Journal of Computational Science 1 (2010) 132–145

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model

Duygu Balcan[a,b], Bruno Gonçalves[a,b], Hao Hu[c], José J. Ramasco[d], Vittoria Colizza[d], Alessandro Vespignani[a,b,d,*]

[a] Center for Complex Networks and Systems Research (CNetS), School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA
[b] Pervasive Technology Institute, Indiana University, Bloomington, IN 47406, USA
[c] Department of Physics, Indiana University, Bloomington, IN 47406, USA
[d] Computational Epidemiology Laboratory, Institute for Scientific Interchange (ISI), Torino, Italy

ARTICLE INFO

Article history:
Received 7 May 2010
Received in revised form 13 July 2010
Accepted 13 July 2010

Keywords:
Computational epidemiology
Complex networks
Multiscale phenomena
Human mobility
Infectious diseases

ABSTRACT

Here we present the Global Epidemic and Mobility (GLEaM) model that integrates sociodemographic and population mobility data in a spatially structured stochastic disease approach to simulate the spread of epidemics at the worldwide scale. We discuss the flexible structure of the model that is open to the inclusion of different disease structures and local intervention policies. This makes GLEaM suitable for the computational modeling and anticipation of the spatio-temporal patterns of global epidemic spreading, the understanding of historical epidemics, the assessment of the role of human mobility in shaping global epidemics, and the analysis of mitigation and containment scenarios.

1. Introduction

The increasing computational and data integration capabilities witnessed in recent years have enabled the development of computational epidemic models of great complexity and realism [36]. Generally accepted methodologies are represented by very detailed agent-based models [17,33,18,19,24,8,34] and large-scale spatial metapopulation models [38,21,25,29,12,16,9,1,2]. These two major classes of computational models have different resolutions and limitations. Agent-based models are stochastic, spatially explicit, discrete-time, simulation models where the agents represent single individuals. The infection can spread among individuals by contacts within household members, within school and workplace colleagues and by random contacts in the general population. One of the key features of the model is the characterisation of the network of contacts among individuals based on a realistic model of the sociodemographic structure of the population (see for instance [27] for a comparison between several models based on this approach). The second scheme relies on metapopulation structured models that considers the system divided into geographical regions defining a subpopulation network where connections among subpopulations represent the individual fluxes due to the transportation and mobility infrastructures [1–3,10,11]. Infection dynamics occurs inside each subpopulation and is described by compartmental schemes that depend on the specific etiology of the disease and the containment interventions considered [38,21]. Agent-based models provide a very rich data scenario but the computational cost and most importantly the need for very detailed input data has limited their use to a few country level scenarios so far [27], up to continent level [34]. On the opposite side, the structured metapopulation models are fairly scalable and can be conveniently used to provide world-wide scenarios and patterns with thousands of stochastic realizations [29,12,16,9,1,2,22]. While on one hand, the level of information that can be extracted in structured metapopulation models is less detailed than those of agent-based models, on the other hand, their computational scalability allows the simulation of disease spreading on the worldwide scale and the use of statistical approaches that leverage on Monte Carlo techniques based on the analysis of a large number of simulation runs exploring the parameter space.

In this paper, we provide a detailed presentation of the Global Epidemic and Mobility (GLEaM) model [2] that uses a structured metapopulation scheme integrating the stochastic modeling of the disease dynamics, high resolution census data worldwide and

* Corresponding author at: Center for Complex Networks and Systems Research (CNetS), School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA.
E-mail addresses: balcand@indiana.edu (D. Balcan), bgoncalv@indiana.edu (B. Gonçalves), hahu@indiana.edu (H. Hu), jramasco@isi.it (J.J. Ramasco), vcolizza@isi.it (V. Colizza), alexv@indiana.edu (A. Vespignani).

1877-7503/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jocs.2010.07.002

human mobility patterns at the global scale. GLEaM makes use of high resolution population data [6,7] that allow for the definition of subpopulations according to a Voronoi decomposition of the world surface centered on the locations of major transportation hubs. This procedure leads to the construction of a metapopulation model consisting of more than 3300 subpopulations across the world connected through a network of more than 16,800 mobility fluxes describing the daily patterns of travel and mobility among subpopulations. In particular GLEaM integrates data obtained from the International Air Transport Association (IATA [30]) and Official Airline Guide (OAG [35]) databases and multimodal mobility data collected and analyzed from more than 30 countries in 5 different continents. This integration results in a worldwide multiscale mobility network spanning several orders of magnitude in intensity and spatio-temporal scales. The disease dynamics is simulated by a fully stochastic compartmental approach defining the temporal equations for each subpopulation [1]. The equations of different subpopulations are then coupled through effective interactions and mechanistic schemes accounting for the mobility of individuals encoded in the multiscale mobility network.

The GLEaM computational model trades off the high realism of agent-based models for the computational scalability of the algorithm implementation and the relatively small amount of input data needed to initialize the model. This allows detailed analysis of epidemic patterns at the worldwide scale. This feature is extremely relevant in evaluating the time pattern of emerging infectious diseases, and cannot be accounted for by agent-based models restricted to country or continent level. For instance, given a set of initial conditions for a local outbreak of a new strain of influenza, the timeline of the arrival of the epidemic in each country and the ensuing activity peak are mainly determined by the human mobility network that couples different regions of the world. By looking at individual countries or a given continent in isolation, any estimate of the epidemic timeline is based on assumptions about imported cases from the rest of the world. This is obtained without an explicit coupling or knowledge of the propagation of the disease in the system outside the boundaries of the country or the continent that is the focus of the model. GLEaM instead explicitly integrates human mobility patterns that allow us to consistently simulate the mobility of infectious individuals on the global scale thus providing *ab initio* estimates of the epidemic timeline in each country or urban area without assumptions on case importation.

Differently from agent-based models, the scalability of GLEaM has also the advantage of making possible the use of statistical methods such as Monte Carlo likelihood analysis to fit epidemic parameters which are usually not known in the case of new emerging diseases, with the aim of understanding the observed pattern and simulate its possible future spread [1]. This is enabled by the possibility of generating large numbers of *in silico* epidemics to allow the self-consistent estimate of all the parameters needed for the simulation of the future propagation of the disease. A large number of computational runs is indeed needed to systematically explore the space of parameters and, for each point in such space, to build a robust statistical ensemble and reduce the fluctuations induced by stochastic effects. The intensive CPU requirements of agent-based models limit the feasibility of large explorations of the space of parameters aimed at estimation procedures, or at performing sensitivity analysis on the parameters included in the models to assess effects in the simulated results induced by their changes [27]. This constraint becomes particularly relevant in the case computational models are used as risk-assessment tools for scenario evaluations of an epidemic emergency in real time.

Here we specify the definition and integration of the different data layers composing the model, and also provide a detailed explanation of the Voronoi tessellation used for the subpopulation definition. The construction of the mobility network and the derivation of the stochastic mobility equations among different subpopulations are described in detail as well. We illustrate the time-scale separation technique that allows for the integration of the mobility processes occurring on small time scales as effective coupling terms. This method reduces the computational cost by simulating in an explicit way only mobility processes occurring on the long time scales. The metapopulation structure and the mobility processes are then integrated in the basic equations describing the time behavior of the disease process within each population. We detail the structure of the equations in the specific case of an influenza-like-illness compartmentalization, although the equations can be generalized to generic compartmental structures according to the disease of interest. The second part of the paper is devoted to the algorithmic implementation of the model. We describe the algorithm structure, inputs and outputs that allow GLEaM to perform the simulation of stochastic realizations of the worldwide unfolding of the epidemic. From these *in silico* epidemics a variety of information can be gathered, such as prevalence, morbidity, number of secondary cases, number of imported cases, hospitalized patients, amounts of drugs used, and other quantities for each subpopulation with a minimal time resolution of 1 day. Finally we provide an example of the results that can be obtained with GLEaM by simulating the 2001–2002 seasonal influenza spreading and comparing the computational results with real data from different surveillance infrastructures.

2. Related work

Many data-driven epidemic models have been proposed, however only a few, mostly based on metapopulation schemes, tackle the spatio-temporal behavior of diseases at the global scale. Agent-based models are to be able to consider individually targeted interventions for the mitigation of an epidemic, as well as the possibility to introduce changes of behavior at the individual level reproducing the adaptation of individuals to the disease spread. This is performed by tracking each agent of the artificial society considered in the model, and applying rules for the behavior of individuals in their virtual space. Therefore, most agent-based models can be very accurate in the description of the spread of a disease in time and spatial scales if it is possible to integrate high quality data at the individual agent level. The difficulties in gathering high quality data worldwide and to the limit imposed by high performance computing, however have restricted the application of agent-based models to local populations or a few countries – such as e.g., the US [24,19,27], the UK [19], Italy [8], Thailand [33,18] – up to the continent of Europe [34]. Among the metapopulation schemes at the global level available in the literature [29,12,16,9,1,2,22], the main differences lie in the accuracy and completeness of the demographic and mobility layers. Indeed, being based on simple homogeneous assumptions inside each subpopulation, the accuracy and realism of these models are found in their ability to capture the distribution of population and the travel flows of individuals from one subpopulation to another. With the airline transportation system being the main and fastest mean of connection between different parts of the world, previous works have included an always increasing portion of the worldwide airport network in the metapopulation approaches considered. Indeed, even in continental Europe that possesses one of the most structured and modern railway network, long-range railway traffic across countries is just one-tenth of the corresponding airline traffic [14]. From samples with 52 airports in Ref. [38,22], 105 airports in Ref. [12], 155 in Ref. [16], 500 in Ref. [29], up to the complete International Air Transport Association (IATA) [30] and Official Airline Guide (OAG [35]) databases incorporated in GLEaM [9,2]. Samples of the worldwide airport network usually correspond to the largest airports, the

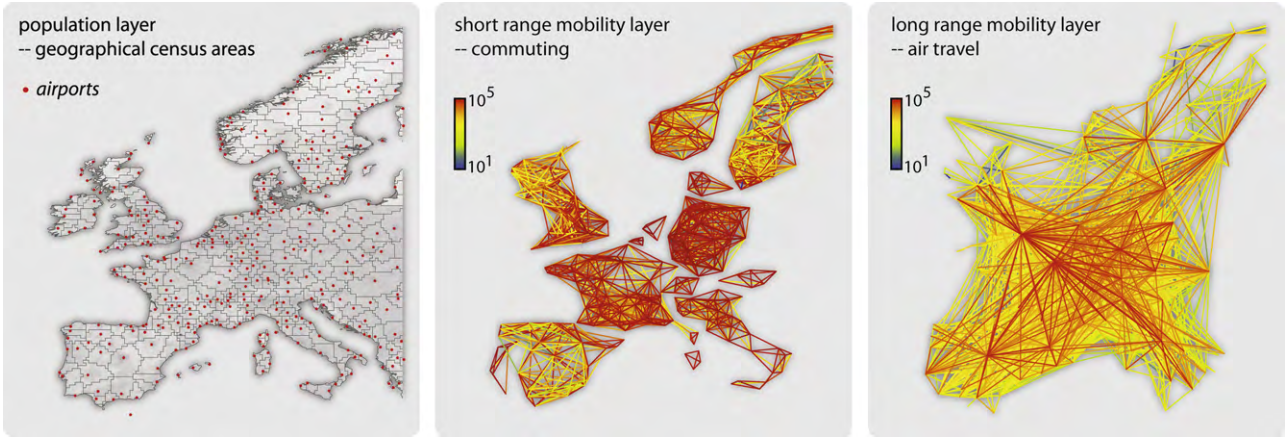


Fig. 1. GLEaM, GLobal Epidemic and Mobility model. The world surface is represented in a grid-like partition where each cell – corresponding to a population value – is assigned to the closest airport. Geographical census areas emerge that constitute the subpopulations of the metapopulation model. The demographic layer is coupled with two mobility layers, the short range commuting layer and the long range air travel layer.

most connected cities, or the most central ones, and therefore they may include a large portion of the total commercial traffic. While including the largest flows of real-world mobility, these samples are limited in their ability to capture the entire network information for a detailed description of the geotemporal evolution of the disease on a city by city basis. The overall paths of spreading may be fairly well reproduced [4], but models based on samples would fail if the question under study focuses on the description of the epidemic behavior at a higher level of detail, such as e.g., country or city level, due to the lack of data on connections and travel fluxes. In addition, the accuracy in reproducing the spreading pattern of diseases is largely challenged by the absence of large fluctuations in the topology of the airline network and in the traffic volumes, and of correlations and non-trivial loops that are responsible for the definition of the geotemporal propagation in the real world [9]. The increase of resolution imposes different requirements in the definition of the population distribution and of additional means of transportation that may become relevant at this level of detail. Previous works considered cities with no geographical reference whose population was obtained from national and international city population databases [29,12,16,9,22], and did not consider coupling effects other than air transportation. The GLEaM computational model presented here takes into account also the short range mobility to capture the daily population displacements from a given geographical census area to its neighboring one. In addition, the model already integrates long-range railway connections indexed by the OAG database and we are making a progressive introduction of detailed railway networks in specific countries. By integrating a multi-scale mobility layer, GLEaM is therefore the world-wide model that consider a finer description of the evolution of the epidemic behavior, with the air travel dictating the pathways of the disease through the large geographical areas, whereas the daily short-range displacements control the timing of spreading within localized regions [2].

3. GLEaM computational model definition

The global epidemic and mobility structured metapopulation (GLEaM) model is based on a metapopulation approach in which the world is divided into geographical regions defining a subpopulation network where connections among subpopulations represent the individual fluxes due to the transportation and mobility infrastructure. GLEaM integrates three different data layers (see Fig. 1). The population layer is based on the high-resolution population database of the "Gridded Population of the World" project of Columbia University [6,7] that estimates the population with a granularity given by a lattice of cells covering the whole planet at a resolution of 15 min × 15 min of arc. The transportation mobility layer integrates air travel mobility obtained from the International Air Transport Association (IATA) [30] and OAG [35] databases that contain the list of worldwide airport pairs connected by direct flights and the number of available seats on any given connection, and commuting patterns as obtained from data collected and analyzed from more than 30 countries in 5 continents. The combination of the population and mobility layers allows for the subdivision of the world into georeferenced census areas defined with a Voronoi tessellation procedure around transportation hubs. GLEaM simulates the mobility of individuals from one subpopulation to another by a stochastic procedure in which the number of passengers of each compartment traveling from a subpopulation j to a subpopulation ℓ is an integer random variable defined by a stochastic process defined on the basis of real mobility data. Short range commuting between subpopulations is modeled with a time scale separation approach that defines the effective force of infections in connected subpopulations. Superimposed on the worldwide population and mobility layers is the epidemic model that defines the disease and population dynamics. The infection dynamics takes place within each subpopulation and assumes the classic compartmentalization in which each individual is classified by one of the discrete states such as susceptible, latent, infectious symptomatic, infectious non-symptomatic or permanently recovered/removed. In the following sections we provide a detailed presentation of each data layer and of the basic equations that defines the computational model.

3.1. Population layer

The dataset of the "Gridded Population of the World" and the "Global Urban-Rural Mapping" projects [6,7] run by the Socioeconomic Data and Application Center (SEDAC) of Columbia University divides the surface of the world into a grid of cells that can have different resolution levels. Each of these cells has assigned an estimated population value. Out of the possible resolutions, we have opted for cells of 15 min × 15 min of arc to constitute the basis of our model. This corresponds to an area of each cell approximately equivalent to a rectangle of 25 km × 25 km along the Equator. The dataset comprises 823,680 cells, of which 250,206 are populated. In order to define the subpopulations that constitute the metapopulation structure of our model we have performed a Voronoi-like tessellation of the Earth surface centered around the airports of the IATA database. In particular, we identify 3362 subpopulations

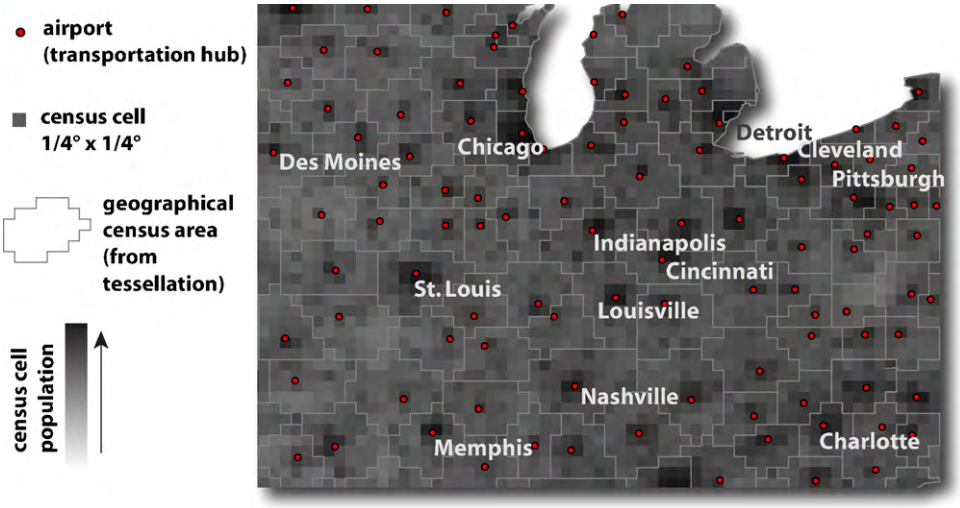


Fig. 2. Population database and Voronoi tessellation around main transportation hubs. The world surface is represented in a grid-like partition where each cell – corresponding to a population values – is assigned to the closest airport. Geographical census areas emerge that constitute the subpopulations of the metapopulation model.

centered around indexed IATA airports in 220 different countries. Since the coordinates of each cell center and those of the airports are known, the distance between the cells and the airports can be calculated. We assign each cell to the subpopulation associated to the closest airport that satisfies the following two conditions: (i) each cell is assigned to the closest airport within the same country and (ii) the distance between the airport and the cell does not exceed 200 km. This cutoff naturally emerges from the distribution of distances between cells and closest airports, and it is introduced to avoid that in barely populated areas such as Siberia we can generate geographical census areas thousands of kilometer wide but with almost no population. It also corresponds to a reasonable upper cutoff for the ground traveling distance expected to be covered to reach an airport before traveling by plane.

In addition, the tessellation procedure needs to take into account that there exist urban areas served by more than one airport. Examples include London with up to six airports, Paris with two, New York City with three and others. This condition is relevant in the tessellation, as the aim of the procedure is to provide geographical census areas that will correspond to the subpopulation of the metapopulation model, where homogeneous mixing is going to be assumed. Given that the mixing between individuals in a given urban area is expected to be high, independently from their choice of the airport for mobility reasons, we first need to proceed to the aggregation of the groups of airports that serve the same urban area, prior to tessellation. We have searched for groups of airports located close to each other and manually processed the identified groups to select those belonging to the same urban area. The airports of the same group are then aggregated in a single "super-hub". An example with the final result of the Voronoi tessellation procedure with cells and airports can be seen in Fig. 2.

3.2. Mobility layers

The geographical census areas obtained with the tessellation procedure define the basic subpopulations of the GLEaM metapopulation structure. The spatio-temporal patterns of the disease spreading are however associated to the mobility flows that couple different subpopulations. These flows constitute the mobility data layer that is represented as a network of connections among subpopulations that identifies the number of individuals that goes from one subpopulation to the others. The mobility network is made by different kind of mobility processes from short-range commuting to intercontinental flights with time-scale and traffic volumes that span several orders of magnitude. In the following we discuss the data integration process and the construction of this multiscale mobility network.

3.2.1. Worldwide Airport Network

The Worldwide Airport Network (WAN) is composed of 3362 commercial airports indexed by the IATA located in 220 different countries. The database contains the number of available seats per year for each direct connection between a pair of these airports. The coverage of the dataset is estimated to be 99% of the global commercial traffic. The WAN can be seen as a weighted graph comprising 16,846 edges whose weight, $\omega_{j\ell}$, represents the passenger flow between airports j and ℓ. The network shows a high degree of heterogeneity both in the number of destinations per airport and in the number of passengers per connection [9,3,10,11].

3.2.2. Commuting networks

Our commuting databases have been collected from the Offices of Statistics of 30 countries in 5 continents. The full dataset comprehends more than 80,000 administrative regions and over five million commuting flow connections between them (see [2]). The definition of administrative unit and the granularity level at which the commuting data are provided vary enormously from country to country. For example, most European countries adhere to a practice that ranks administrative divisions in terms of geocoding for statistical purposes, the so called Nomenclature of Territorial Units for Statistics (NUTS) going from level 1 to 3 plus the Local Administrative Units (LAU) corresponding to the municipalities and that can be further subdivided in Wards (LAU 2). In most of the cases, we obtained the commuting data at the LAU level 1 or 2. The US or Canada, on the other hand, have different standards and report commuting at the level of counties. Not only there are clear differences across countries in the definition of the administrative divisions, but even within the same country the actual extension, shape, and population of the administrative divisions can be strongly heterogeneous, being a result of historical and administrative reasons (Table 1).

In order to overcome the differences in spatial resolution of the commuting data across different countries, we define a worldwide homogeneous standard for GLEaM. We used the geographical census areas obtained from the Voronoi tessellation as the elementary units to define the centers of gravity for the process of

Table 1
Commuting networks in each continent. Number of countries (N), number of administrative units (V) and inter-links between them (E) are summarized.

Continent	N	V	E
Europe	17	65,880	4,490,650
North America	2	6986	182,255
Latin America	5	4301	102,117
Asia	4	4355	380,385
Oceania	2	746	30,679
Total	30	82,268	5,186,186

Table 2
Transitions between compartments and their rates.

Transition	Type	Rate
$S_j \to L_j$	Contagion	λ_j
$L_j \to I_j^a$	Spontaneous	εp_a
$L_j \to I_j^t$		$\varepsilon(1-p_a)p_t$
$L_j \to I_j^{nt}$		$\varepsilon(1-p_a)(1-p_t)$
$I_j^a \to R_j$		μ
$I_j^t \to R_j$		μ
$I_j^{nt} \to R_j$		μ

commuting. This allows to deal with self-similar units across the world with respect to mobility as emerged from the tessellation and not country specific administrative boundaries. We have therefore mapped the different levels of commuting data into the geographical census areas formed by the Voronoi-like tessellation procedure described above. The mapped commuting flows can be seen as a second transport network connecting subpopulations that are geographically close. This second network can be overlaid to the WAN in a multi-scale fashion to simulate realistic scenarios for disease spreading. The network exhibits important variability in the number of commuters on each connection as well as in the total number of commuters per geographical census area. Being the census areas statistically homogeneous we can also extract a general statistical law that allows for the synthetic generation of commuting networks in countries where real data are not available. A full account of the commuting data obtained across different continents and their statistical analysis can be found in Ref. [2].

3.3. Disease model

Each geographical census area corresponds to a subpopulation in the metapopulation model. The infection dynamics within each subpopulation is governed by a disease specific compartmental model in which we assume homogeneous mixing in the population. Although the model can use any compartmental structure, for the sake of clarity we will carry on our discussion by using the explicit example of a typical influenza-like illness (ILI) where we consider a Susceptible-Latent-Infectious-Recovered (SLIR) compartmental scheme. In Fig. 3, a diagram of the compartmental structure with transitions between compartments is shown. The contagion process, i.e., generation of new infections, is the only transition mechanism which is altered by short-range mobility, whereas all the other transitions between compartments are spontaneous and remain unaffected by the commuting. The rate at which a susceptible individual in subpopulation j acquires the infection, the so called force of infection λ_j, is determined by interactions with infectious persons either in the home subpopulation j or in its neighboring subpopulations on the commuting network. In general, the force of infection is assumed to follow the mass action principle for which the infection rate is $\lambda = \beta I/N$ where β is the infection transmission rate and I/N is the density of infected individuals in the population. In the case of asymptomatic individuals the force of infection is usually reduced by a factor r_β. In the case of multiple interacting subpopulations and different classes of infectives the force of infection will be the sum of different contributions as reported in Section 4.3.

Given the force of infection λ_j in subpopulation j, each person in the susceptible compartment (S_j) contracts the infection with probability $\lambda_j \Delta t$ and enters the latent compartment (L_j), where Δt is the time interval considered. Latent individuals exit the compartment with probability $\varepsilon \Delta t$, and transit to asymptomatic infectious compartment (I_j^a) with probability p_a or, with the complementary probability $1 - p_a$, become symptomatic infectious. Infectious persons with symptoms are further divided between those who can travel (I_j^t), probability p_t, and those who are travel-restricted (I_j^{nt}) with probability $1 - p_t$. All the infectious persons permanently recover with probability $\mu \Delta t$, entering the recovered compartment (R_j) in the next time step. All transitions and corresponding rates are summarized in Table 2 and in Fig. 3.

4. Epidemic and mobility dynamics

Once the mobility data layers and the disease dynamics has been defined, the number of individuals in each compartment $[m]$ and subpopulation j follows a discrete and stochastic dynamical equation that reads as

$$X_j^{[m]}(t + \Delta t) - X_j^{[m]}(t) = \Delta X_j^{[m]} + \Omega_j([m]) \quad (1)$$

where the term $\Delta X_j^{[m]}$ represents the change due to the compartment transitions induced by the disease dynamics and the transport operator $\Omega_j([m])$ represents the variations due to the traveling and mobility of individuals. The latter operator takes into account the long-range airline mobility and sets the minimal time scale of integration at 1 day. The mobility due to the commuting flows is

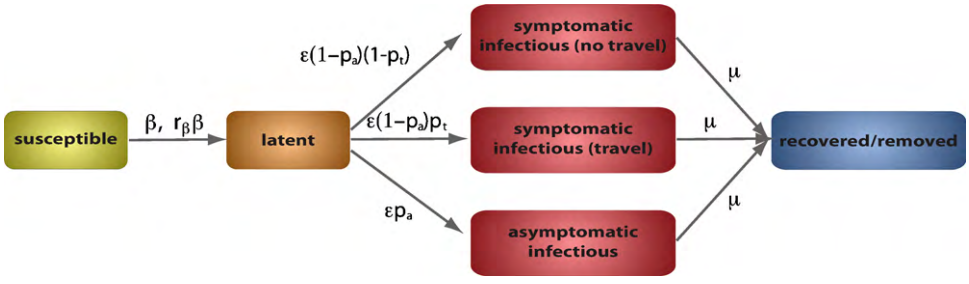


Fig. 3. Compartmental structure of the epidemic model within each subpopulation. A susceptible individual in contact with a symptomatic or asymptomatic infectious person contracts the infection at rate β or $r_\beta\beta$, respectively, and enters the latent compartment where he is infected but not yet infectious. At the end of the latency period ε^{-1}, each latent individual becomes infectious, entering the symptomatic compartments with probability $1 - p_a$ or becoming asymptomatic with probability p_a. The symptomatic cases are further divided between those who are allowed to travel (with probability p_t) and those who would stop traveling when ill (with probability $1 - p_t$). Infectious individuals recover permanently with rate μ. All transition processes are modeled through multinomial processes.

included in the model by an effective force of infection obtained using a time scale separation approximation as detailed in the following sections. The term $\Delta X_j^{[m]}$ can be written as a combination of a set of operators $\mathcal{D}_j([m],[n])$. Each $\mathcal{D}_j([m],[n])$ determines the number of transitions from compartment $[m]$ to $[n]$ occurring in Δt and is simulated as a random variable extracted from a multinomial distribution. The change $\Delta X_j^{[m]}$ is then given by the sum

$$\Delta X_j^{[m]} = \sum_{[n]} \{-\mathcal{D}_j([m],[n]) + \mathcal{D}_j([n],[m])\}. \tag{2}$$

As a concrete example let us consider the evolution of the latent compartment. There are three possible transitions from the compartment: transitions to the asymptomatic infectious, the traveling and the non-traveling symptomatic infectious compartments. The elements of the operator acting on L_j are extracted from the multinomial distribution

$$Pr^{Multin}(L_j(t), p_{L_j\to I_j^a}, p_{L_j\to I_j^t}, p_{L_j\to I_j^{nt}}), \tag{3}$$

determined by the transition probabilities

$$\begin{aligned} p_{L_j\to I_j^a} &= \varepsilon p_a \Delta t, \\ p_{L_j\to I_j^t} &= \varepsilon(1-p_a)p_t \Delta t, \\ p_{L_j\to I_j^{nt}} &= \varepsilon(1-p_a)(1-p_t)\Delta t, \end{aligned} \tag{4}$$

and by the number of individuals in the compartment $L_j(t)$ (its size). All these transitions cause a reduction in the size of the compartment. The increase in the compartment population is due to the transitions from susceptibles into latents. This is also a random number extracted from a binomial distribution

$$Pr^{Bin}(S_j(t), p_{S_j\to L_j}), \tag{5}$$

given by the chance of contagion

$$p_{S_j\to L_j} = \lambda_j \Delta t, \tag{6}$$

and a number of attempts equal to the number of susceptibles $S_j(t)$. After extracting these numbers from the appropriate multinomial distributions, we can calculate the change $\Delta L_j(t)$ as

$$\Delta L_j(t) = -\left[\mathcal{D}_j(L,I^a) + \mathcal{D}_j(L,I^t) + \mathcal{D}_j(L,I^{nt})\right] + \mathcal{D}_j(S,L). \tag{7}$$

4.1. The integration of the transport operator

The transport operator is defined by the airline transportation data which provides the number of available seats $\omega_{j\ell}$ between each pair of airports (j,ℓ). The operator is in general affected by fluctuations coming from the fact that the occupancy rate of the airplanes is not 100%. To take into account such fluctuations, we assume that on each connection (j,ℓ) the flux of passengers at time t is given by a stochastic variable

$$\tilde{\omega}_{j\ell} = \omega_{j\ell}[\alpha + \eta(1-\alpha)], \tag{8}$$

where α denotes the average occupancy rate of the order of 70–90% provided by IATA and η is a random number drawn uniformly in the interval $[-1,1]$ at each time step. The number of individuals in the compartment $[m]$ traveling from the subpopulation j to the subpopulation ℓ is an integer random variable, in that each of the $X_j^{[m]}$ potential travelers has a probability $p_{j\ell} = \tilde{\omega}_{j\ell}\Delta t/N_j$ to go from j to ℓ. In each subpopulation j the numbers of individuals $\xi_{j\ell}$ traveling on each connection $j\to\ell$ at time t define a set of stochastic variables $\{\xi_{j\ell}\}$, which follows the multinomial distribution

$$P(\{\xi_{j\ell}\}) = \frac{X_j^{[m]}!}{(X_j^{[m]} - \sum_\ell \xi_{j\ell})! \prod_\ell \xi_{j\ell}!} \prod_\ell p_{j\ell}^{\xi_{j\ell}} \times \left(1 - \sum_\ell p_{j\ell}\right)^{(X_j^{[m]} - \sum_\ell \xi_{j\ell})}, \tag{9}$$

where $(1-\sum_\ell p_{j\ell})$ is the probability of not traveling, and $(X_j^{[m]} - \sum_\ell \xi_{j\ell})$ stands for the number of non-traveling individuals of the compartment $[m]$. The multinomial distribution provides the correct probability for traveling individuals leaving j to distribute across the possible connections according to $\{p_{j\ell}\}$. We use standard numerical subroutines to generate random numbers of travelers following these distributions. The transport operator in each subpopulation j is therefore written as

$$\Omega_j([m]) = \sum_\ell (\xi_{\ell j}(X_\ell^{[m]}) - \xi_{j\ell}(X_j^{[m]})), \tag{10}$$

where the mean and variance of the stochastic variables are $\langle \xi_{j\ell}(X_j^{[m]})\rangle = p_{j\ell}X_j^{[m]}$ and $\mathrm{Var}(\xi_{j\ell}(X_j^{[m]})) = p_{j\ell}(1-p_{j\ell})X_j^{[m]}$. Direct flights as well as connecting flights up to two-legs flights can be considered. It is worth remarking that on average the airline network flows are balanced so that the subpopulation N_j are constant in time, e.g., $\sum_{[m]}\Omega_j([m]) = 0$.

4.2. Time-scale separation and the integration of the commuting flows

The GLEaM model combines the infection dynamics with long- and short-range human mobility. Each of these dynamical processes operates at a different time scale. The inverse of the rates of the disease dynamics define the time scale of the stochastic process that we can see as the average individual's permanence in a given compartment. For ILIs there are two important intrinsic time scales, given by the latency period ε^{-1} and the duration of infectiousness μ^{-1}, both larger than 1 day. The long-range mobility given by the airline network has a time scale of the order of 1 day, while the commuting takes place in a time scale of approximately $\tau^{-1} \sim 1/3$ day. The explicit implementation of the commuting in the model thus requires a time interval shorter than the minimal time of airline transportation data. To overcome this problem, we use a time-scale separation technique, in which the short-time dynamics is integrated into an effective force of infection in each subpopulation.

We start by considering the temporal evolution of subpopulations linked only by commuting flows and evaluate the relaxation time to an equilibrium configuration. Consider the subpopulation j coupled by commuting to other n subpopulations. The commuting rate between the subpopulation j and each of its neighbors i will be given by σ_{ji}. The return rate of commuting individuals is set to be τ. Following the work of Sattenspiel and Dietz [39], we can divide the individuals original from the subpopulation j, N_j, between $N_{jj}(t)$ who are from j and are located in j at time t and those, $N_{ji}(t)$, that are from j and are located in a neighboring subpopulation i at time t. Note that by consistency

$$N_j = N_{jj}(t) + \sum_i N_{ji}(t). \tag{11}$$

The rate equations for the subpopulation size evolution are then

$$\begin{aligned} \partial_t N_{jj} &= -\sum_i \sigma_{ji} N_{jj}(t) + \tau \sum_i N_{ji}(t), \\ \partial_t N_{ji} &= \sigma_{ji} N_{jj}(t) - \tau N_{ji}(t). \end{aligned} \tag{12}$$

By using condition (11), we can derive the closed expression

$$\partial_t N_{jj} + (\tau + \sigma_j) N_{jj}(t) = N_j \tau, \tag{13}$$

where σ_j denotes the total commuting rate of population j, $\sigma_j = \sum_i \sigma_{ji}$. $N_{jj}(t)$ can be expressed as

$$N_{jj}(t) = e^{-(\tau+\sigma_j)t} \left(C_{jj} + N_j \tau \int_0^t e^{(\tau+\sigma_j)s}\, \mathrm{d}s \right), \tag{14}$$

where the constant C_{jj} is determined from the initial conditions, $N_{jj}(0)$. The solution for $N_{jj}(t)$ is then

$$N_{jj}(t) = \frac{N_j}{1+\sigma_j/\tau} + \left(N_{jj}(0) - \frac{N_j}{1+\sigma_j/\tau} \right) e^{-\tau(1+\sigma_j/\tau)t}. \tag{15}$$

We can similarly solve the differential equation for the time evolution of $N_{ji}(t)$

$$N_{ji}(t) = \frac{N_j \sigma_{ji}/\tau}{1+\sigma_j/\tau} - \frac{\sigma_{ij}}{\sigma_j} \left(N_{jj}(0) - \frac{N_j}{1+\sigma_j/\tau} \right) e^{-\tau(1+\sigma_j/\tau)t} + \left[N_{ji}(0) - \frac{N_j \sigma_{ji}/\tau}{1+\sigma_j/\tau} + \frac{\sigma_{ij}}{\sigma_j} \left(N_{jj}(0) - \frac{N_j}{1+\sigma_j/\tau} \right) \right] e^{-\tau t}. \tag{16}$$

The relaxation to equilibrium of N_{jj} and N_{ji} is thus controlled by the characteristic time $[\tau(1+\sigma_j/\tau)]^{-1}$ and τ^{-1} in the exponentials, respectively. The former term is dominated by $1/\tau$ if the relation $\tau \gg \sigma_j$ holds. In our case, $\sigma_j = \sum_i \omega_{ji}/N_j$, that equals the daily total rate of commuting for the population j. Such rate is always smaller than one since only a fraction of the local population is commuting, and it is typically much smaller than $\tau \simeq 3\,\mathrm{day}^{-1}$ to $10\,\mathrm{day}^{-1}$. Therefore the relaxation characteristic time can be safely approximated by $1/\tau$. This time is considerably smaller than the typical time for the air connections of one day and hence we can approximate the subpopulations $N_{jj}(t)$ and $N_{ji}(t)$ with their equilibrium values,

$$N_{jj} = \frac{N_j}{1+\sigma_j/\tau} \quad \text{and} \quad N_{ji} = \frac{N_j \sigma_{ji}/\tau}{1+\sigma_j/\tau}. \tag{17}$$

This approximation, originally introduced by Keeling and Rohani [32], allows us to consider each subpopulation j as having an effective number of individuals N_{ji} in contact with the individuals of the neighboring subpopulation i. In practice, this is similar to separate the commuting time scale from the other time scales in the problem (disease dynamics, traveling dynamics, etc.). While the approximation holds exactly only in the limit $\tau \to \infty$, it is good enough as long as τ is much larger than the typical transition rates of the disease dynamics. In the case of ILIs, the typical time scale separation between τ and the compartments transition rates is close to one order of magnitude or even larger. Eq. (17) can be then generalized in the time scale separation regime to all traveling compartments $[m]$ obtaining the general expression

$$X_{jj}^{[m]} = \frac{X_j^{[m]}}{1+\sigma_j/\tau} \quad \text{and} \quad X_{ji}^{[m]} = \frac{X_j^{[m]}}{1+\sigma_j/\tau} \frac{\sigma_{ji}}{\tau}, \tag{18}$$

while $X_{jj}^{[m]} = X_j^{[m]}$ and $X_{ji}^{[m]} = 0$ for all the other compartments which are restricted from traveling. These expressions will be used to obtain the effective force of infection taking into account the interactions generated by the commuting flows.

4.3. Effective force of infection

The force of infection λ_j that a susceptible individual of a subpopulation j sees can be decomposed into two terms: λ_{jj} and λ_{ji}. The component λ_{jj} refers to the part of the force of infection which is due to interactions among individuals in j. While λ_{ji} indicates the force of infection acting on susceptibles of j during their commuting travels to a neighboring subpopulation i. The effective force of infection can be estimated by summing these two terms weighted by the probabilities of finding a susceptible from j in the different locations, S_{jj}/S_j and S_{ji}/S_j, respectively. Using the time-scale separation approximation that establishes the equilibrium populations of Eq. (18), we can write

$$\lambda_j = \frac{\lambda_{jj}}{1+\sigma_j/\tau} + \sum_i \frac{\lambda_{ji}\sigma_{ji}/\tau}{1+\sigma_j/\tau}. \tag{19}$$

We will focus now on the calculation of each term of the previous expression. The force of infection (see Table 2) occurring in a subpopulation j is due to the local infectious persons staying at j or to infectious individuals from a neighboring subpopulation i visiting j and so we can write

$$\lambda_{jj} = \frac{\beta_j}{N_j^*} \left(I_{jj}^{nt} + I_{jj}^t + r_\beta I_{jj}^a \right) + \frac{\beta_j}{N_j^*} \sum_i \left(I_{ij}^{nt} + I_{ij}^t + r_\beta I_{ij}^a \right), \tag{20}$$

where β_j is introduced to account for the seasonality in the infection transmission rate (if the seasonality is not considered, it is a constant), and N_j^* stands for the total effective population in the subpopulation j. By definition, $I_{jj}^{nt} = I_j^{nt}$ and $I_{ji}^{nt} = 0$ for $j \neq i$. If we use the equilibrium values of the other infectious compartments (see Eq. (18)), we obtain

$$\lambda_{jj} = \frac{\beta_j}{N_j^*} \left[I_j^{nt} + \frac{I_j^t + r_\beta I_j^a}{1+\sigma_j/\tau} + \sum_i \frac{I_i^t + r_\beta I_i^a}{1+\sigma_i/\tau} \sigma_{ij}/\tau \right]. \tag{21}$$

The derivation of λ_{ji} follows from a similar argument yielding:

$$\lambda_{ji} = \frac{\beta_i}{N_i^*} \left(I_{ii}^{nt} + I_{ii}^t + r_\beta I_{ii}^a \right) + \frac{\beta_i}{N_i^*} \sum_{\ell \in \upsilon(i)} \left(I_{\ell i}^{nt} + I_{\ell i}^t + r_\beta I_{\ell i}^a \right), \tag{22}$$

where $\upsilon(i)$ represents the set of neighbors of i, and therefore the terms under the sum are due to the visits of infectious individuals from the subpopulations ℓ, neighbors of i, to i. By plugging the equilibrium values of the compartment into the above expression, we obtain

$$\lambda_{ji} = \frac{\beta_i}{N_i^*} \left[I_i^{nt} + \frac{I_i^t + r_\beta I_i^a}{1+\sigma_i/\tau} + \sum_{\ell \in \upsilon(i)} \frac{I_\ell^t + r_\beta I_\ell^a}{1+\sigma_\ell/\tau} \sigma_{\ell i}/\tau \right]. \tag{23}$$

Finally, in order to have an explicit form of the force of infection we need to evaluate the effective population size N_j^* in each subpopulation j, *i.e.*, the actual number of people at the location j. The effective population is $N_j^* = N_{jj} + \sum_i N_{ij}$, that in the time-scale separation approximation reads

$$N_j^* = I_j^{nt} + \frac{N_j - I_j^{nt}}{1+\sigma_j/\tau} + \sum_i \frac{N_i - I_i^{nt}}{1+\sigma_i/\tau} \sigma_{ij}/\tau. \tag{24}$$

Note that in these equations all the terms corresponding to compartments have an implicit time dependence.

By inserting λ_{jj} and λ_{ji} into Eq. (19), it can be seen that the expression for the force of infection includes terms of zeroth, first and second order on the commuting ratios (i.e., σ_{ij}/τ). These three term types have a straightforward interpretation: the zeroth order terms represent the usual force of infection of the compartmental model with a single subpopulation. The first order terms account for the effective contribution generated by neighboring subpopulations, and is due to the contacts between susceptible individuals of subpopulation j and infectious individuals of neighboring subpopulations i. This can occur in two ways – either susceptible individuals of j visiting i or infectious individuals of i visiting j. The second

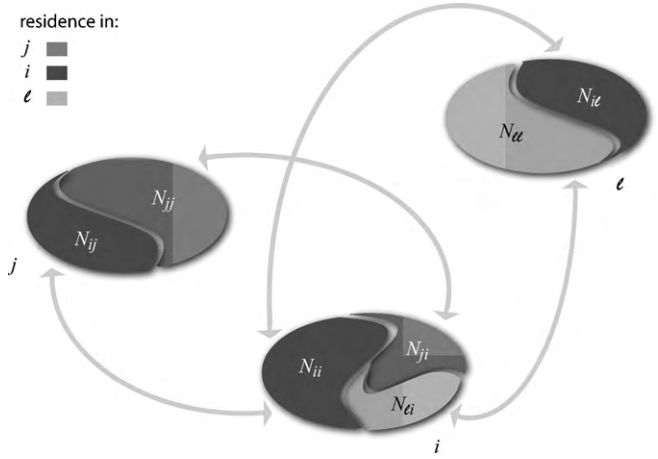


Fig. 4. Schematic representation of the subdivision of the population in each geographical census area. The population in each geographical census area is divided into partial populations N_{xy}, where x represents the subpopulation of residence and y represents the subpopulation of the actual location at time t. Three subpopulations are shown – i, j, ℓ – to represent the various contributions to the force of infection (see Eq. (19)).

order terms correspond to an effective force of infection generated by the contacts of susceptible individuals of subpopulation j meeting infectious individuals of subpopulation ℓ (neighbors of i) when both are visiting subpopulation i (see Fig. 4). This last term is very small in comparison with the zeroth and first order terms, typically around two order of magnitudes smaller, and in general can be neglected.

4.4. Seasonality modeling

To model seasonal variations we follow the approach of Cooper et al. [12] and scale the basic reproduction ratio R_0 by a seasonal function, $s_i(t)$,

$$s_i(t) = \left[\left(1 - \frac{R_{min}}{R_{max}}\right) \sin\left(\frac{2\pi}{365}\left(t - t_{max,i}\right) + \frac{\pi}{2}\right) + 1 + \frac{R_{min}}{R_{max}}\right]\frac{1}{2}, \tag{25}$$

where i stands for the North or South hemispheres. This function is identically equal to 1.0 in the tropical regions. $t_{max,i}$ is the time corresponding to the maximum seasonal effect, Jan 15 in the North and 6 months later in the South. Seasonality has a dual effect, it increases the value of R_0 up to $R_{max} = \alpha_{max} R_0$ with $\alpha_{max} \equiv 1.1$ [26] and reduces it down to $R_{min} = \alpha_{min} R_0$.

4.5. Age structure

In order to achieve refined analysis including the impact of an epidemics on different age groups, it is possible to include a generalization of the basic formalism that takes into account the presence of different contact rates among individuals belonging to different age bracket or more generally specific population groups. We start by distinguishing among different age groups with varying contact rates by using the results by Wallinga et al. [43]. In 2006, Wallinga et al. [43] measured the contact rates using a group of 1813 Dutch survey participants. With such data it is possible to write a contact matrix M, describing how many interactions an individual in one class has with individuals in a different age group. The main characteristic of the contact matrix is its asymmetry. This is easily explained if, for example, one considers children and adults. Children almost always live with adults, but adults do not always live with children. In order to obtain the effective rate of infection, we must multiply the probability of infection by appropriately rescaled rates describing the contacts between different age groups. A full description of the generalization of the formalisms is reported in Appendix A. While the theoretical and computational formalisms are ready to be generalized to the inclusion of age classes in the system, the main limitation to proceed along this direction is in the lack of data. Reliable information can be obtained on the age structure of most of the countries in the world, however detailed data on the contact matrix are limited to specific countries or settings, therefore a data-driven generalization to the whole world is still not available.

5. Algorithms, the simulator and its implementation

The GLEaM simulation toolbox is implemented in a modular way. Each module performs a single function, and they can be combined in different ways to include or remove specific features. In Algorithm 1 we outline the general program flow of a basic GLEaM run.

Algorithm 1. Generic GLEaM program flow.

```
Parse model file
Load data input files:
  population database
  commuting
  flight networks

foreach timestep t:
do
  Flight connections (See Algorithm 2)
  Infect (See Algorithm 3)
  Aggregate results for each detail level.
done

Generate final output
```

5.1. Long distance travel

Each time step represents a full day. At the start of the time step, we use the flight network to move travelers to their destination using Algorithm 2. Travel is assumed to be instantaneous with no transitions being possible on route. Performing this step at the start of the "day", guarantees that incoming travelers will contact with the local inhabitants during that day. As a consequence, the arrival time for the infection is the day at which the first infected traveler arrives and this seed individual is considered to have a full day chance of infecting others. The probability of traveling changes from day to day through fluctuations in the occupancy rate of flights, as shown in Algorithm 2, where α represents the average occupancy rate of the plane, and η is a stochastic random variable uniformly distributed between $[-1, 1]$. The Flight module can be customized in order to consider the effects of generalized or location specific airline traffic reductions.

Algorithm 2. Long distance mobility.

```
foreach city i:
do

  foreach neighbor j ∈ v(i):
  do
    Calculate traffic: ω̃_ij = ω_ij[α + η(1 − α)]
    Traveling probability: p_ij = ω̃_ij / N_i
  done

  distribute travelers among neighbors
  updated population matrix
end
```

5.2. Compartment transitions

The GLEaM framework is conceived in a generic way that facilitates the simulation of an arbitrary compartmental model that is

given as part of the input. The infection module is completely separated from the other modules (like Flight and Aggregation). The module can be customized in order to simulate the effect of policy measures that modify the transmission rates during a specific period of time.

The epidemic model description is processed to generate a directed multigraph, where each node represents a compartment and each edge a transition, following the representation of Fig. 3. Each edge is given a type, a weight and several other attributes. The type identifies whether the edge corresponds to a contagion or a spontaneous transition and the weight is the rate of transition. In the case of contagion transitions, the infectious agent is also identified, as there may be multiple infectious compartments as shown by Fig. 3. This structure provides a convenient way of internally representing arbitrarily complex models as well as facilitating an efficient implementation. The edges contain all the information necessary to calculate the transition probabilities that can then be used directly as arguments of the multinomial function that calculates the number of individuals making the transition.

Algorithm 3. Compartment transitions.

foreach city i:
do
  calculate effective populations due to commuting
  foreach initial compartment x:
  do
    Update transition probability to compartment y using Eq. (22) and Eq. (24).
    For seasonal transitions, scale transition rate by $s(t)$ (Eq. (25))
  done
  Move population between compartments using a multinomial
done

5.3. Aggregation and post-processing

The output produced by each run includes the population of each compartment for each census area at each time step and the number of transitions along each of the edges in the transition graph. The final step performed after each simulated day is a partial aggregation of the results, in order to both simplifying the post processing required to obtain useful results and reducing the already considerable amount of output generated for each run. At this point in the simulation, the populations of each census area and each compartment have already been updated and several quantities of interest can be calculated. In particular, we calculate the number of secondary cases generated during this specific time step and the current incidence at each of the following aggregation levels:

- Census area
- Country
- Region
- Continent
- Hemisphere
- Globe

In the case of some countries, we also consider within-country divisions, such as US states and Australian provinces.

After the run is finished, the output data files are post processed by a series of Python scripts to generate the analysis, figures and animations that are finally used. The advantage of decoupling simulation and analysis is in the flexibility it gives in tailoring the whole process. While some post processing steps (like the generation of epidemic profiles, arrival times and ArgGIS illustrations) are almost always considered, others can be added, removed or customized for specific situations. The full simulation process, containing all the steps described above, is illustrated schematically in Fig. 5.

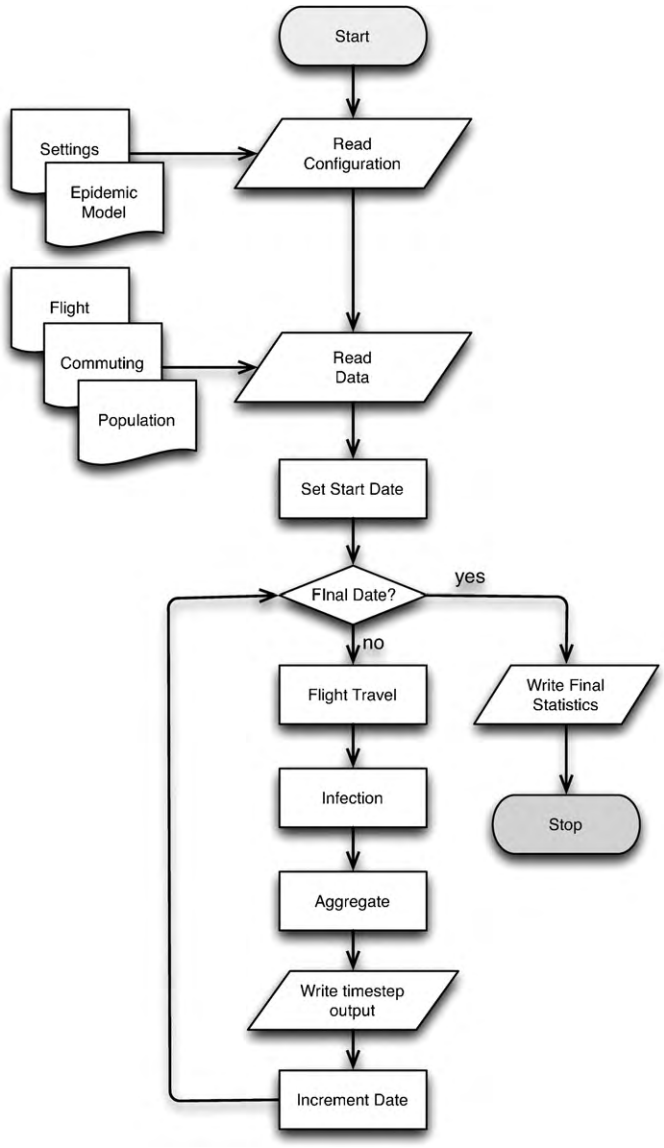


Fig. 5. Full illustration of the procedure used for the GLEaM simulation engine. The left column represents input databases and the right column the data structures that are generated. Program flow occurs along the center. The three steps in the center box are repeated for each simulated day.

6. GLEaM at work: simulation of 2001–2002 seasonal influenza A

In order to present a case study for the use of the GLEaM simulator we consider the spreading of seasonal influenza worldwide. Here we want to show how the model calibration may proceed by using real data from the surveillance and monitoring systems and what parameters are crucial in the description of the disease spread. Every year, seasonal influenza circulates globally and infect from 5% to 15% of the population, resulting in 3–5 million severe cases and ~500,000 deaths worldwide [42,45]. For the sake of simplicity, we focus on one influenza season with one dominant strain, in order to neglect complications arising from the interplay of different strains. This makes the 2001–2002 season a good candidate, which satisfies these criteria, among all the seasons from 1998 to 2006. In the Northern hemisphere, the season 2001–2002 has less than 5% mean proportion of annual A/H3N2 isolates, while in 2001–2002 this proportion is above 60% [20].

6.1. Model calibration and simulation

The main issue in the simulation of the influenza is the parametrization of the model in terms of the transmission rate and the initial condition for the circulation of a given strain at the global level. The origin of annual influenza circulation is still an unknown issue [37], however, from past experiences, new variants of influenza often originate in East-Southeast Asia [37], or Southeast China [13,40,41]. For season 2001–2002, according to the epidemiological records [44], Hong Kong is the only country/region in SE Asia having sporadic A/H3 influenza activity during June and July 2001. We therefore choose Hong Kong as the source of the influenza strain and explore possible starting dates between June and July. We further assume that a fraction equal to 10^{-5} of the city's population is latent, consistently with the literature and with the specific choice for the same season in Ref. [26]. In the case of influenza, we can implement the compartmental structure reported in Fig. 3. For the parameters of the model, we consider a latent period of $\epsilon^{-1} = 1.1$ days, and infectious period of $\mu^{-1} = 2.95$ days. The average generation interval for our choice is around 4 days, a value close to published estimates for the A/H3N2 [5]. Also in agreement with the literature, we assume that only a fraction of $\gamma = 60\%$ of the world population is susceptible to the circulating strain [26]. For the seasonality rescaling, we use the same seasonal rescaling as in Ref. [1]. We fix α_{max} and α_{min} at 1.1 and 0.1, respectively, to reflect the seasonal variabilities of influenza transmission.

The transmissibility of the disease is measured by the basic reproduction number R_0 which is defined as the average number of infected cases generated by the introduction of a single infectious individual into a fully susceptible population. For the compartmentalization used here, R_0 can be obtained in each subpopulation by evaluating the largest eigenvalue of the Jacobian or next generation matrix of the infection dynamics in a disease-free state [15,28], yielding

$$R_0 = \beta\mu^{-1}(1 - p_a + r_\beta p_a). \tag{26}$$

Given the parameters p_a and r_β, the value of R_0 depends on the transmission rate β that fixes the reference reproductive number in each subpopulations. For seasonal influenza, however, since the fraction of initially susceptible population is not one, the reproductive number must be rescaled by the proportion of susceptible individuals and we define an effective reproductive number $R_{eff} = \gamma R_0$.

In order to find a best estimate of the transmissibility and initial start date t_0, we perform simulations of the model for varying values of these two parameters and compare the results with the empirical data on the influenza activity peak in the French regions. The French Sentinelles Network is a surveillance system reported by voluntary and unpaid general practitioners (GP), which keeps a weekly record of ILI consultations since 1984 [23]. From the data, we can obtain for each French region the time of the activity peak t^{emp_peak}. We then perform a latin square sampling in the phase space of the parameters R_{eff} and t_0, constructing the surface representing the χ^2 values obtained by comparing the empirical peak times with the average simulated activity peak times $t_i^{sim\_peak}$ obtained by analyzing 2,000 stochastic GLEaM realizations for each sampled point. This Monte Carlo latin sampling procedure is computationally intensive as for each sampled point 2000 realization of the epidemic propagation worldwide must be generated. We have opted for a trade-off in the accuracy and computational cost samplings the phase space with a resolution $\Delta R_{eff} = 0.03$ and $\Delta t_0 = 7$ days. The best fit for the initial condition and the transmissibility is associated with the minimum of the χ^2 surface. Fig. 6 reports the χ^2 surface as a function of R_{eff} and seeding date t_0. The best fit range for R_{eff} is between 1.47 and 1.53 with the initial date between late June and early July, depending on the R_{eff}. From the analysis of the surface, we find a best estimate corresponding to $R_{eff} = 1.50$ and t_0 = July 11. A more accurate analysis with confidence interval is needed in order to provide a full discussion of these epidemiological results. This is however beyond the scope of this paper, where we want only to provide a practical example of the GLEaM implementation.

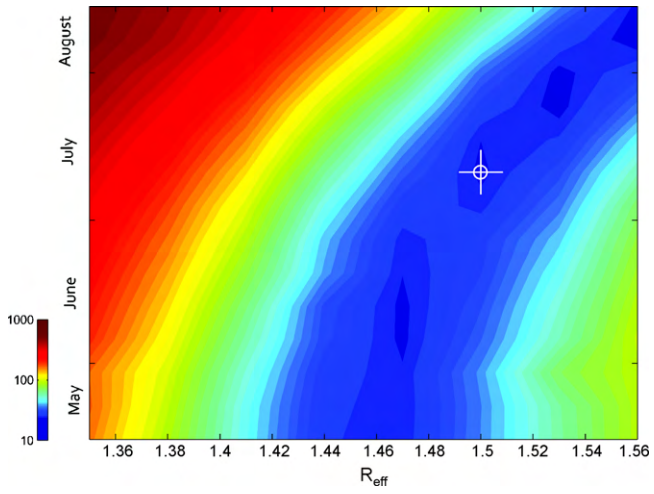


Fig. 6. Monte Carlo latin sampling. χ^2 values as functions of effective reproduction ratio (R_{eff}) and seeding date (t_0) of simulated epidemics obtained by 2000 stochastic runs for each pair of parameter values. Activity peak times of ILI consultations in the various French regions have been selected as probe and were compared with simulation results to obtain χ^2. As seen in the figure, there are 4 local minimums. Parameter values chosen for the analysis in Fig. 7 are shown by the crosshairs.

The best estimate of the parameters is obtained by using data only from a single country, in this case France. In order to provide an example of the accuracy of the GLEaM model in reproducing the spatio-temporal patterns of the disease spreading, we can compare the numerical results obtained with the parameters fitted in France with empirical data in several countries where reliable surveillance data is available. We have chosen a set of countries for which the reported dominant strain is A/H3N2 with a sufficient number of reported cases. Data is obtained from either the national public health agencies or the regional organizations. The full list of selected countries is shown in Table 3.

In Fig. 7, we report the activity peaks for the selected countries and compare our predictions with the 2001–2002 weekly surveillance data. The simulation and empirical data show a good agreement in most of the countries and regions. All data are normalized to 1, which guarantees that activities are shown on the same scale. For the simulated data, the activity peaks are reported with median values from 2000 stochastic simulations, along with the 95% reference range. For the empirical data, in addition to the number of laboratory confirmed cases, we also refer to additional indicators, such as ILI or Acute Respiratory Infection (ARI) consultation rate (per 100,000 population or per 1000 patient visits) which is usually conducted by physicians. For selected countries having only one type of dominant strain, the percentage of ILI is also a good indicator of influenza activity for the seasonal activity.

Table 3
Data sources for ILI% in the 2001/2002 influenza season.

Country	Type	Data source
US	A/H3N2	CDC
Canada	A/H3N2	PHA Canada
UK	A/H3N2	ECDC, UK HPA
Portugal	A/H3N2	ECDC
Spain	A/H3N2	ECDC
Belgium	A	ECDC
Australia	A/H3N2	DHA

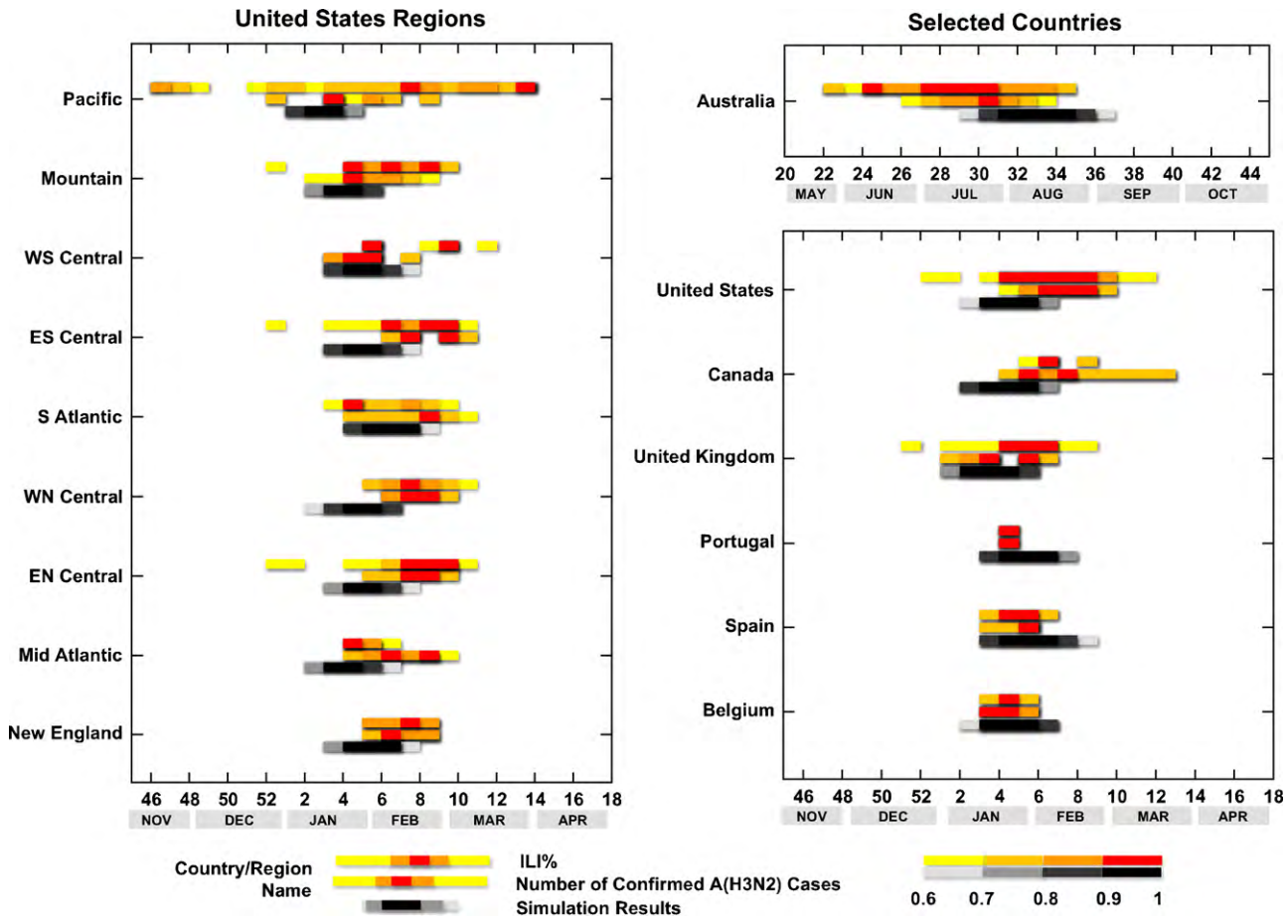


Fig. 7. Comparison of simulation results with the ILI consultations and number of confirmed cases of influenza A(H3N2). Simulations have been run by setting $R_{eff} = 1.5$ and seeding date of July 11th, as marked in Fig. 6. In order to obtain epidemic activity timelines, empirical and each of simulated profiles have been normalized to 1. Then the time windows have been evaluated relative to the peak activities in each case. For instance, lightest yellow bars of empirical data (lightest gray of simulated data) correspond to the time window in which activity is between 60% and 70% of the peak activity. Simulation results correspond to 95% reference range of simulated epidemics. The overlap between the predicted and observed cases is striking. It should be noted that parameter values have been obtained only by fitting the surveillance data in France, which has enabled GLEaM to reproduce the global pattern of the influenza season successfully.

Table 3 shows the dominant virus type and the data source used for individual countries. While the analysis reported here must be considered only as a simple illustration of the GLEaM implementation, the results appear to recover with good agreement the main spatio-temporal pattern of the 2001–2002 season. We want to stress that the timing of the epidemic spreading across different regions of the world is mostly determined by the human mobility patterns that are integrated in the GLEaM model with great accuracy. The best fit of the parameters obtained by the timeline of the epidemic in one or more countries allows the model to self-consistently capture the mobility of infected individuals and case importation that set the epidemic timeline worldwide.

7. Conclusions

Here we have provided a detailed description of the GLEaM simulator that is a discrete stochastic epidemic computational model based on a metapopulation approach in which the world is defined in geographical census areas connected in a network of interactions by human travel fluxes corresponding to transportation infrastructures and mobility patterns. Given the multitude of scales and mobility layers existing in the GLEaM model, the process of interest can be studied on a wide range of scales ranging from small administrative units (counties, municipalities) to worldwide. Although the GLEaM model has been used in the past in the analysis of realistic scenarios and in comparison with real data, also in relation with H1N1 pandemic, here we have presented for the first time all the data integration details, models and algorithms implementation that are under the hood of the GLEaM simulator. It is also worth noticing that while the model is being developed and tested in the context of emerging diseases such as new pandemic strains, it considers different transportation and interaction layers and distinguishes the mobility modeling from the dynamical process mediated by the human dynamics. This allows the integration of different processes of social contagion that are not necessarily of biological origin but occurs taking advantage of the individuals mobility such as information spreading, social behavior, etc. GLEaM has proved to be very flexible and we are working to make the GLEaM platform available to the scientific community at large. In particular we are developing an easy to use interface to the software that allows for the simulation and visualization of the spread of epidemics at a global scale.

Acknowledgements

We are grateful to the International Air Transport Association for making the airline commercial flight database available to us. This work has been partially funded by the NIH R21-DA024259 award, the Lilly Endowment grant 2008 1639-000 and the DTRA-1-0910039 award to AV; the EC-ICT contract no. 231807 (EPIWORK), and the EC-FET contract no. 233847 (DYNANETS) to AV and VC; the ERC Ideas contract n.ERC-2007-Stg204863 (EPIFOR) to VC. The

work has been also partly sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

Appendix A. Generalization including age structure

We now introduce the formalisms that allow for the inclusion of different contact rates among individuals in different age groups.

While we still make the fundamental assumption that the epidemic is governed by a single transmission rate β, we must now rescale it to take into account the different contact rates among different age groups. The contact matrix M, shown in Table A.1 describes how many contacts an individual in one class has with individuals in a different age group. Columns correspond to survey participants, and rows to the people they interacted with. As an example, we use the data gathered in 2006 by Wallinga et al. [43] who measured the contact rates using a group of 1813 Dutch survey participants. For self consistency, we required that the total number of interactions between two age groups must be the same. In other words, so we must have

$$m_{ab}N_b = m_{ba}N_a$$

Symmetrized matrix values are then given by $C_{ab} = m_{ab} \cdot N / N_a$, where N_a is the number of individuals in age group a and N is the total number of individuals. Values of N_a for both the survey participants and the entire Dutch population are given in Table A.2 and the full symmetric matrix C is shown in Table A.3.

While Wallinga considers only 6 age groups, our demographic data, as provided by the US Census Bureau [31] is more fine grained. We make the simplest choice and assume that people are uniformly distributed within each 5-year compartment, thus combining the age groups so that they fit the Wallinga picture.

A change in the way the different populations interact with each other necessarily implies a change in the way the epidemic spreads, requiring modifications to the R_0 calculation. We apply the techniques described in [15,28] to the general age structure case of interest.

Table A.1
Contact matrix M. From Ref. [43].

Age of contacts	Age of survey participants					
	1–5	6–12	13–19	20–39	40–59	60+
0–5	12.26	2.28	1.29	2.50	1.15	0.83
6–12	2.72	23.77	2.80	3.02	1.78	1.00
13–19	2.00	3.63	25.20	5.70	4.22	1.68
20–39	11.46	11.58	16.87	25.14	16.43	8.34
40–59	3.59	4.67	8.50	11.21	13.89	7.48
60+	1.94	1.95	2.54	4.25	5.59	9.19

Table A.2
Wallinga's population structure.

Age group	Participants	Population ($\times 10^3$)
0	0	184
1–5	125	876
6–12	154	1265
13–19	152	1642
20–39	681	4857
40–59	360	3312
60+	341	2477
Total	1813	14,614

Table A.3
Symmetrized contact matrix. From Ref. [43].

Age of contacts	Age of participants					
	1–5	6–12	13–19	20–39	40–59	60+
0–5	169.14	31.47	17.76	34.50	15.83	11.47
6–12	31.47	274.51	32.31	34.86	20.61	11.50
13–19	17.76	32.31	224.25	50.75	37.52	14.96
20–39	34.50	34.86	50.75	75.66	49.45	25.08
40–59	15.83	20.61	37.52	49.45	61.26	32.99
60+	11.47	11.50	14.96	25.08	32.99	54.23

Let us define $\vec{x} = (x_1, \ldots, x_n)$ to be a vector containing the number of individuals in each infected compartment. We have 4 such compartments, $L = x_1$, $I^t = x_2$, $I^{nt} = x_3$ and $I^a = x_4$. The matrix F, defining the rate of creation of new infected cases is then:

$$F \equiv \begin{pmatrix} 0 & \beta & \beta & r_\beta\beta \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

with a simple meaning: Latent cases (first row) are created (from susceptible) with rate β ($r_\beta\beta$) through interaction with $I^{t,nt}$ (I^a). Since these are the only ways in which the disease can spread through a Susceptible population, all other entries in the matrix are null. After infection, the disease progresses through several stages as described by the matrix $V = (v_{ab})$ where element v_{ab} is the number of individuals leaving compartment a to compartment b, minus the number of individuals following the opposite path. For seasonal flu, we have:

$$V \equiv \begin{pmatrix} \epsilon & 0 & 0 & 0 \\ -(1-p_a)p_t\epsilon & \mu & 0 & 0 \\ -(1-p_a)(1-p_t)\epsilon & 0 & \mu & 0 \\ -p_a\epsilon & 0 & 0 & \mu \end{pmatrix}$$

Using these two matrices we can calculate the next generation matrix,

$$\mathcal{N} \equiv FV^{-1}$$

that describes the complete epidemic process and whose interpretation is relatively simple: F is the rate at which new infections are created and V^{-1} is the average duration of each infected compartment. The basic reproductive ratio, R_0 is finally given by the maximum eigenvalue of this matrix that in a model without age structure reads as

$$R_0 = \lambda_{max}(\mathcal{N}) \equiv \frac{\beta}{\mu}[r_\beta p_a + (1-p_a)].$$

Adding age structure results in a proliferation of infected compartments. In the case of the Wallinga's age grouping, we have 6 times as many infected compartments. Fortunately, the fact that we do not consider aging implies that individuals never move between compartments corresponding to different age groups, thus greatly simplifying the analysis. We define the new vector $\vec{x}^\dagger$ to be a concatenation of 6 vectors $\vec{x}$ each corresponding to a different age cohort. Mixing between the different groups results in a susceptible individual becoming latent by interacting with an infectious person from any other group. In matrix notation, and using the previous definitions, the new infection matrix $F^\dagger$ is given by:

$$F^\dagger = M \times F,$$

where $\times$ represents the Kronecker product. After the initial infection, the disease progresses as before with each age group being

isolated from all others. The progression matrix $V^\dagger$ is then:

$$V^\dagger = \mathcal{I} \times V,$$

where $\mathcal{I}$ is the 6×6 identity matrix. The next generation matrix can now be written as:

$$\mathcal{N}^\dagger = M \times FV^{-1}$$

Therefore, the new basic reproductive number can be written as a function of the previous one:

$$R_0^\dagger = R_0 \cdot \lambda_{max}(M) \tag{A.1}$$

This formulation is completely generic and completely generalizable for any number of age groups with only a very small numerical effort. A specific value of R_0 can be set by inverting this expression and calculate the appropriate value of $\beta(R_0)$.

Before we can use this formulation in our global simulation, we must take into account the different demographics of each country or census areas and their change in time. Using the definitions above, we can write:

$$\Delta I_a = \beta \sum_b \frac{m_{ab}}{N_a} S_a I_b \equiv \beta \sum_b c_{ab} S_a I_b \tag{A.2}$$

to describe the increase in the number of people in compartment I_i in a basic SI model. Defining the fraction of individuals in compartment I_a as $\rho_{I_a} \equiv I_a/N$, we rewrite this expression as:

$$\Delta \rho_{I_a} = \beta \rho_{S_a} \sum_j C_{ab} \rho_{I_b}$$

where C_{ab} is the symmetric matrix defined above. Since this expression depends only on the relative fraction of individuals in each compartment and not on the details of how many people are actually in each compartment, we can safely conclude that C_{ab} is the matrix that must be kept constant for every population. We can now identify:

$$C_{ab} \equiv \frac{m_{ab}^\dagger}{N_a^\dagger} N^\dagger \equiv C_{ab}^\dagger$$

or, in other words:

$$m_{ab}^\dagger \equiv C_{ab} \frac{N_a^\dagger}{N^\dagger} \tag{A.3}$$

as the matrix that we must use in Eq. (A.1) and that will differ from country to country. Substituting in Eq. (A.2) we obtain:

$$\Delta I_a = \beta \sum_b C_{ab} \frac{S_a I_b}{N},$$

where N is the total population for the subpopulation considered and C_{ab} is the same for every population. The resulting force of infection is then:

$$\lambda_a = \beta \sum_b C_{ab} \frac{I_b}{N}. \tag{A.4}$$

During the derivation of this expression, and for the sake of clarity, we considered only a *single population*. The expression for the full force of infection including the mobility dynamics Eq. (A.4) can be obtained after the application of the prescription of Section 4. This can be easily done by replacing every term of the form $\beta_i I_i$ by

$$\beta_i \sum_b C_{ab} I_i^b. \tag{A.5}$$

References

[1] D. Balcan, H. Hu, B. Gonçalves, P. Bajardi, C. Poletto, J.J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. van den Broeck, et al., Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility, BMC Med. 7 (2009) 45.

[2] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J.J. Ramasco, A. Vespignani, Multiscale mobility networks and the large scale spreading of infectious diseases, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 21484–21489.

[3] A. Barrat, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 3747–3752.

[4] G. Bobashev, R.J. Morris, D.M. Goedecke, Sampling for global epidemic models and the topology of an international airport network, PLoS One 3 (2008) e3154.

[5] F. Carrat, E. Vergu, N. Ferguson, M. Lemaitre, S. Cauchemez, S. Leach, A. Valleron, Time lines of infection and disease in human influenza: a review of volunteer challenge studies, Am. J. Epidemiol. 167 (2008) 775–785.

[6] Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT), The Gridded Population of the World Version 3 (GPWv3): Population Grids, Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, NY. http://sedac.ciesin.columbia.edu/gpw.

[7] Center for International Earth Science Information Network (CIESIN), Columbia University; International Food Policy Research Institute (IFPRI); The World Bank; and Centro Internacional de Agricultura Tropical (CIAT), Global Rural–Urban Mapping Project (GRUMP), Alpha Version: Population Grids, Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, NY. http://sedac.ciesin.columbia.edu/gpw.

[8] M.L. Ciofi degli Atti, S. Merler, C. Rizzo, M. Ajelli, M. Massari, et al., Mitigation measures for pandemic influenza in Italy: an individual based model considering different scenarios, PLoS One 3 (2008) e1790.

[9] V. Colizza, A. Barrat, M. Barthélemy, A.J. Valleron, A. Vespignani, Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions, PLoS Med. 4 (2007) e13.

[10] V. Colizza, A. Barrat, M. Barthélemy, A. Vespignani, The role of the airline transportation network in the prediction and predictability of global epidemics, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 2015–2020.

[11] V. Colizza, A. Barrat, M. Barthélemy, A. Vespignani, The modeling of global epidemics: stochastic dynamics and predictability, Bull. Math. Biol. 68 (2006) 1893–1921.

[12] B.S. Cooper, R.J. Pitman, W.J. Edmunds, N.J. Gay, Delaying the international spread of pandemic influenza, PloS Med. 3 (2006) e12.

[13] N. Cox, K. Subbarao, Global epidemiology of influenza: past and present, Annu. Rev. Med. 51 (2000) 407–421.

[14] Database of the Statistical Office of the European Commission (Eurostat). http://epp.eurostat.ec.europa.eu/portal/page/portal/transport/data/database.

[15] O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, J. Math. Biol. 28 (1990) 365–382.

[16] J.M. Epstein, D.M. Goedecke, F. Yu, R.J. Morris, D.K. Wagener, G.V. Bobashev, Controlling pandemic flu: the value of international air travel restrictions, PLoS One 2 (2007) e401.

[17] S. Eubank, H. Guclu, V.S. Anil Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczkai, N. Wang, Modelling disease outbreaks in realistic urban social networks, Nature 429 (2004) 180–184.

[18] N.M. Ferguson, D.A.T. Cummings, S. Cauchemez, C. Fraser, S. Riley, et al., Strategies for containing an emerging influenza pandemic in Southeast Asia, Nature 437 (2005) 209–214.

[19] N.M. Ferguson, D.A. Cummings, C. Fraser, J.C. Cajka, P.C. Cooley, D.S. Burke, Strategies for mitigating an influenza pandemic, Nature 442 (2006) 448–452.

[20] B. Finkelman, C. Viboud, K. Koelle, M. Ferrari, N. Bharti, B. Grenfell, Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients, PLoS One 2 (2007) e1296.

[21] A. Flahault, A.-J. Valleron, A method for assessing the global spread of HIV-1 infection based on air-travel, Popul. Stud. 3 (1991) 1–11.

[22] A. Flahault, E. Vergu, L. Coudeville, R. Grais, Strategies for containing a global influenza pandemic, Vaccine 24 (2006) 6751–6755.

[23] P. Garnerin, A.J. VAlleron, The French communicable diseases computer network: A technical view, Computers in Biology and Medicine 22 (1992) 189–200.

[24] T.C. Germann, K. Kadau, I.M. Longini, C.A. Macken, Mitigation strategies for pandemic influenza in the United States, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 5935–5940.

[25] R.F. Grais, J. Hugh Ellis, G.E. Glass, Assessing the impact of airline travel on the geographic spread of pandemic influenza, Eur. J. Epidemiol. 18 (2003) 1065–1072.

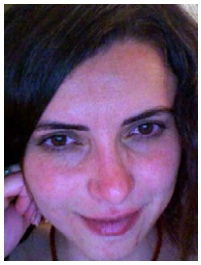
[26] R.F. Grais, J.H. Ellis, A. Kress, G.E. Glass, Modeling the spread of annual influenza epidemics in the U.S.: the potential role of air travel, Health Care Manage. Sci. 7 (2004) 127–134.

[27] M.E. Halloran, N.M. Ferguson, S. Eubank, I.M. Longini, D.A.T. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T.C. Germann, D. Wagener, R. Beckman, K. Kadau, C.A. Macken, D.S. Burke, P. Cooley, Modeling targeted layered containment of an influenza pandemic in the United States, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 4639–4644.

[28] J.M. Heffernan, R.J. Smith, L.M. Wahl, Perspectives on the basic reproductive ratio, J. R. Soc. Interface 2 (2005) 281–293.

[29] L. Hufnagel, D. Brockmann, T. Geisel, Forecast and control of epidemics in a globalized world, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 15124–15129.

[30] International Air Transport Association (IATA). http://www.iata.org.
[31] International data base (idb). http://www.census.gov/ipc/www/idb/. Last accessed January 31, 2009.
[32] M.J. Keeling, P. Rohani, Estimating spatial coupling in epidemiological systems: a mechanistic approach, Ecol. Lett. 5 (2002) 20–29.
[33] I.M. Longini, A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D. Cummings, M.E. Halloran, Containing pandemic influenza at the source, Science 309 (2005) 1083–1087.
[34] S. Merler, M. Ajelli, The role of population heterogeneity and human mobility in the spread of pandemic influenza, Proc. Roy. Soc. B: Biol. Sci. 277 (2010) 557–565.
[35] Official Airline Guide (OAG). http://www.oag.com.
[36] S. Riley, Large-scale spatial-transmission models of infectious disease, Science 316 (2007) 1298–1301.
[37] C. Russell, T. Jones, I. Barr, N. Cox, R. Garten, V. Gregory, I. Gust, A. Hampson, A. Hay, A. Hurt, et al., The global circulation of seasonal influenza A(H3N2) viruses, Science 320 (2008) 340–346.
[38] L.A. Rvachev, I.M. Longini, A mathematical model for the global spread of influenza, Math. Biosci. 75 (1985) 3–22.
[39] L. Sattenspiel, K. Dietz, A structured epidemic model incorporating geographic mobility among regions, Math. Biosci. 128 (1995) 71–91.
[40] K. Shortridge, Is China an influenza epicentre? Chin. Med. J. 110 (1997) 637–641.
[41] R. Snacken, A. Kendal, L. Haaheim, J. Wood, The next influenza pandemic: lessons from Hong Kong, Emerg. Infect. Dis. 5 (1999) 195–203.
[42] K. Stohr, Influenza—who cares, Lancet Infect. Dis. 2 (2002) 517–519.
[43] J. Wallinga, P. Teunis, M. Kretzschmar, Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents, Am. J. Epidemiol. 164 (2006) 936–944.
[44] World Health Organization, Influenza in the world, Weekly Epidemiol. Rec. 76 (2001) 357–364.
[45] World Health Organization, Influenza: Fact sheet (March 2003). http://www.who.int/mediacentre/factsheets/2003/fs211/en.



Duygu Balcan is a research associate at the Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington. Her current research interests involve mathematical and computational modeling of contagion processes with a specific focus on spreading of emergent infectious diseases. She obtained her PhD in Physics from Istanbul Technical University, Turkey, in 2007.

Bruno Gonçalves completed his joint PhD in Physics, MSc in C.S. at Emory University in Atlanta, GA in 2008 following which he joined the Center for Complex Networks and Systems Research at Indiana University as a post-doctoral research associate. His research activity focuses on using computational, visualization and data analysis methods for the study of Complex Systems in a multidisciplinary context. Current projects include detailed epidemic modeling in structured populations; knowledge diffusion on large technological networks; and the study of human behavior through the analysis of proxy social network dynamics.

Hao Hu completed his undergraduate studies at the Department of Physics, University of Science and Technology of China (USTC) in July, 2005. He then went to Indiana University and obtained his physics master's degree in February, 2007. Currently he is a PhD student in the physics department and the biocomplexity institute. During his study he joined the complex system group. His research interests involve the study of complex networks, especially the mathematical modeling of dynamical processes on networks, such as the spreading of diseases and malwares.

José J. Ramasco completed his PhD at the "Universidad de Cantabria" in Santander (Spain). After this, he transferred to Oporto (Portugal) for a two years postdoc in the "Centro de Fisica do Porto", an institute of the University of Oporto. Later he hold a two- year postdoc fellowship at the Physics Department of Emory University in Atlanta, GA. Since 2006, he is a research scientist at the ISI Foundation in Turin, Italy. His research activity focuses on several aspects of complex networks, from theoretical issues to real world applications including realistic modeling of epidemic spreading or of user Web traffic.

Vittoria Colizza is a research scientist at the Institute for Scientific Interchange (ISI Foundation) in Turin, Italy, where she leads the Computational Epidemiology Lab. Her research focuses on the characterisation and modeling of the spread of emerging infectious diseases, through an integrated approach that includes methods of complex systems, statistical physics techniques, computational sciences, and GIS. After obtaining her PhD in Physics at SISSA in Trieste, Italy, in 2004, she held a research position at Indiana University in Bloomington, IN, USA, and joined the ISI Foundation in 2007. She was awarded in 2008 a Career Grant by the European Research Council.

Alessandro Vespignani is currently James H. Rudy Professor of Informatics and Computing and adjunct professor of Physics and Statistics at Indiana University where he is also the director of the Center for Complex Networks and Systems Research (CNetS) and associate director of the Pervasive Technology Institute. Recently Vespignani's research activity focuses on the interdisciplinary application of statistical and simulation methods in the analysis of epi spreading phenomena and the study of biological, social and technological networks. Vespignani is an elected fellow of the American Physical Society and is serving in the board/leadership of a variety of professional association and journals.