

Analysis of COVID-19 in Tokyo: agent-based model and data assimilation

T. Miyoshi^a, S. Richard^{a,b}, C. Sun^{a,c}✉, L. Zhang^c

a) *Data Assimilation Research Team, RIKEN Center for Computational Science (R-CCS), Japan*

b) *Graduate School of Mathematics, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan*

c) *School of Science, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan*

E-mails: *takemasa.miyoshi@riken.jp, richard@math.nagoya-u.ac.jp, chang.sun@a.riken.jp, zhly.ok@gmail.com,*

1 Description of the model

The model consists in an extension of the well-known SEIR model, but adapted in the context of an agent-based model. It is based on seven compartments, as shown in Figure 1 which can be described as follows:

S The compartment of all susceptible agents, they do not play any role yet. The size of this compartment is irrelevant for the model, and thus a few parameters related to this compartment will hold an index ∞ ,

E The exposed agents. Right after infections, agents moved to E , they are not infectious so far,

I_a The infectious agents, not aware of their infectious condition and not taking special precautions. Some of these agents will be asymptomatic (never develop any symptoms), while some will become symptomatic and move to I_s . This later cohort corresponds to pre-symptomatic agents. Note of these agents are not recorded by health authorities so far.

I_s The infectious agents having clear symptoms and taking all necessary precautions. Most of these agents will look for medical support, but some will just quarantine themselves without seeking for medical support. This latter cohort will not be identified by medial system,

H The hospitalized agents, and more generally the treated agents. These agents can be treated in hospitals, at home or in any facilities. They are recorded by health authorities,

D The pass away agents, who were recorded while in H ,

R The agents who have recovered from the disease and were recorded while in H .

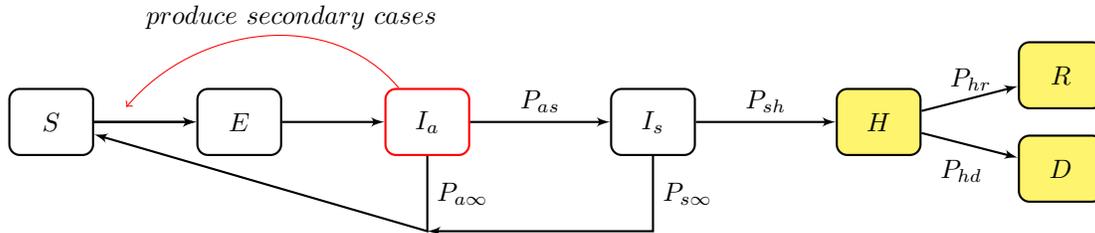


Figure 1: The compartments and the probabilities of each path

The different paths in Figure 1 follow certain probabilities. We provide in the Table 1 the different values and the sources. Note that these values will be discussed again with additional experiments. Note also that the probabilities P_{hd} and P_{hr} will be evaluated by the experiment, and therefore are time (\equiv day) dependent. In addition, one has $P_{hd}(t) + P_{hr}(t) = 1$ for any t . Let us take this opportunity to stress that the unit of time in this work is one day.

Name	Path	Probability of	value (%)	source
P_{as}	$I_a \rightarrow I_s$	becoming symptomatic	83	[1]
$P_{a\infty}$	$I_a \rightarrow S$	being asymptomatic	17	[1]
P_{sh}	$I_s \rightarrow H$	contacting health authorities	78	[4]
$P_{s\infty}$	$I_s \rightarrow S$	not contacting health authorities	22	[4]
P_{hd}	$H \rightarrow D$	dying	Evaluated with data	
P_{hr}	$H \rightarrow R$	recovering after hospitalization	Evaluated with data	

Table 1: Probabilities between compartments

In some of these compartments (Cpt), the number of days spent by agents is important. We list in Table 2 the necessary information about these durations and provide the sources. Some given distributions are provided in Table 3. Note also that the time duration T_h will be evaluated by the experiment, and therefore is time dependent.

Cpt	Name	Description	Value	source
E	T_e	Incubation period	3 days	Estimated period before becoming contagious
I_a	T_{as} $T_{a\infty}$	Time before becoming symptomatic Time before recovering for asymptomatic	Given in Table 3 Given in Table 3	[5]
I_s	T_{sh} $T_{s\infty}$	Time before moving to hospital Time before recovering	Given in Table 3 Irrelevant for the model	[3]
H	T_h	Time spent in hospital	Evaluated with data	

Table 2: Time spent in compartments

days	1	2	3	4	5	6	7	8	9	\mathbb{E}
T_{ar}	0	0	0	0	0.05	0.2	0.5	0.2	0.05	7
T_{as}	0	0.3	0.6	0.05	0.05	0	0	0	0	2.85
T_{sh}	0	0.1	0.2	0.5	0.15	0.05	0	0	0	3.85

Table 3: Distribution for T_{ar} , T_{as} , and T_{sh}

In compartment I_a , asymptomatic agents and pre-symptomatic agents can spread the disease. A simplified daily offspring distribution Od is provided in Table 4, and we refer for example to [6] for more details:

# of s. c.	0	1	2	3	4	5	$\mathbb{E}(\text{s. c.})$
Prob.	$p_0 = 0.5$	$p_1 = 0.35$	$p_2 = 0.12$	$p_3 = 0.01$	$p_4 = 0.01$	$p_5 = 0.01$	0.71

Table 4: Daily offspring distribution Od : number of secondary cases (s.c.)

One delicate question is the relation between the transmission coefficient for asymptomatic agents and the transmission coefficient for symptomatic agents. For our investigations, we shall rely on the result of the systematic review [1] which asserts that the relative risk of asymptomatic transmission is 42% lower than that for symptomatic transmission. As a consequence, we shall fix that the *relative infectivity coefficient* k is 0.58, meaning that the transmission of asymptomatic agents is k times the transmission coefficient of symptomatic agents. This factor k is slightly smaller but of a comparable scale compared to earlier investigations, see for example [2]. This factor will be discussed again with additional experiments.

Let us now set $r_t \in [0, 1]$ for a time dependent multiplicative factor which take into account the real interaction between agents, and which shall be updated on a daily basis. This factor depends clearly on individual behavior but also on non-pharmaceutical interventions. As a consequence, for asymptomatic agents in I_a , the daily production of second generation infections is given by the daily offspring production (for # secondary cases ≥ 1) multiplied by the factor $k \cdot r_t$, while for pre-symptomatic and symptomatic agents in I_a the daily production of second generation infections is given by the daily offspring production (for # secondary cases ≥ 1) multiplied by the factor r_t . As a consequence, the daily expected value for the reproduction number R_t is given by the formula

$$\begin{aligned}
R_t &= (k \cdot P_{a\infty} \cdot \mathbb{E}(T_{a\infty}) + P_{as} \cdot \mathbb{E}(T_{as})) \cdot \mathbb{E}(\text{s. c.}) \cdot r_t \\
&= (0.58 \cdot 0.17 \cdot 7 + 0.83 \cdot 2.85) \cdot 0.71 \cdot r_t \\
&= 2.17 \cdot r_t.
\end{aligned}$$

However, note that the above quantity corresponds just to an expected value, while we shall get later on a distribution for R_t .

2 Evolution process

The data provided by health authorities for Tokyo start on March 6th 2020. However, the epidemic had already started in Tokyo before this date, since the departure of the Diamond Princess from Yokohama took place on January 20th 2020. For this reason, and after several trials, we have fixed the following initial conditions for our experiment: January 17th 2020. This date corresponds to 50 days before the start of the data available for Tokyo.

Description	Initial values
Initial date	January 17, 2020
Initial number of infected agents	Uniformly at random in $\{3, 4, 5, 6, 7\}$
r_0	Uniformly at random in $[0, 1]$
$T_h(0)$	15 days, initial estimation of health authorities
$P_{hd}(0)$	Uniformly at random in $[0, 0.05]$

Table 5: Initial conditions

On a daily basis the parameters have some freedom to change their values. In order to describe their evolution, let us use the notation $\text{TN}(\mu, \sigma; a, b)$ for the truncated normal distribution of mean μ , variance σ^2 , minimum value a and maximal value b . For the evolution of r_t with $t \geq 1$ we choose randomly its value according to the distribution $\text{TN}(r_{t-1}, 0.05; 0, 1)$. The evolution of P_{hd} is also given by using a truncated normal distribution, namely for $t \geq 1$ the value $P_{hd}(t)$ is chosen randomly according to the distribution $\text{TN}(P_{hd}(t-1), 0.0025; 0, 0.05)$.

Since the parameter T_h is integer valued, its evolution is slightly more complicated, but nevertheless follows a scheme similar to the previous two parameters. The mean value of $T_h(t)$ is chosen randomly according to the distribution $\text{TN}(T_h(t-1), 0.75; 4, 19)$, where the minimum and maximum values have been fixed according to some information provided by health authorities. Then, one constructs a distribution supported on the greatest integer less than or equal to $T_h(t)$ and the least integer greater than or equal to $T_h(t)$, and such that its expectation is equal to $T_h(t)$. The agents entering the H on that day are then assigned a time in H chosen at random according to this distribution.

Each agent in the compartment I_a can infect on a daily basis susceptible agents belonging to the compartment S . We now describe this process. Let us denote by

$$\text{MN}(x_j, n; d_j) \equiv \text{MN}(0, 1, 2, 3, 4, 5, n; d_0, d_1, d_2, d_3, d_4, d_5)$$

the multinomial distribution for n trials in the set of values $x_j \in \{0, 1, 2, 3, 4, 5\}$ with the probability of x_j given by d_j . We also denote by the vector $X = (X_0, X_1, X_2, X_3, X_4, X_5)$ one realization of this distribution. Clearly, $\sum_{j=0}^5 X_j = n$, and the expectation value for X_j is nd_j for any $j \in \{0, 1, 2, 3, 4, 5\}$.

On a given day t , assume that the compartment I_a contains n asymptomatic agents and m pre-symptomatic agents. Then, the n agents will infect $\sum_{j=1}^n jX_j$ susceptible agents, where X is one realization of the multinomial distribution $\text{MN}(x_j, n; d_j)$ with $d_j = k \cdot r_t \cdot p_j$ for $j \in \{1, 2, 3, 4, 5\}$ and where p_j was given in the offspring distribution of Table 4. In addition, one has $d_0 = (1 - k \cdot r_t) + k \cdot r_t \cdot p_0$. Similarly, the m agents will infect $\sum_{j=1}^n jY_j$ susceptible agents, where Y is one realization of the multinomial distribution $\text{MN}(x_j, m; d_j)$ with $d_j = r_t \cdot p_j$ for $j \in \{1, 2, 3, 4, 5\}$ and where p_j was given in the offspring distribution of Table 4. In addition, one has $d_0 = (1 - r_t) + r_t \cdot p_0$.

3 Data assimilation

As highlighted in Figure 1, three compartments which are associated with real data, namely H , D and R . Some uncertainties must be attached to these data, and these uncertainties are commonly called observation error and will be denoted by σ . For our model, we shall consider two of these observation errors constant over time, while one will be time dependent. In fact, since the daily variations of H are quite important, the corresponding uncertainties will take them into account. More precisely, if we write $H(t)$ for the number of hospitalized agents at time t , we set

$$\begin{aligned} \sigma_H(t) &= \sqrt{(0.3H(t) + 4|H(t) - H(t-1)|)^2 + 400}, \\ \sigma_R &= 2000, \\ \sigma_D &= 100. \end{aligned}$$

The first expression and the different values have been chosen after several trials. As shown below, even if these values look very large, they lead to a successful selection of the most suitable particles.

On a daily basis, a weight w is assigned to each particle. We denote by N the number of particles, keeping in mind that this number will be kept constant during the simulation. Let i denote the index of the particles, and let $H_i(t)$, $R_i(t)$ and $D_i(t)$ denote the values of these three compartments for this particle at time t . One first computes the unnormalized weight $W_i(t)$ for $t \geq 1$ by the formula

$$W_i(t) := w_i(t-1) \cdot \exp \left(-\frac{(H_i(t) - H(t))^2}{2\sigma_H(t)^2} - \frac{(R_i(t) - R(t))^2}{2\sigma_R^2} - \frac{(D_i(t) - D(t))^2}{2\sigma_D^2} \right),$$

with the convention that $W_i(0) = \frac{1}{N}$. Then, the normalized weight at time t and for the particle i is given by

$$w_i(t) := \frac{W_i(t)}{\sum_{j=1}^N W_j(t)}.$$

If we keep computing the weight with the above formula, it will soon turn out that very few particles will concentrate all the weight, and that nearly all other particles would end up with negligible weights. For dealing with this problem, a process of resampling has to be implemented. For a given time t , this process consists in considering again a multinomial distribution $\text{MN}(i, N, w_i(t-1))$ with $i \in \{1, \dots, N\}$ indexing the set of particles, and by assigning a new weight $\frac{1}{N}$ for time t to the N particles selected by one realization of this distribution. Clearly, some particles will appear several times, but their trajectories will soon diverge due to the randomness involved on a daily basis.

One still has to decide when resamplings have to be organized. For that purpose, let us set the effective number of particles

$$N_{eff}(t) := \frac{1}{\sum_{i=1}^N w_i(t)^2},$$

and observe that if $w_i = \frac{1}{N}$ for $i \in \{1, \dots, N\}$, then $N_{eff} = N$, while if $w_i \approx 1$ for one i , and $w_j \approx 0$ for all $j \neq i$, then $N_{eff} \approx 1$. For this reason, N_{eff} is often used as an indicator of the number of particles still playing a role in the process. As a consequence, we shall use the following rule: If $N_{eff} < \frac{N}{10}$, then a resampling has to take place. In practice, as the model tends to go unstable if resamplings take place too infrequently, a resampling is forced if the most recent resampling took place 15 days ago.

Let us stress that the computation of the weights and the implementation of the resampling start only on day 50 of the simulation, namely on March 6. Indeed, before this date, no data exist, and therefore the particles just evolve freely and independently.

References

- [1] O. Byambasuren, M. Cardona, K. Bell, J. Clark, M.-L. McLaws, P. Glasziou, *Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis*, J. Association of Medical Microbiology and Infectious Disease Canada Volume 5 Issue 4, December 2020, 223–234.
- [2] D. Buitrago-Garcia, D. Egli-Gany, M. J. Counotte, S. Hossmann, H. Imeri, A. M. Ipekci, et al., *Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis*, PLoS Med 17(9): e1003346, 2020.
- [3] <https://www.m3.com/open/iryoIshin/article/849820/>
- [4] Osaka prefecture government, *Citizens awareness and behavior change of measures against COVID-19*, http://www.pref.osaka.lg.jp/hodo/attach/hodo-40479_4.pdf
- [5] A. M. Pollock, J. Lancaster, *Asymptomatic transmission of covid-19*, BMJ 2020;371:m4851.

- [6] X.-K. Xu, X.F. Liu, Y. Wu, S. Taslim Ali, Z.Du, P. Bosetti, E. H. Y. Lau, B. J. Cowling, L. Wang, *Reconstruction of Transmission Pairs for Novel Coronavirus Disease 2019 (COVID-19) in Mainland China: Estimation of Superspreading Events, Serial Interval, and Hazard of Infection*, *Clinical Infectious Diseases*, Volume 71, Issue 12, 15 December 2020, Pages 3163–3167, <https://doi.org/10.1093/cid/ciaa790>.