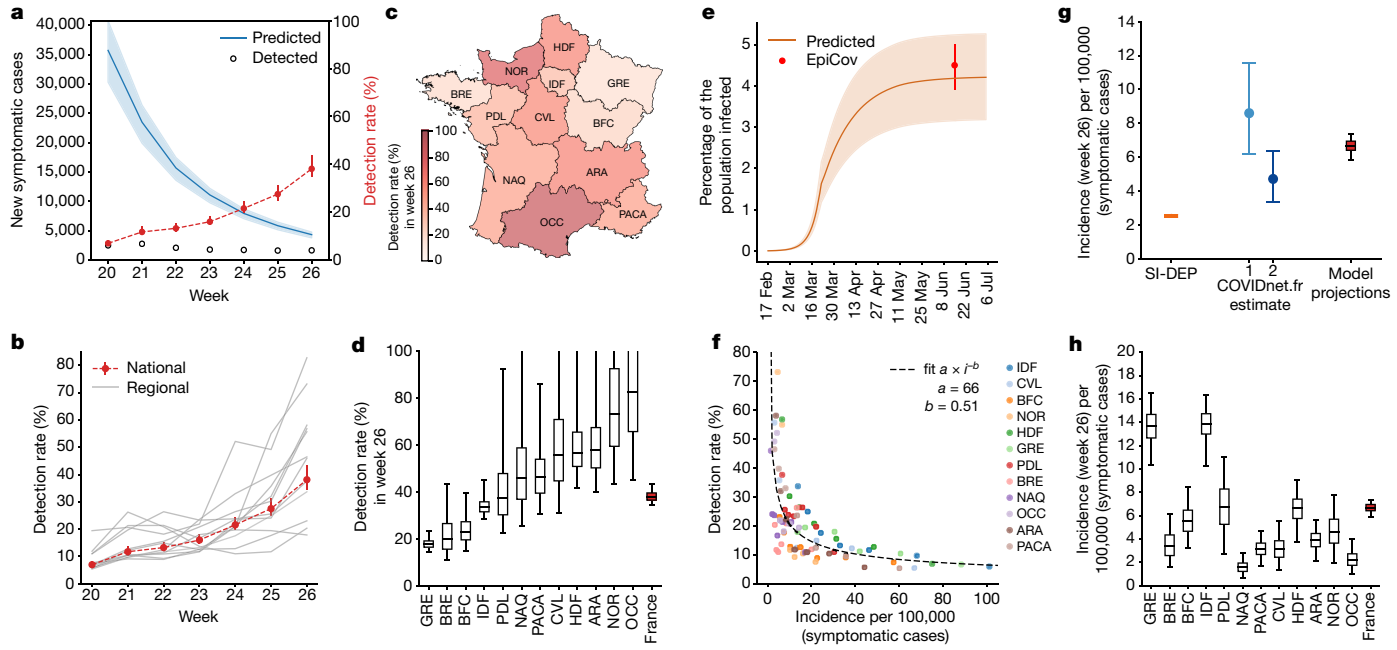


Fig. 3 | Detection rate and incidence. a, Projected number of new symptomatic cases over time (median and 95% confidence interval) and estimated number of virologically confirmed symptomatic cases by week of onset (points) in mainland France (left *y* axis). The estimated detection rate of symptomatic cases (%) is also shown (red points, median and 95% confidence interval; right *y* axis). **b**, Estimated detection rate of symptomatic cases (%) and 95% confidence intervals over time for mainland France (red dots and bars), and for all regions (grey lines, only median values are shown for visualization). **c**, Map of the estimated detection rate (%) by region in week 26 (22–28 June 2020). **d**, Estimated detection rate per region compared to the national estimate. Regions are ranked by increasing median detection rate. Box plots represent the median (line in the middle of the box), interquartile range (box limits) and 2.5th and 97.5th percentiles (whiskers). **e**, Predicted percentage of the population infected (median and 95% confidence interval) compared with estimates from the serological study EpiCov[26] performed on a representative sample of the population in mainland France. **f**, Estimated detection rate of symptomatic cases (%) by region and by week compared with the projected incidence by region and by week. The curve shows the result of a least-square fit to the data with a power-law function, $\pi = a \times i^{-b}$, where π is the detection rate (expressed as a percentage), i is the weekly incidence (cases per 100,000), $a = 66$ (95% confidence interval, 52–85) and $b = 0.51$ (0.41–0.60). **g**, Estimated incidence of symptomatic cases and 95% confidence intervals in mainland France in week 26 from different sources: virological surveillance data (SI-DEP), participatory surveillance data (COVIDnet.fr, with two estimates) and model projections. **h**, Projected incidence per region compared to the national estimate. Regions are ranked as in **d**. Box plots are as defined in **d**. In all panels, medians and 95% confidence intervals for model projections were obtained from $n = 500$ independent stochastic runs.

Red tape might have contributed to low testing rates. Prescription of a test was deemed compulsory in the new testing policy to prevent misuse of diagnostic resources[8]; however, this involved consultation, prescription and a laboratory appointment, which may have discouraged mildly affected individuals who do not require medical assistance. To facilitate access, testing should not require a prescription, as later established by authorities[28]. Some local initiatives emerged over summer that increased the number of drive-through testing facilities, promoted massive screening in certain areas and offered mobile testing facilities to increase proximity to the population[29]. The use of antigen tests will further facilitate access. These initiatives are particularly relevant to counteract socioeconomic inequalities in access to care in populations that are vulnerable to COVID-19[30,31]. However, such strategies should not hinder a testing protocol that targets suspected index cases. Our results show that high testing efforts, measured by low test positivity rates, are not associated with high rates of detection. This was also observed in the UK during the first wave, when detection remained low despite large numbers of tests and a low positivity rate[32]. Without strong case-based surveillance, the risk is to disperse resources towards random individuals without symptoms who are unlikely to be positive. This could saturate the test–trace–isolate system, as observed during summer[33], without slowing down the circulation of SARS-CoV-2 that is required to safeguard the hospital system.

Given presymptomatic transmission, notification of contacts should be almost immediate to enable the effective interruption of transmission chains[22]. For testing to be an actionable tool to control the transmission of SARS-CoV-2, delays should be suppressed and screening rates greatly increased but better targeted. Over May–June, the average weekly number of tests was 250,000—remaining well below the objective that was originally set by authorities (700,000 tests). The number of tests increased over summer, but proportionally to the increased circulation of the virus. The capacity of detection of the test–trace–isolate system scaled as the inverse of the square root of the incidence, already deteriorating rapidly at low incidence levels. More aggressive testing that targets suspected index cases should be performed at low viral circulation to avoid case resurgence. The system was predicted to be able to detect more than two out of three cases (rate >66%) only if the incidence was lower than one symptomatic case per 100,000, a figure that is 50 times smaller than estimated at the exit from lockdown. As detection of at least 50% of cases is needed to control the pandemic while avoiding strict social distancing[2], these results indicate that the system was insufficient to perform comprehensive case-based surveillance, as has been recommended when aiming to phase out restrictions[5]. Current restrictions applied in Europe to curb the second wave offer a second opportunity to improve testing policies and support the lifting of these measures in the upcoming weeks. Failing to do so may lead to a rapid and uncontrolled increase in the number of cases of COVID-19[2,34]. Such risk is even stronger in the winter season and with the existing fatigue with regard to adhesion to the restrictions[18].

Models were region-based and did not consider a possible coupling between regional epidemics caused by mobility. This choice

was supported by stringent movement restrictions during lockdown[30], and by the limited mobility increase in May–June, before important inter-regional displacements took place at the start of the summer holidays in July. Foreign importations of the virus[35] were neglected as France reopened its borders with EU member states on 15 June, and the Schengen area remained closed until July. The COVIDnet.fr cohort is not representative of the general population; however, a previous study on influenza-like illnesses has shown that the adjusted incidence was in good agreement with sentinel estimates[4]. Underdetection may also continue because of the imperfect characteristics of the reverse-transcription PCR tests used to identify infections of SARS-CoV-2[36]. Some cases tested for SARS-CoV-2 could have had false-negative results, for example, because they were tested too early after the infection, thus further increasing the rate of underdetection. Previous work assessed the rate of underdetection in 210 countries[32], but this study mainly focused on the early global dynamics. Our model gives up geographical extent for higher data quality in a specific country, providing a synthesis of data sources that characterizes human behaviour over time and space together with virological and participatory surveillance data to identify the weak links in the pandemic response.

Our findings identify critical needs for the improvement of the test–trace–isolate response system to control the COVID-19 pandemic. Substantially more aggressive and efficient testing that targets suspected cases of COVID-19 needs to be achieved to act as a way to control the COVID-19 pandemic. Associated communication and logistical needs should not be underestimated. These elements should be considered to enable the lifting of restrictive measures that are currently used to curb the second wave of COVID-19 in Europe.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-03095-6.

1. OECD. Testing for COVID-19: a way to lift confinement restrictions. https://www.oecd.org/coronavirus/policy-responses/testing-for-covid-19-a-way-to-lift-confinement-restrictions-89756248/ (2020).
2. Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. *BMC Med.* **18**, 240 (2020).
3. Data.gouv.fr. Données relatives aux résultats des tests virologiques COVID-19 (SI-DEP). https://www.data.gouv.fr/en/datasets/donnees-relatives-aux-resultats-des-tests-virologiques-covid-19/ (2020).
4. Guerrisi, C. et al. The potential value of crowdsourced surveillance systems in supplementing sentinel influenza networks: the case of France. *Euro Surveill.* **23**, 1700337 (2018).
5. WHO. Considerations for implementing and adjusting public health and social measures in the context of COVID-19. https://www.who.int/publications/i/item/considerations-in-adjusting-public-health-and-social-measures-in-the-context-of-covid-19-interim-guidance (2020).
6. Pollán, M. et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* **396**, 535–544 (2020).
7. Carrat, F. et al. Seroprevalence of SARS-CoV-2 among adults in three regions of France following the lockdown and associated risk factors: a multicohort study. Preprint at https://doi.org/10.1101/2020.09.16.20195693 (2020).
8. Ministère des Solidarités et de la Santé. Prise en charge par les médicins de ville des patients atteints de COVID-19 en phase de déconfinement. https://solidarites-sante.gouv.fr/IMG/pdf/prise-en-charge-medecine-ville-covid-19.pdf (2020).
9. Verity, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).
10. Pinotti, F. et al. Tracing and analysis of 288 early SARS-CoV-2 infections outside China: a modeling study. *PLoS Med.* **17**, e1003193 (2020).
11. Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
12. Lavezzo, E. et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* **584**, 425–429 (2020).
13. Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151 (2004).
14. Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Modelling safe protocols for reopening schools during the COVID-19 pandemic in France. *Nat. Commun.* (in the press).
15. Béraud, G. et al. The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. *PLoS ONE* **10**, e0133203 (2015).
16. Ministère de l'Education Nationale de la Jeunesse et des Sports. Déconfinement phase 2: point de situation au 28 mai. https://www.education.gouv.fr/deconfinement-phase-2-point-de-situation-au-28-mai-303813 (2020).
17. Google. Community Mobility Report - COVID-19. https://www.google.com/covid19/mobility?hl=fr (2020).
18. Santé publique France. COVID-19: une enquête pour suivre l'évolution des comportements et de la santé mentale pendant l'épidémie. https://www.santepubliquefrance.fr/etudes-et-enquetes/covid-19-une-enquete-pour-suivre-l-evolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie (2020).
19. Haut Conseil de la Santé Publique. Signes cliniques d'orientation diagnostique du Covid-19. https://www.hcsp.fr/explore.cgi/avisrapportsdomaine?clefr=812 (2020).
20. Santé publique France. COVID-19: point épidémiologique du 25 juin 2020. https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-25-juin-2020 (2020).
21. INSEE. Pyramide des âges 2020 - France et France métropolitaine. https://www.insee.fr/en/statistiques/3696316 (2020).
22. Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
23. Riccardo, F. et al. Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020. *Eurosurveillance* **25**, 2000790 (2020).
24. Salje, H. et al. Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020).
25. Vu, S. L. et al. Prevalence of SARS-CoV-2 antibodies in France: results from nationwide serological surveillance. Preprint at https://doi.org/10.1101/2020.10.20.20213116 (2020).
26. Ministère des Solidarités et de la Santé. En mai 2020, 4,5% de la populationen France métropolitaine a développé des anticorps contre le SARS-CoV-2. Premiers résultats de l'enquête nationale EpiCov. https://drees.solidarites-sante.gouv.fr/IMG/pdf/er1167.pdf (2020).
27. Stringhini, S. et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet* **396**, 313–319 (2020).
28. Ameli. Covid-19: les tests de dépistages sont possibles sans prescription médicale. https://www.ameli.fr/assure/actualites/covid-19-les-tests-de-depistages-sont-possibles-sans-prescription-medicale (2020).
29. Covid-19: une campagne de tests à grande échelle débute dans 32 communes d'Ile-de-France. *Le Parisien* (29 June 2020).
30. Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S. & Colizza, V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. *Lancet Digit. Health* **2**, e638–e649 (2020).
31. Garnier, R., Benetka, J. R., Kraemer, J. & Bansal, S. Socio-economic disparities in social distancing during the COVID-19 pandemic in the United States. Preprint at https://doi.org/10.1101/2020.11.07.20201335 (2020).
32. Russell, T. W. et al. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med.* **18**, 332 (2020).
33. Explosion des délais d'attente des résultats aux tests Covid-19. *Le Figaro* (22 September 2020).
34. Aleta, A. et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **4**, 964–971 (2020).
35. Ruktanonchai, N. W. et al. Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science* **369**, 1465–1470 (2020).
36. Wernike, K. et al. Pitfalls in SARS-CoV-2 PCR diagnostics. *Transbound. Emerg. Dis.* https://doi.org/10.1111/tbed.13684 (2020).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Virological surveillance data

The centralized database SI-DEP for virological surveillance[3] collects all tests performed in France for any reason. In the period under study, guidelines recommended individuals to consult a general practitioner at the first sign of COVID-19-like symptoms and to obtain a prescription for a virological test (a prescription was compulsory to access the test)[8]. In addition, routine testing was performed for patients admitted to the hospital with any diagnosis, healthcare personnel and individuals at other facilities (for example, in some care homes for older people or long-term healthcare facilities). Data include detailed information for the individuals tested in France, including (1) the date of the test; (2) the result of the test (positive or negative); (3) location (region); (4) the absence or presence of symptoms at the time of testing; (5) self-declared delay between onset and test in presence of symptoms. The delay is provided with the following breakdown: onset date occurring 0–1 day before date of test, 2–4 days before, 5–7 days before, 8–15 days before, or more than 15 days before. For some tests, information on points (4) and (5) is missing. The SI-DEP database provided complete information for 23,210 (66%) out of 35,264 laboratory-confirmed cases of COVID-19 tested between week 20 (11–17 May) and week 30 (19–26 July), with an increasing trend of complete information over time (from 49% in week 20 to 76% in week 30) (Extended Data Fig. 1). Among confirmed cases with complete information, 12,716 (55%) showed no symptoms at the time of testing (Extended Data Fig. 1). The study referred to the period from week 20 to week 26. Data up to week 30 were used to consolidate the data in the study period accounting for the delays.

Imputation of asymptomatic versus presymptomatic cases, onset date and missing information

Individuals who tested positive on a given date were recorded in the SI-DEP database as: cases with symptoms at the time of testing, with a self-declared delay from onset of symptoms; cases without symptoms at the time of testing; or cases with no information on presence or absence of symptoms at the time of testing. These three subsets of cases were analysed to account for the presence of presymptomatic individuals among those with no symptoms at the time of testing, imputation of missing data and the estimation of dates of infection or symptom onset.

For laboratory-confirmed cases of COVID-19 who had symptoms at the time of testing, we estimated their date of onset using the information on the date of test and the time interval of onset-to-test delay, which was self-declared by the patients (Fig. 1b). In the time period between weeks 20 and 30, 20% of cases had an onset-to-test delay of ≤1 day, 63% had a delay of ≤4 days, 83% had a delay of ≤7 days and 88% had a delay of ≤15 days (Extended Data Fig. 1). We fitted a Gamma distribution to the onset-to-test delay data with a maximum likelihood approach, using three different periods of time (May, June and July), to account for changes in the distribution of self-declared delays over time (that is, longer delays at the beginning of the study period, shorter delays at its end) (Extended Data Fig. 2). The estimated average delay in May, June and July was 12.9 (95% confidence interval, 7.0–16.1), 5.1 (3.7–6.3) and 2.7 (2.0–3.1) days, respectively. July data were used to consolidate data corresponding to infections with onset in June and tested with delay. Given a confirmed case with symptoms testing on a specific date, we assigned the onset date by sampling the onset-to-testing delay from the fitted distribution for that period, conditional to the fact that the delay lies in the corresponding time interval declared by the patient. We assumed that onset did not occur before the implementation of the national lockdown, on 17 March 2020 (week 12); we therefore truncated the Gamma distribution accordingly, when assigning the date of onset for cases with onset-to-test delay >15 days. The imputation procedure was carried out 100 times. Results were aggregated by week of onset.

For laboratory-confirmed cases of COVID-19 with no symptoms at the time of testing, we assumed that on average 40% of them were asymptomatic[12] (see the 'Transmission model summary' section), whereas the remaining 60% were presymptomatic who tested early thanks to contact tracing. Imputation was done by sampling from a binomial distribution and repeated 100 times. Data on contact tracing could not be used to inform data on infection or symptom onset, because of national regulatory framework on privacy preventing the matching of the two databases (virological tests and contact tracing). Given the low sensitivity of PCR tests in the early phase of the incubation period, we considered that imputed presymptomatic cases belonged to the prodromic phase. Onset date for presymptomatic cases was estimated by sampling from an exponential distribution with a mean of 1.5 days, corresponding to the duration of the prodromic phase in our model (Supplementary Table 1). For imputed asymptomatic, we assumed the same delay from infection to testing as in cases with symptoms. Given the structure of our compartmental model and to match the definition of the time used for symptomatic individuals (week of onset), we considered a delay in the detection of asymptomatic individuals starting from the end of the prodromic phase (corresponding to the symptom onset time for symptomatic infections) to the date of testing. We assigned this date by sampling the delay from the monthly gamma distribution. Imputation of the dates was repeated 100 times.

For laboratory-confirmed cases of COVID-19 with no information on symptoms at the time of testing, missing data were imputed by sampling from a multinomial distribution with probabilities equal to the rate of occurrence of the outcomes (asymptomatic, presymptomatic or symptomatic with five possible time intervals for the onset-to-test delay) reported for cases with complete information and assuming the imputation of cases without symptoms into asymptomatic and presymptomatic, as described above. Imputation was performed by region and by week and repeated 100 times. Presymptomatic and symptomatic individuals were aggregated together by onset date (Fig. 1a) to estimate the rate of detection of symptomatic cases.

Participatory surveillance data and analysis

COVIDnet.fr is a participatory online system for the surveillance of COVID-19, available at https://www.covidnet.fr/. It was adapted from GrippeNet.fr[4] to respond to the COVID-19 health crisis in March 2020. GrippeNet.fr is a participatory system for the surveillance of influenza-like illnesses available in France since 2011 through a collaboration between Inserm, Sorbonne Université and Santé publique France, supplementing sentinel surveillance[4,37]. The system is based on a dedicated website to conduct syndromic surveillance through self-reported symptoms volunteered by participants resident in France. Data are collected on a weekly basis; participants also provide detailed profile information at enrolment[38]. In addition to tracking the incidence of influenza-like illnesses[4,37], GrippeNet.fr was used to estimate vaccine coverage in specific subgroups[39], individual perceptions towards vaccination[40] and healthcare-seeking behaviour[41]. It was also used to assess behaviours and perceptions related to diseases other than influenza[42], including COVID-19[43].

Participants are on average older and include a larger proportion of women compared to the general population[38,44]. The participating population is, however, representative in terms of health indicators such as diabetes and asthma conditions. Despite these discrepancies, trends of the estimated incidence of influenza-like illnesses from GrippeNet.fr reports compared well with those of the national sentinel system[4,37]. All analyses were adjusted by age and sex of participants.

To monitor suspected cases of COVID-19 in the general population, we used the expanded case definition recommended by the High Council of Public Health for systematic testing and described in their

20 April 2020 notice[19], which included either of the two following definitions: (1) (sudden onset of symptoms OR sudden onset of fever) AND (fever OR chills) AND (cough OR shortness of breath OR (chest pain AND age > 5 years old)) or (2) (sudden onset of symptoms) OR (sudden onset of fever AND fever); and one of the three following conditions: (i) (age > 5 years old) AND ((feeling tired or exhausted) OR (muscle/joint pain) OR (headache) OR (loss of smell WITHOUT runny or blocked nose) OR (loss of taste)); or (ii) ((age ≥ 80 years old) OR (age < 18 years old)) AND (diarrhoea); or (iii) (age < 3 months old) AND (fever WITHOUT other symptoms).

Two independent estimates obtained from COVIDnet.fr cohort data for the incidence of symptomatic cases in week 26 are shown in Fig. 3. These estimates were computed as follows. Estimate 1 = (COVIDnet.fr estimated incidence of suspected cases in week 26) × (test positivity rate from SI-DEP in week 26); estimate 2 = (COVIDnet.fr estimated incidence of suspected cases in week 26) × (estimated proportion screened and confirmed as a suspected case of COVID-19 by a physician, and prescribed a test; estimates from COVIDnet.fr) × (test positivity rate from SI-DEP in week 26). The two estimates were used to validate model projections and identify the specific surveillance mechanisms that needed improvement.

Ethics statement

GrippeNet.fr/COVIDnet.fr was reviewed and approved by the French Advisory Committee for research on information treatment in the health sector (that is, CCTIRS, authorization 11.565), and by the French National Commission on Informatics and Liberty (that is, CNIL, authorization DR-2012–024)—the authorities ruling on all matters related to ethics, data and privacy in the country. Informed consent was provided by each participant at enrolment, according to regulations.

Transmission models summary

We used a stochastic discrete age-stratified transmission model for each region based on demographic, contact[15] and age profile data of French regions[21]. Models were region-specific to account for the geographically heterogeneous epidemic situation in the country and given the mobility restrictions limiting inter-regional movement fluxes. The study focused on mainland France where the epidemic situation was comparable across regions, and excluded Corsica, which reported very limited epidemic activity and overseas territories characterized by increasing transmission[20].

Four age classes were considered: [0–11), [11–19), [19–65) and 65+ years old, referred to as children, adolescents, adults and older individuals. Transmission dynamics follows a compartmental scheme specific to COVID-19, in which individuals were divided into susceptible, exposed, infectious and hospitalized (Supplementary Information and Supplementary Figs. 1, 2). We did not consider further progression from hospitalization (for example, admission to intensive care units, recovery or death[2]) as it was not needed for the objective of the study. The infectious phase is divided into two steps: a prodromic phase (I_p) and a phase during which individuals may remain either asymptomatic (I_a, with probability[12] $p_a = 40\%$) or develop symptoms. In the latter case, we distinguished between different degrees of severity of symptoms[9,11,23,24], ranging from paucisymptomatic (I_{ps}), to infectious individuals with mild (I_{ms}) or severe (I_{ss}) symptoms. Prodromic, asymptomatic and paucisymptomatic individuals have a reduced transmissibility $r_\beta = 0.55$, as estimated previously[11], and in agreement with evidence from the field[45–47]. A reduced susceptibility was considered for children and adolescents, along with a reduced relative transmissibility of children, following available evidence from household studies, contact-tracing analyses, serological investigations and modelling works[48–53]. A sensitivity analysis was performed on the relative susceptibility and transmissibility of children, and on the proportion of asymptomatic infections (Supplementary Figs. 10–13). Full details are reported in the Supplementary Information.

The study was not extended to the summer months, because of (1) the challenge of mechanistically parameterizing the contact matrices during summer; (2) the increase of movement fluxes across regions weakening our assumption of region-specific models; and (3) the interruption of COVIDnet.fr surveillance during the summer break, which prevented the identification of the key factors behind case underascertainment.

Contact matrices

Age-stratified transmission uses a social contact matrix that reports the average contact rates between different age classes in France[15]. This refers to the baseline condition, that is, before lockdown. The contact matrix includes the following layers: contacts at home, school, workplace, transport, leisure activities and other activities, and discriminates between physical and non-physical contacts. To account for the change of contact patterns over time, contact matrices are mechanistically parameterized, by region and over time, with different data sources informing on the percentage of students going to school[16], the percentage of workers going to the workplace[17], the compliance to preventive measures[18], with a higher compliance registered in older individuals[18]. Information on the progressive reopening of activities indicates that leisure and other activities were only partially open in the study period. Data, however, are not fine-grained enough to parameterize our model, so we assume a 50% opening of these activities and explore variations in the sensitivity analysis.

School attendance. School reopening was parameterized by considering the percentage of reported attendance at school (pre-school and primary school; middle and high school) provided by the Ministry of Education[16] (Supplementary Fig. 3). The number of contacts in the school matrix was modified to account for the attendance of students in each school level provided by data. That is, attendance of 14.5%, referring, for example, to the attendance registered in Île-de-France in pre-schools and primary schools, corresponds to a reduction of 85.5% in the number of contacts established at school by students belonging to that school level. Contacts for different modes of transport were modified accordingly.

Presence at work. To account for the percentage of individuals at work, given recommendations on remote working and activities that were not yet reopened, we used the estimated variation of presence at workplaces based on mobile phone location data provided by Google Mobility Trends[17]. Contacts at work and for different modes of transport were therefore modified according to this percentage, as described for contacts at school. Household contacts were increased proportionally to each adult staying at home based on statistics comparing weekend versus weekday contacts[15] and the proportion of adults working during the weekend[54], as done previously[2].

Adoption of physical distancing. Our previous work showed that physical contacts during lockdown were fully avoided[2], in agreement with data collected afterwards[18]. To account for individual adoption of preventive behaviour after lockdown, we used the percentage of population avoiding physical contacts estimated from a large-scale survey conducted by Santé publique France (CoviPrev[18]). Data were fitted with a linear regression (Fig. 1) to provide the weekly percentage of individuals avoiding physical contacts. We therefore modified our contact matrices over time, removing the percentage of physical contacts corresponding to the survey estimates for that week.

Increased risk aversion of older individuals. Data from the Santé publique France survey CoviPrev[18] also show that older individuals protected themselves further relative to other age classes. On average, they respected physical distancing 28% more than the other age classes (Supplementary Fig. 4). For this reason, we considered a further

reduction of 30% in contacts for older individuals in the exit phase, informed by survey data.

Inference framework

The parameters of the transmission models to be estimated are specific to each pandemic phase.

Before lockdown, $\{\beta, t_0\}$, where β is the transmission rate per contact and t_0 is the date of the start of the simulation, seeded with 10 infectious individuals.

During lockdown, $\{\alpha_{LD}, t_{LD}\}$, where α_{LD} is the scaling factor of the transmission rate per contact and t_{LD} is the date when lockdown effects on hospitalization data became visible.

After lockdown, $\{\alpha_{exit}, \pi_a(w), \pi_s(w)\}$, where α_{exit} is the scaling factor of the transmission rate per contact, and $\pi_a(w)$ and $\pi_s(w)$ are the proportion of asymptomatic and symptomatic cases tested in week w of the exit phase, respectively. Detected cases in the simulations had their contacts reduced by 90% to mimic isolation, as done in previous studies[2,14].

We used simulations of the stochastic model to predict values for all quantities of interest (500 simulations each time). We fitted the model to the daily count of hospitalizations $H_{obs}(d)$ on day d throughout the period and the number of people testing positive by week of onset, split according to disease status (symptomatic or asymptomatic), denoted $\text{Test}_{s,obs}(w)$ and $\text{Test}_{a,obs}(w)$ in week w of the exit phase. We used hospital admission data up to week 27 (29 June–5 July) to account for the average delay from infection to hospitalization. Data in week 27 were consolidated by waiting for one additional week to account for updates and missing data (week 28, 6–12 July 2020).

We assumed a Poisson distribution for hospitalizations and a binomial distribution for the number of people getting the test, therefore the likelihood function is

$$L(\text{Data}|\Theta) = \prod_{d=t_o}^{t_n} P_{\text{Poisson}}(H_{obs}(d); H_{pred}(d), \beta, t_0, \alpha_{LD}, t_{LD}, \alpha_{exit}, \pi_a(w_d), \pi_s(w_d)) \times \prod_{w \in \text{exit}} P_{\text{Binomial}}(\text{Test}_{s,obs}(w); i_{s,pred}(w), \pi_s(w)) \times \prod_{w \in \text{exit}} P_{\text{Binomial}}(\text{Test}_{a,obs}(w); i_{a,pred}(w), \pi_a(w))$$

where $\Theta = \{\beta, t_0, \alpha_{LD}, t_{LD}, \alpha_{exit}, \{\pi_a(w)\}, \{\pi_s(w)\}\}$ indicates the set of parameters to be estimated, $H_{pred}(d)$ is the model-predicted number of hospital admissions on day d, $i_{s,pred}(w)$ and $i_{a,pred}(w)$ are the model-predicted weekly incidences of symptomatic and asymptomatic cases, respectively, in week w of the exit phase, $P_{Poisson}$ is the probability mass function of a Poisson distribution, $P_{Binomial}$ for a binomial distribution, $[t_0, t_n]$ is the time window considered for the fit, and w is the week in the exit phase (weeks 20–26).

We reduced the required computations with an optimization procedure in two steps, first maximizing the likelihood function in the pre-lockdown and lockdown phase to estimate the first four parameters, and then maximizing the likelihood in the exit phase by fixing the first four parameters that describe the epidemic trajectory before the exit phase to their maximum likelihood estimators (MLEs). This second step was further simplified through an iterative procedure, and we show through simulations that the simplified optimization procedure is consistent and well-defined. The parameter space was explored using NOMAD software[55]. Fisher's information matrix was estimated at the MLE value to obtain the corresponding confidence intervals. Simulations were then parameterized with 500 parameter sets obtained from the joint distribution of transmission parameters at MLE (one stochastic simulation for each parameter set). A Bayesian estimate of the posterior parameter distribution using Markov chain Monte Carlo (MCMC) would also have been an alternative to maximum likelihood and confidence interval estimation. In this case, however, MCMC would have considerably slowed down parameter exploration, with negligible added value to the fitting procedure.

We repeated model fitting starting from several starting points and using different random number streams. Values of fitted parameters and full details on the different steps and the tests performed are reported in the Supplementary Information (Supplementary Figs. 6, 7 and Supplementary Table 3).

Simulation details

Simulations are initialized with 10 infected adults in the I_p compartment at time t_0. We obtained 500 parameter sets from the joint distribution of transmission parameters at MLE and ran one stochastic simulation for each parameter set. Therefore, errors in the detection rates computed in the output account for the variability of the estimate of the parameters, in addition to the stochastic fluctuations of the model. We find that the errors in the estimation of the detection rates obtained including the variability of the parameters are slightly larger than the ones obtained with only stochastic fluctuation, suggesting that the stochasticity of the model is the main source of error in the estimation of the detection rate.

Model selection analysis

To assess the role of the mechanistic modification of the contact matrix informed by the different data sources in the exit phase, we compared our model to a simplified version assuming that contact patterns in the exit phase do not change from pre-epidemic conditions, and that all changes in the epidemic trajectory are explained exclusively by the transmissibility per contact. This is equivalent to normalizing the contact matrix to its largest eigenvalue and estimating the reproductive ratio over time. We compared the two models with the Akaike information criterion and found that accounting for changes in contacts better describes the epidemic trajectory (Supplementary Table 2 and Supplementary Fig. 5).

Comparison with serological estimates

We compared model projections with serological estimates from three independent studies[7,25,26] (Fig. 3e and Extended Data Fig. 6).

Estimates by Carrat et al.[7] used ELISA-S tests and ELISA-NP tests. The sample was not representative of the population, and estimates were weighted to account for this bias. In the comparisons, we used the results from a multiple imputation method performed by the authors and estimating a participant's positivity with a likelihood of positivity based on observed test results and covariates (see ref. [7] for more details).

Estimates by Santé publique France[25] are based on at least one positive result in one of the following three tests: ELISA-S, ELISA-NP and a pseudo-neutralization test that detects the presence of pseudo-neutralizing antibodies, representative of the presence of neutralizing antibodies as conferring protection against infection. Analyses were performed on residual sera obtained from clinical laboratories, and estimates were weighted to account for the lack of representativeness.

Estimates by EpiCoV[26] (Enquête Epidémiologie et Conditions de vie liées à la Covid-19) used ELISA-S tests and further validated these with a seroneutralizing antibody test at higher specificity (see ref. [26] for more details). This was the only seroprevalence survey that was conducted in a representative sample of the population. For this reason, we used it as the reference study.

For all studies, we report in Fig. 3e and Extended Data Fig. 6 the estimates 14 days before the last blood collection to account for the time needed to mount a detectable presence of antibodies. For the EpiCoV survey, we used the last date at which samples were sent back to the laboratory.

Modelling results are in good agreement with the serological estimates at the national level (Fig. 3e) and in the large majority of the regions (Extended Data Fig. 6). Projections tend to be systematically smaller than serological estimates in two regions that were weakly affected by the epidemic (Pays de la Loire and Brittany), although they remained compatible with observations.

Overall differences may be due to the limitations of the methods involved. First, the type of tests, the specificity levels, the samples of the population tested, and the weighting and imputation approaches considered in each serological study could lead to differences across the three investigations. We note, for example, that larger discrepancies are observed between EpiCov and Santé publique France results in those regions that experienced smaller epidemics. We used EpiCov as the reference study as it was the only one study that was conducted on a representative sample of the population. Second, there are limitations to the dataset of hospital admissions used to calibrate the models: the database infrastructure for data collection became operational in mid-March and was filled in retrospectively. Notification biases would inevitably alter the inference of parameters in the pre-lockdown phase. This may have differed region by region; however, we have no way to control for this potential bias; possible errors would have affected regions with small-size epidemics more than others. In support of this hypothesis, we note that a similar but independent mathematical model fitted to regional hospitalization data[24] in the first wave predicted small epidemics in Pays de la Loire and Brittany, similarly to our model.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data used for model parameterization (contact matrices[15], school attendance[16], presence at workplaces[17] and avoidance of physical contacts[18]) are available at the cited sources. Hospitalization data used for model calibration are made available with the code. Virological data[3] are available at the cited source; data on onset-to-test delay are made available with the code. COVIDnet.fr individual data cannot be shared owing to restrictions imposed by the French national data protection authorities. Requests for custom access to aggregated and post-processed data can be made to the GrippeNet.fr/COVIDnet.fr Scientific Committee (https://covidnet.fr/fr/covidnet/confidentialite-et-securite-des-donnees/) through the submission of a scientific proposal describing the aims, methods, data format requested and team proposing the project. Decisions by the GrippeNet.fr/COVIDnet.fr Scientific Committee will be based on pertinence of the scientific proposal to the objectives of COVIDnet.fr and on the constraints on privacy and data treatment imposed by national regulatory authorities. Adjusted data on COVIDnet.fr participants, incidence and healthcare-seeking behaviour are made available with the code. Source data are provided with this paper.

Code availability

Code for the transmission models used for the analyses is available at https://github.com/EPIcx-lab/COVID-19/tree/master/Underdetection_France.

37. Guerrisi, C. et al. Participatory syndromic surveillance of influenza in Europe. *J. Infect. Dis.* **214**, S386–S392 (2016).
38. Debin, M. et al. Evaluating the feasibility and participants' representativeness of an online nationwide surveillance system for influenza in France. *PLoS ONE* **8**, e73675 (2013).
39. Loubet, P. et al. Influenza during pregnancy: incidence, vaccination coverage and attitudes toward vaccination in the French web-based cohort G-GrippeNet. *Vaccine* **34**, 2390–2396 (2016).
40. Boiron, K. et al. Opinion about seasonal influenza vaccination among the general population 3 years after the A(H1N1)pdm2009 influenza pandemic. *Vaccine* **33**, 6849–6854 (2015).
41. Ariza, M. et al. Healthcare-seeking behaviour in case of influenza-like illness in the French general population and factors associated with a GP consultation: an observational prospective study. *BJGP Open* **1**, bjgpopen17X101253 (2018).
42. Mathieu, P. et al. Population perception of mandatory childhood vaccination programme before its implementation, France, 2017. *Euro Surveill.* **24**, 1900053 (2019).
43. Raude, J. et al. Are people excessively pessimistic about the risk of coronavirus infection? Preprint at https://doi.org/10.31234/osf.io/364qj (2020).
44. Cantarelli, P. et al. The representativeness of a European multi-center network for influenza-like-illness participatory surveillance. *BMC Public Health* **14**, 984 (2014).
45. Buitrago-Garcia, D. et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis. *PLoS Med.* **17**, e1003346 (2020).
46. Qiu, X. et al. Defining the role of asymptomatic and pre-symptomatic SARS-CoV-2 transmission - a living systematic review. Preprint at https://doi.org/10.1101/2020.09.01.20135194 (2020).
47. Luo, L. et al. Contact settings and risk for transmission in 3410 close contacts of patients with COVID-19 in Guangzhou, China. *Ann. Intern. Med.* **173**, 879–887 (2020).
48. Zhang, J. et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
49. Davies, N. G. et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26**, 1205–1211 (2020).
50. Zimmermann, P. & Curtis, N. Coronavirus infections in children including COVID-19: an overview of the epidemiology, clinical features, diagnosis, treatment and prevention options in children. *Pediatr. Infect. Dis. J.* **39**, 355–368 (2020).
51. Jiehao, C. et al. A case series of children with 2019 novel coronavirus infection: clinical and epidemiological features. *Clin. Infect. Dis.* **71**, 1547–1551 (2020).
52. Fontanet, A. et al. SARS-CoV-2 infection in primary schools in northern France: a retrospective cohort study in an area of high transmission. Preprint at https://doi.org/10.1101/2020.06.25.20140178 (2020).
53. Fontanet, A. et al. Cluster of COVID-19 in northern France: a retrospective closed cohort study. Preprint at https://doi.org/10.1101/2020.04.18.20071134 (2020).
54. INSEE. Durée et organisation du temps de travail. https://www.insee.fr/fr/statistiques/3303396?sommaire=3353488 (2020).
55. Le Digabel, S. Algorithm 909: NOMAD: nonlinear optimization with the MADS algorithm. *ACM Trans. Math. Softw.* **37**, 44 (2011).

Acknowledgements This study was partially supported by ANR projects DATAREDUX (ANR-19-CE46-0008-03) and EVALCOVID-19 (ANR-20- 706 COVI-0007); EU H2020 grants MOOD (H2020-874850) and RECOVER (H2020-101003589); and REACTing COVID-19 modelling and surveillance grants. We thank P. Crepey, C. Pelat, E. Chatignoux, J. Paireau and D. Levy-Bruhl for discussions, and all participants of the COVIDnet.fr system.

Author contributions V.C. conceived and designed the study. C.T., M.D., C.G., C.K.-K., C.S., T.H., T.B., S.B.-S. and V.C. developed and implemented the protocol of the participatory platform for COVID-19 surveillance. G.P., L.D.D., C.E.S., C.T., J.F., S.V., C.C. and S.B.-S. collected the data. G.P., L.D..D, C.E.S. and C.T. analysed the data. G.P., L.D.D., C.E.S., P.-Y.B. and V.C. developed the inference framework. G.P. and L.D.D. developed the code. G.P., L.D.D., C.E.S. and P.-Y.B. performed the numerical simulations, and analysed the results. G.P., L.D.D., C.E.S., E.V., C.T., M.D., C.G., C.K.-K., C.S., T.H., T.B., P.-Y.B., J.F., S.V., C.C., S.B.-S. and V.C. interpreted the results. V.C. drafted the Article. All authors contributed to and approved the final version of the Article.

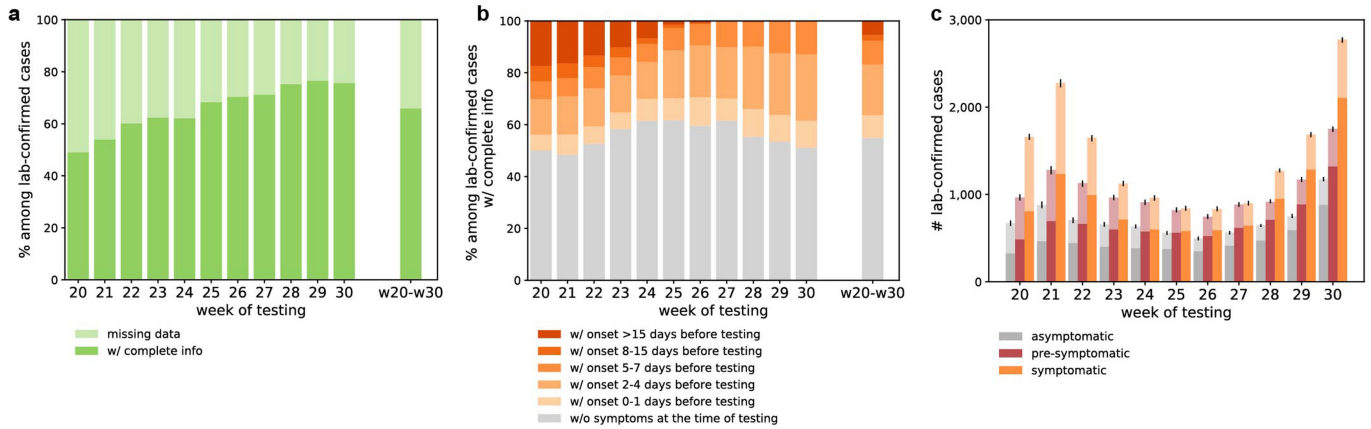
Competing interests The authors declare no competing interests.

Additional information

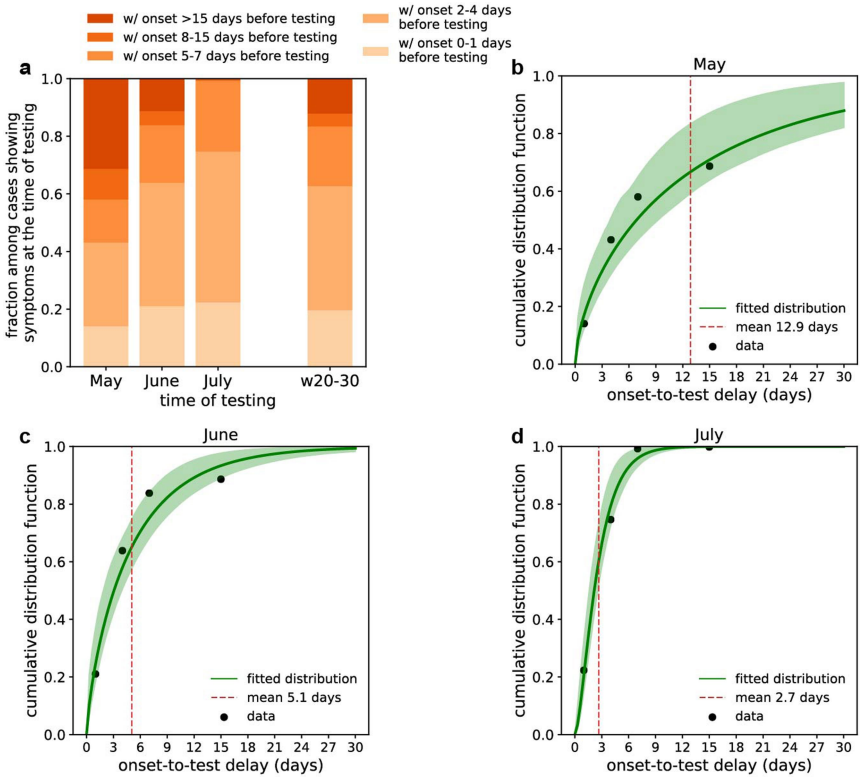
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-020-03095-6.

Correspondence and requests for materials should be addressed to V.C.

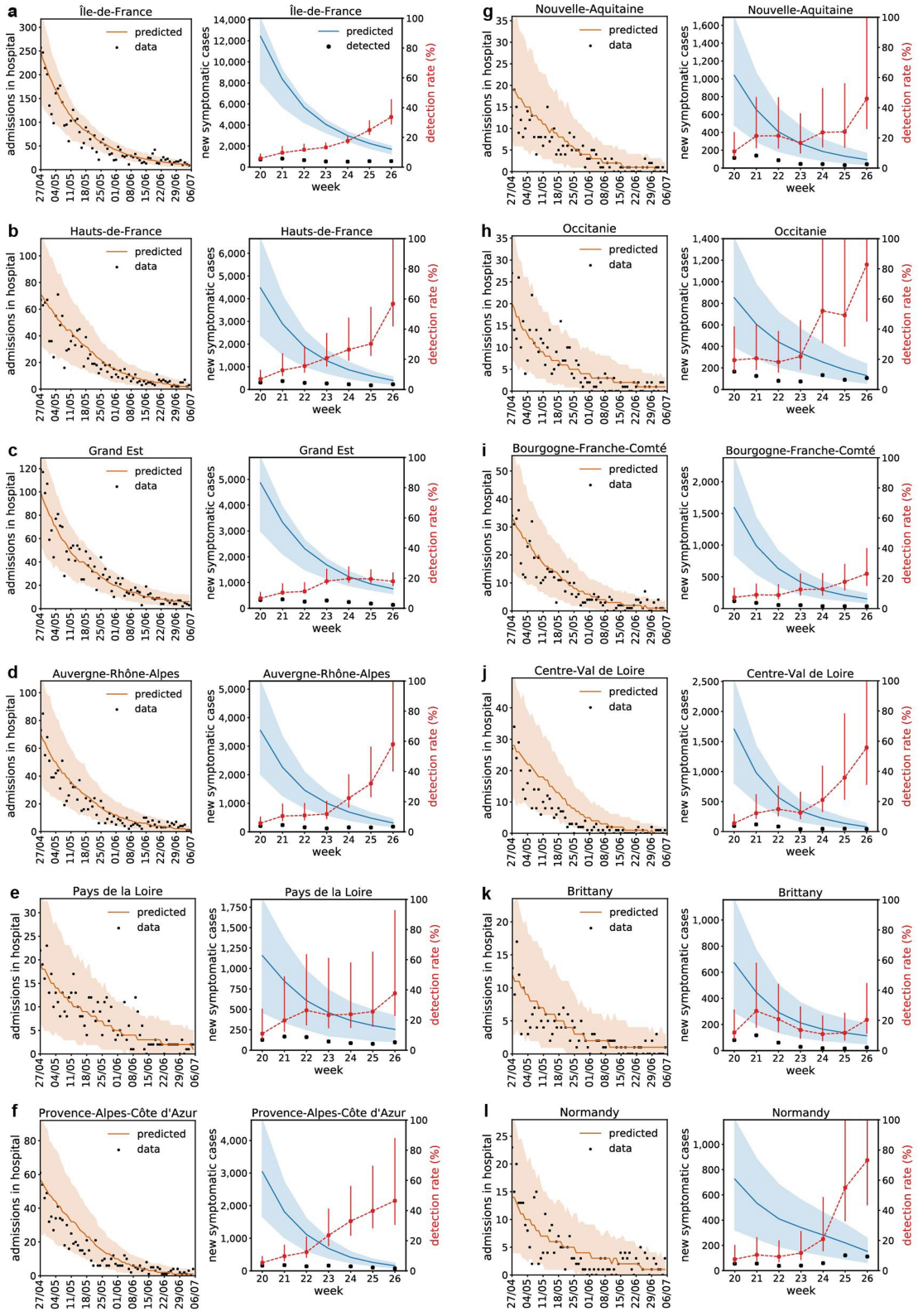
Peer review information *Nature* thanks Ilaria Dorigatti, Niel Hens and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



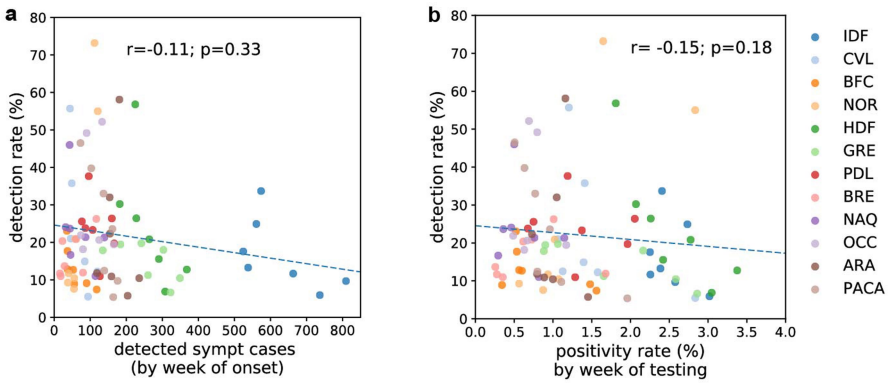
Extended Data Fig. 1 | Information provided by SI-DEP database and imputation of missing data. a, Breakdown of laboratory-confirmed cases of COVID-19 ($n = 35,264$) according to information available. Missing data refer to cases for which information regarding absence or presence of symptoms at the time of testing and self-declared onset-to-test delay were not provided. **b**, Breakdown of laboratory-confirmed cases of COVID-19 with complete information ($n = 23,210$) according to absence or presence of symptoms on the date of testing and onset-to-test delay. **c**, Estimated number of asymptomatic, presymptomatic and symptomatic confirmed cases by week of testing, after imputation of missing data about presence or absence of symptoms. Darker and lighter colours indicate cases with complete information and missing data, respectively. The 95% confidence intervals (black bars) are obtained by applying the imputation procedure 100 times.



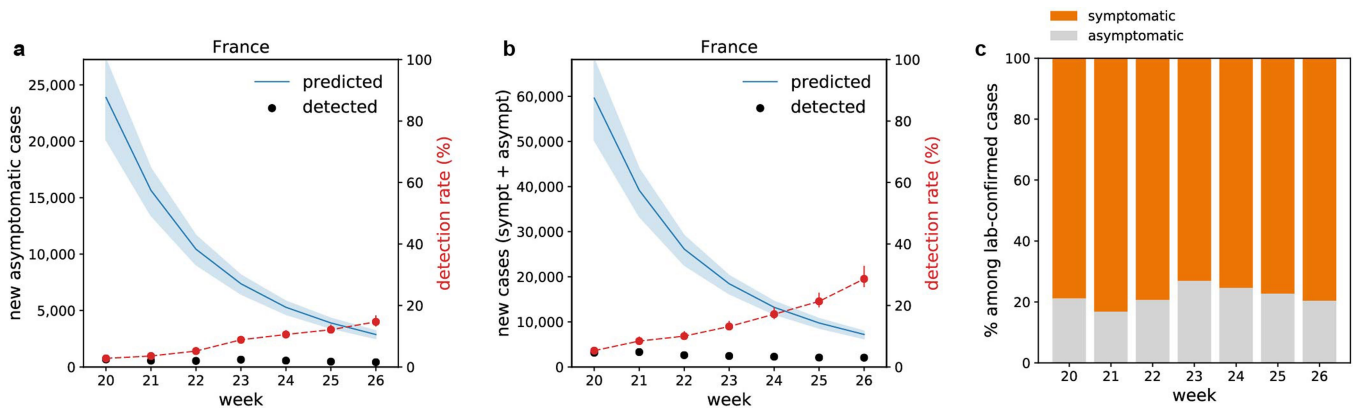
Extended Data Fig. 2 | Estimation of onset-to-test delay distribution. **a**, Data used for the fit. Breakdown of laboratory-confirmed cases of COVID-19 with symptoms (n = 10,494) according to month of testing and the declared time interval of onset-to-testing delay. **b**, Gamma distribution fitted through maximum likelihood to data referring to May (shape parameter 0.61, 0.42–0.80 (median estimate and 95% Wald confidence interval); scale parameter 21.08, 9.41–32.78). **c**, Gamma distribution fitted through maximum likelihood to data referring to June (shape parameter 0.75, 0.51–0.99; scale parameter 6.77, 3.77–9.75). **d**, Gamma distribution fitted through maximum likelihood to data referring to July (shape parameter 1.69, 1.12–2.28; scale parameter 1.57, 0.96–2.17). Shaded areas represent 95% confidence interval of the cumulative distribution function.



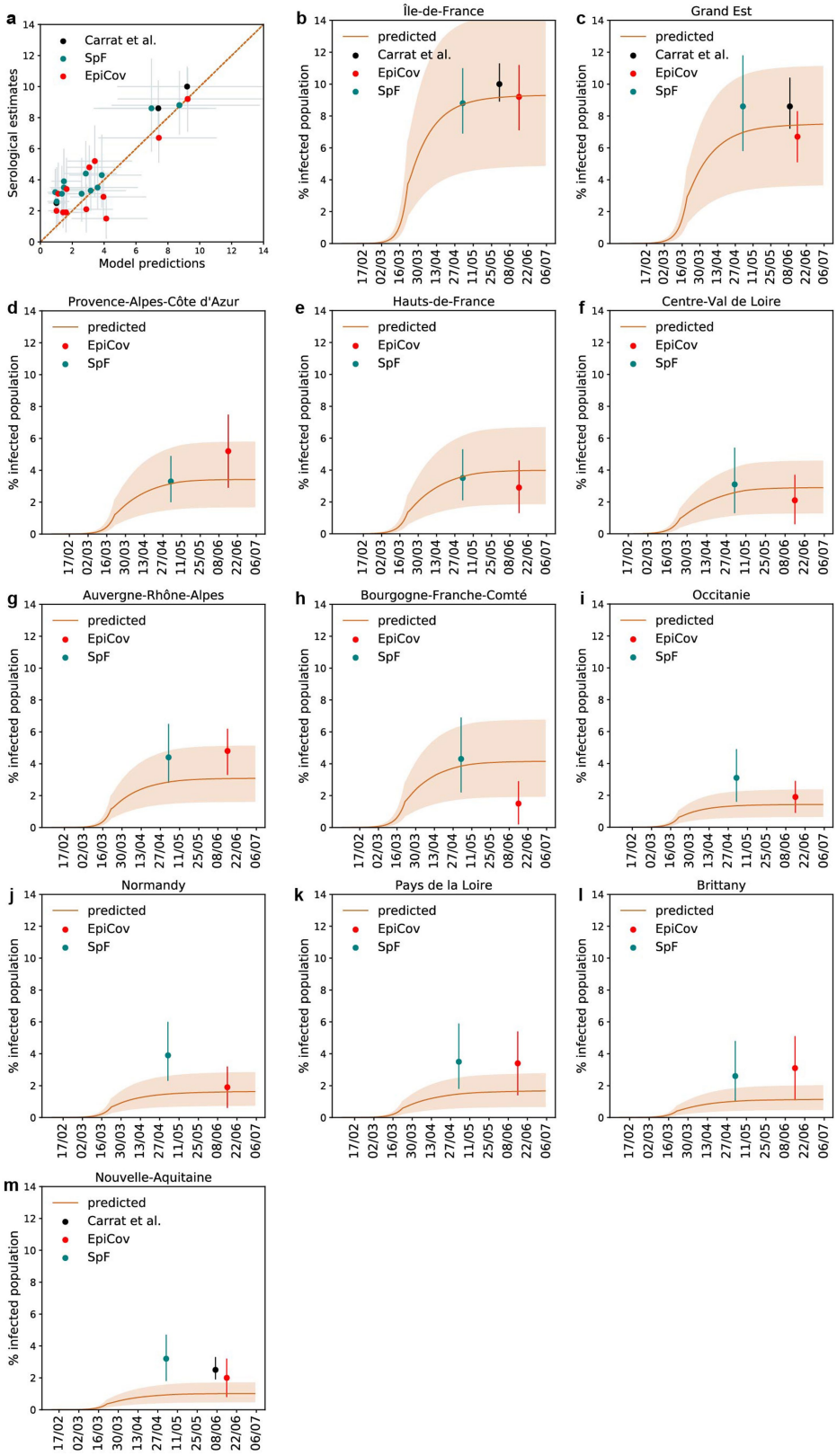
Extended Data Fig. 3 | Hospital admissions and number of new symptomatic cases. a–**l**, For each region, the pair of panels shows the hospital admissions over time (data (points) and simulations (median and 95% confidence interval)) (left) and the projected number of new symptomatic cases over time (median and 95% confidence interval) and the estimated number of virologically confirmed symptomatic cases by week of onset (points) and detection rate (red points; right *y* axis) (right). Medians and 95% confidence intervals are obtained from $n = 500$ independent stochastic runs. Plots are reported for all 12 regions of mainland France: Île-de-France (**a**), Hauts-de-France (**b**), Grand Est (**c**), Auvergne–Rhône–Alpes (**d**), Pays de la Loire (**e**), Provence–Alpes–Cote d'Azur (**f**), Nouvelle Aquitaine (**g**), Occitanie (**h**), Bourgogne–Franche–Comté (**i**), Centre-Val de Loire (**j**), Brittany (**k**) and Normandy (**l**). Hospital admission data up to week 27 (consolidated in week 28) were used to calibrate the models.



Extended Data Fig. 4 | Detection rate versus indicators linked to testing policy and effectiveness. a, Detection rate versus the number of detected symptomatic cases, by week and by region. **b**, Detection rate versus positivity rate by week of testing. Results of a Spearman correlation test are provided (*r*, Spearman correlation coefficient; *P*, *P* value).



Extended Data Fig. 5 | Estimates of the detection rate for symptomatic and asymptomatic cases. a, Projected number of new asymptomatic cases over time (median and 95% confidence interval), estimated number of virologically confirmed symptomatic cases by week of onset (black dots) and detection rate by week (red dots, median and 95% confidence interval) in mainland France. Medians and 95% confidence intervals are obtained from $n = 500$ independent stochastic runs. **b**, As in **a**, but considering both symptomatic and asymptomatic cases. **c**, Percentage of estimated asymptomatic and symptomatic individuals among the estimated number of virologically confirmed cases ($n = 17{,}939$).



Extended Data Fig. 6 | Model predictions versus serological estimates. **a**, Serological estimates per region from three seroprevalence studies versus corresponding model projections. Error bars correspond to 95% confidence intervals. **b**–**m**, For each region, the panel shows the predicted percentage of infected population over time (red curves and shaded areas for median and 95% confidence interval) and serological estimates of EpiCov[26] (red dots), Santé publique France (SpF)[25] (green dots) and ref. [7] (black dots in **b**, **c**, **m**). Medians and 95% confidence intervals for model projections are obtained from $n = 500$ independent stochastic runs. Plots are reported for all 12 regions of mainland France : Île-de-France (**b**), Grand Est (**c**), Provence–Alpes–Cote d'Azur (**d**), Hauts-de-France (**e**), Centre-Val de Loire (**f**), Auvergne–Rhône–Alpes (**g**), Bourgogne–Franche– Comté (**h**), Occitanie (**i**), Normandy (**j**), Pays de la Loire (**k**), Brittany (**l**) and Nouvelle Aquitaine (**m**).

Extended Data Table 1 | Population, confirmed and projected symptomatic cases, and estimated detection rate

Region	**Pop (millions)**	**# lab-confirmed symptomatic cases by week of onset**	**#projected symptomatic cases by week of onset** (median and 95% CI)	**Estimated detection rate (%) of symptomatic cases** (median and 95% CI)
		w20-w26	*w20-w26*	*w20-w26*
Île-de-France (IDF)	12.3	4,405	37,476 [27,692 - 40,786]	12 [11 - 16]
Grand Est (GRE)	5.5	1,797	15,182 [9,887 - 17,377]	12 [10 - 18]
Hauts de France (HDF)	6.0	1,868	12,404 [6,749 - 17,463]	15 [11 - 28]
Auvergne-Rhône-Alpes (ARA)	8.0	1,204	9,749[5,592 - 14,134]	12 [9 - 22]
Occitanie (OCC)	5.9	774	2,793 [1,381 - 4,554]	28 [17 - 56]
Provence-Alpes-Côte d'Azur (PACA)	5.1	952	7,470 [4,250 - 11,041]	13 [9 - 22]
Pays de la Loire (PDL)	3.8	821	4,006 [1,654 - 6,393]	20 [13 - 50]
Bourgogne-Franche-Comté (BFC)	2.8	426	4,264 [2,392 - 6,179]	10 [7 - 18]
Nouvelle Aquitaine (NAQ)	6.0	505	2,789 [1,252 - 4,606]	18 [11 - 40]
Centre-Val de Loire (CVL)	2.6	484	4,066 [1,985 - 5,924]	12 [8 - 24]
Brittany (BRE)	3.3	346	2,037 [954 - 3,478]	17 [10 - 36]
Normandy (NOR)	3.3	480	2,682 [1,245 - 4,340]	18 [11 - 39]
France*	64.6	14,061	103,907 [90,216-116,377]	14 [12 - 16]

Results correspond to the full study period, from 11 May to 28 June 2020.
* Data are for mainland France; Corsica and overseas territories were excluded.

6.3 Conclusions

In this work, we used a stochastic discrete age-stratified transmission model, integrating demographic, age profile, social contact data, mobility data, data on adoption of preventive measures to quantify the underdetection of COVID-19 cases in France in the 7 weeks after the first national lockdown.

A total of 20,777 virologically confirmed cases of COVID-19 on emerging from the first lockdown from 13 May to 28 June 2020 were reported in mainland France by Santé publique France with a weekly number of tests of around 250,000 and a positivity rate < 2%. We used virological data to estimate the number of confirmed cases by onset date, and we measured the weekly rate of underdetection by comparing these estimates with model projections and we found that around 90,000 symptomatic infections, corresponding to 9 out of 10 cases, were not ascertained by the surveillance system, although the test positivity rate was lower than the threshold of 5% recommended by the World Health Organization (WHO). This suggests that test positivity rate is not a good indicator of a system's detection capacity.

The detection rate increased over time in France, ranging from 7% [6%-8%] at the beginning of May to 38% [35%-44%] at the end of June, and it varies widely region by region, showing how the number of cases is not a reliable estimation to monitor viral circulation within regions. Moreover, we found that the capacity of detection of the surveillance system scaled as the inverse of the square root of the incidence, showing that the surveillance system detects more than two out of three cases (66%) only if the incidence is lower than one symptomatic case per 100,000 inhabitants. This suggests that surveillance efforts need to be improved when viral circulation is low.

When strategies such as vaccinations or medical treatments are not available or sufficient, performing a surveillance system is the only option to avoid non-pharmaceutical restrictions. In a previous work, we found that it is necessary to test at least 50% of COVID-19 cases to control local diffusion without putting restrictions in place [81]. Our results, therefore, underlined the importance of improving the performance of the surveillance system even when the positive rate was not alarming. In fact, as expected, at the end of June in France, a detection rate lower than 40% and the relaxation of social distancing involved the resurgence of viral circulation in most regions. This growing incidence strained the testing system and led to the 2nd wave.

In literature static contact matrices have been largely embedded into transmission epidemic model to describe social mixing [89, 91, 93, 95, 191, 192]. In this work, we introduced a method for parametrizing contact matrices and thus embedding into models those changes in time and space of social and mobility behaviour which were due to the ongoing epidemic. Using static matrices from survey data for a pre-pandemic scenario, we parametrized them through mobility and behavioural data in order to model the effective social mixing across age groups during and on emergence from the first lockdown in France. More specifically, we assessed variations in social contacts due to control strategies in terms of partial closure of school, recommendations on working remotely, and avoiding mass gatherings.

Google mobility reports [121] provide the variation in each region of people going to work compared to a pre-pandemic scenario. We used such information to reduce contact at work and increase contact at home accordingly. We also parametrized adaptive behaviours of individuals to the epidemic in terms of senior isolation and avoidance of physical contact. Integrating this information into the epidemic models allowed us to reproduce reliable outputs on new infections and thus assess the

efficacy of the testing system.

In order to assess the impact of using parametrized contact matrices compared to static ones which do not take behavioural changes into account, we performed a model selection analysis comparing fitted simulated outputs using both approaches. We found that accounting for changes in contact patterns over time better describes the epidemic trajectory (See Supplementary Material of the paper). Parametrizing contact matrices through real-time data is thus extremely powerful in developing more realistic transmission models.

Given the reliability of simulated output provided by dynamical matrices, the use of parametrized social contacts in specific settings may be used in the epidemic model for doing predictions and scenario analysis. For instance, they may be used to reproduce social mixing during potential targeted intervention scenarios (e.g. school closure, required teleworking) and to inform predictive epidemic models in order to assess the expected efficacy of different social or travel restrictions [124, 125, 193].

There are however some limitations. We developed region-based models, not taking into account potential re-seeding events caused by mobility to/from other regions. However, given the strong reductions in mobility in France during the first lockdown and the limited individual displacements in May-June, in this context, inter-regional trips were negligible [31]. Moreover, foreign importations were neglected, as France reopened its borders with the European Member States in June, and the Schengen area remained closed till July.

To conclude, as static contact matrices have been computed in several countries over the world [89, 95], dynamical matrices could be parametrized over time in such countries and applied to specific epidemic contexts. Since the approach is fully data-driven, the availability of mobility and behavioural data on a large scale becomes the fundamental ingredient to parametrize social contacts. Mobility data such as Google and Apple mobility reports, however, are available for all countries in the world, providing daily positioning indicators in different settings and it may thus make possible the parametrization of social mixing across countries.

Chapter 7

Conclusions and Perspectives

Over the previous Chapters, I discussed how mobility can be inferred by several sources of data and can be integrated into mathematical models to identify the main mechanisms that drive epidemic transmission and to support the management of an health crisis. In Chapter 2, 3 and 4, particular attention is dedicated to the extraction of mobility fluxes from mobile phone data and their integration into epidemic models. In Chapter 5 and 6, I presented two of the published works in which we integrated air traffic and mobile phone data into epidemic models to help governments and health authorities to fight the COVID-19 pandemic.

In the first work, during the early stage of the COVID-19 pandemic, we assessed the risk of global epidemic importation of COVID-19 into Europe, due to air traffic connections with the most affected areas in China. We were interested in understanding the risk of carrying the virus to other geographically-distant countries, and we looked at the international travel before and after the travel ban in Hubei province. Air traffic data have thus been integrated into a probabilistic model to compute the importation risk.

In the second work, we focused on the assessment of testing policy during the 7 weeks following the first national lockdown in France, estimating the rate of underdetection of COVID- 19 cases. We used positioning indicators from mobile phone data and behavioural data to parametrize social mixing. Moreover, virological and surveillance data were used for model calibration and validation and to compute the rate of detection. Our work highlights the importance of merging real-time mobility, epidemiological and other types of behavioural data in order to improve predictive power, accounting for mobility and social changes due to mobility restrictions, and to spontaneous adaptive behaviours to the epidemic like risk aversion.

The rapidly evolving situation during the COVID-19 pandemic has highlighted the importance of mobility data to monitor individual behaviours and to inform epidemic models. In recent decades, many network operators have started to share their data for the epidemiological studies of various infectious disease epidemics, including e.g. malaria, cholera, Ebola; researchers also validated them as a good proxy for human mobility. During the COVID-19 emergency we witnessed a data-sharing revolution in which network operators such as Orange, Vodafone, Telefonica, big companies like Google, Apple, Facebook, or small ones as Cuebiq, SafeGraph, Unacast, started providing their aggregated mobility data (Call Detail Records, signaling and location history data) extracted from mobile phone traces in real-time to support the outbreak response.

Thanks to the mobile phone data availability, and the past 20 years development of mathematical and computational frameworks to integrate mobility data into models, researchers were therefore able to provide a rapid response to the global heath emergency.

The rapid and wide availability of mobile phone data, however, has also opened many new challenges. Firstly, given the high-resolution in space and time of individual displacements extracted from mobile phone data, it becomes important to assess the level of detail of such information that is relevant in modeling epidemic transmission. In an effort to gain better knowledge of this, I assessed the appropriate definition of the mobility process that is relevant to epidemic spread and the resolution of mobility fluxes extracted from Call Detail Records (CDRs) needed to inform the transmission models. The results which I presented in Section 4.5 show how preserving displacements on the individual trajectories does not capture the epidemiological link between different locations. Moreover, we found that on the national scale commuting mobility is the dominant driver of disease diffusion. In fact, tracking individual activities beyond home and work/school locations (e.g., leisure activities) does not add relevant epidemiologically information.

This work was specifically on CDRs, however during COVID-19 emergency, data based on GPS tracks like location history data from mobile apps started emerging as new powerful sources of mobile phone data. These new data allow the tracing of individual trajectories in even more detail, capturing positioning with a precision in the range of meters over time, contrary to CDRs which capture positioning only when individuals make a call/SMSs and at cell tower resolution. The increasing availability and detail provided by mobile phones thus require further investigation into the process of integration of mobile phone data into models. During COVID-19, however, the higher resolution of such type of data allowed studying individuals' movements and positioning in specific places (e.g. at home, at work, in places of entertainment, transport) for studying the compliance of mobility restrictions and to inform transmission models [106, 121, 122, 194, 195].

Moreover, the mobile phone data extensive sharing resulted in many controversies on privacy issues. In fact, sharing and dealing with such sensitive data capturing individuals' behaviour involves several risks. Risks which might compromise the security of citizens. During the COVID-19 pandemic, mobile phone data have been shared at the national level through ad hoc legal agreements between national network operators/app companies and researchers or shared by companies (e.g., Google, Apple) in online platforms which provide periodical mobility indicators. However, many data providers are still reluctant to share their data, even when aggregated. To drastically reduce concerns about sharing data, standard protocols for sharing them can therefore be envisioned which will strengthen privacy and confidentiality, while at the same time providing data at the resolution needed for in-depth epidemiological studies.

To conclude, collaborative efforts are still needed to exploit the full potential of mobile phone data for crisis to come.

Publication issued from my thesis

- Pullano, G. et al. (2020). Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020. Eurosurveillance, 25(4), 2000057.
- Pullano, G. et al. (2020). Underdetection of cases of COVID-19 in France threatens epidemic control. Nature https://doi.org/10.1038/s41586-020-03095- 6.

Publication related to my thesis

- Gilbert, M., Pullano, G., Pinotti, F., Valdano, E., Poletto, C., Boëlle, P. Y., ... & Colizza, V.. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. The Lancet, 395(10227), 871-877, 19/02/2020.
- Pinotti, F., Di Domenico, L., Ortega, E., Mancastroppa, M., Pullano, G., Valdano, E., ... & Colizza, V. (2020). Tracing and analysis of 288 early SARS-CoV-2 infections outside China: A modeling study. PLoS Medicine, 17(7), e1003193.
- Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S., & Colizza, V. (2020). Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. The Lancet Digital Health, 2(12), e638-e649
- Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P. Y., & Colizza, V. (2020). Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. BMC Medicine, 18(1), 1-13.
- Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P. Y., & Colizza, V. (2021). Modelling safe protocols for reopening schools during the COVID-19 pandemic in France. Nature Communications, 12(1), 1-10.
- Di Domenico, L., Sabbatini, C. E., Pullano, G., Lévy-Bruhl, D., & Colizza, V. (2021). Impact of January 2021 curfew measures on SARS-CoV-2 B. 1.1. 7 circulation in France. Eurosurveillance, 26(15), 2100272.

Bibliography

[1] J. S. Koopman. "Emerging objectives and methods in epidemiology". eng. In: *American Journal of Public Health* 86.5 (May 1996), pp. 630–632. ISSN: 0090-0036. DOI: 10.2105/ajph.86.5.630.

[2] Matt Keeling, Pejman Rohani, and Babak Pourbohloul. "Modeling Infectious Diseases in Humans and Animals". In: *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 47 (Oct. 2008), pp. 864–865. DOI: 10.1086/591197.

[3] Suttikiat Changruenngam, Dominique J. Bicout, and Charin Modchang. "How the individual human mobility spatio-temporally shapes the disease transmission dynamics". en. In: *Scientific Reports* 10.1 (July 2020). Number: 1 Publisher: Nature Publishing Group, p. 11325. ISSN: 2045-2322. DOI: 10.1038/s41598-020-68230-9. URL: https://www.nature.com/articles/s41598-020-68230-9 (visited on 05/18/2021).

[4] A. Barrat et al. "The architecture of complex weighted networks". en. In: *Proceedings of the National Academy of Sciences* 101.11 (Mar. 2004). Publisher: National Academy of Sciences Section: Physical Sciences, pp. 3747–3752. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0400087101. URL: https://www.pnas.org/content/101/11/3747 (visited on 05/03/2021).

[5] Vittoria Colizza et al. "The role of the airline transportation network in the prediction and predictability of global epidemics". en. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.7 (Feb. 2006), pp. 2015–2020. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0510525103. URL: http://www.pnas.org/content/103/7/2015 (visited on 01/11/2018).

[6] Vittoria Colizza et al. "Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study". In: *BMC Medicine* 5.1 (Nov. 2007), p. 34. ISSN: 1741-7015. DOI: 10.1186/1741-7015-5-34. URL: https://doi.org/10.1186/1741-7015-5-34 (visited on 04/13/2021).

[7] D. Balcan et al. "Multiscale mobility networks and the spatial spreading of infectious diseases". en. In: *Proceedings of the National Academy of Sciences* 106.51 (Dec. 2009), pp. 21484–21489. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0906910106. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0906910106 (visited on 01/16/2020).

[8] Aidan Findlater. "Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel". en. In: (), p. 12.

[9] Ana Pastore y Piontti et al. *Charting the Next Pandemic: Modeling Infectious Disease Spreading in the Data Science Age*. en. Springer International Publishing, 2019. ISBN: 978-3-319-93289-7. DOI: 10.1007/978-3-319-93290-3. URL: https://www.springer.com/gp/book/9783319932897 (visited on 06/24/2021).

[10] Paolo Bajardi et al. "Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic". en. In: *PLOS ONE* 6.1 (2011). Publisher: Public Library of Science, e16591. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0016591. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0016591 (visited on 02/16/2021).

[11] Michele Tizzoni et al. "On the Use of Human Mobility Proxies for Modeling Epidemics". en. In: *PLOS Computational Biology* 10.7 (July 2014), e1003716. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003716. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003716 (visited on 01/15/2020).

[12] Sandro Meloni et al. "Modeling human mobility responses to the large-scale spreading of infectious diseases". en. In: *Scientific Reports* 1.1 (Aug. 2011). Number: 1 Publisher: Nature Publishing Group, p. 62. ISSN: 2045-2322. DOI: 10.1038/srep00062. URL: https://www.nature.com/articles/srep00062 (visited on 04/13/2021).

[13] Xin Lu, Linus Bengtsson, and Petter Holme. "Predictability of population displacement after the 2010 Haiti earthquake". en. In: *Proceedings of the National Academy of Sciences* 109.29 (July 2012). Publisher: National Academy of Sciences Section: Physical Sciences, pp. 11576–11581. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1203882109. URL: https://www.pnas.org/content/109/29/11576 (visited on 02/25/2021).

[14] Amy Maxmen. "Can tracking people through phone-call data improve lives?" eng. In: *Nature* 569.7758 (2019), pp. 614–617. ISSN: 1476-4687. DOI: 10.1038/d41586-019-01679-5.

[15] Flavio Finger et al. "Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.23 (June 2016), pp. 6421–6426. ISSN: 0027-8424. DOI: 10.1073/pnas.1522305113. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4988598/ (visited on 01/11/2018).

[16] Amy Wesolowski et al. "Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data". en. In: *Proceedings of the National Academy of Sciences* 112.35 (Sept. 2015), pp. 11114–11119. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1423542112. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1423542112 (visited on 04/14/2021).

[17] Corey Peak et al. "Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data". In: *International journal of epidemiology* 47 (June 2018). DOI: 10.1093/ije/dyy095.

[18] Vincent D. Blondel, Adeline Decuyper, and Gautier Krings. "A survey of results on mobile phone datasets analysis". En. In: *EPJ Data Science* 4.1 (Dec. 2015), p. 10. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-015-0046-0. URL: https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-015-0046-0 (visited on 01/11/2018).

[19] Yves-Alexandre de Montjoye et al. "Unique in the Crowd: The privacy bounds of human mobility". en. In: *Scientific Reports* 3.1 (Mar. 2013). Bandiera_abtest: a Cc_license_type: cc_y Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Applied mathematics;Applied physics;Computational science;Statistics Subject_term_id:

applied-mathematics;applied-physics;computational-science;statistics, p. 1376. ISSN: 2045-2322. DOI: 10.1038/srep01376. URL: https://www.nature.com/articles/srep01376 (visited on 06/24/2021).

[20] Yves-Alexandre de Montjoye et al. "On the privacy-conscientious use of mobile phone data". en. In: *Scientific Data* 5.1 (Dec. 2018). Number: 1 Publisher: Nature Publishing Group, p. 180286. ISSN: 2052-4463. DOI: 10.1038/sdata.2018.286. URL: https://www.nature.com/articles/sdata2018286 (visited on 04/16/2021).

[21] Nuria Oliver et al. "Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle". en. In: *Science Advances* 6.23 (June 2020). Publisher: American Association for the Advancement of Science Section: Editorial, eabc0764. ISSN: 2375-2548. DOI: 10.1126/sciadv.abc0764. URL: https://advances.sciencemag.org/content/6/23/eabc0764 (visited on 02/25/2021).

[22] Marta C. González, César A. Hidalgo, and Albert-László Barabási. "Understanding individual human mobility patterns". en. In: *Nature* 453.7196 (June 2008). Number: 7196 Publisher: Nature Publishing Group, pp. 779–782. ISSN: 1476-4687. DOI: 10.1038/nature06958. URL: https://www.nature.com/articles/nature06958 (visited on 02/18/2021).

[23] J. Candia et al. "Uncovering individual and collective human dynamics from mobile phone records". In: *Journal of Physics A: Mathematical and Theoretical* 41.22 (June 2008). arXiv: 0710.2939, p. 224015. ISSN: 1751-8113, 1751-8121. DOI: 10.1088/1751-8113/41/22/224015. URL: http://arxiv.org/abs/0710.2939 (visited on 01/11/2018).

[24] Balázs Cs Csáji et al. "Exploring the Mobility of Mobile Phone Users". In: *Physica A: Statistical Mechanics and its Applications* 392.6 (Mar. 2013). arXiv: 1211.6014, pp. 1459–1473. ISSN: 03784371. DOI: 10.1016/j.physa.2012.11.040. URL: http://arxiv.org/abs/1211.6014 (visited on 02/25/2021).

[25] Stefania Rubrichi, Zbigniew Smoreda, and Mirco Musolesi. "A Comparison of Spatial-based Targeted Disease Containment Strategies using Mobile Phone Data". In: *arXiv:1706.00690 [physics]* (June 2017). arXiv: 1706.00690. URL: http://arxiv.org/abs/1706.00690 (visited on 01/08/2018).

[26] Amy Wesolowski et al. "Impact of human mobility on the emergence of dengue epidemics in Pakistan". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 112.38 (Sept. 2015), pp. 11887–11892. ISSN: 1091-6490. DOI: 10.1073/pnas.1504964112.

[27] A. Lima et al. "Disease Containment Strategies based on Mobility and Information Dissemination". En. In: *Scientific Reports* 5 (June 2015), p. 10650. ISSN: 2045-2322. DOI: 10.1038/srep10650. URL: https://www.nature.com/articles/srep10650 (visited on 01/11/2018).

[28] Cecilia Panigutti et al. "Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models". en. In: *Royal Society Open Science* 4.5 (May 2017), p. 160950. ISSN: 2054-5703. DOI: 10.1098/rsos.160950. URL: http://rsos.royalsocietypublishing.org/content/4/5/160950 (visited on 01/11/2018).

[29] Manuela Ciddio et al. "The spatial spread of schistosomiasis: A multidimensional network model applied to Saint-Louis region, Senegal". eng. In: *Advances in Water Resources* 108 (Oct. 2017), pp. 406–415. ISSN: 0309-1708. DOI: 10.1016/j.advwatres.2016.10.012.

[30] C Poletto et al. "Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic". en. In: *Eurosurveillance* 19.42 (Oct. 2014). ISSN: 1560-7917. DOI: 10.2807/1560-7917.ES2014.19.42.20936. URL: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES2014.19.42.20936 (visited on 04/15/2021).

[31] Giulia Pullano et al. "Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study". English. In: *The Lancet Digital Health* 2.12 (Dec. 2020). Publisher: Elsevier, e638–e649. ISSN: 2589-7500. DOI: 10.1016/S2589-7500(20)30243-0. URL: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30243-0/abstract (visited on 05/06/2021).

[32] *COVID-19: Belgium analyses telecom data to measure the impact of confinement*. en-US. URL: https://press.telenet.be/covid-19-belgium-analyses-telecom-data-to-measure-the-impact-of-confinement (visited on 04/02/2021).

[33] Mattia Mazzoli José Javier Ramasco. *Ramasco, J. J., and Mazzoli, M., 4° Informe sobre los cambios de movilidad en España debido a las medidas de confinamiento contra la extensión del COVID-19 https://analytics.ifisc.uib-csic.es/es/respuesta-covid-19/?fbclid=IwAR2C8nA_CWa_AgrD4e_DF5UqWVs-QNZpQeDkxGED-HS1gOyDX5aDpr6UZKY Accessed : April 2020*. (Visited on 04/16/2020).

[34] Emanuele Pepe et al. "COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown". In: *Scientific Data* 7 (July 2020). DOI: 10.1038/s41597-020-00575-2.

[35] Eugenio Valdano et al. "Highlighting socio-economic constraints on mobility reductions during COVID-19 restrictions in France can inform effective and equitable pandemic response". In: *Journal of Travel Medicine* 28.4 (May 2021). ISSN: 1708-8305. DOI: 10.1093/jtm/taab045. URL: https://doi.org/10.1093/jtm/taab045 (visited on 08/09/2021).

[36] SPF. *Santé publique France.* fr. URL: https://www.santepubliquefrance.fr/ (visited on 03/09/2021).

[37] *Réseau Sentinelles.* URL: https://www.sentiweb.fr/?page=presentation (visited on 07/08/2021).

[38] Matteo Chinazzi et al. "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak". en. In: *Science* 368.6489 (Apr. 2020). Publisher: American Association for the Advancement of Science Section: Research Article, pp. 395–400. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aba9757. URL: https://science.sciencemag.org/content/368/6489/395 (visited on 03/09/2021).

[39] Marino Gatto et al. "Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures". en. In: *Proceedings of the National Academy of Sciences* 117.19 (May 2020), pp. 10484–10491. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2004978117. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.2004978117 (visited on 04/13/2021).

[40] Ruizhi Zheng et al. "Spatial transmission of COVID-19 via public and private transportation in China". In: *Travel Medicine and Infectious Disease* 34 (2020), p. 101626. ISSN: 1477-8939. DOI: 10.1016/j.tmaid.2020.101626. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7118651/ (visited on 04/14/2021).

[41] Joseph T. Wu, Kathy Leung, and Gabriel M. Leung. "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study". English. In: *The Lancet* 395.10225 (Feb. 2020). Publisher: Elsevier, pp. 689–697. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(20)30260-9. URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30260-9/abstract (visited on 03/31/2021).

[42] Neil M. Ferguson et al. "Strategies for containing an emerging influenza pandemic in Southeast Asia". en. In: *Nature* 437.7056 (Sept. 2005). Number: 7056 Publisher: Nature Publishing Group, pp. 209–214. ISSN: 1476-4687. DOI: 10.1038/nature04017. URL: https://www.nature.com/articles/nature04017 (visited on 03/09/2021).

[43] A. Marm Kilpatrick et al. "Predicting the global spread of H5N1 avian influenza". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.51 (Dec. 2006), pp. 19368–19373. ISSN: 0027-8424. DOI: 10.1073/pnas.0609227103.

[44] Cécile Viboud et al. "Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza". en. In: *Science* 312.5772 (Apr. 2006). Publisher: American Association for the Advancement of Science Section: Report, pp. 447–451. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1125237. URL: https://science.sciencemag.org/content/312/5772/447 (visited on 03/09/2021).

[45] Caroline O. Buckee et al. "Mobile phones and malaria: Modeling human and parasite travel". en. In: *Travel Medicine and Infectious Disease*. Special Issue : Malaria Reviews 11.1 (Jan. 2013), pp. 15–22. ISSN: 1477-8939. DOI: 10.1016/j.tmaid.2012.12.003. URL: https://www.sciencedirect.com/science/article/pii/S1477893913000057 (visited on 04/14/2021).

[46] Amy Wesolowski et al. "Commentary: Containing the Ebola Outbreak – the Potential and Challenge of Mobile Network Dataeffe". English. In: *PLOS Currents Outbreaks* (Sept. 2014). ISSN: 2157-3999. DOI: 10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e. URL: http://currents.plos.org/outbreaks/article/containing-the-ebola-outbreak-the-potential-and-challenge-of-mobile-network-data/ (visited on 01/11/2018).

[47] C. Poletto et al. "Assessment of the Middle East respiratory syndrome coronavirus (MERS-CoV) epidemic in the Middle East and risk of international spread using a novel maximum likelihood analysis approach". eng. In: *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 19.23 (June 2014). ISSN: 1560-7917. DOI: 10.2807/1560-7917.es2014.19.23.20824.

[48] L. Hufnagel, D. Brockmann, and T. Geisel. "Forecast and control of epidemics in a globalized world". en. In: *Proceedings of the National Academy of Sciences* 101.42 (Oct. 2004). Publisher: National Academy of Sciences Section: Biological Sciences, pp. 15124–15129. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0308344101. URL: https://www.pnas.org/content/101/42/15124 (visited on 03/09/2021).

[49] Jean Gaudart et al. "Spatio-temporal dynamics of cholera during the first year of the epidemic in Haiti". eng. In: *PLoS neglected tropical diseases* 7.4 (2013), e2145. ISSN: 1935-2735. DOI: 10.1371/journal.pntd.0002145.

[50] Jennifer L. Kyle and Eva Harris. "Global Spread and Persistence of Dengue". In: *Annual Review of Microbiology* 62.1 (2008). _eprint: https://doi.org/10.1146/annurev.micro.62 pp. 71–92. DOI: 10.1146/annurev.micro.62.081307.163005. URL: https://doi.org/10.1146/annurev.micro.62.081307.163005 (visited on 04/16/2021).

[51] Arnaud Le Menach et al. "Travel risk, malaria importation and malaria transmission in Zanzibar". en. In: *Scientific Reports* 1.1 (Sept. 2011). Number: 1 Publisher: Nature Publishing Group, p. 93. ISSN: 2045-2322. DOI: 10.1038/srep00093. URL: https://www.nature.com/articles/srep00093 (visited on 04/16/2021).

[52] Amy Wesolowski et al. "Quantifying the impact of human mobility on malaria". eng. In: *Science (New York, N.Y.)* 338.6104 (Oct. 2012), pp. 267–270. ISSN: 1095-9203. DOI: 10.1126/science.1223467.

[53] Dirk Brockmann and Dirk Helbing. "The Hidden Geometry of Complex, Network-Driven Contagion Phenomena". en. In: *Science* 342.6164 (Dec. 2013). Publisher: American Association for the Advancement of Science Section: Research Article, pp. 1337–1342. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1245200. URL: https://science.sciencemag.org/content/342/6164/1337 (visited on 04/12/2021).

[54] Pascal Crépey and Marc Barthélemy. "Detecting Robust Patterns in the Spread of Epidemics: A Case Study of Influenza in the United States and France". In: *American Journal of Epidemiology* 166.11 (Dec. 2007), pp. 1244–1251. ISSN: 0002-9262. DOI: 10.1093/aje/kwm266. URL: https://doi.org/10.1093/aje/kwm266 (visited on 04/18/2021).

[55] Xin Lu et al. "Approaching the Limit of Predictability in Human Mobility". En. In: *Scientific Reports* 3 (Oct. 2013), p. 2923. ISSN: 2045-2322. DOI: 10.1038/srep02923. URL: https://www.nature.com/articles/srep02923 (visited on 01/11/2018).

[56] Wouter Van den Broeck et al. "The GLEAMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale". In: *BMC infectious diseases* 11 (Feb. 2011), p. 37. DOI: 10.1186/1471-2334-11-37.

[57] Marianne E. Sinka et al. "A global map of dominant malaria vectors". In: *Parasites & Vectors* 5.1 (Apr. 2012), p. 69. ISSN: 1756-3305. DOI: 10.1186/1756-3305-5-69. URL: https://doi.org/10.1186/1756-3305-5-69 (visited on 06/01/2021).

[58] Andrew J. Tatem, David J. Rogers, and Simon I. Hay. "Estimating the malaria risk of African mosquito movement by air travel". In: *Malaria Journal* 5.1 (July 2006), p. 57. ISSN: 1475-2875. DOI: 10.1186/1475-2875-5-57. URL: https://doi.org/10.1186/1475-2875-5-57 (visited on 05/19/2021).

[59] John H. Ouma et al. "Morbidity in schistosomiasis: an update". en. In: *Trends in Parasitology* 17.3 (Mar. 2001), pp. 117–118. ISSN: 1471-4922. DOI: 10.1016/S1471-4922(00)01877-8. URL: https://www.sciencedirect.com/science/article/pii/S1471492200018778 (visited on 06/01/2021).

[60] *2014-2016 Ebola Outbreak in West Africa | History | Ebola (Ebola Virus Disease) | CDC*. en-us. Mar. 2020. URL: https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html (visited on 04/07/2021).

[61] Robert Roos | May 07 and 2003. *Estimates of SARS death rates revised upward*. en. URL: https://www.cidrap.umn.edu/news-perspective/2003/05/estimates-sars-death-rates-revised-upward (visited on 08/11/2021).

[62] *Middle East respiratory syndrome coronavirus (MERS-CoV)*. URL: https://www.who.int/health-topics/middle-east-respiratory-syndrome-coronavirus-mers (visited on 08/11/2021).

[63] Robert Roos | Aug 08 and 2011. *Study puts global 2009 H1N1 infection rate at 11% to 21%*. en. URL: https://www.cidrap.umn.edu/news-perspective/2011/08/study-puts-global-2009-h1n1-infection-rate-11-21 (visited on 08/11/2021).

[64] *WHO Coronavirus (COVID-19) Dashboard*. en. URL: https://covid19.who.int (visited on 06/01/2021).

[65] Hien Lau et al. "The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak". In: *Journal of microbiology, immunology, and infection = Wei mian yu gan ran za zhi* (Mar. 2020). DOI: 10.1016/j.jmii.2020.03.026.

[66] Moritz U G Kraemer et al. "The effect of human mobility and control measures on the COVID-19 epidemic in China". en. In: (2020), p. 6.

[67] Hamish Gibbs et al. "Changing travel patterns in China during the early stages of the COVID-19 pandemic". en. In: *Nature Communications* 11.1 (Oct. 2020). Number: 1 Publisher: Nature Publishing Group, p. 5012. ISSN: 2041-1723. DOI: 10.1038/s41467-020-18783-0. URL: https://www.nature.com/articles/s41467-020-18783-0 (visited on 06/01/2021).

[68] Andrew Tatem. *Pandemic acceleration*. en-GB. URL: http://campaignforsocialscience.org.uk/news/pandemic-acceleration/ (visited on 03/24/2021).

[69] Vittoria Colizza et al. "Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions". eng. In: *PLoS medicine* 4.1 (Jan. 2007), e13. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.0040013.

[70] Vittoria Colizza and Alessandro Vespignani. "Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations". en. In: *Journal of Theoretical Biology* 251.3 (Apr. 2008), pp. 450–467. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2007.11.028. URL: https://www.sciencedirect.com/science/article/pii/S0022519307005991 (visited on 07/02/2021).

[71] Andrew J. Tatem. "Mapping population and pathogen movements". In: *International Health* 6.1 (Mar. 2014), pp. 5–11. ISSN: 1876-3413. DOI: 10.1093/inthealth/ihu006. URL: https://doi.org/10.1093/inthealth/ihu006 (visited on 05/24/2021).

[72] Michael Walters Dols. *The Black Death in the Middle East*. en. Google-Books-ID: F22DDwAAQBAJ. Princeton University Press, Jan. 2019. ISBN: 978-0-691-19668-8.

[73] Vittoria Colizza, Romualdo Pastor-Satorras, and Alessandro Vespignani. "Reaction–diffusion processes and metapopulation models in heterogeneous networks". en. In: *Nature Physics* 3.4 (Apr. 2007). Number: 4 Publisher: Nature Publishing Group, pp. 276–282. ISSN: 1745-2481. DOI: 10.1038/nphys560. URL: https://www.nature.com/articles/nphys560 (visited on 04/13/2021).

[74] A. Paolo Masucci et al. "Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows". en. In: *Physical Review E* 88.2 (Aug. 2013), p. 022812. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.88.022812. URL: https://link.aps.org/doi/10.1103/PhysRevE.88.022812 (visited on 04/16/2021).

[75] Shengjie Lai et al. "Seasonal and interannual risks of dengue introduction from South-East Asia into China, 2005-2015". en. In: *PLOS Neglected Tropical Diseases* 12.11 (Nov. 2018). Publisher: Public Library of Science, e0006743. ISSN: 1935-2735. DOI: 10.1371/journal.pntd.0006743. URL: https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0006743 (visited on 04/16/2021).

[76] Benjamin D. Dalziel, Babak Pourbohloul, and Stephen P. Ellner. "Human mobility patterns predict divergent epidemic dynamics among cities". In: *Proceedings of the Royal Society B: Biological Sciences* 280.1766 (Sept. 2013). ISSN: 0962-8452. DOI: 10.1098/rspb.2013.0763. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3730584/ (visited on 04/13/2021).

[77] Stephen Eubank et al. "Modelling disease outbreaks in realistic urban social networks". en. In: *Nature* 429.6988 (May 2004). Number: 6988 Publisher: Nature Publishing Group, pp. 180–184. ISSN: 1476-4687. DOI: 10.1038/nature02541. URL: https://www.nature.com/articles/nature02541 (visited on 04/18/2021).

[78] Marius Gilbert et al. "Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study". English. In: *The Lancet* 395.10227 (Mar. 2020). Publisher: Elsevier, pp. 871–877. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(20)30411-6. URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30411-6/abstract (visited on 02/25/2021).

[79] *Archived: WHO Timeline - COVID-19. https://www.who.int/news/item/27-04-2020-who-timeline—covid-19.* en. URL: https://www.who.int/news/item/27-04-2020-who-timeline---covid-19 (visited on 02/16/2021).

[80] Jessica T. Davis et al. "Estimating the establishment of local transmission and the cryptic phase of the COVID-19 pandemic in the USA". In: *medRxiv* (Aug. 2020), p. 2020.07.06.20140285. DOI: 10.1101/2020.07.06.20140285. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7359534/ (visited on 07/03/2021).

[81] Laura Di Domenico et al. "Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies". In: *BMC Medicine* 18.1 (July 2020), p. 240. ISSN: 1741-7015. DOI: 10.1186/s12916-020-01698-4. URL: https://doi.org/10.1186/s12916-020-01698-4 (visited on 02/15/2021).

[82] *spatial spread covid countries - Recherche Google*. URL: https://www.google.com/search?q=spatial+spread+covid+countries&oq=spatial+spread+covid+countries&aqs=chrome..69i57.6265j0j4&sourceid=chrome&ie=UTF-8 (visited on 07/03/2021).

[83] Duygu Balcan et al. "Seasonal transmission potential and activity peaks of the new influenza A(H1N1): A Monte Carlo likelihood analysis based on human mobility". In: *BMC medicine* 7 (Oct. 2009), p. 45. DOI: 10.1186/1741-7015-7-45.

[84] Stefano Merler and Marco Ajelli. "The role of population heterogeneity and human mobility in the spread of pandemic influenza". In: *Proceedings of the Royal Society B: Biological Sciences* 277.1681 (Feb. 2010), pp. 557–565. ISSN: 0962-8452. DOI: 10.1098/rspb.2009.1605. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2842687/ (visited on 04/13/2021).

[85] Anne Ewing et al. "Contact, Travel, and Transmission: The Impact of Winter Holidays on Influenza Dynamics in the United States". In: *Journal of Infectious Diseases* 215 (Dec. 2016), jiw642. DOI: 10.1093/infdis/jiw642.

[86] Eugenio Valdano et al. "Using mobile phone data to reveal risk flow networks underlying the HIV epidemic in Namibia". en. In: *Nature Communications* 12.1 (May 2021). Number: 1 Publisher: Nature Publishing Group, p. 2837. ISSN: 2041-1723. DOI: 10.1038/s41467-021-23051-w. URL: https://www.nature.com/articles/s41467-021-23051-w (visited on 05/24/2021).

[87] Hsiao-Han Chang et al. "Mapping imported malaria in Bangladesh using parasite genetic and human mobility data". In: *eLife* 8 (Apr. 2019). Ed. by Neil M Ferguson, Edward A Wenger, and Oliver Brady. Publisher: eLife Sciences Publications, Ltd, e43481. ISSN: 2050-084X. DOI: 10.7554/eLife.43481. URL: https://doi.org/10.7554/eLife.43481 (visited on 05/07/2021).

[88] Jessica R. Floyd et al. "Malaria Parasite Mobility in Mozambique Estimated Using Mobile Phone Records". In: *ECAI 2020* (2020). Publisher: IOS Press, pp. 2901–2902. DOI: 10.3233/FAIA200444. URL: https://ebooks.iospress.nl/doi/10.3233/FAIA200444 (visited on 05/24/2021).

[89] Joël Mossong et al. "Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases". en. In: *PLOS Medicine* 5.3 (Mar. 2008). Publisher: Public Library of Science, e74. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.0050074. URL: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0050074 (visited on 02/22/2021).

[90] Shweta Bansal, Bryan T Grenfell, and Lauren Ancel Meyers. "When individual behaviour matters: homogeneous and network models in epidemiology". In: *Journal of The Royal Society Interface* 4.16 (Oct. 2007). Publisher: Royal Society, pp. 879–891. DOI: 10.1098/rsif.2007.1100. URL: https://royalsocietypublishing.org/doi/full/10.1098/rsif.2007.1100 (visited on 04/15/2021).

[91] Laura Fumanelli et al. "Inferring the Structure of Social Contacts from Demographic Data in the Analysis of Infectious Diseases Spread". en. In: *PLOS Computational Biology* 8.9 (2012). Publisher: Public Library of Science, e1002673. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002673. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002673 (visited on 04/16/2021).

[92] Juanjuan Zhang et al. "Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China". en. In: *Science* 368.6498 (June 2020). Publisher: American Association for the Advancement of Science Section: Report, pp. 1481–1486. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abb8001.

URL: https://science.sciencemag.org/content/368/6498/1481 (visited on 03/09/2021).

[93] Júlia Koltai et al. "Monitoring behavioural responses during pandemic via reconstructed contact matrices from online and representative surveys". In: *arXiv:2102.09021 [physics, stat]* (Feb. 2021). arXiv: 2102.09021. URL: http://arxiv.org/abs/2102.09021 (visited on 04/15/2021).

[94] Guillaume Béraud et al. "The French Connection: The First Large Population-Based Contact Survey in France Relevant for the Spread of Infectious Diseases". In: *PLoS ONE* 10.7 (July 2015). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0133203. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4503306/ (visited on 02/22/2021).

[95] Kiesha Prem, Alex R. Cook, and Mark Jit. "Projecting social contact matrices in 152 countries using contact surveys and demographic data". en. In: *PLOS Computational Biology* 13.9 (2017). Publisher: Public Library of Science, e1005697. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005697. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005697 (visited on 02/11/2021).

[96] Kim Van Kerckhove et al. "The Impact of Illness on Social Networks: Implications for Transmission and Control of Influenza". In: *American Journal of Epidemiology* 178.11 (Dec. 2013), pp. 1655–1662. ISSN: 0002-9262. DOI: 10.1093/aje/kwt196. URL: https://doi.org/10.1093/aje/kwt196 (visited on 04/15/2021).

[97] Giancarlo De Luca et al. "The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium". In: *BMC Infectious Diseases* 18.1 (Jan. 2018), p. 29. ISSN: 1471-2334. DOI: 10.1186/s12879-017-2934-3. URL: https://doi.org/10.1186/s12879-017-2934-3 (visited on 03/09/2021).

[98] Kiesha Prem et al. "The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study". en. In: *The Lancet Public Health* 5.5 (May 2020), e261–e270. ISSN: 2468-2667. DOI: 10.1016/S2468-2667(20)30073-6. URL: https://www.sciencedirect.com/science/article/pii/S2468266720300736 (visited on 03/25/2021).

[99] Valerio Gemmetto, Alain Barrat, and Ciro Cattuto. "Mitigation of infectious disease at school: targeted class closure vs school closure". In: *BMC Infectious Diseases* 14.1 (Dec. 2014), p. 695. ISSN: 1471-2334. DOI: 10.1186/s12879-014-0695-9. URL: https://doi.org/10.1186/s12879-014-0695-9 (visited on 05/28/2021).

[100] Shweta Bansal et al. "The dynamic nature of contact networks in infectious disease epidemiology". In: *Journal of Biological Dynamics* 4.5 (Sept. 2010). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17513758.2010.503376, pp. 478–489. ISSN: 1751-3758. DOI: 10.1080/17513758.2010.503376. URL: https://doi.org/10.1080/17513758.2010.503376 (visited on 06/24/2021).

[101] Dennis L. Chao, M. Elizabeth Halloran, and Ira M. Longini Jr. "School opening dates predict pandemic influenza A(H1N1) outbreaks in the United States". In: *The Journal of Infectious Diseases* 202.6 (Sept. 2010), pp. 877–880. ISSN: 0022-1899. DOI: 10.1086/655810. URL: https://doi.org/10.1086/655810 (visited on 05/28/2021).

[102] Shweta Bansal et al. "The Shifting Demographic Landscape of Pandemic Influenza". en. In: *PLOS ONE* 5.2 (Feb. 2010). Publisher: Public Library of Science, e9360. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0009360. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0009360 (visited on 07/05/2021).

[103] Anne Ewing et al. "Contact, Travel, and Transmission: The Impact of Winter Holidays on Influenza Dynamics in the United States". In: *The Journal of Infectious Diseases* 215.5 (Mar. 2017), pp. 732–739. ISSN: 0022-1899. DOI: 10.1093/infdis/jiw642. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5853779/ (visited on 04/15/2021).

[104] Shengjie Lai et al. "Effect of non-pharmaceutical interventions to contain COVID-19 in China". en. In: *Nature* 585.7825 (Sept. 2020). Number: 7825 Publisher: Nature Publishing Group, pp. 410–413. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2293-x. URL: https://www.nature.com/articles/s41586-020-2293-x (visited on 02/25/2021).

[105] Seth Flaxman et al. "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe". en. In: *Nature* 584.7820 (Aug. 2020). Number: 7820 Publisher: Nature Publishing Group, pp. 257–261. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2405-7. URL: https://www.nature.com/articles/s41586-020-2405-7 (visited on 03/25/2021).

[106] Serina Chang et al. "Mobility network models of COVID-19 explain inequities and inform reopening". en. In: *Nature* 589.7840 (Jan. 2021). Number: 7840 Publisher: Nature Publishing Group, pp. 82–87. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2923-3. URL: https://www.nature.com/articles/s41586-020-2923-3 (visited on 04/14/2021).

[107] Leo Ferres et al. "Measuring Levels of Activity in a Changing City". en. In: (), p. 11.

[108] *Assessing changes in commuting and individual mobility in major metropolitan areas in the United States during the COVID-19 outbreak*. URL: https://www.networkscienceinstitute.org/publications/assessing-changes-in-commuting-and-individual-mobility-in-major-metropolitan-areas-in-the-united-states-during-the-covid-19-outbreak (visited on 02/25/2021).

[109] *(PDF) The effect of human mobility and control measures on the COVID-19 epidemic in China*. URL: https://www.researchgate.net/publication/339773373_The_effect_of_human_mobility_and_control_measures_on_the_COVID-19_epidemic_in_China (visited on 02/25/2021).

[110] *M. Qian, A. Saunders, D. Pesch, S. Reece, W.D. Lee, X. Dong, R. Lambiotte. Regional Radius of Population movement*. URL: https://grapher.oxford-covid-19.com/grapher/gyration_2020-04-02 (visited on 02/25/2021).

[111] Kim Lyons. *Governments are using cellphone location data to manage the coronavirus*. en. Mar. 2020. URL: https://www.theverge.com/2020/3/23/21190700/eu-mobile-carriers-customer-data-coronavirus-south-korea-taiwan-privacy (visited on 02/25/2021).

[112] David Martín-Calvo et al. "Effectiveness of social distancing strategies for protecting a community from a pandemic with a data-driven contact network based on census and real-world mobility data". en. In: (), p. 13.

[113] Linus Bengtsson et al. "Using mobile phone data to predict the spatial spread of cholera". eng. In: *Scientific Reports* 5 (Mar. 2015), p. 8923. ISSN: 2045-2322. DOI: 10.1038/srep08923.

[114] Augustino Isdory, Eunice W. Mureithi, and David J. T. Sumpter. "The Impact of Human Mobility on HIV Transmission in Kenya". In: *PLoS ONE* 10.11 (Nov. 2015). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0142805. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4657931/ (visited on 01/16/2020).

[115] Emanuele Strano et al. "Mapping road network communities for guiding disease surveillance and control strategies". en. In: *Scientific Reports* 8.1 (Mar. 2018). Number: 1 Publisher: Nature Publishing Group, p. 4744. ISSN: 2045-2322. DOI: 10.1038/s41598-018-22969-4. URL: https://www.nature.com/articles/s41598-018-22969-4 (visited on 05/06/2021).

[116] *Cross-Border Mobility Responses to Covid-19 in Europe: New Evidence from Facebook Data*. en. May 2021. DOI: 10.21203/rs.3.rs-479253/v1. URL: https://www.researchsquare.com (visited on 05/14/2021).

[117] Vincent Blondel et al. "Data for Development: the D4D Challenge on Mobile Phone Data". In: (Sept. 2012).

[118] Sanja Brdar et al. "Unveiling Spatial Epidemiology of HIV with Mobile Phone Data". en. In: *Scientific Reports* 6.1 (Jan. 2016). Number: 1 Publisher: Nature Publishing Group, p. 19342. ISSN: 2045-2322. DOI: 10.1038/srep19342. URL: https://www.nature.com/articles/srep19342 (visited on 05/07/2021).

[119] Andrew M. Kramer et al. "Spatial spread of the West Africa Ebola epidemic". In: *Royal Society Open Science* 3.8 (Aug. 2016). ISSN: 2054-5703. DOI: 10.1098/rsos.160294. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5108957/ (visited on 05/07/2021).

[120] *COVID-19 Community Mobility Report*. URL: https://www.google.com/covid19/mobility?hl=it (visited on 02/19/2021).

[121] *Google.com. COVID-19 Community Mobility Report. COVID-19 Community Mobility Report https://www.google.com/covid19/mobility?hl=fr.* URL: https://www.google.com/covid19/mobility/ (visited on 02/13/2021).

[122] *Facebook Data for Good*. en-US. URL: https://dataforgood.fb.com/ (visited on 05/14/2021).

[123] Laura Di Domenico et al. "Modelling safe protocols for reopening schools during the COVID-19 pandemic in France". en. In: *Nature Communications* 12.1 (Feb. 2021). Number: 1 Publisher: Nature Publishing Group, p. 1073. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21249-6. URL: https://www.nature.com/articles/s41467-021-21249-6 (visited on 02/25/2021).

[124] Giulia Pullano et al. "Expected impact of exit strategies after the second lockdown - France, Nov 2020". en. In: (), p. 6.

[125] Giulia Pullano, Laura Di Domenico, and Vittoria Colizza. "Telework and other measures reducing the presence at work to slow down COVID-19pdm in France (Sept 2020)". en. In: (), p. 8.

[126] Amy Wesolowski et al. "Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data". In: *The Journal of Infectious Diseases* 214 (Dec. 2016), S414–S420. DOI: 10.1093/infdis/jiw273.

[127] Kyra H. Grantz et al. "The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology". en. In: *Nature Communications* 11.1 (Sept. 2020). Number: 1 Publisher: Nature Publishing Group, p. 4961. ISSN: 2041-1723. DOI: 10.1038/s41467-020-18190-5. URL: https://www.nature.com/articles/s41467-020-18190-5 (visited on 04/14/2021).

[128] *The Unacast Social Distancing Scoreboard*. en. URL: https://www.unacast.com/about#team (visited on 04/28/2021).

[129] Ramesh Raskar et al. "Adding Location and Global Context to the Google/Apple Exposure Notification Bluetooth API". en. In: *arXiv:2007.02317 [cs]* (July 2020). arXiv: 2007.02317. URL: http://arxiv.org/abs/2007.02317 (visited on 04/28/2021).

[130] Amy Wesolowski et al. "Heterogeneous Mobile Phone Ownership and Usage Patterns in Kenya". In: *PLoS ONE* 7.4 (Apr. 2012). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0035319. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3338828/ (visited on 01/16/2020).

[131] Andrew J. Tatem et al. "Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning". In: *Malaria Journal* 13 (Feb. 2014), p. 52. ISSN: 1475-2875. DOI: 10.1186/1475-2875-13-52. URL: https://doi.org/10.1186/1475-2875-13-52 (visited on 01/11/2018).

[132] Amy Wesolowski et al. "The Use of Census Migration Data to Approximate Human Movement Patterns across Temporal Scales". en. In: *PLOS ONE* 8.1 (2013). Publisher: Public Library of Science, e52971. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0052971. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0052971 (visited on 05/06/2021).

[133] Pierre Deville et al. "Dynamic population mapping using mobile phone data". en. In: *Proceedings of the National Academy of Sciences* 111.45 (Nov. 2014). Publisher: National Academy of Sciences Section: Physical Sciences, pp. 15888–15893. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1408439111. URL: https://www.pnas.org/content/111/45/15888 (visited on 05/06/2021).

[134] *General Data Protection Regulation (GDPR) – Official Legal Text*. en-US. URL: https://gdpr-info.eu/ (visited on 04/16/2021).

[135] *e-Privacy Regulation*. fr. Text. URL: https://edpb.europa.eu/our-work-tools/our-documents/topic/e-privacy-regulation_fr (visited on 04/16/2021).

[136] Yves-Alexandre Montjoye et al. "Unique in the Crowd: The Privacy Bounds of Human Mobility". In: *Scientific reports* 3 (Mar. 2013), p. 1376. DOI: 10.1038/srep01376.

[137] Antonio Lima et al. "Progmosis: Evaluating Risky Individual Behavior During Epidemics Using Mobile Network Data". In: *arXiv:1504.01316 [physics]* (Apr. 2015). arXiv: 1504.01316. URL: http://arxiv.org/abs/1504.01316 (visited on 05/07/2021).

[138] Chaogui Kang et al. "Exploring human movements in Singapore: A comparative analysis based on mobile phone and taxicab usages". In: vol. 1. Aug. 2013. DOI: 10.1145/2505821.2505826.

[139] Klaus Dietz and J. A. P. Heesterbeek. "Daniel Bernoulli's epidemiological model revisited". en. In: *Mathematical Biosciences* 180.1 (Nov. 2002), pp. 1–21. ISSN: 0025-5564. DOI: 10.1016/S0025-5564(02)00122-0. URL: https://www.sciencedirect.com/science/article/pii/S0025556402001220 (visited on 03/09/2021).

[140] William Heaton Hamer. *The Milroy Lectures on Epidemic Diseases in England: The Evidence of Variability and of Persistency of Type ; Delivered Before the Royal College of Physicians of London, March 1st, 6th, and 8th, 1906*. en. Google-Books-ID: ap09AAAAIAAJ. Bedford Press, 1906.

[141] William Ogilvy Kermack, A. G. McKendrick, and Gilbert Thomas Walker. "Contributions to the mathematical theory of epidemics. II. —The problem of endemicity". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 138.834 (Oct. 1932). Publisher: Royal Society, pp. 55–83. DOI: 10.1098/rspa.1932.0171. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1932.0171 (visited on 03/09/2021).

[142] Nicola Perra. "Non-pharmaceutical interventions during the COVID-19 pandemic: a rapid review". In: *arXiv:2012.15230 [physics]* (Dec. 2020). arXiv: 2012.15230. URL: http://arxiv.org/abs/2012.15230 (visited on 03/09/2021).

[143] Stefano Merler et al. "Pandemic Influenza A/H1N1pdm in Italy: Age, Risk and Population Susceptibility". en. In: *PLOS ONE* 8.10 (2013). Publisher: Public Library of Science, e74785. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0074785. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0074785 (visited on 04/26/2021).

[144] Pietro Coletti et al. *A data-driven metapopulation model for the Belgian COVID-19 epidemic: assessing the impact of lockdown and exit strategies*. en. preprint. Epidemiology, July 2020. DOI: 10.1101/2020.07.20.20157933. URL: http://medrxiv.org/lookup/doi/10.1101/2020.07.20.20157933 (visited on 03/09/2021).

[145] Andrea Apolloni, Chiara Poletto, and Vittoria Colizza. "Age-specific contacts and travel patterns in the spatial spread of 2009 H1N1 influenza pandemic". In: *BMC Infectious Diseases* 13.1 (Apr. 2013), p. 176. ISSN: 1471-2334. DOI: 10.1186/1471-2334-13-176. URL: https://doi.org/10.1186/1471-2334-13-176 (visited on 03/09/2021).

[146] Andrea Apolloni et al. "Metapopulation epidemic models with heterogeneous mixing and travel behaviour". en. In: *Theoretical Biology and Medical Modelling* 11.1 (Jan. 2014), p. 3. ISSN: 1742-4682. DOI: 10.1186/1742-4682-11-3. URL: https://doi.org/10.1186/1742-4682-11-3 (visited on 03/09/2021).

[147] E. REA et al. "Duration and distance of exposure are important predictors of transmission among community contacts of Ontario SARS cases". In: *Epidemiology and Infection* 135.6 (Aug. 2007), pp. 914–921. ISSN: 0950-2688. DOI: 10.1017/S0950268806007771. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2870656/ (visited on 04/15/2021).

[148] Jonas F. Ludvigsson. "Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults". en. In: *Acta Paediatrica* 109.6 (2020). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/apa.15270, pp. 1088–1095. ISSN: 1651-2227. DOI: https://doi.org/10.1111/apa.15270. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/apa.15270 (visited on 03/09/2021).

[149] Enrico Lavezzo et al. "Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'". en. In: *Nature* 584.7821 (Aug. 2020). Number: 7821 Publisher: Nature Publishing Group, pp. 425–429. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2488-1. URL: https://www.nature.com/articles/s41586-020-2488-1 (visited on 03/09/2021).

[150] Flavia Riccardo et al. "Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020". en. In: *Eurosurveillance* 25.49 (Dec. 2020). Publisher: European Centre for Disease Prevention and Control, p. 2000790. ISSN: 1560-7917. DOI: 10.2807/1560-7917.ES.2020.25.49.2000790. URL: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.49.2000790 (visited on 03/09/2021).

[151] Henrik Salje et al. "Estimating the burden of SARS-CoV-2 in France". en. In: *Science* 369.6500 (July 2020). Publisher: American Association for the Advancement of Science Section: Report, pp. 208–211. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abc3517. URL: https://science.sciencemag.org/content/369/6500/208 (visited on 03/09/2021).

[152] Nicholas G. Davies et al. "Age-dependent effects in the transmission and control of COVID-19 epidemics". en. In: *Nature Medicine* 26.8 (Aug. 2020). Number: 8 Publisher: Nature Publishing Group, pp. 1205–1211. ISSN: 1546-170X. DOI: 10.1038/s41591-020-0962-9. URL: https://www.nature.com/articles/s41591-020-0962-9 (visited on 03/09/2021).

[153] Petra Zimmermann and Nigel Curtis. "Coronavirus Infections in Children Including COVID-19". In: *The Pediatric Infectious Disease Journal* 39.5 (May 2020), pp. 355–368. ISSN: 0891-3668. DOI: 10.1097/INF.0000000000002660. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7158880/ (visited on 03/09/2021).

[154] Jiehao Cai et al. "A Case Series of children with 2019 novel coronavirus infection: clinical and epidemiological features". In: *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* (Feb. 2020). ISSN: 1058-4838. DOI: 10.1093/cid/ciaa198. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7108143/ (visited on 03/09/2021).

[155] Arnaud Fontanet et al. *SARS-CoV-2 infection in primary schools in northern France: A retrospective cohort study in an area of high transmission*. June 2020. DOI: 10.1101/2020.06.25.20140178.

[156] Arnaud Fontanet et al. "Cluster of COVID-19 in northern France: A retrospective closed cohort study". en. In: *medRxiv* (Apr. 2020). Publisher: Cold Spring Harbor Laboratory Press, p. 2020.04.18.20071134. DOI: 10.1101/2020.04.18.20071134. URL: https://www.medrxiv.org/content/10.1101/2020.04.18.20071134v1 (visited on 03/09/2021).

[157] Ruiyun Li et al. "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)". en. In: *Science* 368.6490 (May 2020). Publisher: American Association for the Advancement of Science Section: Research Article, pp. 489–493. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abb3221. URL: https://science.sciencemag.org/content/368/6490/489 (visited on 03/09/2021).

[158] *Insee - Institut national de la statistique et des études économiques*. URL: https://www.insee.fr/fr/accueil (visited on 02/24/2021).

[159] *Ministère de l'Education Nationale de la Jeunesse et des Sports. Déconfinement phase 2≤: point de situation au 28 mai. Ministère de l'Education Nationale et de la Jeunesse https://www.education.gouv.fr/deconfinement-phase-2-point-de-situation-au-28-mai-303813.* fr. URL: https://www.education.gouv.fr/deconfinement-phase-2-point-de-situation-au-28-mai-303813 (visited on 02/13/2021).

[160] *CoviPrev : une enquête pour suivre l'évolution des comportements et de la santé mentale pendant l'épidémie de COVID-19*. fr. URL: /etudes-et-enquetes/coviprev-une-enquÃłte-pour-suivre-lâĂŹÃľvolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie-de-covid-19 (visited on 03/09/2021).

[161] Joseph T. Wu and Benjamin J. Cowling. "The use of mathematical models to inform influenza pandemic preparedness and response". eng. In: *Experimental Biology and Medicine (Maywood, N.J.)* 236.8 (Aug. 2011), pp. 955–961. ISSN: 1535-3699. DOI: 10.1258/ebm.2010.010271.

[162] R. F. Grais et al. "Modeling the spread of annual influenza epidemics in the U.S.: the potential role of air travel". eng. In: *Health Care Management Science* 7.2 (May 2004), pp. 127–134. ISSN: 1386-9620. DOI: 10.1023/b:hcms.0000020652.38181.da.

[163] Joshua M. Epstein et al. "Controlling Pandemic Flu: The Value of International Air Travel Restrictions". en. In: *PLOS ONE* 2.5 (2007). Publisher: Public Library of Science, e401. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0000401. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000401 (visited on 03/09/2021).

[164] Ben S. Cooper et al. "Delaying the International Spread of Pandemic Influenza". en. In: *PLOS Medicine* 3.6 (2006). Publisher: Public Library of Science, e212. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.0030212. URL: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0030212 (visited on 03/09/2021).

[165] Lisa Sattenspiel and Klaus Dietz. "A structured epidemic model incorporating geographic mobility among regions". en. In: *Mathematical Biosciences* 128.1 (July 1995), pp. 71–91. ISSN: 0025-5564. DOI: 10.1016/0025-5564(94)00068-B. URL: https://www.sciencedirect.com/science/article/pii/002555649400068B (visited on 05/16/2021).

[166] Matt J. Keeling and Pejman Rohani. "Estimating spatial coupling in epidemiological systems: a mechanistic approach". en. In: *Ecology Letters* 5.1 (2002). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1461-0248.2002.00268.x, pp. 20–29. ISSN: 1461-0248. DOI: https://doi.org/10.1046/j.1461-0248.2002.00268.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1461-0248.2002.00268.x (visited on 05/16/2021).

[167] Benjamin Bolker and Bryan Thomas Grenfell. "Space, persistence and dynamics of measles epidemics". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 348.1325 (May 1995). Publisher: Royal Society, pp. 309–320. DOI: 10.1098/rstb.1995.0070. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.1995.0070 (visited on 05/16/2021).

[168] A. L. Lloyd and R. M. May. "Spatial heterogeneity in epidemic models". eng. In: *Journal of Theoretical Biology* 179.1 (Mar. 1996), pp. 1–11. ISSN: 0022-5193. DOI: 10.1006/jtbi.1996.0042.

[169] D. J. Earn, P. Rohani, and B. T. Grenfell. "Persistence, chaos and synchrony in ecology and epidemiology". eng. In: *Proceedings. Biological Sciences* 265.1390 (Jan. 1998), pp. 7–10. ISSN: 0962-8452. DOI: 10.1098/rspb.1998.0256.

[170] P. Rohani, D. J. Earn, and B. T. Grenfell. "Opposite patterns of synchrony in sympatric disease metapopulations". eng. In: *Science (New York, N.Y.)* 286.5441 (Oct. 1999), pp. 968–971. ISSN: 0036-8075. DOI: 10.1126/science.286.5441.968.

[171] Matt J. Keeling. "Metapopulation moments: coupling, stochasticity and persistence". en. In: *Journal of Animal Ecology* 69.5 (2000). _eprint: https://besjournals.onlinelibrary.wiley 2656.2000.00430.x, pp. 725–736. ISSN: 1365-2656. DOI: https://doi.org/10.1046/j.1365-2656.2000.00430.x. URL: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2656.2000.00430.x (visited on 05/16/2021).

[172] John R. Giles et al. "The duration of travel impacts the spatial dynamics of infectious diseases". In: *Proceedings of the National Academy of Sciences* 117.36 (Sept. 2020), pp. 22572–22579.

[173] Chiara Poletto, Michele Tizzoni, and Vittoria Colizza. "Heterogeneous length of stay of hosts' movements and spatial epidemic spread". en. In: *Scientific Reports* 2.1 (June 2012). Number: 1 Publisher: Nature Publishing Group, p. 476. ISSN: 2045-2322. DOI: 10.1038/srep00476. URL: https://www.nature.com/articles/srep00476 (visited on 05/03/2021).

[174] Chiara Poletto, Michele Tizzoni, and Vittoria Colizza. "Human mobility and time spent at destination: Impact on spatial epidemic spreading". en. In: *Journal of Theoretical Biology* 338 (Dec. 2013), pp. 41–58. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2013.08.032. URL: https://www.sciencedirect.com/science/article/pii/S0022519313004062 (visited on 06/28/2021).

[175] Martin Rosvall et al. "Memory in network flows and its effects on spreading dynamics and community detection". en. In: *Nature Communications* 5.1 (Aug. 2014). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Applied physics;Statistical physics Subject_term_id: applied-physics;statistical-physics, p. 4630. ISSN: 2041-1723. DOI: 10.1038/ncomms5630. URL: https://www.nature.com/articles/ncomms5630 (visited on 06/28/2021).

[176] Matt J. Keeling et al. "Individual identity and movement networks for disease metapopulations". en. In: *Proceedings of the National Academy of Sciences* 107.19 (May 2010). Publisher: National Academy of Sciences Section: Biological Sciences, pp. 8866–8870. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1000416107. URL: https://www.pnas.org/content/107/19/8866 (visited on 03/09/2021).

[177] B. T. Grenfell, O. N. Bjørnstad, and J. Kappey. "Travelling waves and spatial hierarchies in measles epidemics". en. In: *Nature* 414.6865 (Dec. 2001). Number: 6865 Publisher: Nature Publishing Group, pp. 716–723. ISSN: 1476-4687. DOI: 10.1038/414716a. URL: https://www.nature.com/articles/414716a (visited on 03/09/2021).

[178] I. M. Hall et al. "Real-time epidemic forecasting for pandemic influenza". eng. In: *Epidemiology and Infection* 135.3 (Apr. 2007), pp. 372–385. ISSN: 0950-2688. DOI: 10.1017/S0950268806007084.

[179] Fănică Gavril. "Generating the maximum spanning trees of a weighted graph". en. In: *Journal of Algorithms* 8.4 (Dec. 1987), pp. 592–597. ISSN: 0196-6774. DOI: 10.1016/0196-6774(87)90053-8. URL: https://www.sciencedirect.com/science/article/pii/0196677487900538 (visited on 05/17/2021).

[180] Claire Donnat and Susan Holmes. "Tracking network dynamics: a survey of distances and similarity metrics". In: *arXiv:1801.07351 [physics, stat]* (Mar. 2018). arXiv: 1801.07351. URL: http://arxiv.org/abs/1801.07351 (visited on 01/16/2020).

[181] *EpiRisk*. URL: https://epirisk.net/ (visited on 02/19/2021).

[182] *ANSD. Agence Nationale de Statistique et de la Démographie*. URL: https://www.ansd.sn/ (visited on 05/10/2021).

[183] Ousmane Thiam. "L'axe Dakar-Touba (Sénégal) : analyse spatiale d'un corridor urbain émergent". In: (Mar. 2008).

[184] Pietro Coletti et al. "Shifting patterns of seasonal influenza epidemics". In: *Scientific Reports* 8 (Dec. 2018). DOI: 10.1038/s41598-018-30949-x.

[185] M. U. G. Kraemer et al. "Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings". eng. In: *Scientific Reports* 9.1 (Mar. 2019), p. 5151. ISSN: 2045-2322. DOI: 10.1038/s41598-019-41192-3.

[186] *Timeline of ECDC's reponse to COVID-19. https://www.ecdc.europa.eu/en/covid-19/timeline-ecdc-respons*. en. URL: https://www.ecdc.europa.eu/en/covid-19/timeline-ecdc-response (visited on 02/16/2021).

[187] Francesco Pinotti et al. "Tracing and analysis of 288 early SARS-CoV-2 infections outside China: A modeling study". en. In: *PLOS Medicine* 17.7 (2020). Publisher: Public Library of Science, e1003193. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1003193. URL: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1003193 (visited on 02/10/2021).

[188] Rene Niehus et al. "Using observational data to quantify bias of traveller-derived COVID-19 prevalence estimates in Wuhan, China". en. In: *The Lancet Infectious Diseases* 20.7 (July 2020), pp. 803–808. ISSN: 1473-3099. DOI: 10.1016/S1473-3099(20)30229-2. URL: https://www.sciencedirect.com/science/article/pii/S1473309920302292 (visited on 03/30/2021).

[189] Tigist F. Menkir et al. "Estimating internationally imported cases during the early COVID-19 pandemic". en. In: *Nature Communications* 12.1 (Jan. 2021). Number: 1 Publisher: Nature Publishing Group, p. 311. ISSN: 2041-1723. DOI: 10.1038/s41467-020-20219-8. URL: https://www.nature.com/articles/s41467-020-20219-8 (visited on 02/16/2021).

[190] *SPF. COVID-19: une enquête pour suivre l'évolution des comportements et de la santé mentale pendant l'épidémie. https://www.santepubliquefrance.fr/etudes-et-enquetes/covid-19-une-enquete-pour-suivre-l-evolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie*. fr. URL: /etudes-et-enquetes/coviprev-une-enquÃłte-pour-suivre-lâĂŹÃľvolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie-de-covid-19 (visited on 02/13/2021).

[191] Steven Abrams et al. "Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories". en. In: *Epidemics* 35 (June 2021), p. 100449. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2021.100449. URL: https://www.sciencedirect.com/science/article/pii/S1755436521000116 (visited on 07/08/2021).

[192] Rajesh Singh and R. Adhikari. "Age-structured impact of social distancing on the COVID-19 epidemic in India". In: *arXiv:2003.12055 [cond-mat, q-bio]* (Mar. 2020). arXiv: 2003.12055. URL: http://arxiv.org/abs/2003.12055 (visited on 05/28/2021).

[193] Chiara E Sabbatini et al. "Estimated date of dominance of VOC-202012/01 strain in France and projected scenarios". en. In: (), p. 6.

[194] *COVID-19 -Mobility Trends Reports.* it-it. URL: https://www.apple.com/covid19/mobility (visited on 08/09/2021).

[195] Giulia Pullano et al. "Underdetection of cases of COVID-19 in France threatens epidemic control". en. In: *Nature* 590.7844 (Feb. 2021). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7844 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Infectious diseases;Nonlinear phenomena;SARS-CoV-2 Subject_term_id: infectious-diseases;nonlinear-phenomena;sars-cov-2, pp. 134–139. ISSN: 1476-4687. DOI: 10.1038/s41586-020-03095-6. URL: https://www.nature.com/articles/s41586-020-03095-6 (visited on 08/09/2021).