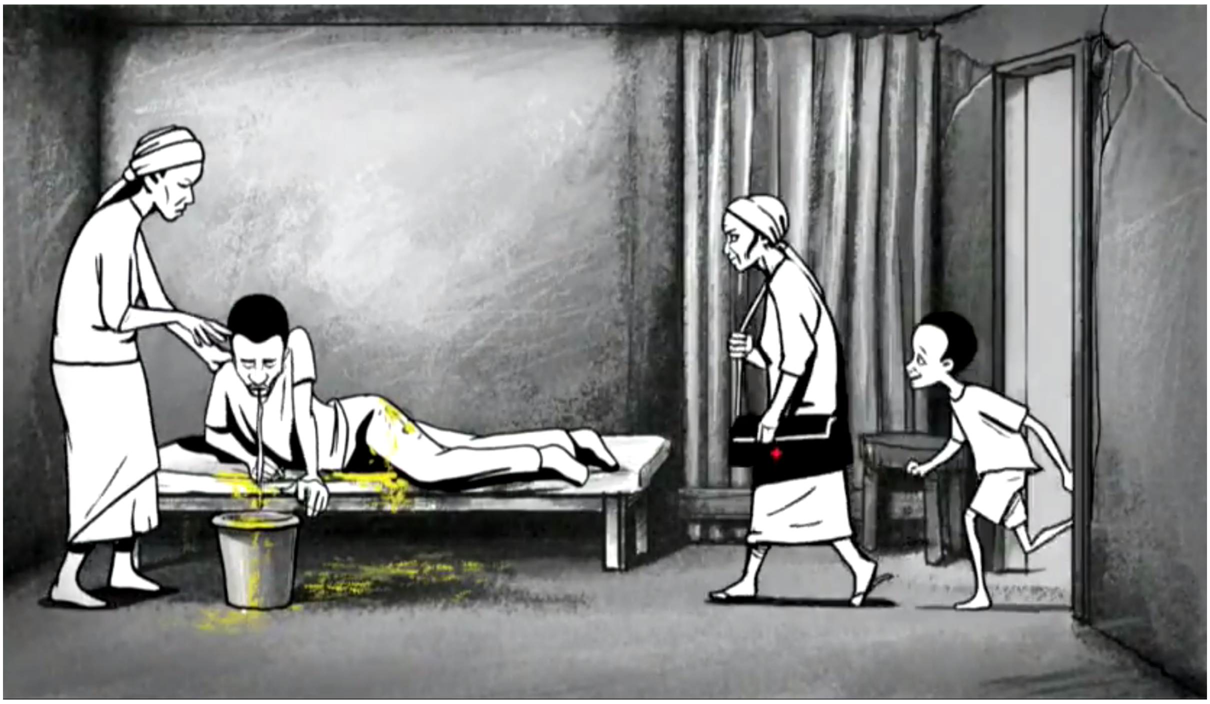
UNIVERSITY OF TWENTE.

CEM MSc Thesis



The role of water in cholera diffusion

Improvements of a cholera diffusion model for Kumasi, Ghana

Author:
Tom Doldersum

Supervisors:
Ir. P.W.M. Augustijn-Beckers
Dr. Ir. D.C.M. Augustijn
Dr. M.S. Krol

June 19, 2013

Title: The role of water in cholera diffusion

Subtitle: Improvements of a cholera diffusion model for Kumasi, Ghana

Thesis for the degree of Master of Science
in Civil Engineering and Management

Author: Tom Doldersum

Institute: University of Twente

Departments: Water Engineering and Management &
Geo-Information Processing

Supervisors: Ir. P.W.M. Augustijn
University of Twente
Department of Geo-Information Processing

Dr. ir. D.C.M. Augustijn
University of Twente
Department of Water Engineering and Management

Dr. M.S. Krol
University of Twente
Department of Water Engineering and Management

Date: June 19, 2013

Cover photo: Under the authority of the Global Health Media Project Goodman and van Dyke (2011) developed the film: "The story of cholera". The photo represents the moment that the nurse starts with taking care for the cholera infected man on the bed. The aim of the film is to inform people about the spreading mechanisms of cholera and how to prevent the cholera to spread. This film shows all the spreading mechanisms that are investigated in this research and shows that doing research to cholera is still relevant and necessary.

Summary

Cholera is still one of the most feared infection diseases, especially in those countries where clean drinking water is not available to the local people. In order to develop an effective strategy against the spread of cholera it is important to understand the behaviour of cholera which can be studied by simulating it. Therefore, Useya (2011) developed an agent based model that simulates the spread of cholera for the outbreak of 2005 in Kumasi, Ghana. Although the developed cholera model was able to reproduce the epidemic curve of cholera, it contained some parts, particularly the hydrological part, that should be improved before it can be used for more practical purposes. This study focused on the improvement of these parts. Therefore the research objective is:

Improve the model of Useya (2011) by implementing the hydrological processes that play a major role in the spread of V. cholerae, to gain more insight in the spread of cholera via water and use the model to evaluate different scenarios to make the strategy against cholera more effective.

An analysis of the original hydrological procedure shows that: (i) the velocity of the water was unrealistically low and (ii) the study area was rather small. To improve this procedure the velocity of water is now based on the formula of Manning and the study area is enlarged to the catchment area. The next step was to calibrate the procedure. Since no discharge data were available, the simulated discharges were compared with the outcomes of a generally accepted model: the Curve Number method. The calibration resulted in a Nash-Sutcliffe of 0.95 and a Relative Volume Error of $\pm 0.2\%$ for the epidemic period. The model was validated for the same period of the years 2006, 2009 and 2010 resulting respectively in a NS of 0.92, 0.94 and 0.94 and for all years the RVE was lower than $\pm 0.3\%$.

The main improvements made to the original cholera model are: (i) removing inconsistencies between the model and the report of Useya (2011) and (ii) adapt the model's procedures to the enlarged study area. This is done to compare the simulated and available data to study the cholera diffusion within the study area. The model was calibrated on the relative contribution of each community (eleven are taken into account) to the total number of cholera cases. Since it is widely accepted that there is a substantial under-reporting of cases, the use of absolute values would not be appropriate. The calibration resulted in a r^2 of 0.87. This means that the model was able to reproduce the geographical distribution well. The resulting epidemic curve and the contribution of the transmission mechanisms are both comparable to what is reported in the literature. The main contribution to the number of cases is caused by transmission of cholera via water.

The model analysis shows that within the cholera model the hydrological procedure is very important, because 75% of the cholera cases get infected via river water that is contaminated by the runoff from the dump sites. Due to the Environment-Human (EH) transmission procedure the model clearly shows a random spatial pattern of the diffusion process while this is not expected from literature. The model is quite sensitive to the scheduling of the daily activities and changes to the bacteria procedure.

The scenario analyses show that there is a strong relation between the epidemic curve and the rainfall. When it rains river water gets infected and the model shows a strong increase in

the number of cholera cases. Crucial for the model performance to simulate the geographical distribution well is a process where the probability to get infected with cholera depends on the living location within the study area. Furthermore, the model shows that removing dump sites that are situated close to the river resulted in a decrease in the number of cholera cases.

Finally recommendations for follow up studies are:

- Pay attention to processes that: contaminate rain water that drops on the dump site and describe the behaviour of V. cholerae in the river.
- Extend the activity model with activities that can explain the upstream movement of the cholera epidemic. For example the mobility of individuals to work, markets or family.
- After the model improvements have been done, perform a validation for either another catchment of Kumasi for which data are available or another cholera outbreak in an area with characteristics similar to this study area.

Preface

"*The man who moves a mountain begins by carrying away small stones*", Confusius. In September 2012 I started ambitiously on this thesis. My subject enables me to improve the world and help to get cholera out of the world. Unfortunately I have to conclude that cholera is still a big threat for many people. In addition, during the last few months I saw that the impact of my results became less and less. Now I realize that the stones Confucius meant are really small.

Although cholera is still a threat, I am proud of the result and glad that I have done my research about the spread of cholera. This would not have been possible without the support I received from my supervisors. Therefore I would like to thank Ellen-Wien for introducing me to Agent Based modelling, for her large number of comments on my draft versions and the discussions we had about the assumptions for the input data. I am grateful to Denie for reminding me to focus on the relevant points of the research and also for his large number of comments on my draft versions. Also I would like to thank Maarten for his comments on my model assumptions and his critical questioning about the methodology I used.

I would like to thank the other residents of the WEM graduation room for the moments of fun and the wisdom walks. Last but definitely not least, I would like to thank my girlfriend, Kimberley, for her support during the research and my parents, Jan en Lida, for their unconditional support during my entire education.

Tom Doldersum

Enschede, June 2013

Contents

List of Figures

List of Tables

List of Abbreviations

ABM Agent Based Model

CN Curve Number

DEM Digital Elevation Model

EH Environment to Human

HEH Human to Environment to Human

HEHD Human to Environment to Human due to contaminated raindrops from dump sites

HEHP Human to Environment to Human due to playing at the dump site

GPS Global Positioning System

HH Human to Human

NS Nash-Sutcliffe coefficient

RVE Relative Volume Error

SCS Soil Conservation Service

TC Total number of Cases

VT Vector Transmission

V. cholerae Vibrio cholerae

List of Symbols

#	Number of ...
$\#C_{com}$	Number of cases in the community
$\#C_{tot}$	Total number of cases in the study area
$\#R$	Number of raindrops on dumpsite
γ_s	Averaged slope in basin
A	Surface of the catchment area
D	Duration excess rainfall
D_f	Maximum distance for flies from a dump site to a household
D_{hd}	Actual distance from household to dump site
D_i	Dump site infection level
D_{imax}	Maximum dump site infection level
F_L	Flow length
GD	Percentage of cholera cases of a community to the total number of cholera cases
HL_{lm}	Switch point low to middle hygiene level
HL_{mh}	Switch point middle to high hygiene level
I	Percentage of rainwater that infiltrates
I_{max}	Maximum amount of infiltration
L	Hydraulic river length
m_c	Manning roughness coefficient
m_{sh}	Manning roughness coefficient sheet flow
P	Precipitation
P_a	Potential abstraction
P_{cw}	Probability to get infected by drinking contaminated water
P_{Feh}	Probability of fetching water contaminated with EH
P_{ff}	Probability of flies on the food
P_{fw}	Probability that a household has to fetch water
P_{if}	Probability to get infected due to contaminated food
P_{max}	Threshold rain before the tapewater stops working
P_{sh}	Probability to get infected due to soiled hands
Q	Discharge
Q_{obs}	Observed discharge
Q_{sim}	Simulated discharge
Q_p	Peak discharge
R	Hydraulic radius
r^2	coefficient of determination
R_f	Risk of getting infected by flies
R_c	Risk of children getting soiled hands
R_{hh}	Risk on Human to Human infection
S	Slope
S_{min}	Minimum slope
S_{sg}	Switch point from sheet to gully flow

S_{gr}	Switch point from gully to river flow
T_c	Duration of patch contamination due to raindrops
T_{lag}	Lag between precipitation and peak discharge
T_p	Time from start of the precipitation till the peak discharge
T_r	Time of the peak discharge to the end
T_{wp}	Threshold value for waterpoints

Chapter 1: Introduction

Cholera is still one of the most feared infection diseases in public health, especially in those countries where clean drinking water is not available to the local people (Alam et al., 2007). Cholera has a long history, before 1195 seven large cholera pandemics had already been reported (Kaper et al., 1995). Nowadays it is estimated that cholera still causes 120.000 deaths in Africa each year, this number increases every year and a large number of this are children (Faruque et al., 1998; WHO, 2011). In addition it is widely accepted that there is a substantial under-reporting of cases (Lee, 2001), this indicates that the real problem is even worse.

This research can contribute to the knowledge about the spreading of cholera, which will help to avoid the spreading of cholera and to reduce the number of cholera cases and deaths in the future. The next section provides a brief description on the background of this research. Section 1.2 presents the research objective and questions, followed by the research strategy and thesis outline in section 1.3.

1.1 Background

In order to develop an effective strategy against the spread of cholera it is important to understand and simulate the behaviour of cholera. Several studies have researched this subject. One study that investigated the simulation of the diffusion of cholera with an Agent Based Model (ABM) is Useya (2011). Her objective was to simulate the diffusion of cholera in Kumasi, Ghana. The broader objective of the research was to evaluate the contributions of different transmission mechanisms to the overall spread during a cholera outbreak.

To achieve this objective Useya (2011) developed an ABM that simulates the diffusion of cholera within the study area. The developed cholera model was able to reproduce the epidemic curve of cholera, but it has some parts that should be improved before it can be useful for more practical purposes. For example the cholera model did not determine the right velocities for the runoff and the runoff in the model accumulated on certain points in the river without flowing to the outlet.

In 1854 John Snow established the link between cholera and contaminated water of a well in London. Later it was found that Vibrio cholerae (V. cholerae), the bacteria that causes cholera, is also a natural member of the aquatic microbial community (Bertuzzo et al., 2008). For this reason it is important that the hydrology of any cholera simulation model represents reality.

1.2 Research objectives and questions

The model of Useya (2011) is state of the art for modelling the diffusion of cholera in ABM. However not all the hydrological processes that play a role in the spread of cholera are implemented yet. When these processes are implemented and the cholera model simulates the diffusion of cholera better. Then it becomes possible to improve the strategy against the spread of cholera by performing scenario analyses. Therefore the objective of this research is:

Improve the model of Useya (2011) by implementing the hydrological processes that play a major role in the spread of V. cholerae, to gain more insight in the spread of cholera via water and use the model to evaluate different scenarios to make the strategy against cholera more effective.

The following questions serve to accomplish the research objective. The questions are distinguished in two categories: Preliminary research questions (P) and Main research questions (M).

- P1. What is V. cholerae and how does it behave in water?
- P2. What processes are implemented in the original cholera model of Useya (2011) and which of them needs to be improved?
- P3. What are the relevant characteristics of the study area?
- M1. What are the relevant hydrological processes and how well can they be modelled in an ABM?
- M2. How can the cholera model be improved and what is the performance after improvement?
- M3. What are the most important mechanisms that explain the spread of cholera according to the model?
- M4. Which input data influence the model outcomes the most?

1.3 Research strategy and thesis outline

Each chapter of this thesis will cover one research question. Figure 1.1 presents a schematic overview of this research and shows how the research questions are related to each other and the objective.

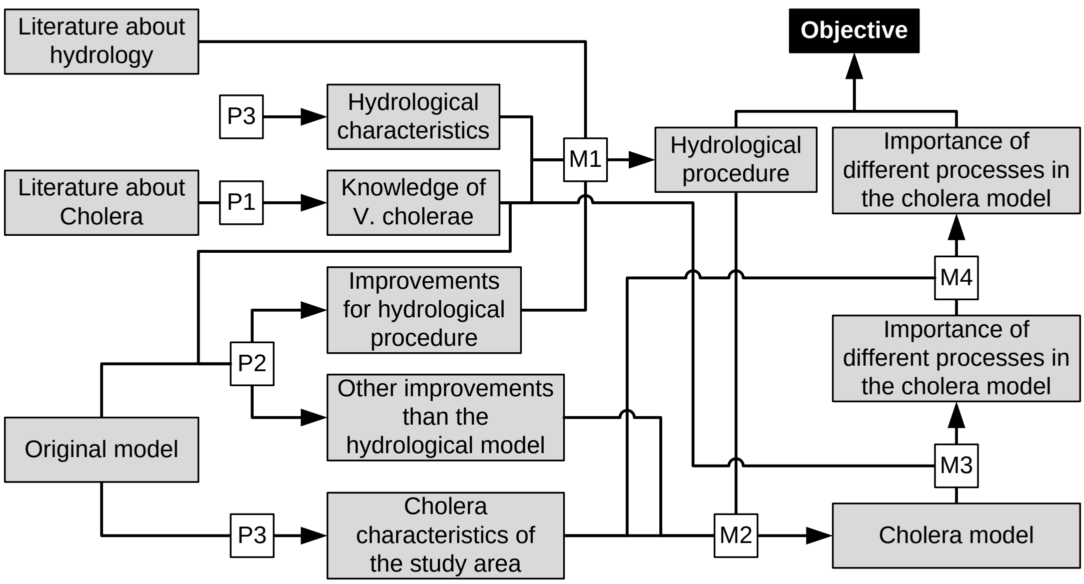


Figure 1.1: *The underlying structure behind this thesis. The numbers correspond to the research questions, which follow the same order as the chapters in this thesis. The answer to each research question is found by combining and comparing the related research objects. The objective is achieved when a verified hydrological procedure is developed and successfully implemented in the cholera model resulting in a good geographical distribution of the cholera epidemic of 2005 in Kumasi, Ghana.*

The objective of P1 is to find which hydrological processes are relevant for the spread of cholera. Gaining more knowledge enables it to have a critical view on the developed model by Useya (2011) to be sure that all relevant aspects are included. Chapter 2 describes the findings of a literature study about the characteristics and behaviour of V. cholerae in general and water.

The objective of P2 is to make an overview of the original model, specifically to gain more insight in the hydrological processes and which other processes in the model can be improved based on the information of P1. The outcomes of this study are presented in chapter 3.

P3 is meant to gather relevant information about the study area. Two types of information are relevant: information that is needed for the hydrological procedure and information that will be used to improve the procedures of the original cholera model. An overview of the study area and the data is presented in chapter 4.

M1 has a direct relation with the first part of the objective, namely to improve the hydrological processes of the original model. The knowledge gained from the preliminary research questions will be used to develop the hydrological procedure. Afterwards the procedure will be calibrated and validated. The whole process from development to verification is described in chapter 5.

M2 is meant to implement the hydrological procedure and improve the other findings that follow from P2. Then the new cholera model will be calibrated. Chapter 6 present the development, implementation and verification of the cholera model.

The aim of M3 is to gain more insight in the implemented processes of the cholera model, to make a distinction between less and more important transmission mechanisms. Furthermore, there will be a check when these matches with the findings from P1. This research question focusses especially on the role of water within the cholera model because this procedure will be developed in this research. The findings are described in chapter 7.

The aim of M4 is to gain more insight in the behaviour of the model and to get a feeling what measures can do to reduce the spread of cholera. Therefore the model will simulate several scenarios. The outcomes of these scenario analyses are described in chapter 8.

Finally, chapter 9 provides a brief answer to the main research questions and the recommendations for further research.

Chapter 2: Cholera

The first step is to perform a literature study on the relevant aspects of cholera regarding the spread of the disease, especially the behaviour of cholera in water. This is used to find the relevant hydrological processes and to be able to critically analyse the cholera model of Useya (2011). Therefore this chapter contains a brief introduction into the cholera disease in section 2.1, followed by a description of the most important habitat conditions of the V. cholerae bacteria with respect to the spread characteristics by water in section 2.2. Section 2.3 contains a discussion about the infection dose of V. cholerae and section 2.4 describes the transmission mechanisms of V. cholerae.

2.1 General information about cholera

Cholera is a bacterial infection of the intestine caused by certain strains of Vibrio cholerae (V. cholerae), see figure 2.1. The bacteria can be found in food or water, so people can get infected after oral ingestion. When V. cholerae is ingested it produces enterotoxins (toxins that act in the gastrointestinal tract) whose actions on the mucosal epithelium are responsible for the characteristic symptoms of cholera, namely: acute watery diarrhoea and vomiting (Lee, 2001).

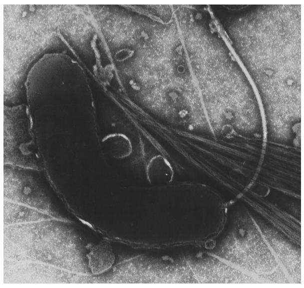


Figure 2.1: *V. cholerae bacteria O395, 0.2-0.3 μm wide and 3-6 μm long (Kirn et al., 2000)*

The most important hallmarks of the disease are: (i) a high degree of clustering of cases by location and season, (ii) highest rates of infection for children from 1 to 5 years old in areas with endemic infection, (iii) antibiotic resistance patterns that frequently change from year to year, (iv) clonal diversity of epidemic strains, (v) protection against the disease by improved sanitation and hygiene and pre-existing immunity (Faruque et al., 1998). In addition, a healthy person may become hypotensive within an hour after the onset of symptoms and may die within 2-3 hours, however in most cases the death will come within one day. In literature much more information can be found about the behaviour of the V. cholerae in the human body, e.g. Kaper et al. (1995) and Chaudhuri and Chatterjee (2009).

Many species of V. cholerae exist. Important distinctions between the different species are made on the basis of: (i) the serogroup, (ii) production of cholera enterotoxin and (iii) the potential for epidemic spread. At the moment around 200 serogroups of V. cholerae have been identified, which are all labelled OX (where X=1−200). All major epidemics in the past were caused by the same serogroup, namely O1. However, the epidemic in South-East Asia of 1992 learned that serogroup O139 is also able to cause a cholera epidemic. The first strain of O139 was found in Bangladesh therefore it is also called 'Bengal' (Alam et al., 2007). Later it was also found on the African continent in the estuarine waters and sediments of Beira in Mozambique (Du Preez et al., 2010). Only those strains of O1 and O139, which produce cholera toxins are associated with epidemic cholera. Other strains of these two types do not produce cholera toxins and are therefore not involved in epidemics. Although some strains that do not belong to O1 or O139 produced occasional outbreaks of cholera, they have not been associated with any large epidemic or extensive pandemics so far. Lastly, it is generally accepted that O139 behaves similarly in the aquatic environment to O1 (Du Preez et al., 2010).

2.2 Habitat conditions of V. cholerae

V. cholerae is a natural member of the aquatic environment (Colwell et al., 1981; Huq et al., 1983; Xu et al., 1982). The most optimal conditions for the survival and growth of V. cholerae O1 are aquatic environments which are rich in nutrients (Borroto, 1997). Temperature has the most direct and significant effect on the ecology of V. cholerae, warmer temperatures enhances growth and multiplication in the environment (Lipp et al., 2002). The optimal salinity is between 5-25‰, however V. cholerae can withstand lower salinities as long as ions of Na^+ are available that are required for growth (Lipp et al., 2002; Singleton and Attwell, 1982). Many times iron, which is used for the growth of bacteria, is the limiting factor due to its low solubility in water. To deal with this problem V. cholerae is able to produce iron-chelating siderophes to take insoluble iron from the environment (Lipp et al., 2002; Patel and Koornhof, 2004). V. cholerae cells will rapidly die when the acidity becomes lower than 6.0, but they are quite tolerant to alkaline conditions (Chaudhuri and Chatterjee, 2009). Furthermore, V. cholerae can live under both aerobic and anaerobic conditions (Epstein, 1993; Kan et al., 2004).

These abiotic factors are important for the growth of V. cholerae. Since these are also basic conditions for the growth of some phytoplankton species and aquatic plants, this are indicators of suitable conditions and also provide food for zooplankton. A correlation was found between the seasonal occurrence of algal blooms and cholera outbreaks, however no evidence is provided that this leads to the growth of toxigenic V. cholerae (Epstein, 1993; Reidl and Klose, 2002). Research showed that V. cholerae attach to the chitin particles of copepods (a group of small crustaceans found in the sea and nearly every freshwater habitat (Wikipedia, 2012)) and other crustaceans survived significantly longer (Lipp et al., 2002).

2.3 Infection by cholera

Cholera is a dose-dependent disease, an infection requires 10^4 cells. One single copepod may carry 10^4 to 10^6 cells of V. cholerae, thus an incidental ingestion of a few copepods (sized 1-2 mm (Wikipedia, 2012)) in untreated drinking water can lead to an infection.

The incubation period of the V. cholerae is 1-5 days (Mintz et al., 2005). This depends on the carrier of the V. cholerae. The symptoms can be identified in about 24-48 hours after the infection (Glass and Black, 1992). Furthermore humans with blood group “O” are more susceptible to getting cholera than other blood groups, the reason for this is not clear (Sack et al., 1998; Harris et al., 2005).

2.4 Transmission mechanisms

Besides the oral ingestion of naturally existing V. cholerae, there are several routes to get infected with V. cholerae. Figure 2.2 presents an overview of these routes. Within this research the transmission routes are defined as: Human to Human (HH), Environment to Human (EH), Human to Environment to Human (HEH) and Vector Transmission (VT). The processes that are included in these transmission routes are presented in table 2.1.

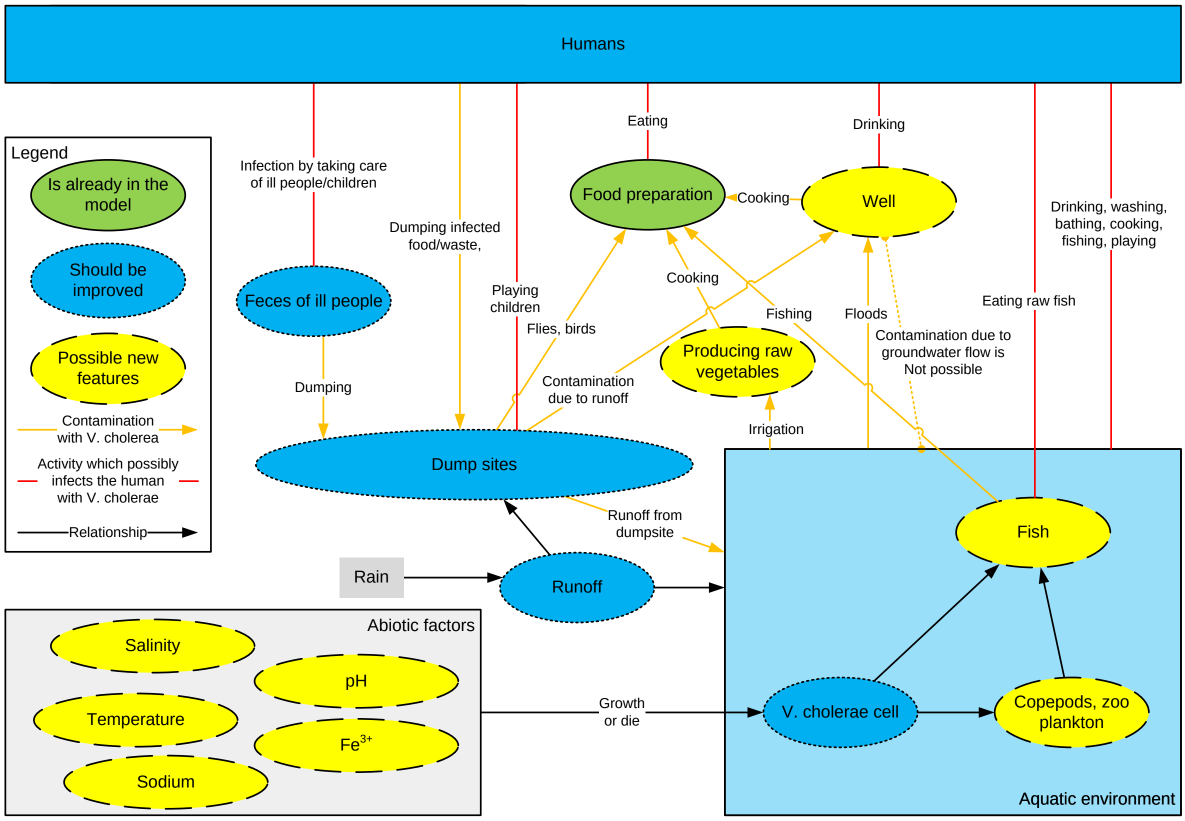


Figure 2.2: *Relation diagram of the cholera spreading mechanisms. The diagram distinguishes processes that are already in the model of Useya (2011), processes that are in the model but that should be improved to come closer to findings in literature and processes that could be possibly add to the model because they play a significant role in the spread of cholera. Chapter 3 goes more into detail about the original cholera model. The relation diagram is based on: Pollitzer et al. (1959), Felsenfeld (1965), Sanyal et al. (1974), Singleton and Attwell (1982),Xu et al. (1982), Huq et al. (1983), Colwell and Spira (1992), Glass and Black (1992), Kaper et al. (1995), Sack et al. (1998), Reidl and Klose (2002), Mintz et al. (2005), Harris et al. (2005), Vezzulli et al. (2010), Osei (2010), Igbinosa et al. (2011), Ferguson et al. (2012) and Adubofour et al. (2013).*

Table 2.1: *Processes that are included in the four transmission routes of the cholera model.*

Route	Definition
HH	**Human to Human transmission:** Individuals get infected due to taking care of a family member.
EH	**Environment to Human transmission:** Individuals get infected due to the ingestion of natural existing V. cholerae fetched from the aquatic environment.
HEH	**Human to Environment to Human transmission:** Individuals get infected due to V. cholerae that is spread by individuals due to dumping waste at dump sites. As a consequence people can get infected due to playing at the dump site or consuming fetched water that is infected with V. cholerae from the dump site.
VT	**Vector transmission:** Individuals get infected due to the spreading of V. cholerae by flies.

Chapter 3: Model Description

Now it is clear what cholera is and how the spreading mechanisms work. This information will be used to define the problems and challenges of the original model in section 3.2. But first section 3.1 describes the original model developed by Useya (2011).

3.1 Existing model

Useya (2011) developed an Agent Based Model (ABM) to simulate the diffusion of cholera. This was done within the software package of Netlogo. Appendix B provides a description about ABM in general and Netlogo. Performing a model simulation consists of two steps: the set-up procedure and the running procedure. These steps are briefly described in this section, for more details is referred to Useya (2011). Furthermore, figure A.1 presents a scheme of Useya's model.

3.1.1 Set-up procedure

The set-up procedures loads the following three types of data:

- *Input variables defined by the user*:
  - Put the different transmission routes on/off.
  - Define the probabilities of HH, children getting infected by playing at the dumpsite and food getting infected by flies.
  - Define the number of ticks (arbitrary time step in Netlogo) that equals a day.
- *Digital geographic layers*: Elevation (using a Digital Elevation Model (DEM)), the house layer and the dump sites locations.
- *Demographic input variables*: the model loads the inhabitants of a predefined synthetic population, who have the following characteristics:
  - Income, which is defined as low, middle or high.
  - Blood type, defined as "O" or "Other"
  - Tap water, defined as having tap water or not
  - Hygiene level, defined as low, middle and high
  - Soiled hands, all agents starts with clean hands
  - Home location, all the individuals are allocated to a house which are defined in the house layer.
  - Age
  - Gender
  - Family.

3.1.2 Running procedure

The cholera model consists of three parts: the activity model, disease model and the hydrological procedure.

The activity model consists of the daily activities of individuals in the model. These activities are: fetching water, dumping waste, children play at the dump site, cooking food, eating food, drinking-water, see figure A.1. The figure shows that these activities are scheduled on fixed time steps.

The disease model consists of the transmission procedures: EH, HEH, HH and VT. When one of these transmission procedures is activated the individual gets an infection. Then the status of an individual is set to “infected” and the time of recovery is determined, this will be between 14 and 16 days. Only during this period HH transmission is possible.

The hydrological model procedure lets it rain every two days. On these days the procedure spread 200 raindrops randomly over the study area during the third time step of a day. Then it allows the water to flow during two days, flowing means that the raindrops move to the lowest neighbouring cell. After these two days raindrops that are not contaminated are removed from the model run to reduce the calculation time.

3.2 Problems and challenges of the present model

This section enumerates the drawbacks of some modelling choices, first for the hydrological procedure and second for the cholera model.

Hydrological procedure

- The existing model does not simulate realistic velocities of the raindrops. On average the velocity of the raindrops is $3.4 \cdot 10^{-4}$ m/s. However a more realistic velocity will be between 0.1 - 1.5 m/s Soong et al. (2012). The consequence is that the V. cholerae cells that are mixed in the water from the dump site have a longer travelling time than they will have in reality.
- The velocity of the raindrops are equal for every cell, which is not expected from physics.
- There is a point in the model where the infected raindrops accumulate and do not flow further downstream. Therefore the individuals that fetch water downstream have no chance to fetch contaminated water and those that fetch at the particular locations have a higher chance.
- A visual comparison between the model and satellite photos from Google Earth shows that the river in the east side of the model cannot be recognized as a river on the satellite photos. The reason for this is probably that the simulated map of 670 by 740 meters is rather small. The consequence for the model is that people fetch water on places where in reality no water flows.
- The input rain data are not based on realistic rainfall. Because on all days with rainfall the rainfall varies between 75 and 85 mm, while the total amount of rainfall for a year is equal to 1400 mm (Ghana Meteorological Services Department, 2012).
- Upstream contamination of the river is not taken into account, although the locations of the dump sites are known.

Cholera model

- The model predicts around 150 cases of cholera for the outbreak in 2005. Based on the data of Osei (2010) just one case was expected. In epidemic research the number of cases is uncertain (Mari et al., 2011; Alam et al., 2007), therefore it is more relevant to simulate the geographical distribution of the cases then reproducing the exact number of cases.
- The model predicted around 20 infections via HH, which is 13% of the total number of simulated cases while literature states that HH is rather unlikely (Glass and Black, 1992).
- The house layer used within the model allocates all the houses copied from a satellite photo of Google Earth, which is not very accurate. This layer is also used within the hydrological model, here the raindrops flow around the houses. This pretends a precision that could not be achieved with this hydrological procedure.
- Based on expert knowledge the individuals are allocated to the houses on the house layer. However, no information was available about the specific household characteristics as for example household size. Therefore this layer contains an error that will influence the outcomes.

The enumeration contains the problems and challenges for the hydrological procedure and the cholera model. The mentioned problems and challenges will be further developed in respectively chapter 5 and 6.

Chapter 4: Study Area

Useya (2011) investigated a relatively small area, although data for a larger study area were available. This chapter contains a brief description of the expanded study area and the data that are available.

4.1 Kumasi

The study area is located in the North-East part of Kumasi, the second largest city of Ghana (Campion and Venzke, 2013). The study area is shown in figure 4.1. It is approximately 40 times bigger than the study area used in Useya (2011) and is now 19.2 km^2, the reason for this will be elucidated in chapter 5. The figure shows the boundaries of the communities and their ID. The boundaries were unknown, therefore they are determined with Thiessen polygons. Table 4.1 provides some general characteristics of Kumasi and the study area.

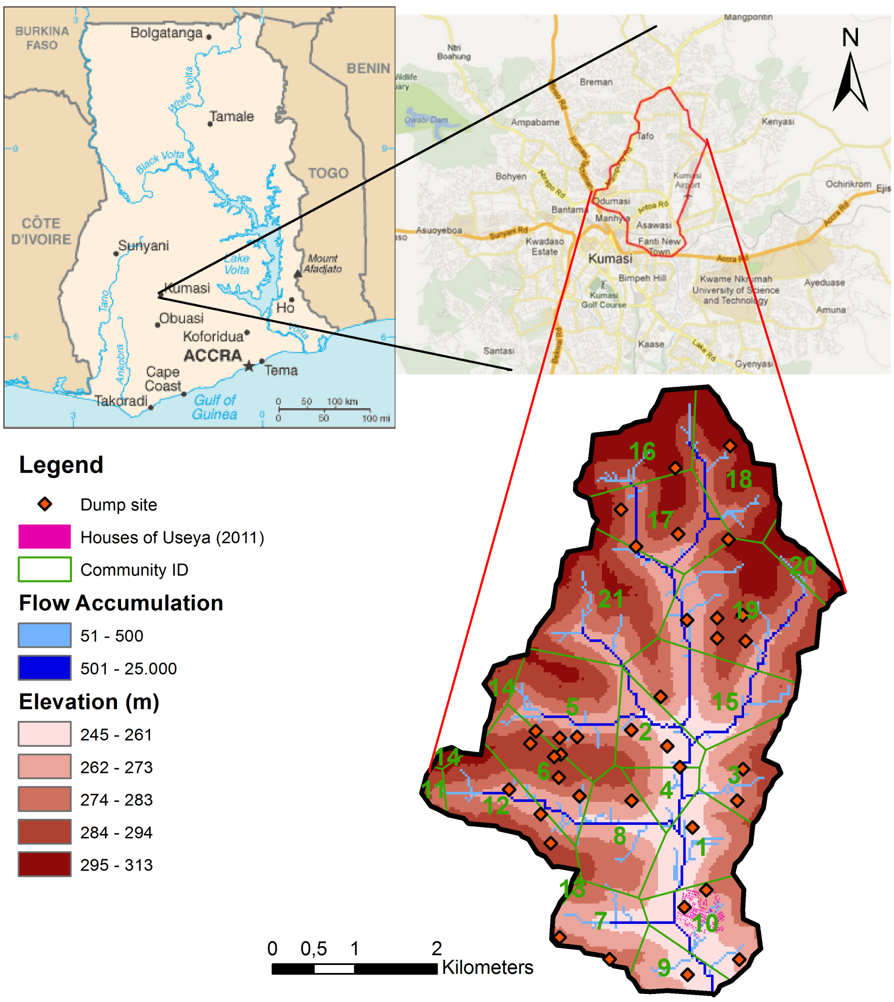


Figure 4.1: *Overview of study area, including: DEM, location of dump sites, community ID's and boundaries (Osei, 2010; U.S. Department of the Interior, 2012; Google, 2013)*

Table 4.1: *General characteristics of the Kumasi area (Osei, 2010; Kumasi Metropolitan Assembly, 2010; GSS, 2012; Ministry of Food & Agriculture, 2011; U.S. Department of the Interior, 2012)*

Characteristic	Kumasi	Study area
Coordinates	6.69^{o}N and 1.62^{o}W	6.72^{o}N and 1.60^{o}W
Inhabitants	2.0 million	± 60.000
Households	512.767	± 15.000
Householdsize	3.8 persons	3.9 persons
Drinkwater	86% of the households has access to safe drinking water	
Temperature	Temperature ranges by average from 21.5°C to 30.7°C	
Precipitation	Yearly precipitation on average 625mm	
Elevation	245 - 320m	245 - 313m

4.2 Available data

The cholera model needs a DEM and rainfall data. The best available DEM for the study area has a resolution of 90 by 90 meters (figure 4.1). Tutiempo Network SL (2005) provides daily rainfall data for the study area during the recorded cholera epidemic from 1 September till 30 November 2005 (figure 5.3).

For this research the cholera data collected by Osei (2010) for a cholera epidemic in 2005 are available. Table 4.2 presents the number of inhabitants and cases for the communities that are used within this research, for an explanation see section 6.4.1. It is remarkable that the number of inhabitants of communities 17 and 21 are exactly the same. This may indicate that the communities are in fact the same. It is widely accepted that in general the number of cases is underreported for cholera epidemics (Lee, 2001). Therefore the objective will not be to reproduce the exact number of cases but the geographical distribution between the communities. The locations of the dump sites, are determined by Osei (2010) using a Global Positioning System (GPS) and shown in figure 4.1.

Finally the cholera model needs demographical data as input. The original model of Useya (2011) used data from the year 2000 collected by Ghana Statistical Service (2008). For this research also the more recent data from 2010 will be used (GSS, 2012).

Table 4.2: *Number of people, cholera cases and the moment of the first and last cholera case per community for the 2005 epidemic, figure 4.1 presents the locations of the communities.*

ID	Name	Population	#Cases	First case	Last case
1	ASAWASE	46243	42	10/15/2005	12/4/2005
2	YENYAWOSO	7254	11	-	-
4	MENHYIA_EXT	21281	1	11/8/2005	11/8/2005
5	MBROM	3337	10	10/14/2005	11/9/2005
6	KROFROM	6373	8	10/13/2005	11/9/2005
8	DICHEMSO	21281	10	10/14/2005	11/29/2005
12	ASH_TOWN	24458	30	10/12/2005	12/4/2005
15	SEPE_APRAMPRAM	8375	3	-	-
17	OLD_TAFO	56417	72	10/12/2005	12/9/2005
19	BUOKROM	12374	11	10/16/2005	11/30/2005
21	EAST_OLD_TAFO	56417	23	-	-

Chapter 5: Hydrological procedure

In chapter 3 the main challenges for the hydrological procedure are defined. It follows that the main objective for the hydrological procedure is: make the travel time of water through the area more realistic. Therefore the water flow will be based on general physical principles rather than flowing each tick to the lowest neighbouring cell. Since no discharge data are available the travel time will be verified by a generally accepted hydrological model.

This chapter describes the development and analysis of the hydrological procedure. First the requirements for the hydrological procedure are described in section 5.1. Section 5.2 describes the method that will be used to improve the model. Section 5.3 presents the conceptual model followed by the implementation in Netlogo in section 5.4. Then the calibration is performed in section 5.5. Section 5.6 presents the verification results and finally section 5.7 presents a brief discussion.

5.1 Features for the hydrological procedure

For the development of the original hydrological procedure some requirements were formulated, which are still important (Useya, 2011):

- The model should describe the transport of V. cholerae cells in water through the study area.
- For convenience the hydrological procedure should be incorporated in Netlogo.

To achieve the objective the hydrological procedure should contain the following features:

- The travel time of the water should be more realistic.
- The fetching water procedure should be able to make a distinction between EH and HEH contaminated water.
- The water in the study may not accumulate on the way to the outlet of the study area as it did in the original model.
- In the original model the raindrops flow around the houses, however this pretends a precision that could not be achieved with this hydrological procedure. Therefore within the hydrological procedure the houses will not be taken into account.
- The study area is enlarged to the entire catchment of the river. This makes it possible to validate the cholera model with more data (spatial distributed cholera cases) and verify the hydrological procedure with the Soil Conservation Service (SCS) Curve Number (CN) method.

5.2 Approach

Many hydrological models have been developed, e.g. HBV, GR4J , WetSpa, TOPMODEL, and VIC. These models have shown good performance under a wide variety of conditions (Lindström et al., 1997; Kobold and Brilly, 2006; Perrin et al., 2003; Pagano et al., 2011; Wang et al., 1996;

Bahremand et al., 2006; Romanowicz, 1993; Cameron et al., 2000; Liang and Lettenmaier, 1994; Andreadis et al., 2009). However, these models require many types of input data which are not available for this study area. It are lumped models that produce results on some predefined points within a study area, but do not contain particle tracking information. For this research the routing of the water is important, this determines whether the water at a fetching point is contaminated by a dump site or not, when households fetch water. In conclusion, this type of model cannot be used for this research because of the lack of available data and the purpose of the hydrological procedure.

The most important requirement for the hydrological procedure is that the timing of the water and therefore the calculated velocity should become more realistic. Because of the lack of data, a simple method based on the slope should be used to determine the velocity for sheet flow, gully and river flow. The most general methods to determine flow velocities for surface flow are the equations of Manning and Chézy. For this research the equation of Manning is used.

The only way to calibrate the hydrological procedure is by comparing the simulated discharges with the discharges calculated by another model, since no discharge data are available. The requirements for this model are that: it is generally accepted, produces good results and the needed input data are available. The SCS CN method meets these requirements. This method is developed for drainage basins where no runoff has been measured (Boonstra, 1994) and is used to calculate storm runoff, peak rate of discharge, hydrographs and storage volumes (SCS, 1986). Section 5.5 provides a more elaborate explanation of the method.

5.3 Conceptual hydrological procedure

This section contains an explanation of the conceptual hydrological procedure presented in figure 5.1. The procedure will be agent based and modelled in Netlogo, therefore appendix B contains a brief general introduction to ABM's and Netlogo. The preparation of the input data and the parameter ranges are described in the next section.

The procedure consists of 2 phases. The first phase, procedures 1 to 5, loads the study area and the model variables into the Netlogo environment. The second phase, procedures 6 to 15, simulates the raindrops flowing to the outlet of the study area.

1. The DEM is loaded.
2. The flow direction layer of the study area is loaded.
3. The flow accumulation (number of upstream cells) layer is loaded.
4. The rainfall data are loaded.
5. The location of the dump sites are loaded.
6. Since it is not realistic that it will rain constantly through the day, the rain is given a limited duration in hours (D), that will be distributed uniformly in time and space and will start at the beginning of each day in which rain falls. The volume of each raindrop is calculated with equation 5.1, where A = area [m^2], P = precipitation [m].

$$\text{Volume of raindrop} = \frac{A \cdot P}{D \cdot \#\text{Raindrops}} \tag{5.1}$$

7. Some rainwater will infiltrate into the soil, the amount depends on the intensity of the rain, the soil type, degree of saturation and the ratio of unpaved/paved area. This information is not available for the study area, therefore it is assumed that a certain percentage of the rain will infiltrate into the soil (I) until a maximum amount of infiltration (I_{max}) is achieved.

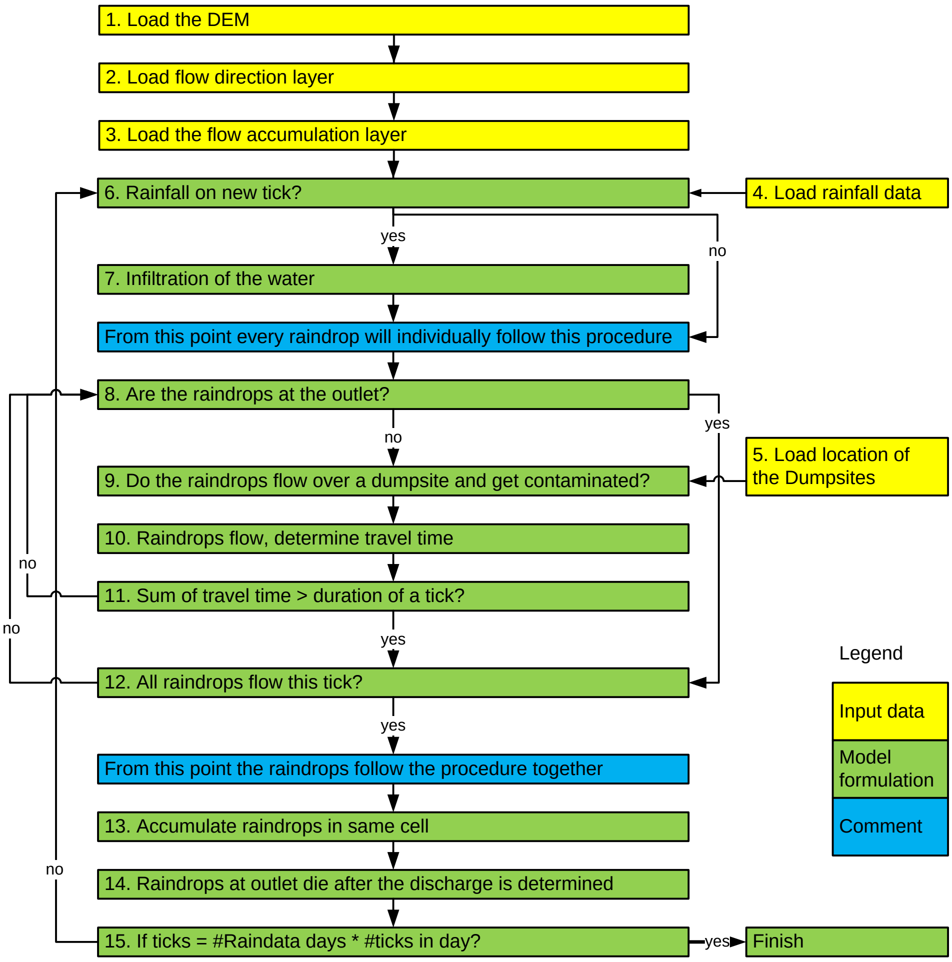


Figure 5.1: *Conceptual hydrological model*

8. When a raindrop is at the outlet of the catchment the raindrop stops flowing, then the next raindrop follows the scheme starting at point 6.
9. When a raindrop is located on a dump site and the dump site infection level (D_i) is above a certain threshold value (D_{imax}), the raindrop gets contaminated.
10. The flow direction (2) determines to which of the 8 neighbouring cells the raindrop flows. It uses the flow accumulation file to determine the type of flow: sheet, gully or river flow. The travel time of the raindrop for the present patch[1] is calculated with the general Manning formulas (SCS, 1986). This results in the following statements:

    - **Sheet flow**: for patches with a flow accumulation $< S_{sg}$ water runoff flows as a sheet and the travel time is calculated with equation 5.2, where m_{sh} = Manning coefficient for sheet flow $[\frac{\text{hours}^{1.25}}{\text{m}^{0.375}}]$, F_L = flow length [m], P = the rainfall [m] and S = slope [m/m]. The formula is only validated for sheet flow shorter than 90 meters (SCS, 1986) and transformed to SI units in appendix C.2.

$$\text{Travel time} = 0.007 \cdot 12^{0.8} \cdot 3600 \cdot \frac{(m_{sh} \cdot F_L)^{0.8}}{(P)^{0.5} S^{0.4}} \tag{5.2}$$

[1] A patch is a grid cell within Netlogo, when referring to ArcGIS still the term grid cell will be used

- **Gully flow**: for patches with a flow accumulation $\geq S_{sg}$ and $< S_{gr}$ small gullies develop and flow will be described with the manning equation for channel flow, see equation 5.3. Where F_L = flow length [m], R = hydraulic radius $R = \frac{\text{Cross sectional flow area}}{\text{Wetted perimeter}}$ [m], S = slope [m/m] and m_c = Manning coefficient for open river channel flow [$m^{\frac{1}{3}}$·hours] (Shaw et al., 2011).

$$\text{Travel time} = \frac{F_L}{\text{Velocity}} = \frac{F_L}{\frac{R^{\frac{2}{3}} \cdot S^{\frac{1}{2}}}{m_c}} \tag{5.3}$$

- **River flow**: for patches with a flow accumulation $\geq S_{gr}$ the travel time is also calculated with equation 5.3, but m_c and R will be different.

R, m_c, S_{sg} and S_{gr} will be determined during the calibration process. The slope of a patch is calculated with equation 5.4, the target patch is determined with the flow direction layer. Some patches have a slope that is equal to 0, to prevent accumulation of the water the slope is in this case set to a minimum slope (S_{min}), this value is determined during the calibration process.

$$S = \frac{\text{Elevation of patch - Elevation of target patch}}{F_L} \tag{5.4}$$

11. A raindrop flows till it travels longer than the duration of a tick.
12. The next raindrop will flow. After executing all raindrops at one tick the procedure continues with the accumulation procedure, step 13.
13. The accumulation procedure is developed to speed up the calculation time. Only patches with a flow accumulation >50 follow the procedure, because this limits the number of patches that is taken into account furthermore the probability that those patches have more than one raindrops is larger than patches with a lower flow accumulation.

    The procedure itself is presented in figure 5.2. The procedure substitutes all raindrops on a patch that have the same amount of time left by one raindrop. Each raindrop has some travel time left because the travel time should be longer than the duration of a tick. Therefore each patch substitutes all raindrops that contain a travel time left between 0 en 60 seconds with one raindrop. This procedures is then repeated for the next 60 seconds, until all raindrops are substituted, this will be after six times. The new raindrop gets the characteristics presented at the ** in the blue box.

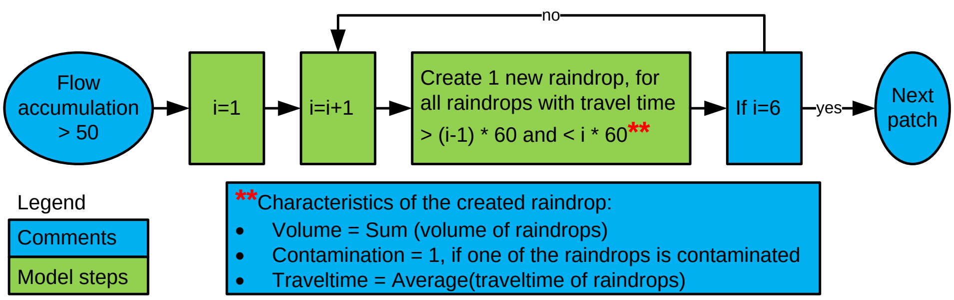


Figure 5.2: *Accumulation procedure*

14. To determine the discharge [m^3/s] at the outlet of the catchment equation 5.5 is used. When the raindrops pass the outflow point they will be removed from the model.

$$\text{Discharge} = \frac{\sum(\text{Volume of raindrops at the outlet})}{\text{Duration of a tick}} \tag{5.5}$$

15. Data are available of an outbreak starting in September until October 2005 in Kumasi Osei (2010). In general the duration of a cholera outbreak on a local scale takes 12-24 weeks (Agheksanterian and Gobbert, 2007). Therefore the simulated period will be from 1 September till 30 November.

5.4 Implementation

The conceptual model is implemented in Netlogo. Due to the limitations of the programme and available data some modelling choices had to be made. First the choices for the input data are elucidated, second the model parameters.

5.4.1 Data preparation

The flow direction and accumulation layer are produced with ArcGIS (Environmental Systems Research Institute, 2012). To produce them, first the gaps in the DEM are filled, using the fill procedure. The raster of 90 by 90 meters is rather coarse, therefore the inverse distance weighting method is used to redefine the grid cells to 30 by 30 meters, followed again by the fill procedure. Then the flow direction procedure defines for each cell to which neighbouring cell the water will flow. Followed by the accumulation procedure that creates a raster of accumulated flow to each grid cell. Based on this raster the outflow point of the area is chosen to determine the size of the catchment area using the Watershed procedure. Finally all layers are resized to the area of the catchment using the clipping procedure. The result is shown in figure 4.1.

The precipitation data are taken from Tutiempo Network SL (2005), which gave the data per day. Since it is not expected that it rains the whole day an assumption for D is made. There are no hourly rainfall characteristics for Kumasi available, only studies based on daily or monthly rainfall data (e.g. Barbé et al., 2002; Laux et al., 2009; Campion and Venzke, 2013). Therefore D will be set to 2 hours at the beginning of the day (0:00-2:00AM) independent of the amount, the influence of this assumption will be analysed in section 5.5.4.

5.4.2 Model parameters

Because there is no information about the hydrological characteristics of the area, the values of the parameters incorporated in the model will be determined during the calibration process. The range of the parameter values will be based on general knowledge about the area and the parameters are applied uniformly to the study area. The ranges used for the calibration process are presented in table C.2. This section contains the reasoning for the ranges of the model parameters.

The CN method is used to verify the outcomes of the hydrological procedure. The CN method as well as the hydrological procedure do not calculate groundwater flows. This is no problem since groundwater is not relevant for the spreading of cholera (Ferguson et al., 2012; Hunt et al., 2010).

Both infiltration parameters (I and I_{max}) are set to 0, for two reasons. First, the CN method accounts for infiltration, but the CN for this study area is estimated to be 85 (see section 5.5.2), this means that infiltration is negligible. Second, people in Kumasi do not collect rainwater (Whittington et al., 1993; GSS, 2012), this means that the infiltration factor and the CN do not have to be compensate for this behaviour.

It is assumed that the flow length for sheet flow is equal to the length of a patch. For the gully and river flow the flow length should be at least equal to the length of one patch. The maximum flow length is set equal to two times the length of a patch, because in practice it is possible that the stream meanders over the patch.

Also the hydraulic radius is taken into the calibration, because no cross sections of the river are available. Measuring the width of the river in Google Earth learns us that the width of the river is about 2-3 meters. However the depth cannot be determined. The width for a gully is also measured and is about 1 meter, but the depth is again unknown. In general the upstream cross sectional area is smaller than the downstream cross sectional area, because the river carries more water downstream. Therefore the hydraulic radius downstream will be bigger than upstream for an equal slope.

An order of magnitude for the Manning coefficient of the sheet flow is taken from SCS (1986), the order of magnitude for the Manning coefficient of the gully and river is based on Chow (1959), Arcement and Schneider (1989), U.S. Department of Tranportation (2011) and Soong et al. (2012). It is hard to define what the characteristics of the gully and the river are these values are also determined during the calibration process.

To make the calibration process less time consuming the Manning coefficient and the hydraulic radius are combined in one variable 'ManningHR', that represents both variables. This is possible since both variables are unknown and used within the same formula. Equation 5.6 presents the use of ManningHR. The minimum value is determined using the minimum Manning coefficient and the maximum hydraulic radius, because this results in the largest velocities and vice versa.

$$\text{Velocity} = \frac{R^{2/3} \cdot S^{1/2}}{m_c} = \text{ManningHR} \cdot S \tag{5.6}$$

Finally, the ranges for S_{sg} and S_{gr} are determined based on the flow accumulation layer compared with the observations from Google Earth.

5.5 Calibration of the hydrological procedure

This section describes the calibration of the hydrological procedure. The first section describes the procedure, the second the CN method, the third presents the results and the fourth discusses the robustness of the model.

5.5.1 Calibration procedure

As an objective function to compare the discharge simulated by the hydrological procedure, with the results of the CN method, the Nash-Sutcliffe coefficient (NS) is used (equation 5.7) (Nash and Sutcliffe, 1970). To check if the water balance is correct the Relative Volume Error (RVE) is calculated with equation 5.8. Because there is no infiltration in the hydrological procedure and the CN method, the water balance can also be checked by comparing Q_{obs} with Q_{sim}.

$$\text{NS} = \frac{\sum_{t=1}^{T} (Q_{obs} - Q_{sim})^2}{\sum_{t=1}^{T} (Q_{obs} - \overline{Q_{obs}})^2} \tag{5.7}$$

$$\text{RVE} = 100 \cdot \frac{\sum (Q_{sim} - Q_{obs})}{\sum Q_{obs}} \tag{5.8}$$

The hydrological model is calibrated using a semi-manual calibration. First, 100 input variables per parameter were randomly and uniformly generated within the ranges of the parameters.

Then the procedure was run for the predefined period, using 5.000 raindrops. Afterwards these results were analysed by plotting the NS against the parameter range and narrowing the range to the best scores for NS. This procedure was repeated until the model shows no further improvement.

5.5.2 Curve number method

The CN method is developed by the US Soil Conservation Service for conditions prevailing in the United States. Since then it has been adapted to conditions in other parts of the world. Some regional research centres have developed additional criteria. However, the basic concept is still widely used all over the world (Boonstra, 1994; Arnold et al., 1998). An explanation of the method is provided in appendix C.

To apply this method the curve number that represents the study area needs to be determined. The soil texture in the study area is sandy clay loam (Agodzo and Adama, 2004), this soil can be classified within the CN method as hydrologic soil group "C". Based on Google Earth and ArcGIS, the study area consists approximately for 30% out of grassland with a good grass cover which has a curve number of 74. The other 70% of the area consists of houses and has a curve number of 90. Thus the weighted curve number for the study area is 85 (SCS, 1986).

5.5.3 Calibration results

Figure 5.3 presents the final hydrograph, which has a NS of 0.95 and a RVE of 0.20%. This result shows that the newly developed hydrological procedure simulates the calculated discharge with the CN method very good. The simulated peak discharges are a bit overestimated, the timing of the peaks and the start of the discharge wave match very well. The backside of the CN modelled discharge ends abruptly, but the modelled hydrograph shows a curvature which may also be expected in reality. The small value of RVE is probably caused by rounding off errors.

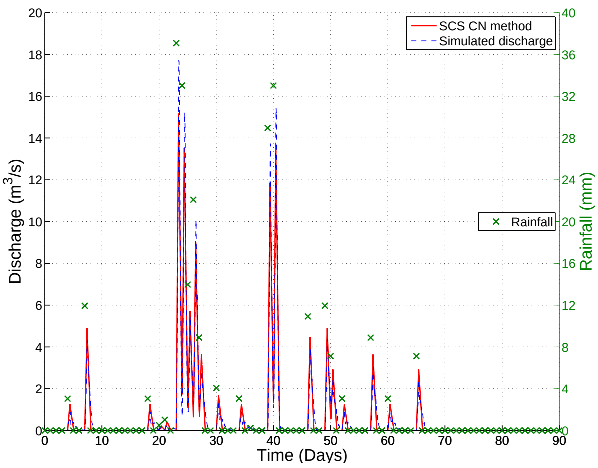


Figure 5.3: *Simulated and CN method hydrograph, resulted in a NS of 0.95. Furthermore the daily rainfall from September to November 2005 for Kumasi, is presented (Tutiempo Network SL, 2005).*

The average velocity for each type of flow is determined and presented in table 5.1. The sheet flow velocity is low, however Riscassi and Schaffranek (2003) found velocities in the same order of magnitude. The velocities for the gully and river flow are in the expected order of magnitude. The corresponding parameter values are presented in table 5.2.

Table 5.1: *Flow characteristics averaged for the different stream profiles*

Type of river	River	Gully	Sheet flow
Velocity (m/s)	0.65	0.35	$3.86 \cdot 10^{-3}$
# Patches	564	7251	13320
Average slope (m/m)	$4.5 \cdot 10^{-1}$	$4.0 \cdot 10^{-2}$	$4.5 \cdot 10^{-2}$

Table 5.2: *Parameter values for the hydrological model after the calibration*

Parameter	**Value**	**Dimension**
F_L	40.7	m
m_{sh}	$6.93 \cdot 10^{-}2$	$\text{hours}^{1.25} \cdot \text{m}^{-0.375}$
S_{sg}	5	#cells
S_{gr}	506	#cells
ManningHRgully	1.83	m/hour
ManningHRriver	6.37	m/hour
S_{min}	$2.3 \cdot 10^{-3}$	m/m

5.5.4 Robustness of the model

The calibration results are promising. However, to be useful the model should also be robust. This means that the model should predict the same outcome when it runs with the same parameter values. When the model is robust, it enables the user to investigate the influence of for example different duration of rainfall on the spreading of cholera.

During the calibration 5.000 raindrops were simulated, however for speeding up the model this number have to be lowered. It was found that the result remains well when the number of raindrops is reduced, but for less than 500 raindrops the NS dropped. The next test was to judge the stability of the model. For this purpose the model run 400 times with 500 and 5000 raindrops, for each run the NS was determined. The final test was changing the duration of the rainfall. All tests show good results and are presented in table 5.3, the hydrograph shows some small differences due to the probabilistic character of the rainfall.

Table 5.3: *Results of the robustness analysis*

#Raindrops	**NS**	**#Raindrops**	**500**	**5000**	**Hours of Raining**	**NS**
250	0.91	#Runs	400	400	1	0.91
500	0.93	Mean NS	0.91	0.93	2	0.95
1000	0.94	Min NS	0.88	0.93	3	0.95
2500	0.95	STDEV NS	$7.2 \cdot 10^{-3}$	$2.1 \cdot 10^{-3}$	4	0.96
5000	0.95					
10000	0.95					

5.6 Validation

The validation of the hydrological model is performed with the rainfall data from 1 September to 30 November of 2006, 2009 and 2010 (Tutiempo Network SL, 2005). The results are good and shown in table 5.4, the hydrographs show the same characteristics as for the calibrated period. The resulting hydrographs are presented appendix C.4.

Table 5.4: *Validation results*

Year	2005	2006	2009	2010
NS	0.95	0.92	0.94	0.94
RVE (%)	0.20	0.11	0.24	-0.18
Total Precipitation (mm)	256	23	186	256

5.7 Discussion

The developed hydrological model has some limitations, which are important to realize when this procedure is used. The aim of the procedure is to calculate realistic flow velocities and not to calculate the water heights. The Manning formula uses the hydraulic radius to determine the velocity. Therefore it is implicitly taken into account, but it is assumed that it will be constant. The consequence of not calculating the water heights is that the river possibly contains no water while it in reality does, see also figure 5.3. A possible drawback of this could be that the V. cholerae coming from the dump sites has a smaller chance to be fetched by the individuals. This will be discussed in chapter 6.

It is important to realize that only the relevant hydrological processes for the spread of cholera are taken into account and the model simulates only the direct runoff, so ground water flows are not taken into account. One aspect that might be important but which is not taken into account is evaporation. There were two reasons to do so. First, this type of data was not available. Second, the model shows that the rainfall is within one day at the outlet of the study area, so the water has a limited time to evaporate which results in a small amount of evaporation. However, when the hydrological procedure will be developed in the future and for example the water height will be incorporated, the importance of the evaporation will increase since the water will stay in the study area for a longer period of time.

Finally the lack of discharge data makes the calibration results disputable, because the simulated discharges are compared with estimated data during the calibration and validation process. The daily rainfall data may be too coarse for a good prediction of the discharge. However, the stability test shows that the duration of the rainfall did not have major consequences for the performance of the hydrological procedure.

Chapter 6: Cholera model

The hydrological procedure has been improved and the challenges for the original model are known. In this chapter the hydrological procedure will be implemented and challenges will be translated to features of the cholera model. Section 6.1 describes the features for the adjusted cholera model. Section 6.2 describes the conceptual cholera model, followed by the description of the input data in section 6.3. Section 6.4 presents the calibration and section 6.5 describes the limitations of the improved cholera model.

6.1 Features for the cholera model

In order to achieve the objective to simulate the spread of cholera for the catchment area, the following features have to be included in the cholera model:

- The study area will be expanded to the size of the catchment of the river. Because it is a logical scale for the hydrological procedure, and the model outcomes can be verified with the data of Osei (2010) that covers a larger area than the original study area.
- For each community the model should compares the contribution to the total number of cases with the collected data by Osei (2010). A relative comparison of the number will be done, since the number of cases of epidemic diseases is always underestimated (Lee, 2001).
- The developed hydrological procedure should be included.
- Some of the original procedures are inconsistent with the report of Useya (2011), therefore these procedures will be improved.
- The VT and HH transmission routes will be kept in the model, although their effect in reality is disputable (Glass and Black, 1992).
- The synthetic population in the model will be allocated based on logical principles instead of getting a permanent house. Because not enough information is available to make a static layer of the synthetic population of the study area.

6.2 Conceptual model

To include the features mentioned in the previous section, the original procedures have to be changed. This section explains the conceptual cholera model presented in figure 6.1 and the way that the procedures are included in the adjusted cholera model. The technical details about the procedures are provided in appendix D.1. Useya (2011) describes the working of the original cholera model's procedures and section 6.3 the development of the input data and parameter values.

The model consists of 2 phases. The first phase, initialisation, loads the study area and the model variables into the Netlogo environment. The second phase simulates the behaviour of the synthetic population.

1. The general and hydrological input variables are loaded (section 5.4.1).

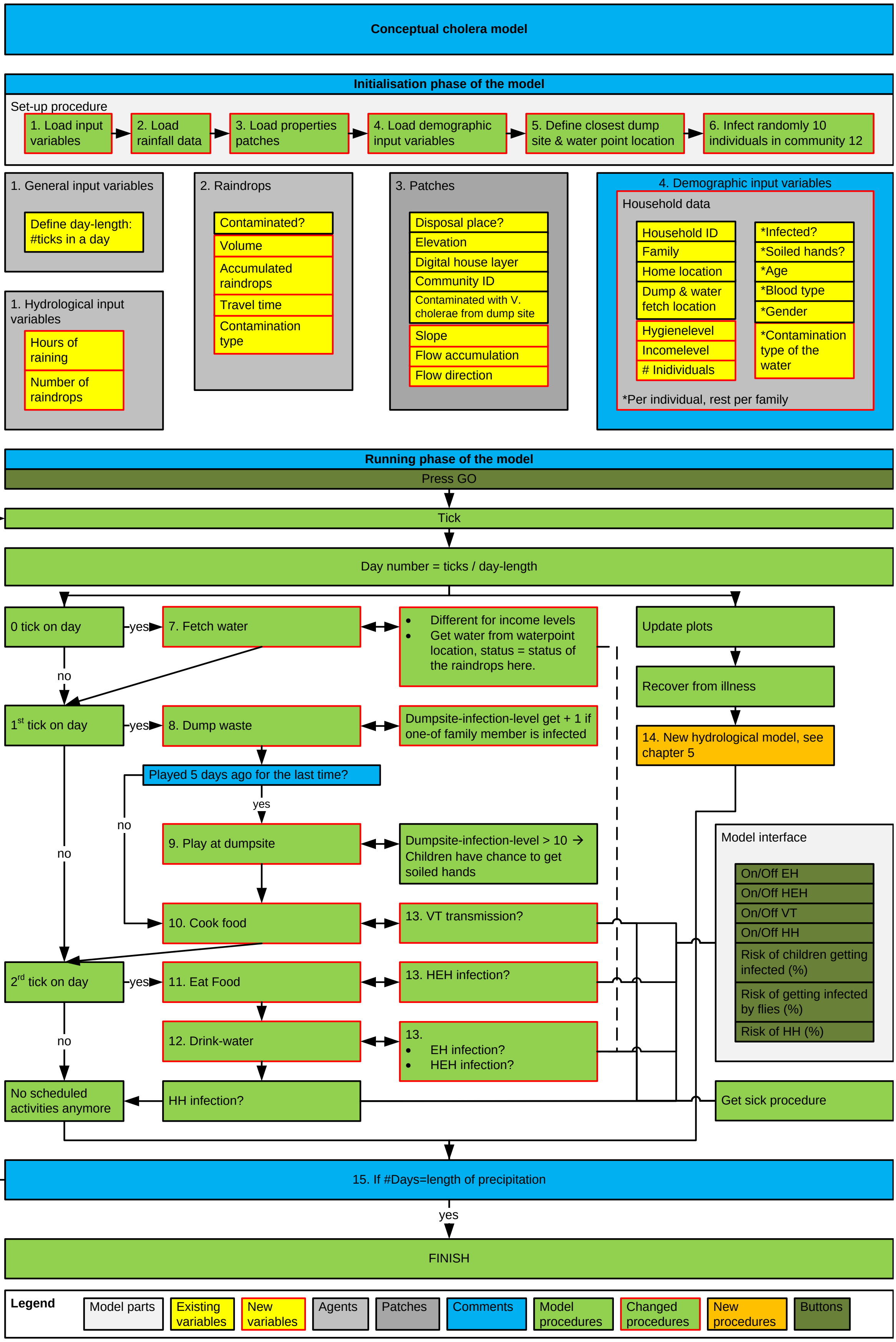


Figure 6.1: *Conceptual cholera model*

2. The rainfall data are loaded and raindrops receive their characteristics (chapter 5).
3. Each patch loads the following information: (i) elevation (from DEM), (ii) flow accumulation, (iii) slope, (iv) flow direction (v) dump sites, (vi) dump site infection level (D_i) (vii) location of the houses and (viii) community ID.
4. The demographic variables are loaded and households are allocated to one of the houses with the same income level. The chance that they are allocated to a particular community depends on the population density of that community. Furthermore, a household receives the individual's characteristics to reduce the calculation time. But the individuals are still treated as if they are single agents in the model.
5. The dump sites are loaded and the water fetch locations are determined, which are every part of the river with a flow accumulation greater than a certain threshold value (T_{wp}). Next, each household determines what the closest dump site and water fetching point is.
6. Ten initial infections are created in community twelve to trigger the cholera model. Specifically in community twelve because the data of Osei (2010) show that the first case was registered in this community.
7. Households that have no access to tap water will fetch or buy water. This depends on their income level. When they buy water no infection will occur otherwise there is a probability of a HEH or EH infection. This procedure accounts for the fact that the tap water will not work on raining days when the rainfall is above a certain threshold value (P_{max}). Therefore there will be a probability for each household, depending on the income level, that they will fetch water on theses days.
8. Households who have infected family members will add one to D_i, when dumping waste.
9. Children from families with a "low" income and hygiene level, play every 5 days at the dump site and have a probability of getting soiled hands (R_c) when D_i is above the maximum dump site infection level (D_{imax}).
10. This procedure determines the chance that there will be flies on the food (P_{ff}) using equation 6.1, where D_f = maximum travel distance of flies from the dump site and D_{hd} = distance of household to the dump site. When there are flies on the food and the D_i is above D_{imax} the food has a chance of R_f to get contaminated with cholera.

$$P_{ff} = \frac{D_f - D_{hd}}{D_f} \tag{6.1}$$

11. Individuals with both blood types ("O" and "other", see section 6.3.2) can get infected with cholera due to eating with soiled hands or contaminated food by flies.
12. The type of water that is fetched determines the source of the cholera infection.
13. The definitions of the transmission routes are defined in table 2.1. An individual can get infected by the ingestion of water or food, the type of transmission depends on the type of contamination of the water or food. The water can contain two types of contamination: EH or HEH, this depends on the water that is fetched. The food can get contaminated during the cooking process via flies or eating with soiled hands. The chance that an individual gets infected with V. cholerae, depends also on the income level, hygiene level and blood type of an individual. An overview of the corresponding chances is provided in table D.1 and D.2 (Useya, 2011).
14. The hydrological model (chapter 5) is implemented in the cholera model. To reduce the calculation time the procedure is changed. Instead of raindrops on the whole study area it will only rain on the dump sites, therefore the number of raindrops can be reduced.

This has two consequences: (i) the discharge of the river cannot be determined in the cholera model. This is no problem because the discharge is not a primary outcome of the cholera model. (ii) There will be less contaminated raindrops in the river. This reduces the probability that households fetch water which is contaminated by a dump site. To compensate for this the patches that a contaminated raindrop passes get a contaminated status for several hours (T_c). Furthermore the raindrops only flow from the dump sites where $D_i > D_{imax}$, because only those raindrops are contaminated.

15. The model will run for the outbreak period set in the rainfall data.

6.3 Input data

This section presents the development of the input data and the ranges for the parameter values. First it describes the development of the digital house layer, second the generation of the synthetic population and third the parameter ranges used in the calibration procedure.

The development of the input files of the hydrological procedure is already explained in section 5.4. The dump site layer is provided as a Shapefile by Osei (2010). This layer is converted with ArcGIS to a grid file with the same resolution as the DEM. The grid cells below a dump site are treated in the model as a dump site.

6.3.1 Digital house layer and income distribution

The cholera model needs the income level and the distribution of the households over the area. The model distinguishes three income levels (low, middle and high) and it allocates households based on their income level. The data of Osei (2010) provides the number of individuals in a community, but no information about the income distribution. Furthermore, the information provided by GSS (2008, 2012) was too general and other sources could not provide detailed information about distribution of the income levels. This scarcity of income data is also shown in researches of Keraita (2003), Akumiah (2007) and Adubofour et al. (2013). They all investigated partly the same study area and needed information about the income distribution, however all researches had to use general assumptions instead of detailed information about the income distribution.

Therefore, a digital house layer is constructed using the topographic layer within ArcGIS. Afterwards, based on this layer and Google Earth the houses were assigned to one of the income levels (figure 6.2). A more detailed description of the assigning process is provided in appendix D.2. Then the number of houses within an income level were counted, resulting in an income distribution of 19% 'low', 52% 'middle' and 29% 'high'. A research of Water & Sanitation for the Urban Poor (2012) showed that 77% of the residents in Kumasi has no access to improved sanitation, this quite comparable to the sum of the low and middle income groups. Therefore this is at least an indication that the distribution of income could be right. Unfortunately there is not more data found to verify this distribution. The result is shown in figure 6.2 and contains 17.000 buildings.

Uncertainties in the geographical layers

It is important to realize that most input was gathered after the cholera epidemic. For example the house layer is based on an ArcGIS base layer of 2012, this may not represent the situation of 2005.

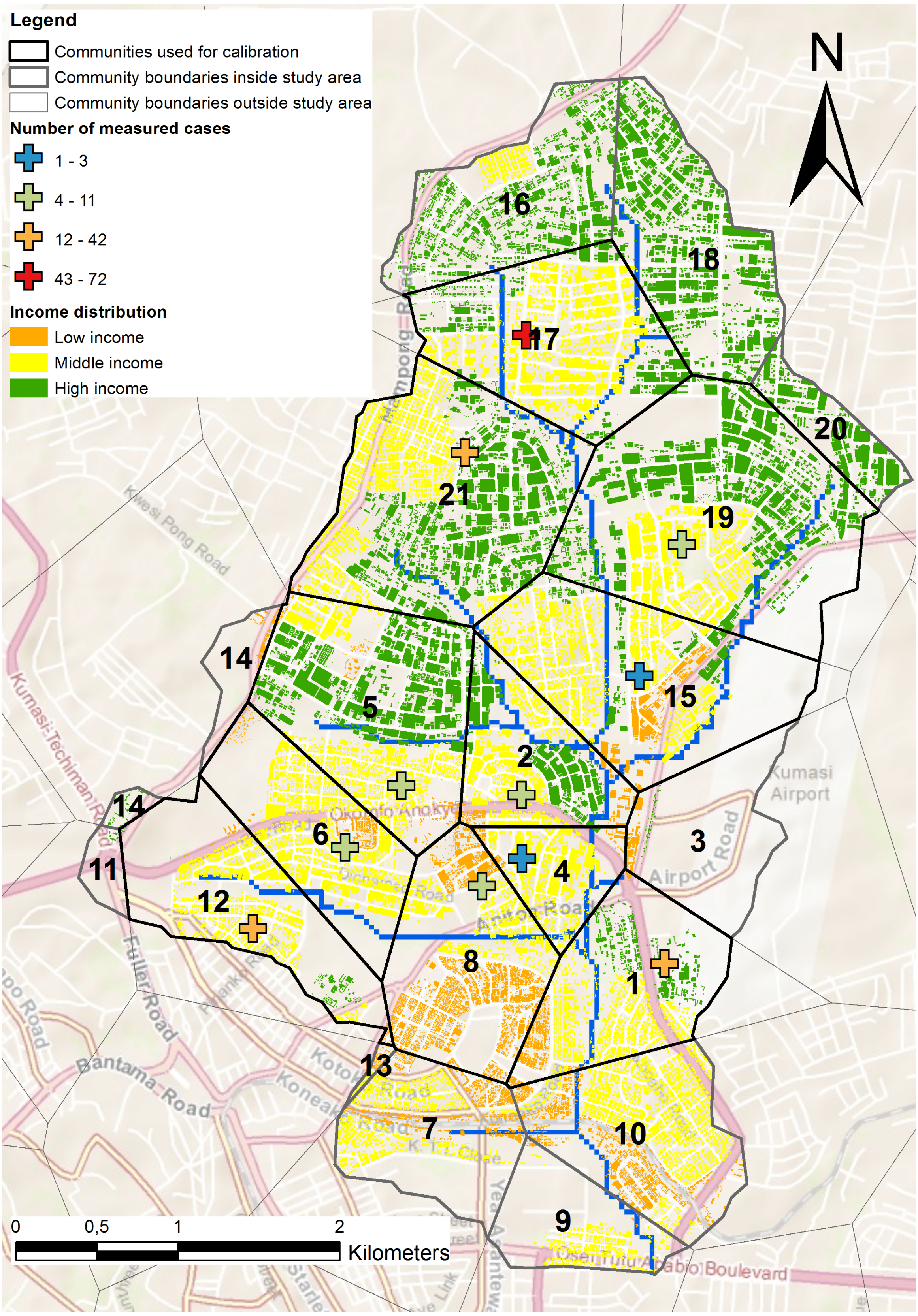


Figure 6.2: *Income areas and boundaries of the communities*

Furthermore, it is unknown whether all the buildings of the house layer are houses, they could also be factories or a block of flats for example. The income level of the houses is based on a visual study of the study area and some general knowledge about the housing conditions in the area. The consequence is a large uncertainty in the house layer.

An important question is: what is the influence of the house layer? For the number of cases per community the precise location is not important, but it is important that the distribution of the households between the different communities represents reality. On the other hand the location of a house determines which fetching point and dump site a household uses, this influences the probability of getting infected. In conclusion it depends on the type of analysis whether this layer will influence the outcomes significantly or not.

6.3.2 Generation of a synthetic population

The cholera model needs a population. Since there is no information available about the location, income distribution and hygiene levels of households, a synthetic population will be generated using the same procedure as Useya (2011) (figure 6.3). To compensate for the uncertainty of this input data, the synthetic population will be generated each model run instead of generating a static population layer. For this reason the population generator is implemented in the cholera model.

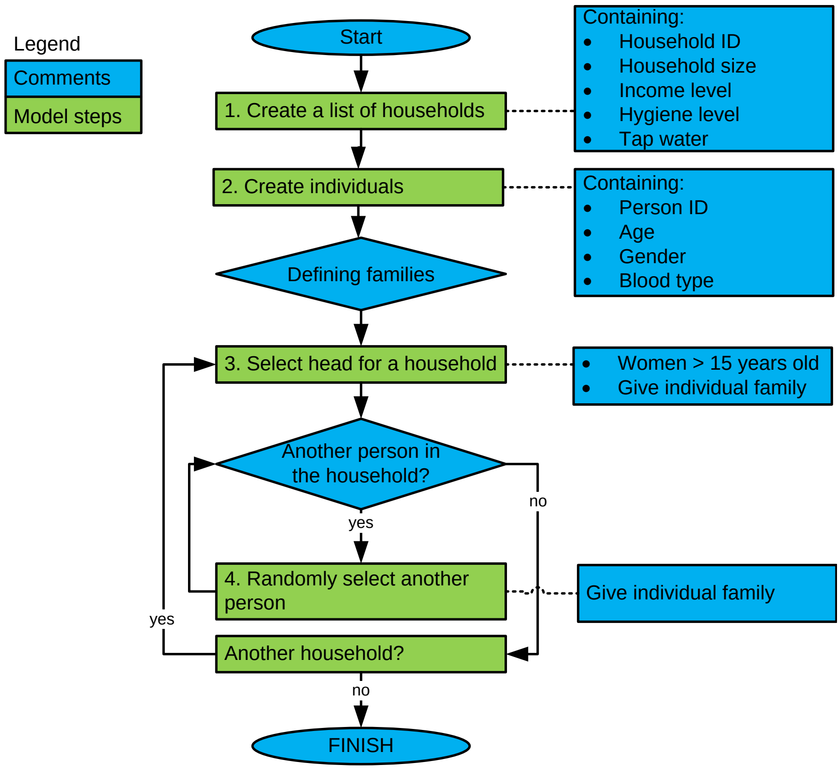


Figure 6.3: *Synthetic population generator after Moeckel (2003)*

1. The user defines the number of households, next each household receives the following characteristics:

    - Household ID.

- Household size: an average household in the study area (Old Tafo, part of Kumasi) exists of 3.9 individuals (GSS, 2012). It is assumed that the household size has a normal distribution with a standard deviation of 2 and an average of 3.9.
- Hygiene level: exists of three groups (low, middle and high). Since there is no data available about the distribution of the hygiene level, the size of the groups will be calibrated, see section 6.3.3.
- Income level: low=19%, middle=52% and high=29% (section 6.3.1).
- Access to tap water: no detailed information about households' source of drinking water in the study area is available. Therefore it is assumed that figure D.2, which presents the source of drinking water for people in Ghana, is representative for the study area. However the cholera model contains two sources of water: (i) tap water, which represents all the sources of safe water in the model and (ii) water from the river, which could possibly be contaminated. For this reason piped water, bore/hole/protected/well/spring water and satchet/bottled water are representing the tap water, the summation of these sources results in a tap water availability of 86% for the households. This percentage has to be divided between the different income levels. Therefore, it is assumed that 100% of the high incomes will have access to tap water, 88% of the middle incomes have tap water and 78% of the low incomes have tap water, on average this results in a tap water availability of 86%. Households that have no access to tap water have a chance to fetch water, see section 6.2 point 7.

2. The summation of the households' size defines the number of individuals, then all individual gets the following characteristics:

   - Personal ID.
   - Age: the distribution is presented in figure D.3 (GSS, 2012). Within each age group the age of the individuals is assumed to be uniformly distributed.
   - Gender: 48.9% men and 51.1% women (GSS, 2012).
   - Blood type: 45% of the people has blood type "O" and 55% "other" (Useya, 2011).

3. The head of the household is assigned to the household ID, namely a women aged from 15 to 60 years (Useya, 2011).
4. The other family members are randomly picked from the list of individuals until the household size is achieved, then this process is repeated for the next household.

Uncertainties of the synthetic population generator

The generation of the synthetic population itself contains many uncertainties. However, this was accepted because it was the only way to create a working model. Bearing this in mind, the generation of the synthetic population contains two assumptions that will contain a large uncertainty: the income distribution and hygiene level. The income distribution is uncertain because it is based on the house layer. The hygiene level is uncertain because there was no information available about the distribution and influence of the hygiene of households on the probability that they get infected with cholera. With a sensitivity analysis the sensitivity of both parameters could be determined, the results of such an analysis will also be an indication of the importance of these parameters.

It has to be noticed that the synthetic generation of the population itself is able to produce an unrealistic family composition, for example 5 women in one family. However this chance is rather small, since the distribution between men and women is nearly equal. Besides there is no data to make the generator more realistic. The influence on the model results will be small because only the transmission mechanisms are taken into account and not a social model for example. Only the age and the type of blood of people influence the chance to get infected. The age determines whether children will play at the dump site or not and the type of blood because individuals with blood type "O" are more susceptible for cholera infection. In reality the type of blood of a child depends on its parents, however in the synthetic population generator this is not taken into account.

6.3.3 Model parameters

The values for the parameters will be determined during the calibration process. The parameter ranges are presented in table D.3. These ranges could not be based on available data. Therefore the ranges of the parameters, including hygiene level, will be based on general knowledge about cholera and the original cholera model of Useya (2011). Appendix D.4 provides a brief reasoning of the parameter ranges and explains the function of the parameter in the cholera model.

The number of households used in the simulations is set to 8.500. In section 6.3.1 17.000 buildings in the study area were determined. Furthermore it is known that there are approximately 263.000 individuals living in the communities that are taken into account (Osei, 2010). This means that there will be about 67.000 households in the area. However, simulating 67.000 households will be too time consuming. Therefore the number of households was tested, it was found that 8.500 households simulated the geographical distribution the same as 17.000 households.

6.4 Calibration of the cholera model

This section describes the calibration of the cholera model. Section 6.4.1 explains the calibration procedure. In section 6.4.2 the robustness of the model is tested. Followed by the results in section 6.4.3.

6.4.1 Calibration procedure

As already mentioned in chapter 4 it is widely accepted that the number of cholera cases is under reported (Morris, 2003; Zuckerman et al., 2007; Rivera et al., 2013). Therefore the aim of the calibration will not be to reproduce the exact number of cases but to simulate the contribution of each community to the total number of measured cases in the communities (GD). GD is calculated with equation 6.2, where $\#C_{com}$ = number of cases in a community and $\#C_{tot}$ = total number of cases in the study area. The measured GDs are then compared to the simulated GDs to determine the accuracy of the model. The accuracy will be expressed as r^2 and is calculated with equation 6.3 (Menard, 2000).

$$GD = \frac{\#C_{com}}{\#C_{tot}} \cdot 100 \tag{6.2}$$

$$r^2 = 1 - \frac{\sum(GD_{\text{mea}} - GD_{\text{sim}})^2}{\sum(GD_{\text{mea}} - \overline{GD_{\text{mea}}})^2} \tag{6.3}$$

The boundaries of the communities are determined with Thiessen polygons. However, not all the communities are entirely situated within the study area. To make a fair comparison possible only the communities situated within the study area are taken into account. Figure 6.2 shows an overview of these areas, and presents the measured cholera cases per community.

The actual calibration is done using a Monte Carlo simulation. The parameter sets were generated by varying for each set every parameter randomly within their range. Then the model simulates for each parameter set the outcomes and the r^2 is calculated. Next the r^2 is plotted against the range of each parameter. Finally the range of the parameters will be narrowed to the range of the best scores for r^2. This procedure will be repeated until the model shows no further improvement.

6.4.2 Robustness

It was found that the model was not really stable. Figure 6.4 shows that the r^2 values vary over a wide range. Here the model is run ten times with the same parameter set. This variance was expected due to the probabilistic character of the model. Besides it was noticed that generating the synthetic population influences the stability of the model runs as well.

However, for the analysis in this research it will be less time consuming when the model outcomes are stable. Therefore the model will be stabilized by averaging the results of several model runs. To determine the number of runs before the model stabilises, the model was run 2700 times with the same parameter set. To reduce the influence of the synthetic population the synthetic population is regenerated every ten runs.

The results are presented in figure 6.5. The graph shows that the range becomes narrower when the number of runs increases. It shows that after 90 model runs the range in r^2 reduces only with small steps, therefore the cholera model will be run 90 times for each parameter set during the calibration process and the analyses afterwards.

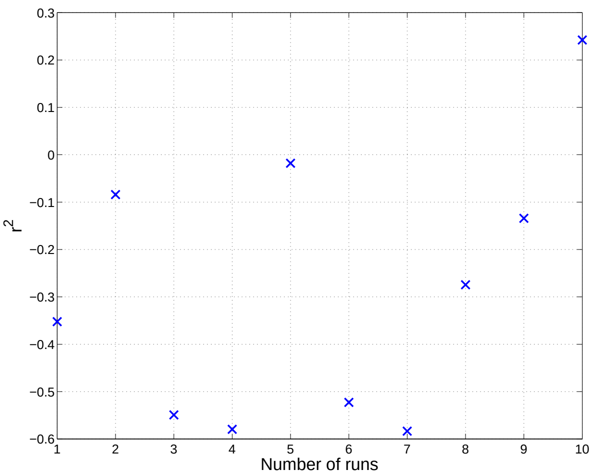


Figure 6.4: *Stability test*

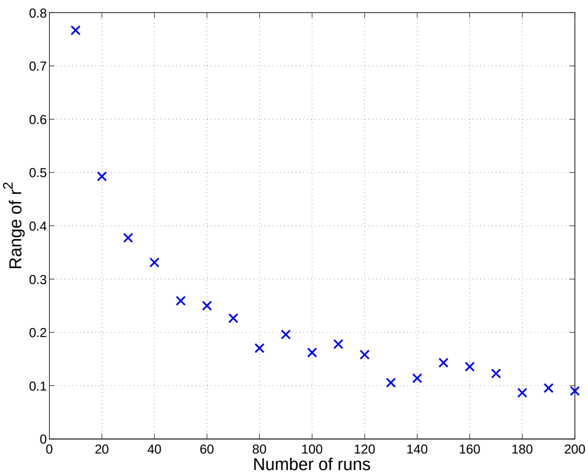


Figure 6.5: *Robustness Analysis*

6.4.3 Results

This section discusses the results of the calibration procedure. First the geographical distribution is presented, second the model parameters, third the epidemic curve and finally the distribution between the different spreading mechanisms.

Figure 6.6 presents the calibration result with a r^2 of 0.87. The figure shows that the cholera model reproduces the geographical distribution well. The simulated contributions for each community are in the same order of magnitude as the measured cases. For community 21, it is slightly overestimated. A possible explanation for this is that community 21 is actually twice as large as the part that is taken into the model simulations (see figure 6.2). The consequence is that the actual density of people in this area will be smaller than the density used in the model.

However with the available data it was not possible to reduce this type of error. In addition it should be emphasized that the boundaries are arbitrarily defined and therefore contain an error.

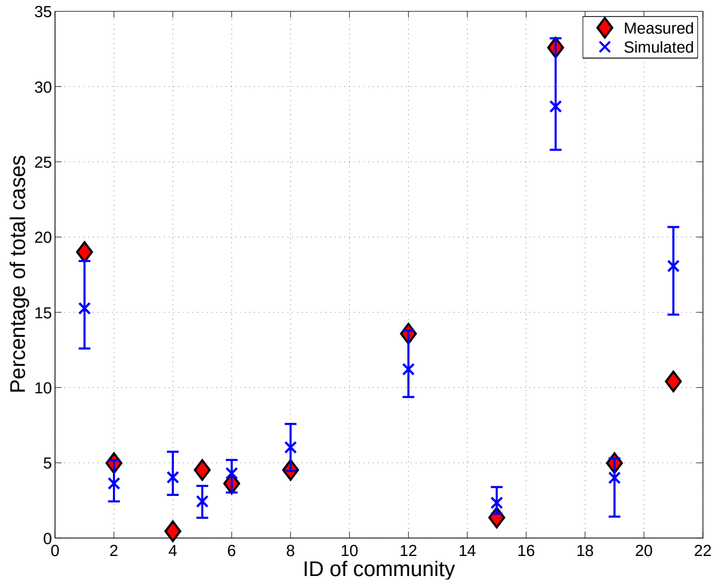


Figure 6.6: *The GD measured and simulated for each community with a r^2 of 0.87. The error bars represents the minimum and maximum GD for each within the 90 runs.*

Table D.4 presents the parameter set that resulted in the best calibration. During the calibration procedure it was noticed that some parameters influence the outcomes more than others. An example is provided in figure 6.7. It can be seen that there was no influence of D_f on r^2, but that T_{wp} shows to be important for the model. This difference might be expected since the number of infections caused by VT is much smaller than the number of HEH transmissions (see table 6.1). Furthermore the graph of T_{wp} shows that there are good results ($r^2 > 0.85$) in a wide range for this parameter, this characteristic was also seen for other model parameters. This may indicate that the model is not only sensitive to the model parameters but that other factors such as the input data may also have a major influence on the outcomes of the model.

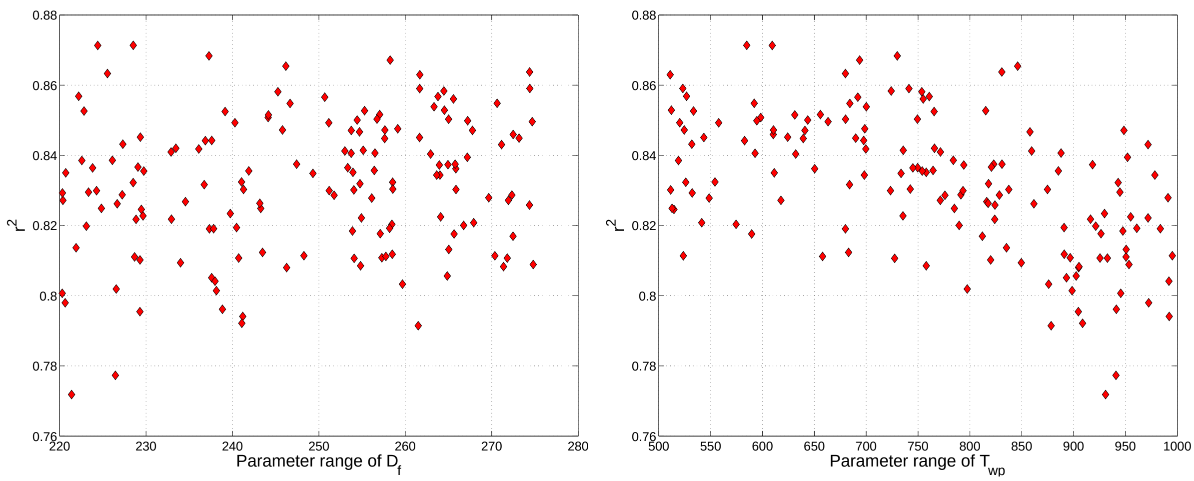


Figure 6.7: *Two parameter ranges plotted against the resulting r^2*

The calibration parameter P_{Feh} (7.2%) is bit high compared to literature, but there are three possible explanations for this relative high value. First, the number of cases is approximately 20 times overestimated, which is probably too much although it is known that the number of cases is underestimated. Second, within the present model EH transmission is the only mechanism that can move V. cholerae to upstream areas, therefore the high value is probably needed to explain the spatial pattern. Third, not all processes are included in the model, for example fish that may contain V. cholerae. It is possible that this transmission mechanism also explains partly this type of processes. For future studies it will be relevant to give the model some restrictions for the number of cases and it might be useful to develop a more extended bacteria procedure.

Figure 6.8 presents the epidemic curve of the final calibration result. The epidemic curve as well as the cumulative curve show the same curvature as the results of Agheksanterian and Gobbert (2007) and Mari et al. (2011). The development in time differs for each cholera epidemic. Agheksanterian and Gobbert (2007) show for 2 types of shredding cholera, low and high, that the curvature is the same however the period differs. Mari et al. (2011) simulated the cholera epidemic of 2000-2001 for KwaZulu-Natal province in South Africa. Here the epidemic curve lasted about 9 months with a peak after approximately 4 months. In all cases the curvature remains the same and are similar to the simulated result. That means initially a gradually increase of the number of cholera cases, then the curve increases rapidly until the peak is reached. After the peak the curve decreases initially rather fast and after a while it decreases more gentle and in this case after 90 days the number of cases is nearly zero.

Furthermore, the epidemic curve seems to have a relation with the discharge. The main increase of the epidemic curve coincides with the period of thigh discharges after 25 days. The other periods with discharge show also an increase in the number of cases. However, this relation may not be based on one rainfall event, therefore this will be further studied in chapter 8.

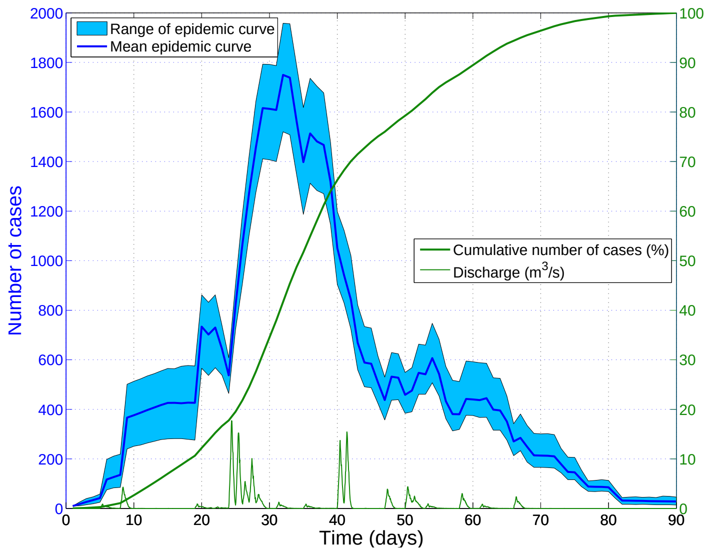


Figure 6.8: *Final epidemic curve after calibration, the range of the epidemic curve represents the minimum and maximum number of cases within the corresponding 90 runs.*

Table 6.1 presents the number of infections caused by the different transmission mechanisms. The distribution between the transmission mechanisms is now more or less equal to the expectations from literature. Glass and Black (1992) stated that VT transmission seems to be less important compared to the other routes of transmission, furthermore they claimed that there is no direct evidence of HH transmission. In line with this claim the model simulated that the contribution of HH and VT is equal to 3.2% of the cases.

The main transmission route of V. cholerae is via water (Glass and Black, 1992; Reidl and Klose, 2002), therefore it is expected that HEH and EH have the highest contribution to the number of cases. In the model a distinction was made between those transmission mechanisms, but in reality no distinction can be made between naturally existing cholera and people who got infected via V. cholerae from the dump site. Probably for this reason no information was found in literature about the distribution between those mechanisms. It is known that an infection requires 10^4 V. cholerae cells (Codeço, 2001; Lipp et al., 2002). It is assumed that EH transmission only occurs via free living V. cholerae cells, in that case there is a lot of water required to get infected. However, a copepod (size 1-2mm (Wikipedia, 2012)) may carry 10^4 to 10^6 V. choleare cells which may lead after ingestion to an infection. In the model this type of infection is not further specified. This means that the model does not include all the processes precise enough and that EH and/or HEH transmission compensate for these missing processes. Therefore it is hard to make a strong statement about the distribution between those transmission mechanisms, but this emphasizes again that the model requires an extended bacteria procedure.

Table 6.1: *Distribution of the cases between the four transmission routes*

Transmission mechanism	**HH**	**HEH**	**EH**	**VT**
Average number of cases	80	2461	683	22
Minimum and maximum number of cases	60-104	2237-2608	595-786	10-38
Contribution to total number of cases (%)	2.5	75.8	21.0	0.7

Finally, Skvortsov et al. (2007) claims that the critical factor for a high fidelity of an epidemiological model is the ability to independently validate its predictive results. However, there is no data available to perform such a test. This is more often a problem in epidemiological research, therefore mathematical models have been developed to validate the general behaviour of an epidemic, see for example Skvortsov et al. (2007), Hartley et al. (2006) and Agheksanterian and Gobbert (2007). These models are already used to compare the epidemic curve. However, for the geographical distribution this type of models was not found and means that the model should be verified on another catchment or study area with more or less the same characteristics.

6.5 Discussion

The cholera model contains many assumptions, the used data contains large uncertainties and there is practically no literature to compare the results with. To place the results in a broader perspective this section contains a discussion about four shortcomings of the cholera model and the consequences of them.

First, there is much knowledge about V. cholerae and the possible transmission mechanisms, but the behaviour of V. cholerae in water is still largely unknown (Bertuzzo et al., 2007). Due to this lack of knowledge the model contains a rather simple procedure, V. cholerae is present or not. The consequence is that individuals can only get infected by drinking contaminated water or eating contaminated food. While in reality it is possible to get infected by the combination of eating and drinking, since cholera is dose-dependent (Glass and Black, 1992; Mintz et al., 2005). More important to realize is that the model creates in some cases probably an infection, while in reality the dose was not high enough.

Within the model the bacteria procedure is only used for the HEH transmission. In reality after several days without rain V. cholerae will die due to dehydration (Colwell and Spira, 1992), while in the model the bacteria are still there. Also during the trip from the dump site to the river V. cholerae has a chance to dehydrate. The consequences of this dehydration are not taken into account but may influence the outcomes of the model. Therefore it will be useful to do more practical research to the transport mechanisms of V. cholerae especially around the dump sites and its survival in the aquatic environment.

In the cholera model it rains only on the dump site. To compensate for the raindrops from the upstream parts of the dump site the infected raindrops contaminates a patch for the duration of T_c. However, this assumes that all the upstream parts are equally sized, because this value is uniform for the model. That may cause a wrong number of cases because too few or too less fetching points are contaminated, which may influence the geographical pattern. It is expected that this effect is not so large looking at the results. But for future purposes it will be important to improve this in combination with a more advanced bacteria procedure.

Second, the cholera model took only the daily processes into account that are responsible for the direct spreading of cholera. In practice individuals will do more activities in a day, they for example visit markets, go to school or their jobs. This probably influences the source of water individuals use, but this is not implemented in the cholera model yet. So the lack of an extended activity model influences the model outcomes. In addition, the present activity model uses the same schedule for every day, while this is not the case in reality. Therefore chapter 7 will explore the consequences of the time people fetch water.

The spatial spread of cholera is mainly caused by the EH transmission mechanism and partly by the HEH transmission mechanism. However, researches show that V. cholerae could also be transmitted by all types of transport for example air planes and cars (Lee, 2001). This type of transmission could be included in a more extended activity model. The the influence of EH transmission on the spread will probably reduced when a more extended bacteria procedure is implemented. Another type of behaviour which is not implemented yet is that individuals in the real world can independently act. For example when individuals notice that the river water is too dirty to fetch, they will fetch water on another place.

Third, the chosen calibration procedure implicitly means that the epidemic event of 2005 is a kind of average event. Because the cholera model is averaged over 90 simulations which causes a more stable outcome but it also reduces the chance on a unique outcome. Furthermore, 16 parameters were used in the calibration, while there were only 11 data points available to verify the outcomes of the model. The risk of this is that the model goes too much in detail which may cause that the model is not useful for other epidemic events. The large number of parameters was needed because the model contains many probabilities that could not be based on literature. Therefore, before the model can be used for more general purposes it should be verified for another cholera outbreak.

Fourth, one of the assumptions of the model is that all processes that are responsible for the spatial pattern are included in the cholera model. When not all these processes are taken into account the model outcomes might be wrong. However, all the processes are distinguished from a literature study, this study showed that at least all relevant processes for the spatial pattern are included. But the model outcomes show that the level of detail of the processes might be improved.

Chapter 7: Model analysis

In this chapter the model outcomes will be analysed. Section 7.1 describes the importance of the hydrological procedure within the cholera model. Section 7.2 describes the spreading of cholera over time and section 7.3 discusses the influence of the initial infections on the model outcome.

7.1 Importance of hydrology in the cholera model

This sections analyses the importance of the hydrological processes in the spread of cholera. First, the number of cases in which the hydrology plays a role is determined. Second, the travelling time of V. cholerae in the study area will be calculated and third, the relationship between the discharge and the number of cases will be pointed out.

7.1.1 Number of cholera cases where hydrology plays a role

Within the cholera model the hydrology plays only a role in the HEH transmission, which causes 76% of the cholera cases. HEH contains in fact two types of spreading, namely: (i) Human to Environment to Human due to playing at the dump site (HEHP) (57 cases) and (ii) Human to Environment to Human due to contaminated raindrops from dump sites (HEHD) (2404 cases). The number of HEHD cases contributes the most to the total number of cholera cases, namely 74%. For this reason it is concluded that the hydrological model is important for the spreading of cholera.

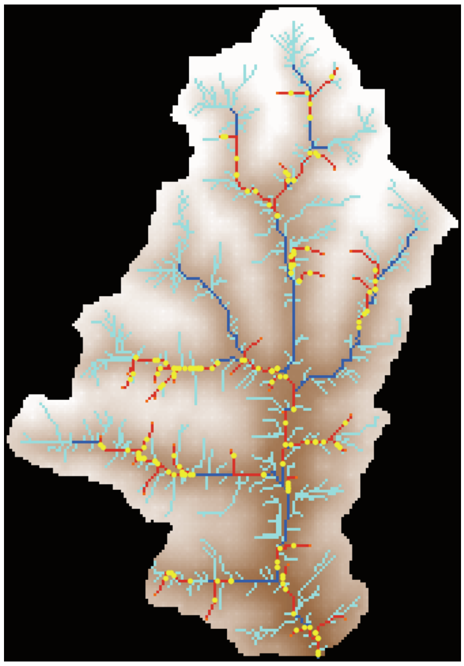


Figure 7.1: *Spread of the V. cholerae one hour after the rainfall event, in the case that all dump sites are contaminated. Yellow circles represent contaminated raindrops from the dump site and red squares represent the patches that are contaminated by the raindrops.*

7.1.2 Travel time of V. cholerae

Since hydrology plays an important role in the spread of cholera, it will be useful to analyse which patches are contaminated after a rainfall event (scheduled at the start of a day) at the moment that individuals fetch water. Figure 7.1 presents the study area and the patches that are contaminated 1 hour after the rainfall, the moment that individuals fetch water.

It was found that on average the raindrops coming from the dump site arrive at the catchment outlet after 6.5 hours. The last raindrop, which comes from the north, arrives after 13 hours at the catchment outlet. The last V. cholerae left the study area after 18 hours, which is logical since T_c is 5 hours. This means that the V. cholerae is flushed out of the system within one day.

It is not shown in figure 7.1 but it will be easy to imagine that after 2 hours more patches are contaminated with V. cholerae, because T_c retards the movement of V. cholerae, while the raindrops are moving to the outlet. This indicates that the moment that people fetch water will influence the number of cholera cases. Therefore in chapter 8 the consequences of rescheduling the time that households fetch water is one of the scenarios that will be investigated.

7.1.3 Discharge and the spread of cholera

In section 6.4.3 a relation between the discharge and the number of cholera cases in time was noticed. Compared with the findings above this might be expected.

Figure 7.2 shows clearly that the main increase of Total number of Cases (TC) = EH + HEH + VT + HH and HEHD coincide with the period of high discharges after 25 days. Also the other days with discharge coincide with an increase in the number of HEHD cases. Furthermore, like the cumulative epidemic the slope of HEHD shows a gradual decline in time. This is because the number of susceptible individuals decreases in time and therefore the number of cases will show a lower increase in time and eventually come to a stop.

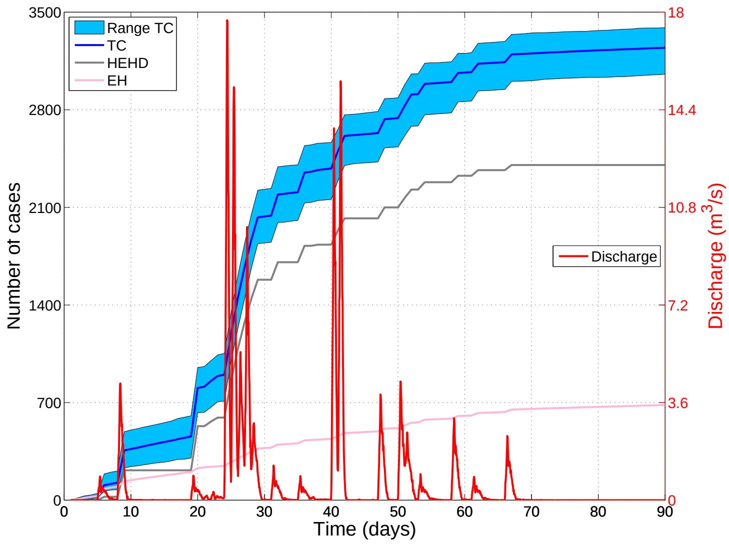


Figure 7.2: *Development of the number of TC, HEHD and EH cases in time*

The figure shows also that the number of EH cases has a less clean relation with the discharge than HEHD. The reason therefore is that when the rainfall is greater than P_{max} more households have to fetch water because of the absence of tap water. The fetching water procedure is modelled as follows, when there is no contaminated water from the dump site a household has a chance of P_{Feh} to fetch water contaminated with EH. Since most households will fetch HEHD contaminated water the number of EH cases will increase less than HEHD cases. However, people with no access to tap water have the same chance (P_{Feh}) on dry days of fetching contaminated water.

After a lot of rain it might be in reality that the bacteria in the aquatic environment flushed out of the system. Whether this is the case is unknown and therefore not included in the model. However, when the bacteria model is extended this type of the system's behaviour could also be included.

7.2 The diffusion of cholera in time and space

In this section the development in time and space of the number of cholera cases will be investigated more in detail. Therefore in the next section the development of the epidemic curve per community is analysed, followed by an analysis about the spread of cholera cases over the study area in time.

7.2.1 Epidemic curve per community

It was concluded in section 6.4.3 that the curvature of the epidemic curve was simulated well. In fact this curve is a summation of the epidemic curves of the communities. Figure 7.3 presents the epidemic curves of several communities. The presented communities are representative for the results of the other communities, the locations of them can be found in figure 6.2.

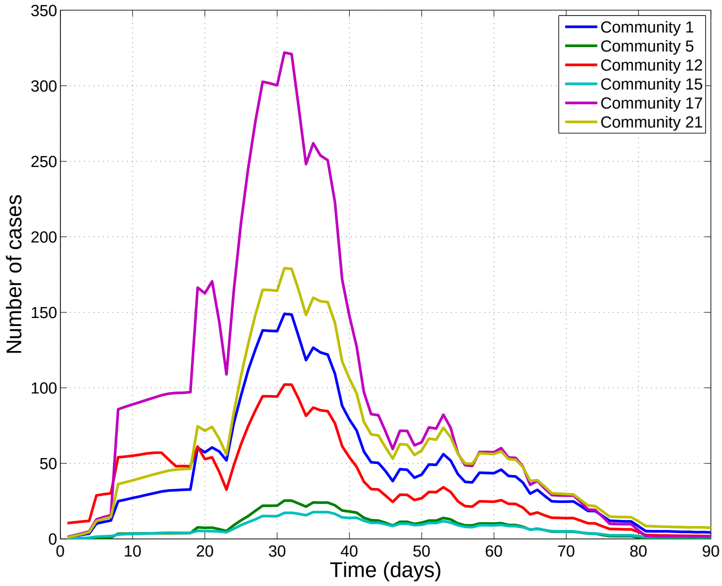


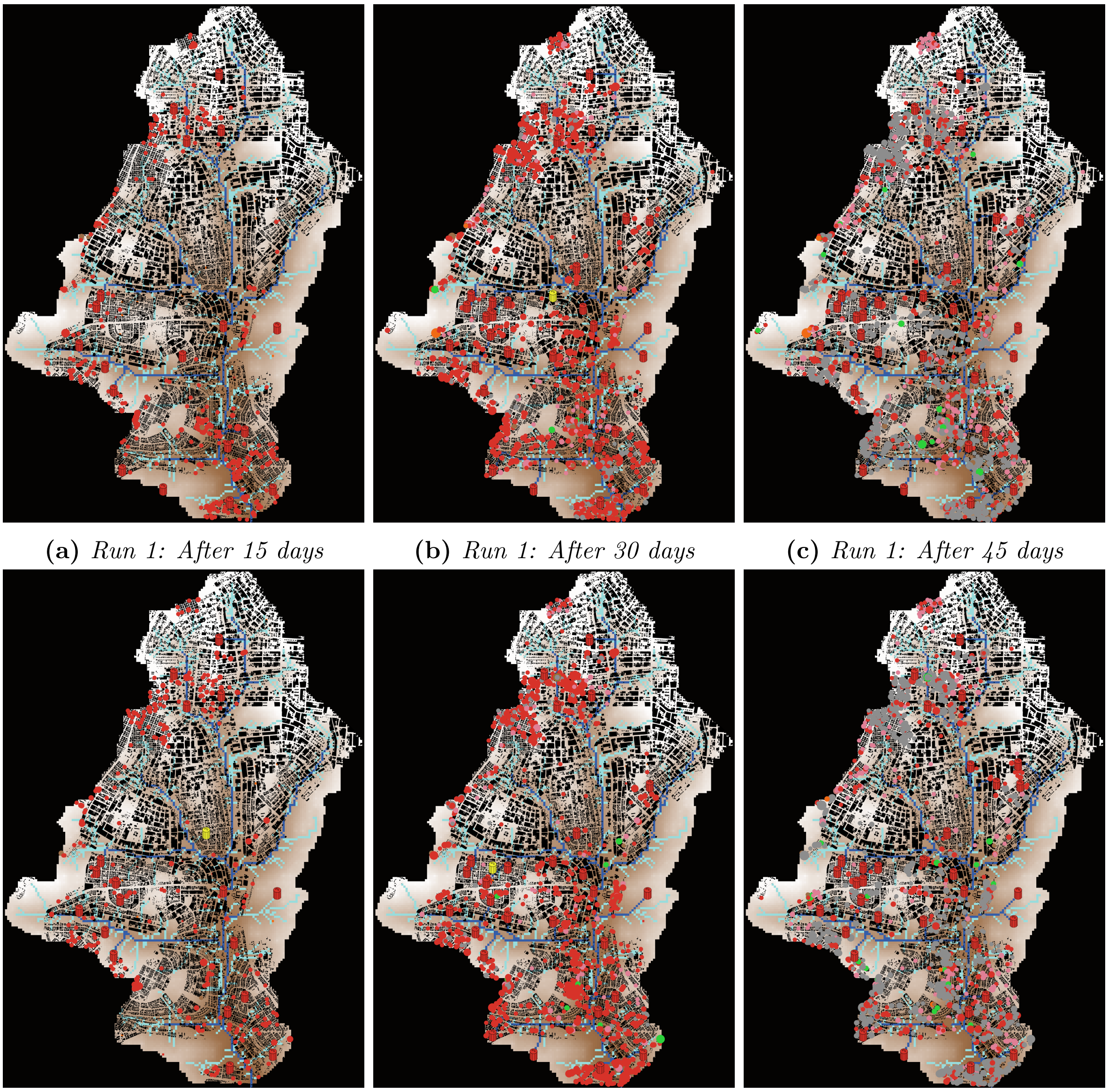
Figure 7.3: *Epidemic curve for several communities*

Most important to notice from the figure is that for all communities the epidemic curve starts and stops more or less at the same moment. This is not what might be expected from reality, however within the model this behaviour is logical. Because most individuals are infected via HEHD transmission which is only possible on days with rainfall, therefore the main increases of the number of cholera cases in the individual communities are at the same moment. The data of Osei (2010) shows that for this epidemic the first case in the communities were registered within 4 days from each other with exception for community 4 which registered the first case a month later, see table 4.2. The last cases in each community are reported within a month from each other. The model is developed to reproduce the geographical spread and not the development in time. This figure makes clear that the model should be improved before the model represents the right duration per community, but then also more data are required before the epidemic curve can be verified per community. The graph clearly shows that the number of cholera cases is related to the number of people in a community. When the number of people in a community is low also the number of cases is low and vice versa.

7.2.2 Distribution of cases over the study area

This section analyses the spatial pattern of cholera diffusion in time over the study area. Normally a simulation result is the average outcome of 90 simulations. It will be too time consuming to analyse all these simulations, therefore just two independent simulations are done and compared with each other. During these runs the model produced for every day an overview of the study area. The result after every 15 days is shown for each run in figures 7.4 and 7.5.

The overall spread of cases is quite similar, comparing figures at the same time steps for run 1 and 2. Most cases are situated in the south, the central east and the north east part of the study area. Comparing both runs, it was found that the infected dump sites are not precisely the same.



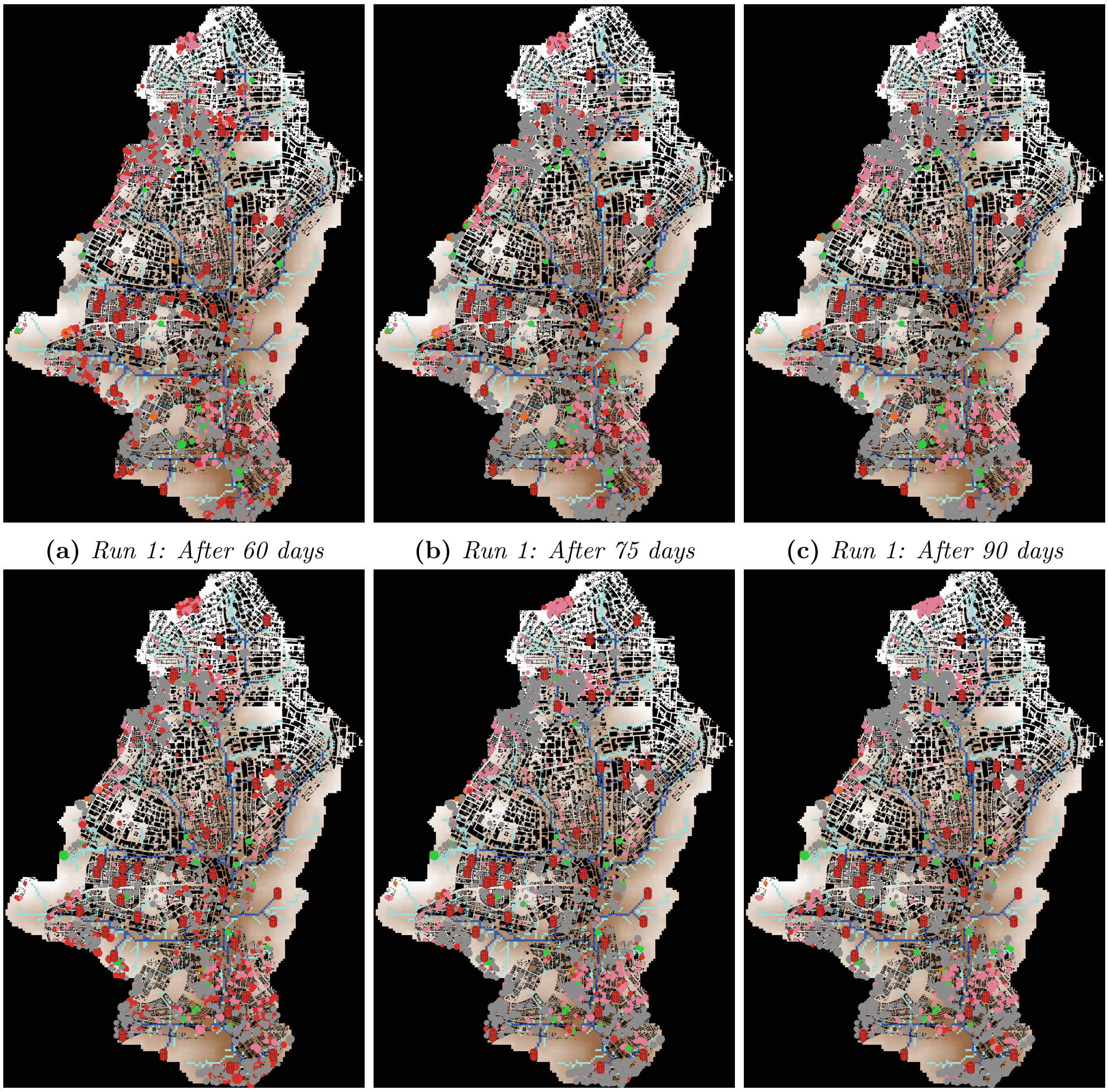
(a) *Run 1: After 15 days* **(b)** *Run 1: After 30 days* **(c)** *Run 1: After 45 days*

(d) *Run 2: After 15 days* **(e)** *Run 2: After 30 days* **(f)** *Run 2: After 45 days*

Figure 7.4: *Two independent simulations after 15, 30 and 45 days. The circles represent the households that are infected, the size represents the number of infected individuals in that household and the colour of it represents the type of transmission: grey=HEHD, brown=HEHP pink=EH, orange= VT, lime=HH. The garbage cans are dump sites that exceed* D_{imax}.

Due to the probabilistic character of the model the infected dump sites in the central part of the study area differ. This is mainly caused by the probabilistic character of the EH transmission. Because the first cases of cholera are already in the north after day one. The location of the household with the infected person determines which dump sites get contaminated, this causes the different sequence of dump site contamination.

Figures 7.5c and 7.5f show the infected cases in the study area at the end of the simulation period. The individuals that got infected via HEHD are distributed over all "low" and "middle" income areas. The individuals that got infected via EH are mainly clustered in the north east, because they have only access to river water that could not be infected by dump sites, their water fetching point is upstream from the point where the dump site contamination enters the



(a) *Run 1: After 60 days* **(b)** *Run 1: After 75 days* **(c)** *Run 1: After 90 days*

(d) *Run 2: After 60 days* **(e)** *Run 2: After 75 days* **(f)** *Run 2: After 90 days*

Figure 7.5: *Two independent simulations after 60, 75 and 90 days. The circles represent the households that are infected, the size represents the number of infected individuals in that household and the colour of it represents the type of transmission: grey=HEHD, brown=HEHP pink=EH, orange= VT, lime=HH. The garbage cans are dump sites that exceed* D_{imax}.

river. The reason for the clustering of EH cases in the south is not clear, while it was seen in multiple runs. The VT, HH and HEHP cases are distributed over the whole study area.

In general the final result shows that the middle and low income areas (figure 6.2) have the most cholera cases. It was noticed that not all the dump sites got a D_i exceeding D_{imax}. These dump sites are situated in the north-western part of the study area, for run 1 it are two dump sites and for run 2 one. Also in other simulations it was found that the dump sites situated in the north-western part do not all exceed D_{imax}. The reason for this is that those dump sites are situated in the high income area and people who dump their waste there are not infected.

7.3 Function of the trigger in the model

The primary trigger for the spread of cholera are 10 forced initial cases in community 12. To investigate their influence a model run is done without initial cases.

It was found that there was no significant effect on the geographical distribution. The epidemic curve shows a bit smaller number of cases between 0 and 20 days, however the peak is exactly the same. The total number of cases is also equal to the calibration result. Therefore it can be concluded that the initial cases within the present model have no effect and might be left out. However when the EH spreading procedure can and will be developed in the future, based on new knowledge, it is likely that the trigger of the model will become more important.

A question that follow from the findings until now: why do cholera epidemics happen not more often when there is such a high chance of an infection via EH transmission? Because this outcome has no direct relation with reality. The model simulates a period for which an epidemic was recorded. Since V. cholera is not endemic in this study area, the percentage of P_{Feh} is not representative for the entire year. Actually the stated question can not be answered with this model, because it requires a model that determines the start of a cholera outbreak, which was not the objective in this research.

Chapter 8: Scenario Analyses

The final part of this research consists of a scenario analysis. In the previous chapters, the cholera model was developed and analysed. This chapter investigates the influence of different scenarios on the number of cases, geographical distribution and epidemic curve. Section 8.1 describes the procedure and the objective, followed by a description of the scenarios in section 8.2. Finally section 8.3 presents the results of the scenarios.

8.1 Objective and procedure

The underlying idea of the scenario analyses is to get a feeling which measures will be useful to avoid or reduce the spread of cholera. Moreover it is meant to study the effect of the different processes and try to get a better understanding of V. cholerae's behaviour. The present model shows good results for the cholera epidemic of 2005 in Kumasi. But, the model is not validated on another area or outbreak within Kumasi. Together with the uncertainty of many model assumptions, this leads to the conclusion that it makes no sense to make strong statements about the effectiveness of measures that could be taken to avoid the distribution of cholera.

The results show that the cholera model can reproduce the geographical distribution rather well. Thus, the influence on the spreading of cholera could be investigated considering the uncertainties that are mentioned in chapter 6. Therefore the objective of the scenario analyses is to gain more insight in the importance of the different types of input data for the spreading of cholera.

In addition, scenarios that reduce the number of cholera cases in this case mean that in according with the model it could be an effective measure. When the reduction is large it indicates that it will have a higher probability to be a useful measure to reduce the number of cases in reality than the other scenarios.

The procedure to analyse a scenario consists of: (i) developing a scenario, (ii) implementing the scenario in the cholera model and (iii) an analysis. This analysis monitors three model outcomes for every scenario: the differences in the geographical distribution, the distribution between the different spreading mechanisms and the epidemic curve.

8.2 Scenarios

This sections explains the developed scenarios: (i) Rainfall data, (ii) Rescheduling the time of fetching water, (iii) Explanation of the spreading without HEHD, (iv) Everybody has access to tap water and (v) Removing dump sites.

8.2.1 Rainfall data

Section 7.1.3 indicates a relation between the discharge and the development in time of the epidemic curve. Therefore the objective of this scenario is to find the relation between the peak of the epidemic curve and the rainfall. The rainfall data of 2006, 2009 and 2010 are used to determine the effects, these data were also used for the verification of the hydrological model.

Although the rainfall can not be changed by any government it will be useful to get an insight in the different behaviour of the cholera for different types of rainfall distribution. Because governments can adapt their strategy to avoid the spread with this knowledge.

8.2.2 Rescheduling the time of fetching water

Section 7.1 shows that the time at which people fetch water could be important for the simulations results. Therefore the effect of shifting the moment that people fetch water will be investigated. There is chosen to shift the time of fetching water instead of shifting the moment that it rains. Because in reality individuals can influence the time they fetch water, but not the moment of a rainfall event. In addition, the daily activities that follow after fetching water (section 6.2) will be equally shifted in time to keep the model structure the same. For this analysis the model runs 90 times per shifting hour and is shifted for 21 hours.

8.2.3 Simulating the diffusion of cholera without HEHD

Section 7.1 concludes that the hydrology is of major importance for the spread of V. cholerae within the cholera model. The outcomes of the previous scenario indicates that the geographical distribution might be explained without HEHD transmission. To make the conclusion more general a scenario should be build that tries to reproduce the geographical distribution without HEHD transmission. There are two possible outcomes: the scenario succeeds or the scenario is not able to reproduce the geographical spreading. In the first case it is more likely that other input data/variables are responsible for the spreading of cholera cases. In the second case it will be more likely that the hydrology explains a significant part of the spreading of cholera. The aim of this scenario is to check which of these possibilities is true.

Therefore the following scenario is developed. The HEHD transmission will be turned off. The probability that people fetch naturally contaminated water will be varied from 20% to 35% (now $P_{Feh} = 6.4\%$), this is done to get the number of infected people in the same order of magnitude as it is in the calibrated model.

8.2.4 All households have access to tap water

The results of the previous scenario show in section 8.3.3 that it will be interesting to investigate the consequences of giving all households access to tap water. These results indicates that the model will be more stable when all households have access to tap water. Therefore in this scenario all households will be given access to tap water. Within the model the tap water will be set to 1 for all households in the synthetic population generator. This means that households only have a chance to fetch water when it rains.

8.2.5 Removing dump sites

For this scenario the dump sites close to water fetching points will be removed. The resulting map of dump sites is shown in figure 8.1, compared to the original map 11 dump sites are removed. Compared to the normal situation this means that more households use the same dump site. The aim of this scenario is to find the effect of removing the dump sites close to the river. It is expected that dump sites with a larger distance to the river will have a smaller contribution to the distribution of cholera and therefore cause less cholera cases.

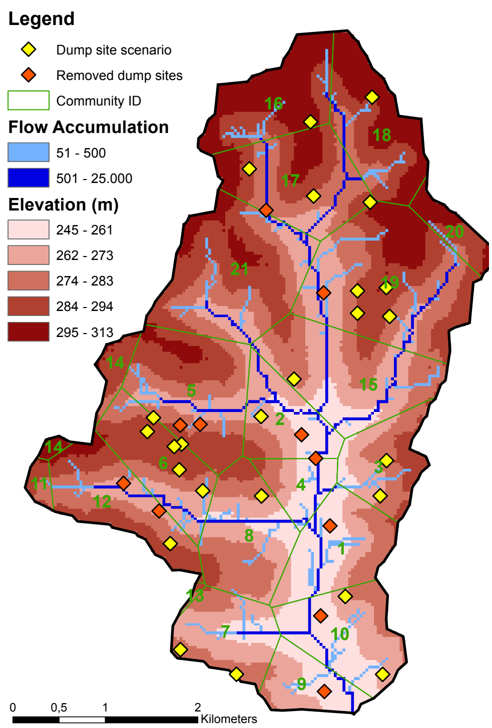


Figure 8.1: *Overview of the removed dump sites*

8.3 Results

This section describes the results of the tested scenarios.

8.3.1 Rainfall data

Table 8.1 presents the characteristics of the different rainfall events.

Table 8.1: *Characteristics of the different rainfall events*

Year	**2005**	**2006**	**2009**	**2010**
Total Precipitation (mm)	256	23	186	256
Number of days > P_{max}	19	2	14	24

The scenario shows that the rainfall did not have a significant influence on the geographical distribution. The r^2 was 0.84, 0.85 and 0.84 respectively for the years 2006, 2009 and 2010. The distribution between the different transmission mechanisms remains the same unless the number of days that the rainfall exceeds P_{max} is low. Then the number of HEHD cases reduces significantly. This is the case in 2006, the result is presented in figure 8.2.

Figure 8.3 confirms that the epidemic curve is mainly driven by the rainfall data. In 2006 there was nearly any rainfall and the main peak arises on the days that there is some rainfall. The amount of rainfall in 2010 was in the same order of magnitude as 2005, resulting in an epidemic curve that shows a more similar curve (figure 8.4). The difference between both epidemic curves is that the peak for 2010 is broader, this is caused by the difference in rainfall, because the rainfall of 2010 has more rainy days but not continuous in contrast with the rainfall of 2005. The result is a more gradually increasing epidemic curve.

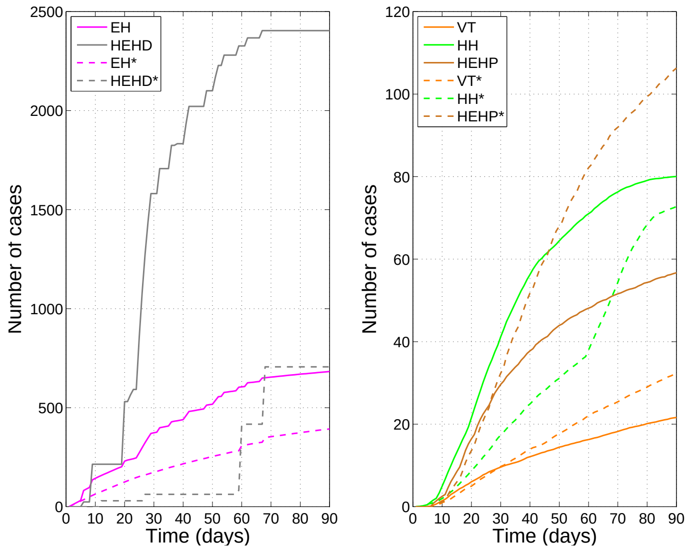


Figure 8.2: *Distribution between the different transmission mechanisms. Where the continuous lines are the calibration results and the dashed lines the results for the rainfall of 2006.*

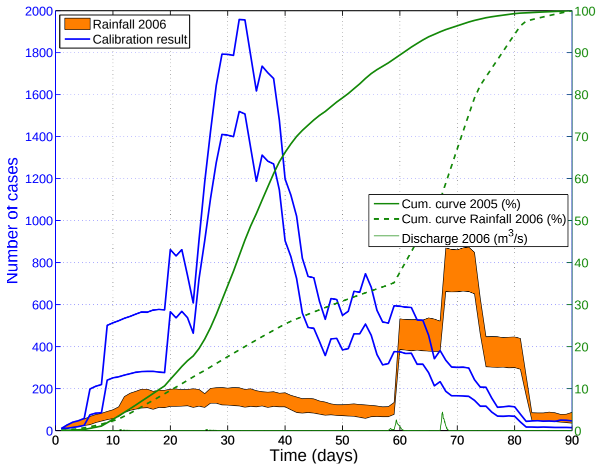


Figure 8.3: *Epidemic curve, rainfall 2006*

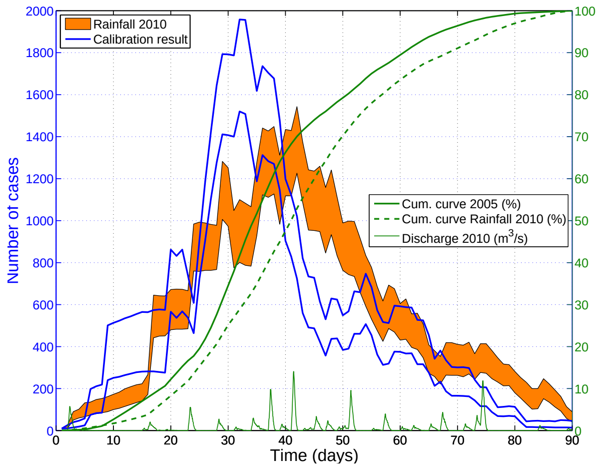


Figure 8.4: *Epidemic curve, rainfall 2010*

8.3.2 Rescheduling the time of fetching water

Figure 8.5 shows the geographical distribution for the hours where the number of HEHD cases shows a significant change, see figure 8.6. The figure shows that when the number of shifting hours increases the contribution of the communities one and four increases too. This is logical since both communities are situated downstream. On the other hand it shows that the contribution of communities 17 and 21 decreases, because they are situated upstream. After 15 hours the contribution of the communities became more equal to the calibration result, with a r^2 of 0.68. This might be an indication that the geographical distribution possibly could be explained without HEHD transmission. Because the model seems to be able to reproduce a fair geographical distribution when no HEHD is possible. Therefore this will be evaluated in the next section.

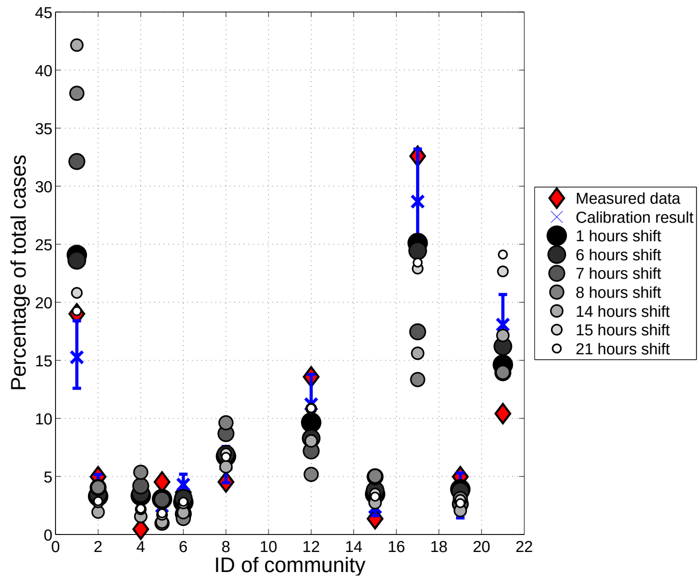


Figure 8.5: *Geographical distribution for shifting the time of fetching water*

The result for the distribution between the transmission mechanisms for every shifted hour is presented in figure 8.6. The graphs show clearly that the number of cases increases when the time is shifted from 1 to 6 hours, which could be explained by the larger number of contaminated patches. For 5, 6 and 7 hours the number of cases is larger than for the calibration result (shift = 0 hours) although the number of contaminated patches is lower. The reason for this is that the V. cholerae is now in places where more people fetch water. The number of EH cases decreases when the HEHD increases and vice versa. It is remarkable that the number of HEHP also decreases when HEHD increases. This means that children with a “low” income level got sick by drinking contaminated water instead of playing at the dump site.

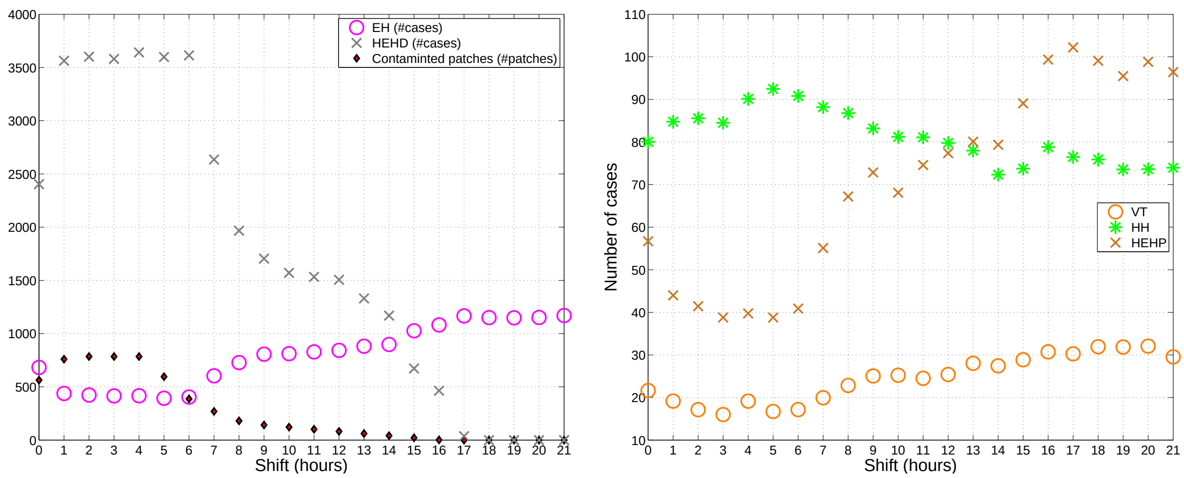


Figure 8.6: *Number of cases per transmission mechanisms for shifting the time of fetching water.*

It was found that the shape of the epidemic curve remains the same. But the number of cases initially increases and after 8 hours it becomes lower than the present epidemic curve, which might be expected based on the results found in section 7.1.3. Furthermore, it is noticeable that after 50 days for all shifting hours the epidemic curve becomes the same.

8.3.3 Simulating the diffusion of cholera without HEHD

It was found that the best result ($r^2 = 0.70$) for simulating the spread of cholera without HEHD was simulated with a P_{Feh} of 30%. This was just slightly better than using the original of P_{Feh}=6.4%, which resulted in a r^2 of 0.65. The objective was to find the value for which P_{Feh} represents the geographical distribution the best, therefore this section presents the results for P_{Feh}=30%.

Figure 8.7 shows that the geographical distribution of cholera cases is worse than the calibration result, with a r^2 of 0.70. It clearly shows that communities 1, 15 and 21 overestimate the number of cases compared to the calibration result. This might be caused by the number of inhabitants in these areas these are relatively high. Furthermore, these communities have large areas of low and middle incomes. This results in a relatively higher contribution because the EH is only based on a probability and does not depend on the location. This is in contrast with the calibration result where those areas show a smaller contribution to the total number of cases due to the fact that they are situated more upstream and the water fetching points are conveniently situated with respect to the dump sites. The figure shows that the number of cases depends mainly on the number of inhabitants in this case, because area 17 en 21 have the same number of inhabitants and also the same number of cases in this simulation.

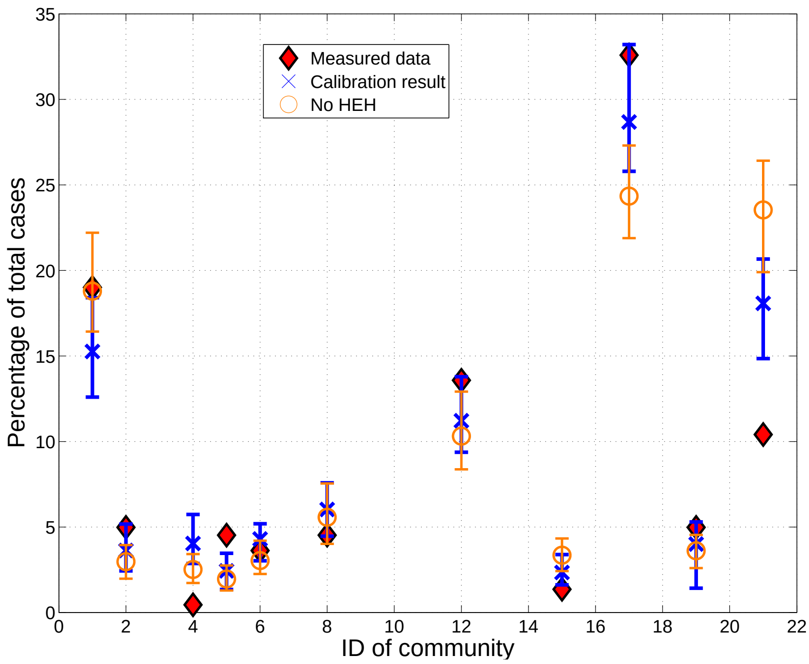


Figure 8.7: *Geographical distribution without HEH*

The number of TC is in the same order of magnitude as the calibration result. The distribution between the different spreading mechanisms are as expected, namely an enormous increase in the number of EH cases. Furthermore, VT, HH and HEHP are in the same order of magnitude as in the calibration result.

The epidemic curve presented in figure 8.8 has three important characteristics. First, the main peak of the curve is described well. This suggests that the peak is not mainly caused by EH or HEH, but it is likely that the initial assumption that households cannot use tap water during a rainy day explains the time of the growth. Therefore, it will be interesting to investigate how the epidemic curve develops when everybody has excess to tap water, so the rain is the main driver for the spreading of cholera. Second, before the main peak a small peak was noticed which is significantly larger than in the calibrated epidemic curve, this is directly caused by the increased probability of P_{Feh}. Third, the graph for the cumulative number of cases is more linear due to the more equal probability in time for getting infected.

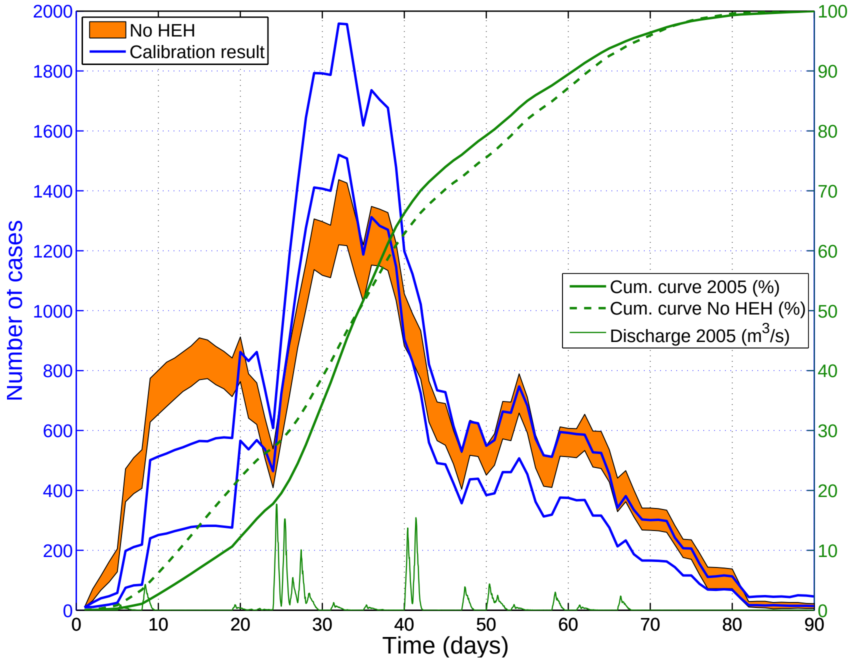


Figure 8.8: *Epidemic curve without HEH*

In conclusion, this scenario shows that the geographical distribution cannot be explained as well as it can with HEHD. This means that for the geographical distribution a location depended process explains a part of the spreading of cholera. It was also shown that the main reason for the existence of the curve is that more households have to fetch water when it rains.

8.3.4 Everybody has excess to tap water

Giving everybody access to tap water resulted in a slightly better geographical distribution, $r^2 = 0.88$. Furthermore, the model showed to be more stable, running the model 30 times resulted in a range for r^2 of 0.04. This is less than the range of the calibration result of 0.1. The reason for this stability is that the model is less dependent on the randomness of the EH transmission because this will happen less frequent.

The number of cases for the different transmission mechanisms are all in the same order of magnitude with the exception of EH. The reason for this is that only when it rains households have to fetch water and therefore the EH transmission that occurs on the dry days did not happen any more in the model.

The epidemic curve, presented in figure 8.9, clearly shows that the epidemic starts more gradually, caused by the reason mentioned in the previous paragraph. It can be seen that the other part of the curve is roughly the same.

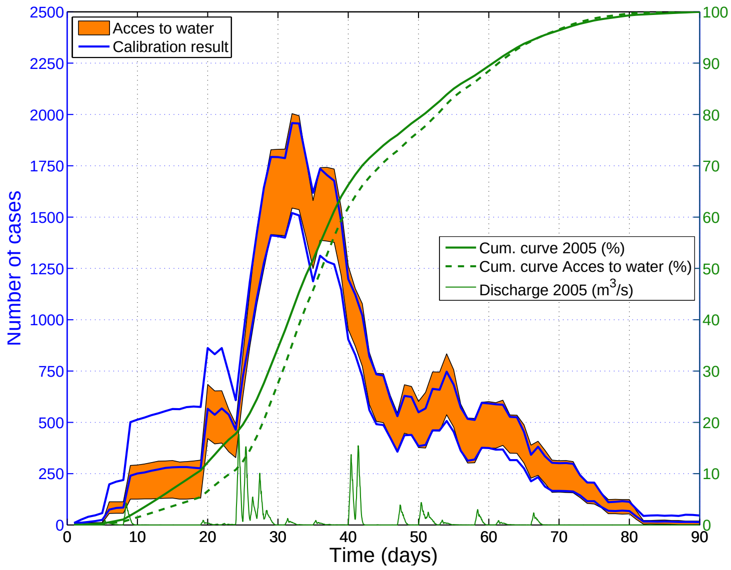


Figure 8.9: *Epidemic curve when all households have access to tap water.*

8.3.5 Removing dump sites

The "removing dump sites close to the river" scenario has some interesting outcomes. The model predicts the geographical distribution better with a r^2 of 0.92 (figure 8.10). It clearly shows that community 17 is simulated better now and that community 21 is closer to the measured value. Noticeable is also the decrease in the value of community 4. The reason for these changes is the decreased number of dump sites. Analysing figure 8.1 shows that communities 4 and 21 lose their dump sites, which means that the water they fetch from the river has a larger chance to be uncontaminated. As a result the relative number and therefore the chance of an infection is larger in community 17. The other communities show less variety in the results, which is logical since the number of dump sites is more equal to the original situation.

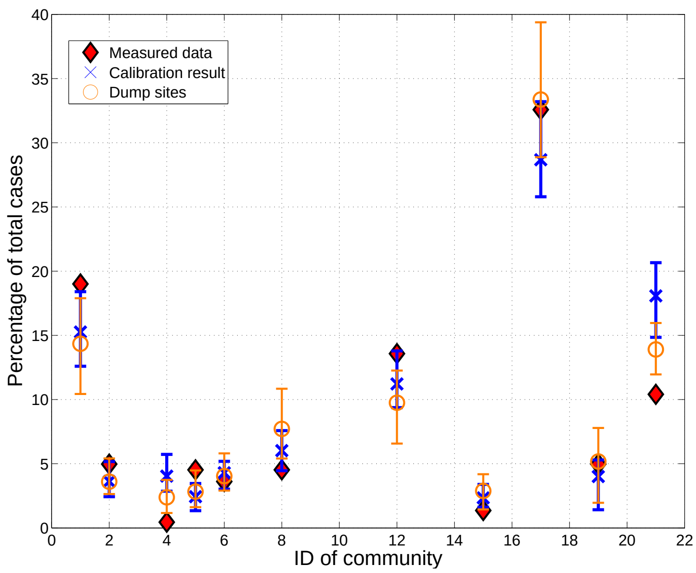


Figure 8.10: *The geographical distribution measured and simulated for each community.*

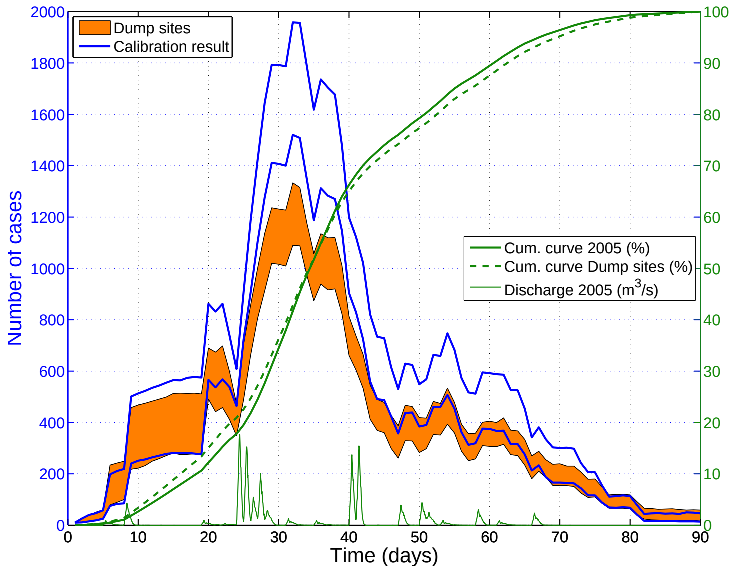


Figure 8.11: *Epidemic curve for the dump site scenario*

It was found that TC reduces with approximately 25%. This decrease is mainly caused by the reduction of HEHD and is slightly compensated by a small increase for the EH, HEHP and VT cases. This increase is caused by the larger amount of susceptible individuals, since the probabilities of a transmission remains the same. The increase of those three transmission routes show the same characteristics as were shown in section 8.3.3.

The curvature of the epidemic curve remains the same (figure 8.11). However the graph clearly shows that the number of cases is significantly less than for the calibrated epidemic curve. This was expected since the probability of HEHD transmission is reduced by removing dump sites. It is clearly noticed that the uncertainty bands are smaller, this is probably caused by the dump sites that achieve their D_{imax} earlier, this reduces the variety between two runs.

Chapter 9: Conclusions & Recommendations

The objective of this research was: "*Improve the model of Useya (2011) by implementing the hydrological processes that play a major role in the spread of V. cholerae, to gain more insight in the spread of cholera via water and use the model to evaluate different scenarios to make the strategy against cholera more effective.*" Four main research questions were formulated as a guideline towards this goal. This chapter will present the answers to these research questions in section 9.1, followed by the recommendations for additional research in section 9.2.

9.1 Conclusions

This section briefly describes the answers to the main research questions, which are the most important conclusions of this research and how they contribute to achieve the objective.

What are the relevant hydrological processes and how well can they be modelled in an ABM?

The most important improvements made to the hydrological procedure are: (i) expansion of the study area to the catchment area of the river, (ii) the velocity of the raindrops is based on physical principles and (iii) all the water flows through the outlet of the study area. The new hydrological procedure is calibrated on the discharges calculated by the CN method because there was no discharge data available.

The calibration (NS=0.95, RVE=0.20%) and validation (NS=0.92, 0.94 and 0.94, RVE$< \pm 0.3\%$) results are good. The velocities became more realistic with an average of 0.65 m/s for the river flow. The robustness of the model was analysed, because of the uncertainty of the input data. The tests show that the hours of rainfall have practically no influence on the performance of the hydrological procedure. The results also remain stable when the number of raindrops decreases from 5000 to 500 to reduce the calculation time.

How can the cholera model be improved and what is the performance after improvement?

The original cholera model is improved by implementing the hydrological model and expanding the study area to the size of the catchment area. During the calibration process it was noticed that the outcomes of the model were unstable due to the probabilistic character of the model, therefore multiple runs are averaged. It turned out that the range of r^2 reduces from 0.7 to 0.1 for respectively averaging over 1 and 90 runs.

The cholera model reproduces the geographical distribution quite well, with a r^2 of 0.87 for the epidemic of 2005. The shape of the epidemic curve was comparable to the curves found by Agheksanterian and Gobbert (2007) and Mari et al. (2011).

The calibration results show that many combinations of parameter values result in a good geographical distribution. This may indicate that the model is not only sensitive to the model parameters, but that other factors such as the input data may also play a major role in the outcomes of the model.

What are the most important mechanisms that explain the spread of cholera according to the model?

The distribution between the different transmission mechanisms is now more comparable with the expectations from literature, which are a large contribution of HEHD (Dump site → water → individual) and EH (naturally existing cholera → water → individual) transmission and just a small contribution (<3.2%) of the VT (flies → food → individual) and HH (individual → individual) transmission (Glass and Black, 1992; Reidl and Klose, 2002).

This means that the hydrological procedure plays an important role, because 75% of the cases is caused by the HEHD transmission. The epidemic curve also has a strong relation with the discharge and thus the rainfall in the study area. When it rains more than P_{max} the number of people that have to fetch river water increases. As a result the number of cholera cases increases because the probability of getting infected remains the same.

The EH transmission causes each run a random spatial pattern in time for cholera diffusion and enables the V. cholerae to move upstream. Because the chance on an infection depends on the availability of free living V. cholerae in the river. In reality V. cholerae's behaviour and places where it exists in the river are unknown, therefore its presence on a water fetching point is represented by a probability. This probability is during the entire simulation period for all fetching points the same. In contrast with the other spreading mechanisms where the probability to get infected with V. cholerae depends on the location in the study area. Therefore it will be important in a follow up study to improve this part of the model. However, at the same time the spatial pattern after 90 days in every run is more or less the same. The reason for this is that the households that have to fetch water always live in the same areas. The model clearly shows that most cases occur in the low and middle income areas.

Lastly, it was found that the trigger (first infection cases) in the cholera model has no significant influence on the outcomes of the model.

Which input data influence the model outcomes the most?

The results of the scenario analyses confirmed the relation between the discharge and the main increase in the number of cases. It was found that the geographical distribution was simulated more or less the same for the different rainfall events and the rainfall influences the peak of the epidemic curve significantly.

The analyses show that the time that people fetch water is important, because the number of cases varied in a wide range when shifting this time. The source of the HEHD cases is the V. cholerae that runs off from the dump sites with the rain that subsequently flows into the river and through the study area. Shifting the time 1-7 hours resulted in a larger number of HEHD cases, because the V. cholerae gets more time to contaminate patches. These patches will be contaminated for 5 hours. Shifting the time 8-18 hours the number of cases decreased, in this case the V. cholerae flows downstream which means that the upstream patches are not contaminated any more as a result the number of contaminated patches reduces and therefore the number of cholera cases. For shifting more than 18 hours there are zero cases caused by HEHD transmission, because all the V. cholerae flushed out of the study area. Whether this is realistic or not depends on the behaviour of the cholera in water. Therefore it is recommended to do further research to the behaviour of V. cholerae in water, because the model clearly shows that the travelling behaviour of V. cholerae influences the outcomes significantly.

Crucial for the model performance to simulate the geographical distribution properly is a process where the probability to get infected with cholera depends on the living location within the study area. Another finding was that the model turned out to be more stable when the randomness of the model is reduced by giving all households access to tap water.

Removing dump sites situated close to the river resulted in a better geographical distribution ($r^2 = 0.92$) than by the calibration. The reason for this is that there are more dump sites removed from communities that overestimated the number of cases, as a result the probability that individuals in those communities get infected due to HEHD transmission reduces. The total number of cases decreases because there are less dump sites consequently less patches are contaminated, therefore less individuals get infected via HEHD transmission. Furthermore, the uncertainty bands were smaller because there was less variability between two runs due to the reduced number of dump sites, the dump sites get faster contaminated and therefore more equal in time comparing two individual runs.

These results are based on the current hydrological procedure and are therefore only valid when this procedure is implemented properly. The hydrological procedure is improved, however, only general formulas about river flow are used, while the spread of V. cholerae starts from the dump sites. In the present model it is assumed that this happens when it rains, however in reality this may be more complex. Therefore it will be useful to study the behaviour of V. cholerae in the aquatic environment and especially in the neighbourhood of dump sites.

Objective

In conclusion, a new hydrological procedure is developed, calibrated, validated and successfully implemented in the cholera model. The expanded and improved cholera model is able to reproduce the geographical distribution of the cholera epidemic of 2005 well. The model and scenario analyses show some interesting insights in the behaviour of the epidemic.

The main improvement to the cholera model is: a more realistic residence time of the water. Furthermore, the study area is expanded to the catchment area which makes the model more useful. The scenario analyses make clear how the model behaves and that the input data are rather important.

The main shortcomings of the present model are: the hydrology around the dump sites and the interactions between the V. cholerae and the aquatic environment, and a more extended activity model which is able to describe the transmission to upstream areas in a more realistic way.

9.2 Recommendations

In the first part of this research the hydrological procedure is developed. This procedure is validated on data from another model. To improve the quality of this procedure, this procedure should be validated on measured discharges.

The cholera model contains too many assumptions and the validity is not proven, since it is only calibrated on one epidemic event. Therefore the model is not ready to make a comparison with reality. However, the ultimate goal of this kind of research is to fully understand the behaviour of cholera and use these models to ban cholera in the end. Therefore it will be interesting to improve this model further. For this reason it is recommended for follow up studies to improve the following processes of the model:

- The hydrology around the dump sites and the interaction between the aquatic environment and V. cholerae. In the present model the same hydrological assumptions are used everywhere. This resulted in a more realistic residence time of the water. However, in reality the amount of V. cholerae flushed from a dump site to the river will depend on the distance to the river and the intensity of the rainfall. Therefore these processes have to be investigated and included in the model.

- Improving the bacteria model in such a way that it describes V. cholerae cells independently and probably includes the dose-dependency of the disease. At the moment the EH transmission is only based on probabilities. To improve the model this procedure should be based on the biological processes that explain the behaviour of V. cholerae in the water. As a result this will determine the probability that a household will fetch contaminated water and makes it possible to implement the dose dependency.
- Extend the present activity model with activities that may play a role in the spatial spread of cholera. In the present model it was tried to include all the relevant activities where individuals can get infected with cholera. The results show that only because of a relative high chance of an infection by the EH transmission the spatial pattern can be explained. It is quite unlikely that this transmission mechanism explains the spatial spread in reality. It is more likely that this can be explained by the interactions between individuals on a day, for example visits to market places, family/friends, work, etc.

After these improvements are implemented it is recommended to improve the validity of the model by:

- Performing a validation for another catchment of Kumasi for which data are available or another cholera outbreak in an area that has characteristics similar to this study area.
- Collecting data about: daily activities, distribution of income levels, time that people do their activities and adaptations of their behaviour when there is a cholera outbreak. This information is used as input data for the cholera model. Improving the input data that resembles reality better will lead to a model that is useful to develop strategies to reduce cholera outbreaks or number of infected people during a cholera outbreak.

Finally, in this research there was a lack of data. Therefore for future researches more practical research has to be done on the cholera outbreak itself. The information needed from a cholera outbreak itself is the location of infected people and the most likely transmission mechanism that caused the infection. This type of information will give more insight in the behaviour of cholera and will improve the models that simulate cholera.

Bibliography

Adubofour, K., Obiri-Danso, K., and Quansah, C. (2013). Sanitation survey of two urban slum Muslim communities in the Kumasi metropolis, Ghana. *Environment & Urbanization*, 25(1):1–19.

Agheksanterian, A. and Gobbert, M. (2007). Modeling the spread of epidemic cholera: an age-structured model. *Maryland; sn, Nov. 2007,*, pages 1–16.

Agodzo, S. and Adama, I. (2004). Bulk density, cone index and water content relations for some Ghanian Soils.

Akumiah, P. O. (2007). *Water management and health in Ghana ; Case study -Kumasi.* PhD thesis, Linköpings Universitet.

Alam, M., Sultana, M., Nair, G. B., Siddique, a. K., Hasan, N. a., Sack, R. B., Sack, D. a., Ahmed, K. U., Sadique, a., Watanabe, H., Grim, C. J., Huq, a., and Colwell, R. R. (2007). Viable but nonculturable Vibrio cholerae O1 in biofilms in the aquatic environment and their role in cholera transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 104(45):17801–6.

Andreadis, K. M., Storck, P., and Lettenmaier, D. P. (2009). Modeling snow accumulation and ablation processes in forested environments. *Water Resources Research*, 45(5):n/a–n/a.

Arcement, G. and Schneider, V. (1989). *Guide for selecting Manning's roughness coefficients for natural channels and flood plains, Paper 2339.* United States Geological Survey Water-supply.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: model development. *Journal of the American Water Resources Association*, 34(1):73–89.

Bahremand, a., Smedt, F., Corluy, J., Liu, Y. B., Poorova, J., Velcicka, L., and Kunikova, E. (2006). WetSpa Model Application for Assessing Reforestation Impacts on Floods in MargecanyHornad Watershed, Slovakia. *Water Resources Management*, 21(8):1373–1391.

Barbé, L. L., Lebel, T., and Tapsoba, D. (2002). Rainfall variability in West Africa during the years 1950-90. *Journal of Climate*, 15:187–202.

Bertuzzo, E., Azaele, S., Maritan, a., Gatto, M., Rodriguez-Iturbe, I., and Rinaldo, a. (2008). On the space-time evolution of a cholera epidemic. *Water Resources Research*, 44(1):1–8.

Bertuzzo, E., Maritan, a., Gatto, M., Rodriguez-Iturbe, I., and Rinaldo, a. (2007). River networks and ecological corridors: Reactive transport on fractals, migration fronts, hydrochory. *Water Resources Research*, 43(4):1–12.

Bonabeau, E. (2002). Agent-based modeling: methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99 Suppl 3:7280–7.

Boonstra, J. (1994). Estimating Peak Runoff Rates. In Ritzema, H., editor, *Drainage principles and applications.*, chapter 4, pages 111–144.

Borroto, R. J. (1997). Ecology of Vibrio cholerae serogroup 01 in aquatic environments. *Revista panamericana de salud pública = Pan American journal of public health*, 1(1):3–8.

Cameron, D., Beven, K., and Naden, P. (2000). Flood frequency estimation by continuous simulation under climate change (with uncertainty). *Hydrology and Earth System Sciences*, 4(3):393–405.

Campion, B. B. and Venzke, J.-F. (2013). Rainfall variability, floods and adaptations of the urban poor to flooding in Kumasi, Ghana. *Natural Hazards*, 65:1895–1911.

Carlin, G., Cook, F., Cropp, R., Hartcher, M., Smajgl, A., Heckbert, S., and Hodgen, M. (2007). Using a catchment contour approach for simulating ground and surface water behaviour within agent-based modelling platforms. *ModSim2007.*

Castiglione, F. (2006). Agent based modeling. *Scholarpedia*, 1(10):1562.

Chaudhuri, K. and Chatterjee, S. (2009). *Cholera Toxins.* Springer Berlin Heidelberg, Berlin, Heidelberg.

Chow, V. T. (1959). *Open-Channel Hydraulics.* Mc Graw-Hill, New York.

Codeço, C. T. (2001). Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC infectious diseases*, 1:1.

Colwell, R. R., Seidler, R. J., Kaper, J., Joseph, S. W., Garges, S., Lockman, H., Maneval, D., Bradford, H., Roberts, N., Remmers, E., Huq, I., and Huq, a. (1981). Occurrence of Vibrio cholerae serotype O1 in Maryland and Louisiana estuaries. *Applied and environmental microbiology*, 41(2):555–8.

Colwell, R. R. and Spira, W. M. (1992). The ecology of vibrio cholerae. In Barua, D. and Greenough III, W. B., editors, *Current topics in infectious disease Cholera*, chapter 6, pages 107–123. Plenum medical book company, New York and London.

Du Preez, M., Van der Merwe, M., Cumbana, A., and Le Roux, W. (2010). A survey of Vibrio cholerae O1 and O139 in estuarine waters and sediments of Beira, Mozambique. *Water SA*, 36(5):615–620.

Environmental Systems Research Institute (2012). ArcGIS 9.2 Desktop Help.

Epstein, P. R. (1993). Algal blooms in the spread and persistence of cholera. *Bio Systems*, 31(2-3):209–21.

Faruque, S. M., Albert, M. J., and Mekalanos, J. J. (1998). Epidemiology, genetics, and ecology of toxigenic Vibrio cholerae. *Microbiology and molecular biology reviews : MMBR*, 62(4):1301–14.

Felsenfeld, O. (1965). Notes on food, beverages and fomites contaminated with Vibrio cholerae. *Bulletin of the World Health Organization*, pages 725–734.

Ferguson, A. S., Layton, A. C., Mailloux, B. J., Culligan, P. J., Williams, D. E., Smartt, A. E., Sayler, G. S., Feighery, J., McKay, L. D., Knappett, P. S. K., Alexandrova, E., Arbit, T., Emch, M., Escamilla, V., Ahmed, K. M., Alam, M. J., Streatfield, P. K., Yunus, M., and van Geen, A. (2012). Comparison of fecal indicators with pathogenic bacteria and rotavirus in groundwater. *The Science of the total environment*, 431:314–22.

Ghana Meteorological Services Department (2012). Weather information for Kumasi.

Ghana Statistical Service (2008). Population and Housing Census 2000. Technical Report August.

Glass, R. and Black, R. (1992). The epidemiology of cholera. In Barua, D. and Greenough III, W. B., editors, *Current topics in infectious disease Cholera*, chapter 7, pages 129–154. Plenum medical book company, New York and London.

Goodman, Y. and van Dyke, D. (2011). THE STORY OF CHOLERA.

Google (2013). Google Maps.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., and DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science (New York, N.Y.)*, 310(5750):987–91.

GSS (2008). Ghana Living Standards Survey, report of the fifth round (GLSS 5). Technical Report Glss 5, Ghana Statistical Service.

GSS (2012). 2010 Population & Housing Census Summary report of Final results. Technical report, Ghana Statistical Service, Accra.

Harris, J., Khan, A., and LaRocque, R. (2005). Blood group, immunity, and risk of infection with Vibrio cholerae in an area of endemicity. *Infection and immunity*, 73(11):7422–7427.

Hartley, D. M., Morris, J. G., and Smith, D. L. (2006). Hyperinfectivity: a critical element in the ability of V. cholerae to cause epidemics? *PLoS medicine*, 3(1):e7.

Hunt, R. J., Borchardt, M. a., Richards, K. D., and Spencer, S. K. (2010). Assessment of sewer source contamination of drinking water wells using tracers and human enteric viruses. *Environmental science & technology*, 44(20):7956–63.

Huq, a., Small, E. B., West, P. a., Huq, M. I., Rahman, R., and Colwell, R. R. (1983). Ecological relationships between Vibrio cholerae and planktonic crustacean copepods. *Applied and environmental microbiology*, 45(1):275–83.

Igbinosa, E. O., Obi, C. L., and Okoh, A. I. (2011). Seasonal abundance and distribution of Vibrio species in the treated effluent of wastewater treatment facilities in suburban and urban communities of Eastern Cape Province, South Africa. *Journal of microbiology (Seoul, Korea)*, 49(2):224–32.

Kan, B., Habibi, H., Schmid, M., Liang, W., Wang, R., Wang, D., and Jungblut, P. R. (2004). Proteome comparison of Vibrio cholerae cultured in aerobic and anaerobic conditions. *Proteomics*, 4(10):3061–7.

Kaper, J. B., Morris, J. G., and Levine, M. M. (1995). Cholera. *Clinical Microbiology*, 8(1):48–86.

Keraita, B. (2003). Influence of urban wastewater on stream water quality and agriculture in and around Kumasi, Ghana. *Environment and Urbanization*, 15(2):171–178.

Kirn, T. J., Lafferty, M. J., Sandoe, C. M., and Taylor, R. K. (2000). Delineation of pilin domains required for bacterial association into microcolonies and intestinal colonization by Vibrio cholerae. *Molecular microbiology*, 35(4):896–910.

Kobold, M. and Brilly, M. (2006). The use of HBV model for flash flood forecasting. *Natural Hazards and Earth System Science*, 6(3):407–417.

Kumasi Metropolitan Assembly (2010). Ghana districts.

Laux, P., Wagner, S., and Wagner, A. (2009). Modelling daily precipitation features in the Volta Basin of West Africa. *International Journal of Climatology*, 29:937–954.

Lee, K. (2001). The global dimensions of cholera. *Global Change & Human Health*, 2(1):6–17.

Liang, X. and Lettenmaier, D. (1994). A simple hydrologically based model of land surface water and energy ïuxes for general circu-lation models. *Journal of geophysical research*, 99(14):415–428.

Lindström, G., Johansson, B., and Persson, M. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of hydrology*, 201:272–288.

Lipp, E., Huq, A., and Colwell, R. (2002). Effects of global climate on infectious disease: the cholera model. *Clinical microbiology reviews*, 15(4):757–770.

Mari, L., Bertuzzo, E., Righetto, L., Casagrandi, R., Gatto, M., Rodriguez-Iturbe, I., and Rinaldo, A. (2011). Modelling cholera epidemics: the role of waterways, human mobility and sanitation. *Journal of the Royal Society, Interface*, 9(67):376–88.

Menard, S. (2000). Coefficients of Determination for Multiple Logistic Regression Analysis. *The American Statistician*, 54(1):17–24.

Ministry of Food & Agriculture (2011). Kumasi Metropolitan.

Mintz, E., Popovice, T., and Blake, P. (2005). Transmission of Vibrio cholerae O1. In Wachsmuth, K., Blake, P., and Olsvik, O., editors, *Vibrio cholerae and Cholera,*, chapter 22, pages 345–356. American Society for Micriobiology, Washington.

Moeckel, R. (2003). Creating a synthetic Population. *Proceedings of the 8th International Conference on Computers in Urban Planning and Urban Management (CUPUM)*, pages 1–18.

Morris, J. G. (2003). Cholera and other types of vibriosis: a story of human pandemics and oysters on the half shell. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 37(2):272–80.

Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models part IA discussion of principles. *Journal of hydrology*, 10:282–290.

Osei, F. (2010). *Spatial statistics of epidemic data: the case of cholera epidemiology in Ghana.* Phd thesis, University of Twente, Enschede.

Pagano, T., Wang, Q., Hapuarachchi, P., and Robertson, D. (2011). A dual-pass error-correction technique for forecasting streamflow. *Journal of Hydrology*, 405(3-4):367–381.

Patel, M. and Koornhof, H. (2004). Survival of Vibrio cholerae in industrially polluted water, with particular reference to iron concentrations. *Water SA*, 30(1):65–70.

Perrin, C., Michel, C., and Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4):275–289.

Pollitzer, R., Swaroop, S., and Burrows, W. (1959). *Cholera.* World Health Organization, Geneva, monograph edition.

Reidl, J. and Klose, K. E. (2002). Vibrio cholerae and cholera: out of the water and into the host. *FEMS microbiology reviews*, 26(2):125–39.

Riscassi, A. and Schaffranek, R. (2003). Flow Velocity Water Temperature, and Conductivity in Shark River Slough, Everglades National Park, Florida: August 2001-June 2002. Technical Report June, U.S. Department of the Interior, Reston.

Rivera, F. P., Medina, A. M., Bezada, S., Valencia, R., Bernal, M., Meza, R., Maves, R. C., and Ochoa, T. J. (2013). Bovine lactoferrin decreases cholera-toxin-induced intestinal fluid accumulation in mice by ganglioside interaction. *PloS one*, 8(4):e59253.

Romanowicz, R. (1993). TOPMODEL as an application module within WIS. *Application of Geographic Information Systems in Hydrology and Water resources*, (211):211–223.

Sack, D. a., Tacket, C. O., Cohen, M. B., Sack, R. B., Losonsky, G. a., Shimko, J., Nataro, J. P., Edelman, R., Levine, M. M., Giannella, R. a., Schiff, G., and Lang, D. (1998). Validation of a volunteer model of cholera with frozen bacteria as the challenge. *Infection and immunity*, 66(5):1968–72.

Sanyal, S., Singh, S., and Tiwari, I. (1974). Role of household animals in maintenance of cholera infection in a community. *Journal of Infectious Diseases*, 130(6):575–579.

SCS (1986). Urban Hydrology for Small Watersheds, Tehnical Release 55. Technical report, United States Department of Agriculture Soil Conservation Service Engineering Division.

Shaw, E. M., Beven, K. J., Chapell, N. A., and Lamb, R. (2011). *Hydrology in Practice.* Spon Press, London and New York, 4th edition.

Singleton, F. and Attwell, R. (1982). Effects of temperature and salinity on Vibrio cholerae growth. *Applied and environmental microbiology*, 44(5):1047–1058.

Skvortsov, A., Connell, R., Dawwon, P., and Gailis, R. (2007). Epidemic modelling: Validation of agentbased simulation by using simple mathematical models. *Proceedings of the Congress on Modelling and Simulation, MODSIM'07, Modelling and Simulation Society of Australia and New Zealand*, pages 657–662.

Soil Conservation Service (1986). Urban Hydrology for Small Watersheds, Tehnical Release 55.

Soong, D., Prater, C., Halfar, T., and Wobig, L. (2012). Manning's Roughness Coefficients for Illinois Streams. Technical report, U.S. Department of the Interior, U.S. Geological Survey, Reston.

Tutiempo Network SL (2005). Rainfall data Kumasi, Ghana.

U.S. Department of the Interior (2012). Global Data Explorer.

U.S. Department of Tranportation (2011). Introduction to Highway Hydraulics.

Useya, J. (2011). *Simulating diffusion of cholera in Ghana*. Master thesis, University of Twente,, Enschede.

Vezzulli, L., Pruzzo, C., Huq, A., and Colwell, R. R. (2010). Environmental reservoirs of Vibrio cholerae and their role in cholera. *Environmental Microbiology Reports*, 2(1):27–33.

Wang, Z., Batelaan, O., and Smedt, F. D. (1996). A distributed model for water and energy transfer between soil, plants and atmosphere (WetSpa). *Physics and Chemistry of the Earth*, 21(3):189–193.

Water & Sanitation for the Urban Poor (2012). Kumasi, Ghana.

Whittington, D., Lauria, D., and Choe, K. (1993). Household sanitation in Kumasi, Ghana:a description of current practices, attitudes, and perceptions. *World Development*, 21(5).

WHO (2011). Cholera cases reported to WHO by year and by continent.

Wikipedia (2012). Copepod.

Wilensky, U. (1999). *Netlogo. http://ccl.northwestern.edu/netlogo/.* Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Xu, H., Roberts, N., and Singleton, F. (1982). Survival and viability of nonculturableEscherichia coli andVibrio cholerae in the estuarine and marine environment. *Microbial Ecology*, 8:313–323.

Zuckerman, J. N., Rombo, L., and Fisch, A. (2007). The true burden and risk of cholera: implications for prevention and control. *The Lancet infectious diseases*, 7(8):521–30.

Appendices

Appendix A: Scheme of the original model

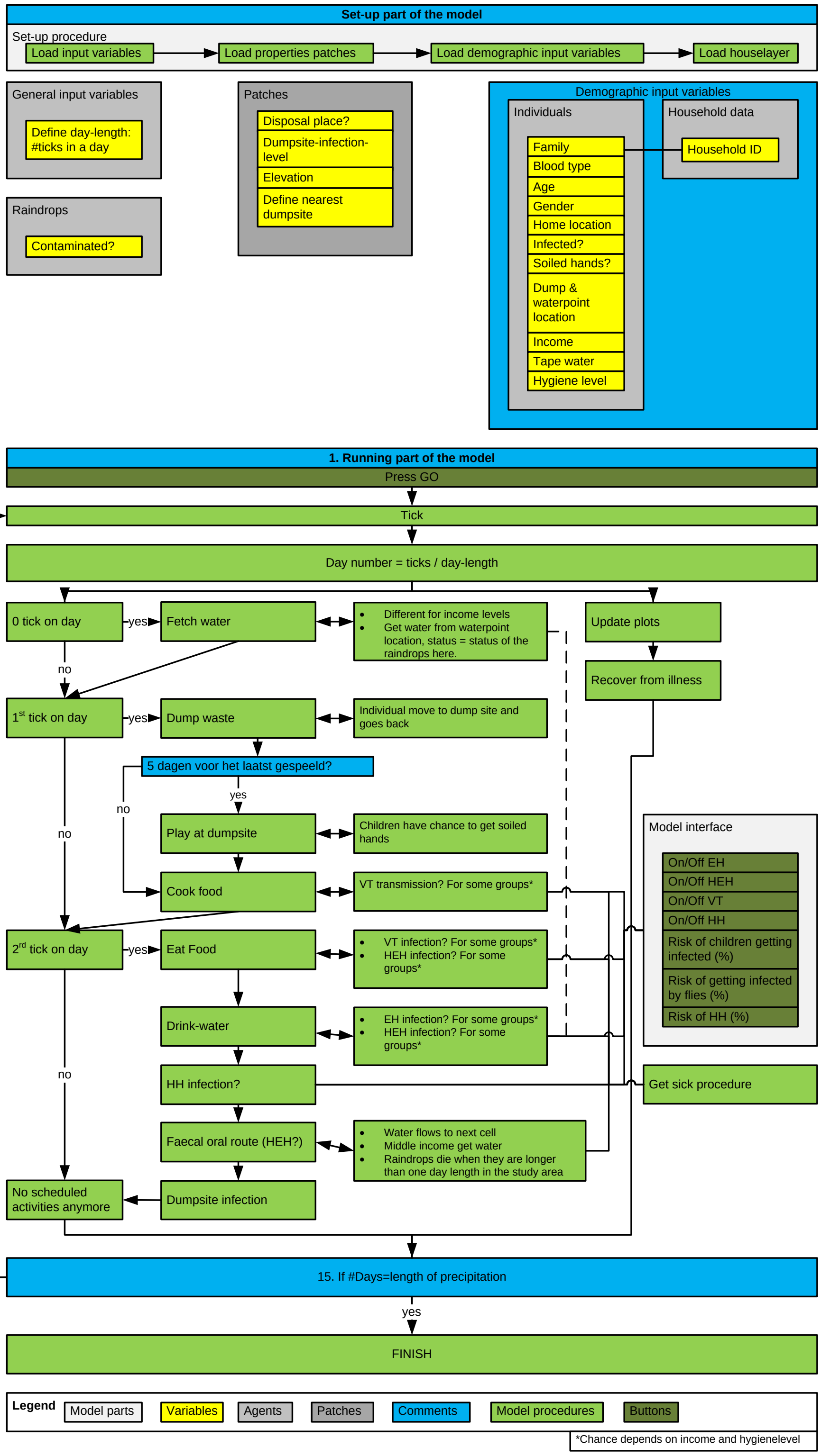


Figure A.1: *Schematic presentation of Useya's (2011) cholera model.*

Appendix B: ABM and Netlogo

The first section contains a brief description about agent based models in general. Then the second section presents a brief introduction to Netlogo and the most important terminology which is also used within this report.

B.1 What is an Agent Based Model?

ABM is a relatively new powerful simulation technique, also called individual-based modelling (Castiglione, 2006). The technique is used to create complex models by integrating well-understood models at a smaller scale. The ABM is used to scale up, then the ABM framework is utilized to coordinate, stimulate and drive the behaviour interactions of the well understood smaller scale models (Carlin et al., 2007). With other words the technique is used for modelling phenomena as dynamical systems of interacting agents (Castiglione, 2006). To do so the following question can be asked: What observed patterns seem to characterize the system and its dynamics, and what variables and processes must be in the model so that these patterns could, in principle, emerge? (Grimm et al., 2005).

In an ABM a dynamic system is modelled as a collection of autonomous decision-making entities called agents. This makes it possible to study the effect of individual made decisions. Compared to other modelling techniques the benefits of ABM can be made in three statements: (i) ABM captures emergent phenomena, (ii) ABM provides a natural description of a system, (iii) ABM is flexible, it is easy to add more agents and it provides a natural framework for tuning the complexity of the agents: behaviour, degree of rationality, ability to learn and evolve and rules of interaction. For these reasons ABM is applied in many fields, including epidemiology (Bonabeau, 2002).

B.2 Netlogo

In the master thesis of Useya (2011) an assessment is made for the ABM program that should be used for the developed cholera model. Netlogo is an ABM tool kit that make up a two-dimensional modelling environment of Turtles (Agents), Patches, Links and the Observer, this enables the turtles to move either in the x or y direction. Each turtle is able to make its own decision by assessing a certain situation on the basis of a set of rules, it is possible to have different type of turtles e.g. drops of water or humans. The patches are the environment of the agents, it forms the ground of the agents. The links take care of the connections between the turtles and finally the observer oversees everything that is going on (Wilensky, 1999).

Netlogo runs the model for the agents that are defined, you can make as many as you want. Therefore most models contain a set-up procedure where the area is defined and the patches get a meaning and some properties. Then the program runs the model, when the program meets a loop or statement, e.g. for-loop or if-statement, it continues the whole loop or statement for a single agent or patch till it reaches the 'end' term then it does it again for the next agent, after this the program continues the model.

Appendix C: Background information about the hydrological model

C.1 Curve Number Method

The CN method calculates the direct runoff. The initial accumulation of rainfall represents interception, depression storage and infiltration before the start of runoff is called initial abstraction. After runoff has started, some of the additional rainfall is lost, mainly in the form of infiltration; this is called actual retention. With increasing rainfall, the actual retention also increases up to some maximum value, the potential maximum retention. The following steps should be taken to calculate the runoff (SCS, 1986; Boonstra, 1994).

1. Use the same rainfall data and duration for the CN method
2. Determine from depth-duration-frequency curves the duration of the rainfall, see equation C.1. Where P = Precipitation [mm], P_a = Potential abstraction [mm], Q = excess rainfall [mm] and CN = Curve Number [-].
3. After this a unit hydrograph for each raining day is designed e.g. in figure C.1. Equation C.3 provides the peak discharge (Q_p [m^3/s]), the duration of the rainfall (D[seconds]) depends on the rainfall data, equation C.4 determines the time to the peak runoff [hours] T_p, time after the peak runoff [hours] T_r with equation C.5 and the time lag between the rainfall and the peak of the discharge [hours] T_{lag} can be calculated with equation C.7. Where P = Precipitation [m], P_a = Potential abstraction [m], CN = Curve Number [-], A = surface of the study area [m^2], L = hydraulic river length = $A^{0.5}$[m] and γ_s = averaged slope in basin [m].
4. The final step is to make a summation of all the individual unit hydrographs.
5. More background information about the method can be found in Boonstra (1994) and SCS (1986).

$$Q = \begin{cases} \dfrac{(P-0.2P_a)^2}{P+0.8P_a} & \text{if } P > 0.2P_a \\ 0 & \text{if } P < 0.2P_a \end{cases} \quad \text{(C.1)}$$

$$P_a = \left(\frac{1000}{CN} - 10\right) \quad \text{(C.2)}$$

$$Q_p = \frac{3}{4} \cdot P \cdot \frac{A \cdot 10^3}{T_p} \quad \text{(C.3)}$$

$$T_p = \frac{1}{2}D + T_{lag} \quad \text{(C.4)}$$

$$T_r = \frac{5}{3} \cdot T_p \quad \text{(C.5)}$$

$$\gamma_s = \frac{\text{Highest point - Elevation basin outlet}}{\text{Maximum length river}} \quad \text{(C.6)}$$

$$T_{lag} = \frac{3.42 \cdot L^{0.8}(1+P_a)^{0.7}}{\gamma_s^{0.5}} \quad \text{(C.7)}$$

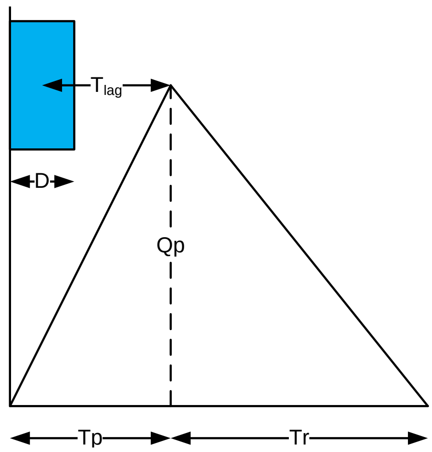


Figure C.1: *Unit hydrograph*

C.2 Sheet flow formula

The sheet flow formula is given in equation C.8 with its dimensions. SCS (1986) did not give a dimension for the Manning coefficient for sheet flow (m_{sh}) and it is unclear whether the m_{sh} is in inches or feets, however it is assumed that is is in inches. Equation C.9 presents the calculation of the travel time with parameters using the SI units.

$$\text{Travel time [hours]} = \frac{0.007(m_{sh}[\frac{\text{hours}^{1.25}}{\text{in}^{0.375}}] \cdot F_L[\text{ft}])^{0.8}}{(P[\text{in}])^{0.5} \cdot S[\frac{\text{ft}}{\text{ft}}]^{0.4}} = \frac{([\frac{\text{hours}^{1.25}}{\text{in}^{0.375}}] \cdot [12\text{in}])^{0.8}}{[\text{in}]^{0.5} \cdot [-]^{0.4}} \tag{C.8}$$

$$\text{Travel time [seconds]} = 0.007 \cdot 12^{0.8} \cdot 3600 \cdot \frac{(m_{sh}[\frac{\text{hours}^{1.25}}{\text{m}^{0.375}}] \cdot F_L[\text{m}])^{0.8}}{(P[\text{m}])^{0.5} S[\frac{\text{m}}{\text{m}}]^{0.4}} \tag{C.9}$$

C.3 Calibration inputs and results

Table C.1: *Ranges of the Manning coefficient and Hydraulic radius*

Parameter	Min	Max	Dimension
R gully	0.1	0.25	m
R river	0.25	0.6	m
m_{sh}	0.01	0.2	hours$^{1.25}$·m$^{-0.375}$
m_c gully	0.1	0.4	m$^{\frac{1}{3}}$·hour
m_c river	0.05	0.3	m$^{\frac{1}{3}}$·hour

Table C.2: *Ranges of the model parameters for the calibration process*

Parameter	Min	Max	Dimension
F_L	30	60	m
m_{sh}	0.01	0.2	hours$^{1.25}$·m$^{-0.375}$
S_{sg}	2	30	#cells
S_{gr}	400	600	#cells
ManningHRgully	0.5	7.9	m/hour
ManningHRriver	1.3	14.2	m/hour
S_{min}	0.001	0.005	m/m

C.4 Verification results

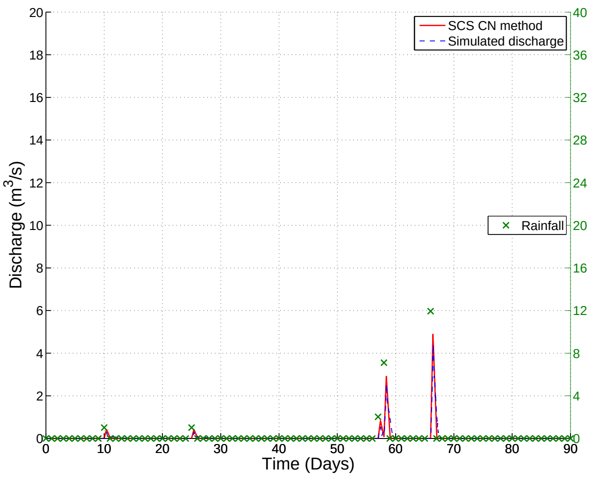


Figure C.2: *Verification result 2006*

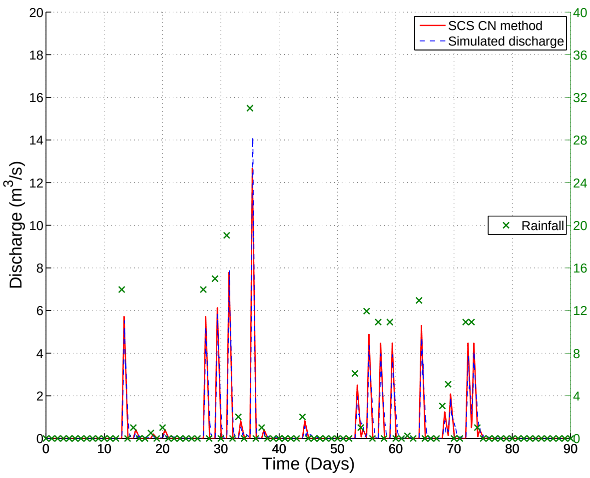


Figure C.3: *Verification result 2009*

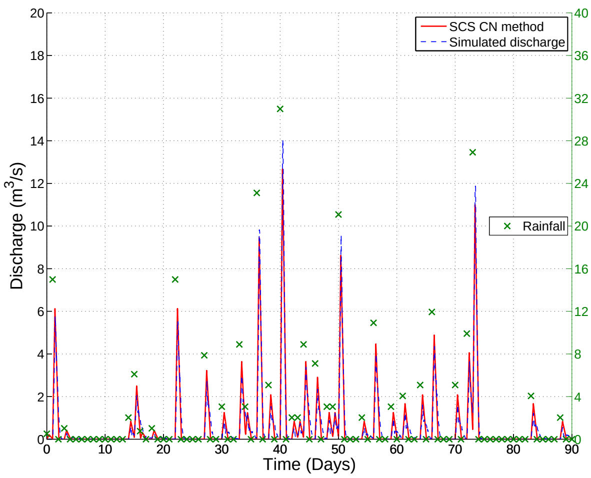


Figure C.4: *Verification result 2010*

Appendix D: Background information about the cholera model

D.1 Changes of the original procedures

This section describes the changes that are made to the original procedures. First the changes to the fetching water procedure are described. Second a list of smaller changes is provided.

D.1.1 Fetching water procedure

This section describes the changes to the fetching water procedure.

Useya (2011) states that individuals with a high income get water 1 out of 3 days. However in the original model when it is the day that they have to fetch water they have also a chance that they would not have to fetch water, this chance is equal to the chance "children getting infected at a dump site". However both statements do not make sense, because fetching water has nothing to do with children furthermore the people need water which is independent of a certain chance. Therefore both statements are removed from the procedure.

Next to this individuals have a chance of 50% to fetch water contaminated with natural existing V. cholerae, but is not implemented within the model. Therefore the fetching water procedure is changed, individuals now fetch water for a household, then the water could be contaminated with V. cholerae from the dump site, when this is not the case there will be a chance that the water get contaminated with natural V. cholerae (P_{fw}).

Further the original model asked every individual above 15 to fetch water, however the water is in practice used by a household. Therefore the procedure is changed and one member of a household fetches water and this water is utilized by all household's members.

Furthermore individuals with a middle income do not fetch water within the fetching water procedure, in the new model they are incorporated in the procedure.

D.1.2 Changes to several original procedures

The numbers correspond with the numbers in figure 6.1.

7. In the cholera model people have to get water from another source when the rain is above P_{max} , because then the tap water will not work. Individuals with a low income will fetch water from the river, middle income will have a chance of 25% that they have money to buy water else they have to fetch it and individuals with a high income have a chance of 75% that they can buy water. Both percentages are arbitrary chosen and not taken into the calibration process.
8. During the dumping waste procedure individuals move to the dump site and move back to their household location and nothing changed in the model. This has been changed, individuals who move to the dump site and have infected family members will add 1 to the D_i this will increase the chance that the raindrops flowing over the dump site get

infected. The patch that represent the dump site will achieve a counter that represents the D_i .

9. The playing-at-dumpsite procedure gives children a probability of 10% that they will play at the dump site every 5 days, however the report stated that children play every 5 days at the dump site. Therefore this chance is removed from the procedure.
10. During the period that individuals prepare the food, flies get the chance to contaminate the food. The procedure controls whether an infected dump site is situated within 500 meters, then the food has a chance of "RiskOffFliesGettingInfected" to get contaminated with cholera. Only when foods get contaminated the model will represent this in a turtle, this saves calculation time.
11. This procedure allows now also that individuals with blood type "O" can get infected with cholera due to flies. Furthermore it distinguishes now two infection routes, namely VT and HEH due to soil hands gathered during playing at the dump site.
12. In the original model the income level determined the source of the cholera, although the type of water that is fetched determines in fact the source of the cholera infection, see table 2.1. Furthermore every household that need fetched water will drink this.
13. In the original model infected individuals with a middle or high income are set as HEH transmission and with a low income as a EH transmission when the individuals are drinking water. Because the type of infection individuals get after drinking fetched water depends now on the source of the cholera instead of income.

Table D.1: *Probabilities that a household has to fetch water*

Income level	**P_{fw} per household**[1]
low	100
middle	50+P_{fw}
high[2]	25+P_{fw}

[1]The chances are the same when the rain exceeds P_{max}
[2]This chance exists only when the rain exceeds P_{max}

Table D.2: *Probabilities for the drinking water and eating food procedures*

Income level	**Hygiene level**	**Blood Type**	**P_{if}**[1]	**P_{cw}**[1,2]
low	low	O	30-P_{if}	30-P_{cw}
low	middle	O	20-P_{if}	10-P_{cw}
middle	low	O	20-P_{if}	30-P_{cw}
high	low	O	10-P_{if}	10-P_{cw}
low	low	Other	15-P_{if}	15-P_{cw}
low	middle	Other	10-P_{if}	-
middle	low	Other	10-P_{if}	15-P_{cw}
high	low	Other	5-P_{if}	5-P_{cw}

[1]Defined per individual depended on Income level, Hygiene level and Blood type
[2]Accounts for EH and HEHD contaminated water

D.2 Development of the house layer

A house layer is developed since there is no detailed information available about the location of the houses within the study area. ArcGIS 10.1 provides a topographic base layer which contains the location of the buildings also for Kumasi. From this base layer a detailed picture (zoomed to 1:1000) is saved, this level of detail was needed to see also the smallest houses on the resulting picture. Next this picture was loaded again into ArcGIS and geo-referenced using the georeferencing tool. It turned out that the green values between 201 and 218 represents the houses, therefore the other values were removed. Then the raster file is converted to a shape file. Finally the buildings that are clearly no houses, e.g. the airport and schools found on google maps, are removed from the layer.

Afterwards the income level of each house was determined, therefore was assumed that all houses within a block will have the same income. Next the compounds who could be easily recognized from the map are identified, these were assigned to the 'middle' incomes, figure D.1 provides a satellite picture of compounds houses. Afterwards the small houses in the study area are marked as informal settlement and are assigned as 'low' incomes. Then the bigger houses with green areas around the house are assigned to the 'high' incomes. For the houses that did not fit into these criteria an educated guess is done.

Figure D.1: *Typical structure of compounds (Google, 2013)*

D.3 Background figures of the cholera model

This section contains some figures that present some background knowledge for the cholera model.

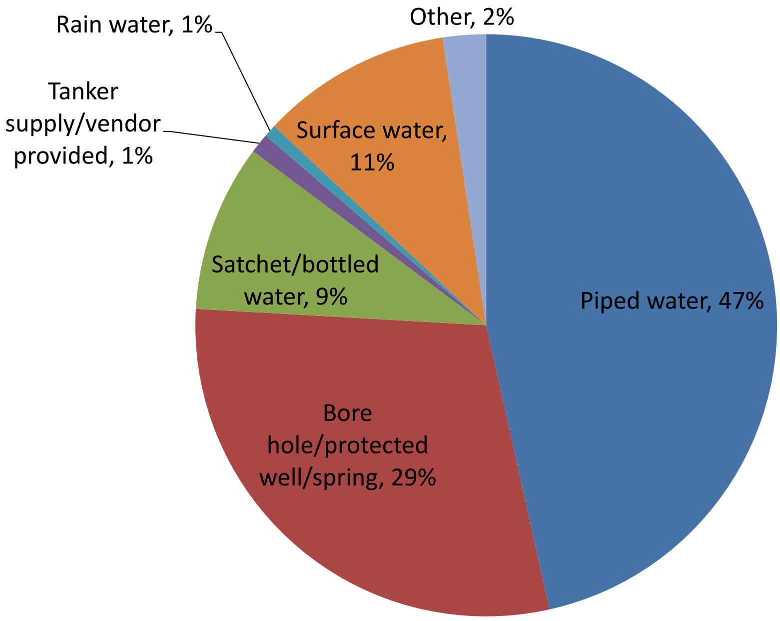


Figure D.2: *Source of drinking water (GSS, 2012)*

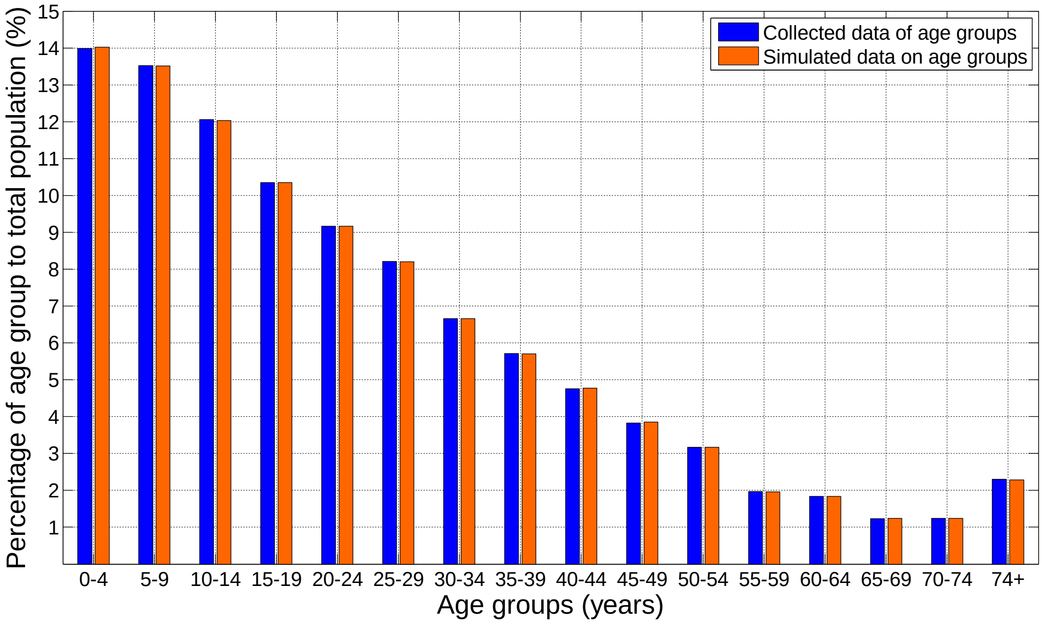


Figure D.3: *Ashanti region population divided by age and group after (GSS, 2012), next the averaged distribution of the age for the synthetic population.*

D.4 Calibration inputs

This section contains a brief explanation of the parameters ranges used in the calibration, furthermore the function of the parameter in the cholera model will be elucidated. In the explanation the parameter symbols are used, see table D.3.

It is assumed that the two switch points between of the hygiene levels do not exceeds 50%, therefore the range for HL_{lm} is set 0-50% and HL_{mh} 50-100%. $\#R$ is a arbitrary parameter, the range is set to 1-8 raindrops per hour when it rains. The parameters R_f and R_c are indirect parameters and make it possible that within the eating procedure the individual get infected, therefore the range will vary between 0-20%. R_{hh} determines the chance for HH infection, this is only possible when one of the household members is already infected, therefore this risk gets also the range 0-20%. D_{imax} is also an arbitrary parameter it is the threshold value before it is possible that individuals get contaminated via HEH infection. The range for this parameter is therefore large 1-500 wastes dumped by households. Each day households with infected family members will dump 1 waste at their dump site. P_{fw} whether people who have no access to tap water have to fetch or buy water. Households with a low income have to fetch water in this case, the middle and high incomes have a chance that they will have enough money 50% and 75% to buy water. P_{fw} will reduce both values, in case P_{fw} is negative it will be added up to these values. The range of parameter P_{Feh} will be set to 0-20% and D_f 100-500m. The parameters P_{if} , P_{sh} and P_{cw} are applied in the same way in the cholera model as P_{fw} , these ranges are set to 0-30%. The range for T_{wp} is set from S_{gr} to 5000 accumulated cells, because this range will represent the start of the main river of which it is assumed that it is also possible to fetch water. It is known that the tap water does not work with intensive rainfall however it at which intensity the tap water stops flowing, therefore the range for P_{max} is set to 1-19 mm/day.

Table D.3: *Calibration ranges of the cholera model parameters*

Explanation of symbol	Symbol	Min	Max	Unit
Swith point low to middle Hygienelevel	HL_{lm}	0	50	%
Swith point middle to high Hygienelevel	HL_{mh}	50	100	%
Number of raindrops on dumpsite	$\#R$	1	8	
Risk of food getting infected by flies	R_f	0	20	%
Risk of children getting soiled hands	R_c	0	20	%
Risk of Human to Human infection	R_{hh}	0	20	%
Dump site infection level	D_{imax}	1	500	
Duration of patch contamination due to raindrops	T_c	1	10	Days
Probability that a household has to fetch water	P_{fw}	-20	20	%
Probability of fetching water contaminated with EH	P_{Feh}	0	20	%
Maximum distance for flies from a dump site to a household	D_f	100	500	m
Probability to get infected due to contaminated food	P_{if}	0	30	%
Probability to get infected due to soiled hands	P_{sh}	0	30	%
Probability to get infected by drinking contaminated water	P_{cw}	0	30	%
Threshold value for waterpoints	T_{wp}	517	5000	
Threshold rain before the tap water stops working	P_{max}	1	19	mm/day

D.5 Calibration results

Table D.4: *Final parameter values*

Parameter	Symbol	Value	Unit
Swith point low to middle Hygienelevel	HL_{lm}	45	%
Swith point middle to high Hygienelevel	HL_{mh}	71	%
Number of raindrops on dumpsite	$\#R$	5	
Risk of food getting infected by flies	R_f	3	%
Risk of children getting soiled hands	R_c	6	%
Risk of Human to Human infection	R_{hh}	0.91	%
Dump site infection level	D_{imax}	22	
Duration of patch contamination due to raindrops	T_c	5	Days
Probability that a household has to fetch water	P_{fw}	7.2	%
Probability of fetching water contaminated with EH	P_{Feh}	5.8	%
Maximum distance for flies from a dump site to a household	D_f	229	m
Probability to get infected due to contaminated food	P_{if}	9	%
Probability to get infected due to soiled hands	P_{sh}	12	%
Probability to get infected by drinking contaminated water	P_{cw}	5.4	%
Threshold value for waterpoints	T_{wp}	585	
Threshold rain before the tap water stops working	P_{max}	2.3	mm/day