

Supplementary Text to ...

Olivier Thomine^{1,*}, Samuel Alizon², Marc Barthelemy³,
Corentin Boennec², Mircea T. Sofonea²

¹ LIS UMR 7020, CNRS, Aix Marseille University, France

² MIVEGEC, CNRS, IRD, Université de Montpellier, France

³ Institut de Physique Théorique, CEA, CNRS-URA 2306, F-91191, Gif-sur-Yvette, France

* Corresponding author, olivier.thomine@univ-amu.fr

Contents

S1 Supplementary Methods	1
S1.1 Geographical structure	2
S1.2 Demographic model	2
S1.3 Infection model	4
S1.4 Simulation specifications	5

S1 Supplementary Methods

The approach used in Epidemap consists in coupling three models. The first model dispatches each agent to a building, depending on the national demographic distribution (from INSEE INSEPop and the buildings properties (from OpenStreetsMap OSM). The second model, based on that from Barthelemy2018, determines the buildings that each agent will daily visit and socials interactions with other agents located in the same distant buildings. Finally, the third model captures the life-history properties of the infection disease in infected agents and is based on that from SofoneaEtAl2020.

21 **S1.1 Geographical structure**

22 We use the freely available OpenStreetMap (OSM) database to extract all the points of interest
23 for this study. Note that the accuracy of this database is partially high in France because the
24 official national ‘Cadastre’ database is merged into the OSM database. These databases are
25 formatted in ASCII XML, with a size of approximately 80 *GB* for France.

26 OSM labels residential buildings specifically. We use the surface of these buildings and their
27 geographical position to allocate each agent to a ‘home’ (residency) building. For the mobility
28 model, we also include all the other types buildings (*e.g.* hospitals, schools, airport, commercial
29 centres), where agents can meet.

30 At this time, the initial database size contains $4.1 \cdot 10^8$ nodes, and $4.8 \cdot 10^7$ buildings to pre-
31 compute.

32 **S1.2 Demographic model**

33 The Institut National de la Statistique et des Etudes Economiques (INSEE) provides us with a
34 high-level resolution of the distribution of the French population. Here, we use the information
35 about the full population in each city in 2016 to allocate agents to different locations. By
36 combining this database and OSM’s, we know the number of residents in each building of
37 each city. More precisely, agents are allocated to buildings proportionally to their floor-surface
38 projection. The number of agent in each building is given by equation S1, with N_k the number
39 of agent in the building k , N_{city} the number of residents in the city considered, S_k the floor-
40 surface projection of building k , $\sum_i S_i$ the total surface of all residential buildings in the city,
41 $|\cdot|$ the entire part of a number, and $\alpha \in [0, 1]$ a scaling parameter such that $\sum_k N_k = N_{city}$.

$$N_k = \left\lfloor N_{city} \frac{S_k}{\sum_i S_i} + \alpha \right\rfloor. \quad (\text{S1})$$

	$\mathbf{lognorm}(\mu = 2, \sigma = 0.88, d = 0.5l)$	INSEE statistic
< 10 km	32.9 %	33.7 %
10 km < X < 20 km	30.6 %	30.5 %
20 km < X < 30 km	15.5 %	16.0 %
30 km < X < 50 km	12.7 %	12.5 %
50 km < X < 100 km	6.8 %	5.8 %
> 100 km	1.5 %	1.5 %

Table S1: Lognormal randomly-chosen values and INSEE statistic of the distance between residential and distant building. Original datas (2) and fitting.

42 The age of each agent is randomly generated and follows the age pyramid of the french resi-
43 dents, given by INSEE.

44 Individual movement patterns have been study in (1). According to this study, we assume that
45 each person visits on average 2 distant buildings per day (where it can meet other agents).

46 To select a distant building, we first select all the buildings located at a given distance of the
47 agent's home (± 50 m), and then select the target with a probability proportional to its floor-
48 surface projection. This approach allows larger buildings to be visited by a higher amount of
49 people than smaller ones. The distance l itself is assumed to follow a log-normal distribution,
50 with $PDF(d) = \mathbf{lognorm}(\mu = 2, \sigma = 0.88, d = 0.5l)$. This parameterisation yields results
51 that are very consistent with the INSEE data (2) (tab. S1.2).

52 If no building is found at the randomly-generated distance, the agent does not interact with any
53 other agent for this movement round.

54 Every day, each agent can interact randomly and non-exclusively with other agents present in
55 the same building at the same time. The maximum number of interactions is limited to 17
56 persons in a distant building and 5 at residential building.

57 **S1.3 Infection model**

58 Epidemap is versatile and can be adapted to simulate many infectious disease epidemics. Here,
59 we focus on COVID-19 spreading in France and parametrize the epidemiological and clinical
60 dynamics following the non-Markovian model of (3). Importantly, this approach allows to take
61 into account memory effects involving the age of the infection of each contaminated individual,
62 e.g. in the case of critical infections, the probability of being hospitalized increases with the
63 time elapsed since infection.

64 For each interaction between a susceptible and an infected agent, we assume a constant proba-
65 bility of contamination $b = 6\%$ PHE multiplied by normalized daily infectiosity **Weibull**($k =$
66 2.24 , $\lambda = 5.42$), with d the number of day since contamination. We can simulate the effect of
67 non-pharmaceutical interventions (e.g. mask-wearing) by decreasing this probability.

68 Upon infection, an individual of age a either develops a mild or a severe infection with proba-
69 bility θ . Independently of the severity, the contagiousness varies every day d post infection. For
70 simplicity, these variations are assumed to follow the serial interval $\zeta(d)$, which is the distribu-
71 tion of the number of days between the onset of the symptoms in a ‘infecting’ host and that in
72 an ‘infected’ host (4). After 15 days, and according to the serial interval used, infectiousness is
73 negligible and mildly-infected hosts move the the recovered.

74 Severely-infected individuals have a daily probability $\eta(d)$ to be hospitalised with two trajecto-
75 ries. A fraction $1\psi_a$ have a very low prognosis with a daily probability to die of $v(d)$. The other
76 fraction will follow an intensive care trajectory with a daily probability $\rho(d)$ to leave the ICU
77 either via recovery (with probability μ) or via death.

78 All recovered hosts are assumed to be immune to the infection until the end of the simulation.

79 Further details about the underlying distributions are available in SofoneaEtAl2020.

Table S2: **Parameters used in the simulations.** LN stands for log-normal, We for Weibull

Name	Description	Value	Reference
N	population size (number of agents)	6.6E7	(5)
l	individual dispersal kernel distance	LN($\mu = 2, \sigma = 0.88, d = 0.5l$)	fit of (2)
N_1	max daily number of agent meet (distant)	17	user-defined
N_2	max daily number of agent meet (home)	5	user-defined
$b(d)$	transmission probability per contact	6% $\zeta(d)$	(6)
$\zeta(d)$	contagiosity, d day after contamination	We($k = 2.24, \lambda = 5.42, d$)	(7)

80 **S1.4 Simulation specifications**

81 The 100 simulations use some parameters to specify for each of the 3 models (agents repartition,
 82 mobility and infection). These parameters are resumed in the tabS2. Concerning the infection
 83 model, any kind of compartmental model can be applied.

84 The computing code is written in Fortran 90 (F90) with Open Multi-Processing (OMP) ap-
 85 proach to parallelize the computation and contains $\simeq 18,000$ lines. An huge effort was made to
 86 reduce the memory print of the code, which runs with less than 64GB Random Access Mem-
 87 ory (RAM) for $6.6 \cdot 10^7$ agents. A full epidemic simulation, which represent approximately 300
 88 days, is performed in 2 hours on a standard personal computer (12-cores AMD Ryzen 9). The
 89 statistics are written in ASCII format and the graphical outputs use the compressed Paraview
 90 format.

91 **References and Notes**

- 92 1. C. Schneider, V. Belik, T. Couronné, Z. Smoreda, M. Gonzalez, *J R Soc Interface* **10** (2013).
- 93 2. INSEE, De plus en plus de personnes travaillent en dehors de leur commune de résidence,
94 <https://www.insee.fr/fr/statistiques/2019022> (2016).
- 95 3. M. T. Sofonea, *et al.*, *Epidemics* in press (2021).
- 96 4. H. Nishiura, N. M. Linton, A. R. Akhmetzhanov, *Int J Infect Dis* **93**, 284 (2020).
- 97 5. INSEE, Populations légales 2017. recensement de la population régions, départements,
98 arrondissements, cantons et communes, [https://www.insee.fr/fr/](https://www.insee.fr/fr/statistiques/4265511)
99 [statistiques/4265511](https://www.insee.fr/fr/statistiques/4265511) (2019).
- 100 6. Variant Technical group, *Public Health England. Technical briefing 3.* (2021).
- 101 7. H. Nishiura, N. Linton, A. Akhmetzhanov, *International Journal of Infectious Diseases* **93**,
102 284 (2020).