Forecast and Control of Epidemics in a Globalized World

L. Hufnagel[†§¶], D. Brockmann[†], T. Geisel[†]

[†]Max-Planck-Institut für Strömungsforschung, Bunsenstrasse 10, 37073 Göttingen, Germany and
[§]Kavli Institute for Theoretical Physics, University of California Santa Barbara, CA 93106, USA

[¶]To whom correspondence should be addressed. E-mail: lars@chaos.gwdg.de

Supporting Online Material

The Master Equation for M Coupled Populations

The dynamics of disease transmission, recovery, and dispersal of individuals is schematically given by

$$\begin{aligned} S_i + I_i &\xrightarrow{\alpha} 2\,I_i \\ I_i &\xrightarrow{\beta} \emptyset \\ S_i &\xrightarrow{\gamma_{ij}} S_j \\ I_i &\xrightarrow{\gamma_{ij}} I_j, \end{aligned} \qquad \textbf{[1]}$$

where the index $i = 1, ..., M$ labels the populations. The entire dynamics is defined by this set of $2M(M+1)$ reactions. The $2M$-component vector $\mathbf{X} = \{S_1, I_1, ...S_M, I_M\}$ specifies the stochastic state of the system, and the probability density function $p(\mathbf{X}, t)$ obeys a combinatorial master equation

$$\partial_t\, p(\mathbf{X}, t) = \sum_{\mathbf{X}'} w(\mathbf{X}|\mathbf{X}')\, p(\mathbf{X}', t) - w(\mathbf{X}'|\mathbf{X})\, p(\mathbf{X}, t), \qquad \textbf{[2]}$$

which is defined by the probability rate $w(\mathbf{X}|\mathbf{X}')$ of making a transition from state $\mathbf{X}'$ to $\mathbf{X}$. For the process defined by Eq. **1**, this master equation reads explicitly

$$\begin{aligned} \partial_t\, p\left(\{S_1, I_1, ..., S_M, I_M\}, t\right) = \\ \alpha \sum_i (S_i + 1)\,(I_i - 1)\, p\left(\{S_1, I_1, ..., S_i + 1, I_i - 1, ..., S_M, I_M\}, t\right) \\ - \alpha \sum_i S_i I_i p\left(\{S_1, I_1, ..., S_M, I_M\}, t\right) \\ + \beta \sum_i (I_i + 1)\, p\left(\{S_1, I_1, ..., S_i, I_i + 1, ..., S_M, I_M\}, t\right) \\ - \beta \sum_i I_i\, p\left(\{S_1, I_1, ..., S_M, I_M\}, t\right) \\ + \sum_{ij} \gamma_{ij}\,(S_i + 1)\,(S_j - 1)\, p\left(\{S_1, I_1, ..., S_i + 1, I_i, ..., S_j - 1, I_j, ..., S_M, I_M\}, t\right) \\ - \sum \gamma_{ij}\, S_i\, S_j\, p\left(\{S_1, I_1, ..., S_M, I_M\}, t\right) \\ + \sum_{ij} \gamma_{ij}\,(I_i + 1)\,(I_j - 1)\, p\left(\{S_1, I_1, ..., S_i, I_i + 1, ..., S_j, I_j - 1, ..., S_M, I_M\}, t\right) \\ - \sum_{ij} \gamma_{ij}\, I_i\, I_j\, p\left(\{S_1, I_1, ..., S_M, I_M\}, t\right). \end{aligned} \qquad \textbf{[3]}$$

Along with an initial condition $p(\mathbf{X}, t = 0)$, this defines the temporal evolution of the process. The first two terms on the right account for disease transmission in each population, the third and fourth terms account for recovery from the disease in each population, and the last four terms describe probabilistic dispersal among the populations. A realization $\mathbf{X}(t)$ of the process defined by Eq. **3** represents the time course of an epidemic in the entire network of populations.

Input Data and Model Structure

Here we present in detail the modeling of severe acute respiratory syndrome (SARS) on the aviation network. The model consists of two parts, as illustrated in Fig. 5: the local infection dynamics and the traveling dynamics on the aviation network.

The aviation network

We have analyzed the global aviation traffic among the 500 largest airports. Table 2 depicts the raw data used to estimate the global traveling of individuals. Each flight f is specified by its departure $i_{\rm D}(f)$ and arrival $i_{\rm A}(f)$ airport, the aircraft type $t(f)$, and the days of the week on which the flight operates. The average number of passengers $m(f)$ that are transported by flight f per day from airport $i_{\rm D}(f)$ to $i_{\rm A}(f)$ is then given by

$$m(f) = \frac{1}{7}\, n_{\rm Op}(f) \times C_{t(f)},$$

where $n_{\rm Op}(f)$ is the number of days the flight operates per week. Apart from long time variations, two international flight schedules alternate: a summer and a winter schedule. The flights within each schedule repeat after 1 week. $C_{t(f)}$ is the capacity (i.e., the number of passengers) of the aircraft type that operates along flight f, which is unambiguously identified by its International Air Transportation Association (IATA) aircraft code. The total number of passengers per week that travel from airport i to j is

$$M_{ij} \;\; = \;\; \sum_f \delta_{ii_D(f)} \delta_{ji_A(f)}\, m(f),$$

where the sum runs over all flights of the aviation network.

Simulation and epidemiological determinants of SARS

The local infection dynamics is based on a stochastic metapopulation compartmental model for each region. A stochastic model is required, because fluctuations in case numbers can be large especially in the early stages of an epidemic. A schematic flow diagram of the transmission model is shown in Fig. 5B. In each region, we classify the population into susceptible, latent, infectious, and removed. Here, the removed class includes recovered, hospitalized, and deceased individuals. Multiple realizations of the stochastic model are required to estimate expectation values and the size of the fluctuations. Susceptible individuals S_i become infected and enter a latent stage L_i. They then progress to an infectious phase I_i. It is assumed that every infected individual eventually enters a hospital, dies, or recovers and thus does not participate in the dynamics any longer. Each region i in the global model uses the same classification of individuals. Furthermore, we assume the dynamics between the classes to be the same for all regions. The dynamical equations are closed by specifying the waiting time distribution in each class and the basic reproduction number ρ_0.

For SARS, the waiting time distributions within the classes have been investigated in great detail for Hong Kong(1, 2). Fig. 6A shows the distributions of times from the initial infection to the onset of symptoms $\tau_{\rm IO}$. The time from the onset of symptoms to admission $\tau_{\rm OA}$ is shown in Fig 6B. Both times together give the infection to admission time $\tau_{\rm IA} = \tau_{\rm IO} + \tau_{\rm OA}$. The distribution of latent times is not well known. We assume it to range from 2 to 3 days.

It remains to be specified how many transmissions an infected causes, on average, while being infectious. In our simulations, we assume that infections occur one at a time and infection times are independent. Then the infection process is characterized by the basic reproduction number ρ_0, which for SARS, lies between $2.2 \leq \rho_0 \leq 3.7$ during the early stages of the outbreak in Hong Kong, February 2003(2). Similar values have also been estimated for Singapore (3). Due to interventions, infection control, and quarantine, the basic reproduction number was reduced from the initial high to a value < 1 (see figure. 2C in ref. (2)).

The starting time of the simulation corresponds to the outbreak of the epidemic in Hong Kong on February 19, 2003. The simulation time is $t = 90$ days corresponding to March 20, 2003. To estimate the average number of infections and the range of the fluctuation for each country, 1,000 realizations of the epidemic were simulated.

Results of the Simulations

The results of our simulations are summarized in Table 3, together with the corresponding World Health Organization (WHO) data. Column 4 contains the outcome of the simulations and should be compared to the WHO data presented in column 2. In addition to the full simulation, we also investigated the spread of SARS for a model that neglects the stochastic infection dynamics outside the initial outbreak area. The results are presented in column 8. It is evident, that for countries with a high number of infecteds, this approximation fails. However, it is a good approximation for countries with a small number of infecteds.

1. Donnelly, C. A., Ghani, A. C., Leung, G. M., Hedley, A. J., Fraser, C., Riley, S., Abu-Raddad, L. J., Ho, L.-M., Thach, T.-Q., Chau, P., *et al.* (2003) *Lancet* **361**, 1761–1766.

2. Riley, S., Fraser, C., Donnelly, C. A., Ghani, A. C., Abu-Raddad, L. J., Hedley, A. J., Leung, G. M., Ho, L.-M., Lam, T.-H., Thach, T. Q., *et al.* (2003) *Science* **300**, 1961–1966.

3. Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., Gopalakrishna, G., Chew, S. K., Tan, C. C., Samore, M. H., *et al.* (2003) *Science* **300**, 1966–1970.

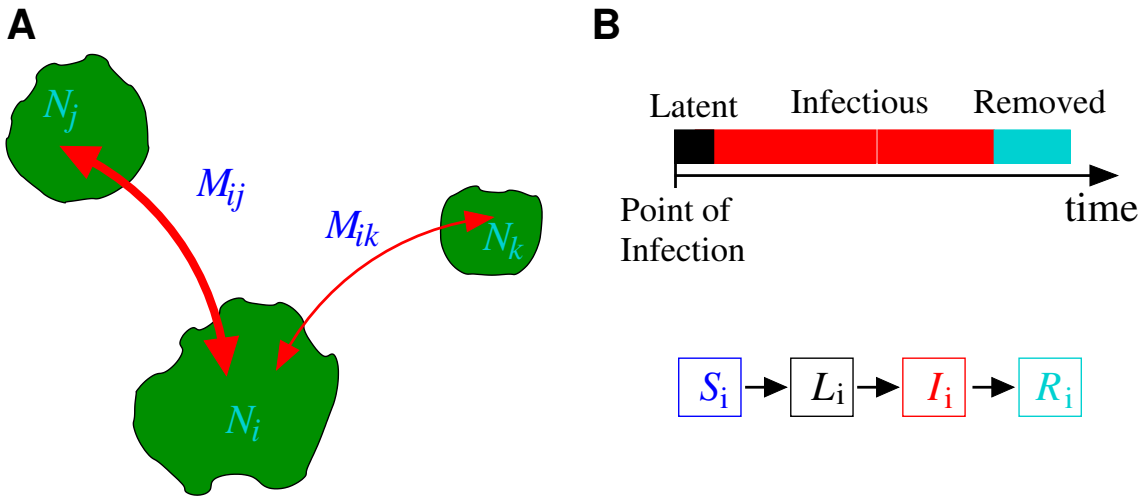

Fig. 5. Structure of the model for the worldwide spread of epidemics. (*A*) The whole population is composed of local urban regions i with N_i individuals. The traveling behavior of individuals between regions i and j is described by the matrix element M_{ij}. (*B*) In each region N_i individuals are in one of four stages: susceptibles, S_i, who can catch the disease; latents, L_i, who carry the disease but are not yet infectious; infecteds, I_i, and the removed, R_i, which are either recovered, immune, isolated, or deceased. The times individuals remain in the latent or infectious phase are drawn from γ shaped delay distribution (see figure 6).

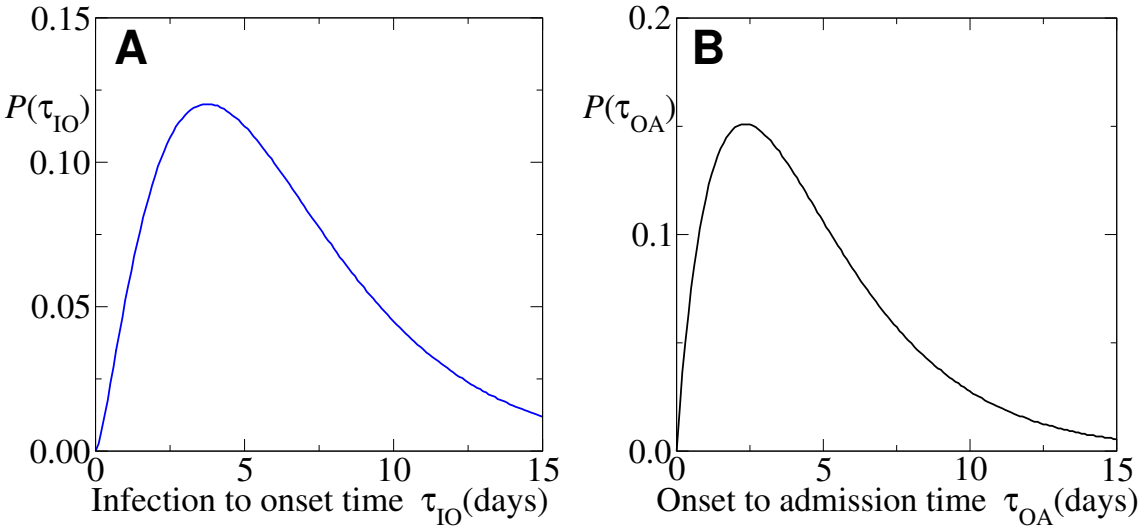

Fig. 6. (*A*) Distribution of infection-to-onset times τ_{IO} and (*B*) onset-to-admission times τ_{OA} used in the stochastic SARS model. The Data are reproduced form refs. 1 and 2.

Table 2. Illustration of the flight data used in the simulations

Days of	Departure		Arrival		Flight	Aircraft
Operation	Airport	Time	Airport	Time	Number	Type
MTWTFSS	ABE	6:40a	ATK	8:53a	*DL5521	CRJ
MTWTFSS	ABE	6:45a	ATL	8:58a	*DL5521	CRJ
MTWTFSS	ABE	12:45p	ATL	3:01p	*DL5534	CRJ
MTWTFSS	ABE	5:30p	ATL	7:40p	*DL5536	CRJ
MTWTFSS	ABE	9:00p	ATL	10:55p	*DL4582	CRJ
MTWTFSS	ABE	9:00p	ATL	11:00p	*DL4582	CRJ
MTWTF**	ABE	6:45p	BOS	8:17p	*US3903	DH8
MTWTF**	ABE	6:45p	BOS	8:17p	*US3903	DH8
MTWTFSS	ABE	7:50a	CLT	9:18a	US1785	319
MTWTFSS	ABE	7:50a	CLT	9:18a	*UA1876	319
****FS*	ABE	7:50a	CLT	9:26a	US684	734
****FS*	ABE	7:50a	CLT	9:26a	*UA1876	734
MTWTFSS	ABE	7:55a	CLT	9:31a	US684	734
MTWTFSS	ABE	7:55a	CLT	9:31a	*UA1876	734
MTWTFSS	ABE	7:55a	CLT	9:31a	US684	734
MTWTFS*	ABE	6:15a	DTW	7:48a	NW1895	DC9
MTWTFS*	ABE	6:15a	DTW	7:50a	NW1895	D9S
MTWTFS*	ABE	6:15a	DTW	7:50a	NW1895	D9S
MTWTFSS	ABE	9:15a	DTW	10:50a	NW1897	D9S

Each flight is characterized by 7 entries: The first column gives the days on which the flight operates; a star indicates nonoperation days. The next two columns specify the departure airport and time. The airport is abbreviated by its three-letter airport location code (see Internation Air Transportation (IATA), www.iata.com) (e.g., ATL, Atlanda). This information is followed by the corresponding arrival data. In column six, the international flight number is shown. The last column specifies the type of aircraft used by its IATA Aircraft Type Codes (e.g., *734*, Boeing 737-400 or *D10* McDonnell Douglas DC10).

Table 3. A comparison of the SARS case reports provided by the World Health Organization (WHO) and the results of our simulation for all countries with a reported case number ≥ 1

	WHO		Simulation					
Country	**5/20/03**	**5/30/03**	**Average**	η	**Minimum**	**Maximum**	**Average**	η_{D}
Hong Kong	1718	1739	1951	0.35	1373.9	2770.4	1890.5	0.32
Taiwan	383	676	318.2	0.55	184.0	550.3	67.6	0.37
Singapore	206	206	136.6	0.68	69.4	268.7	31.7	0.43
Japan	-	-	60.4	0.84	26.6	137.0	13.5	0.61
Canada	140	188	41.8	0.94	16.4	106.6	15.3	0.62
USA	67	66	65.9	0.84	28.4	152.7	10.8	0.72
Vietnam	63	63	49.2	0.86	20.7	116.3	9.7	0.71
Philippines	12	12	30.0	0.97	6.2	50.7	7.6	0.76
Germany	9	10	14.4	1.1	4.8	43.1	3.8	0.79
Netherlands	-	-	5.9	1.09	2.0	17.6	2.1	0.71
Bangladesh	-	-	10	1.15	3.2	31.6	2.6	0.76
Mongolia	9	9						
Italy	9	9	5.3	1.02	1.9	14.6	1.2	0.55
Thailand	8	8	35.4	0.89	14.5	86.8	8.6	0.71
France	7	7	7.6	1.09	2.6	22.6	1.9	0.68
Australia	6	6	27.0	1.05	10.1	72.5	7.9	0.74
Malaysia	7	5	17.7	1.05	6.2	50.7	6.1	0.76
UK	4	4	16.7	1.04	5.9	47.0	4.4	0.82
South Korea	3	3	14.9	1.17	4.6	48.0	5.8	0.78
Sweden	3	3						
India	3	3	6.8	1.13	2.2	21.1	1.9	0.69
Indonesia	2	2	14.9	1.05	4.9	48.3	4.0	0.81
Brazil	2	2						
Finland	1	1						
Colombia	1	1						
Switzerland	1	1	2.5	0.78	1.1	5.5	0.6	0.41
Spain	1	1	0.9	0.53	0.5	1.5	0.1	0.13
South Africa	1	1	5.6	1.21	1.7	18.9	3.7	0.81
Russia	-	1	3.4	0.84	1.4	7.8	0.7	0.44
Romania	1	1						
Ireland	-	1						
New Zealand	1	1	12.0	1.14	3.8	37.6	3.0	0.79
Kuwait	1	1						

The expected number of infecteds predicted by our model is estimated by averaging over 1000 realizations of the stochastic model. The range defined by column 6 and 7 was computed by means of the fluctuation range measure η (i.e., Eq. **8**). Column 8 and 9 show the average number of infected individuals and the estimated fluctuation range η_{D} of a modification of the model that neglects the stochastic infection dynamics outside the initial outbreak area.