# What if COVID-19 Started in Europe

Alex Procelewski (s4272196), Ana Terna (s4396618), Borna Massari (s4490754), Channa Dias Perera (s4400739), Hasti Haddadi (s4460995), Kees Wierenga (s4228456), Tamara Tataru (s4333314)

***Abstract*—The COVID-19 virus put epidemic modelling into the spotlight, becoming an important part of almost every country's public policy-making process. Despite that, creating a model that will accurately predict the spread of the coronavirus disease (2019) is still a challenge. Previous modelling efforts focused on the virus originating in China's Hubei Province and spreading through an Europe that was aware of the virus. We have created a model named WICSIE (What if COVID-19 Started In Europe) which follows a Continuous-Time Random Walk (CTRW) model as well as a Susceptible-Infected-Recovered model which simulates the spread of COVID-19 based on population density. We also take into account that the population may have their immunity reduced over time as well as the affect of of long-distance travel. We believe that this model will be useful when comparing the efficacy of European policy with regard to the spread of Covid, as it allows one to compare the effectiveness of policy, as opposed to no control.**

***Index Terms*—COVID-19, simulation, modelling, continuous-time, random walk, Go**

## I. Introduction

As summarised by Ciotti et al. [1] the severe acute respiratory syndrome coronavirus, better known as SARS-CoV-2 which causes disease knows as COVID-19, is a betacoronavirus of the Sarbecovirus subgenus. These viruses are enveloped, positive-sense single-stranded RNA viruses that bind to the angiotensin-converting-enzyme 2 receptor of human cells. In December 2019, after an uptick of pneumonia in Wuhan, Hubei Province, China, this virus was isolated for the first time from inoculation of human epithelial cells. Epidemiological study traced the cases back to the Huanan Seafood market. Similar to other respiratory viruses, droplet transmission and aerosols are both seen as the leading spreading routes of the virus. Furthermore, the competence of transmission is linked to other respiratory viruses as well, which leads to an estimated reproduction value between 1.4 and 2.5 for the new SARS-CoV variant.

According to World Health Organization expert advisor Prof. David Heymann, the high infection rate is partially due to the increased globalisation of European citizens, since travelling by plane is more accessible and affordable than ever[2]. This made it possible for a middle class citizen to access international travel and spread the virus rapidly around the globe. In this paper we will refer to citizens as agents, to ensure consistency throughout the whole paper. Compared to when SARS-CoV-1 caused a global epidemic in 2003[3], the number of flights has more than doubled [4]. In addition, earlier coronaviruses that caused similar symptoms, likely developed in the same way as the current pandemic. But, contrary to present time, they likely circulated locally and then gradually spread to the neighbouring countries and onward throughout the world [2].

Another factor Prof. David Heymann mentions, is the increased population density. This results in people not only living closer to each other but also closer to various species of animals, either domestic (used for food or as pets) or wild. The aforementioned creates a scenario which grants infectious viruses the circumstance to cross the species barrier that typically separates the humans from the animals [2]. Between the first report, on 31 December 2019 [5] in Wuhan, to the Health organisations concerning dozens of pneumonia cases, and the first confirmed case outside of main land China, 13 January 2020, there are 13 days. This is an evidence of the high spread of the virus. The virus arrived in Europe, more precisely France, on 24th of January. Germany and Italy followed within just a few days. As of January 31$^{st}$, 2022 there have been more than 1.7 million reported deaths and 143 million confirmed infections of COVID-19 across the entirety of Europe [6]. While the European Centre for Disease Prevention and Control had nearly a month to prepare, the European politics had suggested no particular measures, leading to a poorly prepared Europe.

Most countries approached epidemic modelling as a way to predict the behaviour of the highly infectious virus, trying to simulate it's spread in order to determine the effects of various policy decisions. This paper will take a retrospective approach, asking **What if COVID started in Europe**?.

In our simulation, we have selected rural Northern Ireland around Charlestown as the origin of the virus.

## II. Methodologies

Pandemic spread is a complex mathematical model with many variables and parameters. To illustrate, during the COVID-19 pandemic, there have been very different behaviours and case numbers in South Korea and the US and public behaviour have been significantly distinguishable. [7] A simulation of COVID-19 would therefore have to include all those different parameters for all 195 countries in the world. For the sake of this paper, the countries simulated are only in Western and Central Europe. While WICSIE allows for modelling different behaviour in regions and depending on COVID-19 cases, the simulations will only follow one shared behaviour, which this research paper will elaborate on later.

### A. Implementation of the simulation

WICSIE is divided into two phases:

1) Initialisation
2) Step simulation, where we derive the state of the pandemic at the next time step (the time between time steps is 1 day)

*1) Initialisation:* The initialisation step requires a heatmap, as seen in Figure 1. This heatmap has been taken from an Online Platform [8] and then cross verified with each countries' authorities respectively (Statistisches Bundesamt, Eurostate, Centraal Beureau Statistieken, Agence Nationale de Statique et de la Demographie, Instituto Nacional de Estadistica). It is then posterised and blurred interchangeable to reduce the number of colours and remove shades, while still preserving the information the images contain. The colour of a pixel on the heatmap corresponds to the population density within the given geographic region of the heatmap.
Subsequently, the heatmap is translated into ids at each pixel, representing the distinct colours. Those ids get translated into multipliers at each pixel, which are then used to distribute *agents* at each location, resulting in a similar distribution to the countries' respective populations. The process can also be seen in Figure 1.
Finally, we start the virus spreading process by infecting the *agents* from Northern Ireland, around Charlestown, with a 40% chance of infection. The map is initialised with all the *agents* and an epi-centre of the pandemic exists.

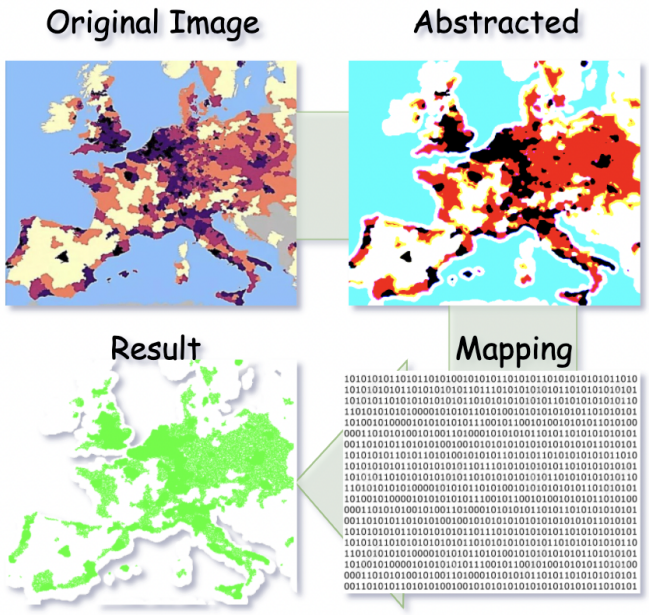


Fig. 1. Population Heatmap to *agent* Placement

*2) Simulation:*
*a) Movement:* To model the movement, we make use of a continuous-time, random-walk system, which estimates distance moved and the time spent in a single location with a probabilistic model:

$$P(\Delta d) \sim |\Delta d|^{-1-\alpha} \tag{1}$$
$$P(\Delta t) \sim |\Delta t|^{-1-\beta} \tag{2}$$

Where $\Delta d$ is the distance moved after waiting and $\Delta t$ is the time spent at a single location. Song et al. [9] found that $\alpha = 0.55 \pm 0.05$ and $\beta = 0.8 \pm 0.1$, which will be the values we use for $\alpha, \beta$ in our movement simulation. After determining the distance, we chose a uniformly random direction for the *agent* to move in.
Note that the $\sim$ notation indicates an asymptotic equivalence between functions. To approximate this equivalence, we model the ratio of left and right hand sides of the equation as a simple ratio. That is:

$$P(\Delta d) = a \cdot |\Delta d|^{-1-\alpha} \tag{3}$$
$$P(\Delta t) = b \cdot |\Delta t|^{-1-\beta} \tag{4}$$

Song et al. suggests that the distance you can reasonably cover in one step is at most 100 km, and the amount of time one would wait for in a single place is at most 17 hours. Thus, to find $a$ and $b$ we have to solve

$$1 = \int_0^{100} a \cdot |\Delta d|^{-1-\alpha} d\Delta d \tag{5}$$
$$1 = \int_0^{17} b \cdot |\Delta t|^{-1-\beta} d\Delta t \tag{6}$$

However, as $-1-\alpha < 0$ and $-1-\beta < 0$, the right hand side of the equations are infinite. Thus, we have to change our lower bound. We have chosen $0.05$ as this bound. Thus, solving these equations, we have $a \approx \frac{1}{9.3}, b \approx \frac{1}{13}$. hence, by taking the inverse functions, we get:

$$|\Delta d| = (P(\Delta d) \cdot 9.3)^{1/(-1.55)} \tag{7}$$
$$|\Delta t| = (P(\Delta t) \cdot 13)^{1/(-1.8)} \tag{8}$$

Forward, to estimate a distance/time jump, we generate a uniformly distributed number in the range $[0, 1]$, which we then insert into formulas 7, 8 to determine the distance/time jump.
To speed up the simulation, we scaled the resulting distance and time steps by a factor of 5.
However, human movement is not always local, as *agents* might travel large distances through air journeys. To account for this, we modelled an "extreme" jump, where the *agent* is transported at a random distance and the distance travelled is uniformly distributed between 0 and the distance of the map. The chance of such a jump occurring in our simulation is 3%, which was arbitrarily chosen.
Finally, to account for the possibility that a traveller will return home, we have allocated a 10% chance for this to occur.
This model does not take into account the affect of travel-policy, as we wish to simulate the beginning of a pandemic, where the governments are unprepared.
*b) Disease spreading:* SARS-CoV-2, like most infectious diseases, follows the SIR Model. WICSIE adopts the SIR Model by adding health states to *agents*, which can be either Healthy, Incubated, (Unknowingly) Infected, Vaccinated or Cured. However, due to the time constraints of the course, the simulation discussed in this paper ignores Unknown Infections and Vaccines. Susceptible (Healthy) *agents* first become incubated after infection for $3 \pm 1$ days, in which they are non-spreading but already infected. The infection period lasts for $10 \pm 5$ days, and afterwards, *agents* have a

natural immunity of $200 \pm 100$ days. Here it should be noted that the simulated *agents* do not die, and neither is there a new population of babies. Moreover, WICSIE is not targeting different age groups differently from others, as *agents* do not contain personal information.

To handle disease spreading, all the *agents* have to know which other *agents* are in the surrounding. Checking every *agent's* position against every other *agent* on the map would result in an $\mathcal{O}(i * n)$ time complexity, where $i$ is the amount of infected *agents*, while $n$ is the total amount. Instead, WICSIE creates a grid-map of 3x3 pixel large cells or clusters, which hold a reference to the *agents* within. Then all infected *agents* can immediately get a reference to all *agents* that are approximately nearby by calculating their position against their respective grid position. Assuming an equal distribution, the time complexity is: $\mathcal{O}(i * \frac{n}{f})$, where $f$ is the amount of cells, for this paper $f = 33333$. Then, within each of these cells, every infectious *agent* has a $10\%$ chance to infect any adjacent *agent*. The resulting algorithm is seen in Algorithm 1.

**Algorithm 1** Global Infection Algorithm

```
1:  procedure INFECTAGENTS(grid, probability, simulation)
2:      for each infected Agent agent do
3:          neighbours = grid.GetNeighbours(agent)
4:          for each Agent neighbour ∈ neighbours do
5:              if randomFloat ≤ propability then
6:                  simulation.Infect(neighbour)
7:              end if
8:          end for
9:      end for
10: end procedure
```

Nonetheless, the algorithm's running time is still significant and would result in hours of CPU time for just a few steps. Therefore, WICSIE is parallelising the spreading in a similar fashion to Google MapReduce using Co-routines. For that, every infected *agent* gets its own co-routine, similar to k-clustering. Next, each *agent* follows a variation to Algorithm 1, which can be seen in Algorithm 2, similar to a map phase. The main difference is the ***ParallelInfect***, which is a thread-safe implementation of ***Infect***. It uses a Thread-safe queue, to which every infected element is added. Note that this algorithm only stores future infections here and does not yet modify them. If those *agents* would become infected, a race condition will be created, as *agents* are not aware whether they are still healthy or just got infected. Instead, once the infection cycle is over, the Queue gets popped and continuously infects *agents*, ignoring double infections, similar to the reduce phase. The algorithm is not a one-to-one mapping of MapReduce as it is missing the shuffle phase, however, MapReduce was a strong inspiration when designing the parallelisation process.

*c) Visualisation:* The last step of WICSIE is Visualisation. When tracking 15 million *agents*, $\mathcal{O}(n)$ operations are time consuming. Hitherto, there were already two $\mathcal{O}(n)$ (or worse) sub algorithms; therefore, instead of iterating again, each grid cell could already hold information about the health status of the *agents* within. The resulting Cell datatype can be seen in Listing 1. Each pixel rectangle can now be colour-coded to the dominant cluster within each cell from the grid information. For this simulation, green, red and grey have been used to represent the respective SIR states. An example result can be seen in Figure 2. Additionally, the combined SIR numbers are stored in a CSV file on every step. The time complexity for the visualisation is a constant, as it is independent of the *agent* amount.

**Algorithm 2** Agent neighbour infection

```
1: procedure FINDANDINFECTNEIGHBOURS(agent, grid, probability, simulation)
2:     neighbours = grid.GetNeighbours(agent)
3:     for each Agent neighbour ∈ neighbours do
4:         if randomFloat ≤ propability then
5:             simulation.ParallelInfect(neighbour)
6:         end if
7:     end for
8: end procedure
```

Listing 1. Cell implementation - proposed structure

```
type Cell struct {
        Agents                  []*Agent
        Healthy, Infected, Cured int
}
```

*3) Implementation specifics and simulation:* WICSIE is implemented in GO; as the language base already implements the CSP paradigm, it has been exceptionally straightforward to implement the parallel aspects of WICSIE. Specifically, *WaitGroups* have been used for the mapping phase of the algorithm. Additionally, GO being a compiled language, offered faster running time than interpreted languages while also offering a Garbage Collector.

Once WICSIE was finished, it was submitted to Peregrine, the RUG's High Computation Cluster, and ran for 30 hours generating just shy of 200 days.

## III. Discussion and Results

The simulation you will see now uses three colors and the alpha channel:

- Green represents predominantly healthy areas
- Red represents predominantly infected areas
- Grey represents predominantly recovered areas

The brighter an area, the more people can be found in it. A frame of the simulation can be seen in Fig, 2.

A video to the full simulation can be found in the appendix. The resulting simulation broadly matches our expectations of the pandemic, where there is initial local spread, followed by international movement resulting in new local spreads.

Running the simulation in both high and low resolutions throughout the Europe provided us similar infection results. Additionally, we noticed a very interesting phenomenon: "virtual waves". During the lower resolution, after a period of time *agents* become susceptible to Covid again. However as there are less *agents* while extreme movement chances are the same, there is a smaller amount of new origins. This leads to local waves, as previously recovered *agents* become susceptible and

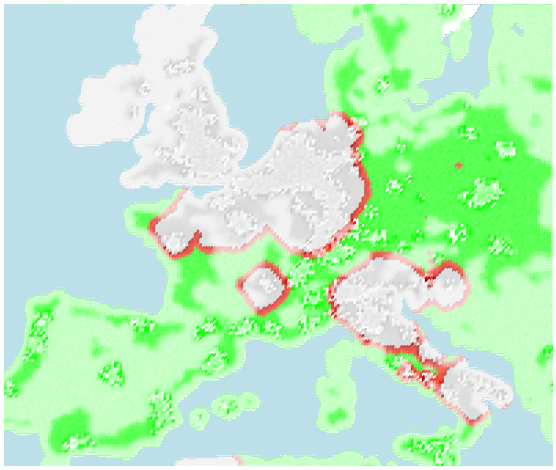

Fig. 2. One state of the pandemic during the simulation

are then infected by newly infected *agents* (who themselves where in the previous time-step just susceptible again).
By looking at the SIR graph (see 4), we noticed that while the graph might suggest herd immunity. On the other hand, the number of newly infected individuals (see 3) shows that the total of active cases increases at all times, sometimes jumping to shy of a million. This shows that herd immunity is not achieved yet.

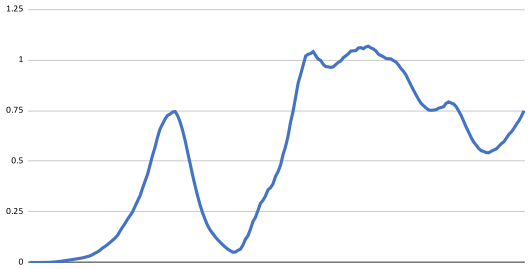


Fig. 3. The number of new infections over time

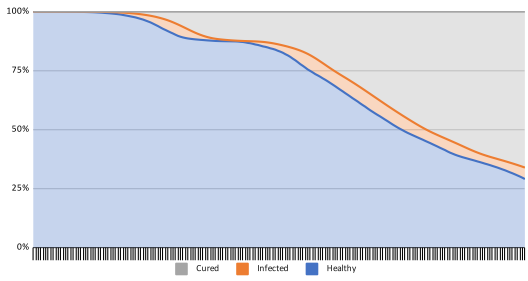


Fig. 4. The amount of infected-recovered-susceptible individuals in the population

The results of the simulation draws similar parallels to the real spread of COVID-19, however the spread is much faster in comparison to real-world data. This indicates that European measures to limit COVID-19 were (at least initially) somewhat successful.

## IV. Future Work

There are various ways in which we can improve our simulation. Below are some recommendations of potential improvements upon our model. In no particular order:

- One could increase the number of *agents* that are simulated. Due to time and resource constraints, we modelled only up to 15 million *agents*, when in reality the population of Europe is on the order of hundreds of millions.
- The probability of an ”extreme” movement was arbitrarily chosen. Potential research into international travel behaviour can be utilised to determine a more realistic probability, or a better model, even.
- More complex constraints can also be placed on the movement model. For example, movement highways can be modelled to represent likely / possible travel paths. In addition, border control can also be simulated, to accurately represent international border entry points.
- The model can be executed with a starting origin in China, and the parameters can be tuned to match existing spread. Then, this tuned model can then be executed upon Europe, for a more ”realistic” simulation.
- The immunity period could be reduced or more extensively modelled. As is, the immunity outlasts the length of the simulation. Furthermore, we could introduce a mortality into our model, as people who lose immunity might be dead and thus, should not be considered susceptible.
- Different choices of origins could be selected, and analyzed based on relative isolation.
- COVID has differing effects based on the infected’s age. This information could be integrated into the simulation as well.

## V. Conclusion

In this paper, we proposed a COVID-19 simulation that predicts the possible outcomes of the virus having its origin in Europe. Our results are visualised and encoded as a video of the SARS-CoV-2 spread throughout the continent with descriptive graphical illustrations and statistical reports. After careful analysis and investigation, we inferred certain conclusions.
To begin with, we are satisfied with the fact that our model produced results comparable to the ones from the coronavirus pandemic. The spread of the infection, cycles and their duration were following similar behaviour to the pandemic we witnesses during 2020 and 2021. That leads us to the conclusion that our assumptions were right and the sources we based our research on were credible.
Nonetheless, there are certain trade-offs. We couldn’t simulate movement of all the potential inhabitants of Europe. Thus, our model is scaled-down to 15 million agents and 3 million agents for the lower resolution respectively. Secondly, our model currently doesn’t make any assumptions regarding overseas travel - they are simply ignored. Those kinds of migrations could be in fact crucial in the case we want to adjust our model, so it is capable of facilitating different strains and types of the virus that usually originate in different parts of the world. Lastly, in this paper, the SIR model is used, which

although is a good base for the simulation, is also very abstract and it is not ideal solution when it comes to diseases where susceptibility period is not very revealing. More accurately, SIR model is more precise for diseases such us Ebola, where infection can be detected within a day or two.[10] On the other hand, a week can pass before the agent infected with coronavirus shows any symptoms or even none at all.

As a matter of fact, creating a 100% precise epidemic model is an unreasonable goal. To begin with, the human factor, involving their movements and decision-making, is hard to predict [11]. Furthermore, there are multiple genetic factors that may influence COVID-19's susceptibility and severity [12]. Nevertheless, our predicted outcome is remarkably similar to COVID-19's behaviour we have seen during the pandemic in 2020 and 2021. Thus, we can only conceptualise about positive outcomes which could have resulted from the governing bodies having access to it beforehand. Aforementioned proves that epidemic modelling has a well-established place in the 21'st century and thus improving WICSIE's accuracy is definitely meaningful. In the future, we are planning on expanding the preciseness of our model by analysing the already existing results of the pandemic and increasing it's efficiency by taking into account the medical, social and governmental resources and also the cultural position that could affect a country's position in the fight against the virus [13].

## Acknowledgements

We thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

## Appendix

To access the codebase for the simulation, please follow the following link: Codebase

To access the video showing the pandemic spreading (with 15 million agents), please use the following link: Simulation with 15 million agents

## References

[1] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, and S. Bernardini, "The covid-19 pandemic," *Critical Reviews in Clinical Laboratory Sciences*, 2020. available under: https://www-tandfonline-com.proxy-ub.rug.nl/doi/full/10.1080/10408363.2020.1783198.

[2] Y. Hewings-Martin, "What who advisor david heymann told us about covid-19." https://www.medicalnewstoday.com/articles/covid-19-interview-with-who-advisor-professor-david-heymann.

[3] P. Yella Hewings-Martin, "How do sars and mers compare with covid-19?." https://www.medicalnewstoday.com/articles/how-do-sars-and-mers-compare-with-covid-19.

[4] IEA, "World air passenger traffic evolution, 1980-2020." https://www.iea.org/data-and-statistics/charts/world-air-passenger-traffic-evolution-1980-2020.

[5] W. H. Organization, "Pneumonia of unknown cause – china." https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229.

[6] W. H. Organization, "Covid-19: Who european region operational update epi weeks 48/2021–4/2022 (1 december 2021–31 january 2022), howpublished = https://www.euro.who.int/__data/assets/pdf_file/0019/525502/covid-19-operational-update-weeks-48-2021-4-2022-eng.pdf, note = "(accessed: 04.04.2022)"."

[7] P. Bump, "The difference in how the pandemic has affected the u.s. and south korea remains staggering." https://www.washingtonpost.com/politics/2020/12/04/difference-how-pandemic-has-affected-us-south-korea-remains-staggering/, Dec 2020.

[8] Ammaren, "R/mapporn - europe population density map by subdivision. all data from various wikipedia sources, ranging from 2015 to 2020 census, as countries don't have censuses every year.." https://www.reddit.com/r/MapPorn/comments/qkcub5/europe_population_density_map_by_subdivision_all/, Nov 2021.

[9] C. Song, T. Koren, P. Wang, and A.-L. Barabasi, "Modelling the scaling properties of human mobility," *Nature Physics*, vol. 6, 10 2010.

[10] S. Moein, N. Nickaeen, A. Roointan, N. Borhani, Z. Heidary, S. H. Javanmard, J. Ghaisari, and Y. Gheisari, "Inefficiency of sir models in forecasting covid-19 epidemic: A case study of isfahan." https://www.nature.com/articles/s41598-021-84055-6, Feb 2021.

[11] K. Naseer, J. Qazi, A. Qazi, B. K. Avuglah, and R. Tahir, "Travel behaviour prediction amid covid-19 underlaying situational awareness theory and health belief model." https://www.tandfonline.com/doi/abs/10.1080/0144929X.2021.1984579?journalCode=tbit20, Oct 2020.

[12] S. Choudhary, K. Sreenivasulu, P. Mitra, S. Misra, and P. Sharma, "Role of genetic variants and gene expression in the susceptibility and severity of covid-19." https://synapse.koreamed.org/articles/1145785, Mar 2021.

[13] C. A. Varotsos and V. F. Krapivin, "A new model for the spread of covid-19 and the improvement of safety." https://www.sciencedirect.com/science/article/pii/S0925753520303593, Aug 2020.