DETECTION AND MANAGEMENT OF VIRUS SOURCES IN AN AQUIFER UNDER EQUILIBRIUM AND KINETIC SORPTION

Thesis submitted in partial fulfillment of the requirements

for the award of the degree of

DOCTOR OF PHILOSOPHY

in

Civil Engineering

by

Bandaru Goutham Rajeev Gandhi
Roll No. 166104014

Under the guidance of

Prof. Rajib Kumar Bhattacharjya
(Professor)

Department of Civil Engineering
Indian Institute of Technology Guwahati
Guwahati - 781039, Assam, India
August, 2022

DETECTION AND MANAGEMENT OF VIRUS SOURCES IN AN AQUIFER UNDER EQUILIBRIUM AND KINETIC SORPTION

Thesis submitted in partial fulfillment of the requirements

for the award of the degree of

DOCTOR OF PHILOSOPHY

in

Civil Engineering

by

Bandaru Goutham Rajeev Gandhi
Roll No. 166104014

Under the guidance of

Prof. Rajib Kumar Bhattacharjya
(Professor)

Department of Civil Engineering
Indian Institute of Technology Guwahati
Guwahati - 781039, Assam, India
August, 2022

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati
Guwahati – 781 039
Assam, INDIA

Phones: 0361-2582428 (O)
0361-2584428 (R)
Fax: 0361-2582440
E-mail: rkbc@iitg.ac.in
rajibkbc@gmail.com

DEPARTMENT OF CIVIL ENGINEERING

Dr. Rajib Kumar Bhattacharjya
Professor

Date: 1 February 2023

CERTIFICATE

This is to certify that the work contained in the thesis entitled "**Detection and Management of Virus Sources in an Aquifer under Equilibrium and Kinetic Sorption**" submitted by **Bandaru Goutham Rajeev Gandhi** to the Department of Civil Engineering, Indian Institute of Technology Guwahati for the award of the degree of Doctor of Philosophy in Civil Engineering is a record of bonafide research work carried out by him under my supervision and guidance. This thesis work, in my opinion, has reached the requisite standard fulfilling the requirements for the degree of Doctor of Philosophy.

The research work contained in this thesis has not been submitted in part or full to any other University or Institute for the award of any degree or diploma.

Place: Guwahati

(Rajib Kumar Bhattacharjya)

DECLARATION

I do hereby declare that the thesis entitled **"Detection and Management of Virus Sources in an Aquifer under Equilibrium and Kinetic Sorption"** has been performed by me for the degree of Doctor of Philosophy in civil engineering under the guidance of Prof. Rajib Kumar Bhattacharjya, Department of Civil engineering, Indian Institute of Technology Guwahati, Assam, India.

This work has not been submitted for the award of any degree or any diploma at this Institute or other Institute. I have also acknowledged whenever the work describes the findings based on the literature review.

Date: 01/02/2023

Place: Guwahati

B. G. Rajeev Gandhi

(Bandaru Goutham Rajeev Gandhi)

Dedicated to

My Family

ACKNOWLEDGEMENTS

I am deeply indebted and grateful to my supervisor Prof. Rajib Kumar Bhattacharjya, without whose voracious knowledge, sage guidance, and invaluable intellectual inputs, this research endeavor would not have been possible.

I would also like to thank the chairman of my doctor committee, Prof. Sajal K. Deb, and the members, Dr. Suresh A. Kartha and Dr. Sreeja P, for their invaluable suggestions and insights regarding my research.

I also extend my gratitude to Dr. Leichombam Sophia, Dr. L.N.V. Satish, and Anant Dubey for assisting me in my work whenever the need arose. I would also like to thank Dr. Anjaneyulu, Dr. Vinay Chembolu, Dr. Dilip Kumar, Dr. Bhrigumani Sharma, Mrs. Mamta Das, and all of my lab mates for their encouragement and support.

Lastly, I thank my mother, B. Vijayeswari, my wife, Bhavagnya Kakara, and my brother, B. Goutham Yasodhar, for accommodating and supporting my research endeavor.

(Bandaru Goutham Rajeev Gandhi)

ABSTRACT

The alarming rate at which groundwater contamination is increasing has motivated several researchers to identify ways to detect and manage the contamination. One such contaminant is a virus that can stay for a prolonged period in the groundwater under favorable conditions. The virus contamination is generally detected by laboratory tests from the samples obtained from monitoring wells installed in the aquifer. However, the virus entering the aquifer is challenging to locate through laboratory tests alone. Therefore, the modeling of virus movement through groundwater aquifer, called a virus transport model, is required to predict the spatial and temporal distribution of virus in an aquifer. A virus as a contaminant is different from the conventional ones as it reacts to the environment to inactivate (disrupt its protein coat) and become harmless. However, a virus is susceptible to change with the slightest environmental changes. Therefore, the model should be capable of handling the changing environmental conditions. Unfortunately, the existing models such as HYDRUS-3D or MT3DMS for contaminant transport are not efficient in handling such changes. Therefore, I have developed a mathematical model that can handle the spatial and temporal variation of contaminant (virus) flow and transport parameters. Once the simulation model is ready, the experimental observations can be matched with the simulated concentrations at the observation locations to identify the virus's sources. However, the sources of the virus should be assumed initially at an arbitrary location and with an arbitrary source strength to obtain the simulated concentrations. Later, minimizing the absolute error between the simulated and the observed concentrations detects the source of the virus in the aquifer. The approach is termed an inverse optimization approach. However, an initial approximation of the source locations accelerates the detection of sources in the optimization model. In order to overcome this limitation, I have developed a methodology to find the probable locations of potential sources of viruses in the aquifer based on the observation data and the flow field of the aquifer in the period of observation. Once the probable locations are identified, the inverse optimization model can be solved using an optimization method. The Genetic Algorithm (GA) is efficient in identifying the discrete variables (locations of contamination), and gradient-based classical optimization is efficient in identifying the continuous variables (source strength/flux). Therefore, I have modified GA to accommodate the initial approximation of locations as probable location operators and combined it with a local location search based on the gradient-based search to detect the virus sources in the aquifer. I have found that the algorithm works for the early detection of contamination sources in the aquifer with limited observation data, which is crucial in the case

of viruses. Once the source of the virus was detected, I simulated the extent of contamination in the aquifer using the developed simulation model. Considering the susceptibility of the parameters to the changing environment, I have studied the parameters in the aquifer closely. I developed a pore network model that can simulate the aquifer properties in the field scale from the pore properties. The properties such as intrinsic permeability, porosity, dispersivity, and tortuosity are the functions of pore properties that affect the aquifer field scale properties. Also, the virus adsorption parameters are functions of the pore properties and network arrangements. I have estimated all the relevant field-scale properties from the pore scale properties with their sensitivities to temperature, ionic strength, pore radius, pore water velocity, the radius of the virus, the surface potential of the pore and the virus, the permittivity of the pore space, and viscosity of the medium. Considering the sensitivity of the virus parameters to the environment, I changed the fluid temperature in the aquifer to increase the inactivation rate of the virus for its faster removal. I have proposed changing the fluid temperature by injecting hot water into the aquifer. Finally, I have proposed a management model that minimizes the volume of hot water injected into the aquifer to reduce the concentration to the desired level. I have given the field scale applicability of the proposed management model by solving two theoretical problems of virus contamination using three methods. One solution is injecting hot water into the aquifer by selecting the injection wells with the highest virus concentration. It is an efficient management strategy, but it might not always be feasible for field application. The second solution is to inject hot water into the existing injection wells in the aquifer without digging new wells. However, the volume of water injected increases drastically by following this solution. Therefore, I have proposed a third solution, where the locations of injection wells are from the existing observation wells. After analyzing the results, I have found that all three solutions are feasible depending on the field conditions for their applicability.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS / NOTATIONS

n, θ	Porosity of the porous medium	-
K_{xx}	Principal hydraulic conductivity in the x-direction	m/s
K_{yy}	Principal hydraulic conductivity in the y-direction	m/s
K_{zz}	Principal hydraulic conductivity in the z-direction	m/s
S_s	Specific storage of the porous medium	m^{-1}
α_L	Longitudinal dispersivity of the REV	m
α_{TH}	Transverse horizontal dispersivity of the REV	m
α_{TV}	Transverse vertical dispersivity of the REV	m
D^*	Molecular diffusion coefficient	m^2/s
K_f	Constant of proportionality of Freundlich isotherm	$m^{3/m}kg^{-1/m}$
m	Exponential constant of Freundlich isotherm	-
ρ_b	Bulk density of the porous medium	kg/m^3
μ_l	Inactivation constant in aqueous phase	d^{-1}
$\mu_{s,eq}$	Inactivation constant in sorbed phase for equilibrium sorption	d^{-1}
$\mu_{s,kin}$	Inactivation constant in sorbed phase for kinetic sorption	d^{-1}
k_{att}	Attachment coefficient	d^{-1}
k_{det}	Detachment coefficient	d^{-1}
$C_{p,solid}$	Specific heat capacity of the solids in porous medium	$kg^{-1}\ ^0C^{-1}$
$C_{p,fluid}$	Specific heat capacity of the fluids in porous medium	$kg^{-1}\ ^0C^{-1}$
ρ_s	Density of the solids in porous medium	kg/m^3
κ	Thermal diffusion coefficient	m^2/s
κ	inverse Debye-Huckel length	m^{-1}
N_A	Avogadro number	-
k_B	Boltzmann constant	J/K
e	Charge of an electron	C
ε_0	Permittivity of vacuum	$C^2/J/m$
ε	dielectric constant of water	-
λ	Characteristic wavelength	m

a	Radius of the nano particle (virus)	m
R_{ij}	Radius of the pore throat 'ij'	m
R_i	Radius of the pore body 'i'	m
vm_{ij}	Mean velocity of the pore throat 'ij'	m/s
ψ_1	Surface potential of the virus surface	J/C
ψ_2	Surface potential of the collector surface	J/C
I	Solution ionic strength	mM
μ	Viscosity of the medium	Nm/s^2
T	Temperature	K
C	Concentration	mg/l
C_0	Initial concentration / Maximum concentration	mg/l
S_{eq}	Equilibrium sorbed concentration ratio	-
S_{kin}	Kinetic sorbed concentration ratio	-
q_s	Discharge of the source/sink	m^3/s
h	Piezometric head	m

LIST OF ABBREVIATIONS

CFT	Colloid Filtration Theory
DNA	Deoxy-Ribose Nucleic Acid
DLVO	Derjaguin-Landau-Verwey-Overbeek
FDM	Finite Difference Method
FEM	Finite Element Method
FVM	Finite Volume Method
GA	Genetic Algorithm
GMS	Groundwater Modeling System
HIV	Human Immuno-deficiency Virus
LLS	Local Location Search
LTS	Longitudinal Transverse Search
MATLAB	Matrix Laboratory
MODFLOW	Modular Finite Difference Groundwater Flow Model
MT3DMS	Modular Three-Dimensional Multi Species Transport Model
NSE	Nash Sutcliffe Efficiency
OC	Observed Concentration
PCG	Preconditioned Conjugate Gradient
PNM	Pore Network Model
REV	Representative Elementary Volume
RMSE	Root Mean Square Error
RNA	Ribose Nucleic Acid
SARS-COV-2	Severe Acute Respiratory Syndrome – Corona Virus – 2
SC	Simulated Concentration
SFLA	Shuffled Frog Leaping Algorithm
USGS	United States Geological Survey
USEPA	United States Environmental Protection Agency

CHAPTER 1: Introduction

1.1 General

The purest and the most abundant form of water in freshwater is groundwater. Groundwater comprises 30.1 % of the freshwater, 0.296 % is in the lakes and rivers, and the rest of the water is in the glaciers and polar ice caps (Chow 2010). The recent technological advancements and rapid urbanization have also increased groundwater contamination at an alarming rate. Groundwater contamination can happen in many ways, such as heavy metal contamination, increased ionic concentrations, pH level variations, radioactive contamination, and bacterial or virus contamination. The contamination at any level with such contaminants poses a health threat to the people using the contaminated groundwater for drinking and various domestic purposes. The groundwater contamination can thus be a health hazard depending on the type of source which contaminated the groundwater.

1.2 Virus Sources as Contaminants

Viruses have existed from the very beginning of evolution since there has been the existence of living cells. Viruses are the most crucial part of evolution as they provide the required genetic diversity among the species through horizontal and vertical gene transfer. However, some viruses are very harmful and can infect any animal and plant species once they enter a living cell. Viruses such as Ebola, HIV, Hepatitis B and C, Adenovirus, Poliovirus, SARS-COV-2, etc., still exist in this environment either in cells of living organisms or in the outer environments such as water, air, etc. Viruses such as smallpox, chicken pox, cholera, and flu viruses have created major outbreaks throughout the world, and some of these diseases have been there for centuries and have been eradicated recently.

The viruses as contaminants enter groundwater through leaking septic tanks and sewer lines, sanitary landfills, and land application of wastewater and sludge from treatment plants. (Seetha et al. 2014). They enter through the vadose zone (unsaturated zone) or the fractured rocks and enter the aquifer region, a fresh groundwater source. The detailed movement of the viruses into the saturated aquifer is given in Figure 1.1. Virus in groundwater does not get multiplied inside the groundwater alone; they replicate by themselves when they find a host in the form of a living cell. Otherwise, they die off frequently due to various reasons such as groundwater's pH

or the environment's temperature. But due to their tiny size (20-350 nm), they can travel very long distances if the conditions are favorable (Loveland et al. 1996).

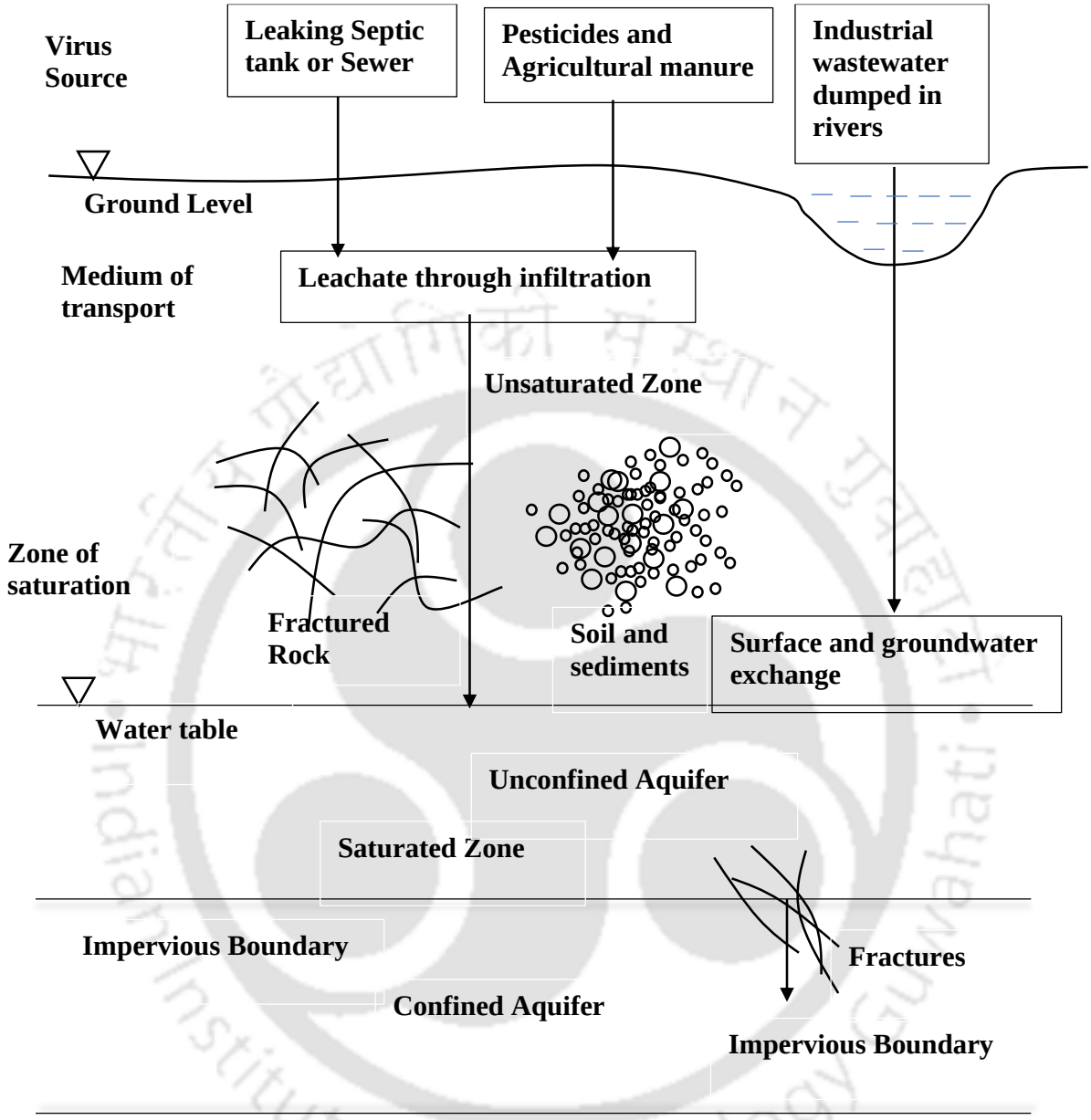

Figure 1.1: Illustrative diagram indicating the movement of virus from different sources to the saturated zone, i.e., unconfined and confined aquifers.

1.3 Modeling the Virus Transport in an Aquifer

The contaminant flow and transport processes in an aquifer can be simulated by solving the governing flow and transport equations. As the closed form solutions are not available for field scale problems, the equations are generally solved using numerical methods such as the Finite Difference Method (FDM), Finite Volume Method (FVM), and Finite Element Method (FEM). Several models are available in the public domain or on commercial platforms (MT3DMS and HYDRUS 3D) (Bedekar et al. 2016; Šimunek et al. 2012) for simulating the flow and transport

processes. However, there are some limitations to the available models as they cannot handle the temporal variation in the flow and transport parameters. Models like HYDRUS 3D account for the parameters' dependency on temperature or water content, but the user cannot specify the dependency. Some contaminants such as viruses or bacteriophages are susceptible to changes in temperature, pH, and other fluid properties in the aquifer medium. The solution chemistry also changes with the aquifer environment temporally. Therefore, there is a need to address the temporal dependency of such sensitive parameters.

1.4 Pore-scale Modeling of Virus Contamination in an Aquifer

Modeling the transport of virus after that gives a perfect idea of the contaminants moving through the aquifer, and the management policy can be modeled using different techniques. But, there is a need to understand the whole transport process of viruses in order to develop a mathematical model. The significant parameters that are different for the virus transport from the primary contaminant transport are inactivation rates in the aqueous and equilibrium sorption phase and attachment and detachment rates in the kinetic sorption phase. All these parameters are highly dependent on many parameters such as flow velocity, viscosity, temperature, solution dielectric constant, solution dielectric strength, and the nanoscale parameters such as pore radius, nanoparticle radius, and nanoparticle surface potential and collector surface potentials (Seetha et al. 2015). However, the mathematical modeling for virus transport can be carried out in the Darcy scale by knowing the exact values of the parameters controlling virus transport. Later the nanoscale transport modeling can be applied to know the interdependencies of the flow parameters in the pore scale to the Darcy scale. The pore-scale modeling can be done using the pore network model considering cylindrical pores and spherical bodies (Raoof and Majid Hassanizadeh 2010).

1.5 Detection of Virus Contamination in an Aquifer

Contaminant transport in groundwater is increasing daily and is causing a severe problem to the groundwater environment. The modeling or management of contaminant transport for field scale problems always faces an issue of contaminant sources. Generally, the modeler is only presented with data that a particular contaminant is found in some observation wells. Based on the data given, the modeler has to identify the location and strength of the contaminant source. Such a problem is called the source identification problem (Leichombam and Bhattacharjya 2019). The problem is solved using the inverse optimization technique, where an arbitrary value is assigned as a contaminant source, and the error between the observed and simulated

concentrations is minimized. However, the source identification needs to address the issues such as the early detection of viruses in the aquifer and finding the source with limited observation data.

1.6 Management of Virus in a Contaminated Aquifer

Remediation of the contaminated aquifer is complex, and removing a contaminant entirely from groundwater is almost impossible. A cost and time-effective management strategy have to be suggested to decontaminate the aquifer from a virus concentration to the desired level. Therefore, there is a need to know the source of the virus, the exact location, and its strength. A practical source identification strategy has to be designed for an aquifer system that can identify viruses' location and flux in the aquifer. Viruses can enter an aquifer by different means, such as dumping pathogenic water into landfill sites, leaks from a sewer pipe, or leaks from septic tanks. Once the virus enters an aquifer, it takes a considerable amount of time to get to a non-harmful concentration as the rate of die-off of viruses is generally very low. The die-off rate of viruses depends on many factors influencing the inactivation rates in the aqueous and attached phase. The die-off rate can be altered by varying the conditions of the aquifer, such as pH and temperature. The present study deals with removing viruses in the aquifer to a specified concentration level by varying the temperature. The virus concentration is calculated by altering the temperature and designing an optimal management strategy in an unconfined aquifer. The management problem is designed as a minimization problem to minimize the volume of water injected into the management policy. The analysis involved solving the Groundwater flow equation, the heat transport equation, followed by the virus transport equation dynamically in every time step.

1.7 Research Objectives

The following objectives are proposed considering the previous research and the problems encountered with the virus transport in groundwater.

- To develop a numerical simulation model in Darcy's scale that accounts for all the heterogeneities in aquifer medium and other factors such as temperature, pH, and solution chemistry, considering the spatial and temporal variation of the parameters.
- To develop a pore-scale model that translates the pore-network characteristics to the Darcy-scale models.

- To translate the parameters in the aquifer medium from the pore-scale to Darcy's scale by solving the pore network flow and correlation equations for a single pore.
- To develop an inverse optimization model that can identify the potential source strengths of the virus.
- To develop an optimal management strategy by altering the parameters that affect viruses' deposition and inactivation through Darcy scale models.

The linkages of the objectives to get the overall objective of my thesis, the detection, and management of virus sources in an aquifer using equilibrium and kinetic sorptions, are given in Figure 1.2.

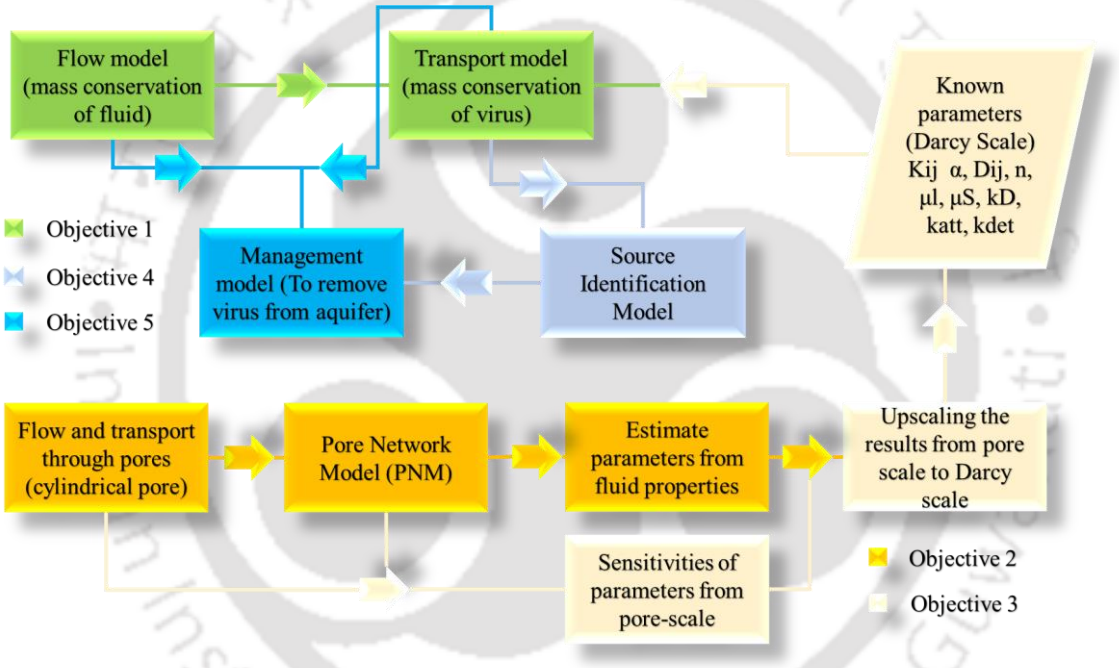

Figure 1.2: Overview of my thesis and the linkages of all the objectives of my thesis

1.8 Organization of the Thesis

There are eight chapters in the thesis, including the present introduction chapter. A brief overview of each of the chapters is given as follows.

Chapter 1: In this chapter, I have concisely discussed the groundwater contamination, the threats the virus poses to the population as a contaminant, and the nuances of modeling the virus transport. The motivation for my research work is presented in this chapter and also enlisted the research objectives.

Chapter 2: In this chapter, I provided a detailed discussion of the previous research works for virus flow and transport. I have showcased the necessity of modeling for both Darcy-scale and pore-scale models and their possible linkages. I have reported the details of the advancement of source identification techniques over recent years. I have given the virus's susceptibility to various environmental factors and provided a possibility for a management model.

Chapter 3: In this chapter, I developed the contaminant flow and transport models in groundwater that can handle the spatial heterogeneities in the aquifer and the temporal changes in the parameters. I have validated the developed Darcy-scaled models with the existing models of MODFLOW and MT3DMS and presented the results.

Chapter 4: In this chapter, I developed a pore-scale model assuming the cylindrical pore throats and spherical pore bodies. Each pore-scale model represents a Representative Elementary Volume (REV), and the model can produce the desired Darcy-scale parameters from the pore-network properties.

Chapter 5: In this chapter, I used the correlation equations of a single pore to estimate the pore-scale properties of virus deposition and adsorption. Later, I upscaled the parameters to the Darcy scale by analyzing the parameter sensitivities with the pore properties.

Chapter 6: In this chapter, I modeled the virus source identification with the information of the observation data of the virus concentration in the aquifer. I have assumed the source identification model's number, location of sources, source strengths, and their period of activity to be unknown. I developed a strategy to find an initial probabilistic space that has the potential to be the source of contamination and used it to modify the genetic algorithm through a probable location operator. I later devised a local search algorithm, which is gradient-based, to pinpoint the location and the source strength of the contamination.

Chapter 7: In this chapter, I proposed an optimal management strategy for concentration reduction by injecting hot water into the aquifer. I have optimized the volume of water injected to a minimum value while obtaining the desired concentration level in the aquifer. I have solved two theoretical problems of industrial effluent and sewer leakage by injecting hot water to remediate the aquifer. I have proposed the management strategy for both the equilibrium and the kinetic sorption cases.

Chapter 8: In this chapter, I have summarized the research and briefly presented all the conclusions of my research. Lastly, I have presented the future scope of the study and the possibility of application of the proposed strategy in the field along with alternative management strategies.

CHAPTER 2: LITERATURE REVIEW

2.1 General

Various researchers have studied viruses as contaminants in groundwater ever since there were outbreaks of epidemics due to virus contamination. Researchers have made several attempts over the past few decades to understand the process of virus contamination transport in groundwater. The following literature highlights the methodologies researchers adopted to understand virus transport, source identification, and management of identified sources.

2.2 Effect of virus-contaminated groundwater

Research investigation done on virus transport confirmed that viruses can travel from the source of contamination through the unsaturated soil to groundwater, where they can enter the drinking water and ultimately infect humans (Gerba 1984; Gerba et al. 1981; Keswick and Gerba 1980). Because the virus can survive in soil with the disposal of wastage, there needs to be a proper evaluation of the risk of virus survival in soil and the factors affecting the virus persistence in soil (Gerba et al. 1981). It was observed that most of the virus originates from human sewage from nearby municipal wastewater discharges, septic tanks and sanitary landfills (Keswick and Gerba 1980). As per centres for disease control and prevention (Gerba 2004), 47 % of diseases are due to unknown actions, 18 % are due to bacteria, 16 % are due to parasites, 11 % are due to chemicals, and 8 % are due to viruses in groundwater. The viruses and bacteria in groundwater are causing almost 26 % of diseases in the people consuming groundwater. Viruses as contaminants are hazardous, and the potable water should be free from viruses up to 99.99% according to the Groundwater rule (the United States Environmental Protection Agency as of 2006).

2.3 Transport of Virus in Groundwater (Darcy Scale)

The transport of viruses can be defined under 4 major processes i) Advection ii) Diffusion iii) Deposition and iv) Inactivation. The first two processes come under a general contaminant transport called transport. The deposition accounts for the transfer of viruses from the bulk fluid to the solid matrix and vice versa, called attachment and detachment. The inactivation is the death of viruses in an aqueous medium or on the solids (Tim and Mostaghimi 1991). The

appropriate mathematical processes must be considered to model the fate and transport of viruses in porous media.

2.3.1 Advection Dispersion Equation

The advection is caused by the natural pore water velocity (v_i) existing in the porous media. The diffusion is governed by Fick's law and is considered hydrodynamic dispersion (Dij), which combines mechanical and molecular diffusion. The Governing equation for the general advection-dispersion equation is given below in equation 2.1 (Hassanizadeh et al. 2000; Yates and Yates 1988).

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x_i}\left(D_{ij} \frac{\partial C}{\partial x_j} \right) - \frac{\partial}{\partial x_i}(v_i C) \qquad 2.1$$

2.3.2 Adsorption

Viruses are in suspension in the bulk fluid within the pores. In transporting these viruses, they tend to attach to the pore walls (soil grains) by a fast or slow process. The fast attachment process is called adsorption and the slow one is called deposition. The adsorption is called equilibrium sorption and occurs in the later stage of transport after the virus enters the pores. For a particular case, this can be defined as a linear isotherm shown in equation 2.2 (Yates and Yates 1988). The isotherms can be nonlinear in nature, namely Freundlich and Langmuir isotherms, which are also defined in equation 2.2.

$$\frac{\partial C}{\partial t} + \frac{\rho_b}{n}\frac{\partial S_{eq}}{\partial t} = \frac{\partial}{\partial x_i}\left(D_{ij} \frac{\partial C}{\partial x_j} - v_i C \right)$$
$$S_{eq} = K_{eq} C,\ S_{eq} = K_f C^m,\ S_{eq} = \frac{K_l \bar{S} C}{1 + K_l C} \qquad 2.2$$

2.3.3 Deposition

The deposition is a kinetic attachment mechanism which is a two-step process. The first step is viruses getting transferred from the bulk fluid towards the sediment grains. In the second step, viruses are attached to the surface due to Physico-chemical interactions (Bos et al. 1999). The governing equation for this process and the advection-dispersion are given below in equation 2.3 (Hassanizadeh et al. 2000; Yates and Yates 1988).

$$\frac{\partial C}{\partial t}+\frac{\rho_b}{n}\frac{\partial S_{kin}}{\partial t}=\frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial C}{\partial x_j}-v_i C\right)$$
$$\frac{\rho_b}{n}\frac{\partial S_{kin}}{\partial t}=k_{att}C-\frac{\rho_b}{n}k_{\det}S_{kin} \qquad 2.3$$

2.3.4 Inactivation

The inactivation is also called as death or die-off of viruses. Viruses die off for many reasons, such as solution composition and pH, temperature, grazing and attachment to sediment surfaces (Hassanizadeh et al. 2000). The equation governs the inactivation of viruses and all the other processes combined is given in equation 2.4.

$$\frac{\partial C}{\partial t}+\frac{\rho_b}{n}\frac{\partial S_{kin}}{\partial t}+\frac{\rho_b}{n}\frac{\partial S_{eq}}{\partial t}=\frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial C}{\partial x_j}-v_i C\right)-\mu_l C-\frac{\rho_b}{n}\mu_{s,kin}S_{kin}-\frac{\rho_b}{n}\mu_{s,eq}S_{eq}$$
$$\frac{\rho_b}{n}\frac{\partial S_{kin}}{\partial t}=k_{att}C-\frac{\rho_b}{n}k_{\det}S_{kin}-\frac{\rho_b}{n}\mu_{s,kin}S_{kin} \qquad 2.4$$

The inactivation and attachment are explained figuratively in Figure 2.1 (Grant et al. 1993). The researchers (Hassanizadeh et al. 2000) have performed various experiments and model simulations to understand the processes of adsorption (equilibrium sorption), deposition (attachment and detachment) and the process of inactivation (die-off). Figure 2.2, Figure 2.3, and Figure 2.4 overview each of the processes involved. Even considering all these active processes, the transport of viruses is a fairly complex process to understand. The researchers (Bales et al. 1991) have observed in laboratory field scale studies that the attachment of viruses follows a slow kinetic rate with slow rising limbs and long tails at the end of the breakthrough curves.

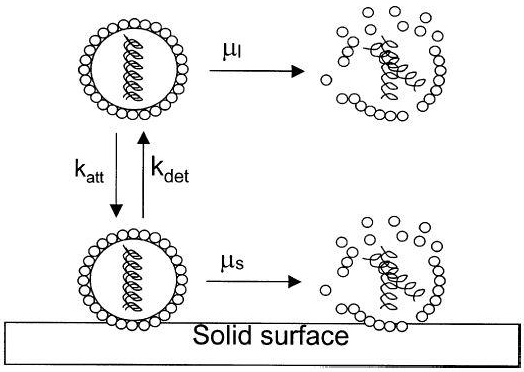

Figure 2.1: Virus inactivation attachment and detachment process (Grant et al. 1993)

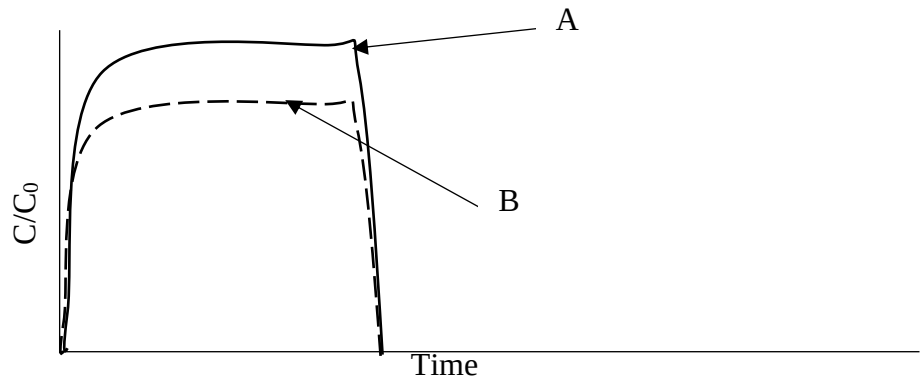

Figure 2.2: Break through curve for A) Conventional salt tracer B) Virus with inactivation and no adsorption (Hassanizadeh et al. 2000)

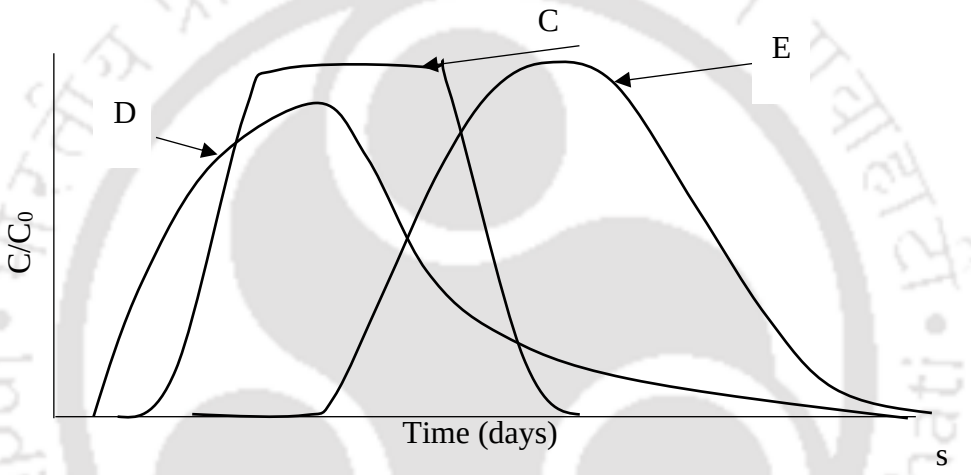

Figure 2.3: Break through curve for C) Virus with equilibrium adsorption and inactivation D) Virus with same adsorption and inactivation as C and more dispersion E) Virus with same inactivation as C with more retardation (Hassanizadeh et al. 2000)

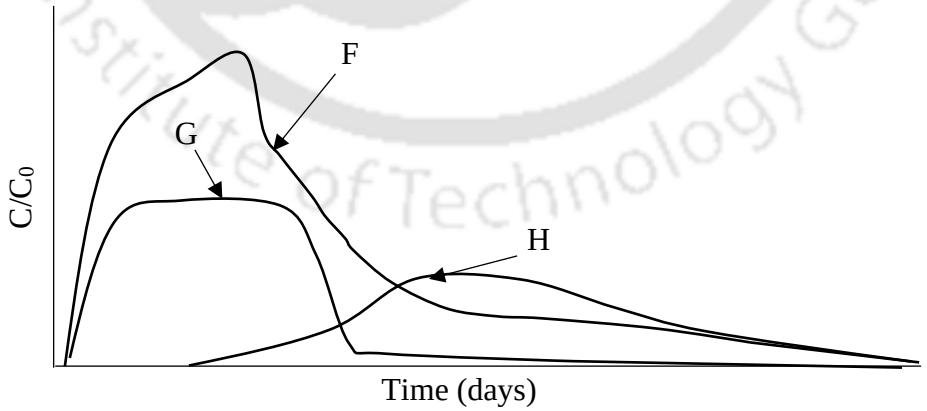

Figure 2.4: Break through curve for F) Kinetic adsorption and inactivation of virus G) same as F with slower kinetic detachment and inactivation H) same as E with higher kinetic attachmment katt/kdet as 10 (Hassanizadeh et al. 2000)

In contrast, some researchers (Hassanizadeh et al. 2000) have shown numerically that modelling the virus under equilibrium and the kinetic sorption might sometimes lead to similar conclusions (shown in Figure 2.2, Figure 2.3, and Figure 2.4). However, there is a need for a better theory to understand the attachment and detachment coefficients. Several theories available are Colloid Filtration Theory (CFT), Derjaguin-Landau-Verwey-Overbeek (DLVO) Theory, Hydrophobic Interactions theory and Blocking theory. Colloid Filtration Theory (CFT) is described based on the researcher (Yao et al. 1971). that the attachment of viruses to the soil surface can be described in two processes i) mass transport to the surface and ii) virus-surface interactions. These interactions are, therefore, accountable on microscale terms where the surface interactions are calculated by collision efficiency (η) and the sticking efficiency (α). According to this theory, a particle suspended may come into contact with the collector (soil surface) by either interception, sedimentation or diffusion (Yao et al. 1971). The attachment rate is defined by equation 2.5, where θ represents the saturation (volumetric), d_c represents the average collision diameter (grain size), the fraction of particles that collide with the collector is η, and v represents the fluid velocity.

$$k_{att} = \frac{3}{2}\frac{(1-\theta)}{d_c}\alpha\eta v \qquad 2.5$$

Generally, experiments have to be conducted to calculate the k_{att} and then the collision efficiency calculated before should be used to calculate the sticking efficiency (α) (Bales et al. 1991; Kinoshita et al. 1993; Penrod et al. 1996). Alternatively, some researchers (Ryan and Elimelech 1996) proposed a set of equations to calculate sticking efficiency by considering irreversible adsorption of viruses through the relative breakthrough curves of conservative salt tracer. DLVO theory proposed by (Derjaguin and Landau 1941; Verwey et al. 1948) suggests that the interactions between the grain surfaces and the viruses are determined by the attractive and the repulsive forces between them. Such interactions can be characterized by a simple interaction energy profile which accounts for the total potential energy for a particle to interact. The total energy is the sum of the double layer potential energy ϕ^{DL}, London van-der Waals potential energy ϕ^{vdW} and the Born potential energy ϕ^{Born}. The total potential energy curve is characterized by an attraction energy well at a minimum separation distance (δ) within a primary minimum (ϕ_{min1}). A repulsive energy barrier exists at ϕ_{max} and a secondary minimum ϕ_{min2} which is shallow attractive energy at a more significant separation distance (Hassanizadeh et al. 2000). Figure 2.5 shows the interaction energy profile as described in DLVO theory with normalized potential energy vs separation distance (m).

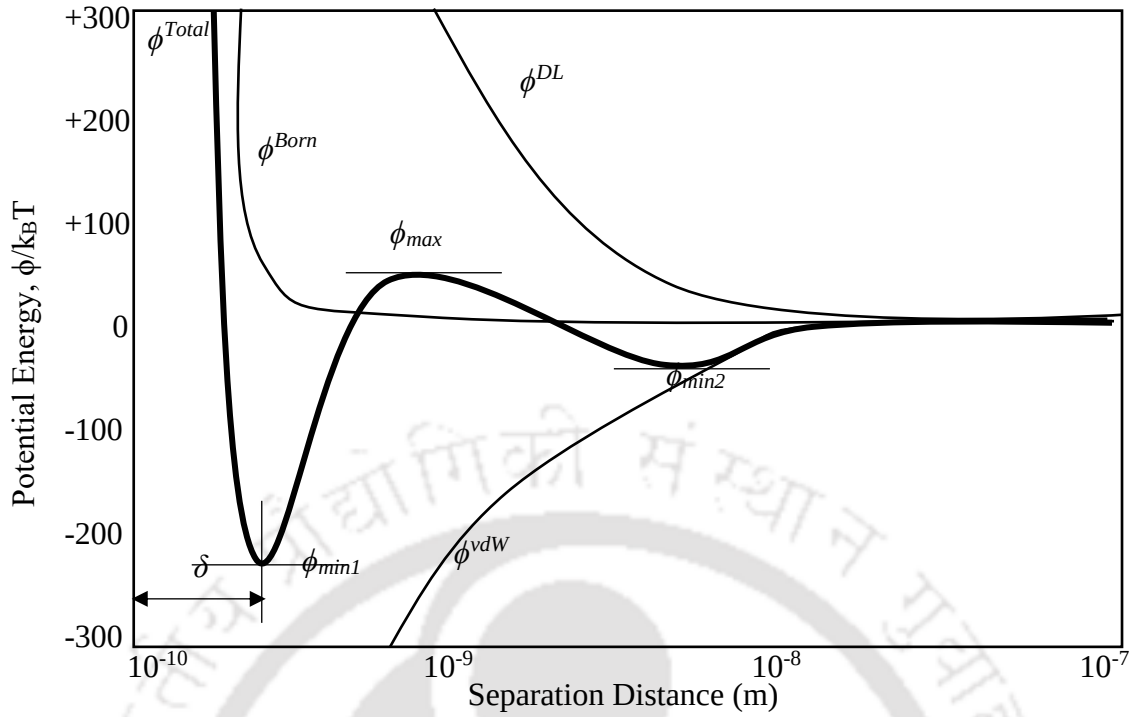

Figure 2.5: DLVO theory interaction energy profile (Ryan and Elimelech 1996)

The particle deposition is predicted based on two theories: Interaction Force Boundary Layer (IFBL) approach (Spielman and Friedlander 1974). This theory suggests that the deposition of particles is only based on the primary minimum, and the secondary minimum is neglected. But experimental studies based on this theory have found that the attachment efficiencies based on the IFBL approach are a few orders lesser than the conventional values obtained theoretically and experimentally (Elimelech and O'Melia 1990; Hahn and O'Melia 2004; Shen et al. 2008). The other approach is the Maxwell approach (Hahn and O'Melia 2004; Shen et al. 2008, 2010). Researchers did all the experiments, and they finally proposed a theory based on the Maxwell approach, which accounts for the primary and secondary minima for viruses smaller than 30 mm and hydrodynamic forces and torque balances for more significant viruses under unfavourable conditions and had better agreement with the experimental values of attachment efficiencies. Hydrophobic interaction theory is not that widely accepted in the research community because it doesn't consider all the factors that are prevailing in the deposition of viruses over solids. The researcher Gerba (Gerba 1984) reviewed that the hydrophobic interactions play a significant role in virus attachment to the surface. When both surfaces are negatively charged at higher pH, hydrophobic interactions may be a significant factor in maintaining the virus attachment (Bales et al. 1991; Gerba 1984; Shields and Farrah 1983). In contrast, Loveland and others (Loveland et al. 1996) have suggested that the increased virus

attachments may be more attributed to a decreased double layer repulsion than hydrophobic expulsion from the solution. Blocking theory suggests that when a particle is attached to a surface of a solid and if the particle-particle interactions are repulsive, the attachment rate decreases with time, and the soil surface becomes progressively occluded as more particles accumulate. Ryan and Elimelech (Ryan and Elimelech 1996) have reviewed the modelling of particle accumulation and concluded that blocking prevails in groundwater, with low ionic strength and hardness levels.

2.4 Transport of Virus in Groundwater (Pore Scale)

Modelling of viruses transport in pore scale gives a better understanding of the processes such as deposition and inactivation, which are quite complex and depends on many factors. The pore-scale visualization of soil samples can be done using X-Ray and Computer Tomography. With the recent advances in non-destructive techniques such as micro Computer Tomography combined with spatial analysis, a complete internal architecture of soil in 3D (Sleutel et al. 2008) along with soil's organic matter can be visualized (Bouckaert et al. 2009). A pore-network model can be complex and computationally challenging but gives a better understanding of all the processes happening in a porous media. A three-dimensional cylindrical pore model is suggested (Seetha et al. 2014), which can simulate flow through a single pore. They have considered DLVO theory for the deposition processes combined with the Maxwell equations for interaction energy calculations. The diffusion predominant in this scale is Brownian diffusion. The mean diameter of the virus particle (assumed spherical in shape) is '*a*' in a pore (cylindrical) of radius R and length L shown in Figure 2.6. The equations governing the process of pore transport in a cylindrical pore for different regions are defined by a researcher (Seetha et al. 2014). The interaction energy ϕ can be calculated as suggested in equation 2.6, based on DLVO theory.

$$\phi^* = \frac{\phi}{k_B T} = \frac{\phi_{EDL}}{k_B T} + \frac{\phi_{VDW}}{k_B T} + \frac{\phi_{Born}}{k_B T} \qquad 2.6$$

Each of the energy calculations is based on different approaches. The Electric double layer energy is calculated based on a Hogg-Healy-Furestenau (HHF) (1966) theory which takes into account the surface potential of the virus and the collector, the dielectric constant of water and vacuum and the ionic strength of the solution (Hogg et al. 1966). The London Van-der-Waals interaction energy is calculated based on the proposed theory, which considers the characteristic wavelength of interaction (Gregory 1981). The Born repulsion energy considers

the collision diameter. This total model has been tested for various sensitivities by researchers and is considered a building block for pore-scale studies in the transport of viruses in a porous medium.

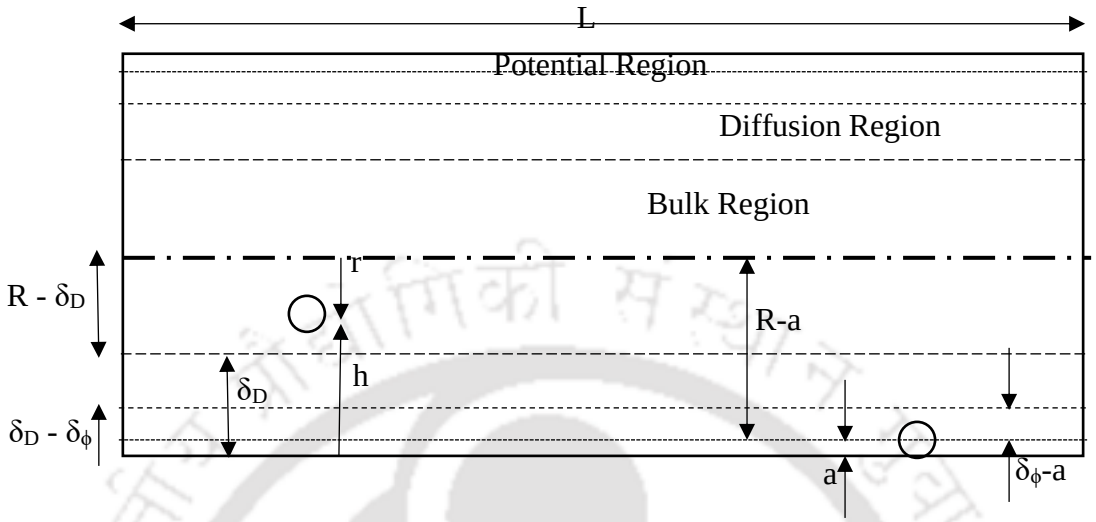

Figure 2.6: A representation of a cylindrical pore and its interaction energy regions (Seetha et al. 2014)

2.5 Transport of Virus in Groundwater (Pore-scale to Darcy scale)

To model the transport of viruses from the pore scale to the Darcy scale, the conversion of the parameters that affect at pore level to the parameters that are dominant at the Darcy level has to be done (Raoof et al. 2010a; Raoof and Majid Hassanizadeh 2010). Seetha and others have done extensive research on nanoparticle transport in a pore scale and have given the correlation equations to upscale the results in a Darcy scale (Seetha et al. 2015). The correlation equations are based on parameters such as flow velocity, viscosity, temperature, solution dielectric constant, solution dielectric strength and the Nanoscale parameters such as pore radius, nanoparticle radius, nanoparticle surface potential and collector surface potentials. These correlation equations hold theoretically for any virus, but the corresponding values are determined by experimental analysis. The collector is assumed to be chemically homogenous in deriving these correlation equations (Seetha et al. 2015). The equations above give the averaged parameters over a single pore, and based on the network of pores and the REV, the results have to be up-scaled to the Darcy scale.

2.6 Virus Source Identification in Groundwater

Accurately identifying and characterizing pollutant sources is the first objective taken into consideration for efficiently remediating the contaminated aquifers. The spatial location of the

source needs to be identified, the duration of the activity of the source and the rate of injection of the flux from the sources (Aral et al. 2001; Datta et al. 2009; Jha and Datta 2013; Mahar and Datta 2001; Singh and Datta 2006). For the past two decades, the source identification problem has been a severe issue, and numerous methodologies have been introduced worldwide. The release history of the contaminant source should be defined to identify pollutant source characteristics. The location of the source contaminant, along with the magnitude of the source, is required (Aral et al. 2001). The transport of contaminants is considered ill-posed as the solution does not satisfy existence, stability and uniqueness. But the first problem is solved as the sources originate from a point. Researchers adopted various techniques to solve the problem of source identification. Due to the ill-posed problem, numerous methods based on simulation optimization were utilized. These methods solved the problem of non-uniqueness and stability by resolving the inverse problem. Over the years, various researchers have introduced numerous methodologies to overcome the problem of identification of source pollutants in groundwater, methodologies such as linear and response matrix approaches (Gorelick et al. 1983), Statistical Pattern recognition technique (Datta et al. 1989), Tikhonov regularization (Chongxuan and Ball 1999; Skaggs et al. 1994), nonlinear optimization with embedded technique (Mahar and Datta 2000, 2001), genetic algorithm based approach (Aral et al. 2001; Mahinthakumar and Sayeed 2006; Singh and Datta 2006), artificial neural network approach (Singh and Datta 2004), classical optimization based approach (Datta et al. 2009).

2.6.1 Response Matrix

The groundwater flow and transport phenomena are used to obtain the response matrix, assuming that the system is linear (Gorelick et al. 1983). The response matrix is then used in the identification of the pollution source. Various works have been conducted in the past for source identification using the response matrix method. But the limitation is that the superposition principle is based on the flow and transport equation. The aquifer parameter needs to be known in advance, and the simulation model needs to be run before the response matrix (Mahar and Datta 2001). Gorelick et al. (1983) used response matrix method techniques and linear programming combined with groundwater solute transport simulation to identify the location and magnitudes of the pollutant sources. They adopted steady-state transport and the second state transient case. Measurement error occurred in the magnitude of flux in the linear and regression model. Another drawback of linear programming is that it takes more time than multiple regressions. Therefore these methods are highly prone to measurement errors. Later Wagner combined contaminant transport with the nonlinear weighted least squares multiple

regression procedure for estimation of transport parameters but showed that the linear source was susceptible to measurement error (Wagner 1995). The response matrix was used to simulate the groundwater flow and transport processes (Datta et al. 1989) and the authors developed an expert system-based statistical pattern recognition technique for identifying pollution sources. The results of the study were found to be satisfying.

2.6.2 Embedded Approach

In the embedded technique, the governing equations of flow and transport describing the physical processes in groundwater are embedded in the optimization model. The identification model is formulated as a nonlinear optimization model. The computer code MINOS is used to solve the nonlinear programming problem. After evaluating all the performance and the measurements, identifying sources with known aquifer parameters and unknown parameters are performed satisfactorily considering single and multiple locations. (Mahar and Datta 2001). This technique requires high computer storage, and it is not efficient when considering large-scale problems (Gorelick et al. 1983). Because of these limitations, various methodology has been developed linking the simulation model with the optimization model.

2.6.3 Linked Simulation Optimization

For efficient identification of source and characterization in large study areas, linked simulation optimization is considered the best choice. The researchers started using a genetic algorithm (GA) coupled with other optimization methods, and it has proved to be a successful approach compared to the previous techniques. Aral and others used the PGA method for solving the nonlinear optimization problem following two approaches (Aral et al. 2001). The PGA performs a linear subdomain characterization of the optimization model. First, in the iteration stage, the groundwater simulation models are run to obtain the approximate optimization model. Later the genetic algorithm is used in searching for the optimal local solution. Singh and Datta used GA and simulated the concentration measurement to obtain the optimum (Singh and Datta 2006). The objective function is to minimize the weighted sum of the absolute difference between observed and simulated concentrations. Bhattacharjya and others have suggested a hybrid optimization approach to estimate the parameters of a one-dimensional virus transport phenomenon (Bhattacharjya et al. 2015). The same approach can be used to link the simulation and optimization models. The approach is a combination of a discrete and continuous variable problem as the number of sources, and their location is discrete variables, and the source strengths are continuous. Some researchers (Hazrati-Yadkoori and Datta 2017)

have suggested an Adaptive Surrogate Model Based Optimization (ASMBO), which mimics the actual flow and transport process happening with a surrogate model which can be easily characterized based on a Self-Organizing Map (SOM). The Linked-Simulation Optimization approach is by far the best method for source identification. A practical algorithm should be designed based on the optimization variables, and the source identification can be made with minimal computational effort.

2.7 Management of Identified Virus Sources

There is a need to know the complete details of concentrations to manage the concentrations of viruses in the aquifer. There should be accurate predictions of the concentrations in the aquifer in future. So, the exact location of the virus sources and the information on the number of sources should be known. Therefore, a source identification model must be developed before the management model's development. Remediation of a Virus contaminated aquifer to its original state is the goal of a management model. The remediation can be done by altering the environmental conditions in the aquifer by studying the survival of viruses. Then, the different techniques based on experiments and simulations can be applied for optimal management policy.

2.7.1 Factors affecting the survival of viruses

Studies conducted on virus survival showed that seeded poliovirus could survive in the soil after irrigation with effluent for 11 days and 96 days during the winter (Tierney et al. 1977). Laboratory studies investigated the factors affecting the survival and migration of viruses in groundwater. The use of actual viruses on a field scale requires a high degree of control to overcome the risks, leading to the use of bacteriophage as an indicator having similar characteristics (Bales et al. 1993; Gerba 1984). The main phenomena responsible for the removal of the virus during soil passage are adsorption and inactivation (Keswick and Gerba 1980; Yates et al. 1988). Bales and others studied the behaviour of bacteriophage transport in sandy soil and fractured tuff and found that the bacteriophage can be used to describe dispersion better than the solute tracer by laboratory experiments and field study (Bales et al. 1989). It was reported that decreasing moisture content enhances virus sorption into the solid by forming a thin film of water around the surface of the solid (Lance and Gerba 1984). But in arguing with it, (Poletika et al. 1995; Powelson et al. 1990) reported that the removal of the virus at low soil moisture content is due to the sorption of viruses onto air-liquid interfaces. Virus sorption at air-liquid interfaces is mainly caused by the virus particle surface hydrophobicity, ionic

solution strength and particle charge (Wan and Wilson 1994). Virus inactivation is controlled by many factors such as temperature, moisture content and pH of the soil water, microbial activity and soil properties (Yates and Yates 1988).

The various range pH value significantly affects the attachment and detachment of virus to the soil. Goyal and others studied the behaviour of various groups of viruses on the attachment and detachment of viruses on soil (Gerba et al. 1981). Bales and others performed field injection at U.S. Geological Survey's toxic substances hydrology research site on Cape Cod and found that the attachment and detachment of PRD1 are pH dependent (Bales et al. 1989). At pH 5.7, the virus was significantly attenuated but remobilized with the injection of a high pH pulse of water. Virus survival is not significantly affected by the concentration of sewage effluent based on the finding for both sterile and non-sterile conditions (Keswick and Gerba 1980), but this is not in agreement that viruses survive longer in sewage-polluted water than in non-polluted water (Hill et al. 1971). Virus removal depends on many factors in a porous medium, such as Temperature, pH, Ionic Strength, UV Irradiation and solution chemistry. The various factors affecting virus management are listed in table 2.1, as suggested by Schijven and Hassanizadeh (Hassanizadeh et al. 2000).

Table 2.1: Factors influencing the survival of viruses (Hassanizadeh et al. 2000)

FACTOR	**INFLUENCE**
Temperature	Prolonged survival at low temperatures and rapid die-off at high temperatures.
Moisture content	Desiccation is detrimental to most microorganisms.
Sunlight	Most rapid die-off at soil surface due to UV irradiation.
pH	Bacteria die off more rapidly in acid soil than in alkaline soils.
Organic carbon content	It increases the survival and may give rise to the regrowth of bacteria.
Cations	Cations enhance virus adsorption to soil and indirectly increase survival.

2.7.2 Removal of the virus by using a management model

The removal of the virus is often designed as a one-dimensional problem as it is challenging to study the 2D or 3D transports through batch or column experiments. Management models have not been developed for virus transport in particular, as the transport process and behaviour of viruses are highly dynamic with the conditions of the aquifer. A groundwater management model has four characteristics: (1) It is stochastic, with the primary source of uncertainty being that associated with the aquifer simulation models; (2) It is nonlinear with respect to the

decision variables; (3) It involves both continuous and discrete decision variables, i.e., it is a mixed-integer programming problem; and (4) It requires the solution of a set of (perhaps nonlinear) coupled partial differential equations describing groundwater flow and transport (Wagner 1995). Whiffen and Shoemaker presented a two-stage method of nonlinear weighted feedback to design aquifer remediation strategies under model uncertainty, a coupled groundwater management model (Whiffen and Shoemaker 1993). Chang and others used differential dynamic programming to demonstrate the time-varying policy for pumping and benefits (Chang et al. 1992). Any management model can be thus taken as a stochastic, dynamic and mixed integer programming problem. The inactivation rates for virus transport in porous media are affected mainly by the changes in temperature and hardness of the aquifer. Alone on the temperature, the inactivation rates vary up to 60% (Yates et al. 1985). The inactivation rates are therefore indirectly related to the time as the temperature changes can be monitored temporally. Similarly, the pH of the solution and the ionic strength also affect the attachment and detachment rates (Bales et al. 1991) which can be related to the change in pH of the system through injection wells and then re-remediation of the de-contaminated aquifer to the natural state has to be done successively. A detailed study can be done based on pore scale modelling and the minor changes in the combinations of temperature, ionic strength and pH can be applied to generate a much efficient management policy.

2.8 Research gap from virus transport to remediation of viruses

- Over the past forty years, various studies have been conducted to understand virus transport in a porous medium in saturated and variably saturated environments. The inherent processes of deposition and inactivation have been studied thoroughly, and there is a good understanding of these processes and their effects on transport under soil heterogeneities, and several other conditions are studied.
- The recent advancements in Magnetic Resonance Imaging and X-Ray Computerized Tomography allow us to understand the microscale transport of porous media, and several models were developed in pore scale approaches which can be up-scaled to a Darcy scale based on pore networks. Several such models were also developed in the past decade.
- Many researchers have not attempted source identification in viruses, whereas many models account for the contaminant source identification. The reason is due to the many

complex phenomena occurring in the virus transport, which are highly sensitive to many parameters of the aquifer medium and soil characteristics.

- Remediation of viruses have only been attempted through the passage of contaminated water through soil, but field-scale in-situ remediation technique for virus-contaminated aquifers have not been attempted so far.

2.9 Summary and Conclusion

This chapter was proposed as state-of-the-art for the research work on detecting and managing virus sources under equilibrium and kinetic sorption. The literature reviewed in this report summarizes the various theories explaining the transport of viruses and the research gap from virus transport to management and remediation. The transport of viruses in the Darcy scale is first reviewed, explaining the processes of macro-dispersion, deposition adsorption and inactivation in the viruses. Then, the review is done for the theories proposed to explain the high sensitiveness of deposition and inactivation, which are affected by many parameters such as soil heterogeneity, temperature, solution chemistry, ionic strength, pH etc. The pore scale model developments in viruses are then reviewed, which accounts for many parameters and gives a better understanding of the physical phenomenon underlying the sensitivity of the parameters in the Darcy scale. The methodologies that upscale the results obtained from the pore scale to the Darcy scale are reviewed so that the parameters can be applied to a field scale equation proposing an adequate simulation model.

The optimization approaches are reviewed in the second part, where different techniques are used to identify the potential sources in groundwater. The linked simulation-optimization approach is the process which combines the physical process with the optimization algorithms to identify the source of contamination. The in-situ management of contaminants is also reviewed based on the algorithms available for optimal management policies. Finally, the motivation and the research gap in linking the experimental studies and the physical processes to formulate an optimal management policy are stated, and the objectives are proposed to fill the research gap.

CHAPTER 3: DEVELOPMENT OF FLOW AND TRANSPORT MODEL IN DARCY SCALE

3.1 Introduction

Fresh water is an essential source of drinking in urban and rural areas. The primary fresh water source is groundwater which comprises 99 percent of the available freshwater (Chow 2010). The rapid urbanization and increased groundwater usage are changing the available groundwater resources and the system's dynamics. The natural soil bed affects groundwater filtration, making the groundwater usable without treatment. Natural filtration occurs through adsorption and deposition of contaminants onto the soil matrix, leaving the slow-moving groundwater much cleaner to use (Schijven 2002). However, the rapid groundwater movement slows down the filtration process, leaving the water relatively contaminated. Also, groundwater contaminants rapidly reach the pumping wells due to overexploitation of groundwater. Specifically, the groundwater contaminated with a virus is a huge threat as it can cause an immediate outbreak in the population who rely on groundwater for domestic use (Hunt et al. 2010; Keswick and Gerba 1980).

Several researchers have attempted to simulate groundwater contaminant transport (Barth and Hill 2005; Burnett and Frind 1987; Gorelick 1990; Jin et al. 2000; Ojha et al. 2011). There exist three-dimensional models for groundwater flow, such as MODFLOW (McDonald and Harbaugh 1988) and HYDRUS 3D (Šimunek et al. 2012), and three-dimensional transport models such as MT3DMS (Bedekar et al. 2016) and HYDRUS 3D (Šimunek et al. 2012). However, none of the models allow the user to input the temporally variable parameters. Some researchers proposed the temporal dependency of the parameters (Sim and Chrysikopoulos 1996) and have developed an analytical solution for the virus transport equation in one dimension with time-dependent inactivation rates. However, the researchers have not extended the model to three-dimensional field scaled problems. Therefore, numerical models need to be developed that account for the temporal variation in the parameters.

Therefore, I have developed a numerical model in this study that solves groundwater flow and transport governing equations using the Finite Volume Method (FVM) to handle the temporal changes in the aquifer parameters. The developed numerical model is validated using the Groundwater Modelling Systems (GMS) with the module of MODFLOW for the flow model and MT3DMS for the transport models. The MT3DMS model cannot simulate the virus

transport process under the time-dependent inactivation and sorption process. Therefore, I have validated the models under the temporally invariable parameters. The developed coupled virus and heat transport models are used to simulate the virus transport process under changing temperatures. The model proves efficient in removing viruses from the aquifer by changing the solution chemistry.

3.2 Methodology

The inactivation rates of viruses, and the adsorption parameters, depend on several factors, such as the temperature of the aquifer media or the solution chemistry. The aquifer media's temperature and solution chemistry are generally assumed to be constant with space and time. I have attempted to show the effects of temperature on virus concentration in this chapter. Therefore, I have considered the model with the variation of aquifer parameters as an effect of changing temperature. The heat transport model generates the temperature of the aquifer with space and time. Researchers have conducted several laboratory experiments to check the virus's dependency on temperature (Bales et al. 1993; Grant et al. 1993; Kinoshita et al. 1993; Sadeghi et al. 2011; Vilker et al. 1980; Yahya et al. 1993). I have used these relations in modeling the aquifer. I have given the example of the virus inactivation coefficient in liquid phase μ_l as a function of temperature in equation 3.1. The same principle applies to any of the parameters' interdependencies with the solution chemistry.

$$\left.\begin{array}{l}\mu_l = \mu_l(T) \\ T = T(x,y,z,t)\end{array}\right\} \Rightarrow \mu_l = \mu_l(x,y,z,t) \qquad 3.1$$

The solution to the heat transport equation gives the temperature in the aquifer as a function of space and time. The inactivation and adsorption parameters in the aquifer are dependent on temperature. Therefore, the parameters are also obtained as functions of space and time. Therefore, by using the flow model, virus transport model, and heat transport model in unison, one can calculate the variation of the concentration in the aquifer with the effect of temperature variations.

I have considered a hypothetical study area with the parameters taken from previous research works of field-scaled models (Molson et al. 1992; Palmer et al. 1992). Then, I developed the flow and transport models in MATLAB. I have tested the developed models with existing and well-established models. The developed models are validated the using MODFLOW for the flow model and MT3DMS for transport model. I have used hypothetical virus sources of MS2 in the aquifer to validate the results from my developed code for virus transport models. I have

used injection wells to induce hot water into the aquifer and validated the results of heat transport models. The parameters for the study used are given in Table 3.1, and Table 3.2. I have used these validated flow and transport models for all the further studies in my thesis. Figure 3.1 shows the developed methodology.

Table 3.1: Aquifer properties for heat transport model

Aquifer Thermal Parameters and Values	
$C_{p,solid} = 800\, JKg^{-1o}C^{-1}$	$\rho_s = 2230.77\, Kgm^{-3}$
$C_{p,fluid} = 4714\, JKg^{-1o}C^{-1}$	$\kappa = 7.10\, X\, 10^{-7} m^2 s^{-1}$

Table 3.2: Aquifer properties for the flow and transport model for viruses

Flow parameters	**Dispersion parameters**	**Sorption parameters**	**Virus inactivation**
$K_{xx} = 9.0 \times 10^{-5}\, ms^{-1}$	$\alpha_L = 2\, m$	$m = 0.8875$	$\mu_l = 0.0550\, d^{-1}$
$K_{yy} = 9.0 \times 10^{-5}\, ms^{-1}$	$\alpha_{TH} = 0.2\, m$	$K_f = 6.67 \times 10^{-7} m^{3/m} kg^{-1/m}$	$\mu_{s,eq} = 0.0367\, d^{-1}$
$K_{zz} = 9.0 \times 10^{-5}\, ms^{-1}$	$\alpha_{TV} = 0.1\, m$	$\rho_b = 1800\, kg\, m^{-3}$	$\mu_{s,kin} = 0.01\, d^{-1}$
$S_s = 0.001\, m^{-1}$	$D^* = 6 \times 10^{-10}\, m^2 s^{-1}$		$k_{att} = 0.12\, d^{-1}$
			$k_{det} = 0.05\, d^{-1}$

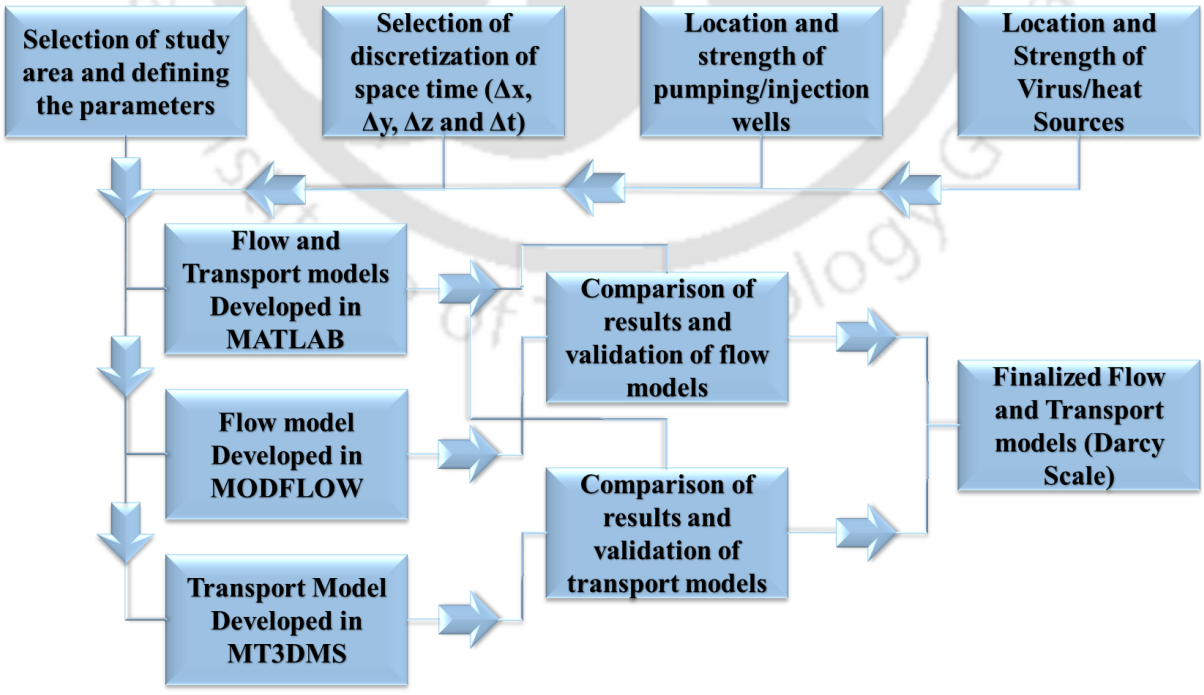

Figure 3.1: Overview of the model development and validation

3.2.1 Development of Flow Model (Darcy Scale)

The flow model is the conservation of mass of the fluid (the continuity equation). Equation 3.2 shows the conservation of the mass equation.

$$\frac{\partial}{\partial x}(\rho q_x)+\frac{\partial}{\partial x}(\rho q_y)+\frac{\partial}{\partial x}(\rho q_z)=\frac{\partial}{\partial t}(\rho n) \tag{3.2}$$

The fluid and the solid compressibility for a non-deformable porous media are combined to obtain the right-hand side of equation 3.2 in terms of specific storage and piezometric heads. The left-hand side of equation 3.2 is converted in terms of the piezometric heads using Darcy's law. The converted form of 3.2 is shown in equation 3.3.

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right)+Q_s=S_s\frac{\partial h}{\partial t} \tag{3.3}$$

Here, K_{xx}, K_{yy}, and K_{zz} are the principal values of the hydraulic conductivity, Q_s is the source or sinks term, S_s is the specific storage term, and h is the piezometric head. Equation 3.3 is formulated by aligning the Eigenvectors of the hydraulic conductivity tensor to the principal axes. Considering a rectangular control volume 'V' and integrating the equation 3.3 over a Representative Elementary Volume (REV) as a control volume 'V' gives equation 3.4.

$$\iiint_V \nabla.\left(\overline{\overline{K}}.\nabla h\right)+\iiint_V Q_s=\iiint_V S_s\frac{\partial h}{\partial t} \tag{3.4}$$

Considering the REV averaged values of hydraulic conductivity, specific storage value, and the piezometric head and using the gauss divergence theorem to convert volume integrals to surface integrals, equation 3.4 is converted into the form of equation 3.5.

$$\sum_m C_{nm}(h_m-h_n)+\sum Q_{n,s}=Q_{STO} \tag{3.5}$$

Here, n and m represent two adjacent cells. C_{nm} is the conductance from cell n to cell m. The heads corresponding to the cells n and m, are h_n and h_m respectively, $Q_{n,s}$ is the source or sink term at cell n, and Q_{STO} is the corresponding storage term for the cell n. Equation 3.5 is solved at each cell to obtain the solution for the flow model.

3.2.1.1 Horizontal Conductance

The term C_{nm} is the conductance, which can occur horizontally or vertically in a three-dimensional model. When C_{nm} is horizontal conductance, the formula for the conductance

between the cells n and m is given by the formula in equation 3.6. Figure 3.2 shows the overview of the conductance between cells *n* and *m*.

$$C_{n,m} = \left(\frac{K_{nm} K_{mn}}{L_{nm} K_{mn} + L_{mn} K_{nm}} \right) \Delta w_{n.m} \left(\frac{\Delta v_n + \Delta v_m}{2} \right) \quad 3.6$$

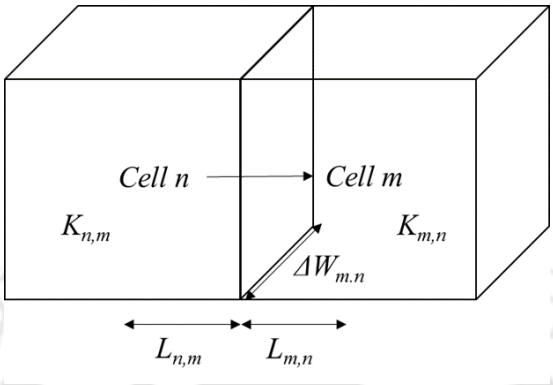

Figure 3.2: Horizontal conductance between cell *n* and *m*

3.2.1.2 Vertical Conductance

The flow between the cells considering vertical conductance is calculated using equation 3.7. Figure 3.3 shows the vertical conductance between the cells n and m.

$$C_{n.m} = \frac{2A_{n,m} K_{n,m} K_{m,n}}{(T_n - B_n) K_{m,n} + (T_m - B_m) K_{n,m}} \quad 3.7$$

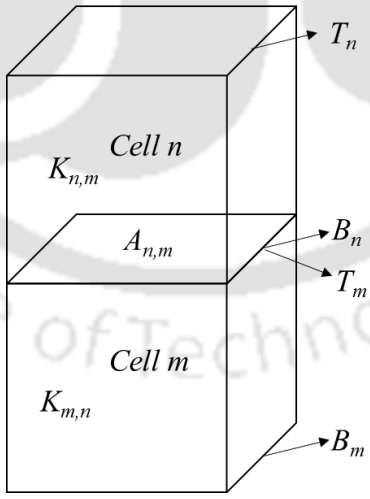

Figure 3.3: Vertical conductance between cell *n* and *m*

3.2.1.3 Storage

The total storage in the aquifer comes from the parts called the specific storage and the specific yield. When a cell is in a confined state (i.e., the piezometric head is more than the thickness of the cell), the cell can only store water. Therefore, the specific storage term is considered

only when the cell is confined. But when a cell is in unconfined condition, there is an additional condition of water table within the cell. Therefore, such cell has two storage terms, one is the term from specific storage and the other is specific yield term. Each cell should be able to convert into confined or unconfined stages throughout the simulation for the field-scaled problems. Therefore, the saturated fraction determines whether the cell is confined or unconfined. Equation 3.8 and equation 3.9 gives the formulae for contributions of specific storage and specific yield to the storage in the aquifer.

$$Q_{ss,n}^{k} = \frac{SC1_n}{t - t_{old}} \left(SF_n^{t_{old}} H_{old,n} - SF_n^{k-1} h_n^{k} \right) \qquad 3.8$$

$$Q_{s_y,n}^{k} = \frac{SC2_n \left(T_n - B_n \right)}{t - t_{old}} \left(SF_n^{t_{old}} - \frac{h_n^{k} - B_n}{T_n - B_n} \right) \qquad 3.9$$

Here $SC1_n = A_n S_{s,n}(T_n - B_n)$ and $SC2_n = A_n S_{y,n}$, Where A_n is the top area of the cell n. Also, saturated fraction SF is the ratio of h_n-B_n to T_n-B_n.

If a cell is confined from the old time-step to the current time-step, the saturated fraction is equal to one in both cases and therefore the second term as in equation 3.9 vanishes. However, if a cell converts from a confined state to an unconfined state from the old time-step to the current time-step, the saturated fraction in current time-step will be less than one, and the equation 3.9 contributes to the storage from the water table condition as well.

3.2.1.4 Source and Sink (Well)

The source and sink term Q_s in equation 3.3 represent any source (the one providing the volume of fluid to the aquifer) or a sink (the one extracting the volume of fluid from the aquifer). The typical sink term for any model is a groundwater pumping well. The discharge of the groundwater pumping well acts as a sink term in the model. However, when a well is penetrating deep into the aquifer of multiple layers, knowing the transmissivity of each layer, the discharge contributing to the total discharge can be calculated by using equation 3.10, where the actual discharge from the well is $Q_{wel,nb}$.

$$Q_{wel,n} = \frac{T_n}{\sum_{no.} T} Q_{wel,nb} \qquad 3.10$$

The term T_n represents the transmissivity of the n^{th} layer in the aquifer, and $Q_{wel,n}$ represents the discharge of the total well penetrated to n layers.

3.2.1.5 Boundary Conditions

When a cell is a boundary cell, it will either be a Dirichlet boundary where the head value is constant or a specified flow boundary (zero in case of no flow). In all such cases, the boundary condition is added to the equation in the summation of the flow term.

3.2.1.6 Solution Procedure

After the substitutions are made per the definitions from equations 3.6 to 3.10, equation 3.5 turns into equation 3.11. Solving equation 3.11 at all the cells in the aquifer gives the piezometric heads at different cells for space and time in the aquifer. The equation has three types of variables, the first term on the left-hand side is represented by the cell at which the equation is applied, i.e., the coefficient of h_n (say $Cofh_n$). The second term is the cross term h_m (say $Cofh_m x h_m$) represented by the six adjacent cells surrounding the central cell. The third is the constant over a time step term on the right-hand side of the equation. The different types of cells that need to be involved in solving the equation at a single cell are given in Figure 3.4.

$$\left[-\sum_m C_{n,m} + \frac{SC1_n SF_n^{k-1} + SC2_n}{t - t_{old}}\right] h_n^k + \sum_m C_{n,m} h_m^k = -Q_{wel,n} + \frac{SF_n^{t_{old}} \left(SC1_n H_{old,n} + SC2_n (T_n - B_n)\right)}{t - t_{old}} + \frac{SC2_n B_n}{t - t_{old}} \quad 3.11$$

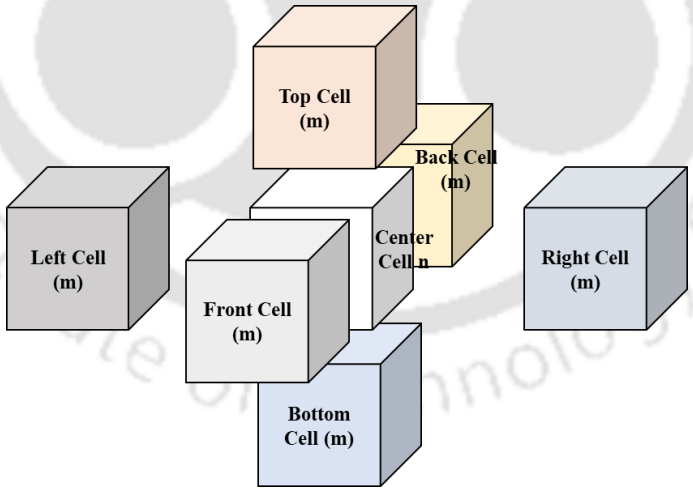

Figure 3.4: Placement of the central cell along with the adjacent cells for flow equation

There will be n set of equations that are in the matrix of the form of $Ax = b$, where A contains the coefficient terms of all the variables that are not constant over a particular time step. The term b contains the variables that are constant over a time step (i.e., the right-hand side of equation 3.11). The matrix is a sparse matrix that is dense over the principal diagonal. The

solution to this equation is obtained using the preconditioned conjugate gradient method, which is explained in the later sections with the transport solution procedure of the transport model.

3.2.2 Development of Virus Transport Model (Darcy Scale)

The virus transport equation is solved using the Finite Volume Method (FVM) on the model domain, which is divided into uniform cuboids with length, width, and height dx, dy, and dz, respectively. Solving the equation over the domain requires satisfying equation 3.12 over the whole domain. That requires the equation to be solved for each of these cuboids. Therefore, the equation is integrated over the whole control volume, which in turn can be divided into a lot of smaller control volumes (each representing a REV). For each cuboid representing a REV, there is a need to define the faces as the Control Surfaces so that these integrals can be converted from volume to surface integrals using the Gauss theorem. Therefore, the control surfaces are named as left face and right face in the x – direction, front face and back face in the y – direction, and top face and bottom face in the z-direction.

The advantage of using the model domain to be defined in this form is that the boundaries can be distinguished very easily. If there is a specified flow from a boundary (Newman boundary condition), then the flow from that specific control surface can be replaced with the flow specified. This makes the computations over complex shapes of irregular boundaries very easy to handle. The governing equation for virus transport considering the equilibrium and kinetic sorption is given in equation 3.12. The first and the third parts of the equation above are to be coupled and solved simultaneously to accommodate for the kinetic sorption part. The first and the second parts of the equation combined together solve the equilibrium sorption part.

$$n\frac{\partial C}{\partial t}+\rho_b\frac{\partial S_{kin}}{\partial t}+\rho_b\frac{\partial S_{eq}}{\partial t}=\frac{\partial}{\partial x_i}\left(nD_{ij}\frac{\partial C}{\partial x_j}-q_iC\right)-n\mu_l C-\rho_b\mu_{s,kin}S_{kin}-\rho_b\mu_{s,eq}S_{eq}-q_s^{|}C-q_sC_s$$

$$S_{eq}=K_fC^{m_f} \qquad 3.12$$

$$\rho_b\frac{\partial S_{kin}}{\partial t}=nk_{att}C-\rho_bk_{\det}S_{kin}-\rho_b\mu_{s,kin}S_{kin}$$

3.2.2.1 Sorption Terms

The equilibrium sorption is a reversible process where the percentage of the solutes attached to the solids is precisely equal to the percentage detached into the aqueous phase. The kinetic sorption follows first-order kinetics, where attachment and detachment rates determine the sorption. The kinetic sorption process converts to equilibrium sorption in controlled conditions after a long time. Therefore, both the equilibrium and kinetic sorption conditions cannot exist

simultaneously (Hassanizadeh et al. 2000). Also, through thorough experiments, the researchers have concluded that the equilibrium sorption condition is the ideal condition and is not that easy to achieve. Therefore, kinetic sorption is the standard and crucial model when considering viruses. As the kinetic sorption requires coupling of equations, the solution procedure is as follows for FVM. The equations 3.13 to 3.16 explain the kinetic sorption coupling with the virus transport equation in detail.

$$\rho_b \frac{\partial S_{kin}}{\partial t} = nk_{att}C - \rho_b k_{\det} S_{kin} - \rho_b \mu_{s,kin} S_{kin}$$

$$\rho_b \frac{S_{kin}^{n+1} - S_{kin}^{n}}{t - t_{old}} = nk_{att}C^{n+1} - \rho_b k_{\det} S_{kin}^{n+1} - \rho_b \mu_{s,kin} S_{kin}^{n+1} \qquad 3.13$$

$$S_{kin}^{n+1} = \frac{nk_{att}C^{n+1} + \left(S_{kin}^{n}\right)\rho_b \Big/ \left(t - t_{old}\right)}{\rho_b \Big/ \left(t - t_{old}\right) + \rho_b k_{\det} + \rho_b \mu_{s,kin}} \qquad 3.14$$

$$n\frac{\partial C}{\partial t} + \rho_b \frac{\partial S_{kin}}{\partial t} + \rho_b \frac{\partial S_{eq}}{\partial t} = L(C) - n\mu_l C - \rho_b \mu_{s,kin} S_{kin} - \rho_b \mu_{s,eq} S_{eq} - q_s^{|} C - q_s C_s$$

$$n\frac{C^{n+1} - C^{n}}{t - t_{old}} + \rho_b \frac{S_{kin}^{n+1} - S_{kin}^{n}}{t - t_{old}} + \rho_b \left(\frac{\partial S_{eq}}{\partial C}\right)^{n} \frac{C^{n+1} - C^{n}}{t - t_{old}} \qquad 3.15$$

$$= L(C) - n\mu_l C^{n+1} - \rho_b \mu_{s,kin} S_{kin}^{n+1} - \rho_b \mu_{s,eq} \frac{S_{eq}^{n}}{C^{n}} C^{n+1} - q_s^{|} C^{n+1} - q_s C_s$$

$$\frac{-n}{t - t_{old}} \left(1 + \frac{\rho_b \left(\frac{\partial S_{eq}}{\partial C}\right)^{n}}{n}\right) C^{n+1} - \left(n\mu_l + \rho_b \mu_{s,eq} \frac{S_{eq}^{n}}{C^{n}} + nk_{att} + q_s^{|}\right) C^{n+1} -$$

$$\left(\frac{n\rho_b k_{att} k_{\det}}{\rho_b \Big/ \left(t - t_{old}\right) + \rho_b k_{\det} + \rho_b \mu_{s,kin}}\right) C^{n+1} \qquad 3.16$$

$$= L(C) - q_s C_s \frac{n}{t - t_{old}} \left(1 + \frac{\rho_b \left(\frac{\partial S_{eq}}{\partial C}\right)^{n}}{n}\right) C^{n} - \left(\frac{\rho_b^2 k_{\det} \Big/ t - t_{old}}{\rho_b \Big/ \left(t - t_{old}\right) + \rho_b k_{\det} + \rho_b \mu_{s,kin}}\right) S_{kin}^{n}$$

$$\rho_b \frac{\partial S_{eq}}{\partial t} = \rho_b \frac{\partial S_{eq}}{\partial C}\frac{\partial C}{\partial t}$$

$$\left(\frac{\partial S_{eq}}{\partial C}\right)^n = m_f^n K_f^n \left(C^n\right)^{m_f^n - 1} \qquad 3.17$$

$$S_{eq}^{n+1} = \frac{S_{eq}^{n+1}}{C^{n+1}} C^{n+1} \approx \frac{S_{eq}^{n}}{C^{n}} C^{n+1} \approx K_f^n \left(C^n\right)^{m_f^n - 1} C^{n+1}$$

The finite volume approximations for the equilibrium sorption with the Freundlich isotherm are given in equation 3.17. Here, the approximation is that the ratio of equilibrium sorption to the concentration at the time step 'n+1' is almost equal to that of the same at the time step 'n'.

3.2.2.2 Advection Terms

The term *L(C)* shown in Equations 3.15 and 3.16 represents the advection and dispersion terms that are much more complicated to express in the form of FVM. The advection term is the easiest among the two terms and is expressed as shown in equation 3.18

$$\iiint_V \nabla.\left(qC\right)dV = \iint_S \left(qC\right).\vec{m}dS = \sum_m \left(qC\right).\vec{m}dS \qquad 3.18$$

As there are 6 faces (left, right, front, back, top, and bottom), the integral is carried over these six faces and then added together. The concentration term is required on each control surface. This can be defined as an upstream weight term between the control volumes (α_m), and its usage can be seen in equation 3.19.

$$\iint_S C q_i n_i dS = \sum_m Q_n \left(\alpha_m C_n + \left(1-\alpha_m\right)C_m\right) \qquad 3.19$$

The term Q_n represents the normal component of the velocity multiplied by its respective surface area. This is equal to the amount of fluid flowing out of the control surface.

3.2.2.3 Dispersion Terms

The second term in *L(C)* of equations 3.15 and 3.16 is the dispersion term which is more complex to express as an integral as the term is a double derivative. The expression for the term is shown in equation 3.20.

$$\iiint_V \nabla.\left(n\bar{\bar{D}}.\nabla C\right)dV = \iint_S \left(n\bar{\bar{D}}.\nabla C\right).\vec{m}dS = \sum_m \left(n\bar{\bar{D}}.\nabla C\right).\vec{m}dS \qquad 3.20$$

Here, the term *D* is a tensor of second order in three dimensions, meaning the resulting term has 9 components. The term combines the molecular diffusion given by D^*_{ij} and the mechanical

dispersion term called D^m_{ij}. This mechanical dispersion tensor is represented by a much more complex term called dispersivity, which is a 4th-order tensor in three dimensions with 81 components. The dispersivity tensor (α) and the Dispersion tensor (D) are related, as shown in equation 3.21, with the vector 'v' being the velocity vector

$$D^m_{ij} = \alpha_{ijkl} \frac{v_k v_l}{|v|} \tag{3.21}$$

The components of the dispersivity term can be brought down to two terms when the tensor itself is isotropic. The terms are longitudinal dispersivity (α_L) and transverse dispersivity (α_T). Considering the practical scenario, this tensor is defined with three components (Burnett and Frind 1987) and is shown in equation 3.22.

$$D^m_{ij} = \begin{pmatrix} \alpha_L \frac{v_x^2}{|v|} + \alpha_{TH} \frac{v_y^2}{|v|} + \alpha_{TV} \frac{v_z^2}{|v|} & (\alpha_L - \alpha_{TH}) \frac{v_x v_y}{|v|} & (\alpha_L - \alpha_{TV}) \frac{v_x v_z}{|v|} \\ (\alpha_L - \alpha_{TH}) \frac{v_x v_y}{|v|} & \alpha_{TH} \frac{v_x^2}{|v|} + \alpha_L \frac{v_y^2}{|v|} + \alpha_{TV} \frac{v_z^2}{|v|} & (\alpha_L - \alpha_{TV}) \frac{v_y v_z}{|v|} \\ (\alpha_L - \alpha_{TV}) \frac{v_x v_z}{|v|} & (\alpha_L - \alpha_{TV}) \frac{v_y v_z}{|v|} & \alpha_{TV} \frac{v_x^2}{|v|} + \alpha_{TV} \frac{v_y^2}{|v|} + \alpha_L \frac{v_z^2}{|v|} \end{pmatrix} \tag{3.22}$$

The terms α_L, α_{TH}, and α_{TV} are longitudinal, transverse horizontal, and transverse vertical dispersivity. v is the velocity vector considering the coordinate axes to be in line with the principal direction with the hydraulic conductivity tensor. Thus, the cross-terms of the dispersion tensor have to be considered, as it would lead to a change in all the vectors calculated in the flow equation if the transformation is considered to be the principal direction of the hydrodynamic dispersion tensor. The last portion of the term in equation 3.20, in the expanded form, is given in equation 3.23.

$$\begin{aligned} \sum_m nD_{ij} \frac{\partial C}{\partial x_j} m_i dS = \sum_m \Big(& nD_{xx} \frac{\partial C}{\partial x} m_x dydz + nD_{xy} \frac{\partial C}{\partial y} m_x dydz + nD_{xz} \frac{\partial C}{\partial z} m_x dydz \\ & + nD_{yx} \frac{\partial C}{\partial x} m_y dxdz + nD_{yy} \frac{\partial C}{\partial y} m_y dxdz + nD_{yz} \frac{\partial C}{\partial z} m_y dxdz \\ & + nD_{zx} \frac{\partial C}{\partial x} m_z dxdy + nD_{zy} \frac{\partial C}{\partial y} m_z dxdy + nD_{zz} \frac{\partial C}{\partial z} m_z dxdy \Big) \end{aligned} \tag{3.23}$$

The terms m_x, m_y, and m_z are the normal vectors in the x, y, and z directions. The nine dispersion tensor terms are D_{xx}, D_{xy}, D_{xz}, D_{yx}, D_{yy}, D_{yz}, D_{zx}, D_{zy}, and D_{zz}. n is the porosity of the porous

medium, and the partial differentials of the concentrations in each direction are given later in equation 3.24. I have taken two summations of cell widths for the partial differentials of concentration terms. One is the summation of the central cell with the adjacent cell (any of the 6 adjacent cells as defined in Figure 3.4). The other is the summation of cells in opposite faces of the central cell, each represented by a forward and backward arrow, as shown in equation 3.24.

$$s_{nmD} = d_n + d_m \text{ and } s_D = \overrightarrow{d_m} + \overleftarrow{d_m} \qquad 3.24$$

The other is the summation of two adjacent cell widths in each direction, x, y, and z. The term gradient of the concentration has to be calculated on every control surface. The gradient terms are shown on one of the surfaces (left face) in equation 3.25.

$$\frac{\partial C}{\partial x} = \left(\frac{2}{s_{nmD,xL}}\right) C_{n,xL} - \left(\frac{2}{s_{nmD,xL}}\right) C_{m,xL}$$

$$\frac{\partial C}{\partial y} = \left(\frac{d_{n,xL}}{s_{nmD,xL}} \times \frac{1}{\left(s^{y}_{D,xL} + d_{n,xL}\right)}\right)\left(C_{m,xLyF} + C_{m,yF}\right) - \left(\frac{d_{n,xL}}{s_{nmD,xL}} \times \frac{1}{\left(s^{y}_{D,xL} + d_{n,xL}\right)}\right)\left(C_{m,xLyB} + C_{m,yB}\right) \qquad 3.25$$

$$\frac{\partial C}{\partial z} = \left(\frac{d_{n,xL}}{s_{nmD,xL}} \times \frac{1}{\left(s^{z}_{D,xL} + d_{n,xL}\right)}\right)\left(C_{m,xLzT} + C_{m,zT}\right) - \left(\frac{d_{n,xL}}{s_{nmD,xL}} \times \frac{1}{\left(s^{z}_{D,xL} + d_{n,xL}\right)}\right)\left(C_{m,xLzB} + C_{m,zB}\right)$$

The equations 3.23 and 3.25 are combined for all the respective faces. Then the coefficient corresponding to each concentration term surrounding the central control volume is calculated in terms of the known parameters.

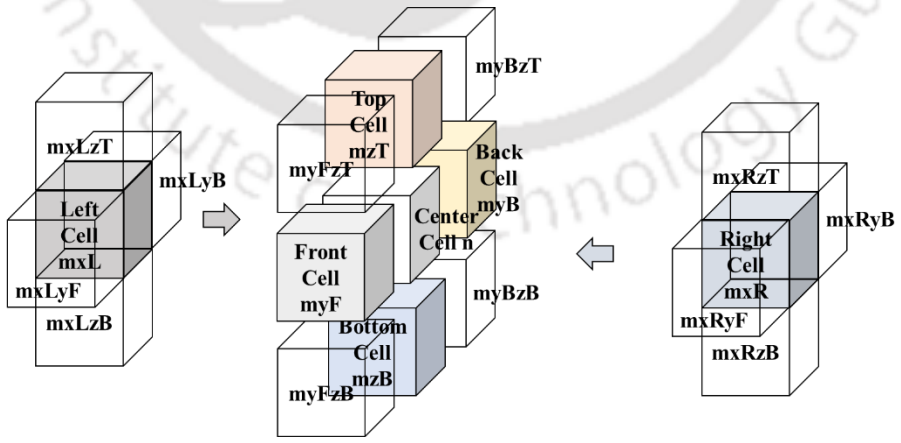

Figure 3.5: The 19 cells needed to define the dispersion term on each of the control surfaces

There is a total of 19 control volumes to be considered to get an expression for equation 3.20. One control volume is central, 6 are adjacent corresponding to each face, and 12 are the ones

that are in between the adjacent faces. These control volumes in consideration are given in detail in Figure 3.5. Each face is named according to its position and the principal directions *x*, *y*, and *z*. The cell adjacent to the left face is named *mxL*, the cell adjacent to the right is named *mxR*, the one adjacent to the top face is named *mzT*, the one adjacent to the bottom face is named *mzB*, the one adjacent to the front face is *myF* and one adjacent to back face is *myB*. These comprise all the six cells that are adjacent to the faces. The cells along the edges are 12 in number and shown as transparent faces named with the common terms of the adjacent faces that were already named. For example, the cell along the edge of the right and the top cells are named mxRzT. These details can are shown in Figure 3.5.

Each cell contributes to the transport of the adjacent cells. However, the contribution of specific cells to incorporate into the coefficients is given in

Table 3.3. It represents the coefficient terms for the corresponding concentration of all the 19 (12+6+1) control volumes. It can be observed from the table that each of the faces contains ten components that contribute to the concentration term of the central volume. In advection, there were only two components for each face. Adding all these components together gives the term *L(C)*, and then adding it to equation 3.16 gives one more final implicit form of the partial differential equation 3.12 into the Finite Volume Form represented in equation 3.26. This equation has to be solved simultaneously for all the control volumes to solve the PDE for the whole domain.

$$\begin{aligned} &A_n^1 C_n^{t+\Delta t} + A_n^2 C_{mxR}^{t+\Delta t} + A_n^3 C_{mxL}^{t+\Delta t} + A_n^4 C_{myF}^{t+\Delta t} + A_n^5 C_{myB}^{t+\Delta t} + A_n^6 C_{mzT}^{t+\Delta t} + A_n^7 C_{mzB}^{t+\Delta t} \\ &+ A_n^8 C_{mxRyF}^{t+\Delta t} + A_n^9 C_{mxRyB}^{t+\Delta t} + A_n^{10} C_{mxRzT}^{t+\Delta t} + A_n^{11} C_{mxRzB}^{t+\Delta t} + A_n^{12} C_{mxLyF}^{t+\Delta t} + A_n^{13} C_{mxLyB}^{t+\Delta t} \\ &+ A_n^{14} C_{mxLzT}^{t+\Delta t} + A_n^{15} C_{mxLzB}^{t+\Delta t} + A_n^{16} C_{myFzT}^{t+\Delta t} + A_n^{17} C_{myFzB}^{t+\Delta t} + A_n^{18} C_{myBzT}^{t+\Delta t} + A_n^{19} C_{myBzB}^{t+\Delta t} = b_n C_n^t \end{aligned} \quad 3.26$$

This equation is also converted into the matrix form of $Ax = b$, the same as the one in the flow equation. The only difference is that the flow model only uses 7 cells to formulate the equation and the transport model uses 19 cells to formulate the matrix equation. Nevertheless, the solution procedure for solving the matrix equation is the same for both the equations which are given in equation 3.27 and equation 3.28.

Table 3.3: Table showing the corresponding components for each face

Coefficient corresponding to the face	*x*, Left	*x*, Right	*y*, Back	*y*, Front	*z*, Bottom	*z*, Top
C_n	m_x,A_1	m_x,A_1	m_y,A_1	m_y,A_1	m_z,A_1	m_z,A_1

$C_{m,xL}$	m_x,A_2		m_x,A_2	m_x,A_2	m_x,A_2	m_x,A_2
$C_{m,xR}$		m_x,A_3	m_x,A_3	m_x,A_3	m_x,A_3	m_x,A_3
$C_{m,yB}$	m_y,A_4	m_y,A_4	m_y,A_4		m_y,A_4	m_y,A_4
$C_{m,yF}$	m_y,A_5	m_y,A_5		m_y,A_5	m_y,A_5	m_y,A_5
$C_{m,zB}$	m_z,A_6	m_z,A_6	m_z,A_6	m_z,A_6	m_z,A_6	
$C_{m,zT}$	m_z,A_7	m_z,A_7	m_z,A_7	m_z,A_7		m_z,A_7
$C_{m,xLyF}$	m_y,A_8			m_x,A_8		
$C_{m,xLyB}$	m_y,A_9		m_x,A_9			
$C_{m,xLzT}$	m_z,A_{10}					m_x,A_{10}
$C_{m,xLzB}$	m_z,A_{11}				m_x,A_{11}	
$C_{m,xRyF}$		m_y,A_{12}		m_x,A_{12}		
$C_{m,xRyB}$		m_y,A_{13}	m_x,A_{13}			
$C_{m,xRzT}$		m_z,A_{14}				m_x,A_{14}
$C_{m,xRzB}$		m_z,A_{15}			m_x,A_{15}	
$C_{m,yFzT}$				m_z,A_{16}		m_y,A_{16}
$C_{m,yFzB}$				m_z,A_{17}	m_y,A_{17}	
$C_{m,yBzT}$			m_z,A_{18}			m_y,A_{18}
$C_{m,yBzB}$			m_z,A_{19}		m_y,A_{19}	

3.2.2.4 Solution Procedure

A sparse matrix with all the 19 components (in the case of the transport model) or with 7 components (in the case of the flow model) is formed with these equations, which would be of the form $Ax = b$. An iterative technique is used to solve the equation. The property of the sparse matrix that the principal diagonal elements of *A* are weighted more than the other nonzero elements is used for the solution. The matrix A is written as *A = D-(D-A)*, with matrix *D* having all diagonal elements of *A*. Equation 3.27 shows the modification of the matrix to an iterative procedure.

$$DC_{i+1}^{n+1} = (D - A)C_i^{n+1} + b \qquad 3.27$$

After some iterations, the solution converges, giving the same value at i^{th} and $i+1^{th}$ iterations. This solution procedure is called the preconditioned conjugate gradient method with the Jacobi pre-conditioner. However, an acceleration procedure is used for the solution to converge much faster, which is described in equation 3.28.

$$\delta^0 = D^{-1}\left(b - Au^0\right)$$

$$\lambda_n = \frac{\delta^n \eta^n}{D^{-1}Ap^n q^n}, \alpha_n = \frac{\delta^n \eta^n}{\delta^{n-1}\eta^{n-1}}$$

$$p^n = \delta^n + \alpha_n p^{n-1}, p^0 = \delta^0$$

$$q^n = \eta^n + \alpha_n q^{n-1}, \mathrm{q}^0 = \delta^0 \qquad 3.28$$

$$u^{n+1} = u^n + \lambda_n p^n$$

$$\delta^{n+1} = \delta^n - \lambda_n D^{-1} A p^n$$

$$\eta^{n+1} = \eta^n - \lambda_n \left(D^{-1}A\right)^T p^n$$

Here u^0 is an arbitrary initialization of the solution vector. The final solution is checked by calculating the error between u^n and u^{n+1}. The solution converges if the absolute error is less than the tolerance (typically 10^{-6}).

3.2.3 Heat Transport Model (Darcy Scale)

The heat transport model is also necessary to model warm water transport in the aquifer. The model is used for managing viruses by varying the water temperature in the aquifer in the later parts of the thesis. The heat transport equation is given in equation 3.29 as follows.

$$\left(1 + \frac{1-n}{n}\frac{\rho_s}{\rho}\frac{C_{p,solid}}{C_{p,\,fluid}}\right)\frac{\partial(nT)}{\partial t} = \nabla.\left[n(\kappa + D).\nabla T\right] - \nabla.(qT) + \sum_{i=1}^{i=N} q_s^i T_s^i \qquad 3.29$$

Where n is the porosity in the aquifer; ρ_s is the density of the solid matrix in the aquifer (ML^{-3}); ρ is the density of the fluid in the aquifer (ML^{-3}); $C_{p,solid}$ and $C_{p,fluid}$ represent the specific heat for solid and fluid respectively ($L^2T^{-2}K^{-1}$); κ is the aquifer thermal diffusivity (L^2T^{-1}); q_s^i is the flow into the aquifer through source or flow out of the aquifer through sink (L^3T^{-1}); T and T_s are the temperatures in the aquifer at any point and the temperature at the source or sink respectively (K).

The heat transport equation is almost the same as the virus transport equation as they are both in the same form. Each of the terms can be written in the form of the other by comparing equation 3.12 and equation 3.29. The parameters of the heat transport equation are to be converted into the equivalent virus transport parameters. The process starts with considering the virus transport process with equilibrium sorption term, as shown in equation 3.30.

$$n\frac{\partial C}{\partial t} + \rho_b \frac{\partial S_{eq}}{\partial t} = \frac{\partial}{\partial x_i}\left(nD_{ij}\frac{\partial C}{\partial x_j} - q_i C\right) - n\mu_l C - \rho_b S_{eq}\mu_{s,eq} - q_s^{|}C - q_s C_s \qquad 3.30$$

$$S_{eq} = K_{eq} C$$

Considering μ_l and $\mu_{s,eq}$ to be zero and substituting the equivalent sorption term and rearranging the terms gives equation 3.31.

$$n\frac{\partial C}{\partial t}+q_s^{|}C+\rho_b\frac{\partial\left(K_{eq}C\right)}{\partial t}=\frac{\partial}{\partial x_i}\left(nD_{ij}\frac{\partial C}{\partial x_j}-q_iC\right)-q_sC_s \qquad 3.31$$

The term $q_s^{|}$ is called the rate of change of transient groundwater storage given by $\frac{\partial n}{\partial t}$ and also by considering the bulk density to be constant over time and splitting the term D_{ij} into the molecular diffusion and mechanical dispersion coefficients D^*_{ij} and D^m_{ij}, the equation converts to the equation 3.32.

$$\left(1+\frac{\rho_bK_{eq}}{n}\right)\frac{\partial(nC)}{\partial t}=\nabla.\left(n\left(D^*+D^m\right).\nabla C-qC\right)+q_sC_s \qquad 3.32$$

Equation 3.32 is similar to the heat transport equation, and by using the appropriate replacement of the variables, the heat transport can also be modeled using the same virus transport model. The term replacements in the model are $\rho_b=(1-n)\rho_s$, $K_{eq}=\frac{C_{p,solid}}{\rho C_{p,\,fluid}}$, $D^*=\kappa$, $C=T$ and $C_s=T_s$.

3.3 Results and Discussion

This chapter's objective is to develop field-scale flow and transport models. Before applying these models, I have validated the models using well-established models, i.e., MODFLOW and MT3DMS models. I have used the flow and transport models per the procedure described in sections 3.3.2 and 3.3.3 of this chapter. I have validated the developed models using the test cases from our previous research (Rajeev Gandhi et al. 2017).

3.3.1 Description of the Problem and Study Area

The area of 4 hectares, as shown in Figure 3.6, is an industrial dumping area with potential source locations. The area has three pumping wells and six injection wells. The locations of pumping wells and the discharges for the corresponding wells are given in Table 3.4. Some industrial dumping sites release wastewater into the ground. The details of the source flux released at three locations, S1, S2, and S3, for 40 days are given in Table 3.5. There are 5 releases in the period, each happening at a time gap of 10 days. The study area is a three-dimensional unconfined aquifer of the cuboidal volume of 400-hectare meters, which has

dimensions of 200m x 200m x 100m. The boundary conditions of the aquifer are a constant head boundary for two opposite sides and a no-flow boundary for two sides opposite to each other and adjacent to the constant head boundary in the lateral directions. There is an impermeable boundary at the bottom of the aquifer, and at the top is the water table. The model study area with all boundary conditions is given in the same Figure 3.6.

Table 3.4: Pumping and injection locations along with discharges

Well No.	Notation	Distance from the top (m)	Distance from the left (m)	Start time (day)	End time (day)	Discharg-e (m^3/s)	Tempera-ture (K)
1.	P1	65	65	0	120	-300	-
2.	P2	95	135	0	150	-500	-
3.	P3	145	95	120	150	-400	-
4.	I1	115	95	50	150	+340	320
5.	I2	165	95	50	150	+115	350
6.	I3	85	55	50	150	+300	370
7.	I4	85	115	50	150	+182	320
8.	I5	65	135	50	150	+174	370
9.	I6	45	155	50	150	+294	370

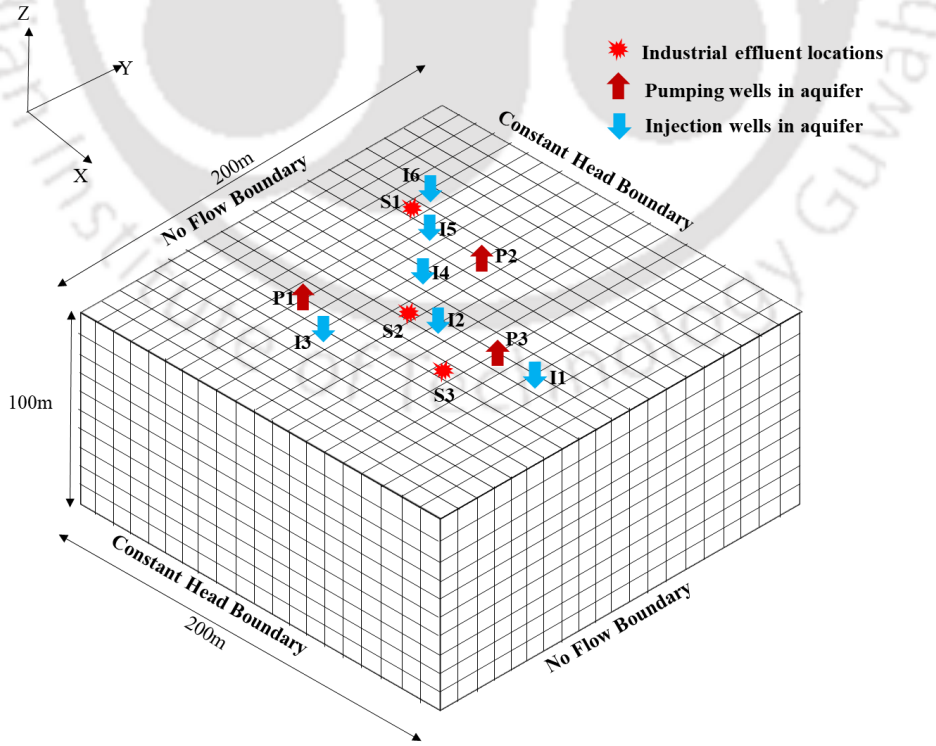

Figure 3.6: The industrial dumping site in the discretized grid

The porosity in the aquifer is assumed to be 0.35, and the specific yield is assumed to be 0.175, as proposed by Johnson (Johnson 1967). He proposed the range of specific yield to be 10% to 35% for sand deposits (fine, medium, and coarse). The longitudinal dispersivity in the referred paper (Palmer et al. 1992) is 0.1 m, and the study area is 40m x 30m x 20 m. The longitudinal dispersivity in the current study is 2 m based on the variation in the study area. The ratio of the transverse horizontal dispersivity and transverse vertical dispersivity was kept the same as taken by Molson and others (Molson et al. 1992; Palmer et al. 1992), as shown in Table 3.2. I have done the spatial discretization for a grid size of 10 x 10 x 10 m for the three principal directions and temporal discretization for 10 days for each time step.

Table 3.5: Source release strengths at locations S1, S2, and S3 from day 0 to day 40

Time	Source flux (Kg/day)		
	Location S1	Location S2	Location S3
Day 0 to Day 10	5000	1000	700
Day 10 to Day 20	1100	1960	1000
Day 20 to Day 30	2350	4000	1500
Day 30 to Day 40	1340	0	3000
Day 40 to Day 50	3500	0	0

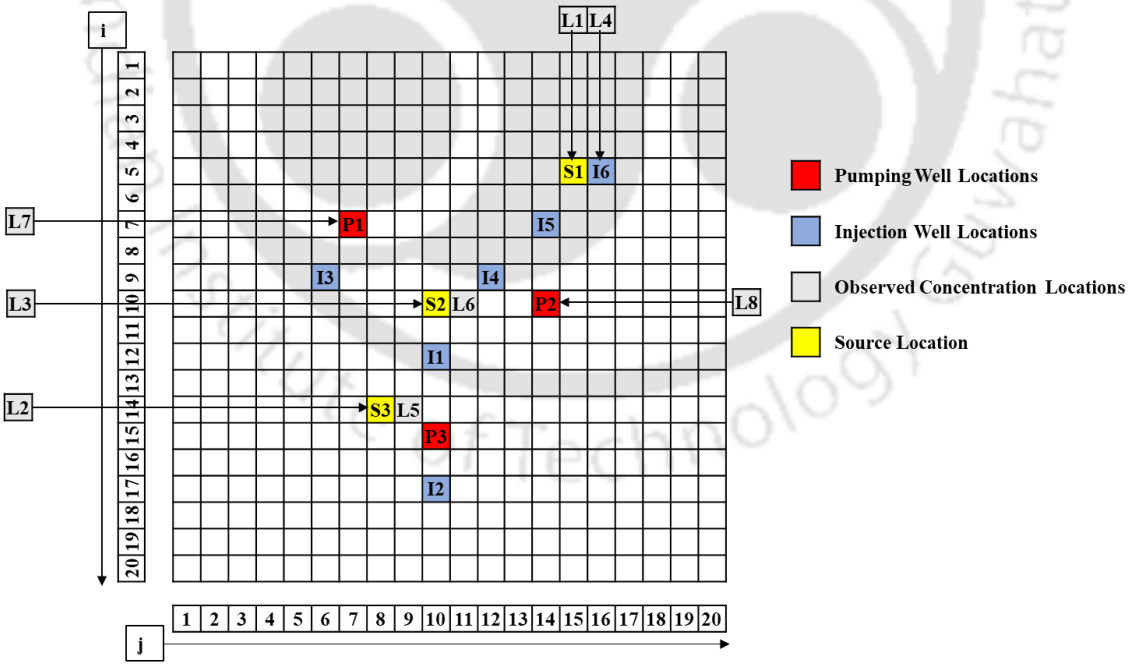

Figure 3.7: Discretized grid with reference locations from the top view indicating various locations

In order to observe the data much carefully, there are some additional points other than that given in Figure 3.6 introduced as observation sites for the concentration movement. Also, the

notations for the discretized grid are used hereafter, and therefore, I have given the details of each site from the top view as a discretized grid in Figure 3.7.

3.3.2 Simulation and Validation of the Flow Model

I have simulated the flow model using the code formulated using the methodology described in section 3.3.1. Also, the same model is generated in MODFLOW using GMS (Groundwater Modelling Systems). I have compared the results from both the models and there is a near identical match between the results. The first set of simulated results is for the pumping case and the second set of results is for the pumping and the injection case.

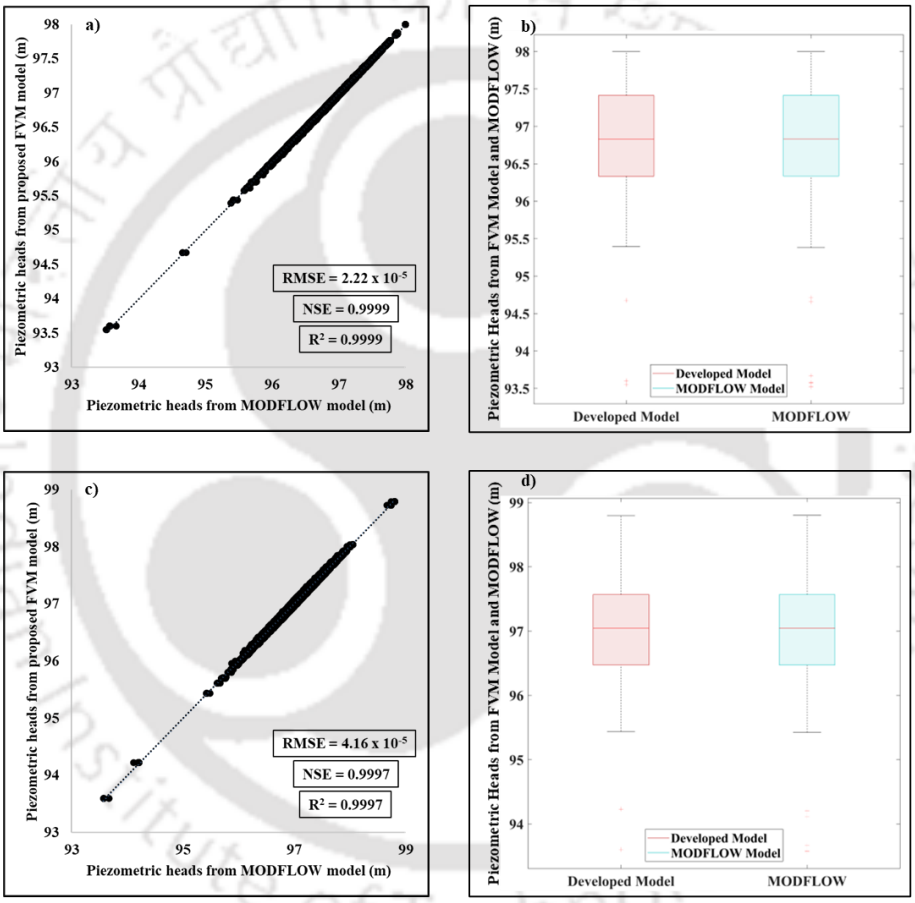

Figure 3.8: Comparison of heads generated from the developed simulation model and MODFLOW a) Scatter plots for pumping case, b) box plot for pumping case, c) scatter plots for injection case, and d) box plot for injection case

The maximum error is 0.071% (pumping) and 0.125% (pumping+injection), the Root Mean Squared Error (RMSE) is 2.22 x 10^{-5} (pumping) and 4.16 x 10^{-5} (pumping+injection), the Nash Sutcliffe Efficiency (NSE) is 0.9999 (pumping) and 0.9997 (pumping+injection), and the squared Pearson's correlation coefficient is 0.9999 (pumping) and 0.9997 (pumping+injection) in between the models. The match between the heads simulated using MODFLOW and proposed MATLAB code is shown in Figure 3.8. Part (a) and part (c) shows a scatter plot of

the pumping and pumping+injection wells results. Part (b) and part (d) show the box plot of the results for pumping and pumping+injection wells obtained from the models.

The contours of piezometric heads are also drawn by interpolating the results and smoothening the results of the head obtained. These contours also show a very close match of the results from both MODFLOW and proposed MATLAB codes. These contours are drawn at multiple time steps.

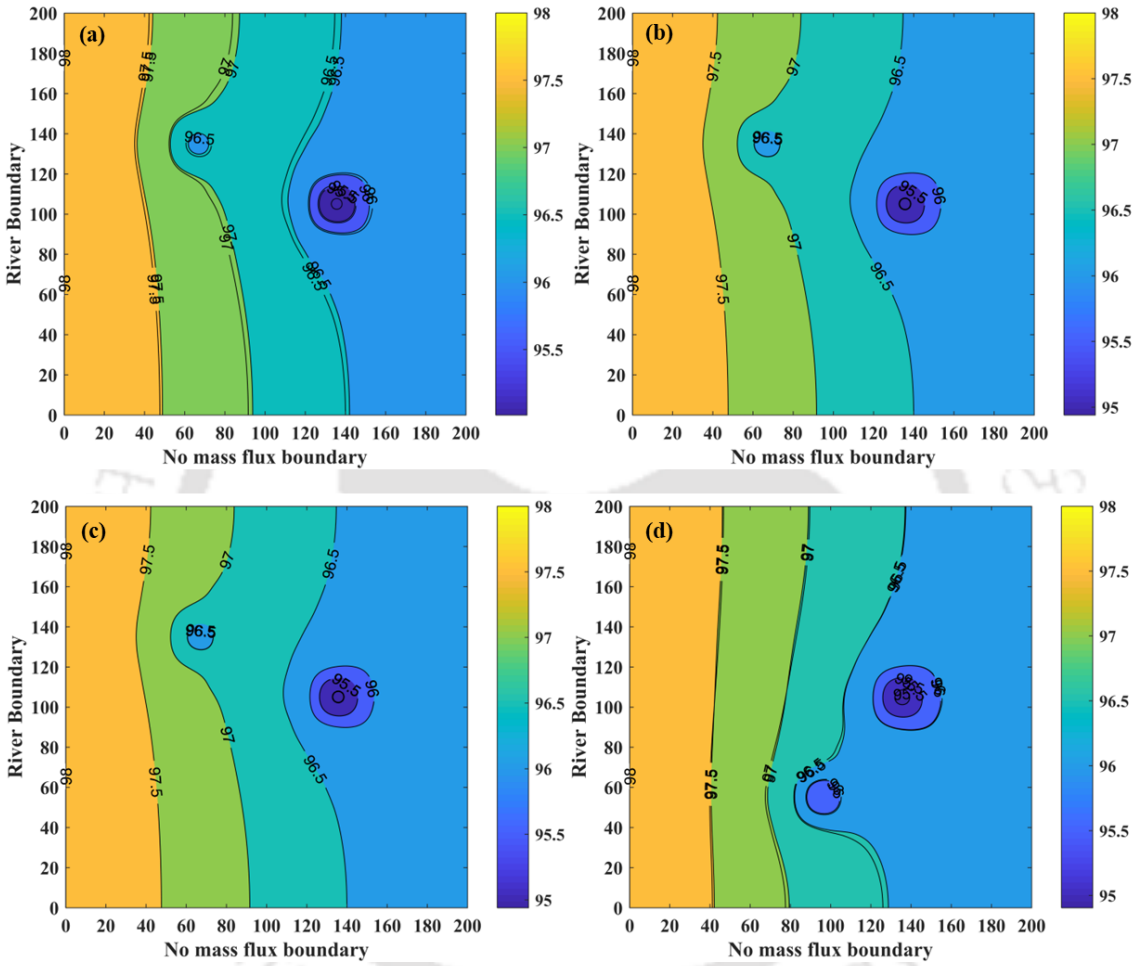

Figure 3.9: Contours of heads from MATLAB and MODFLOW for the pumping case

The contours on the 10th day, 50th day, 100th day, and 130th day are shown in Figure 3.9 from the subplots a, b, c and d. The color-filled contours are the contours from the MODFLOW, and the contours represented as lines are the contours from proposed MATLAB code. It can be observed that there is no significant change in the head distribution on the 50th and 100th day contours. This is because the steady state is reached by then and stays the same until well 2 becomes inactive and well 3 starts its operation. Color-filled contour plots are also generated for the same conditions (Figure 3.9). The contour plots for the heat flow are shown in Figure

3.10. The contours on the 10th day, 60th day, 100th day, and 130th day are shown in Figure 3.10 in subplots a, b, c, and d.

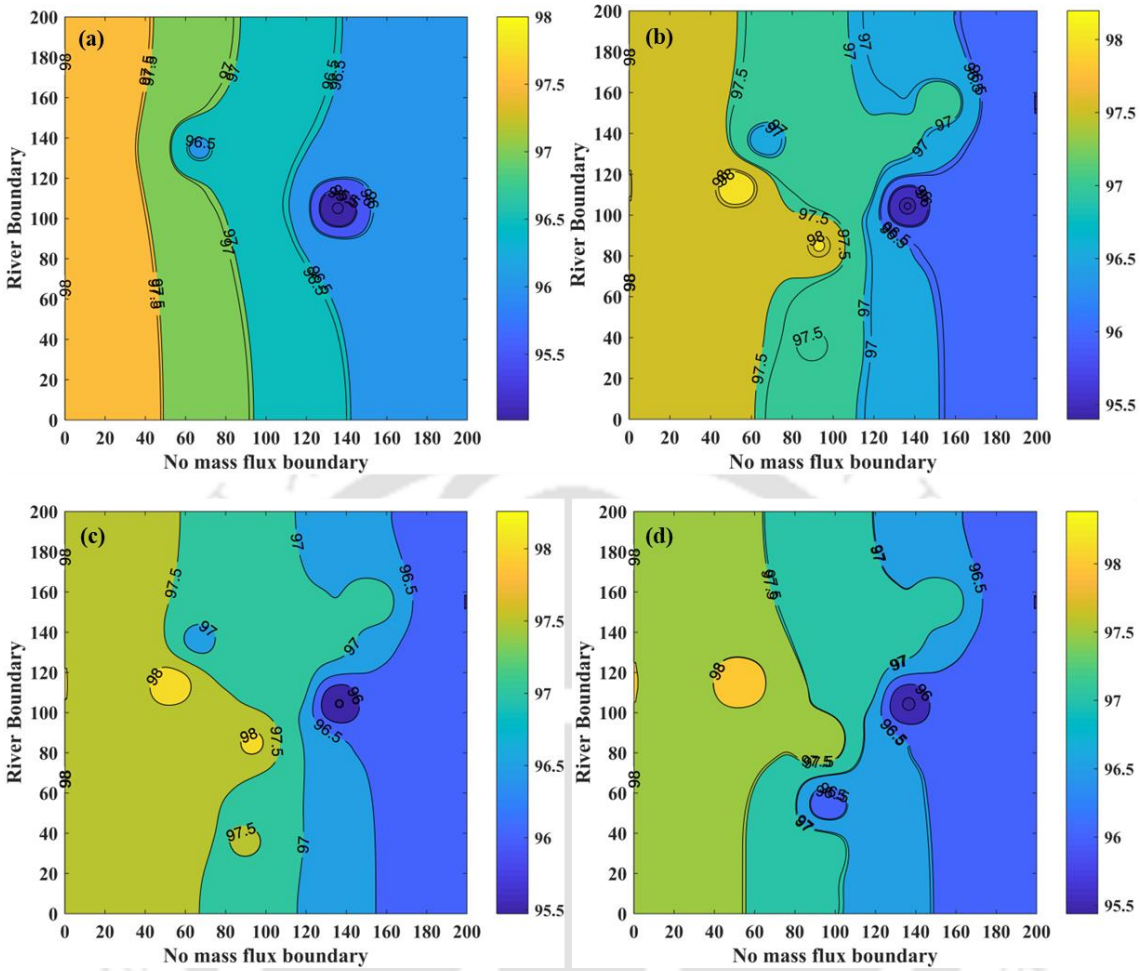

Figure 3.10: Contour of heads from MATLAB and MODFLOW for injection and pumping wells

Here, it can be observed in the contours of Figure 3.9(a) and (d); and Figure 3.10(a), (b) and (d) that the proposed FVM model is not exactly matching the results of the MODFLOW models. This is because of the difference of numerical schemes used in the models. The MODFLOW package uses Modified Incomplete Cholesky (MIC) factorization which works for only symmetric matrices. However, for higher accuracy, I have used Preconditioned Conjugate Gradient (PCG) method with Jacobi preconditioner which changes the results with a minor difference. However, it is evident from the RMSE, and NSE values that this difference is almost negligible for the practical purposes.

3.3.3 Simulation and Validation of the Transport Model

The transport model needs the input of the velocity data from the flow model. Therefore, the validated flow model results are used in the transport simulations. The virus transport model is

developed for equilibrium sorption under nonlinear isotherm conditions. The heat transport model is simulated using the same model conditions as described in section 3.2.3. The model inputs and boundary conditions are replicated in the MT3DMS model. Later, the results are compared for the study area and the problem description as in section 3.3.1. The injection wells are not considered in the aquifer for simulating the virus transport. However, in heat transport simulations, both the pumping and injection wells are considered. The correlation between the two results is given in Figure 3.11. There is a near-perfect validation between the results as per Figure 3.11. However, there are some irregularities due to numerical errors near the pumping locations of the aquifer. The errors are so negligible that the effect is not visible in the correlation of the results in Figure 3.11. The first set of simulated results is for the virus transport and the second set of results is for the heat trasnport. The Root Mean Squared Error (RMSE) is 7.52 x 10^{-5} (virus transport) and 0.001807 (heat transport), the Nash Sutcliffe Efficiency (NSE) is 0.9997 (virus transport) and 0.9972 (heat transport), and the squared Pearson's correlation coefficient is 0.9999 (virus transport) and 0.9973 (heat transport) in between the models.

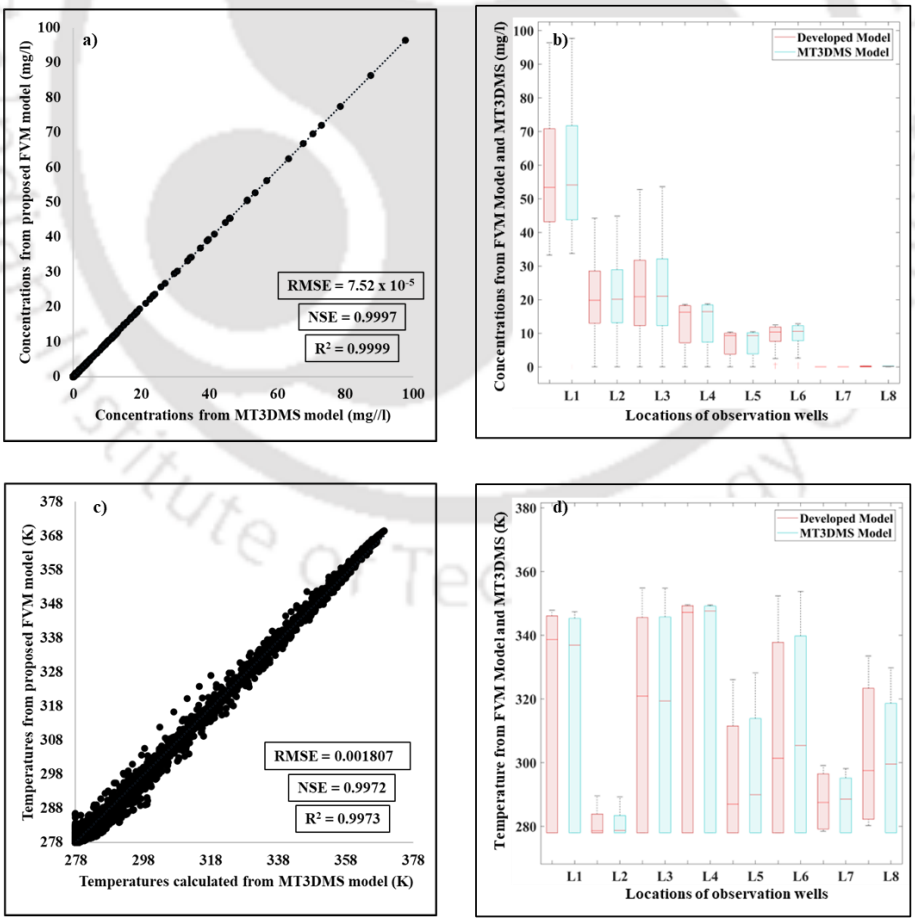

Figure 3.11: Correlation between simulated concentrations between the MT3DMS model and proposed MATLAB Code

The contour plots for the different timesteps in the modeling of virus transport are given in Figure 3.12. The contour plots are plotted for 60 days, 90 days, 120 days, and 150 days of simulation in Figure 3.12(a), Figure 3.12(b), Figure 3.12(c), and Figure 3.12(d), respectively. The color-filled contours are the results of the MT3DMS model, and the overlapping contours that are line filled are results from the MATLAB model. As the contours are so closely matching, the results are similar and indistinguishable in Figure 3.12.

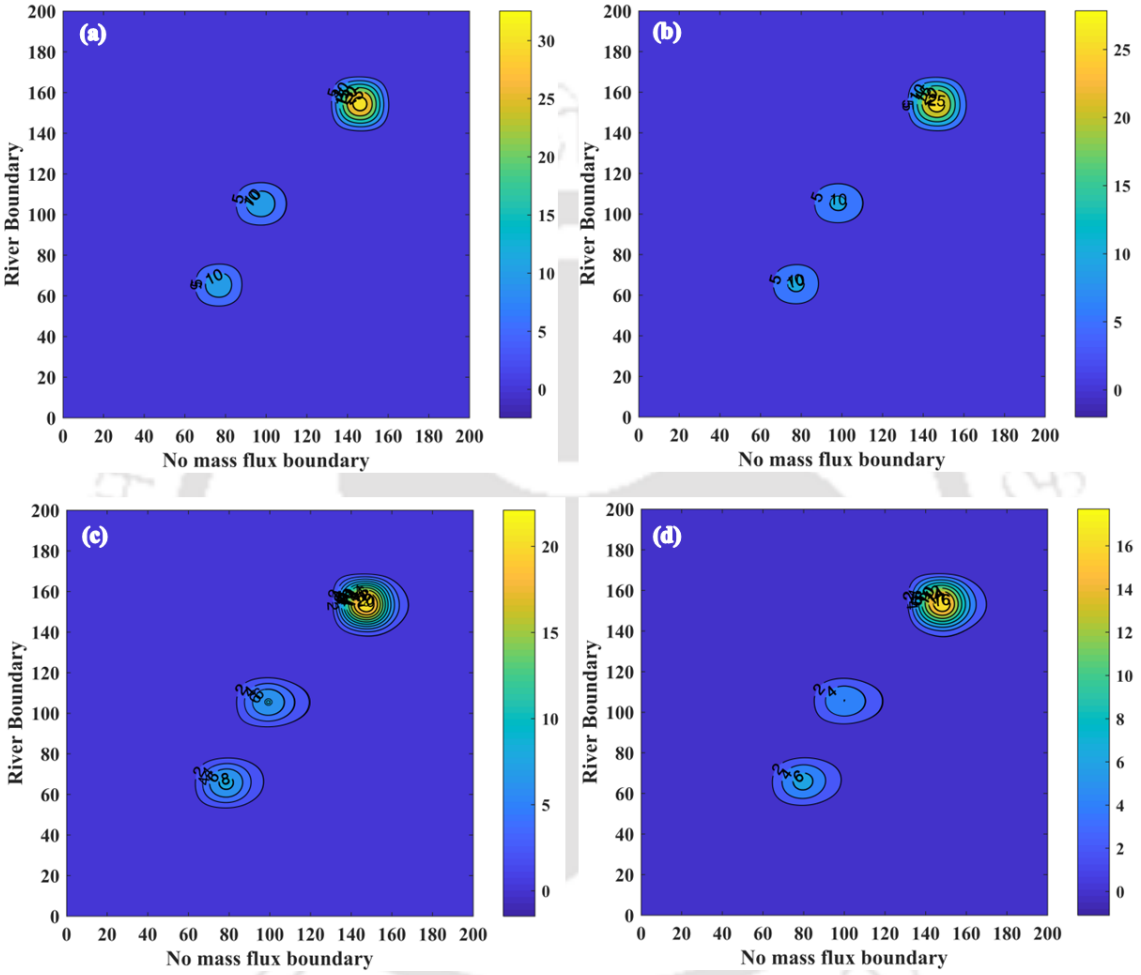

Figure 3.12: Contour plots for the comparison of virus transport model from MT3DMS and proposed MATLAB Code at 60 days, 90 days, 120 days, and 150 days of simulation

Figure 3.13 shows the breakthrough curves of the models comparing the results at eight observation locations. Three locations are exactly the source locations, three locations are just right to the source locations, and the other two are at the pumping well locations. Each location is represented as position (i,j) in the figure, where 'i' represents the grid point location from the top of the model in the downward direction and 'j' represents the grid point location from the model's left and in the right direction.

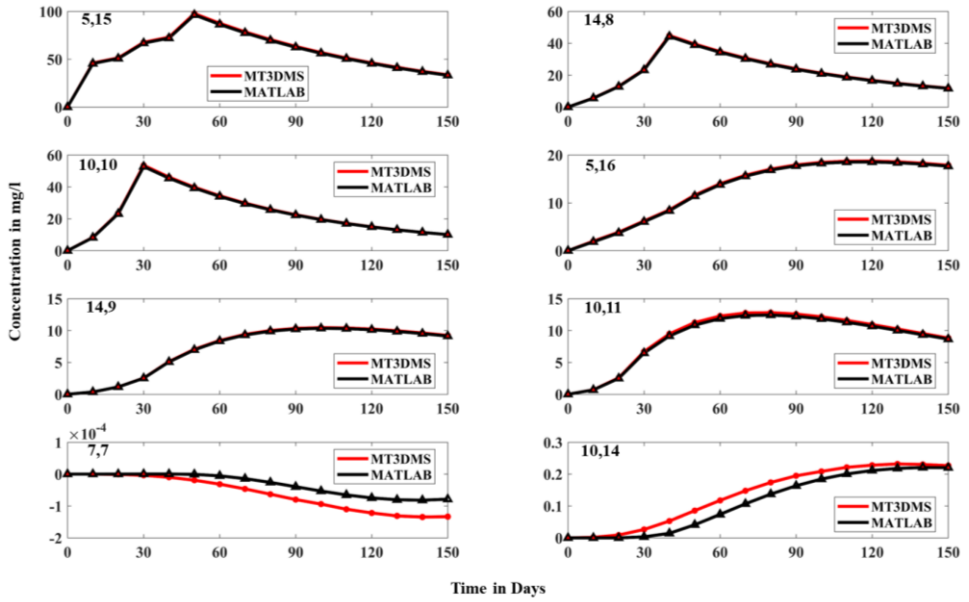

Figure 3.13: Breakthrough curves at different locations of observation wells comparing the results of virus transport model from proposed MATLAB code and MT3DMS

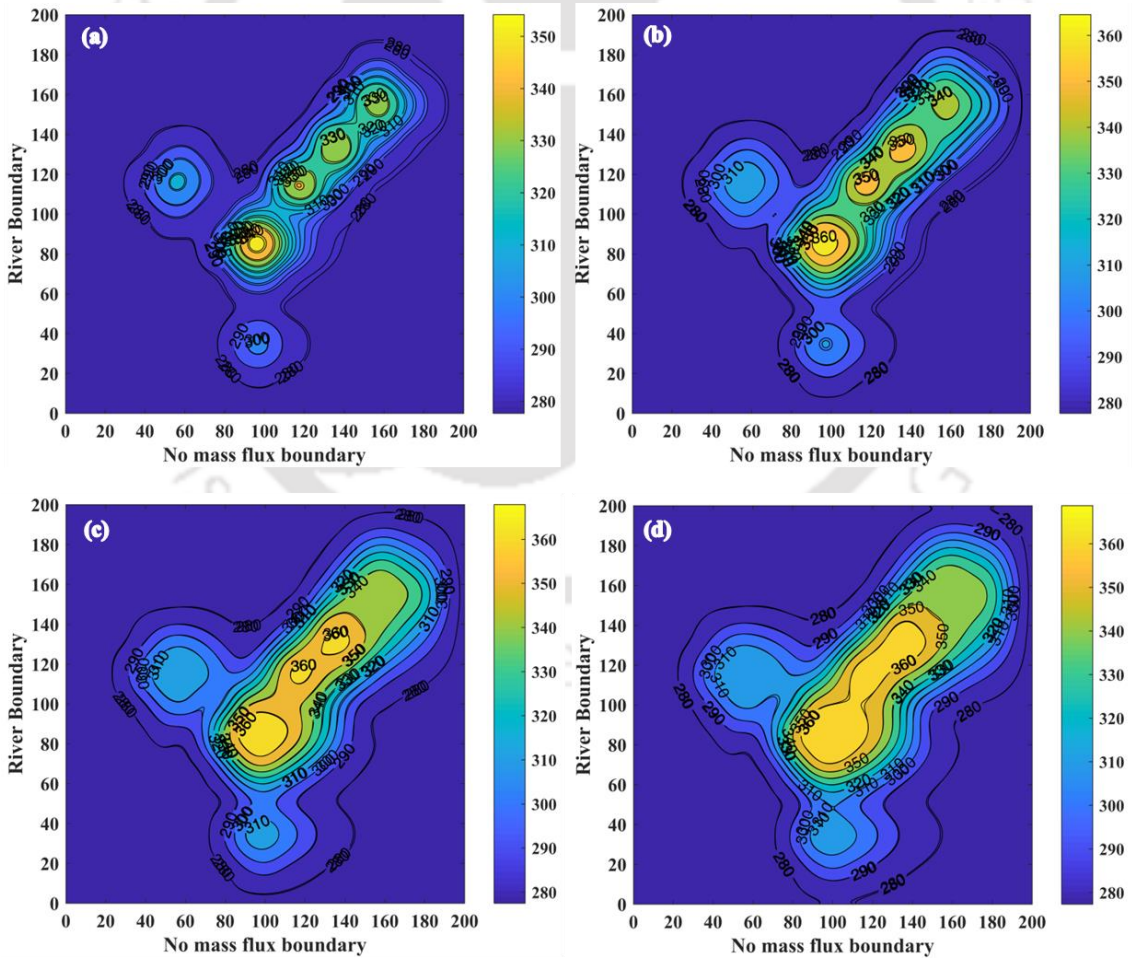

Figure 3.14: Contour plots for the comparison of heat transport model from MT3DMS and proposed MATLAB Code at 60 days, 90 days, 120 days and 150 days of simulation

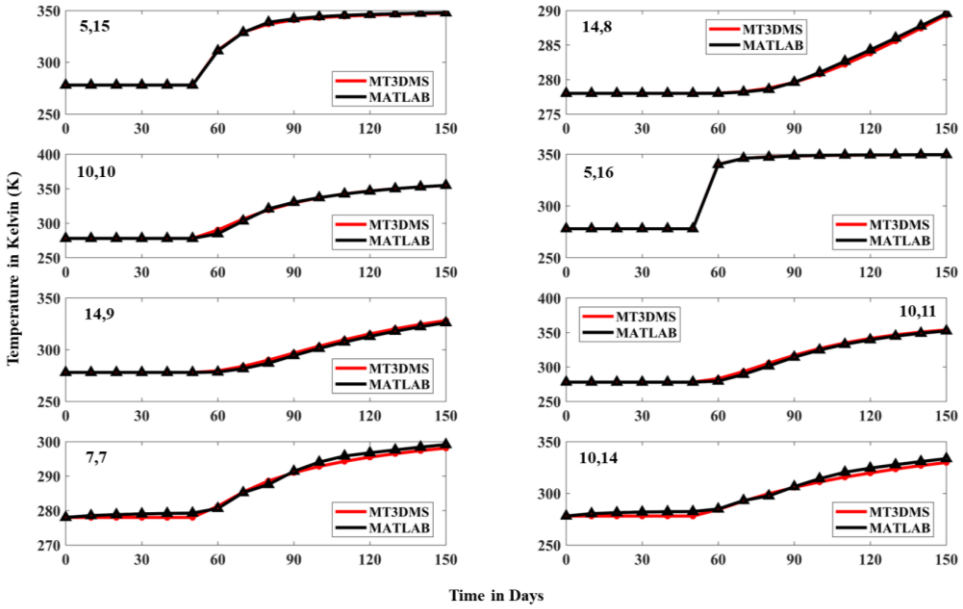

Figure 3.15: Breakthrough curves at different locations of observation wells comparing the results of the heat transport model from proposed MATLAB code and MT3DMS

The contour plots of the heat transport model are shown to compare models in Figure 3.14. The breakthrough curves of both the models plotted at the exact specific locations as that of the virus transport model are given in Figure 3.15.

3.3.4 Applicability of the Coupled Virus and Heat transport models

The models developed can simulate the various conditions of virus transport under the effect of temperature and other fluid parameter changes. To show the effect of virus removal using this concept of variable temperature, I have considered studies of the effect of temperature on inactivation rate (Hurst et al., 1980; Yahya et al., 1993). Schijven and Hassanizadeh (Hassanizadeh et al. 2000; Schijven et al. 1999) have reported that the virus (MS2) changes its inactivation rates with the increase in groundwater temperature, and their rate of adsorption is affected by the change in temperature (Vilker et al. 1980). They further concluded that the inactivation rate follows an exponential trend, and the coefficients for adsorption isotherm follow a linear trend with temperature in a Freundlich isotherm for equilibrium sorption. The equations for these behaviors are given in equations 3.33 and 3.34, respectively, for inactivation and adsorption behaviors. The researchers also observed that the inactivation coefficient changes to the variations in temperature up to 35^0 C. At higher temperatures like 40^0 C or 50^0 C, the change in inactivation coefficient was almost negligible. Therefore, equation 3.33 accounts for the invariability of the inactivation rate at higher temperatures. The threshold temperature (T_a) for the change in behavior of the inactivation rate is found based on many

studies (Hassanizadeh et al. 2000; Yahya et al. 1993; Yates and Yates 1988). The values of T_a and μ_a for this study are 32.72 ^{0}C and 1.10 day$^{-1,}$ respectively. The sigmoid function is known to extrapolate the behavior of any function to a smooth constant value without any discontinuity or non-differentiability. Hence, I have chosen the sigmoid function to account for the change in behavior of the function.

$$\mu_l = \begin{cases} e^{aT+b} & T \le T_a \\ \mu_a + \dfrac{e^{T+T_a}}{1+e^{T+T_a}} & T > T_a \end{cases} \qquad 3.33$$

$$\begin{aligned} K_f &= a_1 T + b_1 \\ m &= a_2 T + b_2 \end{aligned} \qquad 3.34$$

Here, the term μ_l represents the inactivation rate in the aqueous or liquid phase, and the terms K_f and m represent the parameters of the nonlinear Freundlich isotherm in the equilibrium sorption. The terms a, a_1, a_2, b, b_1, and b_2 are the constants of proportionalities obtained through regression analysis which have the values 0.12, 0.0969 x 10^{-6}, -0.0069, -3.5, 0.1825 x 10^{-6} and 0.8875, respectively for MS2 viruses. The inactivation rate in the attached phase (μ_s) also behaves the same as in the aqueous phase. One of the assumptions is that the inactivation in the attached phase is 2/3rd of the inactivation rate of the aqueous phase (Rajeev Gandhi et al. 2017). The effect on inactivation factors and relative sorption with the concentration (for the aqueous concentration of 100 mg/l) concerning temperature is shown in Figure 3.16.

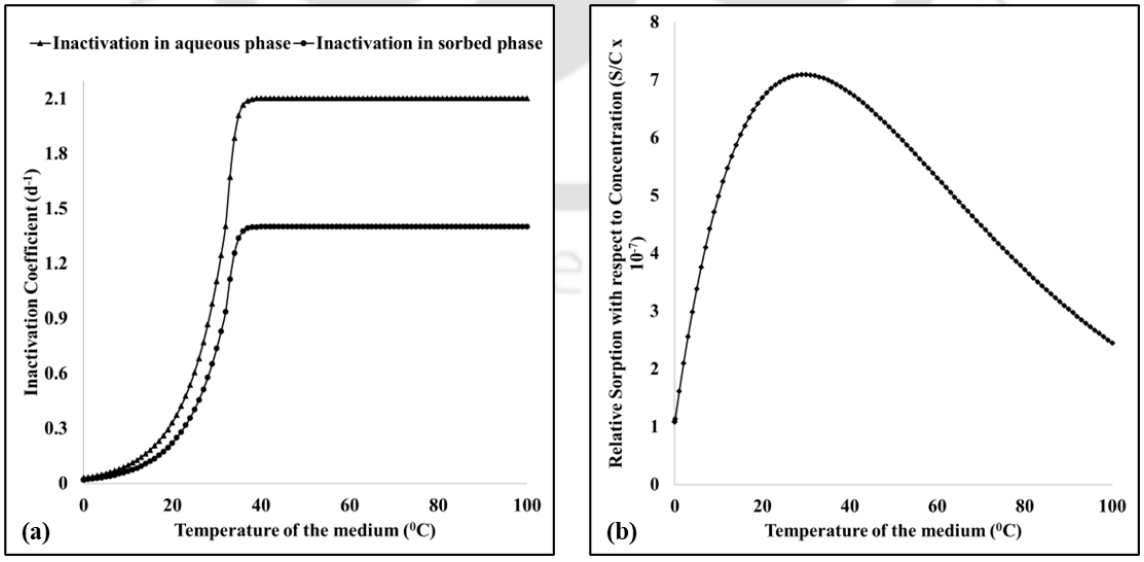

Figure 3.16: Inactivation coefficients and relative sorption (at C = 100 mg/l) with respect to temperature

The overview of the procedure is followed in Figure 3.17. The first part of the methodology is to select any arbitrary well location for the injection of warm water. The flow profile changes with the introduction of a source in the aquifer. Therefore, I have used the flow model with injection wells along with the existing pumping wells. As the flow model doesn't deal with the properties of the fluid used for the injection, the injection wells act as a simple source without any effect on the temperature. The flow profile thus created is a combined effect of the existing pumping wells and the introduced injection wells. The piezometric heads and the corresponding velocity vectors are inputs to the transport of viruses as well as the transport of hot water.

For simulating the transport of viruses under normal conditions, I have used the flow model for only the pumping wells. The simulated virus transport model gives the virus concentration in the aquifer under conditions without treatment. The injection wells and the pumping wells are used in the flow model to simulate heat transport. After I have simulated hot water transport in the aquifer, the temperature is obtained as a function of space and time. The virus inactivation and adsorption are affected by temperature according to the relations in equations 3.33 and 3.34. Therefore, I have obtained the parameters as a function of space and time. Now, I have used the updated inactivation rates and adsorption coefficients to simulate the virus transport. Hence I have obtained the virus concentration with the effect of temperature change.

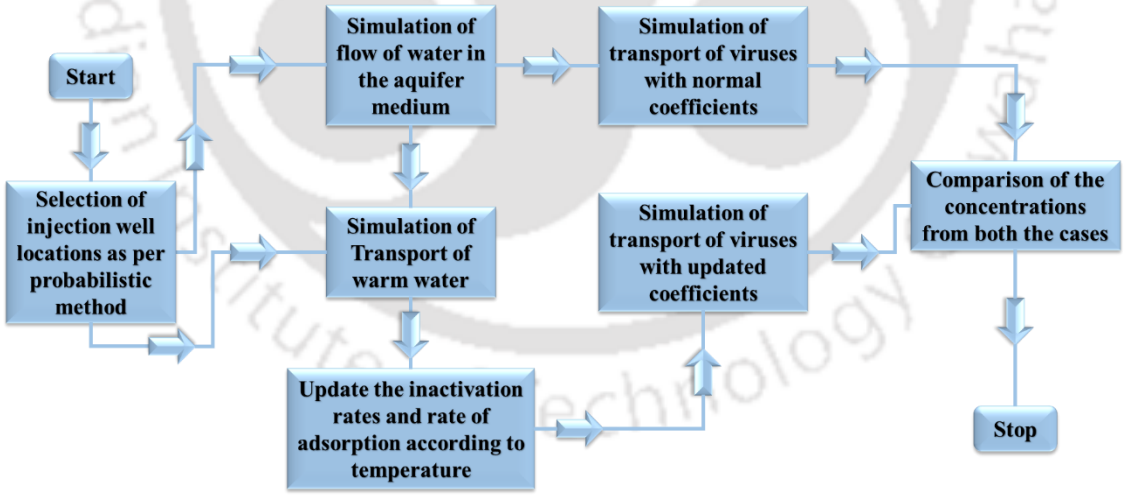

Figure 3.17: Flowchart of the models used to generate the effect of temperature on virus transport

3.3.5 Simulation of the Virus Transport Model with Variable Temperature

I have selected three wells at crucial locations for hot water injection and run the model for a nominal pumping rate of 100 m^3/day at each location for 10 days. The injection water temperature is near boiling, i.e., 100^0 C (373 K), to minimize the usage of temperature

regulators. As explained in section 3.3.3, the simulation model takes the aquifer's flow with the injection wells, heat transport model, and the virus transport model with varied inactivation and sorption coefficients. Therefore, the effect of temperature on viruses with injection wells in the aquifer can be studied quite carefully.

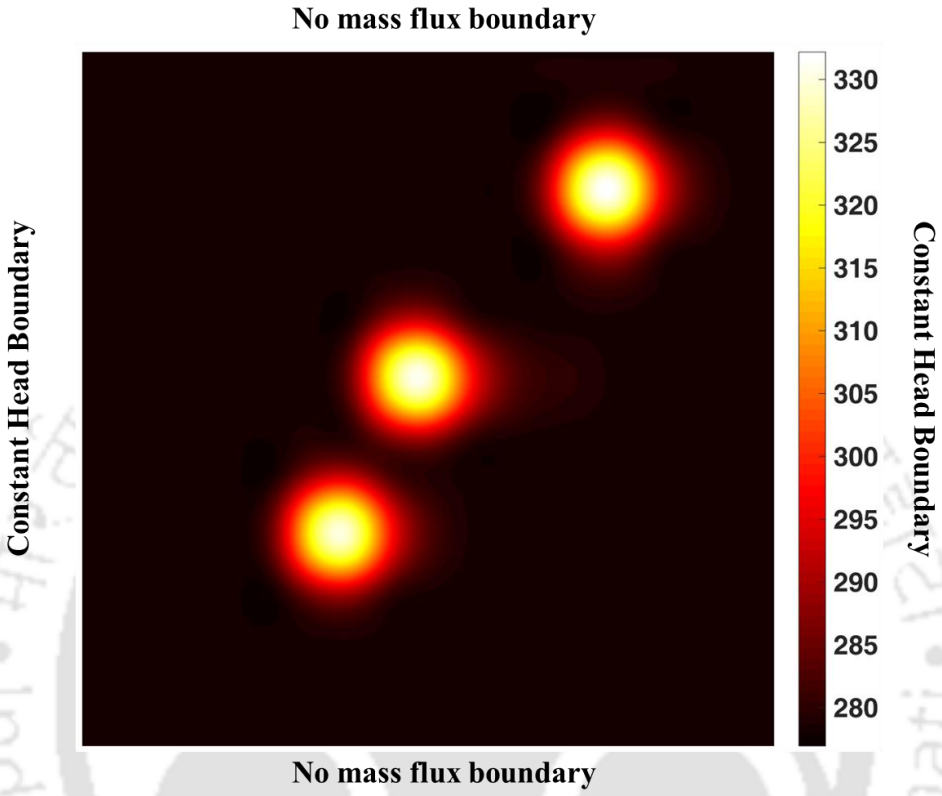

Figure 3.18: Contour plots of the temperature in the model at 150 days of simulation

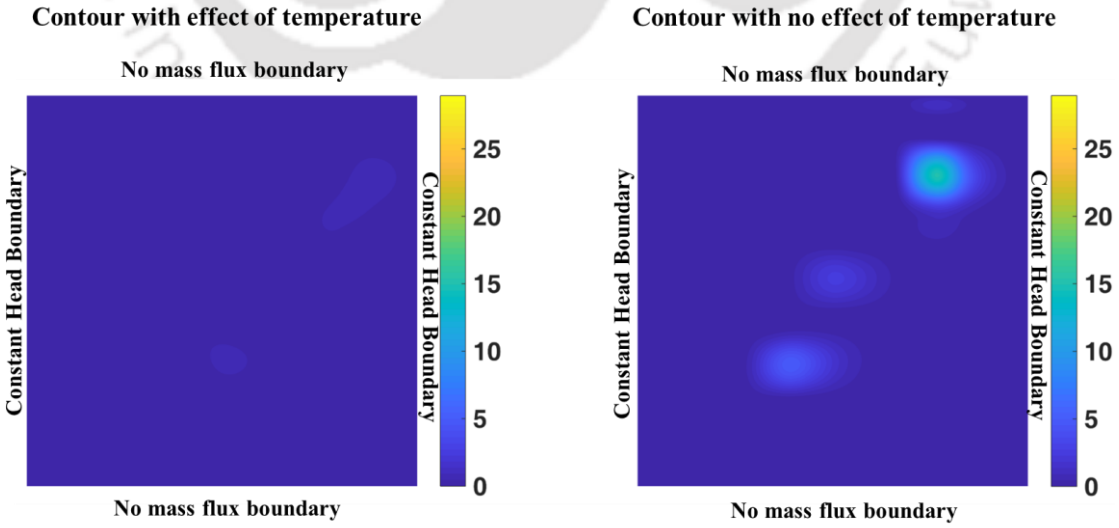

Figure 3.19: Contours of Virus transport model with the effect of temperature and without the effect of temperature

The spatial variation of temperature obtained using the heat transport model is given in Figure 3.18. The results of the virus transport model with the effect of temperature and with no effect of temperature at the end of day 150 are given in Figure 3.19. The effect of temperature at different locations (*i*, *j*) as breakthrough curves and concentration variations at the source locations are given in Figure 3.20. The locations are selected so that all the crucial points of the aquifer are addressed. The resulting variations of the effects of temperature are visible in Figure 3.20.

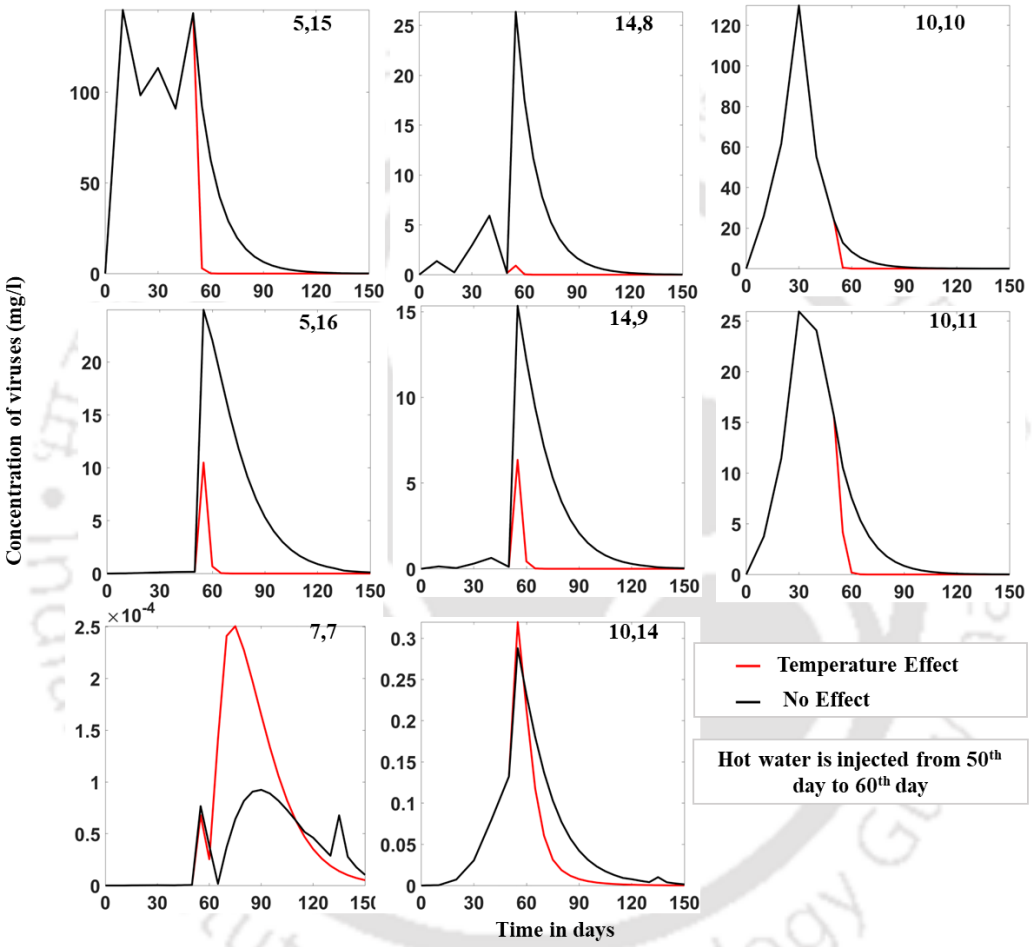

Figure 3.20: Break-through curves at different locations and concentration profile with time at source locations from Virus transport model with the effect of temperature and without the effect of temperature

Here, in the last two sub-plots (i.e., at locations 7,7 and 10,14) it can be seen that the concentration increased at first and then it started to decrease along with time. This is due to a very specific reason of the locations being the pumping well locations. The general flow profile of a pumping well is a sink type profile with all the velocity vectors pointing towards the well around the location. When any injection well is introduced into the aquifer, it changes the flow geometry in such a way that the groundwater table rises at all the locations till where the radius

of influence of the well reaches. With the rise in groundwater table and an existing sink point, the overall hydraulic gradient increases at and around the pumping well. This increases the flow velocity and therefore the advection component increases the concentration arbitrarily. However, as the effect of the increased temperature reaches the location of pumping or when the water with higher concentration leaves the aquifer medium, the concentration decreases and it can be seen in both the sub-figures at the latter half of the break through curves.

3.4 Conclusions

The models developed in this chapter proved that the virus transport can be modeled with variable parameters using the experimental data. In this chapter, I have attempted to inject hot water into injection wells to evaluate its impact on the concentration of the virus. The inactivation and adsorption rates are changed with the change in temperature. I have also observed that introducing new injection wells into the aquifer changes the flow geometry. Therefore, the flow velocities change according to the wells' locations, which changes both the advection terms and the mechanical dispersion terms in the aquifer. The merit of this study is that the models consider the changes in aquifer flow properties before the transport simulations. The demerit is that introduction of hot water into the aquifer media changes the pore properties in the aquifer. Also, water viscosity is sensitive to temperature, and its changes affect the hydraulic conductivity. The molecular diffusion coefficient also depends on temperature and changes due to high temperature. The soil grains might get affected due to the temperature changes by changing the porous matrix properties. I haven’t considered such effects in the aquifer properties in the study. I have assumed that temperature affects only the inactivation rates and adsorption in the aquifer, which is the limitation of this study.

This study shows that the virus concentrations can be effectively reduced by the injection of hot water into the aquifer at appropriate locations. However, there are other methods, such as pump-and-treat or altering the pH in aquifer media. Most of those strategies would take more cost because the injected chemicals or the altered pH in the aquifer should be returned to a normal state for the water to be potable. But while using the effect of temperature, the water does not need to be returned to the original temperature. The effects of not getting the water back to the initial temperatures might change some conditions in the aquifer. In such a case, it takes a minimal injection of normal or cold water without any chemicals, which would be both cost and time-efficient.

CHAPTER 4: DEVELOPMENT OF A PORE NETWORK MODEL

4.1 Introduction

The pore-scale modeling is a random model of pores that randomly represents the arrangement of the pore network, producing comparable properties while averaging to a Darcy scale (Raoof and Majid Hassanizadeh 2010). The model produces the results of pore network arrangements differently for every run randomly, but the overall effect of the randomized network stays the same throughout the REV. Such a pore scale model needs to improve the insight into micro-scale flow and transport processes. The distribution of pore throats, the size and shape of pore bodies, and the geometry of pores play a significant role in the assessment of the overall behavior of the pore network (Hemes et al. 2015; Raoof et al. 2010a; Wan and Wilson 1994; Xiong et al. 2016). The characterization of the pore network with its connections and other eliminations can be calculated by its average coordination number ($\bar{z}$). The coordination number is a value of the number of connections each pore body has with all its adjacent bodies through pore throats (Ioannidis and Chatzis 2000). The range of these coordination numbers is from 3.5 to 4.5 in different types of sandstones (Raoof and Majid Hassanizadeh 2010) but can vary from any value of 0 to 26 theoretically.

The generation of the pore network model is done in this chapter by the procedure followed by Raoof and Hassanizadeh (Raoof and Majid Hassanizadeh 2010) till the generation of the network and mean coordination number etc. However, after the generation of the pore network is done, the upscaling of the pore properties to Darcy scale parameters, such as the permeability, tortuosity, longitudinal and transverse dispersivity, porosity etc., are found using the traditional methodology of Bear and Bachmat (Bear and Bachmat 1986, 1990) for isotropic conditions as assumed for the problems in the previous chapter. The randomization of radii of pore bodies is done using two kinds of pore size ranges with a mean and standard deviation of pore sizes different from each other but in range for the sand pores (Houben et al. 2013; Raoof and Hassanizadeh 2013; Rijnaarts et al. 1996). The pore throats were assumed to be of negligible dimensions for the studies carried out till now, and the volume of water in the system is assumed to be wholly contained in the pore bodies, i.e., the spherical-shaped junctions (Raoof et al. 2010b, 2013; Raoof and Majid Hassanizadeh 2010). But for this study, I have calculated

the volume of the connecting pores throats by the minimum area method described in the methodology section.

The pore network model should generate a REV for any required value of hydraulic conductivity, porosity, and longitudinal and transverse dispersivity to any desired value. Therefore, an optimization model is used with 5 variables, the mean and variance of the two categories of pore bodies and the percentage division of both the categories in the pores. The model used is the shuffled frog leaping algorithm (Gandhi and Bhattacharjya 2020a), the basics of which are explained in the work. The algorithm's results are given in this chapter's result and discussion section.

4.2 Methodology

The objective is to find an optimum pore network model that can define the Darcy scale parameters through the pore properties. The overall methodology is described in the flowchart of Figure 4.1 given below.

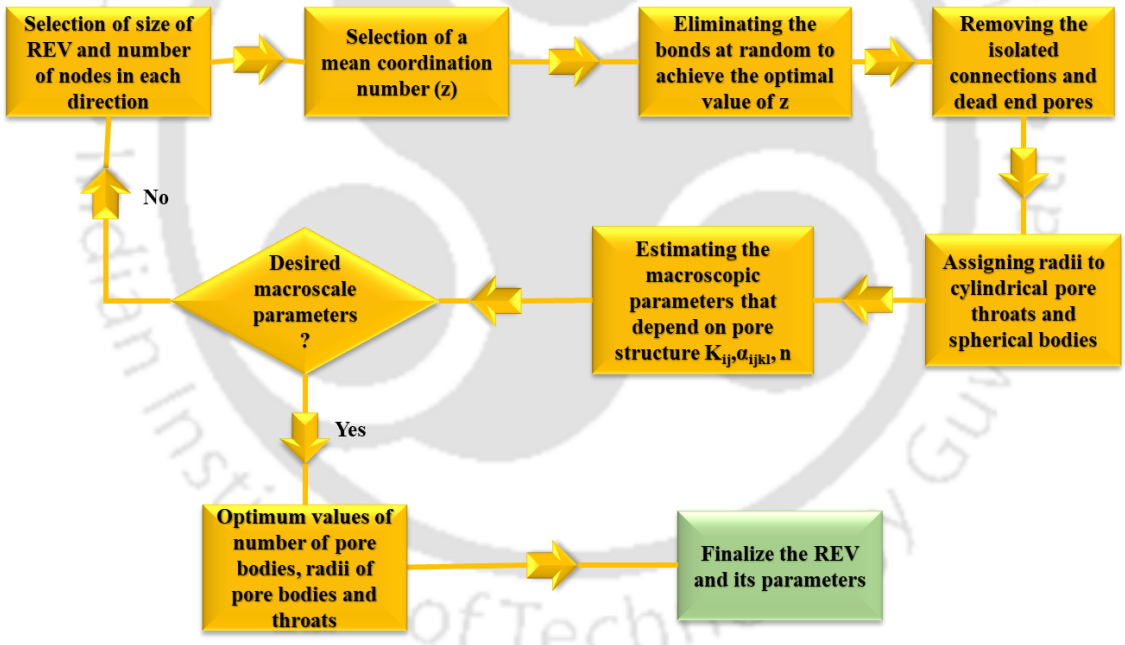

Figure 4.1: Overall methodology of generating a pore network model

The size of REV is the size of the grid that can be used for the Darcy scale simulation. The number of nodes in each direction also depends on the type of soil used for simulation. The mean coordination number is also selected in the next step, which comes from the soil type used. The bonds are then randomly eliminated to get the mean value of the coordination number as described in the subsequent sections. The isolated and dead-end pores should also be removed as they do not conduct any water through them. The radii to the cylindrical pores and

spherical bodies are assigned as shown in the methodology of subsequent sections. The macroscopic parameters are estimated using the procedure given below, and then, based on the desired values of macroscopic parameters, the optimum values of the pore bodies and throat radii are finalized as REV parameters.

4.2.1 Definition of a REV (Representative Elementary Volume)

The definition of a pore network comes in to picture when one chooses a volume of the network on such a scale that it contains both the pores and the solid surfaces. Such a choice of volume where one can clearly define the porosity is called a Representative Elementary Volume (REV). The change in porosity with the choice of volume is shown in Figure 4.2. If one chooses the volume smaller than REV, the porosity fluctuates to either 0 (void space) or 1 (Solid phase) with a smaller volume selection. If one chooses a volume greater than REV, then the porosity blows to a higher value from the constant value in REV. Therefore, the choice of REV should be a volume where the porosity is clearly defined at a constant value.

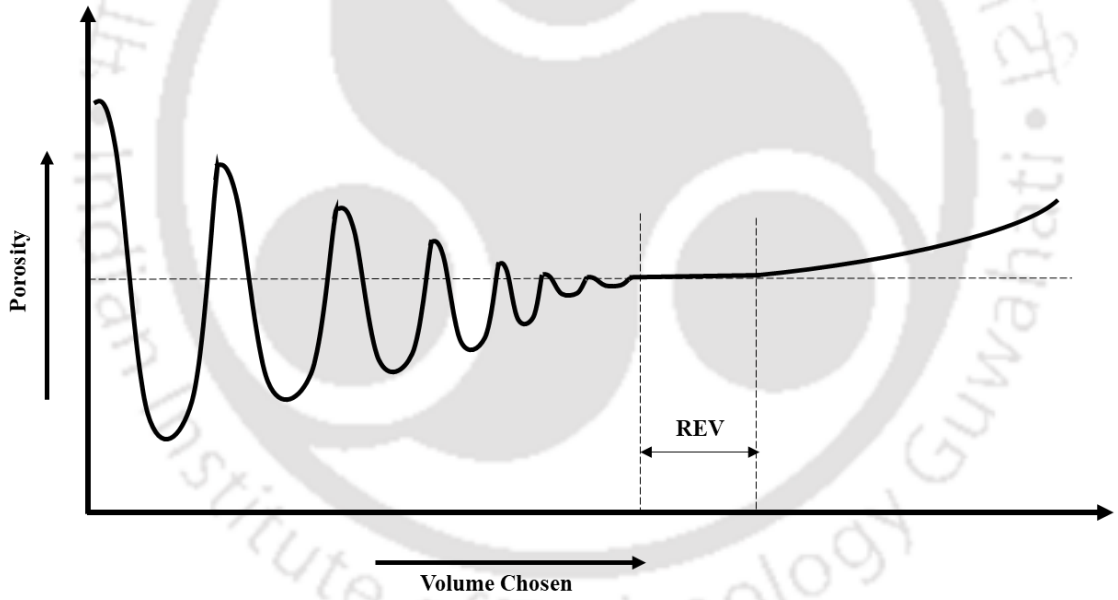

Figure 4.2: Change in porosity with the choice of volumes

4.2.2 Generation of simple pore network model

The pore networks are the network of tubes that are tortuous (curved in the network) connected to either side faces of the Representative Elementary Volume (REV). For simplicity, the pore network is formulated as a cubic lattice with a possibility of a bond in any possible direction where the pore junctions are spheres, and the pore spaces are cylinders. This pore network model with spheres and cylinders is widely accepted and is considered a good conceptualization of the pore network (Raoof and Majid Hassanizadeh 2010). Consider a pore

network model of three layers in each direction as a cubic lattice, as shown in Figure 4.3. The directions of all the possible connections are thirteen. The figure shows the directions of all the central site's forward connections. The number of layers in each direction is three for this example. The layers for each direction are given by the term 'N_i', 'N_j', and 'N_k' for the three directions of i, j, and k, as shown in the figure. One can think of these directions as the global coordinate system and directions 1 to 13 as the local directions of the pore connections. As the total number of sites that would be available in a pore network is given by N_i x N_j x N_k, the number of sites for this example is 27.

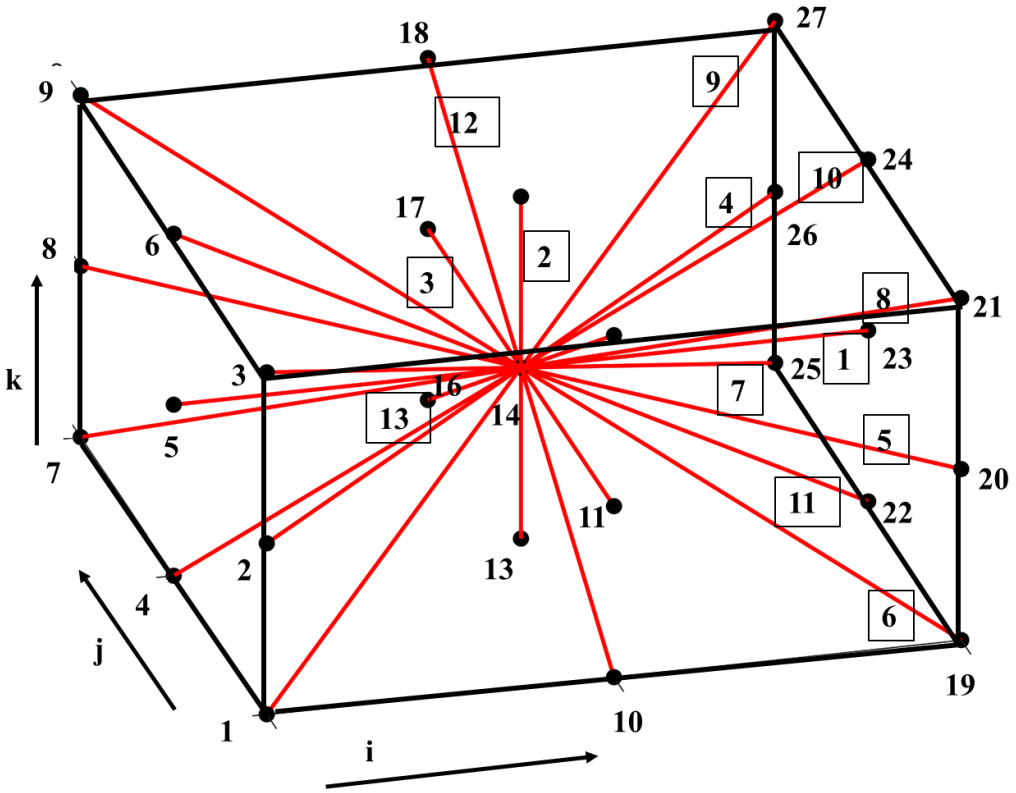

Figure 4.3: A basic cubic lattice with only central node connections

4.2.2.1 Average Coordination Number ($\hat{z}$)

The average coordination number measures the average number of connections of a particular site in the network. In Figure 4.3, the central site, i.e., site number 14, is connected to 26 sites (as it is the only site connected to all the other sites in the cubic lattice). Of these 26 connections, thirteen are forward connections, and thirteen others are backward connections. Therefore, site number 14 has a coordination number (z) of 26. All the other sites have a coordination number of 1. Therefore, the average coordination number ($\hat{z}$) will be 1.925. Generally, for any pore network, the coordination number ranges from 3.5 to 6, according to many studies (Bakke and

Øren 2002; Ioannidis and Chatzis 2000). However, for a particular type of soil and the condition of saturation, the coordination number is pre-specified and can be supported by experimental evidence.

4.2.2.2 Connection Matrix

The connection matrix defines the whole network as a set of connections from one site to another. The matrix is a sparse matrix with entries as the direction in which the sites are connected. Therefore, the matrix has entries only in the places where there is a bond between the two sites. The flow is assumed to be from the left to the right direction. The bonds between those sites have been eliminated to avoid the in-plane flow in the left face and the right face. The whole network with all possible connections is shown in Figure 4.4. The matrix showing these connections is given as a connection matrix shown below in Figure 4.5. As the flow in plane of the left and right face are not allowed, the matrix has no entries in the sites 1 to 9 and 19 to 27. It is to be noted that the matrix is symmetric, and the forward connections are given in the upper triangle, and the backward connections are given in the lower triangle.

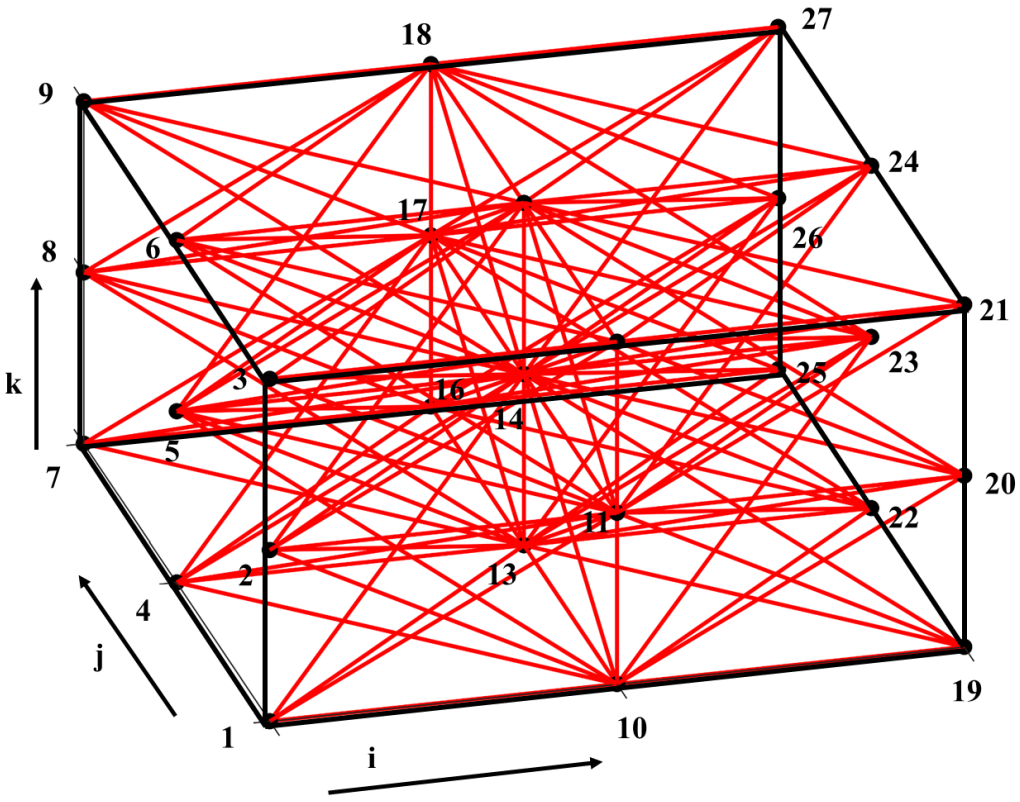

Figure 4.4: Full connections of all the possible bonds for cubic lattice

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1										1	10		4	9													
2										11	1	10	7	4	9												
3											11	1		7	4												
4										5	8		1	10		4	9										
5										6	5	8	11	1	10	7	4	9									
6											6	5		11	1		7	4									
7													5	8		1	10										
8													6	5	8	11	1	10									
9														6	5		11	1									
10	1	11		5	6						2		3	12					1	10		4	9				
11	10	1	11	8	5	6				2		2	13	3	12				11	1	10	7	4	9			
12		10	1		8	5					2			13	3					11	1		7	4			
13	4	7		1	11		5	6		3	13			2		3	12		5	8		1	10		4	9	
14	9	4	7	10	1	11	8	5	6	12	3	13	2		2	13	3	12	6	5	8	11	1	10	7	4	9
15		9	4		10	1		8	5		12	3		2			13	3		6	5		11	1		7	4
16				4	7		1	11					3	13			2					5	8		1	10	
17				9	4	7	10	1	11				12	3	13	2		2				6	5	8	11	1	10
18					9	4		10	1					12	3		2						6	5		11	1
19										1	11		5	6													
20										10	1	11	8	5	6												
21											10	1		8	5												
22										4	7		1	11		5	6										
23										9	4	7	10	1	11	8	5	6									
24											9	4		10	1		8	5									
25													4	7		1	11										
26													9	4	7	10	1	11									
27														9	4		10	1									

Figure 4.5: Connection matrix for the cubic lattice (Raoof and Majid Hassanizadeh 2010)

4.2.2.3 Bond Elimination

The cubic lattice with all the possible connections generates a mean coordination number of 8.74. But there is a requirement for the coordination number to be desirable as per the material for which the pore network model is being generated. Therefore, there is a need to eliminate some bonds in the network to reduce the coordination number to a desirable value. An optimization model can be developed to reduce the coordination number by eliminating the bonds. But, a much easier procedure is given by Raoof and Hassanizadeh, which is explained as follows (Raoof and Majid Hassanizadeh 2010). The first step is to select a desired value of the coordination number and create a vector of the size of 13 rows and one column with a value of the desirable average coordination number divided by 26. Equation 4.1 gives the entries of such vectors from 1 to 13.

$$v_i = \frac{\bar{z}}{26}, i = 1 \, to \, 13 \tag{4.1}$$

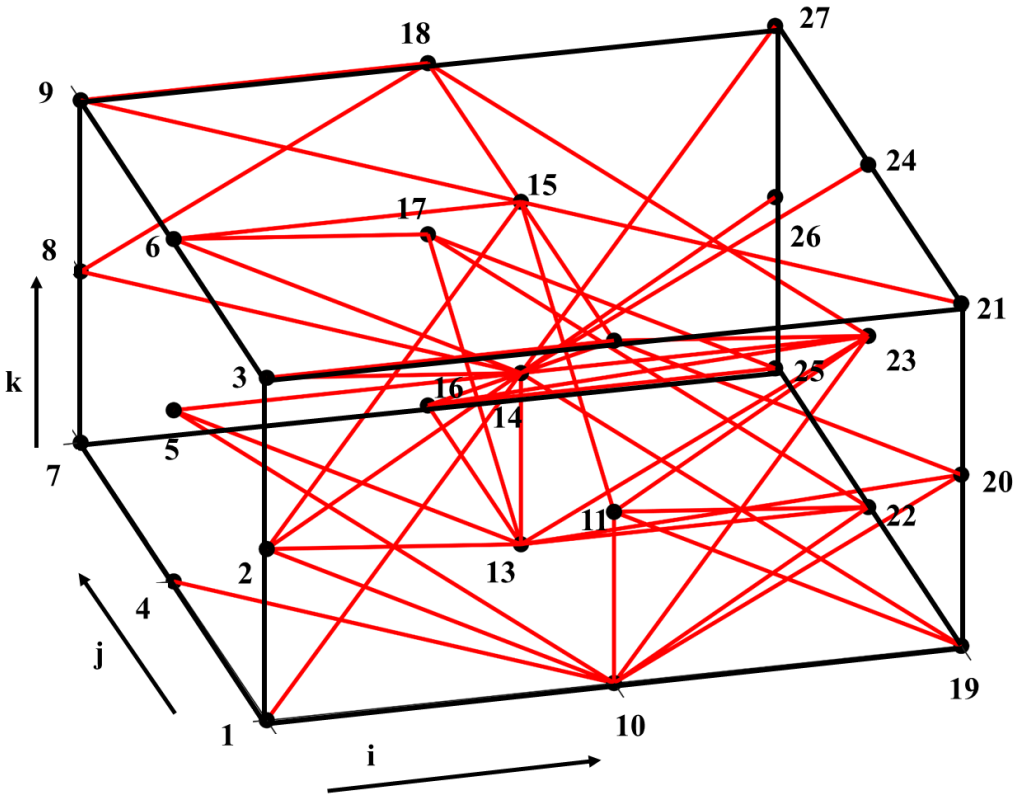

Figure 4.6: Connections in cubic lattice after elimination

The second step generates a random number between 0 and 1 for all the existing bonds. Then, if the random number for a specific bond exceeds the vector entry for that direction of bond v_i, the bond is eliminated. The condition for the bond state is given in equation 4.2.

$$Bond\ State = \begin{cases} Open & if\ rand() < v_i \\ Closed & if\ rand() > v_i \end{cases} \tag{4.2}$$

Initially, it is assumed that all the bond states are open, and then, by following this elimination procedure, the bonds are eliminated. The cubic lattice after eliminating the bonds is given in Figure 4.6. The average coordination number is reduced to a value of 3.62 after elimination.

4.2.2.4 Isolated clusters and dead-end bonds

The isolated clusters are those connections of the pore sites which do not contribute to the flow and transport processes. Dead end bonds in the structure of pores are the ones that do not contribute to the flow of fluid but are very important for the processes of diffusion and dispersion in solute transport. An example of the isolated clusters and dead-end bonds is given

in Figure 4.7a. However, the isolated clusters are of no use in modeling pore networks and pose unnecessary complications for simulating the flow and transport, creating loops in the flow paths. Therefore, these clusters have to be identified and eliminated. The eliminated isolated clusters are shown in Figure 4.7b.

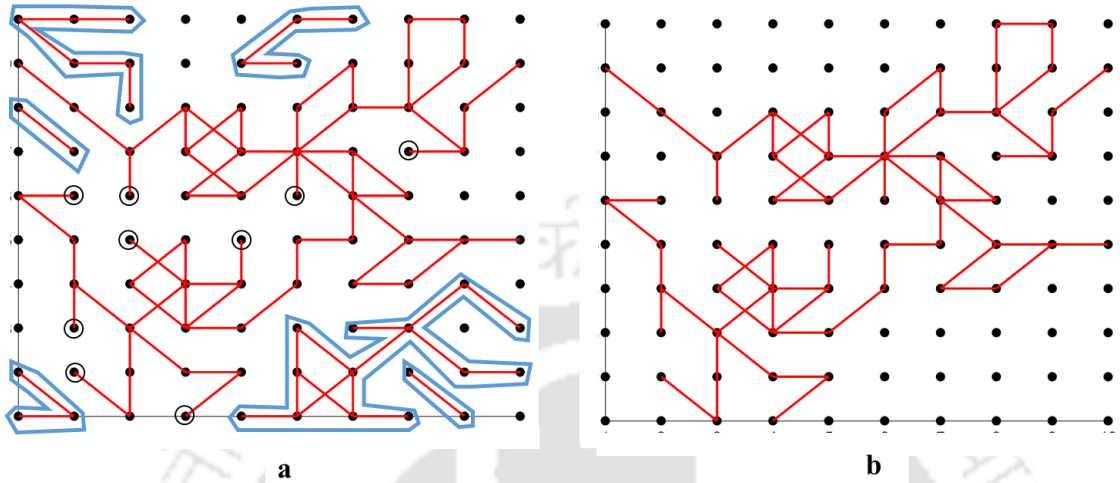

Figure 4.7: Isolated clusters and dead-end bonds (a) before elimination (b) after elimination

The isolated clusters are shown with a highlighted blue color, and the dead-end bonds are indicated in small circles of black in Figure 4.7a. The dead-end bonds are not eliminated from the pore network as they might be necessary for the solute transport processes.

The procedure to eliminate the isolated clusters is as follows. First, we need to start with the face from which the flow is allowed (boundary face); here, the face is taken as the left face. From all the sites in the left face, we need to identify the sites they are connected to the next face and so on. Finally, when we arrive at the right face (the other boundary), the set of all the sites which lead to the destination is called the backbone of the pore network. Once we identify the backbone or backbones of the pore network, we can eliminate all the other bonds as they will be the isolated clusters in the network.

4.2.2.5 The radius of the spherical bodies and cylindrical throats

The pore network is incomplete without the actual assignment of the radii to the pores and the pore connections. I fixed the average radius of the pore bodies based on the experimental evidence and the material for which the pore network model is developed. I have distributed the average radius among the network with all the bodies connected through bonds with an assumed variance. A log-normal distribution is followed for the fixation on radii distribution among the pore connections. Once the radii of the pore bodies are fixed, the radii of pore throats are to be fixed based on the number of connections of each pore body.

If any pore body has a number of connections 'n_i', it can only have its whole surface area shared at most with the total connecting pore throats (cylinders).

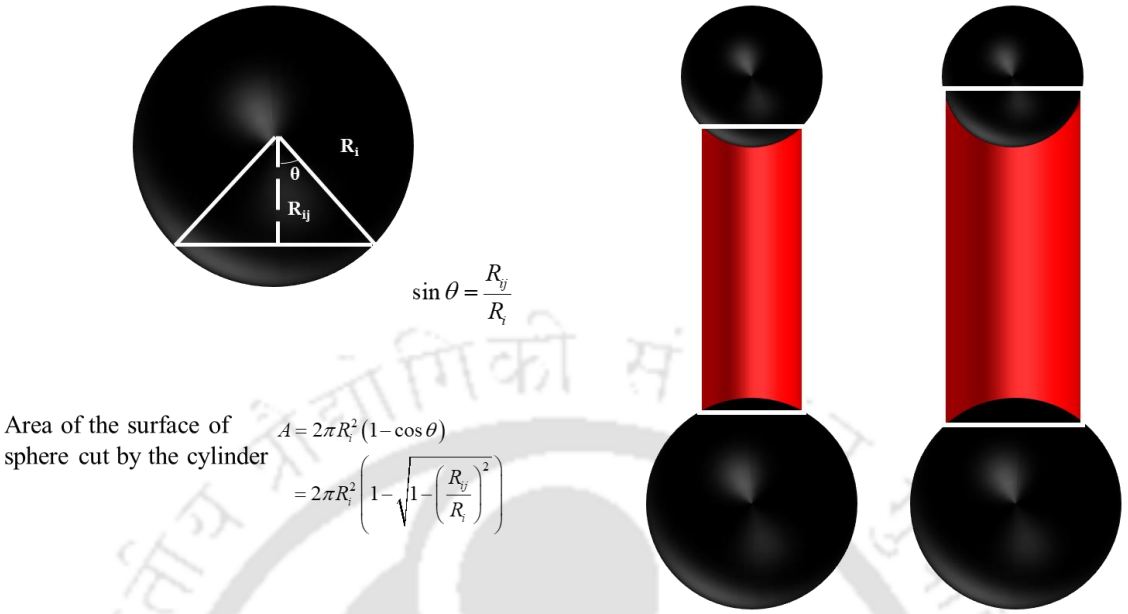

Figure 4.8: Possibility of different radii of pore throats for the same connecting pore bodies

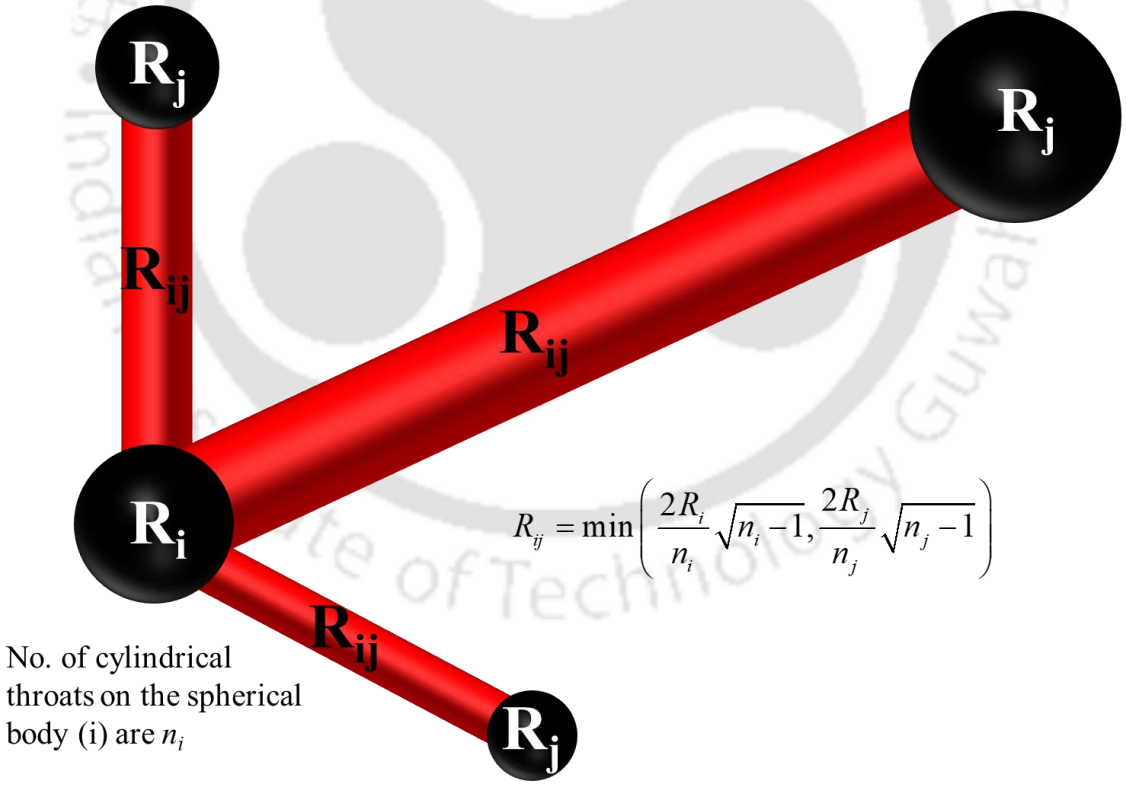

Figure 4.9: Different radii of pore throats from the same pore body

Based on this, the radius of the pore throat can be decided. A pore body '*i*' is connected to another pore body '*j*' through a connecting pore throat '*ij*'. For every pore body, the radius of the cylinders that intersect it can be calculated. For any connection '*ij*', the radius of the pore

throat (cylinder) is the minimum of the possible radius information from the two connecting bodies '*i*' and '*j*'. The calculation of the area of intersection and the possible radius for each body are given in Figure 4.8 and Figure 4.9, respectively. In Figure 4.8, it was shown that two pore bodies (spheres) of the same radii can be connected with different pore throats (cylinders) radii. Each sphere can intersect a cylinder at different surface areas, but the key is to minimize the surface area available for each cylinder to connect to the sphere. In Figure 4.9, the left connection covers a smaller surface of each sphere, but the right-side connection covers a more extensive surface area, allowing the cylinder's radius to be higher. If a pore body has a more significant number of connections, then the radius of each connection has to be smaller. The possibility of a different radius of pore throats from the same body is shown clearly in Figure 4.9.

4.2.3 Estimation of macroscopic parameters

A pore network model is formulated to do the microscopic analysis on the pore scale. However, for the field application of the results, macroscopic analysis of the pore network properties has to be formulated. Bear and Bachmat (Bear and Bachmat 1990) have formulated the conversion of microscopic parameters to macroscopic properties through some parameters. These parameters are porosity (θ_α), tortuosity (T^*_{ij}), permeability (k_{ij}), dispersivity (α_{ijkl}), and fractional surface area (θ^s_α). All these parameters can be estimated as per their definitions given by Bear and Bachmat (Bear and Bachmat 1990).

4.2.3.1 Porosity (θ_α)

The porosity is defined by the ratio of the volume of void spaces to the volume of total soil. For a cuboidal REV, the porosity is given by equation 4.3.

$$\theta_\alpha = \frac{\sum_{m=1}^{N_x} A_{xm} b_{xm} + \sum_{m=1}^{N_y} A_{ym} b_{ym} + \sum_{m=1}^{N_z} A_{zm} b_{zm}}{L_x L_y L_z} \qquad 4.3$$

Here, the number of spherical bodies intersecting opposite faces is given as N_x, N_y, and N_z. The areas of such bodies and the total length of their channels connecting the other face are multiplied to give the volume of the void spaces occupied by the pore bodies and pore throats. Lx x Ly x Lz gives the volume of the REV in total.

4.2.3.2 Fractional Surface Area ($\theta_{s\alpha}$)

The fractional surface area is the ratio of the void spaces' surface area to the REV's total surface area. It is given by equation 4.4.

$$\theta_{\alpha}^{S} = \frac{\sum_{m=1}^{N_x} A_{xm} + \sum_{m=1}^{N_y} A_{ym} + \sum_{m=1}^{N_z} A_{zm}}{L_x L_y + L_y L_z + L_x L_z} \qquad 4.4$$

Here, as the areas are only counted once while intersecting the pair of opposite faces, the denominator contains only half of the surface area of the REV.

4.2.3.3 Tortuosity (T^*_{ij})

The tortuosity of a REV is defined as the total static moment of the oriented elementary surfaces comprising the Solid-fluid surface concerning the planes passing through the centroid of the REV per unit volume of the fluid phase within REV. The equation describing tortuosity is given in equation 4.5.

$$T_{ij}^{*} = \frac{1}{U_{0f}} \int_{S_f} \nu_i x'_j = \frac{\theta_{\alpha}^{S}}{\theta_{\alpha}} \frac{1}{S_{0f}} \int_{S_f} \nu_i \nu_j \qquad 4.5$$

In equation 4.5, the term U_{0f} represents the volume of the REV occupied by the fluid; the term S_{0f} represents the surface area of the fluid inside the REV, the term ν_i represents the normal unit vector outward to the fluid phase, the term x' is the fluctuation of the point x (any point space within REV) for the centroid of REV. Therefore, the ratio $\theta^s_{\alpha}/\ \theta_{\alpha}$ represents the total factor of the tortuous paths inside the REV, and the rest represents the anisotropy of the REV. If one assumes the isotropic behavior of the fluid, the anisotropy term can be reduced to the unit tensor of the second rank. It gives the tortuosity as given in equation 4.6.

$$T_{ij}^{*} = \frac{\theta_{\alpha}^{S}}{\theta_{\alpha}} \delta_{ij} \qquad 4.6$$

Here, δ_{ij} is the Kronecker delta which gives a value of one when i = j and zero otherwise.

4.2.3.4 Permeability (k_{ij}) and Hydraulic Conductivity (K_{ij})

The porous medium's permeability depends upon the media's tortuosity and the anisotropy factor over the solid-fluid interface (a_{ij}). It also depends upon the characteristic distance from the walls of the solid to the interior of the fluid phase (Δ_c) and the macroscopic dimensionless

shape factor (C_f). The equation defining the permeability of the porous media is given by equation 4.7. All the supporting equations are given as a set in equation 4.8.

$$k_{ij} = \frac{\theta_\alpha \Delta_f^2}{C_f}\left(a_{ij}\right)^{-1} T_{ij}^* \qquad 4.7$$

$$a_{ij} = \delta_{ij} - \overline{(v_{f_i} v_{f_j})}^{f-s}$$
$$C_f = \frac{\Delta_f}{\Delta_c}, \Delta_f = \frac{\theta_\alpha U_o}{S_{fs}} \qquad 4.8$$

However, for the isotropic condition, the equation for the intrinsic permeability reduces to a more straightforward form as given in equation 4.9.

$$k_{ij} = \frac{3}{2}\frac{U_0}{S_0}\theta_\alpha \Delta_c \delta_{ij} \qquad 4.9$$

The unit for the intrinsic permeability (equation 4.9) is m^2 in SI units. From this, the hydraulic conductivity can be worked out by using equation 4.10.

$$K_{ij} = k_{ij}\frac{\overline{\rho_f} g}{\overline{\mu}} \qquad 4.10$$

Here, the term ρ_f is the density of the fluid phase, the term μ is the fluid's dynamic viscosity, and the term g is the acceleration due to gravity.

4.2.3.5 Dispersivity (α_{ijkl})

The porous medium's dispersivity is the most difficult to find experimentally, as it is a fourth-rank tensor with 81 components in 3 dimensions. The dispersivity transforms the fluctuations of the velocities inside the porous medium into a measurable parameter in the macroscopic analysis of the flow through the porous medium. To define dispersivity, a new term has to be defined: a second-order tensor that transforms the average velocity into a local velocity. The tensor is given by the symbol T_{ij} and is given in equation 4.11.

$$T_{ij} = \beta^* \frac{dx_i}{ds'}\frac{dx_j}{ds'} \qquad 4.11$$

Here, β^* is the proportionality constant, and *s'* is the direction of the point in porous media to be transformed (local direction). The term x_i is the direction of the global axes. As the property of this tensor is to transform the global average to the local scale of velocity, the average value of this tensor gives the Kronecker delta ($T_{ij}^{avg} = \delta_{ij}$). From this relation, the constant of

proportionality β^* can be found. Therefore, the dispersivity can be given from these terms as given in equation 4.12.

$$\alpha_{ijkl} = l^E \overline{\overset{o}{T}_{ij} \overset{o}{T}_{mk}}^{f} T^*_{ml} \tag{4.12}$$

The symbol ^{o} represents the fluctuating component of the term (i.e., the actual term subtracted by its average). The term l^E is the random statistical length that represents the porous media. Xu and Tartakovsky have explained that the statistical length is a constant multiplied by the average length of the pore throats and doesn't exceed a value of 0.25 (Xu and Tartakovsky 2017). As it is difficult to deal with 81 terms at a time, the isotropic case has to be considered to reduce the full terms to only two. These two terms can be given by α_L and α_T and are explained in equation 4.13. The transverse dispersivity is only given for the values where *i* is not equal to *j*.

$$\alpha_L = \frac{\theta^S_\alpha}{\theta_\alpha} l^E \overline{\overset{o}{T}_{ii} \overset{o}{T}_{ii}}^{f}$$
$$\alpha_T = \frac{\theta^S_\alpha}{\theta_\alpha} l^E \overline{\overset{o}{T}_{ij} \overset{o}{T}_{ij}}^{f} \tag{4.13}$$

All these terms are needed to convert the pore scale parameters to the Darcy scale parameters. The results of a pore network model generated using all the given methodology are given in the subsequent sections.

4.2.4 Methodology of Optimization Model

The shuffled frog leaping algorithm works on the principle of memetic evolution. The algorithm is a metaheuristic algorithm that is used as an optimization algorithm for solving complex problems (Eusuff et al. 2006). The desired macroscopic parameters are given in Table 4.1.

Table 4.1: Desired values of macroscopic parameters

Macroscopic Parameter	**Value**
Porosity (θ_α)	0.35
Hydraulic conductivity (K_{ij})	9.0 x 10^{-5} δ_{ij} m/s
Longitudinal dispersivity (α_L)	2 m

The methodology for this optimization problem is discussed as given in Figure 4.10. The decision variables are the mean and variance of the radii of the pore bodies of two categories (four variables) and the percentage division of the two categories (one variable), comprising a total of five variables. The methodology is simple, where the output function is generated from the macroscopic value calculation from the pore network model.

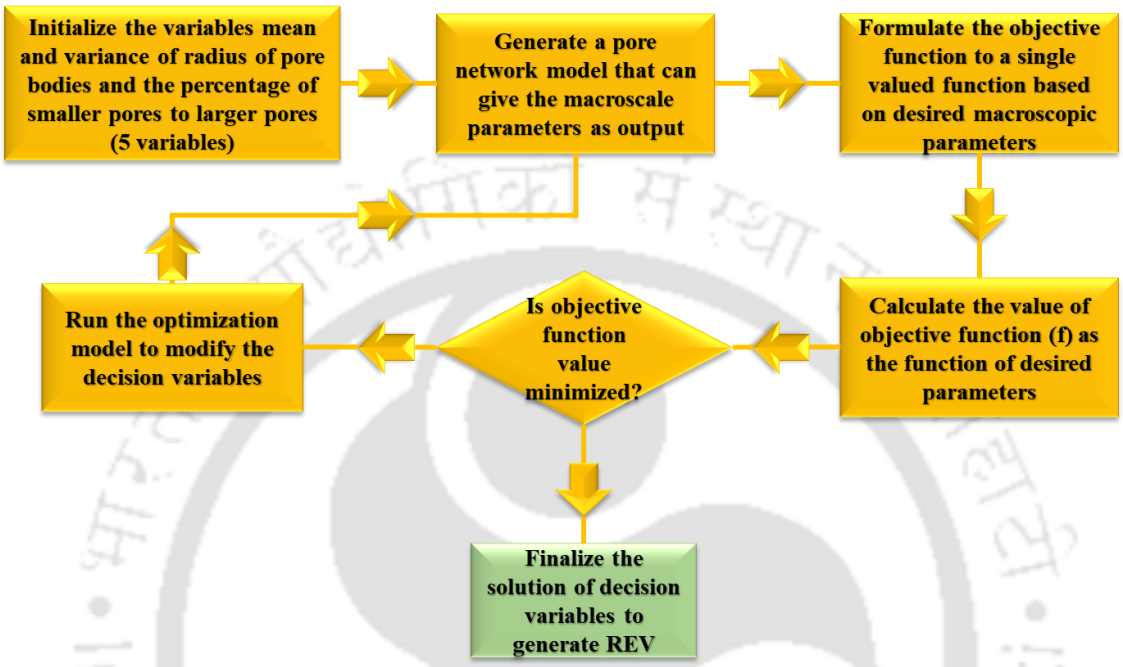

Figure 4.10: Methodology of the optimization model

The objective function is generated as a normalized error function of all the parameters generated from the pore network model as given in equation 4.14.

$$f = \frac{abs\left(\theta_{\alpha}^{gen} - \theta_{\alpha}\right)}{\theta_{\alpha}} + \frac{abs\left(K_{ij}^{gen} - K_{ij}\right)}{K_{ij}} + \frac{abs\left(\alpha_{L}^{gen} - \alpha_{L}\right)}{\alpha_{L}} \qquad 4.14$$

Here, the superscript ^{gen} represents the generated value, and the rest represents the desired value of the parameters. The parameters can be added or deleted based on the user's requirement and reduce the optimization model's complexity.

After calculating the objective function, the termination criteria is checked, and then the optimization algorithm of the shuffled frog leaping algorithm is called to modify the values of the decision variables. Therefore, the final solution contains the objective function's minimum value, which gives the best decision variables values.

4.3 Results and Discussion

The Pore Network Model (PNM) is created using the given methodology for the Number of pore bonds in the three orthogonal directions *i*, *j*, and *k* as N_i, N_j, and N_k are equal to 20. The length, breadth, and height of the cuboidal PNM are equal to 1 mm. The targeted mean coordination number was 5 – 5.5 for the sand as the porous material. After following the total procedure up to the elimination of the isolated clusters, the results are shown in Table 4.2. The PNM as generated is shown in a figurative representation in Figure 4.11.

Table 4.2: Characteristics of the PNM as generated by the model

Characteristic	**Generated by model**
No. of pore bodies	7886
No. of pore throats	21018
Mean Coordination number	5.25
Porosity	0.3045

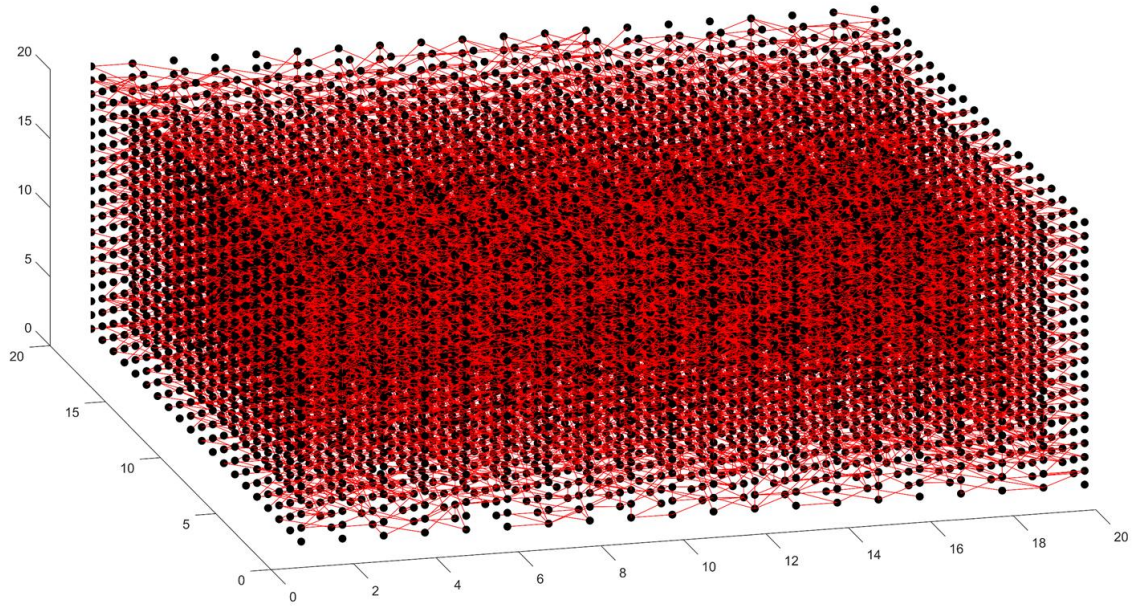

Figure 4.11: Pore Network Model generated for the given characteristics

4.3.1 The radius of pore bodies and pore throats

The radius of the pore bodies is divided into two categories of mean and variance for the log-normal distribution. The first category has an occurrence probability of 95 % and the second category has an occurrence probability of 5 % which are given in Table 4.3.

Table 4.3: Parameters for the radius of the pore bodies' distribution (log-normal)

Category	Percentage of pore bodies	Mean	Variance
Category 1	95 %	0.0035 mm	0.0005 mm^2
Category 2	5 %	0.05 mm	0.005 mm^2

This categorization allows for the possibility of having some macro pores in the PNM, which is more natural to occur in field conditions. The distribution is given in a histogram for each category in Figure 4.12 and Figure 4.13, respectively.

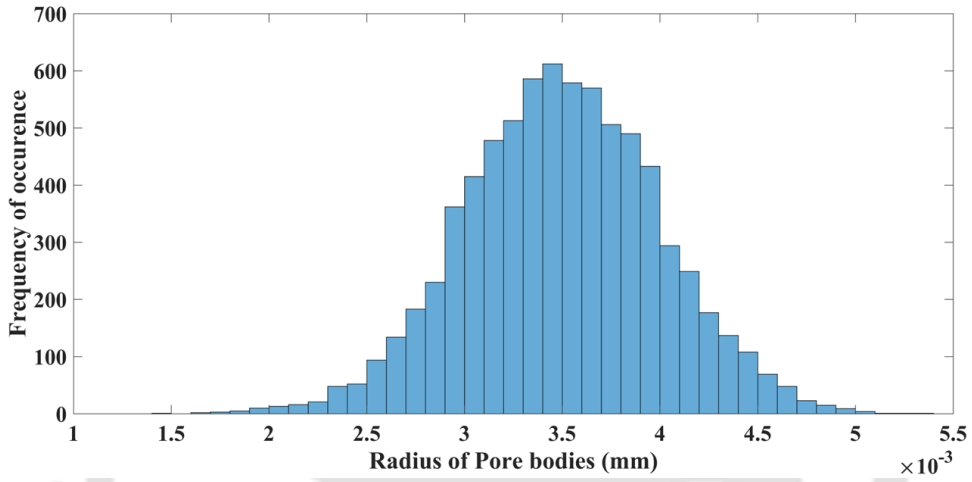

Figure 4.12: Distribution of radius of the pore bodies for category 1

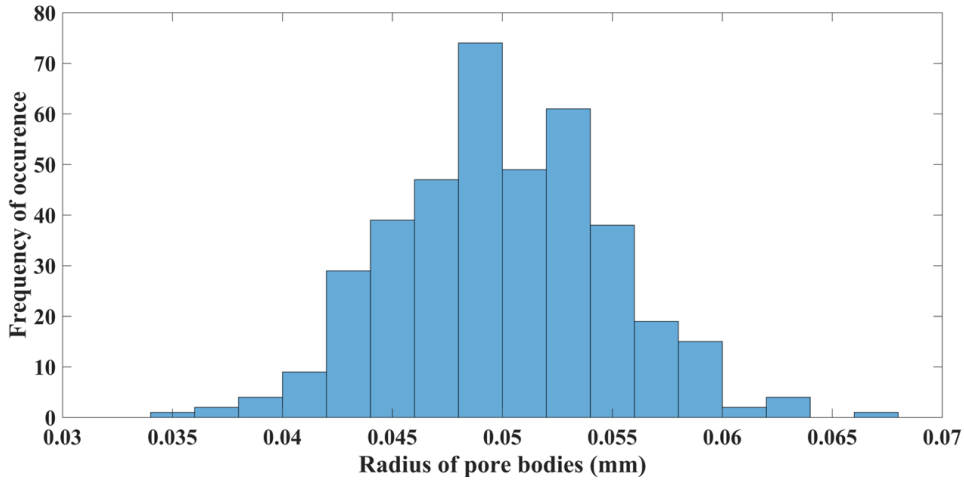

Figure 4.13: Distribution of radius of the pore bodies for category 2

The distributions of the radii of the pore throats are much more complicated, as the combinations of categories are different. If the pore body of category 1 meets the pore body of category 2, the pore body of category 1 dominates the radius of the pore throat. But if the pore body of category 2 meets the pore body of category 2, then only the pore throats would be of the same range of radius of category 2. Therefore, category 2, when it meets itself, doesn't fit

a perfect normal distribution as some of the radii of pore bodies get distributed in the first category combination. These distributions of pore throats are given in Figure 4.14 and Figure 4.15.

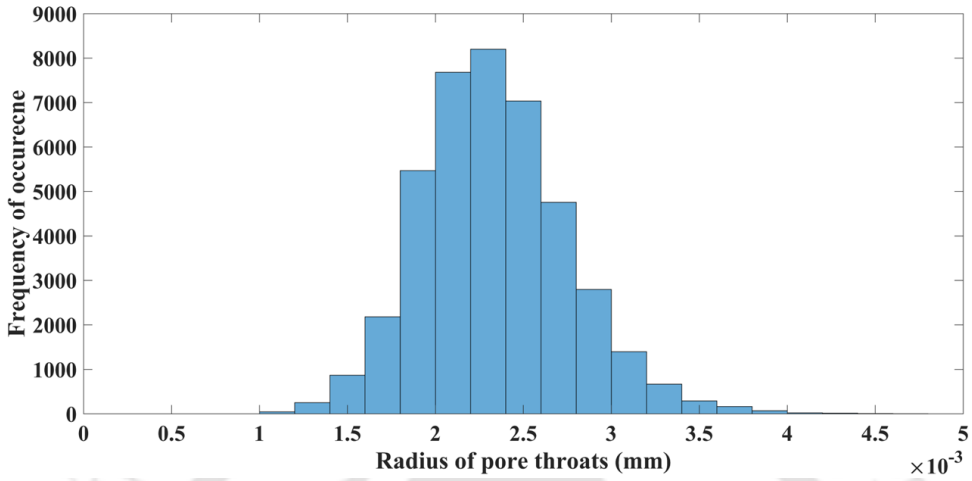

Figure 4.14: Distribution of radii of pore throats for combinations of categories 1,1; 1,2 and 2,1

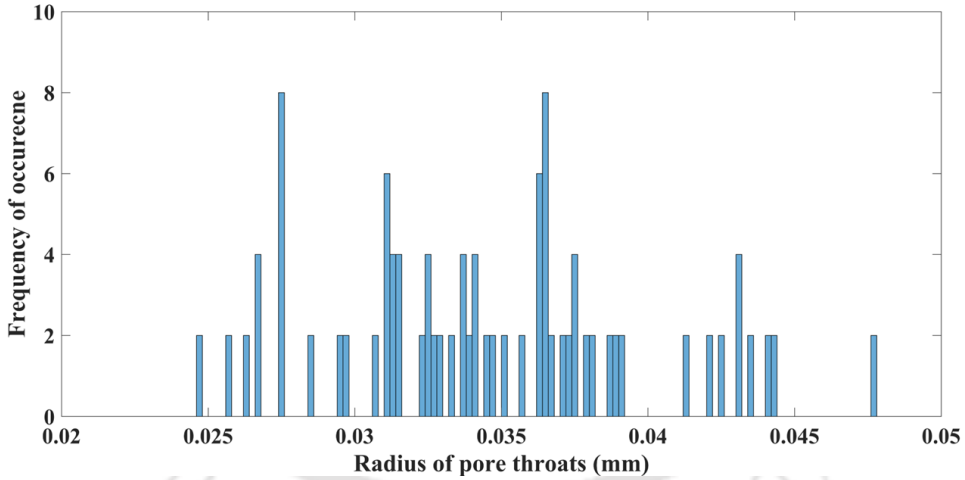

Figure 4.15: Distribution of radii of pore throats for combinations of categories 2, 2

4.3.2 Macroscopic properties of generated PNM

The generated PNM is a microscopic representation of a single REV, which is treated as a point in the Darcy scale modeling. Therefore, the properties of the PNM should be convertible to a larger scale through the parameters that were defined in the methodology chapter. All the properties were assumed to be isotropic for the sake of simplicity. These parameters are listed in Table 4.4. These parameters are used as the modeling parameters while modeling the macroscopic equation of the advective – dispersive equation. The parameters are only dependent on the choice of the number of pore bodies in each direction and the radius of pore

bodies. The rest are all generated by the model. Therefore, the model can generate the desired value of the parameters with some uncertainty and can be applied to the field conditions for heterogeneous porous media where the parameters are not constant over large distances.

Table 4.4: Macroscopic parameters of the PNM

Macroscopic parameter	**Value as generated by PNM**
Porosity (θ)	0.3045
Fractional surface area (θ^S)	0.7212
Tortuosity (T_{ij}^*)	2.3687 δ_{ij}
Permeability (k_{ij})	1.8508 x 10^{-4} δ_{ij} mm^2
Hydraulic conductivity (K_{ij})	0.002 δ_{ij} mm/day
Longitudinal dispersivity (α_L)	0.757 mm
Transverse dispersivity (α_T)	0.519 mm

4.3.3 Optimization Model

The optimization algorithm for the model is the shuffled frog leaping algorithm. The desired results from the optimization model are given in the Table 4.5 and Table 4.6. The minimization of the objective function with iterations is given in the following Figure 4.16.

Table 4.5: Results from the optimization model

Macroscopic Parameter	**Desired Value**	**Estimated Value**
Porosity (θ_α)	0.35	0.3346
Hydraulic conductivity (K_{ij})	9.0 x 10^{-5} δ_{ij} m/s	10.05 x 10^{-5} δ_{ij} m/s
Longitudinal dispersivity (α_L)	2 m	1.7944 m

Table 4.6: Best solution of the optimization model

Variables	**Best value**
Mean Radius (Category 1)	0.00017 m
Variance of Radius (Category 1)	0.00019 m^2
Mean Radius (Category 2)	0.0805 m
Variance of Radius (Category 2)	0.0255 m^2
Percentage of category 1 to total	98.72 %

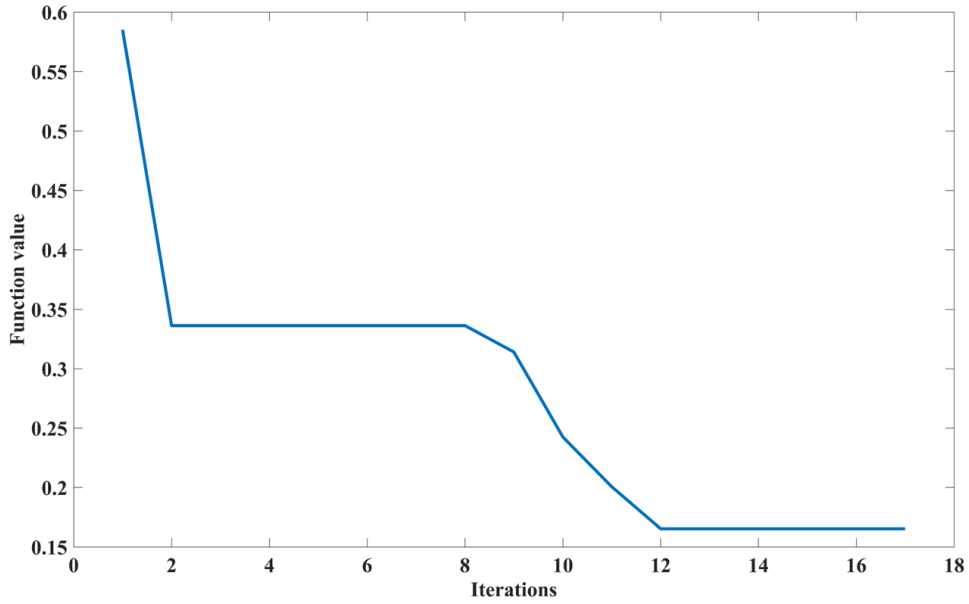

Figure 4.16: Objective function variation with iterations of the optimization algorithm

As the pore network model generation is a randomized algorithm, it is not reasonably possible to exactly generate the results of the desired parameters from the optimization algorithm. However, the relative values of the parameters are enough for the further study the pore network model for the next objective.

4.4 Conclusions

The Pore Network Modelling is generally a well-established type of modeling, but is avoided due to the complications arising from the increase in computational effort. However, the model is almost an accurate representation of the pore scale characterization of the porous media and is very useful for modeling the microscale effects. The model is instrumental in the transport phenomenon of viruses and small colloids as they are very tiny compared to the pore scale. The effects of such tiny particles are more significant on a smaller scale, giving much more accurate predictions for the Darcy scale parameters. Besides, the pore network properties that estimate the Darcy scale geometrical parameters are the key to conceptualizing the representative elementary volume (REV).

CHAPTER 5: ESTIMATION OF DARCY SCALE PARAMETERS FROM PORE PROPERTIES

5.1 Introduction

This chapter aims to estimate the pore-scale parameters for the management model that can account for different pore properties and solution chemistry parameters. As the study aims to the generation of a management model, it should be efficient in removing the virus particles from the porous media effectively. The removal of virus particles is possible only through the inactivation of viruses or through the attachment of virus particles to a solid surface. There has been pervasive research on virus inactivation rates as an effect of solution and the soil parameters (Keswick and Gerba 1980; Sim and Chrysikopoulos 1996; Yahya et al. 1993; Yates et al. 1985; Yates and Yates 1988). However, the researchers haven't focused on the part where the viruses may be attached to the soil surface for different reasons. The rates of virus removal from the aqueous phase might be affected by such deposition.

Virus transport is distinguished from solute transport because viruses undergo different inactivation and adsorption mechanisms (Ratha et al., 2009). The transport of virus in porous media is a complex process that depends on factors such as the nature of the cell, including size, hydrophobicity, motility, and growth phase (Fang and Logan 1999). The transport of virus also depends on medium characteristics such as soil type, grain size, organic matter content (Bales et al. 1989), and water chemistry factors such as pH and ionic strength (Bales et al. 1991). As it is known, the inactivation may happen in the aqueous phase of the viruses (i.e., in the pore space) or the adsorbed phase (i.e., attached to the soil surface). Experimental observations suggest that the inactivation rate is lower for the attached phase than for liquid phase viruses (Gerba 2004; Hurst et al. 1980). The attachment and detachment rates of the virus to the soil surface depend on many factors (Hassanizadeh et al. 2000). In this area, researchers have extensively studied the effect of various soil and aqueous parameters on the deposition rates towards the soil surface (Elimelech and O'Melia 1990; Loveland et al. 1996; Seetha et al. 2014, 2015). The most recent research done in the year 2015 (Seetha et al. 2015) has identified the correlation parameters of the attachment and detachment rates of viruses as a function of nine pore scale parameters: the pore radius, radius of the virus, mean flow velocity, ionic strength of the solution, viscosity, temperature, solution dielectric constant, and the surface potentials of the virus and the soil surface. The researchers have derived the correlation

equations by assuming a nanoparticle transport in a single pore and upscaled the results to a 1-D advective-diffusive equation (Raoof et al. 2010a; Seetha et al. 2014, 2015). The equations resulting from a total of 1350 pore network models, considering different scenarios of the virus deposition to the soil surface (Seetha et al. 2015), are presented. The work done in this chapter is based on the exact correlation equations and the application of such equations in developing an effective management strategy.

I have used the correlation equations to develop the detachment and attachment rates under kinetic sorption and the distribution coefficient under equilibrium sorption. I have used Hagen-Poiseuille equations to derive the mean velocity of a single pore (Raoof et al. 2010a). The developed sorption, attachment, and detachment rates are averaged over the REV using the surface average method as the viruses get deposited on the surface of the pore throat or the pore body. I have then used these averaged values of deposition rates to estimate the Darcy-scaled parameters. I have used the estimated Darcy-scaled parameters in the subsequent chapters.

5.2 Methodology

This chapter aims to estimate parameters in Darcy-scale that can be used to reduce the concentration of viruses in the aquifer with the effect of changing parameters throughout the aquifer. I have taken the correlation equations from research published in 2015 (Seetha et al. 2015) which corresponds to a single pore. These correlation equations account for nine pore parameters. Two parameters correspond to the virus, and seven correspond to the pore medium and geometry. The virus parameters are the radius of the virus particle (a) and the surface potential of the virus (ψ_1). The pore parameters are the permittivity of the pore space (ε), the temperature of the medium (T), the viscosity of the medium (μ), the surface potential of the collector (ψ_2), solution ionic strength (I), the mean velocity of the fluid in the pore (v_{mij}) and the radius of the pore throat (R_{ij}). The correlation equations give three values such as the distribution coefficient for equilibrium sorption (K_D) and the attachment coefficient (k_{att}) and detachment coefficient (k_{det}) for kinetic sorption. These parameters are the critical factors for the variation of virus parameters in terms of sorption.

The inactivation terms are also a function of these parameters and are variable with the pore scale parameters. I have converted these pore parameters to the averaged values in Darcy's scale using the averaging techniques. Then, we studied the effect of changing each parameter in the Darcy scale and produced the results. The overview of the method used in this chapter is in Figure 5.1.

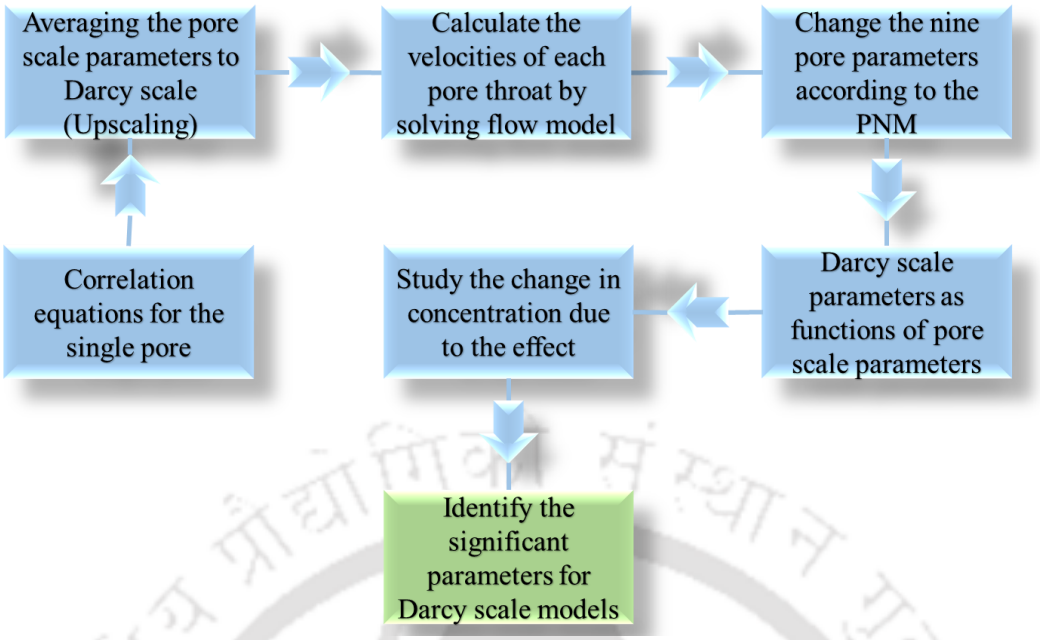

Figure 5.1: Overview of the methodology

5.2.1 Correlation equations for a single pore

The correlation equations for pore averaged deposition rate coefficients proposed by the researchers (Seetha et al. 2015) depend on five non-dimensional numbers (NE_1, N_{DL}, A_{ij}, Pe_{ij} and λ^*) as given in equation 5.1. The deposition rate coefficients using three non-dimensional numbers ($K_D^|$, $Da_{att_{ij}}$ and $Da_{\det_{ij}}$) as given in equation 5.2.

$$NE_1 = \frac{\pi\varepsilon\varepsilon_0 a\left(\psi_1^2+\psi_2^2\right)}{k_B T},\quad N_{DL} = \kappa a,\ \kappa = \sqrt{\frac{2000 N_A Ie^2}{\varepsilon\varepsilon_0 k_B T}},\quad A_{ij} = a/R_{ij}$$
$$Pe_{ij} = \frac{v_m^{ij} R_{ij}}{D_\infty},\ D_\infty = \frac{k_B T}{6\pi\mu a},\ \lambda^* = \frac{\lambda}{a} \qquad 5.1$$

$$K_D^| = \frac{K_D}{R_{ij}},\ Da_{att_{ij}} = \frac{k_{att} R_{ij}}{v_{m_{ij}}},\ Da_{\det_{ij}} = \frac{k_{\det} R_{ij}}{v_{m_{ij}}} \qquad 5.2$$

These parameters are all related to the proposed correlation equations, as shown in equation 5.3.

$$\begin{aligned} K_{D_{ij}}^| &= p_1 NE_1^{p_2} N_{DL}^{p_3} Pe_{ij}^{p4} A_{ij}^{p5} \lambda^{*p_6} \\ Da_{att_{ij}} &= q_1 NE_1^{q_2} N_{DL}^{q_3} Pe_{ij}^{q4} A_{ij}^{q5} \lambda^{*q_6} \\ Da_{\det_{ij}} &= r_1 NE_1^{r_2} N_{DL}^{r_3} Pe_{ij}^{r4} \exp\left(A_{ij} r_5\right) \lambda^{*r_6} \end{aligned} \qquad 5.3$$

The researchers gave the coefficients of these correlation equations for different ranges of the five non-dimensional parameters (Seetha et al. 2015). These correlation equations are used in this research to find the Darcy scale parameters as a function of the pore scale parameters.

5.2.2 Estimation of mean velocity in each pore

The mean velocity in each pore is calculated based on the Hagen-Poiseullie flow equation for simple pressure relations, which is in equation 5.4. I have taken the boundaries of the pore network model to exist at two different pressures in single directional boundaries. Based on those boundary conditions, I have calculated the pressures in all the other pore bodies. I have combined the continuity equation and the Hagen-Poiseullie equation to calculate the pressures in the pore bodies. The idea is that the total inflow into a pore body through all the connecting pore throats equals the total outflow (Raoof et al. 2010a). The Hagen-Poiseullie equation connects the discharges through the pore throat to the pressures at the corresponding pore bodies. Therefore, I have calculated the pressures at all the pore bodies by solving the discharge equations of continuity (equation 5.5). After solving for the pressures, I calculated the mean velocity according to equation 5.5.

$$q_{ij} = \frac{\pi R_{ij}^4}{8\mu l_{ij}}\left(p_j - p_i\right) \qquad 5.4$$

$$\sum_{j=1}^{z_i} q_{ij} = 0,\ v_m^{ij} = \frac{q_{ij}}{\pi R_{ij}^2} \qquad 5.5$$

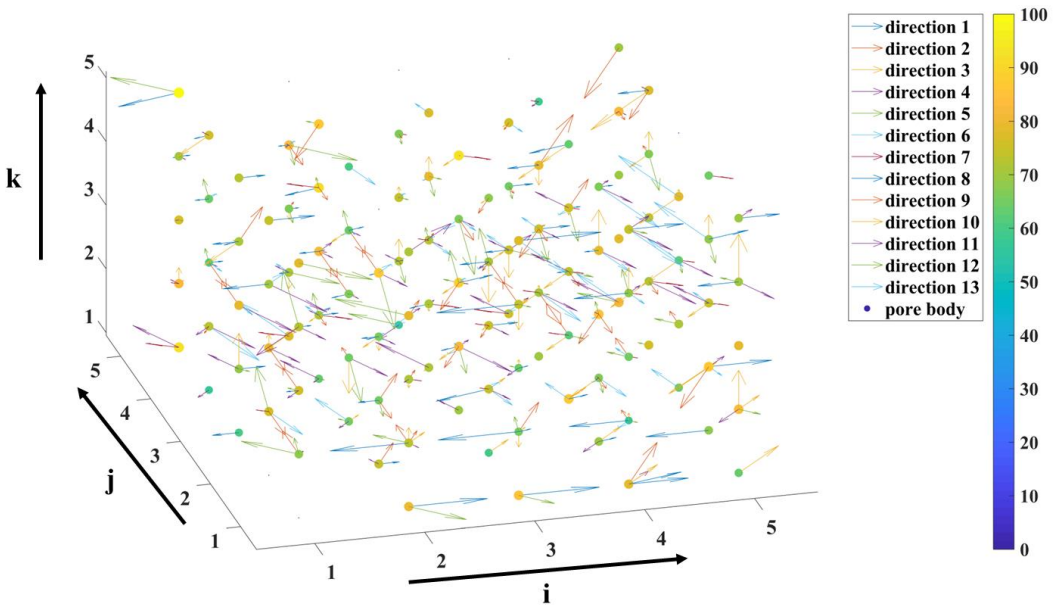

Figure 5.2: Distribution of velocity in the pore throats

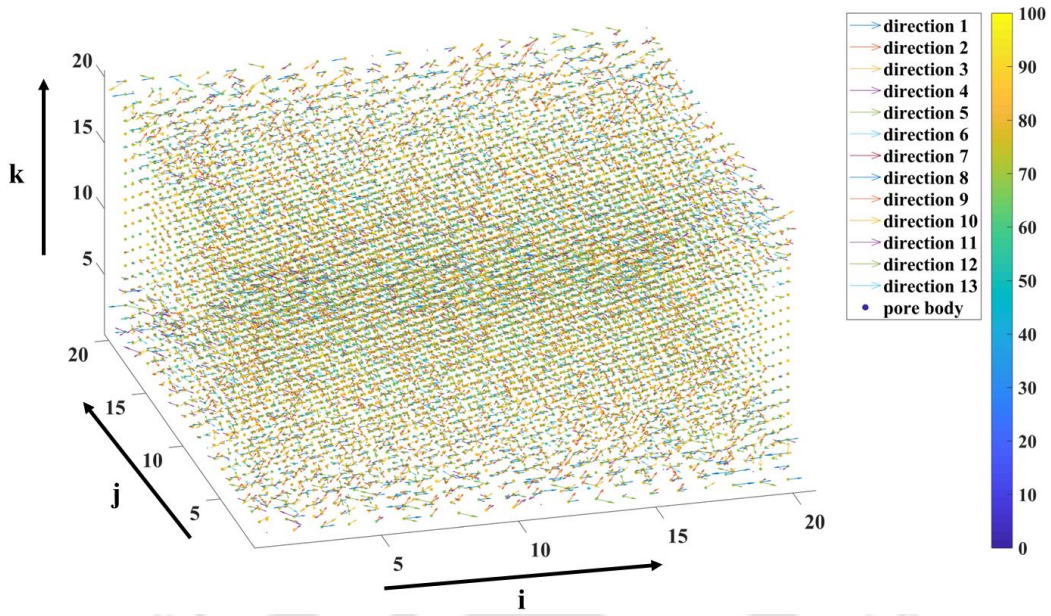

Figure 5.3: Distribution of velocity in pore throats for the full REV

The calculated velocity distribution in the REV of a porous sample is in Figure 5.2. I have given an arrowhead to show the velocity direction from each pore body. It was described in the previous chapter that several pore throats could connect from a single pore body. Therefore, Figure 5.2 shows the arrows emerging in different directions from each pore body. Figure 5.2 also shows the sizes of pore bodies represented as the size and color map. I have shown the representation of the same velocity distribution for the whole REV in Figure 5.3.

5.2.3 Generation of a simple pore network model

As discussed in the previous chapter, I generated a pore network model (PNM) to estimate the pore-scale parameters effectively. The generated model is stable for several runs, giving the same range of values of porosity and intrinsic permeability. The desired value of porosity is 0.35, and the desired value of intrinsic permeability is 8.165×10^{-6} mm^2. I have used the same parameters in the first objective of this study. The pore network model has a tiny volume of 0.1 x 0.1 x 0.1 mm as a REV. The number of nodes in each direction are 10 for the generated PNM. I have studied the variation of correlation parameters of N_{DL}, NE_1, Pe, A, and λ^* over the averaged values for this PNM to generate the results.

5.3 Results and Discussion

The pore network model (PNM) has one variable corresponding to the mean of the radii of the pore bodies, one variable corresponding to the variance of the radii, and one variable for the mean coordination number. I have considered sand as the aquifer material. The mean radius of

the pore throats for sandy soil is 1µm to 300 µm (Minagawa et al. 2008). I have taken the variance from 0.01 µm to 0.01mm. The mean coordination number ranges from 4 to 8 for the optimization model. The shuffled frog leaping algorithm (SFLA) is used, as discussed in chapter 4. The optimum function value is 0.0016 for the objective function, with the solutions of the mean radius of 0.0018 mm, the variance of 9.13 x 10^{-4} mm, and the mean coordination number of 7.2097. I have used the generated PNM to check for the porosity and intrinsic permeability stability by running the model 1000 times. I have shown the results of the 1000 runs as box plots in Figure 5.4.

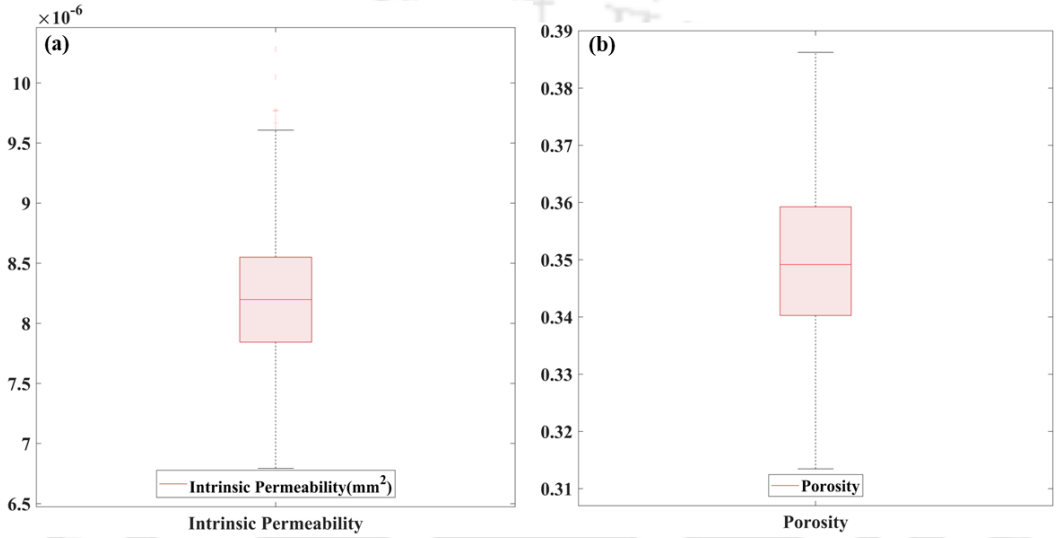

Figure 5.4: Box plots for 1000 runs of PNM for (a) intrinsic permeability and (b) porosity.

The porosity values fluctuate from 0.34 to 0.36 in the total 1000 runs, with the mean value at 0.35. The intrinsic permeability values fluctuate from 7.8 x 10^{-6} to 8.6 x 10^{-6} mm^2 which are reasonable value ranges for the specified problem.

5.3.1 Sensitivities of the parameters as a result of correlation equations over an REV

The parameters corresponding to the correlation equations are related through five non-dimensional numbers. The effect of each non-dimensional number is studied with the specified range of parameters suggested by previous research (Seetha et al. 2014, 2015) in this section. While studying the variation of one non-dimensional number, the rest of the numbers are kept constant. The constant values of non-dimensional numbers are as such NE_1 = 40, N_{DL} = 10, Pe = 20, A = 0.001and λ^* = 1.

5.3.1.1 Sensitivity of NE_1

The range of values for the number NE_1 is 1 to 400. I studied the sensitivities of this parameter on the attachment and the detachment coefficients and produced the results in Figure 5.5. The

attachment coefficients ranged from 0.0703 to 3.0095 d^{-1} and settled at a range of 0.1036 to 0.08 over the value of NE_1 greater than 40. The detachment coefficient ranged from 21.0372 to 714.7833 d^{-1} and settled at 373 to 380 d^{-1} over the value of NE_1 greater than 40. The parameters that affect the values of NE_1 are the dielectric constant of water (ε), the radius of the virus particle (a), the surface potentials of the virus, and the soil surface (ψ_1, ψ_2), and the temperature of the medium (T). The major changing factor in this non-dimensional number is the temperature of the medium. The next major ones are the surface potentials of the virus and the solid surface.

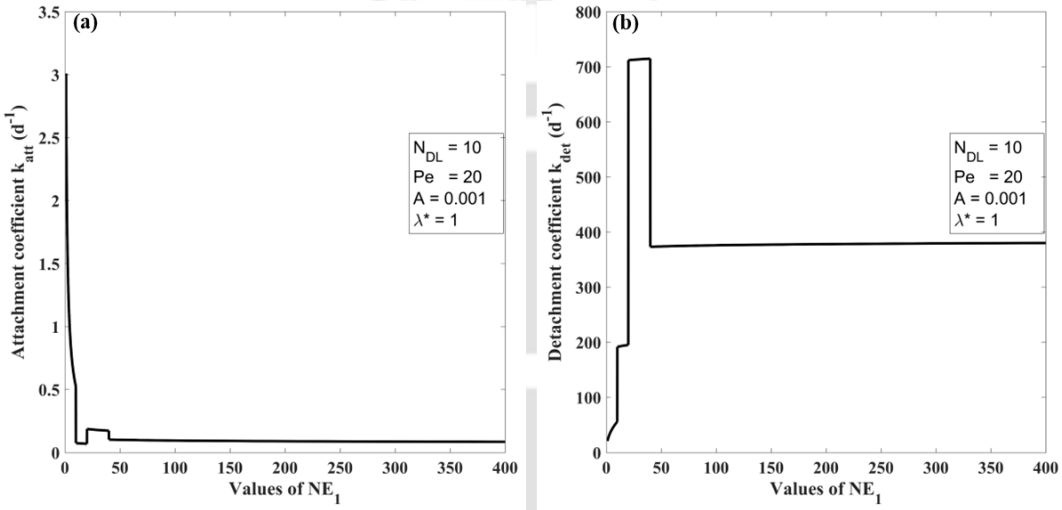

Figure 5.5: Sensitivities of NE_1 on (a) attachment and (b) detachment coefficient

5.3.1.2 Sensitivity of N_{DL}

The range of values for the number N_{DL} is 5.5 to 100. I studied the sensitivities of this parameter on the attachment and the detachment coefficients and produced the results in Figure 5.6. A clear increasing trend is found in the attachment coefficient case. However, the range of the values for the attachment coefficient is 0.0792 to 0.2912 d^{-1} which is a tiny range to consider. Similarly, a decreasing trend can be found in the detachment coefficient case. The maximum and the minimum values for the detachment coefficient are 381.0166 d^{-1} and 345.2631 d^{-1}, respectively. The non-dimensional number N_{DL} depends on the dielectric constant of water (ε), the radius of the virus particle (a), the Ionic strength of the solution (I), and the temperature of the medium (T). Here, the major contributing factor is the ionic strength of the medium, and with equal weightage is the temperature of the medium.

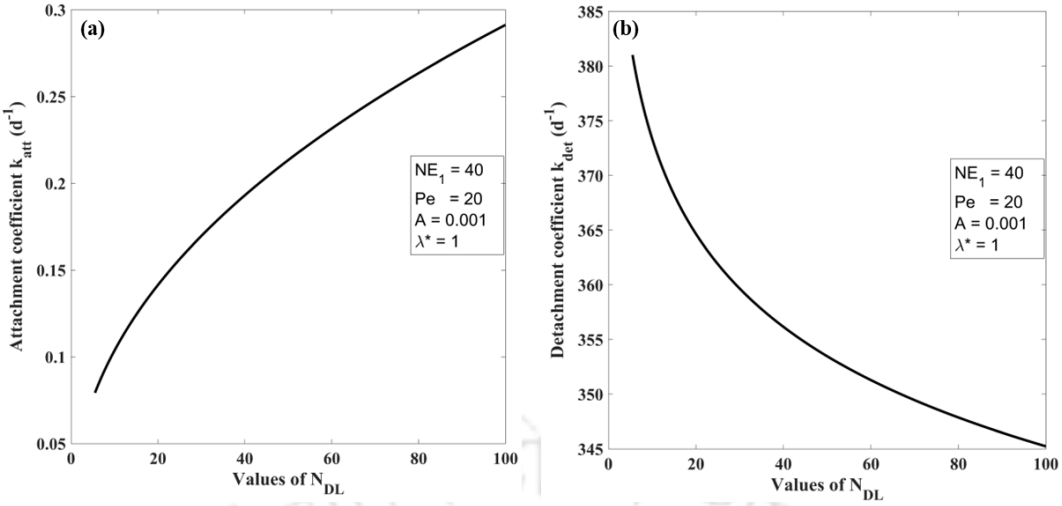

Figure 5.6: Sensitivities of N_{DL} on (a) attachment and (b) detachment coefficient

5.3.1.3 Sensitivity of Pe

The range for the variation of Pe is 0.05 to 50. I studied the sensitivities of this parameter on the attachment and the detachment coefficients and produced the results in Figure 5.7. The Peclet number (Pe) depends upon the velocity of the medium (v_m), the radius of the pore (R), the viscosity of the medium (μ), the radius of the virus (a), and temperature of the medium (T). The range of values for the attachment coefficients is 0.0326 to 200 d^{-1}. The range of values for the detachment coefficient is 163.07 to 84524 d^{-1}. However, the trend is exponentially decreasing, with most values arriving to a minimum at the Peclet number value of 5. The major contributing factor here is the temperature and the mean velocity of the fluid in pores.

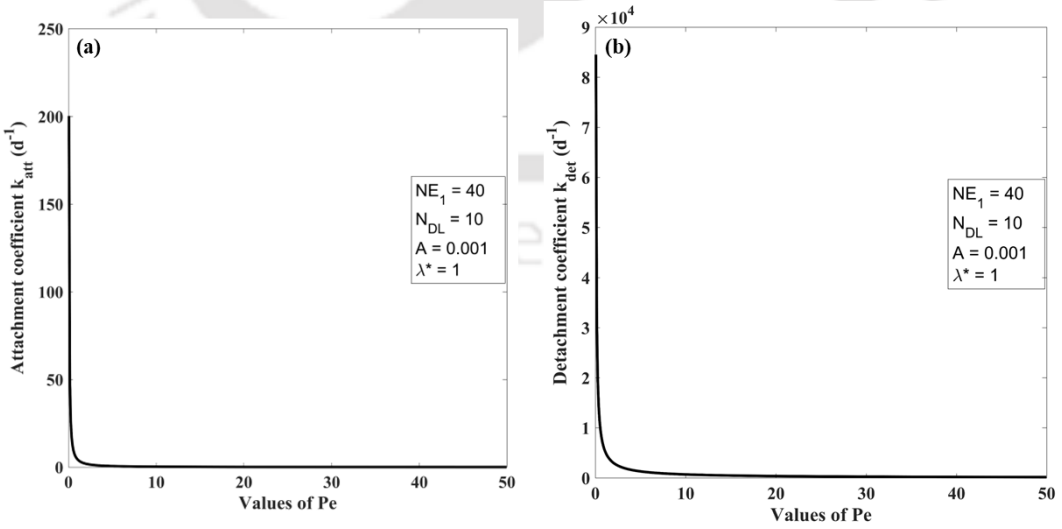

Figure 5.7: Sensitivities of Pe on (a) attachment and (b) detachment coefficient

5.3.1.4 Sensitivity of A

The range for the variation of A is 0.00004 to 0.01. I studied the sensitivities of this parameter on the attachment and the detachment coefficients and produced the results in Figure 5.8. The value A only depends on the radius of the pore (R) and the radius of the virus (a). The attachment and detachment coefficient figures show variations in the behaviors. The variation is because of the way the correlation factors are assigned. The correlation factors for attachment coefficient are different when A < 0.001 to A > 0.001; therefore, the curve switches direction at that point. However, such assignment of coefficients is not given for the detachment coefficient, and therefore the curve behaves the same.

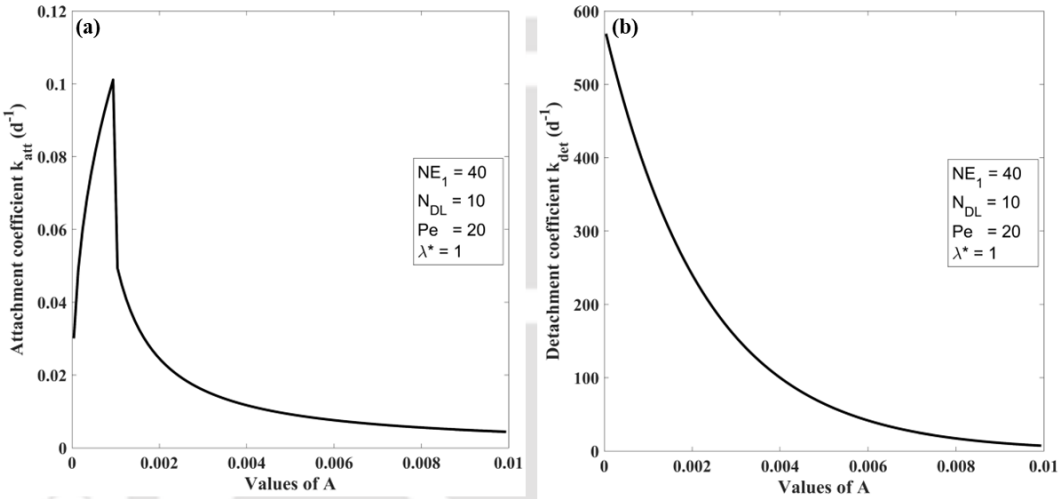

Figure 5.8: Sensitivities of A on (a) attachment and (b) detachment coefficient

5.3.1.5 Sensitivity of λ^*

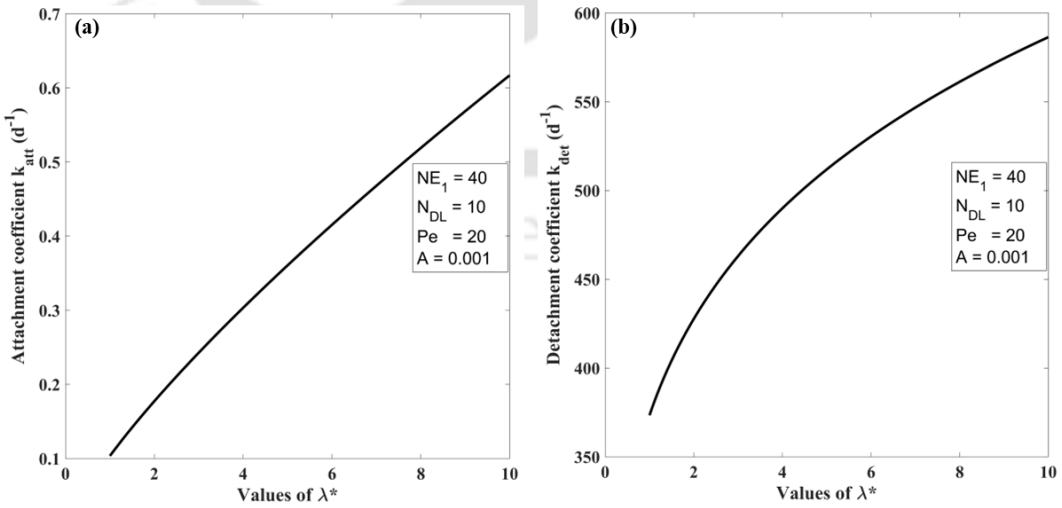

Figure 5.9: Sensitivities of λ^* on (a) attachment and (b) detachment coefficient

The range for the variation of λ^* is 1 to 10. I studied the sensitivities of this parameter on the attachment and the detachment coefficients and produced the results in Figure 5.9. The dependent factor for the λ^* value is only the radius of the virus. Therefore, this factor plays almost no significant role in the design of a management policy.

5.3.2 Sensitivities of the parameters linked over multiple REV

As I have discussed in section 5.3.1, the major contributing factor is the temperature of the medium in the REV, as it occurs in three different non-dimensional numbers. The temperature change also changes the fluid's viscosity, becoming a significant factor. Also, when multiple REVs are involved in the analysis, the mean velocity of each REV might change due to some pumping wells in the aquifer. Therefore, the velocity at each node must be considered for the calculations. Lastly, to check the method's applicability over different virus particles, the radius of the virus becomes an essential factor too. The variation of these four parameters are shown in Figure 5.10 to Figure 5.13. A study area of 5 grid points in each direction is taken, and the PNM is generated for all the REVs in the study area (i.e., 125). At each REV, the averaged correlation equations are found based on the different conditions in the aquifer.

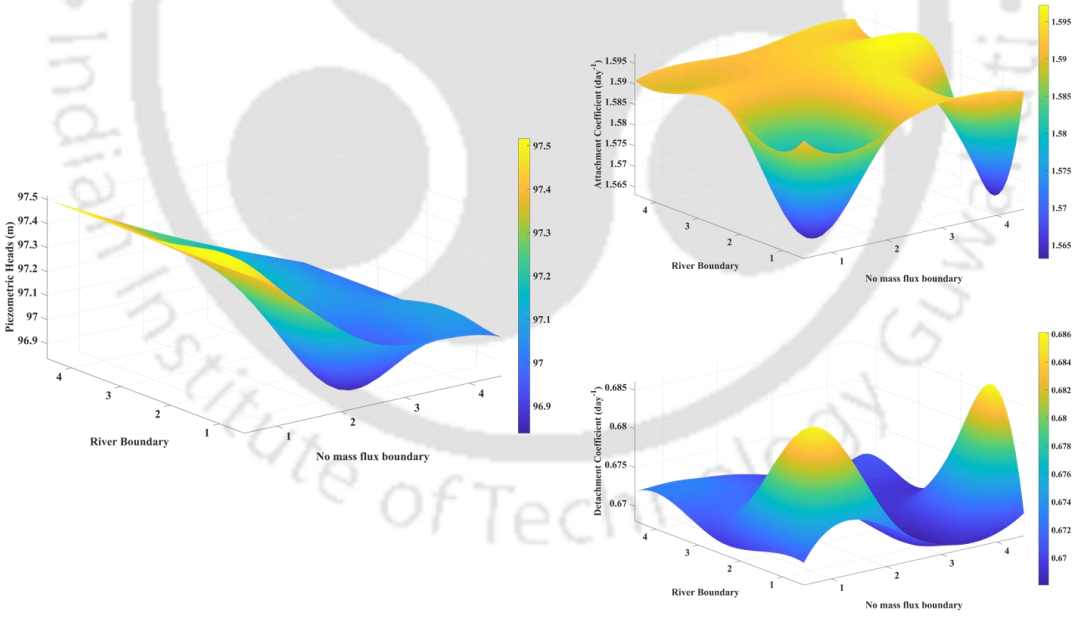

Figure 5.10: Variation of attachment and detachment coefficients with piezometric head

The piezometric head changes because of the introduction of pumping wells into the aquifer. I have installed a pumping well at 2 m from the no mass flux boundary and 3 m from the river boundary. Due to pumping, velocities fluctuate in the aquifer around the pumping well. Those fluctuations result in the variation of the attachment and detachment coefficients in the aquifer.

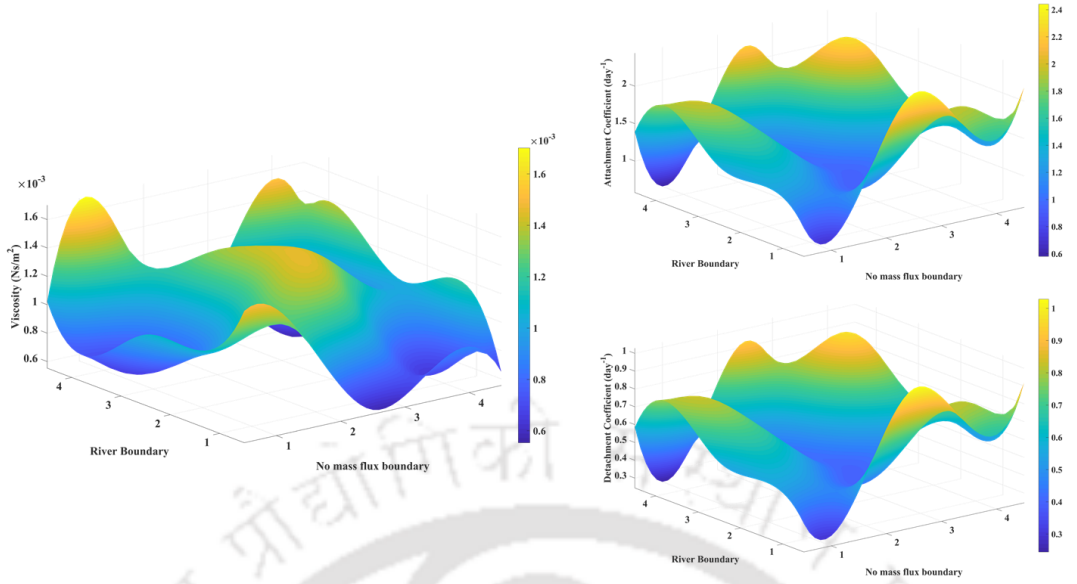

Figure 5.11: Variation of attachment and detachment coefficients with the viscosity of the medium

I have generated the virus's radius, the medium's viscosity, and the medium's temperature as random fields in the aquifer. Therefore, the variations of the attachment and detachment coefficients can be seen as random dependencies, unlike the piezometric head. The random field generation is done to isolate the behavior of each parameter to be independent of the others. Figure 5.11 shows that the variation of attachment and detachment on the viscosity is a negligible value. At the same time, Figure 5.12 shows the effect of temperature to be a much more significant variation as compared to viscosity.

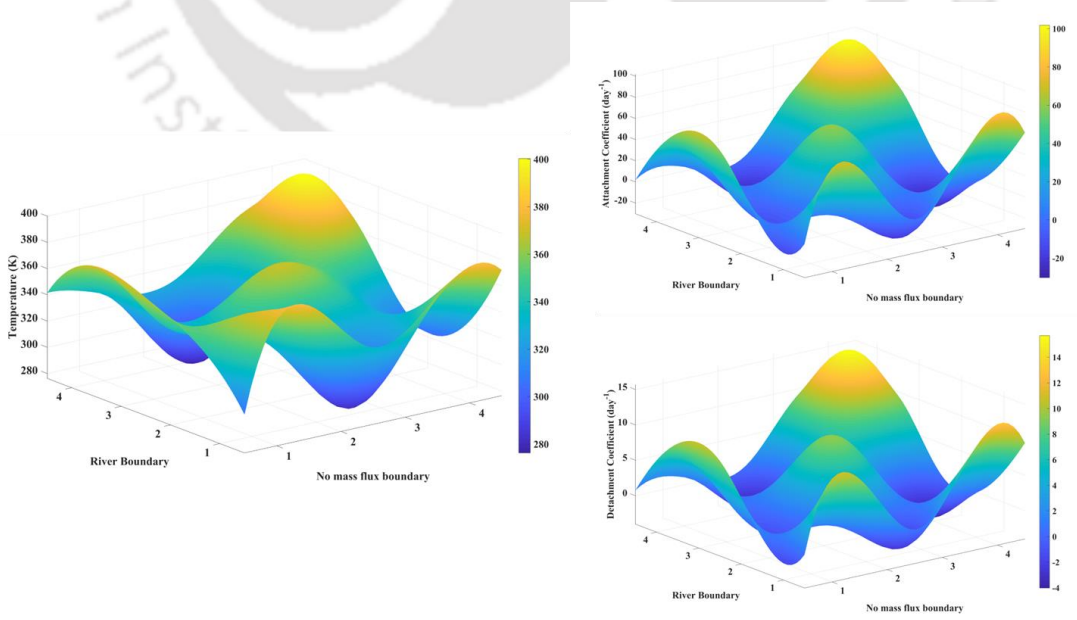

Figure 5.12: Variation of attachment and detachment coefficients with the temperature of the medium

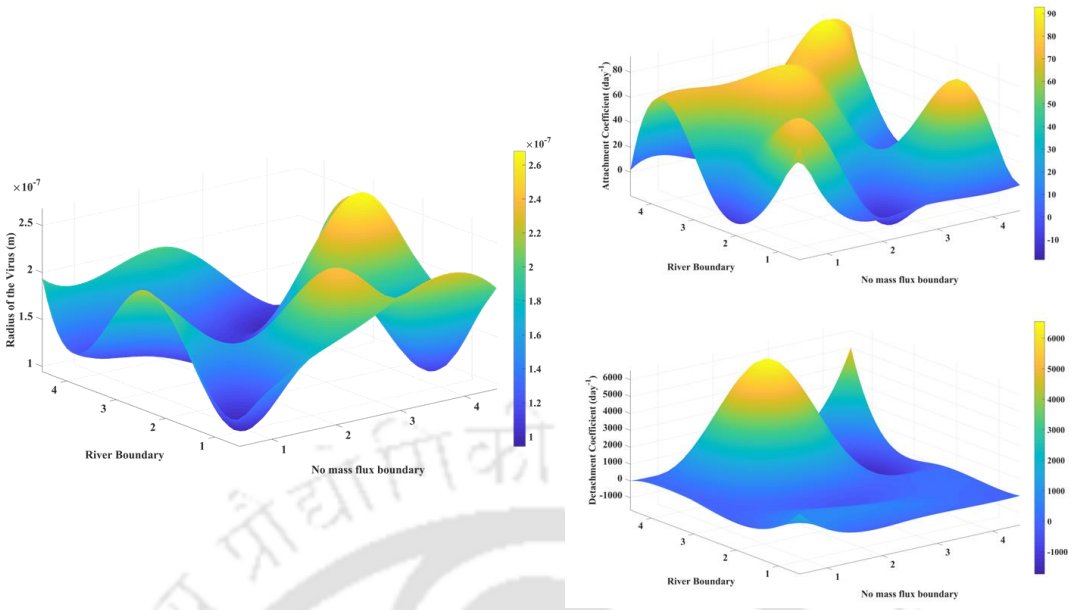

Figure 5.13: Variation of attachment and detachment coefficients with the radius of the virus

It can be observed from Figure 5.13 that the radius of the virus particle has much significant effect on the attachment and detachment behavior. However, a virus cannot change its radius during its transport in porous media unless it mutates to a different protein structure. Therefore, the parameter can be ignored despite its sensitivity towards attachment and detachment behaviors.

5.4 Conclusion

The estimation of parameters from the pore network model (PNM) is a computationally challenging task. Therefore, one needs to select the significance of the parameters before modeling the results from the PNM. This chapter comprehensively explains the significance of each parameter in the porous network. It is concluded from this chapter that temperature plays a significant role in the attachment and detachment of virus to the porous matrix. Later comes the ionic strength, then the mean velocity of the matrix, and then the surface potentials of the virus and the porous matrix. Estimating the properties in Darcy's scale can be done successfully from the pore-scale properties using the approach in this chapter. All the averaging is done based on the surface averaging as the deposition constants correspond to the surface of the porous matrix.

CHAPTER 6: VIRUS SOURCE IDENTIFICATION MODEL

6.1 Introduction

Groundwater quality is deteriorating at an astounding rate in recent years due to the rapid unplanned urbanization and industrialization (Aral et al. 2001; Leichombam and Bhattacharjya 2019). The contaminants that cause the significant source of pollution stay in the groundwater for a prolonged time to cause harm to the humans that consume groundwater. Out of such contaminant is a virus that is a dangerous contaminant that can cause the most harm even when present in the groundwater at a minuscule concentration. Urbanization and industrialization increase the probability of more landfill dumping sites and underground sewers, which cause harmful viruses and bacteria to seep into the groundwater (Gandhi and Bhattacharjya 2020b). The contamination sources must be detected as soon as the concentration occurs in some monitoring wells (Gorelick et al. 1983; Leichombam and Bhattacharjya 2019). Detecting contaminants means knowing the origin of contamination with a specific location and the time the contaminants have entered the aquifer. Identification of such contaminants is called a groundwater source identification problem. Many researchers have addressed the source identification problem with various approaches (Chadalavada et al. 2012a; Jha and Datta 2013; Singh and Datta 2004, 2006). The major approaches are the response matrix approach, the embedded approach and the linked simulation optimization approach (Chadalavada et al. 2011). Out of these approaches, the linked simulation and optimization approach has proven to be the best for problems that contain multiple and mixed variables (Leichombam and Bhattacharjya 2019; Rajeev Gandhi et al. 2017).

Simulating the aquifer condition backwards in time to identify the sources of contamination is computationally challenging. Also, it takes a large amount of input data and produces mixed results most of the time, where the convergence of the solution might not occur (Singh and Datta 2004). However, there is another way to address this source identification problem where the initial condition of the aquifer is initially assumed to be known. Here, the pollution source and strength of pollution are arbitrarily assumed to be any value and the simulation is done for the forward model. The breakthrough curves of observation data at any monitoring well and the simulated data are matched. The error between the observed and simulated data is minimized to find the actual source location in the aquifer. Such an approach is called the inverse optimization approach. (Bhattacharjya et al. 2015; Jha and Datta 2013; Mahinthakumar

and Sayeed 2006; Rajeev Gandhi et al. 2017). Also, the optimization and simulation models are to be linked in this approach. It is also categorized as a linked simulation optimization model. Although the problem seems to be solvable, many computational challenges must be addressed before the solution to the problem is devised. The major challenge for such a problem is the mixed variable nature of the optimization model. The locations of the contaminant sources are generally integer values (as the simulation model is typically any numerical solution, and the source location can be addressed by its element number or cell number depending upon the approach used). The contaminant source flux values are typically continuous variables. Therefore, finding the right algorithm to solve such mixed variable optimization problems creates a problem (Gorelick et al. 1983). Algorithms such as genetic algorithms are very efficient in solving integer variables, and gradient-based search algorithms are efficient in solving the problems of continuous variable nature (Mahar and Datta 2000, 2001). To take advantage of both the algorithms together, researchers (Mahinthakumar and Sayeed 2006) have developed a hybrid genetic algorithm with local search (GA-LS) to identify the sources. Leaping forward, a different approach of using a surrogate-based model is developed by researchers (Hazrati-Yadkoori and Datta 2017) to address the complex nature of the solution to source identification. A recent new methodology (Leichombam and Bhattacharjya 2019) has used the genetic algorithm and the gradient-based search to identify the sources effectively with fewer function evaluations and the corresponding computational burden. However, these methodologies have used the information from the observation wells for a very long period (generally 3 to 5 years) to identify the sources. It might not be the case for identifying virus sources as the damage caused by viruses is much more severe, and the time required to identify the sources is of significant concern while devising the source identification model.

The virus source identification model proposed for this research can identify the sources with limited and minimum information from the observation wells. The locations of the virus sources are unknown, and the source number or the information of virus entry into the aquifer is not known for this model. The virus transport model explained in chapter 3 is used to solve this problem. The problem for which the virus transport model is generated is given in chapter 3, and the same problem is taken for identifying the sources, assuming that the virus locations and the source flux are unknown. The virus sources creep into the aquifer from any source but show concentration in the observation wells for 50 days, increasing the concentration daily. The information from the observation wells is the only input for the source identification model. The model assumes that the source flux values are unknown for the entire 50 days and

the locations are unknown. The proposed methodology can handle such problems efficiently with the lesser computational effort.

6.2 Methodology

The source identification model needs input from the field case where there is a virus detection in the field for a prolonged period at different observation wells. The duration of measurement concentration of viruses and the location of the observation wells can give us an idea of where the sources of the viruses can be located. Thus, the observation wells' location is selected to cover the aquifer's whole area. A monitoring well network is created in the hypothetical unconfined aquifer discussed in chapter – 3. The location of the observation wells is to be selected randomly but with a set of rules. The total aquifer area is divided into equal parts (16 parts in this case), as shown in Figure 6.1.

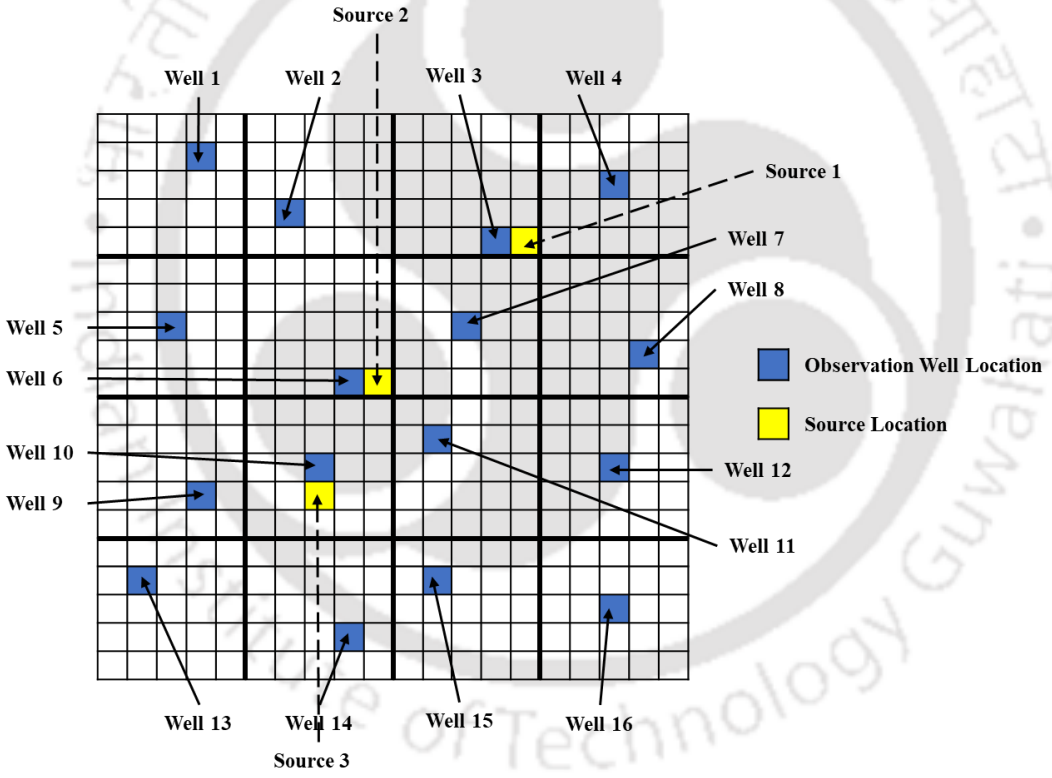

Figure 6.1: Location of observation wells along with the actual source locations

Each part selects one observation well by following the probability distribution of highest at the centre to least at the endpoints. For this current problem, the study area consists of 400 cells in each layer. Therefore, the probable locations of observation wells are 16, each selected from 400/16, i.e., 25 locations. One observation well is selected for each part based on a uniform normal distribution from highest at the centre to least at the edges as specified. As the problem specifies that the virus source locations are industrial dumping sites, all the observation wells

are in the first layer not exceeding 10 m depth from the ground level. The 16 probable well locations and the whole area is given in Figure 6.1. This method of selecting the observation wells covers the aquifer's whole area with observation wells, and the data of concentration from each well can be used to identify the source location and strength.

The source identification model is an optimization model that minimizes the weighted and squared difference between observed and simulated concentrations added throughout the observation wells and the number of time steps. The simulated concentrations are the concentrations in the aquifer after simulating the virus transport by assuming the arbitrary locations and source strengths. The information on the aquifer parameters, the location of the pumping, boundary conditions and the observation wells from where the concentration data is collected in the aquifer with observed concentration at different times is collected before the preparation of the simulation model. Based on this information, the virus transport model for arbitrary sources can be modelled and then the optimization search can be performed on the model to identify the source locations and the flux. The objective of the optimization model is to minimize the error from the observed and the simulated concentrations at the observation wells. Equation 6.1 shows the objective function.

$$Minimize\ f = \sum_{o=1}^{o=oc} w_o \left(\sum_{k=1}^{k=sk} \left(OC_o^k - SC_o^k \right)^2 \right) \qquad 6.1$$

Here *'sk'* represents the total number of time steps and *'oc'* represents the total number of observation wells for concentrations. The source locations are unknown, and the leakage at the source locations per day can also be predicted using this minimization function. Therefore, the observed OC_o^k and simulated concentrations SC_o^k at specific well locations are a function of the leakage at the source per day. The term w_o represents the observation weight of the well calculated in section 6.2.1.1. The observation well data collected is for the period of 50 days; therefore, the source leakages are assumed to be active for the whole of 50 days. The time step for this model is 10 days; therefore, the number of time steps '*sk*' of 5 is used. Also, the number of observation wells described above is 16; therefore, '*oc*' is 16 for this model. The details of the observed concentration are given in Figure 6.2.

As it can be observed from the figure, the locations close to the actual source locations, i.e., Well 4, Well 6 and Well 10 detect some observed concentration, and the rest of the wells are almost negligible in detecting concentration.

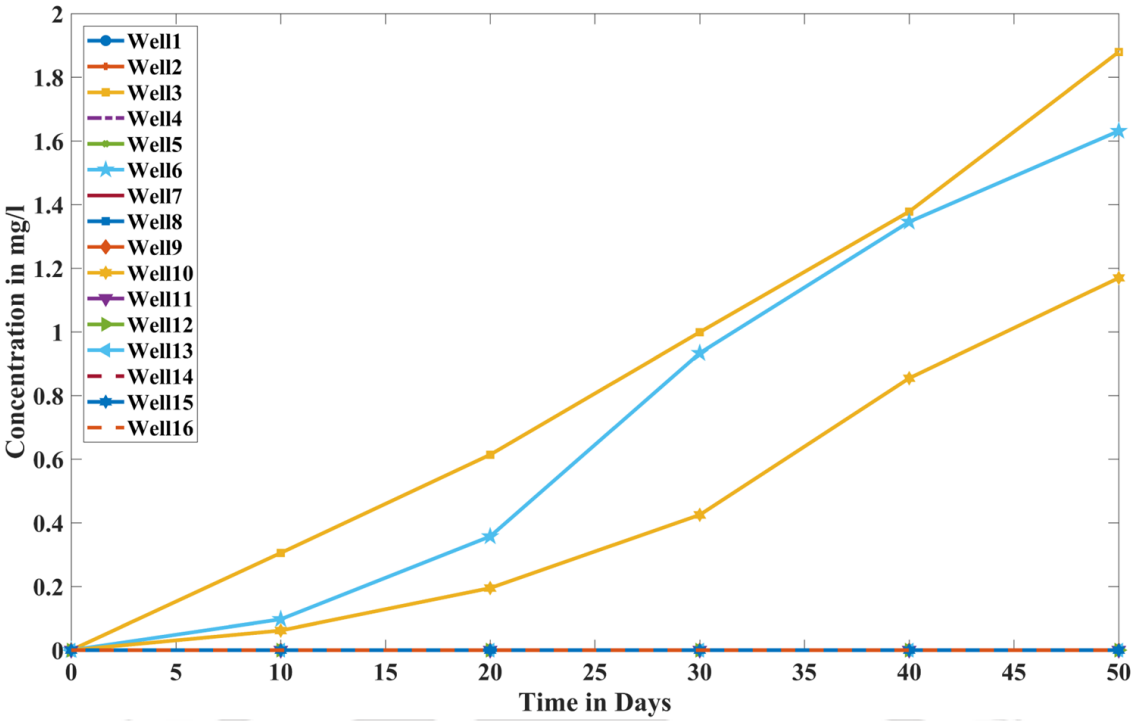

Figure 6.2: Observed concentration at different observation well locations

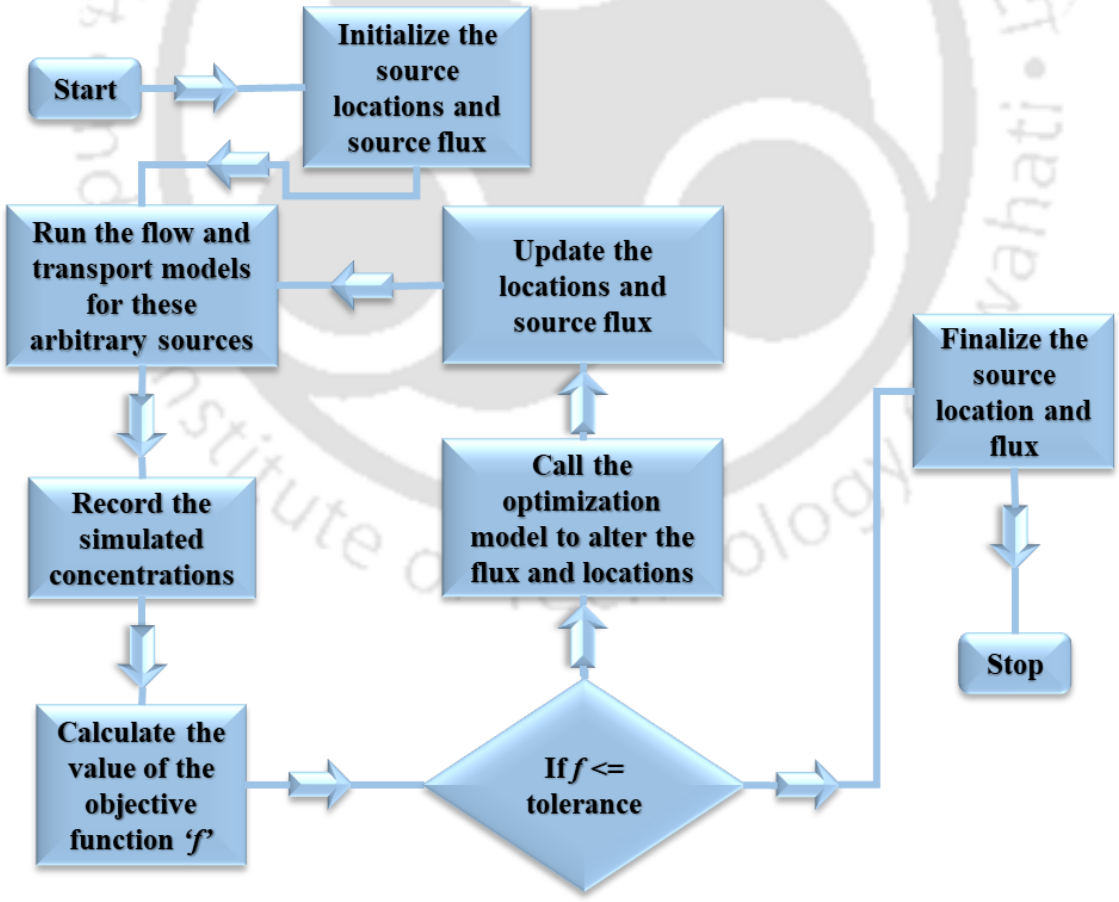

Figure 6.3: Overview of the source identification model

The methodology proposed in this chapter allows us to use such information of observation wells to identify the probable source locations before running the optimization model. The overview of the methodology is given in Figure 6.3. According to the methodology, I have assigned arbitrary source locations and source flux to calculate the objective function value *'f'*. Then, I have called the optimization model after checking the tolerance (usually taken to be 10^{-6}) for the function value. After the optimization model generates the updated locations and the flux values, I repeated the steps to calculate the objective function till the tolerance criterion was satisfied, and the source locations and the flux values are finalized.

For running the optimization model, a new hybrid methodology (Leichombam and Bhattacharjya 2019) is used, which uses a modified genetic algorithm along with the gradient-based search. However, some modifications are made to the established methodology better to identify the source strengths for all the aquifer conditions. First, the proposed methodology applies to two-dimensional flow and transport and needs to be extended to three dimensions. As the philosophy is the same, the extension to three dimensions is not difficult to execute. However, the drawback of the methodology is that it assumes the initial source locations and flux arbitrarily with no prior information of the observation wells taken into consideration. The method's efficiency is highly dependent on the initial guess of the source location. The source flux can be easily identified once the location is known for the source. Therefore, a methodology is proposed in this chapter to add the initial approximation of source location efficiently. Also, I have added a new operator to the modified genetic algorithm to always include the most probable locations as a priority by introducing a new probabilistic operator.

6.2.1 Initial Approximation of Likely Locations

The initial approximation of the locations is an essential step in the source identification with unknown locations as it reduces the algorithm time drastically. The overview of the source identification model in Figure 6.3 initializes the source locations and flux. I have designed this methodology to initialize the locations based on probabilities of likely locations. I have taken into account both advection and dispersion dominant flows to initialize the source locations. The molecular diffusion is almost negligible compared to the dispersion; hence, the flow is termed dispersion dominant flow contrary to the usual term of diffusion dominant flow. The overall methodology to predict the probable locations of sources is given in the Figure 6.4. I have explained the methodology's step-by-step process in detail in the subsequent sections.

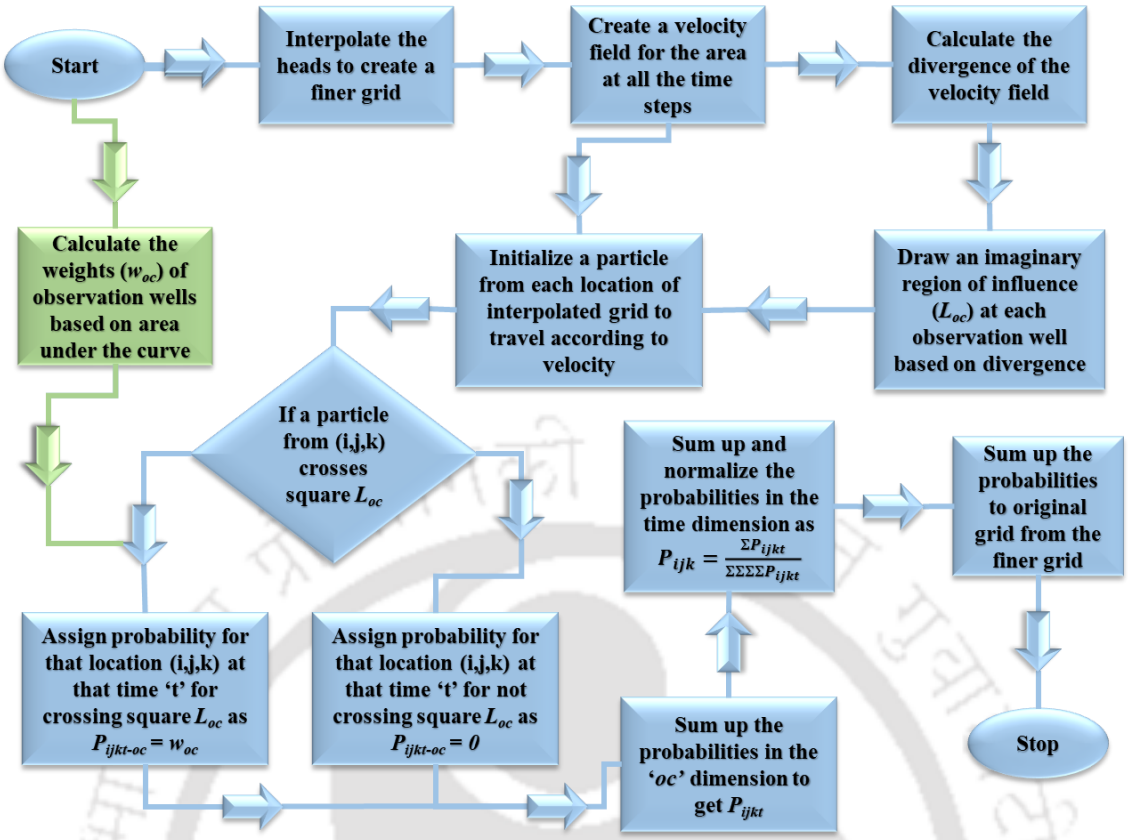

Figure 6.4: Flowchart for initial approximation of probable source locations

6.2.1.1 Calculation of weights based on area under the curve

Each observation well has different concentration values at different times, as shown in Figure 6.4. If the concentration is detected at any observation well, a source location may be near and around that well. However, I have converted that possibility space to a probability space by assigning how strong the concentration values are at the given observation well. Therefore, weights are to be assigned to each observation well for the probability space. I have used the trapezoidal rule to numerically calculate the area under the curve. Later, the area under the curve is normalized to be the weight of each observation well. The formula for the generation of weights is given by equation 6.2. The generated weights at all the observation wells are given in Table 6.1. As discussed in section 6.2 and Figure 6.2 the weights are higher for the well near the source; therefore, that well contributes the highest probability.

$$w_{oc} = \frac{\int_{t=0}^{t=sk\times dt} C_{oc}}{\sum_{oc=1}^{oc=16} \int_{t=0}^{t=sk\times dt} C_{oc}} \qquad 6.2$$

Table 6.1: Normalized weight at each observation well

Observation Well	Normalized Weight	Observation Well	Normalized Weight
Well 1	3.87×10^{-20}	**Well 9**	2.92×10^{-7}
Well 2	6.59×10^{-13}	**Well 10**	0.2160
Well 3	0.4221	**Well 11**	1.20×10^{-5}
Well 4	6.39×10^{-6}	**Well 12**	8.58×10^{-14}
Well 5	1.18×10^{-13}	**Well 13**	9.26×10^{-15}
Well 6	0.3618	**Well 14**	3.26×10^{-15}
Well 7	8.79×10^{-6}	**Well 15**	1.39×10^{-9}
Well 8	1.38×10^{-9}	**Well 16**	6.32×10^{-28}

6.2.1.2 Interpolation to a finer grid

The aquifer area for any practical problem is relatively larger; therefore, forming the grid as per the aquifer area is more practical. The grid formation for this particular problem is 10 m in each of the three directions forming a cube of 10 x 10 x 10 m for each cell.

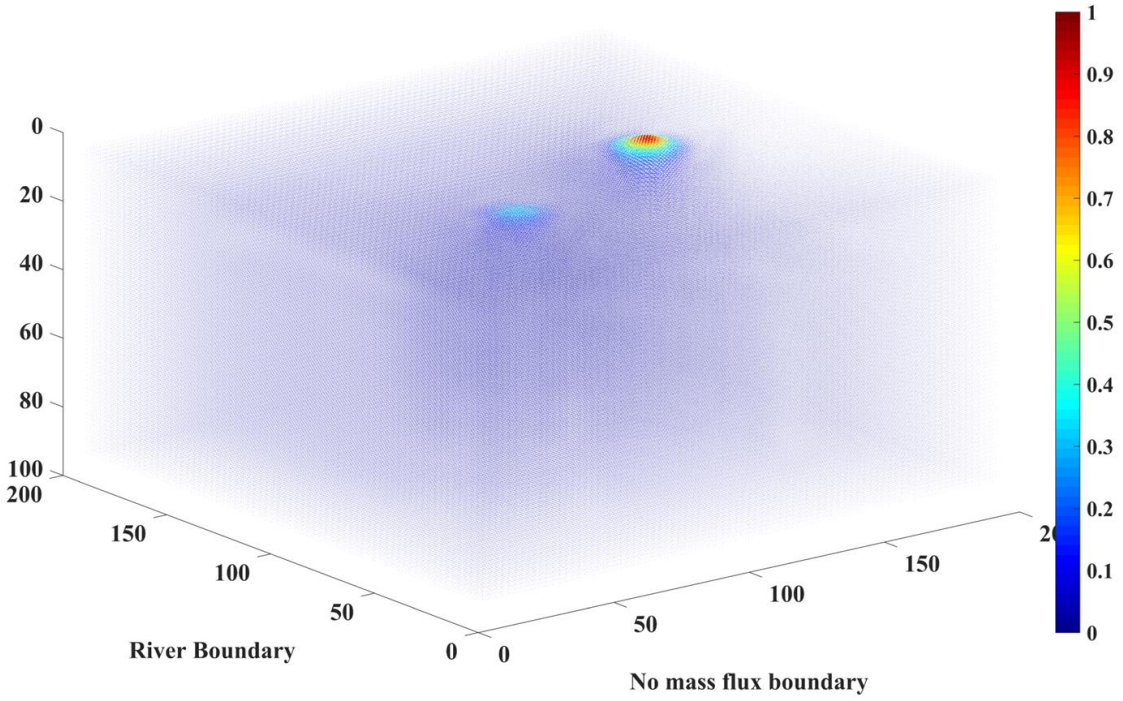

Figure 6.5: Normalized velocity field for the interpolated grid (All the velocities are unit vectors)

However, the velocities of groundwater are much less in any groundwater flow. Therefore, in the given number of time steps (5 in this case) with the step length of a low value (dt = 10 days

in this case) the velocities might not result in a significant movement in any particle. The particles might only travel a distance from one cell to another cell in the whole simulation time (50 days in this case). I have used the prediction for the probability of positions based on velocities for this methodology, and such slow movement might not result in a very accurate prediction of the probabilities. Therefore, I have interpolated the piezometric heads generated from the flow model to a finer grid size and the calculations are made based on the interpolated datasets. I have also interpolated the hydraulic conductivity to the appropriate grid size. In this case, the 20x20x10 grid is converted into a 153x153x73 finer grid size. The division is based on the rule that each cell is divided into 2^3 i.e., 8 parts in each direction. The interpolated velocity field is calculated using Darcy's Law to measure the specific discharge in any direction. The velocity field normalized to a magnitude between 0 and 1 is given in Figure 6.5. The units of the velocity field are in m/day.

6.2.1.3 Divergence of the velocity field

The divergence is a scalar value assigned to every point of space that determines the strength of fluid flow around that region.

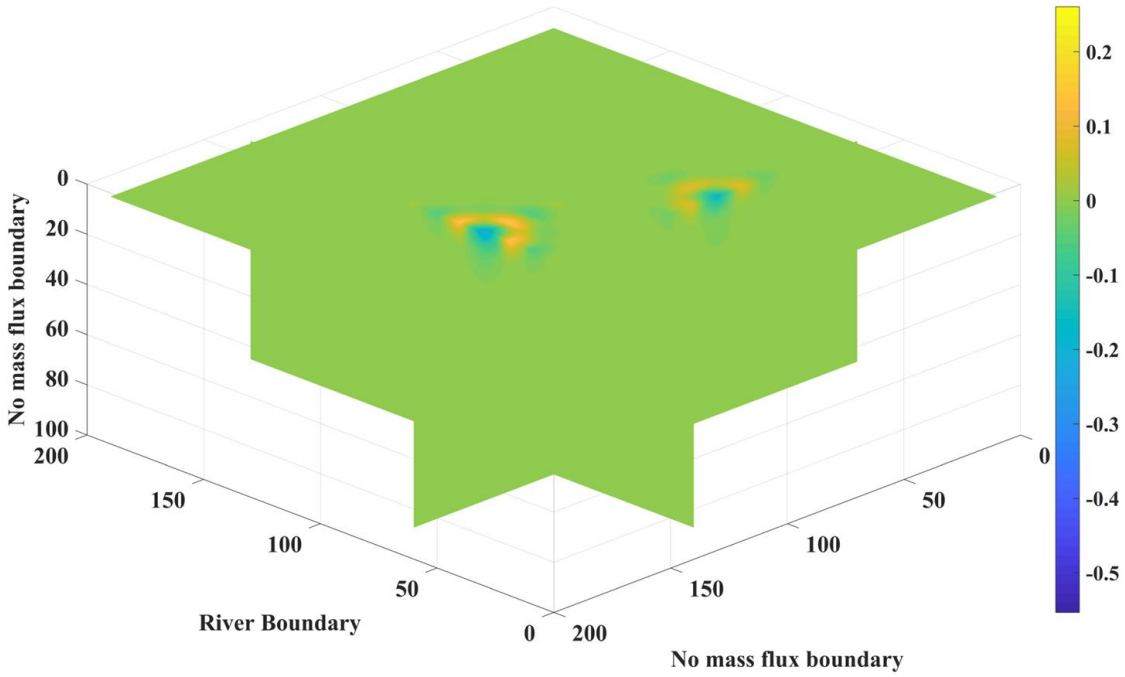

Figure 6.6: Divergence of the velocity field for the slices around the pumping wells

The divergence will be negative values around sinks and positive values for sources and regions where there is a gradient of velocity. In other words, the divergence value determines the strength of velocity around a particular region. The divergence of the velocity fields for the

relevant slices of contours is given in Figure 6.6. The divergence has a lesser value (negative) near the pumping wells and relatively negligible values elsewhere. The divergence of a region is the key to the strength of movement around that region. Therefore, I have considered it for selecting probable source locations. The details of accounting divergence in the source identification are given in section 6.2.1.4.

6.2.1.4 Imaginary region of influence based on divergence

The divergence assigns a region of influence at each observation well. If a contaminant from any particular location (say L) has to pass through this region of influence in its travel time, a probability is assigned to that location 'L' for it to be a potential source of contamination. However, the contaminant moves faster when the absolute divergence is higher and slower when the absolute divergence is lower in any given region. Therefore, the region of influence is assigned a higher or lower volume around each observation well based on divergence. Initially, the divergence is converted into an absolute value and normalized to obtain a probability distribution, as shown in equation 6.3. The probability of the divergence is over the interpolated grid as the base values of divergence are from the interpolated grid. Therefore, the strength of divergence S_{div} at a particular location of observation well is to be cumulatively added over the minute regional interpolated grids as shown in equation 6.4.

$$P_{div} = \frac{\sum_{time} |div|}{\sum_{time}\sum_{x-dir}\sum_{y-dir}\sum_{z-dir} |div|} \tag{6.3}$$

$$S_{div}^{obs} = \sum_{z=1}^{z=2^{k-1}+1} P_{div}\left(x^{obs}, y^{obs}, z\right) \tag{6.4}$$

The size of the region of influence is inversely proportional to the divergence. Therefore, the strength calculated is inverted to calculate the overall probability called region value $R_{\inf}$ as shown in equation 6.5 and normalized to a value between 0 and 1 as shown in equation 6.6.

$$R_{\inf} = \frac{\sum_{obs} S_{div}^{obs}}{S_{div}^{obs}} \tag{6.5}$$

$$R_{\inf}^{norm} = \frac{R_{\inf} - R_{\inf}^{\min}}{R_{\inf}^{\max} - R_{\inf}^{\min}} \tag{6.6}$$

This value of normalized R_{inf} is used to calculate the region of influence at a particular observation well. I have used the factor R_{inf} to calculate the stretch of the region of influence. I have assigned a maximum value of three times the discretized grid and minimum value of one time the discretized grid. The selection is done using equation 6.7 shown for the x-direction. The region of influence should, however, not exceed the study area, and it can be achieved by cutting the region if it exceeds the limits in each direction.

$$I_x^H = 3dx \;\&\; I_x^L = dx$$
$$I_x = I_x^L + R_{\inf}^{norm}\left(I_x^H - I_x^L\right) \qquad 6.7$$

Here, I^H and I^L represent the higher and lower limits of the region of influence in each direction, dx represents the discretized grid in the x-direction, and I_x represents the calculated extent of the region of influence from the observation well based on the normalized influence at that particular observation well.

The respective values of probability through divergence, the strength of probability, region of influence value, the normalized value of the region of influence and the maximum stretch for each direction from the centre of the observation well are given in Table 6.2.

Table 6.2: Table representing the region of influence values for each observation well

Observation Well	S_{div}^{obs}	$R_{\inf}$	$R_{\inf}^{norm}$	I_x	I_y	I_z
Well 1	3.40 x 10^{-7}	250.487	0.3066	16.1312	16.1312	12.246
Well 2	2.91 x 10^{-5}	2.928	0	10	10	7.1295
Well 3	3.98 x 10^{-7}	213.667	0.261	15.2193	15.2193	9.7885
Well 4	1.05 x 10^{-7}	810.47	1	30	30	18.4598
Well 5	3.32 x 10^{-6}	25.589	0.0281	10.5612	10.5612	8.1244
Well 6	7.61 x 10^{-6}	11.17	0.0102	10.2041	10.2041	7.022
Well 7	1.36 x 10^{-6}	62.637	0.0739	11.4788	11.4788	7.2325
Well 8	2.14 x 10^{-7}	397.968	0.4892	19.7838	19.7838	11.9496
Well 9	3.49 x 10^{-6}	24.345	0.0265	10.5304	10.5304	8.0279
Well 10	2.03 x 10^{-5}	4.19	0.0016	10.0313	10.0313	7.1275
Well 11	1.15 x 10^{-5}	7.368	0.0055	10.11	10.11	6.5794
Well 12	5.52 x 10^{-7}	154.046	0.1871	13.7427	13.7427	8.4006
Well 13	4.71 x 10^{-7}	180.4	0.2198	14.3954	14.3954	11.3448
Well 14	1.70 x 10^{-6}	50.128	0.0584	11.169	11.169	7.8971

Well 15	4.47 x 10^{-6}	19.015	0.0199	10.3984	10.3984	6.9926
Well 16	1.20 x 10^{-7}	709.05	0.8744	27.4882	27.4882	16.7338

The limits of the region of influence for each direction are calculated as shown in equation 6.8.

$$I_x^{left} = \max\left(X_{\min}, X^{obs} - I_x\right)$$
$$I_x^{right} = \min\left(X_{\max}, X^{obs} + I_x\right) \qquad 6.8$$

The stretch of the region of influence in the left and right directions is calculated in equation 6.8 by taking the limiting condition of not exceeding the model domain in the left and right directions, respectively. The process can be repeated for the other two directions to get the whole region of influence for the front and back directions and top and bottom directions for Y-direction and Z-direction, respectively. The whole region of influences calculated as cuboids is given in Figure 6.7.

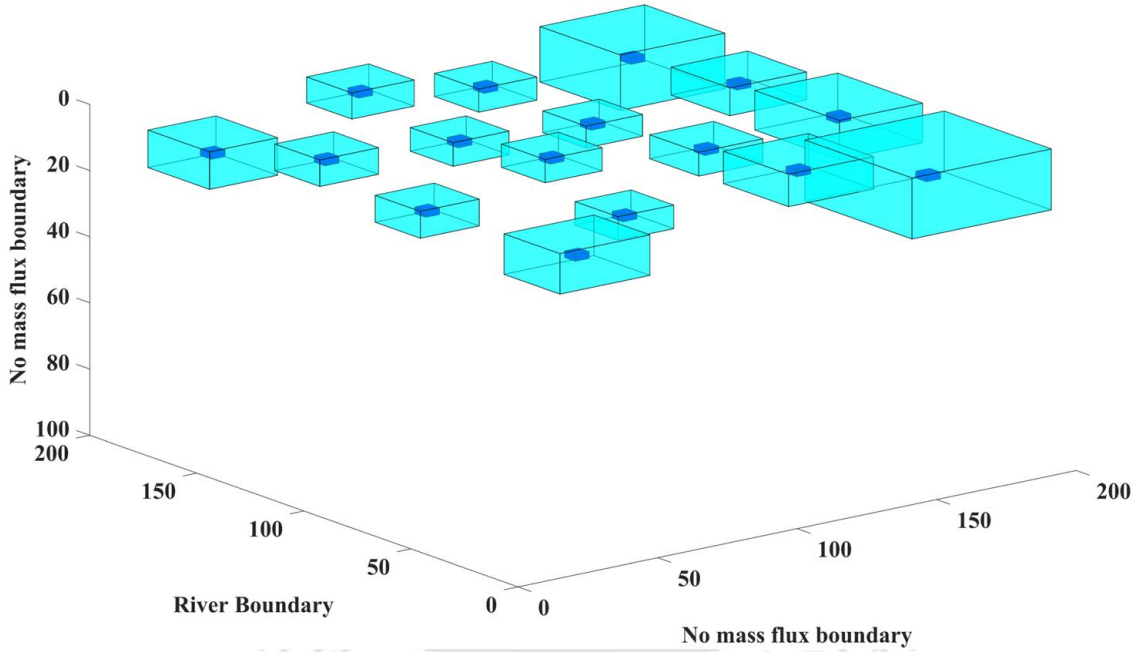

Figure 6.7: Imaginary region of influences calculated based on divergence

6.2.1.5 Assigning the probability of passing through the region of influence

The region of influence for all the observation wells is selected as described in the previous section. The position of all the points in the model area for the interpolated grid are known as X, Y and Z coordinates. The velocity distribution is also known for every time step. Therefore, if a tracer moves from one point to another due to advection, the movement can be calculated using the Lagrangian mechanics by simple distance and time formulation. I have assumed that multiple tracers are located all along the model domain and the tracers move to a new location based on advection. The new location is then stored, and the process is continued with the new

velocities until the end of the time steps. The process of calculating the position of particles is given in equation 6.9. After the new positions are calculated for each time step, the probability of crossing the region of influence is calculated. For that, I have used a unique method to calculate the region of influence.

$$\begin{aligned} X(x,y,z,t) &= X(x,y,z,t-dt) + V_x(x,y,z,t) \times dt \\ Y(x,y,z,t) &= Y(x,y,z,t-dt) + V_y(x,y,z,t) \times dt \\ Z(x,y,z,t) &= Z(x,y,z,t-dt) + V_z(x,y,z,t) \times dt \end{aligned} \qquad 6.9$$

I have taken the history of Each particle history and the probabilities assigned for crossing the region of influence differently for each direction. I have explained the process in Figure 6.8.

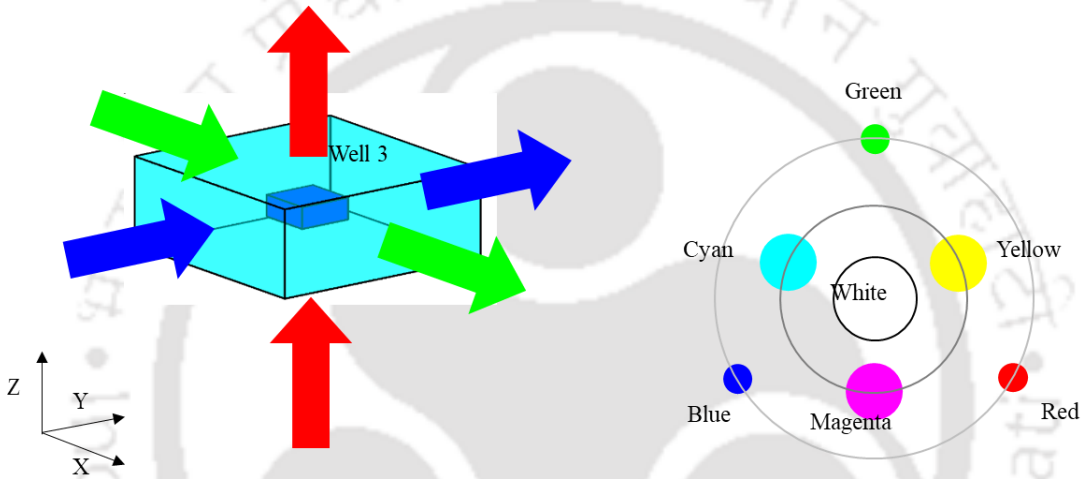

Figure 6.8: Probability assignment for crossing region of influence

Here, I have shown a sample region of influence and assigned the colours green, blue and red for crossing it in x, y and z directions, respectively. I have used an exponential function for the weights assigned to the particle for crossing the regions of influence in each direction. If a particle crosses the region of influence in a single direction, I assigned the weights to a value of $10^{-4}w_{oc}$. The colours blue, green and red represent this category as a symbol in the figure, with the large circle surrounding the colours. If a particle crosses the region of influence in cross directions, i.e., x-y direction, y-z direction, and x-z direction, I assigned the weight of $10^{-2}w_{oc}$. The combination colours of cyan, magenta and yellow represent the combined directions. If a particle crosses the region of influence in all three directions, I assigned the weight precisely equal to the weight of that observation well. The combination colour of white is used to symbolically represent the particles that crossed all the three directional bounds of the walls. If a particle doesn't cross any of the bounds, the weight assigned is just zero. The assignment of probability is unique for each spatial point, each observation well and each time step as well.

Therefore, the probability assigned is in 5 dimensions for this step and is shown in equation 6.10.

$$P(x,y,z,o,t) = k \times w_o$$

$$k = \begin{cases} 10^{-4} & for\ x\ or\ y\ or\ z\ bounds \\ 10^{-2} & for\ xy\ or\ yz\ or\ zx\ bounds \\ 1 & for\ xyz\ bounds \\ 0 & for\ no\ bounds \end{cases} \qquad 6.10$$

The assigned probabilities are initially summed over all the observation wells to get the 4-dimensional probability. Then, the probabilities are normalized over the time dimension to get the sum of all probabilities equal to one and reduce the dimensions to just the spatial locations. Therefore, this gives the overall summary of the probability that each spatial location is a source of pollution.

$$P(x,y,z,t) = \sum_{o=1}^{o=16} P(x,y,z,o,t)$$

$$P_{xyz} = P(x,y,z) = \frac{\sum_t P(x,y,z,t)}{\sum_x \sum_y \sum_z \sum_t P(x,y,z,t)} \qquad 6.11$$

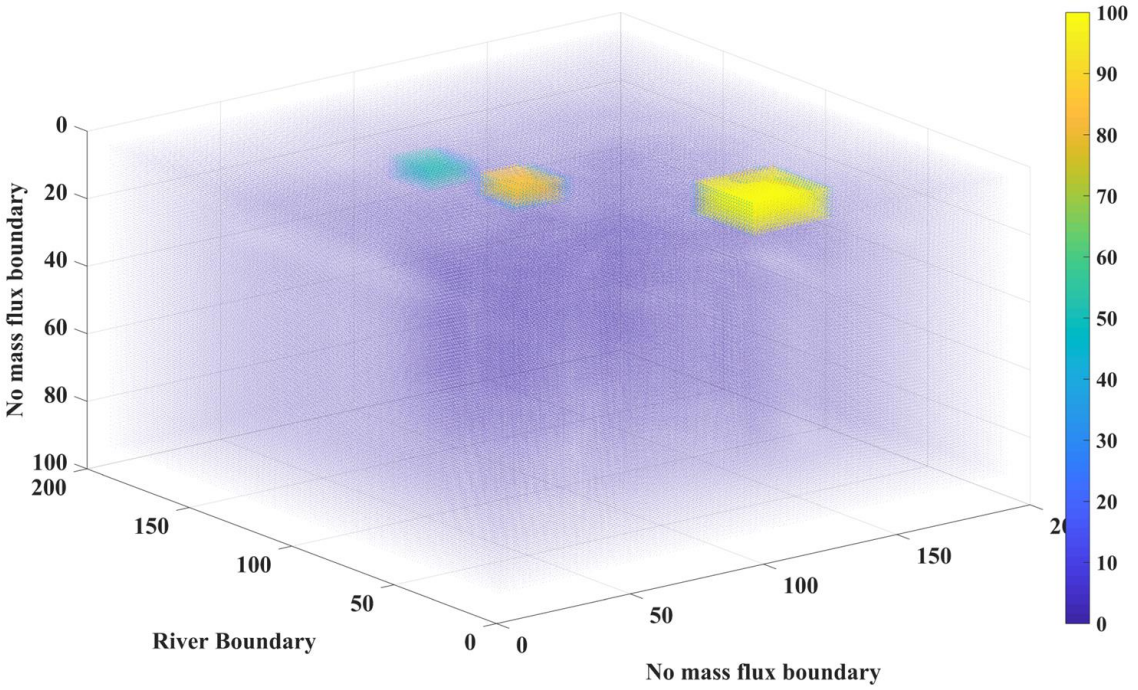

Figure 6.9: Probabilities of the potential source locations

The details of reducing the dimensions are given in equation 6.11. The assigned probability is over the interpolated grid; therefore, the probabilities are distributed very finely in the scattered

plot shown in Figure 6.9. The colour shows the probability percentage, with yellow as 100% probable and blue as 0% probable. All the locations are nearly improbable to be potential source locations. However, three regions are shown as probable locations as per the details in Figure 6.9. The actual source locations fall within the region predicted by this procedure and are explained clearly in section 6.2.1.6.

6.2.1.6 Finalizing the probabilities for each location of the original grid

The original problem contains 20 x 20 x 10 cells, and the interpolated grid region is to be converted to the original cell region to get the actual probabilities for the regions to be potential source locations. I have used the interpolated grid's probabilities to obtain the original grid's probabilities. The cell arrangement for the interpolated to the original grid is given in Figure 6.10 for any single direction. The central part of the cell refers to the cell, which is shown with a blue square in Figure 6.10.

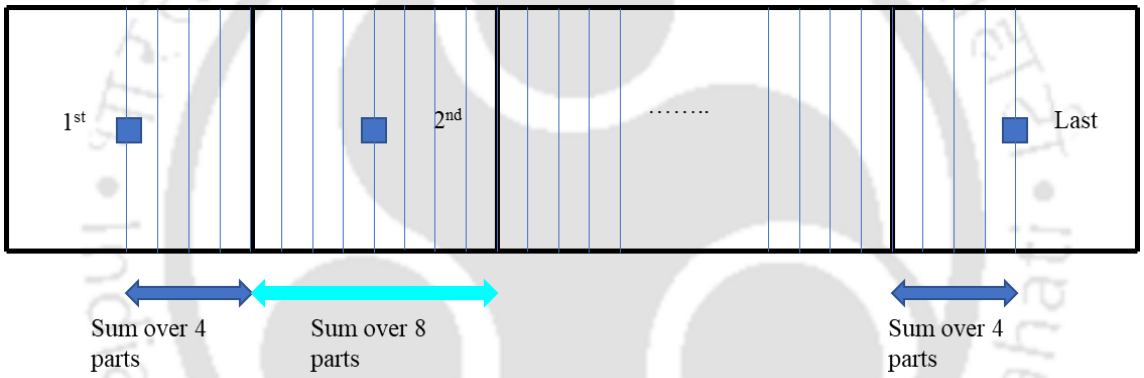

Figure 6.10: Interpolation grid description for any single direction

The number of divisions from cell to cell is 2^3, i.e., 8, as described in section 6.2.1.2. For example, the 20 cells of x-direction convert to 4+8x18+4 parts, i.e., 152 parts creating 153 points in space after interpolation. Therefore, for summing up the probabilities for each cell, the middle cells can be converted by summing over the eight parts, but the first and last cells are to be converted with only half of the parts, i.e., four parts. The process is shown clearly in Figure 6.10. After this process, I have calculated the probabilities for the location of the original grid points to be the potential source locations. The representation of the probabilities for the source locations can be seen in Figure 6.11.

The whole process narrows the source locations to very few cells. The actual sources are also shown in Figure 6.11, which are located with a reasonable probability per the procedure. For comparison, a random distribution of probabilities without any knowledge of the source information would look as shown in Figure 6.12. In both Figure 6.11 and Figure 6.12, the

colour and size of the sphere represent the probability of that location being a potential source. The color ranges from blue to yellow with values of 0 to 100%, representing the probability in percentage.

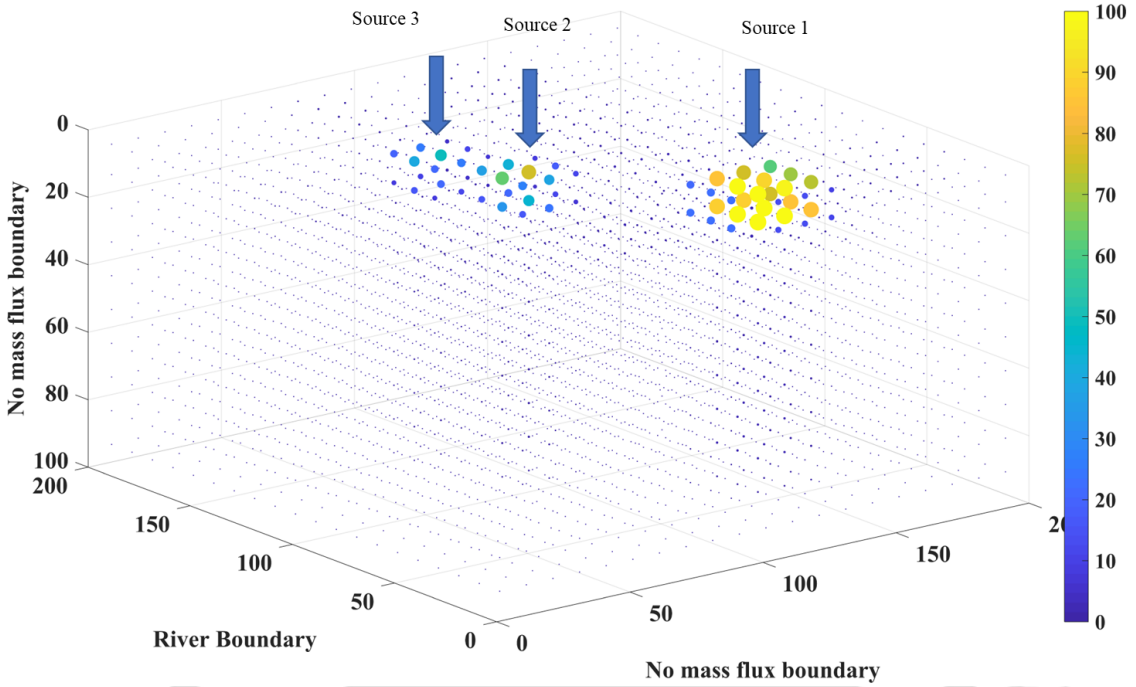

Figure 6.11: Probabilities of each location to be a potential source location

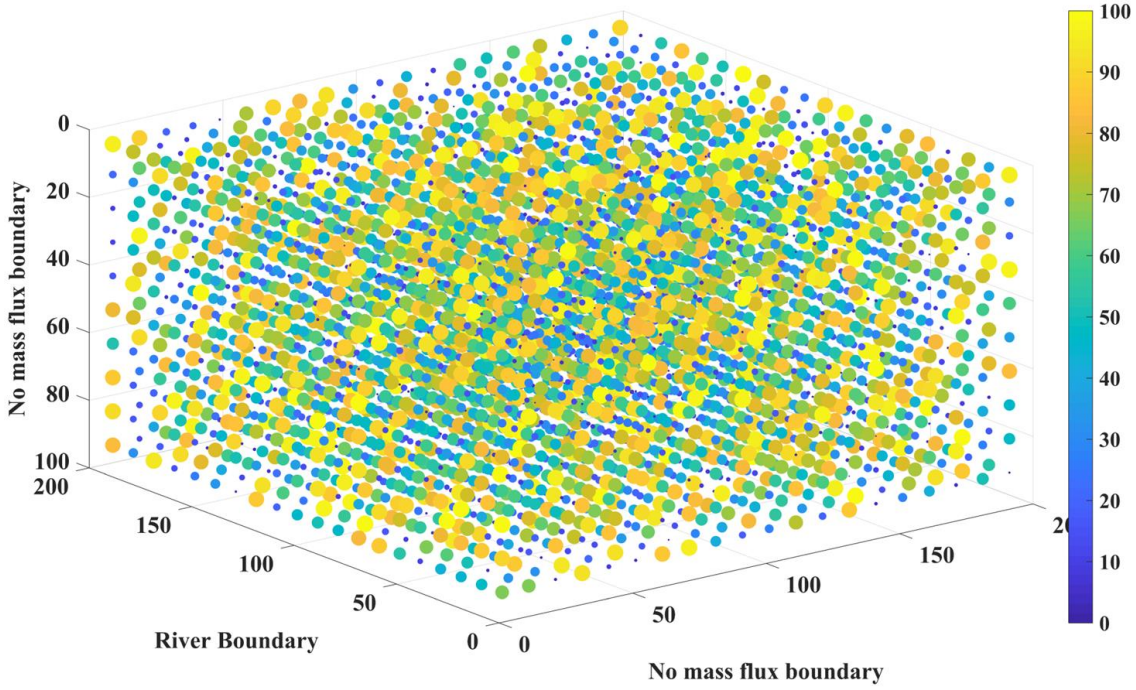

Figure 6.12: Random probability distribution of potential source location

By comparing Figure 6.11 and Figure 6.12, one can see the advantage of using the described procedure. The procedure narrows the search space to a very reasonable level that the initial solution for the potential source locations is much less than the actual source points.

6.2.2 Predicting the probabilities and number of source locations

The described methodology of initializing the most likely locations has the added advantage of predicting the number of potential source locations. It is evident from Figure 6.1 and Figure 6.2 that three potential observation wells signify three potential sources near the observation well. The model prediction from the Figure 6.11 shows that three probable location sources are at a high probability of being the potential source locations. Therefore, the model initially starts with a guess that there are three source locations and proceeds forward. The next step is to rank the potential source locations corresponding to three possible source locations. However, the locations are assigned with a specific number to quickly identify the locations with a unique value. The snake numbering philosophy involves assigning a unique number to each location. Each cell's centre is located in the x, y and z directions with a unique location named i, j and k. Here, i runs from 1 to the number of cells in the x-direction (N_x); j runs from 1 to the number of cells in the y-direction (N_y), and k runs from 1 to the number of cells in the z-direction (N_z). Therefore, as per the snake numbering technique, a snake is assumed to travel from (1,1,1) to (N_x,N_y,N_z), turning at each edge through 180^0 and assigning a value from 1 to N_x x N_y x N_z increasing by a value of one with each cell crossed. The following formula can give the corresponding location assignment in equation 6.12. The same formula can be used to calculate the individual values of i,j and k from a given location 'L' with the modifications made as per equation 6.13

$$L = i + (j-1)N_x + (k-1)N_xN_y \tag{6.12}$$

$$\begin{aligned} k &= 1 + \left\lfloor L \Big/ (N_xN_y + \delta) \right\rfloor \\ j &= 1 + \left\lfloor \left(L - (k-1)N_xN_y\right) \Big/ (N_x + \delta) \right\rfloor \\ i &= L - (k-1)N_xN_y - (j-1)N_x \end{aligned} \tag{6.13}$$

Here, $\lfloor \ \rfloor$ represents the floor function where the value inside the function is rounded off to the largest integer less than the value, δ represents a minimal value (typically 10^{-6}).

The observation well locations are arranged according to their corresponding weights, and the first three locations are selected to be the near locations of potential sources. Then, all the locations adjacent to the observation well locations and the cells next to the adjacent ones are arranged according to the probabilities assigned as per the Figure 6.11. The cells that are to be searched near each observation location are shown in the Figure 6.13. Each cell is then ranked based on its corresponding probability near each observation well location for the first three observation wells selected.

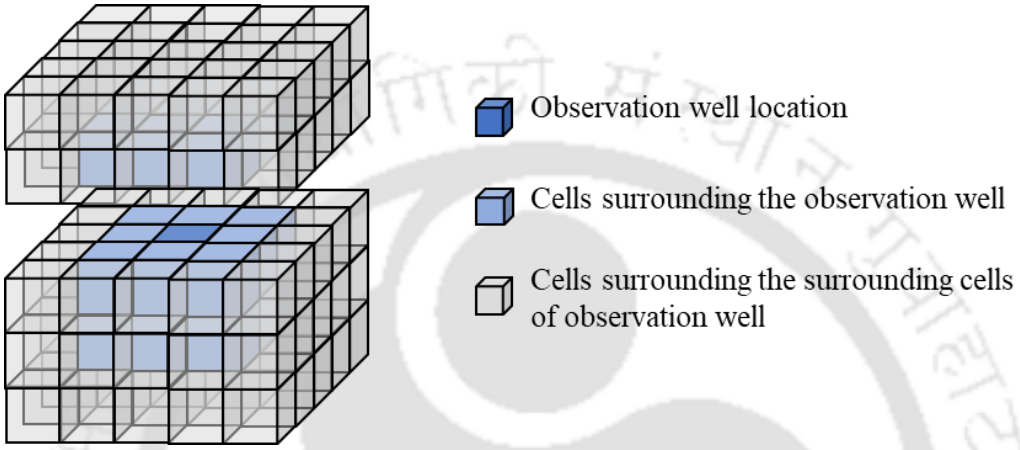

Figure 6.13: Cells to be ranked for potential sources near an observation well

The actual source locations for this problem are S1 - (5,15,1), S2 – (10,10,1) and S3 – (14,8,1), which can be translated to unique locations as S1 – 285, S2 – 190 and S3 – 154. The potential locations (with unique numbering 'L') after the extensive search methodology described are given in Table 6.3.

Table 6.3: Probabilities of the potential source locations for three sources

Source 1 – 285		**Source 2 – 190**		**Source 3 – 154**	
Location	**Probability**	**Location**	**Probability**	**Location**	**Probability**
645	0.0323	170	0.0246	153	0.0158
644	0.0323	150	0.0206	133	0.0128
665	0.0323	570	0.0144	151	0.0116
664	0.0323	171	0.0138	**154**	**0.0083**
244	0.0321	169	0.0122	152	0.0082
245	0.0318	151	0.0116	553	0.0069
264	0.0318	550	0.0110	132	0.0067
265	0.0288	149	0.0086	134	0.0064
646	0.0283	569	0.0079	533	0.0053
666	0.0283	571	0.0069	552	0.0038
684	0.0275	132	0.0067	173	0.0037
685	0.0275	**190**	**0.0057**	554	0.0035
246	0.0275	549	0.0053	191	0.0029

266	0.0247	551	0.0051	532	0.0028
686	0.0244	590	0.0037	534	0.0027
284	0.0233	189	0.0036	172	0.0022
285	**0.0225**	191	0.0029	174	0.0020
286	0.0201	589	0.0024	573	0.0019
225	0.0075	172	0.0022	572	0.0011
263	0.0043	591	0.0017		
223	0.0010				

It is to be noted that the best probability is not always the actual source location according to this methodology. This is evidence that the problem is a non-convex problem, i.e., multiple optimal solutions exist. The reason is that the location near the source location has a higher probability of being the potential source location as it can also produce similar concentration curves without a different value of source fluxes. The problem is discussed in detail in the following sections.

6.2.3 Modified genetic algorithm to identify the near-optimal sources

The genetic algorithm is well known for optimization problems that are mixed and discrete in nature. However, for this particular source identification model, the genetic algorithm needs to be modified to handle the source locations and flux with equal potential and also take the probabilities into account while selecting the source locations. The overview of the modified genetic algorithm is given in the Figure 6.14.

The modified genetic algorithm contains not only the basic operators such as selection, crossover and mutation but also the location probability operator to account for the initial solution's probabilistic locations. Also, the search space for the locations is narrowed down based on the initialization of locations performed previously in section 6.2.2. It is achieved by assigning a number to all the locations that have the probability of being searched for the optimal solution, as shown in Figure 6.15. Therefore, the lower and upper bounds for location – 1 is 1 and 21 respectively and that for location – 2 are 1 and 20 respectively and that of location – 3 are 1 and 19 respectively as there are only the same number of feasible search locations for each of the sources. For example, if location – 1 has identified the number 3 as the optimal location at the end of the algorithm, then the location identified is 665, corresponding to the mapped number 3 from the first part of Figure 6.15. The algorithm continues with the exact mapped locations till the end of the modified genetic algorithm.

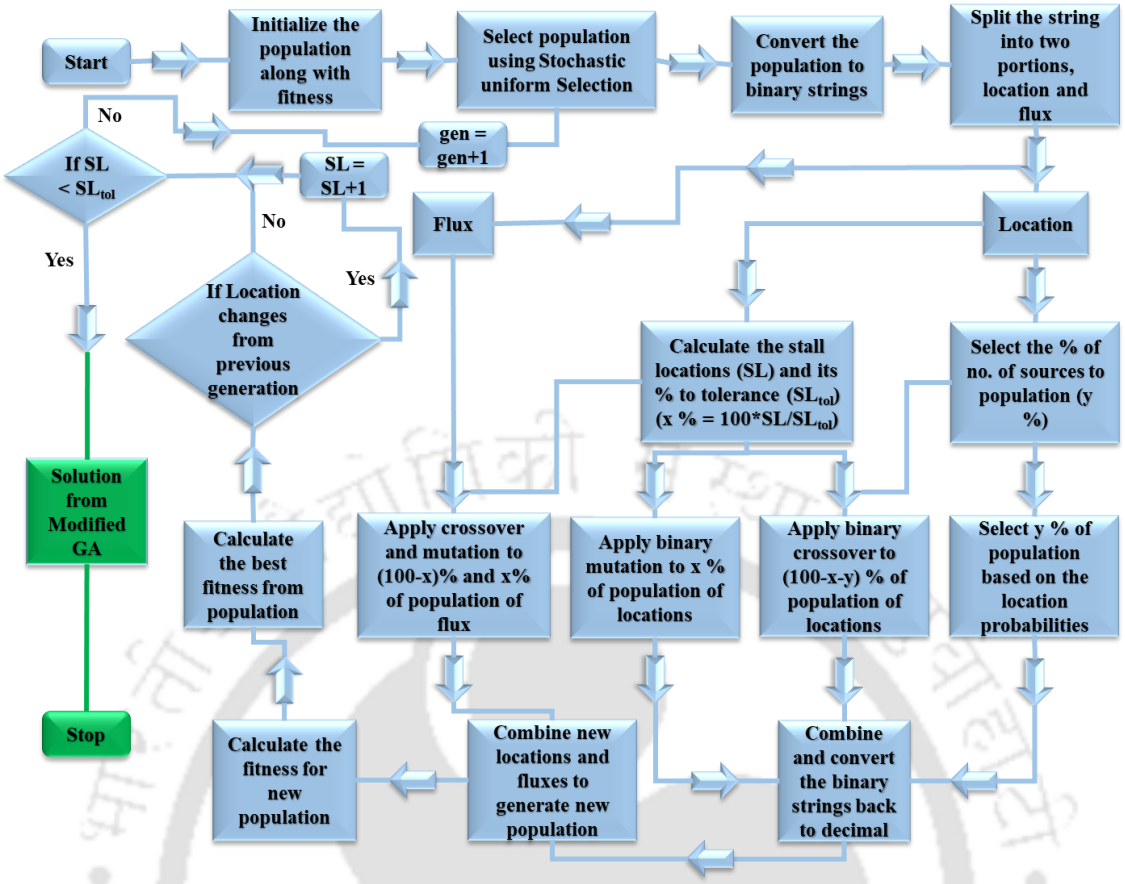

Figure 6.14: Overview of the modified genetic algorithm

Source Location – 1 (Mapping index to locations)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
645	644	665	664	244	245	264	265	646	666	684	685	246	266	686	284	285	286	225	263	223

Source Location – 2 (Mapping index to locations)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
170	150	570	171	169	151	550	149	569	571	132	190	549	551	590	189	191	589	172	591

Source Location – 3 (Mapping index to locations)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
153	133	151	154	152	553	132	134	533	552	173	554	191	532	534	172	174	573	572

Figure 6.15: Source location mapping for modified genetic algorithm

6.2.3.1 Basic genetic operators

The basic operators are the selection, crossover and mutation operators. The selection operator used for the algorithm is the stochastic uniform selection operator which is explained very well by the previous research (Chipperfield 1995; Thierens and Goldberg 1994). The crossover and

the mutation operators are also well defined for the binary strings (Deb 1999). However, a slight variation is adapted for this problem with the percentage of crossover and mutation. Generally, the population of genetic algorithm undergoes mutation more often initially and then as the solution converges towards optima, the mutation reduces, and crossover takes over the solution convergence. Here, I have adapted the strategy opposite to the one generally followed in traditional genetic algorithms for this model. The strategy here is that as the solution converges to a set of locations, there should be some certainty that the converged solution is near the global-optimal solution. To achieve this, we have increased the population percentage for mutation as the location converges. This is important in both cases where the optimal location is near global optima and not near global optima. If the location is near the global optima, the location converges at the same value till stall location tolerance (SL_{tol}). Also, if the location is not near the global optima, the location converges to a different near optimal location due to higher mutation and widened search space. This tweak for the modified genetic algorithm is also explained well in the previous research (Leichombam and Bhattacharjya 2019). However, the flux corresponding to the location is not needed to be modified as such, and it follows the same traditional genetic algorithm. The location and the flux are separated initially and then combined back after the operations are done on the population, as shown in Figure 6.14.

6.2.3.2 Probabilistic location operator

The probability of the locations being the actual source locations from the initialization of the population should be incorporated into the algorithm somehow. Therefore, a new operator is introduced where a select percentage of the population is selected (y%), the locations part of the population is changed according to the operator, and the flux part is left the same.

$$pop = \left[lb + rand\left(ub - lb\right) \right] \tag{6.14}$$

Here, *pop* is the population; *lb* is the lower bound of the location; *ub* is the upper bound of the location; *rand* is a number between 0 and 1. The step function '[]' just rounds off the result to the nearest integer. The same exercise is done for each location with its respective lower and upper bounds considered for all the y% population. This operator randomizes the locations identified sufficiently to trigger a better location near to the global-optimal solution. This process is continued to a maximum number of generations (user-specified) or till the stall location tolerance (SL_{tol} = 50 in this case) criteria is met, and the final output of the modified genetic algorithm is obtained. The optimal solution obtained at the end of the algorithm is a

near-optimal solution and needs to be brought to the global-optimal solution by the following procedure.

6.2.4 Local location search algorithm to find the global-optimal solution

The solution from the modified genetic algorithm is always a near-optimal location with the corresponding flux values that are not optimized to the minimum error between observed and the simulated concentrations. This algorithm is derived from the same philosophy of the previous research (Leichombam and Bhattacharjya 2019), but some additional operators were added to the existing algorithm to accommodate the three-dimensional nature of the problem and to address the limited observation data.

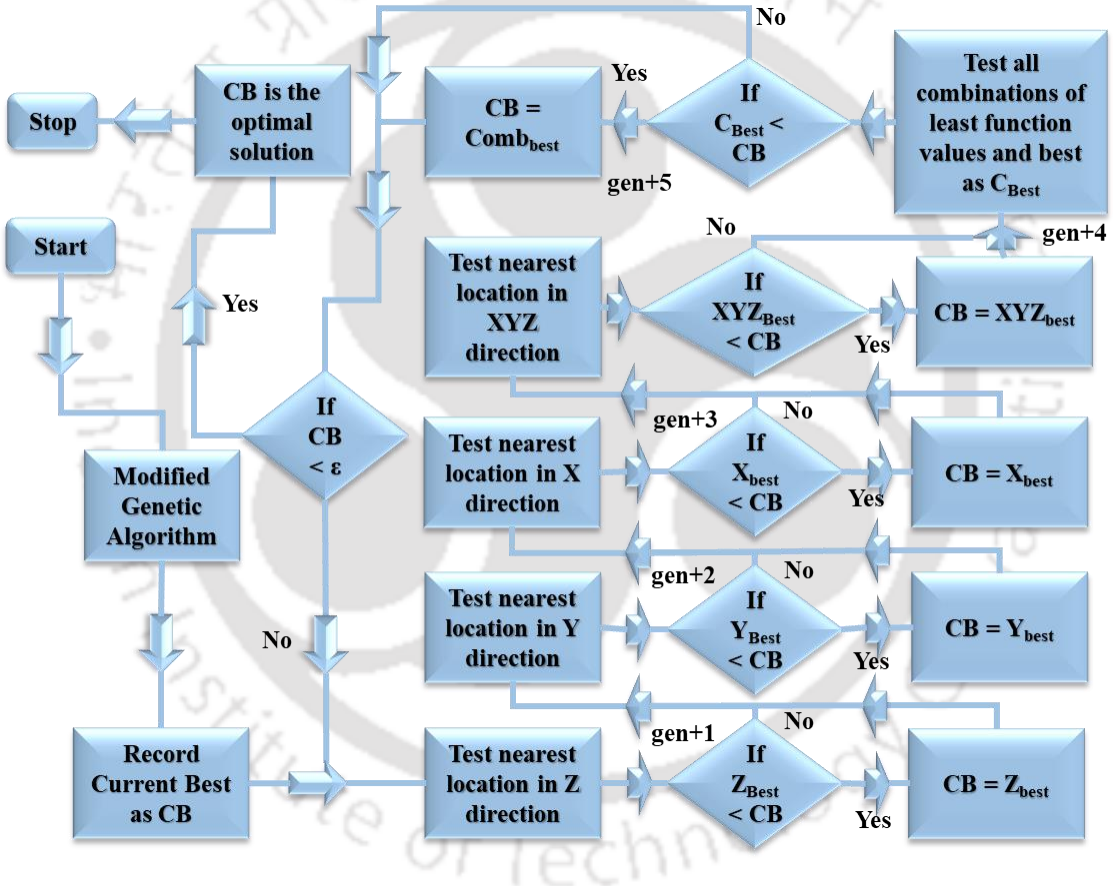

Figure 6.16: Overview of the local location search algorithm

The advantage of this algorithm is that the total time of leakage of sources into the aquifer is considered till the end of simulation time, i.e., the problem addresses the sources that are at three different locations for 50 days, and the simulation is with a time step of 10 days, and therefore there are 5 different leakage rates possible (one for each time-step of 10 days). The algorithm considers the simulation time also to be 50 days; therefore, the observation data is also for 50 days. The approach followed in this chapter is a new and unique approach to the

problem as the researchers have either assumed a single leakage possible at different locations or assumed the observation data to be available well after the leakage has been released into the aquifer. This new approach allows the identification of sources as soon as the observation well starts to detect concentration surges. A detailed overview of the algorithm is given in Figure 6.16.

The local location search algorithm starts after the result from the modified genetic algorithm is obtained, which is called the current best location (CB). The algorithm then continues with the three basic directional searches Z, Y and X directions, respectively. The current best (CB) gets updated after every directional search on the condition that the directional search resulted in a location with a lesser function value (as the objective is to minimize the error function). The subsequent search takes place in the corners of XY, YZ, ZX and XYZ directions which comprises a total of 20 locations in the same function, and the best criteria are checked to finalize the current best for the next phase of the search.

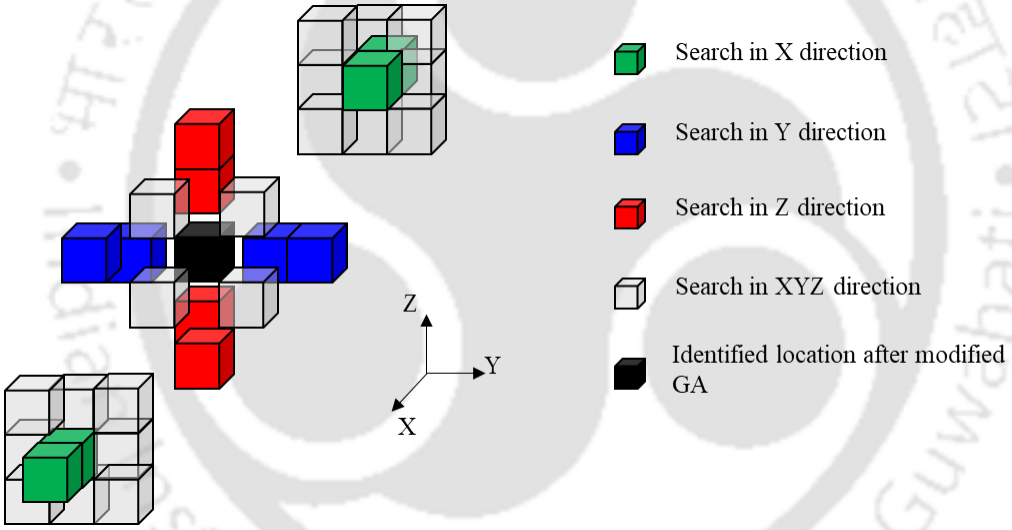

Figure 6.17: Locations to be searched in the local location search algorithm

The last phase of the search is the combination of least function values in the searches that have taken place till now. The typical locations that have been searched are listed for all the three source locations, and the combinations of all the possible locations are found out to be searched. After the best location is identified after the combination checks are performed, the tolerance criterion (ε) for the function value is checked (generally considered 10^{-6}) to finalize the optimal solution. The search spaces to be checked for each identified location for the local location search algorithm are represented in Figure 6.17. The colours green, blue and red represents the locations searched in X, Y and Z directions, respectively. The grey colour represents the locations searched in XYZ directional search.

6.2.4.1 Search in X, Y and Z directions

The result from the modified genetic algorithm is taken as the initial input for the search, and the directional search locations for the Z direction are identified. The search space includes two cells in the positive Z-direction and two cells in the negative Z-direction. After identifying the appropriate and feasible search spaces, the algorithm runs the gradient-based location search algorithm (*fmincon*) and minimizes the objective function till the tolerance criteria are met. The general tolerance criteria include the first order optimality condition (generally taken to be 10^{-2}) and a maximum number of iterations of 150, and maximum function evaluations of the number of variables multiplied by 200 (which is 3000 for this case). After the best location is identified in the Z-direction search, it is named as Z_{best} and compared with the current best (CB). If the condition that $Z_{best} < CB$ is satisfied, then the current best (CB) is replaced with Z_{best} or left as the previous value in case the condition is not satisfied. The algorithm then proceeds to the Y-direction search with the same rules and the X-direction search. The locations to be searched for a given location from modified GA are given in equation 6.15. The terms X_s, Y_s and Z_s represent the search spaces for the three principal directions. After the three algorithms are executed, the current best is taken as the final output.

$$\begin{aligned} X_s &= \begin{bmatrix} L+1 & L-1 & L+2 & L-2 \end{bmatrix} \\ Y_s &= \begin{bmatrix} L+N_x & L-N_x & L+2N_x & L-2N_x \end{bmatrix} \\ Z_s &= \begin{bmatrix} L+N_xN_y & L-N_xN_y & L+2N_xN_y & L-2N_xN_y \end{bmatrix} \end{aligned} \qquad 6.15$$

6.2.4.2 Search in XY, YZ, XZ and XYZ directions

The current best obtained from the previous step is the input for this step, and the search is continued for the best location among all the combined XY, YZ, ZX and XYZ locations which are represented in the grey colour in Figure 6.17. These locations are taken as inputs, and the gradient-based search explained in section 6.2.4.1 is performed for all locations in this step. The locations that are to be searched for optimal location is given in equation 6.16. The first two rows of the matrix in equation 6.16 represent the XY direction, the third and fourth rows represent the XZ direction, and the fifth and sixth rows represent the YZ search locations. These combinations are obtained simultaneously from one incremental or decremental search in two directions. The last four rows represent the XYZ locations obtained from the combination of all three directions with one increment or decrement in each direction. The best location obtained after the search is done called XYZ_{best}. The objective function value for this solution is compared to the function value of the current best solution, and then the current best (CB) is

updated if the function value is less than that of the CB solution. The solution CB from this step serves as an input for the next step.

$$XYZ_s = \begin{bmatrix} L+N_x+1 & L+N_x-1 \\ L-N_x+1 & L-N_x-1 \\ L+N_xN_y+1 & L+N_xN_y-1 \\ L-N_xN_y+1 & L-N_xN_y-1 \\ L+N_x+N_xN_y & L-N_x+N_xN_y \\ L+N_x-N_xN_y & L-N_x-N_xN_y \\ L+N_x+N_xN_y+1 & L+N_x+N_xN_y-1 \\ L+N_x-N_xN_y+1 & L+N_x-N_xN_y-1 \\ L-N_x+N_xN_y+1 & L-N_x+N_xN_y-1 \\ L-N_x-N_xN_y+1 & L-N_x-N_xN_y-1 \end{bmatrix} \quad 6.16$$

6.2.4.3 Search the combinations of least function values

The function values less than the tolerance of 10^{-1} throughout the previous steps are recorded and used in this step. A set of 10 solutions (generally three times the number of sources) is taken, and the combination of locations is generated, including all the possible combinations of the selected solutions. The generation of a combination of locations is given in Figure 6.18.

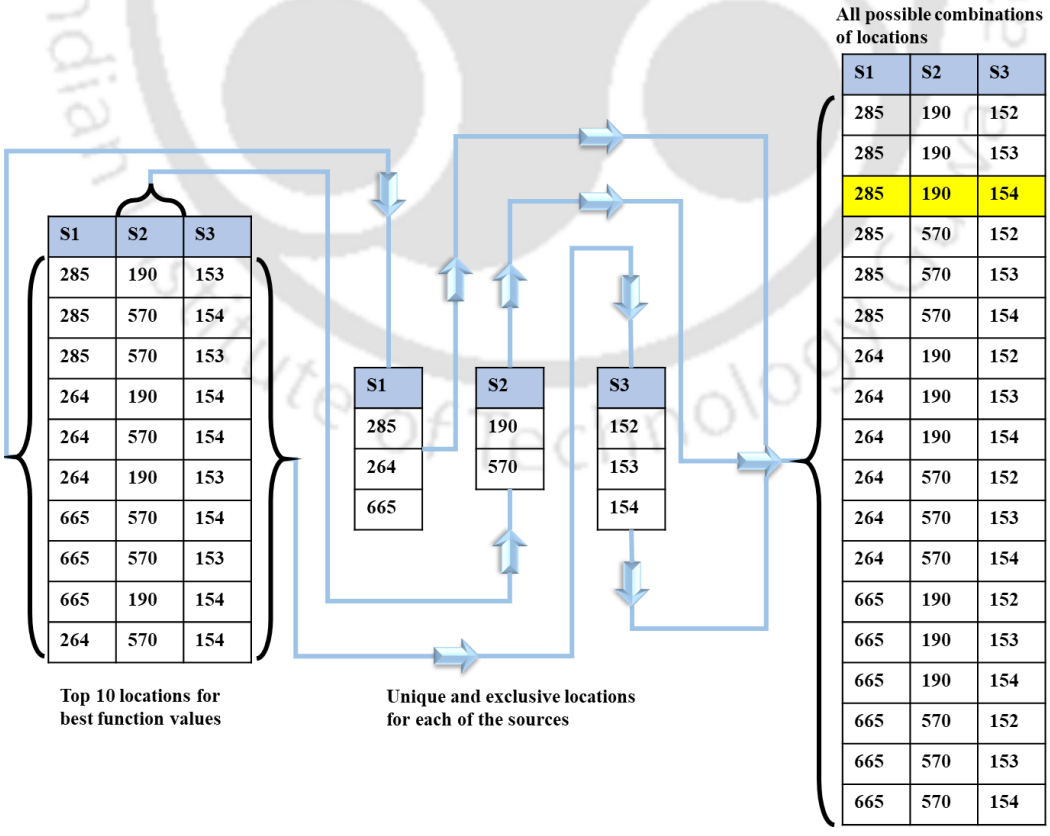

Figure 6.18: Generation of the combination of locations from the best function values

These combinations are all searched with the initial value of the fluxes of the current best (CB) solution, and the best solution of this step is given as C_{best}. The function value of C_{best} is compared with that of CB and the best function value is finalized as the solution for this step. The current best (CB) is then sent back to the same loop of these steps with X, Y and Z directional searches, XYZ search and the combination search until the tolerance criterion of function value of 10^{-6} is met. Once the tolerance criteria are satisfied, the algorithm terminates to give the final solution. The changing of locations for the whole algorithm from the modified genetic algorithm to the end solution is given in Figure 6.19.

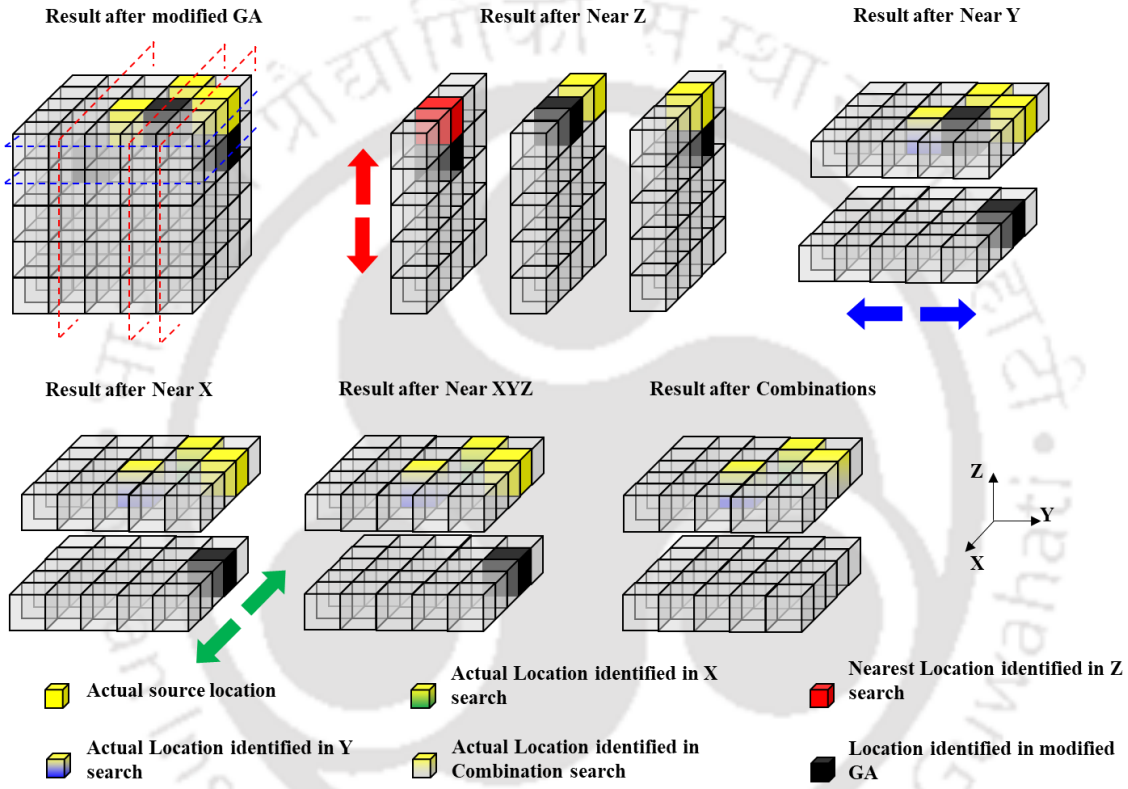

Figure 6.19: Migration of the identified source location through each step of the algorithm till the sources are identified

The locations in black colour are the locations identified in the modified GA algorithm. The actual source locations are represented in yellow colour. After the first step, the search in Z-direction, the result that is nearest to the source location is updated, which is represented in red colour, the result updated one of the identified locations, and the rest of the two locations are the same after the first step. The exact location is updated after the second step; search in Y-direction. The updated location is the actual source location. It is represented in a gradient of yellow and blue colours. Therefore, one of the locations is identified in this step. The next step, search in X-direction, has updated the second location to be identified in the actual source location. The identified source location is represented in a gradient of yellow and green colours.

The results identified in the XYZ search are the same as that of the previous step. However, the other location that is yet to be identified is finally identified in the last step of the combination search, which is explained in Figure 6.18. This whole procedure of local location search has multiple checkpoints of different search directions; therefore, all the source locations are usually identified with this method by the end of all the steps.

6.3 Results and Discussions

The hypothetical problem explained in Chapter – 3 (Rajeev Gandhi et al. 2017) is used to generate the results and check the algorithm's efficiency. The procedure explained in the methodology section 6.2 is applied to the same problem. The model is run for 10 trials to check the algorithm's efficiency. The relative errors expressed percentage for the estimated flux values for 10 trail runs using this model are given in Table 6.4. Here, the terms are abbreviated as AL – Actual Location, EL – Estimated Location and ASF – Actual Source Flux. The estimated locations are ideally identified using this model for all the 10 trail runs. The maximum relative error for the sources is 4.14 x 10^{-3} %, which is within the acceptable limit of error.

Table 6.4: Actual sources and estimated sources using the source identification model

AL	EL	Day	ASF (x 10^2) kg/day	Relative Error for Estimated Source Flux (%) x 10^{-3}									
				1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
285	285	Day-0	50	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
		Day-10	11	0.16	0.12	0.13	0.15	0.12	0.16	0.13	0.16	0.13	0.12
		Day-20	23.5	0.05	0.02	0.02	0.03	0.02	0.05	0.02	0.04	0.02	0.02
		Day-30	13.4	0.02	0.18	0.19	0.07	0.18	0.02	0.19	0.01	0.18	0.18
		Day-40	35	0	0.03	0.03	0.02	0.03	0	0.03	0.01	0.03	0.03
190	190	Day-0	10	0	0.03	0.03	0.01	0.03	0	0.03	0	0.03	0.03
		Day-10	19.6	0.08	1.95	1.95	0.88	1.95	0.08	1.95	0.18	1.95	1.95
		Day-20	40	0.12	2.87	2.86	1.29	2.87	0.12	2.87	0.26	2.87	2.87
		Day-30	0	0	0	0	0	0	0	0	0	0	0
		Day-40	0	0	0	0	0	0	0	0	0	0	0
154	154	Day-0	7	0	0.02	0.02	0.01	0.02	0	0.03	0	0.02	0.02
		Day-10	10	0	0.1	0.1	0.04	0.1	0	0.12	0.01	0.1	0.1
		Day-20	15	0.08	1.86	1.86	0.84	1.86	0.08	1.85	0.17	1.86	1.86
		Day-30	30	0.17	4.14	4.14	1.85	4.14	0.17	4.14	0.37	4.14	4.14
		Day-40	0	0	0	0	0	0	0	0	0	0	0

The best function value in each generation is represented in Figure 6.20 below. The function values are represented in a logarithmic scale to visualize the values with generations better. The mean function value in each generation is represented in a darker black colour.

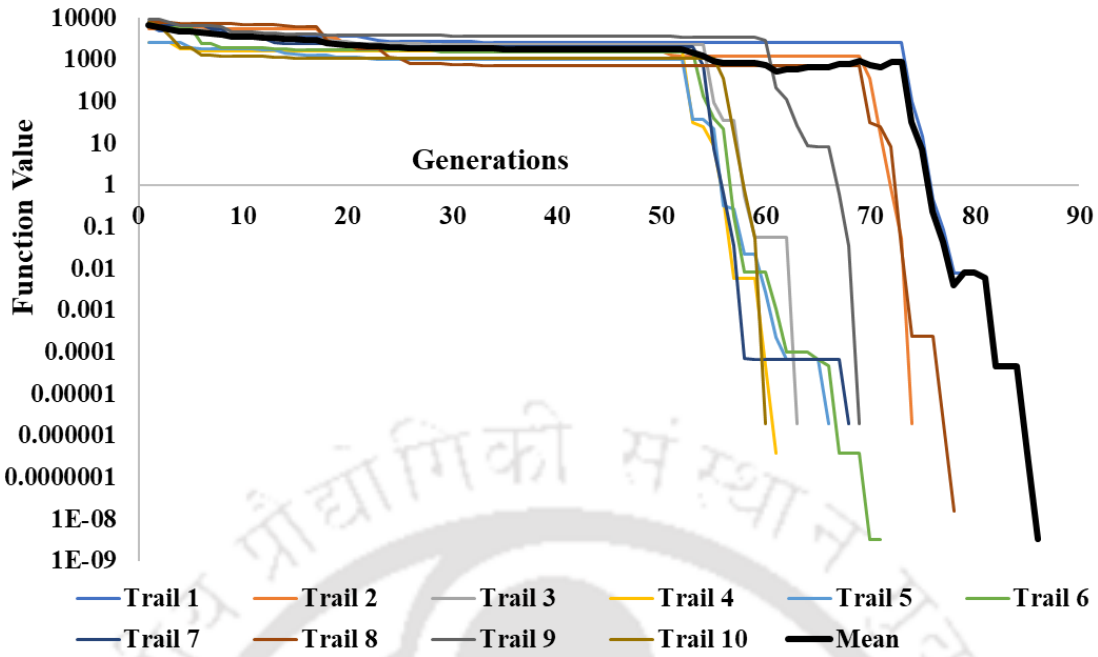

Figure 6.20: Best function value with each generation

The number of function evaluations is also calculated for each trail run and represented in Figure 6.21. The mean of the number of function evaluations for this algorithm is 118475.

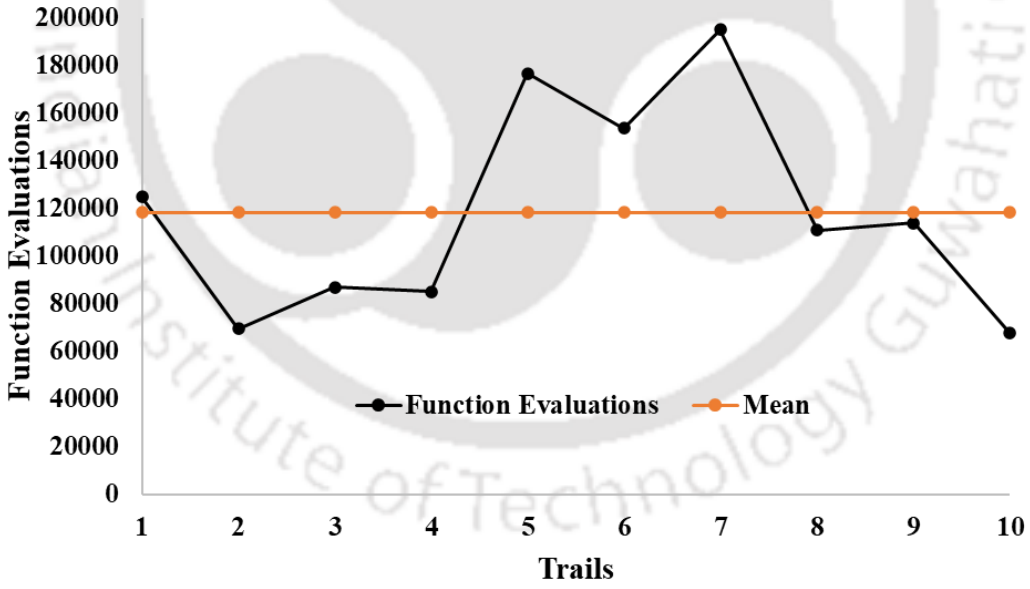

Figure 6.21: The function evaluations for each trail run

The number of function evaluations is higher than the previous algorithms used for source identification (Rajeev Gandhi et al. 2017). However, this algorithm has some advantages over the previous algorithm, as the location of sources is not known in the previous algorithm and the number of time steps for the sources is also fifteen as compared to the time steps for this

algorithm of five steps. Due to all these advantages, this algorithm offers much higher versatility for handling different problems. The change in the locations and flux values for all the three sources that are discussed in this problem are given in Figure 6.22, Figure 6.23 and Figure 6.24, respectively. However, these figures are represented with only one of the trails, as it is difficult to represent the figures for all the trails together. The figures are generated for the 10^{th} trial with the number of generations of 60 and the number of function evaluations of 67639. For this trial run, the modified genetic algorithm runs for 55 generations, and within one cycle of the local location search algorithm, the sources are identified. As each cycle of the local location search algorithm contains five generations, the total number of generations are 60 for this trail run. It can also be observed from the figures that there is a drastic change in the sources found in the last five generations as a result of the local location search algorithm. It can be observed from these figures that the locations that are identified in the modified genetic algorithm rarely change after the first 10 generations. This case is consistent throughout all the ten trail runs. The stall location tolerance is therefore taken as 50, as the algorithm terminates and shifts to the local location search algorithm when the locations identified don't change for 50 generations. The flux values, however, do fluctuate a lot in the process of these 50 generations. This is why the modified GA is efficient in only finding the near-optimal location in a short time; after that, the local location search algorithm will be much more efficient in finding the solution faster.

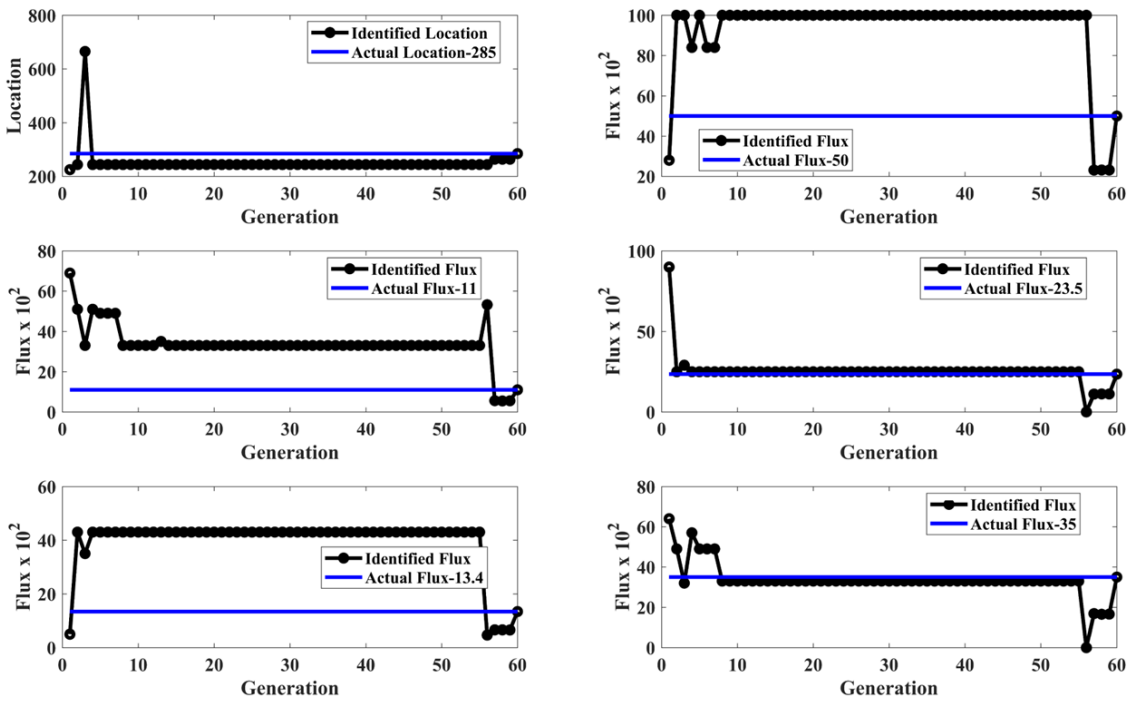

Figure 6.22: Location and flux values with generations for Source-1

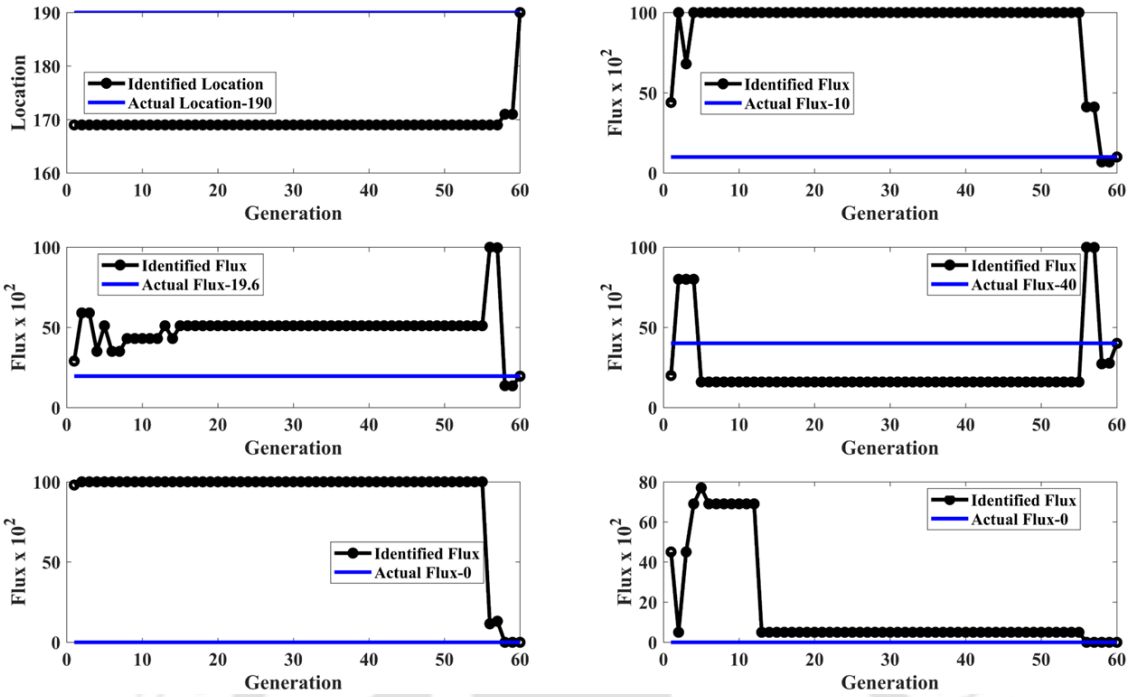

Figure 6.23: Location and flux values with generations for Source-2

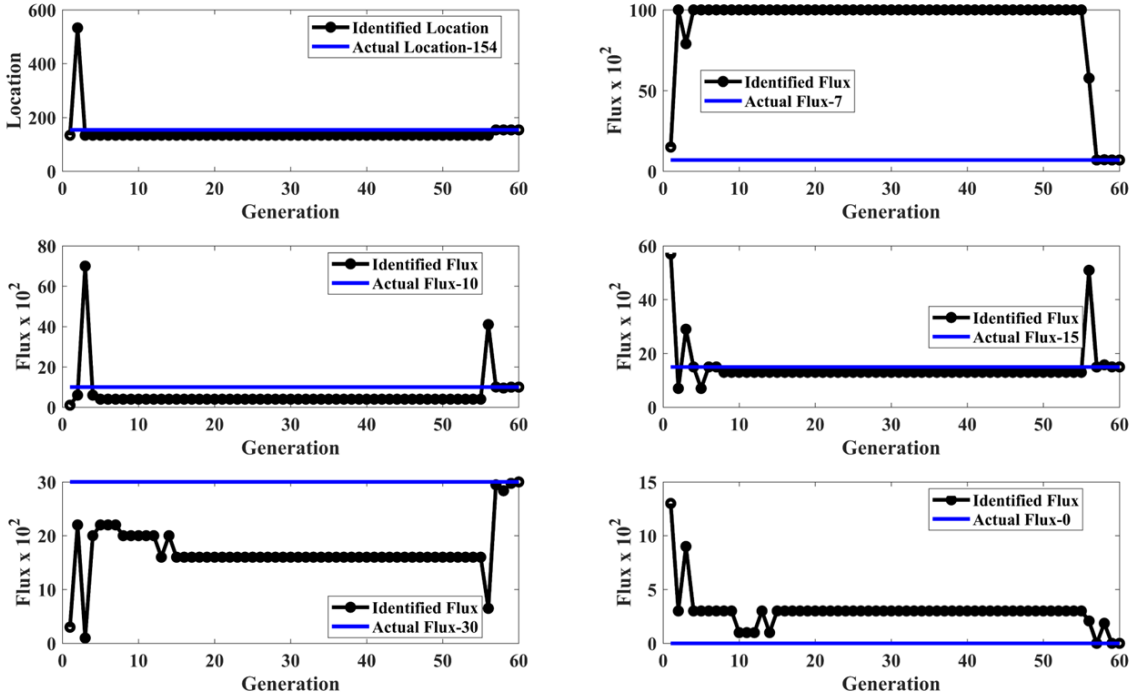

Figure 6.24: Location and flux values with generations for Source-3

6.4 Conclusions

The algorithm discussed in this chapter can handle the limited amount of observation data very efficiently. Also, there is no prior information on the number of sources or the number of time steps that the sources are actively needed for the algorithm to work. The algorithm assumes the

sources to be continuously active for all the time steps and with different flux values for every time step. If any source has been active for a time less than the assumed time, the algorithm generates zero flux value for those particular time steps. The algorithm is also consistent with the results generated and can be relied upon in the actual field experiments. The algorithm's consistency ensures the results are the same with every run, which increases the reliability of the engineers and experimentalists. The part of modified GA is supported with the initial approximation of the solutions based on observation data for this algorithm. The approach is novel in identifying the sources as valuable information is considered instead of randomizing the locations for the initial solution. The local location search algorithm is also added with two new operations that weren't included in the previous study (Leichombam and Bhattacharjya 2019). These operations of search in XYZ principal directions and the combinations of best locations ensure the convergence of the optimization model for the near-optimal solution to the global-optimal solution. However, the drawback of any source identification model is that the solution is not identified for the sources that don't reach the observation wells to generate any observation well data for the observed concentrations. The identical drawback prevails with this algorithm when there is a lack of observation well data to identify the sources. This algorithm is far superior to the existing algorithms in limited observation data, as it can identify the sources with the least amount of observation data possible to identify the source locations and flux values.

CHAPTER 7: VIRUS MANAGEMENT MODEL

7.1 Introduction

The virus pollution in groundwater stays as a plume, spreads due to dispersion, and advection transports it to longer distances. Thus, it has a wide range of spread in terms of pollution and once polluted, groundwater is challenging to use without proper treatment. Treatment of water by pump and treat method would be uneconomical as it would require more treatment time and several cycles (Bredehoeft 1994; Chadalavada et al. 2012b; Wagner 1995). It would be economical and more manageable if there were a method to treat the water without drawing it for treatment, and treatment is done within the aquifer itself. The relative cost analysis for both the methods of pump and treat and in-situ treatment is to be carried out, and a decision has to be made. There are many kinds of research carried out on the effectiveness of pump and treat methods (Boal et al. 2015; Huang et al. 1997; Kazemzadeh-Parsi et al. 2014). Also, particular research showed that the pump and treat method is relatively difficult to perform and practically impossible for hazardous wastes for various reasons (Voudrias and E.A. 2018). However, the situation in India and other developing countries is relatively different as the groundwater is generally used for domestic purposes directly without any particular treatment. In such a case, pumping and treating the aquifer free from viruses will be more time-consuming as well as more costly. The method of in-situ treatment is more effective than the pump and treat method. As such, a suitable management strategy should be adopted to decontaminate the aquifer. After knowing the location of the source and the rate at which the contamination is occurring at the source location, an optimal management policy can be designed, which is straight, dynamic, or stochastic in nature (Chang et al. 1992). The management policy design can be tackled as an optimization problem, which minimizes the concentrations in the aquifer to any desired level. The constraints of the optimization for the design of a management policy can be the cost of management and the time taken for remediation. In order to design an effective management policy, a better understanding of the transport of viruses in a saturated porous media and the factors affecting the virus transport is required. Therefore, a detailed study has been conducted on viruses and the primary factor that affects the concentration of viruses along with simulation models in the previous chapters.

Groundwater management models have been studied and categorized by many researchers (Ahlfeld and Heidari 1994; D.P. Ahlfeld 1986; Gorelick 1983, 1990). However, management

models have not been developed for virus transport in particular, as the transport process and behavior of the virus is highly dynamic with the conditions of the aquifer (Bredehoeft 1994). In general, a groundwater management model has four characteristics: (1) it is a stochastic process, where the primary source of uncertainty is that associated with the aquifer simulation models. The uncertainty associated can be with the assumptions involved in the estimation of parameters; (2) the decision variables considered highly influence the nonlinearity of the model; (3) it involves both continuous and discrete decision variables, i.e., it is a mixed-integer programming problem; and (4) it requires the solution of a set of (perhaps nonlinear) coupled partial differential equations describing groundwater flow and transport (Wagner 1988, 1995). Some researchers have presented a two-stage method of nonlinear weighted feedback to design aquifer remediation strategies under model uncertainty (Chadalavada and Datta 2008; Gorelick 1990; Wagner 1988; Whiffen and Shoemaker 1993), which is a coupled groundwater management model. Some researchers used differential dynamic programming to demonstrate the time-varying policy for pumping and benefits. Some groundwater management models are dealt as multi-objective optimization models (Bhattacharjya and Datta 2009). Any management model can be thus taken as a stochastic, dynamic, and mixed-integer programming problem (Chang et al. 1992). To manage the concentrations of viruses in the aquifer, there is a need to know the complete details of the concentrations over a while, and there should be accurate predictions of the concentrations in the aquifer in the future (Ahlfeld and Heidari 1994; Chadalavada and Datta 2008; Wagner 1995). So, there is a need to have the exact location of the virus sources and the information on the number of sources entering the aquifer. Therefore, a source identification model has to be developed prior to the development of the management model (Leichombam and Bhattacharjya 2019). The development of a source identification model is discussed clearly in the previous research (Rajeev Gandhi et al. 2017)

In this chapter, the removal of viruses like bacteriophage MS2 from an unconfined aquifer by injecting warm water into the aquifer has been attempted. The virus transport simulation model has been constructed as a three-dimensional model in MATLAB capable of varying the inactivation rates over time, as described in chapter 3. The management model is developed in MATLAB by linking the simulation models to the optimization model. The management model is solved as an optimization problem, which minimizes the total volume of injection required to carry out the management and is subject to the constraint that the maximum concentration in the aquifer shall not exceed the desired concentrations in the aquifer by the end of the aquifer management period. The locations of injections are decided based on a new probabilistic

approach, and injecting warm water at these locations according to some schedule is the optimal management policy.

7.2 Methodology

The methodology of this chapter aims to remove virus concentration in an unconfined aquifer and ensure the concentration of viruses is lesser than a specified threshold. The virus MS2 is reactive to temperature, as shown in previous studies (Schijven and Hassanizadeh, 2000). The groundwater temperature affects the rate of inactivation of the viruses and the rate of adsorption. Typically, groundwater stays at an even temperature throughout the area, so the concentration of viruses might not change much over time. In this chapter, the property of the viruses to change the inactivation rate and adsorption rate is used to decrease the concentration of viruses at a faster rate. The management period depends on the threshold to which the concentration of viruses is to be maintained in the aquifer. The methodology in which the concentration of viruses is reduced is given in the flowchart given in Figure 7.1

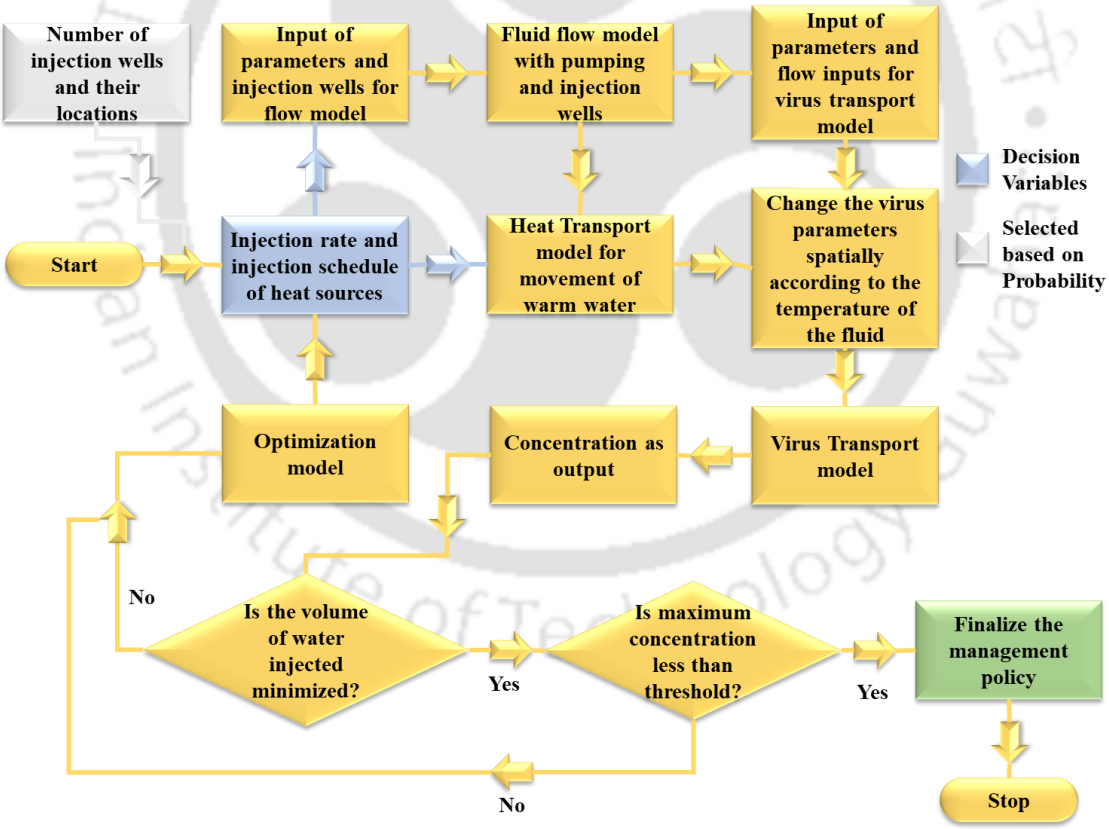

Figure 7.1: Overview of the methodology.

The methodology can be explained as follows. The injection rates and schedules of injection are selected based on the aquifer's concentration conditions before management. At the specified locations, the injection wells are created in the model, and the arbitrary variables are

used as inputs to the flow and heat transport models. After the heat spread in the aquifer is acquired from the model, the inactivation rates are altered in the aquifer as a function of temperature. Then, the virus transport model is run to find the maximum concentration in the aquifer after the specified management period. Suppose the concentration is within the specified threshold. In that case, the constraint is satisfied, and the arbitrary injection rates are then minimized to inject the least possible volume of warm water in the aquifer. To execute the process in this methodology, external input is needed, which gives the number of management locations and the number of schedules along with the actual locations of injection wells in the aquifer. The locations of the injection wells are selected based on a probabilistic procedure, explained in the subsequent sections. However, the number of locations and the number of schedules of injection are selected by the user.

In this methodology, the injection rates and the injection schedule of heat sources are taken as decision variables. The decision variables are the ones that change from the inverse optimization model based on the response from the aquifer. These decision variables of injection rates and schedule are the inputs for the flow model, which gives the flow of groundwater in the aquifer by solving the continuity equation. The flow model also provides the velocities of water at all the regions in the aquifer. These velocities serve as input for the transport model. The transport of warm water from the injection wells is then modeled using the heat transport model by solving the heat equation. The output is the temperature spread all over the aquifer with different temperatures in different regions. Then, the virus transport parameters of inactivation and adsorption rates are modified spatially and temporally all over the aquifer. These parameters serve as inputs for the virus transport model. Then the advection-dispersion equation is solved in the virus transport model resulting in the spatial and temporal variation of concentrations in the aquifer. The whole process is explained in chapter 3 in detail. The maximum concentration is then considered for the threshold. The volume of water injected is calculated from the injection rates and the injection schedule from the decision variables. The volume of water is the objective function, and the maximum concentration is the constraint for optimization. The objective function is minimized using the classical optimization method. As the objective is minimized and the constraint is satisfied, the management policy of injection rates and injection schedules are finalized. The mathematical models required to execute the methodology are categorized as simulation and optimization models. The simulation model consists of a flow model, heat transport model, and virus transport model, which is the main focus of work in chapter 3. The optimization model is explained in the section of the management model in the subsequent sections. However, before the execution of the

management model, the injection well locations need to be finalized. The method to identify the best injection well locations is given in section 7.2.1.

7.2.1 Selection of injection wells

The well locations of injection through which treatment of the aquifer is done are termed as injection wells. The problem discussed in Chapter 3 is used for the simulation and analysis of the selection of injection wells. Two procedures were adapted for the selection of probable injection wells.

7.2.1.1 Probability space for injection wells

The first procedure is that the location of injection wells is selected based on the response of the aquifer from the injection. A location of injection well is selected and the warm water of temperature 373 K is injected at that location with a constant discharge of 100 m^3/s for 10 days. After the scheduled injection period, the concentration of aquifer is then calculated and the sum of concentrations is given as an output of the injected location. Therefore, the function as an input and output are the location linked to the concentration of aquifer. This function is used to calculate the normalized probability of response to the injection of warm water at different locations. This is given in equation 7.1

$$f(x,y) = \sum_{x_1=1}^{x_1=N_x} \sum_{y_1=1}^{y_1=N_y} C(x_1, y_1, z=1, t=t_{end}) \qquad 7.1$$

The function of injection location $f(x,y)$ represents the summation of concentrations of all the spatial locations in the first layer $z=1$ and the end time step $t=t_{end}$. This function is used to find the normalized probability of effectiveness of a location to be an injection well location as given in equation 7.2. This probability has an inverse relation to the function described in equation 7.1. Therefore, the normalization of the probability density function is done accordingly. The probabilistic space obtained is given as a scatter plot with probabilities representing the circles' size and color in Figure 7.2(a).

$$p(x,y) = \frac{\left(1/f(x,y)\right)}{\sum_{x=1}^{x=N_x} \sum_{y=1}^{y=N_y} 1/f(x,y)} \qquad 7.2$$

The second approach is straightforward, i.e., to find the concentrations of viruses in the aquifer before the treatment starts and convert them into a probabilistic space. The normalization of the

concentrations can be carried out in such a way that a probability density function of a cumulative probability of 1 can be generated through them. The formulation given in equation 7.3 shows the generation of the probabilistic space. The time $t = t_1$ represents the time at which the management of viruses is to be started. The probabilistic space such obtained is given as a scatter plot with probabilities represented as size and colour of the circles in Figure 7.2(b).

$$p(x,y) = \frac{C(x,y,z=1,t=t_1)}{\sum_{x=1}^{x=N_x}\sum_{y=1}^{y=N_y} C(x,y,z=1,t=t_1)} \qquad 7.3$$

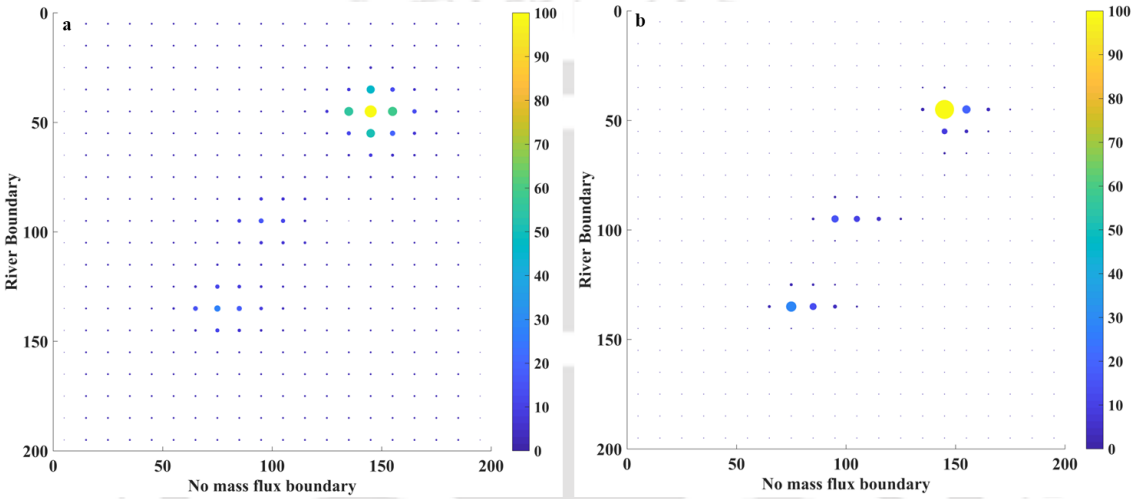

Figure 7.2: Scatter plot of probabilistic space for injection well method (a) and the starting concentration method (b).

The two methods take very different approaches but arrive at almost similar conclusions when the location of injection wells is to be considered. The probable injection wells to be selected are based on the ranking and number of peak concentration locations. The aquifer's treatment will be considered complete only when the maximum concentration of virus in the aquifer is within the permissible limits of safety. Therefore, the injection needs to be done based on the number of peaks and the corresponding ranks to the peaks.

7.2.1.2 Ranking of probable injection locations

Over the discretized grid of this probabilistic space, a 3×3 box is laid over to check whether the center of the 3×3 box has a maximum value over the rest eight boxes. The box to be laid over is shown in Figure 7.3, along with the contour plot of probability space with grid lines. It can be seen that the maximum value of this box, when situated in the center, can be recognized as peaks in the probability space. Such peaks are identified over the probabilistic space. Here in this particular problem, three peaks can be seen.

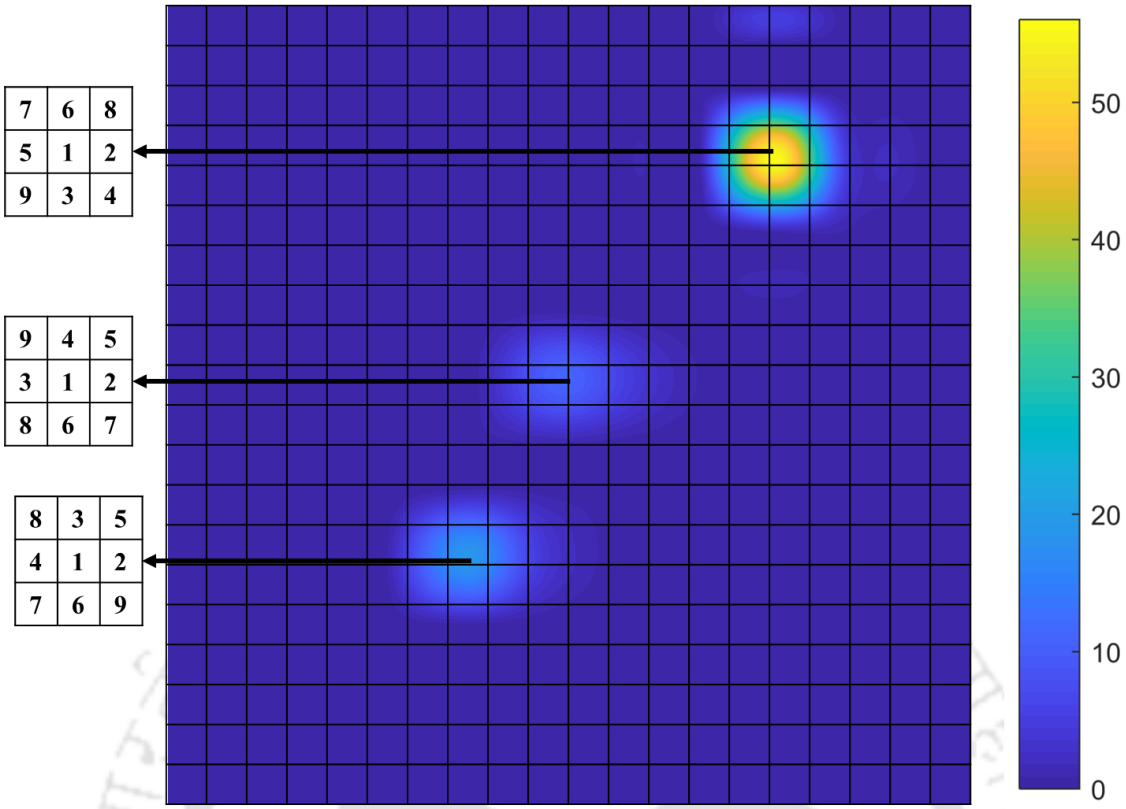

Figure 7.3: Identification of peaks in the plot with a box layover method

Identifying peaks in the initial condition of the problem is crucial for selecting the location of injection wells. Here, as there are three peaks in the probabilistic space, a minimum of three injection wells should be needed to remove viruses from the aquifer. If the number of injection wells selected is less than the number of peaks, the aquifer will not be decontaminated from the viruses at one or more locations. Hence the maximum concentration of viruses will never arrive below the threshold. When a number of injection wells need to be selected, a pool of probabilistic values is created containing a location near each identified peak. Therefore, the pool can contain a minimum of three locations (one arriving from each peak).

For selecting the injection wells more than the number of peaks, a ranking system is given to each of the cells in the peak box (the peak box has a rank one at the center, showing that it has the highest probability). The ranking for each box cell is given based on the probability corresponding to that location. To select the fourth injection location, all the cells with rank two are collected and sorted from maximum to a minimum. The cell with maximum probability under rank 2 is the fourth location of injection, the second cell of rank 2 is the fifth location of injection, and so on. The ranks of each of the boxes are given in Figure 7.3. This way, we can ensure that each new injection well is being selected based on probability and the distribution of sources' locations. If the selection of cells is based on probability alone, then the location

where there is the highest peak will have more injection wells, and the ones with lesser peaks might end up not having even a single injection well. By taking this approach, the locations of injection wells are not considered a decision variable for the design of the optimization model.

7.2.2 Management Model

The management model is an optimization model that minimizes the objective function subject to constraints. The objective function's mathematical form and its constraint are given in equations 7.4 and 7.5.

$$f(Q,T) = \sum_{l=1}^{nL} \sum_{s=1}^{nS} Q_l^s T_l^s dt \qquad 7.4$$

$$g(Q,T) = \max\left(C\left(x, y, z, \mathrm{t}_{end}\right)\right) - Th \qquad 7.5$$

Here, Q represents the discharge at a particular location *'l'* and in a schedule *'s'*, T represents the number of time steps in that particular location *'l'*, and for a particular schedule *'s'*, *dt* is the discretized time step for the flow and transport models, *C* represents the simulated concentration at all the locations of the aquifer, but at a time step at the end of the management period, *Th* represents the threshold to which the concentration in the aquifer is permitted for the whole management period.

To execute this management model, mathematically, the constraint is added to the objective function with a large multiplier 'M', and a new objective function is created, which is minimized. The new objective function generated with the objective and constraint is as shown in equation 7.6.

$$h(Q,T) = f(Q,T) + M \times g(Q,T) \qquad 7.6$$

The number of locations (*nL*) and schedules (*nS*) are decided based on the user's requirements. The pumping schedule is the amount of time for which the pumping is to be carried out at each location.

7.2.3 Optimization Model

The optimization of the new objective function, shown in equation 7.6, is carried out in this model. The optimization procedure for this problem is a purely classical approach done using *fmincon* function of the MATLAB optimization toolbox. The decision variables for optimization are the discharge of injection well and the time to which the discharge must be maintained. For example, if three locations are selected, and each location can have four

schedules, then the total number of decision variables will be 24, as explained in Figure 7.4. The total time schedules will be 4 x 3, i.e., 12 schedules, and each schedule needs a rate of injection, so 12+12 = 24 decision variables are required for this problem. The simple formula for the number of decision variables would be 2 x *nS* x *nL*, where *nS* is the number of schedules, and *nL* is the number of locations, which are 4 and 3 in this example. The user decides the lower and upper bounds of the time schedule (T) and the injection rate (Q) at each location.

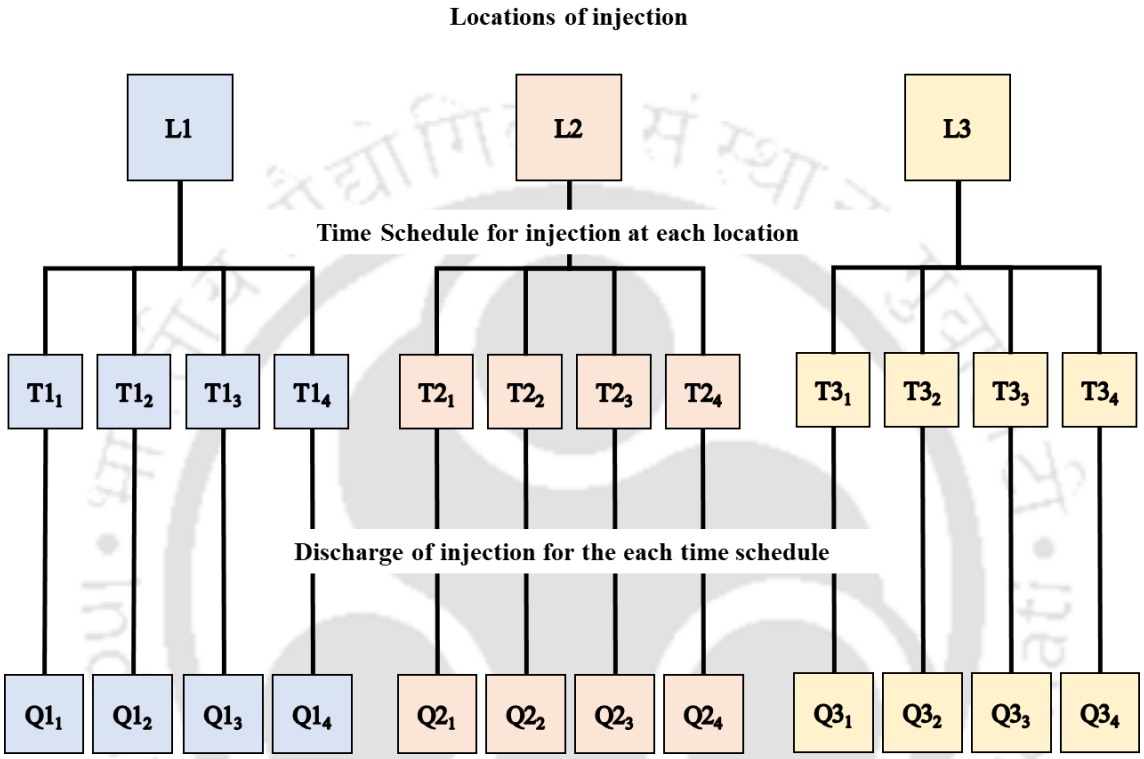

Figure 7.4: Decision variables listed for three locations and four schedules

Here, the schedule is a discrete variable, and the discharge is a continuous variable. The function *fmincon* does not directly allow discrete variables to be a decision variable. Thus, we have rounded off the schedule variables to the nearest integer to account for the mixed variable problem. The classical approach to solving this optimization model is a gradient-based approach, which requires a start point. The initial point is given to be the upper bounds of the decision variables. The termination condition is when the average step length of the variables reaches a value of less than 10^{-5}. There was a final fine-tuning done to the result of *fmincon* algorithm to tweak the sensitivity of each of the selected locations. The fine-tuning algorithm is a pattern-search-based algorithm used in previous research for the source identification of viruses (Rajeev Gandhi et al., 2017).

7.3 Case Scenarios

The case scenarios are all hypothetical test cases that have been derived from the industrial effluent problem from the source identification model of our previous research (Gandhi and Bhattacharjya 2020b; Rajeev Gandhi et al. 2017). These problems are further categorized into three categories based on the type of solution suggested. The first category of solution comprises the case of selecting the injection well anywhere in the aquifer to reduce the concentration of viruses. The second solution category is based on the case where the injection well locations are selected on the existing wells in the aquifer that are distributed anywhere. The third solution category is based on the case where the injection wells are selected based on the existing observation wells in the aquifer. The study area considered is 4-hectare as described in the two categories (Rajeev Gandhi et al. 2017). The aquifer parameters were adapted from (Molson et al. 1992; Palmer et al. 1992).

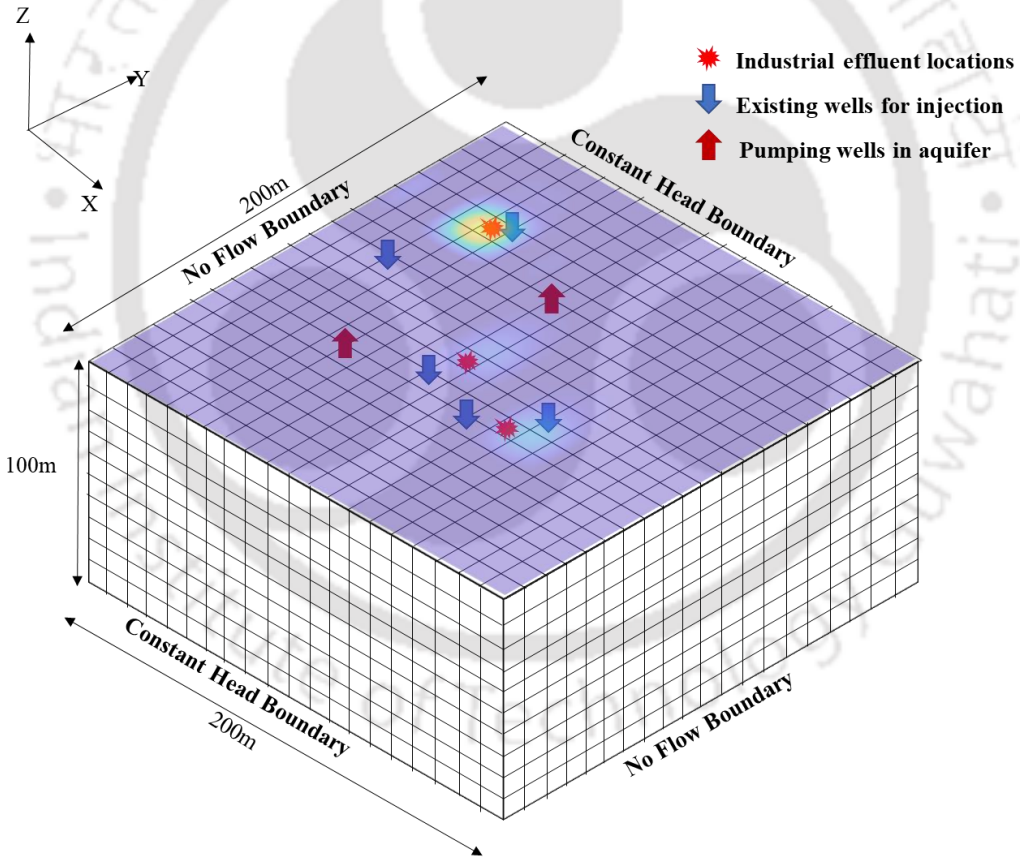

Figure 7.5: The study area for the industrial effluent problem shows the location of sources and the initial simulation condition of the aquifer before management.

The problem is when the viruses are released at different source strengths at different times as slug loads into the aquifer. The case was already discussed, and source identification models were prepared for such cases in our work (Gandhi and Bhattacharjya 2020b; Rajeev Gandhi et

al. 2017) and in chapter 3. The industrial effluent problem is solved using two methods, which fall under solution categories 1 and 2. Three locations of virus releases in the aquifer come from three potential industries. The management of the viruses starts after 50 days from the initial release of viruses. The simulated condition of the aquifer from the initial release of viruses to the final state before the management commences (i.e., 50 days) is given as an initial condition as an overlay on the grid in the Figure 7.5.

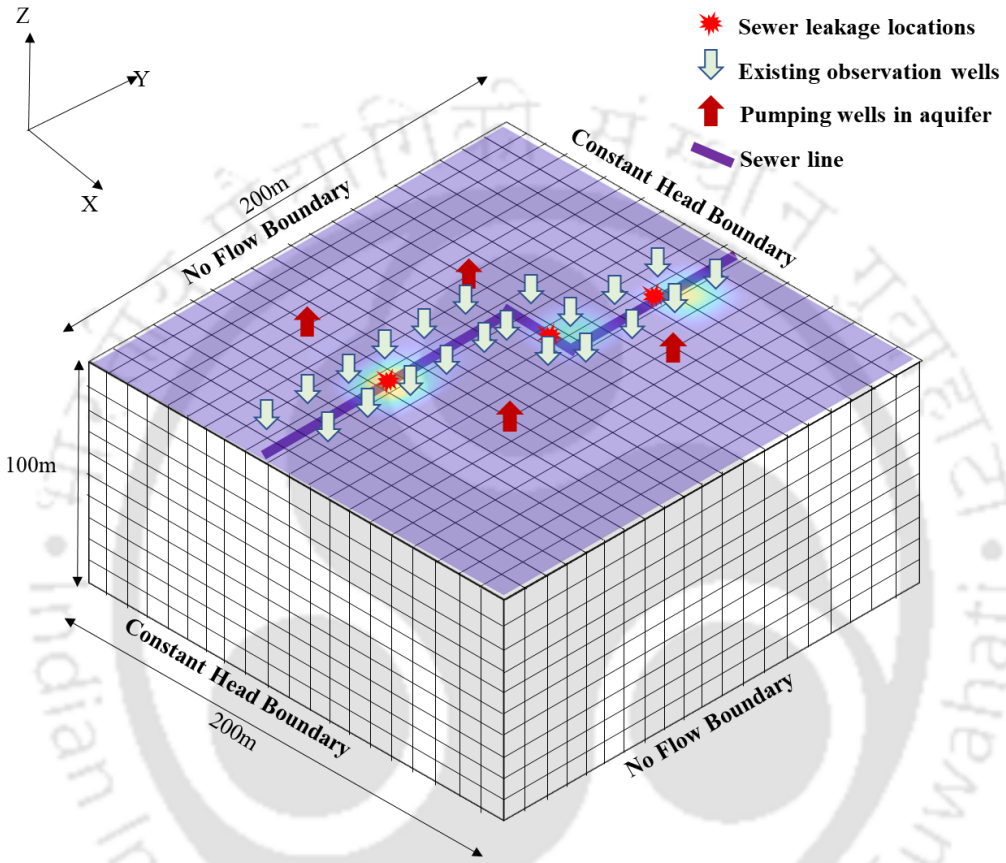

Figure 7.6: The study area for the sewer leakage problem shows the location of sources and the initial simulation condition of the aquifer before management.

The study area selected for solution category 3 is defined in Figure 7.6. This solution category is the one where the injection wells are to be selected from the existing observation wells in the aquifer. The sewer line, three leakage locations, four pumping wells for this problem, and all the probable observation wells (21 in number) along with the initial condition of the aquifer before the management commences are given in Figure 7.6. Here the initial conditions of the aquifer flow model have been tweaked a little just to show that this methodology works for any condition for the selection of injection wells from the set of observation wells. Here, the initial flow model has constant head boundaries of 98 m (left) and 92 m (right) in contrast to the industrial effluent problem's flow model, where the right side's head is 96 m. This increases

advection in the flow model resulting in faster movement of the viruses than the industrial effluents. This would result in a slightly different set of injection wells that are to be selected as the probability space changes due to the higher advection.

7.4 Results and Discussions

The problems are solved by three solution categories with slightly different user inputs in each category. The solution procedures in each category are described in detail in the subsequent sections. The common thing for all the solution categories is injecting hot water. As it is challenging to control the water temperature, we have limited the temperature to barely boiling temperature for water, i.e., at 100^0C or 373 K. All the simulations for these two categories are done using the discrete-time step of 5 days. The management model is to be run by injecting warm water in these given locations. The period of management was capped at a maximum of 20 days. However, using injection schedules ensures that the management period need not be 20 days but can be less than the given time.

7.4.1 Solution Category 1

This solution category is selected in places where there is a possibility of digging an injection well wherever necessary, i.e., places far away from the cities and urban areas. As there is a possibility of digging observation wells anywhere, the whole of the probabilistic space can be used to select injection wells. Initially, the number of locations (*nL*) is selected by the user, and the number of schedules (*nS*) of injection is selected too. Then, the number of injection wells is selected as per the methodology proposed based on the probabilistic selection of injection wells. Here all the probability space is feasible. Therefore, the ranking procedure described in the methodology section is used as it was described. The approach applied to give the ranking to each of the peaks is given in Figure 7.7. Here, the grid arrangement for the probability wave function based on an initial condition is shown on the left, followed by the probability values for all three peaks. Then, the ranking is given based on the descending order of probability and is shown in Figure 7.7. The number of locations selected '*nL'* should be greater than or equal to 3 (as there are three peaks for the initial condition). Therefore, the lower bound for '*nL*' is 3, and the upper bound was limited to 6 for this problem. Here, (i,j) in Figure 7.7 represents the position of the cell at which the probability is shown.

It can be seen from Figure 7.7 that the ranking of the cells around the peaks doesn't follow a single trend. It depends on the position of the virus release, the direction of flow, and the strength of the release of the virus. Now, to select the number of locations for injection wells,

we categorize the ranks concerning peaks and choose the location of each peak based on rank. The chosen list of locations when the number of locations is 3, 4, 5, and 6 are given in detail in Figure 7.8. Also, the selection criteria based on the ranking are given in the same figure by a gradient from the top right to the bottom left, showing that the darker regions are to be selected first, and the lighter regions are to be selected last. This ensures that each time a selection is made, the hierarchy is followed based on the probabilities and the distribution of selected locations. For the number of locations' *nL*' equal to three, the selection is based on rank 1 for each of the peaks. For the number of locations' *nL*' equal to four, five, and six, the selection is made based on peaks for rank 2.

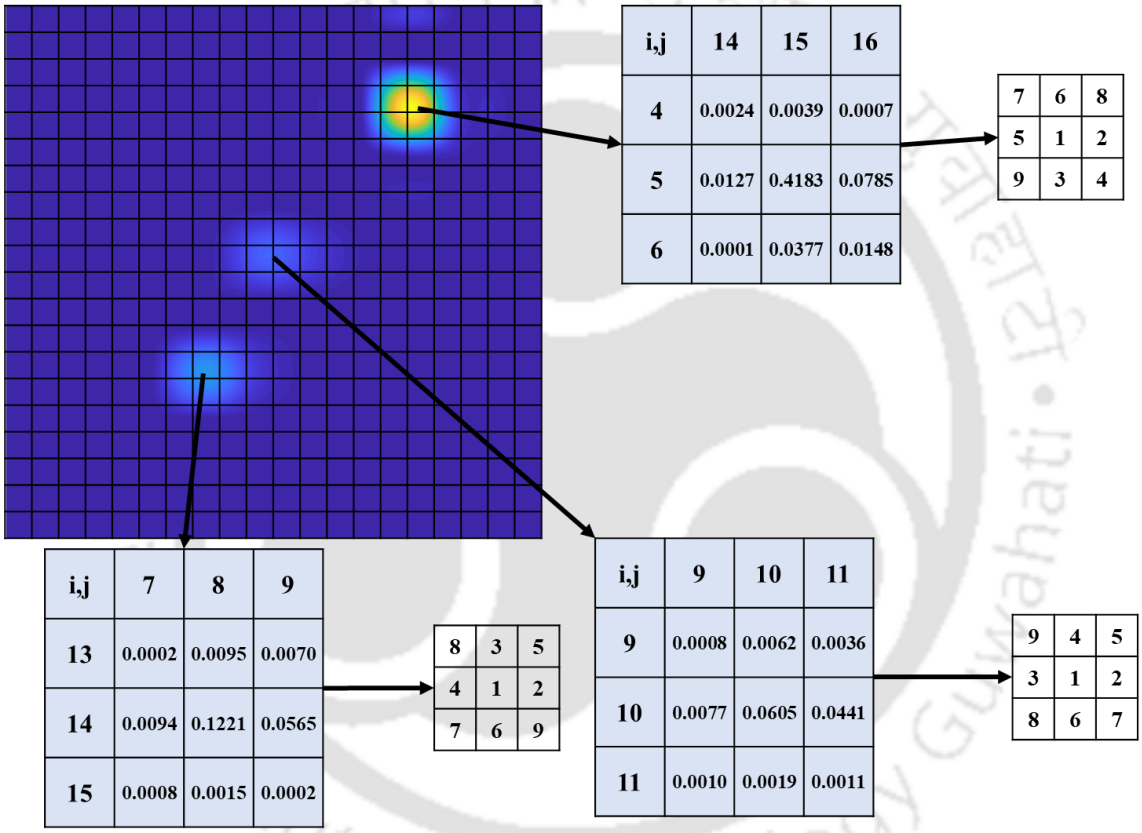

Figure 7.7: Probability space for each of the peaks in solution category 1

Let the position of each location in Figure 7.8 be represented as rank and peak position as (*p*,*r*). For example, when *p* is 1 and *r* is 2, the location is (5,16), which is at that position. To define the hierarchy, a simple mathematical formulation can be made for the basis of selection to be made given in equation 7.7.

$$nL = p + (r-1) \times p_{\max}$$
$$rL = 1 + \left\lfloor \frac{nL}{(p_{\max} + \varepsilon)} \right\rfloor ; pL = nL - (rL-1) \times p_{\max} \quad 7.7$$

Here, $\lfloor \ \rfloor$ it represents the greatest integer function or floor function where the number inside the brackets is reduced to the greatest integer less than the value inside the brackets. ε is a minimal number greater than zero. Using this equation, one can select a unique value of '*pL*' and '*rL*' for a given value of '*nL*'. The rest is to select all the locations up to the point by *pL*, *rL*. For example, if *nL* is 5, then *pL* and *rL* work out to be 2 and 2. Therefore the selection is made up to 14,9 from the first i.e., 5,15. The selected locations for each '*nL*' are also given in Figure 7.8.

Ranks/ Peaks	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	5,15	5,16	6,15	6,16	5,14	4,15	4,14	4,16	6,14
2	14,8	14,9	13,8	14,7	13,9	15,8	15,7	13,7	15,9
3	10,10	10,11	10,9	9,10	9,11	11,10	11,11	11,9	9,9

Number	Locations selected					
nL = 3	5,15	14,8	10,10	-	-	-
nL = 4	5,15	14,8	10,10	5,16	-	-
nL = 5	5,15	14,8	10,10	5,16	14,9	-
nL = 6	5,15	14,8	10,10	5,16	14,9	10,11

Figure 7.8: Peaks and ranks distribution of the critical locations and selections made based on probabilities

The total number of schedules selected is two for each location. Each schedule has a lower limit of 0-time steps and an upper limit of 2-time steps. These limits are decided based on the maximum management period selected by the user. The maximum management period here is 20 days. The upper limit of the time steps for each schedule is given by equation 7.8.

$$uS = \frac{M_{\max}}{nS \times dt} \tag{7.8}$$

M_{max} is the maximum management period, and *dt* is the time step discretization. The results will be more accurate as the discretization of time steps decreases. The discharge's upper limit is 300 m^3/day, and the lower limit is 0 m^3/day for injection wells. The optimization model is

run by giving these user inputs, and the results are given as a Pareto- optimal front for different values of threshold concentration to be achieved in the aquifer in Figure 7.9.

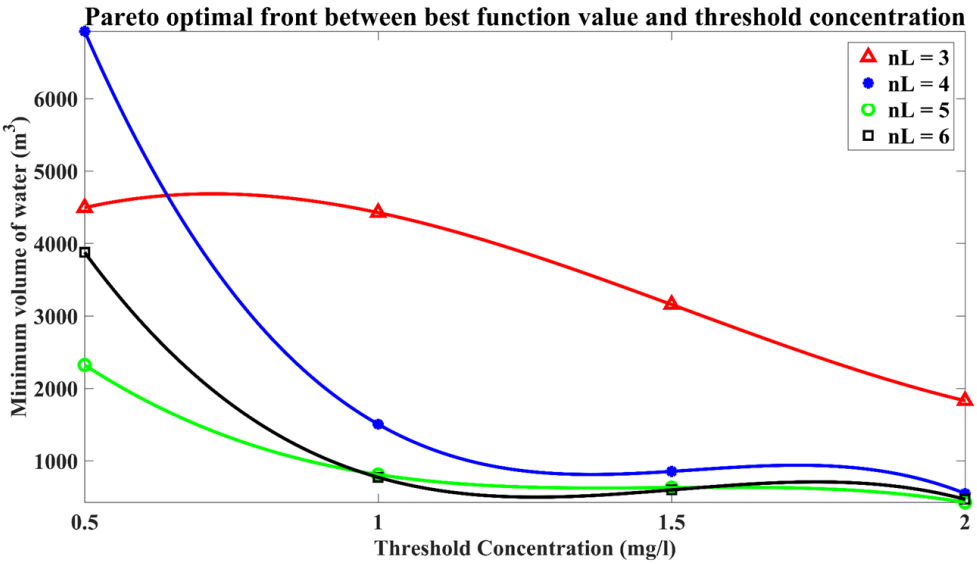

Figure 7.9: Optimization results for different threshold concentrations

The number of locations of injection is represented as '*nL*'. It can be seen that for a value of threshold less than 1 mg/l, the objective function did not converge to satisfy the constraint effectively. Once the constraint is satisfied, then the optimization would result in a converged result of the minimum volume of water that needs to be injected, which has occurred in all the values of '*nL*' for greater concentrations of the threshold value. The objective to minimize the concentrations was achieved with 3, 4, 5, and 6 injection wells. But it can be seen in this figure that the minimum injection volume has not gone down for a lesser number of injection wells, as one might assume. It has gone up, showing that when a more significant number of injection wells are used, a lesser amount of water is required to be boiled and injected into the aquifer. This proves that the correct distribution of injection wells is much more necessary than the total volume of water injected at any location. Figure 7.10 to Figure 7.13 shows the concentration trends in the aquifer for all 6 injection locations and also both the pumping well locations for four threshold concentration values of 2 mg/l, 1.5 mg/l, 1 mg/l, and 0.5 mg/l, respectively. The symbol 'V' in the figure represents the volume of water injected. It can be observed from the figures that the minimum possible volume of injection was achieved when the number of locations was 5 (nL = 5). The viruses will be inactivated with a larger volume of water when management is done with either 3 or 4 injection wells. But, 6 number wells is redundant in this situation as it has not shown any significant improvement over the previous locations of 5.

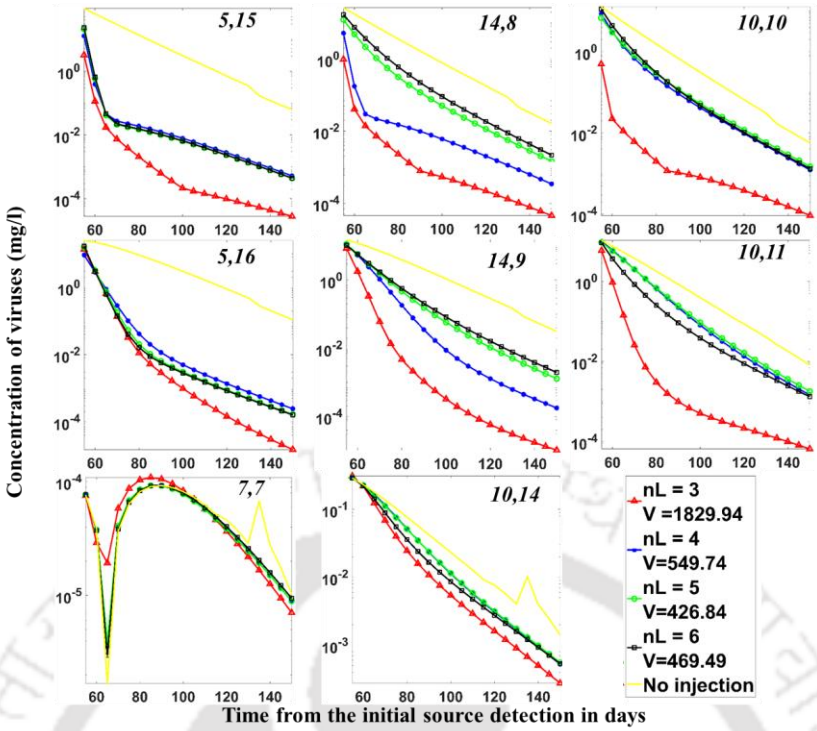

Figure 7.10: Concentration with and without management for a constraint of 2 mg/l in 20 days

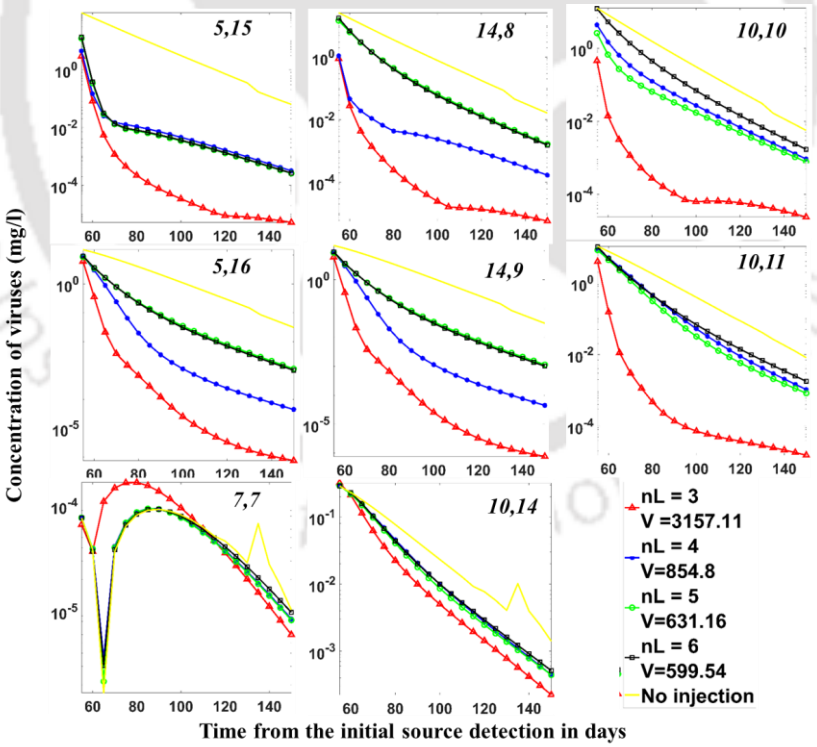

Figure 7.11: Concentration with and without management for a constraint of 1.5 mg/l in 20 days

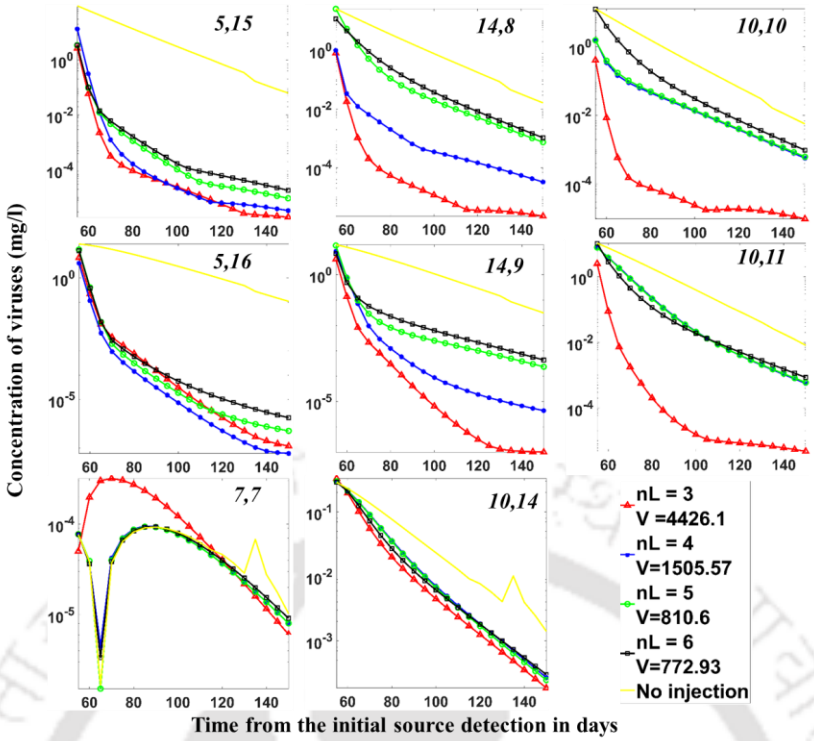

Figure 7.12: Concentration with and without management for a constraint of 1 mg/l in 20 days

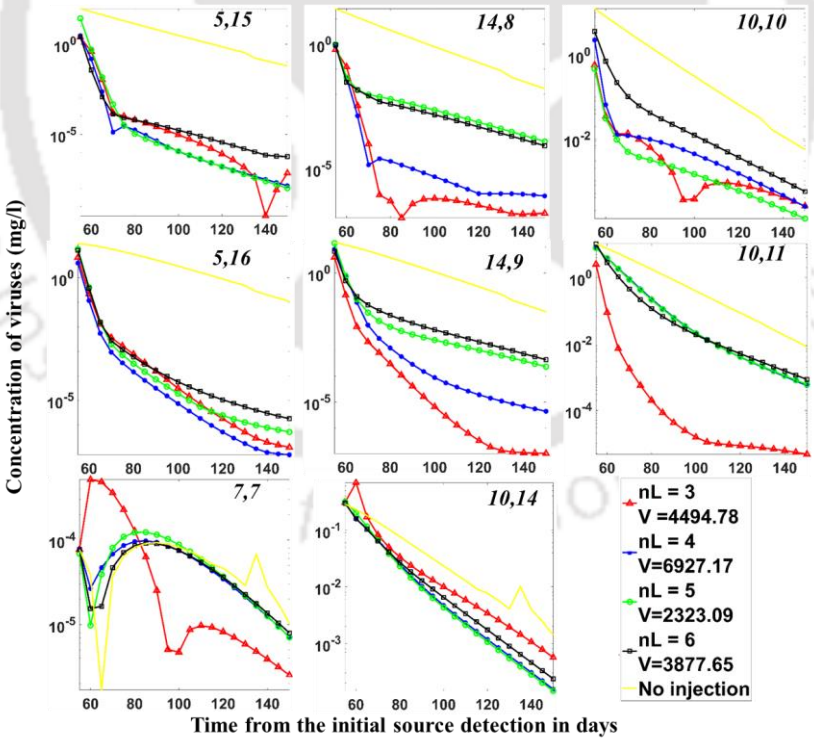

Figure 7.13: Concentration with and without management for a constraint of 0.5 mg/l in 20 days

While designing a management policy, the decision on an optimal solution should be based on the requirements in the field. Here, we had started with the number of locations of 3 and then

stopped when the solution was redundant at the number of locations of 6. The redundancy in adding the injection wells is best identified when the volume of water is not further minimized by increasing the number of locations. At all the locations in the aquifer, the concentration trend is decreasing and is much lesser than the curve with no management. But, at the locations of the pumping wells, the situation is slightly tilted as the well receives much more water than it usually does because of the injection of warm water. The injection wells create a positive curvature in the aquifer system, which drives the viruses to move faster in any direction and disperse faster due to the increased velocities. This phenomenon is explained more clearly in the last section of this chapter.

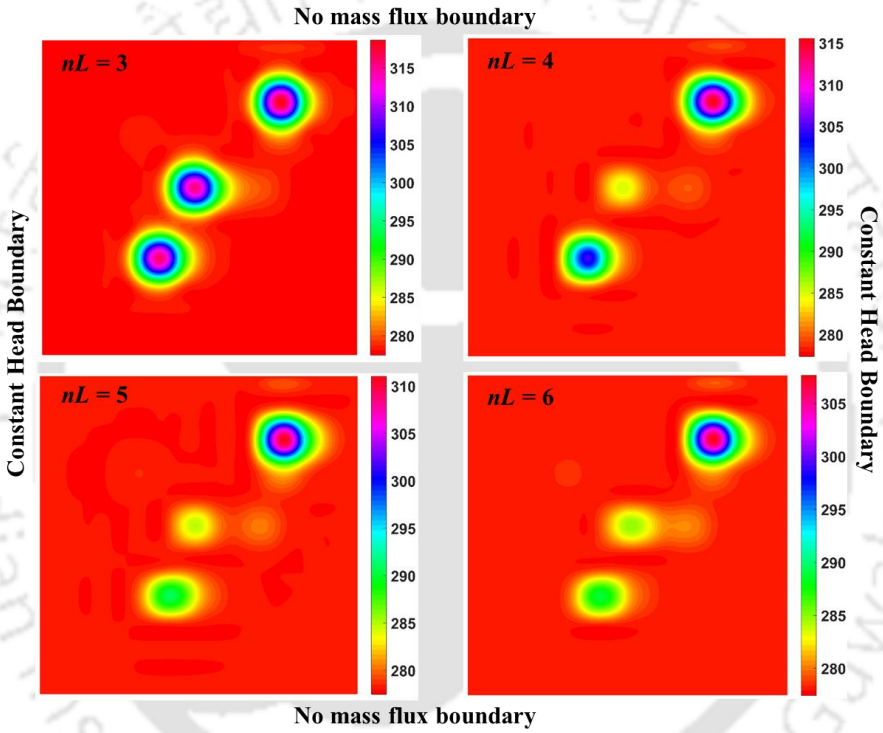

Figure 7.14: Spread of warm water in the aquifer for number of injection wells' nL' 3, 4, 5, and 6

The injection wells to be injected with warm water are located in a space with the highest probability of virus concentration. For the case where the number of injection wells is three, the wells are located at the peaks of the probability of spaces. For the case of four, five, and six injection wells, the injection wells are located at a space where there is the next highest probability of virus concentration to be more, which falls precisely to the right of the three peak locations of highest probability. In Figure 7.14, the peaks are similar in the cases with the number of locations 3 and 6 (as they have one and two injection wells each, respectively), and the cases where the number of injection wells is 4 and 5 have some contrast as some locations have one injection well whereas some locations have two injection wells. As it was clear from

Figure 7.10 to Figure 7.13 that the total volume of injection was lesser for the number of locations of injection of 5, we have only shown the effect on the virus concentrations for '*nL*' of 5.

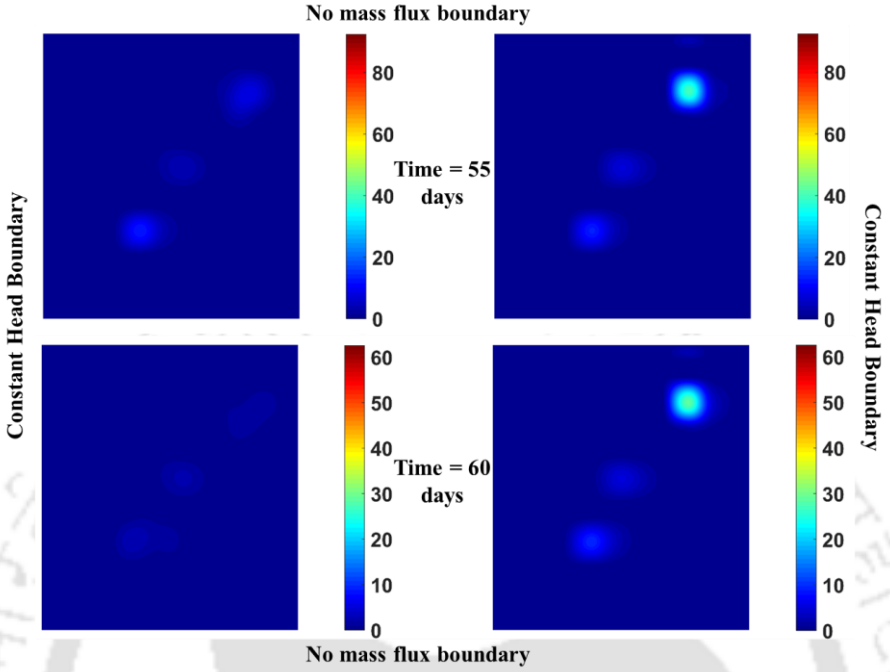

Figure 7.15: Concentration contours with management (left) and without management (right) at 55 days and 60 days.

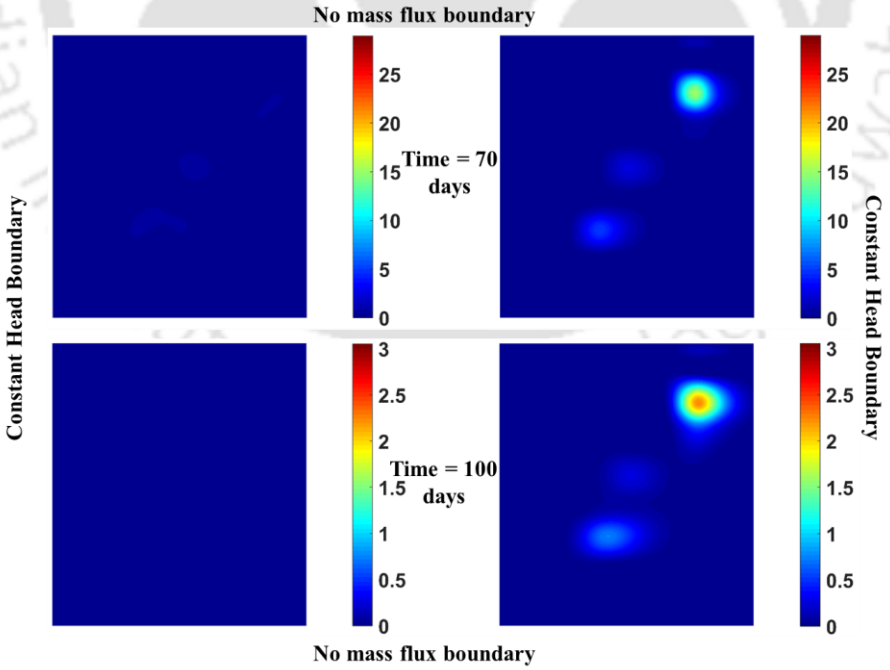

Figure 7.16: Concentration contours with management (left) and without management (right) at 70 days and 100 days.

The contours of concentration of viruses at different times from the initial detection of viruses are given in Figure 7.15 and Figure 7.16. The figures on the right show the concentration of viruses with no management action taken, and the figures on the left show the effect of injection wells on the concentration. Figure 7.15 shows the virus concentration in mg/l at 55 days and 60 days, respectively. Figure 7.16 shows the virus concentration in mg/l at 70 days and 100 days, respectively. The constant head boundary and the no mass flux boundary are common for all the figures, so they are shown at the edge of all the more miniature figures. It can be seen from Figure 7.15 that, as soon as the management of the aquifer starts, the concentration of viruses has drastically come down. In Figure 7.16, it can be seen that the virus concentration is almost negligible even from a management time of just 20 days. The maximum concentration of viruses in the aquifer at the end of 150 days was found to be just 0.006 mg/l, whereas the same when no management policy was undertaken was 0.109 mg/l.

The volume of water injected at different locations in the aquifer as the number of locations increases is given in Figure 7.17 for the threshold value of 1 mg/l.

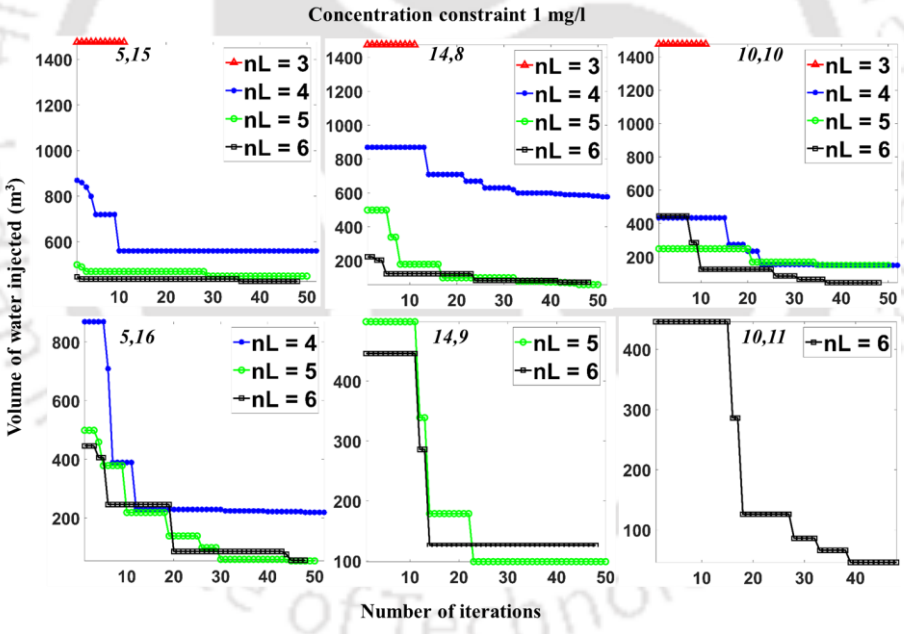

Figure 7.17: The volume of water injected at different locations in the aquifer for a threshold of 1 mg/l.

As the number of locations increases, the volume of hot water that needs to be injected to bring the virus concentration down decreases drastically. This figure is drawn as a result of the optimization model that gives the variation of the volume of water injected with the number of iterations. This figure results from the pattern search algorithm used to fine-tune the result obtained from *fmincon* to check the sensitivity of the locations. It can be observed that there is no change in the volume of water injected in the case in which the number of locations is three.

This is because there is no possibility of decreasing the volume of water injected in this case, which will affect the threshold concentration. In the other cases where the number of locations of injections is greater than the number of peaks, the solution can be improved much more.

It has to be noted that, even though the time-steps in each of the schedules are 2 which is allowed for optimization, the optimal solution has selected the time-steps to be only one constraining the total management time to be effectively 10 days. This is explained in Table 7.1, which gives the optimal management policy solution for the threshold concentration of 1 mg/l.

Table 7.1: The volume of water injected at injection locations for different values of nL

The volume of water injected (m^3) at location		51-55 days	56-60 days	61-65 days	66-70 days	Total volume (m^3)
5,15	nL = 3	737.72	737.72	0	0	1475.45
	nL = 4	124.72	434.72	0	0	559.45
	nL = 5	199.51	249.51	0	0	449.02
	nL = 6	202.99	222.99	0	0	425.98
14,8	nL = 3	737.72	737.48	0	0	1475.21
	nL = 4	302.22	274.72	0	0	576.94
	nL = 5	2.01	57.01	0	0	59.02
	nL = 6	72.99	0	0	0	72.99
10,10	nL = 3	737.72	737.72	0	0	1475.45
	nL = 4	149.72	0	0	0	149.72
	nL = 5	149.51	0	0	0	149.51
	nL = 6	2.99	42.99	0	0	45.98
5,16	nL = 4	104.72	114.72	0	0	219.45
	nL = 5	4.51	49.51	0	0	54.02
	nL = 6	12.99	42.99	0	0	55.98
14,9	nL = 5	9.51	89.51	0	0	99.02
	nL = 6	62.99	62.99	0	0	125.98
10,11	nL = 6	2.99	42.99	0	0	45.98

For some locations, the effective management period is only a one-time step, i.e., 5 days in some cases. This also explains the lesser warm water spread in the heat map in Figure 7.14 for

the locations of 4, 5, and 6. Also, the total volume of water injected is shallow at some locations in some cases.

The Effectiveness of Treatment (ET) term defines the change in concentration at different management stages. The term is just the ratio of the difference of maximum concentration in the aquifer without any management and maximum concentration in the aquifer with management to the maximum concentration in the aquifer without any management. Equation 7.9 gives this ET expressed as a percentage.

$$ET_t = \left(\frac{\max(C_t^{nM}) - \max(C_t^{M})}{\max(C_t^{nM})} \right) \times 100 \qquad 7.9$$

Here, ET_t represents the effectiveness of treatment at a time '*t*', '*nM*' no management, '*M*' represents with management, and '*C*' represents the concentration.

This effectiveness is given in different stages of treatment (ET) expressed as a percentage for different locations for a concentration threshold of 1 mg/l at the end of 5, 10, 20, and 50 days of management Table 7.2.

Table 7.2: Effectiveness of Treatment at different stages for different number of locations

C (mg/l) / Time	C (mg/l) 55 days	ET (%)	C (mg/l) 60 days	ET (%)	C (mg/l) 70 days	ET (%)	C (mg/l) 100 days	ET (%)
No treatment	92.366		62.569		28.940		3.053	
nL = 3	7.557	91.82	3.498	94.41	1.000	96.54	0.087	97.14
nL = 4	13.286	85.62	3.903	93.76	0.996	96.56	0.061	98.01
nL = 5	26.136	71.70	5.461	91.27	0.991	96.58	0.064	97.89
nL = 6	13.152	85.76	4.627	92.60	0.922	96.82	0.079	97.40

It can be observed from this table that the concentration threshold of 1 mg/l has been reached at the end of 20 days of treatment, shown in the second column of Table 7.2. The treatment time has been shown from the timeline of initial detection of concentration; hence, it was shown as 50 + 20 days, i.e., 70 days. It can also be observed that just the initial time step has reduced the concentrations of the viruses to a minimum of 71-91 % effective reduction. The rest is the after-effects of the decline in the concentration curve. The volume of water injected was 4426.1 m^3, 1505.57 m^3, 810.6 $m^{3,}$ and 772.93 m^3 for the number of locations *nL* of 3, 4, 5, and 6,

respectively. It is clear from these numbers that the wastage of resources (energy) is much more in the case of a lesser number of locations.

7.4.2 Solution Category 2

Solution category 2 is selected when existing wells in the aquifer can be used as injection wells. Figure 7.5 in the study area section gives the actual locations of the injection wells in the aquifer. The management model in this category is run for different threshold concentrations from 4 mg/l to 1 mg/l with a decrement of 0.2 mg/l. However, the threshold concentration of 1 mg/l was challenging to achieve in 20 days, and the optimal solution never got converged to it. Therefore, a Pareto-optimal front of the solutions that converged is given in Figure 7.18. An interesting observation can be seen in Figure 7.18 at the concentration threshold of 3.8 and 4 mg/l. There was suddenly a decline in concentration at that point due to a redundant injection well. The redundant injection well is where the aquifer's concentration goes down naturally below the threshold even without injection. Therefore, a drastic decline in the minimum volume of water injected can be seen.

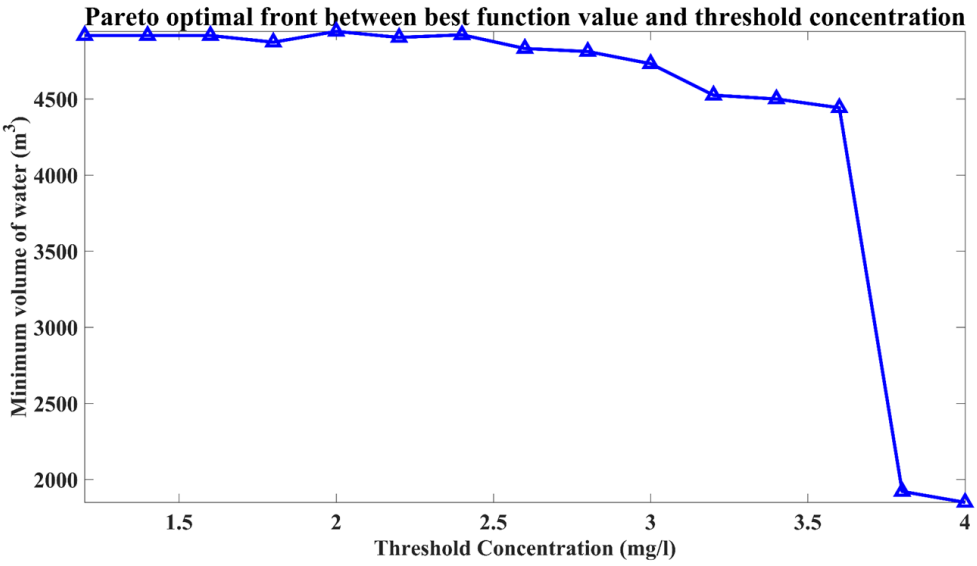

Figure 7.18: Pareto optimal front for the minimum volume of water injected to the concentration threshold

At the location of the actual source detection and the locations of the injection wells, the concentrations of viruses in the aquifer are given concerning the time from the initial source detection in Fig.18. The top three subplots of Figure 7.19 represent the location of sources, i.e. (5,15); (14,8) and (10,10) in the aquifer when represented in a grid distribution of cell size 20x20. The rest five figures are the locations of the injection wells in the aquifer that are pre-existing, i.e. (4,11); (6,15); (10,8); (14,7), and (15,9). These all figures are drawn for the

threshold concentrations of 4 mg/l, 3 mg/l, 2mg/l, and 1 mg/l, respectively. The term '*Ca*' in the legend represents the maximum allowable threshold concentration. The term 'V' represents the volume of water injected into the aquifer to cross the threshold concentration.

The injected warm water is given as a heat map in the aquifer after simulating the heat transport equation in Figure 7.20. The heat maps are given for the concentration constraints of 4 mg/l, 3 mg/l, 2 mg/l, and 1 mg/l, respectively, from top left to bottom right in Figure 7.20. It can be noted that, as the volume of water injected at the threshold of 4 mg/l is very low, the heat map only gives a small amount of heat in three locations in the aquifer. The rest of all the threshold concentrations gives the heat map to be on a large scale extending up to the actual source locations and thus treating the concentrations of viruses. The same can be observed in Figure 7.19, where at the locations 10,10, and 10, 8, the treated concentration breakthrough curve traces the same as the concentration breakthrough curve with no treatment for threshold concentration of 4 mg/l.

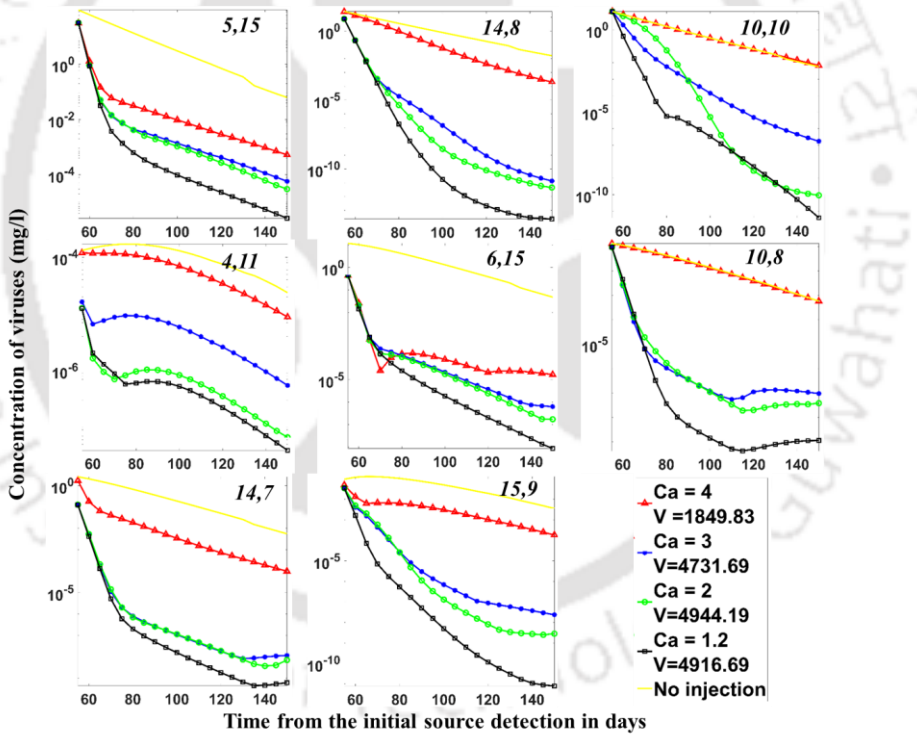

Figure 7.19: Concentration with and without management for a constraint of 4, 3, 2, and 1 mg/l in 20 days of management

The change in the aquifer concentration over time is given in four phases of 55 days, 60 days, 70 days, and 100 days from the initial detection of sources in the aquifer in Figure 7.21 and Figure 7.22.

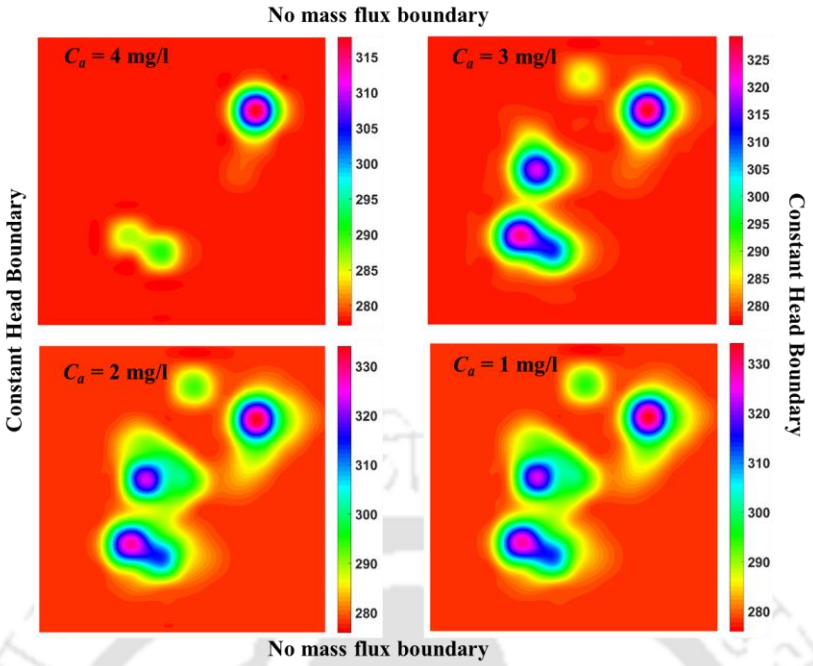

Figure 7.20: Spread of warm water in the aquifer for concentration constraints of 4, 3, 2, and 1 mg/l.

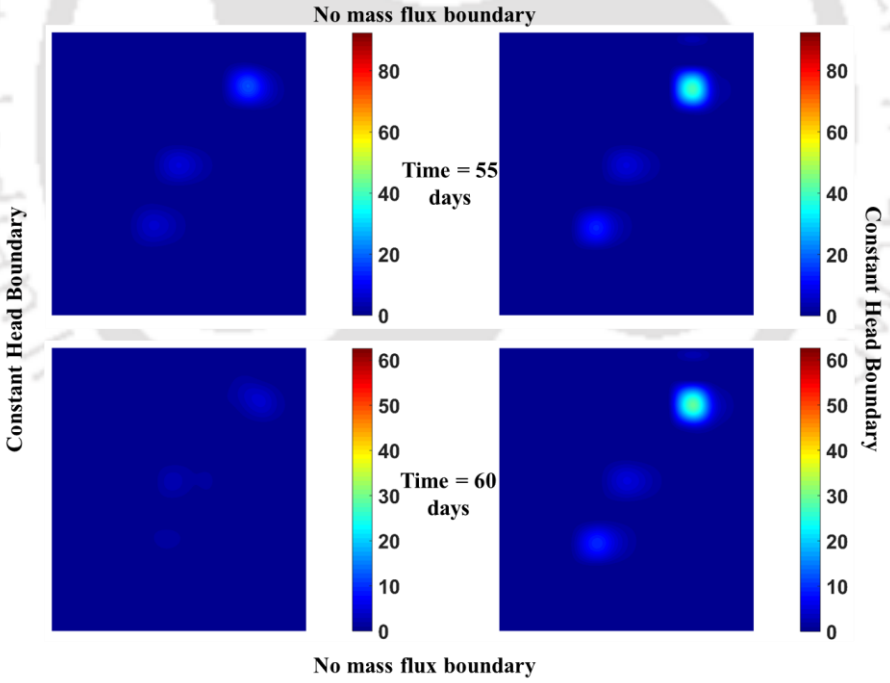

Figure 7.21: Concentration contours with management (left) and without management (right) at 55 days and 60 days

It can be seen from these figures that the virus concentration has reduced to a negligible level in just 70 days and then to almost zero in 100 days. All the figures are generated for a concentration threshold of 2 mg/l. The figure on the left represents the concentration of viruses in the aquifer as a contour with the implemented management policy. The figure on the right

represents the concentration in the aquifer with no management policy applied. Just to see the comparisons between the two cases, the maximum and minimum of the contour were fixed to be the maxima and minima of concentration in the aquifer for no management case. The volume of warm water injected at each location for different threshold concentrations is given in Table 7.3. The concentration constraints in the table are 4, 3, 2, and 1 mg/l. It cannot be identified that obviously that the solution hasn't converged, but it was because the threshold of 2 mg/l and the threshold of 1 mg/l have almost the same values of the total volume of water injected, which indirectly suggests that the solution will result in a concentration value close to 2 mg/l and thus not satisfy the constraint of 1 mg/l.

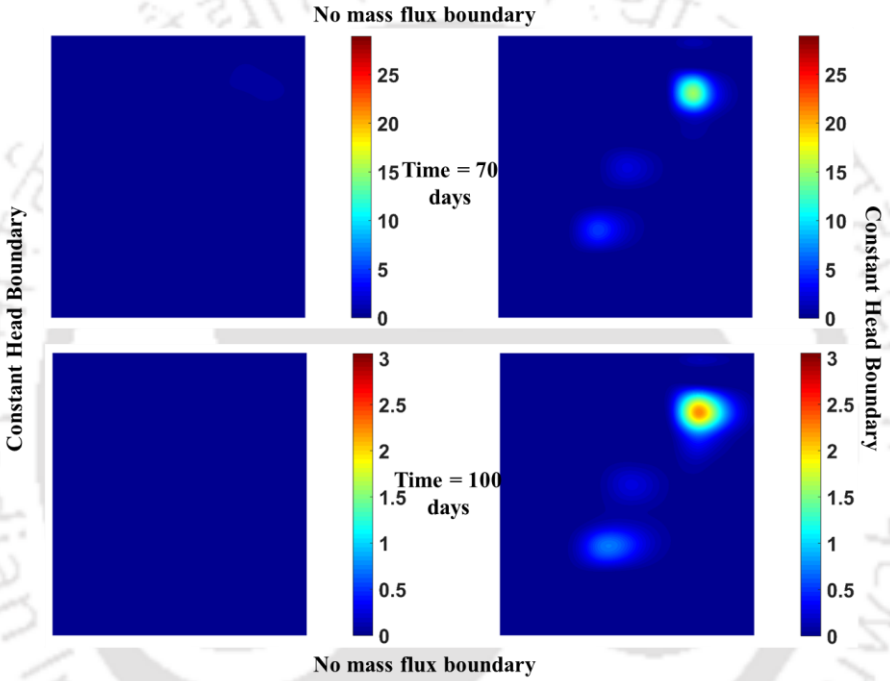

Figure 7.22: Concentration contours with management (left) and without management (right) at 70 days and 100 days

Table 7.3: The volume of warm water injected at different thresholds at injection wells

The volume of water injected (m^3) at location		51-55 days	56-60 days	61-65 days	66-70 days	Total volume (m^3)
4,11	Ca = 4 mg/l	0	0	0	0	0
	Ca = 3 mg/l	171.85	0	0	0	171.85
	Ca = 2 mg/l	385.60	0	0	0	385.60
	Ca = 1 mg/l	435.60	0	0	0	435.60
6,15	Ca = 4 mg/l	766.85	681.85	0	0	1448.70

	Ca = 3 mg/l	766.85	766.85	0	0	1533.70
	Ca = 2 mg/l	766.85	766.85	0	0	1533.70
	Ca = 1 mg/l	766.85	766.85	0	0	1533.70
10,8	Ca = 4 mg/l	1.85	1.85	0	0	3.70
	Ca = 3 mg/l	46.85	766.85	0	0	813.70
	Ca = 2 mg/l	46.85	765.60	0	0	812.45
	Ca = 1 mg/l	46.85	765.60	0	0	812.45
14,7	Ca = 4 mg/l	46.85	116.85	0	0	163.70
	Ca = 3 mg/l	766.85	766.85	0	0	1533.70
	Ca = 2 mg/l	766.85	766.85	0	0	1533.70
	Ca = 1 mg/l	766.85	766.85	0	0	1533.70
15,9	Ca = 4 mg/l	106.85	126.85	0	0	233.70
	Ca = 3 mg/l	391.85	286.85	0	0	678.70
	Ca = 2 mg/l	391.85	286.85	0	0	678.70
	Ca = 1 mg/l	391.85	286.85	0	0	678.70

It can also be seen that the concentration of 4 mg/l results in two redundant locations (4,11) and (10,8), where the volume of warm water injected is almost negligible. Also, at other injection locations, the volume of warm water injected has reduced drastically compared to the other threshold values of concentration.

Table 7.4. Effectiveness of treatment for different threshold concentration at different times

C (mg/l) / Time	**C (mg/l) 55 days**	**ET (%)**	**C (mg/l) 60 days**	**ET (%)**	**C (mg/l) 70 days**	**ET (%)**	**C (mg/l) 100 days**	**ET (%)**
No treatment	92.366		62.569		28.940		3.053	
Ca = 4 mg/l	31.216	66.20	13.678	78.14	3.991	86.21	0.411	86.54
Ca = 3 mg/l	31.416	65.99	7.065	88.71	2.553	91.18	0.201	93.43
Ca = 2 mg/l	31.828	65.54	8.451	86.49	1.154	96.01	0.063	97.93
Ca = 1 mg/l	31.913	65.45	8.454	86.49	1.149	96.03	0.062	97.96

The effectiveness of treatment is given in Table 7.4. The allowable concentrations used were 4, 3, 2, and 1 mg/l to represent the effectiveness of treatment in the aquifer. The effectiveness is given at 5 days, 10 days, 20 days, and 50 days from the start of the treatment in the aquifer.

It can be noted that the values at the thresholds of 2 mg/l and 1 mg/l were almost similar as they have capped to the nearest concentration by the end of 70 days, around 1.15 mg/l. This also gives evidence that the concentration in the aquifer cannot be reduced beyond 1.15 mg/l under the given injection wells with any amount of water to be injected by the end of 20 days.

7.4.3 Solution Category 3

Solution category 3 is the one in which the injection wells are selected from the existing observation wells. As shown in Figure 7.6, there are 21 observation wells located all along the sewer line in a staggered form. The injection wells have to be selected based on the probabilities of each observation wells, the same as the solution category 1. But, here, the probabilities are used only for the observation wells but not the whole aquifer. Therefore, selecting the locations in the aquifer for designing management strategy differs slightly. The probabilities and ranking used for selecting locations are all described in Figure 7.23.

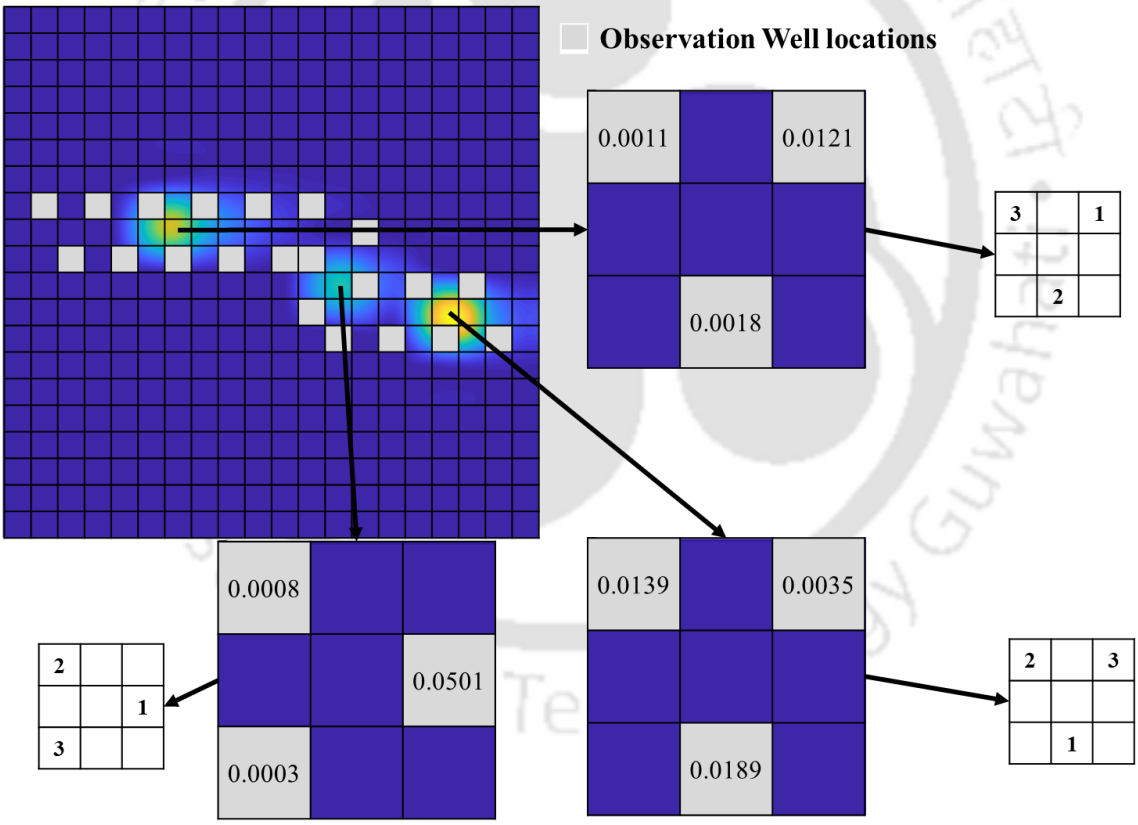

Figure 7.23: Probability and ranking of observation wells for selection of location

The first step is to calculate the probabilities at all the observation wells. The distance between each observation well and the peaks is then calculated as a matrix. Each observation well is then ranked according to its distance to the nearest peak. Finally, a cluster of the observation wells is formed near each of the peaks from which the injection wells are to be selected. Each

such cluster is given in Figure 7.23 as a 3x3 grid consisting of three observation wells each. Now, each cluster is ranked according to its computed probabilities. The final step is to select the number of injection wells according to the ranks.

Here, a different strategy is applied to select the injection wells compared to the procedure from solution category 1. The first selection goes with all the ranks of 1, the same as the solution category 1. But, the second selection set goes from a pool of probabilities that have the selection system from rank 2. For this solution category, three locations are selected to be the minimum as there are three peaks in the aquifer system. Therefore, three locations will be ranked 1, and the first selection is made. When one wants the number of injection wells to be more than 3, they have to select the rest of the injection wells from the pool of probabilities which has ranks ranging from 2 to n (n = 3 in this case), a number selected while deciding the neighborhood points for the peaks. The selection pool and the locations chosen for each case are given in detail in Table 7.5.

Table 7.5: Selected locations based on ranks and probabilities

Selected locations form rank 1			8,8; 11,14; 13,17 – 148; 271; 333			
Location	**Cell Number**	**Probability**	**nL = 3**	**nL = 4**	**nL = 5**	**nL = 6**
11,16	311	0.0139	-	311	311	311
11,18	351	0.0035	-		351	351
10,8	130	0.0018	-			130
8,6	108	0.0011	-			
10,12	230	0.0008	-			
12,12	232	0.0003	-			

I have done the simulations for the number of locations 3, 4, 5, and 6. The number of schedules is capped to have a maximum management time of 20 days with two schedules. Each schedule can have a maximum of 2-time steps, comprising the 20 days for time discretization of 5 days. The concentration thresholds are taken to be 5 mg/l, 4.5 mg/l, 4 mg/l, 3.5 mg/l, 3 mg/l, 2.5 mg/l, 2 mg/l and 1.5 mg/l which are eight in total. Under these constraints, the management model is run for different locations. The results of management in terms of volume of water injected concerning concentration constraints are given in Figure 7.24 as a Pareto-optimal front. It has to be noted that the number of locations of 3 and 4 has not reached the concentration

thresholds with the amount of water injected, as shown in the figure for threshold values of 1.5 mg/l, 2 mg/l, and 2.5 mg/l. The same is true for the number of locations of 5 for the concentration thresholds of 1.5 mg/l and 2 mg/l. Only the number of locations of 6 has satisfied all the eight concentration constraints.

The concentration trends in the aquifer at different locations are given in the set of figures from Figure 7.25 to Figure 7.28. These figures are generated for location of the actual sources, (9,7); (11,13) and (12,17) and also all the 6 injection well locations 8,8; 11,14; 13,17; 11,16; 11,18 and 10,7 comprising a total set of 9 subplots in each figure. The term 'V' represents the volume of warm water injected, and the term 'nL' represents the number of locations. The concentration threshold of 5 mg/l, 4 mg/l, 3 mg/l, and 2 mg/l is used for Figure 7.25, Figure 7.26, Figure 7.27, and Figure 7.28 respectively. These thresholds of concentrations were to be calculated at the end of the management period of 20 days.

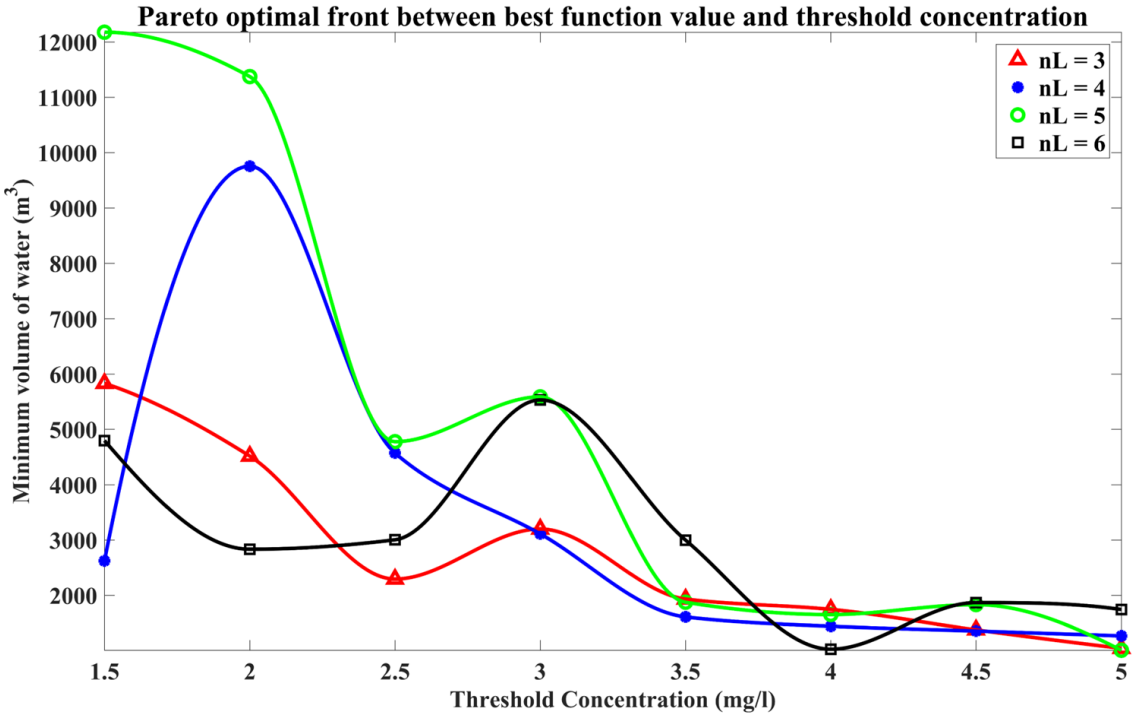

Figure 7.24: Pareto-optimal front for the volume of warm water injected to concentration thresholds for the different number of locations

It can be noted in these figures that the concentration declines rapidly at the point of injection and a slower pace at the actual source locations and the points of no injection of warm water. Also, from the Pareto optimal front in Figure 7.24, it is evident that the concentration threshold was not satisfied for 2 mg/l in the case of the number of locations 3, 4, and 5. The same can be seen in Figure 7.28, where the volume of water injected for these locations is very high as compared to the number of locations 6. The optimal management policy details are discussed

in the subsequent sections. To understand the spread of warm water in the aquifer for different number of locations, heat maps are created for the concentration constraint of 3 mg/l. These heat maps are for the number of locations 3, 4, 5, and 6, as described in the subplots of Figure 7.29.

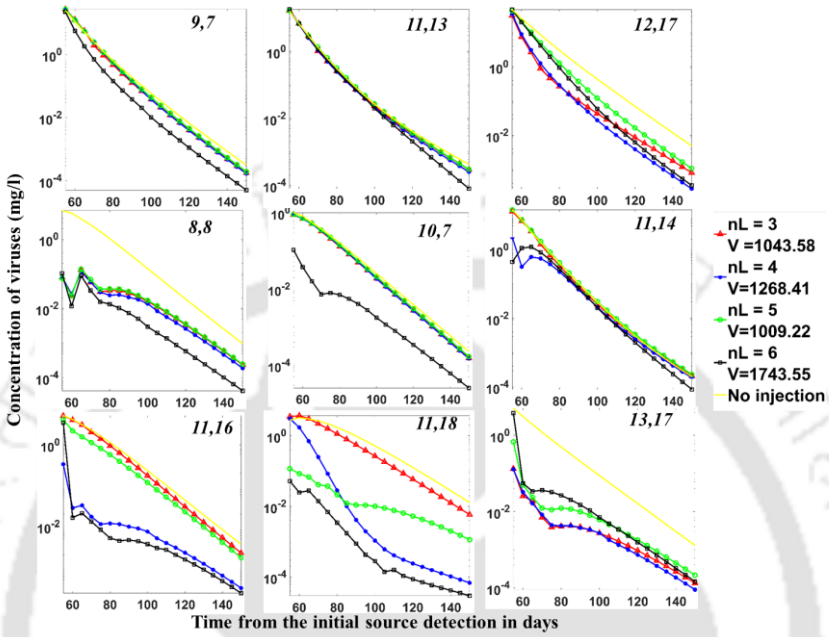

Figure 7.25: Concentration with and without management for a constraint of 5 mg/l in 20 days of management

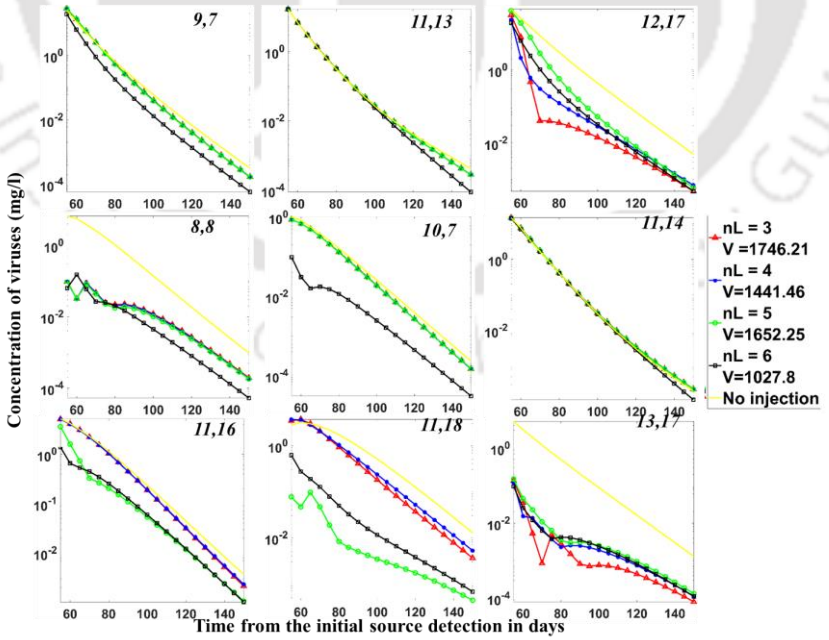

Figure 7.26: Concentration with and without management for a constraint of 4 mg/l in 20 days of management

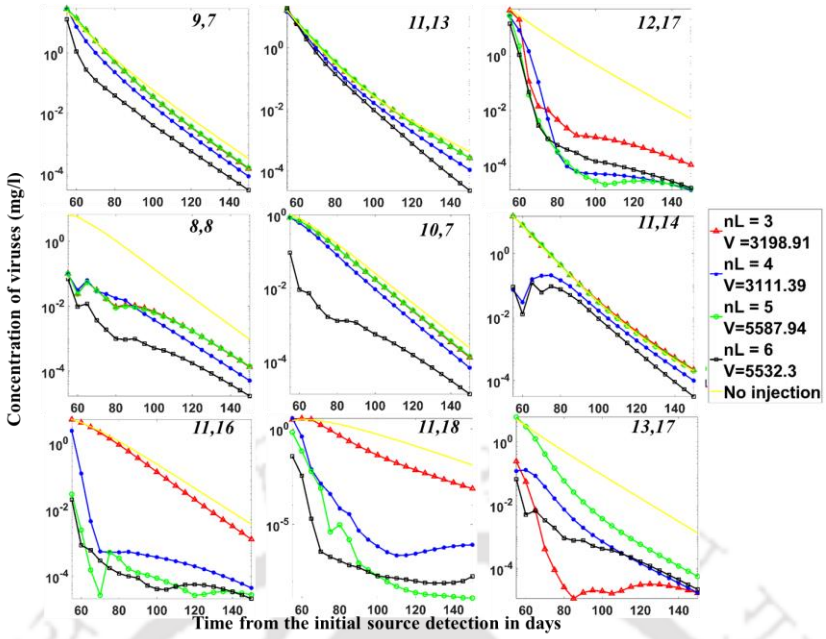

Figure 7.27: Concentration with and without management for a constraint of 3 mg/l in 20 days of management

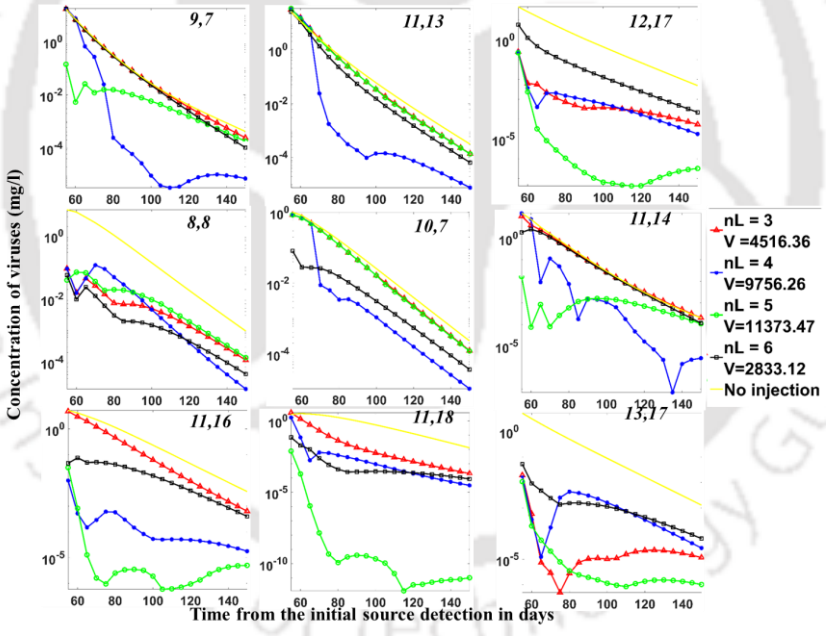

Figure 7.28: Concentration with and without management for a constraint of 2 mg/l in 20 days of management

It can be seen that the volume of injection plays a significant role in the concentration decline curve. However, the location of injection has a more prominent role to play as it is evident from Figure 7.28 that, even though the injection was made in the aquifer at a higher volume for nL = 5, the location of injection has not been good, which have resulted in the infeasible solution.

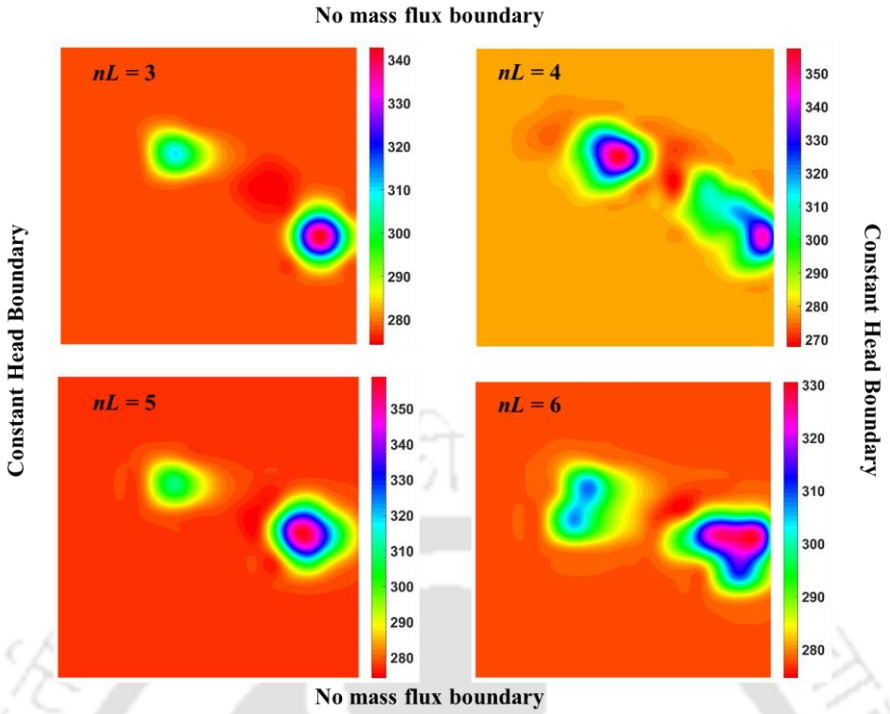

Figure 7.29: Spread of warm water in the aquifer for the concentration constraint of 3 mg/l for the different locations

Interestingly, the number of locations 3 and 4 have almost the same volume of water injected, 3198.91 m^3 and 3111.39 $m^{3,}$ respectively, into the aquifer. But the heat maps they have produced are very different. This is due to the injection locations, which are different for these two cases. Similarly, the same can be observed for nL values of 5 and 6, with the injection volume very close but producing a very different heat map. The injection schedule also plays a prominent role in producing extra heat spread in the aquifer.

The change in concentration of viruses over time with the management and without the management in the aquifer for the times of 55 days, 60 days, 70 days, and 100 days from the initial detection of the source are given for the concentration threshold of 3 mg/l for the number of locations of 6 in Figure 7.30 and Figure 7.31. The figures to the left represent the concentration of viruses in the aquifer with the management policy implemented, and the figures on the right represent no management policy. The concentration contours are plotted on the same scale of maximum and minimum for both the subplots (left and right) to compare the results more quickly.

It can be observed from these figures that, for the virus concentration to reduce to a level negligibly trimmed, it takes around 50 days of management. Only effective management is 10 days as selected by the optimal solution. The rest is only the after-effects of the injected warm

water moving in the aquifer and the viruses. The volume of water injected at different locations in the aquifer is given in Figure 7.32 for the concentration threshold of 3 mg/l.

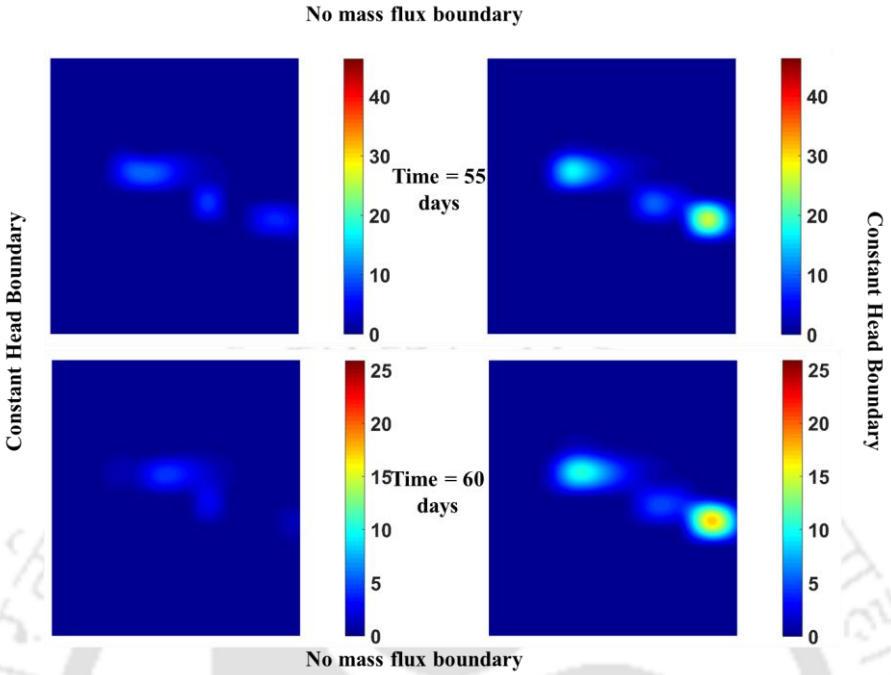

Figure 7.30: Concentration contours with management (left) and without management (right) at 55 days and 60 days

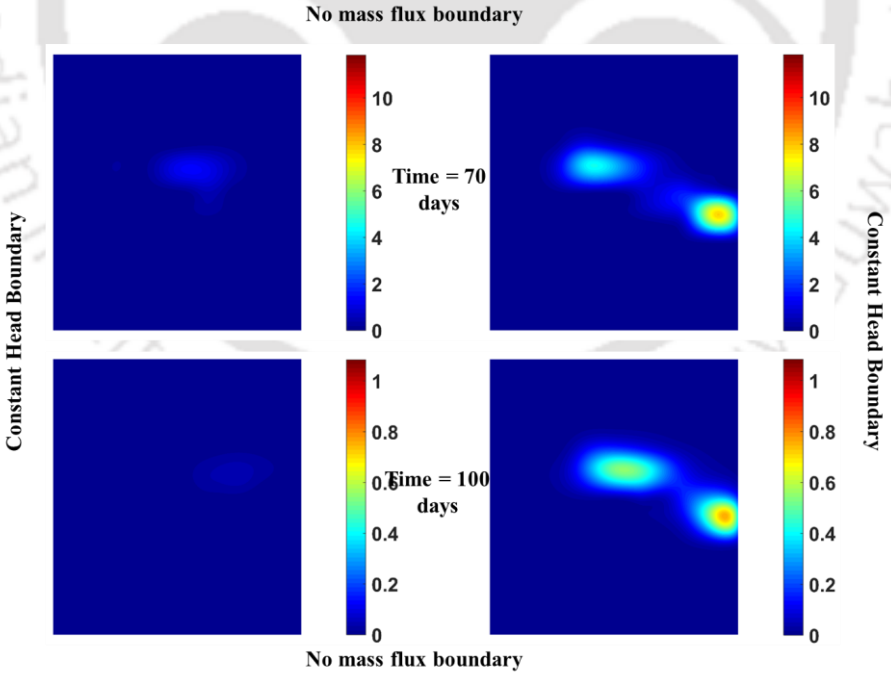

Figure 7.31: Concentration contours with management (left) and without management (right) at 70 days and 100 days

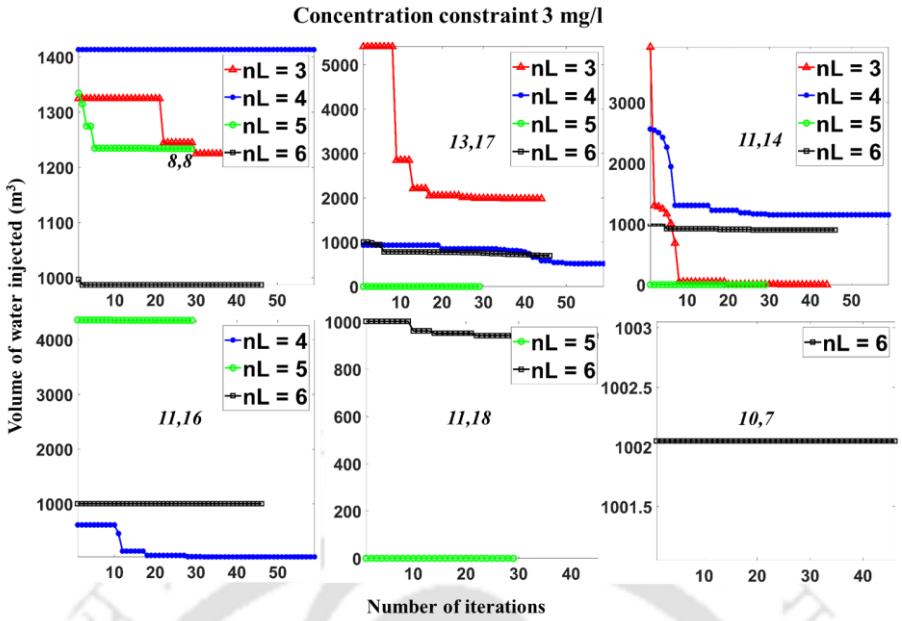

Figure 7.32: The volume of water injected at different locations for a concentration threshold of 3mg/l

Here, it can be observed that the volume of water injected at the top three locations has reduced by a considerable amount, while the volume of water injected at the bottom three locations have not changed much. This is because these locations are crucial for the treatment of the aquifer to a threshold. The top three locations are ranked 1. They are relatively important for treatment, but the volume of water injected, even a tiny amount, would affect the concentration on a large scale and hence can be reduced.

Table 7.6: The volume of water injected at different times for the different number of locations

The volume of water injected (m^3) at location		51-55 days	56-60 days	61-65 days	66-70 days	Total volume (m^3)
8,8	nL = 3	610.49	610.49	0	0	1220.98
	nL = 4	706.74	706.74	0	0	1413.48
	nL = 5	616.91	616.91	0	0	1233.82
	nL = 6	486.02	501.02	0	0	987.04
13,17	nL = 3	71.37	71.37	917.59	917.59	1977.92
	nL = 4	258.23	258.23	0	0	516.46
	nL = 5	0	0	0	0	0
	nL = 6	446.02	251.02	0	0	697.05

11,14	nL = 3	0	0	0	0	0
	nL = 4	575.64	575.64	0	0	1151.28
	nL = 5	0	0	0	0	0
	nL = 6	401.02	501.02	0	0	902.05
11,16	nL = 4	15.08	15.08	0	0	30.17
	nL = 5	1011.82	1011.82	1165.24	1165.24	4354.12
	nL = 6	501.02	501.02	0	0	1002.05
11,18	nL = 5	0	0	0	0	0
	nL = 6	441.02	501.02	0	0	942.05
10,7	nL = 6	501.02	501.02	0	0	1002.05

The same can be seen in Table 7.6 quantitatively. In this table, the total volume of water injected as a schedule is given for each location and at each injection well. The table is prepared for the concentration threshold of 3 mg/l. It can be observed from the table that 10 days of management was enough for many cases except for two locations for two particular subcases. Three injection locations have become redundant for the number of locations of 5, and one has become redundant for the number of locations of 3. This is why the heat map looks similar for the number of locations of 3 and the number of locations of 5. Two redundant injection locations for the number of locations of 5 are 13,17 and 11,18, whereas the injection location of 11,14 was redundant for both locations. This creates an active set of only two injection points in the aquifer for both cases, hence a similar heat map.

The effectiveness of treatment in the aquifer is given in Table 7.7. For this solution category and particularly for this threshold concentration, the treatment was relatively less effective as compared to the previous solution categories.

Table 7.7: Effectiveness of treatment with respect to the number of locations for concentration threshold of 3 mg/l

C (mg/l) / Time	**C (mg/l) 55 days**	**ET (%)**	**C (mg/l) 60 days**	**ET (%)**	**C (mg/l) 70 days**	**ET (%)**	**C (mg/l) 100 days**	**ET (%)**
No treatment	46.319		25.888		11.845		1.084	
nL = 3	41.228	10.99	22.674	12.42	2.999	74.69	0.143	86.84
nL = 4	30.034	35.16	7.831	69.75	1.355	88.57	0.018	98.37
nL = 5	31.060	32.94	14.472	44.10	3	74.67	0.140	87.04
nL = 6	18.908	59.18	7.222	72.10	2.253	80.98	0.069	93.61

The concentration constraint is 3 mg/l for this table. The concentration in the aquifer for times of 55 days, 60 days, 70 days, and 100 days from the initial source identification was compared with the concentrations with treatment to calculate the effectiveness of treatment (ET). Due to the redundant solution, the treatment was less effective for nL = 3 and nL =5. There was 98.37 % reduction in concentration for nL = 4 and 93.61 % reduction in concentration for nL = 6 by the end of 100 days.

In the concentration breakthrough curves for all the solution categories, there can be observed an initial increase in the concentration of virus in the breakthrough curves at some portions. This is happening due to the change in the flow due to the injection wells. The injection of water creates a cone of accretion near the well, increasing the flow velocities considerably near the wells. This increase in velocity can be observed as the difference between Figure 7.33(a) and Figure 7.33(b). The cone of accretion can be observed in the exact figures Figure 7.33(c) and Figure 7.33(d). This increase in velocity changes the dispersion and the advection of the whole flow.

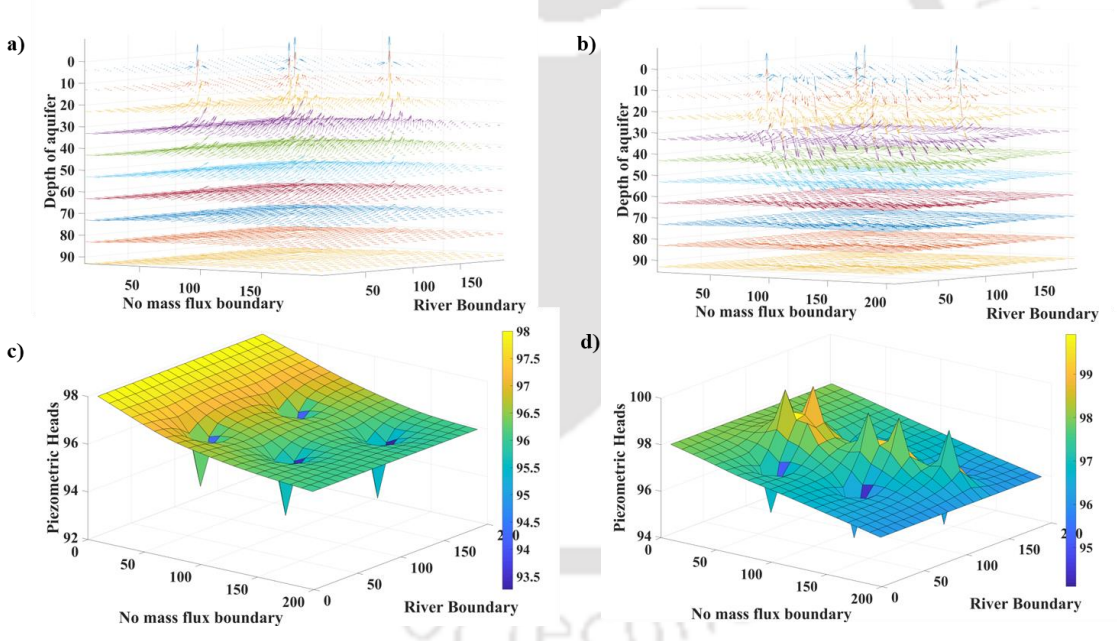

Figure 7.33: Velocity vectors over the aquifer before management (a) and after management (b). Heads in the top layer of the aquifer before management (c) and after management (d)

As the new velocities are much different as compared to the portions before management, the effect can be observed near the pumping well and injection well locations. These effects of the dispersion due to velocity changes can be observed where the change due to change in velocity is more prominent than the change due to the change in inactivation due to temperature. As the change due to change in inactivation is much dominant near the injection wells, the effect of velocity changes can be seen prominently near the pumping well locations. But the effects

won't last long as the virus source is being killed at first with the effect of temperature. Thus, when the flow velocities are increased in the aquifer due to injection wells, the contaminated water moves faster in the porous media, and the pumping wells draw the same contaminated water fast enough. In all the cases near the pumping wells, there is an initial increase in the concentration compared to the concentration without the management, but it gradually decreased as the effect of temperature started to take over the effect of advection and dispersion. Thus, temperature dominates the effects of advection and dispersion over time.

7.5 Conclusions

This chapter attempts to inject hot water into injection wells to minimize the injection volume and constrain the concentration of the virus by changing the inactivation rates according to the temperatures. It is also observed that introducing new injection wells into the aquifer drastically changes the flow geometry and causes an initial spike in the concentration at the pumping wells. However, the concentrations at the pumping wells are well less than the threshold concentration of the aquifer after management, and thus the spike can be ignored. The objective of the management model is to minimize the energy required to treat the aquifer while constraining the concentrations in the aquifer to a specified threshold.

Solution category 1 is where the injection wells can be located anywhere in the aquifer. This management policy applies when the aquifer is in an area isolated from the city, and there is a possibility to dig a well wherever required. The injection wells are used as the existing wells in the case of solution category 2. This management policy is applicable when there is no possibility of digging new wells in the aquifer zone, and the existing wells are only to be used. This management study concludes that hot water can be effectively used in the aquifer if one needs to reduce the concentrations inside the aquifer without the pump-and-treat method. The efficacy of all the models is over 94 % by the end of the total management period, which is a very positive sign for the removal of the virus. Cycles of such management policies can drastically reduce aquifer concentration when organized with utter care. It is also observed that as the number of injection wells increases, the injection volume decreases for the same threshold concentration as targets. However, the number of injection wells that need installation as in solution category 1 causes some amount of resources as an investment for installation. In such a case, a detailed cost and benefit analysis are needed to make a final decision on increasing the number of injection wells.

CHAPTER 8: SUMMARY, CONCLUSIONS, AND FUTURE SCOPE OF THE RESEARCH

8.1 Summary of the thesis

As discussed by many researchers, the virus as a pollutant causes many environmental threats. The virus contamination in groundwater has caused several outbreaks in the past. In such a case, detecting virus sources in groundwater and managing the sources is necessary to reduce the health hazard of the area. Therefore, the main objective of the research is to detect and manage the virus sources in an aquifer. The parameters of virus transport are susceptible to the change in environment in the aquifer. Therefore, I developed virus flow and transport models that can handle the parameters' spatial and temporal variation in this study's first objective. However, as the virus transport processes are very different from the traditional contaminant transport, there is a need to understand the processes much more deeply. Therefore, I developed a pore network model as the second objective of my research. The developed pore network model can produce the desired porosity, intrinsic permeability, and dispersivity values with the linkage to an optimization model. I have used the shuffled frog leaping algorithm (SFLA) as the linking optimization model. The virus transport parameters are susceptible to environmental changes; therefore, the estimation of these parameters is done as a function of the environmental factors in the third objective. The fourth objective of the research is to detect the virus sources in the aquifer. The process for the detection of sources is the linked simulation-optimization technique. The simulated and observed concentrations should be compared to minimize the error in this technique. Researchers have showed that the source identification model is modeled as a mixed-integer problem with discrete variables for locations and continuous variables for source flux values. However, in the case of virus source identification, the information on the location of leakage or the time until the leakage pertains is unavailable. Therefore, I have developed a model that initializes the solution based on the observation data. The information is used in the optimization model of a modified genetic algorithm combined with a local location search algorithm to detect the virus. However, the mere detection of virus sources doesn’t solve the problem of virus contamination. Therefore, a management strategy should be devised to remediate the aquifer from the detected source. Effective management of the virus can be done only by studying the parameters that affect the survival of the virus in groundwater. The management policy should be designed as an

optimization model to remediate the aquifer to the desired and acceptable level of virus concentration. In the fifth and final objective of my research, I have developed a management model that decreases the virus concentration in the aquifer by changing the temperature in the aquifer medium. The parameter, inactivation coefficient, is susceptible to temperature, and thus the removal of virus concentration from the aquifer is modeled. The management model gives the minimum volume of hot water needed for injection to remove viruses from the aquifer effectively.

8.2 Conclusions

The research findings of the work are concluded as follows.

- **Development of the Flow and Transport models in the Darcy Scale**
  - ✓ The models are developed using the Finite Volume Model and produce good correlation values with the MODLFOW model and MT3DMS that solve the same equations using the Finite Difference Model.
  - ✓ The developed models can handle temporally variable parameters, which is an advantage while modeling the virus flow and transport in the aquifer. The existing models, such as MT3DMS and HYDRUS, can be used to input the temporally variable parameter, but the process is quite tedious as the inputs should be generated from another program externally.
  - ✓ The developed model has reduced the computational effort and brought ease to handling the process as the development of these models is in the same environment (MATLAB).
  - ✓ The developed FVM models are used in the further objectives to manage the virus sources in the aquifer.
- **Development of the Pore Network Model**
  - ✓ The pore network model (PNM) development is complex but can produce the desired porosity and other Representative Elementary Volume (REV) properties from random networks.
  - ✓ The PNM is a simple network with only cylindrical pore throats and spherical pore bodies.
  - ✓ The network generated can have micro and macro pores with the user-specified distribution of each of the pores.

- ✓ The desired results are produced by optimizing the mean and variance of the radius of the pore bodies.
- ✓ The model produces consistent properties of the REV while using micro pores only and fluctuating results over several generations considering micro and macro pores.

➢ **Estimation of the Parameters from Pore-scale to the Darcy-scale and their Sensitivities.**

- ✓ The pore parameters depend on nine pore-scale parameters crunched into five non-dimensional numbers.
- ✓ Each non-dimensional is studied carefully for the variation in their ranges with critical emphasis on the significant contributing pore-scale parameter for each number.
- ✓ The analysis is done by running thousands of pore network simulations with variabilities in each non-dimensional number.
- ✓ It has been found that, in decreasing order, temperature, pH, surface potentials of virus and porous matrix, and the mean pore velocity affects the concentration of the virus.

➢ **Virus Source Identification Model**

- ✓ The virus source identification model can be developed using a linked simulation-optimization approach.
- ✓ A new methodology is proposed to initialize the probable location of sources without prior information on the number of sources or the activity time.
- ✓ The Genetic Algorithm (GA) is modified and linked with a local location search algorithm which finds the virus sources as a hybrid GA and gradient-based search.
- ✓ It is observed that the model produces definitive results every time, whereas the existing algorithms seldom converge to the optimum value with this limited data from the wells.
- ✓ The new methodology additions of XY, YZ, and ZX cross-search, the XYZ search, and the best combination search are the methodology's key aspects that guarantee convergence.

➢ **Virus Management Model**

- ✓ The management of virus sources is performed based on the experimental evidence that the inactivation coefficient changes with the change in temperature.
- ✓ The management model is solved for two hypothetical cases of the industrial effluent problem and sewer leakage. However, it can be extended to any field scale problems
- ✓ The solution is proposed in three categories. One is that the injection wells for hot water can be installed anywhere in the aquifer; the second is that the injection wells

can be the existing wells in the aquifer that are randomly distributed, and the third is that the injection wells are the observation wells that are used to find the source in the aquifer.

- ✓ It is observed that the volume of injected water is the least when solutions from category-1 are used, followed by category-3 and then category-2 by the increasing order of volume injected.
- ✓ The proposed model can generate the optimal solution under the cases of desired maximum concentration in the aquifer. The volume of water injected increases as the desired concentration level decreases.

8.3 Future scope of the research

The research work extensively covers the topics related to the virus transport processes. However, there is a scope for the future in this area of research.

- I have done the research based on the experimental datasets from the experiments conducted on bacteriophages, also considered model viruses. However, experiments on viruses can be conducted in safe environments to understand the processes better.
- I have modeled the virus transport processes mathematically and validated them with the existing mathematical models. However, the experimental datasets can be used directly to validate the models, and the parameters can be calibrated based on the experiments.
- I have solved the virus transport and heat transport models as interdependent based only on the temperature inputs to the inactivation rates and adsorption coefficients. The models in the future can be developed as fully-coupled models, where the factors such as viscosity and density of the fluid effects the flow and transport processes of virus and heat transport.
- Electronically calibrated instrumentation tools can be installed in the field that constantly sends the data from the observation wells to identify the sources at the earliest in the aquifer.

REFERENCES

Ahlfeld, D. P., and Heidari, M. (1994). "Applications of Optimal Hydraulic Control to GroundWater Systems." *Journal of Water Resources Planning and Management*, American Society of Civil Engineers, 120(3), 350–365.

Aral, M. M., Guan, J., and Maslia, M. L. (2001). "Identification of Contaminant Source Location and Release History in Aquifers." *Journal of Hydrologic Engineering*, American Society of Civil Engineers, 6(3), 225–234.

Bakke, S., and Øren, P. ˚ A.-E. (2002). "Process Based Reconstruction of Sandstones and Prediction of Transport Properties Article in Transport in Porous Media · Process Based Reconstruction of Sandstones and Prediction of Transport Properties." *Transport in Porous Media*, 46(2–3), 311–343.

Bales, R. C., Gerba, C. P., Grondin, G. H., and Jensen, S. L. (1989). "Bacteriophage Transport in Sandy Soil and Fractured Tuff." *Applied and Environmental Microbiology*, 55(8).

Bales, R. C., Hinkle, S. R., Kroeger, T. W., Stocking, K., and Gerba, C. P. (1991). "Bacteriophage Adsorption during Transport through Porous Media: Chemical Perturbations and Reversibility." *Environmental Science and Technology*, American Chemical Society, 25(12), 2088–2095.

Bales, R. C., Li, S., Maguire, K. M., Yahya, M. T., and Gerba, C. P. (1993). "MS-2 and poliovirus transport in porous media: Hydrophobic effects and chemical perturbations." *Water Resources Research*, John Wiley & Sons, Ltd, 29(4), 957–963.

Barth, G., and Hill, M. C. (2005). "Numerical methods for improving sensitivity analysis and parameter estimation of virus transport simulated using sorptive-reactive processes." *Journal of Contaminant Hydrology*, Elsevier, 76(3–4), 251–277.

Bear, J., and Bachmat, Y. (1986). "Macroscopic modelling of transport phenomena in porous media. 2: Applications to mass, momentum and energy transport." *Transport in Porous Media*, 1(3), 241–269.

Bear, J., and Bachmat, Y. (1990). *Introduction to modeling of transport phenomena in porous media*. *Introduction to modeling of transport phenomena in porous media*.

Bedekar, V., Morway, E. D., Langevin, C. D., and Tonkin, M. J. (2016). "MT3D-USGS version 1: A U.S. Geological Survey release of MT3DMS updated with new and expanded transport capabilities for use with MODFLOW." *Techniques and Methods*.

Bhattacharjya, R. K., and Datta, B. (2009). "ANN-GA-Based Model for Multiple Objective

Management of Coastal Aquifers." *Journal of Water Resources Planning and Management*, American Society of Civil Engineers (ASCE), 135(5), 314–322.

Bhattacharjya, R. K., Srivastava, A., and Satish, M. G. (2015). "Hybrid-Optimization Approach for Estimating Parameters of a Virus Transport Process in Aquifer." *Journal of Hazardous, Toxic, and Radioactive Waste*, American Society of Civil Engineers (ASCE), 19(2), 04014025.

Boal, A. K., Rhodes, C., and Garcia, S. (2015). "Pump-and-Treat Groundwater Remediation Using Chlorine/Ultraviolet Advanced Oxidation Processes." *Groundwater Monitoring and Remediation*, Blackwell Publishing Ltd, 35(2), 93–100.

Bos, R., van der Mei, H. C., and Busscher, H. J. (1999). "Physico-chemistry of initial microbial adhesive interactions – its mechanisms and methods for study." *FEMS Microbiology Reviews*, Oxford University Press (OUP), 23(2), 179–230.

Bouckaert, L., van Loo, D., Sleutel, S., de Neve, S., Jacobs, P., and van Hoorebeke, L. (2009). "Application of X-ray tomography for quantification of the soil pore structure and visualization of soil organic matter." *Geochimica et Cosmochimica Acta Supplement*, 73, A145.

Bredehoeft, J. D. (1994). "Hazardous Waste Remediation: A 21st Century Problem." *Groundwater Monitoring & Remediation*, John Wiley & Sons, Ltd, 14(1), 95–100.

Burnett, R. D., and Frind, E. O. (1987). "Simulation of contaminant transport in three dimensions: 1. The alternating direction Galerkin Technique." *Water Resources Research*, 23(4), 683–694.

Chadalavada, S., and Datta, B. (2008). "Dynamic optimal monitoring network design for transient transport of pollutants in groundwater aquifers." *Water Resources Management*, Springer, 22(6), 651–670.

Chadalavada, S., Datta, B., and Naidu, R. (2011). "Optimisation approach for pollution source identification in groundwater: An overview." *International Journal of Environment and Waste Management*, Inderscience Publishers, 8(1–2), 40–61.

Chadalavada, S., Datta, B., and Naidu, R. (2012a). "Optimal Identification of Groundwater Pollution Sources Using Feedback Monitoring Information: A Case Study." *Environmental Forensics*, Taylor & Francis Group , 13(2), 140–153.

Chadalavada, S., Datta, B., and Naidu, R. (2012b). "Optimal Identification of Groundwater Pollution Sources Using Feedback Monitoring Information: A Case Study." *Environmental Forensics*, 13(2), 140–153.

Chang, L.-C., Shoemaker, C. A., and Liu, P. L.-F. (1992). "Optimal time-varying pumping

rates for groundwater remediation: Application of a constrained optimal control algorithm." *Water Resources Research*, John Wiley & Sons, Ltd, 28(12), 3157–3173.

Chipperfield, A. J. (1995). "The MATLAB Genetic Algorithm Toolbox." *IEE Colloquium on Applied Control Techniques Using MATLAB*, IEE, 1995, 10–10.

Chongxuan, L., and Ball, W. P. (1999). "Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware." *Water Resources Research*, 35(7), 1975–1985.

Chow, V. Te. (2010). *Applied hydrology*.

D.P. Ahlfeld, J. M. M. G. F. P. (1986). "Designing optimal strategies for contaminated groundwater remediation." *Advances in Water Resources*, 9(2), 77–84.

Datta, B., Beegle, J., Kavvas, M., and Orlob, G. (1989). "Development of an expert system embedding pattern recognition techniques for pollution source identification."

Datta, B., Chakrabarty, D., and Dhar, A. (2009). "Simultaneous identification of unknown groundwater pollution sources and estimation of aquifer parameters." *Journal of Hydrology*, Elsevier, 376(1–2), 48–57.

Deb, K. (1999). "An introduction to genetic algorithms." *Sadhana 1999 24:4*, Springer, 24(4), 293–315.

Derjaguin, B., and Landau, L. (1993). "Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes." *Progress in Surface Science*, Pergamon, 43(1–4), 30–59.

Elimelech, M., and O'Melia, C. R. (1990). "Kinetics of Deposition of Colloidal Particles in Porous Media." *Environmental Science and Technology*, American Chemical Society, 24(10), 1528–1536.

Eusuff, M., Lansey, K., and Pasha, F. (2006). "Shuffled frog-leaping algorithm: A memetic meta-heuristic for discrete optimization." *Engineering Optimization*, 38(2), 129–154.

Fang, Y., and Logan, B. E. (1999). "Bacterial Transport in Gas-Sparged Porous Medium." *Journal of Environmental Engineering*, American Society of Civil Engineers (ASCE), 125(7), 668–673.

Gandhi, B. G. R., and Bhattacharjya, R. K. (2020a). "Introduction to Shuffled Frog Leaping Algorithm and Its Sensitivity to the Parameters of the Algorithm." *Modeling and Optimization in Science and Technologies*, Springer, 105–117.

Gandhi, B. G. R., and Bhattacharjya, R. K. (2020b). "Differential Evolution and Its Application in Identification of Virus Release Location in a Sewer Line." *Modeling and Optimization in Science and Technologies*, Springer, 53–72.

Gerba, C. (2004). “Why the Concern About Pathogens in Water?” *Southwest Hydrology*, (December), 14–16.

Gerba, C. P. (1984). “Applied and Theoretical Aspects of Virus Adsorption to Surfaces.” *Advances in Applied Microbiology*, Academic Press, 30(C), 133–168.

Gerba, C. P., Goyal, S. M., Cech, I., and Bogdan, G. F. (1981). “Quantitative Assessment of the Adsorptive Behavior of Viruses to Soils.” *Environmental Science and Technology*, American Chemical Society , 15(8), 940–944.

Gorelick, S. M. (1983). “A review of distributed parameter groundwater management modelling methods.” *Water Resources Research*, 19(2), 305–319.

Gorelick, S. M. (1990). “Large scale nonlinear deterministic and stochastic optimization: Formulations involving simulation of subsurface contamination.” *Mathematical Programming 1990 48:1*, Springer, 48(1), 19–39.

Gorelick, S. M., Evans, B., and Remson, I. (1983). “Identifying sources of groundwater pollution: An optimization approach.” *Water Resources Research*, John Wiley & Sons, Ltd, 19(3), 779–790.

Grant, S. B., List, E. J., and Lidstrom, M. E. (1993). “Kinetic analysis of virus adsorption and inactivation in batch experiments.” *Water Resources Research*, John Wiley & Sons, Ltd, 29(7), 2067–2085.

Gregory, J. (1981). “Approximate expressions for retarded van der waals interaction.” *Journal of Colloid And Interface Science*, Academic Press, 83(1), 138–145.

Hahn, M. W., and O’Melia, C. R. (2004). “Deposition and Reentrainment of Brownian Particles in Porous Media under Unfavorable Chemical Conditions: Some Concepts and Applications.” *Environmental Science and Technology*, American Chemical Society , 38(1), 210–220.

Hassanizadeh, S. M., Schijven, J. F., and Majid Hassanizadeh, S. (2000). “Removal of Viruses by Soil Passage: Overview of Modeling, Processes, and Parameters.” *Critical Reviews in Environmental Science and Technology*, CRC Press LLC, 30(1), 49–127.

Hazrati-Yadkoori, S., and Datta, B. (2017). “Adaptive Surrogate Model Based Optimization (ASMBO) for Unknown Groundwater Contaminant Source Characterizations Using Self-Organizing Maps.” *Journal of Water Resource and Protection*, Scientific Research Publishing, Inc, 09(02), 193–214.

Hemes, S., Desbois, G., Urai, J. L., Schröppel, B., and Schwarz, J. O. (2015). “Multi-scale characterization of porosity in Boom Clay (HADES-level, Mol, Belgium) using a combination of X-ray μ-CT, 2D BIB-SEM and FIB-SEM tomography.” *Microporous and*

Mesoporous Materials, 208, 1–20.

Hill, W. F., Akin, E. W., and Benton, W. H. (1971). "Detection of viruses in water: A review of methods and application." *Water Research*, Pergamon.

Hogg, R., Healy, T. W., and Fuerstenau, D. W. (1966). "Mutual coagulation of colloidal dispersions." *Transactions of the Faraday Society*, The Royal Society of Chemistry, 62(0), 1638–1651.

Houben, M. E., Desbois, G., and Urai, J. L. (2013). "Pore morphology and distribution in the Shaly facies of Opalinus Clay (Mont Terri, Switzerland): Insights from representative 2D BIB-SEM investigations on mm to nm scale." *Applied Clay Science*, Elsevier B.V., 71, 82–97.

Huang, C., Research, A. M.-W. R., and 1997, undefined. (1997). "Pump-and-treat optimization using well locations and pumping rates as decision variables." *Wiley Online Library*, Blackwell Publishing Ltd, 33(5), 1001–1012.

Hunt, R. J., Borchardt, M. A., Richards, K. D., and Spencer, S. K. (2010). "Assessment of sewer source contamination of drinking water wells using tracers and human enteric viruses." *Environmental Science and Technology*, 44(20), 7956–7963.

Hurst, C. J., Gerba, C. P., and Cech, I. (1980). "Effects of environmental variables and soil characteristics on virus survival in soil." *Applied and Environmental Microbiology*, 40(6).

Ioannidis, M. A., and Chatzis, I. (2000). "On the geometry and topology of 3D stochastic porous media." *Journal of Colloid and Interface Science*, Academic Press Inc., 229(2), 323–334.

Jha, M., and Datta, B. (2013). "Three-Dimensional Groundwater Contamination Source Identification Using Adaptive Simulated Annealing." *Journal of Hydrologic Engineering*, American Society of Civil Engineers (ASCE), 18(3), 307–317.

Jin, Y., Chu, Y., and Li, Y. (2000). "Virus removal and transport in saturated and unsaturated sand columns." *Journal of Contaminant Hydrology*, Elsevier, 43(2), 111–128.

Johnson, A. (1967). *Specific yield: compilation of specific yields for various materials*.

Kazemzadeh-Parsi, M. J., Daneshmand, F., Ahmadfard, M. A., Adamowski, J., and Martel, R. (2014). "Optimal groundwater remediation design of pump and treat systems via a simulation–optimization approach and firefly algorithm." *http://dx.doi.org/10.1080/0305215X.2013.858138*, Taylor & Francis, 47(1), 1–17.

Keswick, B. H., and Gerba, C. P. (1980). "Viruses in groundwater." *Environmental Science and Technology*, American Chemical Society , 14(11), 1290–1297.

Kinoshita, T., Bales, R. C., Maguire, K. M., and Gerba, C. P. (1993). "Effect of pH on

bacteriophage transport through sandy soils." *Journal of Contaminant Hydrology*, Elsevier, 14(1), 55–70.

Lance, J. C., and Gerba, C. P. (1984). "Virus movement in soil during saturated and unsaturated flow." *Applied and Environmental Microbiology*, 47(2).

Leichombam, S., and Bhattacharjya, R. K. (2019). "New Hybrid Optimization Methodology to Identify Pollution Sources Considering the Source Locations and Source Flux as Unknown." *Journal of Hazardous, Toxic, and Radioactive Waste*, 23(1), 04018037.

Loveland, J. P., Ryan, J. N., Amy, G. L., and Harvey, R. W. (1996). "The reversibility of virus attachment to mineral surfaces." *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, Elsevier B.V., 205–221.

Mahar, P. S., and Datta, B. (2000). "Identification of pollution sources in transient groundwatersystems." *Water Resources Management*, 14(3), 209–227.

Mahar, P. S., and Datta, B. (2001). "Optimal Identification of Ground-Water Pollution Sources and Parameter Estimation." *Journal of Water Resources Planning and Management*, American Society of Civil Engineers, 127(1), 20–29.

Mahinthakumar, G. K., and Sayeed, M. (2006). "Reconstructing groundwater source release histories using hybrid optimization approaches." *Environmental Forensics*, Taylor & Francis Group , 7(1), 45–54.

McDonald, M., and Harbaugh, A. (1988). *A modular three-dimensional finite-difference ground-water flow model*.

Minagawa, H., Nishikawa, Y., Ikeda, I., Miyazaki, K., Takahara, N., Sakamoto, Y., Komai, T., and Nairta, H. (2008). "Characterization of sand sediment by pore size distribution and permeability using proton nuclear magnetic resonance measurement." *Journal of Geophysical Research: Solid Earth*, John Wiley & Sons, Ltd, 113(B7), 7210.

Molson, J. W., Frind, E. O., and Palmer, C. D. (1992). "Thermal energy storage in an unconfined aquifer: 2. Model development, validation, and application." *Water Resources Research*, John Wiley & Sons, Ltd, 28(10), 2857–2867.

Ojha, C. S. P., Surampalli, R. Y., Sharma, P. K., and Joshi, N. (2011). "Breakthrough curves and simulation of virus transport through fractured porous media." *Journal of Environmental Engineering*, 137(8), 731–739.

Palmer, C. D., Blowes, D. W., Frind, E. O., and Molson, J. W. (1992). "Thermal energy storage in an unconfined aquifer: 1. Field Injection Experiment." *Water Resources Research*, John Wiley & Sons, Ltd, 28(10), 2845–2856.

Penrod, S. L., Olson, T. M., and Grant, S. B. (1996). "Deposition kinetics of two viruses in

packed beds of quartz granular media." *Langmuir*, American Chemical Society, 12(23), 5576–5587.

Poletika, N. N., Jury, W. A., and Yates, M. V. (1995). "Transport of Bromide, Simazine, and MS-2 Coliphage in a Lysimeter Containing Undisturbed, Unsaturated Soil." *Water Resources Research*, 31(4), 801–810.

Powelson, D. K., Simpson, J. R., and Gerba, C. P. (1990). "Virus Transport and Survival in Saturated and Unsaturated Flow through Soil Columns." *Journal of Environmental Quality*, Wiley, 19(3), 396–401.

Rajeev Gandhi, B. G., Bhattacharjya, R. K., and Satish, M. G. (2017). "Simulation-optimization-based virus source identification model for 3D unconfined aquifer considering source locations and number as variable." *Journal of Hazardous, Toxic, and Radioactive Waste*, 21(2).

Raoof, A., and Hassanizadeh, S. M. (2013). "Saturation-dependent solute dispersivity in porous media: Pore-scale processes." *Water Resources Research*, 49(4), 1943–1951.

Raoof, A., Hassanizadeh, S. M., and Leijnse, A. (2010a). "Upscaling Transport of Adsorbing Solutes in Porous Media: Pore-Network Modeling." *Vadose Zone Journal*, 9(3), 624–636.

Raoof, A., Hassanizadeh, S. M., and Leijnse, A. (2010b). "Upscaling Transport of Adsorbing Solutes in Porous Media: Pore-Network Modeling." *Vadose Zone Journal*, Wiley, 9(3), 624–636.

Raoof, A., and Majid Hassanizadeh, S. (2010). "A new method for generating pore-network models of porous media." *Transport in Porous Media*, 81(3), 391–407.

Raoof, A., Nick, H. M., Hassanizadeh, S. M., and Spiers, C. J. (2013). "PoreFlow: A complex pore-network model for simulation of reactive transport in variably saturated porous media." *Computers and Geosciences*, Elsevier, 61, 160–174.

Ratha, D. N., Hari Prasad, K. S., and Ojha, C. P. (2009). "Analysis of virus transport in groundwater and identification of transport parameters." *Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management*, 13(2), 98–109.

Rijnaarts, H. H. M., Norde, W., Bouwer, E. J., Lyklema, J., and Zehnder, A. D. J. B. (1996). "Bacterial deposition in porous media related to the clean bed collision efficiency and to substratum blocking by attached cells." *Environmental Science and Technology*, American Chemical Society , 30(10), 2869–2876.

Ryan, J. N., and Elimelech, M. (1996). "Colloid mobilization and transport in groundwater." *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 107, 1–56.

Sadeghi, G., Schijven, J. F., Behrends, T., Hassanizadeh, S. M., Gerritse, J., and Kleingeld, P.

J. (2011). “Systematic Study of Effects of pH and Ionic Strength on Attachment of Phage PRD1.” *Ground Water*, 49(1), 12–19.

Schijven, J. F. (2002). “Virus Removal by Soil Passage at Field Scale and Groundwater Protection.” *Riverbank Filtration: Understanding Contaminant Biogeochemistry and Pathogen Removal*, Springer Netherlands, 55–84.

Schijven, J. F., Hoogenboezem, W., Hassanizadeh, S. M., and Peters, J. H. (1999). “Modeling removal of bacteriophages MS2 and PRD1 by dune recharge at Castricum, Netherlands.” *Water Resources Research*, 35(4), 1101–1111.

Seetha, N., Majid Hassanizadeh, S., Mohan Kumar, M. S., and Raoof, A. (2015). “Correlation equations for average deposition rate coefficients of nanoparticles in a cylindrical pore.” *Water Resources Research*, Blackwell Publishing Ltd, 51(10), 8034–8059.

Seetha, N., Mohan Kumar, M. S., Majid Hassanizadeh, S., and Raoof, A. (2014). “Virus-sized colloid transport in a single pore: Model development and sensitivity analysis.” *Journal of Contaminant Hydrology*, Elsevier, 164, 163–180.

Shen, C., Huang, Y., Li, B., and Jin, Y. (2008). “Effects of solution chemistry on straining of colloids in porous media under unfavorable conditions.” *Water Resources Research*, 44(5).

Shen, C., Huang, Y., Li, B., and Jin, Y. (2010). “Predicting attachment efficiency of colloid deposition under unfavorable attachment conditions.” *Water Resources Research*, 46(11).

Shields, P. A., and Farrah, S. R. (1983). *Influence of Salts on Electrostatic Interactions Between Poliovirus and Membrane Filters*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*.

Sim, Y., and Chrysikopoulos, C. V. (1996). “One-dimensional virus transport in porous media with time- dependent inactivation rate coefficients.” *Water Resources Research*, 32(8), 2607–2611.

Šimunek, J., Genuchten, M. T. van, and Šejna, M. (2012). “HYDRUS: Model Use, Calibration, and Validation.” *Transactions of the ASABE*, American Society of Agricultural and Biological Engineers, 55(4), 1263–1274.

Singh, R. M., and Datta, B. (2004). “Groundwater pollution source identification and simultaneous parameter estimation using pattern matching by artificial neural network.” *Environmental Forensics*, 5(3), 143–153.

Singh, R. M., and Datta, B. (2006). “Identification of groundwater pollution sources Using GA-based linked simulation optimization model.” *Journal of Hydrologic Engineering*, 11(2), 101–109.

Skaggs, T., Hydrology, Z. K.-J. of C., and 1998, U. (1994). “Limitations in recovering the history of a groundwater contaminant plume.” *Elsevier*.

Sleutel, S., Cnudde, V., Masschaele, B., Vlassenbroek, J., Dierick, M., Van Hoorebeke, L., Jacobs, P., and De Neve, S. (2008). “Comparison of different nano- and micro-focus X-ray computed tomography set-ups for the visualization of the soil microstructure and soil organic matter.” *Computers and Geosciences*, Pergamon, 34(8), 931–938.

Spielman, L. A., and Friedlander, S. K. (1974). “Role of the electrical double layer in particle deposition by convective diffusion.” *Journal of Colloid And Interface Science*, 46(1), 22–31.

Thierens, D., and Goldberg, D. (1994). “Convergence models of genetic algorithm selection schemes.” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer, Berlin, Heidelberg, 866 LNCS, 119–129.

Tierney, J. T., Sullivan, R., and Larkin, E. P. (1977). “Persistence of poliovirus 1 in soil and on vegetables grown in soil previously flooded with inoculated sewage sludge or effluent.” *Applied and Environmental Microbiology*, 33(1).

Tim, U. S., and Mostaghimi, S. (1991). “Model for Predicting Virus Movement Through Soils.” *Groundwater*, 29(2), 251–259.

Verwey, E., Overbeek, J., and Nes, K. Van. (1948). “Theory of the stability of lyophobic colloids: the interaction of sol particles having an electric double layer.”

Vilker, V., Research, W. B.-W., and 1980, U. (1980). “Adsorption mass transfer model for virus transport in soils.” *Elsevier*, 14(7), 783–790.

Voudrias, and E.A. (2018). “Pump-and-Treat remediation of groundwater contaminated by hazardous waste: can it really be achieved?” *Global NEST Journal*, 3(1), 1–10.

Wagner, B. J. (1988). “Optimal groundwater quality management under uncertainty.” Department of Applied Earth Sciences, Stanford University.

Wagner, B. J. (1995). “Recent advances in simulation-optimization groundwater management modeling.” *Reviews of Geophysics*, 33(2 S), 1021–1028.

Wan, J., and Wilson, J. L. (1994). “Colloid transport in unsaturated porous media.” *Water Resources Research*, 30(4), 857–864.

Whiffen, G. J., and Shoemaker, C. A. (1993). “Nonlinear weighted feedback control of groundwater remediation under uncertainty.” *Water Resources Research*, 29(9), 3277–3289.

Xiong, Q., Baychev, T. G., and Jivkov, A. P. (2016). “Review of pore network modelling of

porous media: Experimental characterisations, network constructions and applications to reactive transport." *Journal of Contaminant Hydrology*, The Authors, 192, 101–117.

Xu, Z., and Tartakovsky, A. M. (2017). "Method of model reduction and multifidelity models for solute transport in random layered porous media." *Physical Review E*, 96(3), 1–8.

Yahya, M. T., Galsomies, L., Gerba, C. P., and Bales, R. C. (1993). "Survival of bacteriophages MS-2 and PRD-1 in ground water." *Water Science and Technology*, IWA Publishing, 409–412.

Yao, K. M., Habibian, M. T., and O'Melia, C. R. (1971). "Water and Waste Water Filtration: Concepts and Applications." *Environmental Science and Technology*, 5(11), 1105–1112.

Yates, M., Technology, C. G.-W. S. and, and 1985, U. (1985). "Factors controlling the survival of viruses in groundwater." *iwaponline.com*, 17(4–5), 681–387.

Yates, M. V, Yates & Charles, S. R., and Gerba, P. P. (1988). "Modeling microbial fate in the subsurface environment." *Critical Reviews in Environmental Science and Technology*, 17(4), 307–344.

Yates, M. V, and Yates, S. R. (1988). *VIRUS SURVIVAL AND TRANSPORT IN GROUND WATER. Wat. Sci. Tech.*

LIST OF PUBLICATIONS

Rajeev Gandhi, B. G., Bhattacharjya, R. K., and Satish, M. G. (2017). "Simulation-optimization-based virus source identification model for 3D unconfined aquifer considering source locations and number as variable." *Journal of Hazardous, Toxic, and Radioactive Waste*, 21(2)

Gandhi, B. G. R., and Bhattacharjya, R. K. (2020). "Differential Evolution and Its Application in Identification of Virus Release Location in a Sewer Line." *Modeling and Optimization in Science and Technologies*, Springer, 53–72.

Gandhi, B. G. R., and R. K. Bhattacharjya. "Introduction to shuffled frog leaping algorithm and its sensitivity to the parameters of the algorithm." *Nature-Inspired Methods for Metaheuristics Optimization*. Springer, Cham, 2020. 105-117.