Modeling of COVID-19 epidemic in Italy

Introduction

We provide technical details on the implementation of the Global Epidemic and Mobility (GLEAM) model to describe and project the spreading of the COVID-19 epidemic in Italy.
It is worth mentioning that the current transmission within Italy is generated within the context of the entire globe, including the early transmission in China. The focus on the presentation of the results here is the current situation in Italy, including the re-opening of the country starting on May 4, 2020.

Modeling Framework

Metapopulation approach

The GLEAM framework is based on a metapopulation approach in which the world is divided into geographical subpopulations. The entire planet is divided into cells with resolution of 15 x 15 arc minutes (approximately 25 x 25 kilometers). Subpopulations are constructed from these cells using a Voronoi tessellation of the Earth's surface, with each subpopulation centered around a major transportation hub obtained from the International Air Transport Association (IATA) and OAG database. Hubs generally correspond to major urban areas and airports. By considering the distance between the cells and the transportation hubs, we assign each cell to a specific hub; this process generates over 3,200 subpopulations (census areas) worldwide. We then use highly detailed data on the cell populations to characterize the subpopulations. Other subpopulation attributes, such as the age structure of the population, health infrastructures, etc., are added according to available data.

Human mobility between subpopulations is represented on a network. This mobility data layer identifies the numbers of individuals traveling from one sub-population to another. The mobility network is made up of different kinds of mobility processes, from short-range commuting between nearby subpopulations to intercontinental flights. The airline system layer integrates air travel mobility containing the list of daily passenger flows between airport pairs (origin and destination) worldwide. Individuals in the model travel on airplanes according to an explicit dynamic that considers the probability for each individual in the subpopulation to travel on a specific route.

To model short-range mobility such as commuting or car travel, we rely on databases collected from the Offices of Statistics of 30 countries on five continents. The full dataset includes more than 80,000 administrative regions and over five million commuting flow connections between them. To overcome differences in the spatial resolution of the commuting data across different countries, we define a worldwide homogeneous standard for GLEAM. Using this standardized approach, we are better able to model mobility between countries with a shared border. Where data are not available, the short-range mobility layer can be generated synthetically by relying on the "gravity law" and the more recent "radiation law," both calibrated using the real data available. Briefly, these approaches assume more frequent travel to nearby subpopulations and less frequent travel to more distant subpopulations. The short-range mobility network is overlaid with the airline network forming the mobility system of the GLEAM synthetic world. **Figure 1** shows a representation of the geographical resolution of the model for Italy.

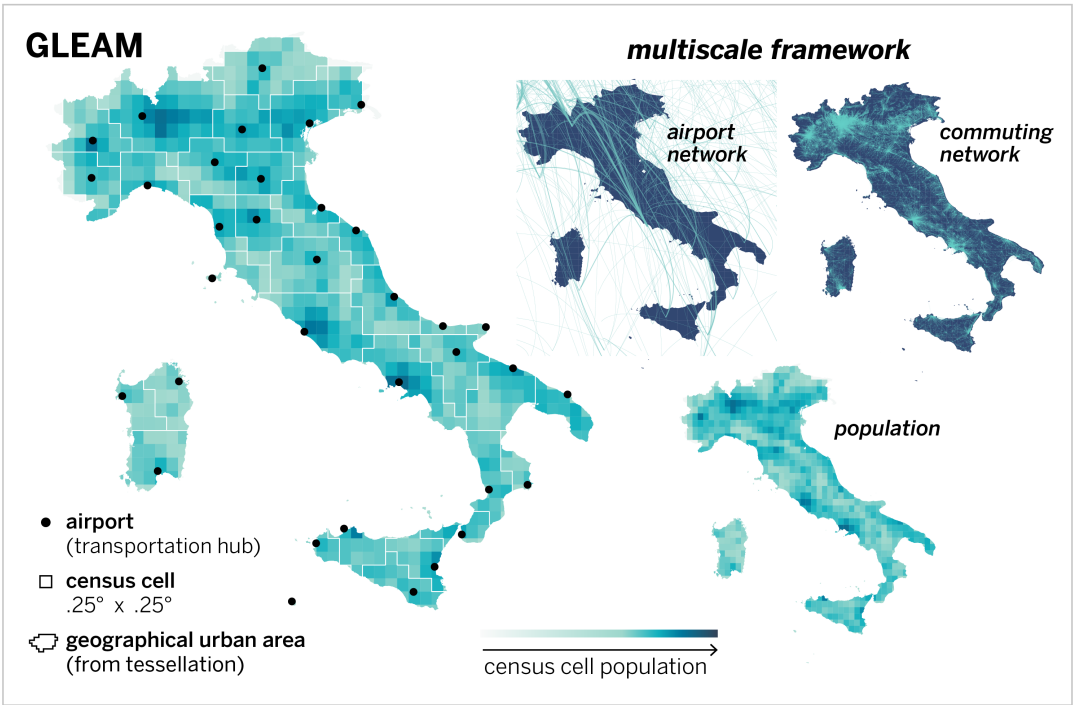

Figure 1. Schematic representation of the Global Epidemic and Mobility (GLEAM) model for Italy. The world surface is divided into census cells that are assigned to subpopulations centered around transportation hubs. The population layer describing the census cells is coupled with two mobility layers, the short-range commuting layer and the long-range air travel layer.

Infection transmission dynamics

Superimposed on the worldwide population and mobility layers is an agent-based epidemic model that defines the infection and population dynamics. Individuals (agents) move and transmit the infection via their interactions with other people. The infection dynamics take place within each subpopulation and assumes a classic SLIR-like compartmentalization scheme for disease progression. Each individual at any given point in time is assigned to a compartment corresponding to their particular disease-related state (being, e.g., susceptible, exposed, symptomatic, hospitalized, recovered). Individuals stochastically transition between compartments based on underlying assumptions about the parameters from the available literature that define the natural history of disease, such as the incubation period and age-specific infection fatality rate. In the GLEAM model, human mobility is allowed to vary as a function of disease status. In many cases, clinical symptoms are associated with reduced or no mobility of the sick individual. Reduced or no mobility also occurs in compartments representing the isolation or quarantine of individuals.

GLEAM defines a synthetic world in which we can simulate the unfolding of epidemics. Initial conditions are set specifying the number and location of individuals capable of transmitting the infection. The GLEAM model is then able to track over time the proportion of the synthetic population in each disease compartment across all subpopulations. At the start of each simulated day, travelers move to their destinations via the flight network. The probability of air travel changes from day to day, varies by age group, and can be generalized to consider the effects of location specific airline traffic reductions. Short-range mobility (e.g. commuting) varies between workdays and weekends, by age group, and by disease status. The arrival time of the infection into a new subpopulation is the day on which the first infected traveler arrives. This seed individual may then go on to infect others in the subpopulation. Each full day is simulated using 12 distinct time steps, and this process is repeated for every simulated day. Individuals and their traveling patterns are tracked as shown in the pseudo-code for the GLEAM algorithm (**Figure 2**).

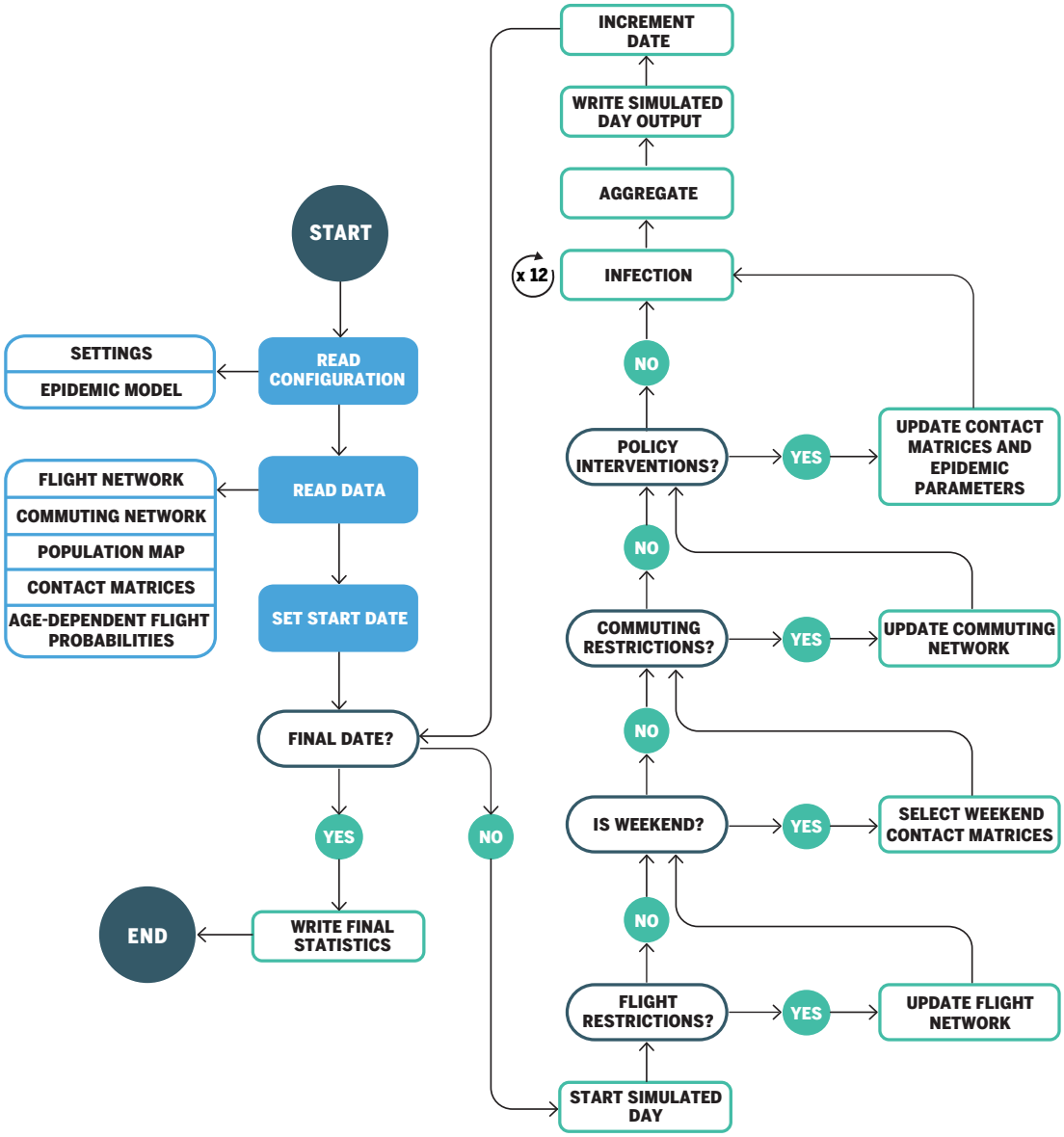

Figure 2. Pseudo-code of GLEAM's algorithm.

The GLEAM model and the detailed presentation of the underlying algorithms have been previously published [1,2] and the model has been used in the analysis and projections of several disease outbreaks including the 2009 H1N1 pandemic [3], the West African Ebola virus epidemic [4], the Zika virus epidemic in the Americas [5].

Transmission contact patterns

Using detailed sociodemographic data from publicly available sources, ranging from macro data, such as census data, to micro data, such as surveys on socio demographic features, we construct representative synthetic populations for different countries around the world. In particular we focus on population features such as age structure, household composition, school structure, and employment rates. Individuals within these populations are assigned realistic age-specific contact patterns. Using the synthetic populations of interacting individuals from the agent-based model, we construct age-stratified contact matrices for the most common social settings in which individuals spend their time. Main settings where individuals can interact include households, schools, workplaces, and the general community. For each location, these age-based contact patterns are encoded in a contact matrix F^k, whose elements F_{ij}^k describes the average frequency of contact between a given individual of age i and individuals of age j in setting k. The methodology used to build the contact pattern from the synthetic populations is detailed in Fumanelli et al. [7] and Mistry et al.[8].

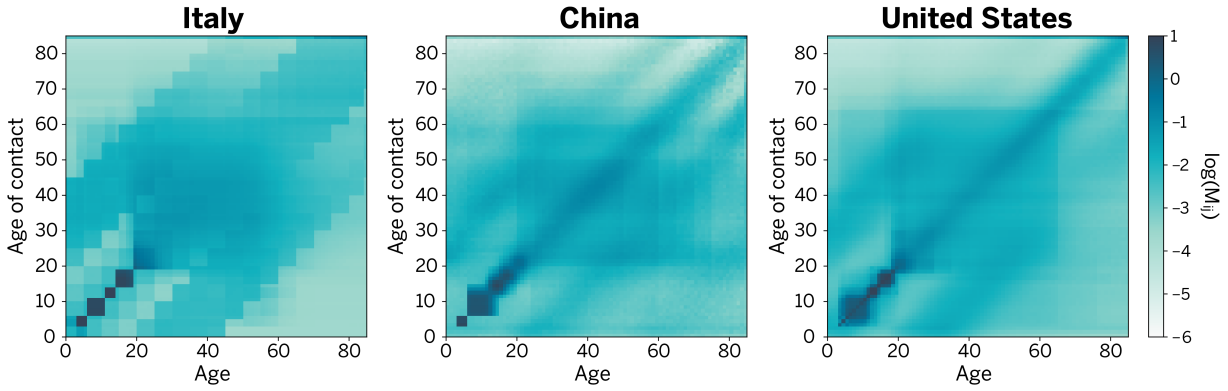

Figure 3. Contact matrices by age built from the synthetic populations, for Italy, China, and United States (among others) when we consider the aggregation of all the settings.

We focus on four settings where individuals interact: households (H), schools (S), workplaces (W), and the general community (C). Here we adopt the frequency dependent (mass action) transmission model, with the implicit assumption that an increased population density has no effect on the per capita contact rate between individuals [6]. In **Figure 3** we provide a visual representation of some of these matrices. From **Fig. 3**, where we show the contact matrices for Italy, China, and United States, we can observe the differences in the age contact patterns of each country. Although all matrices have a main diagonal representing increased contact among people of similar age, usually the couples in a population, the secondary diagonals, representing contact between children and parents, are not always present. This becomes

evident in China, known for their one-child policy in most of its regions, where we can see that children have much more contact with adults than adults with children. As contact patterns can differ even within the same country, using a country level contact matrix is not adequate if we want to capture the heterogeneities in populations.

In our model, the basic reproductive number R_0 not only depends on the disease parameters but also on the contact patterns within a particular region. This means that even when the disease parameters are the same in different places we should expect differences in R_0 and the attack rate driven by differences in contact patterns.

COVID-19 Model

A compartmental representation of the disease under study is used to model human-to-human transmission. Within each subpopulation, individuals can occupy one of the following compartments: Susceptible (S), Latent (L), Infectious (I) and Removed (R). Individuals move through compartments with transitions mathematically defined by chain binomial and multinomial processes to preserve the discrete and stochastic nature of the transmission and disease evolution process. Susceptible individuals can acquire the virus through contacts with individuals in the infectious compartment, and become latent, meaning they are infected but cannot transmit the disease yet. Latent individuals progress to the infectious stage with a rate inversely proportional to the length of the latent period (which we assume to have the same duration as the incubation period). Infectious individuals progress into the recovered stage with a rate inversely proportional to the length of the infectious period. The sum of the mean latent and infectious periods defines the generation time. Disease carriers travel during the entire latent and infectious periods and are not traveling when they enter the removed stage. The removed stage includes those who can no longer infect others because they are recovered, isolated, hospitalized, or dead.

Each subpopulation's disease dynamic is coupled with the other subpopulations through the mechanistically simulated travel and commuting patterns of infected individuals. The model works in discrete time steps, simulating air travel and compartmental transitions (where the force of infection considers both the infection dynamics and the short-range movement of individuals). Data are reported at the scale of single day, aggregating results at the desired level of geographic resolution. All the technical details of the model have been previously published [9].

Parameter ranges for the latent period, infectious period, and generation time are derived from early estimates of the epidemiological characteristics of COVID-19 [10-13]. We have

performed a sensitivity analysis by considering different combinations of average latent and infectious periods, detection rates, initial conditions, and a generation time (Tg) ranging from 6 to 11 days. Details and sensitivity analysis on all parameters are reported in [12]. In the following we report the results for generation time Tg = 6.5 days. For the key time-to-event intervals we use values reported in the literature [14-16]. The obtained posterior distribution provides an average reproductive number R_0 = 2.83 [90% CI 2.80-2.87], and a doubling time measured at Td = 3.3 days [90% CI 3.2-3.4]. The obtained values are in the same range as previous analyses based on early COVID-1 data. Although the calibration obtained for different generation times provides different posterior distributions for R_0, in the early stages of the epidemic the prevalence of infections and case importations are determined by the epidemic growth rate and the obtained results are consistent with other modeling efforts[16-21]. Age-specific infection fatality rates considered here are from reference [14].

Global model calibration

We assume a starting date of the epidemic that falls between 11/15/2019 and 12/1/2019, with 40 cases seeded by zoonotic exposure. For each generation time Tg, we perform an Approximate Bayesian Computation considering the international importations to estimate the posterior distribution of the basic reproductive number R_0 of the outbreak. We simulate epidemics with R_0 in the range 1.5 to 4.0, sampled with a uniform prior. This allows us to calculate the distribution P(D) for the evidence D, and for each value of R_0, the likelihood $P(D|R_0)$. From these distributions we can calculate the posterior probability $P(R_0|D)$ of interest (see Fig. S1B). The evidence D is the growth rate of COVID-19 cases imported to international locations during the exponential growth period of the epidemic. We select the simulated epidemics that match the observed number of cumulative imported cases by January 23, 2020, before the travel ban in Wuhan, China, with a tolerance accounting for the 40% probability of detecting an importation [14]. We consider only statistically independent importation events by date of arrival at international destinations (details available in Ref [9]). Then, $P(R_0 = x|D)$ is computed as the number of simulations where R0 = x and the evidence constraint D is satisfied over the total number of simulations where $R_0 = x$.

COVID-19 epidemic specific mitigation policies in Italy

The simulated epidemics include changes in mobility attributable to the travel ban in Wuhan. We implemented long-range travel restrictions beginning on January 23, 2020, and decreased local commuting patterns starting on January 25, 2020. Travel limitations in mainland China were modeled by using de-identified and aggregated domestic population movement data between Chinese provinces for February 2020 as derived from Baidu Location-Based Services

(LBS). Starting early February 2020, more than 60 airline companies suspended or limited flights to mainland China, and a number of countries including the US, Russia, Australia, and Italy had also imposed government issued travel restrictions.

To estimate the seeding of the epidemic outside mainland China, we must assume the level of detection and isolation of imported infections in each country. There are several recent estimates for the rate of detection of imported infections across countries, suggesting overall values in the range of 30% to 40% [22-24]. We use the classification proposed in Ref [22], stratifying countries in three groups; namely high, medium and low surveillance capacity according to the Global Health Security Index [25]. We report a baseline scenario where high, medium, and low surveillance countries have a 40%, 20% and 10% detection rate, respectively.

After the first case in Italy was confirmed on January 31, 2020, a series of interventions were considered. On January 31, flights to and from China were suspended, and later in February many municipalities in northern Italy identified as epicenters of several clusters were placed under quarantine. The different interventions were implemented at the different times, according to the evolution of the epidemic in different regions:

- Starting in the north of Italy, on February 25, schools were closed and mobility restrictions (50%) were imposed on Lombardy, Veneto, Piedmont, Emilia-Romagna, Liguria, and Friuli.
- On March 8, the Lombardy region was under lockdown, or "stay at home" situation. This kind of interventions were adopted as well partially in the regions of Veneto, Emilia-Romagna, and Piedmont.
- On March 12, the entire country was put under lockdown, without allowing any sort of mobility between regions.
- After March 23, Italy has been under a strict "stay at home" order, where only essential businesses were allowed to keep functioning.

We have an additional calibration step for Italy by using an information loss criterion on the actual deaths data for the model selection. We then explore different exit strategy by modulating the contact matrices for Italy according to the following scenarios:

- **Lockdown:** This scenario has been calibrated on the past epidemiological data and considers that only essential work activities are allowed. This means school are closed, interaction in workplaces is reduced by 80%, and casual social interaction is reduced by 90%. We also assume a 20% transmissibility reduction due to the awareness of population and behavioral changes.

- **Full Lift:** Starting on May 4th, 2020 social distancing measures are lifted. Full mobility resumed. Schools are open according to schedule.

- **Lift 1:** Starting on May 4th, 2020 we assume that schools are closed, interaction in workplaces is reduced by 50%, and a 90% community social interaction reduction. Mobility is reduced of 50%. We also assume a 20% transmissibility reduction due to the awareness of population and behavioral changes. Total transmissibility reduction amounts to 60%.

- **Lift 2:** Starting on May 4th, 2020 e assume that schools are closed, interaction in workplaces is reduced by 30%, and a 50% community social interaction reduction. Mobility is reduced of 50%. We also assume a 20% transmissibility reduction due to the awareness of population and behavioral changes. Total transmissibility reduction amounts to 50%.

School closure and workplace reduced interactions are modeled following [26]. We implement contact matrices at country level resolution. The mobility restrictions are imposed on top of the "social distancing" mitigations described above.

Limitations

These results were obtained under several assumptions. The first is that we use modeling estimates for the effect of school closures, smart working and social distancing effects on the transmissibility of SARS-CoV-2. The model does not include currently pre-symptomatic transmission. All estimates do not consider the likely introduction of additional specific suppression or mitigation policies issued to lower the transmissibility in specific regions that experience elevated epidemic activity. On the other hand, modelling estimates for the impact of stay at home policies is driven by data on contact patterns from China. These results may not be directly generalizable as uptake and implementation of these policies will vary from country to country. Similarly, parameters characterizing the natural history of disease may vary from country to country. The current implementation of the model focuses on the impact of social distancing policies; thus, it does not address the role of other strategies such as contact tracing in reducing transmission. These should still be considered as part of any epidemic response. The model does not consider seasonal drivers on the transmissibility of SARS-CoV-2, such as temperature or humidity. Finally, we are not considering superspreading events and differential transmissibility across age brackets.

Even in the presence of these limitations we hope that the information provided here can be of help in assessing the impact of the COVID-19 epidemic in Italy. There are very large

uncertainties around the transmission of SARS-CoV-2, the effectiveness of different policies and the extent to which the population is compliant to social distancing policies. The presented material is based on modeling scenario assumptions informed by current knowledge of the disease, and subject to change as more data become available. Future decisions on when and for how long to relax policies will need to be informed by ongoing surveillance. Additional modeling and data studies are required to assess the level and effectiveness of additional non-pharmaceuticals interventions required to lift current social distancing interventions.

References

1. Balcan, Duygu, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J. Ramasco, and Alessandro Vespignani. "Multiscale mobility networks and the spatial spreading of infectious diseases." Proceedings of the National Academy of Sciences 106, no. 51 (2009): 21484-21489.
2. Balcan, Duygu, Bruno Gonçalves, Hao Hu, José J. Ramasco, Vittoria Colizza, and Alessandro Vespignani. "Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model." *Journal of computational science* 1, no. 3 (2010): 132-145.
3. Tizzoni, Michele, Paolo Bajardi, Chiara Poletto, José J. Ramasco, Duygu Balcan, Bruno Gonçalves, Nicola Perra, Vittoria Colizza, and Alessandro Vespignani. "Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm." *BMC medicine* 10, no. 1 (2012): 165.
4. Gomes, Marcelo FC, Ana Pastore y Piontti, Luca Rossi, Dennis Chao, Ira Longini, M. Elizabeth Halloran, and Alessandro Vespignani. "Assessing the international spreading risk associated with the 2014 West African Ebola outbreak." *PLoS currents* 6 (2014).
5. Zhang, Qian, Kaiyuan Sun, Matteo Chinazzi, Ana Pastore y Piontti, Natalie E. Dean, Diana Patricia Rojas, Stefano Merler et al. "Spread of Zika virus in the Americas." *Proceedings of the National Academy of Sciences* 114, no. 22 (2017): E4334-E4343.
6. Keeling, M. J. & Rohani, P. Modeling infectious diseases in humans and animals (Princeton University Press, 2011).
7. Fumanelli, Laura, Marco Ajelli, Piero Manfredi, Alessandro Vespignani, and Stefano Merler. "Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread." *PLoS computational biology* 8, no. 9 (2012).
8. Mistry D., Litvinova M, Pastore y Piontti A., et al. "Inferring high-resolution human mixing patterns for disease modeling", arXiv:2003.01214 [q-bio.PE] (2020).
9. Chinazzi, Matteo, Jessica T. Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti et al. "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak." *Science* (2020) DOI: 10.1126/science.aba9757.

10. Backer, Jantien A., Don Klinkenberg, and Jacco Wallinga. "Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020." *Eurosurveillance* 25, no. 5 (2020).
11. Huang, Chaolin, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang et al. "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China." *The Lancet* 395, no. 10223 (2020): 497-506.
12. Li, Qun, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren et al. "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia." *New England Journal of Medicine* (2020).
13. Zhang, Juanjuan, Maria Litvinova, Wei Wang, Yan Wang, Xiaowei Deng, Xinghui Chen, Mei Li et al. "Evolving epidemiology of novel coronavirus diseases 2019 and possible interruption of local transmission outside Hubei Province in China: a descriptive and modeling study." *medRxiv* (2020).
14. Verity, Robert, Lucy C. Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg et al. "Estimates of the severity of coronavirus disease 2019: a model-based analysis." *The Lancet Infectious Diseases* (2020).
15. Ferguson, Neil, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia et al. "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand." (2020).
16. Kissler, Stephen M., Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Marc Lipsitch. "Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period." Science (2020).
17. J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. Lancet **395**, 689–697 (2020). doi:10.1016/S0140-6736(20)30260-9pmid:32014114
18. Gatto, Marino, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Luca Carraro, Renato Casagrandi, and Andrea Rinaldo. "Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures." Proceedings of the National Academy of Sciences (2020).
19. Z. Du, L. Wang, S. Cauchemez, X. Xu, X. Wang, B. J. Cowling, L. A. Meyers, Risk for Transportation of 2019 Novel Coronavirus Disease from Wuhan to Other Cities in China. Emerg. Infect. Dis. 10.3201/eid2605.200146 (2020).doi:10.3201/eid2605.200146pmid:3205347
20. Lai, Shengjie, Nick W. Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R. Floyd, Amy Wesolowski et al. "Effect of non-pharmaceutical interventions to contain COVID-19 in China." (2020).
21. Kucharski, Adam J; Russell, Timothy W; Diamond, Charlie; Liu, Yang; Edmunds, John; Funk, Sebastian; Eggo, Rosalind M; Centre for Mathematical Modelling of Infectious Diseases COVID-1; (2020) *Early dynamics of transmission and control of COVID-19: a mathematical modelling study*. The Lancet. Infectious diseases, 20 (5). pp. 553-558. ISSN 1473-3099 DOI: https://doi.org/10.1016/S1473-3099(20)30144-4
22. R. Niehus, P. M. De Salazar, A. Taylor, M. Lipsitch, Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. www.medrxiv.org/content/10.1101/2020.02.13.20022707v2 (2020).

23. P. M. De Salazar, R. Niehus, A. Taylor, C. O. Buckee, M. Lipsitch, Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases.www.medrxiv.org/content/10.1101/2020.02.04.20020495v2 (2020).
24. K. Sun, J. Chen, C. Viboud , Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. Lancet Digital Health 10.1016/S2589-7500(20)30026-1 (2020). doi:10.1016/S2589-7500(20)30026-1
25. F. Pinotti, L. Di Domenico, E. Ortega, M. Mancastroppa, G. Pullano, E. Valdano, P. Boelle, C. Poletto, V. Colizza, Lessons learnt from 288 COVID-19 international cases: importations over time, effect of interventions, underdetection of imported cases. www.medrxiv.org/content/10.1101/2020.02.24.20027326v1 (2020).
26. De Luca, Giancarlo, Kim Van Kerckhove, Pietro Coletti, Chiara Poletto, Nathalie Bossuyt, Niel Hens, and Vittoria Colizza. "The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium." *BMC infectious diseases* 18, no. 1 (2018): 29.