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 22, 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

[0] [0] [0]

2 Individual contributions

3 Introduction and Motivations

The Shock

12. January 2010, 16.53 local time in Haiti: during one minute the earth is shaking with magnitude 7.0 (on the Moment magnitude scale) with an epicentre 25 Km west of Haitis capital Port-au-Prince. Followed by at least 52 aftershocks this earthquake is the most dramatic in the 21. Century: the Haitian government estimated that 316000 people had died , one million people lost their home and 3.2 million are directly affected.

Geological overview

The quake was located on the active plate boundary with the Caribbean Tectonic Plate shifts eastwards relative to the North American Plate. This causes a highly dangerous strike-slip-fault system here with the southern active Enriquillo-Plantain-Garden-fault. The quake is caused by a rupture of this system, which has been blocked for the last 2500 years, loading stress

Figure 1: Hier steht die Beschreibung des Bildes

Primary impacts

As Haiti is one of the poorest country in the Western Hemisphere the economy is highly vulnerable to natural disasters. In addition to earthquakes, the island of Hispaniola shared by the Dominican Republic and Haiti, has often been stroked by Hurricanes.

The huge devastation and damage in the whole country, but especially in the region of the capital, destroy not only human life, furthermore infrastructure, which would have been necessary to respond to the disaster. Not only all hospitals in the capital, as well air, water and land transport facilities and the communication system was destroyed. As well a lot off the public and government buildings were damaged (e.g. the National Assembly, the Palace of Justice and the Supreme Court)

The time afterwards

In the following months after the earthquake many people slept in the streets because their houses has been destroyed or they were feared that their buildings wont survive upcoming aftershocks. Slow distribution of recourses resulted in violence and some people began with plundering.

The earthquake destroyed thousands of families and made collapse a whole social system and destroyed economic interactions in a country. Anarchy and a right-free area was the result. Haitians dramatic fight for survival began in the silent seconds after the earthquake and still goes on.

Upcoming problem: Cholera

During the period of reconstruction in October 2010 Haiti was confronted by a new dangerous problem: a cholera epidemic broke out!

Cholera is a bacterial infection caused by a bacterium named as Vibro cholerae. The symptoms are mainly watery diarrhoea and vomiting. The transmission occurs generally by drinking infected water. Cholera doesnt have to be lethal but requires an appropriate treatment in a hospital. In Haiti in January 2012 some 7025 deaths by Cholera have been reported The source of the illness still is not clearly identified, but localized in de Artibonite River about 100 Km north of Port-Au-Prince (The affected victims had drunk water from the infected river). Some UN investigators did researches, hoping to find the real source from the Epidemic and they are guessing the initial strain was imported by UN peacekeeper from Nepal . The Nepali soldiers may be the source of the outbreak as wastewater from their outhouses at their base flowed into and contaminated the Artibonite River . However who brought the origin strain to Haiti is not of primary importance, rather than to control the plague and this is still one of the most important priority nowadays in Haiti.

Motivation

By the end of October 2010 the Cholera epidemic accomplished four out of ten Haitians departments: Artibonite, Centre, Nord and Ouest. That includes as well the Capital , especially the slum district Cit Soleil . In the following days Cholera was spread all over the country and infected thousands of peoples. The tragedy of Haiti is not only a simple earthquake; it is more a battle rebuilding a country affected by complicating circumstances such as the Cholera Epidemic.

Our aim is to understand the fast spreading of the dangerous illness Cholera and to implement a mathematical model, which is able to predict the expansion of such an epidemic in space and time. We believe to help getting a deeper understanding of the interaction of a human environment interaction and so to do our part for protecting human lives in further catastrophic events.

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 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 the other way around. Infected and infectious have therefore to be seen as descriptions of the same state. Because in reality those 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 (Kermack and McKendrick, 1927). Kermack and McKendrick and have used 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:

$$\frac{dS}{dt} = -\beta X(t)S(t)$$

$$\frac{dX}{dt} = \beta X(t)S(t) - \gamma X(t]$$

$$\frac{dR}{dt} = \gamma X(t)]$$

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 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 (Verweis zu ihrem paper). 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:

$$\frac{dS}{dt} = \mu N - b_W W S - b_I S X - \mu S$$

$$\frac{dX}{dt} = b_W W S + b_I S X - \gamma I - \mu X$$

$$\frac{dR}{dt} = \gamma X - \mu R$$

$$\frac{dW}{dt} = \alpha X - \xi W$$

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.

The flow diagram gives an overview of the parameters and the roles they play in this model. (Taken from Tien and Earn). The model equations that we used were rescaled as recommended by Tien and Earn. Rescaling has some advantages, one of which being the possibility to directly compare different epidemics with different populations. Another one is the absence of units in S, X, W, R. Used for rescaling is
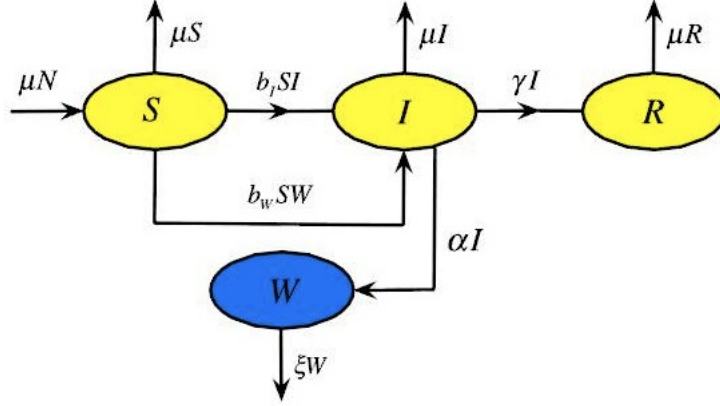
Figure 2: Hier steht die Beschreibung des Bildes

the total population number N. The new properties have the following relationship with the old ones:

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

$$\frac{ds}{dt} = \mu - \beta_W w s - \beta_I s x - \mu s$$

$$\frac{dx}{dt} = \beta_W w s + \beta_I s x - \gamma x - \mu x$$

$$\frac{dr}{dt} = \gamma x - \mu r$$

$$\frac{dw}{dt} = \xi(x - w)$$

Because W was replaced by the expression $wN\frac{\alpha}{\xi}$ which is a bit more complicated 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 $\beta_W = b_W N\frac{\alpha}{\xi}$ 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). For this model the equation

$R_0 = \dfrac{\beta_I + \beta_W}{\gamma + \mu}$ 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 that 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: $\theta_{ij} = \kappa \dfrac{p_i p_j}{d^n}$. 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 10 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 of departments that did not belong to the same part where 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:

$$\frac{ds_i}{dt} = \mu - \beta_W w_i s_i - \beta_I s_i x_i - s_i \sum_{j=1}^{10} \theta_{ij} x_j - \mu s_i$$
$$\frac{dx_i}{dt} = \beta_W w_i s_i - \beta_I s_i x_i + s_i \sum_{j=1}^{10} \theta_{ij} x_j - \gamma x_i - \mu x_i$$
$$\frac{dr_i}{dt} = \gamma x_i - \mu r_i$$
$$\frac{dw_i}{dt} = \xi(x_i - w_i)$$

Another parameter λ is introduced to summarise all the effects of the pathogen like person-to-person and water-to-person transmission within a compartment and person-to-person transmission between compartments. While this may give a better
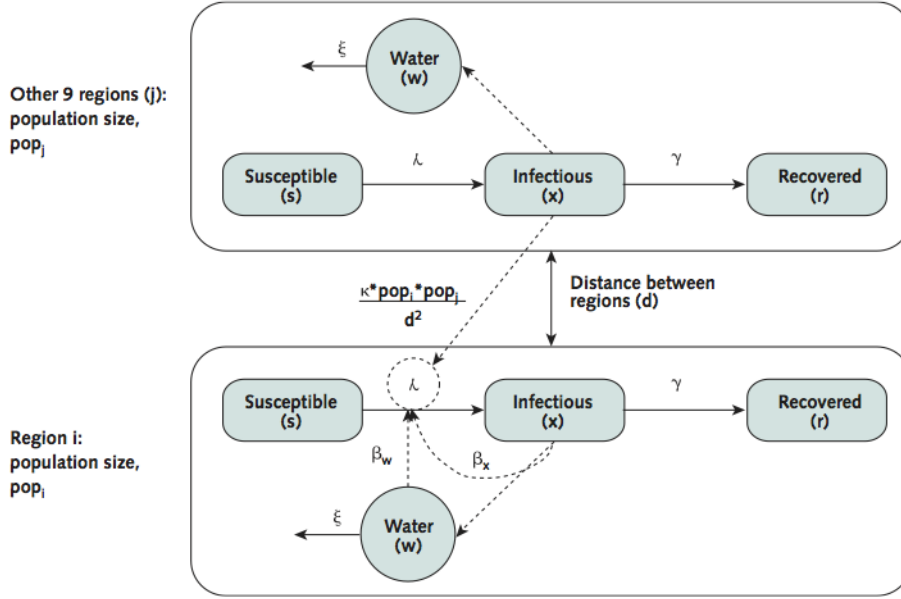
Figure 3: Summary of the used model

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!

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

λ_i represents the total force of infection in the ith department. The model equations with λ included look the following:

$$\frac{ds_i}{dt} = \mu - \lambda_i s_i - \mu s_i$$
$$\frac{dx_i}{dt} = \lambda_i s_i - \gamma x_i - \mu x_i$$
$$\frac{dr_i}{dt} = \gamma x_i - \mu r_i$$
$$\frac{dw_i}{dt} = \xi(x_i - w_i)$$

The figure summarizes the processes and compartments and shows which parameter accounts for which process. Each of the 10 departments consists of the 4 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.

10

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:

$$R_t = \frac{R_0}{(1+f)^t}$$

Here f is a discount factor. (Tuite et al.)

5 Implementation

In this section we will mostly explain 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 explained by Tien and Earn. It seems though, that their rescaled model equations are somewhat paradox. Here are the rescaled model equations as described by Tien and Earn:

$$
\begin{cases}
\dot{s} = \mu - \beta_w sw - \beta_x sx - \mu s & (1) \\
\dot{x} = \beta_w sw + \beta_x si - \gamma x - \mu x & (2) \\
\dot{r} = \gamma i - \mu r & (3) \\
\dot{w} = \xi(x - s) & (4)
\end{cases}
$$

Looking at the dimensions of equation (5.1) 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 can not 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 XXXX. Mainly we will be able to make qualitative statements about the proportions of transmission caused by the local population versus the interdepartemential transmission.

5.2 Data from the MSPP

We collected the raw data from the Ministre de la Sant Publique et de la Population de la Republique d'Haiti (MSPP) CITE. To get our ten districts, we added the numbers from Port au Prince to the departement Ouest. In total we collected the data for 92 days (the months November, December and January 2010) and ten different districts. To compare the original data to the model output we need 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 was taken directly from the data sheets as the number of hospitalisations. The removed data is the sum of deaths in hospital and healed. The susceptible data was calculated as the total population minus the infectious and removed. All the values were divided by the respective total population to receive 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 XXX don't cite any source for their data, we started out by setting all the values for the water quality on day one to zero. 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 uncertainity 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 intercompartimental transmission, we used the same modification as XXX in their SIWR model. The force of infection λ in the ith district is:

$$\lambda_i = \beta_{Xi}x + \sum_{j=0}^{I} \Theta_{ij}x_j \tag{5}$$

The influence of infection of the jth path on the ith patch Θ is calculated the same way as XXX. Also as stated by XXXXX, 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:

$$\begin{cases} \dot{s} = -\lambda s & (6) \\ \dot{x} = \lambda s - \gamma x & (7) \\ \dot{r} = \gamma x & (8) \end{cases}$$

The benefit of this simplified model is that we only need to fit a few coefficients. 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 sum of squares 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 1% 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 simpflification 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 1% 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 suceptible, 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 preliminary by simple nonlinear parameter fitting. Based on these calculations we defined an intervall 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 resultet in $4^3 = 64$ different scenarios. To calculate the final model we chose the coefficients from the scenario with the smallest overall SSQ.

6 Simulation Results and Discussion

6.1 Results of the SIR model

6.2 Comparison with original data

7 Summary and Outlook

Outlook

In response to the cholera outbreak, the Haiti 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 radio and 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 to the mass media campaign, it seems helpful to know how the cholera bacteria would distribute in waterways. With our model it is not yet possible to make an exact forecast about the distribution in water ways. Based on that knowledge the protection of population would be directly improved and should be used when cholera is threatening human beings.

Our model needs more detail information, which nowadays are not available. Hence we recommend a research during an epidemic with the aim to collect data of the consumation and quality of the water. In our case it would be helpful to know the quantity, localisation and exact time of an output (?) from water out of a contamined (?) river.

We belive such a dataset would highly improve the possibility of modelling a system like a cholera epidemic like that one in Haiti 2010.

8   References

References

ja. nei. *jaaj*, 1999.

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.

Tien. Hydro. *Nature*, 1982.