

02443 Stochastic Simulation Project: Virus Outbreaks

Andreas Madsen – s123598
Frederik Wolgast Rørbech - s123956

June 25, 2016

Contents

1	Introduction	2
2	Data	3
2.1	Population	3
2.2	Airport	3
2.3	Aggression	4
3	Modelling	5
3.1	The SIR Model	5
3.1.1	Assumption	5
3.1.2	Governing equations	5
3.2	Multi-region SIR	6
3.3	Stochastic Model	7
3.3.1	Transfer using Dirichlet distribution	8
3.4	Transfer probabilities	10
4	Validation	11
5	Simulation setup	13
5.1	Variance reduction	13
5.1.1	Control Variates	13
5.1.2	Estimating infection rate	13
5.2	Effect of Global Events	14
5.2.1	Motivation: Zika virus	14
5.2.2	Event definition: Olympic Games in Rio 2016	14
6	Results	15
6.1	With and without Olympic Games	15
6.2	Visualization of result	17
7	Conclusion	19

1 Introduction

This report seeks to analyze virus epidemic on a global scale. More specifically, the primary focus will be to analyze how fast a virus spreads and how many people it infects. In particular it will be examined, how a big event like the Olympic Games in Rio 2016 could affect an epidemic.

Because of the many model uncertainties, such as traveling behaviour, infection rate, etc. and the sudden behavioral change associated with the Olympic Games, a stochastic simulation model is created.

2 Data

The basis of a realistic simulation of global virus outbreak is data on

- Geographical population densities
- Travel connections (commuting and airports)

2.1 Population

The population data used in this report comes from NASA [1]. The data is downloadable in a GeoTIFF file-format which is an image file with population encoded in the pixels.

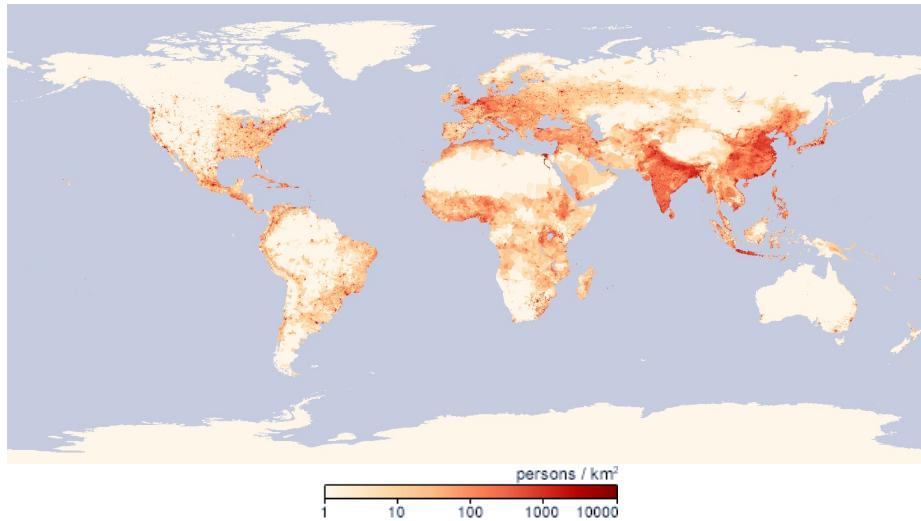


Figure 1: Plot of population in the world, the used resolution is 0.1×0.1 degrees per pixel (3600×1800).

The population dataset encodes 99999.0 as water and 1.0369266 as unknown population on land. Thus to get a population map, a raster mapping of $(x_i < 99999.0 \wedge x_i > 1.0369266) \cdot x_i$ was applied. After this transformation the total world population is just 60,031,128 (without filtering it is 434,130,384,806). This is obviously wrong, thus the data is upscaled such the sum is 7.4 billion people, which is the estimated number of people in March 2016 [2].

2.2 Airport

The airport connection data was taken from OpenFlights [3]. The dataset contains 8021 airports with 37181 airline connections as well as the plane types flying the connection (this could potentially be used for estimating passenger counts). Some airports did not have any connections and were thus removed. After this filtering 3256 airports were left.

Plotting the airport connections we see that Europe, East Coast US and East Cost China are the most connected regions. So if a virus spreads takes hold in these areas we would expect it to cause a global outbreak quickly.

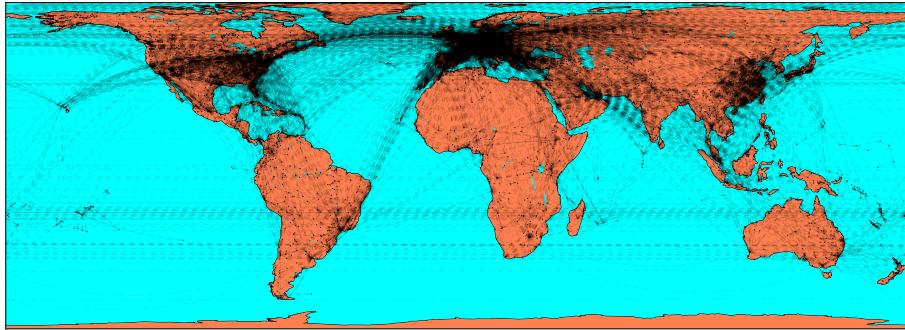


Figure 2: Plot of all airport connections in the dataset. Because there are so many connections each connection is plotted with a low alpha.

2.3 Aggression

Based on the location of the airports a Voronoi partition of the earth's surface is made. From this partitioning one now has geographical regions. The total population of each region can be found by summing over the appropriate pixels in the population dataset. This aggression is the same as done in the GLEaM paper [4].

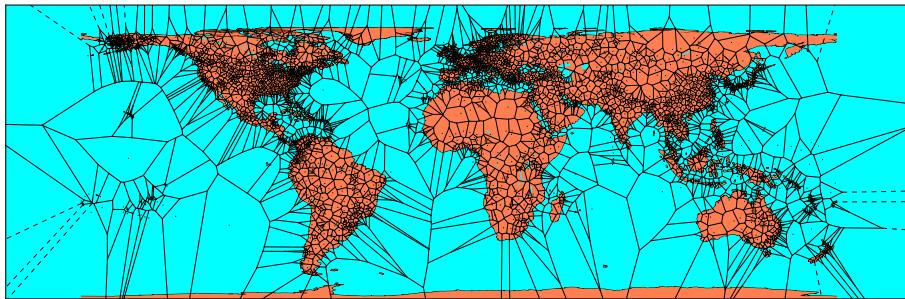


Figure 3: Voronoi tessellation based on airport locations. Airports are marked with a dot, region boundaries with lines and dashed lines.

3 Modelling

This report will analyze how a virus outbreak evolves over time on a global scale. To do so, firstly it will be described how one might model a virus outbreak in a small uniform population. Secondly, this model will be expanded to include multiple population groups and their interactions.

3.1 The SIR Model

3.1.1 Assumption

The Susceptible-Infected-Removed model is a compartmental model assuming that within some subdivision of the population containing N individuals, infection rate, cure rate and spread rates are i.i.d. within each of the 3 “compartments” of individuals:

- Susceptible to the virus
- Infected by the virus
- Recovered and immune to the virus (or alternatively dead)

In reality economic status, job type, air-conditioning etc. will affect how a virus spread [5], however, depending on the heterogeneity of the population these assumptions constitute a good approximation.

3.1.2 Governing equations

The SIR model is governed by 3 differential equations [6]:

$$\frac{dS(t)}{dt} = -\beta \frac{I(t)}{N} S(t) \quad (3.1)$$

$$\frac{dI(t)}{dt} = \beta \frac{I(t)}{N} S(t) - \gamma I(t) \quad (3.2)$$

$$\frac{dR(t)}{dt} = \gamma I(t) \quad (3.3)$$

$S(t)$, $I(t)$ and $R(t)$ are functions describing the number of susceptible, infected and removed (recovered and immune or dead) individuals at time t , β is the rate of infection, γ the rate of removal (dead or cured) and N the total number of individuals.

From equation (3.1) and (3.2) one can see that susceptible individuals become infected by some ratio β and the ratio of already infected individuals. From (3.2) and (3.3) one can see that infected individuals recover with a constant factor γ .

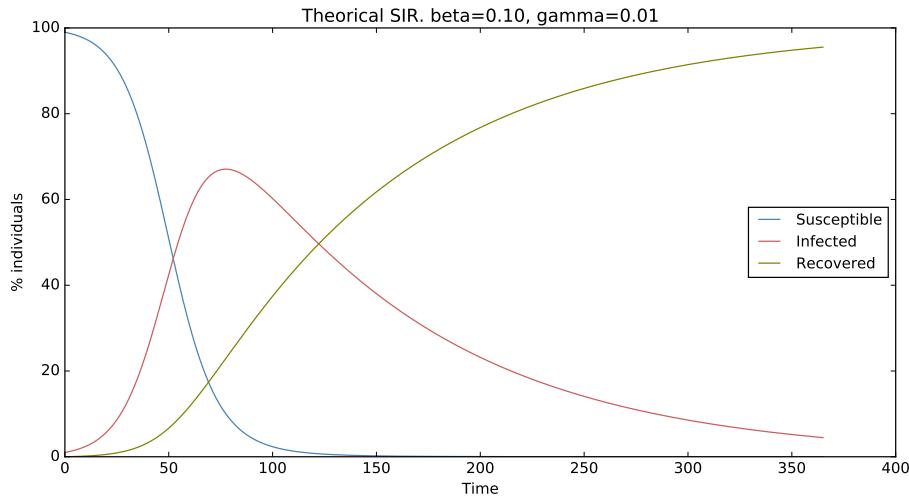


Figure 4: Numerical solution to the nonlinear ODE system defining the SIR model.

From figure 4 it seen how the number of infected individuals increases quickly in the beginning, until a point where there are not enough susceptible to sustain the infection rate. This is also seen from the differential equations:

$$\frac{dI(t)}{dt} = \frac{\beta}{N} I(t)S(t) - \gamma I(t) = 0 \Leftrightarrow \beta \frac{S(t)}{N} = \gamma \quad (3.4)$$

In this hypothetical case almost everybody got infected, but that is not necessarily the case and depends entirely on γ and β .

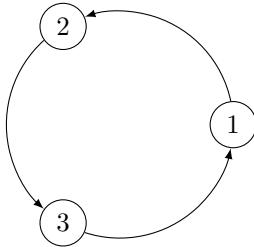
3.2 Multi-region SIR

Let it now be given that instead of having a single population, one has K populations each with initial $N_k, S_k(0), I_k(0)$ and $R_k(0)$. Each of these populations have a probability of transferring individuals to other populations regardless of whether the individual is susceptible, infected or recovered. The governing equations are modified to add the transfers and become

$$\begin{aligned} \frac{dS_k(t)}{dt} &= -\frac{\beta}{N_k} I(t)S(t) + \sum_{i=1}^K (S_i(t)\tau_{i,k} - S_k(t)\tau_{k,i}) && \forall k \in [1, K] \\ \frac{dI_k(t)}{dt} &= \frac{\beta}{N_k} I(t)S(t) - \gamma I(t) + \sum_{i=1}^K (I_i(t)\tau_{i,k} - I_k(t)\tau_{k,i}) && \forall k \in [1, K] \\ \frac{dR_k(t)}{dt} &= \gamma I(t) + \sum_{i=1}^K (R_i(t)\tau_{i,k} - R_k(t)\tau_{k,i}) && \forall k \in [1, K] \end{aligned} \quad (3.5)$$

, where $\tau_{i,j}$ is the probability density of transferring from population i to j per individual.

As an example of the above system let $K = 3$ and the non zero transfer probabilities be defined by the graph:



Solving this system numerically with the outbreak starting in region 1 one yields the following curves:

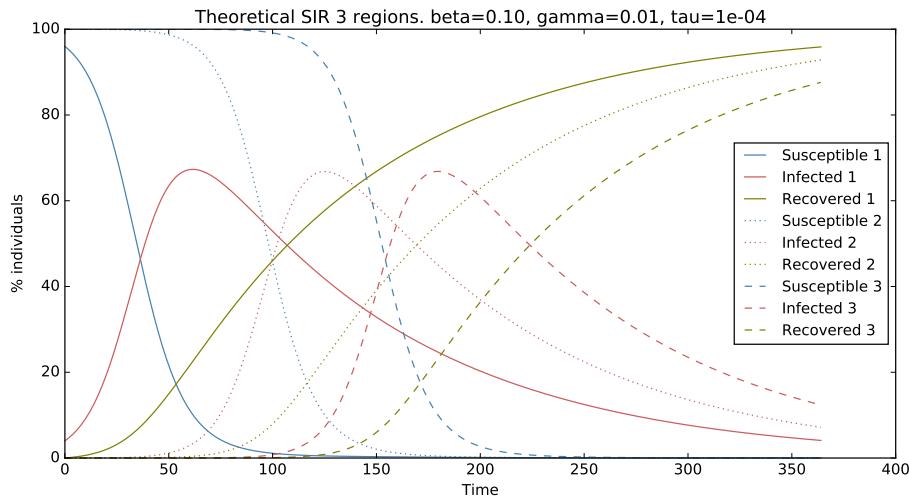


Figure 5: Numerical solution of the nonlinear ODE system defining a 3-region SIR model.

From the above plot it's seen that the virus first spreads in region 1. After a certain time enough infected individuals have been transferred to region 2 where the infection then accelerates. This continues to region 3.

3.3 Stochastic Model

To simulate the multi-region SIR model stochastically, a binomial distribution is used to sample how many people get infected, removed and transferred. The binomial probability is chosen such that the distribution's expectation is the

same as the terms in the differential equations,

$$\begin{aligned}\Delta I_{k,t} &= \text{BINOM}\left(S_{k,t}, \beta \frac{I_{k,t}}{N_k}\right) \\ \Delta R_{k,t} &= \text{BINOM}(I_{k,t}, \gamma) \\ S_{k,t} &= S_{k,t-1} - \Delta I_{k,t} + \sum_{i=1}^K (\text{BINOM}(S_{i,t-1}, \tau_{i,k}) - \text{BINOM}(S_{k,t-1}, \tau_{k,i})) \\ I_{k,t} &= I_{k,t-1} + \Delta I_{k,t} - \Delta R_{k,t} + \sum_{i=1}^K (\text{BINOM}(I_{i,t-1}, \tau_{i,k}) - \text{BINOM}(I_{k,t-1}, \tau_{k,i})) \\ R_{k,t} &= R_{k,t-1} + \Delta R_{k,t} + \sum_{i=1}^K (\text{BINOM}(R_{i,t-1}, \tau_{i,k}) - \text{BINOM}(R_{k,t-1}, \tau_{k,i}))\end{aligned}$$

, where $\tau_{i,j}$ now is the discretized transfer probability.

3.3.1 Transfer using Dirichlet distribution

A problem with this approach, is the small risk that $\text{BINOM}(S_{i,t-1}, \tau_{i,k}) = 0$ and $\text{BINOM}(S_{k,t-1}, \tau_{k,i}) = S_k$ or similar unfortunate combination which would cause $S_{k,t}$ to become negative (i.e. there is a small non zero probability that if transfers are drawn independently that an unfeasible state would occur). In practice, this is a big problem if the number of regions is large. The only way to avoid this problem is to sample all the transfers for a region simultaneously such that $\sum_{i=1}^K \text{BINOM}(S_{k,t-1}, \tau_{k,i}) \leq S_{k,t}$.

As there does not exist a multivariate version of the binomial distribution, the Dirichlet distribution is used as an approximation. This is convenient because $\sum_i X_{k,i} = 1$ if $\mathbf{X}_k \sim \text{DIR}(\boldsymbol{\alpha}_k)$. In addition the Dirichlet distribution has support $X_{k,i} \in [0, 1]$, thus by using $[N_k X_{k,i}]$ as the transfer amount we are guaranteed that each transfer is between 0 and N_k and that the sum of transfers is N_k .

Furthermore to reduced the amount of sampling, the total transfer T is sampled instead of sampling S , I and R independently.

To approximate the binomial distribution using a Dirichlet distribution the expectation is set to be equal the expectation of the corresponding binomial distribution. To simplify notation region k is omitted from the subscript and the summation of τ_k and α_k is introduced:

$$\tau_s = \sum_{i=1}^K \tau_i, \quad \alpha_s = \sum_{i=1}^{K+1} \tau_i \tag{3.6}$$

$$\mathbb{E}[T_i] = \mathbb{E}[N X_i] \Leftrightarrow N \tau_i = N \frac{\alpha_i}{\alpha_s} \quad \forall i \in K \tag{3.7}$$

To allow $\sum_i X_{i,k} \leq 1$ the Dirichlet distribution is extended with another variable $X_{k,K+1}$, which is how many individuals region k should not transfer.

$$\mathbb{E}[T_{K+1}] = \mathbb{E}[N X_{K+1}] \Leftrightarrow N(1 - \tau_s) = N \frac{\alpha_{K+1}}{\alpha_s} \tag{3.8}$$

This is a linear system with n variables and n equations, but it turns out that it doesn't have full rank. To add another equation the variance of T_{K+1} is also set to be equal.

$$\text{VAR}[T_{K+1}] = \text{VAR}[NX_{K+1}] \Leftrightarrow N\tau_s(1 - \tau_s) = N^2 \frac{\alpha_{K+1}(\alpha_s - \alpha_{K+1})}{\alpha_s^2(\alpha_s + 1)} \quad (3.9)$$

The trick to solving this set of equations, is to isolate the sum α_s from (3.8) and insert it into (3.7) and (3.9).

$$\alpha_s = \frac{\alpha_{K+1}}{1 - \tau_s} \quad (3.10)$$

$$N\tau_i = N \frac{\alpha_i}{\frac{\alpha_{K+1}}{1 - \tau_s}} \quad \forall i \in K \quad (3.11)$$

$$N\tau_s(1 - \tau_s) = N^2 \frac{\alpha_{K+1} \left(\frac{\alpha_{K+1}}{1 - \tau_s} - \alpha_{K+1} \right)}{\left(\frac{\alpha_{K+1}}{1 - \tau_s} \right)^2 \left(\frac{\alpha_{K+1}}{1 - \tau_s} + 1 \right)} \quad (3.12)$$

α_{K+1} can now be isolated from (3.12) yielding

$$\alpha_{K+1} = (N - 1) - \tau_s(N - 1) \quad (3.13)$$

, inserting this into (3.10) gives:

$$\alpha_i = \tau_i(N - 1) \quad (3.14)$$

Using the α_i solution α_{K+1} can now be reformulated as

$$\alpha_{K+1} = (N - 1) - \sum_{i=1}^K \alpha_i \quad (3.15)$$

, which is computationally slightly more convenient.

Using this method for choosing α , gives the distribution as seen in figure 6. As seen it has only minor errors when comparing to the binomial distribution.

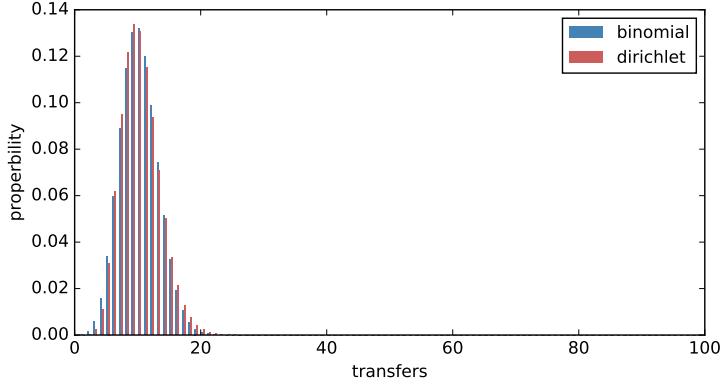


Figure 6: Comparison of the marginal probability $P([NX_1])$ with the binomial distribution. The Dirichlet distribution was scaled up by N , $\tau = (0.1, 0.1)$ and $N = 100$.

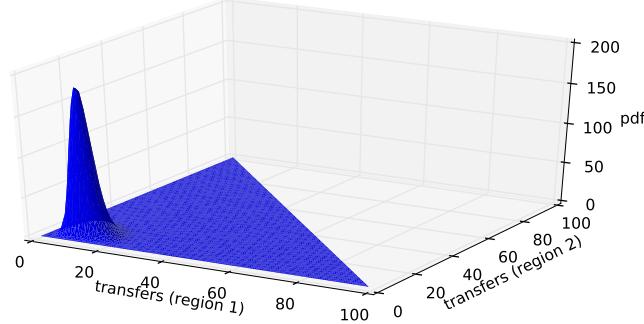


Figure 7: Shows the Dirichlet pdf with $X_3 = 1 - X_1 - X_2$. The Dirichlet distribution is scaled with N , $\tau = (0.1, 0.1)$ and $N = 100$.

3.4 Transfer probabilities

The GLEaM model [4] uses a gravity law for transferring people between neighboring regions within the same country and airline data for transferring people between airports. Unfortunately the GLEaM paper does not specify all the parameters for the gravity law equation and our airline dataset does not contain the traveling frequency. Instead a much simpler model, where the proportional transfer expectation is fixed to some constant, is used:

$$\sum_{i=1}^{K_k} \tau_{k,i} = p_{transfer} \quad (3.16)$$

Using an equal transfer probability one gets $\tau_{k,i} = \frac{p_{transfer}}{K_k}$. To ensure that the population doesn't change over time, this value is used to transfer both from k to i and from i to k , the equivalent transfer probability of this is:

$$\tau_{k,i} = \frac{p_{transfer}}{K_k} + \frac{p_{transfer}}{K_i} = p_{transfer} \left(\frac{1}{K_k} + \frac{1}{K_i} \right) \quad (3.17)$$

Finally one should be wary of transferring more people to the destination than what is already in the destination. To prevent this the transfer probability is scaled with the relative population.

$$\tau_{k,i} = p_{transfer} \frac{N_i}{N_k + N_i} \left(\frac{1}{K_k} + \frac{1}{K_i} \right) \quad (3.18)$$

Using these transfer probabilities the total population for each region will be unchanged since the transfer expectation is symmetric:

$$\mathbb{E}[N_k \tau_{k,i}] = \mathbb{E}[N_i \tau_{i,k}] \Leftrightarrow N_k \tau_{k,i} = N_i \tau_{i,k} \quad (3.19)$$

4 Validation

To validate our implementation of the stochastic simulation model, the 3 region problem previously solved using differential equations, is now solved using the stochastic model.

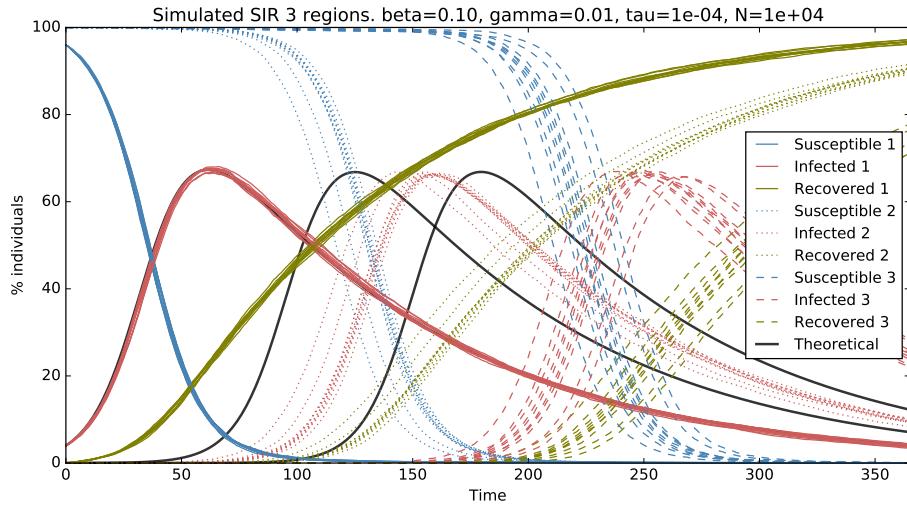


Figure 8: 10 solutions using the stochastic simulation model, on the same 3-region problem as in the theoretical case.

Figure 8 shows a big discrepancy between the solution obtained using differential equations and those obtained using stochastic simulation. This is most likely because the differential equation model allows for transferring people partially. This means regions that has a very low probability of getting infected early in the stochastic model, starts their infection immediately in the differential equation model. This hypothesis can be validated by increasing the number of people in each region (N), as this will cause the discreteness of the stochastic model to be less important. The result of this can be seen in figure 9.

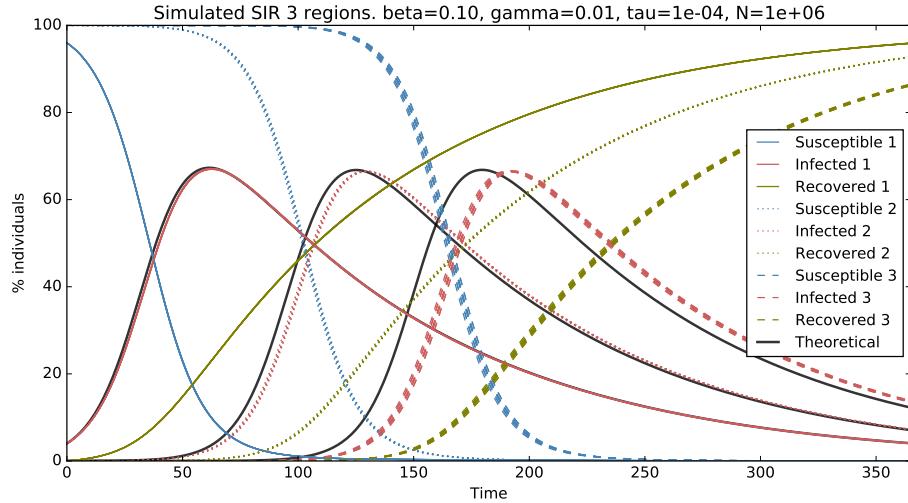


Figure 9: 10 solutions using the stochastic simulation model, on the same 3-region problem as in the theoretical case.

As seen in figure 9 having more people results a much smaller discrepancy. This discrepancy caused by using a discrete model is not necessarily a failure of the stochastic model, but rather it is the differential equation model that overestimates the infection start for initially healthy regions, when the region populations are small.

5 Simulation setup

5.1 Variance reduction

Because the simulation can be time consuming to run, we want to lower the variance of the estimated parameters as much as possible relative to the computational resources used.

5.1.1 Control Variates

Control variates is a variance reduction technique which can be used to reduce the variance of the estimated expected value $\hat{\mu} = \mathbb{E}[Y]$. It works by drawing/-calculating a variable X_i such that $Y_i \approx f(X_i)$. One can then estimate $\hat{\mu} = \mathbb{E}[Z]$ where Z is given by:

$$Z_i = Y_i - c(X_i - \mathbb{E}[X]) \quad (5.1)$$

Choosing $c = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$ gives the largest variance reduction of:

$$\frac{\text{Var}[Z]}{\text{Var}[Y]} = \frac{\text{Var}[Y]}{\text{Var}[Y]} - \frac{\text{Cov}(X, Y)^2}{\text{Var}[X]\text{Var}[Y]} = 1 - \rho_{x,y}^2 \quad (5.2)$$

In our case there is no analytical expression for the covariance or variance of the random variables. Instead c is estimated from the dataset itself. This introduces some overfitting, thus one should take care to reduce the degrees of freedom in the variance, covariance estimates:

$$\hat{c} = \frac{\text{Cov}_{N-1}(X, Y)}{\text{Var}_{N-1}(X)}, \quad \hat{\sigma}^2 = \text{Var}_{N-2}(Z) \quad (5.3)$$

The subscript denotes the degrees of freedom, and $\hat{\sigma}^2$ has $N - 2$ degrees of freedom, because \hat{c} is estimated from the dataset.

5.1.2 Estimating infection rate

From (5.2) it seen that to get a high variance reduction, X_i just need to be heavily correlated with Y_i . In our case we would like to estimate the infection peak time and how many people where infected at the peak. From the theoretical differential equation model those values should be correlated with the infection rate β .

Assuming the stochastic model have a behaviour somewhat similar to the differential equation model, there exists a fairly simple method for estimating β and γ [7].

$$\gamma = \frac{R'_{max}}{I_{max}} \text{ and } \beta = -\frac{\ln(p)}{1-p}\gamma \quad \text{where: } p = \frac{N - R_\infty}{N} \quad (5.4)$$

R'_{max} is the maximum removed difference, and I_{max} is the maximum infected. Due to the construction of the differential equation, these are found at the same time t . R_∞ is how many people were removed in the end, thus p is the survival rate.

5.2 Effect of Global Events

We want to estimate what effect a global event can have on a virus outbreak. We define an event e as simply being a transfer of people from one or more regions r_i to region r_t at a time t_0 , and the reverse transfer of people from r_t to r_i at time t_1 .

5.2.1 Motivation: Zika virus

Recently, a new virus has been observed in South and Latin America. The Zika virus affects pregnant women's fetuses and can cause serious birth defects. The Zika virus is primarily spread by mosquitoes, but can also spread from a man to his sex partners (and from mother to fetus). Some political discussions have been made as to whether or not the Olympic Games should be canceled to prevent the spreading of the virus.

In this report we will run a "hypothetical" simulation, showing how a virus could spread assuming inter-human infections. As such this simulation is primarily an attempt of showing how such a model would work and not a realistic Zika virus predictor.

5.2.2 Event definition: Olympic Games in Rio 2016

In our simulation we define the Olympic Games to take place during $t_0 = '2016-07-03'$ and $t_1 = '2016-07-21'$. According to The Gaurdian [8] Rio expects to receive 380,000 tourist. We choose to transfer people from all over the world to Rio, such that the transfer is proportional to the population size of each region:

$$\text{Transfer}_0(i, j) = 380000 \cdot \frac{\text{population}_i}{N} \quad (5.5)$$

$$\text{Transfer}_1(j, i) = \text{transfer}_0(i, j) \quad (5.6)$$

where $\text{Transfer}_0(i, j)$ denotes the number of people transferred from region i to region j at the start of the Olympics and $\text{Transfer}_1(j, i)$ the people transferred from j to i at the end of the Olympics.

6 Results

6.1 With and without Olympic Games

Choosing infection rate $\beta = 2$, removal rate $\gamma = 0.5$ and an transfer probability $p_{transfer} = 0.005$ we can run a simulation with and without the Olympic Games occurring. We start the simulation at the first day of the games in Rio De Janeiro and set the initial infected population in Rio to 1000. In the following we will look at the SIR curves of 6 major cities.

With the Olympics occurring we get the following curves:

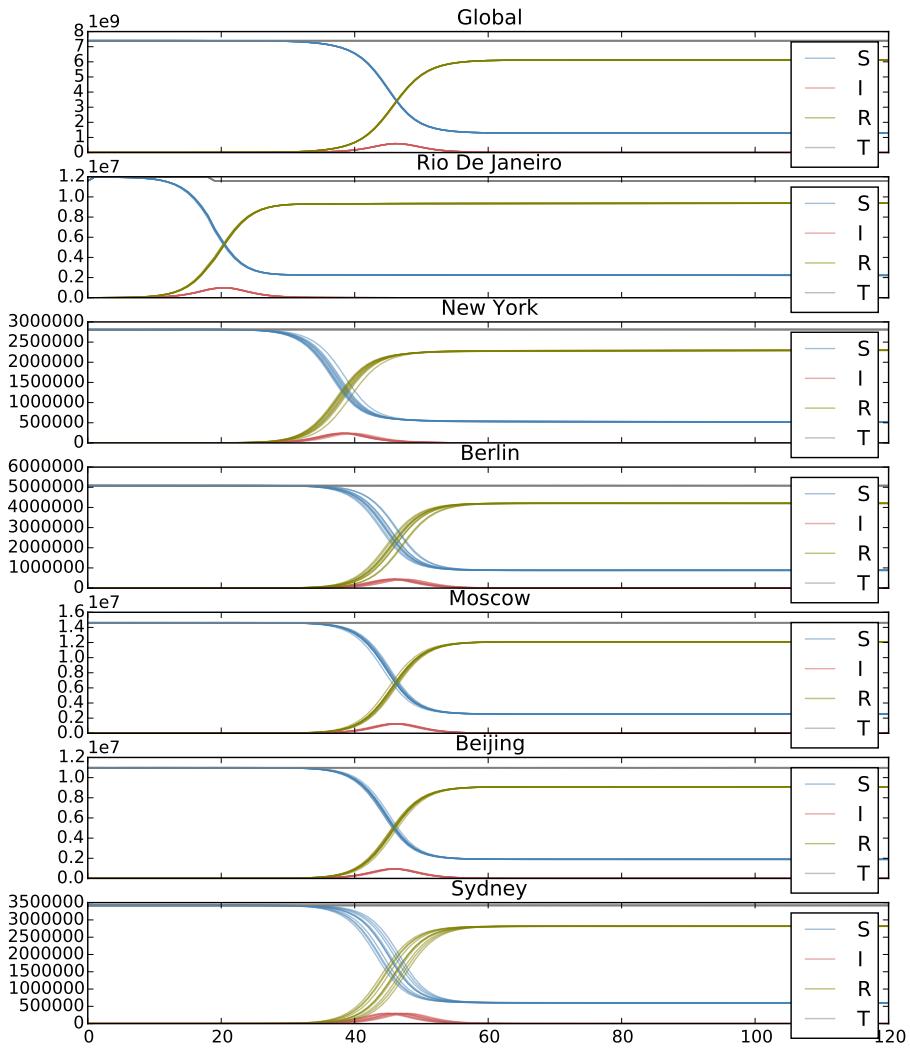


Figure 10: 10 simulations with Olympic Games. Time in days on the x-axis and number of people on the y-axis. Shown is susceptible (S), infected (I), removed (R) and total population (T).

From figure 10 it's seen that except for Rio where the infection started and New York which seems to be infected earliest, the other major cities all seem to reach its infection peak at approximately the same time.

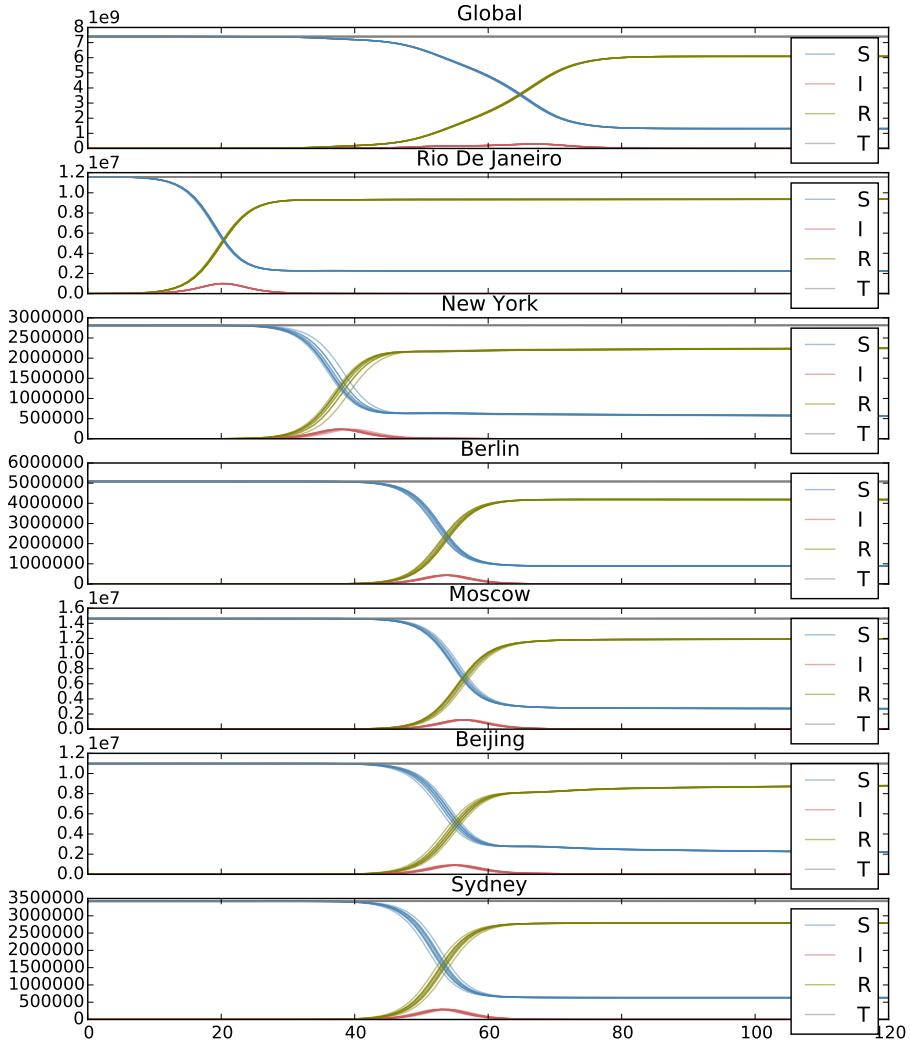


Figure 11: 10 simulations without Olympic Games. Time in days on the x-axis and number of people on the y-axis. Shown is susceptible (S), infected (I), removed (R) and total population (T).

When there is no Olympic Games, as seen in figure 11, the infection peaks occur later, especially for Moscow, Beijing and Sydney. This is emphasized in the summary table 1.

Observation	With OL	Without OL
Peak amount Global [million]	$590.3 \pm 0.88(0.91)$	$291.6 \pm 1.95(1.98)$
Peak time Global	$46.0 \pm 0.00(0.00)$	$67.0 \pm 0.00(0.00)$
Peak time Rio	$20.2 \pm 0.30(0.17)$	$20.0 \pm 0.00(0.00)$
Peak time New York	$38.6 \pm 0.50(0.52)$	$38.5 \pm 0.38(0.38)$
Peak time Berlin	$46.5 \pm 0.38(0.32)$	$54.0 \pm 0.00(0.00)$
Peak time Moscow	$46.0 \pm 0.34(0.18)$	$56.3 \pm 0.35(0.37)$
Peak time Beijing	$46.3 \pm 0.35(0.26)$	$54.9 \pm 0.23(0.24)$
Peak time Sydney	$46.6 \pm 0.69(0.35)$	$53.3 \pm 0.48(0.40)$

Table 1: Results of 10 simulations with and without Olympic Games. Table contains the peak times and amounts for the number of infected individuals. The standard 95%-confidence interval is marked with \pm and the confidence interval using control variates is shown in the parenthesis.

As mentioned, in table 1 one can see that when the Olympic Games occur the peak time is much less spread and generally happens earlier. This is particularly apparent when comparing with the global peak time, which in the case without Olympic Games, happens statistically significantly later than the shown cities. However, in the Olympic Games case, the city peak times are not statistical different from the global peak time (except for the New York and Rio).

The fact that Rio doesn't have a significantly difference in peak time makes sense because this is where the infection starts. Adding 380000 people (out of 12 million people) doesn't make much of a difference. The New York case is also interesting as there isn't a significant change. This is likely because New York being in north America and Rio being in South America is much more closely connected through airlines, compared to cities which are across the Pacific or Atlantic Ocean.

Using control variates generally appear to improve the variance, up to a factor of two in some cases, which seams reasonable. In some cases the control variate confidence interval is larger compared to crude confidence interval. This is because there is less degree of freedom when estimating the variance, which is a result of the extra parameter estimation of c . Thus if the control variates aren't sufficiently correlated, the variance of the estimate won't improve.

Finally table 1 also shows that the number of people infected globally is almost twice as big when including the Olympic Games compared to not including the Olympic Games.

6.2 Visualization of result

When studying a virus outbreak it is interesting to see exactly how the virus spreads. We have thus created an animation on a world map where each region is represented as a dot. Each dot is placed at the region center (airport), the dot size is scaled according to the population size, and the color represents the fraction infected.

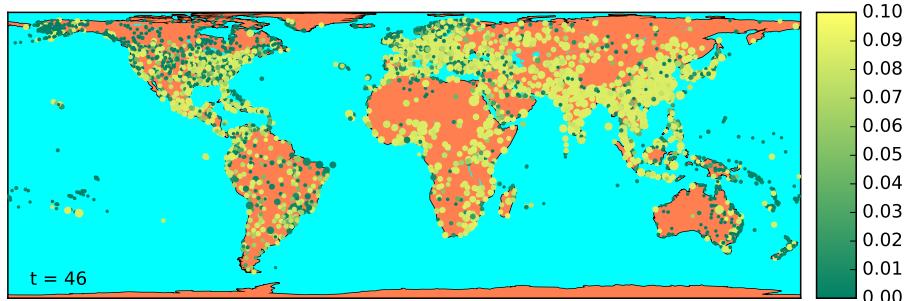


Figure 12: Timestep 46, with Olympic Games. The full animation can be viewed at <https://andreasmdsen.github.io/course-02443-stochastic-virus-outbreaks/>

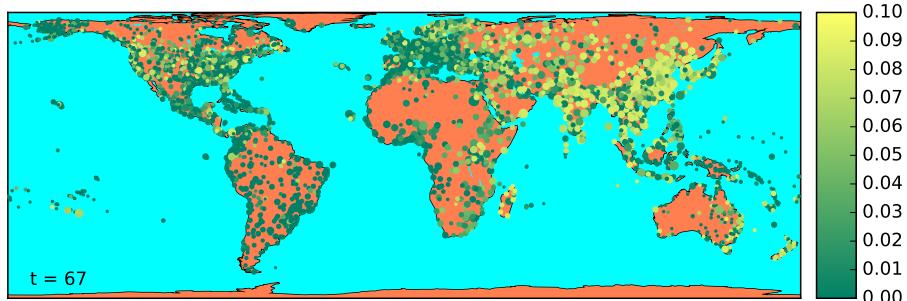


Figure 13: Timestep 67, without Olympic Games. The full animation can be viewed at <https://andreasmdsen.github.io/course-02443-stochastic-virus-outbreaks/>

Animations based on simulations with the Olympic Games shows a high activation across the global at the same time. In simulations without the games it is much more apparent how the virus first spread to regions in the following order:

1. South and Latin America
2. East coast of the USA and then Europe
3. Africa
4. Asia
5. Australia

This ordering is sensible because some regions are more connected than others. E.g. when we start the virus outbreak in Rio de Janeiro, but don't hold the Olympics, the virus naturally first spreads to nearby regions and to regions with which a lot of plane traffic is shared.

7 Conclusion

This report extended the deterministic SIR-model to a multi-region stochastic SIR-model. The model was tested on a small example to validate that it accurately reflects the SIR-model. Using real population and region connection data a hypothetical virus outbreak originating in Rio de Janeiro was simulated. Both through estimated key figures such as peak infection times and peak infection amounts and visual inspection of animations it was shown the that hosting the Olympic Games in a region with a virus outbreak can significantly speed up the outbreak.

References

- [1] N. --. E. Observations. (2016). Population density, [Online]. Available: http://neo.sci.gsfc.nasa.gov/view.php?datasetId=SEDAC_POP (visited on 06/12/2016).
- [2] Wikipedia, *World population — Wikipedia, the free encyclopedia*, [Online; accessed 12-June-2016], 2016. [Online]. Available: https://en.wikipedia.org/wiki/World_population.
- [3] OpenFlights. (2016). Airport, airline and route data, [Online]. Available: <http://openflights.org/data.html> (visited on 06/12/2016).
- [4] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, “Supplementary information multiscale mobility networks and the large scale spreading of infectious diseases”, *Proceedings of the National Academy of Sciences*, [Online]. Available: http://ifiscuibcsic.es/~jramasco/text/sup_gleam.pdf.
- [5] A. Perkins, A. Siraj, C. Warren Ruktanonchai, M. Kraemer, and A. Tatem, “Model-based projections of zika virus infections in childbearing women in the americas”, *bioRxiv*, 2016. DOI: 10.1101/039610. eprint: <http://biorxiv.org/content/early/2016/05/19/039610.full.pdf>. [Online]. Available: <http://biorxiv.org/content/early/2016/05/19/039610>.
- [6] M. J. Keeling and L. Danon, “Mathematical modelling of infectious diseases”, und, *British Medical Bulletin, Br. Med. Bull., Br Med B, Brit Med Bull, Br Med Bull*, vol. 92, no. 1, pp. 33–42, 2009, ISSN: 14718391, 00071420. DOI: 10.1093/bmb/ldp038. [Online]. Available: <http://bmb.oxfordjournals.org.globalproxy.cvt.dk/content/92/1/33>.
- [7] Wikipedia, *Sir modellen — Wikipedia, the free encyclopedia*, [Online; accessed 12-June-2016], 2016. [Online]. Available: <https://da.wikipedia.org/wiki/SIR-modellen>.
- [8] T. Guardian. (2015). Rio 2016: 'the olympics has destroyed my home', [Online]. Available: <https://www.theguardian.com/world/2015/jul/19/2016-olympics-rio-de-janeiro-brazil-destruction> (visited on 06/12/2016).