Combining Epidemiological Models with Fate and Transport Models for Waterborne Pathogens

Undergraduate Honors Thesis

By Yukinobu Tanimoto
Advisers Steven C. Chapra, David M. Gute

Tufts University
Civil and Environmental Engineering Department
Medford, Massachusetts
May 2014

Abstract

Billions of people around the world today lack safe access to clean water sources and improved sanitation facilities. Many of these people also meet their daily water needs from surface waters that are susceptible to open defecation; thus, there is a great risk that these waters may be contaminated by waterborne pathogens.

We have combined two well-established models: a pathogen fate and transport model and an epidemic model to predict the outbreak and progression of diseases caused by waterborne pathogens along an urbanized river channel. The fate and transport model predicts the transport and evolution of the pathogen in the river system, and the epidemic model predicts the outbreak of the disease once populations along the river have ingested that contaminated water. The communities then act as pseudo-incubators for the disease, effectively increasing the amount of pathogen in the river channel. A combined model provides a more holistic view of the waterborne infectious disease paradigm through the inclusion of a river and a human population component.

We provide a case study for this model by examining the Cholera outbreak in Haiti in October 2010, and calibrating the model to the Artibonite River that runs through Haiti. This case study has provided confirmation of our model results to a certain extent.

The model can serve as a decision support system to determine the best management practice and public health interventions, and also may be used to in guiding response to bioterrorism attacks. If used effectively, these hydroepidemiological models will lead to improved access to safe water and sanitation worldwide by serving as a tool to educate and guide decision making for water resource engineers and public health practitioners alike.

Acknowledgements

I would like to thank my phenomenal advisors Professors Steve Chapra and David Gute for their insights, guidance, and reassurance throughout this whole process. This is the largest project that I have worked on in my life, and I could not have gotten here without them. Through the duration of this work I've not only learned about the topic of hydroepidemiological models, but also about research methods, how to effectively convey a message, and most importantly how we can contribute to society as engineers.

I would also like to thank Professor Daniele Lantagne and her colleges at the CDC for all their help with the Haiti case study. A lot of the necessary information for the case study would have been very difficult to obtain without her first-hand experience with Haiti. Also, thanks to all my fellow classmates at Tufts who have supported me and offered their two cents, especially Carel Voltaire (E'14) for translating the Creole documents for the case study, Emi Komatsu (A'15) for her help and insight regarding the thesis proposal, and Abby Barker (E'14) for her help and feedback throughout the writing process.

I owe a lot to Mr. John A. Cataldo (E'46) and the Cataldo Research Scholarship through the Department of Civil and Environmental Engineering for giving me a reason to start this research. The experience has been extremely rewarding, and I truly appreciate the opportunity.

Thank you all very much,

Yukinobu Tanimoto

Table of Contents

List of Figures

Chapter 1. Introduction

1.1 Current Situation of Waterborne Diseases Globally

The WHO estimates that approximately 10% of the total burden of disease worldwide could be prevented by improvements in drinking water quality, better sanitation practices, and water resource management (WHO 2008). Water related diseases such as intestinal nematode infections, lymphatic filariasis, schistosomiasis, dengue fever, and cholera are a few examples of the many water-related pathogens that still affect billions of people around the globe (WHO 2008). In 2011, it was reported that while an approximate 64% of the world population relied on improved sanitation facilities, 2.5 billion people still lacked access to such a resource (WHO 2013). Furthermore, an estimated 185 million people rely heavily on surface waters i.e. rivers, streams and lakes to meet their daily water needs. 15% of the world population still practices open defecation, where human feces are deposited into fields, forests, open bodies of water, and other open spaces (WHO 2013).

Although these statistics demonstrate the scale of access problems to water and related sanitation services, there have been notables examples of success such as Vietnam where open defecation rates have dropped from 40% in 1990 to 3% in 2011 (WHO 2013). The WHO predicts that significant reductions in diarrheal frequency can be expected from a combination of specific interventions that target areas such as hygiene, sanitation, water supply, and water quality (WHO 2008).

The remaining questions are why these issues have not been solved when solutions are known; and further why global progress is off track to meet the sanitation targets outlined in the Millennium Development Goals. One of the main reasons is simply, cost. The appropriate

interventions have been identified, but it is economically and politically unfeasible to implement programs for all populations that would beneficial. Such constraints require the development of decision support tools to assist policy makers in determining the best management practices on a case by case basis. This thesis will explore how computer models and similar system oriented frameworks would be able to provide improved insight into water resource management practices as a result of a comprehensive and holistic perspective (Chapra 2011).

1.2 Background on Traditional Models

Environmental Fate and Transport Model

Environmental Fate and Transport Models, or water quality models are mathematical models that predict the behavior of pollutants in environmental systems. These models illustrate the response of a system that is stressed, via loading of a pollutant and other stimuli, while factoring in the physics, chemistry, and biology of the entire system (Chapra 2011). Since the first water quality model, developed by W. Streeter and E. Phelps in the 1920s, which was designed to predict the critical dissolved oxygen concentration downstream of a defined pollution source (measured concentration, as opposed to a calculation concentration), much progress has been made in the field (Streeter and Phelps 1925).

With advancements in hardware and available computational power, more complex models with non-closed form expressions are presently used for water quality modeling. Following in Streeter-Phelps' paths, models were developed for dissolved oxygen, bacteria, nutrients, organics, metals, acids and many more pollutants. Simultaneously, models became increasingly flexible, offering capabilities to handle 1D, 2D, and 3D, as well as advanced integration of both the artificial and natural components of the model (Chapra 2011).

While the modeling of different pollutants has a long and rich literature, what is still lacking today is the modeling of waterborne pathogens (Chapra 2010). This may result, in part, because in most developed countries the civil infrastructure is well established so that the drinking water quality is adequate, and the modeling of waterborne pathogens is no longer a top priority. For example, in the United States, the last major waterborne pathogen incident was the cryptosporidium outbreak in Milwaukee, Wisconsin in 1993 (Mac Kenzie et al. 1994), which was 20 years ago.

Epidemiological Model

Roughly about the same time as Streeter-and Phelps developed the dissolved oxygen model, W.O. Kermack and A.G. McKendrick in Edinburgh developed a model similar to Lotka–Volterra predator-prey models that compartmentalize a population into disease status categories to predict the spread of an epidemic (Kermack and McKendrick 1927). The so called SIR model in its most basic form distributes the population into three compartments, namely: susceptible healthy individuals (S), infected individuals with the disease (I), and recovered capita (R). The model was originally intended to be run with the initial condition being that a set number of infected individuals would enter the community or contract the transmittable disease, and based upon such interactions, more of the susceptible population would become infected and move into the infected compartment. With time the infected population would either recover and get placed in the recovered compartment; die and be removed from the total population; or in a disease where an individual does not acquire immunity and thus can be affected more than once, get placed back into the susceptible group. In most cases, over time the system will reach equilibrium state where the number of individuals in each of the compartments achieves steady state.

Although the SIR model in its most fundamental and basic form provides adequate insight into the spread of an epidemic, there are two ways that the model can be modified: 1) further divide the three compartments and 2) to alter the infection source.

Mwasa and Tchuenche (2011) for example divided the population into seven different compartments based on common public health interventions: susceptible, educated, vaccinated, quarantined, infected, treated, and removed individuals. While this scheme may seem to more accurately describe the risk status of an actual population, quantifying the rate at which individuals move between the different classes may not be trivial. For example, both educated and susceptible capita may opt to get vaccinated, but it is difficult to quantify the rate at which people are vaccinated as that depends on many variables (accessibility of vaccine, price, or if the vaccination extends beyond a single dose). One obvious advantage however with the further compartmentalization is that the model will be capable of assessing the feasibility of different public health interventions. If the best intervention can be determined by the model and the economic costs are known, it will be a superior decision support system.

A second way to modify the SIR model is to change the infection source. The classic model assumes that the infections start with an infected individual who in turn infects members within the susceptible class, but the problem found with waterborne diseases is that the infected individual is no longer the trigger to an outbreak. The waterborne pathogen will reach a community via the water source, and susceptible individuals that come in contact with the compromised water source will get infected. This water source could be a contaminated drinking water reservoir (Codeco 2001, Hartley et al. 2005, Joh et al. 2008), or a river system (Bertuzzo et al. 2007, McBride and Chapra 2011).

Hydroepidemiological Model

This thesis will couple the environmental fate and transport model and the epidemiological model to create a hydroepidemiological model for the spreading of waterborne pathogens for communities that share access to a given river. The approach inherent in the water quality model will be used to predict the behavior of the pathogen within the river channel. Various communities can be placed along the course of the river, with the assumption being that the communities withdraw water directly from the river to meet their daily needs. If the river water is infected with some waterborne pathogen such as cholera, the communities that use the contaminated water for drinking will become infected, and an outbreak of cholera in that community will result. By starting off the entire community in the susceptible class, and through the use of a dose-response relationship, we can characterize the infections in a community using a SIR model. If these communities practice open defecation so that infected hosts will discharge their excrements containing more of the pathogen back into the river, the downstream communities will be affected by the initial wave of pathogen, and also by the increased amount of pathogen resulting from the infected community upstream. Infected communities will likely function as incubators that can cause the pathogen to multiply rapidly.

The coupling of these two vastly different and well-established models will yield a model that provides greater insight into the spreading of waterborne diseases. As a decision support mechanism, decision makers will be able to evaluate the effectiveness of changing parameters in the river model, or the public health interventions. Because the two, the river model and the epidemic model, are dependent on each other, it is only logical that they be coupled together.

1.3 Literature Review

Role of Aquatic Reservoirs

Codeco (2001) expanded the classic SIR model by adding an environmental reservoir contaminated with cholera. Susceptible individuals that came in contact with the contaminated water i.e. ingestion would get infected based on a logistic dose response curve. Excretions from infected hosts would be returned to the reservoir as mass loading, and the reservoir was treated as a continuous stirred-tank reactor (CSTR).

Cholera in the aquatic reservoir died off at an extinction rate, but as Codeco reported this value has been reported to have a wide range from 0.02 to 3 day^{-1}. All the environmental factors that may influence the growth or decay of the cholera in the water are encompassed in this one parameter, which could be expanded by combining a fate and transport model. Codeco also explores the possibility of oscillating 1) the contact rate of the susceptible population to the water; 2) the contamination rate per capita; and 3) the cholera extinction rate. The oscillations explored by Codeco were based on a sinusoid curve with a period of one year, and this of course could be done more accurately and at a higher temporal resolution by considering the population's water usage habits and parameters on a daily basis. Codeco's research pioneered the coupling of an environmental reservoir of pathogen with an SIR model, where the source of the infection comes from outside the SIR compartments.

Since then there have been many others (e.g. Bertuzzo, Hartley, and Chapra) that have expanded on Codeco's seminal work to better model the spread of waterborne pathogens. These advances are described in the following section.

Space-Time Evolution

Bertuzzo et al. (2007) expanded Codeco's framework and replaced the aquatic reservoir with a river network system to give the model a spatial dimension. Bertuzzo used an orientated graph to simulate the river network, but the general premise that susceptible individuals would get infected after drinking the contaminated water, and that infected individuals would shed the pathogen back into the river was the same as Codeco's. This model was also expanded so that communities along the river could also receive drinking water that was not obtained from the river such as a rainwater collection system, or a nearby stream that was disconnected and distinct from the infected network. Bertuzzo was able to produce satisfactory results by calibrating his model to the 2000 cholera outbreak in the KwaZulu-Natal province of South Africa (Bertuzzo et al. 2007).

Bertuzzo recognized however that there were some factors that were not included such as seasonality and its impact on certain environmental parameters, and the hyperinfectivity of the cholera, which would be a focus for future research. Similar to Codeco's model, Bertuzzo's model, although adding a spatial dimension, encompasses all the environmental parameters into a single in-stream attenuation rate, and there is a need for a model with the added spatial dimensionality and also a complete environmental fate and transport model.

Hyperinfectivity

Hartley et al. (2006) expanded Codeco's model by adding in the attribute of hyperinfectivity for cholera. Laboratory experiments and observations have shown that freshly excreted, or shed, cholera is much more infective compared to cholera that has been outside the host for a longer period of time due to the deleterious impact of the ambient environment. Experiments on mice have shown that freshly shed hyperinfective cholera can out infect lab grown cholera by as much

as 700 times. Hyperinfectivity only lasts only about 1 day, after which time a traditional dose response curve can be used to characterize infections. This is most important in cases where open defecation is practiced and proper sanitation and hygiene measures are not implemented, as individuals that come into contact with hyperinfective cholera have a much higher probability of getting infected.

Hyperinfectivity provides a valuable lesson for disease control activities. This being that even if reducing the contact rate with the contaminated water is impossible (i.e. a population requires it as a water source), simple hygiene steps to keep freshly shed cholera quarantined could help keep the cholera outbreak from becoming explosive.

River Model for Agricultural Watersheds

McBride and Chapra (2011) developed a hydroepidemiological model for indicator organisms and zoonotic pathogens in agricultural areas. The chosen or index zoonotic pathogen was campylobacter, which can be spread both by vectors such as flies, and also through ingestion of contaminated water. The McBride and Chapra model, much like the aforementioned models, only has a constant in-stream attenuation rate to account for the decay of the pathogen in the water.

Although this model was developed to model the spread of an infection for cattle animal farms, the design is similar to the hydroepidemiological model proposed in this manuscript. This framework has inputs for the stream hydrology, which will influence the physics that drives the computation of the model. The model also allows for farms to be placed at arbitrary reaches along a river channel, and a similar design will be adopted for placing human communities along the river in this model.

Chapter 2. Model Methodology

2.1 Environmental Fate and Transport Model

River Set Up

The river channel is modeled by segmenting the river into a sequence of CSTR cells. The river channel as a whole should act as a plug flow reactor when there is little dispersion, but because the river is split into CSTRs, this causes a smearing in the concentration of the pathogen within each reactor cell. This smearing is similar to dispersion, although it is not intentionally added into the model. Hence the model river approximates a mixed-flow reactor, as both advection and dispersion govern the fate and transport (Chapra 1997).

The model is implemented so that certain reaches in the river can be modeled with reactors of smaller or larger lengths. This functionality adds flexibility the model, so that smaller tributaries and larger rivers can be modeled as a single channel. The user may choose to model the smaller tributary with smaller increments and model the larger part of the river with larger reactor lengths.

The channel geometry in the model can be specified manually, or estimated through rating curves and discharge coefficients. If a user knows the channel flow, velocity, width, or depth, they can manually input those values for the respective reaches to accurately calibrate the river model. However, in many countries where there is a lack of information regarding the hydrogeometry, discharge coefficients can be used to estimate the velocity, height, or width, if the channel flow is known (Leopold and Maddock, 1953). Three equations proposed by Leopold and Maddock estimate the velocity (U), mean depth (H), and width (B) of the channel:

$$U = aQ^b$$

$$H = \alpha Q^\beta$$

$$B = cQ^f$$

Empirical constants $a, b, \alpha, \beta, c, f$ are estimated from log-log plots of the two respective parameters in each of the equations and are therefore related, such that the summation of the exponents should be equal to unity (Chapra 1997).

The model can also be easily modified to handle Manning's equation combined with the continuity equation ($Q = UA$) to represent the channel geometry if the user so chooses.

Water Balance

The model river requires a boundary flow condition to initiate the river flow. The boundary flow rate is applied to the very upstream cell, and the flow coming out of the first cell is the flow going into the second cell, and so on. For all cells, the volumetric outflow of water for that specific cell is calculated with a flow balance. This outflow is then the amount of water that flows into the next cell. Knowing the volumetric flow of water going in and out of each cell is crucial for solving a backward space mass balance equation.

The model is also designed to handle point source withdrawals and discharges; diffuse withdrawals and discharges; and the water withdrawals and discharges from cities along the river channel. Point source withdrawals and discharges either remove or add water at a specified location along the river. The model determines the cell where the point source is located, and removes or adds water from that cell. If a point source is specified at the boundary of two cells, it is placed into the downstream cell.

Diffuse withdrawals and discharges occur over a specified length along the river channel. The starting and end locations can be specified, along with the total flow rate for the entire

distributed source. The model determines the corresponding starting cell and ending cell for the diffuse source and distributes the flow accordingly. For example, if a diffuse source is 1 cell long and spans across 2 cells and starts halfway in the first cell, the model will evenly distribute the flow so that half the flow is in the first cell, and the second cell gets the other half. The model determines a flow per unit length for the diffuse source and also how much (physical length) of the reactor is within the diffuse source to distribute the flow accordingly.

Discharges and withdrawals for cities are handled similarly to point sources and diffuse sources. A city in the model may withdraw and discharge water through a point source, or a diffuse source. The amount of water required by a city's population can be estimated with a few parameters regarding water consumption.

$$Q = pop * PC$$

where *PC* is the per capita total water use.

Calculating the discharges from the city requires additional parameters, and can be estimated as:

$$Q = pop * (PC - PCh) * (1 - Fcg) + Pch * (1 - Fch)$$

where

Pch is the per capita human consumption,

Fcg is the fraction used for grey water, and

Fch is the fraction used for human consumptive use.

As mentioned previously, the flow requirements and discharges for cities can be treated as point or diffuse withdrawals and sources. Point sources may better model communities where a single pipe drawn from the main river provides water for the community, and a diffuse source may be more appropriate for situations where individuals directly collect water or discharge wastes.

Loadings

Loadings represent the discharge of pathogens to the river, effectively increasing the concentration of the pathogen in the water channel. The loadings from the various sources (point, diffuse, city, or boundary condition) are stored in an array where each cell corresponds to a river reactor cell. By having a specific array for loading, we can take the loadings from the different sources and spatially distribute them accordingly. Some of the loadings are constant, such as point sources and diffuse sources, while others such as city discharges and boundary condition loadings can be time dependent.

Point source loadings are specified by a reach (location along the river), a flow rate, and a pathogen concentration. The loading is calculated as the product of the flow rate and the pathogen concentration to yield a mass per time rate. This loading rate is applied to the differential equation to calculate the pathogen concentration for that cell.

Diffuse source loadings are specified by a starting and ending reach, a flow rate, and a pathogen concentration. The model calculates the fraction of the flow going into reach of the reactors, and the loading is a product of this distributed fractional flow rate and the pathogen concentration. By distributing the flow evenly throughout the range of the diffuse source, we ensure that the mass of pathogen is distributed evenly.

Boundary condition loadings are used to initiate a flow of a specific mass amount of pathogen for a specified time frame. Thus the boundary condition for loading requires 2 vectors of the same lengths (or a table, which is essentially two vectors side by side); where one column would have discrete times and the corresponding element in the other column will have the concentration of the pathogen at that time. As the model runs the simulation through time, the

boundary condition table can be read by the model with linear interpolation to determine the boundary concentration at a specific time step.

The loading from a city can be expressed as a function of the number of infected individuals in the city. If the number of infected individuals and a shedding rate (the rate at which an infected individual excretes the pathogen) is known, the total mass of the pathogen cumulating from all the infected individuals can be calculated as

$$m_{shed} = infected * \rho$$

where ρ is the shedding rate, or rate of excretion in units of organism/capita/time.

This mass can then be divided by the total discharge of water from the city as discussed previously. The discharge may be simulated as a point discharge or a diffuse discharge and handled with the same techniques mentioned earlier in this chapter.

Decay

The mass balance of pathogen over time can be expressed as a differential equation. By numerically integrating this differential equation, we can calculate the concentration of the pathogen over time as flows, loadings, and the decay rate changes. A differential equation can be written for each of the reactor cells, but it is obvious that the cells are coupled because the differential equation relies on the flow and concentration coming out of the adjacent upstream reactor. Because transport depends on the upstream element, this type of modeling is called a backward space difference model.

$$V\frac{dC}{dt} = Q_{in}C_{in} - Q_{out}C - kVC + W$$

The decay of the pathogen is captured in the term *k*, which is a combination of settling, attenuation by natural mortality, and by light effects (Chapra 1997).

The settling effects depend on how pathogens sorb onto sediments in the river channel. We can calculate the fraction of the total pathogen that sorbs into the sediment by assuming local equilibrium. Using a linear sorption isotherm, the fraction of the total pathogen that is attached to particles can be calculated as (Chapra 1997):

$$F_p = \frac{k_d m}{1 + k_d m}$$

where

k_d is the linear partition coefficient, and

m is the concentration of the suspended solids.

The total loss due to settling then can be represented as a combination of sorption and settling in each of the idealized cells. The overall settling rate can be expressed as (Chapra 1997):

$$k_{bs} = F_p \frac{v_s}{H}$$

where

v_s is the settling velocity of the particles, and

H is the mean channel depth.

Natural mortality can be calculated as a function of salinity and temperature (Mancini 1978, Thomann and Mueller 1987), but if neither of these parameters is known, a single in-stream attenuation value can be used instead.

The decay due to light can be expressed as a function of the light energy and depth of the channel (Thomann and Mueller 1987). The Beer-Lambert law can be used to calculate the extinction of light in a mixed water body as follows:

$$I(z) = I_0 e^{-k_e z}$$

where

I_o is the incoming light energy,

k_e is the extinction coefficient, and

z is the depth measured below the surface.

The extinction coefficient is a function of the channel sediments and calculated as (Di Toro et al. 1981):

$$k_e = 0.55m$$

The Beer-Lambert law can be integrated over the depth of the channel to yield:

$$k_{bi} = \frac{I_0}{k_e H}(1 - e^{-k_e H})$$

The light energy intensity will vary daily depending on the physical location on Earth and time of year, and will also vary hourly depending on the position of the sun in the sky. Specifics on calculating the sunrise and sunset, and the insolation are listed in Appendix A.

2.2 Bridging the Two Models

As explained earlier, the mass of pathogen entering a city can be computed as either a point or a diffuse withdrawal. However, we need a mechanism to translate this mass of pathogen to an infection probability for the epidemic model.

Dose-response assessment is considered to be an integral part of risk assessment. Dose-response relationships analyze and describe the relationship between the amount of toxin that an individual is exposed to, and the probability that individuals will get infected, or show a response to the pathogen. These relationships are derived by fitting curves to experimental data regarding a dose of a toxin and the response of individuals. One important factor in dose-response relationships is the threshold, a certain level of toxin where the individual does not show any

adverse effects (NIH 2002). This is mainly because the body is able to handle, or break down and destroy, a certain amount of the toxin.

Hass et al (2000) proposed a general Beta-Poisson equation to model the dose-response of waterborne diseases that has two parameters, the D_{50}, which is the dose at which 50% of individuals will get infected, and the slope factor α. The equation is as follows, where P_I is the probability of an individual getting infected when they come in contact with contaminated water with a concentration D of the pathogen:

$$P_I = 1 - [1 - \frac{D}{D_{50}}\left(2^{1/\alpha} - 1\right)]^{-\alpha}$$

From observation of the equation, it can be seen that as α reaches infinity, the function will resemble an exponential function. If α is decreased, the dose response becomes more linear with less of an obvious threshold value. A more linear response function would mean that the effect of each subsequent dose of toxin is much less compared to a dose response function with a threshold because the derivative of the function never gets steep.

Codeco (2001) proposed a logistic dose response function that is similar to a Michaelis–Menten model for cholera. This dose response function only requires the dose and the D_{50} in the following way:

$$P_I = \frac{D}{D + D_{50}}$$

However, with the case of waterborne pathogens such as cholera, it is not known if the inoculum size affects the case-to-infection ratio, or if other parameters such as the condition of the host; the condition of the cholera, or whether it is hyperinfective or not is more important (Nelson et al. 2009, Hartley et al. 2006).

By using dose-response relationships, we can characterize the probability of an infection based on the dose, which is a function of the concentration of the pathogen in the water being consumed. Per capita water consumption values can easily be estimated or taken from the literature, and if the concentration of the pathogen is known, the mass of the pathogen ingested can easily be calculated.

2.3 SIR MODEL

The SIR model is a slight modification of the model proposed by Codeco. The three differential equations can be written as:

$$\begin{Bmatrix} \frac{dS}{dt} = -k_i S(t) \\ \frac{dI}{dt} = k_i S(t) - (k_r + k_m) I(t) \\ \frac{dR}{dt} = k_r I(t) \end{Bmatrix}$$

where

k_i is the infection rate by which individuals move from the susceptible class to the infected class,

k_r is the recovery rate by which individuals move from the infected class to the recovered class, and

k_m is the mortality rate in which infected individuals do not survive and are taken out of the population.

This model assumes that reinfection is not possible, as individuals in the recovered class can never get assigned to another class. It can be observed that when the morality rate is zero ($k_m = 0$) the total population remains constant.

The infection rate k_i is calculated through a dose response function. The concentration of the water being ingested can be calculated by the fate and transport model. If we know how much of this water an individual drinks over a period of time (e.g. a day), we can calculate the mass of pathogen that is ingested on a per time basis and relate that to a probability for infection.

2.4 Numerical Methods

As analytical, closed form solutions are not available for our 4 differential equations, we need to use numerical methods to approximate the result. One family of methods is called Runge-Kutta methods, after two German mathematicians in the early 1900s (Chapra and Canale 2009). Runge-Kutta methods have varying degrees of complexities and accuracies, but they all have the general form:

$$y_{i+1} = y_i + \varphi h$$

where a slope estimate of φ is used to extrapolate a new value of y_i over a step size of h.

With higher order Runge-Kutta methods, the slope estimates get more complex and refined, reducing the associated error. The 1st order Runge-Kutta method is known as Euler's Method and uses a single derivative estimate as the slope. Mathematically, if we have a derivative function f which is dependent on both y and t, the derivative at a specific point can be used as the slope estimate such that:

$$\varphi = f(y_i, t_i)$$

Substituting this into the generic Runge-Kutta equation, we have:

$$y_{i+1} = y_i + f(y_i, t_i)h$$

While Euler's Method will give us a decent approximation, the estimation can be crude at times because only 1 derivative is used, and especially if the step size h is too large.

The present model uses a 4th order Runge-Kutta method, otherwise known as the Classical Fourth-Order Runge-Kutta method (RK4), which is mathematically expressed as:

$$y_{i+1} = y_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)h$$

where the k_i's are slope estimates within the step size, and can be expressed as:

$$k_1 = f(t_i, y_i)$$

$$k_2 = f\left(t_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1 h\right)$$

$$k_3 = f\left(t_i + \frac{1}{2}h, y_i + \frac{1}{2}k_2 h\right)$$

$$k_4 = f(t_i + h, y_i + k_3 h)$$

By taking a weighed estimate of the 4 slopes, we can get an improved estimate of the slope needed to solve the differential equations. Graphically the classical Runge-Kutta method can be depicted as (Chapra and Canale 2009):

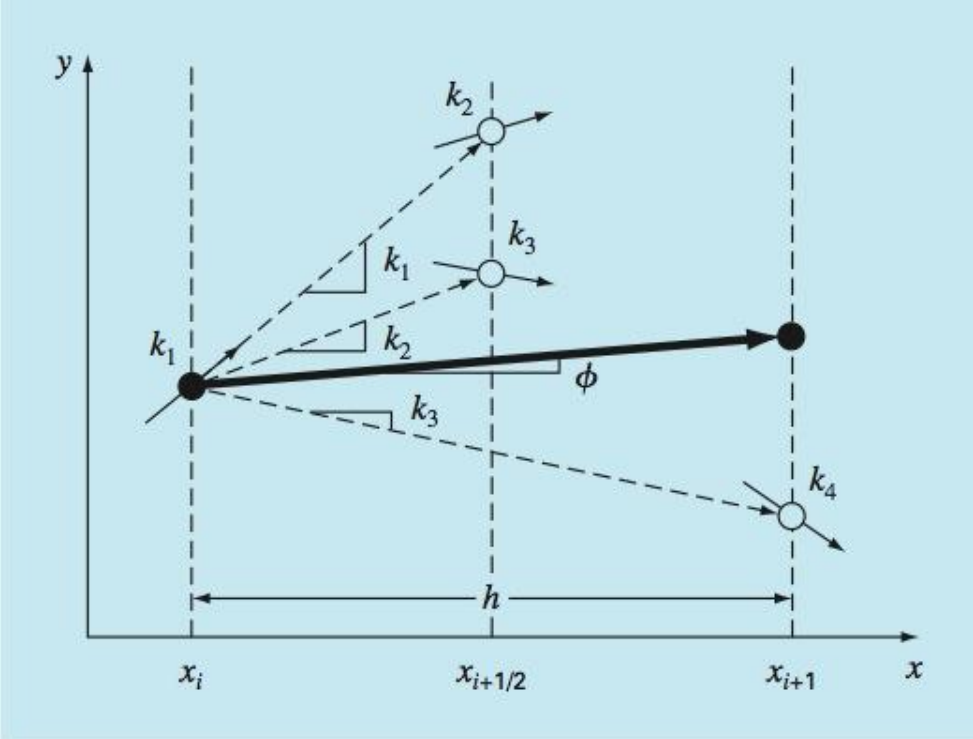

Figure 1 Graphical Representation of RK4

The difference between the results of Euler's method and RK4 can be seen in the below figure, which shows the result of the model run with the two integration techniques for a point source. Note that the figure has been zoomed in closely to show the small deviation in RK4 compared to Euler's Method.

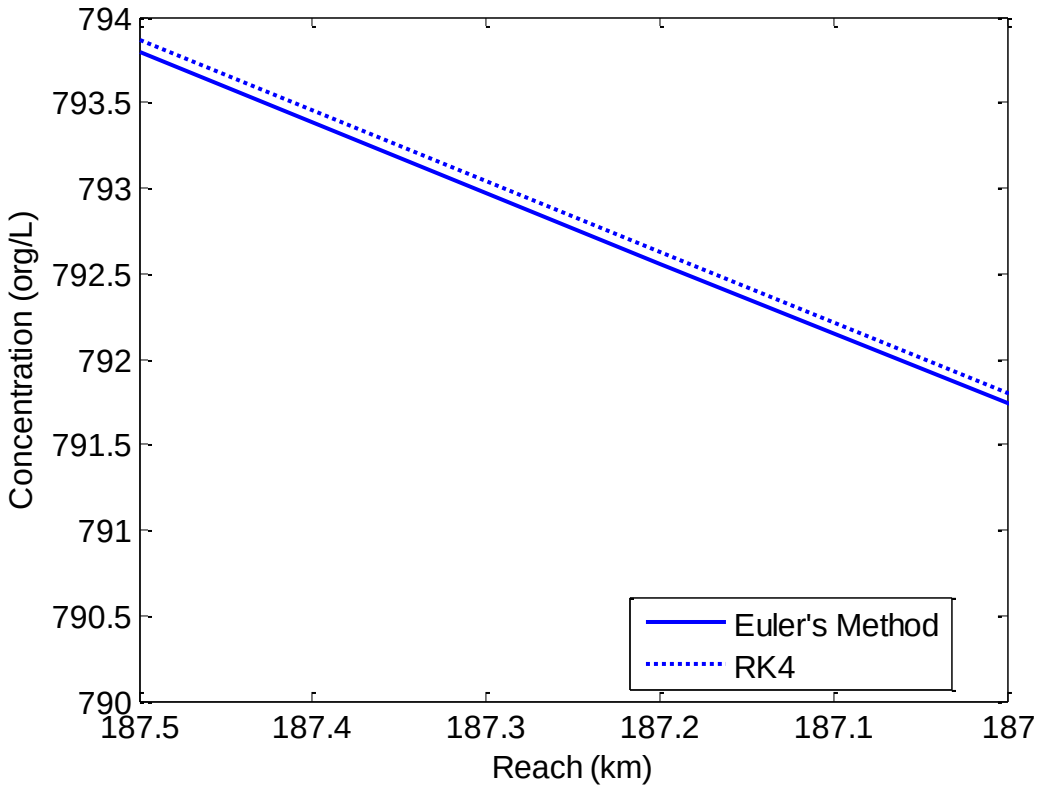

Figure 2 Euler's Method vs RK4

Chapter 3. Verification and Analysis

This chapter will focus on the verification of the model as well as sensitivity analysis. Verification of the model will be achieved by comparing the numerical integration with RK4 to an analytical, closed form solution. The analytical solution provides us with the true result, allowing us to verify our model calculations and numerical integration. The sensitivity analysis will focus on altering the flow in the river channel, and will illustrate the importance of the flow in pathogen transport.

3.1 Verification of Numerical Integration by Analytical Solutions

Point Source Loading

The analytical solution for a point source loading with advection and dispersion is as follows (Chapra 1997):

$$c = \frac{W}{Q\sqrt{1+4\eta}} e^{\frac{U}{2E}\left(1+\sqrt{1+4\eta}\right)x} \quad x \geq 0$$

where

E is the dispersion coefficient,

x is the location downstream, and

η is the modified decay term, $\eta = k_d E/U^2$.

E in our case is limited to numerical dispersion, and can be calculated as:

$$E_n = \frac{\Delta x}{2} U - \frac{\Delta t}{2} U^2$$

where

Δx is the length of each reactor,

Δt is the user specified calculation time increment, and

U is the channel velocity at that reactor.

The analytical solution can be plotted alongside the numerical solution as in the figure below:

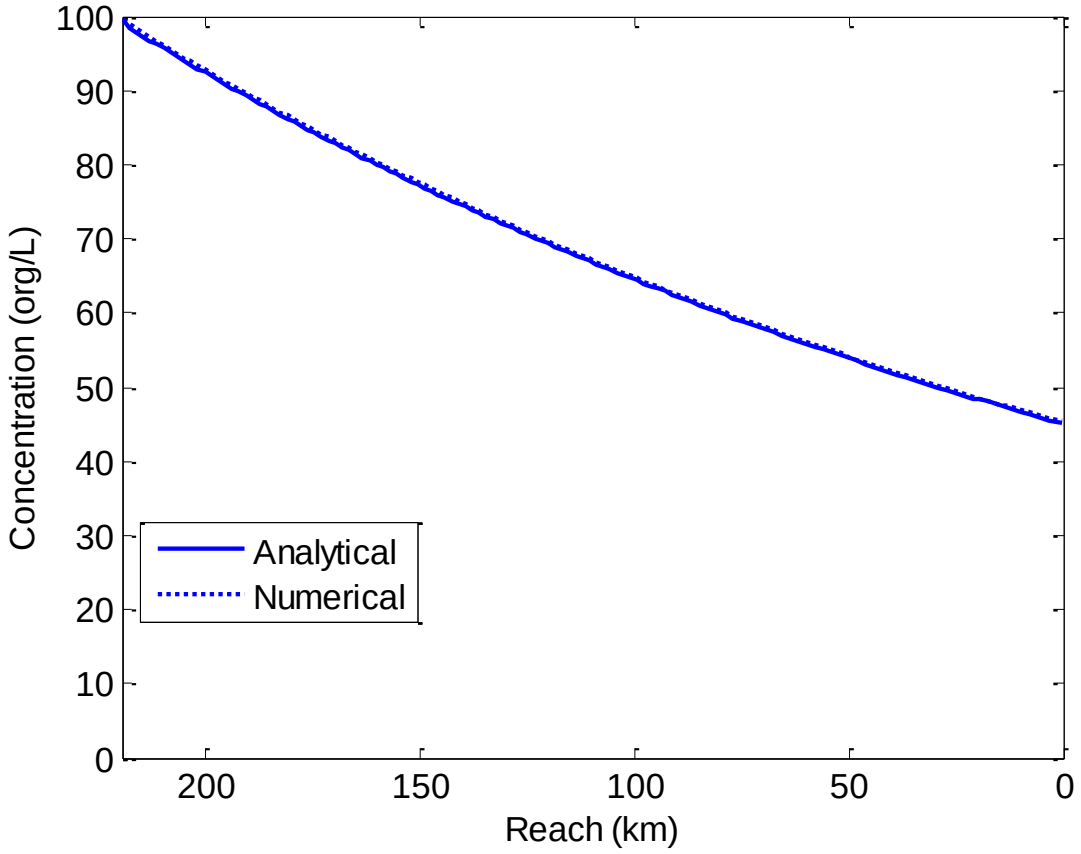

Figure 3 Point Source Verification

From the figure, the analytical solution and the numerical solution align so well that it is difficult to distinguish between the solid and the dotted line.

Finite Time Loading

The boundary finite time step input computation can be compared against the closed form solution for a system with advection and dispersion (O'Loughlin and Bowmer 1975).

$$c(x,t) = \frac{c_0}{2}\left\{\begin{array}{l} e^{\frac{Ux}{2E}(1-\Gamma)}\left[erfc\left(\frac{x-Ut\Gamma}{2\sqrt{Et}}\right) - erfc\left(\frac{x-U(t-\tau)\Gamma}{2\sqrt{E(t-\tau)}}\right]\right. \\ \left. +e^{\frac{Ux}{2E}(1+\Gamma)}\left[erfc\left(\frac{x+Ut\Gamma}{2\sqrt{Et}}\right) - erfc\left(\frac{x+U(t-\tau)\Gamma}{2\sqrt{E(t-\tau)}}\right]\right. \end{array}\right\}$$

where

E = the dispersion coefficient, numerical equation as aforementioned,

$\eta = k_d E/U^2$,

$\Gamma = \sqrt{1+4\eta}$,

τ = the end of the step input, assuming that the input begins at t=0, and

erfc is the error function complement: 1 – *erf*.

By simple inspection of the term in the equation:

$$2\sqrt{E(t-\tau)}$$

We can see that for $t \leq \tau$, when the loading is still occurring, this term yields an imaginary number, or leads to a 0 in the denominator. To correct for this, the closed form solution needs to be altered to be time dependent, so that while the loading is still occurring the solution is for a continuous spill model.

$$c(x,t) = \begin{cases} t \leq \tau & \frac{c_0}{2}\left[e^{\frac{Ux}{2E}(1-\Gamma)} erfc\left(\frac{x-Ut\Gamma}{2\sqrt{Et}}\right) + e^{\frac{Ux}{2E}(1+\Gamma)} erfc\left(\frac{x+Ut\Gamma}{2\sqrt{Et}}\right)\right] \\ t > \tau & \frac{c_0}{2}\left\{\begin{array}{l} e^{\frac{Ux}{2E}(1-\Gamma)}\left[erfc\left(\frac{x-Ut\Gamma}{2\sqrt{Et}}\right) - erfc\left(\frac{x-U(t-\tau)\Gamma}{2\sqrt{E(t-\tau)}}\right]\right. \\ \left. +e^{\frac{Ux}{2E}(1+\Gamma)}\left[erfc\left(\frac{x+Ut\Gamma}{2\sqrt{Et}}\right) - erfc\left(\frac{x+U(t-\tau)\Gamma}{2\sqrt{E(t-\tau)}}\right]\right] \end{array}\right\} \end{cases}$$

The analytical solution can be plotted alongside the numerical solution. The figure below shows the analytical and numerical solution at downstream locations (from left to right) at 1, 10, 25, 50, 75, and 100 km from the boundary of the river, over time.

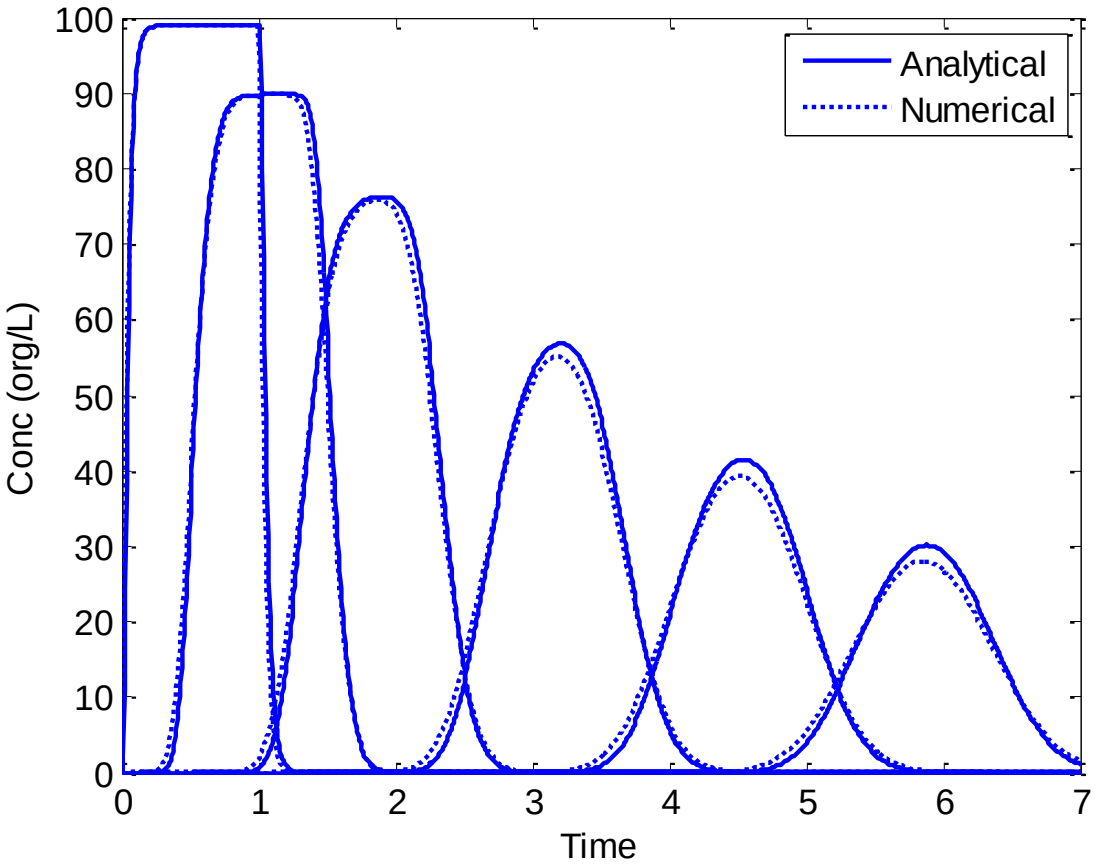

Figure 4 Finite Time Loading Verification

Similar to the plot for the point source, the analytical solution and the numerical solution align well, confirming that our numerical solution is correct. With increasing space and time, we also see smaller deviations in the two results, which can be attributed to numerical dispersion and smearing in each of the river reactor cells.

3.2 Flow Sensitivity Analysis

One of the major components of the river model is the flow of the river. The flow can affect both the concentration of the pathogen, and also the travel time in the river channel. Both these factors are important for a decision maker who is trying to determine the best intervention strategy. The

concentration of the pathogen will dictate the probability of an individual getting infected after they ingest the water, as discussed in Chapter 2. The travel time on the other hand will affect how rapidly the disease would spread along the river channel, so the intervention would need to target a wider area.

In our analysis, we halved and doubled the flow in the river, based on a starting flow of 90 m^3/s. As expected, the concentration of the pathogen was much higher when the flow was lower, and lower when the flow was higher, as shown in the figure below. This simulation was run by placing communities along the river channel to replicate the situation along the Artibonite River, Haiti. Specifics on the model input will be discussed in further detail in Chapter 4.

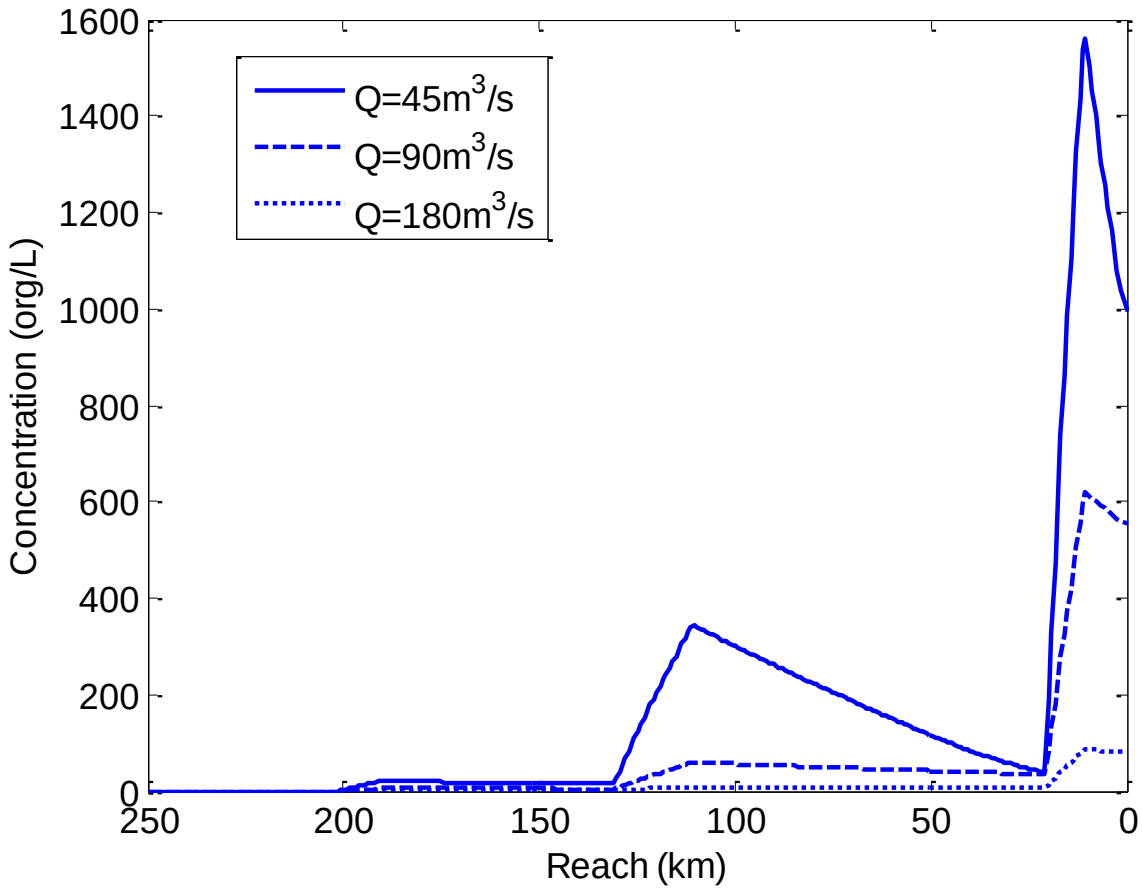

Figure 5 Pathogen Concentration at Different Flows

We can also examine the effects of alternate flows on the travel time down a river channel by plotting the calculated concentrations at a specific point over time. The figure below is a plot of the concentrations over time for a point 150 km away from the upstream boundary. A finite time step loading is applied to the system, and we can see that the pathogen is reaching this point much faster with increased flow.

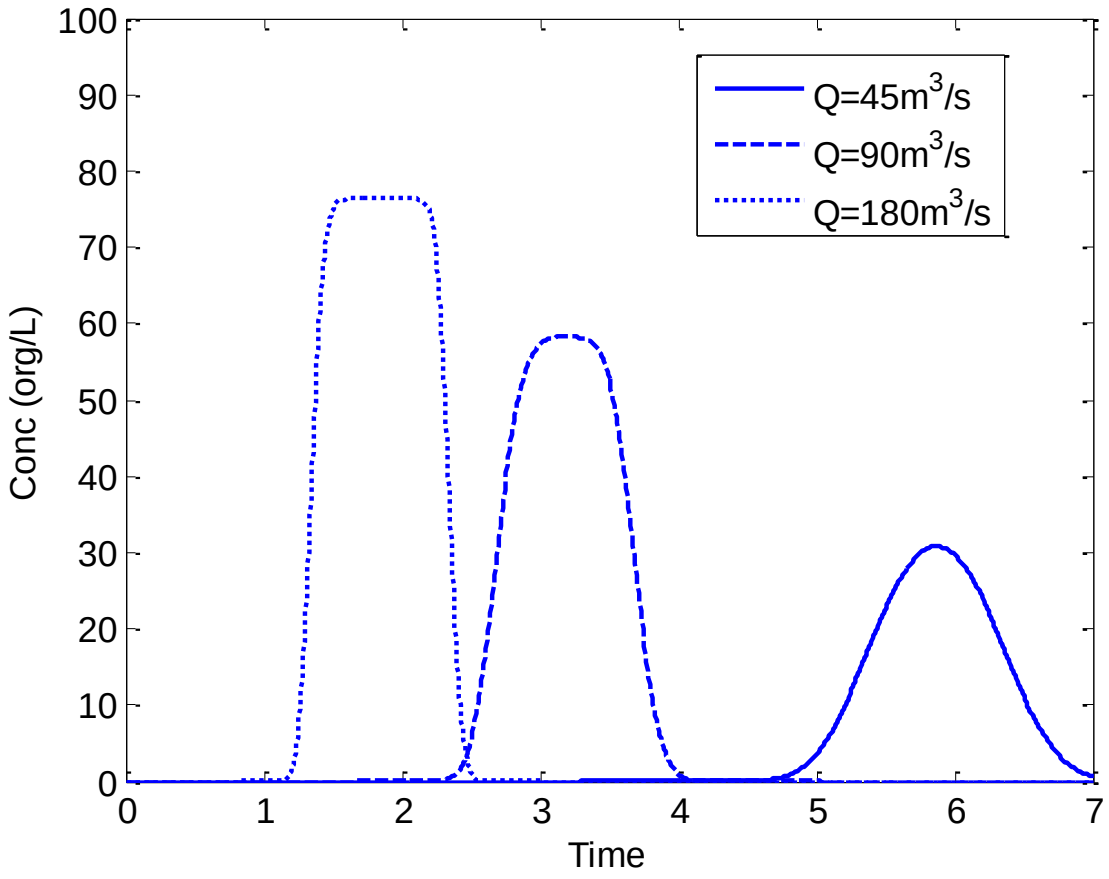

Figure 6 Pathogen Travel Times at Different Flows

The ability of the model to show the difference in concentration with different flows also speaks to the usefulness of the model as a decision support system. Different scenarios, such as opening up a floodway or dam, and their effects on the transport of the pathogen could easily be assessed by this model.

Chapter 4. Haiti Cholera Outbreak Case Study

4.1 Background

Haiti is located on the western third of the island of Hispaniola in the Caribbean Sea, and is neighbored by the Dominican Republic (Library of Congress 2006). Haiti is the poorest country in the Western Hemisphere, and according to the 2010 census, 55% of the population lives below the world poverty line of 1.25 US dollars per day (Lantagne et al. 2013). Prior to social upheaval and infrastructure destruction by the 2010 earthquake it was estimated that 52% of the rural population had access to improved drinking water, 25% of the rural population had access to improved sanitation, and only 26% of the rural households had a water supply (Office of the Secretary General 2013).

On January 12, 2010, a magnitude 7.0 earthquake struck Haiti, resulting in the death of more than 220,000 with over 2.1 million inhabitants being affected through injury, loss of livelihoods and residences (OCHA 2013). This means that approximately ¼ of the population was severely affected in that the 2010 population of Haiti was 9.6 million people. The earthquake crippled infrastructure of all types including transportation, thus making humanitarian aid efforts difficult. A U.S. Centers for Disease Control and Prevention (CDC) report in March 2010 described the situation as the, "current water, sanitation, and hygiene infrastructure in Haiti would certainly facilitate the transmission of cholera (and many other illnesses)." While it was the CDC's position that cholera was "unlikely" to appear, in October 2010 cholera was reported in post-earthquake Haiti. This was the first time that cholera was reported in Haiti in nearly 100 years (Cravioto et al. 2011).

Although the origin of the cholera epidemic has not been determined yet, it is a widely held belief in Haiti that the cholera was introduced by Nepalese MINUSTAH (US) forces stationed

near the Meye Tributary in Mirebalais. At the MINUSTAH base, human waste was contained in six 2500-L fiberglass containers, and when these containers were full, a truck would come to pump out the tank and empty the contents. The waste was then transported to a black water disposal pit near the base, well within the floodplain of the Meye Tributary. It was noted that this disposal pit was not fenced off and children and animals were often seen near the pit. The poor sanitation situation and waste management could have led to the human waste from the base getting into the Meye tributary, then into the Artibonite River to initiate the cholera epidemic. (Cravioto et al. 2011)

4.2 Vibrio cholerae

Cholera is a diarrheal disease that results in severe loss of fluid and dehydration that can lead to circulatory collapse and, if untreated, to rapid death. The disease is caused by a gram-negative facultative pathogen, *Vibrio cholerae*, which releases a harmful toxin known as the cholera toxin (Nelson et al. 2009).

The disease is introduced to the host by ingestion of contaminated food or water, and after an incubation period of 12 to 72 hours (depending on the individual) diarrheal symptoms are exhibited (Nelson et al. 2009). Individuals will also shed vibrios in their stool, regardless of whether they are symptomatic or not. Susceptibility and whether an individual will be symptomatic depend on host factors, as some individuals may prove to be more resilient and less susceptible than others.

Because cholera is shed by infected individuals regardless whether they are symptomatic or not, if the feces are close to a river there is a chance that some of the freshly shed *Vibrio cholerae* may make its way back into the river. There are many mechanisms that may facilitate this, as the feces do not need to be in direct contact with the river. For example, if individuals in

a community defecate into a hole in the ground, under the right groundwater flow conditions, the leachate may find its way into a nearby river and be capable for disease transmission.

4.3 Applicability of the Model

The Artibonite River is the largest in Haiti, and runs from the Dominican Republic to a delta near Saint-Marc. Below is a map of the Artibonite River.

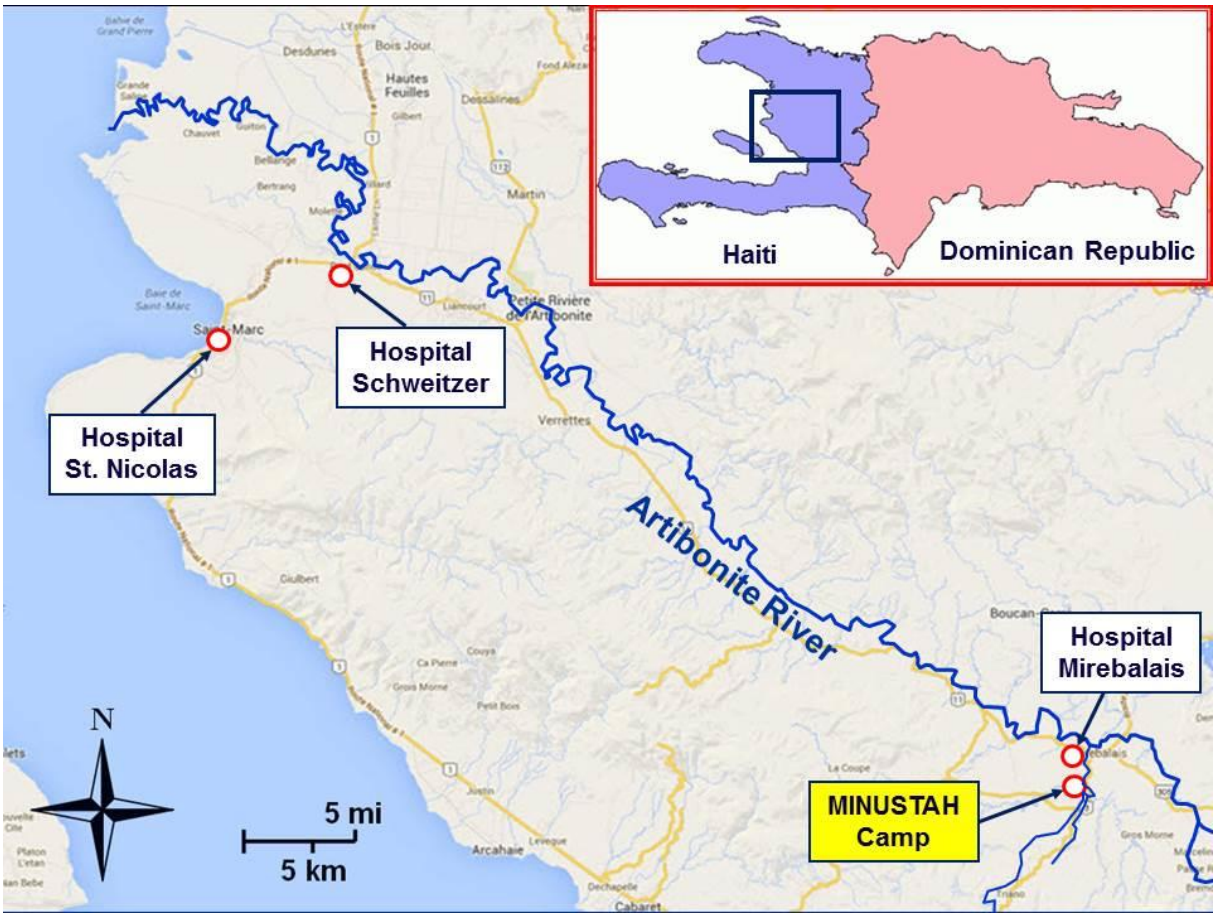

Figure 7 Map of Haiti

The first case of cholera was documented on October 17, 2010 in the Mirebalais Government Hospital, roughly 200 km up steam of Saint-Marc. The Albert Schweitzer Hospital in Deschapelle reported a first case of severe diarrhea requiring hospitalization on October 20,

2010. The Albert Schweitzer Hospital is located two-thirds the distance between Mirebalais and Saint-Marc, and serves as a referral hospital for the communities inhabiting the Artibonite Valley. A third hospital, the St. Nicolas Government Hospital in Saint-Marc saw an explosive increase in the number of patients suffering from diarrhea on October 20, 2010. (Cravioto et al. 2011)

The hospitalization data suggests that the diarrheal disease may have traveled down the Artibonite River channel, and that this was the main source of transmission. The timing of the hospitalization cases also corresponds roughly to the travel time of the water in the Artibonite River, as reported by engineers that operate the dam upstream of the river (Cravioto et al. 2011). It has been reported that tens of thousands of proximal residents rely on the Artibonite River for washing, bathing, drinking, and recreation, thus providing a direct route for infection (Lantagne et al. 2013).

The data suggests that the Haiti cholera outbreak was propagated down the river channel, with people contracting infection as a result of consuming contaminated water from the river. All of these factors make our model very applicable to this situation.

4.4 Model Calibration

To calibrate the model to the Artibonite River in Haiti, river parameters such as the flow, velocity, depth, and width of the channel were gathered. The population along the river was also estimated based on the hospital's catchment area through the use of available 2010 census data, and other sources.

River Data

According to the ODVA (Organization for the Development of the Artibonite Valley) engineering staff, the flow rate in the dry season is roughly 50 m^3/s (Cravioto et al. 2011). The

cholera outbreak was in October however, which is in the wet season. According to the Haiti Bureau of Mines and Energy, flows in the wet season exceed 100 m^3/s (Bureau Des Mines 1992). These two values are approximately the same magnitude as each other, and the increased flow due to the rainy season is not unexpected as deforestation in Haiti is a major issue to the extent that forests covered 60% of the country in 1923 and only 2% in 2006 (Lantagne et al. 2013). Consequently, the precipitation lost to infiltration may currently be much lower, increasing the amount of runoff.

The ODVA dam operators also reported that the water transport from the Peligre Dam (upstream of Mirebalais) to the coast takes roughly three days (Cravioto et al. 2011). From Google Earth, we were able to estimate the length of the Artibonite River from Mirebalais to the coast as 200 km, and with a travel time of three days, we were able to estimate the flow velocity to be around 0.8 m/s.

To attain an estimate for the width of the channel, we took around 50 measurements of the width of the channel from Google Earth. As the Artibonite River is a meandering river channel, there was a relatively large variation in the river channel width. This average value came out to 35 meters in width.

The flow, width, depth, and velocity data for the Meye Tributary near Mirebalais was provided by Dr. Lantagne. Dr. Lantagne obtained these data as the result of her serving on the United Nation's independent review board to investigate the origin of the cholera outbreak (personal communication, October 22, 2013).

Population Data

The at-risk populations needed to be estimated for each of the three hospitals in our network of health institutions reporting cases of diarrhea.

The population in Mirebalais was estimated through data contained on the websites of three different Non-Governmental Organizations (NGOs), Partners in Health (2014); Great Commission Alliance (2010); and Global Giving (2014). These three organizations had data about the demographics of Mirebalais as the result of their interest in sponsoring the construction of a new hospital. The population totals were reported as 185,000; 200,000; and 160,000, yielding an average of around 180,000. For the model an estimate of 200,000 was used as it was estimated that more people had migrated to Mirebalais after the earthquake due to the availability of resources there (Global Giving 2014). The population of Mirebalais was taken as a good proxy for the hospital service area population of the Mirebalais Government Hospital.

The population for the Albert Schweitzer Hospital was the most difficult to estimate, as the hospital serves as a referral hospital to over 300,000 in the Artibonite Valley (Cravioto et al. 2011). However the UN report also notes that because this hospital is not located in close proximity to any population center, there were fewer cases reported here in general Also, because the hospital is a referral hospital (or tertiary hospital), it makes sense that cases of a rather uncomplicated medical condition such as diarrhea would be handled by less sophisticated treatment environments. There were also communities dispersed over the Artibonite River between the hospital and Mirebalais (D. Lantagne, personal communication, October 22, 2013). To arrive at a conservative estimate, the population that would (a) drink the river water and (b) go to the Albert Schweitzer Hospital if they had severe diarrhea was estimated as 100,000.

Lastly, the population for the St. Nicolas Government Hospital in St. Marc was estimated through demographic data provided by the Direction des Statistiques Démographiques et Sociales (DSDS) and Institut Haïtien de Statistique et d'Informatique (IHSI) (2012). We used the

rural population of Saint-Marc, based on the assumption that the rural population was more likely to come in contact with the river water. This population estimate was 250,000 residents.

DSDS and IHSI census data could not be used for Mirebalais because Mirebalais is only a small town and the census is not specific enough, and the data could also not be used to estimate the population for Albert Schweitzer as it is a referral hospital that spans several counties. These counties encompass the river, but also extend into the mountainous regions where the source of water would not be the Artibonite River.

With all of these estimates, it is very challenging to assess the population for the model inputs and also to map the locations of their water intakes. For Mirebalais and Saint-Marc, the withdrawals and discharges were estimated as being 10 km long, whereas the withdrawal and discharge for the middle part of the valley was modeled as a distributed withdrawal/loading spanning over 50 km.

Dr. Lantagne also provided information that the average family in Haiti consisting of 5 individuals would utilize roughly 20-L of water on a daily basis from the river, and this number was included in the model as a per capita water consumption rate (personal communication, October 22, 2013).

4.5 Model Results and Hospitalization Data

The model was run from October 15, 2010, with the initial pathogen dose starting on October 16, 2010 and lasting for a day. The figure below shows the concentration of the cholera at various reaches on the Artibonite River at the end of the 7 day running time of the model.

At 200 km, 130 km, and 10 km, spikes in the concentrations of pathogen are visible as depicted in Figure 8, generated by the model. This is because these are the locations in which cities discharge waste water containing pathogens into the river channel. Also, note that

pathogen concentrations increase as progress is made downstream because there are more infected individuals contributing to the amount of cholera in the river channel; in essence the communities are acting as incubators to amplify the cholera.

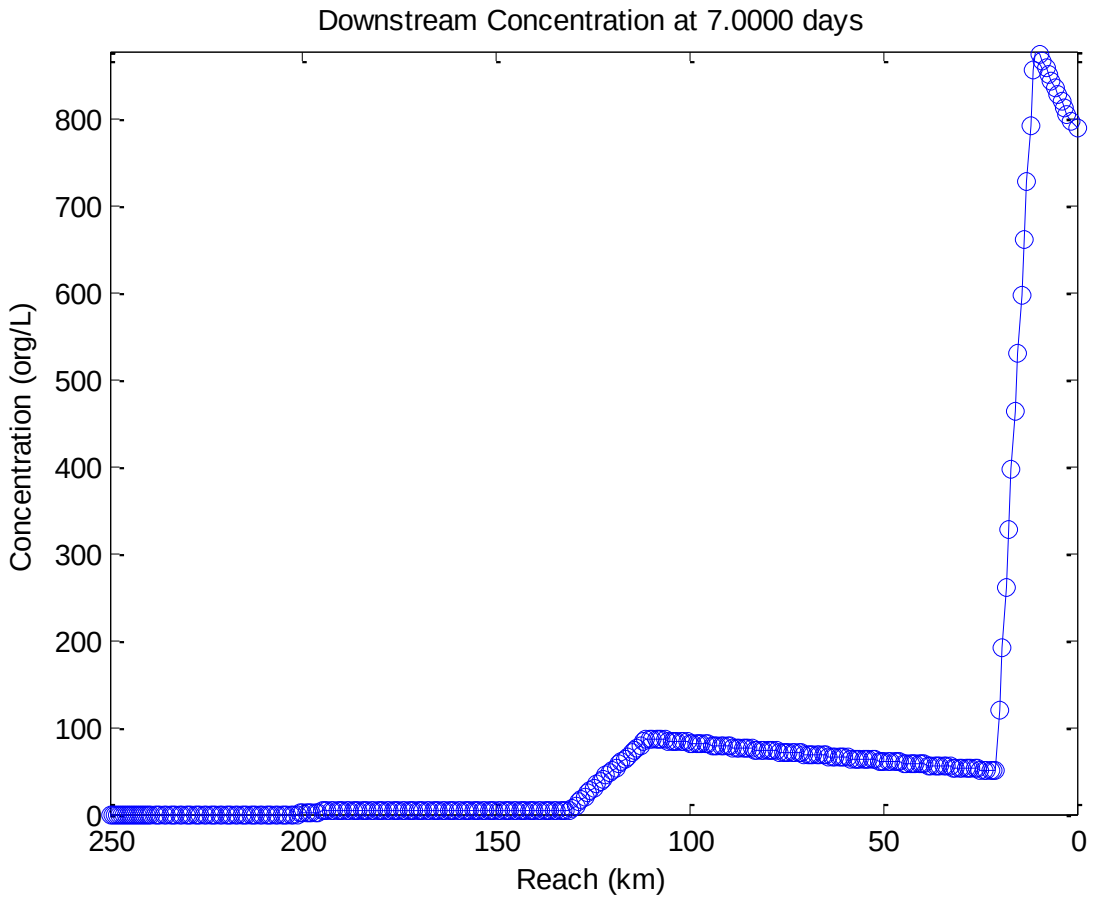

Figure 8 Model Predicted Concentrations

We can also compare the hospitalization data with the infected population calculated by the model, for each of the three communities along the river. Figure 9 shows the projected number of infected individuals (solid lines) plotted alongside the actual hospitalizations as reported by the three hospitals in bar graphs. The number of infected individuals increases as the outbreak travels to downstream communities. This increase is likely due to a cumulative impact of increased pathogen load coupled with the presence of susceptible individuals. The model results

are close to the hospitalization data, but there are biases for both the calculated number of individuals, and the number of hospitalization cases.

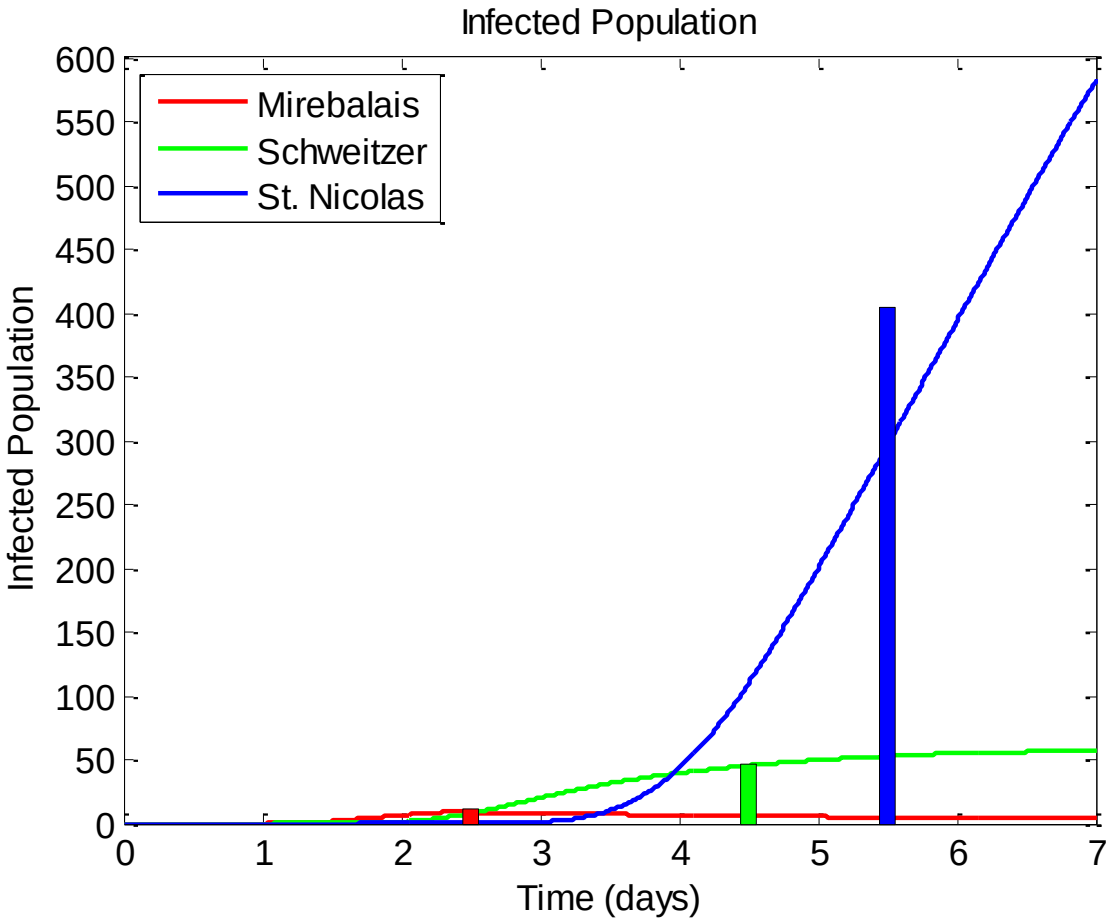

Figure 9 Model Predicted Infective Population

The model results are only as good as the quality of the input parameters, and in this case the population estimates are crucial. We do not actually know how many people rely on water from the Artibonite between Mirebalais and Schweitzer, and if this number were different, it would affect both the number of individuals infected in the Albert Schweitzer Hospital but also in St. Nicolas Hospital.

Another key parameter is how much of the infected excrement load was actually going back into the river channel. For example, such introduction back into the river is dependent on the

behavior of community residents which either favors waste water deposition of feces or alternatively using land-based points of defecation without an ability to immediately enter the river. Although the community may or may not be aware of this, but keeping the stool separate and distant from the river will essentially help to contain the cholera to that community. The downstream communities are still likely to experience a diminished outbreak from the initial cholera in the river channel, but the outbreak is likely to be localized and contained.

The bias in the hospitalization data also needs to be addressed. First, the temporal resolution of the hospitalization data is not too fine, as the cases were only reported on a daily basis. This means that even if patients were hospitalized at 1AM or 11PM, they would get counted as being in the same day. Another important fact to consider is that not all infected individuals would have been hospitalized. It is possible that individuals that did not exhibit severe symptoms did not seek medical attention, or could not go to the hospital for other reasons (financial cost, geographic location, etc.). Nelson et al. (2009) also reports the incubation time of cholera to range between 12 and 72 hours, which is not accounted for in the model. Lastly, it is possible that once individuals had gotten infected, they would return to their homes due to the illness, so the location of infection and where the infection was reported might not be the same.

4.6 Summary

The Haiti cholera outbreak was in many ways one of the most severe public health crisis that could have developed following the earthquake. The event however did provide a real world example where a waterborne pathogen was propagated down a river channel, which could be assessed through our model.

This case study has provided confirmation of our model results to a certain extent. While the model can be further calibrated, as a “first cut” attempt at modeling the situation, the results are

promising. More data of higher quality would definitely help to make the model more useful. It is also clear that there is definitely room for further work to be done in the evolving field of hydroepidemiological models.

Chapter 5. Discussion

5.1 Importance of Model

As outlined in Chapter 1, a hydroepidemiological model that combines a fate and transport model and an epidemic model can provide improved and dynamic insight into the outbreak of waterborne diseases. Previous attempts at combining these two types of models have been rudimentary. A combined model provides a more holistic view of the waterborne infectious disease paradigm through the inclusion of a river and a human population component.

One additional application of the model will be to act as a decision support system to evaluate potential interventions as to how best to contain an epidemic outbreak. The advantage of this model as opposed to others is that river and epidemic parameters can be altered to determine the most efficient intervention. For example, the optimal intervention could be found through the alteration of river characteristics, such as increasing the flow by opening up a dam; or on the human epidemic side, such as starting a vaccination program or a public education program to reduce exposure to the pathogen. Because the epidemic model and river model fully depend on one another, a single parameter can affect both models in significant ways.

The model can also be used in outbreak situations such as Haiti where the source of a waterborne disease may be unknown. By appropriately calibrating the fate and transport and epidemic model parameters, different sources for the disease can be assessed. In our case study discussed in Chapter 4, we assumed that there was one individual in the Meye Tributary leading

into the Artibonite River that was shedding cholera. We could similarly alter the initial loading condition so that there was very low number of *Vibrio cholerae* which underwent a mutation with increased infectivity. Such a scenario could be critically evaluated through the use of the model presented here. The hydroepidemiological model offers the potential to provide a means of assessing plausible sources of an outbreak.

Lastly, the model may be a useful tool in responding to possible bioterrorism attacks that employ the willful introduction of waterborne pathogens. By calibrating and running the model, first responders can get an estimate of when the disease is likely to reach a certain community, and what the most effective response will be. It is possible that in some cases allowing one community to get infected and working on containment may be the most effective way to stop the epidemic. Such cases can be run in the model as well, so the responders can also get an estimate of what the outcome of their actions will be. As bioterrorism attacks using waterborne pathogens have not been documented in recent years, we can only speculate on how useful these models will be in responding to such events.

One part of the model that is still missing is the validation on the fate and transport side of the model. The pathogen loadings resulting from the infected populations are only done by a mass balance now, and it would be helpful if there was river data available for calibration and validation. Along the same lines, there is also much uncertainty in mapping out the populations. The model may not be as accurate when the populations are more distributed along the river channel.

5.2 Open Source

We have developed a comprehensive model that can serve as a decision support system, but the model is not useful if people do not have access to it. The model provides a generic framework

that other researchers can build on and extend the model capabilities. For these reasons, we have decided to make the model open source, so that many others can add to the model and make it better. We hope that by making it open source, we will see innovative applications and new functionalities that we have not even considered at this point. The full source code is available in Appendix B.

5.3 Future Research

By implementing a hydroepidemiological and doing a case study on Haiti, we realized that there are certain aspects of the model that would have been useful had they been implemented. There is definitely more work to be done for hydroepidemiological models, and in this section we will address 4 potential topics: 1) adding tidal dispersion to the river model to add flexibility so that the model can be applied to estuaries, 2) partitioning the pathogen in both a hyperinfective state and a less infective state as proposed by Hartley et al. (2006), 3) addressing the possibility of domestic transmission of the disease in households as detailed by Jensen et al. (2002), and 4) exploring the possibility of further compartmentalizing the population as proposed by Mwasa and Tchuenche (2011).

Tidal Dispersion

Estuaries are regions where free-flowing surface water bodies such as rivers and streams meet the ocean. Because of the tidal influence of the ocean, there is cyclic dispersion in the estuary (Chapra 1997). Extending the capability of the model to include tidal dispersion would make the fate and transport model better suited for estuaries.

Another consequence of implementing tidal dispersion for estuaries would be that we could introduce a salinity gradient in the water. Adding salinity to the model would also add to the decay term of the pathogen, as the saltwater could affect the fate or natural mortality of the

pathogen (Chapra 1997). This would mean that at locations closer to the ocean, the concentration of the pathogen will be reduced. However it is also important to account for the fact that people might not be drinking brackish water because of the salinity.

Hyperinfectivity

As discussed briefly in Chapter 1, recent laboratory experiments have shown that freshly shed cholera are more infective than cholera that has been outside the host for a longer period of time (Hartley et al. 2006). This hyperinfectious *V. cholera* has a lower D_{50} value, increasing the probability of infection for individuals that come in contact with contaminated water. In a hydroepidemiological model where there is a space and time dimensionality to the model, it is possible to partition the cholera into a hyperinfectious and less infectious state, similar to how the population is compartmentalized in the epidemic model.

Hyperinfectivity will increase the probability of an individual getting infected, but this may or may not mean that the cholera will necessarily intensify into a full-blown epidemic. If the distance between two communities along the river is sufficient that all the cholera excreted from the first community will be in a lower infectious stage by the time it reaches the next community, the virulence may not be as high and the outbreak will likely be contained. Compared to our current design where the cholera maintains its virulence independent of time, with hyperinfectivity the magnitude of the outbreak will greatly depend on the proximity of susceptible populations in both time and space to the freshly shed cholera.

Domestic Transmission

Jensen et al. (2002) conducted a 5 week intervention study to evaluate the effects of household contamination in rural communities in Pakistan that relied on surface or well water as their main water source. This study found that even if the water source was not contaminated, there were

cases where the pathogen would get into the domestic water supply due to poor sanitation practices. If the initial water source is contaminated, domestic transmission may be less significant because that household is likely to get infected regardless, but if the water source was not contaminated; domestic transmission becomes much more significant. In many cases, Jensen et al. (2002) found that water pitchers with a narrow opening led to the least number of disease transmissions because individuals were unable to physically touch the water, preventing contamination. In many cases, children were deemed responsible for touching the water with hands that may have been in contact with fecal matter.

Expanding the capability of the model to include domestic transmission will be challenging, as cultural practices and normative behavior of the population will need to be studied to properly to calibrate the model. In terms of infections, this may mean that the susceptible population is still capable of being infected even if their water source is not contaminated, thus introducing an alternate source for disease transmission.

Public Health Interventions

Mwasa and Tchuenche (2011) explored a model setting in which they divided the population into 7 different classes and performed additional mathematical analysis. Their 7 classes were susceptible, educated, vaccinated, quarantined, infected, treated and removed individuals. The cholera was represented similarly to Codeco's model, with an aquatic reservoir.

Further dividing the classic susceptible, infected, and recovered classes may yield more accurate results, as the different classes would consume water differently, and their contribution to the mass of cholera in the water would be different. For example, it can be assumed that quarantined individuals would not be shedding cholera back into the river because they would be placed in a restricted area where their stools would be disposed of correctly.

However, one challenge in implementing such a treatment of different behaviors would be in specifying the rate in which individuals would move from class to class. Education and vaccination rates would be difficult to quantify, as there is no actual measurement of how much any given individual is educated. Even population estimates of this attribute may be incomplete and inaccurate. As for vaccination, for example, cholera requires a two dose vaccination, which further complicates the situation (CDC 2013). It is possible that some individuals do not get the second dosage of the vaccine, which ends up leaving them susceptible to the infection. Such process step details would make the quantifying and subsequent modeling much more challenging.

There are a number of possible topics that can be addressed and incorporated to the model. Here we have presented only four options out of a large number of possible topics. It is also important to realize that for certain rivers and communities, these may not be the most important additions to the model. The proposed framework is very flexible and is meant to supplement the development of a generic model, and a part of calibrating the model further may require the user to add in the appropriate functionality to best model the required situation.

Chapter 6. Conclusion

Currently, billions of people are affected by waterborne pathogens and the lack of sufficient sanitation services. Here we have presented a hydroepidemiological model to help combat these issues by providing feedback obtained from a dynamic and holistic perspective. The coupling of the river fate and transport model with the epidemic model is relatively new and there is still much work to be done, but we have proposed a comprehensive framework that can be used to develop understanding and insight.

We extended the work pioneered by Codeco (2001); incorporated the spatiality explored by Bertuzzo et al. (2007); and the river framework presented by McBride and Chapra (2011). In Chapter 5, we proposed additional parameters to be added to the model. Such additions would likely render it more accurate and useful. These proposed additions are only a few out of the many possible options, and users could choose to extend the model in the manner which is most conducive to understanding the given problem at hand. The model is meant to be a very flexible and generic tool that can be used, extended, and evaluated by others.

No model is completely accurate, but we have also provided verification of the model by analytical solutions in Chapters 3, and confirmation in the Haiti case study covered in Chapter 4. The successful verification by the analytical solutions offers evidence that the river model is accurate and is able to handle different loading conditions successfully. The confirmation results from the case study, specifically how the model predicted infected populations and the alignment with the reported hospitalization cases demonstrates the model's ability to, in this initial application, to predict the magnitude and severity of a given disease outbreak. There is a large bias to be considered in the hospitalization data, as well as large uncertainties in the estimation of model parameters as discussed in Chapter 4, the ability of the model to predict the outbreak in the correct order of magnitude and with an accurate approximation of the disease count data is encouraging.

Waterborne disease outbreaks are (thankfully) a relatively rare occurrence (particularly in high income nations of the world) that strikes when no one is expecting it. Therefore we can only usually derive limited data from these events. Luckily we have hospitalization data from Haiti, and it is no surprise that there were no measurements in the river channel, as no one was expecting the outbreak.

The immediate need for hydroepidemiological models is evident. While there still remains a substantial potential for future work in this field, my hope is that this model will act as a framework that others can expand upon by continuing to improve its accuracy and completeness. In the future, such models will likely play a role in pioneering prevention measures against waterborne diseases, instead of being used retroactively to assess an outbreak. If used effectively, hydroepidemiological models will lead to improved access to safe water and sanitation worldwide by serving as a tool to educate and guide decision making for water resource engineers and public health practitioners alike.

References

Bertuzzo, E., Azaele, S., Maritan, A., Gatto, M., Rodriguez-Iturbe, I., & Rinaldo, A. (2008). On the Space-Time Evolution of a Cholera Epidemic. *Water Resources Research, 44*(1), -W01424. doi:10.1029/2007WR006211

Bras, R. L. (1990). *Hydrology: An Introduction to Hydrologic Science*. Reading, MA: Addison-Wesley Reading.

Bureau Des Mines Et De L'Energie. (1992). *Inventaire Des Ressources Minieres De La Republique D'Haiti.* (No. Department De L'Artibonite). Port-Au-Prince: Direction De La Geologie Et Des Mines.

CDC. (2010). *Acute Watery Diarrhea and Cholera: Haiti Pre-Decision Brief for Public Health Action.* Atlanta, GA: Centers for Disease Control and Prevention.

CDC. (2013). Cholera - Vibrio Cholerae Infection. Retrieved April 15, 2014, from http://www.cdc.gov/cholera/general/

Chapra, S. C., & Canale, R. (2009). *Numerical Methods for Engineers* McGraw-Hill Education.

Chapra, S. C. (1997). *Surface Water-Quality Modeling* McGraw-Hill New York.

Chapra, S. C. (2011). Rubbish, Stink, and Death: The Historical Evolution, Present State, and Future Direction of Water-Quality Management and Modeling FAU. *Environ Eng Res, 16*(3), 113-119. doi:10.4491/eer.2011.16.3.113

Chapra, S. C. (April 2010). *MIT Freeman Lecture: Rubbish, Stink and Death: The Historical Evolution, Present State and Future Direction of Water-Quality Management and Modeling*

Codeco, C. (2001). Endemic and Epidemic Dynamics of Cholera: The Role of the Aquatic Reservoir. *BMC Infectious Diseases, 1*(1), 1-14. doi:10.1186/1471-2334-1-1

Cravioto, A., Lanata, C. F., Lantagne, D., & Nair, G. B. (2011). *Final Report of the Independent Panel of Experts on the Cholera Outbreak in Haiti.* New York, NY: Presented to the Secretary-General of the United Nations.

Di Toro, D. M., O'Connor, D. J., Thomann, R. Y., & St. John, J. P. (1981). *Analysis of Fate of Chemicals in Receiving Waters, Phase 1* (Prepared by HydroQual, Inc., Mahwah, N.J. Trans.). Washington D.C.: Chemical Manufacturing Association.

Direction des Statistiques Démographiques et Sociales, & Institut Haïtien de Statistique et d'Informatique. (2012). *Population Totale, Menages, Superficie Et Densite Estimes En 2012 Au Niveau Des Differentes Unites Geographiques* . Republique D'Haiti.

Global GIving. (2014). Build Mirebalais Hospital in Haiti with PIH. Retrieved April 1, 2014, from http://www.globalgiving.org/projects/build-mirebalais-hospital/

Haas, C. N., Thayyar-Madabusi, A., Rose, J. B., & Gerba, C. P. (2000). Development of a Dose-Response Relationship for Escherichia Coli O157:H7. *International Journal of Food Microbiology, 56*(2–3), 153-159. doi:http://dx.doi.org/10.1016/S0168-1605(99)00197-X

Hartley, D. M., Morris, J. G. Jr, & Smith, D. L. (2006). Hyperinfectivity: A Critical Element in the Ability of V. Cholerae to Cause Epidemics? *PLoS Medicine, 3*(1), e7. doi:10.1371/journal.pmed.0030007

Jensen, P. K., Ensink, J. H., Jayasinghe, G., van der Hoek, W., Cairncross, S., & Dalsgaard, A. (2002). Domestic Transmission Routes of Pathogens: The Problem of in-House Contamination of Drinking Water during Storage in Developing Countries. *Tropical Medicine & International Health: TM & IH, 7*(7), 604-609.

Joh, R., Wang, H., Weiss, H., & Weitz, J. (2009). Dynamics of Indirectly Transmitted Infectious Diseases with an Immunological Threshold. *Bulletin of Mathematical Biology, 71*(4), 845-862. doi:10.1007/s11538-008-9384-4

Kermack, W. O., & McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, 115*(772), 700-721.

Lantagne, D., Balakrish Nair, G., Lanata, C. F., & Cravioto, A. (2013). The Cholera Outbreak in Haiti: Where and how did it Begin? *Current Topics in Microbiology and Immunology,* doi:10.1007/82_2013_331

Leopold, L. B., & Maddock, T. (1953). *The Hydraulic Geometry of Stream Channels and some Physiographic Implications.* (US Geological Survey Professional Paper No. 252). Washington D.C.:

Library of Congress. (May 2006). *Country Profile: Haiti* Research and Reference Services - Federal Research Division.

Mac Kenzie, W. R., Hoxie, N. J., Proctor, M. E., Gradus, M. S., Blair, K. A., Peterson, D. E., . . . Rose, J. B. (1994). A Massive Outbreak in Milwaukee of Cryptosporidium Infection Transmitted through the Public Water Supply. *The New England Journal of Medicine, 331*(3), 161-167. doi:10.1056/NEJM199407213310304

Mancini, J. L. (1978). Numerical Estimates of Coliform Mortality Rates Under various Conditions. *Journal (Water Pollution Control Federation),* 2477-2484.

McBride, G. B., & Chapra, S. C. (2011). New Hydroepidemiological Models of Indicator Organisms and Zoonotic Pathogens in Agricultural Watersheds. *Ecological Modelling, 222*(13), 2093-2102. doi:http://dx.doi.org/10.1016/j.ecolmodel.2011.04.008

Meeus, J. H. (1991). *Astronomical Algorithms*, Willmann-Bell, Incorporated.

Mwasa, A., & Tchuenche, J. M. (2011). Mathematical Analysis of a Cholera Model with Public Health Interventions. *Bio Systems, 105*(3), 190-200. doi:10.1016/j.biosystems.2011.04.001

National Institute of Health. (2002). *Dose-Response relationships.* Unpublished manuscript.

Nelson, E. J., Harris, J. B., Morris, J. G. Jr, Calderwood, S. B., & Camilli, A. (2009). Cholera Transmission: The Host, Pathogen and Bacteriophage Dynamic. *Nature Reviews. Microbiology, 7*(10), 693-702. doi:10.1038/nrmicro2204; 10.1038/nrmicro2204

OCHA. (2013). *Haiti Humanitarian Response*. Geneva: United Nations Office for the Coordination of Humanitarian Affairs.

Office of the Secretary-General's Special Adviser on Community-Based Medicine & Lessons from Haiti. (2013). Water and sanitation. Retrieved April 10, 2013, from http://www.lessonsfromhaiti.org/about-haiti/water-sanitation/

O'Loughlin, E. M., & Bowmer, K. H. (1975). Dilution and Decay of Aquatic Herbicides in Flowing Channels. *Journal of Hydrology, 26*(3), 217-235.

Partners in Health. (2014). Hôpital Universitaire de Mirebalais. Retrieved April 1, 2014, from http://www.pih.org/pages/mirebalais

Streeter, H. W., & Phelps, E. B. (1925). A Study of the Pollution and Natural Purification of the Ohio River. *Public Health Bulletin, 146* (Washington, DC: United States Public Health Service)

Strous, L. (2011). Position of the Sun. Retrieved June 26, 2013, from http://users.electromagnetic.net/bu/astro/sunrise-set.php

Thomann, R., & Mueller, J. *Principles of Surface Water Quality Modeling and Control*, 1987. Happer & Row Publishers, New York.

Tennessee Valley Authority. (1972). *Heat and Mass Transfer Between a Water Surface and the Atmosphere.* (Water Resources Research, Laboratory Report No. 14). Norris TN: Division of Water Control Planning, Tennessee Valley Authority.

World Health Organization. (2008). *Safer Water, Better Health: Costs, Benefits and Sustainability of Interventions to Protect and Promote Health.* Geneva, Switzerland: WHO Press.

World Health Organization. (2013). *Progress on Sanitation and Drinking-Water - 2013 Update.* Geneva, Switzerland: WHO Press.

Appendix

A: Solar Calculations

In the following sections, we will outline how the 1) average insolation, 2) sunrise, sunset, photoperiod, and 3) daytime insolation levels are calculated in the model.

Average Solar Insolation

First we need to calculate the average insolation for a specific location on earth (specified by a latitude and longitude) for a specific day. These steps are outlined in Bras (1990).

The extraterrestrial radiation is computed as (TVA 1972):

$$I_0 = \frac{W_0}{r^2}\sin(\alpha)$$

where

W_0 is the solar constant (1367 W/m^2), and

r is the normalized radius of the earth's orbit (i.e. the ratio of actual earth-sun distance to mean earth-sun distance).

α is the solar altitude (radians) that can be computed as follows:

$$\sin(\alpha) = \sin(\delta)\sin(\phi) + \cos(\delta)\cos(\phi)\cos(\tau)$$

where

δ is the declination of the sun,

ϕ is the local latitude, and

τ is the local hour angle.

This model assumes that the cosine of τ is equal to unity. This assumption can be made because the local hour angle is a very minor adjustment and will not vary the insolation that significantly.

The normalized radius of the earth's orbit, r, can be calculated as:

$$r = 1.0 + 0.017\cos\left[\frac{2\pi}{365}(186 - D)\right]$$

where D is the Julian day ($1 \leq D \leq 365$ or 366), the number of days from the start of the year, so February 1 would be $D = 32$.

The declination of the sun, δ, can be approximated by (TVA 1972):

$$\delta = \frac{23.45\pi}{180}\cos\left[\frac{2\pi}{365}(172 - D)\right]$$

Bras (1990) also discusses the possibility of including attenuation terms for solar radiation for reasons such as cloud cover, the type of ground cover, and atmospheric scattering, but these were chosen not to be included in this model for simplicity sake.

Calculating Solar Position, Sunrise, and Sunset

Calculations for sunrise and sunset are based on the equations from Astronomical Algorithms, by Jean Meeus, with a slight modifications made by Dr. Louis Strous, a professional astronomer who has worked for the New Jersey Institute of Technology, the National Solar Observatory, and for the Lockheed Martin Solar and Astrophysics Laboratory. (Strous 2011)

First the Julian number of days from January 1, 2000 is calculated and is added to the number 2451545, which is the Julian day of January 1, 2000 from January 1, 4713 BC, based on the proleptic Julian calendar. Then the Julian cycle is calculated for that day as:

$$n = round\left((J_{date} - 2451545 - 0.0009) - (\frac{l_w}{360})\right)$$

where

J_{date} is the Julian numbers of day from 1/1/2000, and

l_w is the local latitude where west is in the positive direction.

We can then approximate the Julian date for solar noon, J_*, for a specific day as:

$$J_* = 2451545 + 0.0009 + \left(\frac{l_w}{360}\right) + n$$

Using this approximate value we can then compute the mean solar anomaly *M* as:

$$M = [357.5291 + 0.98560028 \times (J_* - 2451545] \bmod 360$$

where mod is the modulus division operator. We then can calculate the center of the sun *C* as:

$$C = (1.9148 \times \sin(M)) + (0.0200 \times \sin(2M)) + (0.0003 \times \sin(3M))$$

Then with C and M, we can calculate the ecliptical longitude of the sun λ as:

$$\lambda = [M + 102.9372 + C + 180] \bmod 360$$

With the ecliptical longitude and mean solar anomaly known, we can calculate the accurate Julian date for solar noon, $J_{transit}$ as:

$$J_{transit} = J_* + (0.0054 \times \sin(M)) - (0.0060 \times \sin(2\lambda))$$

In order to calculate the hour angle, we can calculate the declination of the sun δ as:

$$\delta = \operatorname{asin}(\sin(\lambda)\sin(23.45))$$

We can then calculate the hour angle of the sun *H* as:

$$H = \operatorname{acos}\left(\frac{[\sin(-0.83) - \sin(l_n)\sin(\delta)]}{\cos(l_n)\cos(\delta)}\right)$$

where l_n is the local latitude where north is in the positive direction.

We can then use the hour angle to approximate the Julian date for solar noon again as:

$$J_{**} = 2451545 + 0.0009 + \left(\frac{H + l_w}{360}\right) + n$$

The sunset then can be calculated as:

$$J_{set} = J_{**} + 0.0053\sin(M) - 0.0069\sin(2\lambda)$$

Then the sunrise can be calculated easily, if we make the assumption that solar noon is halfway between sunrise and sunset, which will also be the assumption used to calculate the insolation at a specific hour of the day.

$$J_{rise} = J_{transit} - (J_{set} - J_{transit})$$

Photoperiod and Daytime Insolation

Now that we have calculated the average solar insolation and the time of sunrise, sunset and solar noon, we can calculate the insolation at a specific time of day between sunrise and sunset. The peak insolation for a day can be calculated based on the average insolation and photoperiod by the following relationship (Chapra 1997):

$$I = I_m \frac{2f}{\pi}$$

where *I* is the average insolation, *Im* is the peak insolation, and *f* is defined as:

$$f = \frac{t_s - t_p}{T_p}$$

where

t_s is the sunset time

t_r is the sunrise time, and

T_p is the daily period of 24 hrs.

The insolation at a specific time can then be calculated as:

$$I(t) = I_m \sin[w(t - t_r)]$$

where *t* is the current time, and *w* is the angular frequency defined as:

$$w = \frac{\pi}{fT_p}$$

The insolation at a specific time can then be used in conjunction with the Beer-Lambert Law to calculate the attenuation by light effects.

B: Model Source Code

```
% Filename: Main.m
% Hydroepidemiological Model for Waterborne Pathogens
% Undergraduate Honors Thesis
% Calibrated for Haiti Case Study
% Yukinobu Tanimoto
% Tufts University
% Department of Civil and Environmental Engineering
% Last modified April 18, 2014

clear; clc; tic;

%----------------- Physical Parameters and Rates--------------------------
% Urban Population Characteristics
PC = 10;         % per cap water use (L/cap/day)
PCh = 2;         % per cap human consumption (L/cap/day)
Fcg = 0.1;       % fraction graywater use
Fch = 0.1;       % fraction human consumptive use

% Pathogen Specific
Kd = 0.2;        % 1/day - pathogen in stream attenuation
D50 = 10^5;      % org/day
shed = 10^10;    % shedding rate (org/cap/day)
Kds = 0.05;      % linear sorption coeff

% SIR specifics
km = 0.0;        % pop Infected mortality rate (I->death)
kre = 0.2;       % pop recovery rate (I->R)

% Urban setup
Vd = 0.05;       % distribution sys Vol (m^3)
Vc = 0.05;       % collection sys Vol (m^3)

% River sediment info
m = 150;         % mg/L

% Geographic Information
Long = 72.78;    % Longitude
Lat = 19.25;     % Latitude

GMToffset = -4; % EST is UTC -4 hours
%-----------------------------------------------------------------------

%------------------------------- Time variables --------------------------
% CRITICAL dt FOR En = 0
%dt = 2^(round(log10(xlength(1)/U(1))/log10(2))); % day

dtuser = 0.01;  % day
ti = 0;          % day
tf = 7;          % day

format long
dt = 2^(round(log10(dtuser)/log10(2)));
```

```
format short

tstart = '15-October-2010';    % model starting day
%-------------------------------------------------------------------------

%-------------------------------Flow Distribution--------------------------
% River Setup
L = 250;                    % Length in km
nReach = 3;
xReach = [L L-10 0];
ns = [20 300];              % nReach -1 = number of sections
maxReach = xReach(1);
nel = 0;
xedn(1:sum(ns)+1) = 0;
xedn(1) = maxReach;         %xedn: reach for each reactor
xlength(1:sum(ns)) = 0;     %xlength: length of each reactor

lastj = 0;
for i=2:nReach
    dx = (xReach(i-1) - xReach(i))/ns(i-1);
    for j=1:ns(i-1)
        nel = nel + 1;
        k=j+lastj;
        xedn(k+1) = xedn(k) - dx;
        xlength(k) = xedn(k) - xedn(k+1);
    end
    lastj=j;
end

% nel = number of reactors
% xedn = 100,98,96,...
% xlength = 2,2,2,2,2,...
%-------------------------------------------------------------------------

% --------------------------- Cities-----------------------------------------
nurb = 3;               % number of cities along river

pop(1:nurb) = 0;        % urban population (cap)
Surb(1:nurb) = 0;       % susceptible population (cap)
xurbpta(1:nurb) = 0;    % city withdrawal location
xurbpt(1:nurb) = 0;     % city discharge location

Tdist(1:nurb) = 1;      % distribution residence time (day)
Tcoll(1:nurb) = 1;      % collection sys residence time (day)

% city populations
pop(1) = 200000;
pop(2) = 150000;
pop(3) = 300000;

% city withdrawal locations (beginning and end)
xurbptaup(1) = 239;
xurbptadw(1) = 235;

xurbptaup(2) = 190;
```

```
xurbptadw(2) = 140;

xurbptaup(3) = 60;
xurbptadw(3) = 50;

% city discharge locations(beginning and end)
xurbptup(1) = 200;
xurbptdw(1) = 190;

xurbptup(2) = 130;
xurbptdw(2) = 110;

xurbptup(3) = 20;
xurbptdw(3) = 10;
% ------------------------------------------------------------------------

% ---------------------------------Point Sources--------------------------
nptt = 1;       % number of pt sources

xptt(1:nptt)=0;         % pt source location (km)
Qptta(1:nptt)=0;        % pt withdrawal flow rate (m^3/s)
Qptt(1:nptt)=0;         % pt discharge flow rate (m^3/s)
Cptt(1:nptt)=0;         % pt discharge conc (org/m^3)

xptt(1) = 240;
Qptt(1) = 100;
cptt(1) = 0;
% ------------------------------------------------------------------------

% ------------------------------Diffused Sources--------------------------
ndiffu = 0; % number of diffused sources

% xdup = upstream loc (km)
% xddw = downstream loc (km)
% Qdiffa = diffused withdrawal (m^3)
% Qdiff = diffused discharge (m^3)
% Cdiff = conc. of discharge (org/m^3)

xdup(1) = 209.7;
xddw (1) = 209.1;
Qdiffa(1) = 50;
Qdiff(1) = 0;
Cdiff(1) = 0;
% ------------------------------------------------------------------------

% ---------------------------------Initialize Arrays----------------------
Qpta(1:nel) = 0;    % outflow per reactor
Qpt(1:nel) = 0;     % inflow per reactor
wpt(1:nel) = 0;     % loading per reactor
% ------------------------------------------------------------------------

% ---------------------------------Flow Distribution----------------------
% Add in/assign point sources
for i=1:nptt
```

```
    for j=2:nel
        if(xedn(j-1)>xptt(i) & xptt(i)>=xedn(j))
            if(Qptta(i)~=0) %if withdrawal is NOT 0, so if withdrawal
                Qpta(j) = Qpta(j) + Qptta(i);
            else  % inflow
                Qpt(j) = Qpt(j) + Qptt(i);
                wpt(j) = wpt(j) + cptt(i)*Qptt(i);
            end
        end
    end
end

% Add in/assign diffused sources. Locate start/end reactor cells
for i=1:ndiffu
    diffux  = 0;
    for j=2:nel+1
        if(xedn(j-1)>=xdup(i) & xdup(i)>xedn(j))
            upns(i) = j-1;
            diffux = [xdup(i)-xedn(j)];
        elseif xdup(i)>xedn(j) & xedn(j)>xddw(i)
            diffux = [diffux xedn(j-1)-xedn(j)];
        elseif(xedn(j-1)>xddw(i) & xddw(i)>=xedn(j))
            dwns(i) = j-1;
            if (xddw(i)==xedn(j))
                diffux = [diffux xlength(j)];
            else
                diffux = [diffux xddw(i)-xedn(j)];
            end
        end

        if(xedn(j-1)>=xdup(i)&xddw(i)>=xedn(j))
            diffux =[xdup(i)-xddw(i)];
            upns(i) = j-1;
            dwns(i) = j-1;
            break;
        end

    end

    if Qdiffa(i)~=0
        if(diffux==0)
            Qpta(upns(i):dwns(i)) = Qpta(upns(i):dwns(i)) + Qdiffa(i);
        else
            Qperx = Qdiffa(i)/(xdup(i)-xddw(i));
            Qpta(upns(i):dwns(i)) = Qpta(upns(i):dwns(i)) + Qperx*diffux;
        end
    end

    if Qdiff(i)~=0
        if(diffux==0)
            Qpt(upns(i):dwns(i)) = Qpt(upns(i):dwns(i)) + Qdiff(i);
        else
            Qperx = Qdiff(i)/(xdup(i)-xddw(i));
            Qpt(upns(i):dwns(i)) =  Qpt(upns(i):dwns(i)) + Qperx*diffux;
            wpt(upns(i):dwns(i)) = wpt(upns(i):dwns(i)) +
Cdiff(i)*Qperx*diffux;
```

```
        end
    end

end
%-------------------------------------------------------------------------

%--------Urban Flows to be treated similarly to diffused flows-------------
% First loop for urban withdrawals (diffused or point)
for i=1:nurb
    diffux  = 0;
    for j=2:nel+1
        if(xedn(j-1)>=xurbptaup(i) & xurbptaup(i)>xedn(j))
            upns(i) = j-1;
            diffux = xurbptaup(i)-xedn(j);
        elseif xurbptaup(i)>xedn(j) & xedn(j)>xurbptadw(i)
            diffux = [diffux xedn(j-1)-xedn(j)];
        elseif(xedn(j-1)>xurbptadw(i) & xurbptadw(i)>=xedn(j))
            dwns(i) = j-1;
            if (xurbptadw(i)==xedn(j))
                diffux = [diffux xlength(j)];
            else
                diffux = [diffux xurbptadw(i)-xedn(j)];
            end
        end

        if(xedn(j-1)>=xurbptaup(i)&xurbptadw(i)>=xedn(j))
            diffux =xurbptaup(i)-xurbptadw(i);
            upns(i) = j-1;
            dwns(i) = j-1;
            break;
        end
    end

    if(diffux==0)
        Qpta(upns(i):dwns(i)) = Qpta(upns(i):dwns(i)) +
pop(i)*PC/1000/24/3600;
    else
        Qperx = pop(i)*PC/1000/24/3600/(xurbptaup(i)-xurbptadw(i));
        Qpta(upns(i):dwns(i)) = Qpta(upns(i):dwns(i)) + Qperx*diffux;
    end
end

% second loop for urban discharges (diffused or point)
for i=1:nurb
    diffux = 0;
    for j=2:nel+1
        if(xedn(j-1)>=xurbptup(i) & xurbptup(i)>xedn(j))
            upns(i) = j-1;
            diffux = [xurbptup(i)-xedn(j)];
        elseif xurbptup(i)>xedn(j) & xedn(j)>xurbptdw(i)
            diffux = [diffux xedn(j-1)-xedn(j)];
        elseif(xedn(j-1)>xurbptdw(i) & xurbptdw(i)>=xedn(j))
            dwns(i) = j-1;
            if (xurbptdw(i)==xedn(j))
                diffux = [diffux xlength(j)];
            else
```

```
                diffux = [diffux xurbptdw(i)-xedn(j)];
            end
        end

        if(xedn(j-1)>=xurbptup(i)&xurbptdw(i)>=xedn(j))
            diffux =[xurbptup(i)-xurbptdw(i)];
            upns(i) = j-1;
            dwns(i) = j-1;
            break;
        end
    end

    if(diffux==0)
        Qpt(upns(i):dwns(i)) =  Qpt(upns(i):dwns(i)) + pop(i)*(PC-PCh)*(1-
Fcg)/1000/24/3600;
        Qpt(upns(i):dwns(i)) =  Qpt(upns(i):dwns(i)) + pop(i)*PCh*(1-
Fch)/1000/24/3600;
    else
        Qperx = (pop(i)*(PC-PCh)*(1-Fcg)+pop(i)*PCh*(1-
Fch))/1000/24/3600/(xurbptup(i)-xurbptdw(i));
        Qpt(upns(i):dwns(i)) =  Qpt(upns(i):dwns(i)) + Qperx*diffux;
    end


end
%-------------------------------------------------------------------------

% --------------------------------Hydraulics--------------------------------
% set up down and out flows for reactors
Qout(1:nel) = 0;    % m^3/s;
Qdown(1:nel) = 0;   % m^3/s;

% calibrated for Artibonite River/Meye Tributary
r1 = ns(1);
U(1:r1) = 0.4;
B(1:r1) = 2.5;
H(1:r1) = 0.3;
A(1:r1) = B(1:r1).*H(1:r1);
Qdown(1) = U(1)*B(1)*H(1);

Qb = Qdown(1);      % m^3/s
Qout(1) = Qb;       % m^3/s
Qdown(1) = Qb;      % m^3/s

r2 = r1;
U(r2+1:nel) = 0.77;
B(r2+1:nel) = 35;
H(r2+1:nel) = 4;
A(r2+1:nel) = B(r2+1:nel).*H(r2+1:nel);

VOL = A.*xlength*1000;

for i=2:nel
    Qdown(i) = Qdown(i-1) + Qpt(i) - Qpta(i);
end
```

```
U = U*24*3600;
Qdown = Qdown*24*3600;
Qout = Qdown + Qpta*24*3600;
wpt = wpt*24*3600;

% assuming original infection shed for 3 days
target = 3*shed/Qdown(1);

% BC Concentration: matrix of row=1 of times, row=2 of corresponding conc.
bc = [ 0     0.999    1      1.999       2      400
    0     0      target   target     0      0   ]';

if (dt<min(VOL./Qdown))
    fprintf('dt of %.9f days is stable\n\n',dt);
else
    error('chosen dt is unstable');
end

fprintf('Running RK4 Integrator...');
% Rk4 Integration
[time, conc, Surb, Iurb, Rurb, Is] =
Rk4Integrator(ti,dt,tf,tstart,Lat,Long,GMToffset,ns,VOL,H,bc,Qdown,Qout,Kd,wp
t,nurb,pop,xurbptaup,xurbptadw,xurbptup,xurbptdw,xedn,nel,PCh,D50,kre,km,shed
,Kds,m,PC,Fcg,Fch);
fprintf('done\n');toc;
conc = conc/1e3;  % bring to org/L

plotincr = 1;
% plotincr = input('Plot every how many calc:');

% Safety in case all conc = 0
if max(conc(:,:))==0
    error('all conc is 0!!')
end

xedn(sum(ns))=[];
plotstart = ceil(2.6/dt);
plotend=ceil(tf/dt);
maxY = max(max(conc(plotend,:)));

% Plot first time step for conc vs reach
clf;
plot(xedn,conc(1,:),'o-');
set(gca,'XDir','reverse');
ylim([0 maxY]); xlim([0 L]);
xlabel('Reach (km)','FontSize',12);
ylabel('Concentration (org/L)','FontSize',12);
title('Downstream Concentration at 0 days','FontSize',12);
set(gca,'FontSize',12);
hold on; pause();

% loop though time steps
for j=2:plotincr:length(conc(:,1))
    clf; plot(xedn,conc(j,:),'o-');
```

```
    set(gca,'XDir','reverse');
    xlabel('Reach (km)','FontSize',12);
    ylabel('Concentration (org/L)','FontSize',12);
    set(gca,'FontSize',12);
    ylim([0 maxY]);xlim([0 L]);
    title(sprintf('Downstream Concentration at %.4f days',time(j-
1)),'FontSize',12);
    pause(0.000001);
end

% clc;
% if (dt<min(VOL./Qdown))
%     fprintf('dt of %.9f days is stable\n\n',dt);
% end

% figure();
% % Display SIR plot
% for i=1:nurb
%     subplot(nurb,1,i);
%     plot(time,Surb(i,:))
%     ylim([0 pop(i)]);
%     hold on
%     plot(time,Iurb(i,:),'g')
%     plot(time,Rurb(i,:),'r')
%     plot(time,Surb(i,:)+Iurb(i,:)+Rurb(i,:),'-k')
%     legend('Susceptible','Infected','Recovered','Total');
%     title(sprintf('City #%1.0f',i),'fontsize',16);
%
% end

% Plot of Infected
pause(); figure();
plot(time,Iurb(1,:),'r')
hold on
plot(time,Iurb(2,:),'g'); plot(time,Iurb(3,:),'b');
ll=legend('Mirebalais','Schweitzer','St. Nicolas','location','NorthWest');
title('Infected population','FontSize',12);
xlabel('Time (days)','FontSize',12);
ylabel('Infected Capita','FontSize',12);
set(gca,'YTick',[0:50:700]);
set(ll,'FontSize',12);
set(gca,'FontSize',12);
xlim([ti tf]);
ylim([0 500]);
```

```
function [time, conc, Surb, Iurb, Rurb, Is] =
Rk4Integrator(ti,dt,tf,tstart,Lat,Long,GMToffset,ns,Vol,H,bc,Qdown,Qout,Kd,wp
t,nurb,popu,xurbptaup,xurbptadw,xurbptup,xurbptdw,xedn,nel,PCh,D50,kre,km,rho
,Kds,m,PC,Fcg,Fch)
% rk4 classic runge-kutta integrator
% Yukinobu Tanimoto
% Tufts University
% Filename: Rk4Integrator.m
% Integrates hydroepidemiological model by non-adaptive classic RK4
```

```
% Input:
%  ti = simulation starting time
%  dt = time increment/step size
%  tf = simulation ending time
%  tstart = simulating starting day. ie '15-October-2010';
%  Lat = local latitude in degrees
%  Long = local longitude in degrees
%  GMToffset = difference from UTC. ie -4 for Boston;
%  ns = array of number of cells per reach
%  Vol = array of volumes of cells
%  H = array of height of cells
%  bc = boundary condition table
%  Qdown = array of flows going into downstream cell
%  Qout = array of total flows coming out of cell
%  Kd = pathogen in stream attenuation (1/day)
%  wpt = array of loadings per cell
%  nurb = number of communities along river
%  popu = array of population for nurb
%  xurbptaup = city withdrawal location (beginning)
%  xurbptadw = city withdrawal location ()
%  xurbptup = city discharge location (beginning)
%  xurbptdw = city discharge location (end)
%  xedn = vector of reaches
%  nel = number of reactors, should be same as sum(ns)
%  PCh = per cap human consumption (L/cap/day)
%  D50 = infectious dose 50 (org/day)
%  kre = pop recovery rate (I->R)
%  km = pop Infected mortality rate (I->death)
%  rho = shedding rate (org/cap/day)
%  Kds = linear sorption coeff
%  m = river sediment load (mg/L)
%  PC = per cap water use (L/cap/day)
%  Fcg = fraction graywater use
%  Fch = fraction human consumptive use
% Output:
%  time = vector of times stepped through by model
%  conc = array of concentrations (time x reach)
%  Surb = array of susceptible individuals over time (time x nurb)
%  Iurb = array of infected individuals over time (time x nurb)
%  Rurb = array of recovered individuals over time (time x nurb)
%  Is = vector of all insolation calculated at every step

% Initialize Vars
t = ti;
conc (1:sum(ns))=0;
c(1:sum(ns))=0;
tr = 0;
count = 0;
Is = 0;

Surb = popu;         Surb = Surb';
Iurb(1:nurb) = 0;    Iurb = Iurb';
Rurb(1:nurb) = 0;    Rurb = Rurb';

dSdt(1:nurb) = 0;
dIdt(1:nurb) = 0;
```

```
dRdt(1:nurb) = 0;

Surbt = Surb;
Iurbt = Iurb;
Rurbt = Rurb;

trise = 0; tnoon = 0; tset = 0;
Im = 0; Io = 0; w=0;
days=0;

%RK4 Additions
cmid(1:sum(ns))=0;      cend(1:sum(ns))=0;
smid(1:nurb) = 0;       send(1:nurb) = 0;
imid(1:nurb) = 0;       iend(1:nurb) = 0;
rmid(1:nurb) = 0;       rend(1:nurb) = 0;

k1dCdt(1:sum(ns)) = 0;   k2dCdt(1:sum(ns)) = 0;
k2dCdt(1:sum(ns)) = 0;   k3dCdt(1:sum(ns)) = 0;

k1dSdt(1:nurb) = 0;   k2dSdt(1:nurb) = 0;
k2dSdt(1:nurb) = 0;   k3dSdt(1:nurb) = 0;

k1dIdt(1:nurb) = 0;   k2dIdt(1:nurb) = 0;
k2dIdt(1:nurb) = 0;   k3dIdt(1:nurb) = 0;

k1dRdt(1:nurb) = 0;   k2dRdt(1:nurb) = 0;
k2dRdt(1:nurb) = 0;   k3dRdt(1:nurb) = 0;

% Keep track of real time
ta = datenum(tstart);       % model actual starting time
t1 = datestr(tstart);       % model actual starting time vector

tf = tf+dt;
% Do calculations
while(1)

    if(t+dt>=tf)
        dt = tf-t;  %Prevents overstepping loop
    end

    % Lin-terp BC
    c(1) = linterp(bc(:,1),bc(:,2),t);

    % on start of each day
    if mod(count,round(1/dt))==0
        [Io,dayLen1] = InsolationCalc(t1,Lat,Long);
        [trise, tnoon, tset] = SunTimes(t1,Lat,Long,GMToffset);
        dayLen2 = dayLength(tset,trise);
        f = dayLen2/24;             % fraction of day that w/ light
        Im = Io * pi() / (24*2*f);  % peak light for the day
        w = pi()/(f*24);            % angular freq
    end

    % need k1-k4 slopes:
```

```
    % smaller step time variables
    kt = t;          % same as dt, to protect t
    h = dt;          % same as dt, to protect dt
    kta = ta;        % same as ta, to protect ta
    kt1 = datestr(kta);

    %k1 step;
    [k1dCdt k1dSdt k1dIdt k1dRdt I] =
deriva(kt,kt1,trise,tset,c,ns,Vol,H,bc,Qdown,Qout,Kd,wpt,nurb,popu,xurbptaup,
xurbptadw,xurbptup,xurbptdw,xedn,nel,PCh,D50,kre,km,rho,Surbt,Iurbt,Rurbt,Kds
,m,Im,w,PC,Fcg,Fch);
    cmid = c     + k1dCdt*h/2;
    smid = Surbt + k1dSdt*h/2;
    imid = Iurbt + k1dIdt*h/2;
    rmid = Rurbt + k1dRdt*h/2;

    %k2 step, need h/2 times.
    kelapsed = floor(h/2*24*3600*1e3);
    kta = addtodate(kta,kelapsed,'millisecond');
    kt1 = datestr(kta);

    [k2dCdt k2dSdt k2dIdt k2dRdt I] =
deriva(kta,kt1,trise,tset,cmid,ns,Vol,H,bc,Qdown,Qout,Kd,wpt,nurb,popu,xurbpt
aup,xurbptadw,xurbptup,xurbptdw,xedn,nel,PCh,D50,kre,km,rho,Surbt,Iurbt,Rurbt
,Kds,m,Im,w,PC,Fcg,Fch);
    cmid = c     + k2dCdt*h/2;
    smid = Surbt + k2dSdt*h/2;
    imid = Iurbt + k2dIdt*h/2;
    rmid = Rurbt + k2dRdt*h/2;

    %k3 step, need h
    [k3dCdt k3dSdt k3dIdt k3dRdt I] =
deriva(kta,kt1,trise,tset,cmid,ns,Vol,H,bc,Qdown,Qout,Kd,wpt,nurb,popu,xurbpt
aup,xurbptadw,xurbptup,xurbptdw,xedn,nel,PCh,D50,kre,km,rho,Surbt,Iurbt,Rurbt
,Kds,m,Im,w,PC,Fcg,Fch);
    cend = c     + k3dCdt*h/2;
    send = Surbt + k3dSdt*h;
    iend = Iurbt + k3dIdt*h;
    rend = Rurbt + k3dRdt*h;

    %increment k times by h/2
    kelapsed = floor(h/2*24*3600*1e3);
    kta = addtodate(kta,kelapsed,'millisecond');
    kt1 = datestr(kta);

    [k4dCdt k4dSdt k4dIdt k4dRdt I] =
deriva(kta,kt1,trise,tset,cend,ns,Vol,H,bc,Qdown,Qout,Kd,wpt,nurb,popu,xurbpt
aup,xurbptadw,xurbptup,xurbptdw,xedn,nel,PCh,D50,kre,km,rho,Surbt,Iurbt,Rurbt
,Kds,m,Im,w,PC,Fcg,Fch);
    % Weigh the separate slope estimates for classic RK4
    dCdt = (1/6)*(k1dCdt + 2*(k2dCdt + k3dCdt) + k4dCdt);
    dSdt = (1/6)*(k1dSdt + 2*(k2dSdt + k3dSdt) + k4dSdt);
    dIdt = (1/6)*(k1dIdt + 2*(k2dIdt + k3dIdt) + k4dIdt);
    dRdt = (1/6)*(k1dRdt + 2*(k2dRdt + k3dRdt) + k4dRdt);
```

```
    %keep track of real time
    elapsed = floor(dt*24*3600*1e3);
    ta = addtodate(ta,elapsed,'millisecond');
    t1 = datestr(ta);

    % Implement derivatives
    t = t + dt;
    c = c + dCdt*dt;
    Surbt = Surbt + dSdt*dt;
    Iurbt = Iurbt + dIdt*dt;
    Rurbt = Rurbt + dRdt*dt;

    % add to cummulative list
    conc = [conc;c];
    Surb = [Surb Surbt];
    Iurb = [Iurb Iurbt];
    Rurb = [Rurb Rurbt];
    tr = [tr t];
    Is = [Is I];

    %Terminates loop at tf
    if (t>=tf), break, end
    count = count+1;

end

% Get rid of first row placeholders
tr(1) = []; Is(1) = [];
Surb(:,1) = []; Iurb(:,1) = []; Rurb(:,1) = [];
time = tr;

end
```

```
function [dCdt dSdt dIdt dRdt I] =
deriva(t,ta,trise,tset,c,ns,Vol,H,bc,Qdown,Qout,Kd,wpt,nurb,pop,xurbptaup,xur
bptadw,xurbptup,xurbptdw,xedn,nel,PCh,D50,kre,km,rho,Surbt,Iurbt,Rurbt,Kds,m,
Im,w,PC,Fcg,Fch)
% derivative function
% Yukinobu Tanimoto
% Tufts University
% Filename: deriva.m
% Calculates derivatives needed for hydroepidemiological model
% Input:
%  t = model time, spans from ts:tf with step dt
%  ta = t in datestr format
%  trise = sunrise in Gregorian format '05-Jul-2013 00:00:00'
%  tset = sunset in Gregorian format '05-Jul-2013 00:00:00'
%  c = vector of concentration for all cells, with c(1) from bs
%  ns = array of number of cells per reach
%  Vol = array of volumes of cells
%  H = array of height of cells
%  bc = boundary condition table
%  Qdown = array of flows going into downstream cell
%  Qout = array of total flows coming out of cell
```

```
%  Kd = pathogen in stream attenuation (1/day)
%  wpt = array of loadings per cell
%  nurb = number of communities along river
%  pop = array of population for nurb
%  xurbptaup = city withdrawal location (beginning)
%  xurbptadw = city withdrawal location ()
%  xurbptup = city discharge location (beginning)
%  xurbptdw = city discharge location (end)
%  xedn = vector of reaches
%  nel = number of reactors, should be same as sum(ns)
%  PCh = per cap human consumption (L/cap/day)
%  D50 = infectious dose 50 (org/day)
%  kre = pop recovery rate (I->R)
%  km = pop Infected mortality rate (I->death)
%  rho = shedding rate (org/cap/day)
%  Surbt = vector of susceptible individuals, length nurb
%  Iurbt = vector of infected individuals, length nurb
%  Rurbt = vector of recovered individuals, length nurb
%  Kds = linear sorption coeff
%  m = river sediment load (mg/L)
%  Im = peak insolation for that day
%  w = angular frequency of sunrise/sunset for the day
%  PC = per cap water use (L/cap/day)
%  Fcg = fraction graywater use
%  Fch = fraction human consumptive use
% Output:
%  dCdt = vector of concentration derivatives ns long
%  dSdt = susceptible derivatives nurb long
%  dIdt = infected derivatives nurb long
%  dRdt = recovered derivatives nurb long
%  I = insolation at that instant

%initialize vars
dCdt(1:sum(ns)) = 0;
if nurb>0
    dSdt(1:nurb) = 0;
    dIdt(1:nurb) = 0;
    dRdt(1:nurb) = 0;
end

%Settling
Fp = (Kds*m)/(1+Kds*m);     %fraction pathogen sorbed
Vs = 0.4;                   % settling velocity

% Photolysis
alpha =1;
Ke = 0.55*m;
Kel = 0;
I = 0;

% Time calculations
sunrise = datenum(trise);
timenow = datenum(ta);
sunset = datenum(tset);
elapsedhrs = 0;
```

```
% Apply sunlight if sun is up
if (sunrise<=timenow & timenow<=sunset)
    z = 'daytime!';
    elapsedhrs = etime(datevec(ta),datevec(trise))/3600;
    I = Im*sin(w*elapsedhrs);
end

% -------------calculate derivatives-----------------------
% start with i=2 because of indexing from 1 in Matlab

% River concentration
for i=2:sum(ns)
    Kes = Fp*Vs/H(i); %Sorption k
    if I>0
        Kel = alpha*I/24*(1-exp(-Ke*H(i)))/(Ke*H(i));
    end
    dCdt(i) = (1/Vol(i)) * (Qdown(i-1).*c(i-1)- (Kd+Kes+Kel)*Vol(i)*c(i) -
Qout(i).*c(i))+ wpt(i)./Vol(i);
end

curb(1:nurb) = 0;

% Determine concentration of water by city
for i=1:nurb
    diffux  = 0;
    for j=2:nel+1
        if(xedn(j-1)>=xurbptaup(i) & xurbptaup(i)>xedn(j))
            upns(i) = j-1;
            diffux = [xurbptaup(i)-xedn(j)];
        elseif xurbptaup(i)>xedn(j) & xedn(j)>xurbptadw(i)
            diffux = [diffux xedn(j-1)-xedn(j)];
        elseif(xedn(j-1)>xurbptadw(i) & xurbptadw(i)>=xedn(j))
            dwns(i) = j-1;
            if (xurbptadw(i)==xedn(j))
                diffux = [diffux  xedn(j) - xedn(j+1)];
            else
                diffux = [diffux xurbptadw(i)-xedn(j)];
            end
        end

        if(xedn(j-1)>=xurbptaup(i)&xurbptadw(i)>=xedn(j))
            diffux =[xurbptaup(i)-xurbptadw(i)];
            upns(i) = j-1;
            dwns(i) = j-1;
            break;
        end

    end

    Qtotal = pop(i)*PC/1000/24/3600;
    if(diffux==0)
        dosemass = sum(Qtotal*c(upns(i):dwns(i)));
    else
        Qperx = Qtotal/(xurbptaup(i)-xurbptadw(i));
        dosemass = sum(c(upns(i):dwns(i)).*Qperx.*diffux);
```

```
    end

    curb(i) = dosemass/Qtotal;

end

% Calculate SIR derivatives
dose = PCh*curb/1000;
ki = dose./(D50+dose);
dSdt = -ki'.*Surbt;
dIdt = ki'.*Surbt-(kre+km)*Iurbt;
dRdt = kre*Iurbt;

% City Discharge/Loadings
for i=1:nurb
    diffux  = 0;
    for j=2:nel+1
        if(xedn(j-1)>=xurbptup(i) & xurbptup(i)>xedn(j))
            upns(i) = j-1;
            diffux = [xurbptup(i)-xedn(j)];
        elseif xurbptup(i)>xedn(j) & xedn(j)>xurbptdw(i)
            diffux = [diffux xedn(j-1)-xedn(j)];
        elseif(xedn(j-1)>xurbptdw(i) & xurbptdw(i)>=xedn(j))
            dwns(i) = j-1;
            if (xurbptdw(i)==xedn(j))
                diffux = [diffux  xedn(j) - xedn(j+1)];
            else
                diffux = [diffux xurbptdw(i)-xedn(j)];
            end
        end

        if(xedn(j-1)>=xurbptup(i)&xurbptdw(i)>=xedn(j))
            diffux =[xurbptup(i)-xurbptdw(i)];
            upns(i) = j-1;
            dwns(i) = j-1;
            break;
        end
    end

    if(diffux==0)
       dCdt(j) = dCdt(j) + rho*Iurbt(i)/Vol(j);
    else
        exitmass = rho*Iurbt(i);
        Qtotal = (pop(i)*(PC-PCh)*(1-Fcg)+pop(i)*PCh*(1-Fch))/1000/24/3600;
        Qperx = Qtotal/(xurbptup(i)-xurbptdw(i));

        % factor in loading from cities into concentration derivatives
        dCdt(upns(i):dwns(i)) = dCdt(upns(i):dwns(i)) +
exitmass.*(Qperx.*diffux)./Qtotal./Vol(upns(i):dwns(i));
    end

end
```

```
function [Io,f] = InsolationCalc(t1,Lat_d,Long_d)
% mean insolation radiation sunlight
```

```
% Yukinobu Tanimoto
% Tufts University
% Filename: InsolationCalc.m
% Calculates daily average insolation for a specific time and date + location
% Input:
%  t1 = time and date in format '05-Jul-2013 00:00:00'
%  Lat_d = latitude in degrees
%  Long_d = longitude in degrees
% Output:
%  Io = mean average insolation in Langley/day
%  f = photoperiod calculated in hours

ta = datevec(t1);

t0 = ta; t0(2:3) = 1; t0(4:6) = 0; % t0 = Jan 1, same year as t
Dy = etime(ta,t0)/3600/24+1;       % Day in year

% Function written for radian coordinates, convert to rad
Lat = Lat_d*pi()/180;
Long = Long_d*pi()/180;

Wo = 1362;  % Solar const (langley/d)

% zeta - solar declination(radians)
zeta = 23.45*cos(2*pi()/365*(172-Dy))*pi()/180;

% r - normalized radius of earth's orbit
r = 1+0.017*cos(2*pi()/365*(186-Dy));

% Sin(alpha) where alpha = solar altitude in radians
sinalph = sin(zeta)*sin(Lat)+cos(zeta)*cos(Lat);

Io = Wo * sinalph / r^2;

% Photoperiod
L = -tan(zeta)*tan(Lat);
f = (2/15)*(180/pi())*(atan(-L/sqrt(1-L*L))+2*atan(1));

end
```

```
function[trise, tnoon, tset] = SunTimes(t0,Lat_d,Long_d,GMToffset)
% sunrise sunset calculator solar position
% Yukinobu Tanimoto
% Tufts University
% Filename: SunTimes.m
% Function that determines the sunrise, sunset, and solar noon
% for a latitude and longitude
% Input:
%  t0 = value of the date in format '05-Jul-2013 00:00:00'
%  Lat_d = latitude in degrees
%  Long_d = longitude in degrees
%  GMToffset = difference from UTC. ie -4 for Boston;
% Output:
```

```
%  trise = sunrise in Gregorian format '05-Jul-2013 00:00:00'
%  tnoon = solar noon in Gregorian format '05-Jul-2013 00:00:00'
%  tset = sunset in Gregorian format '05-Jul-2013 00:00:00'

Lat_deg = Lat_d;
Long_deg = Long_d;

ta = datevec(t0);
t1 = ta; t1(4:6) = 0;   % sets date to midnight for Julian day calc
Jd = (JulianDays(t1));
Jdn = round(JulianDays(t1)+0.5);

% Calculate the Julian Cycle
nstar = (Jdn - 2451545 - 0.0009) - Long_deg/360;
n = round(nstar);

% Approximate the Julian date for solar noon
Jstar = 2451545 + 0.0009 + Long_deg/360 + n;

% Calculate mean solar anomaly
M = mod(357.5291 + 0.98560028 * (Jstar - 2451545),360);

% Calculate equation of center of sun
C = (1.9148 * sind(M)) + (0.0200 * sind(2 * M)) + (0.0003 * sind(3 * M));

% Calculate ecliptical longitude of the sun
lam = mod(M + 102.9372 + C + 180,360);

% Accurate Julian date for solar noon
Jtrans = Jstar+(0.0053*sind(M)) - (0.0069*sind(2*lam));

% Declination of the sun
dec = asind(sind(lam)*sind(23.45));

% Hour angle of the sun
H = acosd((sind(-0.83)-sind(Lat_deg)*sind(dec))/(cosd(Lat_deg)*cosd(dec)));

% Approximate the Julian date for solar noon again
Jss = 2451545 + 0.0009 + ((H + Long_deg)/360) + n;

% Calculate sunset and sunrise
Jset = Jss + (0.0053 * sind(M)) - (0.0069 * sind(2 * lam));
Jrise = Jtrans - (Jset - Jtrans);

% convert Julian days times to Gregorian
trise = JulianToGregorian(Jrise,t0,GMToffset);
tnoon = JulianToGregorian(Jtrans,t0,GMToffset);
tset = JulianToGregorian(Jset,t0,GMToffset);
```

```
function [lightTime] = dayLength(tset,trise)
% day length time difference
% Yukinobu Tanimoto
% Tufts University
```

```
% Filename: dayLength.m
% Function that calculates the elapsed time between sunrise and sunset
% Input:
%  tset = sunset time in MATLAB datestr format i.e. 'Jan-25-1992 07:30:00'
%  trise = sunrise time in MATLAB datestr format i.e. 'Jan-25-1992 07:30:00'
% Output:
%  lightTime = elapsed hours between trise and tset

e = etime(datevec(tset),datevec(trise));
lightTime = e/3600;

end
```

```
function[jdays] = JulianDays(t)
% Julian days gregorian converter date time
% Yukinobu Tanimoto
% Tufts University
% Filename: Juliandays.m
% Calculates Julian Days from a Gregorian Date
% Input:
%  t = time and date in format [yyyy mm dd hr:min:s]
% Output:
%  jdays = Julian date equivalent of time t

year = t(1);
month = t(2);
day = t(3);

leapyear = 0;
jtotal =2451545;

for i=2000:year-1
    if mod(i,400)==0
        leapyear = 1;
    elseif mod(i,100)==0
        leapyear = 1;
    elseif mod(i,4)==0
        leapyear = 1;
    else
        leapyear = 0;
    end
    mnthdays = [31 28+leapyear 31 30 31 30 31 31 30 31 30 31];

    jtotal = jtotal+sum(mnthdays);
end

 for i=1:month-1
     jtotal = jtotal + mnthdays(i);
 end

 jtotal = jtotal + day;
 jdays = jtotal-1.5;
```

```
function [gregor] = JulianToGregorian(Jd,t0,GMToffset)
% julian days gregorian calendar
% Yukinobu Tanimoto
% Tufts University
% Filename: JulianToGregorian.m
% Function that  converts Julian Days to the Gregorian calendar with
% specified GMT Offset
% Input:
%  Jd = Julian Days (days from midnight of year) to be converted to Gregorian
%  t0 = midnight value of the year of Jd in format '05-Jul-2013 00:00:00'
%  GMToffset = difference from UTC. ie -4 for Boston;
% Output:
%  gregor = Gregorian day in format '05-Jul-2013 00:00:00'

ta = datenum(t0);
Jdo = JulianDays(datevec(datestr(ta)));

temp = Jd - Jdo;
temp = temp * 24;
temp = temp+GMToffset;
hour = floor(temp);
minute = round(60*mod(temp,1));

gregor = datestr(addtodate(ta,minute+hour*60,'minute'));

end
```

```
function [yy] = linterp(x,y,xx)
% linear interpolation
% Yukinobu Tanimoto
% Tufts University
% Filename: linterp.m
% Function that will use linear interpolation to return a value
% Input:
%  x = vector of x values
%  y = vector of y values
%  xx = desired x value for which function will return y value
% Output:
%  yy = interpolated value at x=xx

if xx>max(x)|xx<min(x)
    fprintf('out of range\n');
end

for i=1:length(x)-1
    if xx>=x(i) & xx <=x(i+1)
        yy = y(i) + (y(i+1)-y(i))*(xx-x(i))/(x(i+1)-x(i));
    end
end

end
```