

# A spatio-temporal network epidemiological agent-based modelling approach for mapping risk and exposure of viruses

Lars de Kwant — 6958680

Master Computing Science at Utrecht University

Supervisors: Dr. Deb Panja (first) and Dr. Erik Jan van Leeuwen (second)

July 10, 2024

## Abstract

With the globalization of the world, the ability to contain viruses becomes more important. Modeling the spread of a virus is a great tool for understanding how to prevent future epidemics from happening. This thesis continues on an existing model of the COVID-19 pandemic in The Netherlands, which is an epidemiological agent-based model on a spatio-temporal network. The model is adjusted and improved in order to create risk and exposure maps of The Netherlands at a municipality level, per demographic group. The risk maps show the total infections after three weeks after infection of five agents in a certain municipality, while the exposure maps show where the exposures to the virus happened. In order to obtain the desired risk maps, the original model is extended with extra epidemiological parameters and features that allow for infections to be initialised with high precision. The resulting risk maps show that the risk of a municipality depends mostly on the average mobility per agent, as well as the location of the municipality in the underlying complex network of mobility, demography and behavior. As expected, the Randstad is clearly the most risky area in The Netherlands for the high mobility demographic groups. The major municipalities like Amsterdam, Rotterdam and Utrecht do not stand out as higher risk compared to other nearby municipalities, however they are still responsible for the vast majority of exposures to the virus. The results show possibilities for future research into new intervention measures, to discover the efficiency of significantly reducing mobility to these exposure locations, no matter where the infections originated from.

## Introduction

Modeling is one of the main tools to understand the spread of viruses. However, creating such models on a global level makes it hard to properly encapsulate the specifics of each country's mobility patterns, demography and behaviors. In order to advice the local governments in their intervention policies, most research projects were modeling the spread of SARS-CoV-2 for a specific country to predict how the transmission of the virus would evolve locally. The types of models used in these research projects can vary immensely. Some studies choose to use purely mathematical models based on differential equations [1, 2, 3], while other studies used models based on spatial networks [4, 5]. A recent paper by Dekker et al. [5] highlights the flexibility of agent-based models for modeling possible local interventions at the municipality level. An agent-based model is a model that uses individual agents, each with their own characteristics, to simulate interactions between people in a stochastic manner. These models allow for detailed characteristics of each possibly unique agent, while modeling the movements and interactions of these agents on highly complex underlying networks. Dekker et al. uses a spatio-temporal network of The Netherlands to model the inter-municipality mobility of the population, created from real mobility and demography data. Models like this are becoming more and more important with the increasing globalization of the world, in order to quickly contain a virus

before it spreads further spatially.

Having a model that can capture all the spatial heterogeneities of a population opens up new interesting views on virus containability. Throughout the COVID-19 pandemic there was again a big focus on the quantity  $R_0$  and contact tracing.  $R_0$  entails the number of new infections created by every existing infection, however a quantity like  $R_0$  is not as invariant as many say it is. This value depends heavily on the heterogeneities of the population, like demography, population density and mobility patterns. Contact tracing is often used to determine the value of  $R_0$  and to control outbreaks by quarantining infected individuals as well as their contacts, however its effectiveness is limited. A recent study showed that the spread of SARS-CoV-2 could only be stopped if new cases are isolated within 2-3 days, as well as quarantining 80% of their contacts [6]. These numbers are clearly difficult to obtain consistently, especially when not every individual gets tested. To spatially contain a virus, it is important to get ahead of a possible spread, instead of waiting for results of contact tracing and testing for infections. SARS-CoV-2 already spread quickly across the globe, however this spatial containment becomes even more important for a virus that has an incubation time that is longer than its latent time. The incubation time of a virus is the time between exposure and a person becoming symptomatic, while the latent time is the time between exposure and becoming infectious. Therefore, when the incubation time is longer than the latent time, a virus can spread before an individual shows symptoms, which makes containment even more difficult.

This thesis aims to explore more into possible containability metrics and policies, by using and improving the spatio-temporal network agent-based epidemiological model by Dekker et al. The results of the paper by Dekker et al. are first reproduced with a synthetic mobility dataset, due to unavailability of the original Meuzo dataset. Subsequently, the model is improved by the addition of new epidemiological parameters and features to allow for more precise modeling of viruses. Finally, the improved model is used to create risk and exposure maps of The Netherlands. The risk of every municipality in The Netherlands is computed by observing the total amount of infections after three weeks that would be caused by an infection started in that municipality. The exposure maps are created by locating where every infection in the risk maps was caused. Other studies have been performed on the risk mapping of different countries, like Portugal [7] and Australia [8], however these studies did not use agent-based models and contain no further spatial epidemiological information like locations of exposure. The strength of the model by Dekker et al. lies in the fact that it combines epidemiological models with spatio-temporal networks, which allows patterns between risk and certain heterogeneities in the population like demography and mobility to be brought to light. Consequently, these patterns can be useful in researching novel containability policies in the future. This thesis will answer the question whether the model by Dekker et al. can be improved and adjusted to create risk maps of The Netherlands, and whether the information provided by these risk maps leads to new ideas of future interventions to contain a virus. During this process it also provides a robustness check for the model by Dekker et al.

The resulting risk maps show that the Randstad area in The Netherlands is the most risky for the students and middle-age working demographic groups. This matches with expectations, since these groups have high inter-municipality mobility in the Randstad area. On the other hand, the primary-school children group does not show any patterns of increased risk in areas, which matches with the mobility patterns of this group. An interesting observation is that the major municipalities like Amsterdam, Rotterdam, and Utrecht do not have an increased risk compared to their surrounding municipalities in the Randstad. However, the exposure maps show that these major municipalities are still responsible for the majority of the exposure to the virus. This observation leads to ideas for future research into novel intervention measures. This future research could show the effectiveness of significantly reducing mobility from or to the major exposure locations after a certain number of days, no matter where the initial infections occurred.

The remaining sections of this thesis are organized as follows. The section on methodology will explain the model by Dekker et al. and any changes made to this model, after which the methods used to create the risk maps will be discussed. The results section will contain figures that show the robustness of the model by Dekker et al., followed by the produced risk maps and make observations on these maps. Finally the thesis ends with a detailed discussion of the observations, possible limitations of the model, and ideas for future work based on the information provided by the results.

## Methodology

This section will discuss the methodology of this research, starting with an explanation of the model created by Dekker et al. [5]. Followed by listing and explaining the improvements that have been made to the model, and finally describing the methods used to generate the risk maps and other results.

### Model

First of all, an overview of the full model can be seen in figure 1. The model consists of two parts, which are called the mobility model and the transmission model. The first part of the model uses CBS datasets containing the population size and demography of each municipality, as well as a dataset of inter-municipality mobility data. From these datasets, the model computes weekly schedules for every agent in the model. These weekly schedules contain the position at a municipality level with hourly time resolution. To determine these weekly schedules, the model starts of by computing the fractions of time spent in every municipality by drawing these fractions from a Dirichlet distribution created from the inter-municipality mobility data. The distribution is drawn under the assumption that working people and students spend roughly 25% of their time in other municipalities, and all other non-working people spend about 5% of their time in other municipalities. Next, the drawn fractions are converted into integer hours of time spent in each municipality. Finally these hours are organized into days by spreading time spent in their home municipality evenly around midnight, for example an agent that spends 16 hours in their home municipality will do so from 16:00-24:00 and 00:00-08:00. The remaining time spent in other municipalities is placed in a random order throughout the day. If an agent spends multiple hours in a certain municipality, these hours are placed in sequence with each other. This procedure of creating daily schedules is repeated seven times for each agent, to obtain all weekly schedules. Dekker et al. were unable to publish the actual Mezero mobility dataset, therefore they created a dataset of mobility data that was generated by a gravity model, which was instead published alongside their paper. The gravity model was calibrated using the real Mezero mobility data in order to obtain mobility data that would achieve similar results on the full model, however the code of the gravity model is unavailable. Gravity models are often used in social sciences to estimate networks of interactions. The strength of a connection between two nodes in the gravity model is proportional to the population sizes and the inverse of the distance, as shown in equation 1

$$I_{ij} = k \frac{P_i P_j}{D_{ij}^b} \quad (1)$$

where  $I_{ij}$  is the the strength of the connection between nodes  $i$  and  $j$ .  $D_{ij}$  is the distance between nodes  $i$  and  $j$ , and  $b$  determines the strength of the decay of the connection over distance.  $P_i$  and  $P_j$  are the population sizes of the nodes  $i$  and  $j$ . The values the constants  $k$  and  $b$  can vary depending on the specific data set [9, 10]. After the first step, the model now has the weekly schedules with hourly time resolution of every agent in the model, using mobility data from a calibrated gravity model as well as demographic population data.

The second part of the full model computes the transmission of the virus across the country using the mobility patterns computed in step one of the model. The virus infection is initialised by infecting certain agents in one or more municipalities. In the paper of Dekker et al. these initial infections

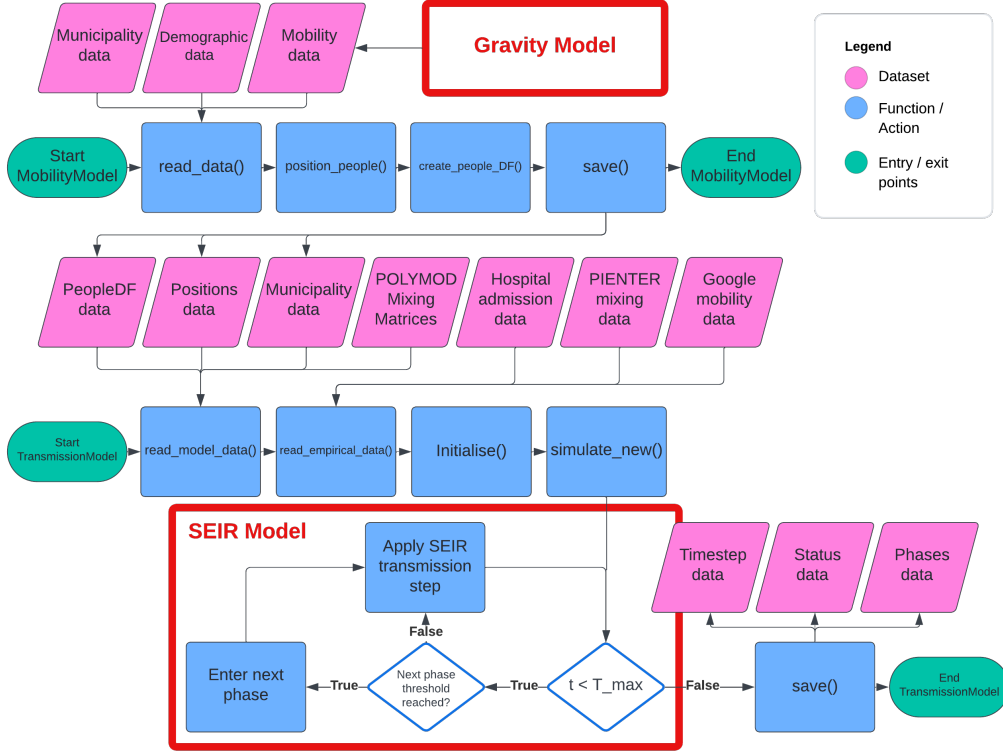


Figure 1: Diagram of the full 2-step model from the paper by Dekker et al. [5]. In the first step weekly schedules are created for every agent. These schedules contain the position of every agent at an hourly time resolution. This is done by using CBS data of each Dutch municipality’s population size and demographic distribution, as well as inter-municipality mobility data. The mobility data can come from a real data set, or from a model that approximates real mobility. These weekly schedules are then used as input to the second step, called the transmission model. If the transmission model is used to model the first COVID-19 wave, it requires additional data of hospital admissions, behavioral change and mobility change as input to accurately model the interventions that were imposed on the country. After creating the initial infection in certain municipalities, the spread of the virus is computed using the SEIR model. The final output contains the SEIR state of every agent at every time step.

are determined based on the early hospitalisation data provided by the NICE registration. From this hospitalisation data, the number of originally infected people in each municipality are approximated. The initial infections are chosen at random in each municipality with respect to these approximated infection numbers, where each agent represents 100 inhabitants. After this initial infection, the SEIR model is applied to compute the state of every agent at every time step in the model. In the SEIR model, an agent can be in one of four states: susceptible (S), exposed (E), infectious (I) or recovered (R). At the start of the model, every agent is in the susceptible state, except for the initially infected agents, who are in the infectious state. At every time step the model determines which agents should transition to a new state as follows. First of all, the model computes the force of infection  $\lambda(g, m, t)$  at time step  $t$  for a certain demographic group  $g$  in a certain municipality  $m$  as shown in equation 2

$$\lambda(g, m, t) = h(g) \cdot \beta_t \cdot \bar{s}(t) \cdot \sum_{g' \in G} n_{g, g'} \cdot \frac{I(g', m, t)}{N(g', m, t)} \quad (2)$$

where  $h(g)$  denotes the susceptibility of an agent from demographic group  $g$  to the virus.  $\beta_t$  is the transmission parameter of each phase  $t$ , which is used to calibrate the model to observed hospitalisation data.  $\bar{s}(t)$  is a parameter that entails how transmission changes during the day due to the day-night cycle. The last component determines the infection pressure based on the fraction of infectious agents in a municipality  $m$  and time step  $t$  and mixing between demographic groups  $G$ . This is computed by a sum over all demographic groups  $g' \in G$ , where  $n_{g,g'}$  determines the mixing between the demographic groups  $g$  and  $g'$ , and the fraction denotes the number of infectious agents amongst the total number of agents present of group  $g'$  in municipality  $m$  at time step  $t$ . Based on this force of infection in a municipality  $m$  at time step  $t$ , every susceptible agent of demographic group  $g$  that is present in that municipality at that time step, has a chance equal to  $\lambda(g, m, t)$  to become exposed to the virus. Once an agent becomes exposed, the model determines the latent time for this agent, which is the time between exposure and becoming infectious. This latent time is obtained by sampling a Weibull distribution with mean  $\mu = 4.6$  days and shape  $\lambda = 20$ . Once agents become infectious, another Weibull distribution with mean  $\mu = 5$  days and shape  $\lambda = 1$  is sampled to obtain the time that this agent will need to enter the recovered state. All these parameters came from the paper by Dekker et al. The final output of the model is the state of every agent at every time step. The model itself does not have any way of tracking hospitalisations, these are computed from the number of new infectious agents in the model at every time step. Every new infectious agent has a certain chance  $p_{hos}(g)$ , depending on their demographic group  $g$ , to be hospitalised after a time period that is again sampled from a Weibull distribution, with mean  $\mu = 14$  days and scale  $k = 10$ . However, in python code, the Weibull distribution is parameterised by the shape parameter  $\lambda$ , the derivation of which is shown in equation 3. Filling in the provided mean and scale parameters, the shape parameter comes out as  $\lambda = 14.7$ .

$$\mu = \lambda \Gamma\left(1 + \frac{1}{k}\right) \rightarrow \lambda = \frac{\mu}{\Gamma\left(1 + \frac{1}{k}\right)} \quad (3)$$

To confirm that the synthetic mobility data generated by the gravity model indeed achieves similar results to the real Mezero mobility data, the described model was used with the synthetic data to reproduce a fraction of the results from Dekker et al. The first figure that has been reproduced, is figure 2a, which reproduces the first COVID-19 wave in the Netherlands and matches it to observed hospitalisations. The second figure that has been reproduced is figure 3a, which shows the cumulative hospital admissions of different intervention measures compared to the observed hospitalisations. Another strong robustness check for the model by Dekker et al. is to check how the model output compares between using the synthetic mobility data versus a null mobility model. This null mobility model uses mobility patterns that are generated by a gravity model which ignores the distance between municipalities. Consequently, the mobility patterns only rely on the population sizes of the municipalities, and have no decay over distance. This can show how the spatial heterogeneity is captured well inside the model, when using mobility data representative of real mobility. To make sure the total mobility from and to every municipality remains the same in the null mobility model, the mobility data is calibrated to the synthetic mobility data, such that the row and column sums are identical between these mobility matrices. The comparison is visualised by showing the factor difference of the total infections in each municipality in a spatial plot. These values are computed by starting an infection of five agents in each of the 11 demographic groups in Amsterdam, after which the total infections for each demographic group are combined using a weighted sum based on the size of each demographic group. This is done for three runs for both the synthetic and null model, finally the spatial plot shows the total infections in the null model divided by the infections in the synthetic model. Furthermore, the comparison between the synthetic and null model is visualised by plotting the total infections over time and the total affected municipalities over time, where a municipalities is considered as affected when 0.05% of its population has been infected by the virus.

## Improvements to the model

The original model has been improved upon by fixing existing bugs and adding more features. This section will go through all of these changes, starting with the bugs that have been fixed. First of all there was an issue with the initialisation of the model. The code that determined the number of infections in each municipality at the start of the simulation, by computing this from the hospitalisation data, was crashing with index-out-of-bounds errors. The bug was fixed by rewriting the code to produce the intialisation as described in the paper by Dekker et al. The second bug that was found, concerned the mobility patterns of very small municipalities. This bug was caused by empty fields in the synthetic mobility matrix. When this matrix was imported, these fields were read as *NaN* (Not a Number), which caused all mobility patterns for this municipality to be *NaN* after attempting to compute them. The bug had a visibly unnoticeable impact on the full model, since it only affected very small municipalities, which played a small role in the spread of the virus. However, for the use case of the model in this thesis, the bug became visibly noticeable due to every municipality separately being used initialisation for the infection. This problem was fixed by making sure the empty cells in the mobility matrix were just read as zero, which represents a mobility of zero between those municipalities, which is a reasonable value since it is mobility between two very small municipalities.

To make the model slightly more flexible, a couple of new features have been added as well. The first change is the implementation of incubation time, which is accompanied by parameters that control the Weibull distribution and a parameter that sets the percentage of agents that will stay at home once they become symptomatic. The addition of these parameters allows for modeling any virus based on its latent, incubation and infection time, as well as representing the behavior of symptomatic people by making a percentage of agents with symptoms stay at home. This allows for even more precise modeling of behavior and virus transmission. The total number of infections over time for the combinations of latent and incubation times of three, six and nine days have been plotted to show how different combinations of these parameters can significantly affect the spread of a virus. During these simulations, 80% of symptomatic agents were set to stay at home.

The second addition to the model creates the possibility to initialize an infection of  $n$  agents in a specific municipality and demographic group. If there are less than  $n$  agents of that demographic group in that municipality, the model adds the missing agents, such that it ends up with  $n$  infected agents in total. The addition of these missing agents is however not trivial, because all mobility patterns of all agents are pre-computed during the first step of the full model. In order to accomplish this, the first step creates an additional dataset of extra agents for every municipality and demographic group, which enables these extra agents to be added whenever needed during the transmission model. This new features allows for a more detailed look into how the virus spreads when introduced in a specific demographic group in a specific municipality. This feature will be used to create risk maps of The Netherlands for three important demographic group.

## Generation of risk and exposure maps

The risk maps that will be created of The Netherlands, show for a specific demographic group how risky each municipality is. This risk value is obtained from simulation data by summing the total amount of infected agents across the country after an initial infection of five agents of a certain demographic group in that municipality. The goal of the risk value is to encapsulate the location of a municipality in the complex network of mobility between municipalities, as well as mixing within those municipalities. The results are obtained by first generating five unique mobility pattern datasets from the first step of the full model. The final risk map value for each of the 380 municipalities consists of the average risk value over 25 runs. These 25 runs entail five runs for each of the five unique mobility dataset. This is done for three important demographic groups: primary school children, students and middle-age working people. Besides the risk maps, other maps can be created from the model output data. For

each of the three risk maps, a corresponding exposure map is created, showing where exposure to the virus happened for each infected agent. This is done by checking when the status of an agent changed to exposed, after which the location of the agent at that time step can be looked up in the weekly schedules dataset. The exposure locations are averaged over all 25 runs, before being accumulated over all 380 municipalities and finally plotted. All data is obtained by running the models in parallel on the Gemini UU cluster in batches of 16 jobs. There are 24 batches of 16 jobs in total, each batch took around three hours to finish. Depending on how active the cluster was, the results of one risk map took about 1.5 days to compute.

## Results

The results section consists of three subsections, starting with results showing to the robustness of the model by Dekker et al., followed by results that display the use for the additional incubation parameters. Finally, this section ends with observation about the risk and exposure maps.

### Original model

The results of reproducing figures 2a and 3a from Dekker et al. using the synthetic mobility data can be found in figures 2 and 3. The figures show that the synthetic mobility data obtains the same epidemic pattern as the results from Dekker et al. However, the total number of cases is higher than the original results using real Meuzo mobility data. An important observation is that the results are still within the same order of magnitude. The difference can be caused by variance in the mobility data, since the Meuzo mobility data had to be used in a way that it is not possible to be reverse engineered. In order for the full model to closely match the hospital admissions as observed during the COVID-19 pandemic, it has to be calibrated using transmission parameters  $\beta_t$ . This calibration has not been done for the synthetic mobility data, and is not necessary for this thesis since the goal is to find important patterns using risk mapping, which doesn't require the model to exactly match the COVID-19 pandemic anymore.

Figure 4 shows the factor differences between the null model and the synthetic model. The infections in the null model spread throughout the country rapidly, as is to be expected with distance of mobility being ignored. Over time the infections in the synthetic model start spreading more throughout the country, slowly catching back up to the null model in the outer parts of The Netherlands. Figure 5 shows the difference in exponential spread of infections between the null and the synthetic mobility models, with doubling times of respectively 3.88 days and 4.03 days. Finally figure 6 shows the difference between the spatial spread of the null and synthetic mobility models, where a municipality is affected when 0.05% of its population is infected. As expected, the null mobility model affects more municipalities over time.

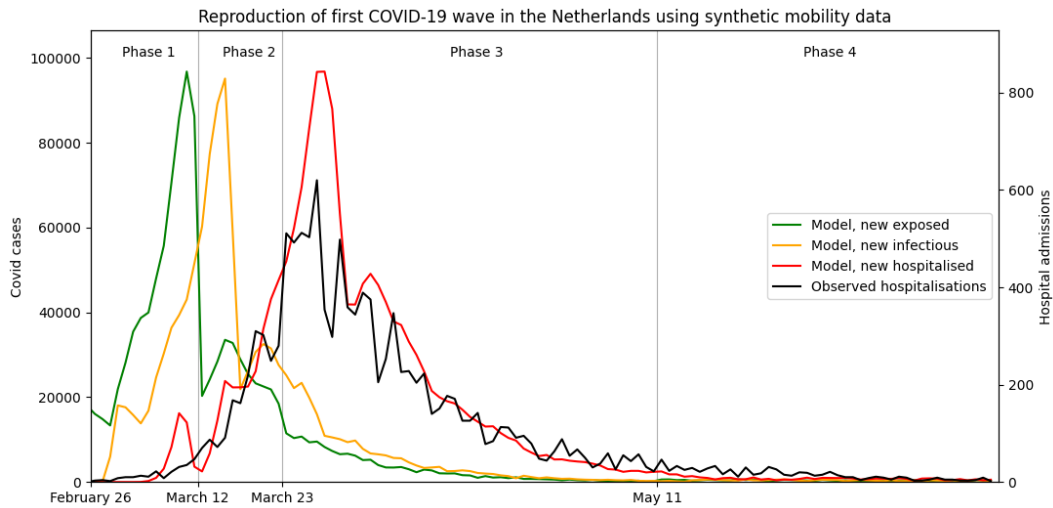


Figure 2: Reproduction of figure 2a from the paper by Dekker et al. The patterns of the first COVID-19 wave clearly match visually with the original results by Dekker et al. However the total number of exposed and infectious cases in the model using synthetic mobility data is close to 100.000 compared to the 80.000 in the original paper, which is about a 25% increase in total cases. The same can be observed for the hospital admissions.

Reproduction of cumulative hospital admissions using synthetic mobility data

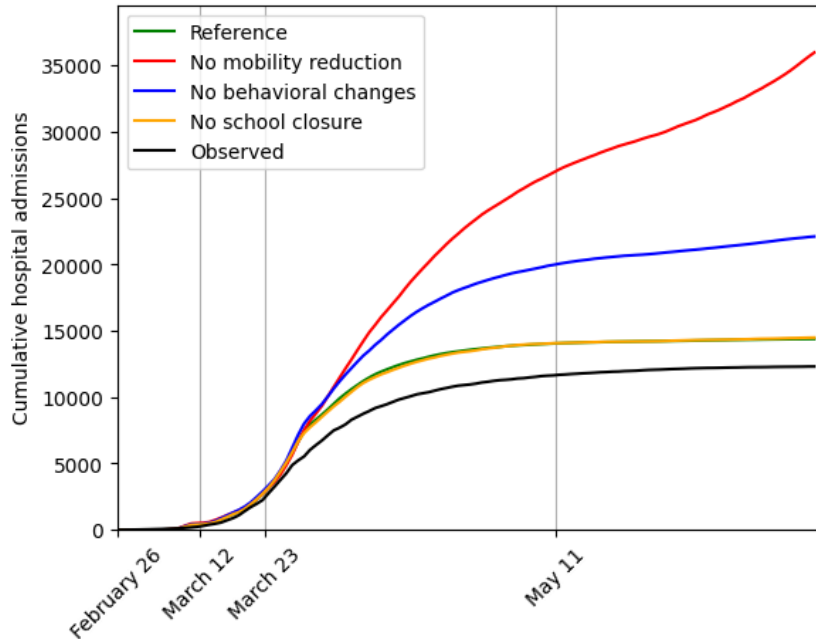


Figure 3: Reproduction of figure 3a from the paper by Dekker et al. The correlation between all different intervention measures remain the same as in Dekker et al., however the total values are again about 25% higher.



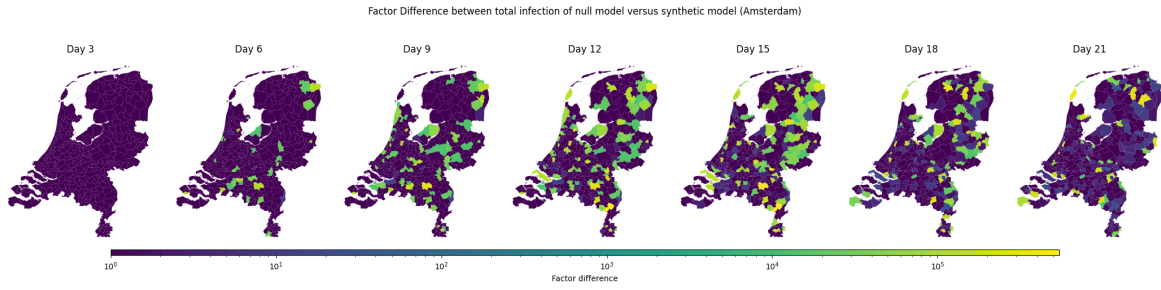


Figure 4: The factor difference between the null and synthetic models for a virus starting in Amsterdam. The null model spreads the virus to the far corners of The Netherlands almost immediately, because the distance factor of mobility is ignored. Over time, the synthetic model slowly starts reaching the outer corners of the country as well, which reduces the extreme factor difference in these outer corners slowly over time. This shows how the mobility patterns are captured well inside the model.

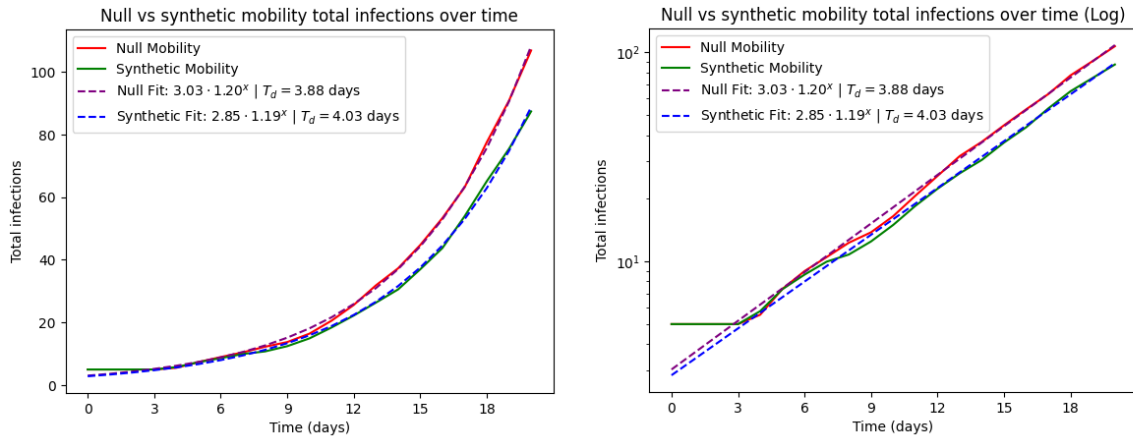


Figure 5: Plots of the total infections of the null model versus the synthetic model over time for a virus starting in Amsterdam. The total infections grow exponentially over time, with the null model growing slightly faster. This is visible in the doubling times  $T_d$ . The doubling time of the null model is 3.88 days, while the doubling time of the synthetic model is 4.03 days.

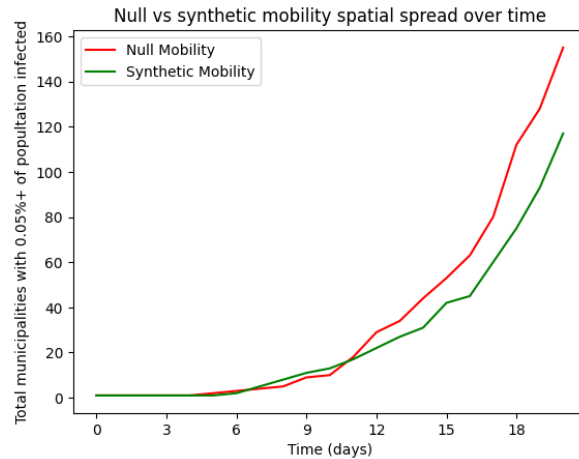


Figure 6: Comparison of the spatial spread of the virus between the null and synthetic mobility models. The infection was started in Amsterdam and a municipality is considered as affected when 0.05% of its population has been infected by the virus. The null model spreads faster spatially throughout the country than the synthetic model. Only at the very beginning the synthetic model has a slightly larger spread, because its mobility is more focused around Amsterdam, which causes it to reach the 0.05% infection rate slightly faster in those neighbouring municipalities.

## Improved model

Figure 7 shows how different combinations of virus parameters in the model indeed have a serious impact on the spread of the virus. In the research by Dekker et al. the incubation time was not relevant due to SARS-CoV-2 having an shorter incubation time than latent time on average, however the addition of this parameter allows for more viruses to be modeled and for changing agents behavior once they have become symptomatic. When the incubation time is shorter or equal to the latent time, the virus does not spread. This is as expected, since 80% of symptomatic agents stay at home, causing the spread to halt. However, in the case where the incubation time is longer than the latent time, the virus spreads faster based on the difference between incubation time and latent time. This is because, when the difference is longer, the infectious agents have more time to freely spread the virus before possibly changing their behavior after becoming symptomatic.

Exposed cases over time for multiple combinations of latent time vs incubation time

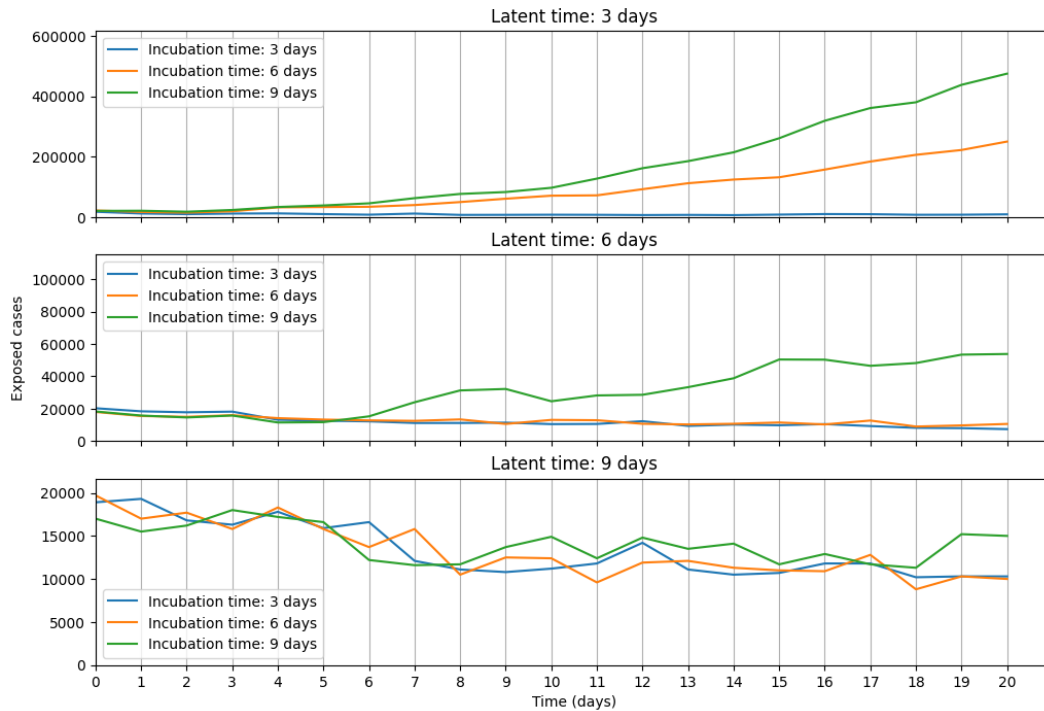


Figure 7: Graphs confirming how the interplay between the latent time and incubation time makes a big difference for virus spread over the first three weeks in the model. All graphs use an infection time of five days. In the top-most graph, the latent time is only three days, which means on average an agent becomes infectious after 3 days. The plotted lines represent incubation times of three, six and nine days, which is the time until an agent becomes symptomatic. In all graphs 80% of symptomatic agents stay at home. As expected, the virus spreads more quickly when there is a bigger difference between the latent time and incubation time. If the latent time is longer or equal to the incubation time, the majority of agents will stay home before they have a chance to spread the virus, which reduces the chances of a virus spreading exponentially.

## Risk and exposure maps

The results from the model have been visualised as risk and exposure maps in figures 8, 9 and 10. The first observation about the risk maps is that the risk map of the primary school children has lower risk values overall than the students and middle-age working groups. On top of that, the risk of every municipality in the children risk map is spread out quite evenly throughout the country, with no big differences between the municipalities. As opposed to the students and working group risk maps, where a higher risk area is clearly visible, which is the Randstad area. However, the biggest municipalities like Amsterdam, Rotterdam, The Hague, Utrecht do not have noticeably higher risk than their surrounding municipalities in the Randstad. Another noticeably higher risk area is in the north of the country above Groningen, which is visible in both the students and middle-age working risk maps. Finally, the students risk map is overall slightly brighter than the middle-age working risk map, indicating that, on average, the risk is slightly higher when an infection starts amongst students. When looking at the exposure maps, a clear pattern is visible across all three exposure maps. The exposures happen in the municipalities where the total incoming and outgoing mobility is highest, which is in the biggest municipalities, like Amsterdam, Rotterdam, The Hague, Utrecht,

but also Eindhoven, Tilburg, Breda, Groningen and more are clearly visible. Out of all three exposure maps, the most exposures happened across the students, followed by the middle-age working group and finally the primary school children.

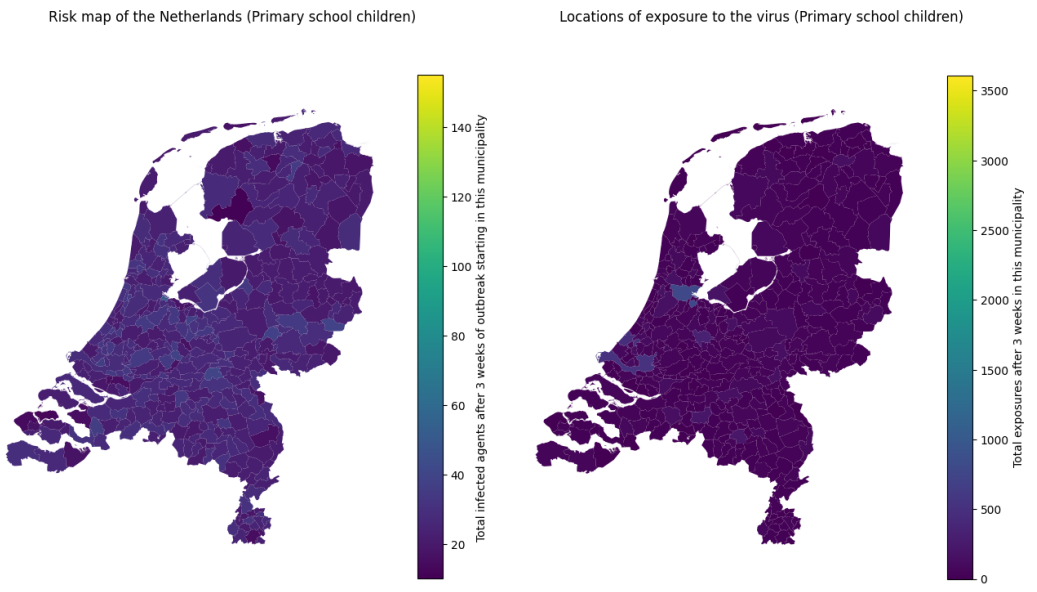


Figure 8: Risk map (left) and exposure map (right) of The Netherlands when the initially infected agents are primary school children. The risk map does not indicate any area as risky, as the risk seems similar for all municipalities in The Netherlands. On the exposure map it is visible that most exposures to the virus do happen in the four biggest municipalities in The Netherlands: Amsterdam, Rotterdam, The Hague and Utrecht.

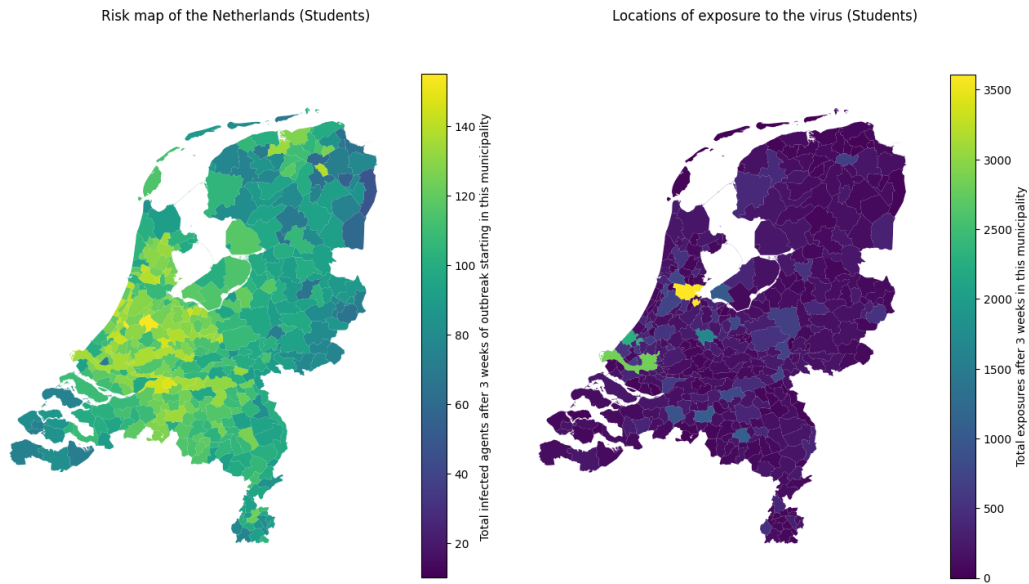


Figure 9: Risk map (left) and exposure map (right) of The Netherlands when the initially infected agents are students. The risk map indicates clearly that the Randstad is the most risky area for a virus to be introduced among students. Another small area of higher risk is visible in the north of The Netherlands. The exposure map is similar to the children exposure map, however the total number of exposures is significantly higher. Therefore most highly connected municipalities in The Netherlands are visible on the exposure map.

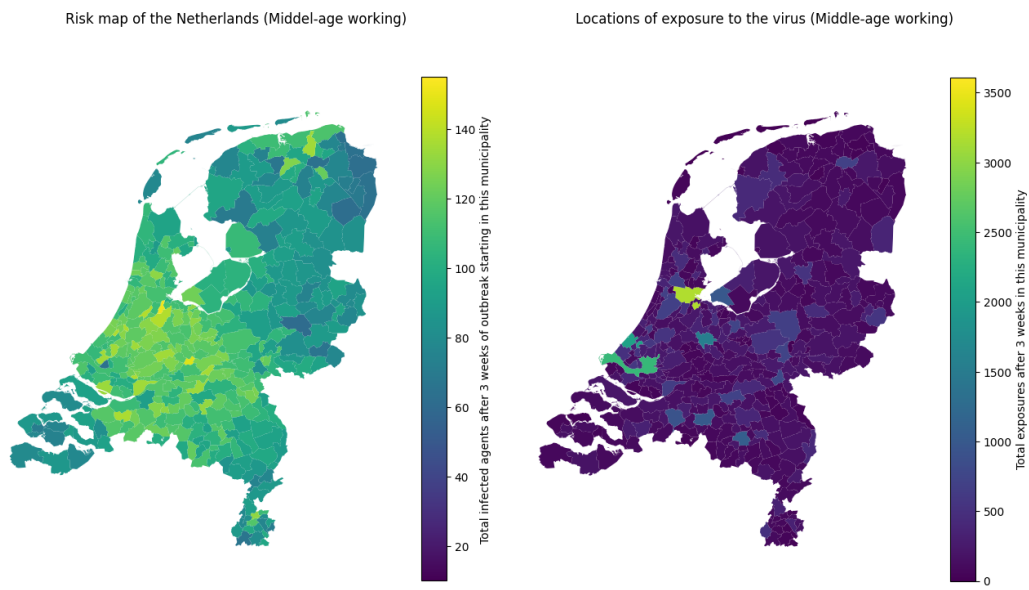


Figure 10: Risk map (left) and exposure map (right) of The Netherlands when the initially infected agents are of the middle-age working demography. Both the risk and the exposure maps are similar to those of the students, however with slightly lower values in general.

Figure 11 shows the average mobility per student. The Randstad area has visibly higher average

mobility than other parts of The Netherlands, however the big municipalities like Amsterdam, Utrecht, Eindhoven and Groningen have lower average mobility than their surrounding municipalities. There is one clear outlier visible in the color yellow, which is the municipality Vlaardingen. This municipality has extremely high mobility numbers to Rotterdam. Another higher average mobility area is visible in the North of The Netherlands.

Map of average mobility per student agent in the Netherlands

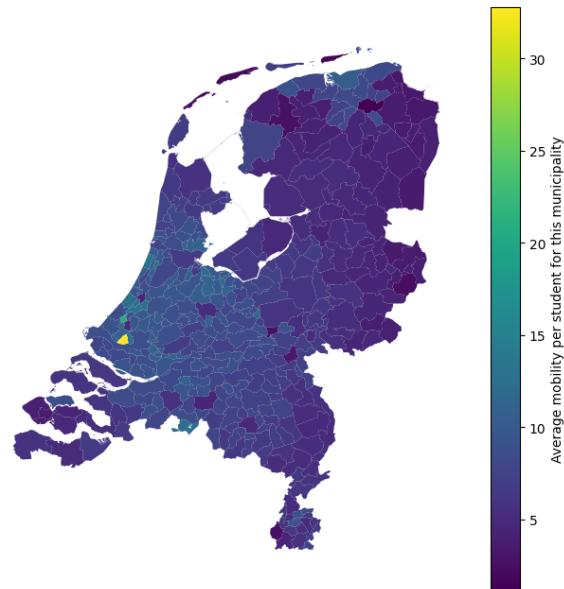


Figure 11: Map showing the average mobility per student in each municipality of The Netherlands. In municipalities like Amsterdam, Utrecht, Groningen, Tilburg, Eindhoven and more, this average is visibly lower than in the surrounding municipalities. A clear outlier is the municipality Vlaardingen, which has extremely high mobility numbers, but only to Rotterdam.

Figure 12 shows four isolated exposure maps from infections that were started in the corners of The Netherlands. Each of these exposure maps have some random municipalities with exposure, however they all have the big municipalities as common exposure locations. It can be observed that municipalities like Amsterdam, Rotterdam, The Hague and Utrecht consistently create exposure to the virus, even when the virus starts in corners of the country.

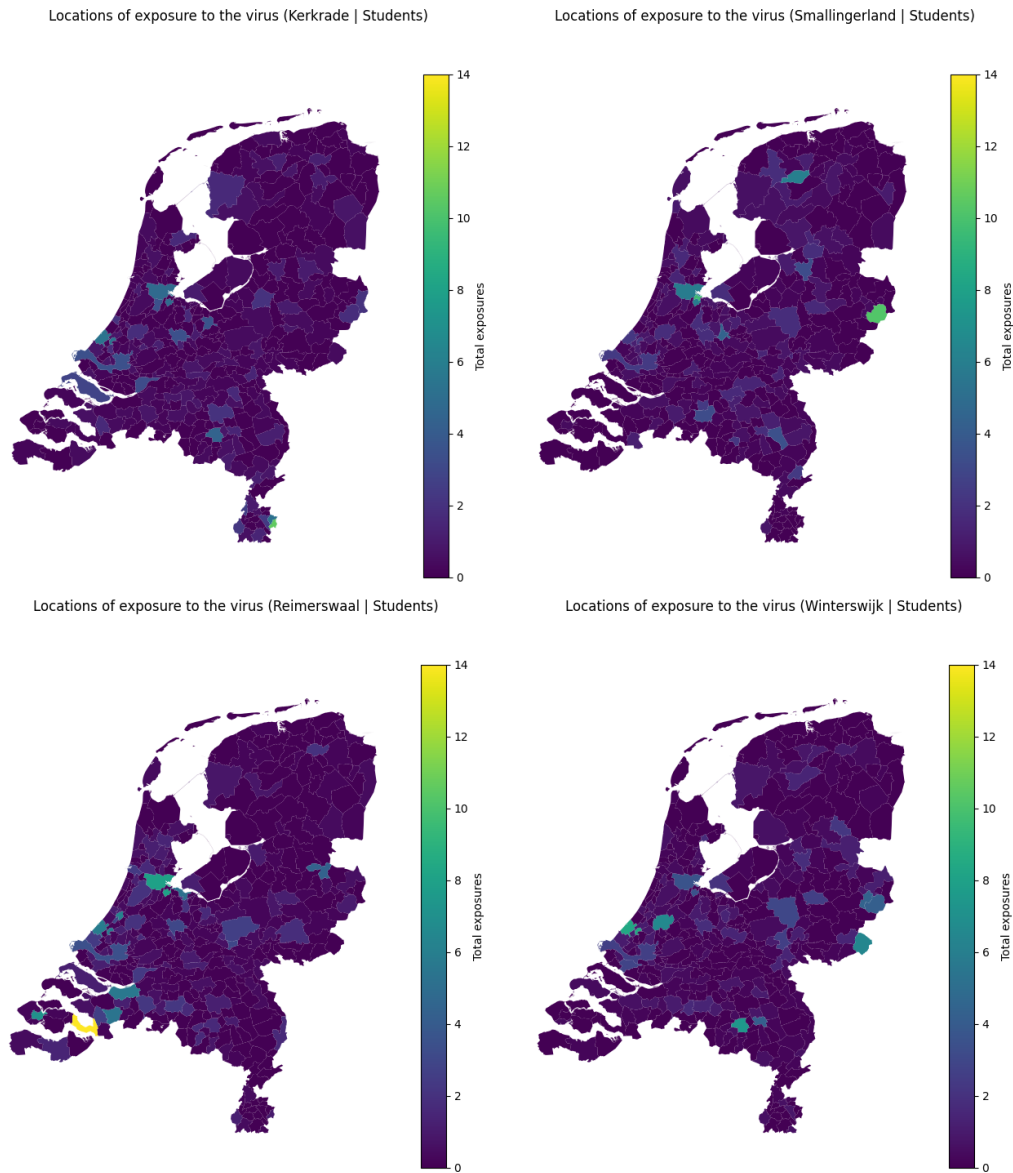


Figure 12: Four samples of exposure maps taken from the students exposure maps dataset. These samples show that the large municipalities like Amsterdam, Rotterdam, The Hague and Utrecht consistently cause large amounts of exposure of the virus. Every sample has slightly different locations where exposure happens, but these large municipalities are always amongst the important exposure locations.

## Discussion

The risk map results bring light to some interesting patterns, on one hand the risk maps look as expected, since the Randstad is considered higher risk, but on the other hand the largest municipalities like Amsterdam, Rotterdam, The Hague and Utrecht are not higher risk than other nearby municipalities. This can be explained by looking at the force of infection as formulated in equation 2, which is directly correlated to the ratio of infected agents to total agent in an area. The chance of infection

is proportional to the force of infection, and this chance is rolled independently for every agent. This means that all  $N$  agents have an infection chance proportional to  $\frac{I}{N}$ , therefore the total number of newly exposed agents is proportional to the current number of infectious agents  $I$ , since the two population sizes  $N$  cancel each other out. From this information it can be concluded that the size of the municipality where the virus initially started, plays no direct role in the spread of the virus, considering that the risk values of these big municipalities are not higher than their surrounding municipalities.

A different quantity that gives a much better indication of how risky a municipality is, is the average mobility per agent of each municipality, more specifically the average mobility per agent of the demographic group that started the infection. If the average mobility per agent is high among a demographic group in a municipality, it has a high chance to quickly spread the virus to other municipalities. Of course municipalities of the likes of Amsterdam and Rotterdam have a lot of total mobility because they are the biggest municipalities of The Netherlands, however as it turns out, they do not have the highest average mobility per agent. Figure 11 showed the average mobility per students in The Netherlands. On this map it becomes clear why the north of The Netherlands had a higher risk area, because these municipalities have high mobility to Groningen on average. Another observation is that a few of the largest municipalities have considerably lower mobility per student than their surrounding municipalities. Since every municipality starts with five infections, this means that the spread of the virus from these municipalities is also slower on average. The reason why the average mobility of these big municipalities might be lower, is because this is a map of inter-municipality mobility. The majority of mobility for these big municipalities is most likely largely within the municipality itself, since most of these people also work or study there. Therefore, they don't have to travel outside of this municipality as much, which means that this mobility is not visible in the inter-municipality mobility data.

On the other hand, municipalities that are close to these large municipalities, have much more mobility per agent on average, because these agents travel for example to work or university in a nearby bigger municipality. However, average mobility per agent is not the only metric that is important to determine the risk of a municipality, this also largely depends on how well-spread this mobility is. For example, Vlaardingen has really high average mobility per agent, but has quite low risk on the risk map. This is because the vast majority of the mobility from Vlaardingen is to Rotterdam, which leaves almost no variety of where agents move in the simulations. Due to this, an infection initialised in Vlaardingen most likely turns into an infection initialised in Rotterdam, but delayed by the time it takes to expose agents in Rotterdam to the virus, and for those agents to then become infectious.

Even though the biggest municipalities might not lead to the most infections after three weeks as given by the risk maps, this fact still seems very counter-intuitive. As explained before, the force of infection does not lead to a higher number of expected exposures if a municipality has a larger population, which also clearly shows some limitations of the model. The model does for example not take into account how densely populated a municipality is, or the amount of mobility that takes place within a municipality. This mobility within municipalities will be significantly higher for those bigger municipalities, which is logical since they have a bigger population, but these values will likely also be higher on average, because these people can travel to important places and resources like work and universities within the municipality itself. Currently, the model uses transmission parameters  $\beta_t$  in the force of infection computation to estimate the transmission of the virus at certain points in time, however these parameters  $\beta_t$  are the same for every municipality. The parameters are calibrated to match the hospital admissions of the first COVID-19 wave when looking at the country as a whole, but when looking at the scale of each municipality, there is no guarantee that these transmission parameters are correct. Further improvements to the model could therefore be the introduction of parameters for each municipality that encapsulate how the population density and mobility within a municipality affects the transmission rates of a virus in that specific municipality. It is also important to note that the mobility data in general has a high impact on the results of the model. This research was based on a synthetic dataset, which was calibrated on the Meuzo data set and confirmed to produce similar



patterns to the real data, however using real data like the Mezero dataset that was used by Dekker et al. would be preferred.

As was clearly visible in the exposure maps, the municipalities with the highest total mobility still played an extremely important role in the exposure of agents to the virus. One might think that these big municipalities just show up on the exposure map due to the exposure values from other municipalities from the Randstad all being accumulated. However, even for relatively small municipalities in the far corners of the country, these core municipalities still consistently play a big role in the exposure to the virus. Figure 12 showed four exposure maps in isolation of municipalities in the corners of The Netherlands that were pulled from the exposure map data. In each of these samples, the municipalities of Amsterdam, Rotterdam, Utrecht, The Hague, etc are still clearly consistently locations of exposure, which means that these are the key points of spread for the virus, no matter where the virus starts initially. This opens up an interesting discussion about future research of possible intervention measures. During the COVID-19 pandemic we saw that the government tried to close down North-Brabant early on in the pandemic, because this is where the initial infections had happened. However, wouldn't a better intervention measure be to significantly reduce the mobility allowed to these big exposure locations instead? Closing down the locations where infections have already happened, just causes the intervention measures to be running after the facts. To be able to contain a virus spatially, it is important to prevent the virus from spreading further through these exposure centres. Future research could experiment with the effects of reducing mobility towards and/or from these exposure centres, and see how the effects of these interventions change with the number of days between the initial infection and imposing the intervention. It could also be interesting to experiment with still closing down the originally infected municipality and possibly some surrounding municipalities. One might be able to show that this no longer has any effect after a certain amount of days, depending on the properties of the virus like latent and incubation time. For SARS-CoV-2 this is already known to be around two or three days, as mentioned in the introduction. When intervention measures like reducing mobility to the big municipalities are implemented, it is important to account for the changes in mobility due to a shift in behavior. For example, during the COVID-19 lockdowns people started travelling to Belgium to go shopping.

Other possible directions to take the research are for example using the model with a completely different virus and see how that impacts the spread throughout the country, by for example taking a virus that has a much longer latent time, or a virus where the incubation time is longer than the latent time. With the recent high alerts over a Monkeypox outbreak in Congo [11], future research could also focus on adjusting the model to use a network of sexual contacts, to predict the spread of a possible outbreak of sexually transmitted diseases. Furthermore, another interesting research would be to check whether the same patterns, namely the majority of exposure happening in the big municipalities, can be found when applying the model to other countries. The Netherlands is a densely populated and well-connected country, which causes a virus to spread quite quickly across the whole country, but this might be different for other countries, especially non-western countries. Mobility patterns play an important role in the spread of a virus, so looking at a country which is less densely populated and/or has worse connections could lead to interesting findings as well.

To conclude, this thesis has shown that the model by Dekker et al. can indeed be improved and used for the generation of risk maps. These risk maps seem to match with expectations of the Randstad being more risky than the rest of The Netherlands, because the average mobility per inhabitant is higher in this area. However the risk maps also showed some possible flaws of the model. Due to only inter-municipality mobility being considered, the spread of a virus within the biggest municipalities might be lower than what happens in reality. To improve these possible flaws of the model in the future, some additional parameters have been suggested to add to the model. The exposure maps showed that the majority of exposures to a virus happens in the big municipalities of The Netherlands, independent of the initially infected municipality. This result brought up the idea to create intervention measures

that significantly reduce mobility to the exposure locations ahead of the spread the virus. This idea requires experimentation in future research in order to determine the effectiveness of these measures depending on the amount of mobility reduction, and the time lag between the initial infections and the intervention measures being imposed.

## **Acknowledgements**

The author of this thesis thanks Dr. Deb Panja and PhD student Matthijs Romeijnders for project supervision and discussion about the contents of this research, and Dr. Michiel van Boven for helpful epidemiological insights about the contents of this research.

## References

- [1] Martijn H. H. Schoot Uiterkamp, Martijn Gösgens, Hans Heesterbeek, Remco van der Hofstad, and Nelly Litvak. The role of inter-regional mobility in forecasting sars-cov-2 transmission. *Journal of The Royal Society Interface*, 19(193):20220486, 2022. ISSN 2054-5703. doi: 10.1098/rsif.2022.0486. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2022.0486>.
- [2] Martijn Gösgens, Teun Hendriks, Marko Boon, Wim Steenbakkens, Hans Heesterbeek, Remco van der Hofstad, and Nelly Litvak. Trade-offs between mobility restrictions and transmission of sars-cov-2. *Journal of The Royal Society Interface*, 18(175):20200936, 2021. ISSN 2054-5703. doi: 10.1098/rsif.2020.0936. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2020.0936>.
- [3] Manotosh Mandal, Soovoojeet Jana, Swapan Kumar Nandi, Anupam Khatua, Sayani Adak, and T.K. Kar. A model based study on the dynamics of covid-19: Prediction and control. *Chaos, Solitons Fractals*, 136:109889, 2020. ISSN 0960-0779. doi: <https://doi.org/10.1016/j.chaos.2020.109889>. URL <https://www.sciencedirect.com/science/article/pii/S0960077920302897>.
- [4] Josh A. Firth, Joel Hellewell, Petra Klepac, Stephen Kissler, Mark Jit, Katherine E. Atkins, Samuel Clifford, C. Julian Villabona-Arenas, Sophie R. Meakin, Charlie Diamond, Nikos I. Bosse, James D. Munday, Kiesha Prem, Anna M. Foss, Emily S. Nightingale, Kevin van Zandvoort, Nicholas G. Davies, Hamish P. Gibbs, Graham Medley, Amy Gimma, Stefan Flasche, David Simons, Megan Auzenbergs, Timothy W. Russell, Billy J. Quilty, Eleanor M. Rees, Quentin J. Leclerc, W. John Edmunds, Sebastian Funk, Rein M. G. J. Houben, Gwenan M. Knight, Sam Abbott, Fiona Yueqian Sun, Rachel Lowe, Damien C. Tully, Simon R. Procter, Christopher I. Jarvis, Akira Endo, Kathleen O'Reilly, Jon C. Emery, Thibaut Jombart, Alicia Rosello, Arminder K. Deol, Matthew Quaife, Stéphane Hué, Yang Liu, Rosalind M. Eggo, Carl A. B. Pearson, Adam J. Kucharski, Lewis G. Spurgin, and CMMID COVID-19 Working Group. Using a real-world network to model localized covid-19 control strategies. *Nature Medicine*, 26(10):1616–1622, 2020. ISSN 1546-170X. doi: 10.1038/s41591-020-1036-8. URL <https://doi.org/10.1038/s41591-020-1036-8>.
- [5] Mark M. Dekker, Luc E. Coffeng, Frank P. Pijpers, Debabrata Panja, and Sake J. de Vlas. Reducing societal impacts of sars-cov-2 interventions through subnational implementation. *eLife*, 12:e80819, 2023. ISSN 2050-084X. doi: 10.7554/eLife.80819. URL <https://doi.org/10.7554/eLife.80819>.
- [6] Carl-Etienne Juneau, Anne-Sara Briand, Pablo Collazzo, Uwe Siebert, and Tomas Pueyo. Effective contact tracing for covid-19: A systematic review. *Global Epidemiology*, 5:100103, 2023. ISSN 2590-1133. doi: <https://doi.org/10.1016/j.gloepi.2023.100103>. URL <https://www.sciencedirect.com/science/article/pii/S2590113323000068>.
- [7] Leonardo Azevedo, Maria João Pereira, Manuel C. Ribeiro, and Amílcar Soares. Geostatistical covid-19 infection risk maps for portugal. *International Journal of Health Geographics*, 19(1): 25, 2020. ISSN 1476-072X. doi: 10.1186/s12942-020-00221-5. URL <https://doi.org/10.1186/s12942-020-00221-5>.
- [8] Cameron Zachreson, Lewis Mitchell, Michael J. Lydeamore, Nicolas Rebuli, Martin Tomko, and Nicholas Geard. Risk mapping for covid-19 outbreaks in australia using mobility data. *Journal of The Royal Society Interface*, 18(174):20200657, 2021. ISSN 2054-5703. doi: 10.1098/rsif.2020.0657. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2020.0657>.
- [9] James O. Wheeler. Geography. In Kimberly Kempf-Leonard, editor, *Encyclopedia of Social Measurement*, pages 115–123. Elsevier, New York, 2005. ISBN 978-0-12-369398-3. doi: <https://doi.org/10.1016/B0-12-369398-5/00277-2>. URL <https://www.sciencedirect.com/science/article/pii/B0123693985002772>.

- [10] S.J. Rey. Mathematical models in geography. In Neil J. Smelser and Paul B. Baltes, editors, *International Encyclopedia of the Social Behavioral Sciences*, pages 9393–9399. Pergamon, Oxford, 2001. ISBN 978-0-08-043076-8. doi: <https://doi.org/10.1016/B0-08-043076-7/02516-X>. URL <https://www.sciencedirect.com/science/article/pii/B008043076702516X>.
- [11] Leandre Murhula Masirika, Jean Claude Udahemuka, Pacifique Ndishimye, Gustavo Sganzerla Martinez, Patricia Kelvin, Maliyamungu Bubala Nadine, Bilembo Kitwanda Steeven, Franklin Kumbana Mweshi, Léandre Mutimbwa Mambo, Bas B. Oude Munnink, Justin Bengenhya Mbiribindi, Freddy Belesi Siangoli, Trudie Lang, Jean M. Malekani, Frank M. Aarestrup, Marion Koopmans, Leonard Schuele, Jean Pierre Musabvimana, Brigitte Umutoni, Ali Toloue, Benjamin Hewins, Mansi Dutt, Anuj Kumar, Alyson A. Kelvin, Jean-Paul Kabemba Lukusa, Christian Gortazar, David J Kelvin, and Luis Flores. Epidemiology, clinical characteristics, and transmission patterns of a novel mpox (monkeypox) outbreak in eastern democratic republic of the congo (drc): an observational, cross-sectional cohort study. *medRxiv preprint*, 2024. doi: 10.1101/2024.03.05.24303395. URL <https://www.medrxiv.org/content/early/2024/03/05/2024.03.05.24303395>.

## Code and data availability

All code and data required to run the model can be found on GitHub under the following repository: <https://github.com/larsdekwant/covid-gravity-model>. The *main* branch contains the model by Dekker et al. to reproduce the first COVID-19 wave and interventions, while the *contact-heatmap* branch contains the improved model and risk mapping code. This repository includes instructions on how to run the model locally and provides scripts for running the models in parallel on a computational cluster like Gemini UU.