



EPIDEMIC PREDICTABILITY IN META-POPULATION MODELS WITH HETEROGENEOUS COUPLINGS: THE IMPACT OF DISEASE PARAMETER VALUES

V. COLIZZA^{*,†}, A. BARRAT^{*,‡}, M. BARTHÉLEMY[§] and A. VESPIGNANI[†]

^{*}*Complex Networks Lagrange Laboratory (CNLL),
Institute for Scientific Interchange (ISI),
10133 Turin, Italy*

[†]*School of Informatics and Biocomplexity Institute,
Indiana University, Bloomington, IN 47408, USA*

[‡]*Laboratoire de Physique Théorique,
(UMR du CNRS 8627), Université de Paris-Sud,
Orsay, France*

[§]*CEA-DIF Centre d'Etudes de Bruyères-Le-Château,
BP12, F-91680, France*

Received June 16, 2006; Revised September 12, 2006

We study the predictability of epidemic forecasts in a data-driven meta-population model considering the complete air transportation system and the associated urban areas. We define the predictability as the robustness of the system evolution with respect to the stochastic fluctuations. As a quantitative measure of predictability we consider the information similarity of the time series characterizing different epidemic outbreaks with the same initial conditions. We study the predictability as a function of the parameters describing the basic susceptible-latent-infected and recovered disease dynamics. We find that the overall predictability is determined by the level of sampling of the underlying travel pattern by infected and latent individuals.

Keywords: Epidemiology; complex networks; stochastic noise; predictability.

1. Introduction

Large scale epidemic forecast is crucially dependent upon the accurate and realistic modeling where the movement of individuals at various levels is taken into account. In the case of spatially extended systems, modeling approaches have evolved into meta-population schemes which explicitly include spatial structures and consist of multiple subpopulations coupled by the movement of people among those [Anderson & May, 1984; May & Anderson, 1984; Bolker & Grenfell, 1993, 1995;

Lloyd & May, 1996; Grenfell & Bolker, 1998; Keeling & Rohani, 1995; Ferguson *et al.*, 2003]. The heterogeneities and details of the population structure have become increasingly important features and the introduction of large scale agent based models has enabled the simulation of the propagation of an infectious disease at the level of the single individual [Chowell *et al.*, 2003; Eubank *et al.*, 2004; Ferguson *et al.*, 2005; Longini *et al.*, 2005]. If one focuses on the world-wide level, the large scale and geographical impact of infectious diseases on

populations in the modern era is mainly due to commercial air travel. For this reason, in the early 80's, models aimed at forecasting the geographical spread of epidemics by using information on the airplane passenger fluxes among cities have been developed [Rvachev & Longini, 1985], capitalizing on previous studies on the Russian airline network [Baroyan *et al.*, 1969]. Similar modeling approaches have been used to study specific outbreaks such as pandemic influenza [Longini, 1988; Grais *et al.*, 2004], HIV [Flahault & Valleron, 1991], and SARS [Hufnagel *et al.*, 2004]. More recently, the availability of complete datasets [Guimerà & Amaral, 2004; Barrat *et al.*, 2004; Guimerà *et al.*, 2005] has allowed to build modeling frameworks for the computational study of epidemics taking into account the full air transportation systems and the census data of the corresponding urban areas [Colizza *et al.*, 2006a, 2006b]. In these approaches the only parameters in the models are the disease transition rates describing the etiology of the disease. All the couplings due to transportation terms, urban area size and the network connectivity pattern are fixed elements of the problem obtained from real data.

While the ability to forecast the spread of an infectious disease is inevitably entangled with the accuracy and the level of realism introduced in the modeling approach, the impact of the parameters and model features on the reliability of the predicted scenarios is difficult to assess. Here we investigate how different communicable diseases — characterized by different epidemiological parameters regarding infectiousness and transition rates — might affect our ability to predict their geographical and temporal propagation. In order to tackle this issue we consider the meta-population model defined in [Colizza *et al.*, 2006a, 2006b] for the forecast of the world-wide spread of emerging diseases. We perform a systematic analysis of the effect of the disease parameters on the statistical similarity of the epidemic behavior evolution obtained with the same initial conditions and different noise realizations. This amounts to the assessment of the robustness of the obtained predictions with respect to the inherent randomness of the epidemic evolution. The analysis shows that the overall predictability of the epidemics is determined by competing effects related to the number of infected individuals and the heterogeneity of the traveling pattern.

The paper is organized as follows: we recall the modeling framework in Sec. 2; Sec. 3 presents the results regarding the impact of disease parameter values on quantitative forecasting of epidemic outbreaks. Finally, we consider some conclusions and directions for future work in Sec. 4.

2. Meta-Population Model

We consider the general modeling framework presented in [Colizza *et al.*, 2006a, 2006b], which describes the world-wide propagation of a disease through a meta-population approach, in the same spirit as in the models of [Longini, 1988; Grais *et al.*, 2003; Flahault & Valleron, 1991; Hufnagel *et al.*, 2004]. The spatial structure is explicitly included in the model by considering multiple sub-populations coupled by movements of individuals. More specifically, the sub-populations correspond to urban areas and individuals in each urban area are allowed to travel according to the routes of the airline transportation system.

2.1. Intra-city level: SEIR infection dynamics

The epidemic evolution inside each urban area is described by a fully mixed population with a basic standard compartmentalization in which each individual can only exist in one of the following discrete states such as susceptible (S), latent (E), infected (I) or permanently recovered (R). This set of states do not correspond strictly to any particular disease but encompasses the most relevant features and parameters of a variety of different virus transmission, by including specific stages of the disease dynamics such as e.g. the latency period. More refined and extended compartmentalizations can be considered.

In each city j the population is N_j and the numbers of individuals in the various classes at time t are denoted as $S_j(t)$, $E_j(t)$, $I_j(t)$ and $R_j(t)$ respectively. By definition $N_j = S_j(t) + E_j(t) + I_j(t) + R_j(t)$. Individuals change compartment according to the following rules: during the time interval dt , the probability that a susceptible individual acquires the infection from any given infected individual, and becomes latent, is proportional to βdt , where β is the transmission parameter that captures the etiology of the infection process. Latents become infected with probability ϵdt , and infected individuals recover with a probability μdt , where ϵ^{-1} and μ^{-1} are the average latency time and average infection duration.

The populations N_j of the large metropolitan areas served by the airports are obtained by different census sources freely available on the Internet. For large cities, N_j represents the population of the whole metropolitan area. The final data set contains the population of the 3100 largest airports.

2.2. Transport operator

In addition to the infection dynamics taking place inside each urban area, the epidemic evolution at the global level is governed by the travel of individuals from one city to another by means of the airline transportation network, therefore infecting different regions of the world. The network is obtained by considering the International Air Transport Association (IATA)¹ database, which contains the world list of airport pairs connected by direct flights and the number of available seats on any given connection. The resulting world-wide air-transportation network is therefore a weighted graph comprising $V = 3100$ vertices denoting airports in 220 different countries and $E = 17182$ weighted edges whose weights represent the passenger flows between airports. This dataset accounts for 99% of the world-wide traffic and has been complemented by the population of the large metropolitan area served by the airport as obtained by different sources. The obtained network is highly heterogeneous both in the connectivity pattern and in the traffic capacities [Guimerà & Amaral, 2004; Barrat *et al.*, 2004].

In practice, the travel dynamics is described by the transport operator $\Omega_j(\{X\})$ representing the net balance of individuals in a given class X (S , E , I or R) that entered and left each city j . This operator is a function of the traffic flows $w_{j\ell}$ per unit time, the city populations N_j , and might also include transit passengers on connecting flights (see [Colizza *et al.*, 2006b] for details). The number of passengers $\xi_{j\ell}(X_j)$ of each compartment X traveling from a city j to a city ℓ is an integer random variable, extracted from a multinomial distribution which considers $p_{j\ell} = w_{j\ell}\Delta t/N_j$ as the probability of traveling from j to ℓ in the time interval Δt . Mean and variance of $\xi_{j\ell}(X_j)$ are given by $\langle \xi_{j\ell}(X_j) \rangle = p_{j\ell}X_j$ and $\text{Var}(\xi_{j\ell}(X_j)) = p_{j\ell}(1 - p_{j\ell})X_j$, respectively, and the transport operator for each city j can be written as

$$\Omega_j(\{X\}) = \sum_{\ell} (\xi_{\ell j}(X_{\ell}) - \xi_{j\ell}(X_j)). \quad (1)$$

2.3. Global stochastic epidemic model

In each city we work directly with the master equations for the dynamical rules given in Sec. 2.1. Under the assumption of large populations, we obtain the Langevin equations for the numbers of susceptible, latent, infected and recovered individuals and we associate to each reaction process a noise term with amplitude proportional to the square root of the reaction term (for details, see [Colizza *et al.*, 2006b; Gardiner, 2004; Marro & Dickman, 1998]). The noise term accounts for the stochastic nature of the infection dynamics of directly transmitted diseases, with intrinsic fluctuations in the infectious contact process, the generation of new cases and the removal of infectious individuals. These stochastic nonlinear differential equations are then coupled through the transport operator, yielding the following set of stochastic discretized differential equations for the worldwide spread of the infection:

$$\begin{aligned} S_j(t + \Delta t) - S_j(t) &= -\beta \frac{I_j S_j}{N_j} \Delta t + \sqrt{\beta \frac{I_j S_j}{N_j}} \Delta t \eta_{j,1}(t) \\ &\quad + \Omega_j(\{S\}) \end{aligned} \quad (2)$$

$$\begin{aligned} E_j(t + \Delta t) - E_j(t) &= +\beta \frac{I_j S_j}{N_j} \Delta t - \epsilon E_j \Delta t - \sqrt{\beta \frac{I_j S_j}{N_j}} \Delta t \eta_{j,1}(t) \\ &\quad + \sqrt{\epsilon E_j \Delta t} \eta_{j,2}(t) + \Omega_j(\{E\}) \end{aligned} \quad (3)$$

$$\begin{aligned} I_j(t + \Delta t) - I_j(t) &= +\epsilon E_j \Delta t - \mu I_j \Delta t - \sqrt{\epsilon E_j \Delta t} \eta_{j,2}(t) \\ &\quad + \sqrt{\mu I_j \Delta t} \eta_{j,3}(t) + \Omega_j(\{I\}) \end{aligned} \quad (4)$$

$$\begin{aligned} R_j(t + \Delta t) - R_j(t) &= +\mu I_j \Delta t - \sqrt{\mu I_j \Delta t} \eta_{j,3}(t) + \Omega_j(\{R\}). \end{aligned} \quad (5)$$

Here $\eta_{j,1}$, $\eta_{j,2}$ and $\eta_{j,3}$ are statistically independent normal random variables with zero mean and unit variance. The model is thus a compartmental system of 3100×4 stochastic differential equations whose integration provides the disease evolution in the corresponding urban areas (see [Colizza *et al.*, 2006b] for more implementation details). A quantity commonly used to describe the epidemic behavior in time is the prevalence in a city j , which is

¹<http://www.iata.org>.

given by the fraction I_j/N_j of infectious individuals with respect to the city population. The global prevalence at time t is defined as the world-wide density of infectious individuals, $\sum_j I_j/\sum_j N_j$, thus providing a measure of the global level of infection.

3. Predictability

A central issue in modeling the spread of epidemic diseases is to assess the reliability of the predicted patterns. The interplay of different aspects involved in the process — infection dynamics, travel, stochasticity, transportation network, etc. — might have a crucial role in the level of accuracy and reliability provided by the obtained epidemic forecasts. In [Colizza et al., 2006a, 2006b], we have investigated the role of the air-transportation network in the emergence of spatiotemporal propagation patterns and proposed to use a similarity measure to quantify the predictability of epidemic outbreaks. Indeed, even when global average quantities have small fluctuations, different sub-populations might be affected in very different ways, resulting in very different epidemic scenarios. We therefore need to assess the reliability of the epidemic forecast at the local level however considering the complete set of sub-populations and geographical areas. More precisely, the predictability of the epidemic pattern is given by the statistical similarity of the detailed history of epidemics outbreaks starting with the same initial conditions and subject to different noise realizations. Let us define the *active* individuals A_j in city j as the sum of latent and infectious individuals, $A_j = E_j + I_j$, thus representing the infected individuals still active, i.e. not yet recovered. We then consider the full vector $\boldsymbol{\pi}(t)$ whose components are $\pi_j(t) = A_j(t)/\sum_\ell A_\ell$; i.e. the probability that an active individual is in city j .² The similarity between two outbreak realizations is quantitatively measured by the statistical similarity of two realizations (I and II) of the global epidemic characterized by the vectors $\boldsymbol{\pi}^I$ and $\boldsymbol{\pi}^{II}$ respectively. Among possible measures of similarity, we consider here the Hellinger affinity

$$\text{sim}_H(\boldsymbol{\pi}^I, \boldsymbol{\pi}^{II}) = \sum_j \sqrt{\pi_j^I \pi_j^{II}}. \quad (6)$$

This measure belongs to the interval $[0, 1]$, with $\text{sim}(\boldsymbol{\pi}^I, \boldsymbol{\pi}^{II}) = 1$ when the two distributions are

identical and $\text{sim}(\boldsymbol{\pi}^I, \boldsymbol{\pi}^{II}) = 0$ when there is no overlap at all. The normalization of $\boldsymbol{\pi}(t)$ however means that two realizations differing only of a global multiplicative factor will generate the same vector $\boldsymbol{\pi}$. In order to correctly take into account the possible difference in the total epidemic prevalence between the two realizations, it is thus necessary to consider as well the quantity $\text{sim}(\mathbf{a}^I, \mathbf{a}^{II})$ where $\mathbf{a} = (a, 1 - a)$ and a is the worldwide density of active individuals $\sum_j A_j/\mathcal{N}$ (\mathcal{N} is the total population). This vector thus takes into account the similarity of the global prevalence in two stochastic realizations. In summary, the overlap function between two different stochastic realizations of an epidemic with the same starting initial conditions and disease parameters is defined as

$$\Theta(t) = \text{sim}_H(\mathbf{a}^I, \mathbf{a}^{II}) \times \text{sim}_H(\boldsymbol{\pi}^I, \boldsymbol{\pi}^{II}). \quad (7)$$

The overlap is maximal ($\Theta(t) = 1$) when the very same cities have the very same number of active individuals in both realizations, and $\Theta(t) = 0$ if the two realizations do not have any common infected cities at time t .

Here we focus on the influence of disease parameters such as the latency time ϵ^{-1} , the infectiousness β , and the initial percentage of susceptible individuals on the predictability. Different values of β and ϵ^{-1} point to different viruses or different strains of the same virus, with variations in the virus infectiousness and average latency period; changes in the percentage of initially susceptible population are related to acquired immunity or vaccination campaigns. The average infectious period μ^{-1} is kept fixed and equal to 3 days, as estimated for seasonal influenza. In the following, we will study how the variation of the disease parameters affects the evolution of the overlap over time. As initial conditions we consider five infected individuals in Frankfurt (which is the airport with the largest number of connections). Results are obtained as averages over 500 different realizations of the stochastic noise. We consider a set of parameter values and initial conditions such that the epidemic outbreak always occurs and the spread of the infection out of the initially infected city is observed. Cases in which the initial outbreak dies out in the seeding city due to fluctuations are not considered in this analysis.

²Note that one could as well consider instead $\pi_j^I(t) = I_j(t)/\sum_l I_l$, i.e. only infected individuals. Results are not affected by the different choice.

3.1. Transmission rate

We first focus on the influence of the transmission rate, at fixed latency time and within a fully susceptible population. This corresponds to a change in the virus infectiousness, i.e. its ability to be transmitted from an infectious individual to a susceptible one, and affects the value of the reproductive ratio R_0 , the key epidemiological parameter for the description of a disease. In an SEIR model, the reproductive ratio is given by $R_0 = \beta/\mu$ and represents the average number of secondary infections generated by a typical infectious individual in a fully susceptible population. Increasing values of R_0 can be achieved by increasing the transmission rate or increasing the average infectious period μ^{-1} , leading to the same qualitative results. In the following we will focus on changes in the virus infectiousness β while keeping μ constant. Larger values of β correspond to faster propagation, higher prevalence and larger number of cases, as shown in Fig. 1(a).

The evolution of the overlap Θ is shown in Fig. 1(b). Starting from a value equal to 1, due to setting identical initial conditions, Θ decreases in the early stage of the outbreak, when the number of cases is still small. As discussed in [Colizza *et al.*, 2006a, 2006b], this decrease is essentially due to the fact that the initially latent or infected may travel and go out of the initially infected city through various routes. In particular, the overlap decreases more if the seed of the infection is a airport hub, from which many different possible paths for the virus propagation can be taken. The overlap then increases again and reaches a maximum at the epidemic peak: in the absence of containment measures, this peak corresponds indeed to a relatively uniform and predictable pattern of infection. The gradual decrease of the global prevalence can on the other hand be quite different from one realization to the other, with the epidemic dying out in different cities at different times, so that Θ decreases again towards 0.

Although the global evolution of Θ always follows the same qualitative pattern, a strong quantitative effect is observed when the disease parameters are changed. A first observation is simply that the time scale of the spatial propagation depends obviously on the time scale of the local infection dynamics inside each urban area: the evolution is globally slower for smaller β . In the initial stages however, the picture is more complicated: while the

overlap profile at short times is similar for all values of β , the decrease of Θ is more pronounced for smaller β , as shown in Fig. 1(b). Such trend can be explained by the fact that, since the transmission rate is reduced the number of active individuals generated per unit time is lower. This has *a priori* two consequences which both tend to lower the overlap: first of all, even in the absence of any travel term, the relative stochastic fluctuations associated to the random contamination process are larger; moreover, the fluctuations due to travel are enhanced as well, since the smaller number of infected individuals can spread the disease in very different “diffusion” patterns for different realizations. In other terms, the evolution of a large number of individuals on the network is statistically more predictable than that of a small number. The effect can be quite drastic, since the minimal value reached by Θ in the initial epidemic stages varies from 70% to approximately 30% for β varying from 0.50 to 1.055.

In order to understand the separate roles of these two effects, we have isolated the two aspects. In a first experiment we have computed Θ for an initially infected *isolated* city, i.e. a dynamical setup in which the travel term is absent. This corresponds to a single-population epidemic model with internal fluctuations. The overlap values obtained are extremely high, with $1 - \Theta$ of the order 10^{-4} . In a second experiment, we have monitored the evolution of Θ when only the diffusion process is considered, with no infection dynamics. This corresponds to set the disease parameter β , ϵ and μ equal to 0. $\Theta(t)$ decreases exponentially at short times and reaches a plateau which depends on the number of individuals diffusing on the network. For larger number of individuals, the travel pattern fluctuates less and the overlap is larger.

At finite β , two competing effects are therefore determining the overlap profile: the diffusion of individuals decreases Θ , while the exponential increase in the number of active individuals tends to increase it. This competition leads to the observed minimum of $\Theta(t)$, which occurs at larger times for decreasing β (since the second effect is slower), and thus at lower values of Θ . After the minimum, the epidemic growth takes over so that the overlap increases; the values of Θ reached at the epidemic peak are essentially independent of β : although the prevalence is smaller for smaller β , it is in any case large enough to have a large reproducibility.

Let us finally note that, for smaller β , the fluctuations associated to the average overlap profile

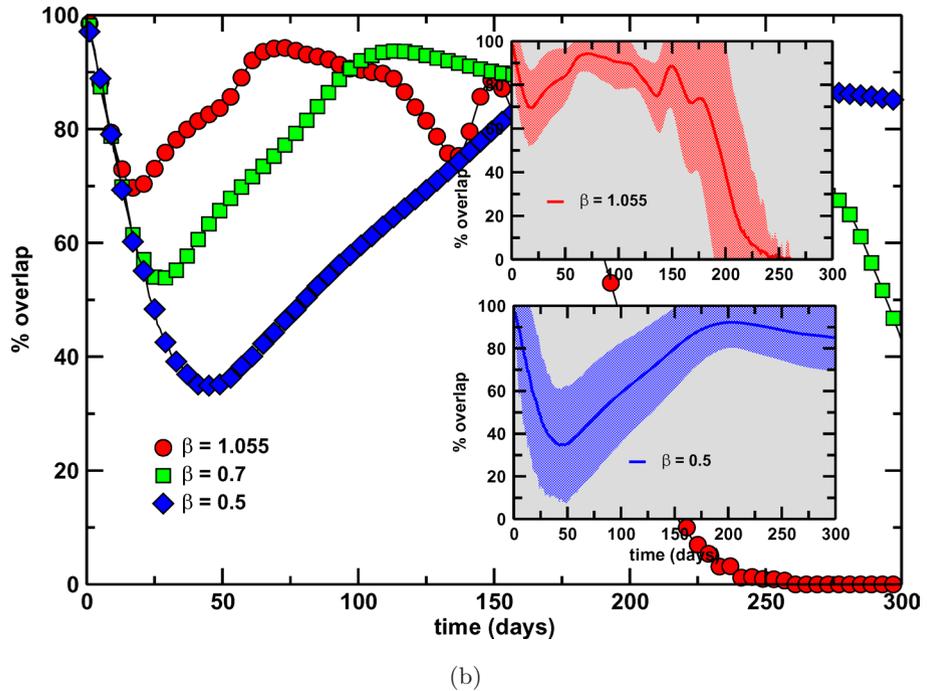
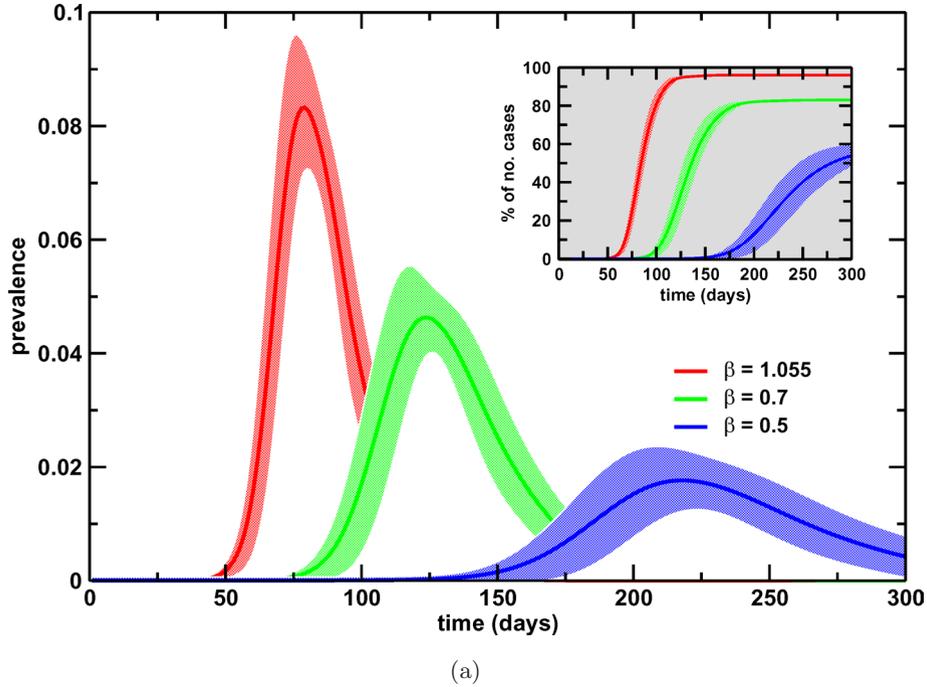


Fig. 1. Results obtained for different values of the virus infectiousness β , assuming $\epsilon^{-1} = 1.9$ days and $\mu^{-1} = 3$ days. Sub-populations are considered to be fully susceptible. (a) Full lines and (b) symbols correspond to average values and shaded areas, representing 95% confidence intervals. (a) Prevalence profiles. Inset: percentage of the cumulative number of cases with respect to the total population. (b) Overlap profiles. Insets: overlap profiles with corresponding fluctuations for $\beta = 1.055$ (top) and $\beta = 0.5$ (bottom).

increase, as shown in the insets of Fig. 1(b), due to the smaller numbers of active individuals, which lead to larger fluctuations from one couple of realizations to the other.

3.2. Initial immunity

We next consider the influence of changes in the initial fraction of susceptible individuals, taking this fraction uniform across cities for

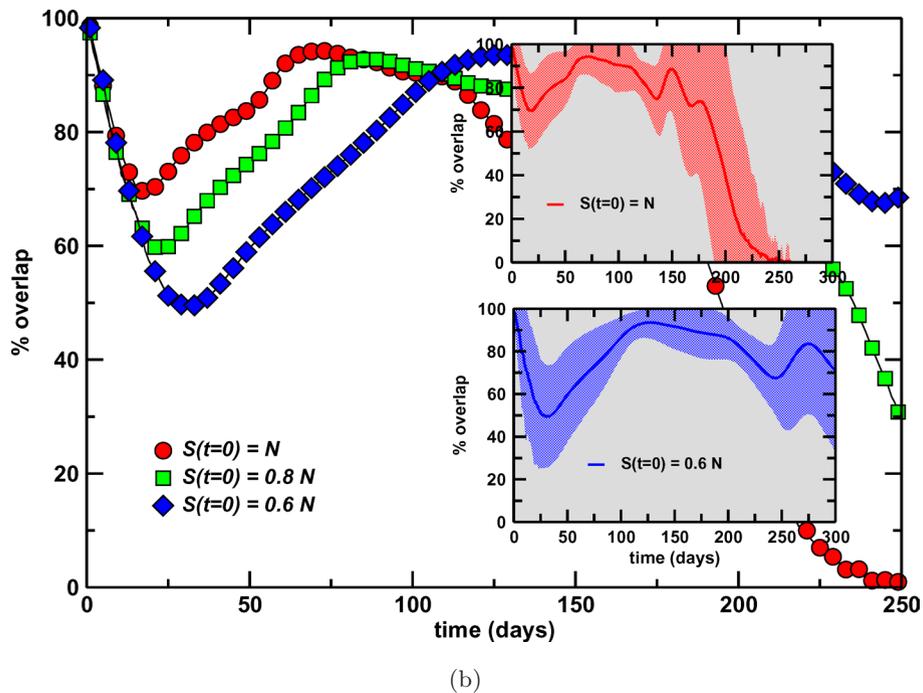
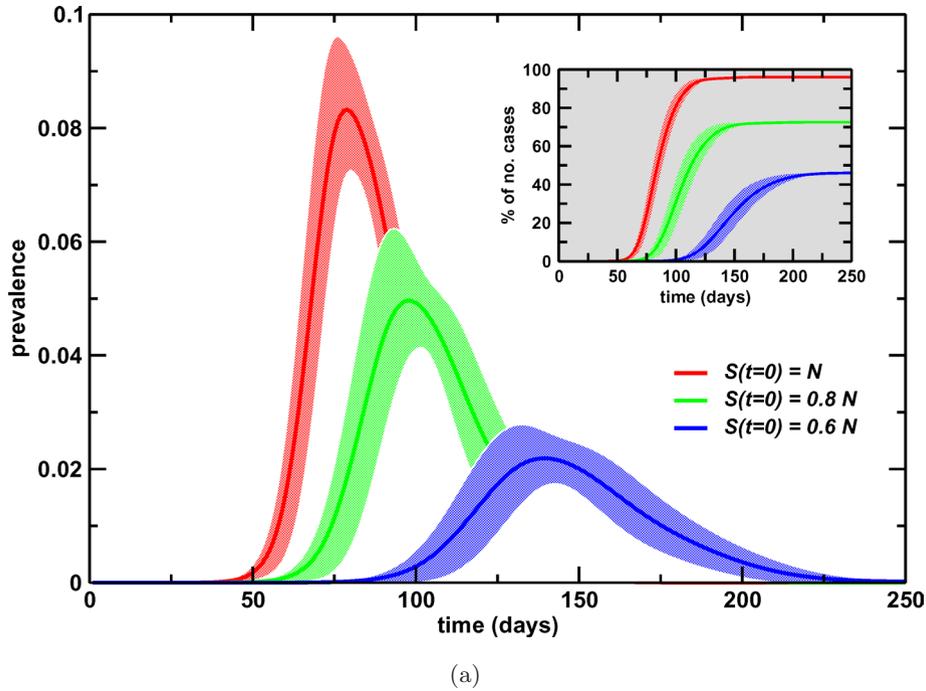


Fig. 2. Results obtained for different values of the initial fraction of susceptible population $S(t = 0) = \alpha N$, assuming $\beta = 1.055$, $\epsilon^{-1} = 1.9$ days and $\mu^{-1} = 3$ days. (a) Full lines and (b) symbols correspond to average values and shaded areas representing 95% confidence intervals. (a) Prevalence profiles. Inset: percentage of the cumulative number of cases with respect to the total population. (b) Overlap profiles. Insets: overlap profiles with corresponding fluctuations for $S(t = 0) = N$ (top) and $S(t = 0) = 0.6N$ (bottom).

simplicity: $\alpha = S_j(t = 0)/N_j$. Smaller values of α corresponds to situations in which a larger fraction of individuals in each sub-populations is immune to the virus at the start of the epidemic. The

prevalence profile assumes smaller values and the time scale is longer for larger α [see Fig. 2(a)], as expected since the virus finds a reduced pool of available susceptible individuals.

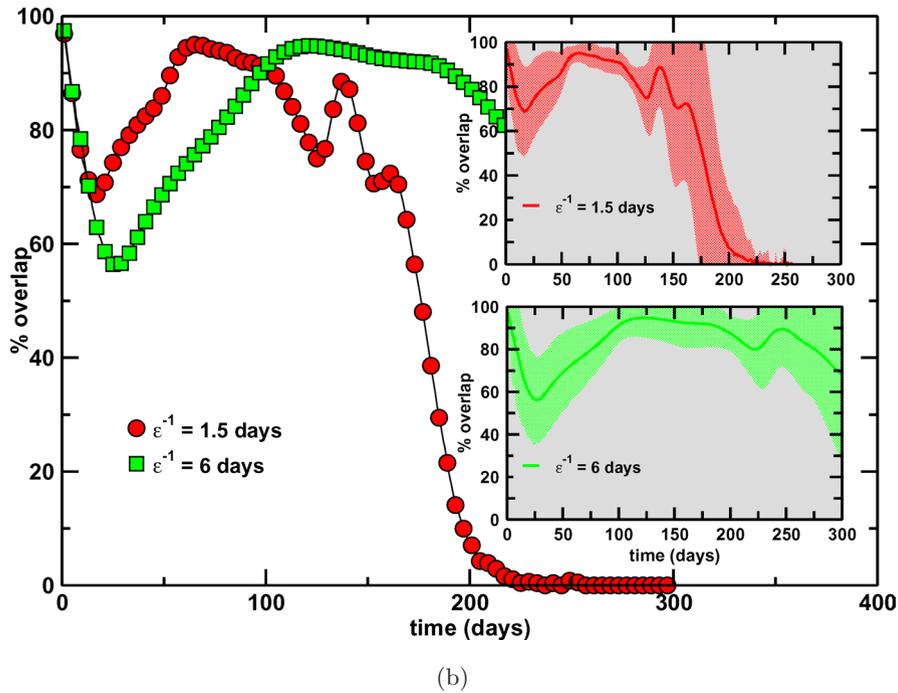
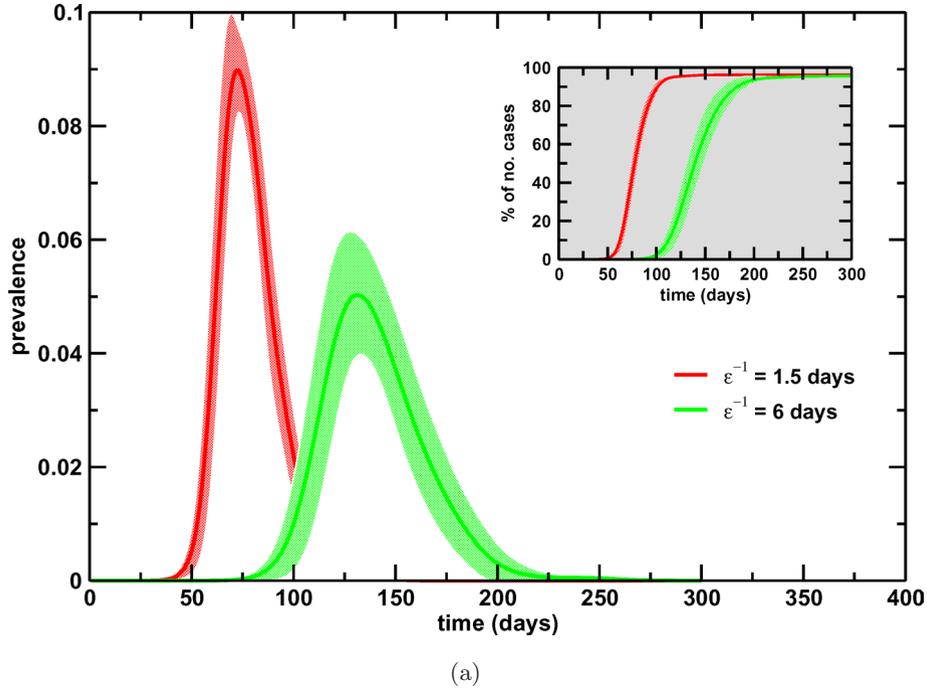


Fig. 3. Results obtained for different values of the average latency period ϵ^{-1} , assuming $\beta = 1.055$, $\mu^{-1} = 3$ days and a fully susceptible population. (a) Full lines and (b) symbols correspond to average values and shaded areas representing 95% confidence intervals. (a) Prevalence profiles. Inset: percentage of the cumulative number of cases with respect to the total population. (b) Overlap profiles. Insets: overlap profiles with corresponding fluctuations for $\epsilon^{-1} = 1.5$ days (top) and $\epsilon^{-1} = 6$ days (bottom).

As shown in Fig. 2(b), a decrease in α has essentially the same effect as a decrease in the value of β , and can be traced back to the same reason: a smaller number of cases leads to stronger decrease

of the predictability, due essentially to the larger fluctuations in the diffusion pattern. The relevant epidemiological parameter is indeed $R_{1-\alpha} = \alpha\beta/\mu$ which reduces to the reproductive ratio R_0 in a fully

susceptible population ($\alpha = 1$). Changes in α will affect $R_{1-\alpha}$ in the same way as changes in the virus infectiousness.

In summary, smaller values of the infectiousness β , or larger values of the initial immunity probability, correspond to less dangerous epidemic outbreaks but at the same time more difficult to forecast. Note that we keep here β and α constant over the whole propagation, while the application of containment measures can be modeled through an effective reduction of β over time. Numerically reproducing data of a successfully contained pandemic in order to validate an epidemic model may therefore be more challenging because the number of cases is drastically reduced thanks to efficient measures.

3.3. Latency time

Let us now consider, at fixed β and $S(t = 0)$, the influence of the latency time ϵ^{-1} . We show data for $\epsilon^{-1} = 1.5$ days and six days corresponding typically to seasonal influenza or SARS-like diseases, respectively.

As Fig. 3(a) shows, increasing latency times lead to slower propagation and a longer duration of the epidemic. The number of cumulative cases is unchanged [see inset of Fig. 3(a)] since the reproductive ratio is constant, not depending on ϵ .

Since the prevalence is lower at any given time for larger latency times, the trend observed is the same as when β decreases: larger fluctuations in the travel pattern are observed as ϵ^{-1} increases. However, a supplementary effect takes place in this case: for larger values of ϵ^{-1} latent individuals spend on average more time in traveling from one region to another before changing compartment and thus becoming infectious. The overall time since infection to recovery is longer, so that the disease will very likely propagate to different regions in different realizations. In summary, the faster the onset of symptoms and infectiousness, the more predictable the propagation pattern.

4. Conclusions

In this paper, we have considered the problem of the predictability of epidemic scenarios, within the global stochastic modeling framework developed in [Colizza *et al.*, 2006a, 2006b]. This approach allows to consider the propagation of epidemics on a worldwide scale by coupling evolution equations inside the various cities through transport terms.

The predictability of the disease propagation pattern is quantified by the overlap between realizations having the same initial conditions but different noise realizations. While the fluctuations due to the internal noise of the disease evolution inside each city are less relevant in affecting the reproducibility of an epidemic outbreak, the availability of a huge number of different traveling paths for latent and infectious individuals lead to a strong decrease of the predictability at short times. The overlap reaches a minimum during the initial stage of the outbreak and then increases until the epidemic peak is reached. Such a profile is due to the presence of two competing effects: on the one hand, given a certain number of individuals diffusing along the transportation network from a given initial city, the overlap between two realizations decreases because of the possible diversity in diffusion patterns; on the other hand, an increase in the number of individuals diffusing (which occurs as the epidemics evolves) leads to smaller fluctuations around the average pattern. In other words, when higher is the number of infected and latent individuals, the smaller will be the effect of the noise in the diffusion process. This readily implies that when the infectiousness is lower, or the latency time larger, providing a slower injection of infected individuals in the system, the diffusion noise has a larger impact and the predictability decreases. These results are a first assessment of the predictability that can be achieved with large scale computational approaches to epidemic forecasts and indicates that a careful theoretical analysis of the specific modeling framework is needed in order to define the appropriate confidence limits for the obtained predictions.

Acknowledgments

We thank the International Air Transport Association (IATA) for providing us with the world-wide airline database. A. Barrat and A. Vespignani are partially funded by the European Commission — contract 001907 (DELIS). A. Vespignani is partially funded by the NSF IIS-0513650 award.

References

- Anderson, R. M. & May, R. M. [1984] “Spatial, temporal and genetic heterogeneity in host populations and the design of immunization programs,” *IMA J. Math. Appl. Med. Biol.* **1**, 233–266.
- Baroyan, O. V., Genchikov, L. A., Rvachev, L. A. & Shashkov, V. A. [1969] “An attempt at large-scale

- influenza epidemic modeling by means of a computer," *Bull. Int. Epidemiol. Assoc.* **18**, 22–31.
- Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. [2004] "The architecture of complex weighted networks," *Proc. Natl. Acad. Sci. USA* **101**, 3747–3752.
- Bolker, B. M. & Grenfell, B. T. [1993] "Chaos and biological complexity in measles dynamics," *Proc. R. Soc. Lond. B* **251**, 75–81.
- Bolker, B. M. & Grenfell, B. T. [1995] "Space persistence and dynamics of measles epidemics," *Phil. Trans. R. Soc. Lond. B* **348**, 309–320.
- Chowell, G., Hyman, J. M., Eubank, S. & Castillo-Chavez, C. [2003] "Scaling laws for the movement of people between locations in a large city," *Phys. Rev. E* **68**, 066102.
- Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. [2006a] "The role of the airline transportation network in the prediction and predictability of global epidemics," *Proc. Natl. Acad. Sci. USA* **103**, 2015–2020.
- Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. [2006b] "The modeling of global epidemics: Stochastic dynamics and predictability," *Bull. Math. Bio.*, DOI 10.1007/s11538-006-9077-9.
- Eubank, S., Guclu, H., Anil Kumar, V. S., Marathe, M. V., Srinivasan, A., Toroczkai, Z. & Wang, N. [2004] "Modeling disease outbreaks in realistic urban social networks," *Nature* **429**, 180–184.
- Ferguson, N. M., Keeling, M. J., Edmunds, W. J., Gani, R., Grenfell, B. T., Anderson, R. M. & Leach, S. [2003] "Planning for smallpox outbreaks," *Nature* **425**, 681–685.
- Ferguson, N. M., Cummings, D. A. T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S. & Burke, D. S. [2005] "Strategies for containing an emerging influenza pandemic in Southeast Asia," *Nature* **437**, 209–214.
- Flahault, A. & Valleron, A.-J. [1991] "A method for assessing the global spread of HIV-1 infection based on air-travel," *Math. Pop. Studies* **3**, 1–11.
- Gardiner, W. C. [2004] *Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences*, 3rd edition (Springer, Berlin).
- Grais, R. F., Ellis, J. H. & Glass, G. E. [2003] "Assessing the impact of airline travel on the geographic spread of pandemic influenza," *Europ. J. Epidemiol.* **18**, 1065–1072.
- Grais, R. F., Ellis, J. H., Kress, A. & Glass, G. E. [2004] "Modeling the spread of annual influenza epidemics in the US: The potential role of air travel," *Health Care Manag. Sci.* **7**, 127–134.
- Grenfell, B. T. & Bolker, B. M. [1998] "Cities and villages: Infection hierarchies in a measles meta-population," *Ecol. Lett.* **1**, 63–70.
- Guimerà, R. & Amaral, L. A. N. [2004] "Modeling the world-wide airport network," *Europ. Phys. J. B* **38**, 381–385.
- Guimerà, R., Mossa, S., Turtschi, A. & Amaral, L. A. N. [2005] "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proc. Natl. Acad. Sci. USA* **102**, 7794–7799.
- Hufnagel, L., Brockmann, D. & Geisel, T. [2004] "Forecast and control of epidemics in a globalized world," *Proc. Natl. Acad. Sci. USA* **101**, 15124–15129.
- Keeling, M. J. & Rohani, P. [1995] "Estimating spatial coupling in epidemiological systems: A mechanistic approach," *Ecol. Lett.* **5**, 20–29.
- Lloyd, A. L. & May, R. M. [1996] "Spatial heterogeneity in epidemic models," *J. Theor. Biol.* **179**, 1–11.
- Longini, I. M. [1988] "A mathematical model for predicting the geographic spread of new infectious agents," *Math. Biosci.* **90**, 367–383.
- Longini, I. M., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D. A. T. & Halloran, M. E. [2005] "Containing pandemic influenza at the source," *Science* **309**, 1083–1087.
- Marro, J. & Dickman, R. [1998] *Nonequilibrium Phase Transitions and Critical Phenomena* (Cambridge University Press, Cambridge, UK).
- May, R. M. & Anderson, R. M. [1984] "Spatial heterogeneity and the design of immunization programs," *Math Biosci.* **72**, 83–111.
- Rvachev, L. A. & Longini, I. M. [1985] "A mathematical model for the global spread of influenza," *Math. Biosci.* **75**, 3–22.