Lecture with Computer Exercises:
Modelling and Simulating Social Systems with MATLAB

Project Report

Cholera Epidemic in Haiti 2010

Nicolas Stocker

Benedict Borer

Benjamin Pluess

Lukas Buehler

Zurich

May 24, 2012

Agreement for free-download

We hereby agree to make our source code for this project freely available for download from the web pages of the SOMS chair. Furthermore, we assure that all source code is written by ourselves and is not violating any copyright restrictions.

Nicolas Stocker

Benjamin Pluess

Benedict Borer

Lukas Buehler

Contents

1 Abstract

2 Individual Contributions

- Conception and design of model: Nicolas Stocker, Benedict Borer

- Implementation: Benedict Borer

- Analysis and interpretation of the data: Benedict Borer, Nicolas Stocker

- Writing the report: Benedict Borer, Nicolas Stocker, Lukas Buehler **FRAGE: muessen wir Angeben wer welcher Text geschrieben hat?**

- Revision of the report: Nicolas Stocker

- Administrative efforts: Lukas Buehler

- Layouting: Lukas Buehler

- Collection and assembly of data: Benjamin Pluess

- visualization tool: Benjamin Pluess

3 Introduction and Motivation

3.1 Introduction to Haiti, the earthquake and cholera

The shock

12 January 2010, 16.53 local time in Haiti: for an entire minute the earth shook, with a magnitude of 7.0 (on the Moment magnitude scale), with an epicenter 25 km west of Haiti's capital, Port-au-Prince. [1] Followed by at least 52 aftershocks, this earthquake was the most dramatic in the 21st century: the Haitian government estimated that 316,000 people died [2], one million people lost their homes, and 3.2 million were directly affected. [3]

Geological overview

The earthquake was located on the active plate boundary, with the Caribbean Tectonic Plate shifting eastwards relative to the North American Plate. This

Figure 1: Hier steht die Beschreibung des Bildes

4

caused a highly dangerous strike-slip-fault system here with the southern active Enriquillo-Plantain-Garden-fault. The earthquake was caused by a rupture of this system, which has been blocked for the past 2500 years, loading stress. [4]

Primary impacts

As Haiti is one of the poorest countries in the western hemisphere, the economy is highly vulnerable to natural disasters [5]. In addition to earthquakes, the island of Hispaniola, shared by the Dominican Republic and Haiti, has often been struck by hurricanes. The huge devastation and damage in the whole country, but especially in the region of the capital, destroyed not only human life but also infrastructure that would have been necessary to respond to the disaster. All hospitals in the capital were destroyed, as were the air, water and land transport facilities and the communication system. Many of the public and government buildings were damaged (e.g. the National Assembly, the Palace of Justice and the Supreme Court) **CITE**.

The aftermath

In the months following the earthquake, many people slept in the streets because their houses had been destroyed or they feared that their buildings would not survive further aftershocks. Slow distribution of resources resulted in violence and some people began plundering. **CITE**
The earthquake destroyed thousands of families, made a whole social system collapse, and destroyed economic interactions in the country. Anarchy and a rights-free area were the result **CITE**. The Haitians' dramatic fight for survival began in the silent seconds after the earthquake and still goes on.

The approaching problem: cholera

During the period of reconstruction in October 2010, Haiti was confronted by a new and dangerous problem: a cholera epidemic broke out! Cholera is a bacterial infection caused by a bacterium named *Vibro cholerae* **CITE**. The symptoms are mainly watery diarrhea and vomiting, generally caused by drinking infected water. Cholera need not be lethal but requires appropriate treatment in a hospital. In Haiti, some 7025 deaths from cholera had been reported by January 2012. [6]. The source of the disease is still not clearly identified, but localized in the Artibonite River about 100 km north of Port-Au-Prince (The affected victims had drunk water from the infected river.) Some UN investigators did research, hoping to find the real source of the

epidemic, and hypothesized that the initial strain was imported by a UN peacekeeper from Nepal [7]. The Nepali soldiers may be the source of the outbreak as wastewater from the outhouses at their base flowed into and contaminated the Artibonite River. However, who brought the original strain to Haiti is not of primary importance, but rather to control the plague, and this is still one of the most important priorities today in Haiti.

3.2 Motivation

By the end of October 2010, the cholera epidemic had four out of ten Haitian departments in its grip: Artibonite, Centre, Nord and Ouest, including the capital, especially the slum district of Cité Soleil [8]. In the following days cholera was spread all over the country and infected thousands of people. The tragedy of Haiti is not only a simple earthquake rather it is the battle to rebuild a country affected by complicating circumstances such as the cholera epidemic. Our aim is to understand the fast spread of the cholera and to implement a mathematical model, which is able to predict the expansion of such an epidemic in space and time.

4 Description of the Model

A general model for epidemics was established by Kermack and McKendrick over 80 years ago (1927). It is a representative of the so-called SIR models. SIR stands for the three groups a population is divided into: susceptible, infected and removed persons for the particular disease. The model makes the assumptions that the population size is constant, no incubation period exists - which means that individuals fall ill the moment they are infected, and that the time of infectivity is the same as the duration of the clinical disease **(powerpoint presentation MSSSM)**. This implies that infected individuals are always infectious and vice versa. Infected and infectious have therefore to be seen as descriptions of the same state. Because in reality these expressions are not equivalent, either one of them is used in a logical, appropriate manner. Furthermore, after having been infected once, individuals become completely immune to the disease in question [9]. Kermack and McKendrick and have used the plague as an example of an epidemic which can be modeled by their work. Since then other epidemics have been modeled by similar frameworks that can be counted as part of the SIR models. The Kermack-McKendrick model equations are the following:

$$
\begin{cases}
\dfrac{dS}{dt} = -\beta X(t)S(t) & (1) \\[2ex]
\dfrac{dX}{dt} = \beta X(t)S(t) - \gamma X(t) & (2) \\[2ex]
\dfrac{dR}{dt} = \gamma X(t) & (3)
\end{cases}
$$

S, X and R are the numbers of susceptible, infectious and removed persons, respectively. Removed means that those people have either died from the disease or they have recovered and as a result are perfectly immune now. Either way, they are no longer candidates to become infected and they do not belong to the infectious population anymore. Here β stands for the infection rate and γ for the immunity (or death) rate. It is the inverse value of the average duration of infectiousness. If β and γ are constants, the spreading of an epidemic that is modelled with this concept can only decrease or be stopped when enough individuals have been moved from the susceptible to the removed compartment (becoming infectious and then being removed because of death or recovery); a decrease in the value of the susceptible compartment (S) is going to lower the term $\beta X(t)S(t)$ which leads to a decrease of the value of infectious (X), eventually stopping the epidemic.

For diseases that are transmitted through water and person-to-person contact, an extension of the SIR model was developed by Tien and Earn [10]. The main difference is the introduction of a fourth compartment W which represents the waterbody. The water compartment can be contaminated by infectious people and susceptible individuals can be infected by the contaminated water compartment. This water compartment can improve the quality of predictions for diseases that are mainly transmitted through ingestion of contaminated water. Another new feature of the model is the introduction of natural death terms in every compartment with humans (all compartments expect W) and a corresponding birth term in the susceptible compartment. These terms are chosen so that the total number of individuals in all compartments with humans stays constant. The model equations are the following:

$$
\begin{cases}
\dfrac{dS}{dt} = \mu N - b_W WS - b_I SX - \mu S & (4) \\[2ex]
\dfrac{dX}{dt} = b_W WS + b_I SX - \gamma I - \mu X & (5) \\[2ex]
\dfrac{dR}{dt} = \gamma X - \mu R & (6) \\[2ex]
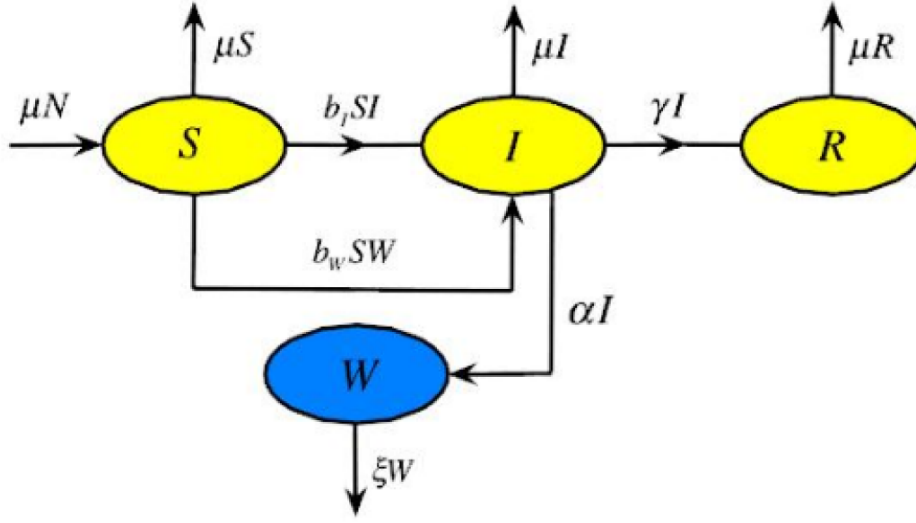\dfrac{dW}{dt} = \alpha X - \xi W & (7)
\end{cases}
$$

Figure 2: Hier steht die Beschreibung des Bildes

μ is the parameter for natural deaths and the birth rate. Natural deaths occur in every human compartment, while all babies are born into the susceptible compartment. b_W represents the rate of infection by water-to-person contact whereas b_I is the rate of infection for person-to-person contact. $\frac{1}{\gamma}$ is the mean infectious period. So γ stands for the rate of infectious persons to either die or recover and become immune. In both cases those individuals change from the X to the R compartment. α is the pathogen shedding rate from infectious persons into the water compartment while ξ denotes the decay rate of the particular pathogen in water.

Figure 2 gives an overview of the parameters and the roles they play in this model [10]. Tien and Earn also recommend rescaling the model equations. Rescaling has some advantages, one of which being the possibility to directly compare different epidemics with different populations. Used for rescaling is the total population number N. The new properties have the following relationship with the old ones:

$$s = \frac{S}{N}, \; x = \frac{X}{N}, \; r = \frac{R}{N}, \; w = \frac{W}{N}\frac{\xi}{\alpha}$$

$$\begin{cases} \dfrac{ds}{dt} = \mu - \beta_W ws - \beta_I sx - \mu s & (8) \\[2mm] \dfrac{dx}{dt} = \beta_W ws + \beta_I sx - \gamma x - \mu x & (9) \\[2mm] \dfrac{dr}{dt} = \gamma x - \mu r & (10) \\[2mm] \dfrac{dw}{dt} = \xi(x - w) & (11) \end{cases}$$

Because W was replaced by the expression $wN\frac{\alpha}{\xi}$, which is a bit more complex than just wN, parameters in expressions with a former capital W are replaced by new parameters. The differential equation for the water compartment has become simpler as only one parameter ξ is needed from now on. Two new parameters are introduced to replace b_W and b_I by $\beta_W = b_W N\dfrac{\alpha}{\xi}$ and $\beta_I = b_I N$ respectively.

> The basic reproductive number R_0 is defined as the expected number of secondary infections that result from introducing a single infected individual into an otherwise susceptible population. R_0 is a fundamental quantity in mathematical epidemiology, which - in the deterministic limit - dictates whether a newly invading pathogen will cause a disease outbreak (direct quote Tien and Earn, 2010 [10]).

For this model, the equation $R_0 = \frac{\beta_I + \beta_W}{\gamma + \mu}$ holds. This model can be applied to the case of a population that is spatially separated into different parts. In our case these were the ten departments of Haiti. The only additional term which was added for this approach gives the person-to-person infections by infectious individuals of different departments. So the susceptible population of every department could be infected by the infectious population or contaminated water of their own department or infectious individuals of other departments. The last term accounts for people who move or travel around. To consider the distances between different compartments a gravity-model was used. Parameter θ depends on the population of the two departments in question and on the distance between their capital cities. The concept is essentially the same as for the gravity equation: $\theta_{ij} = \kappa \frac{p_i p_j}{d^n}$. p_i and p_j are population sizes of different departments, d is given by the distance between the capital cities of the two departments in question, n influences how strong transmission depends on distance and κ determines how important the transmission between departments is thought to be. With this equation a between department transmission rate θ_{ij} can be calculated for every possible constellation. With ten departments, this results in 90 different θ_{ij}, one for every possible combination. Because of the shape of Haiti, it

was divided into a northern and a southern part when determining these distances. Ouest was thought to be the link between those two parts, where all travellers had to pass through. So distances between two departments that did not belong to the same part were taken as the sum of distances between the capital city of the department in the northern part and the capital city of Ouest and the distance between the capital cities of Ouest and the department in the southern part. The equation for the model with different departments (indicated by the index i) are the following:

$$\begin{cases} \dfrac{ds_i}{dt} = \mu - \lambda_i s_i - \mu s_i & (12) \\[2mm] \dfrac{dx_i}{dt} = \lambda_i s_i - \gamma x_i - \mu x_i & (13) \\[2mm] \dfrac{dr_i}{dt} = \gamma x_i - \mu r_i & (14) \\[2mm] \dfrac{dw_i}{dt} = \xi(x_i - w_i) & (15) \end{cases}$$

The force of infection λ is introduced to summarize all the effects of the pathogen, such as person-to-person and water-to-person transmission within a compartment and person-to-person transmission between compartments. While this may give a better overview about the model it hides the non-linear terms. It is important to keep in mind that λ includes parameters as well as variables!

$$\lambda_i = \beta_W w_i + \beta_I x_i + \sum_{j=1}^{10} \theta_{ij} x_j$$
(16)

The figure illustrates the processes and compartments and shows which parameter accounts for which process. Each of the ten departments consists of the four compartments susceptible (s), infectious (x), removed (r) and water (w). Birth and natural deaths (μ) are left out in the figure. n in the gravity model is set to 2 here. Right calibration of every model is crucial to produce results that are close to reality. As the SIWR model has several parameters, this task is quite difficult. With the onset of disease-control interventions the basic reproductive number R_0 is bound to decrease. It can be taken into account by making R_0 time-dependent. This effective reproductive number R_t could look like this with f a discount factor [11].
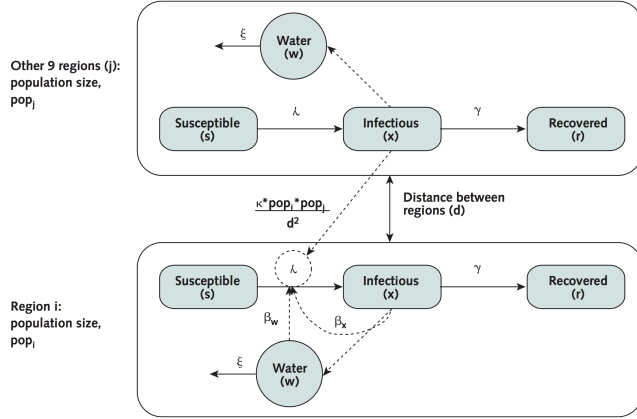
$$R_t = \frac{R_0}{(1+f)^t} \tag{17}$$

Figure 3: Illustration of the model used

5 Implementation

In this section we will describe the implementation of the SIR model in Matlab. To explain why we decided on the SIR and not the SIWR model, we will also look at some of the difficulties of implementing the SIWR model.

5.1 Problems with rescaling

We first tried to rescale our model as proposed by Tien and Earn [10]. It seems, though, that their rescaled model equations are somewhat paradoxical. Looking at the dimensions of equation (4) there already seems to be some contradiction. The dimension of $\dot{s}$,s,i and r is km^{-2}. The dimension of μ, β_w and β_i is declared as day^{-1}. This shows that the dimensions on the left and right side of the equation cannot be the same. Because of this, we decided to let our susceptible, infectious and removed compartments have the dimension of percentages. As a consequence, we do not know what dimensions our coefficients have (e.g. day^{-1} or $cells * ml^{-3}$). We will not be able to make the same conclusions as [11]. Primarily, we will be able to make qualitative statements about the proportions of transmission caused by the local population versus the interdepartmental transmission.

5.2 Data from the MSPP

We collected the raw data from the Ministère de la Santé Publique et de la Population de la République d'Haiti (MSPP) [6] . To get our ten districts, we added the numbers from Port-au-Prince to the department Ouest. In total, we collected the data

for 92 days (the months November and December 2010 and January 2011) and ten different districts. To compare the original data to the model output we needed the amount of susceptible, infectious and removed in each department. Whilst extracting all the data we found some inconsistencies within the data sheets themselves. For this reason we chose to take the data from what we considered the most reliable on the sheets (mainly the data from hospitals). The infectious data were taken directly from the data sheets as the number of hospitalizations. The removed data are the sum of deaths in hospital and the healed. The susceptible data were calculated as the total population minus the infectious and removed. All the values were divided by the respective total population to produce the percentages.

One of the bigger problems with the SIWR model is the lack of knowledge about cholera and the water quality during the epidemic. Because we found no data available in the internet and Tuite et al. [11] don't cite any source for their data, we started out by setting all the values for the water quality on day 1 to 0. The amount of cholera bacteria would rise due to the bacterial shedding rate from the infected to the water compartment. After some trial runs we discovered that the SIWR model is extremely sensitive to changes in the initial conditions of the water quality. This was one of the main reasons we decided to use the SIR model. Continuing with the SIWR model would result in a high uncertainty about the variability caused by the false initial conditions of the water quality versus the parameters fitted by the solver.

5.3 Differential equations

In order to observe intercompartmental transmission, we used the same modification as Tuite et al. [11] in their SIWR model. The force of infection λ in the ith district is:

$$\lambda_i = \beta_{Xi}x + \sum_{j=0}^{I} \Theta_{ij}x_j \tag{18}$$

The influence of infection of the *jth* district on the *ith* district Θ is calculated the same way as Tute et al.. Also as stated by Tuite et al. [11], within the short time period that we observe the natural birth/death rate is of no significance CITE. This reduces our SIR model equations to:

$$\begin{cases} \dot{s} = -\lambda s & (19) \\ \dot{x} = \lambda s - \gamma x & (20) \\ \dot{r} = \gamma x & (21) \end{cases}$$

The benefit of this simplified model is that we only need to fit a few coefficients.

12

As a consequence, our fit will most probably be worse than the fit by the SIWR model.

5.4 Creation of the solver

Our goal was to build a solver which, starting from our given initial conditions, tries to find the optimal values of the coefficients to reduce the sum of the squared residuals (SSQ). The SSQ is calculated by squaring the difference of the predicted percentage of a compartment (e.g. susceptible, infected or removed) and the observed data and summing them up over all days and districts. The solver would stop calculating if either the difference of the two following SSQ was smaller than 0.001% or if the solver has calculated more than 1000 cycles. This was introduced to exclude any infinite loops. In each cycle of the solver, all the coefficients were slightly increased and decreased by 2% of its value. These modified coefficients were then combined to create all possible scenarios. This results in 2^k scenarios where k is the number of coefficients that need to be fitted. Because of our simplification to the model we only needed to fit κ, β_x and γ. In each cycle our solver calculated 8 different scenarios. Table 1 shows the different scenarios calculated.

Table 1: Table with the 8 scenarios, + representing a 1% increase and - a 1% decrease of the coefficient

κ	β_x	γ
+	+	+
+	+	-
+	-	+
+	-	-
-	+	+
-	+	-
-	-	+
-	-	-

The solver computes the SSQ starting with these new coefficients and chooses the ones where the SSQ was reduced the most. The next cycle is then started with these coefficients.
We decided to build our own solver mainly for two reasons:

1. We did not find any adequate solver in the internet for the format of our data

2. We thought it might be an interesting challenge to build one ourselves as a learning experience

Because this is not an established solver there are a few things one has to keep in mind. We tested the consistency of the solver by running the program up to 50 times with the same initial conditions. In each case the result was the same. To test the sensitivity of the solver to different initial conditions we let the whole program run with different starting values of coefficients. This is explained in section 5.5. Another weakness of the solver is that it never lets a parameter stay the same but either increases or decreases its value. This could be excluded by adding a third scenario value, where the parameter does not change. The formula for the total amount of scenarios is $scenarios = n^k$ where n is the number of different value changes and k the number of coefficients being fitted. This can result in a large amount of calculations. We decided that the possible inaccuracy of 2% of the coefficients has a small effect on the overall fit and proceeded with only two value changes in the solver.

5.5   Initial conditions

In our model we not only have initial conditions for the compartments susceptible, infected and removed, but also for the coefficients fitted by the solver. The starting values for the compartments were simply taken from the observed data day 1 (so 1st November 2010). The coefficients of the model were estimated beforehand by simple nonlinear parameter fitting. Based on these calculations we defined an interval of possible values for κ, β_x and γ. Because we decided on a different rescaling method we could not rely on the parameter values given by XXXXX. Similar to the method used in our solver, we created different scenarios with all the combinations of starting values. This resulted in $4^3 = 64$ different scenarios. To calculate the final model we chose the fitted coefficients from the scenario with the smallest overall SSQ.

6   Simulation results and discussion

6.1   Results of the SIR model

All the values for the compartments s, x and r calculated by the model are within the possible intervals. The best fit values for β_x, γ and κ are 3.9114, 3.8816 and 7.5279e-12, respectively. The basic reproductive number $R_0 = \frac{\lambda S}{\gamma}$ is always smaller than 1 because the product of λS is always smaller than γ. This would predict a non spreading disease or a very mild epidemic (CITE). Figures 4 and 5 show the mean over all districts for the susceptible and removed compartments, respectively. One can see that the slope of the curve is steadier for the calculated model compared to the observed data. The observed curve of the infectious compartment in figure 6 is highly nonlinear and the magnitude is slightly earlier than our calculated data. It is

important not only to look at the means but the distribution of s, x and r within Haiti.
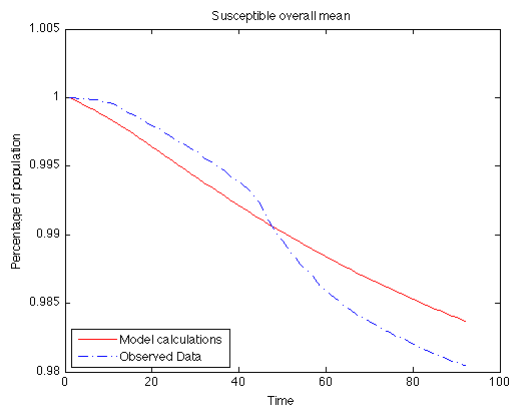

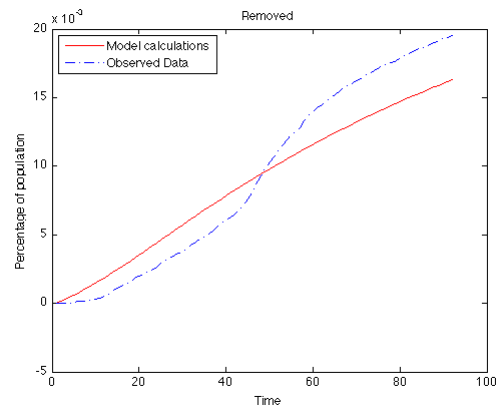Figure 4: Mean over all districts for the susceptible compartment

Figure 5: Mean over all districts for the removed compartment

The plots 7, 8 and 9 show the observed and calculated values of all ten compartments. The districts Nord, Nord-Est and Grand Anse show an incredible increase and decrease in the removed and susceptible compartments, respectively. These compartments also have unlikely observed data for the infectious compartment, as seen in figure 9. In the district Grand Anse the amount of infected people decreases from 0.1% to 0.04% within around 3 days. On the other hand, there are some extremely good fits, for instance in the departments Nord-Ouest and Centre.

One of the benefits of our simple model is the possibility to compare the effect of transmission within the same district versus transmission between districts. The force of infection λ consists of two parts, as seen in equation (lambda). Here $\beta_x x$ represents the transmission within the same district and $\sum_{j=0}^{I} \Theta_{ij} x_j$ the transmission between compartments. Figure XX shows the interdepartmental transmission coefficients over time. The best fit of $\beta_x = 3.9114$ is always around 10^4 times bigger. This predicts that in our model the transmission within the same district is much more important than the interdepartmental transmission.

Figure 10 shows the spread of the removed compartment on the 1st of January 2011. Here the observed data is shown on the left while our calculated data is on the right. Our model predicts the spatial spread of the disease to some extent quite well. After the change of initial conditions in the infectious compartment for the district Grand Anse we were able to also predict the spread in the far west. Our calculated least-infected districts are Sud and Nippes. The observed data show that

15

they are further to the east, namely Ouest and Sud-Est. Our model also calculates the highest-infected district to be Artibonite, whereas the observed data shows that they are Nord-Est and Nord.

6.2 Discussion

The reliability and accessibility of data have proved to be a crucial part of building this model. Because of the lack of water quality data we have decided to construct an ordinary SIR model instead of an SIWR. Even with our simplified SIR model equations we were able to predict the epidemic over 3 months to some extent. In our model, the interdepartmental transmission is of no importance to the spread of the disease. There are two main reasons we believe why this might be in our model. Tien and Earn [10] put the percentage of infected, susceptible and infectious populations in relation to the area observed in the model. For us this would mean we put the respective percentages in relation to the area of the districts. The result would be a density, for instance percent infected per square kilometer. For districts with many inhabitants in a relatively small area this would result in much higher values for s,x and r compared to sparsely inhabited districts. When using the gravity model as proposed by Tuite et al. [11] the districts with a high x would have a much greater influence on the force of infection λ.
Another reason might be the gravity model we used in itself. Because we chose a different way of rescaling our values, s,x and r are much higher than those calculated by Tuite et al. [11]. As a result the relation $\frac{pop_i * pop_j}{d^n}$ might be unsuited for our dimensions. We tried to even this out by fitting κ.

7 Summary and Outlook

7.1 Outlook

In response to the cholera outbreak, the Haitian government and partner agencies initiated emergency public health response activities aimed at treating suspected cholera cases and preventing new ones. Response activities included mass media cholera campaigns through the radio, hygiene promotion activities by community health workers, distribution of water purification tablets and soap, and limited distribution of oral rehydration solution (ORS) sachets. (Quelle: http://www.ncbi.nlm.nih.gov/pmc/articles/
Regarding the mass media campaign, it would be more effective if we understood the distribution of the cholera bacteria in waterways. With our model it is unfortunately not yet possible to do so. Our model needs more detailed information, which is currently not available. In the case of a future epidemic it would be important not only to gather data about the infected and recovered populations, but also to

16

collect information on water quality, i.e. cholera bacteria concentrations, and the consumption of drinking water from surface water reservoirs.
We believe such a dataset would greatly improve the possibility of modelling a system such as a cholera epidemic, like the one in Haiti in 2010.

8 References

References

[1]

[2]

[3]

[4]

[5]

[6]

[7]

[8]

[9] WO Kermack and AG McKendrick. Contribution to the mathematical theory of epidemics. *PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON SERIES A-CONTAINING PAPERS OF A MATHEMATICAL AND PHYSICAL CHARACTER*, 115(772):700–721, AUG 1927.

[10] Joseph H. Tien and David J. D. Earn. Multiple Transmission Pathways and Disease Dynamics in a Waterborne Pathogen Model. *BULLETIN OF MATHEMATICAL BIOLOGY*, 72(6):1506–1533, AUG 2010.

[11] Ashleigh R. Tuite, Joseph Tien, Marisa Eisenberg, David J. D. Earn, Junling Ma, and David N. Fisman. Cholera Epidemic in Haiti, 2010: Using a Transmission Model to Explain Spatial Spread of Disease and Identify Optimal Control Interventions. *ANNALS OF INTERNAL MEDICINE*, 154(9):593–U59, MAY 3 2011.
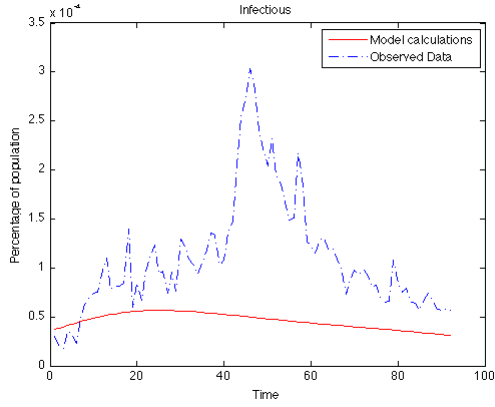
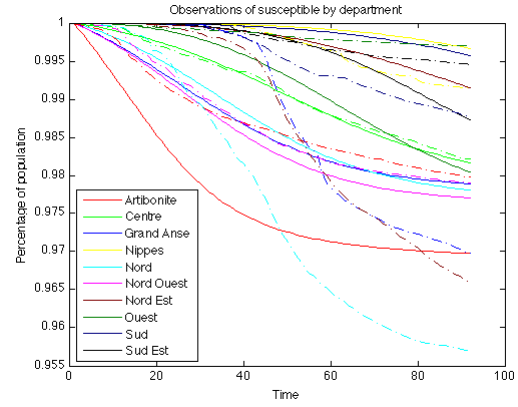Figure 6: Mean over all districts for the infectious compartment

Figure 7: Observed and calculated data for all districts of the susceptible compartment
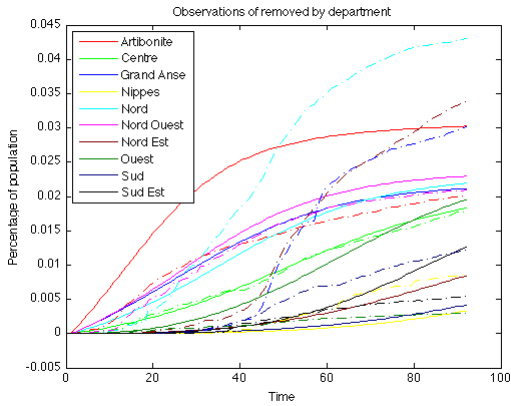

Figure 8: Observed and calculated data for all districts of the removed compartment
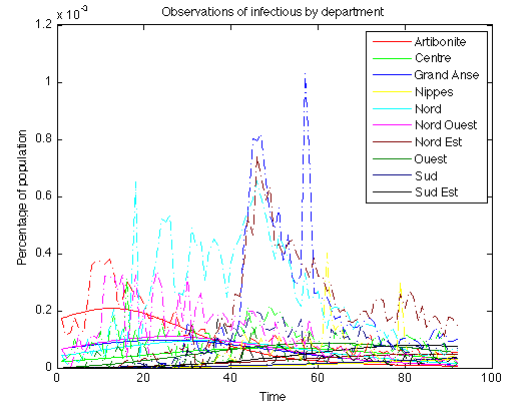

Figure 9: Observed and calculated data for all districts of the infectious compartment

18

Figure 10: Geographical display of observed and calculated removed data at t=92