Simulation Foundations, Methods and Applications

S.K. Basu
Naveen Kumar *Editors*

Modelling and Simulation of Diffusive Processes

Methods and Applications

Springer

Simulation Foundations, Methods and Applications

Series Editor
Louis G. Birta

For further volumes:
www.springer.com/series/10128

S. K. Basu • Naveen Kumar
Editors

Modelling and Simulation of Diffusive Processes

Methods and Applications

Springer

Editors
S. K. Basu
Banaras Hindu University
Varanasi
Uttar Pradesh
India

Naveen Kumar
Banaras Hindu University
Varanasi
Uttar Pradesh
India

ISSN 2195-2817 ISSN 2195-2825 (electronic)
ISBN 978-3-319-05656-2 ISBN 978-3-319-05657-9 (eBook)
DOI 10.1007/978-3-319-05657-9
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014934657

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This edited book on "Modeling and Simulation of Diffusive Processes: Methods and Applications" contains contributions from authors with a variety of academic backgrounds. It is an outgrowth of the International Conference on Simulation and Modeling of Diffusive Processes and Applications (ICMSDPA12) organized in Banaras Hindu University, India during October 9–12, 2012. There are contributors from outside ICMSDPA12 also to make the book more broad-based. This book addresses some of the issues in simulation modeling and simulation over a number of application areas. There are fifteen chapters in the book.

The first chapter is on diffusive processes and modeling: an introduction by Naveen Kumar and S. K. Basu. It deals briefly with a number of processes which are intimately connected with the diffusion processes, advection–diffusion equation (ADE) in different coordinate systems, mentioning different transformations generally used, different analytical and numerical methods. The effect of fractional order space derivative with skewness parameter on the mass transport has been explained through simulation using ADE for a simplified wound healing problem. Lastly, simulation study about the effect of ionic diffusion on the controlled release of nutrients from a coated spherical fertilizer granule is explained.

The second chapter on diffusion and transport of molecules in living cells by Ruchi Gaur, Lallan Mishra, and Susanta K. Sen Gupta deals with diffusion and different models of it, and relevance of different transport phenomena in living cells.

The third chapter on modeling diffusion and transport of suspended sediment in open channels, using two-phase flow theory by Sanjeev Kumar Jha and Fabián A. Bombardelli deals with a general framework of sediment transport in open channels as a two-phase flow, composed of mass and momentum equations for both phases (water and sediment). The authors discuss two levels of model complexity based on the nature of the terms involved in modeling: the complete two-fluid model (CTFM), and a partial two-fluid model (PTFM).

The fourth chapter on mathematical modeling of peristaltic transport of nanofluids by Dharmendra Tripathi and O. Anwar Bég reviews the challenges and potential of mathematical modeling in biofluid mechanics. The fundamentals of peristaltic transport and nanofluid dynamics have also been described qualitatively. A novel mathematical model has additionally been presented by the authors, to simulate

the influence of nanofluid and thermo-diffusive/diffuso-thermal characteristics on peristaltic heat and mass transfer in a two-dimensional axisymmetric channel for simulation of nanofluid peristaltic drug delivery systems.

The fifth chapter on numerical study on isotachophoretic separation of ionic samples in microfluidics by Partho P. Gopmandal and S. Bhattacharyya deals with a high resolution numerical algorithm to analyze two-dimensional isotachophoresis (ITP) of electrolytes of different mobility in a wide micro-channel based on a finite volume method over a staggered grid arrangement along with a higher-order upwind scheme. The model is based on equations for conservation of mass and charge and also electro-neutrality condition.

The sixth chapter on thermal characterization of non-homogeneous media by Helcio R. B. Orlande, Carolina P. Naveira-Cotta, Henrique Massard da Fonseca, Diego Knupp, Renato M. Cotta, and Olivier Fudym presents application of a Markov chain Monte Carlo (MCMC) method, within the Bayesian framework, for the identification of non-homogeneities or inclusions in a medium through the solution of an inverse heat conduction problem. They present two different approaches in conjunction with the MCMC method. A nodal approach which locally linearizes the inverse problem by using temperature measurements for the computation of the sensitivity matrix, and an expansion of unknown spatially-dependent thermophysical properties in terms of eigen functions, which is used in conjunction with the Generalized Integral Transform Technique (GITT).

The seventh chapter on scale dependent porous dispersion resulting from the cumulative effects of velocity fluctuations by Wynand S. Verwoerd deals with semi-analytical stochastic model of the dispersion effects of macroscopic drift velocity fluctuations leading to significant insights like enhancement of intrinsic dispersion by a fluctuation, beyond the value associated with flow at the mean drift velocity. This enhancement manifests as a factor multiplying the spatial variance of the solute plume, so that the effects of a sequence of fluctuations accumulate as a product, implying an exponential rise of dispersion with the distance travelled as a solute plume traverses the fluctuation sequence. This behavior is tempered by an annealing effect downstream of a velocity step, which has a length scale related to plume extension.

The eighth chapter on modeling nitrogen fate and transport at the sediment-water interfaceby M. M. Hantush, and L. Kalin deals with analytical models describing transport and fate phenomena at media interfaces. The first problem discussed is modeling of nitrogen cycling at the sediment-water interface at the bottom of lakes. The second is modeling atmospheric input of oxygen into under-saturated lakes. The third model describes polychlorinated biphenyl redistribution at the sediment-water interface.

The ninth chapter on modeling groundwater flow in unconfined aquifers by S. N. Rai deals with groundwater flow equations to describe two dimensional groundwater flows in inhomogeneous anisotropic unconfined aquifer, inhomogeneous, isotropic unconfined aquifer, in leaky unconfined aquifer, in homogeneous isotropic sloping aquifer in response to intermittently applied time varying recharge and/or pumping from multiple basins of rectangular shapes and wells, respectively along

with the initial and boundary conditions and methods of their solutions. The governing flow equations are used for the development of analytical/numerical models to predict water table fluctuations in the flow system under consideration.

The tenth chapter on two-dimensional solute transport from a varying pulse type point source by Premlata Singh, Sanjay Kumar Yadav, and Alexander V. Perig deals with solute transport originating from a source through a heterogeneous horizontal medium assuming temporal dependence of velocity and dispersivity.

The eleventh chapter on the problems of futile cycles in metabolic flux modeling: flux space characterization and practical approaches to its solution by Wynand S. Verwoerd and Longfei Mao deals with metabolic capabilities and behaviours of an organism by development of flux models of genome scale with flux balance analysis (FBA). For elimination of futile cycles in the FBA results, the authors introduce a simple notion to cut off the circulating flux layer while obtaining the same objective value. To comprehensively elucidate the alternate optimal solutions without the interference of futile values, they present flux variability analysis with target flux minimization, a combined pipeline approach based on FBA and flux variability analysis.

The twelfth chapter on contaminant concentration prediction along unsteady groundwater flow by Mritunjay Kumar Singh and Priyanka Kumari deals with the contaminant concentration pattern of one-dimensional advection-dispersion equation along a homogeneous semi-infinite aquifer with pulse type boundary condition for different forms of velocity expressions.

The thirteenth chapter on wavelet-multigrid method for solving modified Reynolds equation modeling synovial fluid flow in a normal human knee joint by S. C. Salimath deals with modified Reynolds equation, incorporating surface roughness and poroelastic nature of articular cartilage enabling bio-medical engineers in selecting suitable design parameters, giving deeper understanding of the lubrication of knee. The results obtained could guide the new material experimentation for knee replacement with mechanical characteristics.

The fourteenth chapter on a basic concept on modeling soil organic carbon by Nimai Senapati, Subhadip Ghosh, Heiko Daniel, and Amitava Rakshit discusses SOC models as important means of improving our understanding of C turnover process as well as underlying C stabilization mechanisms in soil. The SOC models often simulate the dynamics of different macro- and micro-nutrients along with SOC dynamics inadequately. They also often do not account soil pH and do not simulate the whole process of soil aggregation and the dynamics of soil biota explicitly. Inclusion of all these process/factors/parameters in the SOC models could represent the complex real life systems in a better way and might improve the overall model performance.

The fifteenth chapter on crop growth simulation modeling by Avnish Kumar Bhatia deals with crop growth models emphasizing crop physiology, weather parameters, soil parameters, and management practices to simulate growth and yield of crops. Crop simulation models compute growth values on a day to day basis using the relations among values of crop growth and weather parameters. A generic model can be developed using common crop physiological processes. Validating and fine

tuning of crop model is an important step before using it for actual prediction tasks. The author opines that future crop models should rely on improving the mechanism of interacting with environment and society.

The editors have attempted, through these chapters from different contributors, to put in one place wide ranging areas where simulation-modeling techniques are being used for better understanding of the underlying processes. The editors feel that this volume would be quite useful for researchers and advance graduate students from multiple disciplines where simulation-modeling is of major interest.

S. K. Basu
Banaras Hindu University Naveen Kumar

Acknowledgements

We wish to express our sincere gratitude to the contributors, whose contributions made possible publishing of this edited book. Thanks are also due to those contributors who responded to our requests, but their contributions could not be accommodated because of slack fitting with the title of the book. The reviewers of the chapters deserve cordial and sincere thanks.

We are also thankful to our institution, Banaras Hindu University for providing the ambience for completing the manuscript. The first editor is fortunate to have the complete support of his wife, Nilanjana and son, Subhaditya in this endeavour. The second editor has the complete support of his wife, Shashi and daughter, Paridhi. We appreciate their commendable and sincere cooperation for our work.

Contents

Chapter 1
Diffusive Processes and Modelling: An Introduction

Naveen Kumar and S. K. Basu

1.1 Introduction

Diffusion is a very common natural process occurring everywhere in physical, chemical, biological, geological systems. Considering the centrality of the diffusive process, understanding the effects of diffusion on different systems are of outmost importance. Assessment and management control of the degradation of our environment due to solute mass transport from a variety of sources of pollutants is a growing discipline. Mathematical modelling and computer simulation of these processes is nowadays one of the important approaches in quantitative analysis of different aspects of the discipline. One of the important analytical tools in this regard is the use of the advection–diffusion equation.

When a certain mass of solute is introduced in a medium, experience shows that the solute particles gradually spreads and occupies an increasing portion of the domain. If the medium is advective then this spreading is faster. This mixing and spreading is known as diffusive phenomenon (if there is no flow in the domain) or advective–diffusive phenomenon in the presence of advection. When a blob of ink is dropped in a glass of water, the water becomes coloured eventually; spreading of gas from the leakage point in the direction of wind, these are simple examples of diffusive and advective–diffusive processes, respectively.

Pollution can be classified on the basis of the medium in which it is occurring, such as air pollution, soil pollution, surface water pollution, and groundwater pollution, etc. Its source may be natural or anthropogenic. There are varieties of sources of pollution due to human activities [1–4]. One type of the source of these pollutions is a point source. Stationary point sources include volcanoes, factories, electric power plants, mineral smelters, petroleum refineries, and different small scale

N. Kumar (✉)
Department of Mathematics, Banaras Hindu University, 221005 Varanasi, India
e-mail: navkumar50s@gmail.com

S. K. Basu
Department of Computer Science, Banaras Hindu University, 221005 Varanasi, India
e-mail: swapankb@gmail.com

S. K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_1,

industries; while mobile point sources include all sorts of transport vehicles moving on road, rail, or in the air. A point source is a specific site in a medium (air, water, or soil) where the discharge of pollutant's solute particles in the form of effluents or particulate matters from such a source, enters the environment, and is transported away from the source due to diffusion and advection.

Groundwater pollution occurs due to infiltration of wastes from garbage disposal sites, septic tanks, mines, discharge from surface water bodies polluted due to industrial and municipal wastes [5–6]. Medium of advective–diffusive transport may be porous (soil field, aquifer, oil reservoir) or open medium (air, surface water bodies). In the real cases, medium is seldom homogeneous. Instead, it is heterogeneous. In the former case, the transport properties (porosity in porous medium), hydraulic conductivity, permeability remain uniform with position. In the latter case, these become position dependent or spatially dependent**.** Similarly, if these do not depend upon direction, the medium is isotropic, otherwise the medium is anisotropic. The source of advective–diffusive transport may be a point source (for example, garbage disposal sites, mines, etc.), or line source (for example, interface of sea water in aquifer), or surface source (for example, along agriculture field with high doses of chemical fertilizers) [7].

A point source may be of continuous type or pulse type. In either case, the point source may be uniform or of varying nature. In the presence of the source of pollution, the input concentration may be uniform or of increasing nature. As soon as the source of pollution is eliminated, the input concentration becomes zero or starts decreasing. Solution of a dispersion problem for a pulse type point source is useful in predicting the rehabilitation time period of a polluted domain once its source is eliminated. Smokes coming out of a chimney, wastes from a drainage system reaching a particular location in rivers, lakes, etc. are examples of uniform pulse-type point sources. As soon as the source is eliminated, the input becomes zero. Infiltration from surface point sources reaches groundwater level or oil reservoirs, degrading their quality (an example of varying pulse-type point source). As soon as the source is eliminated, the input starts decreasing, instead of becoming zero. The pollutant's solute transport from a source along the flow field through a medium of air or soil or water is described by a partial differential equation of parabolic type derived on the principle of conservation of mass, and is known as the advection–diffusion equation, also written as ADE in abbreviated form [8–11].

This chapter discusses various diffusive processes, develops the ADE equation and illustrates the use of this equation with fractional derivatives and skewness parameter for wound healing, and ionic diffusion of nitrogen phosphorus potassium (NPK) release from coated fertilizer granules.

1.2 Diffusive Processes

There are a number of processes which are intimately connected with the diffusion processes. These are Brownian motion, chemotaxis, osmosis, random walk. We briefly describe these processes in this section.

1.2.1 Brownian Motion

In the early 1800s, Robert Brown a botanist, studied pollen samples with a microscope and noted that the pollen grains exhibited movement. The phenomenon recorded by Brown is known as the Brownian motion. In essence, all atoms, ions, and molecules are in constant random motion, even those within a solid. The molecular motion in solids is not very much; the molecules simply vibrate in place. This movement is due to the collision of invisible water molecules with those particles. For example, a sphere of 1 μm in diameter in air is subjected to 10^{16} collisions per second. The distance of the particle from its initial position increases with time, although at any given moment the displacement may be either forward or backward with equal probability. This result was derived by Einstein in 1906. The diffusion coefficient (D) for a spherical particle in a liquid is related to its root mean squared displacement (x_{rms}) from the initial position at time t by $x_{rms} = \sqrt{2Dt}$. Suspended particles undergo Brownian motion and so these particles tend to move from regions of high concentration to ones with low concentration (diffusion) and this makes the concentration of suspended particles uniform over a long time.

1.2.2 Diffusion

Diffusion is the passage of particles from a region of higher concentration to a region of lower concentration. Unless physically blocked, diffusion will always occur. More technically, diffusion can be defined by chemical potential. Chemical potential is the measure of free energy available to do work to move a mole of particles from one location to another. Another way of stating diffusion is molecular movement from regions of higher chemical potential to areas of less chemical potential.

A continuous time stochastic process with (almost surely) continuous sample paths having the Markovian property is called diffusion. The simplest and most fundamental diffusion process is Brownian motion $B(t)$ (which is sometimes called the Wiener process $W(t)$). $B(t)$ is Brownian motion if it is a diffusion process satisfying (i) $B(0) = (0)$, (ii) expected value of variance $B(t) = 0$, variance $B(t) = \sigma^2 \times t$, (iii) $B(t)$ has stationary, independent increments.

1.2.3 Chemotaxis

Chemotaxis is the process where cells, bacteria, and other single-cell or multicellular organisms direct their movements according to certain chemicals in their environment. This is important for bacteria to find food by moving towards the highest concentration of food or to go away from harmful molecules. In multicellular organisms, chemotaxis is crucial for early and subsequent phases of development as well as in normal function. Leukocytes in blood move towards a region of bacterial

infection. The movement is up a chemical gradient caused by infection. There are numerous examples in biology for this chemotactic movement. In diffusion, the movement is down a concentration gradient, while in chemotaxis the movement is up a concentration gradient.

1.2.4 Osmosis

Osmosis is the movement of water molecules from a region of higher water potential to a region of lower water potential through a semipermeable membrane. Semipermeable membranes allow the passage of some materials but block the passage of others. Water potential is a measure of chemical potential of water molecules. Pure water under atmospheric pressure has potential zero. When solute is added to water, its potential becomes negative. The direction of water flow is then from greater water potential to less water potential, that is, from pure water side to the side with solutes.

Biological membranes in many cases are semipermeable allowing passage of some molecules/ions and blocking passage of others. Transport of molecules/ions across the membrane is controlled by many factors such as particle size (smaller-sized molecules/ions have greater chance of crossing the membrane), concentration of molecules/ions (the more the concentration, the greater the chance), temperature (higher temperature gives higher energy to the molecules/ions for crossing the membrane), electrical charge, pressure on the particles.

The osmotic pressure ascribed to the suspended particles is given by $pV = RTz$, where V / z, is sufficiently large, z is gram molecule of a nonelectrolyte dissolved in a volume V at temperature T [12]. Suspended particles undergo irregular movement on account of the molecular movement of the liquid according to molecular–kinetic theory of heat. The solvent exerts pressure on the suspended particles as given by:

$$p = \frac{RT}{V}\frac{n}{N} = \frac{RT}{V}C, \tag{1.1}$$

where n is the number of suspended particles present in volume V, N is the actual number of molecules contained in a gram molecule, and C is the concentration. Let the suspended particles of spherical shape with average radius a be in a liquid with viscosity μ, and D, denotes the coefficient of diffusion of the suspended particles. As a result of diffusion, $-D(\partial C / \partial x)$ particles pass across unit area in unit time, where D satisfies the diffusion equation. In dynamic equilibrium D is given by:

$$D = \frac{RT}{N}\frac{1}{6\pi a\mu}. \tag{1.2}$$

1.2.5 Random Walk

A drunken person comes out of a bar and takes discrete steps of equal lengths on the street in front of the bar. The person may take a step towards left or right randomly with equal probability. Decision for moving left or right is taken independently. The total number of different paths that a drunken person can take, given the condition that the first return be at the $2k^{\text{th}}$ step is:

$$\text{TP} = \frac{2}{k}\begin{pmatrix} 2k-2 \\ k-1 \end{pmatrix}. \tag{1.3}$$

The probability that the drunken man reaches the bar for the first time after $2k^{\text{th}}$ steps is $\text{TP}/\,2^{2k}$. Given that the drunken man is at the bar at step $2k$, the probability that this is his first return visit is $1/(2k-1)$.

Diffusion processes are intimately related to random walks. Let us consider the random walk on R (a set of real numbers). Let the initial position of a particle be $x_0 = 0$. Tossing a fair coin, if we get heads then we set $x_{i+1} = x_i - 1$, otherwise we set $x_{i+1} = x_i + 1$. This can be seen to be associated with a partial difference equation satisfied by the distributions of positions that the random walk passes at successive time steps:

$$\begin{aligned}\text{Prob}[x_i = k] - \text{Prob}[x_{i-1} = k] = \frac{1}{2}\{&\text{Prob}[x_{i-1} = k-1] \\ &+ \text{Prob}[x_{i-1} = k+1] - 2 \times \text{Prob}[x_{i-1} = k]\}.\end{aligned}$$

This is a discrete diffusion process. The heat or diffusion equation

$$\frac{\partial p_t(x)}{\partial t} = \frac{\partial^2 p_t(x)}{2(\partial x)^2} \tag{1.4}$$

causes suitable functions from R to R to evolve as a function of time [13]. The operator $\partial^2 / 2(\partial x)^2$ is the infinitesimal generator of one-dimensional Brownian motion. If $p_0(x)$ is a density function, the distribution $p_t(x)$ obtained by solving the heat equation is also the probability density of the Brownian motion witnessed at time t, if its position at zero time was chosen according to the density $p_0(x)$. Thus, Brownian motions are related to continuous time diffusions.

1.3 Advection–Diffusion Equation (ADE)

Let us consider a small cubical element of volume $dxdydz$ of sides, $PQ = dx$, $PS = dy$, $\text{PA} = dz$, surrounding a position $P(x, y, z)$ in a Cartesian three-dimensional frame of reference as shown in Fig. 1.1 in a moving fluid containing solute. Let the concentration of solute at this position be denoted by $c(x, y, z)$.

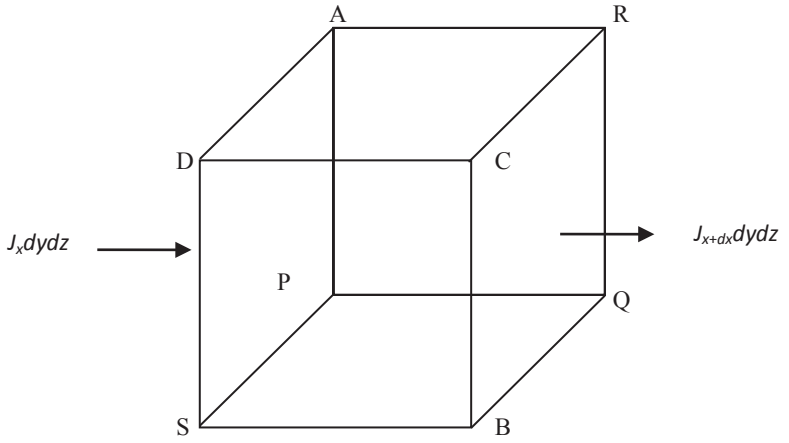

Fig. 1.1 Mass gain inside an elementary volume

Solute mass entering the elemental volume through the face PADS is $J_x dydz, J_x$, is the flux along the *x*-axis. Solute mass leaving the elemental volume through the face QBCR is $J_{(x+dx)} dydz$. Net gain inside the elemental volume along the *x*-axis is:

$$dydz(J_x - J_{(x+dx)}) = dydz[J_x - (J_x + dxJ_x' + \ldots)] = dxdydz(-\partial J_x / \partial x).$$

In the Taylor series expansion, the infinite series is truncated after the first order derivative term. Similarly, that along the y- and z-axes are $dxdydz(-\partial J_y / \partial y)$, and $dxdydz(-\partial J_z / \partial z)$, respectively. Total gain inside the elemental volume is:

$$-dxdydz\left(\frac{\partial \bar{J}_x}{\partial x} + \frac{\partial \bar{J}_y}{\partial y} + \frac{\partial \bar{J}_z}{\partial z}\right).$$

By Fick's first law of diffusion, the flux J_x is proportional to the concentration gradient, $J_x \propto -(\partial c / \partial x)$, where negative sign occurs because the positive x-axis direction is from higher concentration to lower concentration. Similarly, we have in the other two directions, $J_y \propto -(\partial c / \partial y)$ and $J_z \propto -(\partial c / \partial z)$, respectively. The diffusive current densities $\bar{J}_{diff}(J_x, J_y, J_z)$ and convective current density $\bar{J}_{conv}$ through the elemental volume are given as follows:

$$J_x = -D_x \frac{\partial c}{\partial x};\ J_y = -D_y \frac{\partial c}{\partial y};\ J_z = -D_z \frac{\partial c}{\partial z} \text{ and } \bar{J}_{conv.} = \bar{v}c, \tag{1.5}$$

respectively, where D_x, D_y, D_z are the proportionality constants, known as diffusion or dispersion parameters along the x-, y-, and z-axes, respectively and

$\bar{v}(u,v,w)$ is the flow velocity. Total current density through the elemental volume is given by:

$$\bar{J} = \bar{J}_{diff} + \bar{J}_{conv}. \tag{1.6}$$

According to conservation of mass, net rate of change of solute mass inside the elemental volume is equal to the net gain in the mass inside the volume, that is:

$$dxdydz\frac{\partial C}{\partial t} = -dxdydz\left(\frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z}\right) \tag{1.7}$$

$$\text{or} \quad \frac{\partial C}{\partial t} + \frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z} = 0. \tag{1.8}$$

Using the appropriate expressions from above, we have:

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D_x\frac{\partial c}{\partial x}\right) + \frac{\partial}{\partial y}\left(D_y\frac{\partial c}{\partial y}\right) + \frac{\partial}{\partial z}\left(D_z\frac{\partial c}{\partial z}\right) - \frac{\partial}{\partial x}(uc) - \frac{\partial}{\partial y}(vc) - \frac{\partial}{\partial z}(wc). \tag{1.9}$$

Equation 1.9 is known as the ADE in general form in three dimensions. The coefficients $D_x, D_y,$ and $D_z,$ may be the function of position, time as well as concentration. If it is not so, then these components are called dispersion coefficients. In case any one of the six coefficients is function of independent variables, the partial differential equation (Eq. 1.9) remains linear. In case any one of the coefficients depends upon the dependent variable, $c,$ the partial differential equation is nonlinear. If the velocity depends upon time at a particular position, it is said to be unsteady or temporally dependent. If it varies with position at a particular time, the velocity is said to be nonuniform or spatially dependent. If the medium is porous, velocity vector $\bar{v}$ in the ADE satisfies Darcy's law. If the medium is not porous, it satisfies the laminar conditions of flow. In case all the coefficients in Eq. 1.9 are constants, the ADE becomes:

$$\frac{\partial c}{\partial t} = D_x\frac{\partial^2 c}{\partial x^2} + D_y\frac{\partial^2 c}{\partial y^2} + D_z\frac{\partial^2 c}{\partial z^2} - u\frac{\partial c}{\partial x} - v\frac{\partial c}{\partial y} - w\frac{\partial c}{\partial z}. \tag{1.10}$$

The advection–diffusion Eq. 1.9 in one dimension along the x-axis, in general form, may be written as:

$$\frac{\partial c}{\partial t} + \frac{1-n_e}{n_e}\frac{\partial S}{\partial t} = \frac{\partial}{\partial x}\left(D_0 f_1(x,t)\frac{\partial c}{\partial x} - u_0 f_2(x,t)c\right) - \mu_1 c + \mu_2, \tag{1.11}$$

where c is the solute concentration at a position x at time $t,$ in liquid/air phase of the medium, S is the adsorbed concentration on the solid matrix of the porous medium, D_0 represents the solute diffusivity parameter, u_0 is the velocity of the

medium transporting the solute particles, n_e is the porosity of the medium. The first and second terms on the left hand side represent the rate of change in concentration in the elementary volume of the liquid and solid phases, respectively, and the terms on the right hand side represent transport due to diffusion, that due to advection, decay in concentration of first order, and growth in concentration in liquid phase of zero order, respectively. The two concentrations c and S may be related by an isotherm:

$$S = k_1 c^p + k_2,\ p = 1, \text{ represents a linear isotherm.} \tag{1.12}$$

The partial differential equation (1.11) is of parabolic type [14]. To solve it analytically or numerically three conditions are required. The partial derivative with respect to the time variable is of first order, so we need only one condition in the time domain, which is defined at only one point of the time domain. Hence, it is called an initial condition. It may be of homogeneous or nonhomogeneous types. The condition is defined usually at $t = 0$. If it is of homogeneous type, it means the domain is initially solute free; otherwise the domain is not solute free. The partial differential Eq. 1.11 has second-order derivative in space variable, so two conditions in the $x-$ domain are needed to get the particular solution. Both the conditions are usually defined at two different points, hence are termed as the boundary conditions. The first condition is usually introduced at the origin, $x = 0$ of the domain. This condition is called *input condition*. It is of first type (solution type) and of inhomogeneous nature in case the input is of uniform nature. The input condition may be of continuous nature or of pulse type. It is of the third type (mixed type) in case the input is of varying nature. In the uniform pulse-type input, the input concentration at the origin of the domain is considered uniform up to certain time period, beyond which it is assumed zero. It may also happen that the source of the input remains uniform up to certain time and after its elimination forever, the input concentration becomes zero. Such situations occur in the case of pollution sources in air and surface water bodies. The smoke coming out of a chimney may be uniform up to certain time, but as soon as the source of the smoke stops working, the input becomes zero. It may also happen that the input increases in a certain time domain due to a variety of reasons and once the source is eliminated, the input starts decreasing instead of becoming zero at once. This type of input source may occur in groundwater reservoirs, whose source of pollution is on the earth's surface and the pollutants infiltrate through the soil to reach the groundwater. This situation may be described by varying pulse-type input. The second boundary condition is introduced at the other end of the domain. It may be of the first, second (flux type), or of the third type.

The term 'free-boundary value problem' is commonly used when the boundary is stationary. Moving boundaries, on the other hand, are associated with time-dependent problems and the position of the boundary has to be determined as a function of time and space. Moving boundary problems are often called Stefan problems, with reference to the early work of J. Stefan who, around 1890, was interested in the melting of the polar ice cap [15]. Alloy solidification problems differ from the classical Stefan problems in that the melting temperature is not known in advance;

it depends on the composition of the alloy. Typically, an alloy is considered to comprise a pure substance containing small concentrations of one or more impurities. The solidification of an alloy calls for a simultaneous study of the processes of the heat flow and the diffusion of impurities.

1.3.1 Transformation Equations

To solve the ADE with constant coefficients, some transformations are used. Moving coordinate transformation equations

$$X = x - ut, \quad T = t \tag{1.13}$$

reduce the one-dimensional advection–diffusion equation

$$\frac{\partial c}{\partial t} = D_x \frac{\partial^2 c}{\partial x^2} - u \frac{\partial c}{\partial x} \tag{1.14}$$

into the diffusion equation

$$\frac{\partial K}{\partial T} = D_x \frac{\partial^2 K}{\partial X^2}. \tag{1.15}$$

where $K(X,T)$ is the new dependent variable in the new space and time variables. The same diffusion equation in the same independent variables may also be obtained by applying another transformation

$$c(x,t) = K(x,t) \exp\left(\frac{ux}{2D} - \frac{u^2 t}{4D} \right), \tag{1.16}$$

on the ADE (Eq. 1.14). A transformation $X = x - \lambda t$ (similar to that given in Eq. 1.13) reduces the ADE (Eq. 1.14) into an ordinary differential equation [16]

$$D_x \frac{d^2 c}{dX^2} - (u - \lambda) \frac{dc}{dX} = 0. \tag{1.17}$$

A transformation known as similarity transformation

$$X = \frac{x}{\sqrt{t}} \tag{1.18}$$

reduces the diffusion equation (Eq. 1.15) into an ordinary differential equation

$$D_x \frac{d^2 K}{dX^2} + X \frac{dK}{dX} = 0 \tag{1.19}$$

for $D_x = D_y$, the two-dimensional diffusion equation in Cartesian system of coordinates

$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2}\right) \tag{1.20}$$

reduces into the radially symmetric polar system

$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial r^2} + \frac{1}{r}\frac{\partial c}{\partial r}\right) \tag{1.21}$$

by using the transformation $x^2 + y^2 = r^2$. Similarly, a transformation $x^2 + y^2 + z^2 = r^2$ reduces the three-dimensional diffusion equation

$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2}\right) \tag{1.22}$$

into radially symmetric diffusion equation in spherical system of coordinates

$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial r^2} + \frac{2}{r}\frac{\partial c}{\partial r}\right) \tag{1.23}$$

A new independent variable X is introduced, using an operator:

$$\frac{\partial}{\partial X} = f_1(x,t)\frac{\partial}{\partial x} - f_2(x,t). \tag{1.24}$$

Operating it on a dependent variable ϕ, we get a linear first order partial differential equation as:

$$f_1(x,t)\frac{\partial \phi}{\partial x} - \frac{\partial \phi}{\partial X} = f_2(x,t)\phi. \tag{1.25}$$

It is equivalent to a system of three ordinary differential equations:

$$\frac{dx}{f_1(x,t)} = \frac{dX}{-1} = \frac{d\phi}{f_2(x,t)}. \tag{1.26}$$

One solution of it is:

$$X = -\int \frac{dx}{f_1(x,t)} \quad \text{or} \quad \frac{dX}{dx} = -\frac{1}{f_1(x,t)}. \tag{1.27}$$

Introduction of this transformation with a suitable form of $f_1(x,t)$ helps to reduce the variable coefficients of ADE (Eq. 1.3) into constant coefficients; hence enables

us to use the Laplace integral transformation technique (LITT), which is the simplest and most useful among the analytical methods being used. The transformation may be modified according to the need, for example, negative sign may be omitted [17–18].

In another diffusive process, known as chemotaxis in which the mass movement is from lower concentration towards higher concentration, the flux components in Eq. 1.5 will be positive, and may be referred to as chemotactic flux. Hence, the mass conservation Eq. 1.15, in the presence of both diffusion and chemotaxis, may be derived as:

$$\frac{\partial c}{\partial t} = D_x \frac{\partial^2 c}{\partial x^2} - \rho_x \frac{\partial^2 c}{\partial x^2}, \tag{1.28}$$

where the dependent variable c and the coefficient ρ_x may be termed as the cell density and chemotactic coefficient, respectively.

1.3.2 Dispersion Theories

There are three theories which relate the dispersivity and velocity parameters occurring as the two coefficients in the one-dimensional ADE (Eq. 1.11), mostly applicable in porous medium. These are as follows:

i. Ebach and White [19], Bear [8] in their one-dimensional analysis suggested that D is proportional to u.
ii. Taylor [20] in his one-dimensional analysis obtained D proportional to u^2. Scheidegger [21] summarized his analysis on the two possible relationships between D and u according to the role played by molecular diffusion: (a) $D \approx \alpha u^2$, where α, a constant of the porous medium alone (dynamic dispersivity), is derived by a dynamic procedure applicable if there is enough time in each flow channel for appreciable mixing to take place by molecular transverse diffusion; (b) $D \approx \beta u$, where β, another constant of the porous medium (geometric dispersivity), is derived by a geometric procedure applicable where there is no appreciable molecular transverse diffusion from one streamline into another. Thus, in all the models in which the combined effect of a velocity distribution across a channel and transverse molecular diffusion are considered [20], the coefficient of dispersion is proportional to u^2. Disregarding molecular diffusion for the situation where only mean motion in a channel is considered and mixing occurs at junctions connecting different channels, one obtains $D \approx u$. Later, Freeze and Cherry [22] modified these dispersion theories and considered dispersion parameter proportional to the velocity raised to a power n, where n ranges between 1 and 2.
iii. According to Matheron and de-Marsily [23], some large subsurface formations exhibit variable dispersivity properties described by a variable D as a function of position or time variables, while the flow domain remains uniform. Such

variations may be caused, for example, by multiple length scales in these formations. Such formations are often modelled by a scale-dependent dispersion coefficient in the diffusive flux term in the transport equation. This theory was fully supported in the later works [23–26].

1.3.3 Why Modelling?

The idea of simulating real system or process on a computer rapidly spread among researchers since 1960s. It is accepted as one of the powerful tools for understanding processes and systems for predicting functional or operative conditions. The standard approach is numerically solving a mathematical model that governs a chosen process or system. Generally, the analytical solution is not known or difficult to reach for such situations. Solving mathematical models strongly depends on the used computational techniques and resources.

Shannon defined model as a representation of an object, a system, or an idea in some form other than that of the entity itself. We generally distinguish between two types of models: physical such as scale models, prototype plants, etc., and mathematical models such as partial differential equations (PDE), queuing models, etc. Simulation of a system is the operation of a model, which is a representation of that system. The model is amenable to manipulation which would be impossible, too expensive, or too impractical to perform on the system which it portrays. The operation of the model can be studied, and, from this, properties concerning the behaviour of the actual system can be inferred. Simulation is imitation of the operation of a real-world process or system over time. It generates an artificial history of a system; based on the observation of that artificial history, inferences concerning the operating characteristics of the real system can be drawn. A simulation can be only as good as the simulation model is. A simulation model makes a set of assumptions concerning the operation of the system and is expressed as mathematical, logical, or symbolic expressions between the entities (objects of interest) of the system. From the simulation, data are collected as if a real system was being observed. There are many applications such as designing and analysing manufacturing systems, determining ordering policies for an inventory system, designing communications systems and message protocols, drug design, analysing financial systems, and many more.

1.3.4 Review of Modelling Efforts in Diffusive Processes

Pollutants originating from a variety of natural and anthropogenic sources (volcano, industries, factories, refineries, sewage system, garbage disposal sites, mines, etc.) are major causes of degradation of the environment, air, surface water, soil and groundwater. Mathematical modellers use the ADE to describe the concentration levels at different positions and time, away from its source, through its analytical

and numerical solutions. It is possible to solve this equation analytically only in some particular cases. In a more general situation, numerical techniques are required. A number of analytical methods are reviewed in [27]. Most of the analytical solutions for advective–diffusive transport problems in ideal conditions with growth and decay terms, subject to various initial and boundary conditions in semi-infinite or finite media have been compiled in [28–30].

The mechanisms of solute transport and reaction have been studied for many years, and a wide variety of numerical techniques have been developed and successfully applied in many settings. Most of these studies are based on the application of the advection–dispersion–reaction equation (ADRE). Many researchers, recognizing the importance of appropriately treating the hyperbolic part of the ADRE, have utilized various types of characteristic-based solution techniques that do not suffer from nonphysical behaviour such as oscillations at the sharp fronts [31–34]. The two-dimensional multispecies reactive transport in saturated and unsaturated porous media was simulated using Eulerian–Lagrangian localized adjoint methods (ELLAM) [35]. Its applicability and efficiency were assessed by comparing the results with those obtained using a numerical model based on the combination of discontinuous Galerkin and multipoint flux approximation methods.

Following the theories in [20–22] relating D and u in one-dimensional ADE, the number of mass transport studies has increased considerably. Many such models concern homogeneous media, but in reality the ability of the mass to permeate though the medium of air, soil or groundwater varies with position, which is referred to as heterogeneity. Early efforts to describe heterogeneity were achieved by making use of stratification and defining porosity–distance relationship [36–40]. In the former situation, the larger number of layers makes it difficult to get the desired analytical solution. In the latter situation, a numerical method is the only option to deal with the dispersion problems with most of the porosity–distance relationships. Later scale-dependent dispersion has been attributed to heterogeneity. According to the theory [41], some large subsurface heterogeneous formations exhibit variable dispersivity as a function of position or time variable. Based on such observations, analytical solutions to solute transport problems in a semi-infinite medium were obtained [42–45], where the dispersion parameter depends on distance and increases up a limited value. In the third problem, first order reaction coefficient was considered space dependent in the one-dimensional transport of solute through soil. Later an integral expression for a similar problem without using the modified Bessel functions was proposed [45]. Other authors have used simplified one- and two-dimensional models that incorporate variable coefficients to some extent [46–49].

A numerical ADE model with a hyperbolic asymptotic distance-dependent function (HAD) for the dispersion coefficient was proposed and used [50–51]. In a latter study, HAD is adopted and incorporated into the general ADE for describing scale-dependent solute transport in porous media. The problem is solved analytically by applying extended power series method coupled with Laplace transform. The quadrupole method was implemented in order to simulate the effects of heterogeneities on one-dimensional ADE of a passive solute in porous media [52]. Exact solutions of the linear advection–diffusion transport equation with constant and variable

coefficients with different forms for both, transient and steady-state regimes was presented by Guerrero and colleagues [27] by using the generalized integral transform technique (GITT). In approximate solutions [53–54], a time-dependent dispersion coefficient was used. Usually, solute transport models assume a constant dispersion coefficient that is calibrated separately for each different downstream sample location, resulting in different dispersion coefficients for the same flow problem. In an attempt to overcome this, the dispersion coefficient as a function of the mean travel distance was used successfully [55]. Another approach is to model dispersivity as a time-dependent function. Based on the numerical results [56], dispersivity has been suggested to have a time-dependent behaviour which reaches asymptotic values after a long time. Analytical solution for time-varying dispersion coefficients have been presented in one dimension [57–58], in two dimensions [59–60], and in three dimensions [61–62]. Presently, the fractional advection–diffusion equation (FADE) is being used to model anomalous transport [63–66].

The importance of ADE is not confined to hydrology and soil sciences only. It has equal importance in other fields as well, some of which are mentioned below:

i. Petroleum engineering: Displacement of oil with gas; petroleum and natural gas production [67].
ii. Chemical engineering: Flow in packed columns involving chemical reactions or separation of chemical components; pore diffusion of gases, chromatography, ion-exchange [68–69].
iii. Modelling of flood waves: Modelling of flood waves in free-flowing rivers, which are more commonly bulk waves. Ferrick [70] classified these waves into (a) diffusion wave and (b) kinetic wave. A diffusion wave is governed by ADE (Eq. 1.11), where dependent variable, c may be replaced by a suitable variable y representing depth of flow; $u =$ wave celerity; $D =$ diffusion coefficient. In case, $(\partial^2 y / \partial x^2) \approx 0$ but $(\partial y / \partial x) > 1$ but, Eq. 1.11 reduces to kinetic wave governed by the kinetic equation [71]. It is to be noted that presence of $(\partial y / \partial x)$ is responsible of inducing diffusion, thus leading to attenuation of a flood wave.
iv. Wound Healing: In case of dermal wound, the variables in ADE (Eq. 1.11) will be: $c =$ cell density, or chemical concentration at position x, and time t, $u =$ convection due to formation of extra cellular matrix (ECM). Readers can find mathematical models using ADE for wound healing in [72–73].

1.4 Adevection-Diffusion with Fractional Derivatives

In order to simulate the memory formalism [74], the partial differential equation of fractional order space derivatives may be more useful. Skewness in diffusion may also be considered through a suitable parameter. For example, in wound-healing process, the density of cells and the chemical concentration at the centre and at a position close to the centre of the wound are decided by those at the preceding

adjacent position towards the wound margin. Also, a cell may take a skewed path instead of curvilinear path, and it may be captured through a distinct parameter. Following the work in [75–80], we may use fractional order derivative in place of integer order derivative as:

$$\frac{\partial}{\partial x} \approx \left\{ \frac{1+\beta}{2} \frac{\partial^{\alpha-1}}{\partial (x)^{\alpha-1}} - \frac{1-\beta}{2} \frac{\partial^{\alpha-1}}{\partial (-x)^{\alpha-1}} \right\}$$

$$\text{So} \quad \frac{\partial^2}{\partial x^2} \approx \left\{ \frac{1+\beta}{2} \frac{\partial^{\alpha}}{\partial (x)^{\alpha}} - \frac{1-\beta}{2} \frac{\partial}{\partial (-x)} \frac{\partial^{\alpha-1}}{\partial (-x)^{\alpha-1}} \frac{d(-x)}{d(x)} \right\} \tag{1.29}$$

$$\text{or} \quad \frac{\partial^2}{\partial x^2} \approx \left\{ \frac{1+\beta}{2} \frac{\partial^{\alpha}}{\partial (x)^{\alpha}} + \frac{1-\beta}{2} \frac{\partial^{\alpha}}{\partial (-x)^{\alpha}} \right\}, \tag{1.30}$$

where α is the fractional order of the derivative, and β is the skewness parameter. For the *n*th integer order, α is in the range $n-1 \le \alpha \le n$. For $\alpha = 2$ and $\beta = 0$ (no skewness), approximation sign ($\approx$) may be replaced by equality sign. As α decreases from 2, the Leúy probability distribution (LPD) deviates from the Gaussian distribution and the tail of the Leúy distribution becomes heavier. If β is less than zero, the dispersion is skewed backward representing a slowly evolving contaminant plume followed by a heavy tail. For β greater than zero, the dispersion is skewed forward describing a fast evolving contaminant plume followed by a light tail. Many other works on space-fractional partial differential equations refer to the same equation with minor changes. The finite difference approximations of the fractional order derivatives [81–82] are:

$$\frac{\partial^{\alpha} y}{\partial (+x)^{\alpha}} = \frac{1}{h^{\alpha}} \sum_{k=0}^{i+1} g_k y_{i-k+1}, \quad \frac{\partial^{\alpha} y}{\partial (-x)^{\alpha}} = \frac{1}{h^{\alpha}} \sum_{k=0}^{K-i+1} g_k y_{i+k-1} \tag{1.31}$$

where $g_k = \dfrac{\Gamma(\alpha+1)}{\Gamma(k+1)\Gamma(\alpha-k+1)}$, Γ is the gamma function, and h is the uniform size of the intervals into which the spatial axis is divided.

The classical advection–diffusion equation is mathematically identical to the diffusion equation with drift, and furthermore, the same random walk model underlies both. The mean jump size determines the velocity *v* of the advective drift. This connection between random walk and diffusion is due to Einstein. When the variance of the particle jumps is infinite, the resulting plume follows a stable concentration curve, the solution to an ADE with space derivatives having fractional order. This plume has skewness and a power law leading edge. Random waiting times do not affect the eventual shape of the plume as long as the waiting times have finite mean. When the mean waiting time is infinite, the time derivative has fractional order in the ADE.

1.4.1 Application of Fractional Order Derivative in Wound Healing

We demonstrate the effect of fractional order space derivative and skewness parameter on the dependent variable of the ADE. For this, we consider the problem of epidermal wound healing. The epidermal wound area is assumed to be circular. The healing takes place due to convective and diffusive cell migration. The source is at the boundary of the circular wound region. The growth factors (chemicals) infuse inside the wound region, as a result of which one annular domain, ECM scaffolding, takes place adjacent to the wound boundary. The convection is due to movement of the ECM front towards the centre. As the wound is supposed to be confined to epidermis, the role of blood vessels, oxygen, chemotaxis, etc. is not taken into account. The production term in the convective diffusive equation is assumed to be a function of chemical concentration, while that in the diffusive equation is supposed to be a function of cell density. The loss terms in both the equations are supposed to depend upon the respective dependent variables. The partial differential equations for the diffusion of chemicals (concentration c) and the cells (density n) in polar coordinate system and in nondimensional form are

$$\frac{\partial c}{\partial T} = D\frac{\partial^2 c}{\partial R^2} + \frac{D}{R}\frac{\partial c}{\partial R} + \lambda_c f_c(c) f_1(n) - \mu_n f_2(c),\ 1 > R > 0, \tag{1.32}$$

and

$$\frac{\partial n}{\partial T} = \frac{\partial^2 n}{\partial R^2} - Pe\frac{n}{R} - \left(Pe - \frac{1}{R}\right)\frac{\partial n}{\partial R} + \lambda_n f_n(n) f_3(c) - \mu_n f_4(n),\ 1 > R > 0. \tag{1.33}$$

The nondimensional variables used in above equations are (the asterisks are omitted in the PDEs):

$$n^* = n/n_0,\ c^* = c/c_0,\ R = r/a,\ T = D_{n0}t/a^2,\ Pe = au_0/D_{n0},\ D = D_{c0}/D_{n0},$$
$$\lambda_n^* = \lambda_n a^2/D_{n0},\ \lambda_c^* = \lambda_c a^2/D_{n0},\ \mu_n^* = \mu_n a^2/D_{n0},\ \mu_c^* = \mu_c a^2/D_{n0},$$

where n_0 is the cell density at the unwounded state, c_0 is the chemical concentration at a position adjacent to the wound boundary just after the inflammatory phase (which is at t_0), D_{c0} and D_{n0} are the uniform diffusion parameters of chemicals and the cells, λ and μ with appropriately suffixed, are scalars controlling the production and the loss of cell density and growth factor, respectively, r is the radial direction from the centre ($r = 0$) towards the boundary ($r = a$) of the circular wound of radius a. Cell density and chemical concentration inside the wound region are zero just after the occurrence of wound $(t = 0)$. At the boundary, cell density is the same as if there was no wound. As soon as the wound has occurred, chemicals start infusing inside the region, its level being the maximum adjacent to the boundary by the end of the inflammatory phase. With the progress in the healing process,

chemical concentration level lowers down at the boundary and at the interior positions. The input chemical concentration is formulated in this way. At the other end (at the centre), the gradients in both the concentrations will be zero. During the time $0 < t < t_0$, chemical concentration level at the boundary attains a value higher than c_0. This chemical concentration (growth factors) is instrumental in the formation of ECM front near the boundary and its movement towards the centre of the wound area. Thus, the initial and boundary conditions for both the partial differential equations may be written as:

$$n(R,0) = 0; \quad c(R,0) = 0; \quad 1 > R \geq 0, \tag{1.34}$$

$$n(1,T) = 1; \quad c(1,T) = \exp(1 - T/T_0); \quad T > 0, \tag{1.35}$$

$$\frac{\partial n}{\partial R} = 0; \quad \frac{\partial c}{\partial R} = 0, \text{ at } R = 0; \quad T \geq 0, \tag{1.36}$$

where T_0 is the nondimensional form of t_0. We assume that $f_c(c)$ in Eq. 1.32 increases as $c \to c_0$, and also $f_n(n)$ in Eq. 1.33 increases as $n \to n_0$. The functions in production and loss terms of both the Eqs. 1.32 and 1.33 are assumed to be of increasing nature. An exponential function is a general form to represent the changes; as it may be reduced to linear, quadratic, or other higher degree expressions under different approximations as required in modelling of many natural processes. Since the nondimensional dependent variables (n, c) in our model are always in the range 0–1, this limits the rapid growth of exponential functions. We consider:

$$f_n(n) = \frac{1}{n - n_0}, \quad f_1(n) = \exp(\sqrt{n}), \quad f_2(c) = \sqrt{c},$$

$$\text{and} \qquad f_c(c) = \frac{1}{c - c_0}, \quad f_3(c) = \exp(\sqrt{c}), \quad f_4(n) = \sqrt{n}.$$

To demonstrate the effects of fractional order derivative (α) and skewness (β) on the dependent variable, we performed computational experiments to study the two effects on the cell migration from the wound boundary towards its centre in the domain ($a \geq r \geq 0$). The integer order space derivatives in Eqs. 1.32 and 1.33 are replaced by fractional order derivatives as in Eqs. 1.29 and 1.30. Their approximations are used from Eq. 1.31. The input values are chosen as: wound radius $a = 1.0$ cm, cell density at unwounded state $n_0 = 1.0$, chemical concentration at the end of inflammatory phase $c_0 = 1.0$, convective velocity $u_0 = 0.00001$ cm/day, proliferation period $t^* = 21$ days, inflammatory period, $t_0 = 3.5$ days, all the λs, and μs, are assigned a uniform value $0.001(\text{day})^{-1}$, as this is our baseline model, the uniform step sizes along the $R-$ and $T-$ axes are chosen as $h = 0.1, \Delta T = 0.0001$ respectively, satisfying the stability criterion. The diffusion coefficient values for the cell and the chemical concentration are taken from some of the works cited above. Biologically plausible values of D_{n0} are in the range 3.5×10^{-11} to $6.9 \times 10^{-9}\,\text{cm}^2/\text{sec}$ [83], depending upon the type of the cell. Higher value of $D_{n0} = 5 \times 10^{-7}\,\text{cm}^2/\text{sec}$

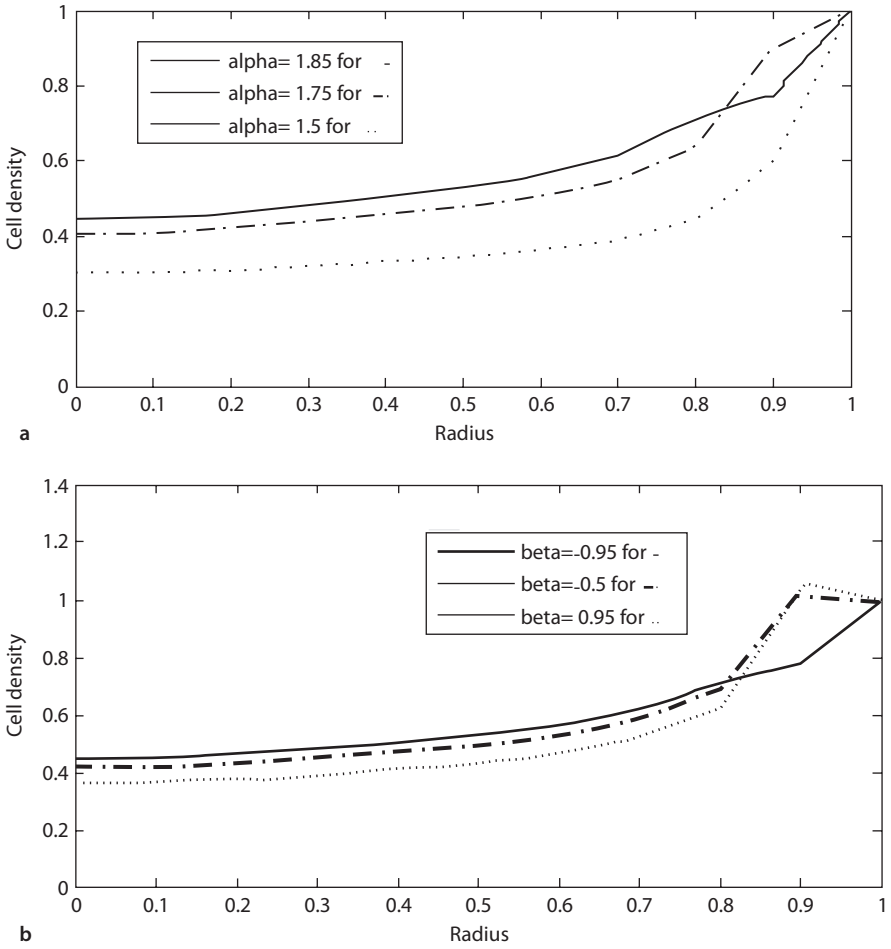

Fig. 1.2 **a** Effect of fractional order of space derivatives on cell migration. **b** Effect of skewness parameter on cell migration

has been considered elsewhere [84]. The value of diffusion coefficient for chemical concentration (growth factors) is suggested to be higher than that of the cells. We have used $D_{n0} = 0.0005\text{cm}^2/\text{day} \approx 0.0005/100000 \approx 5.0\times10^{-9}\,\text{cm}^2/\text{sec}$, and 0.0005 cm^2/day. The effect of the fractional order α is studied for $\alpha = 1.85$, 1.75, and 1.5 and is shown in Fig. 1.2. As the order of the space derivative approaches 2 (LPD is closer to the Gaussian distribution), wound at a particular position is healed in a better way. This trend is not observed in the vicinity of the wound boundary (where cell density is 1.0). This is because we have considered an ideal situation by assuming cell diffusion and convective parameters uniform. In the real situation, both will depend upon chemical concentration; in a recent work (to be communicated soon) we have not found this reversal. The effect of skewness on the

cell density is studied by performing the simulations for $\beta = -0.95, -0.5, +0.95$, and is shown in Fig. 1.3. It may be observed that the cell density has higher values in the backward skewness domain $(-1 \le \beta \le 0)$ compared with those in the forward skewness domain $(0 \le \beta \le 1)$. This effect of β is indistinguishable close to the wound boundary, whereas it is pronounced towards the centre of the wound region. The reversal in this trend near the wound boundary is due to the same reason stated above.

1.5 Ionic Diffusion

We illustrate ionic diffusion with the example of NPK release from coated fertilizer granules. Let us consider a spherical coated fertilizer granule which contains three types of nutrients, one is a nonelectrolyte, other is a weak electrolyte, and the third is a strong electrolyte, the diffusion pattern of all of these three substances are different. The unsteady diffusion equations in a spherical domain for the two species, ions and molecules of a weak electrolyte may be written as [85]:

$$\frac{\partial C_1}{\partial t} = D_1\left(\frac{\partial^2 C_1}{\partial r^2} + \frac{2}{r}\frac{\partial C_1}{\partial r}\right) - 2k, \tag{1.37}$$

and

$$\frac{\partial C_2}{\partial t} = D_2\left(\frac{\partial^2 C_2}{\partial r^2} + \frac{2}{r}\frac{\partial C_2}{\partial r}\right) + k, \tag{1.38}$$

respectively, where k is the rate of formation of the molecules (the dimmers), C_1 and C_2 are the concentrations; D_1 and D_2 are the diffusion coefficients of the ions and the molecules, respectively. The diffusing coefficient of the ions is expressed in terms of diffusion coefficients of both the ions as:

$$D_1 = \frac{|z_c| + |z_a|}{\dfrac{|z_a|}{D_c} + \dfrac{|z_c|}{D_a}}. \tag{1.39}$$

In the case of a non 1-1 electrolyte, where $|z_c|$ and $|z_a|$ are the ionic charges; D_c and D_a are the diffusion coefficients, of the cation and anion, respectively. In the case of a 1-1 electrolyte, where $|z_c| = |z_a|$, we have:

$$D_1 = \frac{2}{\dfrac{1}{D_c} + \dfrac{1}{D_a}}. \tag{1.40}$$

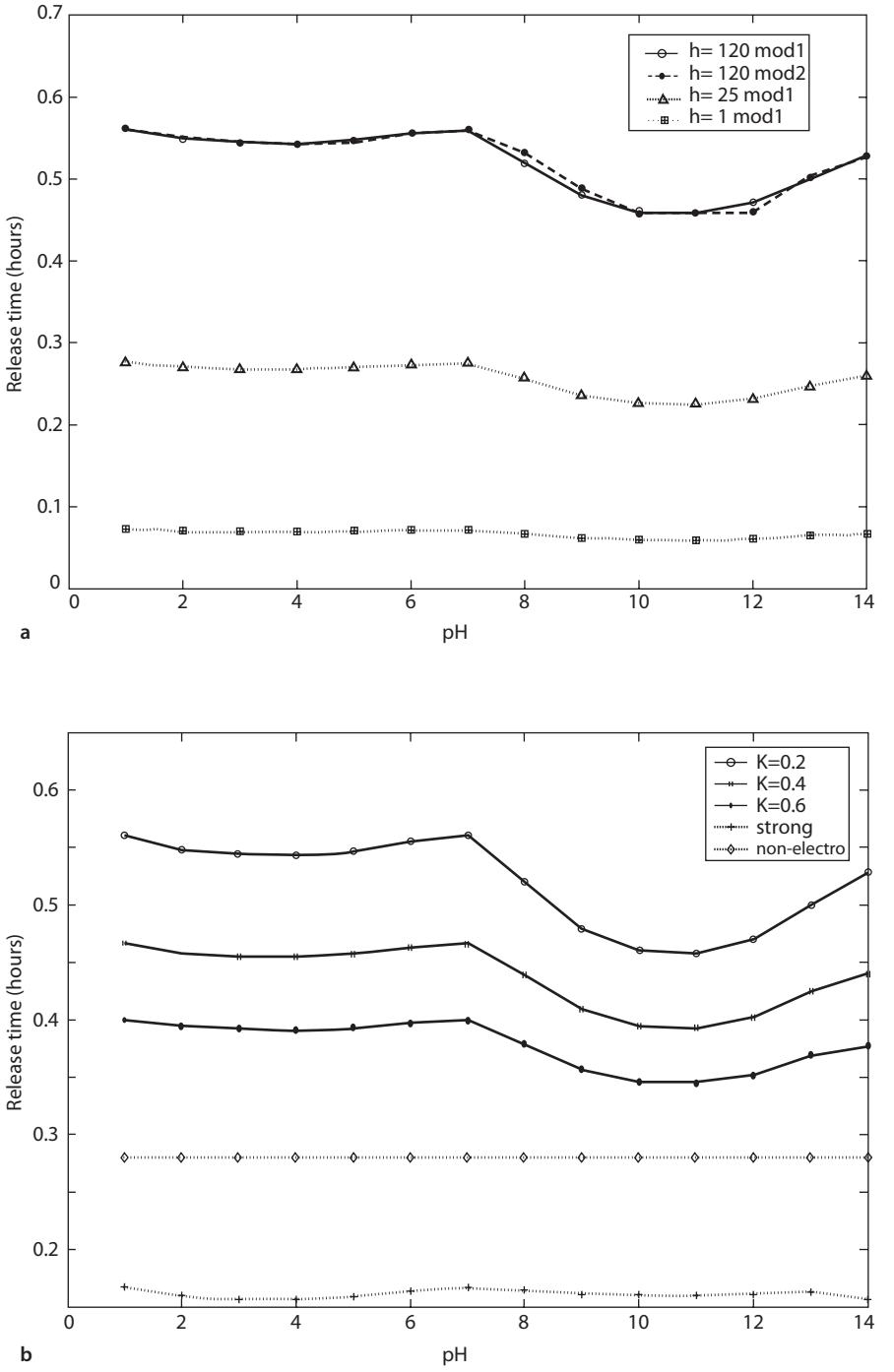

Fig. 1.3 **a** Effect of surface contact area on release time with pH. **b** Effect of association constant of electrolyte nutrient on release time with pH

In an isodesmic model [85]

$$C_2 = KC_1^2, \tag{1.41}$$

where K is an association constant for the diffusing species, which is independent of the size of the aggregate. The total concentration, C_w for weak electrolyte may be written [85] as:

$$C_w = C_1 + 2C_2. \tag{1.42}$$

Using these equations the diffusion equation in terms of one dependent variable, C_1 may be written as:

$$(1+4KC_1)\frac{\partial C_1}{\partial t} = (D_1 + 4KD_2C_1)\left(\frac{\partial^2 C_1}{\partial r^2} + \frac{2}{r}\frac{\partial C_1}{\partial r}\right) + 4KD_2\left(\frac{\partial C_1}{\partial r}\right)^2. \tag{1.43}$$

This is a nonlinear partial differential equation. It suggests strong dependence of the apparent diffusion coefficient of the diffusing substance on the ionic or molecular concentration due to significant interactions among the diffusing species. The linear diffusion equation for a nonelectrolytic substance for diffusion parameter, i.e. $D(C) = D$, in spherical coordinate system may be written as:

$$\frac{\partial C}{\partial t} = D\left(\frac{\partial^2 C}{\partial r^2} + \frac{2}{r}\frac{\partial C}{\partial r}\right). \tag{1.44}$$

A strong 1-1 electrolyte ionizes completely producing equal number of cations and anions. Although the concentrations of cations and anions C_c and C_a, respectively, may vary through the solution, the concentrations and the concentration gradients of these species are equal everywhere because of electro-neutrality, that is:

$$C_c = C_a \quad \text{and} \quad \nabla C_c = \nabla C_a. \tag{1.45}$$

In the case of strong non 1-1 electrolyte, constraints on the concentration and flux at zero current are:

$$z_c C_c + z_a C_a = 0 \quad \text{and} \quad z_c \nabla C_c + z_a \nabla C_a = 0. \tag{1.46}$$

In the case of 1-1 electrolyte, $z_c = -z_a$. The total concentration of strong electrolyte, C_s is

$$C_s = C_c / |z_a| = C_a / |z_c|. \tag{1.47}$$

The diffusion equation of a strong electrolyte is

$$\frac{\partial C_s}{\partial t} = D_s\left(\frac{\partial^2 C_s}{\partial r^2} + \frac{2}{r}\frac{\partial C_s}{\partial r}\right), \tag{1.48}$$

where the diffusion coefficient, D_s has either of the two expressions of Eq. 1.39 or Eq. 1.40, depending on whether the electrolyte is a non 1-1 or a 1-1 electrolyte. The diffusion equation for a nonelectrolyte is the same as Eq. 1.44 with appropriate meaning of the variables.

Diffusion coefficients of cation and anion depend upon the pH of the soil water. As a result, release time of an electrolytic nutrient depends upon pH. The authors have studied the effects of different parameters such as radius of the granule, its surface area in contact with the soil determined by h, association constant (K), pH on the release time in their recent work [86]. The two figures from this work are being given here as Figs. 1.3a and 1.3b for the readers understanding about the ionic diffusion. In Fig. 1.3a. $h = 1$ represents the basal form of granule application (the granule is totally below the soil surface); hence, the release time is the minimum, while $h = 120$ corresponds to the almost point contact of the granule with the soil.

1.6 Summary

Diffusion is not confined to a particular discipline. Its various forms in different disciplines are explained. The derivation of advection–diffusion equation describing the mass transport through a medium is given. Its forms in different coordinate systems are also given. Different transformations being used by various workers in a variety of disciplines are mentioned. How the heterogeneity of the medium and unsteadiness of advection are addressed by the two parameters of the ADE for reactive and nonreactive solute mass, using the different dispersion theories with different analytical and numerical methods are explained. Most importantly, the effect of fractional order space derivative with skewness on the mass transport, have been explained through the wound healing problem. Lastly, the effect of ionic diffusion on the controlled release of nutrients from a coated spherical fertilizer granule is given.

References

1. Bear J (1979) Hydraulics of groundwater. McGraw-Hill, New York
2. Bear J (1988) Dynamics of fluids in porous media. Dover Publications, Inc, New York
3. Fried JJ (1975) Groundwater pollution. Elsevier Scientific Publishing Co., New York
4. Todd DK (1980) Groundwater hydrology. Wiley, New York
5. Lohman SW (1972) Groundwater hydraulics, U.S. geological survey, Prof. Paper 708,70
6. Bouwer H (1979) Groundwater hydrology. McGraw-Hill, New York
7. Marshall TJ, Holmes JW, Rose CW (1996) Soil physics. Cambridge University Press, Cambridge
8. Bear J (1972) Dynamics of fluids in porous media. American Elsevier Co., New York
9. Bear J, Bachmat Y (1990) Introduction to modeling of transport phenomena in porous media. Kluwer, Netherlands
10. Cushman JH (1990) Dynamics of fluid in hierarchical porous media. Academic Press, London

11. Nield DA, Began A (1992) Convection in porous media. Springer—Verlag, Berlin
12. Einstein A (1956) Investigations on the theory of the Brownian movement. Edited with notes by R. Fürth, Translated by A. D. Cowper, Dover Publications
13. Itô K, McKean HP Jr (1965) Diffusion processes and their sample paths. Grundlehren Math. Wiss. 125. Springer—Verlag, Berlin
14. Guenther RB, Lee JW (1988) Partial differential equations of mathematical physics and integral equations. Prentice Hall Englewood Califfs, New Jersey
15. Crank J (1984) Free and moving boundary problems. Oxford University Press, Oxford
16. Manoranjan VS, Stauffer TB (1996) Exact solutions for contaminant transport with kinetic langmuir sorption. Water Resour Res 32:749–752
17. Kumar A, Jaiswal DK, Kumar N (2009) Analytical solutions to one—ssional advection—diffusion equation with variable coefficients in finite domain. J Earth Syst Sci 118:539–549
18. Jaiswal DK, Kumar A, Kumar N, Singh MK (2011) Solute transport along temporally and spatiall dependent flows through horizontal semi-infinite media: dispersion proportional to square of velocity. ASCE J Hydrol Eng 16(3):228–238
19. Ebach EH, White R (1958) Mixing of fluid flowing through beds of packed solids. J Am Inst Chem Eng 4:161–164
20. Taylor GI (1953) Dispersion of soluble matter in solvent flowing slowly through a tube. Proc Royal Soc Lond A219:186–203
21. Scheidegger A (1957) The physics of flow through porous media. University of Toronto Press, London
22. Freeze RA, Cherry JA (1979) Groundwater. Prentice-Hall, New Jersey
23. Dagan G (1988) Time—dependent macro-dispersion for solute transport in anisotropic heterogeneous aquifers. Water Resour Res 24:1491–1500
24. Barry DA, Sposito G (1989) Analytical solution of a convection—dispersion model with time dependent transport coefficients. Water Resour Res 25:2407–2416
25. Gelhar LW (1992) Stochastic subsurface hydrology. Prentice Hall Englewood Cliffs, New Jersey
26. Kulasiri D, Verwoerd WS (2002) Stochastic dynamics—modeling solute transport in porous media. Elsevier, New York
27. Guerrero JSP, Pimentel LCG, Skaggs TH, van Genuchten MTh (2009) Analytical solution of the advection-diffusion transport equation using a change-of-variable and integral transform technique. Int J Heat Mass Transf 52:3297–3304
28. Javandel I, Doughty C, Tsang CF (1984) Groundwater transport: handbook of mathematical models. Am Geoph Union Water Resources Monograph Series 10, American Geophysical Union, Washington
29. van Genuchten MTh, Alves W J (1982) Anaytical solutions of the one—dimensional convective—dispersive solute transport equation, USDA ARS Technical Bulletin Number 1661, U.S. Salinity Laboratory
30. Domenico PA, Schwartz FW (1998) Physical and chemical hydrogeology. Wiley, New York
31. Pinde GF, Cooper HH (1970) A numerical technique for calculating transient position of saltwater front. Water Resour Res 6(3):875
32. Russell TF, Celia MA (2003) An overview of research on Eulerian–Lagrangian Localized Adjoint Methods (ELLAM). Adv Water Resour 25:1215
33. James AI, Jawitz JW (2007) Modeling two—dimensional reactive transport using a godunov—mixed finite element method. J Hydrol 338:28–41
34. Rubio AD, Zalts A, El Hasi CD (2008) Numerical solution of the advection—reaction—diffusion equation at different scale. Environ Model Softw 23:90–95
35. Ramasomanana F, Younes A, Fahs M (2012) Modeling 2D multispecies reactive transport insaturated/unsaturated porous media with the Eularian—Lagrangian Localized Adjoined Method. Water Air Soil Pollut 223:1801–1813
36. Coats KH, Smith BD (1964) Dead-end pore volume and dispersion in porous media. Soc Petrol Eng J 4:73–84

37. Lin SH (1977) Nonlinear adsorption in porous media with variable porosity. J Hydrol 35:235–243
38. Lin SH (1977) Nonlinear adsorption in layered porous media flow. J Hydraul Division 103:951–958
39. Shamir UY, Harleman DRF (1967) Dispersion in layered porous media. J Hydraul Division 95:237–260
40. Valocchi AJ (1989) Spatial moment analysis of the transport of kinetically adsorbing solute through stratified aquifers. Water Resour Res 25:273–279
41. Matheron G, De Marsily G (1980) Is transport in porous media always diffusive?, a counter example. Water Resour Res 16:901–917
42. Yates SR (1990) An analytical solution for one—dimensional transport in heterogeneous porous media. Water Resour Res 26:2331–2338
43. Logan JD (1996) Solute transport in porous media with scale dependent dispersion and periodic boundary conditions. J Hydrol 184:261–276
44. Flury M, Wu QJ, Wu L, Xu L (1998) Analytical solution for solute transport with depth-dependent transformation or sorption coefficients. Water Resour Res 34:2931–2937
45. Pang L, Hunt B (2001) Solutions and verification of a scale—dependent dispersion model. J Contam Hydrol 53:21–39
46. Wang ST, Mcmillan AF, Chen BH (1978) Dispersion of pollutants in channels with non-uniform velocity distribution. Water Resour Res 12:389–394
47. Basha HA (1997) Analytical model of two-dimensional dispersion in laterally non-uniform axial velocity. J Hydraul Eng 123:853–862
48. Yeh GT, Tsai Y J (1976) Dispersion of water pollutants in a turbulent shear flow. Water Resour Res 12:1265–1270
49. Chen JS, Liu CW, Liao CM (2003) Two-dimensional laplace-transformed power series solution for solute transport in a radially convergent flow field. Adv Water Resour 26:1113–1124
50. Chen J-S, Ni C-F, Liang C-P, Chiang C-C (2008) Analytical power series solution for contaminant transport with hyperbolic asymptotic distance—dependent dispersivity. J. Hydrol 362:142–149
51. Mishra S, Parker JC (1990) Analysis of solute transport with a hyperbolic scale dependent dispersion Model. Hydrol Proc 4:45–47
52. Dideerjean S, Maillet D, Moyne C (2004) Analytical solutions on one dimensional macro-dispersion in stratified porous media with quadrupole method: convergence to an equivalent homoheneous porous medium. Adv Water Resour 27:657–667
53. Banks RB, Jerasate S (1962) Dispersion and adsorption in porous media Flow. J Hydraul Division 88:1–21
54. Warrick AW, Bigger JW, Nielson DR (1971) Simultaneous solute and water transfer for unsaturated soil. Water Resour Res 7:1216–1225
55. Pickens JF, Grisak GE (1981) Scale—dependent dispersion in a stratified granular aquifer. Water Resour Res 17:1191–1211
56. Suresh Kumar G, Sekhar M, Misra D (2008) Time—dependent dispersivityof linearly sorbing solutes in a single fracture with matrix diffusion. ASCE J Hydrol Eng 13(4):250–257
57. Kumar N (1983) Unsteady flow against dispersion in finite porous media. J. Hydrol 63:345–358
58. Yadav RR, Vinda RR, Kumar N (1990) One-dimensional dispersion in unsteady flow in an adsorbing porous medium: an analytical solution. Hydrol Proc 4:189–196
59. Aral MM, Liao B (1996) Analytical solution for two-dimensional transport equation with time dependent dispersion coefficients. ASCE J Hydrol Eng 1(1):20–32
60. Singh MK, Singh P, Singh VP (2010) Analytical solution for two dimensional solute transportin finite aquifer with time dependent source concentration. ASCE J Eng Mech 136(10):1309–1315
61. Zou S, Ma J, Koussis AD (1996) Analytical solutions to non—fickian subsurface dispersion in uniform groundwater flow. J Hydrol 179:237–258

62. Zoppou C, Knight JH (1999) Analytical solution of a spatially variable coefficient advection—diffusion equation in up to three dimensions. Appl Math Model 23:667–685
63. Liu F, Anh V, Turner I (2004) Numerical solution of the space fractional fokker-planck equation. J Comput Appl Maths 166(1):209–219
64. Kim S, Kavvas ML (2006) Generalized fick's law and fractional ade of pollution transport in a river: detailed derivation. ASCE J Hydrol Eng 11(1):80–83
65. Huang Q, Huang G, Zhan H (2008) A finite element solution for the fractional advection-dispersion equation. Adv Water Resour 31:1578–1589
66. Zurigat M, Momani S, Odibat Z, Alawneh A (2010) The homotopy analysis method for handling systems of fractional differential equations. Appl Math Model 34:24–35
67. Todd A (1989) On the simulation of incompressible miscible displacement in a naturally fractured petroleum reservoir (French Summary). RAIRO Math Model Numer Anal 23:5–51
68. Hiester NK, Vermeulen T (1952) Saturation performance of ion-exchange and adsorption columns. Chem Eng Proc 48:505–516
69. Houghton PA, Madurawe RU, Hatton TA (1988) Convective and dispersive models for dispersed phase axial mixing, the significance of poly-dispersivity effects in liquid-liquid contactors. Chem Eng Sci 43:617–639
70. Ferrick MG (1985) Analysis of river waves types. Water Resour Res 21:209–220
71. Lighthill MJ, Witham G B (1955) On kinetic waves-i: flood movement in long rivers. Proc Royal Soc Lond 229 A:281–316
72. Schugart RC, Friedman A, Zhao R, Sen CK (2008) Wound angiogenesis as a function of tissue oxygen tension: a mathematical model. PNAS 105:2628–2633
73. Dale PD, Sherrat JA, Maini PK (1996) A mathematical model for collagen fiber formation during foetal and adult dermal wound healing. Proc Royal Soc Lond 263B:653–660
74. Caputo M, Cametti C (2008) Diffusion with memory in two cases of biological interest. J Theor Biol 254:697−703
75. Huang Q, Huang G, Zhan H (2008) A finite element solution for the fractional advection−diffusion equation. Adv Water Resour 31:1578−1589
76. Gao G, Zhan H, Feng S, Huang G (2009) Comparison of alternative models for simulating anomalous solute transport in a large heterogeneous soil column. J Hydrol 377:391−404
77. Chen W, Sun H, Zhang X, Korošak D (2010) Anomalous diffusion modeling by fractional derivatives. Computer Math Appl 59:1754–1758
78. Zurigat M, Momai S, Odibat Z, Alawneh A (2010) The homotopy analysis method for handling systems of fractional differential equations. Appl Math Model 34:24−35
79. Jin-liang W, LI H (2011) Surpassing the fractional derivative: concept of the memory-dependent derivative. Computer Math Appl 62:1562–1567
80. Saadatmandi A, Dehgan M (2011) A tau approach for solution of the space fractional diffusion equation. Computer Math Appl 62:1135–1142
81. Meerschaert MM, Tadjeran C (2006) Finite difference approximations for two-sided space—fractional partial differential equations. Appl Numer Math 56:80−90
82. Yang Q, Liu F, Turner I (2010) Numerical methods for FPDEs with Riesz space fractional derivative. Appl Math Model 34:200−218
83. Sherratt JA, Murray JD (1990) Models of epidermal wound healing. Proc. R. Soc. Lond 241B:29−36
84. Arnold JS, Adam JA (1999) A simplified model of wound healing II: the critical size defect in two dimensions. Math Comp Model 30:47−60
85. Cussler EL (1997) Diffusion: mass transfer in fluid systems. Cambridge University Press, Cambridge
86. Basu SK, Kumar N, Srivastav JP (2010) Modeling NPK release from spherically coated fertilizer granules. J SIMPAT 18:820–835

Chapter 2
Diffusion and Transport of Molecules In Living Cells

Ruchi Gaur, Lallan Mishra and Susanta K. Sen Gupta

2.1 Introduction

Diffusion, derived from the Latin word *diffundere* meaning *to spread out,* is a mass transport phenomenon in both fluids (without requiring bulk fluid motion) and solids. Both macroscopic or phenomenological and microscopic or atomistic and molecular approaches are employed to introduce the concept of *diffusion*. According to the former approach, the diffusion transport goes from regions of high concentration to regions of low concentration, whereas according to the latter diffusion is a result of the random walk of the particles. In molecular diffusion, moving molecules are self-propelled by thermal energy.

In this chapter we discuss the historical developments, diffusion in cells and the current status of research , basic mathematical models of diffusion, and osmosis and its importance for living systems.

2.1.1 Historical Perspective

The history of diffusion goes back to several centuries B.C. Mechanisms of many technical processes in use over centuries are in fact controlled by diffusion, for example cementation used in gold or silver refining, carbon diffusion in elemental iron for steel making by the cementation process since medieval times or earlier, diffusion soldering of gold artefacts, colouring of glasses, earthenware, or chinaware, all happened long before the development of any theory of diffusion. In

S. K. Sen Gupta (✉) · R. Gaur · L. Mishra
Department of Chemistry, Faculty of Science, Banaras Hindu University, 221005 Varanasi, India
e-mail: ruchiuno@gmail.com

L. Mishra
e-mail: lmishrabhu@yahoo.co.in

S. K. Sen Gupta (✉)
e-mail: sksbhuchem@yahoo.co.in

S.K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_2,

the seventeenth century, Robert Boyle demonstrated diffusion of zinc into a copper coin. The well-known Brownian motion, never-ending movement of particles in suspension in a fluid, discovered in 1827 by Robert Brown, a botanist, is a manifestation of the random walk of microscopic particles suspended in a fluid. This motion has been described on one hand as the wanderings of a drunken sailor and on the other hand as the zoom image of molecular movements. Interestingly, Brownian motion is a mathematical object treated in many text books as well as a physical one allowing us to rationalize natural facts as varied as the flight of birds or mosquitoes, the spread of diseases, dissemination of pollutants, the properties of biological membranes, the brain imaging by nuclear magnetic resonance (NMR) spectroscopy, etc. In his theory of Brownian motion, Albert Einstein developed the microscopic theory of diffusion of particles at sufficiently low concentration in a liquid in 1905. Significant contributions towards this approach were also made by Marian Smoluchowski and Jean-Baptiste Perrin. However, long before in 1858, James Clerk Maxwell developed the first microscopic theory of transport in gases based on gas kinetics; the concept of mean free path was introduced by Rudolf Clausius in the same year. Ludwig Boltzmann developed the atomistic backgrounds of the macroscopic transport processes and introduced the Boltzmann transport equation in 1872. The equation has been serving mathematics and physics with a source of transport process ideas and concerns over the last 140 years [1, 2].

The phenomenological approach was introduced by Adolf Fick as a 26-year-old assistant in anatomy and physiology, in his famous papers in 1855 establishing the now classical Fick's equations, governing the mass transport through diffusive means. It is worth mentioning that later as a chair professor of physiology he authored the first treatise on medical physics, the first book of this kind. He remains a well-known name in the history of cardiology. Fick's approach was inspired by Thomas Graham's famous experimentation on diffusion of salts in water for investigating and comparing the diffusibility of different salts in 1850 and his earlier work on diffusion in gases in 1833. Fick used the law of conservation of matter and the deep analogy between diffusion and hydraulic flow (Darcy's law), heat conduction (Fourier's law), or charge transport (Ohm's law), to develop his fundamental laws for diffusion. He used Graham's method to design his experiments on the measurements of concentrations and fluxes of salts diffusing between two reservoirs through tubes of water. Fick's work although originally concerned with diffusion in fluids, his laws later also became the core of understanding diffusion in solids. Today, Fick's laws are the most popularly used laws for diffusion in gases, liquids, and solids [1, 2].

The successful use of Fick's laws to solid state diffusion was demonstrated first in 1896 by William Chardler Roberts-Austen, a longtime associate of Thomas Graham, while extending Graham's work to diffusion of gold in lead. George de Hevesy studied and measured self-diffusion of radioactive isotopes of lead in liquid and solid lead in 1920–1921. In 1926, Yakov Frenkel introduced the idea of diffusion in crystals through local defects (vacancies and interstitial atoms) and concluded that the diffusion in process in condensed matter is an ensemble of elementary jumps and quasi-chemical interactions of particles and defects. He proposed several mechanisms of diffusion and found rate constants from experimental data.

Sometime later, Carl Wagner and Walter H. Schottky developed Frenkel's ideas about mechanisms of diffusion further. It is now universally recognized that atomic defects are necessary to mediate diffusion in crystals. In 1922, Saul Dushman and Irving Langmuir applied Arrhenius' law to determine the coefficient of diffusion of thorium through tungsten and found satisfactory results. Henry Eyring and his coworkers applied his theory of absolute reaction rates to Frenkel's quasi-chemical model of diffusion in 1935. The analogy between reaction kinetics and diffusion leads to various nonlinear versions of Fick's law. Nonlinear models are also required for diffusion on catalyst surfaces [3, 4].

Diffusion is a widely applicable concept. It applies to any field involving random walks in the ensembles of individuals. In fact, the concept of diffusion is used across diverse fields stretching from physics, chemistry and biology to sociology, economics, and even finance.

2.1.2 Diffusion In Living Cells

The efficient delivery of proteins, drugs, and other products to their correct locations within a cell (transport) is of prime importance to the normal cellular function and development [5]. On the other hand, cytoplasm and other aqueous intracellular organelles such as mitochondria, nucleus, etc. are crowded with solutes, soluble macromolecules, skeletal proteins, and membranes inside the cell. In order to maintain homeostasis and cellular functions in the cell, most of the physiological processes depend on selective exchanges of metabolites between the cell and its exterior [6]. Substances such as liquids, nutrients, hormones and other signaling molecules, and waste products are routinely transported (received and delivered) across the cell plasma membranes.

Transportation of materials inside and outside of cells can be described in two ways: passive transport and active transport. In passive transportation, movement of substances does not require energy (adenosine triphosphate, ATP). Types of passive transportation include simple diffusion, facilitated diffusion, osmosis, and filtration. The diffusion process can continue until the concentration of solute in the extracellular and intracellular spaces is attaining the equilibrium. In facilitated diffusion, solute particles move from higher to lower concentration via cell surface channels. Diffusion and facilitated diffusion involve transport of solutes, while osmosis involves movement of water (or solvents) through a membrane. The movement of some transported proteins and glucose inside the cell membrane is highly selective so that movement across the cell membrane occurs only when assisted by the concentration gradient, a type of carrier-assisted transport known as facilitated diffusion. Both simple diffusion and facilitated diffusion are driven by the potential energy differences of a concentration gradient. Simple diffusion is a nonselective process by which any molecule capable of dissolution in the phospholipid bilayer is able to cross the plasma membrane and equilibrate between inside and outside of the cell. Thus, only small hydrophobic molecules are able to diffuse across a phospholipid bilayer at significant rates.

The gases such as oxygen and carbon monoxide, hydrophobic molecules like benzene, and small polar but uncharged molecules hydrogen and ethanol, are able to transport across the plasma membrane by simple diffusion. Other larger uncharged polar molecules, such as glucose, are unable to cross the plasma membrane by passive diffusion. Facilitated diffusion involves the movement of molecules in the direction determined by their relative concentrations inside and outside of the cell without any external energy. However, facilitated diffusion differs from simple diffusion in that the transported molecules do not dissolve in the phospholipid bilayer. Therefore, facilitated diffusion allows only polar and charged molecules, such as nucleosides, carbohydrates, amino acids, and ions to cross the plasma membrane.

Facilitated diffusion is mediated by two classes of proteins such as carrier proteins and channel proteins. Carrier proteins attach with specific molecules for transportation to the other side of the membrane and undergo conformational changes, allowing the molecule to pass through the membrane. Carrier proteins are responsible for the facilitated diffusion of sugars, amino acids, and nucleosides across the plasma membranes of most cells. Channel proteins create open pores through the membrane and allow free diffusion of any molecule of the appropriate size and charge.

Active transport is the movement of a substance across a cell membrane against its concentration gradient. There are three main types of active transport. Active transport in a cell requires energy, usually in the form of ATP. It includes transportation of large molecules (non-lipid soluble) and the sodium–potassium pump. In the case of active transport, the proteins and other molecules move against the concentration gradient. Primary active transport directly uses ATP. Secondary active transport does not directly use ATP. It takes advantage of a previously existing concentration gradient (via carriers). In a sodium–potassium pump, Na^+ is maintained at low concentrations inside the cell and K^+ is found at higher concentrations in nerve cells. When a nerve message is propagated, the ions are transported across the membrane, and a new message is generated. The ions must be actively transported back to their starting positions across the membrane using ATP as the carrier energy.

Active transports are classified as: uniport transport, cotransport, and vesicle-mediated transport. In uniport transport, only one solute movement takes place at a time. It is a facilitated low resistance diffusion and thermodynamically favoured process. It is a reversible process and accelerates the reaction by low concentration gradient. It takes place in glucose transportation, impulse transmission in neurons, primary insulin regulated glucose transportation in muscle, and adipose tissue. In cotransport, movement of one substrate down the gradient is coupled with the movement of another substrate against the gradient at the same time. In 1960, Robert K. Crane introduced the term cotransport in his discovery of the sodium–glucose cotransport as the mechanism for intestinal glucose absorption for the first time [7]. It can transport different numbers of molecules in different directions at the same time. It is also known as secondary active transport or coupled transport. It is divided into two types:(i) symporter—both the substrate (the solute and a cotransported solute) go in the same direction against its electrochemical gradient. It is found in glucose symporter SGLT1, which cotransports one glucose

(or galactose) molecule into the cell for every two sodium ions it imports into the cell. This symporter is located in the small intestine, trachea, heart, brain, testis, and prostate, and S3 segment of the proximal tubule in each nephron in the kidneys [8], and (ii) antiporter—the molecules of the solute go in (or out) and the cotransported solute go in the opposite direction across the membrane. It is found in the sodium–calcium exchanger, in which three sodium ions allowed into the cell to transport one calcium ion out [9].

In vesicle-mediated transport, vesicles and vacuoles fused with the cell membrane is utilized to transport or release chemicals out of the cell. It is also categorized as: (i) exocytosis, in which transport is out of the cell, and (ii) endocytosis, in which a molecule causes the cell membrane to bulge inward, forming a vesicle.

Diffusion also plays a fundamental role in every biochemical process in living cells. Characterizing and distinguishing the cytoskeletal migration, which includes all motor protein-mediated transport within a cell, is critical to understanding cellular function and is regulated by the diffusion process [10]. The rate of diffusion in the cell depends on several factors, such as concentration gradient, thickness of the exchange surface, and the surface area. The most well-known example of diffusion is gas exchange in living organisms. Oxygen gas is transferred from lungs to red blood cells and vice versa via diffusion. Carbon dioxide is produced by all cells as a result of cellular metabolic processes. Since the source is inside the cell, the concentration gradient is constantly being replenished/reelevated, thus the net flow of CO_2 is out of the cell. In photosynthesis, the gas exchange process (carbon dioxide from air to leaf and oxygen from leaf to air) follows diffusion process. Acetylcholine neurotransmitter is transported from presynaptic to postsynaptic membrane at a synapse via diffusion. In alveolus, macrophages eating viruses swim and tumble but can be modelled as diffusion. Cellular Ca^{2+} dynamics involves the exchange of Ca^{2+} ions between intracellular stores and the cytosol, the interior and exterior of a cell or between cells, as well as transport by diffusion and buffering due to the binding of Ca^{2+} to proteins.

There are two ways in which substances can enter or leave a cell. According to this, diffusion may be classified as: (i) *intracellular diffusion or self-diffusion,* (ii) *interdiffusion*. The process of spontaneous mixing of molecules taking place in the absence of concentration (or chemical potential) gradient is known as self-diffusion. The self-diffusion is diffusion in one-component material, when all atoms that exchange positions are of the same type. In the absence of external forces, the displacements of the particles result from their thermal agitation. This is an incoherent process leading to random motions of the particles in any state of aggregation. Nowadays, this process is usually denoted as self-diffusion [11]. Some authors denote self-diffusion in liquid mixtures as intra-diffusion [12], retaining the term self-diffusion only for diffusive processes in pure liquids. Unkel and others described variety in intracellular diffusion during the cell cycle [13].

The self-diffusion coefficient of neat water is: $2.299\times10^{-9}\,m^2.s^{-1}$ at 25 °C and $1.261\times10^{-9}\,m^2.s^{-1}$ at 4 °C [14]. Topgaard and others demonstrated a new method for the characterization of water-swollen biological porous structures using NMR to determine the amount and self-diffusion of water within the porous objects [15].

The method shows slower diffusion of water in starch granules in comparison to cellulose fibres and it is attributed to the smaller amount of freezable water and the pore geometry.

Drugs diffuse through various barriers for metabolism and excretion when administered to the body. Some drugs diffuse through the skin, gastric mucosa. Parenteral drugs must diffuse through muscle, connective tissue, and so on, to get to the site of action; even intravenous drugs must diffuse from the blood to the site of action. Considering all the diffusion processes that occur in the body (passive, active, and facilitated), it is not surprising that the laws governing diffusion are important to drug delivery systems. In the dissolution of the particles of drug, the dissolved molecules diffuse away from the individual particle body.

2.1.3 Current Status of Research: Diffusion In Cell

Over the last three decades, the accelerating growth of publications in the area of ion transport has witnessed the interest in this area of research in current science [16–18]. The ion exchange process intensifies, increasing research interest among chemists, chemical engineers, and biologists in understanding the transport processes occurring across the natural and artificial membranes [19]. As artificial ligand models, several types of macrocyclic molecules have been prepared specifically to transport alkali, alkaline earth, and organic ammonium ions with high selectivities, useful in active transport process [20, 21]. Carrier-mediated transport (pertraction) of metal ions by soluble macromolecules has so far been investigated using the neutral or functionalized polymers and the effectiveness of such macro-ionophores has been demonstrated [22, 23]. For the maintenance of homeostasis in the cell, newly synthesized products from the nucleus are transported to other intracellular medium or the cell membrane via a microtubular network from centrosomes. Various animal viruses including HIV were described as taking advantage of microtubule-based transport in order to reach the nucleus from the cell surface and releasing their genomic material through the nuclear pores [24].

The challenges of cellular transport are particularly great for neurons (brain cells), which are amongst the largest and most complex cells in biology with regard to the efficient transport of newly synthesized proteins from the cell body. In microrheology, the transport and motion of probes especially fluorescent molecules is tracked and recorded over a time period; local mechanical properties in the vicinity of the probes are deduced depending upon the driving mechanisms. Cells respond to the external conditions through internal structural, compositional, and functional modifications; their transport and mechanical properties get altered. Thus, the measurements of changing particle-transport properties indicate an evolution of the internal structure of the cell after administering nocodazole as expected during microtubule dissociation [25]. Extensive theoretical and experimental studies based on the carrier-mediated transport through liquid membranes have been reviewed in literature [26, 27].

Verkman demonstrated diffusion of solutes and macromolecules in aqueous cellular compartments. This is required for numerous cellular processes including metabolism, second messenger signaling, and protein–protein interactions [28]. Recently, Videcoq and others described the diffusion of two pectin methylesterases enzyme (PMEs) with different origins and modes of action and characterized with a multi-scale approach in different media consisting of pectin macromolecular solutions and physical gels [29]. Another useful method for quantitative measurement of the translational diffusion of fluorophores and fluorescently labelled macromolecules is fluorescence recovery after photo bleaching. It is more sensitive than diffusion-weighted imaging (DWI) and permits cell-level spatial resolution. In this method, fluorescently labelled molecules are introduced inside the cells by microinjection or incubation, or by targeted expression of green fluorescent protein (GFP) chimeras. In spot photo bleaching, fluorophores in a defined volume of a fluorescent sample are irreversibly bleached by a short intense light pulse. Using an attenuated probe beam, the diffusion of unbleached fluorophores into the bleached volume is measured as a quantitative index of fluorophore translational diffusion. A variety of optical configurations, detection strategies, and analysis methods have been used to quantify diffusive phenomena in photo bleaching measurements [30].

To measure diffusion, modulated gradient spin-echo method (MGSE) is used, in which pulsed gradients are not necessarily applied [31]. It gives information about diffusion in the frequency domain. By using a periodically oscillating phase factor, the MGSE experiment results in signal attenuation, which is proportional to the spectrum of the velocity autocorrelation function (VAF) of the spin-bearing particles. The diffusion spectrum probed by MGSE is related to the mean square displacement, and analogously to the time-dependent diffusion coefficient, and contains information about the morphology [32]. The MGSE technique, which enables spectral characterization of diffusion with chemical-shift resolution, has been introduced. The use of spin echoes instead of gradient echoes [33] is advantageous in reducing the effects of field inhomogeneity and susceptibility artefacts. The technique is particularly suitable for in vitro studies of samples, where diffusion of several compounds with different chemical shifts is of interest. Whereas, optical tweezers (optical tweezer is a device that allows for manipulation of nano- and microscopic particles by a focused laser beam) are capable of studying diffusion at short timescales, optical particle tracking is typically used for studying diffusion and transport processes at larger timescales (greater than 0.01 s); a combination of these methods yields quite a large frequency range [34].

2.2 Basic Models of Diffusion

The following is an outline of the basic models of linear diffusion put forward by Fick, Einstein, Teorell, and Onsagar.

2.2.1 Fick's Law

The first prominent equation of diffusion is Fick's first law which expresses the diffusion flux, J in mol.m^{-2}.s^{-1} as proportional to the anti-gradient of the local concentration, c (in mol. m^{-3}) at a position vector $\mathbf{r}$ at time t:

$$\mathbf{J} = -D\nabla c\ (\mathbf{r}, t) \tag{2.1}$$

where D symbolizes diffusion coefficient (in m^2. s^{-1}) and ∇ is the del operator. Equation 2.1 applies to ideal mixtures and postulates that the flux goes from regions of higher concentration to regions of lower concentration. In one dimension, Eq. 2.1 reduces to:

$$J_i = -D\frac{\partial c}{\partial x_i} \tag{2.2}$$

where the subscript i denotes the i-th position (in m).For systems other than ideal solutions or mixtures, the concentration gradient $(\partial c / \partial x_i)$ in one dimension is replaced by $(a / RT)\,(\partial \mu / \partial x_i)$, where μ is the chemical potential of the species (in Jmol^{-1}), a is the activity of the species (in mol. m^{-3}), R is the universal gas constant (in J.K^{-1}. mol^{-1}), and T is the temperature (in Kelvin). It may be noted that the anti-gradient of chemical potential, $-\nabla\mu$, the driving force of diffusion is not necessarily a truly real force. It represents the spontaneous tendency of the molecules to disperse as a consequence of the second law of thermodynamics and the hunt for maximum entropy. For ideal or near-ideal solutions and mixtures, a becomes c. The form of Fick's law in one dimension becomes:

$$J_i = -D\frac{c}{RT}\frac{\partial \mu}{\partial x_i} \tag{2.3}$$

The corresponding diffusion equation predicting how diffusion changes concentration with time is Fick's second law, according to which the time derivative of the concentration is the negative divergence of the flux, J. Using Eq. 2.1, we have:

$$\frac{\partial c(\mathbf{r}, t)}{\partial t} = -\nabla \cdot \mathbf{J} = D\nabla^2 c(\mathbf{r}, t) \tag{2.4}$$

where ∇^2 is the Laplacian operator. In one dimension, Eq. 2.4 becomes:

$$\frac{\partial c(x, t)}{\partial t} = D\frac{\partial^2 c(x, t)}{\partial x^2} \tag{2.5}$$

In all these expressions, D has been assumed constant, independent of concentration, position, and direction. The linear diffusion equation (Eq. 2.4 or 2.5) applies

to isotropic diffusion only and is in fact a special case of the generalized diffusion equation:

$$\frac{\partial c(\mathbf{r},t)}{\partial t} = -\nabla \cdot \left[D(c,\mathbf{r}) \nabla c(\mathbf{r},t) \right] \tag{2.6}$$

where $D(c,\mathbf{r})$ is the collective diffusion coefficient for the concentration c at the location $\mathbf{r}$. The expression for Fick's first law, Eq. 2.1 would be written as:

$$\mathbf{J} = -D(c,\mathbf{r}) \nabla c(\mathbf{r},t) \tag{2.7}$$

More generally, when D is a symmetric positive definite matrix, Eq. 2.6 describes anisotropic diffusion which is written as:

$$\frac{\partial c(\mathbf{r},t)}{\partial t} = \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{\partial}{\partial x_i} \left[D_{ij}(c,\mathbf{r}) \frac{\partial c(\mathbf{r},t)}{\partial x_j} \right] \tag{2.8}$$

The nonlinear equation (Eq. 2.6) would obviously reduce to the linear form (Eq. 2.4), when D becomes a constant independent of $c(\boldsymbol{r})$. The generalized equation (Eq. 2.6 or Eq. 2.8) applies to inhomogeneous media, where D varies in space; anisotropic media, where D depends on the direction or inhomogeneous anisotropic media, where D depends upon both position and direction.

The linear diffusion equation (Eq. 2.4) can be solved using Fourier transformation with the initial condition $c(\mathbf{r},0) = c_0 \delta(\mathbf{r})$, c_0 is $c(0, 0)$; that is, all the solute molecules are at the origin initially. The solution is:

$$c(\mathbf{r},t) = c_0 \left(\frac{1}{4\pi D t} \right)^{3/2} e^{-r^2/(4Dt)} \tag{2.9}$$

At any time, the distribution is three-dimensional Gaussian function. Clearly, as the time passes the mixture becomes uniform. Using Eq. 2.9, the mean square displacement from the origin at time t is calculated as:

$$< r(t)^2 > = 6Dt = 2dDt \tag{2.10}$$

where d is the spatial dimension.

The square-root-of-time dependence of the distance travelled is characteristic of diffusive motion and Eq. 2.10 may be considered as a practical definition of the diffusion coefficient. The length $\sqrt{2dDt}$ is defined as the diffusion length. This is the diffusion law derived by Albert Einstein [1, 35–38].

Diffusion results from Brownian motion, the random battering of solute molecules by the solvent molecules. Application of the one-dimensional random walk

model leads, interestingly, to the same expression for $<x(t)^2>$ in one dimension, that is:

$$< x(t)^2 >= \frac{l^2 t}{\tau} \tag{2.11}$$

where l is the step length and τ is the time interval between two successive displacements with the identification of $l^2 / \tau = 2D$. The solutions of diffusion and random walk problems thus become identical. Equation 2.10 is, in fact, a connector between a macroscopic entity D and the microscopic one $<r^2>$ [1, 38].

2.2.2 Einstein's Mobility

While developing the theory of Brownian motion, Albert Einstein compared the motion of particles at sufficiently low concentration in a liquid under a constant force $\boldsymbol{F}$ with diffusion. For a given $\boldsymbol{F}$ (in N), each particle has the average velocity $u\boldsymbol{F}$ (in m.s^{-1}),where u is the mobility (the ratio of the particle's terminal drift velocity to the applied force) of the particle (in m. s^{-1}. N^{-1}) and obtained the connection

$$D = uk_B T \tag{2.12}$$

known as the Einstein–Smoluchowski relation, the relation being revealed independently by Marian Smoluchowski, where k_B is the Boltzmann constant (in JK^{-1}molecule^{-1}). The mobility u is referred as the *Einstein mobility* [1, 38]. This is an early example of the famous fluctuation–dissipation theorem, which bridges microscopic fluctuations with macroscopic transport coefficients. A special case of the Einstein–Smoluchowski relation is the Einstein–Stokes equation for diffusion of spherical particles through a liquid in the limit of low Reynolds number:

$$D = \frac{k_B T}{6\pi\eta\rho} \tag{2.13}$$

where η is the viscosity coefficient of the liquid (in Nm^{-2}s), ρ is radius of the particle (in m), and $6\pi\eta\rho$ is the Stokes' frictional coefficient. In the case of rotational diffusion, Eq. 2.12 becomes:

$$D_{\text{rot}} = \frac{k_B T}{8\pi\eta\rho} \tag{2.14}$$

Using Eq. 2.13, one can estimate the diffusion coefficient (D) of spherical particles from measurements of the viscosity coefficient (η) of a liquid [1, 38]. However, for real solutions at sufficiently high concentrations, the expression for D in terms of u (Eq. 2.12) is modified to:

$$D = u(RT + 2Bc + 3Mc^2 + ...) \tag{2.15}$$

where B, M, … are the second, third, and higher order Virial coefficients [38].

2.2.3 Teorell Formula

In 1935, Torsten Teorell used the mobility-based approach for studying diffusion of ions through a membrane and formulated the essence of his approach as:

$$Flux = mobility \times concentration \times force\ per\ gram\ ion \tag{2.16}$$

This formula called Teorell formula [4, 39] ignores heat effects, special membrane effects, and chemical reactions. The force under isothermal conditions has two components:

a. Diffusion force caused by concentration gradient: $-\nabla\mu = -RT\nabla \ln(c / c^{eq})$, where c^{eq} is the equilibrium concentration.
b. Electrostatic potential gradient: $q\nabla\phi$, where q is the charge and ϕ is the electric potential.

The Teorell formula for flux J is thus:

$$\mathbf{J} = uc\left(-\frac{RT}{c}\nabla c + q\nabla\phi\right) \tag{2.17}$$

This expression allows us to find the concentration jumps and the electric potential across the membrane caused by the combined action of diffusion and the electric field, when mobilities of various components are different. For nonideal systems under isothermal conditions, Teorell equation becomes:

$$\mathbf{J} = ua(-\nabla\mu + \text{external force per gram particle}) \tag{2.18}$$

where a, the activity measures the effective concentration of a species in a nonideal mixture and $a = c / c^0 + o(c / c^0)$, where c^0 is the standard state concentration; the second term is a small correction, the activity coefficient. Equation 2.18 is the main analogue of Fick's law for monomolecular diffusion in non-perfect media.

The time derivative of a or c (as normalized dimensionless quantity) in the Einstein–Teorell approach, for small value of c becomes:

$$\frac{\partial(c / c^0)}{\partial t} = \nabla \cdot [ua(\nabla\mu - \text{external force per gram particle}) \tag{2.19}$$

2.2.4 Onsager's Linear Phenomenology and Equations for Multicomponent Diffusion

In 1931, Lars Onsagar included multicomponent diffusion in the general context of the linear nonequilibrium thermodynamics:

$$\mathbf{J}_i = \sum_j L_{ij}\mathbf{X}_j \tag{2.20}$$

where $\mathbf{J}_i$ is the flux of the *i*-th component and $\mathbf{X}_j$ is the *j*-th thermodynamic force (for pure diffusion, it is the space anti-gradient of the *j*-th chemical potential, μ_j divided by *T*, that is $-\nabla(\mu_j / T)$). After linearization near equilibrium, this approach gives for perfect systems (where deviations of c_j from c_j^{eq} are assumed small) under isothermal conditions:

$$\mathbf{X}_j = -\frac{1}{T}\nabla\mu_j \quad (j > 0) \tag{2.21}$$

$$\mathbf{X}_j = -\frac{R}{c_j^{\text{eq}}}\nabla c_j \quad (j > 0) \tag{2.22}$$

$$\mathbf{J}_i = -\sum_j L_{ij}\frac{R}{c_j^{\text{eq}}}\nabla c_j \quad (i, j > 0) \tag{2.23}$$

$$\frac{\partial c_i}{\partial t} = R\sum_j L_{ij}\frac{\nabla^2 c_j}{c_j^{\text{eq}}} \quad (i, j > 0) \tag{2.24}$$

The matrix of kinetic coefficients L_{ij} are symmetric, $L_{ij} = L_{ji}$: Onsager reciprocal relations and their symmetry follow from microscopic reversibility and statistical mechanics of fluctuations and their decay.

The Onsager form of diffusion equations (Eqs. 2.22–2.24) is correct near the equilibrium but violates the obvious physical requirement: $\mathbf{J}_i$ is zero if c_i has zero value. The Teorell approach (Eqs. 2.17–2.19) satisfies this requirement. Fick's laws (Eqs. 2.1–2.5) also satisfy this requirement in the sense: if for nonnegative smooth $c(x)$, the concentration vanishes at some points, then at these points the flux vanishes too (because these points are minimizers of concentration and the gradient vanishes there).

For isotropic non-perfect systems, the thermodynamic driving forces in Onsager's form for isothermal diffusion in the linear approximation near the equilibrium (using Eq. 2.21) are:

$$\mathbf{X}_j = -\frac{1}{T}\sum_k \left(\frac{\partial\mu_j}{\partial c_k}\right)_{c=c^{\text{eq}}} \nabla c_k, \quad (j,k > 0) \tag{2.25}$$

and the diffusion equations become:

$$\mathbf{J}_i = \sum_j L_{ij}\mathbf{X}_j = -\frac{1}{T}\sum_j L_{ij}\left[\sum_k \left(\frac{\partial\mu_j}{\partial c_k}\right)_{c=c^{\text{eq}}} \nabla c_k\right], \quad (j,k > 0) \tag{2.26}$$

$$\frac{\partial c_i}{\partial t} = -\nabla \cdot \mathbf{J}_i = \frac{1}{T}\sum_k \left[\left(\sum_j L_{ij} \frac{\partial \mu_j}{\partial c_k} \right)_{c=c^{eq}} \right] \nabla^2 c_k, \quad (i,j,k>0) \tag{2.27}$$

The matrix of diffusion coefficients, D_{ik} becomes:

$$D_{ik} = \frac{1}{T}\sum_j L_{ij} \left(\frac{\partial \mu_j}{\partial c_k} \right)_{c=c^{eq}} \quad (i,j,k>0) \tag{2.28}$$

The intrinsic arbitrariness in the definitions of $\mathbf{X}_j$ and L_{ij} is to be noted. These are not measurable separately and only their combinations $\sum_j L_{ij} X_j$ can be measured. Thus, the Onsager's formalism of linear irreversible thermodynamics gives the system of linear diffusion equations in the form:

$$\frac{\partial c_i}{\partial t} = \sum_j D_{ij} \nabla^2 c_j \quad (i,j>0) \tag{2.29}$$

If the matrix D_{ij} is diagonal, this system of equations is simply a collection of decoupled Fick's equations for various components [4, 40–42].

Non-diagonal diffusion must be nonlinear. Diffusion preserves the positivity of concentrations. If the diffusion is non-diagonal and linear, e.g. $D_{12} \neq 0$, for a state where $c_2 = ... = c_n = 0$, the diffusion equation would be:

$$\frac{\partial c_2}{\partial t} = D_{12} \nabla^2 c_1(x) \tag{2.30}$$

If $D_{12}\nabla^2 c_1(x) < 0$ at some points, $c_2(x)$ becomes negative at these points in a short time. Therefore, linear non-diagonal diffusion does not preserve the positivity of concentrations and consequently, non-diagonal equations of multicomponent diffusion are nonlinear [4].

2.2.5 Teorell Formula For Multicomponent Diffusion

The Teorell formula with combination of Onsager's definition of the diffusion force gives:

$$\mathbf{J}_i = u_i a_i \sum_j L_{ij} \mathbf{X}_j \tag{2.31}$$

For isothermal perfect systems:

$$\mathbf{J}_i = -u_i c_i R \sum_j L_{ij} \frac{1}{c_j} \nabla c_j \tag{2.32}$$

Thus, the Einstein–Teorell approach gives the following generalization of the Fick's law for multicomponent diffusion:

$$\frac{\partial c_i}{\partial t} = \sum_j \nabla \cdot \left(D_{ij} \frac{c_i}{c_j} \nabla c_j \right) \tag{2.33}$$

where D_{ij} is the matrix of coefficients [4]. It should be stressed that these physical models of diffusion are different from the toy models (Eq. 2.29), which are valid for very small deviations from uniform equilibration.

For anisotropic multicomponent diffusion coefficients (for example, in crystals) one needs 4-index quantities, for example, $D_{ij\alpha\beta}$, where *i, j* are related to the components and $\alpha, \beta = 1, 2, 3$ correspond to the space coordinates.

2.3 Nonlinear Diffusion

As discussed earlier, the linear diffusion equation has limitations. There are many nonlinear diffusion models. We discuss a few of these below.

2.3.1 Diffusion of Reagents on the Surface of a Catalyst: Jumps on the Surface

Alexander N. Gorban and his coauthors proposed a model of diffusion in monolayers of reagents on the surface, which is based on the jumps of the reagents on the nearest free places. This model has been used for oxidation of CO on platinum under low gas pressure.

The system includes several reagents $A_1, A_2, ..., A_n$ on the surface. Their surface concentrations are $c_1, c_2, ..., c_n$, respectively. The surface is a lattice of the adsorption sites. Each reagent molecule fills a place on the surface. Some of the places are free. A_0 symbolizes a free place and its concentration is $c_0 (= z)$. It follows that $\sum_{i=0}^{i=n} c_i = b$, a constant representing the density of adsorption places. According to the jump model, the diffusion flux of A_i ($i = 1,2, \ldots, n$) is:

$$\mathbf{J_i} = -D_i \left(z \nabla c_i - c_i \nabla z \right) \tag{2.34}$$

and the corresponding diffusion equation is:

$$\frac{\partial c_i}{\partial t} = -\nabla \cdot \mathbf{J_i} = D_i \left(z \nabla^2 c_i - c_i \nabla^2 z \right) \tag{2.35}$$

Due to conservation of the places on the surface $z = b - \sum_{i=1}^{i=n} c_i$ we have a system of *n* diffusion equations:

$$\mathbf{J_i} = -D_i\left[\left(b - \sum_{i=1}^{n} c_i\right)\nabla c_i + c_i \nabla \sum_{i=1}^{n} c_i\right] \tag{2.36}$$

$$\frac{\partial c_i}{\partial t} = D_i\left[\left(b - \sum_{i=1}^{n} c_i\right)\nabla^2 c_i + c_i \nabla^2 \sum_{i=1}^{n} c_i\right] \tag{2.37}$$

It may be noted that for one component:

$$(b-c)\nabla c + c\nabla c = b\nabla c \tag{2.38}$$

$$(b-c)\nabla^2 c + c\nabla^2 c = b\nabla^2 c, \tag{2.39}$$

and the diffusion equation becomes linear, the Fick's law (Eq. 2.29). Obviously, for two or more components, the equations are nonlinear. When $c_i \geq 0$ for all x, $(\partial c_i / \partial t) \geq 0$ for all $c_i = 0$, which is necessary for the preservation of positivity.

If all particles can exchange their positions with their closest neighbours, a simple generalized equation follows:

$$\mathbf{J_i} = -\sum_j D_{ij}\left(c_j \nabla c_i - c_i \nabla c_j\right) \tag{2.40}$$

$$\frac{\partial c_i}{\partial t} = \sum_j D_{ij}\left(c_j \nabla^2 c_i - c_i \nabla^2 c_j\right) \tag{2.41}$$

where $D_{ij} = D_{ji} \geq 0$ is a symmetric matrix of coefficients which characterize the extent of jumps [4].

2.3.2 Diffusion In a Porous Medium

The basic equation describing diffusion in porous media, porous medium equation (PME), is a nonlinear parabolic-type evolution equation:

$$\frac{\partial c}{\partial t} = \nabla^2 c^m = \nabla \cdot [D(c)\nabla c] \tag{2.42}$$

where $c = c(r,t) > 0, m > 1$, and $D(c)$, the concentration-dependent diffusivity is given as $D(c) = mc^{m-1}$. PME applies to a number of processes such as the flow of an isentropic gas through a porous medium $(m \geq 2)$, infiltration of ground water $(m = 2)$, heat radiation in plasmas $(m \geq 4)$ [43].

2.3.3 Phase Separation: Cahn–Hilliard Equation

The process by which the two components of a binary fluid spontaneously separate and form domains pure in each component is a problem of nonlinear diffusion and is described by the Cahn–Hilliard equation:

$$\frac{\partial c}{\partial t} = D\nabla^2\left(c^3 - c - \gamma\nabla^2 c\right) \tag{2.43}$$

where c is the concentration of the fluid, $c = \pm 1$ indicates domains, D is the diffusion coefficient, and γ is the square of the of the length of the transition regions between the domains. The quantity $(c^3 - c - \gamma\nabla^2 c)$ is identified as a chemical potential μ. The term $-\gamma\nabla^2 c$ is derived from a component of the free energy modelling the interface energy and it regularizes the solutions of the equation [4, 44]. In the phase separation problem, the components are definitely non-perfect and necessary corrections by the activity coefficients can be incorporated.

2.3.4 Diffusion In Solids: Eyring's Quasi-Chemical Model

Diffusion in solids takes place through the movement of defects, for example point defects (vacancies, interstitial atoms) are responsible for lattice diffusion. Diffusion occurs when an atom jumps from a normal lattice site into an adjacent vacant one or from an interstitial site to one of the neighbouring interstitial ones. The interstitial mechanism which involves significant lattice distortion is, however, favoured when interstitial atoms are sufficiently small as compared to normal lattice atoms, for example light atoms like H, C, N, O interstitially dissolved in metals. When lattice distortion becomes too large for the interstitial mechanism to be probable, the interstitialcy mechanism (where an interstitial atom pushes one of its nearest neighbours on a normal lattice site into another interstitial position and itself occupies the lattice site of the displaced atom) is favoured.

The quasi-chemical theory of diffusion in solids was initially developed by Yakov Frenkel and later improved by F. C. Frank and D. Turnbull [1, 4, 45]. This was further developed by Henry Eyring and coauthors who applied their famous activated complex theory (ACT) for chemical reactions to diffusion in solids. The diffusion in the treatment is represented by an ensemble of elementary events, each of which is represented by the creation or destruction of an activated complex (transition state). The rate of the elementary process is given by the concentration of the activated complex multiplied by the rate of its decomposition. It is hypothesized that the complex is in quasi-equilibrium with the stable components and the concentration of the complex can be calculated using equilibrium statistical thermodynamics [3].

Eyring's approach to diffusion problems is illustrated here for vacancy diffusion in cubic lattice of an elemental solid. The coefficient for such diffusion may be expressed as:

$$D = \alpha a_0^2 w N_d \tag{2.44}$$

where α is a geometrical factor (=1 for *bcc* and *fcc* lattices), a_0 is the lattice parameter, w is the frequency of a jump to an adjacent site, and N_d is the fraction of vacancies in the lattice.

To obtain an expression for the temperature dependence of D (Eq. 2.44), it needs considering how N_d and w change with temperature. The diffusing atoms may make a jump, only when a neighbouring lattice site is vacant. If ΔG_d is the Gibbs free energy of vacancy formation, N_d may be expressed as:

$$N_d = e^{-\Delta G_d/(RT)} = e^{-(\Delta H_d/RT)} e^{\Delta S_d/R} \tag{2.45}$$

where ΔH_d and ΔS_d are the corresponding enthalpy and entropy terms.

ΔH_d is generally positive and N_d increases with temperature. The rate at which an atom jumps between neighbouring sites in a lattice may be written following ACT as:

$$\omega = \nu e^{-\Delta G_m/RT} = \nu e^{\Delta S_m/R} e^{-\Delta H_m/RT} \tag{2.46}$$

where ΔG_m, ΔH_m, and ΔS_m are the free energy, enthalpy, and entropy changes associated with the movement of the atom from the initial equilibrium condition to the activated complex at the top of the potential energy barrier, which the atom has to surmount during the jump to another equilibrium site. ν in Eq. 2.46 represents the vibration frequency and is of the order of 10^{13} Hertz. ΔS_m is generally small ~ 10 JK^{-1} mol^{-1}.

The temperature dependence of D for vacancy diffusion in a cubic lattice of an elemental solid may now be expressed as [3, 45]:

$$D = \alpha a_0^2 \nu e^{(\Delta S_d + \Delta S_m)/R} e^{-(\Delta H_d + \Delta H_m)/(RT)} \tag{2.47}$$

The experimental D versus T data leads to the following equation:

$$D = D_0 e^{-Q/RT} \tag{2.48}$$

where Q is the activation energy.

On comparing Eq. 2.48 with Eq. 2.47, the activation energy consists of:

$$Q = \Delta H_d + \Delta H_m \tag{2.49}$$

and the corresponding D_0 is given by:

$$D_0 = \alpha a_0^2 \nu e^{(\Delta S_d + \Delta S_m)/R} \tag{2.50}$$

2.4 Osmosis

Transport processes occurring via artificial membranes separating different salt solutions are of great interest to chemists, chemical engineers, and biologists. To understand the mechanism of transport is a thrust area of research for chemists and chemical engineers; they are interested in fabricating membranes of any desired properties. However, biologists would like to use them as simple models for understanding the properties of complex cell membranes [46]. The transport of solvent molecules through semipermeable membranes is known as osmosis. If two solutions of different concentrations are separated by a semipermeable membrane, the solvent tends to transport across the membrane from the less concentrated to the more concentrated side. Osmosis is a selective diffusion process driven by the internal energy of the solvent molecules. It is convenient to express the available energy per unit volume in terms of osmotic pressure.

The movement of a pure solvent is driven for reducing the free energy of the system by equalizing solute concentrations on each side of a membrane, generating osmotic pressure. Osmosis is driven by the imbalance in water concentration. Osmosis is vital to life, because of its function in maintaining equilibrium inside and outside of a cell. In the human body, osmosis is used by the kidneys to cleanse the blood. Osmosis is of great importance in biological processes where the solvent is water. The transport of water and other molecules across biological membranes is essential to many processes in living organisms. The cell membrane functions as a semipermeable barrier; it allows selective passage of molecules through it. Osmosis depends upon concentration of solute particles, ionization of solute particles, hydration of solute particles, and temperature.

Both plant and animal living cells are enclosed by semipermeable membranes, called the cell membranes that regulate the flow of liquids and of dissolved solids and gases into and out of the cell. The cell membrane forms a selective barrier between the cell and its environment so that not all substances can pass through the membrane easily. Without this selectivity, toxic materials from the surroundings would enter the cell. If blood or other cells are placed in contact with an isotonic solution, they will neither shrink nor swell. If the cell is placed in a hypertonic solution, it will lose water and shrink (plasmolysis) and shows exosmosis. If a cell is placed in a hypotonic solution (or if a pure solvent is used), the cells swell and show endosmosis. Hence, the osmotic pressure so developed inside the cell may even be great enough to rupture the cell membrane. In plants, osmosis is at least partially responsible for the absorption of soil water by root hairs and for the up pull of the liquid to the leaves of the plants. However, plants wilt when watered with saltwater or treated with too much fertilizer, since the soil around their roots becomes hypertonic.

The phenomenon of osmosis takes place in the absorption of water by plant roots, reabsorption of water by the proximal and distal convoluted tubules of the nephron, reabsorption of tissue fluid into the venule ends of the blood capillaries, and the absorption of water by the alimentary canal, stomach, small intestine, and colon.

Purification or desalination of water is also carried out by osmosis. Currently, two types of osmosis are used in the purification of water: forward osmosis (FO), reverse osmosis (RO). FO is a manipulated osmosis or engineered osmosis. It involves low-cost energy processes and is one of the emerging membrane technologies, as it has the ability to desalinate seawater or brackish water naturally [47]. RO is currently the most commonly used water purification technology because of its merits over other conventional thermal desalination technologies. In RO, an applied pressure is used to overcome osmotic pressure, a colligative property that is driven by chemical potential, a thermodynamic parameter [48].

In osmotic drug delivery system, the osmotic pressure of drug or other solutes (osmogens or osmagents) is used for the controlled delivery of drugs. In drug delivery, the two most critical properties considered for the selection of osmogen are osmotic activity and aqueous solubility. Osmotic agents or osmogens for drug delivery are classified as: (i) *inorganic water soluble osmogens:* magnesium sulphate, sodium chloride, sodium sulphate, potassium chloride, sodium bicarbonate, etc. (ii) organic polymeric osmogens: sodium carboxymethyl cellulose (NaCMC), hydroxypropylmethycellulose (HPMC), hydroxyethylmethyl cellulose (HEMC), etc. and (iii) *organic water soluble osmogens:* sorbitol, mannitol, etc.

Cellulose acetate is a commonly used semipermeable membrane for the preparation of osmotic pumps in biological systems. Some other examples of polymers used in semipermeable membrane are agar acetate, amylose triacetate, beta-glucan acetate, poly (vinylmethyl) ether copolymers, poly (orthoesters), polyacetals, poly (glycolic acid), and poly (lactic acid) derivatives.

Osmosis has several implications in medical science. If blood cells were stored in water, osmosis would cause them to swell and eventually burst. Osmosis is closely related to dialysis, which is critical to the survival of many victims of kidney diseases. Dialysis is the process by which an artificial kidney machine removes waste products from the patients' blood, performing the role of a healthy, normally functioning kidney.

Osmosis is classified based on the solution in which the cell is placed: endosmosis, exosmosis. When a cell is placed in hypotonic solution, water enters into the cell from the outer (hypotonic) solution. This process of diffusion of water into a cell from the outside is called endosmosis. When a cell is immersed in a hypertonic solution, water diffuses out of the cell because the concentration of water molecules in the cell is more than the outer solution. This is known as exosmosis.

2.4.1 Historical Perspective

In 1748, osmosis phenomenon through semipermeable membranes was first observed by Cleric Jean-Antoine Nollet [49]. The term "osmosis" descends from the Greek words meaning "endosmose" and "exosmose", which were coined by the French physician René Joachim Henri Dutrochet. The general term osmose (now osmosis) was introduced in 1854 by a British chemist, Thomas Graham [50]. Experimental work was conducted primarily with membranes of animals and plants.

In 1867, the first artificial semipermeable membrane of inorganic compound copper ferrocyanide was prepared by Traube [51]. Osmosis was first thoroughly studied in 1877 by a German plant physiologist Wilhelm Pfeffer by successful quantitative measurement of osmotic effect. Pfeffer measured the effect by utilizing a membrane, which is selectively permeable to water but impermeable to sugar. The membrane separated sugar solution from pure water. Pfeffer observed flow of water into the sugar solution, which stopped when a pressure p was applied to the sugar solution. Pfeffer postulated that this pressure, the osmotic pressure π of the sugar solution is proportional to the solution concentration and absolute temperature. Van't Hoff established an expression that is analogous to the Pfeffer results and the ideal gas laws [52] in 1886 as $\pi = n_2 RT$, where n_2 represents the molar concentration of sugar (or other solute) in the solution, R is the gas constant and T is the absolute temperature.

The most carefully documented use of pressure as a driving force for membrane filtration was published in 1907 by Bechold [53]. Osmotic drug delivery uses the osmotic pressure of drug or other solutes (osmogens or osmagents) for controlled delivery of drugs. Osmotic drug delivery has come a long way since the Australian physiologists Rose and Nelson developed an implantable pump in 1955 [54]. Drug itself may act as an osmogen showing good aqueous solubility (e.g. potassium chloride pumps). If the drug does not possess an osmogenic property, osmogenic salt and other sugars can be incorporated in the formulation.

2.4.2 Current Status of Research

New therapeutically active molecules for the treatment and prevention of diseases are currently being developed. It is a primary requirement for the therapeutically active molecules to reach its site of action, hence novel drug delivery systems (NDDS) have been recognized as an attraction for the pharmaceutical and health industry [55]. Many conventional drug delivery systems have been designed by various scientists to modulate the release and transport of drugs over an extended period of time. The rate and extent of drug absorption and release may depend on the factors such as physicochemical properties of the drug, presence of excipients, physiological factors such as presence or absence of food, and pH of the gastrointestinal tract (GI) [56]. Drugs can be delivered in a controlled pattern over a long period of time by the process of osmosis and this is known as the osmotic drug delivery process. Osmotic drug delivery uses the osmotic pressure of drugs or other solutes (called osmagents) for controlled delivery of drugs. Drug delivery systems are independent of the different physiological factors of gastrointestinal tract; the release characteristics can be predicted easily from the known properties of the drug and the dosage form [57]. It is a cost-effective method for the drug delivery. The major advantage of osmotic drug delivery are: (i) the delivery rate of zero order (which is most desirable) is achievable with osmotic systems, (ii) desired drug delivery may be delayed or pulsed, (iii) for oral osmotic systems, drug release is independent of gastric pH and hydrodynamic conditions, which is mainly attributed to the unique properties of semipermeable membrane employed in coating of osmotic formulations.

2.5 Summary

Diffusion is a widely applicable concept. It applies to any field involving random walks in ensembles of individuals. In fact, the concept of diffusion is used across diverse fields stretching from physics, chemistry, and biology to sociology, economics, and finance even. The diffusion behaviour in many of the cases follows linear Fick's laws. However, there are a good number of instances where diffusive flows become non-Fickian. For diffusion on catalyst surfaces, in crystal lattices, porous media, phase-segregation, etc., a more general nonlinear approach is required.

The biological cells use different mechanisms for transportation of substrates, products, waste, etc. to maintain homoeostasis. Diffusion and osmosis are involved in active metabolism in living cells. The implications of restricted molecular diffusion for cell function remain a major unresolved issue. Diffusion has played a key role in extending the applicability of MRI technique and given rise to new MRI techniques. One such technique, diffusion-weighted imaging MR (DW-MRI) relies on the determination of random microscopic motion of free water molecules in tissues with clinical applications to a wide range of pathological conditions. Another technique, diffusion tensor imaging MRI (DT-MRI) is based on the fact that mobilities of water molecules in directionally ordered cellular structures such as cell membranes and myelin become directionally dependent. DT-MRI characterizes this directional nature of water motion and thereby provides structural information that cannot be obtained by standard anatomical imaging. On the other hand, osmosis has also several implications in medical care, particularly for the storage of red blood cells. For drug delivery in the living systems, scientists are pursuing development of osmosis that has better absorptive and pharmacokinetic properties. Osmotic drug delivery uses the osmotic pressure of drugs or other solutes (called osmagents) for controlled delivery of drugs in biological systems.

References

1. Philibert J (2006) One and a half century of diffusion: Fick, Einstein, before and beyond. Diffus Fundam 4:1–19
2. Mehrer H, Stolwijk NA (2009) Heroes and highlights in the history of diffusion. Diffus Fundam 11:1–32
3. Kincaid JF, Eyring H, Stearn AE (1941) The theory of absolute reaction rates and its application to viscosity and diffusion in the liquid state. Chem Rev 28:301–365
4. Gorban AN, Sargsyan HP, Wahab HA (2011) Quasichemical models of multicomponent nonlinear diffusion. Math Model Nat Phenom 6:184–262
5. Alberts B, Johnson A, Lewis J, Raff M, Walter KRAP (2008) Molecular biology of the cell, 5th edn. Garland, New York
6. Ellis RJ (2001) Macromolecular crowding: an important but neglected aspect of the intracellular environment. Curr Opin Struct Biol 11:114–119
7. Crane RK, Miller D, Bihler I (1961) The restrictions on possible mechanisms of intestinal transport of sugars. In: Kleinzeller A, Kotyk A (eds) Membrane transport and metabolism. Proceedings of a Symposium Prague, Czech Academy of Sciences, pp. 439–449

8. Wright EM (2001) Renal Na(+)-glucose cotransporters. Am J Renal Physiol 280(1):F10–F18
9. Carafoli E, Santella L, Branca D, Brini M (2001) Generation, control, and processing of cellular calcium signals. Crit Rev Biochem Mol Biol 36(2):107–260
10. Regner BM, Vucinic D, Domnisoru C, Bartol TM, Hetzer MW, Tartakovsky DM, Sejnowsk TJ (2013) Anomalous diffusion of single particles in cytoplasm. Biophys J 104:1652–1660
11. Weingärtner H (1994) NMR studies of self-diffusion in liquids. Annu Rep Prog Chem Sect C Phys Chem 91:37–69
12. Albright JG, Mills R (1965) A study of diffusion in the ternary system, labeled urea-ureawater, at 25° by measurements of the intradiffusion coefficients of urea. J Phys Chem 69:3120–3126
13. Selhuber-Unkel C, Yde P, Berg-Sørensen K, Oddershede LB (2009) Variety in intracellular diffusion during the cell cycle. Phys Biol 6:025015–025023
14. Holz M, Heil SR, Sacco A (2000) Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate ^{1}H NMR PFG measurements. Phys Chem Chem Phys 2:4740–4742
15. Topgaard D, Söderman O (2002) Self-diffusion of nonfreezing water in porous carbohydrate polymer systems studied with nuclear magnetic resonance. Biophys J 83:3596–3606
16. McBride DW, Izatt RM, Lamb JD, Christensen JJ (1984) Inclusion compounds. In: Atwood JL, Davies JFD, MacNicol DD (eds) Inclusion compounds, vol 3. Academic, Orlando, pp 571–628
17. Guyon F, Parthasarathy N, Buffle J (1999) Mechanism and kinetics of Cu(II) transport through diaza-crown-ether-fatty acid supported liquid membrane. Anal Chem 71:819–826
18. El Aamran FZ, Kumar A, Beyer L, Florido A, Sastre AM (1999) Mechanistic study of active transport of silver(I) using sulfur containing novel carriers across a liquid membrane. J Membr Sci 152:263–275
19. Goering RM, Bowman CN, Koval CA, Noble RD, Ashley ME (2000) Complexation structure and transport mechanism of 1,5-hexadiene and 1-hexene through silver facilitated transport membranes. J Membr Sci 172:49–57
20. Choy EM, Evans DF, Cussler EL (1974) A selective membrane for transporting sodium ion against its concentration gradient. J Am Chem Soc 96:7085–7090
21. Rebek J, Wattley RV (1980) Allosteric effects: remote control of ion transport selectivity. J Am Chem Soc 102:4853–4854
22. Wódzki R, Świątkowski M, Łapienis G (2001) Mobile macromolecular carriers of ionic substances, 2. Transport rates and separation of some divalent cations by poly(oxyethylene) phosphates. Macromol Chem Phys 202:145–154
23. Eyal A, Kislik V (1999) Aqueous hybrid liquid membrane—a novel system for separation of solutes using water-soluble polymers as carriers. J Membr Sci 161:207–221
24. Damm EM, Pelkmans L (2006) Systems biology of virus entry in mammalian cells. Cell Microbiol 8:1219–1227
25. Weihs D, Mason TG, Teitell MA (2007) Effects of cytoskeletal disruption on transport, structure, and rheology within mammalian cells. Phys Fluids 19:103102–103106
26. Ho WSW, Sirkar KK (1992) Membrane handbook. Chapman and Hall, New York
27. Boyadzhiev L, Lazarova Z (1995) Liquid membranes (Liquid Pertraction). In: Noble RD, Stern SA (eds) Membrane separations technology. Elsevier, Amsterdam, pp 283–352
28. Verkman AS (2002) Solute and macromolecule diffusion in cellular aqueous compartments. Trends Biochem Sci 27(1):27–33
29. Videcoq P, Steenkeste K, Bonnina E, Garnier C (2013) A multi-scale study of enzyme diffusion in macromolecular solutions and physical gels of pectin polysaccharides. Soft Matter 9:5110–5118
30. Verkman AS, Vetrivel L, Haggie P (2001) Diffusion measurements by photobleaching recovery methods. In: Periasamy A (ed) Methods in cellular imaging. Oxford University , Oxford, pp 112–127
31. Stepišnik J (2002) A new view of the spin echo diffusive diffraction in porous structures. Europhys Lett 60:453–459
32. Stepišnik J (1993) Time dependent self-diffusion by NMR Spin-echo. Phys B 183:343–350

33. Parsons EC, Does MD, Gore JC (2003) Modified oscillating gradient pulses for direct sampling of the diffusion spectrum suitable for imaging sequences. Magn Reson Imaging 21:279–285
34. Tolic-Nørrelykke I, Munteanu E–L, Thon G, Oddershede L, Berg-Sørensen K (2004) Anomalous diffusion in living yeast cells. Phys Rev Lett 93:078102–078105
35. Graham T (1850) The Bakerian lecture: on the diffusion of liquids. Phil Trans R Soc Lond 140:1–46
36. Fick A (1855) Ueber Diffusion von. Poggendorff's Annalen 94:59–86
37. Crank J (1980) The mathematics of diffusion. Oxford University , Oxford
38. Bromberg S, Dill KA (2002) Molecular driving forces: statistical thermodynamics in chemistry and biology. Garland Science, New York
39. Teorell T (1935) Studies on the 'diffusion effect' upon ionic distribution: some theoretical considerations. Proc Natl Acad Sci U S A 21:152–161
40. Onsager L (1931) Reciprocal relations in irreversible processes. I. Phys Rev 37:405–426
41. De Groot SR, Mazur P (1962) Non-equilibrium thermodynamics. North-Holland, Amsterdam
42. Landau LD, Lifshitz EM (1980) Statistical physics, vol 5, 3rd edn. Butterworth-Heinemann, Oxford
43. Vazquez JL (2007) The porous medium equation: mathematical theory. Oxford University , Oxford
44. Cahn JW, Hilliard JE (1958) Free energy of a non-uniform system. I. Interfacial free energy. J Chem Phys 28:258–267
45. Mehrer H (2007) Diffusion in solids: fundamentals, methods, materials, diffusion-controlled processes. In: Springer series in solid state sciences, vol 155. Springer, Heidelberg
46. Lakshminarayanaiah N (1965) Transport phenomena in artificial membranes. Chem Rev 65:491–565
47. Kim I–C, Ahn S–H, Jin Y–S, Kima B–S, Parka Y–I, Jegala J, Leeb S–H, Kwonc Y–N, Rhee H–W (2013) Preparation of newly synthesized forward osmosis membrane. Desalination Water Treat 51:5191–5195
48. Greenlee LF, Lawler DF, Freeman BD, Marrot B, Moulinc P (2009) Reverse osmosis desalination: water sources, technology, and today's challenges. Water Res 43:2317–2348
49. Nollet JA (1748) Lecons de physique experimental. Hippolyte-Louis Guerin and Louis-FranciosDelatour, Paris
50. Graham T (1854) The Bakerian lecture: on osmotic force. Phil Trans R Soc Lond 144:177–288
51. Traube M (1867) Experimente zur Theorie der Zelleenbildung und Endosmose Physiologie und WisserschlaftlicheMedicin. In: Reischert, DuBois-Reynolds (eds) Archiv fur Anatomic, Veitetcomp, Leipzig 87–165
52. Van't Hoff JH (1887) The function of osmotic pressure in the analogy between solutions and gases. Z Phys Chem 1:481–508
53. Bechold HZ (1907) Kolloidstudienmit der filtrations method. Phys Chem 60:257–318
54. Singh K, Walia MK, Agarwal G, Harikumar SL (2013) Osmotic pump drug delivery system: a novel approach. J Drug Deliv Therapeutics 3(5):156–162
55. Verma RK, Mishra B, Garg S (2000) Osmotically controlled oral drug delivery. Drug Dev Ind Pharm 26(7):695–708
56. Gupta BP, Thakur N, Jain NP, Banweer J, Jain S (2010) Osmotically controlled drug delivery system with associated drugs. J Pharm Pharm Sci 13(3):571–588
57. Verma RK, Krishna DM, Garg S (2002) Formulation aspects in the development of osmotically controlled oral drug delivery systems. J Control Release 79:7–27

Chapter 3
Modeling the Diffusion and Transport of Suspended Sediment in Open Channels, Using Two-Phase Flow Theory

Sanjeev Kumar Jha and Fabián A. Bombardelli

3.1 Introduction

Traditional approaches for the description of sediment-laden, open-channel flows rely on the concept of a flow of a mixture of sediment and water. The sediment is assumed to move with the same stream-wise velocity as the carrier fluid. The velocity of the mixture is obtained from the semi-logarithmic law as:

$$\frac{U}{u_*} = \frac{1}{\kappa}\ln\left(\frac{z}{z_0}\right), \tag{3.1}$$

$$\text{with } z_0 = \frac{K_s}{30} \text{ for rough boundary,} \tag{3.2a}$$

$$\text{and } z_0 = \frac{\nu}{9.0\, u_*} \text{ for smooth boundary,} \tag{3.2b}$$

where U is the time-averaged velocity, u_* is the shear velocity, z indicates the distance from the bed in the wall-normal direction, K is the von-Kármán constant, K_S is the grain roughness equivalent height, and v is the kinematic viscosity, respectively. Several attempts have been made to modify the value of von-Kármán constant to include the influence of sediment concentration [1–6], but there is no widely accepted result to date.

S. K. Jha (✉)
Connected Waters Initiative Research Center, University of New South Wales, Sydney, Australia

National Center for Groundwater Research and Training, Adelaide, Australia
e-mail: jha.sanj@gmail.com

F. A. Bombardelli
Department of Civil and Environmental Engineering, University of California, Davis, USA

S. K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_3,

In the simplest model, the transport of sediment can be described by the advection-diffusion equation for the sediment concentration (C). The sediment diffusivity in the wall-normal direction due to turbulence is assumed to balance the effect of gravity. This assumption leads to the famous Rousean formula [7] to estimate the distribution of sediment in the wall-normal direction. The following balance is achieved by neglecting the horizontal gradients of the transport terms in the advection-diffusion equation [8]:

$$CW_s + D_d \frac{dC}{dz} = 0, \tag{3.3}$$

where D_d is the sediment diffusivity and W_s is the settling (fall) velocity. When the flow carries relatively large amount of sediment particles, a model proposed by Hunt [9] is more suitable than the Rousean balance to represent the physics of the problem. Hunt's model is obtained from:

$$CW_s(1-C) + D_d \frac{dC}{dz} = 0. \tag{3.4}$$

It is clear that the Rousean balance is a special case of Hunt's model in which the fraction of water $(1-C)$ can be approximated as unity when the mixture is dilute. D_d is customarily modeled as [10]:

$$D_d = \beta D_c, \tag{3.5}$$

where D_c is the diffusivity of momentum of the mixture, i.e., the eddy viscosity, and β is the damping coefficient. D_c is usually parameterized as:

$$D_c = \kappa u_* \frac{z}{h}(h-z) \tag{3.6}$$

which is consistent with the semi-logarithmic velocity law [11]. The Rousean formula has been used extensively since the seminal experiments by Vanoni [12] with β=1, in what it is considered its standard form [13–14]. In practice, one needs to adjust β to match the predictions to the experimental values of the sediment concentration.

Diverse authors have documented several limitations of the standard Rousean equation for predicting the sediment concentration in open channels. Many researchers [15–17] have stated, through analysis of different datasets, that the "single-phase model" clearly exemplified by Rouse's approach is too simple from the phenomenology of the sediment-transport problem (a similar limitation has been found with Hunt's model [18].). The observations in [16] pointed to the fact that the slopes in the sediment-concentration curves were flatter than those given by the Rousean curve. Since this can be associated with larger diffusion, Greiman et al. [15] attempted to explain the additional diffusion using the concept of drift.

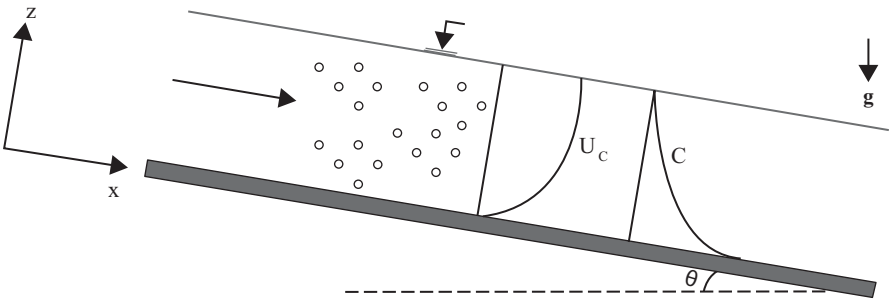

Fig. 3.1 Schematic of sediment-laden, open-channel flow (*y* axis points in the direction perpendicular to the page, adapted from [17])

According to [19], the Rousean distribution has the merit of being rather simple since it contains few parameters. However, the classical approach cannot incorporate in detail the effects of particle inertia, particle–particle interaction, and particle–fluid interaction. Greimann et al. [15] and Jiang et al. [19] stated that the Rousean distribution cannot address the velocity lag between water and the sediment in a direct manner. Further, the Rousean profile does not provide any information about the turbulence distribution in the vertical direction and cannot be extended to assess cases of non-dilute mixtures, where the sediment concentration is large. Hunt's balance does not incorporate the effects of particle inertia, particle–particle interaction, and particle–fluid interaction, or the velocity lag between phases either. These facts provide motivation for seeking a more comprehensive model.

Through the relatively recent findings of experiments [1, 20–22], it is evident that even in the case of dilute flows, the stream-wise velocity of the suspended sediment in open channels is smaller than that of the carrier fluid. By ignoring the velocity lag between the two phases (sediments and water), the suspended-sediment load calculation can be overestimated by 37% in some cases [23]. As the concentration becomes large, the velocity lag may become more significant [24]. These issues call for more general theories than simply solving a vertical balance for the sediment, as done in the Rousean formula. Two-phase or two-fluid models seem to be a natural alternative.

Figure 3.1 shows the phases in sediment-laden flows in open channels; water acts as the carrier phase, and the solid particles constitute the disperse phase. In two-phase-flow models, there are mainly two approaches in use currently: "trajectory" (Lagrangian) approaches, and "two-fluid" (Eulerian) approaches. In trajectory approaches, the motion of each particle is followed in order to describe the distribution of the disperse phase. In the two-fluid approaches, the carrier and disperse phases are described by a set of "continuum" equations representing conservation of mass and momentum of each phase within a fixed elemental volume of the mixture (conservation equations of energy can be employed as well depending on the nature of the problem). The details of these approaches can be found in [25]. The purpose of this chapter is to present a general framework

based on the two-fluid approach to model the transport of suspended sediment in open channels.

We start by presenting a framework for non-dilute transport of suspended sediment; dilute conditions appear as a special case. As expected, difficulties are encountered in solving a large number of equations interconnected by interaction forces and in the definition for diffusivities. The poor understanding of several unknown correlation terms embedded in two-phase models (TFM) requires further simplifications. We present a hierarchy of models, which are as follows: (i) one which solves the two-fluid flow equations (which we call the complete two-fluid model, CTFM), (ii) a second one, based on solving mixture equations, which can be obtained by combining the mass and momentum equations of both phases (we call it the partial two-fluid model, PTFM). After presenting the equations, we turn to discussing potential closures for the correlation terms and their importance, based on different flow conditions and different mass loadings. We finally summarize the key findings of our research on this topic.

3.2 Mathematical Models Based on the Multi Component Fluid Theory

3.2.1 General Equations for a Complete Two-Fluid Model (CTFM)

Drew et al. [26] presented the rigorous derivation of the conservation of mass, momentum, and energy for multiphase flows starting from the very basics of continuum mechanics. They employed the "ensemble" averaging procedure. According to [27], the ensemble can be thought as a large number of realizations of a physical experiment carried out under identical external conditions. The ensemble average of an arbitrary variable φ over several realizations is defined by $\langle \varphi \rangle = \int_{\sum \varepsilon} \varphi d\upsilon$, where υ indicates the density for the measure (probability) on the set of realizations. Based on the continuum assumption enforced for the phases, [26] presented the general TFM consisting of mass, momentum, and energy equations for the carrier and disperse phases. The TFM can be written as:

Conservation of mass

$$\frac{\partial}{\partial t}(\alpha_p \tilde{\rho}_p) + \frac{\partial}{\partial x_j}(\alpha_p \tilde{\rho}_p \tilde{U}_{j,p}) = \Gamma_p. \tag{3.7}$$

Conservation of momentum

$$\frac{\partial}{\partial t}(\alpha_p \tilde{\rho}_p \tilde{U}_{i,p}) + \frac{\partial}{\partial x_j}(\alpha_p \tilde{\rho}_p \tilde{U}_{i,p} \tilde{U}_{j,p}) = -\alpha_p \frac{\partial \tilde{P}_p}{\partial x_i}$$
$$+ \frac{\partial}{\partial x_j}\left[\alpha_p (\tilde{T}_{ij,p} + T_{ij,p}^{EA})\right] + \alpha_p \tilde{\rho}_p g_i \pm (F_{\text{int}})_i. \quad (3.8)$$

In the above equations, the subscript p stands for both phases (which could be c or d for the carrier and the disperse phases, respectively); α represents the volume fraction; $\tilde{\rho}, \tilde{U}$, and $\tilde{P}$ represent the ensemble-averaged density, velocity, and pressure, respectively; $\tilde{T}_{ij,p}$ refers to the ensemble-averaged deviatoric stresses of phase p; the stresses with the superscript EA are the remainder of the process of ensemble averaging; g_i is the ith component of the acceleration due to gravity; $(F_{\text{int}})_i$ includes all the interactive forces, typically, due to the drag, lift, virtual mass, and turbulent dispersion; and x and t are the space and time coordinates, respectively. The indices i and j vary from 1 to 3 and the sum is implied in repeated indices (Einstein convention).

The first and second terms on the left-hand side of Eq. 3.7 are the rate of change of mass in a point and the net rate of convective mass flux. The term on the right-hand side of Eq. 3.7 represents the mass exchange between the phases, if any. The first and second terms on the left-hand side of Eq. 3.8 are the net rate of change of momentum in a point and the net rate of momentum transfer by convection, respectively; the first term on the right-hand side is the ensemble-averaged pressure gradient; the second term represents the deviatoric stresses; the third term includes the body forces such as gravity; and the last term includes all types of interactive forces.

The TFM is the most general theory for two-phase flows. The equations explained in [26] have been used in several studies. Buscaglia et al. and Bombardelli et al. [28, 29] used these equations to derive mixture equations to model large-scale bubble plumes. Lopez De Bertodano et al. [30], in turn, developed a two-fluid model for two-phase jets for nuclear reactor safety. In sediment-transport modeling, TFM has been applied in different forms and simplifications by many researchers [15–17, 19, 31–38].

It could be thought that the TFM equations, together with the appropriate boundary conditions, are the "exact" equations that need to be solved in order to represent the intrinsic mechanics of two-phase flows. However, these equations present some mathematical shortcomings under some conditions that need further analysis. Not only no proof of existence, uniqueness and smoothness with boundary conditions have been given (as happens with the Navier-Stokes equations), but also [26] showed that even a simple case of 1D, unsteady two-phase flow is mathematically ill-posed. As a consequence, care should be taken when evaluating these equations.

Ensemble averaging, encompasses all scales of turbulence, is assumed in the derivations of Eqs. 3.7 and 3.8. Therefore, the effects of turbulence are taken into account in those equations through the turbulent dispersion force. We believe that

the ensemble average only encompasses scales of the order of the particle-to-particle distance. Under this assumption, the length and time scales associated with ensemble averaging are smaller than intermediate to large scales developed due to turbulence [17, 28, 29, 39]. Therefore, we argue that a second averaging procedure is needed to get a Reynolds-averaged Navier-Stokes (RANS) formulation, which would lead to a more involved momentum equation for each phase [33, 40, 41]. To perform a second averaging over time, we first introduce the Reynolds decomposition to each instantaneous variable as:

$$\alpha_p = \overline{\alpha_p} + \alpha_p^{'}; \tilde{U}i,p = \overline{U_{i,p}} + u_{i,p}^{'}; \tilde{P} = \overline{P} + p'; \tilde{T}_{ij,p} = \overline{T_{ij,p}} + t_{ij,p}^{'};$$
$$T_{ij,p}^{EA} = \overline{T_{ij.p}^{EA}} + t_{ij,p}^{EA'}; \left(F_{\text{int}}\right)_i = \left(\overline{F_{\text{int}}}\right)_i + \left(F_{\text{int}}^{'}\right)_i, \tag{3.9}$$

where the uppercase letters with overbar represent the mean components of the variable, and the lowercase letters with superscript indicate the fluctuations of the variable. By inserting Eq. 3.9 into Eqs. 3.7 and 3.8, we obtain:

Conservation of mass

$$\frac{\partial}{\partial t}\left[\tilde{\rho}_p\left(\overline{\alpha_p} + \alpha_p^{'}\right)\right] + \frac{\partial}{\partial x_j}\left[\tilde{\rho}_p\left(\overline{\alpha_p} + \alpha_p^{'}\right)\left(\overline{U_{j,p}} + u_{jp}^{'}\right)\right] = 0. \tag{3.10}$$

Conservation of momentum

$$\frac{\partial}{\partial t}\left[\tilde{\rho}_p\left(\overline{\alpha_p}+\alpha_p^{'}\right)\left(\overline{U_{i,p}}+u_{i,p}^{'}\right)\right] + \frac{\partial}{\partial x_j}\left[\tilde{\rho}_p\left(\overline{\alpha_p}+\alpha_p^{'}\right)\left(\overline{U_{i,p}}+u_{i,p}^{'}\right)\left(\overline{U_{j,p}}+u_{j,p}^{'}\right)\right]$$
$$= -\left(\overline{\alpha_p} + \alpha_p^{'}\right)\frac{\partial}{\partial x_i}\left[\overline{P} + p'\right] + \frac{\partial}{\partial x_j}\left[\left(\overline{\alpha_p} + \alpha_p^{'}\right)\left(\overline{T_{ij,p}} + t_{ij,p}^{'} + \overline{T_{ij,p}^{EA}} + t_{ij,p}^{EA'}\right)\right]$$
$$+\tilde{\rho}_p\left(\overline{\alpha_p} + \alpha_p^{'}\right)g_i + \left(\overline{F_{\text{int}}}\right)_i + \left(F_{\text{int}}^{'}\right)_i. \tag{3.11}$$

By performing averaging over Eq. 3.10, the following equation holds true (we have removed the symbol of ensemble average from the equation):

$$\frac{\partial}{\partial t}\left[\overline{\alpha_p}\rho_p\right] + \frac{\partial}{\partial x_j}\left[\rho_p\,\overline{\alpha_p U_{j,p}} + \rho_p\,\overline{\alpha_p^{'} u_{j,p}^{'}}\right] = 0. \tag{3.12}$$

Similarly, by performing averaging over Eq. 3.11, we obtain:

$$\frac{\partial}{\partial t}\left[\rho_p \overline{\alpha_p U_{i,p}} + \rho_p \overline{\alpha_p' u_{i,p}'}\right] + \frac{\partial}{\partial x_j}\begin{bmatrix} \rho_p \overline{\alpha_p}\,\overline{U_{i,p}}\,\overline{U_{j,p}} + \rho_p \overline{\alpha_p}\,\overline{u_{i,p}' u_{j,p}'} + \\ \rho_p \overline{U_{i,p}}\,\overline{\alpha_p' u_{j,p}'} + \rho_p \overline{U_{j,p}}\,\overline{\alpha_p' u_{i,p}'} + \\ \rho_p \overline{\alpha_p' u_{i,p}' u_{j,p}'} \end{bmatrix} = -\overline{\alpha_p}\frac{\partial}{\partial x_i}\left[\overline{P}\right] + \frac{\partial}{\partial x_j}\begin{bmatrix} \overline{\alpha_p}\,\overline{T_{ij,p}} + \overline{\alpha_p}\,\overline{T_{ij,p}^{EA}} + \\ \overline{\alpha_p' t_{ij,p}'} + \overline{\alpha_p' t_{ij,p}^{EA'}} \end{bmatrix} + \overline{\alpha_p}\rho_p g_i + \left(\overline{F_{\text{int}}}\right)_i \tag{3.13}$$

Equation 3.13 can be approximated as:

$$\frac{\partial(\rho_p \overline{\alpha_p U_{i,p}})}{\partial t} + \frac{\partial}{\partial x_j}(\rho_p \overline{\alpha_p U_{i,p} U_{j,p}}) = -\overline{\alpha_p}\frac{\partial \overline{P}}{\partial x_i} + \frac{\partial}{\partial x_j}\left[\overline{\alpha_p}\left(\overline{T_{ij,p}} + \varsigma_{ij,p}\delta_{p,d} + \overline{T_{ij,p}^{\text{Re}}}\right)\right] + \overline{\alpha_p}\rho_p g_i \pm (F_{\text{int}})_i. \tag{3.14}$$

Following [42] in the derivation of Eq. 3.14, we have neglected several correlations of fourth order and third order, such as $\overline{\alpha_p' u_{i,p}' u_{j,p}'}$. We also collapse two time derivatives of Eq. 3.13 into a single term (first term in Eq. 3.14). We consider that the terms $\rho_p \overline{\alpha_p}\,\overline{u_{i,p}' u_{j,p}'} + \rho_p \overline{U_{i,p}}\,\overline{\alpha_p' u_{j,p}'} + \rho_p \overline{U_{j,p}}\,\overline{\alpha_p' u_{i,p}'}$ represent the Reynolds stresses and Reynolds solid fluxes. In Eq. 3.14, this term is represented by $\overline{T_{ij,p}^{\text{Re}}}$. We neglect the pressure fluctuation term on the right-hand side of Eq. 3.13. We further assume that the terms $\overline{\alpha_p}\,\overline{T_{ij,p}^{EA}} + \overline{\alpha_p' t_{ij,p}'} + \overline{\alpha_p' t_{ij,p}^{EA'}}$ are included in the Reynolds stresses as well. In Eq. 3.14, the $\overline{T_{ij,p}} + \varsigma_{ij,p}\delta_{p,d} + \overline{T_{ij,p}^{\text{Re}}}$ term is assumed to contain contributions of the viscous deviatoric stresses, the ensemble-averaged stresses as well as the contribution from the turbulent fluctuations. In the cases of non-dilute flows, additional stresses develop due to interparticle collisions. Similar to [43], we added a term $\varsigma_{ij,p}\delta_{p,d}$ in Eq. 3.15, where $\delta_{p,d}$ is 0 when $p = c$, and 1 when $p = d$. Equations3.12 and 3.14 constitute the CTFM.

3.2.2 General Equations for a Partial Two-Fluid Model (PTFM)

Extending the works presented in [28, 29, 44, 45], the mixture equations can be derived by combining the mass and momentum equations of the two phases. This

approach has the advantage in reducing the complexities involved in solving nonlinear and interdependent equations of the CTFM. We introduce the mixture variables as:

$$\rho_m U_m = \alpha_c \tilde{\rho}_c \tilde{U}_c + \alpha_d \tilde{\rho}_d \tilde{U}_d; \tag{3.15}$$

$$\rho_m = \alpha_c \tilde{\rho}_c + \alpha_d \tilde{\rho}_d; \tag{3.16}$$

$$T_m = \alpha_c \tilde{T}_c + \alpha_d \tilde{T}_d; \tag{3.17}$$

where the subscript m stands for mixture and ~ denotes ensemble averaging. By combining Eqs. 3.7 and 3.9 and using the mixture variables, the following equations for the mixture can be obtained:

Conservation of mass for the mixture:

$$\frac{\partial \rho_m}{\partial t} + \frac{\partial}{\partial x_j}(\rho_m U_{j,m}) = 0. \tag{3.18}$$

Conservation of momentum for the mixture:

$$\frac{\partial \rho_m U_{i,m}}{\partial t} + \frac{\partial}{\partial x_j}\left(\rho_m U_{i,m} U_{j,m}\right) + \frac{\partial P}{\partial x_i} = \frac{\partial T_{ij,m}}{\partial x_j} + \rho_m g_i$$
$$- \frac{\partial}{\partial x_j}\left[\alpha_d \tilde{\rho}_d \left(1 - \frac{\alpha_d \tilde{\rho}_d}{\rho_m}\right)\left(\tilde{U}_{i,c} - \tilde{U}_{i,d}\right)\left(\tilde{U}_{j,c} - \tilde{U}_{j,d}\right)\right]. \tag{3.19}$$

As discussed in the previous section, we perform time averaging of Eqs. 3.18 and 3.19 by decomposing the variables in their mean and fluctuation components as follows:

$$U_{i,m} = \overline{U_{i,m}} + u'_{i,m}; \; T_m = \overline{T_m} + t'_m, \tag{3.20}$$

Where the uppercase letters with overbar represent the mean component of the variable and the lowercase letters with superscript indicate fluctuating components of the variable. The resulting equations are (after removing the symbol of ensemble average from the equations):

Conservation of mass for the mixture:

$$\frac{\partial \rho_m}{\partial t} + \frac{\partial}{\partial x_j}\left(\rho_m \overline{U_{j,m}}\right) = 0. \tag{3.21}$$

Conservation of momentum for the mixture:

$$\frac{\partial\left(\rho_m \overline{U_{i,m}}\right)}{\partial t} + \frac{\partial}{\partial x_j}\left(\rho_m \overline{U_{i,m}}\,\overline{U_{j,m}} + \rho_m \overline{u'_{i,m} u'_{j,m}}\right) + \frac{\partial \overline{P}}{\partial x_i}$$
$$= \frac{\partial \overline{T_{ij,m}}}{\partial x_j} + \rho_m g_i - \frac{\partial}{\partial x_j}\left\{\overline{\rho_d\left[\overline{\alpha_d} + \alpha'_d\right]\left[1 - \frac{\rho_d\left(\overline{\alpha_d} + \alpha'_d\right)}{\rho_m}\right]\left[\left(\overline{U_{i,c}} + u'_{i,c} - \overline{U_{i,d}} - u'_{i,d}\right)\left(\overline{U_{j,c}} + u'_{j,c} - \overline{U_{j,d}} - u'_{j,d}\right)\right]}\right\}. \tag{3.22}$$

Equation 3.22 is still involved and contains terms which require further closure and it adds uncertainty and complexity to the models. The last term of Eq. 3.22 possesses several correlations of second, third, and fourth orders. The current knowledge on this topic does not allow us to propose a certain specific closure to these unknown terms [42]. Thus, we propose to model the momentum equation in the following simplified form:

$$\frac{\partial\left(\rho_m \overline{U_{i,m}}\right)}{\partial t} + \frac{\partial}{\partial x_j}\left(\rho_m \overline{U_{i,m}}\,\overline{U_{j,m}}\right) + \frac{\partial \overline{P}}{\partial x_i}\frac{\partial}{\partial x_j}\left(\overline{T_{ij,m}} + \overline{T_{ij,m}^{\mathrm{Re}}}\right) + \rho_m g_i$$
$$- \frac{\partial}{\partial x_j}\left\{\left[\overline{\alpha_d}\,\rho_d\left(\left(1 - \frac{\overline{\alpha_d}\,\rho_d}{\rho_m}\right)\left(\overline{U_{i,m}} - \overline{U_{i,d}}\right)\left(\overline{U_{j,m}} - \overline{U_{j,d}}\right)\right)\right]\right\}. \tag{3.23}$$

It is worth mentioning that the last term of Eq. 3.23 disappears when a dilute mixture is considered. Further simplification to the PTFM model is introduced by employing the following algebraic relation instead of the momentum equation for the disperse phase:

$$\overline{U_{j,d}} = \overline{U_{j,m}} + \overline{W_{d,j}}, \tag{3.24}$$

where W_d is the relative velocity vector of the disperse phase. Equations 3.21, 3.23, and 3.24 constitute the PTFM. In the CTFM model, the velocity of the disperse

phase in the wall-normal direction is obtained by solving the momentum equation. But in the case of PTFM, the wall-normal velocity of the sediments is computed from an algebraic model.

3.2.3 Closure to the Models

In the above model, we need closures for the Reynolds fluxes, Reynolds stresses, eddy viscosity of the carrier and disperse phases, stresses due to interparticle collisions, and for the interactions forces. Following our previous papers on this topic, we summarize the closures in Table 3.1. In Eq. T1, the Reynolds fluxes are defined according to the gradient of the time-averaged solid fraction (gradient-transport hypothesis [11]). In Eq. T2, $\nu_{T,c}$ is the eddy viscosity of the carrier phase and Sc is the Schmidt number. Note that the Schmidt number also appears in Eq. 3.5 in the standard Rousean model as the inverse of β parameter. Several authors have proposed different formulae to determine the Schmidt number [17, 35, 46, 47]. There is no consensus whether the Schmidt number should be larger or smaller than one for a given flow condition. Cellino et al. [48] and Yoon et al. [49] obtained *Sc* greater than one even for dilute cases.

The Reynolds stresses of the carrier phase ($\overline{u'_{i,c}u'_{j,c}}$) can be obtained either rigorously by solving the transport equations for the components of the stress tensor leading to the Reynolds stress model (RSM) [50], or using Boussinesq model as presented in Eq. T3.

The Reynolds stress of the disperse phase ($\overline{u'_{i,d}u'_{j,d}}$) can also be approximated by using Eq. T3. Equation T3 involves $\nu_{T,c}$ and $\nu_{T,d}$ which needs further closure. The $\nu_{T,c}$ can be defined in terms of turbulent kinetic energy (K) and its dissipation rate (ε). The transport equations of K and ε are presented in [18]. For dilute flows, the Reynolds stresses in the disperse phase can be assumed as negligible; thus the problem of defining $\nu_{T,d}$ does not arise. Jha et al. [18] has tested the definition as proposed in Eqs. T5 and T6 for simulation of non-dilute flows in open channels.

Some researchers argue that the transport equations for the closure of Reynolds stresses of the carrier phase has to be extended for two-phase flows. Jha et al. [51] proposed extensions for the RSM, $K-\varepsilon$ and $K-\omega$ models. In the case of dilute flows, they found that the use of the RSM does not improve the model prediction beyond the accuracy of the $K-\varepsilon$ model when compared with the data of [21, 52, 53]; however, different turbulence closures needed different values of the Schmidt number. The extensions proposed in the $K-\varepsilon$ model were found to have negligible effect on the model results. The accuracy of the application of extended equations for turbulence closures in the case of non-dilute flow is still under debate. Some researchers have also advocated the idea of modifying the coefficients used in the $K-\varepsilon$ model [35, 54].

For non-dilute flows, further complexity arises because of the need of closure for stresses due to interparticle collisions. Jha et al. [18] and Greimann et al. [34] successfully applied the model originally proposed by [43] as presented in Eq. T7.

Table 3.1 Closures of correlation terms in the CTFM and PTFM equations

	Equation number
Solid-fraction Reynolds fluxes: $-\overline{\alpha_d' w_c'} = -\overline{\alpha_d' w_d'} = D_d \frac{\partial \alpha_d}{\partial z}$	T1
Sediment diffusivity: $D_d = \frac{\nu_{T,c}}{Sc}$	T2
Reynolds stresses: $T_{ij}^{\mathrm{Re}} = \rho_c \left[\nu_T \left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} \right) - \frac{2}{3} \delta_{ij} K \right]$	T3
Eddy viscosity of the carrier phase: $\nu_{T,c} = C_\mu \frac{K^2}{\varepsilon}$	T4
Eddy viscosity of the disperse phase: $\nu_{T,d} = l_{md}^2 \left\| \frac{dU_d}{dz} \right\|$	T5
Mixing length: $l_{md} = \kappa z \left(1 - \frac{z}{h} \right)^{0.5(1+\beta \alpha_d / \alpha_{d,b})}$	T6
Stresses due to interparticle collisions: $\alpha_d \left(\zeta_d \right)_{ij} = -2\alpha_d^2 \rho_d g_o (1+e) \tau \delta_{ij} - \frac{4}{5} \alpha_d^2 \rho_d g_o (1+e) \left(M_{ij} - \tau \delta_{ij} \right) + \frac{4}{5} \alpha_d^2 \rho_d d_p g_o (1+e) \sqrt{\frac{\tau}{\pi}} \left(2(S_d)_{ij} + \left(\frac{\partial (U_d)_m}{\partial x_m} \right) \delta_{ij} \right)$	T7
Drag force: $\overrightarrow{F_D} = \left[\frac{3}{4} \alpha_d \rho_m C_D \sqrt{(U_m - U_d)^2 + (W_m - W_d)^2} \right] \left((U_m - U_d) \vec{i} + (W_m - W_d) \vec{k} \right)$	T8
Lift force: $\overrightarrow{F_L} = C_L \alpha_d \rho_m \left[-W_d \frac{\partial U_m}{\partial z} \vec{i} + \left(U_m - U_d \right) \frac{\partial U_m}{\partial z} \vec{k} \right]$	T9
Virtual mass force: $\overrightarrow{F_{VM}} = \frac{\alpha_d \rho_m}{2} \left[\left(\frac{\partial \left(U_m - U_d \right)}{\partial t} - W_d \frac{\partial U_d}{\partial z} \right) \vec{i} - \left(\frac{\partial W_d}{\partial t} + W_d \frac{\partial W_d}{\partial z} \right) \vec{k} \right]$	T10
Coefficient of drag: $C_D = \frac{24}{\mathrm{Re}_r} \left(1 + 0.15 \mathrm{Re}_r^{0.687} \right)$	T11

In Eq. T7, $g_o = (1 - \alpha_d / \alpha_{d,\max})^{-2.5 \alpha_{d,\max}}$; τ is the granular temperature, which can be defined as $\tau = 2K_d / 3$; K_d is the turbulence kinetic energy of the disperse phase; $\alpha_{d,\max}$ is the maximum packing concentration which has the value of 0.53 as an upper limit; and e is the coefficient of restitution of particle collision.

Equations T8, T9, T10 represent the interaction forces of drag, lift, and virtual mass, respectively. Based on the nature of the flow, a combination of these forces

is necessary to consider in the modeling equations. These force terms contain variables of both the carrier phase and the disperse phase. Therefore, they make the equations coupled.

3.3 Assessment of 1D Versions of the CTFM and PTFM in Modeling Dilute and Non-Dilute Flows

The observations presented in this section are based on our findings resulting from the application of 1D versions of the CTFM and PTFM in its non-dilute and dilute versions. We compared our simulation results with experimental data available in the literature for the mean velocity of the carrier and disperse phases, profiles of concentration of sediment in the flow, and turbulent statistics. Here we present the conclusions of our work. We refer readers to our papers for details.

3.3.1 Model Complexity Necessary to Represent Mean Velocity

The complete solution of the CTFM is cumbersome. Even for a 1D case, the number of basic governing equations is six (mass, momentum in stream-wise direction and momentum in wall-normal direction for each phase). Further, the closure for the turbulence adds more equations to the list. For example, in the $K-\varepsilon$ model, there are two transport equations. So there are total eight transport equations which are interconnected. In the case of PTFM, the number of equations is the same as in the CTFM but there is no need to solve the momentum equation of the disperse phase in the wall-normal direction (Eq. 3.24). In case of dilute flows, we showed in [17] that although the CTFM should, in principle, provide more accurate results given its complexity and "rigor," in practice, it does not provide significant improvement over the PTFM. When the magnitude of the maximum sediment concentration at the river bed boundary exceeds 10 %, the CTFM is found to be necessary to obtain accurate predictions of mean velocities. In Fig. 3.2, we show the profiles of stream-wise velocity obtained from CTFM and PTFM compared with S13, S14, S16 datasets of [55]. The maximum concentration at the bed boundary in S13, S14, and S16 are 13, 14, and 23 %, respectively. It is clear from this figure that as the maximum concentration of sediment at the bed boundary increases, results obtained from PTFM begin to deviate. It is very important that the modeler determines the level of complexity necessary in a given study and selects the most accurate approach.

3.3.2 Importance of Forces

In non-dilute flows, the relative magnitude of the lift and virtual mass forces with respect to the drag force are found to be smaller than 5 and 25 %, respectively. In case of dilute flows, the lift force can be disregarded as opposed to the drag force.

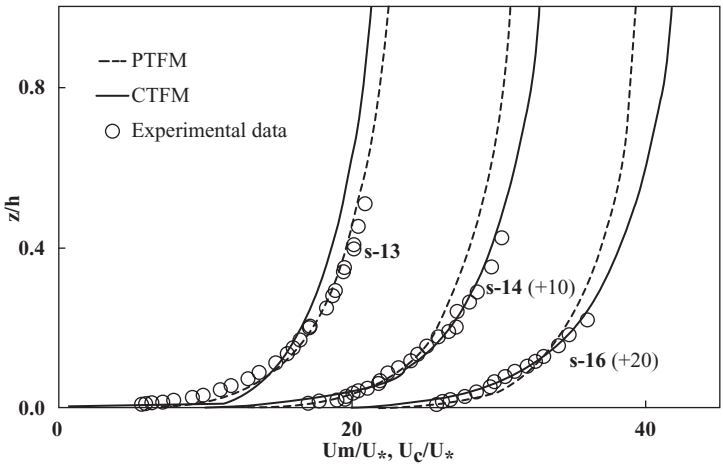

Fig. 3.2 Comparison of simulated values of the stream-wise velocity obtained from the 1D PTFM and the 1D CTFM (profiles shifted to fit into one plot, adapted from [18])

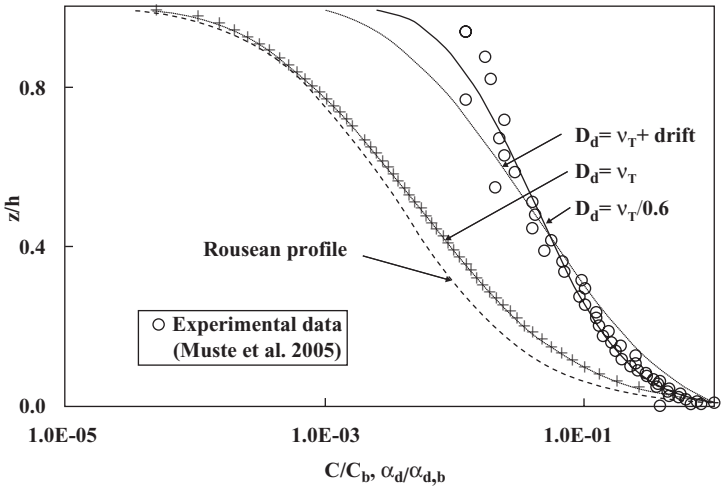

Fig. 3.3 Comparison of simulated values of the concentration of sediment with the experimental data of [53] (simulations performed using 1D PTFM together with different formulations for the sediment diffusion coefficient)

3.3.3 Sediment Concentration

Most of our models were found to represent the velocity relatively accurately within a range of 5–10% difference with data. In predicting the sediment volume fraction in dilute and non-dilute flows, two-phase flow equations bring a substantial improvement with respect to the Rousean distribution, albeit with the need of adjusting the Schmidt number (*Sc*). In Fig. 3.3, we show the results of applying the 1D

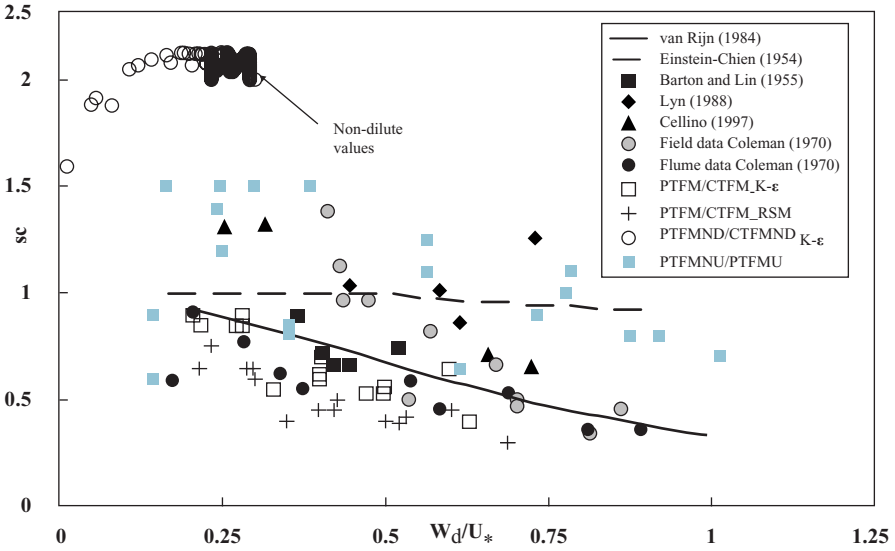

Fig. 3.4 Comparsion of values of the Schmidt number (*Sc*) obtained from runs with the *K*-*ε* model (for the dilute and non-dilute models) and those obtained from observations (PTFMND, CTFMND refer to the non-dilute version of PTFM and CTFM, PTFMNU and the PTFMU refer to PTFM model applied to test cases with sediment mixtures of non-uniform and uniform sizes, respectively, adapted from [45])

version of PTFM for simulating dilute transport of sediments. The reference data is obtained from [53]. We observe that the Rousean distribution with $Sc=1$ does not provide results close to the data. The *Sc* number was found to be less than one to approximate the concentration profile with the data ($Sc=0.6$). We carried out several simulations of dilute and non-dilute flows and our results indicate that the value of *Sc* is less than one for dilute flows and greater than one for non-dilute flows. These findings indicate that the diffusivity of momentum of the carrier phase is smaller than the diffusivity of sediment in dilute flows. In non-dilute flows, the values of *Sc* imply that the diffusivity of the sediment is reduced in a sizeable way with respect to the flow eddy viscosity. Figure 3.4 shows that our findings agree with observations reported elsewhere.

3.4 Concluding Remarks

This chapter summarizes a hierarchy of models based on multi-phase flow theory applied to investigating dilute and non-dilute transport of sediments in open channels. The following insights are highlighted from our findings:

i. Traditional mixture flow model is too simplistic in formulation to account for all the mechanisms involved in sediment transport in open channels.
ii. The TFM model presented in various textbooks are supplemented with a second averaging that modifies the way turbulence is treated.
iii. The resulting model requires further definition of several correlation terms which do not have well-established formulations. It is left up to the modeler to apply proper assumptions and reliable closures. However, our work indicates the necessity and/or advantage for using some formulations over alternatives.
iv. Based on the concentration of the sediment present in the flow, some simplifications can be introduced in the two-phase flow equations.
v. Further understanding of velocity lag between the sediment and water, and turbulence modulation can only be possible through more experimental tests usable for validation of complex two-phase flow models.

3.5 List of Symbols

C	Concentration of suspended sediment
D_c	Diffusivity of momentum of the mixture
D_d	Vertical diffusivity for sediment
d_p	Diameter of particles
e	Coefficient of restitution
F_{int}	Interaction forces
g	Acceleration due to gravity
h	Depth of the channel
K	Turbulent kinetic energy
K_s	Grain roughness equivalent height
l_{md}	Mixing length
P	Pressure
S	Source term
S_b	Slope of the channel bed
T	Deviatoric stresses
U	Mean velocity in the stream-wise direction
u'	Velocity fluctuations in the stream-wise direction
u_*	Wall-friction (shear) velocity
W	Mean velocity in the wall-normal direction
w'	Velocity fluctuations in the wall-normal direction
W_S	Settling velocity
x	Stream-wise coordinate
y	Transverse coordinate
z	Wall-normal coordinate
t	Time coordinate

3.6 Greek symbols

Sc	Schmidt number
κ	von-Kármán constant
ε	Dissipation rate of turbulent kinetic energy
α	Phase fraction
β	Damping coefficient
ρ	Density
Γ	Mass sources (or sinks) due to the convective or molecular fluxes
ς	Stress due to interparticle collisions
δ_{ij}	Kronecker delta
τ	Granular temperature
ν	Eddy viscosity
θ	Angle between the bed of the channel and a horizontal line

3.7 Subscripts

c	Relative to the carrier phase
d	Relative to the disperse phase
m	Relative to the mixture
p	Phase

3.8 Superscripts

`	Fluctuation
Re	Explicit particle Reynolds number

Acknowledgments The results presented in this chapter were obtained during the graduate studies of the first author at the University of California, Davis. Thanks to the grant awarded to Fabián Bombardelliby, the Cooperative Institute for Coastal and Estuarine Environmental Technology (CICEET), the California Department of Water Resources (DWR), and the California Water Resources Control Board (WRCB). The book chapter was supported through a post-doctoral research fellowship awarded to the first author as part of the National Center for Groundwater Research and Training (NCGRT), a joint initiative between the Australian Research Council and the National Water Commission.

References

1. Best J, Bennett S, Bridge J, Leeder M (1997) Turbulence modulation and particle velocities over flat sand beds at low transport rates. J Hydraul Eng 123(12):1118–1129

2. Chien N, Wan Z (1999) Mechanics of sediment transport. ASCE Press, Reston
3. Guo J, Julien PY (2001) Turbulent velocity profiles in sediment-laden flows. J Hydraul Res 39(1):11–23
4. Lyn D (1991) Resistance in flat-bed sediment-laden flows. J Hydraul Eng 117(1):94–114
5. Muste M, Yu K, Fujita I, Ettema R (2009) Two-phase flow insights into open-channel flows with suspended particles of different densities. Environ Fluid Mech 9(2):161–186
6. Nezu I, Azuma R (2004) Turbulence characteristics and interaction between particles and fluid in particle-laden open channel flows. J Hydraul Eng 130(10):988–1001
7. Rouse H (1937) Modern conceptions of the mechanics or fluid turbulence. Transact ASCE 102(1):463–505
8. Parker G (2004) 1D sediment transport morphodynamics with applications to rivers and turbidity currents, E-Book. Natl. Cent. For Earth Surf. Dyn., Minneapolis, Minn. Http://Cee.Uiuc.Edu/People/Parkerg/. Accessed 03 June 2013
9. Hunt J (1954) The turbulent transport of suspended sediment in open channels. Proc Royal Soc London Series A. Math Phys Sci 224(1158):322–335
10. Vanoni VA (1984) Fifty years of sedimentation. J Hydraul Eng 110(8):1021–1057
11. Pope SB (2000) Turbulent flows. Cambridge University Press, Cambridge
12. Vanoni VA (1946) Transportation of suspended sediment by water. Trans ASCE 111(1):67–102
13. Garcia MH (2008) Sediment transport and morphodynamics. Sediment Eng Process Manag Model Pract ASCE Manual no 110:21–163
14. Yalin MS (1977) Mechanics of sediment transport. Pergamon, Oxford
15. Greimann BP, Muste M, Holly FM Jr (1999) Two-phase formulation of suspended sediment transport. J Hydraul Res 37(4):479–500
16. Hsu TJ, Jenkins JT, Liu PLF (2003) On two-phase sediment transport: dilute flow. J Geoph Res Oceans (1978–2012) 108(C3)
17. Bombardelli FA, Jha SK (2009) Hierarchical modeling of the dilute transport of suspended sediment in open channels. Environ Fluid Mech 9(2):207–235
18. Jha SK, Bombardelli FA (2010) Toward two-phase flow modeling of nondilute sediment transport in open channels. J Geoph Res 115(F3):F03015
19. Jiang J, Law AW-K, Cheng N-S (2005) Two-phase analysis of vertical sediment-laden jets. J Eng Mech 131(3):308–318
20. Rashidi M, Hetsroni G, Banerjee S (1990) Particle-turbulence interaction in a boundary layer. Int J Multiphase Flow 16(6):935–949
21. Muste M, Patel V (1997) Velocity profiles for particles and liquid in open-channel flow with suspended sediment. J Hydraul Eng 123(9):742–751
22. Kiger KT, Pan C (2002) Suspension and turbulence modification effects of solid particulates on a horizontal turbulent channel flow. J Turbulence 3(019):1–21
23. Aziz NM (1996) Error estimate in einstein's suspended sediment load method. J Hydraul Eng 122(5):282–285
24. Cheng N-S (2004) Analysis of bedload transport in laminar flows. Adv Water Resour 27(9):937–942
25. Loth E (2008) Computational fluid dynamics of bubbles, drops and particles. Cambridge University Press, Cambridge
26. Drew DA, Passman SL (1999) Theory of multicomponent fluids. Springer, New York
27. Zdunkowski W, Bott A (2003) Dynamics of the atmosphere: a course in theoretical meteorology. Cambridge University Press, Cambridge
28. Buscaglia GC, Bombardelli FA, García MH (2002) Numerical modeling of large-scale bubble plumes accounting for mass transfer effects. Int J Multiphase Flow 28(11):1763–1785
29. Bombardelli FA (2004) Turbulence in multiphase models for aeration bubble plumes. University of Illinois At Urbana-Champaign, Illinois
30. Lopez De Bertodano MA (1998) Two fluid model for two-phase turbulent jets. Nuclear Eng Design 179(1):65–74
31. Drew D (1975) Turbulent sediment transport over a flat bottom using momentum balance. J Appl Mech 42:38

32. Mctigue DF (1981) Mixture theory for suspended sediment transport. J Hydraul Di 107(6):659–673
33. Cao Z, Wei L, Xie J (1995) Sediment-laden flow in open channels from two-phase flow viewpoint. J Hydraul Eng 121(10):725–735
34. Greimann BP, Holly FM Jr (2001) Two-phase flow analysis of concentration profiles. J Hydraul Eng 127(9):753–762 (2001
35. Amoudry L, Hsu TJ, Liu PF (2005) Schmidt number and near-bed boundary condition effects on a two-phase dilute sediment transport model. J Geoph Res Oceans (1978–2012) 110(C9)
36. Chauchat J (2007) Contribution to two-phase flow modeling for sediment transport in estuarine and coastal zones. Phd Thesis, University of Caen, France (In French)
37. Chauchat J, Guillou S (2008) On turbulence closures for two-phase sediment-laden flow models. J Geoph Res Oceans (1978–2012) 113(C11)
38. Toorman EA (2008) Vertical mixing in the fully developed turbulent layer of sediment-laden open-channel flow. J Hydral Eng 134(9):1225–1235
39. Bombardelli FA, González AE, Niño YI (2008) Computation of the particle basset force with a fractional-derivative approach. J Hydraul Eng 134(10):1513–1520
40. Hrenya CM, Sinclair JL (1997) Effects of particle-phase turbulence in gas-solid flows. Aiche Journal 43(4):853–869
41. Prosperetti A, Zhang D (1995) Finite-particle-size effects in disperse two-phase flows. Theoret Comp Fluid Dynamics 7(6):429–440
42. Elghobashi S, Abou-Arab T (1983) A two-equation turbulence model for two-phase flows. Phys Fluids 26:931
43. Enwald H, Peirano E, Almstedt A-E (1996) Eulerian two-phase flow theory applied to fluidization. Int J Multiphase Flow 22:21–66
44. Bombardelli FA, Buscaglia GC, García MH (2003) Parallel computations of the dynamic behavior of bubble plumes. ASME, USA
45. Jha SK, Bombardelli FA (2011) Theoretical/numerical model for the transport of non-uniform suspended sediment in open channels. Adv Water Resour 34(5):577–591
46. Czernuszenko W (1998) Drift velocity concept for sediment-laden flows. J Hydraul Eng 124(10):1026–1033
47. Van Rijn LC (1984) Sediment transport, Part I: Bed load transport. J Hydraul Eng 110(10):1431–1456
48. Cellino M, Graf W (2000) Experiments on suspension flow in open channels with bed forms. J Hydraul Res 38(4):289–298
49. Yoon J-Y, Kang S-K (2005) A numerical model of sediment-laden turbulent flow in an open channel. Can J Civil Eng 32(1):233–240
50. Laín S, Aliod R (2000) Study on the eulerian dispersed phase equations in non-uniform turbulent two-phase flows: discussion and comparison with experiments. Int J Heat Fluid Flow 21(3):374–380
51. Jha SK, Bombardelli FA (2009) Two-phase modeling of turbulence in dilute sediment-laden, open-channel flows. Environ Fluid Mech 9(2):237–266
52. Lyn D (1988) A similarity approach to turbulent sediment-laden flows in open channels. J Fluid Mech 193(1):1–26
53. Muste M, Yu K, Fujita I, Ettema R (2005) Two-phase versus mixed-flow perspective on suspended sediment transport in turbulent channel flows. Water Resour Res 41(10):W10402
54. Squires KD, Eaton JK (1990) Particle response and turbulence modification in isotropic turbulence. Phys Fluids A Fluid Dyn 2:1191
55. Einstein HA, Chen N (1955) Effects of heavy sediment concentration near the bed on velocity and sediment distribution. M. R. D. Sdiment Series. N0. 8, Univ. California, Berkeley, USA

Chapter 4
Mathematical Modelling of Peristaltic Pumping of Nano-Fluids

Dharmendra Tripathi and O. Anwar Bég

4.1 Introduction

4.1.1 Mathematical Modelling

Modelling a system is a Herculean task as the vast majority of systems which are studied, are extremely complex. A system may be *open,* i.e. the factors influencing it are numerous and are affected by the surroundings. Simpler to simulate are *closed* systems, in which, given justifiable assumptions, all components are precisely determinable. This type of system can be modelled with confidence and accuracy. This makes the identification of an exact problem more challenging. Hence, any attempt to model a system is founded on some inherent physical assumptions and some degree of simplification which makes the system theoretically "closed" and renders a robust formulation for the processes involved in the system practicable. We, therefore, make attempts to model a particular phenomenon of a system by initially ignoring the parameters with less influence on the phenomenon. The model is improved by gradually incorporating more and more parameters and applying advanced mathematical know-how, reinforced with experimentation and observation, where possible. Nevertheless, the parameters with meager effects remain neglected. The reliability of a model is dependent on the degree of exactness. Modelling is, therefore, a pragmatic attempt to simulate a reduced complexity and reality. In biological systems, such as transport processes in the human body [1] and external aerodynamics of natural fliers [2], mathematical modelling provides insights which may not be realizable with practical experimentation. In conjunction with computers, mathematical simulation in biology [3] now provides one of the

D. Tripathi (✉)
Department of Mathematics, National Institute of Technology Delhi, 110077 Delhi, India
e-mail: dtripathi@nitdelhi.ac.in

O. A. Bég
Gort Engovation Research (Propulsion, Nanofluids and Biomechanics), Southmere Ave., Bradford, BD73NU UK
e-mail: gortoab@gmail.com

S. K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_4,

most challenging, intriguing, and rewarding areas of scientific endeavour. In the same way that the twentieth century was the era of nuclear science and high-speed transport (aircraft, trains, automobiles and planetary exploration rovers) in which engineers mimicked and attempted to improve upon nature [4], the twenty-first century has emerged as a new frontier for biological modelling.

There has been a tremendous fusion of biologists, mathematicians, biomechanical engineers, biochemists and biophysicists, collectively focusing on resolving many types of problem at different scales [5]. In the same way that engineers mimicked biological mechanisms in the twentieth century (e.g. flight, naval propulsion, tensile structures), biology is now implementing "smart" technologies developed for astronautics, nanotechnology, microelectronics, lubrication technology (tribology), seismic bearings for bridge structures, etc. [6]. This transfer of engineering and scientific technology into medical and biological simulation has accelerated developments in many exciting and critical areas. Paramount among these has been the application of fluid mechanics to medical systems [7], i.e. *biofluid mechanics*. Blood is often termed the "fluid of life" and hemodynamics has received most attention from engineering scientists and mathematicians [8–11]. However, "biofluid mechanics" has infiltrated into a much wider spectrum of biomechanical problems. Excellent examples in this regard are hydroelastic flows mimicking marine swimmers [12], fish school group propulsion used to design optimal wind turbine farm layouts [13], mimicry of flapping wings for micro-unmanned air vehicles (mUAVs) [14], magnetic control of surgical extracorporeal blood flow circuits [15], hydroelastic vibration of inner ear membranes [16], peristaltic propulsion in the gastric system [17–19] and haemodialysis simulations [20–21]. Further, recent developments in the application of fluid mechanics to biomedical systems include viscous flow analysis of ophthalmic diseases [22], haematological purification devices [23], cerebrospinal flows [24], marine plankton dynamics [25], smart magnetic lubrication for prosthetics [26–27], larynx dynamics [28], biomagnetic response during astronaut re-entry [29–30], artificial heart-valve mechanics [31], and flows through capillaries and small blood vessels [32] and nasal ventilation aerodynamics [33]. Many of these applications have benefited from developments in chemical, aerospace, civil, mechanical, and computer engineering, and of course mathematics and physics. Numerical methods, applied mathematics, theoretical chemistry and physics, smart mechanics, laser Doppler anemometry (LDA), particle image velocimetry (PIV), lasers and computer science (imaging, scanning, simulation) are just some of the areas, which although originally developed to solve engineering problems have now found their way into the rapidly expanding domain of mathematical biosciences. The use of mathematical simulation in particular has resulted in garnering new insights into cerebral cortex formation, cancer spread, cartilage degeneration, epidemic disease prediction, pharmacology and even psychology. Many of these topics have utilized aspects of bio-fluid mechanics.

A particularly rich area which has emerged recently is that of nanotechnology, of which nano-fluid [34] is an example. Nano-fluids were developed by Choi [34] at the Argonne Energy Labs, Illinois in the 1990s. Applications have penetrated almost every area of engineering including rocket propellants, solar energy collectors,

lubricant design, automotive and electronic cooling systems and medicine [35]. Coupled with this, many developments have taken place in micro-electromechanical systems and nano-systems including peristaltic pumps for medical applications [36]. In this discussion we, therefore, focus on recent progress in mathematical simulation of nano-fluid propulsion by peristaltic mechanisms. This topic has immense applications in surgical exploration [37], drug delivery [38] hyperthermia cancer medication deployment [39], wound healing [40] and gastric pharmacological drug targeting [41, 42]. Recently the leading Swiss medical engineering corporation, Levitronix [43] has explored the fabrication of peristaltic nano-fluid pumps for a range of pharmaceutical applications where the peristaltic transport mechanism has been shown to achieve maximum reliability, long life, and superior ability to pump precious fluids in the harshest of environments, compared with any other type of micropump. The drawbacks of nanoparticle agglomeration, dilution and dosing as well as filtration difficulties have now largely been eradicated in modern nano-fluid peristaltic devices. This field is, therefore, very promising and will doubtlessly stimulate increased attention from the mathematical modelling community. We shall briefly review the mechanism of peristalsis, and then review the thermophysics and dynamics of nano-fluids. Finally, a new model simulating double-diffusive pumping of nano-fluids in peristaltic transport is presented with future recommendations for new simulations.

4.1.2 Peristaltic Transport

Peristalsis is a physiological mechanism (pumping process) in which physiological fluids are propelled (pumped) within living organs by contraction of circular smooth muscle behind the fluids and relaxation of circular smooth muscle ahead of it. Bayliss and Starling [44] historically first observed this phenomenon over a century ago. This type of pumping was first observed in physiology regarding:

1. Food movement in the digestive tract
2. Frine transportation from the kidney to the bladder through ureters
3. Semen movement in the vas deferens
4. Movement of lymphatic fluids in the lymphatic system
5. Bile flow from the gall bladder into the duodenum
6. Spermatozoa in the ductus efferent of the male reproductive tract
7. Ovum movement in the fallopian tube
8. Blood circulation in small blood vessels

Historically, however, the engineering analysis of peristalsis was initiated much later than physiological studies. A lucid summary of developments was presented by the modern “father of biomechanics” Fung in the early 1970s [45]. Latham [46] initiated modern fluid mechanics simulations of peristalsis using both theoretical and experimental techniques. Significant work was also reported by Shapiro and others [47] who delineated different zones for pumping. Fung and Yih [48]

presented a model on peristaltic pumping using perturbation techniques. Burns and Parkes [49] studied the flow of Newtonian fluid through a channel and a tube by considering sinusoidal vibrations in the walls along the length. Barton and Raynor [50] studied the peristaltic motion in a circular tube by using long wavelength approximation for intestinal flow. Chaw [51] reported the solution for axisymmetric flow with initially nonstationary flow. Jaffrin [52] studied the effects of inertia and curvature on peristaltic pumping. Applications of peristalsis in industrial fluid mechanics usually involve peristaltic pumping of extremely hazardous liquids such as aggressive chemicals, high solids slurries, noxious fluid (nuclear industries), and other materials. Roller pumps, hose pumps, tube pumps, finger pumps, heart–lung machines, blood pump machines and dialysis machines are all engineered on the basis of peristalsis. In such applications and also medical flows, transport fluids are generally non-Newtonian. In recent years, therefore, researchers have developed new mathematical models utilizing a variety of viscoelastic rheological and microstructural models to simulate physiological fluids more accurately, demonstrating better correlation with clinical data than classical Newtonian viscous flow models, as examined in [45–52]. Relevant studies in this regard include Bohme and Fridrich [53] who employed a Walters-B model. Tsiklauri and Beresnew [54] used a Maxwell viscoelastic model. Tripathi [55] used Stokes' couple-stress models, and also employed fractional Maxwell models [56] and generalized Oldroyd-B viscoelastic models [57] to study peristaltic propulsion under various body forces. Hayat and others [58] have employed a Johnson–Segalman model to simulate elastic effects in peristaltic rheological flow. Tripathi and others [59] have also employed a Jeffery's elasto-viscous model to investigate gastric flow and heat transfer in swallowing. Bég and others [60] have used the electrically conducting Williamson non-Newtonian model and also a differential transform algorithm to study peristaltic flow in a tube. Bhargava and others [61] have used a finite element method to study peristaltic waves in micropolar flow in a deformable conduit. Peristaltic nano-fluid dynamics has also recently received some attention. We shall review works in this area in due course.

4.1.3 Nano-Fluids

Nano-fluids [34] are fluids containing nanoparticles (nanometer-sized particles of metals, oxides, carbides, nitrides or nanotubes). Nano-fluids exhibit enhanced thermal properties, notably higher thermal conductivity and convective heat transfer coefficients compared to the base fluid. Nano-fluids are therefore a new class of fluids designed by dispersing nanometer-sized materials (nanoparticles, nanofibers, nanotubes, nanowires, nano-rods, nano-sheet, or droplets) in base fluids. They may also be regarded as nanoscale colloidal suspensions containing condensed nanomaterials. They are *two-phase* systems with one phase (solid phase) in another (liquid phase). Nano-fluids have been found to also exhibit enhanced thermal diffusivity and viscosity compared to those of base fluids like oil or wa-

ter. In many engineering simulations, including computational fluid dynamics (CFD), nano-fluids can be assumed to be *single-phase* fluids. The classical theory of single-phase fluids can be applied, where physical properties of nano-fluids are taken as a function of properties of both constituents and their concentrations. In recent years, a number of mathematical models have been proposed for nano-fluids. These have largely focused on the mechanism for thermal conductivity enhancement. A popular model is the Tiwari–Das [62] formulation, which has the advantage of not requiring a separate species diffusion equation for the nanoparticle volume fraction. This approach has been successfully utilized in a number of recent studies including Rashidi and others [63] for axisymmetric boundary layer flow from a cylinder with the homotopy analysis method and Bég and others [64] for transport in porous media. Rana and others [65] also very recently employed the Tiwari–Das model to simulate nano-fluid convection from an inclined cylindrical solar collector. Other models have also been developed aimed at further elucidating the properties of nano-fluids. Pre-eminent among these has been the Buongiorno model [66] in which multiple mechanisms are identified for the convective transport in nano-fluids using a *two-phase nonhomogenous* approach. In his two-component four-equation nonhomogeneous equilibrium model for mass, momentum, and heat transport in nano-fluids, he emphasized the following mechanisms: *inertia, Brownian diffusion, thermophoresis, diffusiophoresis, the Magnus effect, fluid drainage* and *gravity*. Of all of these mechanisms, however, only Brownian diffusion and thermophoresis were found to be important in the absence of turbulence effects. It was also suggested that the boundary layer has different properties owing to the effect of temperature and thermophoresis. Taking Brownian motion and thermophoresis into account, Buongiorno [66] developed a correlation for the Nusselt number which was compared to the data from Xuan and Li [67] and Pak and Cho [68] and which correlated best with the latter [68] experimental data. The literature on the thermal conductivity and viscosity of nano-fluids has been reviewed by Eastman and others [69], Wang and Mujumdar [70] and Trisaksri and Wongwises [71]. In addition, a succinct review on applications and challenges of nano-fluids has also been provided by Wen and others [72] and Saidur and others [73]. Recently, the Buongiorno [66] model has been used by Kuznetsov and Nield [74] to study the natural convection flow of nano-fluid over a vertical plate and their similarity analysis identified four parameters governing the transport process. The Kuznetsov–Nield formulation has proved immensely popular in computational thermo-sciences. It has been deployed in many subsequent studies including double-diffusive free convection [75], Rayleigh–Benard nano-fluid instability [76–79], tube nano-fluid flows [80], boundary layers on translating sheets [81], vertical plate convection [82], nano-convection from a sphere in porous media [83], stagnation-point nano-fluid flow in electronic components [84], unsteady radioactive hydromagnetic nano-fluid materials processing [85], nano-fluid flows in geothermal systems [86] and nano-fluid oxytactic bio-convection in hybrid microbial fuel cells [87]. These studies have all considered the nano-fluid to be Newtonian. However, recently progress has also been made in non-Newtonian nano-fluid convection including simulations with an Ostwald–de Waele power law

model [88]. The present discourse is restricted, however, to *Newtonian* nano-fluid dynamics, i.e. rheological features are discarded.

4.2 Mathematical Modelling

We now consider the peristaltic flow of a nano-fluid with double-diffusive convection in a deformable channel. A number of studies have appeared in the past few years on nano-fluid peristaltic fluid mechanics including endoscopic effects [89]. These works have generally employed the Kuznetsov–Nield model although the boundary conditions employed are debatable. Akbar and others [90] developed closed-form solutions for stream function and pressure gradient for the peristaltic flow of a nano-fluid in an asymmetric channel with wall slip effects, under long wavelength and small Reynolds number assumptions. Further studies include Akbar and others [91] who used the homotopy perturbation method to compute temperature and nanoparticle concentration for the effects of Brownian motion number, thermophoresis, local thermal Grashof number, and local nanoparticle Grashof number for five different peristaltic waves. They observed that pressure rise is reduced with increasing thermophoresis number whereas an increase in the Brownian motion parameter and the thermophoresis parameter enhances temperatures. Further studies include Mustafa and others [92] who considered viscous heating, Akbar and Nadeem [93] who used the Phan-Thien-Tanner rheological model for Jeffrey–Hamel nano-fluid peristaltic flow and Mustafa and others [94] who considered wall slip in nano-fluid peristaltic transport. In many drug-delivery applications [95] double-diffusive convection is significant. Thermal diffusion is the transport of the components of gaseous mixtures or solutions when subjected to a temperature gradient. If the temperature difference is held constant, thermal diffusion in a mixture will produce a concentration gradient. The production of such a gradient causes classical *species* diffusion. An excellent treatment of double-diffusion phenomena is provided by Gebhart and others [96]. These effects are also sometimes known as *cross-diffusion effects* or Soret–Dufour effects [97–99]. Very few investigations have been conducted on peristaltic pumping of nano-fluids with double-diffusive (thermal and concentration) convection in nano-fluids. Therefore, this chapter aims to examine the peristaltic flow of nano-fluids with Soret–Dufour (double-diffusive) convection through a two-dimensional deformable channel. The analysis is performed under the well-established long wavelength and low Reynolds number approximations. A detailed mathematical formulation is presented and numerical computations reported. Mathematica software is employed to achieve visualization of the stream lines and trapping phenomenon. The influence of Brownian motion parameter, thermophoresis parameter, thermal Grashof number, concentration Grashof number, nanoparticle Grashof number, Soret parameter, Dufour parameter and peristaltic wave amplitude on nanoparticle fraction, temperature, pressure gradient, velocity and trapping are depicted.

4.2.1 Peristaltic Flow Geometry

The constitutive equation for the peristaltic wall geometry due to propagation of a train of waves, considered in the present investigation, takes the form:

$$\tilde{h}(\tilde{\xi},\tilde{t}) = a + b\sin\frac{2\pi}{\lambda}(\tilde{\xi} - c\tilde{t}) \tag{4.1}$$

Here $\tilde{h}, \tilde{\xi}, \tilde{t}, a, b, \lambda,$ and c represent transverse vibration of the wall, axial coordinate, time, half width of the channel, amplitude of the wave, wavelength, and wave velocity, respectively. The values of temperature (T), solute concentration (C) and nanoparticle fraction (F) at the centreline $\eta = 0$ and the wall of the channel $\eta = h$ are taken as T_0, C_0, F_0 and $T_1, C_1, F_1,$ respectively.

4.2.2 Governing Equations

Employing the Oberbeck–Boussinesq approximation, the governing equations for conservation of mass, momentum, thermal energy, solute concentration, and nanoparticle fraction [75] may be formulated thus:

$$\frac{\partial\tilde{u}}{\partial\tilde{\xi}} + \frac{\partial\tilde{v}}{\partial\tilde{\eta}} = 0, \tag{4.2}$$

$$\rho_{\mathrm{f}}\left(\frac{\partial}{\partial\tilde{t}} + \tilde{u}\frac{\partial}{\partial\tilde{\xi}} + \tilde{v}\frac{\partial}{\partial\tilde{\eta}}\right)\tilde{u} = -\frac{\partial\tilde{p}}{\partial\tilde{\xi}} + \mu\left(\frac{\partial^2\tilde{u}}{\partial\tilde{\xi}^2} + \frac{\partial^2\tilde{u}}{\partial\tilde{\eta}^2}\right)$$
$$+\left[g(1-F_0)\rho_{\mathrm{f0}}\left\{\beta_T(T-T_0) + \beta_C(C-C_0)\right\} - (\rho_{\mathrm{p}} - \rho_{\mathrm{f0}})(F-F_0)\right], \tag{4.3}$$

$$\rho_{\mathrm{f}}\left(\frac{\partial}{\partial\tilde{t}} + \tilde{u}\frac{\partial}{\partial\tilde{\xi}} + \tilde{v}\frac{\partial}{\partial\tilde{\eta}}\right)\tilde{v} = -\frac{\partial\tilde{p}}{\partial\tilde{\eta}} + \mu\left(\frac{\partial^2\tilde{v}}{\partial\tilde{\xi}^2} + \frac{\partial^2\tilde{v}}{\partial\tilde{\eta}^2}\right)$$
$$+ g\left[\left(1-F_0\right)\rho_{\mathrm{f0}}\left\{\beta_T\left(T-T_0\right) + \beta_C\left(C-C_0\right)\right\} - \left(\rho_{\mathrm{p}} - \rho_{\mathrm{f0}}\right)\left(F-F_0\right)\right], \tag{4.4}$$

$$(\rho c)_{\mathrm{f}}\left(\frac{\partial}{\partial\tilde{t}} + \tilde{u}\frac{\partial}{\partial\tilde{\xi}} + \tilde{v}\frac{\partial}{\partial\tilde{\eta}}\right)T = k\left(\frac{\partial^2 T}{\partial\tilde{\xi}^2} + \frac{\partial^2 T}{\partial\tilde{\eta}^2}\right) + D_{TC}\left(\frac{\partial^2 C}{\partial\tilde{\xi}^2} + \frac{\partial^2 C}{\partial\tilde{\eta}^2}\right)$$
$$+(\rho c)_{\mathrm{p}}\left\{D_B\left(\frac{\partial F}{\partial\tilde{\xi}}\frac{\partial T}{\partial\tilde{\xi}} + \frac{\partial F}{\partial\tilde{\eta}}\frac{\partial T}{\partial\tilde{\eta}}\right) + \frac{D_T}{T_0}\left(\left(\frac{\partial T}{\partial\tilde{\xi}}\right)^2 + \left(\frac{\partial T}{\partial\tilde{\eta}}\right)^2\right)\right\} \tag{4.5}$$

$$\left(\frac{\partial}{\partial \tilde{t}}+\tilde{u}\frac{\partial}{\partial \tilde{\xi}}+\tilde{v}\frac{\partial}{\partial \tilde{\eta}}\right)C = D_S\left(\frac{\partial^2 C}{\partial \tilde{\xi}^2}+\frac{\partial^2 C}{\partial \tilde{\eta}^2}\right)+D_{CT}\left(\frac{\partial^2 T}{\partial \tilde{\xi}^2}+\frac{\partial^2 T}{\partial \tilde{\eta}^2}\right), \tag{4.6}$$

$$\left(\frac{\partial}{\partial \tilde{t}}+\tilde{u}\frac{\partial}{\partial \tilde{\xi}}+\tilde{v}\frac{\partial}{\partial \tilde{\eta}}\right)F = D_B\left(\frac{\partial^2 F}{\partial \tilde{\xi}^2}+\frac{\partial^2 F}{\partial \tilde{\eta}^2}\right)+\frac{D_T}{T_0}\left(\frac{\partial^2 T}{\partial \tilde{\xi}^2}+\frac{\partial^2 T}{\partial \tilde{\eta}^2}\right), \tag{4.7}$$

where $\rho_f, \rho_p, \tilde{u}, \tilde{v}, \tilde{\eta}, \tilde{p}, \mu, g, \beta_T, \beta_C, (\rho c)_f, (\rho c)_p, k, T, F, C, D_B, D_T, D_S, D_{TC}$ and D_{CT} denote the fluid density, nanoparticle mass density, axial velocity, transverse velocity, transverse coordinate, pressure, fluid viscosity, acceleration due to gravity, volumetric thermal expansion coefficient of the fluid, volumetric solute expansion coefficient of the fluid, heat capacity of fluid, effective heat capacity of nanoparticle, thermal conductivity, temperature, nanoparticle volume fraction, solute concentration, Brownian diffusion coefficient, thermophoretic diffusion coefficient, solute diffusivity of the porous medium, Dufour diffusivity and Soret diffusivity.

4.2.3 Non-Dimensionalization and Boundary Conditions

To facilitate analytical solutions, we introduce the following non-dimensional parameters:

$$\left.\begin{aligned}
&\xi=\frac{\tilde{\xi}}{\lambda},\ \eta=\frac{\tilde{\eta}}{a},\ t=\frac{c\tilde{t}}{\lambda},\ u=\frac{\tilde{u}}{c},\ v=\frac{\tilde{v}}{c\delta},\ p=\frac{\tilde{p}a^2}{\mu c\lambda},\ h=\frac{\tilde{h}}{a}=1+\phi\sin(2\pi\xi),\\
&\delta=\frac{a}{\lambda},\ \phi=\frac{b}{a},\ \nu=\frac{\mu}{\rho_{\mathrm{f0}}},\ \theta=\frac{T-T_0}{T_1-T_0},\ \gamma=\frac{C-C_0}{C_1-C_0},\ \Phi=\frac{F-F_0}{F_1-F_0},\ \mathrm{Re}=\frac{\rho_{\mathrm{f}}ca}{\mu},\\
&P_r=\frac{\nu(\rho c)_{\mathrm{f}}}{k},\ Gr_{\mathrm{T}}=\frac{\beta_T g a^3\left(1-F_0\right)\left(T_1-T_0\right)}{\nu^2},\ Gr_{\mathrm{C}}=\frac{\beta_C g a^3\left(1-F_0\right)\left(C_1-C_0\right)}{\nu^2},\\
&Gr_{\mathrm{F}}=\frac{g a^3\left(\rho_{\mathrm{p}}-\rho_{\mathrm{f0}}\right)\left(F_1-F_0\right)}{\rho_{\mathrm{f0}}\nu^2},\ N_b=\frac{(\rho c)_{\mathrm{p}}D_B\left(F_1-F_0\right)}{k},\\
&N_t=\frac{(\rho c)_{\mathrm{p}}D_T\left(T_1-T_0\right)}{kT_0},\ N_{TC}=\frac{(\rho c)_{\mathrm{f}}D_{TC}\left(C_1-C_0\right)}{k\left(T_1-T_0\right)},\ N_{CT}=\frac{D_{CT}\left(T_1-T_0\right)}{D_S\left(C_1-C_0\right)},
\end{aligned}\right\} \tag{4.8}$$

where $\delta, \phi, \nu, \theta, \gamma, \Phi, \mathrm{Re}, Gr_{\mathrm{T}}, Gr_{\mathrm{C}}, Gr_{\mathrm{F}}, P_r, N_b, N_t, N_{TC}$, *and* N_{CT} are wave number, amplitude ratio, kinematic viscosity, dimensionless temperature, dimensionless solutal (species) concentration, rescaled nanoparticle volume fraction, Reynolds number, thermal Grashof number, solutal Grashof number, nanoparticle Grashof

number, Prandtl number, Brownian motion parameter, thermophoresis parameter, Dufour parameter and Soret parameter, respectively. For low Reynolds number ($Re \to 0$) and long wavelength $\lambda >> a$, Eqs. 4.2–4.7 reduce to:

$$\frac{\partial u}{\partial \xi} + \frac{\partial v}{\partial \eta} = 0, \tag{4.9}$$

$$\frac{\partial p}{\partial \xi} = \frac{\partial^2 u}{\partial \eta^2} + Gr_T \theta + Gr_C \gamma - Gr_F \Phi, \tag{4.10}$$

$$\frac{\partial p}{\partial \eta} = 0, \tag{4.11}$$

$$\frac{\partial^2 \theta}{\partial \eta^2} + N_{TC} \frac{\partial^2 \gamma}{\partial \eta^2} + N_b \frac{\partial \theta}{\partial \eta} \frac{\partial \Phi}{\partial \eta} + N_t \left(\frac{\partial \theta}{\partial \eta} \right)^2 = 0, \tag{4.12}$$

$$\frac{\partial^2 \gamma}{\partial \eta^2} + N_{CT} \frac{\partial^2 \theta}{\partial \eta^2} = 0. \tag{4.13}$$

$$\frac{\partial^2 \Phi}{\partial \eta^2} + \frac{N_t}{N_b} \frac{\partial^2 \theta}{\partial \eta^2} = 0. \tag{4.14}$$

The following boundary conditions are prescribed:

$$\Phi\big|_{\eta=0} = 0,\ \Phi\big|_{\eta=h} = 1, \gamma\big|_{\eta=0} = 0, \gamma\big|_{\eta=h} = 1, \theta\big|_{\eta=0} = 0, \theta\big|_{\eta=h} = 1, \frac{\partial u}{\partial \eta}\bigg|_{\eta=0} = 0, u\big|_{\eta=h} = 0. \tag{4.15}$$

4.2.4 Analytical Solutions

Integrating Eq. 4.13 twice, with respect to η and using the first and second boundary conditions of Eq. 4.15, the *nanoparticle fraction field* is obtained as follows:

$$\Phi = \frac{N_t}{N_b} \left(\frac{e^{-n\eta} - 1}{1 - e^{-nh}} \right) + \left(\frac{N_b + N_t}{N_b} \right) \frac{\eta}{h}. \tag{4.16}$$

Double integrating Eq. 4.14 with respect to η and using the third and fourth boundary conditions of Eq. 4.15, the *solute concentration field* is obtained as:

$$\gamma = N_{CT}\left(\frac{e^{-n\eta}-1}{1-e^{-nh}}\right)+(1+N_{CT})\frac{\eta}{h}. \tag{4.17}$$

Using Eqs. 4.16 and 4.17 in Eq. 4.12 and integrating it twice with respect to η and using the fifth and sixth boundary conditions of Eq. 4.15, the *temperature field* is obtained as follows:

$$\theta = \left(\frac{e^{-n\eta}-1}{e^{-nh}-1}\right). \tag{4.18}$$

Using Eqs. 4.16–4.18 in Eq. 4.10 and integrating it with respect to η and using the seventh boundary condition of Eq. 4.15, the *axial velocity gradient* is obtained as follows:

$$\frac{\partial u}{\partial \eta} = \frac{\partial p}{\partial \xi}\eta - A\left\{\frac{e^{-n\eta}}{n}+\eta-\frac{1}{n}\right\}-\frac{B}{2}\eta^2, \tag{4.19}$$

where

$$A = \frac{1}{1-e^{-nh}}\left(Gr_{\mathrm{T}} - N_{CT}Gr_C + \frac{N_t}{N_b}Gr_F\right), \tag{4.20}$$

$$B = \frac{1}{h}\left((1+N_{CT})Gr_C - \frac{N_b+N_t}{N_b}Gr_F\right), \tag{4.21}$$

$$n = \frac{N_b+N_t}{(1-N_{TC}N_{CT})h}. \tag{4.22}$$

Integrating Eq. 4.19, and deploying the eighth boundary condition of Eq. 4.15, the *axial velocity* is then obtained as:

$$u = \frac{1}{2}\frac{\partial p}{\partial \xi}(\eta^2-h^2) - A\left\{\frac{1}{2}(\eta^2-h^2)-\frac{1}{n^2}(e^{-n\eta}-e^{-nh})-\frac{1}{n}(\eta-h)\right\}-\frac{B}{6}(\eta^3-h^3), \tag{4.23}$$

4.2.5 Volumetric Flow Rate

The volumetric flow rate is given by integrating across the channel width:

$$Q = \int_0^h u d\eta \tag{4.24}$$

Using Eq. 4.23 in Eq. 4.24 and solving the integral yields:

$$Q = -\frac{h^3}{3}\frac{\partial p}{\partial \xi} - A\left\{\frac{1}{n^3}(e^{-nh}(1+nh)-1) + \frac{h^2}{2n} - \frac{h^3}{3}\right\} + \frac{B}{8}h^4. \tag{4.25}$$

Manipulating Eq. 4.25, the *pressure gradient* is obtained as:

$$\frac{\partial p}{\partial \xi} = -\frac{3Q}{h^3} - A\left\{\frac{3}{n^3h^3}(e^{-nh}(1+nh)-1) + \frac{3}{2nh} - 1\right\} + \frac{3B}{8}h. \tag{4.26}$$

The transformations between a wave frame ($\tilde{X}, \tilde{Y}$) moving with velocity c and the fixed frame ($\tilde{\xi}, \tilde{\eta}$) are now introduced as follows:

$$\tilde{X} = \tilde{\xi} - c\tilde{t}, \tilde{Y} = \tilde{\eta}, \tilde{U} = \tilde{u} - c, \tilde{V} = \tilde{v}, \tag{4.27}$$

where $(\tilde{U}, \tilde{V})$ and $(\tilde{u}, \tilde{v})$ are the velocity components in the wave and fixed frame, respectively.

The volumetric flow rate in the fixed frame is given by:

$$Q = \int_0^h (U+1)dY, \tag{4.28}$$

which, on integration, yields:

$$Q = q + h. \tag{4.29}$$

Averaging volumetric flow rate along one time period, we have:

$$\bar{Q} = \int_0^1 Q dt = \int_0^1 (q+h)dt. \tag{4.30}$$

Equation 4.30 yields:

$$\bar{Q} = q + 1 = Q + 1 - h. \tag{4.31}$$

From Eq. 4.26 and Eq. 4.31, the *pressure gradient* is expressed in term of averaged flow rate as:

$$\frac{\partial p}{\partial \xi} = -\frac{3(\bar{Q}-1+h)}{h^3} - A\left\{\frac{3}{n^3h^3}(e^{-nh}(1+nh)-1) + \frac{3}{2nh} - 1\right\} + \frac{3B}{8}h, \tag{4.32}$$

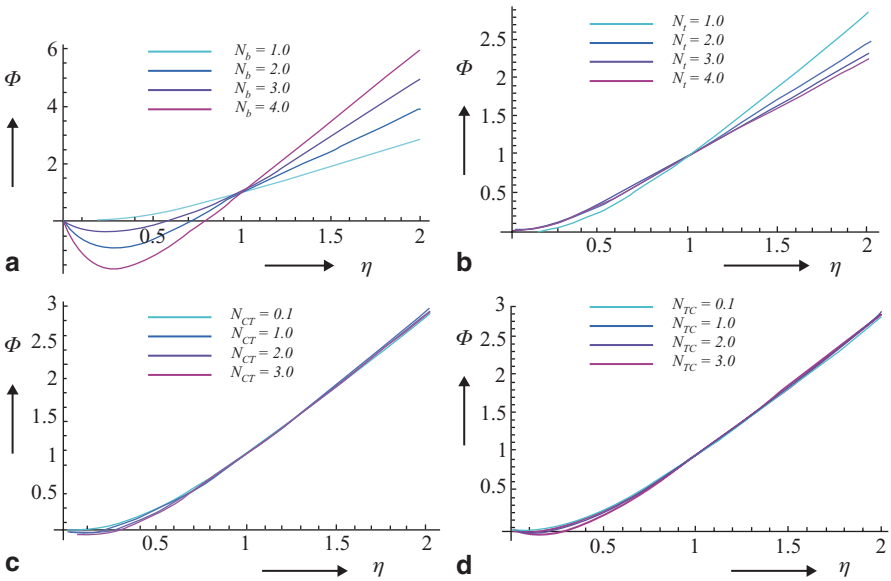

Fig. 4.1 Nanoparticle fraction profiles $\Phi(\eta)$ at ϕ=0.5, x=1.0 for: **a** N_t =1.0, N_{TC}=0.1, N_{CT}=0.1, N_b=1.0, 2.0, 3.0, 4.0. **b** N_b=1.0, N_{TC}=0.1 N_{CT}=0.1, N_t=1.0, 2.0, 3.0, 4.0. **c** N_t=1.0, N_b=1.0, N_{TC}=0.1, N_{CT}=0.1, 1.0, 2.0, 3.0. **d** N_t=1.0, N_b=1.0, N_{CT}=0.1, N_{Tc}=0.1, 1.0, 2.0, 3.0

The *pressure difference across one wavelength* (Δp) is computed using:

$$\Delta p = \int_0^1 \frac{\partial p}{\partial \xi} d\xi, \tag{4.33}$$

Using Eq. 4.23 and the transformations of Eq. 4.27, the *stream function* in the wave frame (Cauchy–Riemann equations, that is, $U = \frac{\partial \psi}{\partial \eta}$ and $V = -\frac{\partial \psi}{\partial \xi}$) is obtained as follows:

$$\psi(\xi,\eta) = \frac{1}{2}\frac{\partial p}{\partial \xi}\left(\frac{\eta^3}{3} - h^2\eta\right) - A\left\{\frac{1}{n^2}\left(\frac{e^{-n\eta}}{n} + e^{-nh}\eta + \frac{1}{n}\right) + \frac{1}{2}\left(\frac{\eta^3}{3} - h^2\eta\right) - \frac{1}{n}\left(\frac{\eta^2}{2} - h\eta\right)\right\}$$
$$- \frac{B}{6}\left(\frac{\eta^4}{4} - h^3\eta\right) - \eta \tag{4.34}$$

4.3 Numerical Results and Discussion

Numerical and computational results of the mathematical model are discussed in this section. Mathematica is used to integrate the solutions due to the complicated definite integrals and plot (Figs. 4.1–4.7). The influences of the thermo-physical

parameters characterizing double-diffusive convection in nano-fluids on the peristaltic flow patterns are also depicted.

The effects of Brownian motion parameter (N_b), thermophoresis parameter (N_t), Soret parameter (N_{CT}) and Dufour parameter (N_{CT}) on nanoparticle fraction profile ($\Phi(\eta)$) are presented through the Figs. 4.1a–d. N_b arises in both the dimensionless temperature and nanoparticle fraction conservation equations, i.e. Eqs. 4.12 and 4.14 in the mixed derivative term, $N_b(\partial\theta/\partial\eta)(\partial\Phi/\partial\eta)$ in the former, and the second order temperature derivative, $(N_t / N_b)(\partial^2\theta/\partial\eta^2)$, in the latter. N_b is a key parameter dictating the diffusion of nanoparticles. With an increase in Brownian motion parameter (N_b), there is a strong reduction initially in nanoparticle fraction profile, $\Phi(\eta)$. This effect is shown in Fig. 4.1a. The nano-fluid is a *two-phase* fluid in nature and random movement of the suspended nanoparticles *enhances* energy exchange rates in the fluid but initially decreases nanoparticle concentrations in the flow regime. As dimensionless transverse coordinate, η, is increased there is change in the effect of the Brownian motion parameter-nanoparticle fraction (Φ) is distinctly increased with a *divergence* in profiles. Thermophoretic parameter (N_t) effects are depicted in Fig. 4.1b. A slight increase in $\Phi(\eta)$ values is caused as N_t from 1 to 4, for some distance from the channel centre line ($\eta=0$); however, as η is further increased, there is a switch and thermophoresis is found to depress fraction Φ values. As with the Brownian motion parameter, N_t also features in both energy and nanoparticle fraction conservation Eqs. 4.12 and 4.14, respectively. Although it features in the same term in the latter as the N_b parameter, in the former (Eq. 4.12) it appears in a separate term, consistent with the original formulation of Buongiornio [66] and Kuznetsov and Nield [74], viz $N_t(\partial\theta/\partial\eta)^2$. Hence, nanoparticle fraction diffusion is found to be initially assisted by thermophoresis but subsequently opposed by it. This pattern is also consistent with *macroscopic convection* flows (non-nano-fluids). The influence of Soret parameter (N_{CT}) and Dufour parameter (N_{TC}) on nanoparticle fraction (Φ) are provided in Figs. 4.1c and d.

When heat and mass transfer occur simultaneously in a moving fluid, an energy flux can be generated not only by temperature gradients but by composition gradients also. The energy flux caused by a composition gradient is termed the *Dufour or diffusion-thermo effect*. On the other hand, mass fluxes can also be created by temperature gradients and this embodies the *Soret or thermal-diffusion effect*. Such effects are significant when density differences exist in the flow regime. For example, when species are introduced at a surface in a fluid domain, with a different (lower) density than the surrounding fluid, both Soret (thermo-diffusion) and Dufour (diffuso-thermal) effects can become influential. Soret and Dufour effects are important for intermediate molecular weight fluids in coupled heat and mass transfer in fluid binary systems, often encountered in biophysical processes. N_{CT} (Soret number) represents the effect of temperature gradients on mass (species) diffusion. N_{TC} (Dufour number) simulates the effect of concentration gradients on thermal energy flux in the peristaltic flow domain. These parameters arise in the energy and species conservation equations, Eqs. 4.12 and 4.13, in the terms $N_{TC}(\partial^2\gamma/\partial\eta^2)$ *and* $N_{CT}(\partial^2\theta/\partial\eta^2)$, respectively.

However, they do not arise in the nanoparticle volume fraction (Eq. 4.14). As such both parameters will exert a minor role on Φ distributions. Inspection of

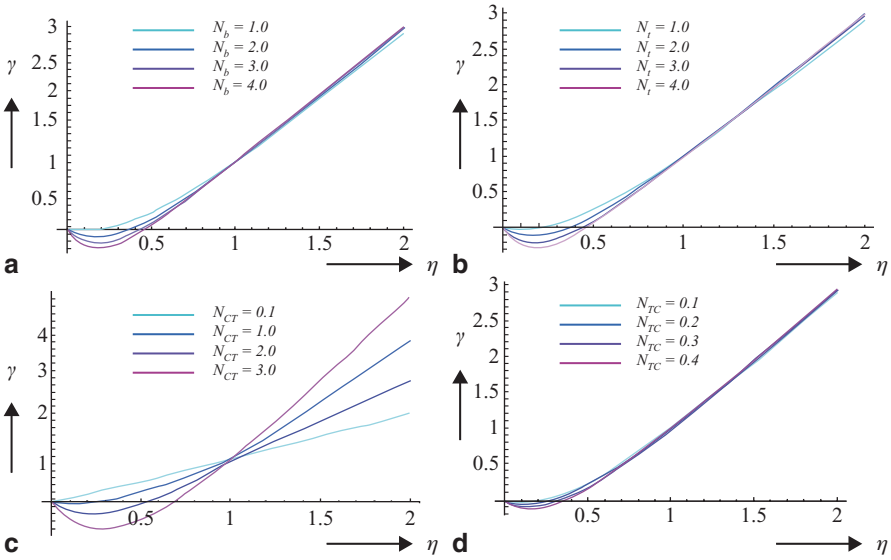

Fig. 4.2 Concentration profiles $\theta(\eta)$ at $\Phi=0.5$. $x=1.0$ for: **a** $N_t=1.0$, $N_{TC}=0.1$, $N_{CT}=1.0$, $N_b=1.0$, 2.0, 3.0, 4.0. **b** $N_b=1.0$, $N_{TC}=0.1$, $N_{CT}=1.0$ $N_t=1.0$, 2.0, 3.0, 4.0. **c** $N_t=1.0$, $N_b=1.0$, $N_{TC}=0.1$, $N_{CT}=0.1$, 1.0, 2.0, 3.0. **d** $N_t=1.0$, $N_b=1.0$, $N_{CT}=0.1$, $N_{TC}=0.1$, 0.2, 0.3, 0.4

Fig. 4.1c shows that a *small decrease* is induced in Φ by a *strong increase* in N_{CT} from 0.1 to 3; subsequently there is, however, a marginal increase in Φ. An almost identical response is sustained by the nanoparticle volume fraction profiles with an increase in the Dufour number (N_{TC}).

Figures 4.2a–d show the concentration profile ($\gamma(\eta)$) for the effects of Brownian motion parameter (N_b), thermophoresis parameter (N_t), Soret parameter (N_{CT}) and Dufour parameter (N_{TC}). With increasing N_b and N_t, species concentration values are significantly reduced. A much more potent response is, however, observed with a change in Soret (N_{CT}) and Dufour (N_{TC}) parameters. Species concentration is found to be very strongly reduced with increasing Soret number (Fig. 4.2c) for some distance from the channel centre; with further distance from the channel centre, as we approach the periphery of the channel, this trend is noticeably reversed and *thermo-diffusion* is observed to accentuate concentration, i.e. *enhance* diffusion of the species. Figure 4.2d also confirms that Dufour number exerts a much weaker effect on species diffusion than the Soret effect, a slight reduction in concentration values is caused, and there is no significant alteration in the effect of Dufour number with transverse distance.

Figures 4.3a–d illustrate the evolution of the temperature profile ($\theta(\eta)$) under the effects of Brownian motion parameter (N_b), thermophoresis parameter (N_t), Soret parameter (N_{CT}), and Dufour parameter (N_{TC}). A very different distribution of profiles is observed compared with the concentration profiles. Increasing Brownian motion parameter initially strongly elevates temperatures in the vicinity of the channel centreline; further away this effect is reversed. It is also apparent that with strong

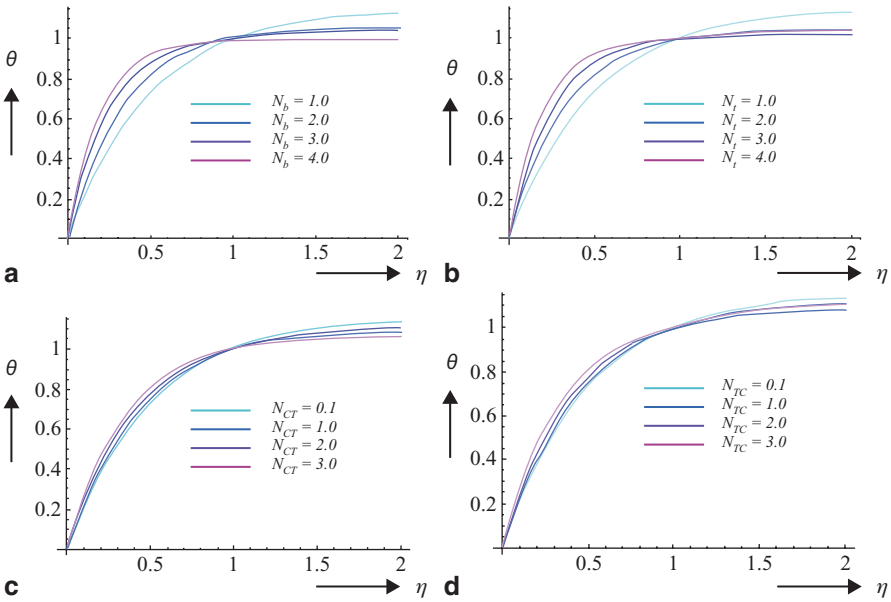

Fig. 4.3 Temperature profiles $\theta(\eta)$ at ϕ=0.5, x=1.0 for: **a** N_t=1.0, N_{TC}=0.1, N_{CT}=0.1, N_b=1.0, 2.0, 3.0, 4.0. **b** N_b=1.0, N_{TC}=0.1, N_{CT}=0.1, N_t=1.0, 2.0, 3.0, 4.0. **c** N_t=1.0, N_b=1.0, N_{TC}=0.1, N_{CT}=0.1, 1.0, 2.0, 3.0. **d** N_t=1.0, N_b=1.0, N_{CT}=0.1= N_{TC}=0.1, 1.0, 2.0, 3.0

Brownian motion (N_b=4.0) effects, temperature profile stabilizes and becomes approximately parallel, i.e. eventually becomes invariant with transverse distance; this effect is clearly visible in Fig. 4.3a. Parabolic trends are only retained at large η values for weak Brownian diffusion (N_b=1.0). A similar evolution in temperature profiles is computed for the influence of thermophoresis parameter, N_t, in Fig. 4.3b. With increasing Soret and Dufour numbers, the profiles for temperature are similar to those in Figs. 4.3a and b; however, they ascend more smoothly. The temperature is found to be enhanced both with Soret and Dufour numbers, initially; with further distance from the channel centre, both effects serve to reduce temperatures.

Figures 4.4a–g illustrate the influence of Brownian motion parameter (N_b), thermophoresis parameter (N_t), Soret number (N_{CT}) Dufour number (N_{TC}) thermal Grashof number (Gr_T), concentration Grashof number (Gr_C), and nanoparticle Grashof number (Gr_F) on the axial velocity profile ($u(\eta)$) across the channel semi-width. Axial velocity, *u,* is generally *negative* for all Brownian motion parameters throughout the channel half-space defined by $0 \leq \eta \leq 1$; flow reversal that is *strong backflow* is therefore taking place. Maximum velocities are always located at the channel centre, decaying smoothly to zero at the periphery (channel wall). Figure 4.4a indicates that an increase in Brownian motion parameter, N_b, decreases magnitudes of the axial velocity, i.e. opposes backflow *u* values, therefore, become *more* positive. The flow is, therefore, actually decelerated with Brownian motion. A substantially different response is computed for the effect of thermophoresis parameter in Fig. 4.4b. At low N_t value (=1.0) negative axial velocity is observed;

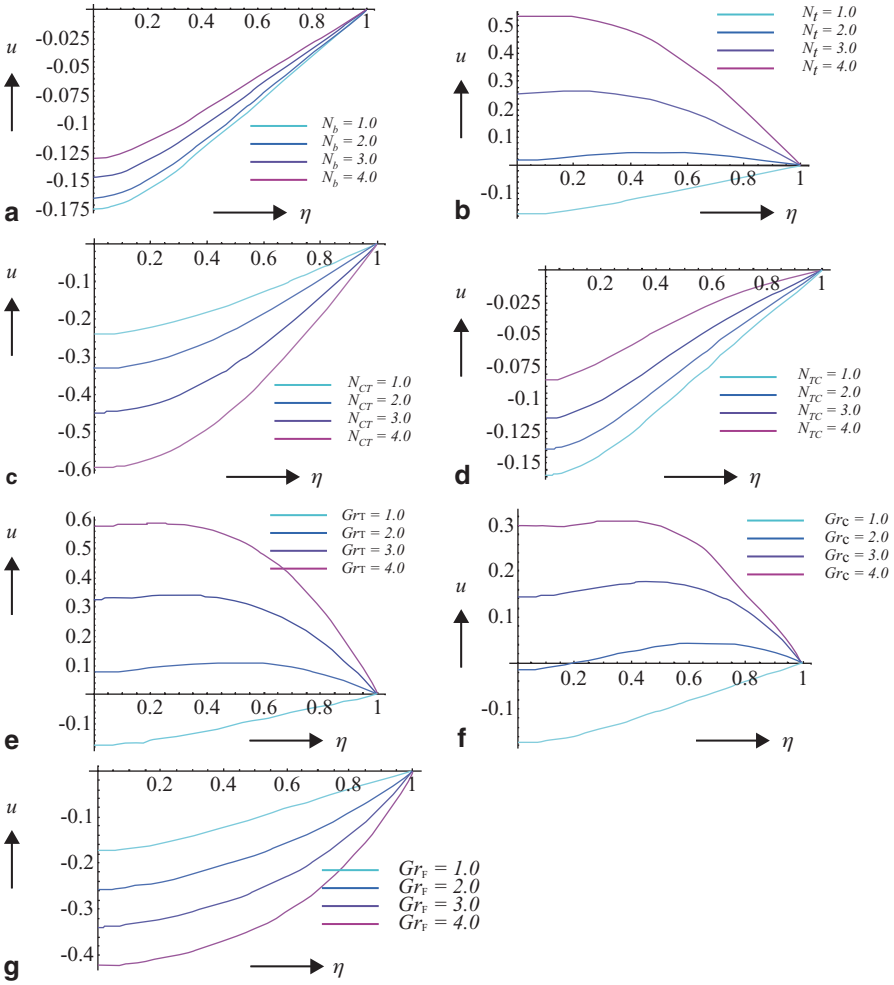

Fig. 4.4 Velocity profiles $u(\eta)$ at ϕ=0.5, ξ=1.0, $\partial p/\partial\xi$=1.0 for: **a** N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_b=1, 2, 3, 4. **b** N_b=1, N_{TC}=0.1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_t=1, 2, 3, 4. **c** N_b=1, N_t=1, N_{TC}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_{CT}=1, 2, 3, 4. **d** N_b=1, N_t=1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_{TC}=1, 2, 3, 4. **e** N_b=1, N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_c=1, Gr_F=1, Gr_T=1, 2, 3, 4. **f** N_b=1, N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_T=1, Gr_F=1, Gr_c=1, 2, 3, 4. **g** N_b=1, N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_r=1, Gr_c=1, Gr_F=1, 2, 3, 4

however, as N_t is increased to 2 and then 3 and the maximum value of 4.0, velocity becomes positive, i.e. *backflow is completely eliminated* across the channel half-space. The profiles also descend for N_t>1, from a maximum at the channel centre to a minimum at the channel wall. The rate of descent is also enhanced with greater thermophoresis parameter. There is an order of magnitude difference also in the values of axial velocity between Figs. 4.4a and b; velocities are much large in Fig. 4.4b. With increasing Soret number, velocities are caused to become more

negative in Fig. 4.4c, i.e. flow deceleration is induced and backflow is accentuated. The contrary response is computed for increasing Dufour number in Fig. 4.4d, where velocities are found to become less negative, i.e. flow reversal is inhibited. The velocity magnitudes in Fig. 4.4c are evidently much greater than those in Fig. 4.4d. Figure 4.4e shows the effect of thermal Grashof number (Gr_T) on axial velocity distribution. This parameter arises in the momentum conservation equation (4.10), in the term, ($Gr_T\theta$). This parameter signifies the relative influence of thermal buoyancy force and viscous hydrodynamic force. For $Gr_T<1$, the peristaltic regime is dominated by viscous forces and vice versa for $Gr_T>1$. For the intermediate case of $Gr_T=1$ both thermal buoyancy and viscous forces are of the same order of magnitude, as described by Gebhart and others [96]. Velocity magnitudes are generally reduced with increasing thermal Grashof number. At low Gr_T velocities are negative, i.e. back flow exists. However, for $Gr_T>1$, backflow is negated and a strong acceleration induced in the peristaltic axial flow. Figure 4.4f reveals that a similar response is induced by the concentration Grashof number, G_{rC} however, there is still some minor backflow at $Gr_C=2$; with larger concentration Grashof number as with thermal Grashof number, the backflow is completely eliminated and a *strong acceleration* achieved in the axial flow. Gr_C represents the ratio of species buoyancy force to the viscous hydrodynamic force; it is the species diffusion analogy to thermal diffusion Grashof number. For the case where both forces are the same, i.e. $Gr_C=1$, axial velocity magnitude is minimized. The same response is observed for thermal Grashof number. Figure 4.4g shows that increasing the nanoparticle Grashof number (Gr_F) exacerbates the axial velocity *back flow,* i.e. increases negative values.

Figures 4.5a–g present the variation of pressure gradient ($\partial p / \partial \xi$) with axial coordinate (ξ) under the influence of Brownian motion parameter (N_b), thermophoresis parameter (N_t), thermal Grashof number (Gr_T), concentration Grashof number (Gr_C), nanoparticle Grashof number (Gr_F). In all cases we have prescribed the wave amplitude and averaged volumetric flow rate as $\phi=0.5$, $\bar{Q}=0.5$ respectively, which is characteristic of the actual physiological regimes as expounded in benchmark peristaltic studies by Shapiro and others [47]. In all profiles the strong periodic behaviour and fluctuations in pressure gradient caused by peristaltic motion are clearly visible. Effectively, Brownian motion is found to slightly enhance pressure gradient (Fig. 4.5a). A significantly greater accentuation in pressure gradient is generated with increasing thermophoresis parameter (Fig. 4.5b). Conversely, Soret number acts to strongly depress pressure gradient values, whereas there is a slight enhancement in them with increasing Dufour number. Thermal Grashof number is observed to depress pressure gradients, whereas the species (concentration) Grashof number and nanoparticle Grashof number distinctly enhance pressure gradients in the peristaltic flow regime in the channel. In all profiles the respective trends indicated above are consistent across all axial coordinate values.

Figures 4.6a–g display the variation of pressure difference across one wavelength (Δp) with averaged flow rate ($\bar{Q}$) under the respective influences of Brownian motion parameter (N_b), thermophoresis parameter (N_t), thermal Grashof number (Gr_T), concentration Grashof number (Gr_C), nanoparticle Grashof number (Gr_F).

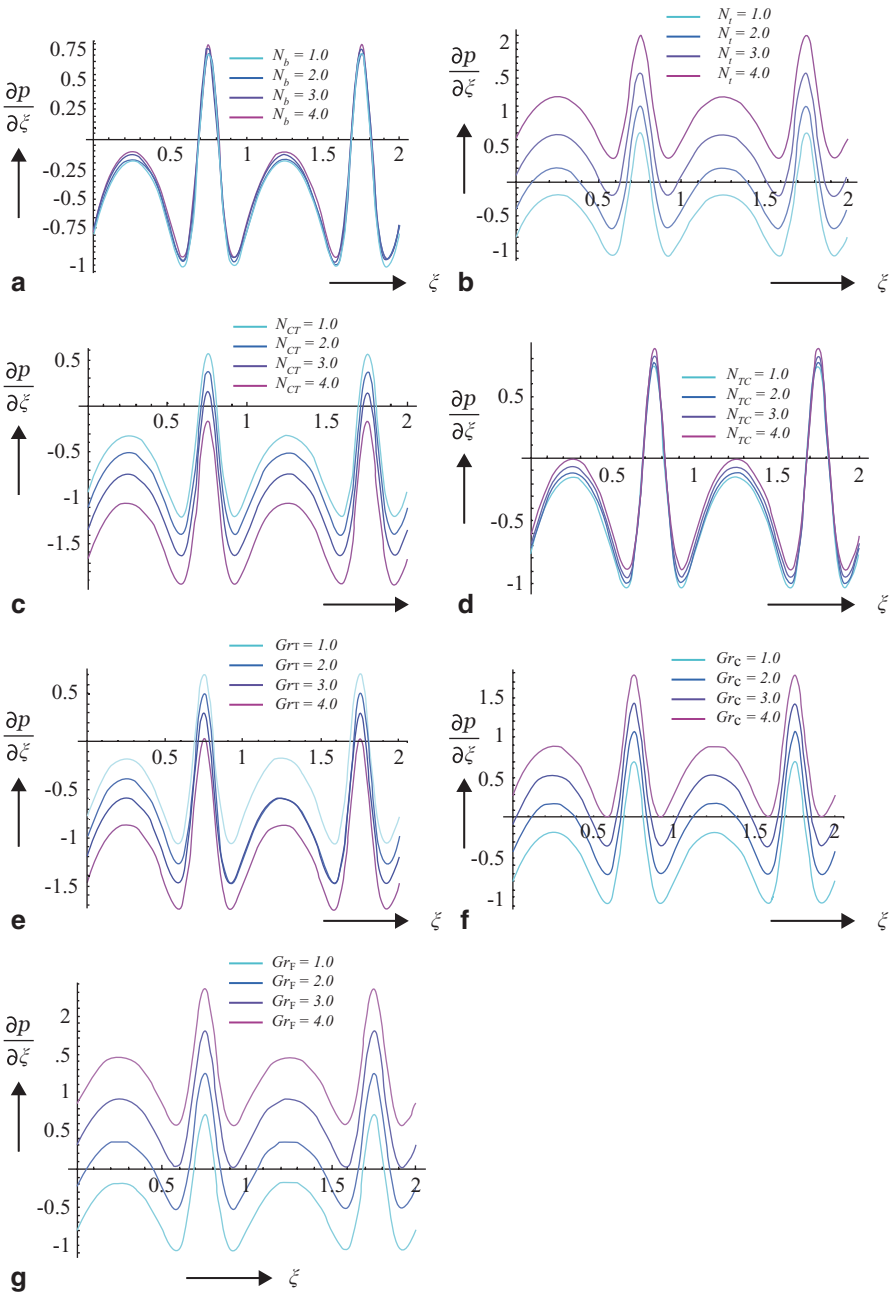

Fig. 4.5 Pressure gradient ($\partial p/\partial \xi$) vs. axial distance (ξ) at ϕ=0.5, $\overline{Q}$ =0.5, **a** N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_b=1, 2, 3, 4. **b** N_b=1, N_{TC}=0.1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_t=1, 2, 3, 4. **c** N_b=1, N_t=1, N_{TC}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_{CT}=1, 2, 3, 4. **d** N_b=1, N_t=1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_{TC}=1, 2, 3, 4. **e** N_b=1, N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_c=1, Gr_F=1, Gr_T=1, 2, 3, 4. **f** N_b=1, N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_T=1, Gr_F=1, Gr_c=1, 2, 3, 4. **g** N_b=1, N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, 2, 3, 4

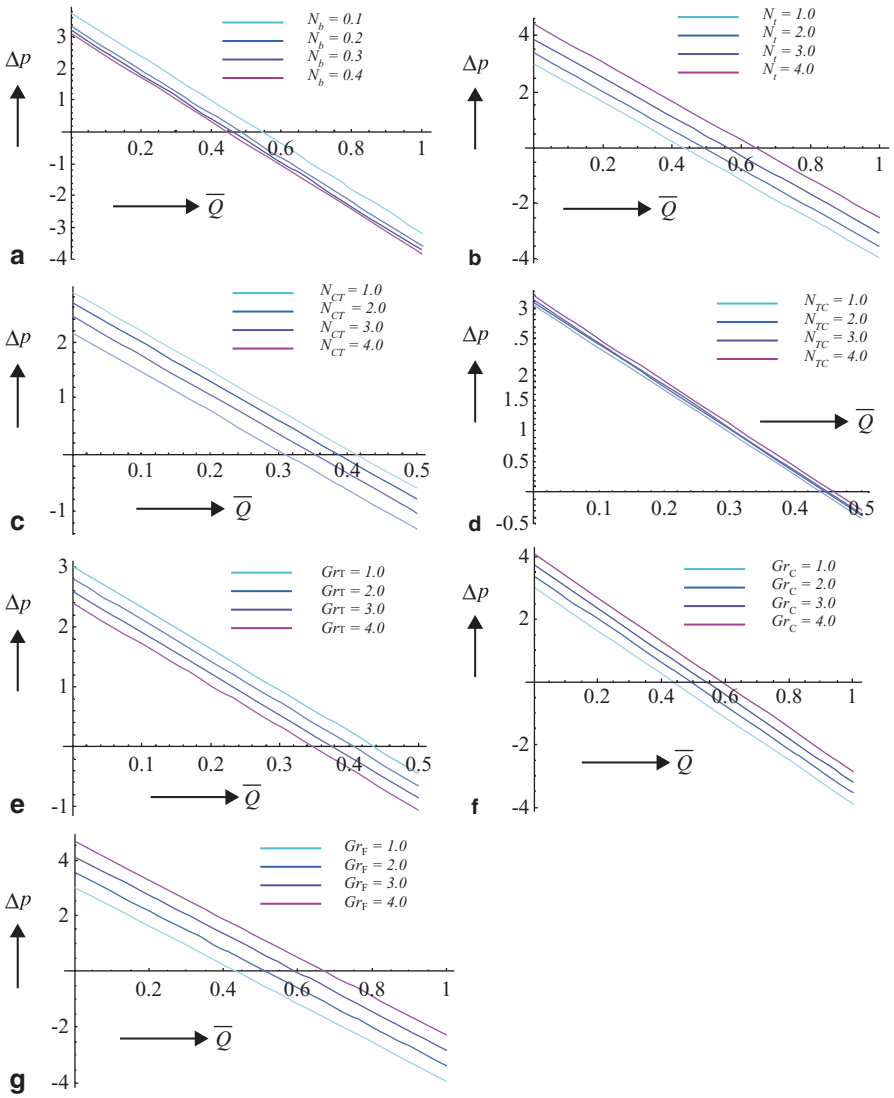

Fig. 4.6 Pressure difference across one wavelength (Δp) vs. averaged volume flow rate ($\overline{Q}$) at ϕ=0.5 for: **a** N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_b=0.1, 0.2, 0.3,0.4. **b** N_b=1, N_{TC}=0.1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_t=1, 2, 3, 4. **c** N_b=1, N_t=1, N_{TC}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_{CT}=1, 2, 3, 4. **d** N_b=1, N_t=1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, N_T=1, 2, 3, 4. **e** N_b=1, N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_c=1, Gr_F=1, Gr_T=1, 2, 3, 4. **f** Gr_T=1, Gr_F=1, N_b=1, N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_c=1, 2, 3, 4. **g** N_b=1, N_t=1, N_{TC}=0.1, N_{CT}=0.1, Gr_T=1, Gr_c=1, Gr_F=1, 2, 3, 4

Invariably linear distributions are observed. Three ranges of pumping are possible namely (a) $\Delta p > 0$, i.e. *pumping* region (b) $\Delta p = 0$ i.e. *free pumping* region, (c) $\Delta p < 0$, i.e. *co-pumping* region, and we have considered all three. Increasing Brownian motion parameter (N_b) reduces pressure difference, an effect which is clearly of

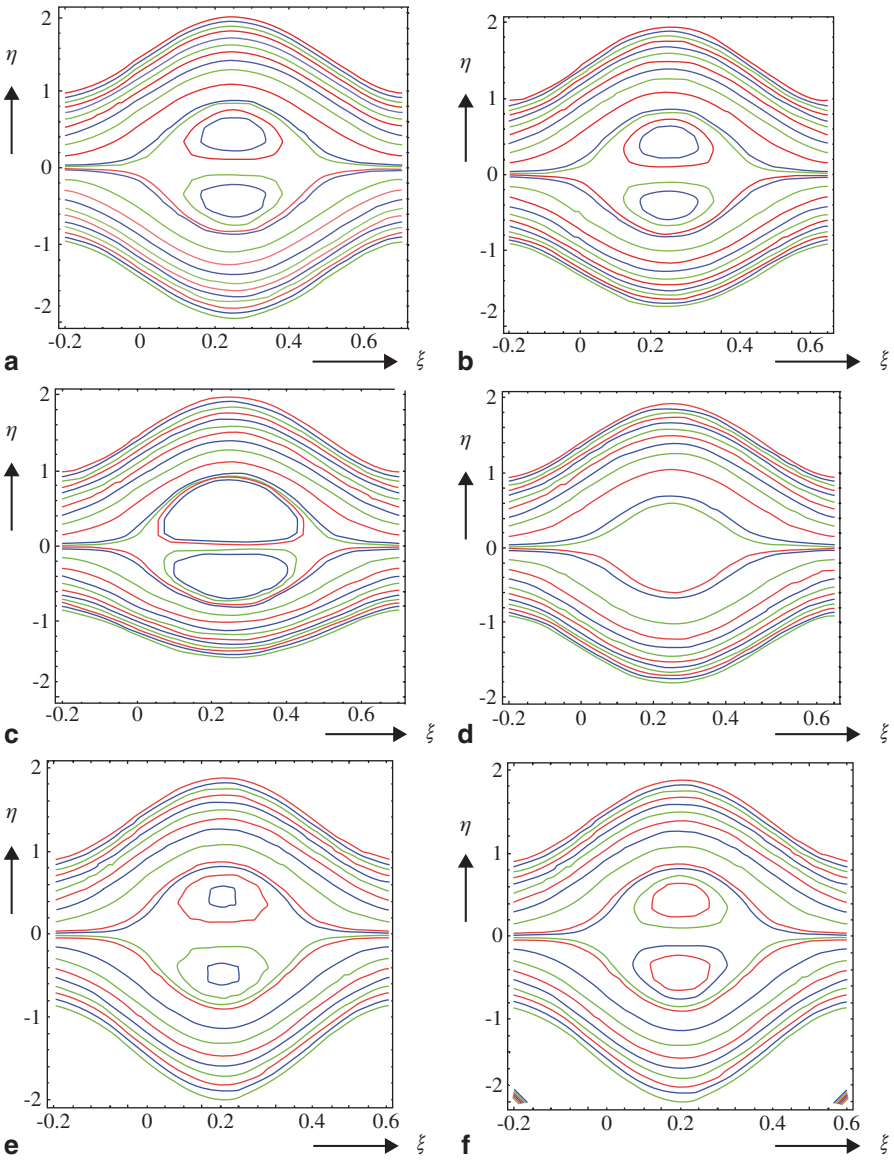

Fig. 4.7 Streamlines in the wave frame at $\phi=0.5$, $\bar{Q}=0.6$ and **a** $N_b=1$, $N_t=0.1$, $N_{TC}=0.1$, $N_{CT}=1$, $Gr_T=0$, $Gr_c=0$, $Gr_F=0$. **b** $N_b=1$, $N_t=0.1$, $N_{TC}=0.1$, $N_{CT}=1$, $Gr_T=0.1$, $Gr_c=1$, $Gr_F=1$. **c** $N_b=1$, $N_t=0.1$, $N_{TC}=0.1$, $N_{CT}=1$ $Gr_T=0.1$, $Gr_c=1$, $Gr_F=3$. **d** $N_b=1$, $N_t=0.1$, $N_{TC}=0.1$, $N_{CT}=1$, $Gr_T=0.1$, $Gr_c=3$, $Gr_F=1$. **e** $N_b=1$, $N_t=0.1$, $N_{TC}=0.1$, $N_{CT}=1$, $Gr_T=0.5$, $Gr_c=1$, $Gr_F=1$. **f** $N_b=1$, $N_t=1$, $N_{TC}=0.1$, $N_{CT}=1$, $Gr_T=0.1$, $Gr_c=1$, $Gr_F=1$. **g** $N_b=1$, $N_t=0.1$, $N_{TC}=0.1$, $N_{CT}=1$, $Gr_T=0.1$, $Gr_c=1$, $Gr_F=1$. **h** $N_b=1$, $N_t=0.1$, $N_{TC}=0.1$, $N_{CT}=3$, $Gr_T=0.1$, $Gr_c=1$, $Gr_F=1$. **i** $N_b=1$, $N_t=0.1$, $N_{TC}=0.5$, $N_{CT}=1$, $Gr_T=0.1$, $Gr_c=1$, $Gr_F=1$

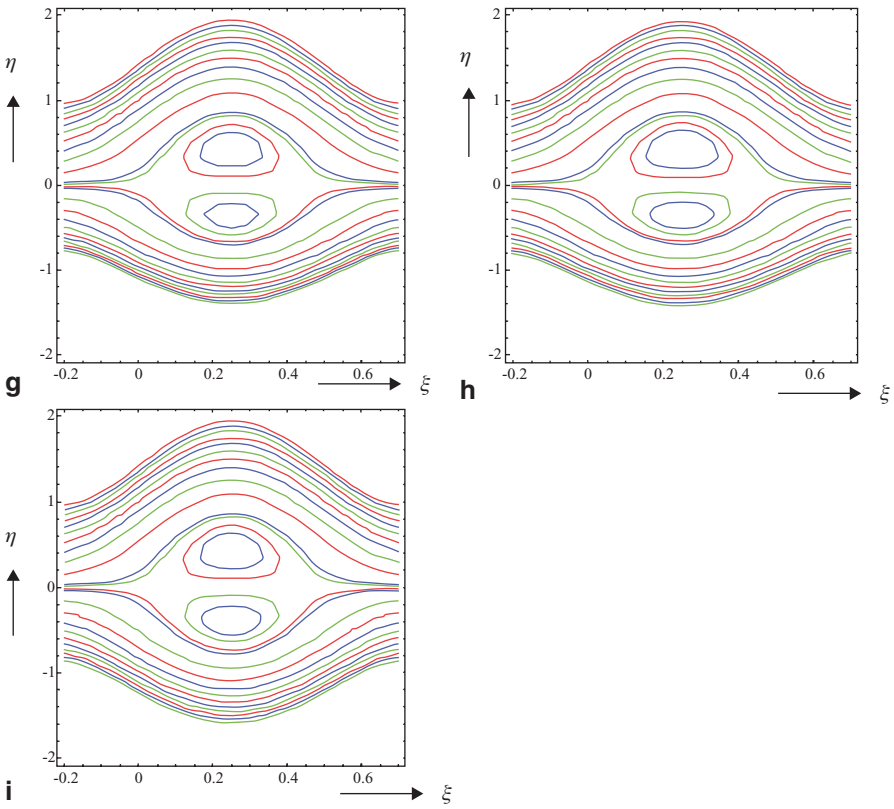

Fig. 4.7 (continued)

significance in *nano-fluid drug-delivery systems*. Conversely, increasing thermophoretic parameter (N_t) strongly increases the pressure difference. In both the cases this pattern is sustained for all values of averaged volume flow rate, $\bar{Q}$. In other words, in *nano-peristaltic* pumps, a pressure difference *drop* or *rise* can be maintained with increasing Brownian diffusion effect or increasing thermophoretic effect at *all operating* flow rates. Figures 4.6c and d show that increasing Soret number strongly decreases the pressure difference, whereas increasing Dufour number acts to slightly increase it. Figures 4.6e–g demonstrate that increasing thermal Grashof number, pressure gradient is curtailed, whereas it is strongly elevated with increasing species (concentration) Grashof number and nanoparticle Grashof number, for all flow rates. In all Figs. 4.6a–g, at higher volumetric flow rates pressure difference becomes negative.

Trapping is an inherent phenomenon of peristaltic motion in which an *internally circulating bolus of fluid* is formed by *closed streamlines* and this trapped bolus is pushed ahead along with the peristaltic wave. The effects of Brownian

motion parameter (N_b), thermophoresis parameter (N_t), thermal Grashof number (Gr_T), concentration Grashof number (Gr_C), nanoparticle Grashof number (Gr_F) on streamlines and trapping phenomenon are therefore also depicted in Figs. 4.7(a–i). Nine streamline distributions are illustrated. A single parameter has been varied for each pair. In all cases, amplitude ratio (ϕ) is fixed at 0.5 and averaged volume flow rate, $\bar{Q}$ constrained at 0.6. The streamlines on the centerline in the wave frame are found to compartmentalize under specific conditions in order to enclose a bolus of fluid particles circulating along closed streamlines. This phenomenon is known as *trapping,* which is a characteristic of peristaltic motion. Since this bolus appear to be trapped by the wave, the bolus moves with the same speed as that of the wave (celerity). Comparison of the appropriate figures shows that magnitude of trapped bolus *clearly reduces* with increasing the magnitude of different Grashof numbers. Brownian motion parameter decreasing acts to reduce the number of trapped boluses. With decreasing thermophoretic parameter, the magnitude of boluses is slightly enhanced. With decreasing thermal Grashof number, the bolus size is amplified. Increasing species Grashof number reduces the multiple bolus structure to a single bolus. Increasing species Grashof number, therefore, exerts a similar effect on streamlines and trapping to decreasing Brownian motion parameter (Figs. 7b and d). Nano-fluid characteristics, therefore, undeniably exert a significant influence on peristaltic flow patterns. Where numbers of boluses are unchanged, the magnitudes are clearly affected by nano-fluid dynamic characteristics. Conversely, there is very little influence detected for a change in Soret and Dufour parameters. Comparing Figs. 4.7b and 4.7h, where the Dufour parameter is increased from 1 to 3, or Figs. 4.7b and 4.7i, where the Soret parameter is increased, the streamline patterns are almost indistinguishable.

4.4 Summary

In this chapter we have briefly reviewed the challenges and potential of mathematical modelling biofluid mechanics. The fundamentals of peristaltic transport and nano-fluid dynamics have also been described qualitatively. A novel mathematical model has additionally been presented to simulate the influence of nano-fluid and thermo-diffusive/diffuso-thermal characteristics on peristaltic heat and mass transfer in a two-dimensional axisymmetric channel, as a simulation of nano-fluid peristaltic drug-delivery systems. The study has been motivated by applications in novel nano-fluid pharmacological delivery. Numerical computations have shown that:

a. Brownian and thermophoresis parameters exert a strong influence on nanoparticle fraction profile, $\Phi(\eta)$ and temperature profile, $\theta(\eta)$.
b. Axial velocity is strongly affected by Soret and Dufour parameters as is the species concentration distribution and temperature evolution through the channel.
c. Pressure difference is increased weakly with Brownian motion, whereas it is very strongly enhanced with increasing thermophoresis parameter.

d. Increasing Soret number considerably reduces the pressure gradient values, whereas increasing Dufour number slightly elevates pressure difference.
e. Thermal Grashof number is observed to depress pressure gradients, whereas the species (concentration) Grashof number and nanoparticle Grashof number markedly elevate pressure gradients in the peristaltic flow regime.
f. Streamline patterns illustrating the trapping of boluses are also found to be more strongly affected with Brownian and thermophoretic parameters and also all Grashof numbers (thermal, species, nanoparticle) than Soret and Dufour effects which exert almost a negligible influence on streamline profiles.

References

1. Mazumdar J (1999) An introduction to mathematical physiology and biology, 2nd edn. Cambridge University Press, UK
2. Shyy W, Lian Y, Tang J, Viieru D, Liu H (2008) Aerodynamics of Low Reynolds Number Flyers. Cambridge aerospace series (No. 22). Cambridge University Press, UK
3. Hou TY, Stredie VG, Wu TY (2007) Mathematical modeling and simulation of aquatic and aerial animal locomotion. J Comput Phys 225:1603–1631
4. Bar-Cohen Y (2006) Biomimetics-using nature to inspire human innovation. Bioinspir Biomim 1:P1–P12
5. Schwarz R (2008) Biological modelling and simulation-a survey of practical models, algorithms and numerical methods. MIT, Cambridge
6. Anderl R, Eigner M, Sendler U, Stark R (2012) Smart engineering … show all 4 hide. Springer, Berlin
7. Skalak R, Ozkaya N, Skalak TC (1989) Biofluid mechanics. Ann Rev Fluid Mech 21:167–200
8. Fung YC (1997) Biomechanics: circulation. Springer, New York
9. Bathe KJ, Zhang H, Ji S (1999) Finite element analysis of fluid flows fully coupled with structural interactions. Computer Structures, 72:1–16
10. Pedley TJ, Hung TK, Skalak R (1981) Fluid mechanics of cardiovascular flow. In: Reul H, Ghista DN, Rau G (eds) Perspectives in biomechanics. Harvard Academic Publishers, Aachen, 1, pp 113–226
11. Hung TK, Tsai TMC (2004) Nonlinear pulsatile flows in rigid and distensible arteries. J Mech Med Biol 4:419–434
12. Pavlov VV (2006) Dolphin skin as a natural anisotropic compliant wall. Bioinspir Biomim 1:31–40
13. Whittlesey RW, Liska S, Dabiri JO (2010) Fish schooling as a basis for vertical axis wind turbine farm design. Bioinspir Biomim 5:035005
14. Von Ellenreider KD, Parker K, Soria J (2008) Fluid mechanics of flapping wings. Exp Therm Fluid Sci 32:1578–1589
15. Bég OA, Parsa AB, Rashidi MM, Sadri SM (2013) Semi-computational simulation of magneto-hemodynamic flow in a semi-porous channel using optimal homotopy and differential transform methods-a model for surgical blood flow control. Comput Biol Med 43(9):1142–1153
16. Gan RZ, Cheng T, Dai C, Yang F, Wood MW (2009) Finite element modeling of sound transmission with perforations of tympanic membrane. J Acoust Soc Am 126:243–253
17. Tripathi D, Pandey SK, Siddiqui A, Bég OA (2012) Non-steady peristaltic propulsion with exponential variable viscosity: a study of transport through the digestive system. Comput Meth Biomech Biomed Eng. doi:10.1080/10255842.2012.703660

18. Tripathi D, Bég OA (2012) Magnetohydrodynamic peristaltic flow of a couple stress fluid through coaxial channels containing a porous medium. J Mech Med Biol 12:1250088
19. Tripathi D, Bég OA, Curiel-Sosa JL (2012) Homotopy semi-numerical simulation of peristaltic flow of generalised oldroyd-b fluids with slip effects. Comput Meth Biomech Biomed Eng. doi:10.1080/10255842.2012.688109
20. Bég TA, Rashidi MM, Bég OA, Rahimzadeh N (2012) Differential transform semi-numerical analysis of biofluid-particle suspension flow and heat transfer in non-darcian porous media. Comput Meth Biomech Biomed Eng doi:10.1080/10255842.2011.643470
21. Ronco C (2007) Fluid mechanics and cross filtration in hollow-fiber hemodialyzers, hemodiafiltration. In: Canaud B, Aljama P (eds) Contributions to nephrology. Basel, Switzerland, 158, p 34–49
22. Andrews JC (2004) Intralabyrinthine fluid dynamics: Meniere disease. Curr Opin Otolaryngol Head Neck Surg 12:408–412
23. Rashidi MM, Keimanesh M, Bég OA, Hung TK (2010) Magnetohydrodynamic biorheological transport phenomena in a porous medium: a simulation of magnetic blood flow control and filtration. Int J Numer Meth Biomed Eng 27:805–821
24. Fin L, Grebe R (2003) Three-dimensional modeling of the cerebrospinal fluid dynamics and brain interactions in the aqueduct of sylvius. Comput Meth Biomech Biomed Eng 6:163–170
25. Guasto JS, Rusconi R, Stocker R (2012) Fluid mechanics of planktonic microorganisms. Ann Rev Fluid Mech 44:373–400
26. Bég OA, Rashidi MM, Bég TA, Asadi M (2012) Homotopy analysis of transient magneto-bio-fluid dynamics of micropolar squeeze film in a porous medium: a model for magneto-bio-rheological lubrication. J Mech Med Biol 12:1250051.1–1250051.21
27. Zueco J, Bég OA (2010) Network numerical analysis of hydromagnetic squeeze film flow dynamics between two parallel rotating disks with induced magnetic field effects. Tribol Int 43:532–543
28. Becker S, Kniesburges S, Müller S, Delgado A, Link G, Kaltenbacher M, Döllinger M (2009) Flow-structure-acoustic interaction in a human voice model. J Acoust Soc Am 125:1351–1361
29. Bhargava R, Sharma S, Bég OA, Zueco J (2010) Finite element study of nonlinear two-dimensional deoxygenated biomagnetic micropolar flow. Commun Nonlinear Sci Numer Simul 15:1210–1233
30. Bég OA, Bhargava R, Rawat S, Takhar HS, Halim MK (2008) Computational modeling of biomagnetic micropolar blood flow and heat transfer in a two-dimensional non-darcian porous medium. Meccanica 43:391–410
31. Dasi LP, Simon HA, Sucosky P, Yoganathan AP (2009) Fluid mechanics of artificial heart valves. Clin Exp Pharmacol Physiol 36:225–237
32. Norouzi M, Davoodi M, Bég OA, Joneidi AA (2013) Analysis of the effect of normal stress differences on heat transfer in creeping viscoelastic dean flow. Int J Thermal Sci. doi:org/10.1016/j.ijthermalsci. 2013.02.002
33. Hörschler I, Meinke M, Schroder W (2003) Numerical simulation of the velocity field in a model of the nasal cavity. Comput Fluids 32:39–45
34. Choi SUS (1995) Enhancing thermal conductivity of fluid with nanoparticles. In: Siginer DA, Wang HP (eds) Developments and application of non-newtonian flows, vol 66. ASME, New York, pp. 99–105
35. Keblinski P, Eastman JA, Cahill DG (2005) Nanofluids for thermal transport. Materials Today June:36–44
36. Hilt JZ, Peppas NA (2005) Microfabricated drug delivery devices. Int J Pharm 306:15–23
37. Patel GM, Patel GC, Patel RB, Patel JK, Patel M (2006) Nanorobot: a versatile tool in nanomedicine. J Drug Targeting 14:63–67
38. Emerich DF, Thanos CG (2006) The pinpoint promise of nanoparticle-based drug delivery and molecular diagnosis. Biomol Eng 23:171–184
39. Su D, Ma R, Zhu L (2011) Numerical study of nanofluid infusion in deformable tissues for hyperthermia cancer treatments. J Med Biol Eng 49:1233–1240

40. Burygin GL (2009) On the enhanced antibacterial activity of antibiotics mixed with gold nanoparticles. Nanoscale Res Lett 4:794–801
41. Coco R, Plapied L, Pourcelle V, Jérôme Ch, Brayden DJ, Schneider YJ, Préat V (2013) Drug delivery to inflamed colon by nanoparticles: comparison of different strategies. Int J Pharm 440:3–12
42. Paolino D, Fresta M, Sinha P, Ferrari M (2006) Drug delivery systems. . In: Webster JG (ed) Encyclopedia of medical devices and instrumentation 2nd edn. Wiley, New York
43. Bég OA (2013) Peristaltic pumps- FSI modelling. Technical report, Gort Engovation, BIO-FSI/02-13., February, pp 142
44. Bayliss WM, Starling EH (1899) The movements and innervation of the small Intestine. J Physiol (London) 24:99–143
45. Fung YC (1971) Peristaltic pumping: a bioengineering model. In Proceedings of Workshop Hydrodynamics. Upper UrinaryTract, Natl Acad Sci, Washington DC
46. Latham TW (1966) Fluid motion in peristaltic pump. MS thesis. MIT, USA
47. Shapiro AH, Jafferin MY, Weinberg SL (1969) Peristaltic pumping with long wavelengths at low Reynolds number. J Fluid Mech 37:699–825
48. Fung YC, Yih CS (1968) Peristaltic transport. ASME J Appl Mech 35:669–675
49. Burns JC, Parkes T (1967) Peristaltic motion. J Fluid Mech 29:731–743
50. Barton C, Raynor S (1968) Peristaltic flow in tubes. Bull Math Biophys 30:663–680
51. Chaw TS (1970) Peristaltic transport in a circular cylindrical pipe. ASME J Appl Mech 37:901–905
52. Jafferin MY (1973) Inertia and streamline curvature effects on peristaltic pumping. Int J Eng Sci 11:681–699
53. Bohme G, Friedrich R (1983) Peristaltic flow of viscoelastic liquids. J Fluid Mech 128:109–122
54. Tsiklauri D, Beresnev I (2001) Non-Newtonian effects in the peristaltic flow of a Maxwell fluid. Phys Rev E 64:036303-1–036303-5
55. Tripathi D (2011) Peristaltic flow of couple-stress conducting fluids through a porous channel: applications to blood flow in the micro-circulatory system. J Biol Syst 19:461–477
56. Tripathi D (2011) Peristaltic transport of fractional maxwell fluids in uniform tubes: application of an endoscope. Comput Math Appl 62:1116–1126
57. Tripathi D (2011) Numerical and analytical simulation of peristaltic flows of generalized Oldroyd-B Fluids. Int J Numer Meth Fluids 67:1932–1943
58. Hayat T, Mahomed FM, Asghar S (2005) Peristaltic flow of a magnetohydrodynamic Johnson-Segalman fluid. Nonlinear Dyn 40:375–385
59. Tripathi D, Pandey SK, Bég. OA (2013) Mathematical modelling of heat transfer effects on swallowing dynamics of viscoelastic food bolus through the human oesophagus. Int J Therm Sci 70:41–53
60. Bég OA, Keimanesh M, Rashidi MM, Davoodi M (2013) Multi-Step dtm simulation of magneto-peristaltic flow of a conducting Williamson viscoelastic fluid. Int J Appl Math Mech 9:1–24
61. Bhargava R, Sharma S, Takhar HS, Bég TA, Bég OA, Hung TK (2006) Peristaltic pumping of micropolar fluid in porous channel—model for stenosed arteries. J Biomech 39:S649–S669
62. Tiwari RK, Das MK (2007) Heat transfer augmentation in a two-sided lid-driven differentially heated square cavity utilizing nanofluids. Int J Heat Mass Transf 50:2002–2018
63. Rashidi MM, Bég OA, Mehr NF, Hosseini A, Gorla RSR (2012) Homotopy simulation of axisymmetric laminar mixed convection nanofluid boundary layer flow over a vertical cylinder. Theor Appl Mech 39:365–390
64. Bég OA, Gorla RSR, Prasad VR, Vasu B, Prashad RD (2011) Computational study of mixed thermal convection nanofluid flow in a porous medium. 12th UK National Heat Transfer Conference, Chemical Engineering Department, University of Leeds, West Yorkshire, Session 8, ID 0004
65. Rana P, Bhargava R, Bég OA (2013) Finite element modeling of conjugate mixed convection flow of al2o3-water nanofluid from an inclined slender hollow cylinder. Phys Scripta 87:15

66. Buongiorno J (2006) Convective transport in nanofluids. ASME J Heat Transf 128:240–250
67. Xuan Y, Li Q (2003) Investigation on convective heat transfer and flow features of nanofluids. ASME J Heat Transf 125:151–155
68. Pak BC, Cho Y (2003) Hydrodynamics and heat transfer study of dispersed fluids with submicron metallic oxide particles. Exp Heat Transf 11:151–170
69. Eastman JA, Phillpot SR, Choi SUS, Keblinski P (2004) Thermal transport in nanofluids. Ann Rev Mater Res 34:219–146
70. Wang XQ, Majumdar AS (2007) Heat transfer characteristics of nanofluids: a review. Int J Therm Sci 46:1–19
71. Trisaksri V, Wongwises SP (2007) Critical review of heat transfer characteristics of nanofluids. Renew Sust Energ Rev 11:512–523
72. Wen D, Lin G, Vafaei S, Zhang K (2009) Review of nanofluids for heat transfer applications. Particuology 7:141–150
73. Saidur R, Leong KY, Mohammad HA (2011) A review on applications and challenges of nanofluids. Renew Sust Energ Rev 15:1646–1668
74. Kuznetsov AV, Nield DA (2010) Natural convection boundary layer flow of nanofluids past a vertical plate. Int J Therm Sci 49:243–247
75. Kuznetsov AV, Nield DA (2011) Double-diffusive natural convective boundary-layer flow of a nanofluid past a vertical plate. Int J Therm Sci 50:712–717
76. Nield DA, Kuznetsov AV (2011) The onset of double-diffusive convection in a nanofluid layer. Int J Heat Fluid Flow 32:771–776
77. Nield DA, Kuznetsov AV (2010) The onset of convection in a horizontal nanofluid layer of finite depth. Eur J Mech B/Fluids 29:217–223
78. Nield DA, Kuznetsov AV (2009) Thermal instability in a porous medium layer saturated by a nanofluid. Int J Heat Mass Transf 52:5796–5801
79. Nield DA, Kuznetsov AV (2010) The effect of local thermal non-equilibrium on the onset of convection in a nanofluid. ASME J Heat Transf 132:052405-1-7
80. Kolade B, Goodson KE, Eaton JK (2009) Convective performance of nanofluids in a laminar thermally-developing tube flow. ASME J Heat Transf 131:052402-1-8
81. Bachok N, Ishak A, Pop I (2010) Boundary-layer flow of nanofluids over a moving surface in a flowing fluid. Int J Therm Sci 49:1663–1668
82. Rashidi MM, Bég OA, Asadi M, Rastegari MT (2012) DTM- Padé modeling of natural convective boundary layer flow of a nanofluid past a vertical surface. Int J Therm Environ Eng 4:13–24
83. Bég OA, Bég TA, Rashidi MM, Asadi M (2012) Homotopy semi-numerical modelling of nanofluid convection boundary layers from an isothermal spherical body in a permeable regime. Int J Micro Nano Therm Fluid Transp Phenom 3:237–266
84. Rashidi MM, Bég OA, Rostami B, Osmond L (2013) DTM- Padé simulation of stagnation-point nanofluid mechanics. Int J Appl Math Mech 9:1–29
85. Bég OA, Ferdows M, Khan Md S (2013) Numerical study of transient magnetohydrodynamic radiative mixed convection nanofluid flow from a stretching permeable surface. Proceedings of IMECHE–Part E: J Proc Mech Eng
86. Bég OA, Bég TA, Rashidi MM, Asadi M (2013) DTM- Padé semi-numerical simulation of nanofluid transport in porous media. Int J Appl Math Mech 9:80–107
87. Bég OA, Prasad VR, Vasu B (2013) Numerical study of mixed bio-convection in porous media saturated with nanofluid containing oxytactic micro-organisms. J Mech Med Biol 13:1350067
88. Uddin MJ, Yusoff NHM, Bég OA, Ismail AI (2013) Lie group analysis and numerical solutions for non-Newtonian nanofluid flow in a porous medium with internal heat generation. Physica Scripta 87:14
89. Akbar NS, Nadeem S (2011) Endoscopic effects on peristaltic flow of a nanofluid. Commun Theor Phys 56:761
90. Akbar NS, Nadeem S, Hayat T, Hendi AA (2012) Peristaltic flow of a nanofluid with slip effects. Meccanica 47:1283–1294

91. Akbar NS, Nadeem S, Hayat T, Hendi AA (2012) Peristaltic flow of a nanofluid in a non-uniform tube. Heat Mass Transf 48:451–459
92. Mustafa M, Hina S, Hayat T, Alsaedi A (2012) Influence of wall properties on the peristaltic flow of a nanofluid: analytic and numerical solutions. Int J Heat Mass Transf 55:4871–4877
93. Akbar NS, Nadeem S (2012) Peristaltic flow of a Phan-Thien-tanner nanofluid in a diverging tube. Heat Transf-Asian Res 41:10–22
94. Mustafa M, Hina S, Hayat T, Alseadi A (2013) Slip effects on the peristaltic motion of nano-fluid in a channel with wall properties. ASME J Heat Transf 135:041701-1-7
95. Kleinstreuer C, Li J (2010) Chapter 5 Microfluidic devices in nanotechnology. Microfluidic devices for drug delivery. Wiley, New York
96. Gebhart B (1988) Buoyancy-induced flows and transport. Hemisphere, Washington
97. Bég OA, Bhargava R, Rawat S, Kahya E (2008) Numerical study of micropolar convective heat and mass transfer in a non-Darcy porous regime with Soret and Dufour diffusion. Emer J Eng Res 13:51–66
98. Bég OA, Prasad VR, Vasu B, Reddy NB, Li Q, Bhargava R (2011) Free convection heat and mass transfer from an isothermal sphere to a micropolar regime with Soret/Dufour effects. Int J Heat Mass Transf 54:9–18
99. Prasad VR, Vasu B, Bég OA (2013) Thermo-diffusion and diffusion-thermo effects on free convection flow past a horizontal circular cylinder in a non-Darcy porous medium. J Porous Media 16:315–334

Chapter 5
Numerical Study on Isotachophoretic Separation of Ionic Samples in Microfluidics

Partha P. Gopmandal and S. Bhattacharyya

5.1 Introduction

Motion of suspended charged particles or macromolecules in an aqueous medium under the action of an applied electric field is known as electrophoresis. Under the influence of an applied electric field E, the ionic species will move with a velocity v as $v = \mu E$, where μ is termed as mobility of the ionic species. Differences in motilities cause differences in velocities and by utilizing this effect the ionic species can be separated. Capillary electrophoresis (CE) is an electrophoretic separation technique performed in a capillary [1–4]. It has a wide range of applications in different fields such as chemical, biotechnological, biomedical, colloidal, and environmental sciences. Depending on the choice of electrolytes present in the system, several mode of CE occurs, such as zone electrophoresis (ZE), moving boundary electrophoresis (MBE), isoelectric focusing (IEF), and isotachophoresis (ITP).

ITP is a powerful electrokinetic technique for the preconcentration, separation, purification, and quantification of ionic analytes. In ITP, a target sample mixture is placed between a leading electrolyte (LE) with higher electrophoretic mobility and a terminating electrolyte (TE) with electrophoretic mobility lower than that of the sample species. When a constant electric field is applied along the axis of the channel, the ions of the analytes are arranged based on the electrophoretic mobilities in increasing order. This phenomenon of consecutive stacking of ions in order of their electrophoretic mobilities is referred as ITP. Once a steady state has been reached, all species move at the same speed, known as isotachophoretic velocity, U^{ITP}. The cationic or anionic ITP corresponds the stacking of cations or anions, respectively. The electric field in each stack is constant. This leads to a formation of a sharp interface (or transition zone) between the stacks in which steep gradients or step jump in ionic concentration and electric field develop. An accurate resolution of such sharp transition zones is a challenge in ITP modeling and has several technical applications in the Lab-on-a-Chip technology [5, 6].

S. Bhattacharyya (✉) · P. P. Gopmandal
Department of Mathematics, Indian Institute of Technology Kharagpur, 721302 Kharagpur, India
e-mail: somnath@maths.iitkgp.ernet.in

S. K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_5,

Depending on the amount of the sample present in the system, we can have two modes of ITP, namely peak mode and plateau mode [7, 8]. In the plateau mode ITP, the interface widths are negligible compared with the sample zone width. Unlike the plateau mode ITP, the sample ionic concentration in a peak mode ITP may vary with the axial position and the sample zone appears as a sharp peak between two adjacent electrolytes, that is, LE and TE.

Kohlrausch [9] first reports the mathematical formulation of ITP by introducing the concept of a regulating function (KRF) through which the conditions at the interface between two different ions with the same counter ion can be established. In ITP modeling, electroneutrality is assumed and the electric field is determined from the charge-conservation equation. Several authors [8, 10, 11] concluded that the electroneutrality is valid if the Debye length, based on ionic concentration of ionic species, is much smaller than the characteristic width of the transition zone.

The step changes in electric field and concentration of ions within the interface zones lead to a strong hyperbolic characteristic of the advection–diffusion equations for ion transport. Several authors, for example [12], described the ITP transport as a similar phenomenon as shock-wave propagation in gas dynamics. Recently, in a review article, Thormann and others [13] provided a state-of-the-art on the available computer simulation software for ITP separation. However, those numerical methods are based on 1-D analysis, which may not be suitable to analyze a dispersed ITP.

The spatial difference in fluid velocity may arise due to an unwanted pressure gradient or a nonuniform electro-osmotic flow (EOF) of electrolytes. The nonuniform convection produces dispersion in ITP. The EOF refers to the convection of electrolytes pasting a charged surface as a consequence of an applied electric field. A detailed discussion on mathematical modeling of EOF is made in our earlier chapter [14]. The nonuniform EOF of ions, which arises due to the variation of electric field across the interface and electro-osmotic mobility, leads to the development of an induced pressure gradient [15]. Unwanted pressure gradients may arise due to the complex geometrical shape of an ITP channel, for example, constricted channels [16] or dog-leg channels or near a cross-channel junction. A detailed knowledge of the impact of dispersion on the ITP transition zone is important. The dispersion should be minimized for an efficient sample stacking in ITP. Thus, it is important to understand the impact of dispersion on the form of interface, distribution of ionic species, and electric field. In this chapter we refer ITP with no dispersion effect as ideal ITP.

The impact on the ITP transient region due to an EOF and/or pressure-driven flow was addressed by several authors [8, 12, 15–20]. Based on the asymptotic expansion of the cross-sectional averaged variables, Saville and others [17] studied the imposed convection effect on ITP. Schonfeld and others [18] made a combined numerical and experimental study on the electro-osmotic effect on ITP near an interface zone. Upon computing the 1-D advection–diffusion equation, Bercovici and others [12] studied the dispersion in ITP due to an external pressure gradient. Shim and others [16] imposed a counter flow in the ITP process in order to stop the migration of interface zones. Garcia–Schwartz and others [8] studied both experimentally

and numerically the dispersion effect on transition zone in peak mode ITP due to an on-uniform EOF of electrolytes. They developed an area-averaged analytical model for sample dispersion based on the Taylor–Aris dispersion coefficient. Baier and others [15] developed an analytical approximation for the flow field in the vicinity of an ITP transition zone between electrolytes of different electro-osmotic velocity. The numerical studies presented in [8, 15, 18] use finite element-based commercial package COMOSOL. Recently, Bhattacharyya and others [20] studied a particular mode of ITP which employs a pressure-driven flow to counter the electromigration of a sample in order to anchor a sample zone at a specific position along a channel or capillary.

In this work, we present a computer model to analyze the ion dynamics in ideal ITP as well as ITP with convective dispersion. We solve numerically the Nernst–Planck equations for ion distribution which are coupled with the charge conservation equation for electric potential and Navier–Stokes equations for fluid flow. A finite volume-based second order accurate upwind scheme Quadratic Upwind Interpolation for Convective Kinematics (QUICK) is adopted to resolve efficiently the sharpness which arise in the ITP transport. The present algorithm is tested for accuracy by comparing with the analytical solution for a plateau mode ideal ITP. In peak mode ITP, the sample zone smears out and the concentration of the sample may not follow the KRF. The preconcentration dynamics and its dependence on several parameters in peak mode ITP is different from the plateau mode ITP. The dispersion in distribution of ionic species is measured through the second- and third-order moment analysis.

5.2 Mathematical Model

We assume the electrolytes to consist of monovalent LE of cationic concentration C_l, monovalent trailing electrolyte (TE) of cationic concentration C_t, cationic concentration of sample species C_s, and a common anion of concentration C_0. Here, the solvent, that is, water is in excess and its ionization is assumed to be negligible. The migration of charged species in an electrolyte under an external electric field is governed by diffusion, convection, and electromigration. We consider x-axis along the wide channel and y-axis perpendicular to it. The mass conservation of the ionic species leads to the Nernst–Planck equation:

$$\frac{\partial c_i}{\partial t} + \nabla . N_i = 0, \tag{5.1}$$

where $i = t, s, l,$ and 0. The concentration of common ion C_0 is obtained through electroneutrality assumption, that is, $C_0 = \sum_{i \neq 0} z_i C_i$. The molar flux of i-th species is:

$$N_i = C_i \mu_i z_i E + C_i q - D_i \nabla C_i, \tag{5.2}$$

where valance $z_i = +1$ for cations, and $z_0 = -1$ for common anion. The mobility and diffusivity are related via the Nernst–Einstein relation $z_i = D_i F / RT$, where F is the Faraday constant, R is the gas constant, and T is absolute temperature. The convective speed is governed by $\mathbf{q}$ and $\mathbf{E} = -\nabla\phi$ is the electric field. The electric current density and charge density are defined, respectively, as $j = F\sum_i z_i N_i$, and $\rho_e = \sum_i z_i C_i$. Conservation of electric charge leads to:

$$\frac{\partial \rho_e}{\partial t} + \nabla . j = 0. \tag{5.3}$$

Under electroneutrality condition, that is, $\rho_e = \sum z_i C_i = 0$, the unsteady term and convective terms in Eq. 5.3 vanishes. The electric field $\mathbf{E}$ can be obtained from the charge conservation equation as:

$$\nabla .(\nu \nabla \phi) = \nabla .(F\sum_i D_i z_i \nabla c_i), \tag{5.4}$$

where ionic conductivity is given by $\nu = F\sum_i z_i^2 C_i$. The term in the right-hand side of Eq. 5.4 is the diffusion current and its contribution is insignificant at all locations, except at the transition zones. We consider a long wide channel so that the motion of electrolytes can be considered as 2-D. The Cartesian coordinates are nondimensionalized by H and velocity is scaled by U^{ITP}. We nondimensionalize the concentration by the bulk concentration of LE, that is, C_l^{∞} and potential by $\phi_0 = RT / F$. The charge conservation equation for electroneutral solution in nondimensional form obtained from Eq. 5.4 as:

$$\begin{aligned} &\nabla .\left[\left(c_l + \frac{D_s + D_0}{D_l + D_0} c_s + \frac{D_t + D_0}{D_l + D_0} c_t\right)\nabla \phi\right] = \\ &-\left(\frac{z_l D_l + z_0 D_0}{D_l + D_0}\right)\nabla^2\left(c_l + \frac{z_s D_s + z_0 D_0}{z_l D_l + z_0 D_0} c_s + \frac{z_t D_t + z_0 D_0}{z_l D_l + z_0 D_0} c_t\right), \end{aligned} \tag{5.5}$$

where the gradient operator is given by:

$$\nabla = \frac{\partial}{\partial x}\hat{e}_x + \frac{\partial}{\partial y}\hat{e}_y \tag{5.6}$$

and $\hat{e}_x$ and $\hat{e}_y$ are the unit vectors along the x and y directions. We consider that the convection effects on ITP transport of ions are either due to nonuniform EOF or pressure driven. Under the bulk fluid flow with average speed $\bar{u}$, the transition zone velocity becomes $\bar{U} = U^{ITP} + \bar{u}$. We consider that the coordinate is moving with the average speed of the ions, that is, $\bar{U}$. This will allow us to consider a truncated computational domain. A variable transformation is introduced as:

$$\bar{x} = x - \bar{U}t,\ \bar{y} = y,\ \bar{t} = t. \tag{5.7}$$

The Nernst–Planck equations in nondimensionalized form can be expressed as:

$$\frac{\partial c_i}{\partial t} - (1+\alpha)\frac{\partial c_i}{\partial x} + (\mathbf{q}.\nabla)c_i - \frac{D_i}{D_l}\frac{1}{Pe}\nabla^2 c_i - z_i\frac{D_i}{D_l}\frac{1}{Pe}\nabla(c_i.\nabla\phi) = 0. \tag{5.8}$$

We drop the overline from the transformed independent variables, that is, $\bar{x}, \bar{y},$ and $\bar{t}$. Here time is scaled by $\tau = H / U^{ITP}$. Different parameters arising in the nondimensional Nernst–Planck equations are Reynolds number $\mathrm{Re} = \rho H U^{ITP} / \eta$, and Peclet number $Pe = U^{ITP} H / D_l = \mathrm{Re}.Sc,$ and $\alpha = \bar{u} / U^{ITP}$. The electric field is determined by solving the charge conservation equation (Eq. 5.5).

We impose a fixed potential drop $E_0 = \phi_0 L$ along the channel. Due to discontinuous conductivity of the electrolytes, the electric field will vary. All the species migrate with the same ITP velocity U^{ITP} in the steady state. Due to a distinct electrophoretic mobility of each ionic species, the electric field strength of each zone will be different. The electric field in each zone is adjusted so that the zones migrate at a constant speed, U^{ITP}. If $E_i (i = t,s,l)$ are the strengths of electric field in each separated zone occupied by only one species, then using the relation $U^{ITP} = E_i \mu_i$, we get:

$$E_i = E_0 \frac{1/\mu_i}{\sum_i l_i / \mu_i}, \tag{5.9}$$

where the summation is taken over all the species, and $l_i = L_i / L$ is the portion of the length L filled by the i-th species. We can also express E_t and E_l in terms of the applied current density j_0 as follows:

$$j_0 = F(\mu_t + \mu_0)C_t^{\infty} E_t = F(\mu_s + \mu_0)C_s^{\infty} E_s = F(\mu_l + \mu_0)C_l^{\infty} E_l. \tag{5.10}$$

From the fact that at steady state all species moves at a constant speed $U^{ITP} = E_i \mu_i$, the above relation leads to the relationships for the bulk concentration of TE, LE, and all the sample species as:

$$\frac{c_t^{\infty}}{c_l^{\infty}} = \frac{\mu_t}{\mu_l}\frac{\mu_l + \mu_0}{\mu_t + \mu_0}, \frac{c_s^{\infty}}{c_l^{\infty}} = \frac{\mu_s}{\mu_l}\frac{\mu_l + \mu_0}{\mu_s + \mu_0}, \tag{5.11}$$

where suffix ∞ stands for the bulk value of the concentrations. We consider a long wide channel so that the motion of electrolytes can be considered to be 2-D.

In solving the Nernst–Planck equations, the net flux through the channel walls are set to be zero, that is, $N_i.\hat{n} = 0$ for $i = t,s,l$ and $\hat{n}$ is the unit outward normal. The electric potential is subjected to insulating boundary conditions along the wall, that is, $\nabla\phi.\hat{n} = 0.$ The left and right boundaries of the computational domain are placed sufficiently far away from the transition zones. The concentration far away from the transition zone is governed by the Kohlrauch's condition (Eq. 5.11). The condition for electric field along the inlet and outlet boundaries of the computational domain

is governed by Eq. 5.9. To study the convection effect on the transition zone, we consider the ideal ITP solution as the initial condition.

The velocity field q due to mismatch of EOF of electrolytes is governed by the Navier–Stokes equations as:

$$\rho[\frac{\partial q}{\partial t}+(q.\nabla)q]=-\nabla p+\eta\nabla^2 q+\varepsilon_e\nabla\varphi\nabla^2\varphi \tag{5.12}$$

where $p, \eta, \text{and } \rho$ are the fluid pressure, viscosity, and density, respectively. The last term in the above equation is important near the transition zone. This electrical stress arises due to the spatial variation of ϕ across the channel, whether the electrolyte solution is electroneutral or not. Garcia–Schwartz and others [8] found that the inclusion of electric body force produces a significant effect for relatively high-applied current density. Taking $\tau=\eta U^{ITP}/H$ as a pressure scale, the nondimensional form of the Navier–Stokes equations can be written as:

$$\frac{\partial q}{\partial t}+(q.\nabla)q=-\frac{1}{\mathrm{Re}}\nabla p+\frac{1}{\mathrm{Re}}\nabla^2 q+B\nabla\varphi\nabla^2\varphi, \tag{5.13}$$

where $B=(\varepsilon_e\phi_0^2/\eta H)\,\mathrm{Re}U^{ITP}$. The nondimensional equation of continuity for incompressible fluid is given by:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0, \tag{5.14}$$

where u and v are the velocity components along the x and y directions, respectively. We can either assume a no-slip condition along the channel wall or a free-slip velocity condition, based on the Smoluchowski equation, along the outer edge of the channel induced Debye layers. Far upstream and downstream of the transition zones, the velocity is assumed to be known. We solve the Navier–Stokes equations, which are coupled with the ion concentration and electric field equations, subjected to the prescribed boundary conditions to determine the convective velocity of ions. We considered the convection along the direction of ITP transport, that is, $\bar{u}>0$.

5.2.1 Numerical Methods

The finite-volume method (FVM) method is used for solving the governing equations for mass, momentum, and electric potential distribution. In this method, the equations, which are cast into conservation law form, are integrated over a control volume. The variables on the control volume interfaces are estimated by a linear interpolation between the two neighboring cells to either sides of the control volume interface. This approach enables us to compute the jump discontinuity as part of the solution. Thus, the finite volume discretization is more efficient than the finite difference scheme when a sharp gradient or step jump is expected. Here, we have

used the staggered grid arrangement [21] for allocating the variables, because of its advantages in considering the influence of the pressure gradients on fluid flow.

In ITP problems, the variables may suffer a step change (plateau mode) or steep gradient (peak mode) across the transition zones. The traditional finite volume method for computation of such cases is unsuitable, as the numerical solution may produce wiggles. The Nernst–Planck equations, based on convection, electromigration, and diffusion of ions may exhibit hyperbolic characteristics near a transition zone. To circumvent this problem, generally upwind schemes are used.

To illustrate the discretization of the transport equations, we express the general convection–diffusion–electromigration equations for the generic variable c as:

$$\frac{\partial c}{\partial t}+\left[\frac{\partial}{\partial x}(Fc)+\frac{\partial}{\partial y}(Gc)\right]-\nabla^2 c=0, \tag{5.15}$$

where the convective and electromigration terms are included in F and G, respectively. The first and last terms of Eq. 5.15 are due to the time dependency and diffusion, respectively. The computational domain is subdivided into a number of elementary rectangular cells, Ω_P with area $d\Omega_P$ whose sides are dx_P and dy_P. Equation 5.15 when integrated over a cell Ω_P (Fig. 5.1) yields the discretized form to advance the solution from k-th time step to $(k+1)th$ time step as:

$$\begin{aligned}&\frac{c_P^{k+1}-c_P^k}{dt}dx_P dy_P+(F_e c_e-F_w c_w)^{k+1}dy_P+(G_n c_n-G_s c_w)^{k+1}dx_P\\&-\left(\left[\frac{\partial c}{\partial x}\right]_e-\left[\frac{\partial c}{\partial x}\right]_w\right)^{k+1}dy_P-\left(\left[\frac{\partial c}{\partial y}\right]_n-\left[\frac{\partial c}{\partial y}\right]_s\right)^{k+1}dx_P=0.\end{aligned} \tag{5.16}$$

Here $e, w, n,$ and s refer to the eastern, western, northern, and southern faces of the cell (Fig. 5.1). An implicit first-order scheme is used to discretize the time derivatives present. The diffusion terms are discretized through a central difference scheme. The diffusion flux at the interfaces "e" and "w" are evaluated as:

$$\left[\frac{\partial c}{\partial x}\right]_e=\frac{c_E-c_P}{0.5(dx_P+dx_E)} \quad \text{and} \quad \left[\frac{\partial c}{\partial x}\right]_w=\frac{c_P-c_W}{0.5(dx_P+dx_W)}.$$

Similar procedure is adopted to estimate the diffusion flux at the other cell faces "n" and "s." Note that the big letter subscripts denote the cell centers in which variables are stored and small letter subscripts denote the corresponding cell faces.

The variable c on the control volume interface is obtained by a linear interpolation between its values on the two neighboring cells on either side of the interface. If we consider that the interface is midway of the corresponding cell centers, then the value of c at interfaces "e" and "w" are evaluated as:

$$c_e=0.5(c_E+c_P),\ c_w=0.5(c_P+c_W).$$

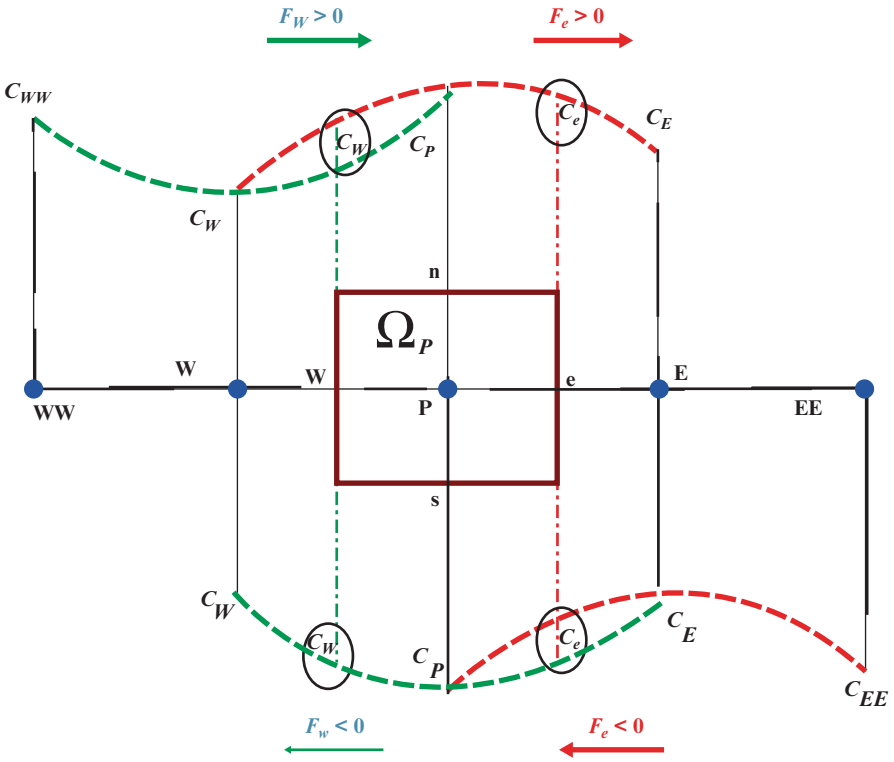

Fig. 5.1 Schematic diagram for control volume Ω_P (e, w, n, and s are the cell faces of the cell centered at P, interpolation for a variable c based on QUICK scheme)

Similar procedure is adopted to estimate c_n and c_s. Therefore, the discretized Eq. 5.15 can be written as:

$$a_P c_P^{k+1} = a_E c_E^{k+1} + a_W c_W^{k+1} + a_N c_N^{k+1} + a_S c_S^{k+1} + a_P^0 c_P^k, \tag{5.17}$$

where the coefficients a_i (i=E, W, P, N, and S) are as follows:

$$a_E = D_e - \frac{1}{2}F_e,\ a_W = D_w + \frac{1}{2}F_w$$

$$a_N = D_n - \frac{1}{2}F_n,\ a_S = D_s + \frac{1}{2}F_s$$

$$a_P = a_E + a_W + a_N + a_S + (F_e - F_w) + (F_n - F_s) + a_P^0$$

with $a_P^0 = \dfrac{dx_P dy_P}{dt}$. The contribution of the diffusion terms are included in D. The discretization procedure, as discussed above, is central difference type and second

order accurate. Stability of this scheme requires the cellular Peclet number to be less than 2. This scheme may produce wiggles near a step change in variable values [22]. An improvement to this scheme can be made by using upwind differences to the first-order derivatives in Eq. 5.15. In this scheme, the variable c at the cell interfaces in discretized equation (Eq. 5.16) is estimated as:

$$c_e = \begin{cases} c_P, & F_e > 0; \\ c_E, & F_e < 0; \end{cases}$$

$$c_w = \begin{cases} c_W, & F_w > 0; \\ c_W, & F_w < 0; \end{cases}$$

Similar procedure is adopted to estimate c_n and c_s. Replacing the values of c in Eq. 5.15 at the cell interfaces (e, w, n, and s) by these approximation, the coefficients a_i (i=E, W, P, N, and S) become:

$$a_E = D_e + [F_e, 0], a_W = D_w + [F_w, 0]$$
$$a_N = D_n + [F_n, 0, a_S = D_s + [F_s, 0]$$
$$\text{and} \quad a_P = D_e + [F_e, 0] + D_w + [F_w, 0] + D_n + [-F_n, 0] + D_s + [F_s, 0] + a_P^0,$$

where the operator $[a,b]$ yields the larger of a and b. The above discussed upwind scheme is first order accurate. The diffusion terms are discretized through a second-order scheme involving a linear interpolation of variable values at the neighboring cells. An improvement on accuracy of the upwind scheme can be made through a QUICK scheme. The QUICK scheme [23] uses a three-point upstream-weighted quadratic interpolation for the c at the cell faces. The main advantage of the QUICK scheme is that it reduces the numerical dissipation. In this scheme, the cell-face values of c_e and c_w can be obtained as:

$$c_e = \begin{cases} \left(\frac{3}{8}c_E + \frac{3}{4}c_P - \frac{1}{8}c_W\right), & F_e > 0 \\ \left(\frac{3}{4}c_E + \frac{3}{8}c_P - \frac{1}{8}c_{EE}\right), & F_e < 0 \end{cases}$$

$$c_w = \begin{cases} \left(\frac{3}{8}c_P + \frac{3}{4}c_W - \frac{1}{8}c_{WW}\right), & F_w > 0 \\ \left(\frac{3}{4}c_P + \frac{3}{8}c_W - \frac{1}{8}c_E\right), & F_w < 0. \end{cases}$$

and similarly the values of c at interfaces "n" and "s."

At any time level, we use an iterative procedure for the computation of ion transport equations, as these equations are coupled with the charge conservation equation and velocity field. The iteration procedure starts with an assumption for potential at each cell and velocity field. At every iteration, the electric field is determined by solving the reduced elliptic equation for charge conservation equation, that is,

Eq. 5.5. The elliptic equation (Eq. 5.5) is solved by a line-by-line iterative method along with the successive-overrelaxation (SOR) technique.

The convective velocity of ions is governed by the Navier–Stokes equation. At every iteration, the Navier–Stokes equations are discretized by using a control volume approach [21] over a staggered grid system, that is, the velocity components are stored at the cell interfaces to which it is normal. Here, we use the QUICK scheme to discretize the convective terms.

The governing discretized equations are solved through a pressure-correction-based algorithm, SIMPLE [21]. The pressure link between the continuity and momentum equations is accomplished by transforming the discretized continuity equation into a Poisson equation for pressure correction. This Poisson equation implements a pressure correction for a divergent velocity field. The pressure Poisson equation is given by:

$$\frac{1}{dx_P}\left[\frac{p_E^c - p_P^c}{0.5(dx_P + dx_E)} - \frac{p_P^c - p_W^c}{0.5(dx_P + dx_W)}\right] + \frac{1}{dy_P}\left[\frac{p_N^c - p_P^c}{0.5(dy_P + dy_N)} - \frac{p_P^c - p_S^c}{0.5(dx_P + dx_S)}\right] = \frac{1}{dt}\left[\frac{u_P^* - u_W^*}{dx_P} + \frac{v_P^* - u_S^*}{dy_P}\right]$$

The variable p^c denotes the pressure correction and u^* and v^* are the velocity components obtained by solving the momentum equations. Thus, at any time step a single iteration in SIMPLE algorithm consists of the following sequential steps:

1. An implicit calculation of the momentum equations is performed. The equations are discretized through the scheme as discussed above. Because of the coupling nature, the resulting system of algebraic equations is solved through a block elimination method.
2. The Poisson equation for pressure correction is solved using the successive under relaxation method.
3. The velocity field at each cell is updated using the pressure correction.

At the end of every iteration, the velocity on the channel wall is obtained by (Eq. 5.29) using the updated values for concentration distribution and electric field. The pressure-correction iteration at each time step is continued until the divergence-free velocity field is obtained. However, for this purpose, the divergence in each cell is towed below a preassigned small quantity. The iterations are continued until the absolute difference between two successive iterations becomes smaller than the tolerance limit 10^{-6} for concentration as well as potential distribution across the channel.

A steady state solution is achieved by taking sufficient time steps until the concentrations and velocity field remain unchanged with time. The initial condition for dispersed ITP is governed by the solution of the corresponding ideal ITP case.

As the variables along the transition zone change more rapidly than elsewhere, a nonuniform grid along the *x*-direction and a uniform grid distribution along

the y-direction are considered. The grid size dx within the transition zone is considered to be small and it is increased by a constant value as we move away from the transition zone. The smallest dx is chosen such that the Courant–Friedrichs–Lewy (CFL) criterion is satisfied. A series of test runs are made to capture the interface characteristics with various grid sizes for ITP with or without the convection case. We have tested our code for different values of dx with coarse grid size varied from 0.01 to 0.005. When $D_r = 0.7,$ the total number of grid points considered are $1{,}000 \times 100$, $1{,}500 \times 100$, $1{,}500 \times 100$ where the first and the second numbers are the number of grid points along the x and y directions, respectively. Our test results suggest that $1{,}000 \times 100$ grid points for the computational domain having nondimensional channel length 10 with height 1 is optimal for ITP without the convection case. For the dispersion case, the length of the domain has been adjusted accordingly by increasing the number of uniform grid points within the dispersed transition zone.

5.3 Results and Discussions

In order to make quantitative predictions using this model, the electrophoretic mobilities of all species, concentration of the LE, voltage drop across the channel and height of the channel must be specified. We take $\phi_0 = 0.02586\text{V}$, $\varepsilon_e = 695.39 \times 10^{-12}$ C/Vm, $K_B = 1.381 \times 10^{-23}$ J/K, $e = 1.602 \times 10^{-19}$ C, $T = 300\text{K}$, $\eta = 10^{-3}$ Pa/s and $\rho = 10^3$ Kg/m^3. We consider that the diffusivity of LE and common ion are equal, that is, $D_l = D_0 = 7 \times 10^{-10}$ m^2/s and vary the diffusivity of TE by changing the value of the diffusivity ratio D_r, where $D_r = D_t / D_l$. The imposed electric field strength is considered to be $E_0 = 10^5$ V/m and the electric fields at the two ends are obtained by using Eq. 5.5. The electro-osmotic mobility of LE is varied between 1.35×10^{-8} and 2.03×10^{-7} m^2/Vs. The bulk concentration of LE is taken as $C_l^{\infty} = 0.001M$, so that the Debye layer thickness (~ 10 nm) is much smaller than the height of the channel considered here $(10 - 50\ \mu\text{m})$. Unless stated otherwise, we consider the channel height as $H = 25\ \mu\text{m}$.

5.3.1 ITP Without Dispersion (Ideal ITP)

We first describe an analytical solution for the case of ITP without dispersion (ideal ITP). Nernst–Planck equation for i-th ionic species under the electroneutrality condition in 1-D form is given by:

$$\frac{d}{dx}\left[-U^{ITP}C_i - D_i\frac{dC_i}{dx} + z_i\mu_i\frac{d}{dx}(C_iE_x)\right] = 0, \tag{5.18}$$

where E_x is the axial potential gradient. Integrating Eq. 5.18 and using the fact that thermodynamic potential at TE and LE ends remain constant along with constant values of concentrations at both the ends, yields:

$$\frac{d}{dx}\left[\ln\left(\frac{C_l}{C_t}\right)\right] = \frac{E_t - E_l}{\phi_0}. \tag{5.19}$$

Integrating Eq. 5.19 and considering the origin of the coordinate system in the region where the concentration of both the species are equal, that is, $C_l = C_t$, we get:

$$\ln\left(\frac{C_l}{C_t}\right) = \frac{\Delta E}{\phi_0} x, \tag{5.20}$$

where $\Delta E = E_t - E_l$, the difference in electric field strengths at TE and LE. We define the transition zone as the region where the concentration ratio C_l / C_t changes from e^2 to $1/e^2$. The width of the transition zone where most change in concentration occurs, is given by:

$$\Delta X = \frac{4\phi_0}{\Delta E}. \tag{5.21}$$

Under the electroneutrality condition, the axial electric field E_x can be obtained from the charge conservation equation (neglecting the diffusion current) as:

$$\frac{d}{dx}(\nu E_x) = 0. \tag{5.22}$$

Integrating Eq. 5.21 and using the relation given in Eq. 5.20, the axial electric field is given by:

$$\frac{E_t}{E_x} = \frac{C_t}{C_t^{\infty}}\left[1 + \frac{\mu_l + \mu_0}{\mu_t + \mu_0} e^{\frac{\Delta E}{\phi_0}}\right] \tag{5.23}$$

Using the above relation and integrating Eq. 5.18 for $i = t$, concentration profile of TE can be expressed through the hypergeometric series

$$\frac{C_t(x)}{C_t^{\infty}} = F\left[1, \frac{E_t}{\Delta E} 1 + \frac{E_t}{\Delta E}, \frac{\mu_l + \mu_0}{\mu_t + \mu_0} e^{\frac{\Delta E}{\phi_0}}\right], \tag{5.24}$$

and the concentration profile for LE can be obtained by using the Eq. 5.21.

In order to check the time dependency of the concentration profile for planer ITP, the computed result for the logarithmic ratio of the concentrations of LE and TE is compared with the analytical solution given by Eqs. 5.24 and 5.20. The results show (Fig. 5.2) that the concentration distribution becomes steady after a short transition.

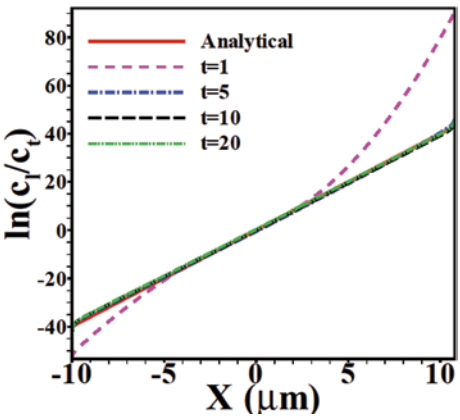

Fig. 5.2 Comparison of logarithmic ratio of TE and LE concentrations with analytical solution showing how the steady state is achieved ($\mu_l = \mu_0 = 2.71\times10^{-8} m^2$ V/s, $\mu_t = 0.7\mu_l$, $E_0 = 10^5$ V/m, $l_l = l_t = 0.5$, $C_l^{\infty} = 0.001M$, $j_0 = 445.03 Amp/m^2$)

Here the solid line represents the analytical result given in Eq. 5.20 and the dotted lines are computed solutions for different nondimensional time. If the width of the sample zone is large compared with the transition zone width (plateau-mode ITP), the variables near the interface can be estimated by the analytical expressions (Eqs. 5.20, 5.23, and 5.24).

To test the accuracy of our algorithm for 2-D ITP, we make a comparison of ionic concentration and the electric field at different values of LE-to-TE mobility ratio (D_r) in Figs. 5.3a and b. Our computed results agree well with the analytical predictions. The electric field gradually decreased as we move from low conductivity TE to high conductivity LE. The jump in voltage between neighboring zones results in the formation of permanently sharp boundaries between the TE and LE and are typical for ITP (Fig. 5.3a). The sharp gradient in E_x, which develops in the transition zone, smear out as the mobility ratio (or diffusivity ratio, D_r) of the electrolytes increases. Figure 5.3b shows that low mobility ratio leads to a sharp transition zone, whereas a high mobility ratio widens the transition zone width and the lateral distribution of E_x is relatively smooth. We consider the zone length occupied by the sample species negligible compared with the channel length, that is, $l_s \to 0$ with $l_l = l_t = 0.5$. Figure 5.4a and b shows that our numerical code can successfully resolve the sharp interfaces between successive zones in both peak and plateau mode ITP.

In a plateau mode ITP (Fig. 5.4a), the concentration distribution in each stack on either side of an interface can be treated independently, that is, consisting of two electrolytes say, TE and LE; whereas in peak mode (Fig. 5.4b), the transition zone width is of the order of the sample zone width.

Figure 5.5a through c shows the concentration profile in peak mode ITP for different values of sample-to-LE mobility ratio k_2 when LE-to-TE mobility ratio is fixed $k_1 = 3$. As the sample amount occupies a negligible portion of a long channel, the local electric field outside the sample zone is not influenced by the presence of the sample ions. We find from these results that the analyte penetrates more toward the LE (or TE) side, if the analyte mobility is near LE (or TE). The tails of the

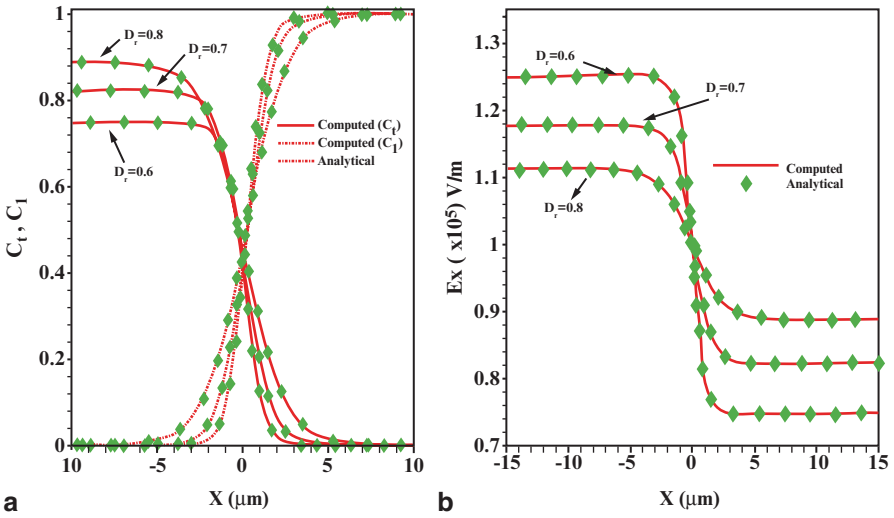

Fig. 5.3 Comparison of our computed results with the analytical results for ideal ITP for different diffusivity ratio values Dr ($D_l = D_0 = 7\times10^{-10} m^2/s$, $E_0 = 10^5 V/m$, $H = 25\mu m$ with $D_r = 0.6,\ 0.7,\ 0.8$ **a** concentration of TE (c_t) and LE (c_l), **b** axial electric field (E_x))

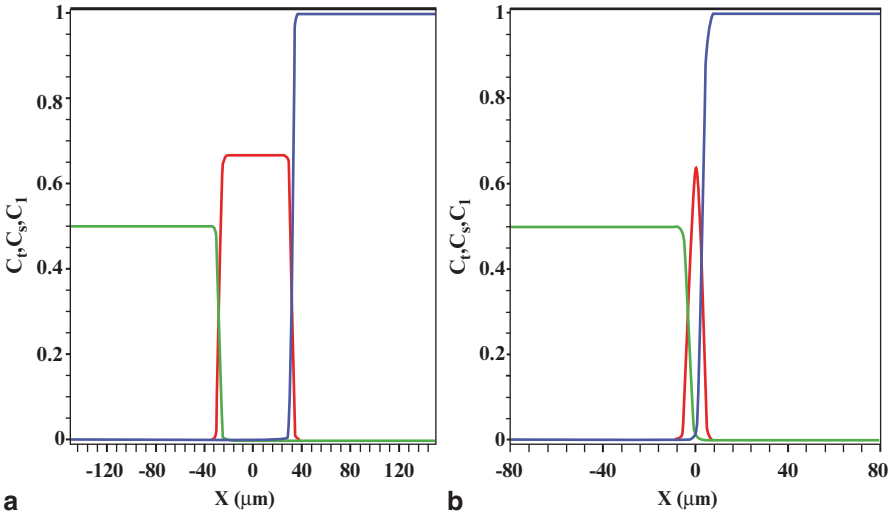

Fig. 5.4 Distribution of sample ions in the micro-channel ($H = 25\mu m$ for $D_l = D_0 = 7\times10^{-10}$ m^2/s $D_s = D_l/2$, $D_t = D_l/3$ when initial molar mass **a** $M_s = 40\mu m \times HC_l^{\infty}$, **b** $M_s = 4\mu m \times HC_l^{\infty}$. The bulk concentration of LE $C_l^{\infty} = 0.001M$, electric field strength $E_0 = 10^5 V/m$, and current density $j_0 = 270.2 Amp/m^2$)

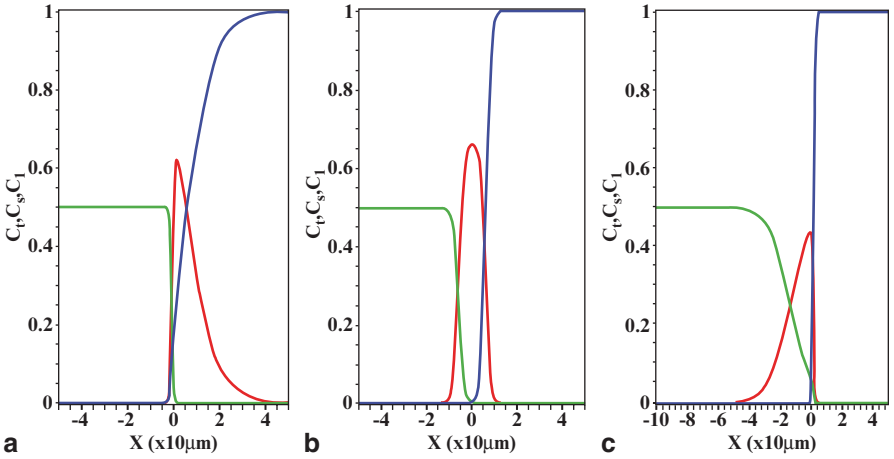

Fig. 5.5 Profiles of ionic concentration in peak mode ITP for fixed $k_1(=\mu_l/\mu_t)=3$ (**a** $k_2(=\mu_l/\mu_s)=1.1$, **b** $k_2(=\mu_l/\mu_s)=2$, **c** $k_2(=\mu_l/\mu_s)=2.9$, channel height $H=25\mu m$, bulk concentration of LE $C_l^{\infty}=0.001M$, electric field strength $E_0=10^5 V/m$, current density $j_0=270.2 Amp/m^2$)

sample distribution lies in the region where the electric field is relatively uniform. A symmetry pattern for analyte distribution results when its mobility is close to an average of the LE and TE mobilities.

5.3.2 *Effect of Convection on Sample Zone in ITP*

We investigate the impact of convective dispersion due to uniform pressure-driven flow and/or EOF on the ion distribution in ITP. The dispersion of the sample zone sandwiched between two adjacent electrolytes, namely LE and TE can be measured through the variance of the concentration of sample species. We define the axial standard deviation as $\sigma=\sqrt{Var(c_s(x,y))}$, where $\mathrm{Var}(c_s(x,y))=\frac{1}{M_s}\iint_D (x-M_x)^2 c_s(x,y)dxdy$ with $M_x=\frac{1}{M_s}\iint_D xc_s(x,y)dxdy$ and the molar mass present in the system $M_s=\iint_D c_s(x,y)dxdy$. The strong dispersion due to convection may produce an asymmetry in the sample distribution. For this, we have estimated the skewness of the sample zone. We define skewness of the concentration profile of the sample species as:

$$\gamma_1=\frac{1}{M_s}\iint_D\left(\frac{x-M_x}{\sigma}\right)^3 c_s(x,y)dxdy.$$

The value of skewness may be positive, negative, or zero depending on whether mass are skewed to the left (negative skew) or symmetric (zero skew) or to the right (positive skew) of average location of the mass. Qualitatively, for a mass distribution having negative skew indicates that the tail on the left side is longer than the right side, and vice versa for positive skew. Zero skew indicates that the mass are distributed in a symmetrical manner.

Effect of Pressure Driven Flow

We consider a constant pressure difference along the axial direction of the channel. Due to the imposed pressure gradient, the fully developed Poiseuille flow along the electromigration direction of the ionic species with average flow $\overline{u}$ develops; that is,

$$u(y) = 6\overline{u}\,\frac{y(1-y)}{H^2}, \tag{5.25}$$

where H is the depth of the channel and U^{ITP} is the velocity of ITP transport for ideal ITP case. Under the bulk fluid flow with average speed $\overline{U}$, the transition zone velocity becomes $\overline{U} = U^{ITP} + \overline{u} = (\alpha+1)U^{ITP}$, where $\alpha = 0$ corresponds to the ideal ITP case.

Figure 5.6a shows the concentration distribution of sample within the channel for $k_1 = 3$, $k_2 = 2$ when the bulk fluid flows with average speed $\overline{u} = 3.24\times10^{-3}\,m/s$ along the favorable direction of ITP. In this case, the ITP zone moves at a faster rate compared with the plane ITP. Though the distribution of the sample species and TE in ideal ITP is *y*-independent, it becomes 2-D when a convection is considered. The electromigration leads to sample stacking, while diffusion works against it. With convection along the ITP direction, the ITP zone widens and smears the sharpness of the electric field and ionic concentration. The combined effect of convection and lateral diffusion would result in a dispersion of the species concentration. Dispersion effect on the transition zone is also evident from the distribution of conductivity (Fig. 5.6b).

The variation of standard deviation (σ) and skewness (γ_1) of the dispersed sample zone by changing its electrophoretic mobility for fixed TE and LE is presented in Figs. 5.7a and b. It is clear from Fig. 5.7a that the sample dispersion is large when its mobility is close to the mobility of one of the adjacent electrolytes. If the mobility of the sample species is close to the harmonic mean of adjacent electrolytes, the dispersion will be minimum and the skewness of the concentration distribution of the sample species is close to zero (Fig. 5.7b). The dispersed sample zone deviates from Gaussian profile when the mobility of sample species is close to LE or TE.

Dispersion Due to EOF of Electrolytes

If the Debye layer thickness is considered to be much lower than the height of the channel, then the slip velocity condition can be assumed on the channel wall.

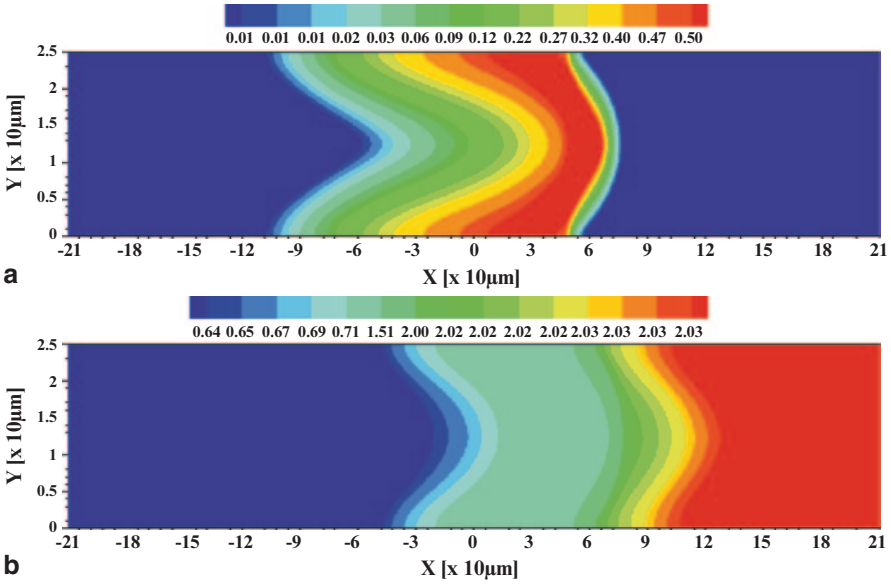

Fig. 5.6 Distribution of **a** concentration of sample species, and **b** conductivity under bulk fluid flow (average speed $\bar{u} = 3.24 \times 10^{-3} m/s$ along the favorable direction of ITP $k_1 = 3, k_2 = 2$, bulk concentration of LE $C_l^{\infty} = 0.001M$, applied electric field strength $E_0 = 10^5 V/m$, channel depth $H = 25\mu m$, and the initial molar mass of the sample species $i\, M_s = 40\mu m \times HC_l^{\infty}$ Mol/m)

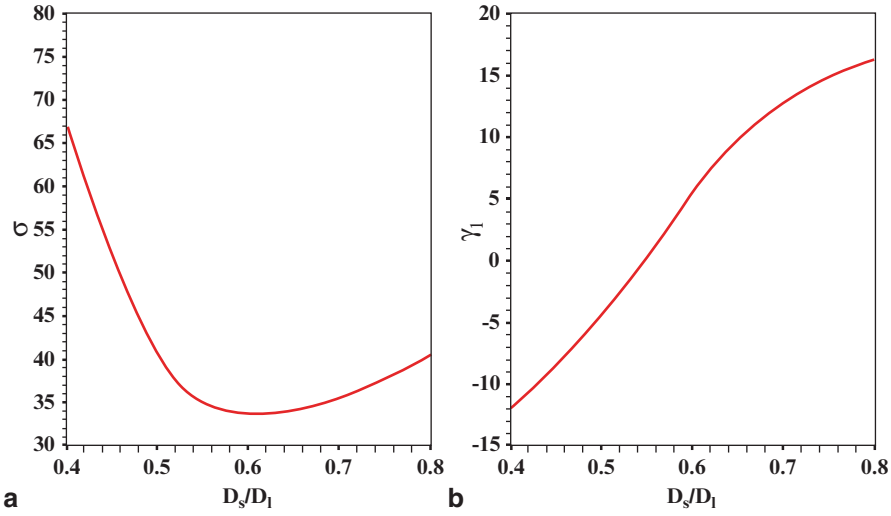

Fig. 5.7 Variation of **a** standard deviation (σ), and **b** skewness (γ_1) of the sample distribution for fixed value of $D_t = D_l / 3$ and $D_l = D_0 = 7 \times 10^{-10} m^2/s$ under bulk fluid flow with average speed $\bar{u} = 3.24 \times 10^{-3} m/s$ along the favorable direction of ITP (channel height $H = 25\mu m$, bulk concentration of LE $C_l^{\infty} = 0.001M$, electric field strength $E_0 = 10^5 V/m$)

We denote the electro-osmotic slip velocity along the wall for LE, sample, and TE by u_l^{EOF}, u_s^{EOF}, and u_t^{EOF}, respectively. The slip velocity u_i^{EOF} is determined by the Smoluchowski's equation $u_i^{EOF} = \varepsilon_e \zeta_i E_i / \eta = \mu_i^{EOF} E_i$, where ε_e is the dielectric permittivity of the medium, ζ_i is the zeta-potential, and μ_i^{EOF} is the electro-osmotic (EO) mobility of the ith species. We define the ratio of the EO slip velocities between LE and TE by $v_r = u_t^{EOF} / u_l^{EOF}$. If $u_l^{EOF} = u_s^{EOF} = u_t^{EOF}$, then the ITP transition zone is unperturbed due to EOF. The difference in EO slip velocity of the electrolytes induces axial pressure gradient. Sufficiently far away from the transition zone, the velocity of the electrolytes considered to be the superposition of EOF velocity and Poiseuille flow due to a constant pressure gradient [18] as:

$$u_i(y) = u_i^{EOF} + 6u_{p,i} \frac{y(1-y)}{H^2}, \tag{5.26}$$

where $6u_{p,i}$ is the average flow due to a constant pressure gradient on i-th zone.

If the net pressure drops across the length L is zero, then $u_{p,l} l_l + u_{p,s} l_s + u_{p,t} l_t = 0$ along with the conservation of mass leads to $u_{p,i} = (\bar{u} - u_i^{EOF})$ and the velocity at the individual zone can be written as:

$$u(y) = (\bar{u} - u_i^{EOF}) \frac{y(1-y)}{H^2} + u_i^{EOF}, \tag{5.27}$$

where the average electro-osmotic slip velocity is:

$$\bar{u} = l_l u_l^{EOF} + l_s u_s^{EOF} + l_t u_t^{EOF}. \tag{5.28}$$

The slip velocity along the wall is:

$$u^{EOF}(x; y = 0, H) = \mu^{EOF}(x) E(x; y = 0, H), \tag{5.29}$$

where $\mu^{EOF}(x)$ and $E(x; y = 0, H)$ are the wall mobility and axial electric field on the wall at different axial positions x, respectively. The approximate wall mobility is given by [15]:

$$\mu^{EOF}(x) = \frac{\sum \mu_i^{EOF} C_i(x; y = 0, H)}{\sum C_i(x; y = 0, H)}, \tag{5.30}$$

where $C_i(x; y = 0, H)$ are the concentration of the i-th ionic species on the wall. The bulk electro-osmotic mobilities of LE, sample, and TE are denoted by μ_l^{EOF}, μ_s^{EOF}, and μ_t^{EOF}, respectively.

The distribution and dispersion of the sample electrolytes sandwiched between LE and TE depend on the choice of the ratio of its electro-osmotic mobility with the adjacent electrolytes. A large dispersion in sample distribution occurs when the electro-osmotic mobility of the sample species deviates by a large margin from the adjacent electrolytes (LE or TE). As the electric field varies in each zone, an equal EO mobility of sample and the adjacent electrolytes lead to a nonuniform EO speed.

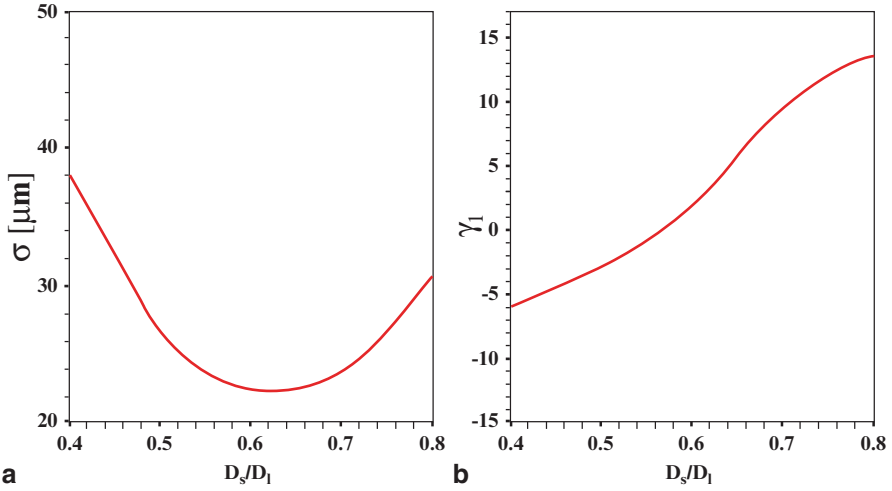

Fig. 5.8. Variation of **a** standard deviation (σ) and **b** skewness (γ_1) of the sample distribution for fixed value of $D_t = D_l / 3$ and $D_l = D_0 = 7 \times 10^{-10} m^2 / s$ by changing the diffusivity of sample species (EO mobility of TE, sample, and LE species are taken to be same as $\mu_l^{EOF} = \mu_s^{EOF} = \mu_t^{EOF} = 3.59 \times 10^{-8} m^2 / Vs$, channel height $H = 25 \mu m$, bulk concentration of LE $C_l^{\infty} = 0.001 M$, electric field strength $E_0 = 10^5 V / m$)

Figure 5.8 a and b shows the standard deviation (σ) and skewness (γ_1), respectively as a function of the sample diffusivity D_s when the diffusivity of both TE and LE are kept fixed as D_r. Here, we have taken the EO motilities of all the three electrolytes as equal. We find that the stack width is minimum and the skewness is almost zero when the diffusivity of the sample is close to the average of D_t and D_t. The sample distribution shows asymmetry and the variance is large when the diffusivity of the sample is either close to LE or TE. However, the dispersion is smaller when sample diffusivity is close to LE $(D_s / D_l > 0.5)$ compared with the case when it is close to TE, that is, $D_s / D_l < 0.5$. It may be noted that as the sample diffusivity changes, the electric field and hence the electro-osmotic velocity varies. Thus, the dependence of dispersion on diffusivity for EOF interacted ITP is qualitatively different from that of the case where ITP is interacted by a constant pressure gradient.

5.4 Summary

A high resolution numerical algorithm is developed to analyze the 2-D ITP of electrolytes of different mobility in a wide microchannel. The model is based on equations for conservation of mass, charge, and electroneutrality condition. In addition, the convective transport of electrolytes is governed by the Navier–Stokes equations. The present numerical algorithm is based on a finite volume method over a staggered grid arrangement along with a higher-order upwind scheme, QUICK. In ITP,

a thin region (transition zone) between two adjacent analytes develops in which either a step change (plateau mode) or steep gradient (peak mode) of variables may occur. Presented numerical method can efficiently capture these sharp boundaries between adjacent anlaytes. In a peak-mode ITP, depending on the mobility ratio of the analytes, the distribution of analytes may show asymmetry. The imposed convection speeds up the ITP separation, however, the transition zone widens and the sharpness of the variables across the transition zone smears- out. The distribution of analytes, in plateau shape in absence of convection, shows a peak when imposed convection is considered. Dispersion of analytes is measured through second- and third-order moment analysis.

References

1. Li SFY (1992) Capillary electrophoresis: principles, practice, and applications. Elsevier, Amsterdam
2. Kuhn R, Hoffstetter-Kuhn S (1993) Capillary electrophoresis: principles and practice. Springer, Berlin
3. Camilleri P (1993) Capillary electrophoresis: theory and practice. CRC, Boca Raton
4. Landers JP (1994) Handbook of capillary electrophoresis. CRC, Boca Raton
5. Chen L, Prest JE, Fielden PR, Goddard NJ, Manz A, Day PJR (2006) Miniaturized isotacophoresis analysis. Lab-on-a-Chip 6:474–487
6. Gebauer P, Mala Z, Bovcek P (2011) Recent progress in analytical capillary isotachophoresis. Electrophoresis 32:83–89
7. Khurana TK, Santiago JG (2008) Preconcentration, separation, and indirect detection of nonfluorescent analytes using fluorescent mobility markers. Anal Chem 80:279–286
8. Garcia-Schwarz G, Bercovici M, Marshall LA, Santiago JG (2011) Sample dispersion in isotachophoresis. J Fluid Mech 679:455–475
9. Kohlrausch F (1897) Uber Concentrations-Verschiebungen durch Electrolyse im Inneren von Losungen und Losungsgemis. Ann Physik 62:209–239
10. Saville DA, Palusinski OA (1986) Theory of electrophoretic separations, Part I: Formulation of a mathematical model. AlChE J 32:207–214
11. Su Y, Palusinski OA, Fife PC (1987) Isotachophoresis: analysis and computation of the structure of the ionic species interface. J Chromatogr 405:77–85
12. Bercovici M, Lelea SK, Santiago JG (2009) Open source simulation tool for electrophoretic stacking, focusing, and separation. J Chromatogr A 1216:1008–1018
13. Thormann W, Breadmore MC, Caslavska RA, Mosher RA (2010) Dynamic computer simulations of electrophoresis: A versatile research and teaching too. Electrophoresis 31:726–754
14. Bhattacharyya S, Zheng Z, Conlisk AT (2005) Electroosmotic flow in two-dimensional charged micro- and nano-channels. J Fluid Mech 540:247–267
15. Baier T, Schonfeld F, Hardt S (2011) Analytical approximations to the flow field induced by electroosmosis during isotachophoretic transport through a channel. J Fluid Mech 682:101–119
16. Shim J, Cho M, Dutta P (2011) A method to determine quasi-steady state in constant voltage mode isotachophoresis. Electrophoresis 32:988–995
17. Saville DA (1990) The effects of electroosmosis on the structure of isotachophoresis boundaries. Electrophoresis 11:899–902
18. Schonfeld F, Goet G, Baier T, Hardt S (2009) Transition zone dynamics in combined isotachophoretic and electro-osmotic transport. Phys Fluids 21:092002

19. Bercovici M, Lelea SK, Santiago JG (2010) Compact adaptive-grid scheme for high numerical resolution simulations of isotachophoresis. J Chromatogr A 1217:588–599
20. Bhattacharyya S, Gopmandal PP, Baier T, Hardt S (2013) Sample dispersion in isotachophoresis with Poiseuille counter flow. Phys Fluids 25:022001–022015
21. Fletcher CAJ (1998) Computational methods for fluid dynamics, vol 2. Springer, Berlin
22. Jaluria Y (2003) Computational heat transfer. Taylor & Francis, New York
23. Leonard BP (1979) A stable and accurate convective modelling procedure based on quadratic upstream interpolation. Comput Meth Appl Mech Eng 19:59–98

Chapter 6
Thermal Characterization of Nonhomogeneous Media

Diego C. Knupp, Henrique Massard Fonseca, Carolina P. Naveira-Cotta, Helcio R. B. Orlando, Renato M. Cotta and Olivier Fudym

6.1 Introduction

Inverse heat transfer problems deal with the estimation of unknown quantities appearing in the mathematical formulation of physical processes in thermal sciences by using measurements of temperature, heat flux, radiation intensities, etc. Originally, inverse heat transfer problems have been associated with the estimation of an unknown boundary heat flux by using temperature measurements taken below the boundary surface of a heat-conducting medium [1]. On the other hand, recent technological advancements often require the use of involved experiments and indirect measurements within the research paradigm of inverse problems. Nowadays, inverse analyses are encountered in single- and multi-mode heat transfer problems dealing with multi-scale phenomena. Applications range from the estimation of constant heat transfer parameters to the mapping of spatially and timely varying functions, such as heat sources, fluxes, and thermophysical properties. Systematic methods for the solution of inverse problems have developed significantly during the last 20 years and have become a powerful tool for analysis and design in engineering [1–12]. Inverse analysis is nowadays a common practice in which the groups involved with experiments and numerical simulation synergistically collaborate throughout the research work, in order to obtain the maximum information

R. M. Cotta (✉) · D. C. Knupp · H. M. da Fonseca · C. P. Naveira-Cotta · H. R. B. Orlando
Department of Mechanical Engineering, Escola Politécnica/COPPE, Federal University of Rio de Janeiro, UFRJ, Cid. Universitaria,
Cx. Postal 68503, Rio de Janeiro, RJ, 21941–972, Brazil
e-mail: cotta@mecanica.coppe.ufrj.br

H. R. B. Orlando
e-mail: helcio@mecanica.ufrj.br

O. Fudym
Centre RAPSODEE, Université de Toulouse; Mines Albi; CNRS, Campus Jarlard, 81013, Albi, France
e-mail: fudym@mines-albi.fr

S.K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_6,

regarding the physical problem under study. Similarly, powerful measurement techniques have been undergoing significant advancements, such as the use of infrared (IR) cameras for temperature measurements. Such measurement technique is quite powerful, because it can provide accurate nonintrusive measurements, with fine spatial resolutions and at large frequencies [10].

The mapping of thermophysical properties of nonhomogeneous materials from thermal images provided by an IR camera is a difficult inverse problem, due to the large amount of data to be processed and large number of parameters to be estimated, as well as the low signal-to-noise ratio. It is thus of great interest to implement estimation approaches of low computational cost that can accurately cope with such inherent difficulties [13–25].

In this chapter, we make use of the so-called statistical inversion approach for the solution of an inverse problem that involves the identification of nonhomogeneities or inclusions in a heat-conducting media, through the identification of its spatially varying thermophysical properties. The statistical inversion approach is based on the following principles [7]:

1. All variables appearing in the mathematical model of the physical problem are modeled as random variables.
2. The randomness describes the degree of information concerning their realizations.
3. The degree of information concerning these values is coded in probability distributions.
4. The solution of the inverse problem is the posterior probability distribution.

The statistical inversion approach falls within the Bayesian statistical framework, in which (probability distribution) models for the measurements and the unknowns are constructed separately and explicitly. The solution of the inverse problem is recast in the form of statistical inference from the posterior probability density, which is the model for the conditional probability distribution of the unknown parameters given the measurements. The measurement model incorporating the measurement error model and the related uncertainties is called the likelihood, i.e., the conditional probability of the measurements given the unknown parameters. The model for the unknowns that reflects all the uncertainty of the parameters without the information conveyed by the measurements is called the prior model [7, 9–12, 26–29].

We briefly present below the statistical approach for the solution of inverse problems, as well as the Metropolis–Hastings algorithm (MH) for the implementation of the Markov chain Monte Carlo (MCMC) method [7, 9, 11, 12, 26–29]. Such method is then applied to the solution of the inverse problem of estimating the spatial variation of thermophysical properties. The MCMC method is used in conjunction with two different techniques in order to obtain low-cost inverse problem solutions, namely: (1) A nodal approach, as advanced in [13–17] and (2) An eigenfunction expansion of the unknown properties and solution of the direct problem with the Generalized Integral Transform Technique (GITT), as advanced in [18–23].

6.2 Statistical Approach for the Solution of Inverse Problems

Consider the vector of parameters appearing in the physical model formulation as:

$$\mathbf{P}^T \equiv [P_1, P_2, ..., P_N], \tag{6.1a}$$

where

N is the number of parameters.

For the solution of the inverse problem of estimating **P,** we assume availability of the measured temperature data as given by:

$$\mathbf{Y}^T = (\vec{Y}_1 \ , \ \vec{Y}_2 \ , ... , \vec{Y}_i) \ , \tag{6.1b}$$

where

$\vec{Y}_i$ contains the measured temperatures for each of the S sensors at time t_i, $i = 1,\ldots, I$,

i.e.,:

$$\vec{Y}_i = (Y_{i1} \ , \ Y_{i2} \ , ... , Y_{iS}) \qquad \text{for } i = 1,\ldots, I, \tag{6.1c}$$

so that we have $M = S\ I$ measurements in total. Bayes' theorem can then be stated as [7, 9–12, 26–29]:

$$\pi_{\text{posterior}}(\mathbf{P}) = \pi(\mathbf{P}|\mathbf{Y}) = \frac{\pi_{\text{prior}}(\mathbf{P})\pi(\mathbf{Y}|\mathbf{P})}{\pi(\mathbf{Y})}, \tag{6.2}$$

where

$\pi_{\text{posterior}}(\mathbf{P})$ is the posterior probability density, i.e., the conditional probability of the parameters **P** given the measurements **Y**;
$\pi_{\text{prior}}(\mathbf{P})$ is the prior density, i.e., the coded information about the parameters prior to the measurements;
$\pi(\mathbf{Y}|\mathbf{P})$ is the likelihood function, which expresses the likelihood of different measurement outcomes **Y** with **P** given; and
$\pi(\mathbf{Y})$ is the marginal probability density of the measurements, which plays the role of a normalizing constant.

If we assume the parameters and the measurement errors to be independent Gaussian random variables, with known means and covariance matrices, and that the measurement errors are additive, a closed form expression can be derived for the $\pi_{\text{posterior}}(\mathbf{P})$. In this case, the likelihood function can be expressed as [7, 9–12, 26–29]:

$$\pi(\mathbf{Y}|\mathbf{P}) = (2\pi)^{-M/2} |\mathbf{W}|^{-1/2} \exp\left[-\frac{1}{2}[\mathbf{Y} - \mathbf{Y}_e(\mathbf{P})]^T \mathbf{W}^{-1}[\mathbf{Y} - \mathbf{Y}_e(\mathbf{P})]\right], \tag{6.3}$$

where

$\mathbf{Y}e$ is the vector of estimated variables, obtained from the solution of the direct (forward) model with an estimate for the parameters **P,**
$M = S\,I$ is the number of measurements, and
W is the covariance matrix of the errors in **Y.**

Similarly, for the case involving a prior normal distribution for the parameters we can write:

$$\pi(\mathbf{P}) = (2\pi)^{-N/2}\left|\mathbf{V}\right|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{P}-\boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{P}-\boldsymbol{\mu})\right], \tag{6.4}$$

where

$\boldsymbol{\mu}$ and **V** are the known mean and covariance matrix for **P,** respectively.

By substituting Eqs. 6.3 and 6.4 into Bayes' theorem, Eq. 6.2, except for the normalizing constant in the denominator we obtain:

$$\ln\left[\pi(\mathbf{P}\mid\mathbf{Y})\right] \propto -\frac{1}{2}\left[(M+N)\ln 2\pi + \ln\mid\mathbf{W}\mid + \ln\mid\mathbf{V}\mid + S_{MAP}(\mathbf{P})\right], \tag{6.5}$$

where:

$$S_{\mathrm{MAP}}(\mathbf{P}) = \left[\mathbf{Y}-\mathbf{Y}_e(\mathbf{P})\right]^T\mathbf{W}^{-1}\left[\mathbf{Y}-\mathbf{Y}_e(\mathbf{P})\right] + (\mathbf{P}-\boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{P}-\boldsymbol{\mu}). \tag{6.6}$$

Equation 6.5 reveals that the maximization of the posterior distribution function can be obtained with the minimization of the objective function given by Eq. 6.6, denoted as the *maximum a posteriori objective function* [4–13]. Equation 6.6 clearly shows the contributions of the likelihood and of the prior distributions in the objective function, given by the first and second terms on the right-hand side, respectively.

On the other hand, if different *prior* probability densities are assumed for the parameters, the posterior probability distribution may not allow an analytical treatment such as that presented above. In this case, MCMC methods are used to draw samples of all possible parameters, so that inference on the posterior probability becomes inference on the samples [7, 9–12, 26–29].

In order to implement the Markov Chain, a density $q(\mathbf{P}^*, \mathbf{P}^{(t-1)})$ is required, which gives the probability of moving from the current state in the chain $\mathrm{P}^{(t-1)}$ to a new state **P***.

The MH [7, 9–12, 26–29] was used in this work to implement the MCMC method. It can be summarized in the following steps:

1. Sample a *candidate point* **P*** from a proposal distribution $q(\mathbf{P}^*, \mathbf{P}^{t-1})$.
2. Calculate:

$$\alpha = \min\left[1, \frac{\pi(\mathbf{P}^*\mid\mathbf{Y})\,q(\mathbf{P}^{(t-1)},\mathbf{P}^*)}{\pi(\mathbf{P}^{(t-1)}\mid\mathbf{Y})\,q(\mathbf{P}^*,\mathbf{P}^{(t-1)})}\right]. \tag{6.7}$$

3. Generate a random value U which is uniformly distributed on (0, 1).
4. If $U \le \alpha$, define $\mathbf{P}^{(t)} = \mathbf{P}^*$; otherwise, define $\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)}$.
5. Return to step 1 in order to generate the sequence $\{\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \ldots, \mathbf{P}^{(n)}\}$.

In this way, we get a sequence that represents the posterior distribution and inference on this distribution is obtained from inference on the samples $\{\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \ldots, \mathbf{P}^{(n)}\}$. We note that values of $\mathrm{P}^{(i)}$ must be ignored until the chain has converged to equilibrium. For more details on theoretical aspects of the MH and MCMC methods, the reader should consult references [7, 9–12, 26–29].

6.3 Nodal Approach for Estimating Spatially Varying Thermal Diffusivity and Heat Source

The physical problem examined here involves two-dimensional transient heat conduction in a thin plate, with spatially varying thermal conductivity and volumetric heat capacity. Lateral boundaries are supposed to be insulated and partial lumping is used across the plate. Internal steady heat generation and surface convective heat losses are taken into account in the formulation. The initial temperature within the medium is nonuniform. The mathematical formulation for this problem is given by:

$$C(x,y)\frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left[k(x,y)\frac{\partial T}{\partial x}\right] + \frac{\partial}{\partial y}\left[k(x,y)\frac{\partial T}{\partial y}\right] - h(x,y)\left(T - T_\infty\right) + g(x,y),$$
$$0 < x < L_x,\ 0 < y < L_y,\ \text{for } t > 0 \qquad (6.8a)$$

$$\frac{\partial T}{\partial x} = 0 \qquad \text{at } x = 0 \text{ and } x = L_x,\ \text{for } t > 0, \qquad (6.8b)$$

$$\frac{\partial T}{\partial y} = 0 \qquad \text{at } y = 0 \text{ and } y = L_y,\ \text{for } t > 0, \qquad (6.8c)$$

$$T(x,y,0) = T_0(x,y) \qquad \text{for } t = 0,\ \text{in } 0 < x < L_x, 0 < y < L_y, \qquad (6.8d)$$

where

$C(x, y)$ is the local volumetric heat capacity
$k(x, y)$ is the local thermal conductivity
$h(x, y)$ is the local convective heat transfer coefficient divided by the thickness of the plate, and
$g(x, y)$ is the local heat source.

In the direct problem associated with the mathematical formulation of such physical problem, the thickness of the medium and the spatial variations of the volumetric heat capacity, thermal conductivity, heat transfer coefficient, heat source, and initial temperature are known. The objective of the direct problem is then to determine the transient temperature variation within the medium. On the other hand, the inverse problem is concerned with the identification of the spatially dependent thermal parameters, by making use of the nodal strategy advanced in [13, 14]. For the application of such strategy, we rewrite Eq. 6.8a in the following nonconservative form:

$$\frac{\partial T}{\partial t} = a(x,y)\nabla^2 T + \frac{1}{C(x,y)}\left[\frac{\partial k(x,y)}{\partial x}\frac{\partial T}{\partial x} + \frac{\partial k(x,y)}{\partial y}\frac{\partial T}{\partial y}\right] - H(x,y)\left(T - T_\infty\right) + G(x,y), \tag{6.9}$$

where

$$a(x,y) = \frac{k(x,y)}{C(x,y)} \qquad H(x,y) = \frac{h(x,y)}{C(x,y)} \qquad G(x,y) = \frac{g(x,y)}{C(x,y)}. \tag{6.10a–c}$$

An explicit discretization of Eq. 6.9 using finite differences results in:

$$Y_{i,j}^{n+1} = L_{i,j}^n a_{i,j} + Dx_{i,j}^n \delta_{i,j}^x + Dy_{i,j}^n \delta_{i,j}^y - \Delta t (T_{i,j}^n - T_\infty) H_{i,j} + \Delta t G_{i,j}, \tag{6.11}$$

where

the subscripts (i, j) denote the finite-difference node at $x_i = i\Delta x,\ i = 1 \ldots n_i$ and

$y_j = j\Delta y,\ j = 1 \ldots n_j,$ and

the superscript n denotes the time $t_n = n\Delta t,\ n = 0 \ldots (n_t - 1).$

The other quantities appearing in Eq. 6.11 are given by:

$$Y_{i,j}^{n+1} = T_{i,j}^{n+1} - T_{i,j}^n, \tag{6.12a}$$

$$L_{i,j}^n = \Delta t \left(\frac{T_{i-1,j}^n - 2T_{i,j}^n + T_{i+1,j}^n}{(\Delta x)^2} + \frac{T_{i,j-1}^n - 2T_{i,j}^n + T_{i,j+1}^n}{(\Delta y)^2} \right), \tag{6.12b}$$

$$Dx_{i,j}^n = \frac{\Delta t}{2\Delta x}(T_{i+1,j}^n - T_{i-1,j}^n) \text{ and } Dy_{i,j}^n = \frac{\Delta t}{2\Delta y}(T_{i,j+1}^n - T_{i,j-1}^n), \tag{6.12c, d}$$

$$\delta_{i,j}^x = \frac{1}{C(x,y)}\frac{\partial k}{\partial x} \text{ and } \delta_{i,j}^y = \frac{1}{C(x,y)}\frac{\partial k}{\partial y}. \tag{6.12e, f}$$

Equation 6.12a defines the forward temperature difference in time. Equation 6.12b approximates the Laplacian of temperature at time t_n and node (i, j). Equation 6.12c, d is a finite difference approximations to the temperature gradient components.

By writing Eq. 6.11 for a given node (i, j) and all time steps, we obtain:

$$\mathbf{Y}_{ij} = \mathbf{J}_{ij}\mathbf{P}_{ij}, \tag{6.13}$$

where

$$\mathbf{J}_{ij} = \begin{bmatrix} L_{i,j}^{0} & Dx_{i,j}^{0} & Dy_{i,j}^{0} & -\Delta t(T_{i,j}^{0} - T_{\infty}) & \Delta t \\ L_{i,j}^{1} & Dx_{i,j}^{1} & Dy_{i,j}^{1} & -\Delta t(T_{i,j}^{1} - T_{\infty}) & \Delta t \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ L_{i,j}^{n_t-1} & Dx_{i,j}^{n_t-1} & Dy_{i,j}^{n_t-1} & -\Delta t(T_{i,j}^{n_t-1} - T_{\infty}) & \Delta t \end{bmatrix}, \tag{6.14a}$$

and

$$\mathbf{Y}_{ij} = \begin{bmatrix} Y_{i,j}^{1} \\ Y_{i,j}^{2} \\ \vdots \\ Y_{i,j}^{n_t} \end{bmatrix} \quad \mathbf{P}_{ij} = \begin{bmatrix} a_{i,j} \\ \delta_{i,j}^{x} \\ \delta_{i,j}^{y} \\ H_{i,j} \\ G_{i,j} \end{bmatrix}. \tag{6.14b, c}$$

In the nodal strategy, the sensitivity matrix, Eq. 6.14a, is approximately computed with the measurements. Therefore, for the implementation of the MH the uncertainties in the computation of the sensitivity matrix need to be taken into account. By assuming that **P** and **J** are independent random variables, the sought $\pi_{\text{posterior}}(\mathbf{P})$ is then given by:

$$\pi(\mathbf{P}, \mathbf{J}|\mathbf{Y}) \propto \pi(\mathbf{Y}|\mathbf{P}, \mathbf{J})\pi(\mathbf{P})\pi(\mathbf{J}), \tag{6.15}$$

where

$\pi(\mathbf{J})$ is the *a priori* distribution for the sensitivity matrix **J.**

Since the sensitivity matrix is computed with the measurements, its prior distribution can be obtained from the distribution of the measurement errors.

We consider here a test case dealing with the identification of a crack (k=0.0257 Wm^{-1} K^{-1}, C=1211025 Jm^{-3} K^{-1}, a=2.12×10^{-8} m^2/s) within a homogeneous material (k=10 Wm^{-1} K^{-1}, C=3.75×10^6 Jm^{-3} K^{-1}, a=2.66×10^{-6} m^2/s) as depicted in Fig. 6.1a. Heat is generated with strength 5×10^6 W/m^3 so that G=1.33 K/s inside the square region shown in Fig. 6.1b, G=4.129 K/s inside the crack, and G=0 K/s outside the square region. For the MH, the prior distributions for the thermal diffusivity field $a(x, y)$ are assumed to be uniform within the interval [5×10^{-9}, 9×10^{-6}] m^2/s. Note that such interval encompasses most poor conductors

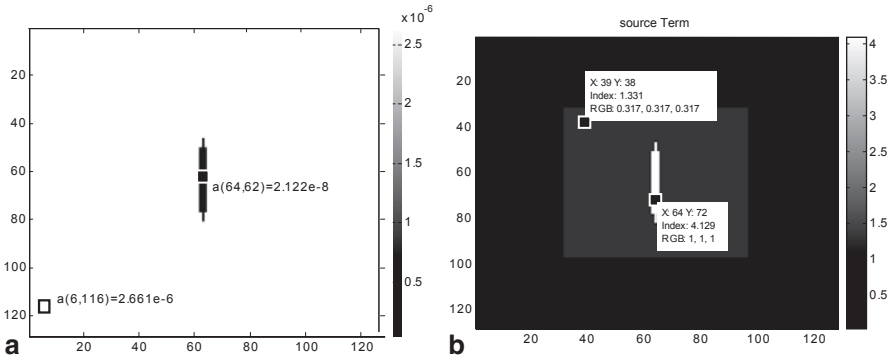

Fig. 6.1 **a** Exact thermal diffusivity distribution. **b** Exact distribution of the source term

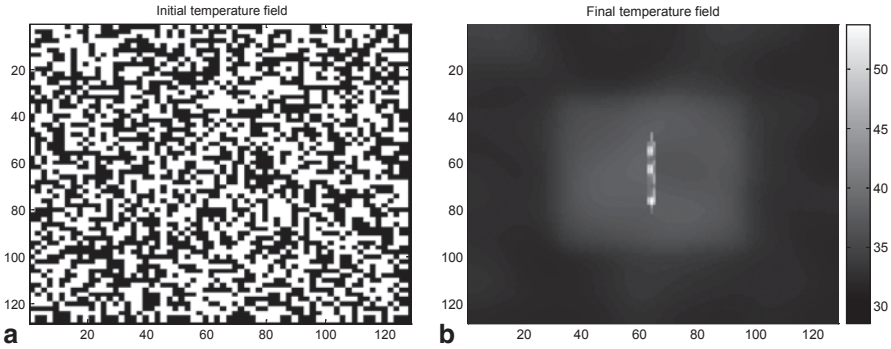

Fig. 6.2 **a** Initial temperature field. **b** Final temperature field

and does not provide any practical restriction to the values to be estimated for the parameters. The prior distributions for the heat source parameters are assumed to be uniform within the interval [0, 10] K/s. The plate is initially randomly excited with a pseudo random binary sequence (PRBS), hence the initial temperature field is randomly distributed as 20 °C or 40 °C. This distribution is chosen in order to enhance the thermal gradients within the plate, and consequently the local sensitivity coefficients. The temperature fields for $t=0$ s and $t=2$ s, are presented in Fig. 6.2a and b, respectively.

In order to avoid using the same direct solution for the generation of the simulated measurements and for the solution of the inverse problem, a finite volume solution is developed that yields the simulated data. For the solution of the inverse problem, the plate is discretized with $n_i = n_j = 128$ internal nodes, where the values of the parameters in $\mathbf{P}_{ij}$ are estimated. The space step is chosen according to the pixel size: $\Delta x = \Delta y = 200\mu\text{m}$. Therefore, the width and length of the plate are 0.0256 m. The simulated measurement errors are Gaussian, additive, uncorrelated, with zero mean, and a constant standard deviation of 0.03 °C.

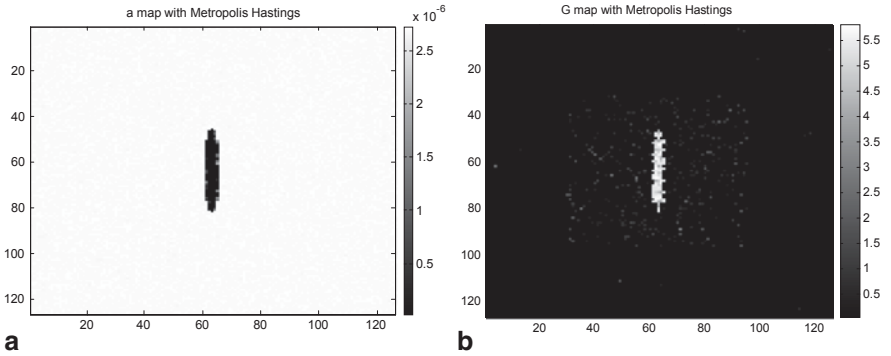

Fig. 6.3 **a** Estimated thermal diffusivity distribution. **b** Estimated distribution of the source term

The estimated thermal diffusivity and heat source term (G) are presented in Fig. 6.3a and b, respectively. The MH presented above is used for the estimation of these quantities in each of the nodal points. The Markov chain for each point consists of 1,500 states, where the first 500 is neglected for the computation of the statistics for the parameters. Generally, the acceptance ratio in the MH is of the order of 10%. A comparison of Figs. 6.1 and 6.2 reveals an excellent agreement between exact and estimated quantities, despite the fact that the measurements are directly used for the computation of the sensitivity matrix. Note that the above parameter estimation problem (Eq. 6.13) is linear. Therefore, its computation is very fast and appropriate for use with computationally intensive techniques like the MCMC method.

6.4 Identification of Thermophysical Properties of Nanocomposites

We consider in this example the following one-dimensional diffusion equation, with respective initial and boundary conditions:

$$w(x)\frac{\partial T(x,t)}{\partial t}=\frac{\partial}{\partial x}\left[k(x)\frac{\partial T(x,t)}{\partial x}\right]-d(x)T(x,t)+P(x,t),\quad 0<x<L,t>0, \tag{6.16a}$$

$$T(x,0)=f(x),\quad t=0,\quad 0<x<L, \tag{6.16b}$$

$$\alpha_0 T(x,t)-\beta_0 k(x)\frac{\partial T(x,t)}{\partial x}=\phi_0(x,t),\ x=0,\ t>0, \tag{6.16c}$$

$$\alpha_1 T(x,t)+\beta_1 k(x)\frac{\partial T(x,t)}{\partial x}=\phi_1(x,t),\ x=L,\ t>0. \tag{6.16d}$$

A formal exact solution of the direct problem given by Eq. 6.16 is achievable, accounting for the exact integral transformation when the specific eigenvalue problem with all the space variable coefficients is chosen, such as in [18–22]. This solution is based on the application of the Classical Integral Transform and the GITT [30–32], by writing:

$$T(x,t) = T_f(x;t) + \sum_{i=1}^{\infty} \tilde{\psi}_i(x)\overline{T}_i(t);\ \overline{T}_i(t) = \int_0^L w^*(x)\tilde{\psi}_i(x)T^*(x,t)dx; \qquad (6.17a, b)$$

where

$T_f(x;t)$ is a proposed analytical filtering solution, and
$\overline{T}_i(t)$ the transformed potentials in the above eigenfunction expansion, defined with the integral transformation operation above.

The eigenvalues μ_i and normalized eigenfunctions $\tilde{\psi}_i(x)$, are obtained from the chosen eigenvalue problem in each case, which specifies the weighting function to be adopted, $w^*(x)$, as described in further details in [18–22].

The coefficients $w(x)$, $k(x)$, and $d(x)$ are also expanded in terms of eigenfunctions in the following form, exemplified for $w(x)$:

$$w(x) = w_f(x) + \sum_{k=1}^{\infty} \tilde{\Gamma}_k(x)\overline{w}_k, \qquad \text{inverse} \qquad (6.18a)$$

$$\overline{w}_k = \int_0^{L_x} \hat{w}(x)[w(x) - w_f(x)]\tilde{\Gamma}_k(x)dx, \text{ transform} \qquad (6.18b)$$

where

$w_f(x)$ is a filtering solution used to enhance convergence of the expansion and
$\hat{w}(x)$ is the weighting function for the chosen normalized eigenfunction $\tilde{\Gamma}_k(x)$ [18–22].

For the solution of the inverse problem, the unknown quantities are the spatially dependent thermal properties and the effective heat transfer coefficients which are expressed as eigenfunction expansions, as well as the applied heat flux, which is parameterized as explained below. Thus, the total number of parameters N to be estimated is given by the sum of parameters in each expansion, including the number of parameters in each filter function, and the number of parameters in the heat flux expression.

By following the methodology advanced in [21], the solution of the inverse problem is obtained within the transformed temperature field. The experimental temperature data are integrally transformed by using Eq. 6.17b, thus compressing the experimental measurements in the spatial domain into a few transformed temperature modes and allowing for much faster computation of the inverse problem solution.

The experimental setup presented in Fig. 6.4a employs temperature measurements obtained from an IR camera FLIR SC660, with 640 × 480 image resolution, and −40 to 1,500 °C temperature range. The main components of the setup are: (1)

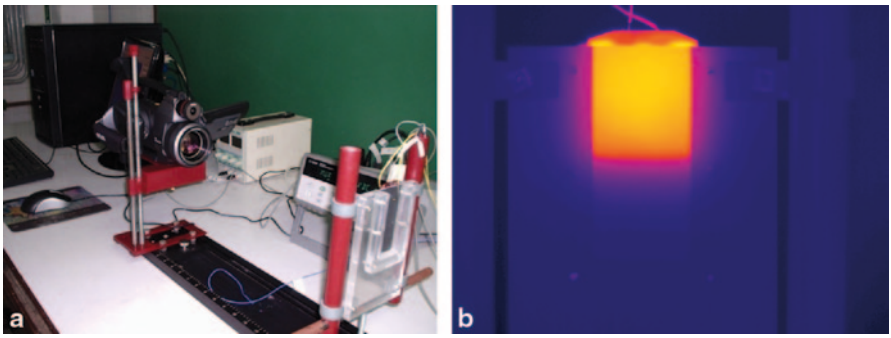

Fig. 6.4 **a** General view of the experimental setup. **b** Image produced by the FLIR SC660 camera after some elapsed heating time

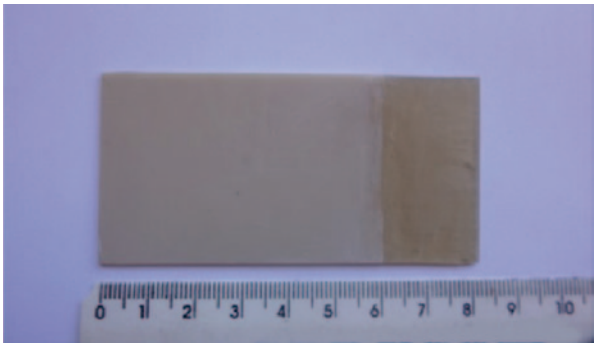

Fig. 6.5 Polyester resin-alumina nanocomposite plate used in the experiment with dimensions 1.51 mm (thickness), 40 mm (width), and 80 mm (length)

IR camera (FLIR SC660); (2) camera stand for vertical experiment configuration; (3) frame with the sample and the heater; (4) sample support; (5) data acquisition system (Agilent 34970-A); (6) microcomputer for data acquisition. Figure 6.4b illustrates an image produced by the FLIR SC660 camera after some elapsed heating time.

Figure 6.5 shows the nanocomposite plate used in this experiment, which is composed of polyester resin as matrix and alumina nanoparticles as filler, manufactured in such a way that three-fourth of the plate's length has 28.5 % of alumina nanoparticles in mass and the other one-fourth of the plate's length is composed only by polyester resin, with no addition of filler. The plate's thickness is 1.51 mm and its lateral and vertical dimensions are 40×80 mm. An electrical resistance (38.2 Ω) is employed for the heating of the plate, with the same lateral dimensions as that of the plate but half the length (40×40 mm), here joined at the upper half of the plate's height with the aid of a thermal compound paste. As a reference case, we use a pair of homogeneous polyester resin plates in a plate-heater-plate sandwich setup; this case is hereafter called Case 1. For the nanocomposite plate we use a plate-heater-insulation sandwich setup and investigate two cases by varying the position of the plate with respect to the electrical resistance: First the portion with no addition of

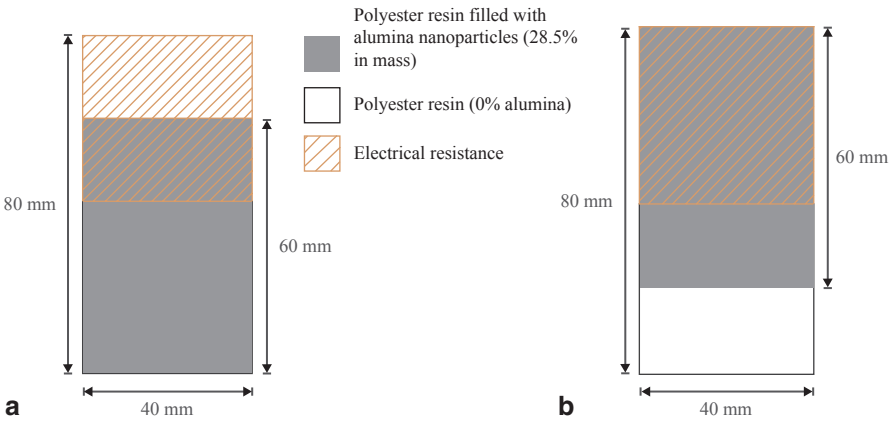

Fig. 6.6 Schematic representation of the experimental setup for **a** Case 2, and **b** Case 3

filler is placed in contact with the electrical resistance and for the last case the plate is turned upside down. These experimental setups are hereafter called Cases 2 and 3, respectively, and are schematically represented in Fig. 6.6. In order to reduce uncertainty in the IR camera readings, the plate surface that faces the IR camera is painted with a graphite ink, which brought its emissivity to $\varepsilon = 0.97$.

Considering a lumped formulation across the sample thickness, the direct problem is represented by averaging the multidimensional heat conduction equation, which results in the one-dimensional formulation with initial and boundary conditions given by Eq. 6.16, and with the following parameters:

$$d(x) = \frac{h_{eff}(x)}{L_z};\ P(x,t) = -\frac{h_{eff}(x)T_\infty}{L_z} + \frac{q_w(x,t)}{L_z};$$
$$f(x) = T_\infty;\ \alpha_0 = \alpha_1 = 0;\ \beta_0 = \beta_1 = 0;\ \phi_0 = \phi_1 = 0; \qquad (6.19a–f)$$

where

$h_{eff}(x)$ is the effective heat transfer coefficient,
$q_w(x,t)$ is the applied heat flux, and
L_z is the plate's thickness, respectively.

It has been verified that the appropriate identification of thermophysical properties through the proposed experimental setup requires simultaneous estimation of the time variation of the applied heat flux, resulting from the thermal capacitance of the heater, as well as of the heat transfer coefficient. For the time variation of the applied heat flux the following parameterization is employed:

$$q_w(x,t) = q(x)f(t) \qquad f(t) = c - ae^{-bt}. \qquad (6.20a, b)$$

Table 6.1 Prior information and estimated parameters for the thermal conductivity (Case 2—linear filter)

Parameter	Prior information	Estimates (mean values)
k_{x0} [W/m°C]	N (0.16, 15%)	0.1621
k_{xL} [W/m°C]	N (0.193, 15%)	0.201
$\overline{k_1}$	U (−0.0279, 0.0279)	0.00427344
$\overline{k_2}$	U (−0.00698, 0.00698)	0.0002019
$\overline{k_3}$	U (−0.00931, 0.00931)	−0.0007205
$\overline{k_4}$	U (−0.00349, 0.00349)	−0.0003482
$\overline{k_5}$	U (−0.00559, 0.00559)	−0.0004592
$\overline{k_6}$	U (−0.00232, 0.00232)	0.0002434
$\overline{k_7}$	U (−0.00399, 0.00399)	0.0001065

Table 6.2 Prior information and estimated parameters for the heat capacity (Case 2—linear filter)

Parameter	Prior information	Estimates (mean values)
w_{x0} [J/m³°C]	N (1.595×10^6)	1.601×10^6
w_{xL} [J/m³°C]	N (1.736×10^6)	1.74×10^6
$\overline{w_1}$	U (−255921.0, 255921.0)	17048.43
$\overline{w_2}$	U (−63980.3, 63980.3)	−1479.72
$\overline{w_3}$	U (−85307.0, 85307.0)	−1193.36
$\overline{w_4}$	U (−31990.1, 31990.1)	−3341.41
$\overline{w_5}$	U (−51184.2, 51184.2)	−477.34
$\overline{w_6}$	U (−21326.8, 21326.8)	668.28
$\overline{w_7}$	U (−36560.2, 36560.2)	−1002.42

In order to have reference values for the thermophysical properties of the polyester resin employed in this study, we perform the inverse analysis of Case 1 with the homogeneous pure polyester resin sample. In this case, we adopt normal priors with 15% standard deviation, centered in the literature values [33], obtaining the estimates: $k = 0.16 \pm 0.02$ W/m °C and $w = (1.57 \pm 0.08) \times 10^6$ J/m³ °C, for the thermal conductivity and heat capacity, respectively.

We now consider the inverse analysis for the estimation of the space varying properties of the nanocomposite plate with the setup of Case 2 (Fig. 6.6a). We assume the filter for the thermal conductivity and heat capacity in the form of linear functions that vary between the end values k_{x0} and k_{xL}, and w_{x0} and w_{xL}, respectively. For these parameters we adopt normal priors, with 15% standard deviation, centered in the literature values for the polyester resin [33] and in the Lewis-Nielsen's formula prediction [34] for the region filled with alumina nanoparticles. Seven terms are used for the expansions in Eq. 6.5 of $w(x)$ and $k(x)$; three terms are used in the expansion of $d(x)$. Tables 6.1, 6.2, 6.3 and 6.4 present the parameters

Table 6.3 Prior information and estimated parameters for the heat transfer coefficient (Case 2—linear filter)

Parameter	Prior information	Estimates (mean values)
h_{x0} [W/m²°C]	N (15.03, 5%)	14.286
h_{xL} [W/m²°C]	N (11.63, 5%)	12.041
$\overline{h}_1$	N (−0.23, 5%)	−0.2492
$\overline{h}_2$	N (0.69, 5%)	0.7197
$\overline{h}_3$	N (−0.044, 5%)	−0.0437

Table 6.4 Prior information and estimated parameters for the applied heat flux (Case 2—linear filter)

Parameter	Prior information	Estimates (mean values)
a	N (0.19, 5%)	0.197
b	N (0.00332, 5%)	0.00315
c	N (0.66, 5%)	0.64

and type of prior information, i.e., adopted in this inverse analysis (N=Gaussian distribution, U=noninformative uniform distribution) as well as the estimated mean values obtained for each one of the parameters.

Markov chains of $N_{\text{MCMC}} = 60,000$ states are used, and the statistics are computed by neglecting the first $N_{\text{burn-in}} = 20,000$ states needed for the warm up of the chains. For the sake of illustration, Fig. 6.7a and b show, respectively, the evolution of the Markov chains for the parameter k_{xL} and for the first coefficient in the expansion of the thermal conductivity, $\overline{k}_1$, where one can clearly observe the convergence of the states. Figure 6.8a and b present, respectively, the estimated spatial variations for the thermal conductivity and heat capacity, with their 99% confidence intervals, as well as the initial guess employed in this test case. One must observe that with the linear filter functions used for the thermophysical properties in this case, which does not provide informative priors regarding their functional forms, the proposed methodology is able to identify a transition of the space varying properties from the end values at $x = 0$ towards constant values. Figure 6.8a and b show that this transition is centered around $x = 0.02$ m, where, in fact, a sharp interface exists between the two materials that compose the plate. Figure 6.9a depicts the residuals between calculated and experimental quantities for the first five transformed temperature modes.

Although presenting some correlation, which results from the integral transformation procedure, i.e., truncated at the low order of 10 modes, the residuals are small. Similar behavior can be observed for the residuals in the temperature field, presented in Fig. 6.9b at three different positions. Some tests are performed with more than seven terms in the expansions of the sought coefficients, but no improvement is observed in the estimated quantities because of the small sensitivity coefficients concerning the higher-order expansion coefficients.

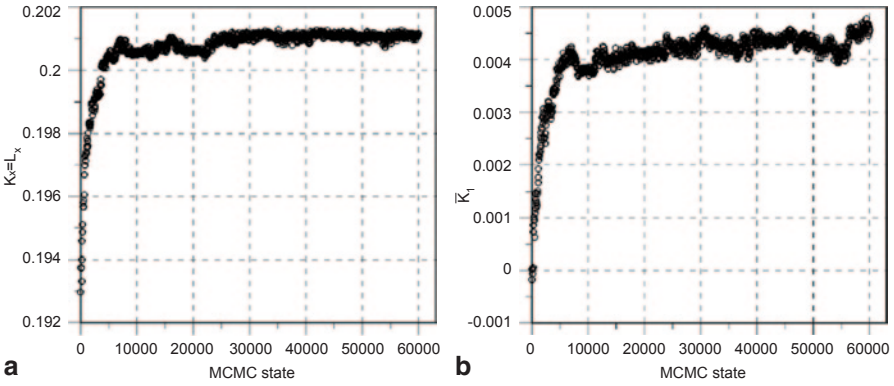

Fig. 6.7 Markov chain evolution of: **a** the parameter k_{xL} in the linear filter, **b** the first coefficient, $\bar{k}_1$, considering a linear filter

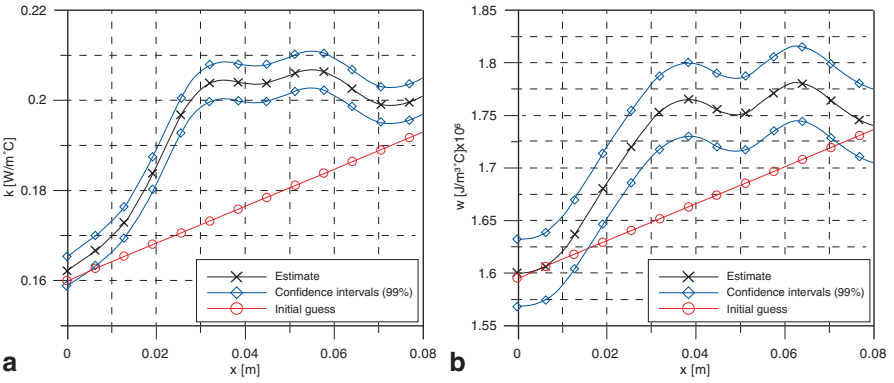

Fig. 6.8 Estimated **a** thermal conductivity and **b** heat capacity, with linear filter and 99 % confidence intervals

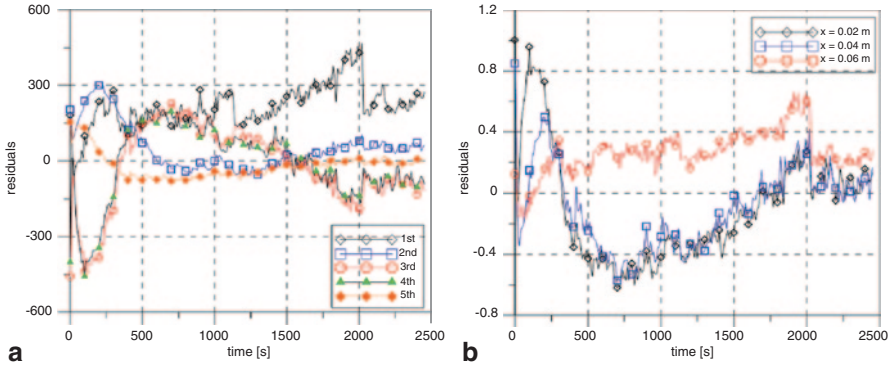

Fig. 6.9 Linear filter case: **a** residuals between calculated and experimental quantities for the first five transformed temperature modes, **b** residuals (°C) between calculated and experimental temperatures at three different positions

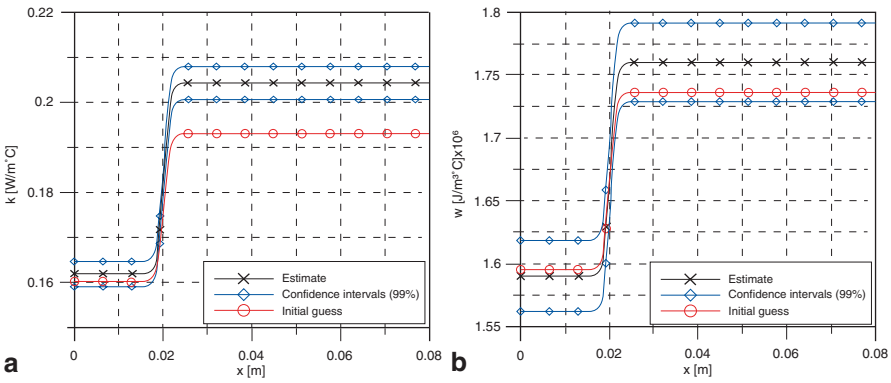

Fig. 6.10 Estimated **a** thermal conductivity and **b** heat capacity, with step function filter and 99 % confidence intervals

On the basis of the observations described above, we now consider a filter function for the thermophysical properties that approximates a step transition between the end values at $x=0$ and $x=L_x$, in the form:

$$k_f(x) = k_{x0} + (k_{xL} - k_{x0})\delta(x),\ w_f(x) = w_{x0} + (w_{xL} - w_{x0})\delta(x), \qquad (6.21a, b)$$

where

$$\delta(x) = \frac{1}{1 + e^{-\gamma(-x+x_t)}}. \qquad (6.21c)$$

In Eq. 6.21c, γ is a parameter that controls the transition sharpness and x_t is the transition point. Both are considered as fixed parameters, not to be estimated with the inverse analysis, the values of which are taken as $\gamma = 1500\,\mathrm{m}^{-1}$ and $= 0.02\,\mathrm{m}$ (based on the above observations with the linear filter function).

Figure 6.10a and b show the estimated spatial variations for the thermal conductivity and heat capacity obtained with the experimental arrangement of Case 2 (Fig. 6.6a), but with the filter function given by Eq. 6.21. The estimated 99 % confidence intervals and the employed initial guesses are also presented in these figures. Similar to the case examined above with the linear filter functions, the values used for the end parameters are obtained from normal priors, with 15 % standard deviation, centered in literature values for the polyester resin, and in Lewis–Nielsen's formula prediction [34] for the region filled with alumina nanoparticles.

In order to provide a better assessment on the estimates obtained, we now compare the experimental arrangements of Case 2 and Case 3 (Fig. 6.6). The approximate step filter function given by Eq. 6.21 is used for this comparison, with $\gamma = 1500\,\mathrm{m}^{-1}$ and x_t given by the transition point of materials in the manufactured plate. The remaining quantities, such as the number of expansion terms used in the eigenfunction approximations and the prior distributions, are the same as specified above. Table 6.5 presents the estimated thermophysical properties obtained with the

Table 6.5 Prior information and estimated thermophysical properties in the Cases 1, 2, and 3 (a step filter is used for Cases 1 and 2)

Property	Material	Prior	Estimates (mean values)
Case 1			
k [W/m°C]	Polyester	N (0.16, 15%)	0.159
w [J/m³°C]	Polyester	N (1.595×10^6, 15%)	1.566×10^6
Case 2			
k [W/m°C]	Polyester	N (0.16, 15%)	0.162
	Polyester+alumina	N (0.193, 15%)	0.204
w [J/m³°C]	Polyester	N (1.595×10^6, 15%)	1.59×10^6
	Polyester+alumina	N (1.736×10^6, 15%)	1.760×10^6
Case 3			
k [W/m°C]	Polyester	N (0.16, 15%)	0.149
	Polyester+alumina	N (0.193, 15%)	0.203
w [J/m³°C]	Polyester	N (1.595×10^6, 15%)	1.529×10^6
	Polyester+alumina	N (1.736×10^6, 15%)	1.743×10^6

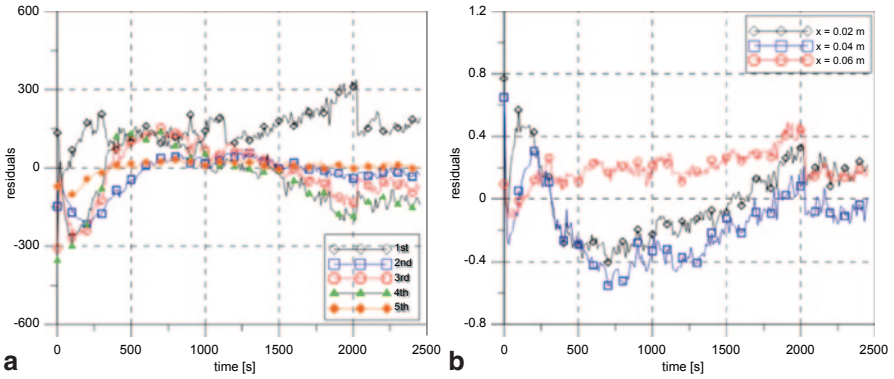

Fig. 6.11 Step function filter case: **a** residuals between calculated and experimental quantities for the first five transformed temperature modes, **b** residuals (°C) between calculated and experimental temperatures at three different positions

experimental arrangements of Cases 1, 2, and 3, where $N_{\text{MCMC}} = 120{,}000$ states in the MCMC method are generated; the first $N_{\text{burn-in}} = 40{,}000$ states neglected in order to achieve the equilibrium of the chains (Figs. 6.11 and 6.12).

One may observe that the estimates obtained for the polyester resin properties of the nanocomposite plate in Case 2 are very close to those obtained in Case 1, where the experimental setup involved homogeneous polyester resin plates. On the other hand, the estimated parameters for the polyester in Case 3 are not in good agreement with those estimated for Case 1. This behavior is because, for Case 3, the portion of the nanocomposite plate composed of polyester resin without addition of filler is placed away from the applied heat flux. This result is somehow expected, since

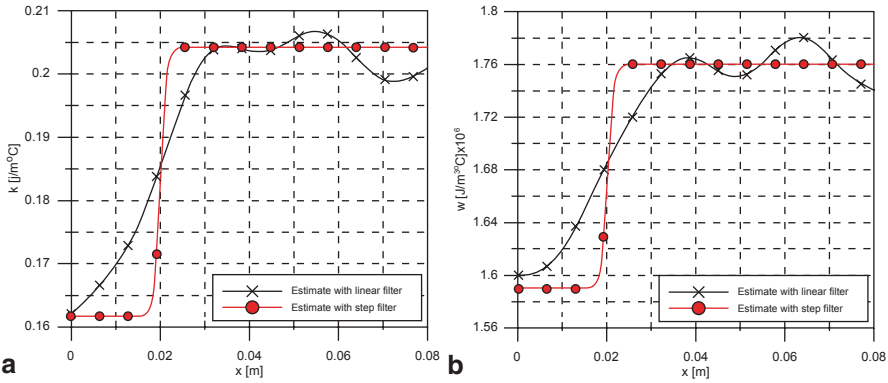

Fig. 6.12 Comparison of the mean values estimated for the a heat conductivity and b heat capacity curves (obtained with the linear filter and with the step function filter)

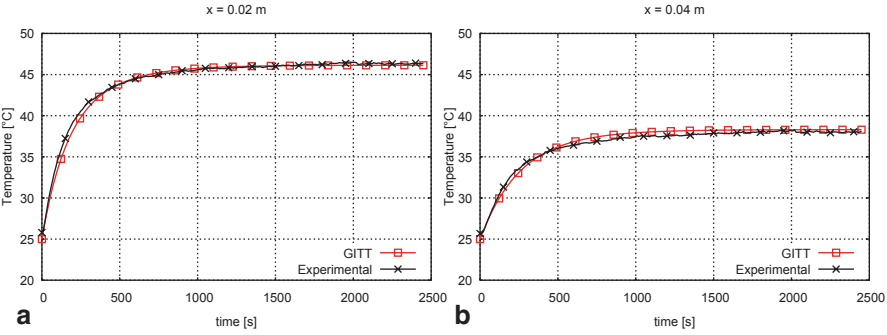

Fig. 6.13 Comparison of measured and simulated (GITT–UNIT code) temperatures using estimated parameters: time evolution of the temperature at **a** $x = 0.02$ m and **b** $x = 0.04$ m

the homogeneous portion of the plate suffers a smaller variation of temperature during the experiment, yielding locally low sensitivity coefficients for the parameters. For the portion of the nanocomposite plate corresponding to the polyester resin filled with alumina nanoparticles, it may be observed that Cases 2 and 3 yield estimates very close to each other. The estimated values for the thermal conductivity are slightly higher than those provided by the Lewis–Nielsen formula [34].

In order to further illustrate the accuracy of the proposed approach for the identification of the spatially varying thermophysical properties, we present a comparison of experimental temperature measurements and predictions obtained with the utilization of the estimated parameters in an independent computer program, known as the UNIT code [35, 36]. The measurements and the simulations refer to Case 2 and the parameters used in the simulation are those estimated with the linear filter function. Figure 6.13a and b present the experimental and simulated time evolutions of the temperatures at $x = 20$ mm and $x = 40$ mm, respectively, up to steady state.

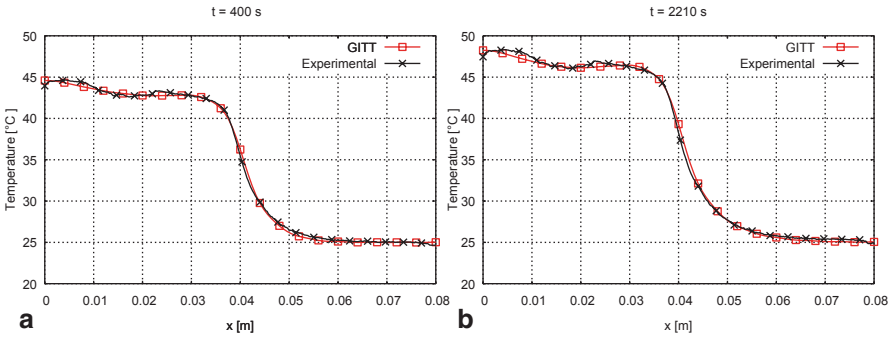

Fig. 6.14 Comparison of measured and simulated (GITT–UNIT code) temperatures using estimated parameters: vertical spatial distribution of temperatures at **a** $t=400$ s and **b** $t=2210$ s

Figure 6.14a and b show the vertical spatial distribution of the temperature at times $t=400$ s and $t=2210$ s, respectively. An excellent agreement is observed in Figs. 6.13 and 6.14, between the experimental data and the simulated temperatures calculated with the parameters estimated with the present inverse analysis.

6.5 Summary

The MCMC method is used for the solution of inverse problems dealing with the characterization of nonhomogenous media, in terms of spatially varying thermophysical properties and source terms. Two approaches are then presented and illustrated with practical examples, in order to reduce the computational cost associated with the MCMC method, namely: (1) a nodal approach, which locally linearizes the inverse problem by using temperature measurements for the computation of the sensitivity matrix and (2) an expansion of unknown spatially dependent thermophysical properties in terms of eigenfunctions, which is used in conjunction with the GITT. The results obtained with both the approaches are accurate and stable, thus demonstrating their capabilities of dealing with the ill-posed character of the inverse problem under consideration, which involves a large number of spatially distributed parameters.

Acknowledgments The authors would like to thank the Brazilian agencies for fostering of science, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and for the financial support for this work.

References

1. Beck JV, Blackwell B, St. Clair CR (1985) Inverse heat conduction: ill-posed problems. Wiley Interscience, New York
2. Tikhonov AN, Arsenin VY (1977) Solution of ill-posed problems. Winston & Sons, Washington, DC
3. Beck JV, Arnold KJ (1977) Parameter estimation in engineering and science. Wiley Interscience, New York
4. Alifanov OM (1994) Inverse heat transfer problems. Springer, New York
5. Alifanov OM, Artyukhin E, Rumyantsev A (1995) Extreme methods for solving ill-posed problems with applications to inverse heat transfer problems. Begell House, New York
6. Woodbury K (2002) Inverse engineering handbook. CRC Press, Boca Raton
7. Kaipio J, Somersalo E (2004) Statistical and computational inverse problems. Applied mathematical sciences, vol 160. Springer, Berlin (2004)
8. Ozisik MN, Orlande HRB (2000) Inverse heat transfer: fundamentals and applications. Taylor and Francis, New York
9. Tan S, Fox C, Nicholls GI (2006) Inverse problems, course notes for physics 707. University of Auckland, Auckland
10. Orlande HRB, Fudym O, Maillet D, Cotta RM (2011) Thermal measurements and inverse techniques. CRC Press, Boca Raton
11. Kaipio J, Fox C (2011) The Bayesian framework for inverse problems in heat transfer. Heat Transf Eng 32:718–753
12. Orlande HRB (2012) Inverse problems in heat transfer: new trends on solution methodologies and applications. J Heat Transf 134:031011
13. Batsale JC, Battaglia JL, Fudym O (2004) Autoregressive algorithms and spatially random flash excitation for 3D non destructive evaluation with infrared cameras. Quant Infrared Thermogr J 1:5–20
14. Bamford M, Batsale JC, Reungoat D, Fudym O (2006) Two dimensional velocity and diffusion mapping in the case of three dimensional transient diffusion: flash method and infrared image sequence analysis. Paper presented at the 8th conference on quantitative infrared thermography, QIRT 2006, Padova, Italy, 27–30 June 2006
15. Fudym O, Orlande HRB, Bamford M, Batsale JC (2008) Bayesian approach for thermal diffusivity mapping from infrared images with spatially random heat pulse heating. J Phys Conf Ser (Online) 135:012042
16. Massard H, Fudym O, Orlande HRB, Batsale JC (2010) Nodal predictive error model and Bayesian approach for thermal diffusivity and heat source mapping. Comptes Rendus Mécani 338:434–449
17. Massard H, Orlande HRB, Fudym O (2012) Estimation of position-dependent transient heat source with the Kalman filter. Inverse Probl Sci Eng 20:1079–1099
18. Naveira-Cotta CP, Cotta RM, Orlande HRB, Fudym O (2009) Eigenfunction expansions for transient diffusion in heterogeneous media. Int J Heat Mass Transf 52:5029–5039
19. Naveira-Cotta CP, Cotta RM, Orlande HRB (2010) Estimation of space variable thermophysical properties. In: Orlande HRB, Fudym O, Maillet D, Cotta RM (eds) Thermal measurements and inverse techniques. CRC Press, Boca Raton, pp 675-707 (Chap. 20)
20. Naveira-Cotta CP, Orlande HRB, Cotta RM (2010) Integral transforms and Bayesian inference in the identification of variable thermal conductivity in two-phase dispersed systems. Num Heat Transf Part B Fundam 57(3):173–203
21. Naveira-Cotta CP, Cotta RM, Orlande HRB (2011) Inverse analysis with integral transformed temperature fields for identification of thermophysical properties functions in heterogeneous media. Int J Heat Mass Transf 54:1506–1519
22. Naveira-Cotta CP, Orlande HRB, Cotta RM (2011) Combining integral transforms and Bayesian inference in the simultaneous identification of variable thermal conductivity and thermal capacity in heterogeneous media. J Heat Transf 133(11):1301–1311

23. Knupp DC, Naveira-Cotta CP, Ayres JVC, Orlande HBO, Cotta RM (2012) Space-variable thermophysical properties identification in nanocomposites via integral transforms. Bayesian inference and infrared thermography. Inverse Probl Sci Eng 20:609–637
24. Knupp DC, Naveira-Cotta CP, Orlande HRB, Cotta RM (2013) Experimental identification of thermophysical properties in heterogeneous materials with integral transformation of temperature measurements from infrared thermography. Exp Heat Transf 26(1):1–25
25. Knupp DC, Naveira-Cotta CP, Ayres JVC, Cotta RM, Orlande HRB (2012) Theoretical-experimental analysis of heat transfer in nonhomogeneous solids via improved lumped formulation, integral transforms and infrared thermography. Int J Therm Sci 62:71–84
26. Gamerman D, Lopes HF (2006) Markov chain Monte Carlo: stochastic simulation for Bayesian inference, 2nd edn. Chapman & Hall/CRC, Boca Raton
27. Lee PM (2004) Bayesian statistics. Oxford University Press, London
28. Winkler R (2003) An introduction to Bayesian inference and decision. Probabilistic Publishing, Gainsville
29. Zabaras N (2004) Inverse problems in heat transfer. In: Minkowycz WJ, Sparrow EM, Murthy JY (eds) Handbook of numerical heat transfer, 2nd edn. Wiley, Hoboken (Chap. 17)
30. Mikhailov MD, Ozisik MN (1984) Unified analysis and solutions of heat and mass diffusion. Wiley, New York
31. Cotta RM (1993) Integral transforms in computational heat and fluid flow. CRC Press, Boca Raton
32. Cotta RM, Mikhailov MD (1997) Heat conduction: lumped analysis, integral transforms, symbolic computation. Wiley Interscience, New York
33. Mark JE (2007) Physical properties of polymers handbook. Springer, New York
34. Lewis T, Nielsen L (1970) Dynamic mechanical properties of particulate-filled polymers J Appl Polym Sci 14(6):1449–1471
35. Sphaier LA, Cotta RM, Naveira-Cotta CP, Quaresma JNN (2011) The UNIT algorithm for solving one-dimensional convection-diffusion problems via integral transforms. Int Commun Heat Mass Transf 38:565–571
36. Cotta RM, Knupp DC, Naveira-Cotta CP, Sphaier LA, Quaresma JNN (2013) Unified integral transform algorithm for solving multidimensional nonlinear convection-diffusion problems. Numer Heat Transf Part A 63(11):840–866

Chapter 7
Scale-Dependent Porous Dispersion Resulting from the Cumulative Effects of Velocity Fluctuations

Wynand S. Verwoerd

7.1 Introduction

The advection–dispersion equation (ADE) [1] that is widely used to model the dispersion of solutes in fluid flow through porous media is based on splitting the carrier fluid velocity into a macroscopic mean (the 'drift' velocity) and microscopic random fluctuations. Using plausibility arguments, a Fickian assumption is made to represent solute transport as a result of the fluctuations. This leads to a diffusion-like transport equation but with the diffusion constant replaced by a medium-dependent dispersion constant D.

Diffusive behaviour (as follows from Fick's law) is exemplified by the linear time dependence of a Gaussian concentration plume variance σ^2 with initial value S^2:

$$\sigma^2(t) = S^2 + 2Dt, \tag{7.1}$$

where $D = dv$ is defined as the dispersion coefficient, d as the dispersivity, and v is the constant drift velocity.

Since the publication of collected experimental measurements of the dispersion of solutes in aquifers by Lallemand and Peaudecerf [2], Gelhar [3], and others, it has been known that the longitudinal dispersivity of natural porous media varies over several orders of magnitude with the scale of the experiment. While this dependence is roughly linear over a moderate range, it becomes nonlinear for the complete range of 5 orders of magnitude covered by available data as shown in Fig. 7.1.

Studies by many authors suggest that this is due to inhomogeneity of the hydraulic conductivity and other properties of real porous media. With the purpose of improving the understanding of such phenomena, we study a model where stochasticity is introduced at the fundamental level of the path that a fluid element follows through a porous medium, rather than stochastic variations of medium properties as done in most of the literature referred to.

W. S. Verwoerd (✉)
Center for Advanced Computational Solutions (C-fACS), WF & Molecular Biosciences Department, Faculty of Ag & Life Sciences, Lincoln University, Christchurch, New Zealand
e-mail: wynand.verwoerd@lincoln.ac.nz

S.K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_7,

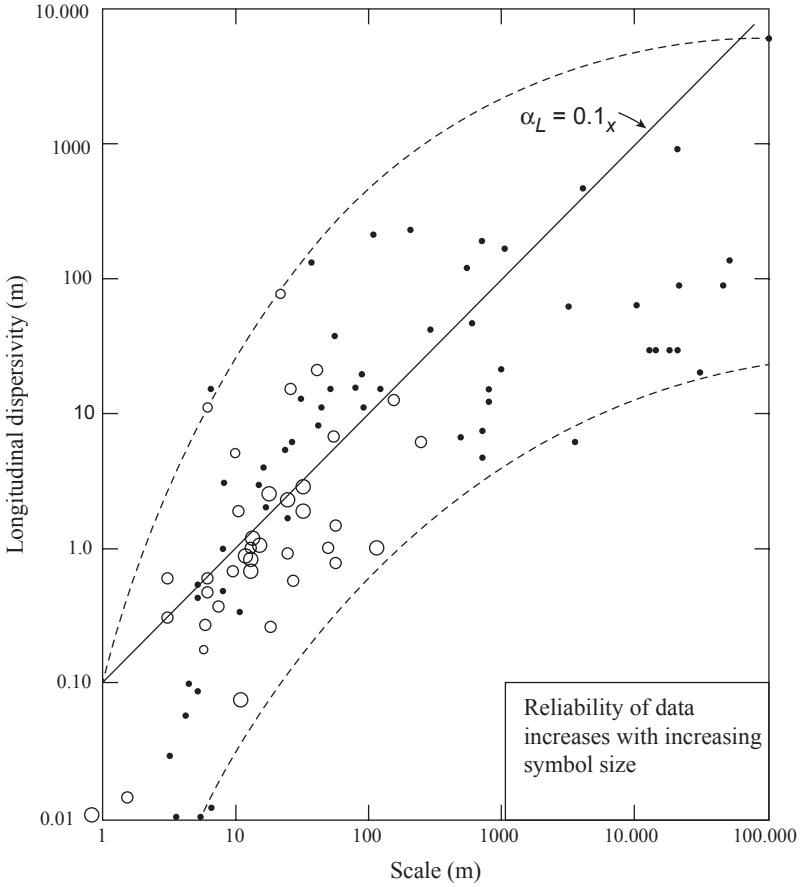

Fig. 7.1 Field-measured values of longitudinal dispersivity as a function of the scale of measurement (*largest circles* represent the most reliable data). (Source: Gelhar [3], reproduced from [1])

Hence, we describe the stochastic path followed by a single fluid element through the porous medium by the equation:

$$dx = u(x)dt + \gamma dB(x,t,\theta). \tag{7.2}$$

Here $u(x)$ is the macroscopic carrier fluid velocity, as derived from an appropriate flow equation such as Darcy's law, and will in general depend on the hydraulic head differential as well as medium properties such as hydraulic conductivity and porosity. The second term represents the pore-scale (microscopic) scattering represented as a stochastic perturbation of the fluid velocity; $B(x, t, \theta)$ is a Wiener process with θ labelling individual realisations and γ is the amplitude that regulates the extent to which the path is modified. Only 1D flow is modelled.

Equation 7.2 is a stochastic differential equation (SDE) and needs to be solved by using Ito calculus and other methods from SDE theory as set out, for example, by Øksendal [4]. Each individual realisation of the solution represents a possible path of a fluid element through the porous structure, and macroscopic dispersion is described by calculating statistics over all realisations.

Instead of numerically calculating individual realisations, our approach is to calculate expectation values over all realisations, of a suitably chosen function. Using Dynkin's formula [4] this in effect replaces the SDE by a deterministic partial differential equation of the form:

$$u(x)\frac{\partial f}{\partial x}+\frac{1}{2}\gamma^2\frac{\partial^2 f}{\partial x^2}=\frac{\partial f}{\partial t}. \tag{7.3}$$

A key step in this approach is using a spectral expansion (also called a Karhunen–Loeve expansion [5]) that builds up a space- and time-dependent Wiener process $B(x,t,\omega)$ with a known spatial covariance $C(x_1, x_2)$ by superposition of a set of simpler time-dependent Wiener processes $b_n(t, \omega)$ according to the definition:

$$B(x,t,\omega)=\sum_{n=0}^{\infty}\sqrt{\lambda_n}\,f_n(x)b_n(t,\omega). \tag{7.4}$$

Comparing expectation values from Dynkin's formula with those calculated from a probability distribution for the position of a fluid element as a function of time, an integral equation for the probability density is derived. This is solved by a Gaussian functional form, hence allowing calculation of dispersivity and hence the evolution of an initial solute concentration from the following expression:

$$C(x,t)=\frac{1}{u(x)}\int_{-\infty}^{\infty}dx'C(x',t_0)u(x')P_{t_0}(x'\mid x,t). \tag{7.5}$$

Here $P_t'(x'|x,\ t)$ is the probability density with respect to x', a fluid element which is found at the position x at time t, originated from position x' at the earlier time $t'<t$, and reduces to a Dirac delta function for deterministic flow (in the absence of stochastic perturbations of the path).

The work reported in this chapter was presented at the Modelling & Simulation of Diffusive Process and Its Applications (ICMSDPA)-12 conference [6] and more details are set out in the research literature [7–9].

7.2 Dispersion in Accelerating Flow

A first application, treated in more detail elsewhere [7], is to the case of a linearly increasing or decreasing drift velocity:

$$u(x)=v_0\pm\mu^2 x. \tag{7.6}$$

Equation 7.3 is solved exactly in this case by:

$$f_+(z,T) = \exp\left[-T - z^2 - aze^{-T} - \frac{1}{4}a^2(1-e^{-2T})\right], \tag{7.7}$$
$$f_-(z,T) = \exp\left[-aze^{-T} + \frac{1}{4}a^2(1-e^{-2T})\right]$$

where variables have been transformed to the dimensionless pair:

$$z = \pm\frac{\mu}{\gamma}(x \pm \frac{v_0}{\mu^2}); \quad T = \mu^2 t. \tag{7.8}$$

The outlined procedure in this case leads to an integral equation that may also be solved exactly, and hence it is found that an initial Gaussian solute concentration $C(x,0) = \mathcal{G}(\xi,s)$ (where ξ is the mean and s is the standard deviation), evolves with time as given by:

$$C(x,t) = M(x-X,T)\mathcal{G}\left[X, s^2e^{\pm 2T} \pm \frac{\gamma^2}{2\mu^2}(e^{\pm 2T}-1)\right]. \tag{7.9}$$

Here $X = X(t)$ is the position that the concentration peak would have in a deterministic (that is plug) flow at velocity $u(x)$, and Eq. 7.9 shows that the Gaussian part of the plume still propagates at the deterministic (but changing) velocity.

However, there is a very significant change in the way that the plume disperses with time. This information is contained in the time-dependent variance of the Gaussian:

$$\sigma^2(t) = s^2e^{\pm 2T} \pm \frac{\gamma^2}{2\mu^2}(e^{\pm 2T}-1). \tag{7.10}$$

The modulation factor M in the Eq. 7.9 is a complicated rational expression and has the qualitative effect of introducing skewness into the peak. Except near stagnation points of the flow, it is slowly varying with x compared with the exponential variation of the peak, and can be taken to a first approximation as the fixed value at the peak position. This turns out to be $M(0,T) = 1$.

Although this is not immediately apparent from the formula, in the limit where $\mu \to 0$ (that is constant velocity flow) the variance grows linearly with time, i.e. the diffusive behaviour of Eq. 7.1 is regained. Equation 7.9 reduces in this limit to:

$$C(x,t) = \frac{1}{\sqrt{2\pi(s^2+\gamma^2 t)}}\exp\left(-\frac{(x-v_0t)^2}{2(s^2+\gamma^2 t)}\right). \tag{7.11}$$

This is identical to the expression obtained when Eq. 7.3 is solved directly for constant flow.

By contrast, Eq. 7.10 shows that for a linearly changing flow velocity, the variance changes exponentially with time. Dispersion in the presence of a gradient in

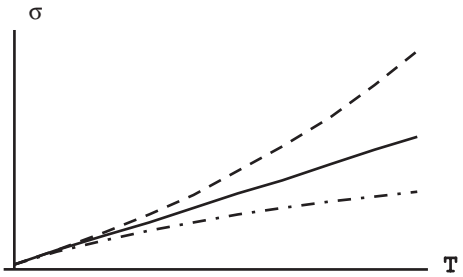

Fig. 7.2 Evolution of actual dispersion with time for accelerating flow (*dashed line*) and decelerating flow (*dash-dot line*) in comparison to diffusive dispersion superimposed on kinematic dispersion (*solid line*)

the carrier fluid velocity is a more complex phenomenon, which cannot be adequately described by a diffusive model.

The first term in Eq. 7.10 represents exponential growth even in the absence of pore scattering (the limit $\gamma \to 0$); this is easily shown to be a reversible kinematic effect. The second term represents the intrinsic dispersion and is not reversible. Figure 7.2 shows its effects in comparison with diffusive dispersion for the same reasonably chosen numerical parameters.

The figure calls into question the convention of describing dispersion by a linear time coefficient (the dispersivity) but characterising this as changing with scale (that is time). Instead, dispersion should simply be recognised as nonlinear, except in the special case of a constant drift velocity.

Also, bearing in mind that the drift velocity can only realistically increase or decrease over a limited distance, the opposite effects of acceleration and deceleration that the figure shows raises the question of what happens to a solute plume that traverses a macroscopic velocity fluctuation. Will these opposite effects cancel, allowing only the mean velocity to be taken into account? The nonlinearity makes that implausible, but a firm answer requires direct investigation of such fluctuations.

7.3 Piecewise Constant Drift Velocities

A basic objective of this work is to find analytical solutions, even if approximate, rather than numeric ones to benefit from the more fundamental insights into the underlying mechanisms that become possible in this way. Pursuant to that aim, while still maintaining flexibility in the qualitative behaviour of the drift velocity profile, a progression of increasingly complex piecewise constant velocity profiles are now studied.

In the case of a flow velocity that is piecewise constant, in a number of adjoining regions, Eq. 7.11 applies in each of these separately by taking an appropriate value for $u = v_0$ in each region. For simplicity, we assume that u remains positive throughout. Let u_i indicate the flow velocity in region i, defined by its entrance boundary x_{i-1} and exit boundary x_i.

To describe the transmission of a peak from one region to the next, we need to formulate the solution of a boundary value problem, rather than the initial value problem discussed so far. The underlying notion is that transmission of a peak from

region i to region $i+1$ delivers solute concentration at the boundary x_i as a time profile $C(x_i, t)$; and this is equivalent (inside region $i+1$) to injecting the same time profile from an external source at $x=x_i$. It seems straightforward to modify Eq. 7.5 appropriately:

$$C_{i+1}(x,t) = \frac{u(x_i)}{u(x)} \int_{-\infty}^{t} dt' C_i(x_i,t') P_{x_i}(t' \mid x,t), \tag{7.12}$$

where the time-dependent probability density used in the boundary value problem is obtained from the spatial density in Eq. 7.5 by an appropriate transformation of variables. Equation 7.12 is indeed equivalent to Eq. 7.5 as a formulation of solute mass conservation in the case of deterministic flow, as may be confirmed by use of the appropriate δ-function expression for P.

However, in the stochastic case, the situation is more subtle. The time profile of an external source injecting solute at position x_i is not the same any more as the time profile in the flow at position x_i. This is because there is a finite probability of solute dispersing upstream from the injection point as well as downstream. There is also a mathematical difficulty in extrapolating the injection profile backwards in time, as required by the lower integration limit of Eq. 7.12. This necessitates the introduction of a time cutoff, t_c. The effect of these difficulties is illustrated by noting that Eq. 7.12, applied to the trivial case when $u_i=u_{i+1}$, does not reduce to transmission through the nonexistent step according to Eq. 7.11, as it should.

To rectify the matter, a modifying factor $g(x, t, x, t, u)$ is introduced by which a desired concentration profile in the flow needs to be multiplied in order to find the required injection profile. It turns out that it is possible to solve for this function, in the case of a Gaussian peak propagating through a step, by requiring conformation to Eq. 7.11 for the case of a nonexistent step; the result is:

$$g(x',t',x,t,u) = \left(\frac{t-t'}{t'-t_c}\right)\left(\frac{x'-ut_c}{x-x'+u(t-t')}\right) + \frac{x-x'}{x-x'+u(t-t')}. \tag{7.13}$$

Incorporating these modifications into Eq. 7.12, the final expression for the concentration in region $i+1$, given as Gaussian peak in region i is:

$$c_{i+1}(x,t) = \left(\frac{u_i}{u_{i+1}}\right)\frac{1}{4\pi\gamma^2}\int_{t_c}^{t} dt' g(x_i,t',x,t,u_{i+1}) \frac{x-x_i+u_{i+1}(t-t')}{(t-t')^{3/2}\sqrt{t'-t_c}} \times,$$
$$\times \exp\left(\frac{(x-x_i-u_{i+1}(t-t'))^2}{2\gamma^2(t-t')} + \frac{(x_i-u_{i+1}t')^2}{2\gamma^2(t'-t_c)}\right) \tag{7.14}$$

and the appropriate cut-off time is given by $\gamma^2 t_c = -s^2 \text{ or } s^2 + \gamma^2 t = \gamma^2(t-t_c)$, in the notation of Eq. 7.11.

7.4 Dispersion at a Velocity Step

A full description of the approximations and procedures used to solve the integral in Eq. 7.14 analytically is beyond the scope of the present chapter, but has been detailed elsewhere [8]. The main result obtained is that when a Gaussian concentration plume propagates through a stepwise velocity change from V_1 to V_2, the profile remains approximately Gaussian but the variance is now given by the expression:

$$\sigma^2(t) = \frac{\gamma^2 (t - t_c)}{a(T)}, \quad t_c = \frac{-S^2}{\gamma^2}. \tag{7.15}$$

For a constant velocity $a(T) = 1$, we recover the linear time relationship; as before, γ is the amplitude of the stochastic term in the underlying SPDE, and a comparison of Eq. 7.15 in the constant velocity case with Eq. 7.1 shows that apart from proportionality constants, γ^2 then reduces to the dispersion coefficient D.

The most significant feature of Eq. 7.15 is that beyond a step, the linear time dependence is modified by the multiplicative factor $1/a(T)$. Here, as in Eq. 7.8, T is a dimensionless time parameter but now expressed by a formula given below in terms of the time θ at which the plume reaches the step, and $a(T)$ is approximated by:

$$a(T) \approx 1 - \frac{4\alpha^2 \Delta}{T^2 (\alpha + \Delta)^2}. \tag{7.16}$$

In Eq. 7.16, $\Delta = (V_2 - V_1)/(V_2 + V_1)$ is a dimensionless parameter that characterises the height of the velocity step from V_1 to V_2, and α is the ratio of the plume variances at the entry and exit points of the velocity regime that is terminated at the step under consideration. The approximation holds for $\Delta < 0.3$.

T is defined such that when the peak reaches the step, $T = 1$. In the absence of dispersion $\alpha = 1$ and then $a(1)$ reduces to the kinematical value $(1 - \Delta)^2 / (1 + \Delta)^2$. With dispersion, Eq. 7.16 only applies at T sufficiently larger than 1 that the plume has fully penetrated the step. At such a time $a(T)$ approaches the kinematical compression (stretching) of the plume at a downwards (upwards) velocity step but then decays back to a value of 1 for large times (that is at positions far beyond the step).

This behaviour is illustrated by the time dependence of the multiplicative factor, for a chosen step size, as shown in Fig. 7.3.

Of most interest is the interpretation of the decay from the kinematic to diffusive values. In the case of the upwards step, the initial kinematic compression produces larger concentration gradients than in a constant velocity, diffusive peak; consequently, it disperses faster until it has eventually reached the same extension as the diffusive peak. The opposite holds for a downwards velocity step.

This result implies that the effects of a single velocity step on dispersion remain significant only over a limited length scale, which can be shown to be proportional to the initial Gaussian variance as well as the step size Δ.

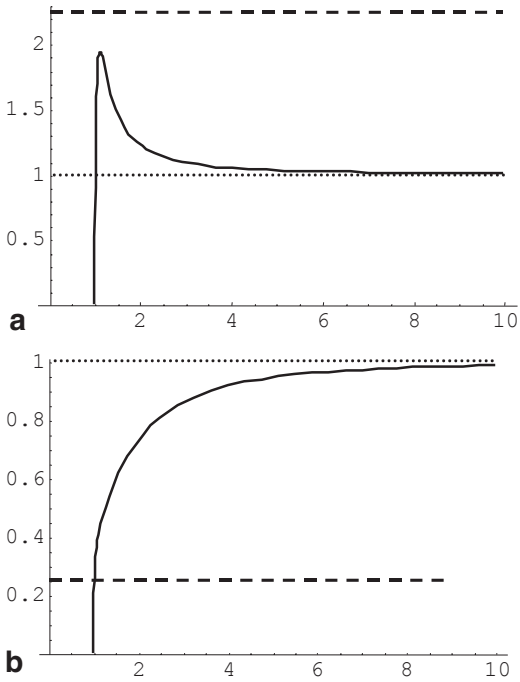

Fig. 7.3 Time development of the factor that multiplies diffusive dispersion, after penetration of **a** an upwards, and **b** a downwards velocity step (diffusive model value (=1) shown as *dotted lines*, and the kinematic values as *dashed lines*)

7.5 Stepwise Fluctuation Sequences

The cases of linear velocity growth over an indefinite period, and of a velocity step, both demonstrate the principle of nonlinear dispersion growth, but neither represents a realistic scenario. Instead, an actual flow through an aquifer would be expected to show fluctuations of the flow velocity around the average that is deduced from total flow volume measurements.

To model that, the single-step results are now extended to a sequence of multiple steps, such as may be used to describe a series of fluctuations of the flow velocity about an average value. Clearly, such a piecewise constant velocity description is still highly idealized, but has the advantage that with a simple formula such as Eq. 7.16, it is feasible to perform a detailed calculation analytically and hence gain insight into the important processes and variables that are involved.

In the detailed analysis of the single step, it is found that a Gaussian incident concentration peak is somewhat distorted when it penetrates the step, and this is expressed by a slowly varying modulation factor that multiplies the transmitted peak. As a simplifying assumption, we now ignore the modulation and hence use the output Gaussian, with its nonlinear time-dependent variance, as the input Gaussian for the next velocity step.

The variables that are needed to describe the effect of the *m*th step are then given by the following expressions that generalise the corresponding expressions for a single step:

$$T_m = \sqrt{\frac{t - t_c}{\theta_m - t_c}}, \quad \alpha_m = \frac{\theta_m - t_c}{\theta_{m-1} - t_c}. \tag{7.17}$$

In a similar way, Δ acquires an index and is calculated by appropriately assigning the applicable velocity values for the particular step.

While the output Gaussian from step m propagates to step $(m+1)$, its variance changes and so the input variance for step $(m+1)$ is found by evaluating Eq. 7.16 for $a_m(T_m)$ at the T_m value obtained by setting $t = \theta_{m+1}$, hence reducing T_m to $\sqrt{\alpha_{m+1}}$. When the peak emerges from step $(m+1)$, the variance acquires an additional factor $1/a_{m+1}(T_{m+1})$, which once more is evaluated at $T_{m+1} = \sqrt{\alpha_{m+1}}$ to give the input Gaussian for step $(m+2)$.

In this way, the cumulative effect of a sequence of M steps is to multiply the variance by an enhancement factor F_M defined by using Eq. 7.16 as:

$$F_M = \prod_{m=1}^{M} \frac{\alpha_{m+1}(\alpha_m + \Delta_m)^2}{\alpha_{m+1}(\alpha_m + \Delta_m)^2 - 4{\alpha_m}^2 \Delta_m}. \tag{7.18}$$

In this expression, strictly speaking the Mth factor and hence also F_M should still be a function of an undetermined T_M parameter. But in most applications, one envisages an infinite sequence of steps. M is simply the number of steps that the plume has penetrated in moving from the origin to a position x. Then it is appropriate to evaluate the effect of the first M steps at the time when it enters the $(M+1)$th, and Eq. 7.18 is obtained.

The approximations leading to Eq. 7.18 can be tested by first applying it to a "staircase" of regularly spaced small discrete steps, all with the same value for Δ. For many small steps this approaches linear acceleration, a case exactly solved as described in a previous section.

Suppose first that there is no dispersion, i.e. the deterministic limit, and so all $\alpha = 1$. Then, the product expression becomes a simple power law

$$F_M = \left(\frac{1+\Delta}{1-\Delta}\right)^{\frac{1+\Delta}{\Delta}\mu^2 t}, \tag{7.19}$$

and in the limit of large M, that is small Δ, this reduces to $e^{\mu^2 t}$ by virtue of the well-known identity $(1+\frac{x}{n})^n \xrightarrow[\infty]{n} e^x$. This exponential growth or decline is identical to that calculated directly for the deterministic evolution of a Gaussian plume in a flow with a constant velocity gradient, as displayed in the first term of Eq. 7.10.

Returning to the full dispersion expression, the θ_{m+1} may be expressed in terms of Δ and μ and used to substitute α_m in Eq. 7.18. The resulting product is complicated but may be expressed in terms of Gamma functions, and these approximated by the Stirling approximation. In this way, similar exponential factors as for the

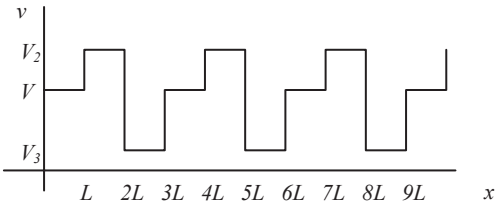

Fig. 7.4 Step model for velocity fluctuations on a periodic grid of length L

deterministic case are retrieved. Again, the result can be compared to the exact result for stochastic dispersion in a constant velocity gradient, both terms of Eq. 7.10. The analytical expressions are in this case not identical, as a result of the approximations involved in Eq. 7.16, but contain similar exponential terms and give a very similar numerical behaviour.

Moving on to a fluctuating drift velocity, a simple model is first investigated where the steps are located on a linear grid with spacing L, as shown in Fig. 7.4.

From the observation of a decay length associated with a single step, it can be anticipated that length-scale effects are bound to follow for stepwise fluctuations. The choice of a grid with a well-defined characteristic length facilitates investigation of such effects. The value V represents the average flow velocity. The value of V_2 is fixed by freely choosing the step size parameter Δ to determine the amplitude of the fluctuation about V, i.e. $V_2 = V(1+\Delta)/(1-\Delta)$. Then the choice $V_3 = V(1+\Delta)/(1+3\Delta)$ ensures that the average has value V. So the fluctuations are fully characterised by the amplitude parameter Δ and the fluctuation length 3 L, which is in effect the periodic repeat length.

In the absence of dispersion, the only effects of the steps are the kinematical compression or stretching of the Gaussian solute peak required by flux conservation, and these obviously cancel over the combination of the 3 steps that make up a single fluctuation. It therefore makes sense to also collect the three steps together when describing the evolution of the plume in the presence of dispersion. Any deviation from 1, in the combined multiplicative factor, can then clearly be ascribed to a fluctuation effect on dispersion.

The first step in this calculation is to find the arrival times of the peak at each step, as $\theta_i = \Theta_i(L/V)$ where the dimensionless arrival times Θ_i are $\Theta_1 = 1$, $\Theta_2 = 2/(1+\Delta)$, and $\Theta_3 = 3$. Putting these into the definitions of the α_i as in Eq. 7.17, we find:

$$\alpha_i = \frac{\Theta_i - \frac{Vt_c}{L}}{\Theta_{i-1} - \frac{Vt_c}{L}}. \tag{7.20}$$

This equation exhibits a crucial fact that pervades all of the results on the fluctuation model, namely that all lengths are measured with respect to a common length scale defined by:

$$\Lambda = -Vt_c = VS^2/\gamma^2. \tag{7.21}$$

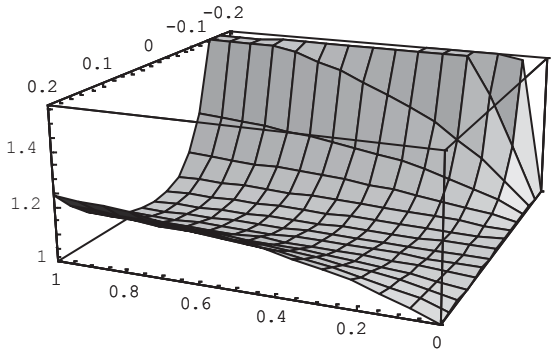

Fig. 7.5 The single fluctuation enhancement factor for $-0.2 < \Delta < +0.2$ and $0 < \lambda < 1$

This is applied by replacing L in the expression for the Θ_i by the scaled grid spacing $\lambda = L / \Lambda$.

The effect of the first fluctuation that the plume encounters is then given by an enhancement factor $f = F_3$ in the notation of Eq. 7.18, and it is easily seen that this is a function of only the two dimensionless parameters Δ and λ that characterise the fluctuation amplitude and length, respectively.

The analytical form obtained for $f(\Delta, \lambda)$ is very complicated, but its numerical behaviour is rather simple, as shown in Fig. 7.5.

Plausibly, there is no enhancement for either $\Delta = 0$ or $\lambda = 0, f = 1$.

However, a very significant feature seen in the figure is that for all other combinations of Δ and λ values, f is greater than 1, which means that dispersion is *increased* relative to the diffusive or Fickian values by a fluctuation. It is not obvious that this will happen, as a single upwards velocity step suppresses dispersion while a downwards step increases it, and without dispersion these effects cancel exactly across a fluctuation.

The parameters displayed in the figure cover a generous range, allowing for the maximum and minimum velocities to differ by ≈40 % from the average, and as will be seen below only $\lambda << 1$ is expected to be physically relevant.

This result adds substantially to the plausibility of the simplified model under discussion where all fluctuations are assumed to be identical. If f turned out to be less than 1 for some fluctuations, it might have been argued that in a realistic system fluctuations over a range of amplitudes and lengths will be present and could cancel each other. As this is not the case, the present model where, in effect, the ranges are represented by their averages and the effects from subsequent fluctuations are allowed to accumulate, appears quite reasonable as a first approximation.

The next step is to include not just multiple steps but multiple *fluctuations* in the calculation. We modify the notation slightly, henceforth, using m as an index counting fluctuations rather than steps. Equation 7.18 is accordingly modified by replacing the single-step factor that appears as the subject of the product, by a product of the three-step factors that make up the mth fluctuation. Analytical calculation of this single fluctuation enhancement factor $f_m(\lambda, \Delta)$ proceeds as outlined above for the

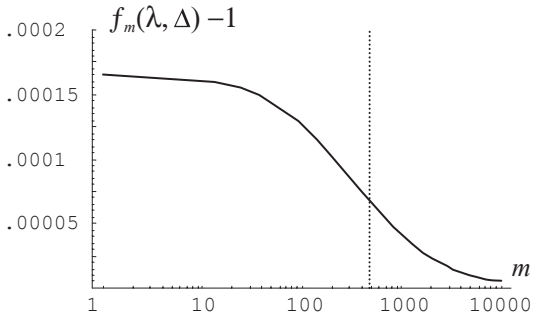

Fig. 7.6 Dependence of the single fluctuation enhancement (*solid line*) on the fluctuation number m ($\Delta = 0.1, \lambda = 0.001,$ *dotted line* corresponds to the scale length Λ)

special case, now written as $f_1(\lambda, \Delta)$. Once more a very complicated but rational algebraic expression is obtained for $f_m(\lambda, \Delta)$. The nature of its m dependence is illustrated in Fig. 7.6.

The enhancement encountered by the plume as it traverses a sequence of fluctuations decreases slowly at first, but when it reaches the m value that corresponds to the scale length Λ, there is a sharp decline and a further asymptotic decrease beyond that.

The parameter values chosen to display this behaviour are found by taking a Peclet number between 1 and 10, the initial plume extension of the order of 10^{-2} m and the pore diameter 10^{-4} m. These are merely meant as plausible order of magnitude estimates and lead to a Λ value in the order of 10 m or larger. Then the assumed value $\lambda = 0.001$ implies a physical fluctuation length of about 3 cm, which appears as reasonable estimate of the scale on which inhomogeneity appears in natural aquifers.

The crucial role played by Λ is seen to be as a *transition point* between distinct short-range and long-range behaviours of the enhancement factor. Such a distinction between short- and long-range behaviours is also noticeable in experimental observations of the dispersivity that extend over a sufficient spatial scale, such as in Fig. 7.1, and it is gratifying to find it arising as a natural consequence of the mathematics of the fluctuation model.

However, before dispersivity can be calculated, it is first necessary to perform the repeated product of enhancement factors required by Eq. 7.18. Unfortunately, this is not feasible with the complicated form obtained for f_m so far and we resort to an approximate expression for it.

The strategy used to find such an approximation is to make three series expansions of f_m about the points $m \to \infty$, $\lambda = 0$, and the point $m = 1/2\lambda$ that falls inside the transition range; then a simple function is guessed that has the same dominating terms in its series expansions about the same points. A surprising feature of the dominating terms in all three expansions is that the Δ dependence is separated out into an identical rational polynomial expression that we designate as $Q(\Delta)$. The approximation found in terms of this is:

$$f_m(\Delta, \lambda) \approx 1 + \frac{3\lambda Q(\Delta)}{1 + 3m\lambda}. \tag{7.22}$$

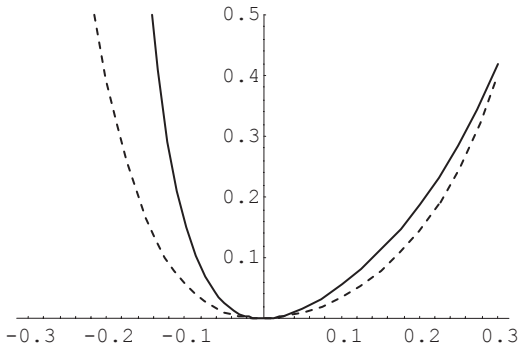

Fig. 7.7 Q as a function of step-size parameter Δ, for the regular grid fluctuation model (*solid line*) and for a non-grid model (*dashed line*)

This amazingly simple expression reproduces the behaviour of the full formula for f_m to such an extent that it is indistinguishable from the original curve for the parameter values shown in Fig. 7.5; only when λ approaches 1, it does underestimate the exact value by a few per cent in the low range, while the high-range values are still very close.

A plot of the Q function for a range of Δ values is shown in Fig. 7.7. The calculations described so far have also been done for a slightly more general fluctuation model, in which the restriction of steps to lie at grid points is relaxed. It turns out that while the full expression for f_m is different, Eq. 7.22 still holds but a somewhat different expression for $Q(\lambda)$ is obtained.

Both curves are plotted in Fig. 7.7, and while Q is seen to be moderately sensitive to the details of the shape of the assumed fluctuation profile, Q values are less than about 0.4 for reasonable fluctuation amplitudes. A more elaborate model in which fluctuations on two different macroscopic scales are superimposed is also investigated, with the result that the shorter-scale fluctuations are found to dominate. This confirms that the simple model with a well-defined single-fluctuation length is an adequate representation.

Equation 7.22 is simple enough that when substituted into Eq. 7.18, the expression for the cumulative enhancement factor F_M can be calculated analytically. It is obtained as a ratio of two Pochhammer functions, which are themselves defined as a ratio of Gamma functions. The Stirling approximation is applied to these, and after simplification, we obtain the following expression for the cumulative enhancement factor:

$$F_M = \left(1+\frac{3Q\lambda}{3M\lambda+1}\right)^M \left(1+\frac{3M\lambda}{3Q\lambda+1}\right)^Q \left(1-\frac{9MQ\lambda^2}{(3M\lambda+1)(3Q\lambda+1)}\right)^{\frac{1}{2}+\frac{1}{3\lambda}}. \qquad (7.23)$$

When the plume has traversed a distance x, the number M of fluctuations that it has encountered is given by $x=3LM$ and the time t taken is $x=Vt$.

It can be seen that the dispersivity is obtained from the calculated enhancement factor by multiplying it by a constant, namely the initial (laboratory scale) dispersivity. In other words, the enhancement factor F can be considered as merely the

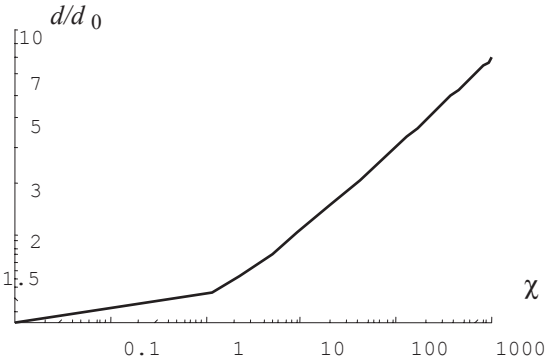

Fig. 7.8 Logarithmic plot of scaled dispersivity as a function of scaled traversal length ($\lambda = 0.001$, Q=0.3, calculated from Eq. 7.23)

(dimensionless) scaled dispersivity. As the measured values are expressed as a function of the traversal length x rather than time t, we eliminate M in Eq. 7.23 in favour of the scaled traversal length $\chi = x / \Lambda$ by the relation:

$$\chi = 3\lambda M. \tag{7.24}$$

The resulting behaviour of $F(\chi)$ is shown in Fig. 7.8 over 5 orders of magnitudes of the traversal length, similar to the range of experimental values in Fig. 7.1.

The figure shows that there is a distinct transition from a low-range behaviour to a high-range behaviour when the traversal length reaches the scale length Λ.

Such a transition is also seen in the experimental values, at a length of 10–100 m.

The high-range behaviour is easily extracted from the analytic expression in Eq. 7.23 by taking the large M limit, as M is proportional to χ. In this limit, the first factor reduces to e^Q (a constant) and the last factor also becomes constant, so that the entire χ dependence is carried by the second factor which simplifies to $(1+\chi)^Q \rightarrow \chi^Q$. This power law becomes a straight line with slope Q in the logarithmic plot, which agrees with the observed high-range behaviour. Even more significant, the measured values suggest a slope of approximately 0.3, and this agrees with the restricted range of Q values derived from the mathematical properties of the fluctuation model as shown in Fig. 7.4.

However, there are also two points on which Fig. 7.8 differs markedly from the observations. First, the low-range trend in the figure is a slower rate of growth than in the high range while the reverse holds for the measurements. Secondly, as a result of this, the overall increase in the figure is only by a factor of 10, while an increase of 10^4 is found for measured values.

This vast numerical discrepancy hinges on a single detail of the low-range behaviour. For $\chi << 1$, i.e. $M << 1/3\lambda$, the increase of F with χ is dominated by the first factor in Eq. 7.23, and this increases roughly exponentially with χ. It saturates to the value e^Q at the transition point and with any reasonable value of Q this is far too small to produce the observed rise of 3 orders of magnitude in dispersivity up to the transition.

7.6 Further Exploration of Fluctuation Effects Using a Schematic Model

The simple ad hoc change of replacing the coefficient of Q in the first term of the Eq. 7.23 by a value increased by a factor of about 25 is found to remove both discrepancies and give a dispersivity curve in rather good agreement with the experimental values. There is no obvious justification for such a change within the confines of the stepped-fluctuation model presented, but this does suggest that it may be possible to modify the details of the model to obtain agreement while retaining key features such as the dispersive length scale that divides short- and long-range behaviour.

As a working hypothesis, it is assumed that the 1D stepped fluctuation model underestimates the dispersion enhancement by a single fluctuation, but correctly predicts the decline of this enhancement with distance as measured against the length scale Λ, the behaviour shown in Fig. 7.6.

Various arguments can be advanced why such an underestimate is plausible. Within the confines of the 1D step model, one reason is the neglect of the non-Gaussian distortion of the plume by each step, mentioned to arrive at Eq. 7.17. The assumption of abrupt steps in the 1D step model may also contribute to this because it has been shown [10] that, for example, linearly slanted steps have associated exponential effects on dispersion.

Finally, it has been established that in higher dimensional systems, transverse variations in flow velocity (such as stratified flow [11]) can also enhance longitudinal dispersion.

Based on this observation, and a number of trials guided by the idea that in a more elaborate model the Q function in the Eq. 7.22 may become a function of m and attain a higher value in the low range, a schematic model is proposed where the single fluctuation enhancement in the Eq. 7.22 is modified to the form:

$$f_m(\Delta, \lambda) = 1 + \frac{3\lambda Q(\Delta)(w + 3\lambda m)}{(1 + 3\lambda m)^2}. \tag{7.25}$$

Here w is a new constant to be empirically determined. In effect Q is assumed to have a product form, consisting of the factor $Q(\Delta)$ previously determined, and a factor that interpolates smoothly between a value enhanced by a factor w for $m = 0$, and 1 at $m \to \infty$, with the transition taking place at the scale length Λ. The trials show that the calculated dispersion is not sensitive to the functional form assumed for this interpolation, provided that it does not drop too steeply from the value w at $m = 0$.

Inserting Eq. 7.25 into Eq. 7.18 evaluation of the products once more leads to the Pochhammer functions, but this time with irrational arguments so that simplification by use of the Stirling approximation is no longer applicable. The expression can nevertheless be calculated numerically.

To compare with experiment, parameters are chosen by the following procedure. Inspection of the measured values suggests that the transition between low- and high-range behaviours takes place at approximately 10 m, so this is chosen as the length scale Λ. To obtain a smooth buildup of the dispersivity to the observed large value at the transition point without an excessive enhancement by a single fluctuation, a large number of factors are needed in the product appearing in Eq. 7.18, which means that λ must be quite small. We choose $\lambda = 0.0035$, but this value can be adjusted by an order of magnitude either way without affecting the results significantly.

The combination of the chosen Λ and λ values imply that the physical length of the dominating fluctuation is about 9 cm, which also appears very reasonable from a physical point of view. The value of Q is simply the slope of the high-range dispersivity on a logarithmic plot, and from the data a value of $Q=0.32$ has been chosen. That leaves only w to be empirically fitted, and the value $w=24$ produces good agreement with experimental values as shown in Fig. 7.9.

This w value represents a large increase in the effect of fluctuations in the low traversal length range for the schematic model, compared with the detailed predictions of the original 1D stepped fluctuation model. It may be questioned whether this is plausible. To put that in perspective, note that for the 1D step model, the enhancement by the first fluctuation would be 0.1 %; this is increased to 2.4 % by $w=24$, and smaller values for downstream fluctuations. Plausibly, even for w as large as 24, the effect of any single fluctuation remains only a small perturbation of the background diffusive plume growth.

There is obviously a reciprocal relationship between this behaviour and the value of λ. For example, at the extreme of choosing $\lambda = 1/3$, only a single fluctuation would fit into the interval covered by the length scale. Hence, not only would there be a single discontinuous jump in the dispersivity, but also the enhancement would have to be the absurd value of 1,000 rather than a few percent. This argument shows that whatever one considers to be a plausible maximum for the single fluctuation enhancement, places an upper limit to the value that can be chosen for λ. The values presented above shows that it is possible to reconcile the enhancement value and the fluctuation length in a plausible way.

Another plausibility test of the model is to compare the value of the length scale Λ as inferred from the data with the expression in terms of flow parameters that is given by Eq. 7.21. The right hand side of that equation may be expressed in terms of pore size p and Peclet number P, as shown in [8]. As a ballpark figure, we take the pore diameter as $p=10^{-4}$ m and the initial Gaussian plume variance as given by $S=10^{-2}$ m. Then the Λ value of 10 m as used above implies a longitudinal Peclet number of $P=5$, which compares well with measured values typically in the range 1–10. Most likely the actual values of P, p, and S in the various experiments from which data are collected for Figs. 7.1 and 7.9 were all different, which may account for some of the variability in the measured values.

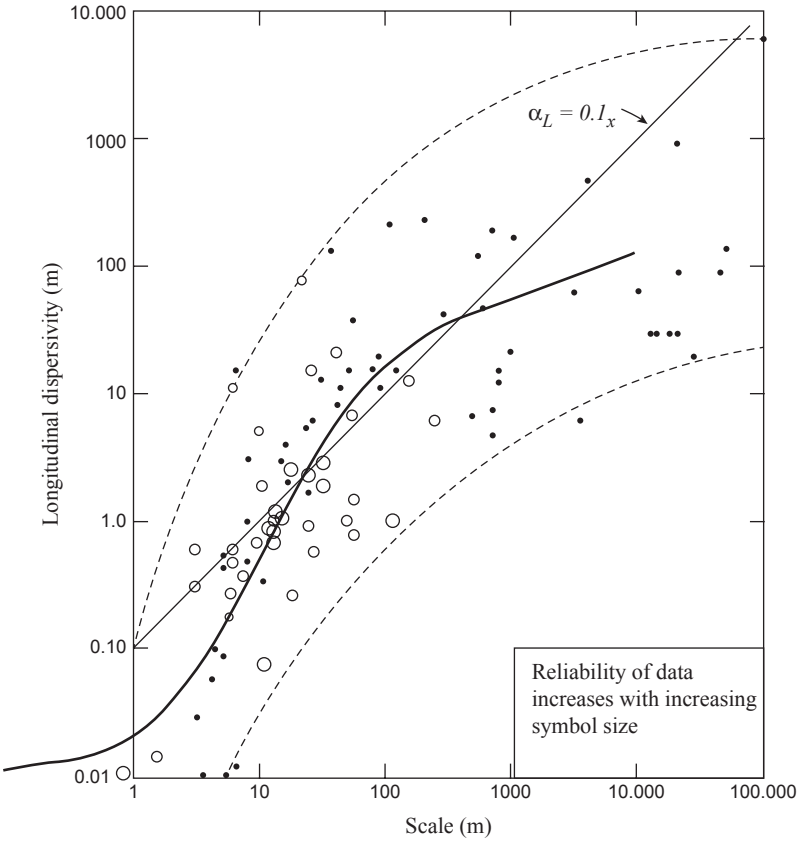

Fig. 7.9 Scaled dispersivity (*solid curve*) as a function of scaled traversal length ($\lambda = 0.0035, Q = 0.32,\ w = 24$, calculated from Eq. 25, compared with experimental data of Fig 1)

7.7 Conclusion

The development as summarized in this chapter of a detailed, semi-analytical stochastic model of the dispersion effects of macroscopic drift velocity fluctuations, leads to significant insights.

One of these is that intrinsic dispersion is increased by a fluctuation, beyond the value associated with flow at the mean drift velocity. A second is that this enhancement manifests as a factor multiplying the spatial variance of the solute plume, so that the effects of a sequence of fluctuations accumulate as a product. This implies an exponential rise of dispersion with the distance travelled as a solute plume traverses the fluctuation sequence.

However, a third observation is that this behaviour is tempered by an annealing effect downstream of a velocity step, which has a length scale related to plume

extension. This happens because of the way concentration gradients are changed by kinetic effects associated with velocity change, and so is not specific to a sharp step change. As a result, the enhancement produced by a fluctuation depends on its length in relation to the plume extension, and this relationship causes the enhancement produced by each fluctuation to decline with traversal length. The decline has a reverse sigmoid shape, characterised by an overall traversal length scale Λ.

The combined effects of the productwise accumulation of declining-dispersion enhancements, is that dispersion at first rises exponentially, but when the traversal length approaches Λ there is a transition to a slower power law rise.

The described behaviour and interpretation is demonstrated in detail specifically for the 1D stepwise velocity fluctuation model. This is obviously highly simplified, but the qualitative and considerable quantitative agreement with experimental observations of dispersion in large-scale natural aquifers suggests that the simple model captures the essential mechanisms. The interpretations set out above do not rely on the presence of sharp steps; indeed key features are also displayed by the exactly solved case of a linear drift velocity growth or decline. This is further supported by the observation that broad agreement is obtained between the exact solution and a stepped "staircase" approximation to it.

Quantitatively, the 1D step model exhibits one major shortcoming: it underestimates the extent of dispersivity growth in the low traversal length range. To support the hypothesis that this does not invalidate the broader interpretation, it was shown that one free parameter can be introduced into the model, in a way that is compatible with its overall structure, and is enough to eliminate the discrepancy.

Each of the parameters in the model has a particular physical interpretation, allowing them to be fixed at physically meaningful values. The implication of the value obtained for the only freely adjustable parameter introduced for the final fit is that the simplified model underestimates the effect of fluctuations encountered early in the sequence.

Causes for this underestimate may plausibly include both simplifying assumptions made to facilitate the analysis, and intrinsic inadequacy of a 1D model. This is seen as the main outstanding issue to be addressed in future elaborations of the model. Some refinements of the 1D model are still possible, but it seems likely that the major source of larger enhancement will be from additional mechanisms for interaction between dispersion and velocity fluctuations coming into play in 2- and 3D systems.

References

1. Fetter CW (1999) Contaminant Hydrogeology. Prentice-Hall, London
2. Lallemand-Barres P, Peaudecerf P (1978) Bulletin, Bureau de Recherches Géologiques et Miniéres 3/4. pp 277–284
3. Gelhar LW (1986) Stochastic hydrology: from hydrology to application. Water Resour Res 22:135–145
4. Øksendal B (1998) Stochastic differential equations—an introduction with applications. Springer Verlag, Berlin

5. Ghanem RG, Spanos PD (1991) Stochastic finite elements: a spectral approach. Springer Verlag, Berlin
6. Verwoerd WS (2012) Longitudinal scale effects on solute dispersion in porous flow, resulting from the cumulative effects of velocity fluctuations. In: Kumar N (ed) International conference on modeling and simulation of diffusive processes and applications, vol 1. BHU, Varanassi pp 1–10
7. Verwoerd WS (2009) New stochastic model for dispersion in heterogeneous porous media: 1. Application to unbounded domains. Appl Math Model 33, 605–625
8. Verwoerd WS (2011) New stochastic model for dispersion in heterogeneous porous media: 2. Gaussian plume transmission across stepwise velocity fluctuations. Appl Math Model 25, 3355–3386
9. Kulasiri D, Verwoerd W (2002) Stochastic dynamics—modeling solute transport in porous media. In: Achenbach JD (ed) North Holland series in applied mathematics and mechanics, vol 44. Elsevier, New York
10. Verwoerd W, Kulasiri D (2002) Solute dispersion in porous flow with a constant velocity gradient. In: Ubertini L (ed) IASTED applied simulation and modelling. Actapress, Calgary, pp 501–505
11. Gelhar LW, Gutjahr AL, Naff RL (1979) Stochastic analysis of macrodispersion in a stratified aquifer. Water Resour Res 15: 1387–1397W. S. Verwoerd

Chapter 8
Modeling Nitrogen Fate and Transport at the Sediment-Water Interface

Mohamed M. Hantush and Latif Kalin

8.1 Introduction

Diffusive mass transfer at media interfaces exerts control on the fate and transport of pollutants originating from agricultural and urban landscapes and affects the conditions of water bodies. Diffusion is essentially a physical process affecting the distribution and fate of various environmental pollutants such as nutrients, pesticides, metals, PCBs, PAHs, etc. Environmental problems caused by excessive use of agricultural chemicals (e.g., pesticides and fertilizers) and the improper discharge of industrial waste and fuel leaks are all influenced by the diffusive nature of pollutants in the environment. Eutrophication is one such environmental problem where the sediment-water interface exerts a significant physical and geochemical control on the eutrophic condition of the stressed water body. Exposure of streams and lakes to contaminated sediment is another common environmental problem whereby transport of the contaminant (PCBs, PAHs, and other organic contaminants) across the sediment water can increase the risk for exposure to the chemicals and pose a significant health hazard to aquatic life and human beings.

Eutrophication is a worldwide problem that is caused by excessive nutrient loading of water bodies. Dissolved and particle-bound nutrients washed away by agricultural and urban runoff ultimately find their way to the bottom of rivers, lakes, and oceans. The degradation of organic matter in bed sediments under aerobic and anaerobic conditions—a process also known as mineralization—produces porewater mineral nitrogen and phosphorus, which upon release diffuse to the water column and contribute to algal bloom and eutrophication of the water body. Aerobic and anaerobic decomposition of organic matter and oxidation reactions in bed sediments, such as nitrification and methane oxidation, deplete oxygen in the sediments

M. M. Hantush (✉)
U.S. EPA National Risk Management Laboratory, Cincinnati, USA
e-mail: hantush.mohamed@epa.gov

L. Kalin
School of Forestry and Wildlife Sciences, Auburn University, Auburn, USA
e-mail: kalinla@auburn.edu

S.K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_8,

and exert oxygen demand on the overlying water column (i.e., sediment oxygen demand, SOD) [1]. Depletion of oxygen in the water body ultimately leads to the well-known hypoxia problem.

The sediment-water interface also plays a significant role in nitrogen transformation and cycling in wetlands which are recognized for their significant ecologic and economic values, owing to the wide variety of services they offer at ecosystem and watershed levels. The importance of wetlands to water quality purification, flood control, wildlife habitats, and biodiversity is well known [2, 3]. The physical and biochemical processes governing the behavior of nitrogen in wetland soils mixed with plant biomass detritus are to some extent similar to the processes in eutrophic-lake bed sediments.

This chapter presents analytical and numerical models describing the fate and transport phenomena at the sediment-water interface in freshwater ecosystems, with the primary focus on nitrogen cycling and the applicability of the models to real-world environmental problems and challenges faced in their applications. The first model deals with nitrogen cycling at the bottom of lakes, a problem that is relevant to the eutrophication of water bodies. The second model addresses nitrogen dynamics in flooded wetlands. For a thorough treatment of analytical solutions for various diffusion problems, interested readers may refer to [4].

8.2 Nitrogen Cycling in Bed Sediments

8.2.1 Sediment Nitrogen Processes

Figure 8.1 depicts processes of nitrogen loading, transformation, and diffusion in bed sediments. The figure shows an active sediment layer wherein the top oxygenated layer separates the overlying water from the underlying anaerobic zone. Usually, the aerobic layer is much thinner than the anaerobic layer. The thickness of the aerobic layer is determined according to the penetration depth of oxygen into the sediments [1]. The sediment layer receives a flux of particulate organic matter, part of which decomposes—a process referred to as mineralization to produce dissolved (pore-water) ammonium ion (NH_4^+). NH_4^+ generated in the anaerobic layer diffuses upward into the aerobic layer, partly undergoes nitrification to produce nitrate, and partly escapes to the water column by diffusion through the boundary layer. Nitrate (NO_3^-) produced by nitrification in the aerobic layer partly diffuses into the underlying anaerobic layer and removed in the process of denitrification, and partly escapes into the water column by diffusion across a resistive boundary layer (d). Advection, bioturbation, and resuspension cause physical mixing and could influence the transport and fate of nitrogen at bed sediments. In the model presented below, these physical processes are not accounted for explicitly, rather approximately by exaggerating diffusive mass transfer across the sediment-water interface.

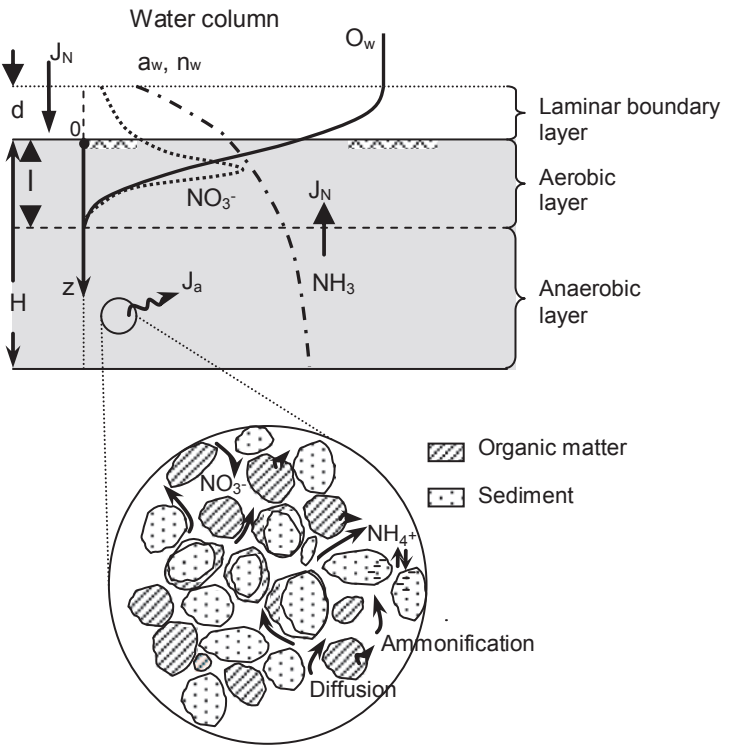

Fig. 8.1 Schematic illustration of nitrogen processes at the sediment-water interface (the active sediment layer has two distinct zones, a thin aerobic zone overlying a thicker anaerobic zone, from [6] with permission from ASCE, NH_3 denotes total ammonia nitrogen)

8.2.2 Ammonia Nitrogen Model

Ammonia formation (ammonification), oxidation (nitrification), and diffusion in a two-layer system may be described by the coupled boundary-value problem [5]:

$$\phi D_a \frac{\partial^2 a_1}{\partial z^2} - \phi k_a a_1 = 0, \quad 0 \le z \le l, \tag{8.1}$$

$$\phi R \frac{\partial a_2}{\partial t} = \phi D_a \frac{\partial^2 a_2}{\partial z^2} + J_a(t), \quad l \le z \le H, \tag{8.2}$$

subject to the following initial and boundary conditions:

$$-\phi D_a \frac{\partial a_1}{\partial z}(0,t) = D_a^* \frac{a_w(t) - a_1(0,t)}{d} \tag{8.3a}$$

$$\frac{\partial a_1}{\partial z}(l,t) = \frac{\partial a_2}{\partial z}(l,t) \tag{8.3b}$$

$$\frac{\partial a_2}{\partial z}(H,t) = 0 \tag{8.3c}$$

$$a_1(l,t) = a_2(l,t) \tag{8.3d}$$

$$a_2(z,0) = a_0(z), \tag{8.3e}$$

where $a_1(z, t)$ is pore-water ammonia-nitrogen ($NH_4^+ + NH_3$) concentration in the aerobic layer [ML^{-3}], $a_2(z, t)$ is pore-water ammonia-nitrogen concentration in the anaerobic layer [ML^{-3}], $a_0(z)$ is initial pore-water ammonia-nitrogen concentration in the active sediment layer [ML^{-3}], z is sediment depth [L], t is time [T], J_a is ammonia-nitrogen rate of production in the sediment layer [$M\,T^{-1}$], R is the retardation factor, k_a is first-order reaction rate for nitrification [T^{-1}], D_a^* is free-water diffusion coefficient of ammonia nitrogen [L^2T^{-1}], D_a = is the sediment ammonia diffusion coefficient [L^2T^{-1}], a_w (t) is ammonia-nitrogen concentration in the water column [ML^{-3}], l is thickness of the aerobic layer [L], d is thickness of laminar boundary layer [L], and H is thickness of the active sediment layer [L].

The retardation factor R accounts for partitioning of NH_4^+ onto negatively charged sediment particles and is given by [5]:

$$R = 1 + f\rho_b \frac{K_d}{\phi} \frac{[H^+]}{[H^+] + K_{eq}}, \qquad K_{eq} = \frac{[NH_3][H^+]}{[NH_4^+]}, \tag{8.4}$$

where $\rho_b = (1-\phi)\rho_s$ is the sediment bulk density [ML^{-3}], ρ_s is the sediment particle density [ML^{-3}], f is the fraction of sediment with negatively charged surfaces (e.g., clay fraction), K_d is the distribution coefficient [L^3M^{-1}], $[H^+] = 10^{-pH}$ is the hydrogen ion concentration, and K_{eq} is equilibrium constant. K_{eq} may be related to the temperature empirically [7, 8].

The gradient boundary condition (Eq. 8.3a) maintains continuity of diffusive mass flux across a resistive boundary layer; it accounts for ammonia concentration in the overlying water, $a_w(t)$, and implicitly assumes a linear concentration profile within the boundary layer.

8.2.3 Nitrate Model

Nitrate production and diffusion in the aerobic layer is given by:

$$\phi D_n \frac{\partial^2 n_1}{\partial z^2} + \phi k_a a_1 = 0, \quad 0 \le z \le l. \tag{8.5}$$

Similarly, assuming diffusion across a laminar boundary, conservation of mass flux at the sediment surface requires:

$$-\phi D_n \frac{\partial n_1}{\partial z}(0,t) = D_n^* \frac{n_w(t) - n_1(0,t)}{d} \tag{8.6a}$$

$$n_1(l,t) = 0 \tag{8.6b}$$

where $n_1(z, t)$ is the pore-water nitrate-nitrogen concentration in the aerobic layer [ML^{-3}], D_n is the sediment nitrate diffusion coefficient [L^2T^{-1}], D_n^* is the free-water nitrate diffusion coefficient [L^2T^{-1}], and $n_w(t)$ is the nitrate-nitrogen concentration in the water column [ML^{-3}]. The premise of boundary condition (B. C.) (Eq. 8.6b) is that nitrate is completely removed from the anaerobic layer.

A more comprehensive model including methane production, oxidation, and gas formation was presented and solved in [5]. In addition to accounting for transient oxic zone thickness, the challenge there was modeling a dynamically evolving methane saturation zone which propagates from the bottom of the active sediment layer upward towards the sediment-water interface.

8.2.4 Organic Nitrogen Model

Mass balance of the organic matter delivered to the active sediment layer is:

$$H \frac{dM_i}{dt} = f_i J_N - k_{m_i} H M_i, \quad i = 1, 2 \tag{8.7}$$

where M is the concentration of decomposable organic matter nitrogen [ML^{-3}], J_N is the rate at which particulate organic matter nitrogen is delivered to the active sediment layer [$ML^{-2}T^{-1}$] including organic nitrogen delivered by the death of nitrogen-fixing microorganisms, k_m is the first-order rate of organic matter decomposition [T^{-1}]. The subscripts $i = 1, 2$ refer to fast and slowly decaying organic matters having fractions f_1 and f_2, respectively. Ammonia flux, J_a is related to the rate at which organic matter is mineralized:

$$J_a(t) = k_{m_1} M_1(t) + k_{m_2} M_2(t). \tag{8.8}$$

The solutions of the coupled boundary value problem (Eqs. 8.1–8.8) are obtained by first applying the Laplace transformation then using Green's functions. From the analytical solutions, closed expressions for fluxes of *a* and *n* are obtained. For brevity and to conserve space only the steady-state solutions are shown below. The readers may refer to [5] for the detailed analysis.

Ammonia

$$a_1(z) = -\frac{H-l}{\phi D_a \delta_a} J_a \sinh \delta_a (l-z) + \left(\frac{\gamma_a}{\psi_4} a_w + \frac{H-l}{\phi D_a \delta_a} \frac{\psi_3}{\psi_4} J_a \right) \cosh \delta_a (l-z) \tag{8.9}$$

$$a_2(z) = \frac{\gamma_a}{\psi_4} a_w + \frac{H-l}{\phi D_a \delta_a} \frac{\psi_3}{\psi_4} J_a + \frac{1}{2} \frac{J_a}{\phi D_a} \left[(H-l)^2 - (H-z)^2 \right] \tag{8.10}$$

where

$$\psi_3 = \gamma_a \sinh \delta_a l + \delta_a \cosh \delta_a l \tag{8.11}$$

$$\psi_4 = \gamma_a \cosh \delta_a l + \delta_a \sinh \delta_a l \tag{8.12}$$

$$\gamma_a = \frac{D_a^*}{\phi D_a d} \tag{8.13}$$

$$\delta_a = \sqrt{\frac{k_a}{D_a}}. \tag{8.14}$$

Nitrate

$$n_1(z,t) = \frac{1}{1+\gamma_n l} \left\{ \gamma_n n_w(t) + \frac{k_a}{D_n \delta_a} (\psi_2 \Lambda_1 - \psi_1 \Lambda_2) \right\} (l-z) - \frac{k_a}{D_n \delta_a^2} \left\{ \Lambda_1 \sinh \left[\delta_a (l-z) \right] + \Lambda_2 (\cosh \left[\delta_a (l-z) \right] - 1) \right\} \tag{8.15}$$

where

$$\gamma_n = \frac{D_n^*}{\phi D_n d} \tag{8.16}$$

$$\psi_1 = \frac{\gamma_n}{\delta_a} \left[1 - \cosh(\delta_a l) \right] - \sinh(\delta_a l) \tag{8.17}$$

$$\psi_2 = \cosh(\delta_a l) + \frac{\gamma_n}{\delta_a} \sinh(\delta_a l) \tag{8.18}$$

$$\Lambda_1 = -\frac{H-l}{\phi D_a \delta_a} J_a \tag{8.19}$$

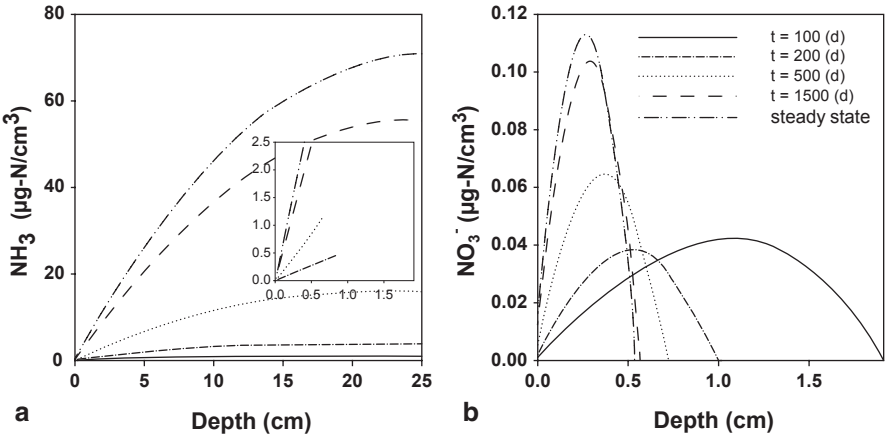

Fig. 8.2 Simulated concentration profiles in time: **a** total ammonia nitrogen in active sediment layer (*inner panel* is a blowout of simulated profiles in the aerobic layer), and **b** nitrate in the aerobic layer (adapted from [5]), NH_3 denotes total ammonia nitrogen

$$\Lambda_2 = \frac{\gamma_a}{\psi_4} a_w + \frac{H-l}{\phi D_a \delta_a} \frac{\psi_3}{\psi_4} J_a. \tag{8.20}$$

Numerical evaluation of the transient analytical solutions for a and n is iterative due to the dependence of l on time, and are computationally demanding as they require nonlinear roots identification and the evaluation of infinite series. A simpler non-iterative model is presented below to compute $l(t)$ independent of the solutions for a and n, but at the expense of an additional calibration parameter.

Figure 8.2 shows the evolution in time of a and n concentrations with depth for a hypothetical sediment layer using the input and parameter values reported in the literature. The concentration profiles asymptotically approach the steady-state solutions given above. The concentration profiles in the inner window in Fig. 8.2a are for a in the aerobic layer. Visually, the interface $z=l$ is shifting to the left and the aerobic layer is getting thinner in time. For each profile, the concentration of a and its derivative at $z=l$ appear continuous. Figure 8.2b depicts concentrations of n in a shrinking aerobic layer; note the zero concentration at $z=l$, $n(l,t)=0$ (B.C. Eq. 8.6b).

8.3 Sediment Oxygen Dynamics

The oxygen demand exerted by the sediment layer due to oxidation reactions in the aerobic layer, $SOD(t)$, is equal to the oxygen flux from the water column to the sediment layer:

$$SOD(t) = -\phi D_o \frac{\partial O}{\partial z}(0,t), \tag{8.21}$$

in which D_o is the sediment pore-water diffusion coefficient of oxygen [L^2T^{-1}], O and O_w are oxygen concentration in interstitial pore-water and the water column, respectively. A mathematical expression for the thickness of aerobic zone, l, can be obtained by conserving the oxygen mass in the laminar and oxic layers [5]. However, a simpler model is presented here where all oxygen-consuming processes are lumped in a zero-order reaction rate term.

The boundary value problem for oxygen penetration is:

$$\phi D_o \frac{\partial^2 O(z,t)}{\partial z^2} - S = 0, \tag{8.22}$$

$$-\phi D_o \frac{\partial O}{\partial z}(0,t) = D_o^* \frac{O_w(t) - O(0,t)}{d} \tag{8.23a}$$

$$\frac{\partial O}{\partial z}(l,t) = 0, \quad O(l,t) = 0 \tag{8.23b}$$

where S is the sediment oxygen consumption rate [$ML^{-3}T^{-1}$]. The above boundary value problem assumes quasi-steady oxygen diffusion in the aerobic layer and can be solved for $l(t)$ [6],

$$l(t) = -\tau\phi d + \sqrt{\tau^2\phi^2 d^2 + \frac{2\phi D_o O_w(t)}{S}}. \tag{8.24}$$

The reaction rates in all of the equations above generally increase with temperature and are updated according to the Arrhenius equation: $k(T_2) = k(T_1)\theta^{(T_2 - T_1)}$ where θ is a constant. The integration of Eq. 8.22 from $z = 0$ to $z = l$ and making use of boundary conditions (Eqs. 8.23a and 8.23b) yields the obvious relationship:

$$SOD(t) = Sl(t). \tag{8.25}$$

8.4 Application to Chesapeake Bay Sediment Flux Data

In this section, the nitrogen sediment-flux model described above is applied to several sites in the Chesapeake Bay, USA. The sediment flux data was collected between March 1985 and December 1988 and is tabulated in [1]. To better understand the dominant processes in the Bay, identify the most sensitive parameters, and estimate the uncertainty in the estimated fluxes, the generalized sensitivity analysis (GSA) first proposed by Spear and Hornberger [9] can be applied. For this purpose, prior probability distributions can be proposed for all model parameters by relying on information from the literature. Monte Carlo (MC) simulations are conducted by generating 100,000 parameter sets from the assumed prior distributions and simulating a corresponding number of ammonia, nitrate, and *SOD* flux time series. The Nash-Sutcliffe efficiency criterion (E_{NS}) [10] is used to evaluate the model

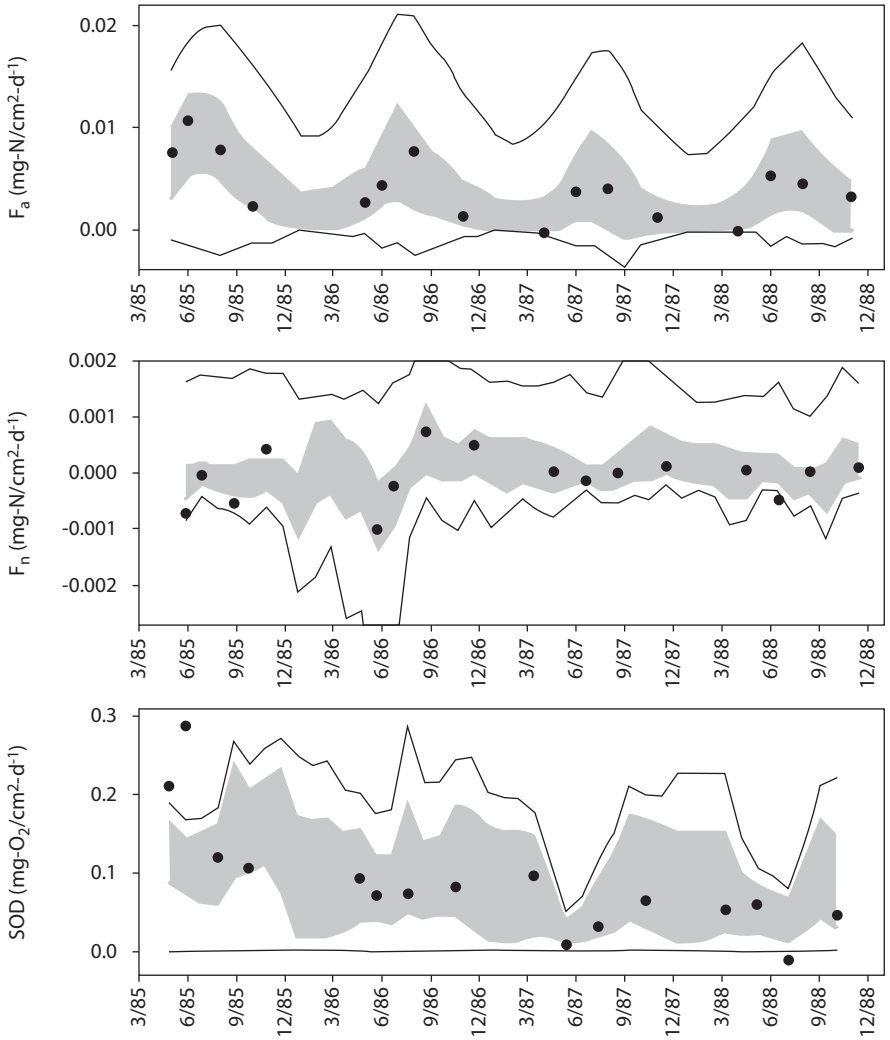

Fig. 8.3 Simulated and measured fluxes of ammonia (F_a), nitrate (F_n), and sediment oxygen demand (*SOD*) at site PP (*gray band* denotes the behavior set results, lines bound the entire set of (100,000) MC simulations)

performance. Parameter sets that produced the top 1,000 model performances (i.e., having the highest E_{NS}) were considered as behavioral (B). The remaining parameter sets are classified as nonbehavioral (B').

Figure 8.3 compares MC-simulated time series with observed ammonium, nitrate, and *SOD* fluxes (F_a, F_n, and *SOD,* respectively) at a selected site (PP). In each figure, the gray area corresponds to the ensemble of behavior parameter set (B) simulations. These simulations involved fine-tuning of the source loading rate (J_N) which is treated as unknown since no direct measurements of this variable are reported.

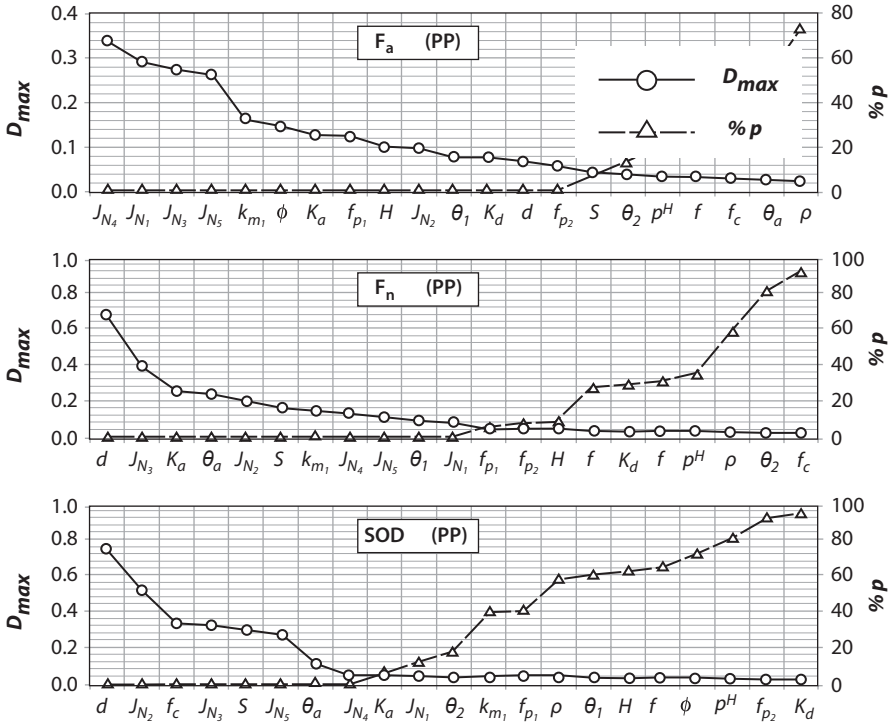

Fig. 8.4 Order of parameter sensitivities for ammonia (F_a), nitrate (F_n), and sediment oxygen demand (*SOD*) at site PP, based on the K–S test

Overall, the model successfully captures the variability in the measured ammonia and nitrate fluxes at the sediment-water interface ($z=0$). The model especially does a good job in predicting F_a. Note that the observed data shows changes in direction of the nitrate flux. At times, the flux is from the water column to the sediment, and at other times from the bed sediments to the water column. The model is able to capture such flux reversals. The discrepancy between simulated F_n and measurements could potentially be attributed to the measurement errors and model deficiencies, such as exaggerating nitrate removal by denitrification in the anaerobic zone (i.e., assuming zero concentration therein) $n_1(l, t)=0$, and/or neglecting denitrification in the aerobic layer itself; for example, in anoxic microsites where the interior of aggregate organic particle may be anaerobic even if the exterior is aerobic [1, 11].

The ability of the model to capture variability of measured *SOD* is relatively poorer. The model-computed uncertainty bands appear also wider in *SOD* predictions, compared to F_a and F_n. It is not clear if the discrepancy between laboratory-measured *SOD* and in situ real values has contributed to such a poor performance. However, the challenges often encountered in accurate on-site measurements of *SOD* are known [12].

Figure 8.4 shows the sensitive parameters for F_a, F_n, and *SOD* for the same site. Order of sensitivities decreases from left to right; this is based on the D_{max} param-

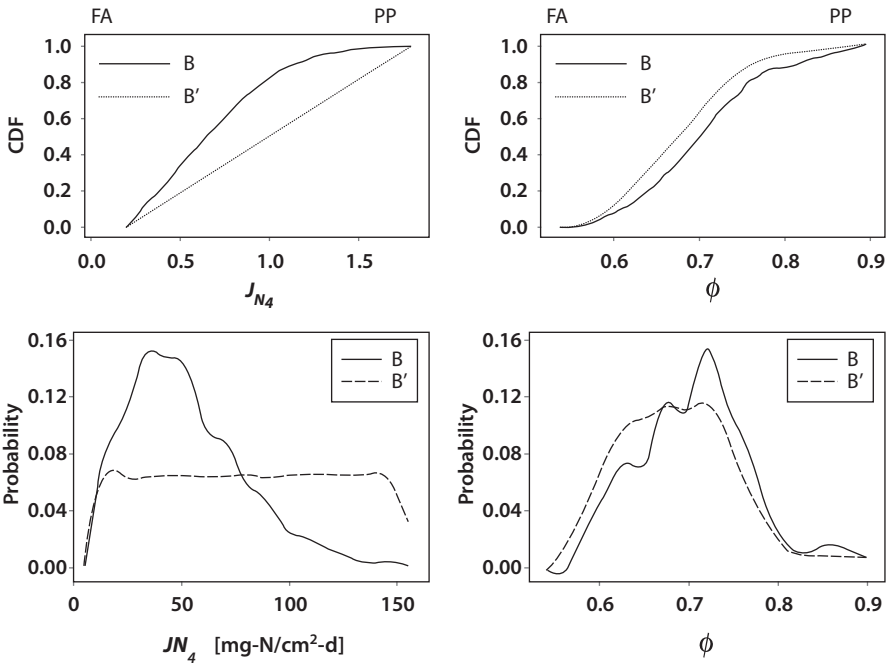

Fig. 8.5 Example CDFs (*top panel*) and PDFs (*bottom panel*) of behavior (*B*) and nonbehavior (*B*') parameter sets for ammonia (F_a) at site PP

eter of the nonparametric Kolmogorov–Smirnov (K–S) test. D_{max} is the maximum deviation between the cumulative distributive functions (CDFs) of the behavior and nonbehavior parameter sets. Figure 8.5 shows two example CDFs, and the corresponding probability density functions (PDFs) for F_a. It is obvious that J_{N4} is a more sensitive parameter than ϕ, because D_{max} of J_{N4} is $>D_{max}$ of ϕ. PDFs of *B* and *B*' also differ significantly in J_{N4} compared to ϕ. In Fig. 8.4 J_{N1}, J_{N2}, J_{N3}, and J_{N4} are parameters related to organic matter deposition in years 1, 2, 3, and 4, respectively (1985–1988). As can be seen, the model is consistently very sensitive to organic matter deposition rates. This is not too surprising as organic matter introduced to the sediment layer through deposition and decay of nitrogen-fixing microorganisms is the primary source of nitrogen to the system. The laminar boundary layer thickness, *d* is the most sensitive parameter with F_n and *SOD,* which indicates that the movement of nitrate and oxygen between the water column and the sediment layer is dominated by diffusive processes.

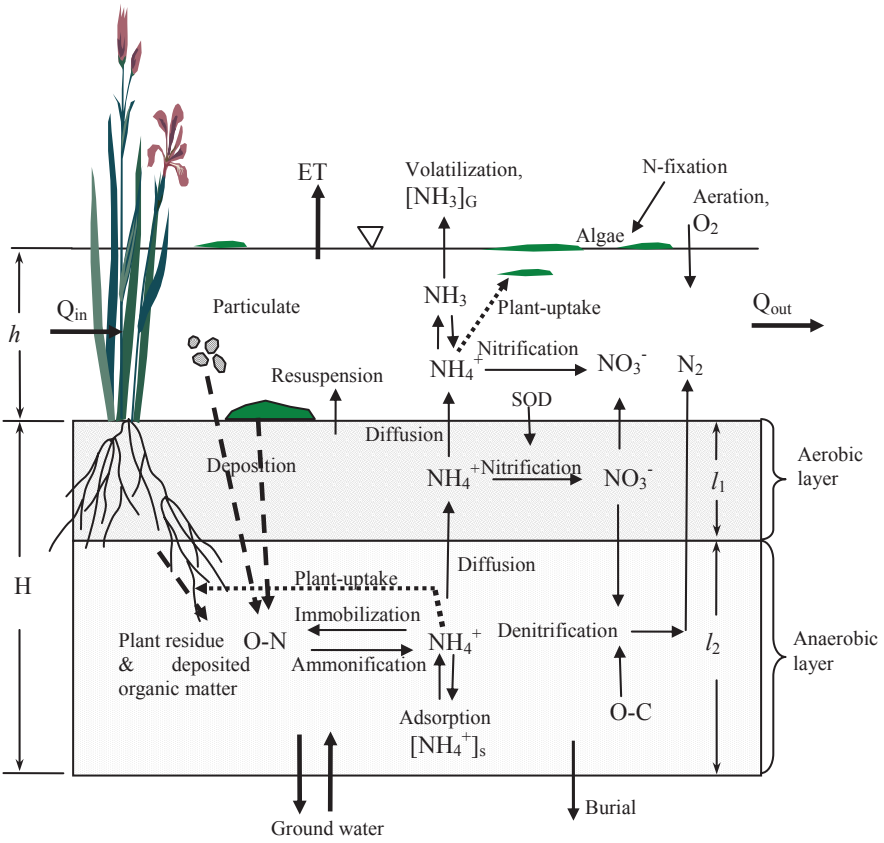

Fig. 8.6 Schematic diagram of nitrogen processes in wetlands: water column, aerobic soil layer, and reduced lower soil layer. ([13] with permission from ASCE)

8.5 Wetland Nitrogen Cycling

8.5.1 Nitrogen Processes

Figure 8.6 depicts the conceptual model for complete biogeochemical pathways of mineralization of organic matter to ammonia and subsequent transport, retention, uptake, and removal (denitrification, volatilization, and burial) in flooded wetlands [13]. The model partitions a wetland into three basic compartments, the water column (free water), wetland soil layer, and plant biomass. The soil layer is further partitioned into aerobic and anaerobic zones [3, 14]. The aerobic layer at the soil-water interface is not a fixed layer and its thickness is determined by the supply of oxygen to the soil surface and consumption of oxygen in the soil [15].

While particulate organic nitrogen mineralizes slowly (stable) and fast (labile), a fraction may resist decomposition. The interactions between the free-water and

soil compartments occur as a result of settling and resuspension of particulate matter, and advective and diffusive mass exchange of dissolved constituents. Burial is caused by net sediment accumulation and thus modeled with advection (at the burial velocity) across the bottom of the active sediment layer [1].

Sources of ammonia and nitrate to the wetland water column include agricultural and urban runoff, groundwater discharge, mineralization of suspended organic nitrogen, sediment feedback (diffusion and resuspension), and atmospheric depositions. Atmospheric deposition and nitrogen gas (N_2) fixation by microorganisms are other sources of nitrogen inputs to wetlands, especially for bogs in the northeast of the USA [16]. Nitrification of ammonium nitrogen occurs in the aerobic part of the soil and the water column, whereas nitrate removal by denitrification in this analysis is confined to the underlying anaerobic zone of the active soil layer. However, both nitrification and denitrification can coexist in the aerobic layer; for example, the anaerobic interior of aggregate organic particles [1]. Nitrification also occurs near roots in the rhizosphere of wetland plants and can be as significant as at the soil surface [17]. Dissociation of NH_4^+ into ammonia gas (NH_3) and subsequent volatilization to the atmosphere is a significant loss pathway for nitrogen under conditions of high alkalinity [15]. In addition to influent concentrations, nitrate (NO_3^-) is produced by oxidation of NH_4^+ in the water column and oxidized soil layer.

Although the primary objective of the wetland model is nutrient cycling, productivity can be modeled using a generic mass balance for free-floating algae and rooted plants with relatively simple growth and death processes.

8.5.2 Wetland Model Equations

Hydrology, mass balance of nitrogenous species, sediment and oxygen dynamics in both free-water and sediment compartments are accounted for in the ordinary differential equations that are presented below [13].

Hydrology

Surface flow routing in a wetland system can be described using a simple flow continuity equation:

$$\phi_w \frac{dV_w}{dt} = Q_i + Q_g - Q_o - AE_T + Ai_p \tag{8.26}$$

where V_w is the water volume of wetland surface water [L^3], A is the wetland surface area [L^2], Q_i is the volumetric inflow rate [L^3T^{-1}], Q_g is groundwater discharge (negative for infiltration) [L^3T^{-1}], Q_o is wetland discharge (outflow) rate [L^3T^{-1}], i_p is the precipitation rate [LT^{-1}], E_T is the evapotranspiration rate [LT^{-1}], and ϕ_w is the effective porosity of wetland free water (since biomass occupies part of the submerged wetland volume).

Outflow-depth relationship (rating curve) of the form $Q_o = \rho h^{\varepsilon}$, where $h = \phi_w V_w / A$, can be used to route flow out of the wetland for a given inflow event, Q_i. The specific case of $\varepsilon = 1$ corresponds to a linear-reservoir model.

Nitrogen Cycling

Water Column

$$\begin{aligned}\phi_w \frac{d(V_w N_{aw})}{dt} &= Q_i N_{awi} + i_p A N_{ap} - \phi_w V_w f_N k_{nw} N_{aw} \\ &+ \beta_{a1} A(N_{a1} - N_{aw}) + F^w_{N_a g} \\ &- k_v \phi_w A(1 - f_N) N_{aw} + \phi_w V_w k_{mw} N_{ow} \\ &- Q_o N_{aw} - f_{aw} a_{na} k_{ga} a + A q_a\end{aligned} \tag{8.27}$$

$$\begin{aligned}\phi_w \frac{d(V_w N_{nw})}{dt} &= Q_i N_{nwi} + i_p A N_{np} + \phi_w V_w f_N k_{nw} N_{aw} \\ &+ \beta_{n1} A(N_{n1} - N_{nw}) + F^w_{N_n g} \\ &- f_{nw} a_{na} k_{ga} a - Q_o N_{nw} + A q_n\end{aligned} \tag{8.28}$$

in which

$$k_{nw} = k^*_{nw}(1 - e^{-\lambda_w O_w}) \tag{8.29}$$

$$F^w_{xg} = \begin{cases} Q_g x_1, & Q_g > 0 \\ Q_g x_w, & Q_g < 0 \end{cases}, \quad x : N_a, N_n \tag{8.30}$$

where $N_{aw} = \left[\mathrm{NH_4^+}\right] + \left[\mathrm{NH_3}\right]$ is the total ammonia-nitrogen concentration in free water [$\mathrm{ML^{-3}}$], N_{nw} is the nitrate-nitrogen concentration in free water [$\mathrm{ML^{-3}}$], N_{ow} is the particulate organic nitrogen concentration in free water [$\mathrm{ML^{-3}}$], O_w is the oxygen concentration in free water [$\mathrm{ML^{-3}}$], a is the mass of free-floating plant [M *Chl a*], N_{awi} and N_{nwi}, respectively, are concentrations of total ammonia nitrogen and nitrate nitrogen in incoming inflow [$\mathrm{ML^{-3}}$], N_{a1} and N_{n1}, respectively, are pore-water concentrations of total ammonia nitrogen and nitrate nitrogen in oxygenated top soil layer (aerobic layer in Fig. 8.6) [$\mathrm{ML^{-3}}$], N_{ap} and N_{np}, respectively, are concentrations of total ammonia nitrogen and nitrate nitrogen in precipitation [$\mathrm{ML^{-3}}$], q_a and q_n, respectively, are dry depositional rates of total ammonia nitrogen and nitrate [$\mathrm{ML^{-2}T^{-1}}$], S is the rate of nitrogen fixation by microorganisms [$\mathrm{ML^{-2}T^{-1}}$], $F^W_{N_a g}$ and $F^W_{N_n g}$, respectively, are groundwater source/loss for total ammonia nitrogen and nitrate nitrogen [$\mathrm{MT^{-1}}$], β_{a1} and β_{n1}, respectively, are diffusive mass-transfer rates of total ammonia nitrogen and nitrate between wetland water and aerobic soil

layer [LT^{-1}], and f_N is the fraction of total ammonia in ionized form. For the definition of all other parameters in the above equations and below one may refer to [13].

Equation 8.29 limits nitrification rate to oxygen levels in the water column [18]. In this equation, λ_w is the first-order nitrification inhibition coefficient (≈ 0.6 L mg^{-1}). Hantush and others [13] provided the following expression relating ammonia volatilization rate k_v to wind speed (U_w) using a two-film resistance model and known relationships between liquid-film and gaseous-film exchange coefficients [19]:

$$k_v = \frac{1.17\alpha}{1+12.07\alpha U_w^{\eta-1}} U_w^{\eta} \tag{8.31}$$

where α and η are empirical parameters. In Eq. 8.31, both U_w and k_v are in md^{-1}. For example, for open water bodies such as lakes, $\alpha = 0.864$ [20] and $\eta = 1$. Due to wind shielding nature of green cover, it might be reasonable to assume that $\alpha < 0.864$. The dependence of k_v on temperature and pH was considered by Wang and others [21]. In Eq. 8.27 and as shown by Eq. 8.31, the effective volatilization rate expression $k_v(1-f_N)$ accounts for wind speed in addition to the above two parameters.

A general expression for the diffusion related parameter β can be obtained by conserving diffusive mass transfer between the two porous layers:

$$\beta = \frac{2\phi_1\phi_2 D_1 D_2}{\phi_2 D_2 l_1 + \phi_1 D_1 l_2}, \qquad D_i = \tau_i D^*, \quad i = 1,2 \tag{8.32}$$

where D^* is free-water diffusion coefficient [L^2T^{-1}], and τ_i is the tortuosity of the layer i.

Aerobic Soil Layer

$$\begin{aligned} \phi V_1 R_s \frac{dN_{a1}}{dt} = &-A\beta_{a1}(N_{a1} - N_{aw}) + F^1_{N_a g} - f_{a1} a_{na} k_{gb} f_1 b \\ &-\phi A v_b N_{a1} - \phi V_1 f_N k_{ns} N_{a1} + A\beta_{a2}(N_{a2} - N_{a1}) \\ &+V_1 k_{mr} N_{or} + V_1 k_{ms} N_{os} \end{aligned} \tag{8.33}$$

in which

$$R_s = 1 + \frac{m_s K_d f_N}{\phi} \tag{8.34}$$

$$k_{ns} = k^*_{ns}(1 - e^{-\lambda_s O_w}) \tag{8.35}$$

and

$$\begin{aligned} \phi V_1 \frac{dN_{n1}}{dt} = &-A\beta_{n1}(N_{n1} - N_{nw}) + F^1_{N_n g} + \phi V_1 f_N k_{ns} N_{a1} \\ &- A\beta_{n2}(N_{n1} - N_{n2}) - f_{n1} a_{na} k_{gb} f_1 b - v_b \phi A N_{n1} \end{aligned} \tag{8.36}$$

in which

$$F_{xg}^{1}=\begin{cases}Q_g x_2-Q_g x_1, & Q_g>0\\ Q_g x_1-Q_g x_w, & Q_g<0\end{cases},\quad x: N_a, N_n \tag{8.37}$$

where N_{or} is the concentration of rapidly mineralizing organic nitrogen in wetland soil [ML^{-3}], N_{os} is the concentration of slowly mineralizing organic nitrogen in wetland soil [ML^{-3}], V_1 is the volume of aerobic soil [L^3], R_s is the total ammonia retardation factor in wetland soil, ϕ is the wetland soil porosity, N_{a2} is the total ammonia-nitrogen pore-water concentration in lower anaerobic layer [ML^{-3}], N_{n2} is the nitrate-nitrogen pore-water concentration in lower anaerobic layer [ML^{-3}], b is the mass of the rooted plants [M *Chl a*], v_b is the burial velocity [LT^{-1}], $f_1=l_1/(l_1+l_2)$ is the volumetric fraction of aerobic soil layer, l_1 is thickness of aerobic soil layer [L], l_2 is thickness of anaerobic soil layer [L], $F_{N_a g}^{1}, F_{N_n g}^{1}$ are, respectively, groundwater source/loss of total ammonia nitrogen and nitrate in the aerobic layer [MT^{-1}], m_s is the soil bulk density [ML^{-3}], and K_d is the NH_4^+ distribution coefficient [L^3M^{-1}].

Anaerobic Soil Layer

$$\begin{aligned}\phi V_2 R_s\frac{dN_{a2}}{dt}&=-A\beta_{a2}(N_{a2}-N_{a1})-\phi A v_b(N_{a2}-N_{a1})+F_{N_a g}^{2}\\&\quad+V_2 k_{mr}N_{or}+V_2 k_{ms}N_{os}-f_{a2}a_{na}k_{gb}f_2 b\end{aligned} \tag{8.38}$$

$$\begin{aligned}\phi V_2\frac{dN_{n2}}{dt}&=A\beta_{n2}(N_{n1}-N_{n2})-\phi V_2 k_{dn}N_{n2}-\phi A v_b(N_{n2}-N_{n1})\\&\quad+F_{N_n g}^{2}-f_{n2}a_{na}k_{gb}f_2 b\end{aligned} \tag{8.39}$$

in which

$$F_{xg}^{2}=\begin{cases}Q_g x_G-Q_g x_2, & Q_g>0\\ Q_g x_2-Q_g x_1, & Q_g<0\end{cases},\quad x: N_a, N_n \tag{8.40}$$

where $N_{a\mathrm{G}}$ is the total ammonia-nitrogen concentration in groundwater [ML^{-3}], $N_{n\mathrm{G}}$ is the nitrate-nitrogen concentration in groundwater [ML^{-3}], $f_2=l_2/(l_1+l_2)$ is volumetric fraction of anaerobic layer, V_2 is the volume of anaerobic soil [L^3], β_{a2} and β_{n2}, respectively, are diffusive mass-transfer rates of total ammonia nitrogen and nitrate between aerobic and anaerobic soil layers [LT^{-1}], and $F_{N_a g}^{2}, F_{N_n g}^{2}$ are groundwater source/loss of total ammonia nitrogen and nitrate in the anaerobic layer [MT^{-1}].

Oxygen Dynamics

Water Column

Oxygen variations in the water column can be described by the following equation:

$$\begin{aligned}\phi_w \frac{d\left(V_w O_w\right)}{dt} = {} & Q_i O_{wi} + i_p A O_p + K_o \phi_w A\left(O^* - O_w\right) \\ & - r_{on,m}\phi_w V_w k_{mw} N_{ow} - r_{on,n}\phi_w V_w f_N k_{nw} N_{aw} \\ & + a_{oc} r_{c,chl}\left\{\left(k_{gb} - k_{db}\right) f_{bw} b + \left(k_{ga} - k_{da}\right) a\right\} \\ & - Q_o O_w - A S_O - \phi_w V_w S_w - E_T A O_w \end{aligned} \tag{8.41}$$

where O_{wi} is the concentration of oxygen in the incoming water [ML^{-3}], O_p is the concentration of total oxygen in precipitation [ML^{-3}], O^* is oxygen concentration in the air (assumed at saturation) [ML^{-3}], K_o is oxygen mass-transfer coefficient [LT^{-1}], S_O is the wetland soil oxygen depletion rate per unit area [$ML^{-2}T^{-1}$], and S_w is the volumetric oxygen consumption rate in water by other processes [$ML^{-3}T^{-1}$].

Wetland Soil

Using the equivalence of two-film theory and assuming linear drop of oxygen concentration from the free-water level (O_w) to zero at depth l_1 below soil surface, conservation of oxygen mass flux across the soil-water interface yields the following expression modified for constricted (porous) wetland surface water [5]:

$$S_O = \phi\tau D_o^* \frac{O_w}{l_1 + \dfrac{\phi\tau}{\phi_w \tau_w}\delta} \tag{8.42}$$

where δ is the thickness of a diffusive boundary layer situated on top of the soil-water interface [L], τ is the wetland soil tortuosity factor, τ_w is effective tortuosity of the flooded area above soil, and D^*_0 is the free-water oxygen diffusion coefficient [L^2T^{-1}]. Typical thickness of the diffusive boundary layer, δ, in natural waters (streams, lakes, oceans) is in the order of millimeters [22, 23]. For low energy environment and shallow wetland waters, the boundary layer is relatively thicker and may be approximated as $\delta \approx h/2$.

Oxygen consumption in the aerobic layer can be related to the processes of nitrification, aerobic decomposition of organic matter (mineralization), and other unknown oxidation processes. Conservation of oxygen mass in wetland soil requires:

$$S_O = l_1 r_{on,n}\phi f_N k_{ns}(O_w) N_{a1} + l_1 r_{on,m}(k_{ms} N_{os} + k_{mr} N_{or}) + l_1 S_s \tag{8.43}$$

where S_s is the oxygen removal rate per unit volume of aerobic layer by other processes, $r_{on,m}$ is gram of oxygen consumed per gram of organic nitrogen mineralized (=15.29), and $r_{on,n}$ is gram of oxygen consumed per gram of total ammonium nitrogen oxidized by nitrification (=4.57).

The use of Eqs. 8.42 and 8.43 leads to:

$$l_1 = -\phi\tau\delta + \sqrt{(\phi\tau\delta)^2 + 2\phi\tau D_o^* O_w / \Omega} \tag{8.44}$$

in which

$$\Omega = r_{on,n}\phi f_N k_{ns}(O_w) N_{a1} + r_{on,m}(k_{ms} N_{os} + k_{mr} N_{or}) + S_s. \tag{8.45}$$

For the simple case of $\delta = 0$, Eq. 8.44 reduces to $l_1 = \sqrt{2\phi\tau D_o^* O_w / \Omega}$. Equation 8.44 predicts the thickness of the top oxic soil layer l_1, which is typically much smaller than l_2 (varies from a few mm to 1–2 cm), e.g., [24], thus, $l_2 \approx H$ where H is the thickness of the active soil layer [L].

The equations for N_{ow}, N_{or}, N_{os}, *a, b,* and sediment transport are not displayed here for brevity. On the related equations and definitions of parameters and coefficients, the readers may refer the original source [13].

8.5.3 Wetland Model Application

The nitrogen model outlined above is evaluated using hydrologic and water quality data from a small restored wetland located on Kent Island, Maryland, on the eastern shores of the Chesapeake Bay. Removal of nutrients and suspended solids from this restored wetland, which received unregulated inflows from a 14-ha agricultural watershed, were monitored through automated flow-proportional sampling. Water flow, sediment and nutrient concentrations entering and leaving the wetland were measured from 8 May 1995 through 12 May 1997 [25]. The developed wetland model described above runs at daily time scale. Simulations started on 5/9/1995 and ended on 5/12/1997. Therefore, all the input data are required at daily time scale. The model internally divides the 1-day time interval into a smaller time interval (in this case 0.01 day) for numerical integration. Daily data are interpolated to generate inputs at higher than daily-time resolution. Initial concentrations required to initiate the model are taken from the values of day one, which is May 9th 1995. Weekly (typically 5–8 days) flow averaged nitrate nitrogen, total ammonia nitrogen, organicnitrogen, and total suspended solids (TSS) concentrations in runoff are available from [25]. Conversion of weekly data and reconstruction of missing ones over the monitoring period were carried by [26].

8.5.4 Model Assessment

To evaluate model sensitivity to various parameters, dotty plots and GSA [9] were conducted in [26]. Dotty plots are simple scatter plots where values of a model parameter are plotted against either a selected model output or a model performance measure (e.g., mass balance error) after a MC simulation. Dotty plots not only provide information about sensitive parameters, but also depict the range in which the model is most sensitive to a given parameter. Further, such dotty plots also disclose the optimal ranges or the values of each parameter where the model performs best. In other words, they can be helpful during model calibration. Nash-Sutcliffe coefficient E_N [10] is used as the model performance measure. Dotty plots only provide qualitative measures for model sensitivity, whereas the GSA method, which is applicable only when observed data are available, is a quantitative approach for performing sensitivity analysis. Model sensitivity to a parameter could vary depending on where the parameter is perturbed, which is why GSA is highly recommended over local sensitivity analysis [27]. Kalin et al. [26] combined the generalized likelihood uncertainty estimation (GLUE) [28] and GSA approaches and applied them to the described wetland model to simultaneously perform sensitivity and uncertainty analysis. In the GSA, model parameter sets generated by sampling from proposed prior distributions of the parameters are separated into behavioral B and nonbehavioral B' parameter sets using a threshold E_N value, often subjectively selected. For example, if the threshold value is 0.2, B is composed of parameters sets whose $E_N \geq 0.2$. Parameters sets in B are valid simulators of the wetland system being modeled, whereas parameters sets in B' are not. The former defines the ensemble of model prediction, which is only a measure of predictive uncertainty and thus should be distinguished from probabilistic-based confidence intervals. The latter can be inferred using formal Bayesian analysis.

Figure 8.7 shows the dotty plots of E_N values for nitrate (top row) and ammonia (bottom row) loadings for some selected parameters. Each figure contains 100,000 points. Nitrate, as expected, is very sensitive to the denitrification parameter, k_{dn}, which indicates that denitrification is a major nitrogen loss pathway in this study wetland. Nitrate also exhibits high degree of sensitivity to the Arrhenius coefficient θ, with higher temperatures stimulating the nitrate reactions and rate of diffusion. Diffusion of nitrate between the water column and the sediment layer is another important process in the study wetland as captured by the parameters β_{n1} and l_2. For total ammonia, the model seems to perform better for higher a_{na} (gram of nitrogen per gram of chlorophyll a in algae/plants) and f_N (fraction of total ammonia nitrogen as ammonium nitrogen) values. The parameter a_{na} is linked to plant uptake of ammonia while higher f_N indicates that the system is high in NH_4 and low in NH_3 (high pH). Dotty plots further reveal that mineralization of organic matter into ammonia and diffusion of ammonia from bed sediments to the water column are the other likely key processes in the study site. Note that the striking difference between the left two figures and the right two figures on each row is caused by the selected prior PDFs of the parameters. Left ones are generated from a uniform distribution

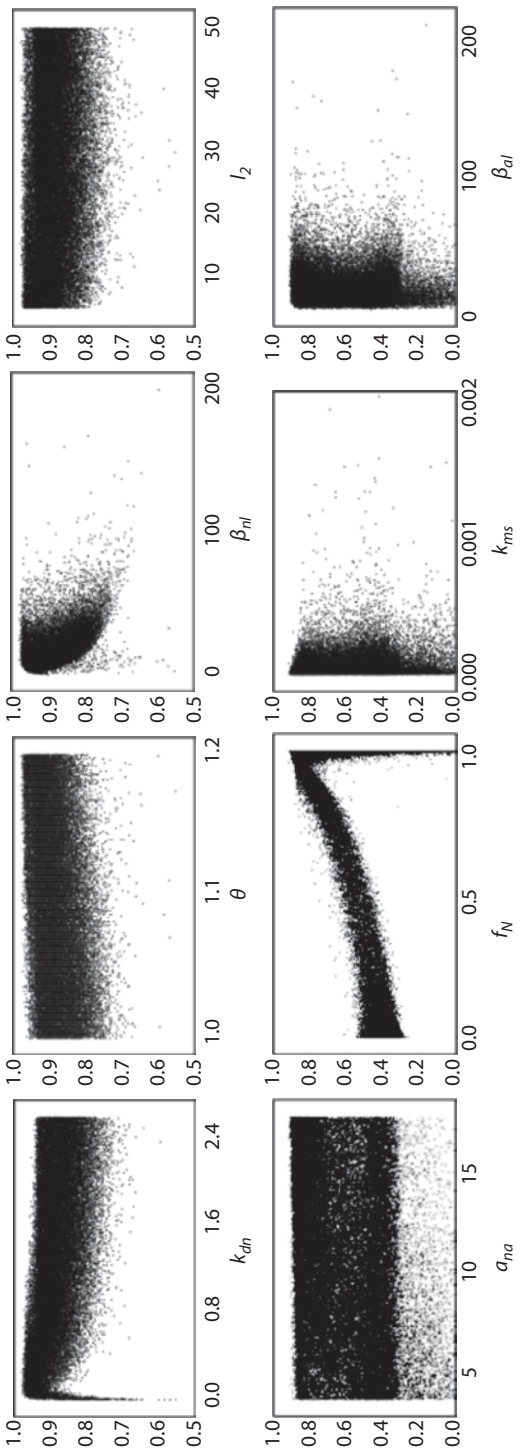

Fig. 8.7 Dotty plots of E_N values for nitrate (N_{nw}) (*top panel*) and total ammonia (N_{aw}) (*lower panel*) loads versus selected sensitive model parameters. (from [26] with permission from ASCE)

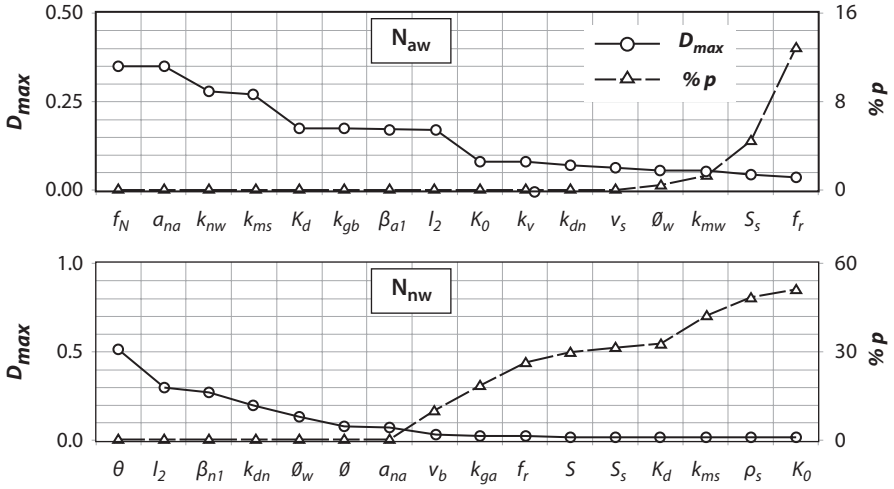

Fig. 8.8 Summary of the Kolmogorov–Smirnov tests and the order of sensitivities for nitrate (N_{aw}) and total ammonia (N_{nw}) loads. (from [26] with permission from ASCE)

and the figures to the right are generated from a log-normal distribution in which most of the data points are clustered and there is a long tail. Thus, although useful, it is difficult to see the full spectrum in dotty plots when the underlying probability distribution is nonuniform.

Results of the K–S test and the ensuing order of parameter sensitivities for nitrate and total ammonia are shown in Fig. 8.8. Compared to nitrate, total ammonia has more sensitive parameters at 5 % confidence level (15–7). This is because there are more processes affecting the fate and transport of ammonia compared to nitrate such as mineralization, adsorption, and volatilization. For ammonia, f_N is the most sensitive parameter along with a_{na}. Note that f_N is a function of pH and temperature. Next important processes are nitrification of ammonia to nitrate in the water column and ammonification. Diffusion of ammonia from the sediment layer to the water column appears to be a significant source of ammonia (see parameters β_{a1} and l_2). For nitrate, the Arrhenius coefficient for temperature adjustment (θ) is the most sensitive parameter, which is rather surprising. As mentioned earlier, many reaction rates and physiological parameters vary with temperature and are sensitive to θ, reflecting this fact. In terms of processes, diffusion of nitrate from the water column to bed sediments (β_{n1} and l_2) and denitrification in the anaerobic sediment layer are the two most important processes. The water and soil porosity parameters, ϕ_w and ϕ also appear sensitive. These two parameters dictate flow-accessible volumes in water and soil, respectively, and thus affect dissolved constituent concentrations.

Figure 8.9 compares the model-simulated total ammonia and nitrate concentrations and loads to observed data. Figures show both behavior (gray band) and nonbehavior sets. Model performance is very good with both constituents. The model is especially doing a good job in predicting loads. The average E_N value of the behavior sets is 0.51 for N_{aw} and 0.85 for N_{nw}. Mass balance errors are less than 6 % for all

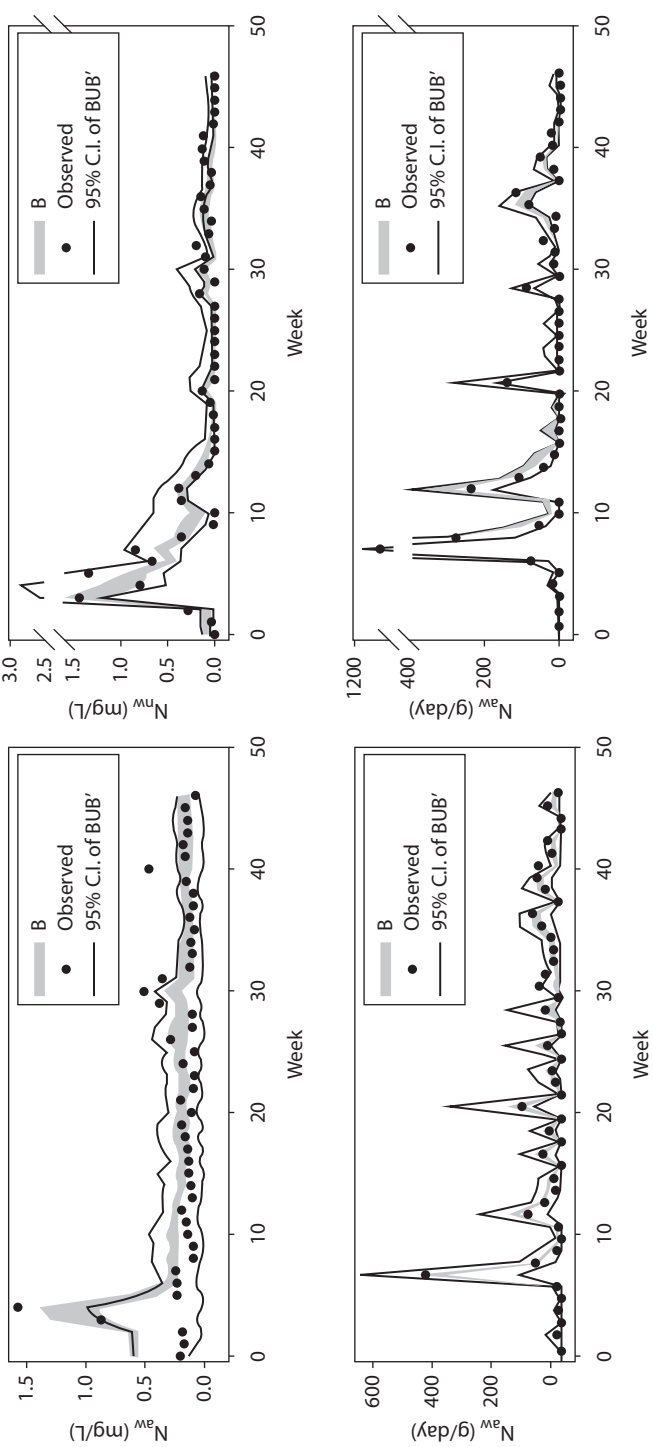

Fig. 8.9 Model generated 95% confidence intervals (C.I) from 100,000 MC simulations versus observed concentrations (*top row*) and loads (*bottom row*) for ammonia and nitrate. *B* behavior set, *B'* nonbehavior set. (from [26] with permission from ASCE)

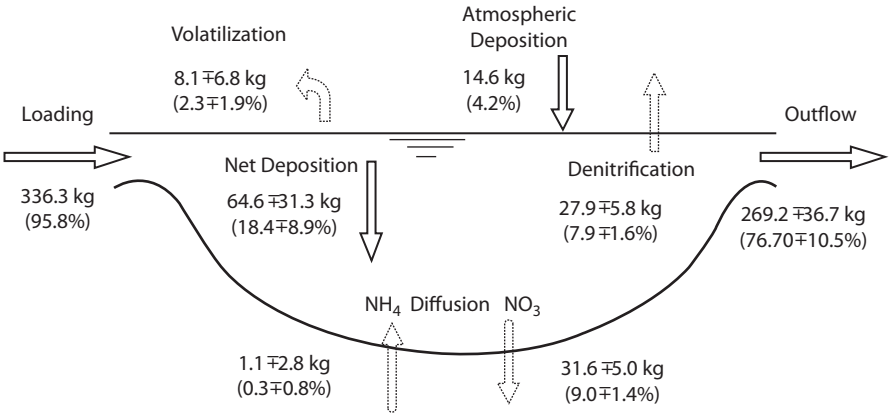

Fig. 8.10 Nitrogen budget in the study wetland with major sources and losses. Numbers in *parentheses* are values normalized with runoff + atmospheric deposition loading. (from [26] with permission from ASCE)

the behavioral simulations. The gray bands in the figure that denote the prediction intervals from the behavior sets are relatively narrow. This means that uncertainty is small, which is a desirable thing.

Figure 8.10 shows computed nitrogen budget of the wetland over the 2-year study period. About 77 % of the incoming total nitrogen load (runoff from the watershed + atmospheric deposition) leaves the system as hydrologic export, which means the wetland system removed about 23 % of the incoming nitrogen. Diffusion to the sediment layer does not seem to be a major loss pathway for ammonia. This does not mean that diffusion is not an important process. What the figure reveals is that the amount of ammonia diffused to the sediment layer and the one diffused from the sediment layer balance out. On the contrary, diffusion of nitrate from the water column to the sediment layer is a major nitrogen retention mechanism, which accounted for roughly 9 % of the incoming nitrogen load.

8.6 Summary

Although processes governing the environmental fate and transport of particular contaminants vary greatly, all contaminants, however, are subject to diffusion caused by the combined effect of random motion of molecules (molecular diffusion) and mechanical mixing associated with deterministic and/or random velocity variations in the medium of concern (e.g., free water, porous media, and air). Of particular interest is diffusive mass exchange of pollutants at the sediment-water interface, a problem that is relevant to lake eutrophication and nitrogen cycling in wetlands, and to management of contaminated sediments accumulating at the bottom of streams, lakes, and estuaries. Mathematical models—conceptual or

physical—of environmental processes that govern the fate and transport of pollutants are valuable tools for testing scientific hypotheses and informing risk-based environmental decision making. While mathematical models (analytical, semi-analytical, and numerical) are useful for understanding processes and quantifying the relative importance of various physical and biogeochemical processes, their application to real-world environmental problems, however, is often very challenging. This is primarily attributed to the complexity of the environment (e.g., heterogeneity, irregular boundaries, multimedia, etc.), lack of observational data, and imperfect understanding of the underlying processes. Aside from the challenge of obtaining exact and numerical solutions for the governing boundary or initial-value problems, parametric sensitivity (i.e., identifying most important processes) and identification (i.e., calibration), quantifying model predictive uncertainty remain even more challenging and more relevant to risk-based environmental decision making.

This chapter presents analytical and numerical models describing the fate and transport of nitrogen in two different, freshwater ecosystems, eutrophic water bodies, and wetlands. Unlike previous models [1], the analytical model that was applied to interpret the Chesapeake Bay sediment flux data considered unsteady-state transport and vertical resolution in bottom sediments. The wetland model on the other hand is among the first models that considered oxygen dynamics and accounted for the impact of a thin oxidizing layer on nitrogen transformation in wetland soils. The emphases were on diffusive transfer and geochemical transformation at the sediment-water interface. In both applications, the GSA method is implemented and the equifinality concept of Beven and coworkers (i.e., different parameter sets are potentially equally likely good simulators of the system) is embraced. The MC and GSA methods reveal the most sensitive parameters and dominant processes in both the Chesapeake Bay and treatment wetland applications. The equifinality, which is reflected by the behavioral parameter set, provides an estimate for model predictive uncertainty caused by measurement and modeling errors. In both cases, diffusive mass transfer plays a significant role in determining nitrogen concentrations. In the treatment wetland application, diffusion of nitrate from the water column to the sediment layer is a significant nitrogen retention mechanism, whereas denitrification is a relatively major loss pathway.

References

1. Di Toro DM (2001) Sediment flux modeling. Wiley, New York
2. EPA (2007) Wetlands and water quality trading: review of current science and economic practices with selected case studies. EPA/600/R-06/155, Ground Water and Ecosystems Restoration Division, Ada, Oklahoma
3. Mitsch WJ, Gosselink JG (2000) Wetland. Wiley, New York
4. Choy B, Reible DD (2000) Diffusion models of environmental transport. Lewis Publisher, Boca Raton
5. Hantush MM (2007) Modeling nitrogen-carbon cycling and oxygen consumption in bottom sediments. Adv Water Resour 30:59–79

6. Kalin L, Hantush MM (2007) Predictive uncertainty and parameter sensitivity of a sediment-flux model: nitrogen flux and sediment oxygen demand. World environment and water resources congress: restoring our natural habitat, Tampa
7. Chapra SC (1997) Surface water-quality modeling. McGraw-Hill, New York
8. Emerson K, Russo RC, Lund RE, Thurston RV (1975) Aqueous ammonia equilibrium calculations: effect of pH and temperature. J Fish Res Board Can 32:2379
9. Spear R, Hornberger G (1980) Eutrophication in peel inlet 2. Identification of critical uncertainties via generalized sensitivity analysis. Water Res 14:43–49
10. Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—a discussion of principles. J Hydrol 10:282–290
11. Jorgensen BB (1977) Bacterial sulfate reduction within reduced microniches of oxidized marine sediments. Mar Biol 41:7–17
12. Hatcher KJ (1986) Introduction to part 1: sediment oxygen demand processes. In: Hatcher K (ed) Sediment oxygen demand: processes, modeling & measurement. University of Georgia, Athens
13. Hantush MM, Kalin L, Isik S, Yucekaya A (2013) Nutrient dynamics in flooded wetlands. I: model development. J Hydrol Eng 18:1709–1723
14. Faulkner SP, Richardson CJ (1989) Physical and chemical characteristics of fresh-water wetland soils. In: Hammer DA (ed) Constructed wetlands for wastewater treatment, municipal, industrial and agricultural. Lewis Publishers, Inc., Boca Raton, pp 41–72
15. Reddy KA, Delaune RD (2008) Biogeochemistry of wetlands: science and applications. CRC Press, Boca Raton
16. Hammer DA (1989) Wetland ecosystems: natural water purifiers? In: Hammer DA (ed) Constructed wetlands for wastewater treatment, municipal, industrial and agricultural. Lewis Publishers, Inc., Boca Raton, pp 5–19
17. Kirk GJD, Kronzucker HJ (2005) The potential for nitrification and nitrate uptake in the rhizoshere of wetland plants: a modeling study. Ann Bot 96:639–646
18. Brown LC, Barnwell TO Jr (1987) The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: documentation and users manual. U.S. Environmental Protection Agency, Athens, GA, Report EPA/600/3-87/007
19. Jørgensen SE, Bendoricchio G (2001) Fundamentals of ecological modeling, 3rd edn. Elsevier, Amsterdam
20. Broecker HC, Petermann J, Siems W (1978) The influence of wind on CO_2 exchange in a wind-wave tunnel. J Marine Res 36:595–610
21. Wang Y, Zhang J, Kong H, Inamori Y, Xu K, Inamori R, Kondo T (2009) A simulation model of nitrogen transformation in reed constructed wetlands. Desalination 235:93–101
22. Joergensen BB, DesMarais DJ (1990) The diffusive boundary layer of sediments: oxygen microgradients over a microbial mat. Limnol Oceanogr 35:1343–1355
23. Nakamura Y, Stefan HG (1994) Effect of flow velocity on sediment oxygen demand: theory. J Environ Eng 120:996–1016
24. Di Toro DM, Paquin PR, Subburamu K, Gruber DA (1990) Sediment oxygen demand model: methane and ammonia oxidation. J Environ Eng 116:945–986
25. Jordan TE, Whigham DF, Hofmockel KH, Pittek MA (2003) Nutrient and sediment removal by a restored wetland receiving agricultural runoff. J Environ Qual 32:1534–1547
26. Kalin L, Hantush MM, Isik S, Yucekaya A, Jordan T (2013) Nutrient dynamics in flooded wetlands. II: model evaluation. J Hydrol Eng 18:1724–1738
27. Saltelli A, Sobol IM (1995) About the use of rank transformation in sensitivity analysis of model output. Reliab Eng Syst Saf 50:225–239
28. Beven K, Binley A (1992) The future of distributed models—model calibration and uncertainty prediction. Hydrol Process 6:279–298

Chapter 9
Modeling Groundwater Flow in Unconfined Aquifers

Shivendra Nath Rai

9.1 Introduction

Water is distributed within the earth, on its surface, and also in the atmosphere in liquid, solid, and gaseous forms, respectively. Of the total available water on the earth, 97% is saline stored in the oceans and the remaining 3% is only in the form of fresh water. Out of this 3% of fresh water, groundwater constitutes only 30.1%. The remaining is in the form of icecap and glaciers (68.7%), liquid surface water (0.3%), and in atmosphere and living being (0.9%). Of the liquid surface fresh water, 87% is stored in lakes, 11% in swamps, and only 2% flows into the rivers (source: Wikipedia-a free encyclopedia). Accordingly, groundwater constitutes the second largest reserve of fresh water available on the earth. The main source of groundwater is precipitation. Precipitated water falls on the ground surface and enters below it. This entering process is known as infiltration. Infiltrated water moves downward and gets stored in pores of subsurface geological formations or in geological structures such as fractures, faults, joints, etc. in the case of hard rocks. This leads to the evolution of groundwater regime below the earth's surface. Geological formations capable of storing groundwater and allowing its movement from one place to another place under ordinary field conditions are known as aquifers. Sands, sandstone, weathered mantle, highly fractured rock, etc. are examples of aquifers. On the other hand, there are some geological formations such as massive basalt and granite units which can neither store the groundwater nor allow the movement of groundwater. Such geological formations are known as aquifuge. There is another category of geological formation such as clay which can store good amount of groundwater but does not allow the movement of groundwater because of lack of interconnectivity of its pores. Such formations are known as aquiclude. Vertical distribution of groundwater is characterized into two zones: unsaturated and saturated. In the unsaturated zone the entire pores contain both water and air, whereas

S. N. Rai (✉)
CSIR-National Geophysical Research Institute, Uppal Road,
500007, Hyderabad, India
e-mail: snrai_ngri@yahoo.co.in

S. K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,*
Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_9,

in the saturated zone the entire pore space is filled with water. If the saturated zone is bounded by two impermeable formations from top and bottom, it is called a confined aquifer.

In confined aquifer water is stored at more than atmospheric pressure. That is why water level in the well penetrating a confined aquifer is at a higher elevation than the elevation of the upper boundary of the confined aquifer. Elevation of the water level in the well penetrating confined aquifer is called piezometric head and is measured from a reference stratum. If the upper boundary of the saturated zone is the water table (or phreatic surface), it is called unconfined aquifer. On the water table, the pressure is equal to the atmospheric pressure. An unconfined aquifer (or part of it) that rests on a semi-pervious layer is a leaky unconfined aquifer. Similarly, a confined aquifer (or part of it) that has at least one semi-pervious layer containing stratum is called a leaky confined aquifer. A schematic diagram of the aquifer's type and zones of vertical distribution of groundwater are shown in Fig. 9.1.

The advantage of unconfined aquifers over confined aquifers to serve as a subsurface reservoir is that the storage of groundwater in large quantity is possible only in unconfined aquifer. This is because the storativity of the unconfined aquifer is linked to the porosity and not to the elastic properties of water and the solid matrix, as in the case of the confined aquifer [1]. Also, the vast surface area of the unconfined aquifer above the water table is available to receive the surface applied recharge, whereas in the case of confined aquifer, only a small open area exposed to the ground surface or leaky portion of the aquifer boundary is available to receive the recharge as shown in Fig. 9.1. Sources of surface water are not available everywhere. Therefore, their use to meet the demand of water supply for irrigation, industrial, and domestic purposes is restricted to those areas where water can be transported through canals from these sources. Also surface water bodies such as rivers, lakes are more vulnerable to contamination. On the other hand, groundwater resources are distributed globally and are less vulnerable to contamination compared to surface water bodies. Therefore, groundwater plays a major role in augmenting water supply to meet the ever-increasing demand, especially in developing counties such as India, where agriculture sector provides job opportunity to a large rural population and is the main source of income to them. Increasing dependence of water supply on groundwater resources is resulting in increasing use of aquifers as a source of fresh water supply and subsurface reservoir for storing excess surface water.

Natural replenishment of aquifers occurs very slowly. Therefore, withdrawal of groundwater at a rate greater than the natural replenishment rate causes declining of groundwater level, which may lead to decrease in water supply, contamination of fresh water by polluted water from nearby sources, seawater intrusion into the aquifer of coastal areas, etc. To increase the natural replenishment, artificial recharging of the aquifer is becoming increasingly important in groundwater management. In many cases, excess recharging also leads to the growth of water table near the ground surface and causes several types of environmental problems, such as water logging, soil salinity, etc. In such a situation, proper management of ground-

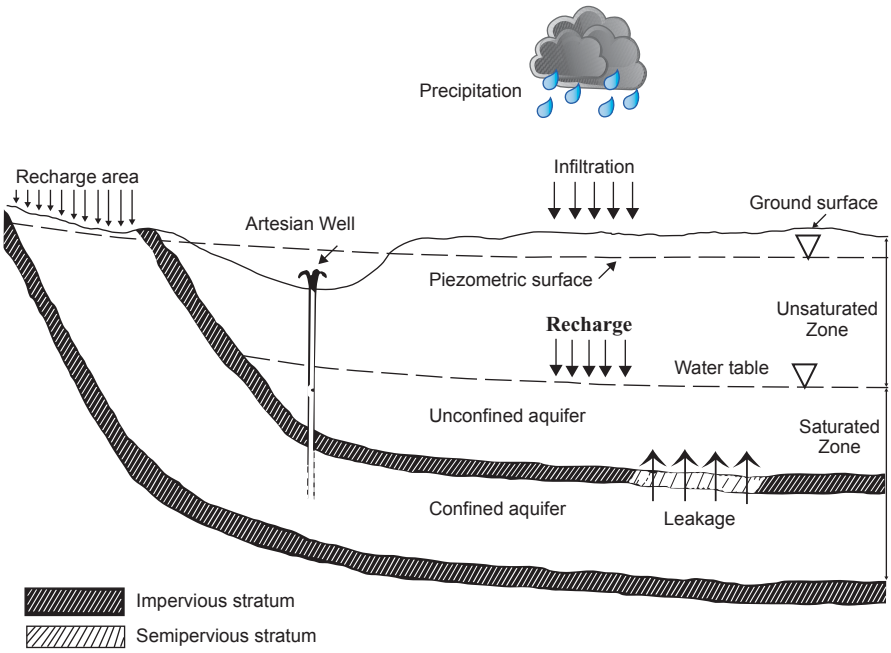

Fig. 9.1 Aquifer types and vertical distribution of groundwater zones

water resources is needed to overcome the shortage of water supply on one hand and to prevent the environmental problems on the other hand. In order to address the management problem, one must be able to predict the response of the aquifer system to any proposed operational policy of groundwater resources development such as artificial recharging and pumping. Such problems are referred to as forecasting problems. Its solution will provide the new state of the groundwater system. Once the new state is known, one can check whether the related recharging and/or pumping scheme is feasible to meet the preset objectives of the sustainable development and management of groundwater resources. Such problems can be tackled by applying mathematical modeling techniques. Mathematical models help in making judicious selection of an appropriate development scheme such as designing of recharging and pumping schemes out of many proposed development schemes by comparing responses of different proposed recharge/pumping schemes in order to select the best scheme without resorting to the expensive field works. This chapter deals with mathematical modeling of groundwater flow in unconfined aquifer and related problems. Mathematical modeling needs simplification of complex geohydrological environ and processes based on assumptions to make it amenable to the mathematical treatment without compromising the physical characteristics of the problem. One such simplification is the hydraulic approach.

9.2 Hydraulic Approach

In general, flow through a porous medium is 3-D . However, as the geometry of most aquifers is such that they are thin relative to their horizontal dimension on regional scale, a simpler approach called hydraulic approach is introduced for modeling purpose. According to this approach, it is assumed that the flow in the aquifer is essentially horizontal everywhere neglecting its vertical component [1–2]. The approximation of horizontal flow in unconfined aquifer is the basis of Dupuit assumption which will be discussed later. However, this assumption fails in regions where the flow has a large vertical component, for example, in the vicinity of partially penetrating wells, or at boundaries with open water bodies such as lakes, rivers, etc.

9.3 Mathematical Modeling

Modeling of groundwater flow begins with a conceptual understanding of the physical problem. The next step is translating the physical problem into a mathematical framework in the form of a set of mathematical equations governing groundwater flow, boundary, and initial conditions (in the case of unsteady state flow). Its solutions are used to describe the dynamic behavior of water table in the flow system under consideration in response to the hydraulic stresses such as recharging, pumping, leakages, stream aquifer interaction etc. Mathematical model may be deterministic, statistical, or some combination of the two. Deterministic models retain a good measure of physical insight while permitting a number of problems of the same class to be tackled with the same model. Our discussion is confined to the development of governing groundwater flow equations, methods of solutions, and related deterministic models used for predicting water table fluctuation induced by recharging and/or pumping, which are the essential components of groundwater resources development. Formulations of groundwater flow equations are based on the conservation principles dealing with mass and momentum. These principles require that the net quantity of mass (or momentum) entering or leaving a specified volume of aquifer during a given time interval be equal to the change in the amount of mass (or moment) stored in the volume. Groundwater flow equations for specific aquifer systems are formulated by combining the equation of motion in the form of Darcy's law, which follows principle of conservation of momentum with the mass balance equation, also known as mass conservation equations or continuity equations, which follows the principle of conservation of mass.

9.3.1 Darcy's Law

Consider flow of water under confined condition through a sand filled cylinder of cross sectional area A and length L as shown in Fig. 9.2. Cylinder is representing

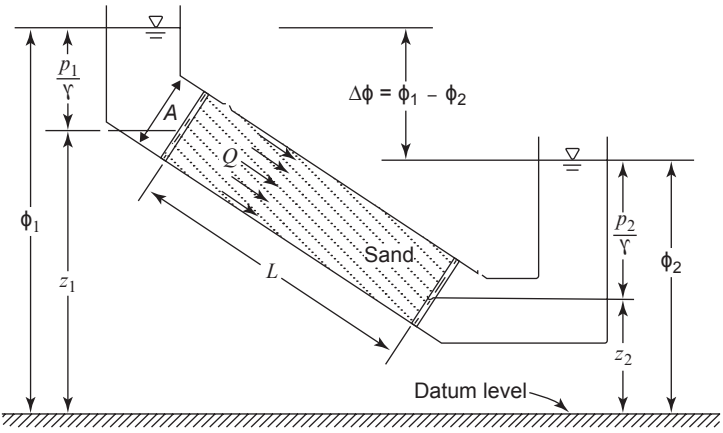

Fig. 9.2 Flow of water through an inclined porous media

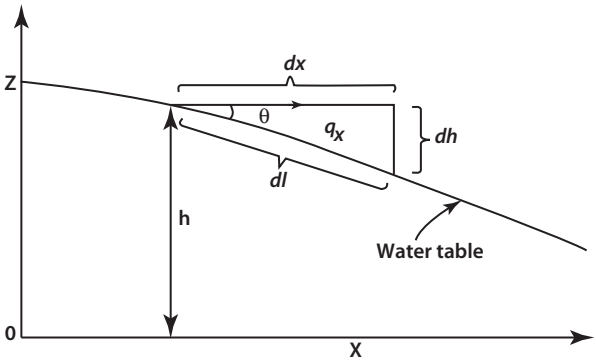

Fig. 9.3 Diagrammatic representation of the Dupuit assumption

a propos media through which water is flowing from one end to another end under gravitational flow because of elevation difference between piezometric heads at both ends. According to Darcy's law, the rate of flow (volume of water per unit time), Q, is proportional to the crosssectional area A of the porous media, proportional to the piezometric heads difference between two points ($\phi_1 - \phi_2$), and inversely proportional to the length of the porous media, L as shown in Fig. 9.3. Mathematically, it can be expressed as:

$$Q = KA\frac{\phi_1 - \phi_2}{L}, \tag{9.1}$$

in which K is the coefficient of proportionality, called hydraulic conductivity. K depends on the properties of fluid as well as solid matrix and is expressed by $K = k\rho g/\mu$ in which k is the solid medium permeability, ρ is the density of fluid (taken 1 in case of water), g is the gravitational acceleration, and μ is the dynamic viscosity. The

piezometric head is expressed as $\phi = z + p/\gamma$, in which z is the elevation head, p is atmospheric pressure, and γ is specific weight of water [1]. Equation 9.1 can also be rewritten as:

$$q = \frac{Q}{A} = K\frac{\phi_1 - \phi_2}{L}, \tag{9.2}$$

where q is the specific discharge defined as the volume of water flowing through unit crosssectional area of the cylinder. Writing $(\phi_1 - \phi_2)/L$ in differential form by defining $(\phi_1 - \phi_2) \rightarrow \mathrm{d}\phi$ and $L \rightarrow \mathrm{d}L$ and introducing a minus sign to indicate that flow is in the direction of decreasing ϕ, Eqs. 9.1 and 9.2 can be expressed as:

$$Q = -KA\frac{d\phi}{dL} \tag{9.3}$$

and

$$q = -K\frac{d\phi}{dL}. \tag{9.4}$$

In case of unconfined aquifer, Eqs. 9.3 and 9.4 can be written as:

$$Q = -KA\frac{dh}{dL} \tag{9.5}$$

$$q = -K\frac{dh}{dL}, \tag{9.6}$$

where *dh* is difference of water table heights between two points separated by distance *dl*. Eqs. 9.5 and 9.6 are known as the equation of motion. It is evident from Eq. 9.6 that for $dh/dl = 1$, $K = q$. Groundwater flow equation for an unconfined aquifer is derived by combining the equation of motion modified by the Dupuit assumption with the mass balance equation.

9.3.2 Dupuit Assumption

Consider a vertical cross section of unconfined groundwater flow as shown in Fig. 9.4. Dupuit assumption is based on the field observation that the slope of the water table, θ, is generally very small on regional scale. It implies that the flow is almost horizontal and $dL \approx dx$ (Fig. 9.3). Replacing *dL* by *dx*, Eq. 9.6 becomes:

$$q = -K_x\frac{dh}{dx}. \tag{9.7}$$

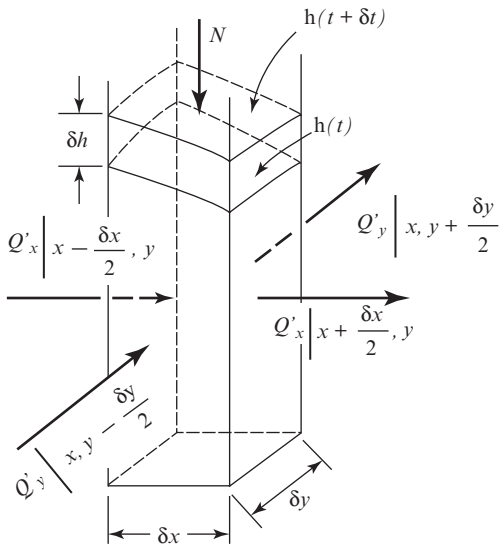

Fig. 9.4 Groundwater flow in a control box of an unconfined aquifer

If groundwater flow takes place through a saturated vertical column of thickness h, then Eq. 9.7 can be rewritten as:

$$Q_x^{'} = -K_x h \frac{dh}{dx}. \tag{9.8}$$

9.3.3 Mass Balance Equation

To derive the mass balance equation, consider groundwater flow through a control box in an unconfined aquifer (Fig. 9.4). The box is bounded by vertical surfaces at $(x - \delta x/2, y)$ and $(x + \delta x/2, y)$. parallel to the y-axis and at $(x, y - \delta y/2)$ and $(x,\ y + \delta y/2)$ parallel to the x-axis. The box has a horizontal impervious base and the water table forms its upper boundary. The control box receives vertical recharges with $N(x, y, t)$ rate. The rate of recharge is the volume of water added to the water table in unit time through unit cross-sectional area and has the unit of velocity. In principle $N(x, y, t)$ is the sum of all recharge rates from distributed sources (recharge basins, ponds, streams, etc.) and withdrawal rates from distributed sinks (wells, leakage sides, etc.).Here, $N(x, y, t)$ is considered as recharge rate only from a single source to simplify the derivation of the mass balance equation.

Because of the excess mass inflow during time, δt, the water table rises from the initial height $h(t)$ to a new height $h(t+\delta t)$. The mass balance equation based on the Dupuit assumption can now be written as:

$$\delta t\{\delta y[Q_x^{'}(x-\frac{\delta x}{2},y)-Q_x^{'}(x+\frac{\delta x}{2},y)] + \delta x[Q_y^{'}(x,y-\frac{\delta y}{2})-Q_y^{'}(x,y+\frac{\delta y}{2})] + \delta x\delta y N(x,y,t)\} = S_y\ \delta x\delta y\left[h(t+\delta t)-h(t)\right]. \tag{9.9}$$

where $Q_x^{'}$ and $Q_y^{'}$ are discharges per unit width in the *x* and *y* directions, respectively; $h(t)$ and $h(t+\delta t)$ are the water table heights at times t and $(t+\delta t)$, respectively, ρ is the density of water which is normally taken as one and *Sy* is the specific yield which is defined as the volume of water added to (or released from) the aquifer per unit horizontal area of aquifer and per unit rise (or decline) of water table. *Sy* is dimensionless aquifer parameter. By expanding $Q_x^{'}, Q_y^{'}$ and $h(t+\delta t)$ about *x, y*, and *t*, respectively, by Taylor series and dropping all terms containing second and higher order derivatives gives:

$$Q\left(x+\frac{\delta x}{2}\right)=Q_x+\frac{\partial Q_x}{\partial x}\frac{\delta x}{2} \tag{9.10}$$

$$Q\left(x-\frac{\delta x}{2}\right)=Q_x-\frac{\partial Q_x}{\partial x}\frac{\delta x}{2} \tag{9.11}$$

$$Q_y^{'}\left(x,y+\frac{\delta y}{2}\right)=Q_y+\frac{\partial Q_y}{\partial y}\frac{\delta y}{2} \tag{9.12}$$

$$Q_y^{'}\left(x,y-\frac{\delta y}{2}\right)=Q_y-\frac{\partial Q_y}{\partial y}\frac{\delta y}{2} \tag{9.13}$$

$$h(t+\delta t)=h(t)+\frac{\partial h}{\partial t}\frac{\delta h}{2}. \tag{9.14}$$

Substituting these values in Eq. 9.9, and thereafter dividing both sides of Eq. 9.9 by $\delta x, \delta y, \delta t$ and letting $\delta x, \delta y$ and $\delta t \rightarrow 0$, we obtain the following mass balance equation for an inhomogeneous and anisotropic unconfined aquifer:

$$-\frac{\partial}{\partial x}(Q_x^{'})-\frac{\partial}{\partial y}(Q_y^{'})+N(x,y,t)=S_y\frac{\partial h}{\partial t}. \tag{9.15}$$

Inserting the expressions for *Qx'* and *Qy'* from Eq. 9.8 into Eq. 9.15 yields:

$$\frac{\partial}{\partial x}\left(K_x h\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_y h\frac{\partial h}{\partial y}\right)+N(x,y,t)=S_y\frac{\partial h}{\partial t}. \tag{9.16}$$

For an inhomogeneous isotropic aquifer $K=K(x, y)$, Eq. 9.16 becomes:

$$\frac{\partial}{\partial x}\left(Kh\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(Kh\frac{\partial h}{\partial y}\right)+N(x,y,t)=S_y\frac{\partial h}{\partial t}. \tag{9.17}$$

9.3.4 Groundwater Flow Equation for a Leaky Unconfined Aquifer

In this case, an unconfined aquifer is separated from an underlying confined aquifer by a partly semi-pervious layer as shown in Fig. 9.2. The mass balance equation for a control box in inhomogeneous anisotropic unconfined aquifer, taking into account a leakage of rate q_{L} between the aquifers is given by:

$$\begin{aligned}\delta t\{&\delta y[Q_x'(x-\frac{\delta x}{2},y)-Q_x'(x+\frac{\delta x}{2},y)]\\&+\delta x[Q_y'(x,y-\frac{\delta y}{2})-Q_y'(x,y+\frac{\delta y}{2})]\\&+\partial x\partial y N(x,y,t)+\partial x\partial y q_L\}=S_y\,\delta x\delta y\big[h(t+\delta t)-h(t)\big],\end{aligned} \tag{9.18}$$

where h is water table height in an unconfined aquifer, ϕ is piezometric head in underlying confined aquifer. q_{L} for $h<\phi$ is expressed as:

$$q_L=\frac{\phi-h}{\sigma}, \tag{9.19}$$

in which $\sigma=B'/K', B'$ being the thickness and K' the hydraulic conductivity of the semi-pervious layer. For $h>\phi$, q_L becomes negative because groundwater outflows from the unconfined aquifer. For $h<\phi$, groundwater flows into the unconfined aquifer and hence q_{L} becomes positive. After simplification, Eq. 9.18 becomes:

$$\frac{\partial}{\partial x}\left(K_x h\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_y h\frac{\partial h}{\partial y}\right)+N(x,y,t)+\frac{\phi-h}{\sigma}=S_y\frac{\partial h}{\partial t}. \tag{9.20}$$

Equation 9.20 is the desired governing equation for groundwater flow in a leaking unconfined aquifer. The equation for an inhomogeneous isotropic aquifer ($K=K(x, y)$) and a homogeneous isotropic aquifer ($K=$constant) can be obtained from Eq. 9.20 as in the previous cases.

9.3.5 Linearization of Groundwater Flow Equation

Generally, Eqs. 9.16, 9.17, and 9.20 are used for development of groundwater flow models. These are nonlinear second order partial differential equations and their exact solutions are difficult to obtain. The nonlinearity is because of the presence of h

as coefficient in the partial derivatives on the left hand side. Therefore, linearization of these equations is essential to obtain analytical solutions. We consider Eq. 9.17 to describe the linearization procedures an example. For homogenous isotropic aquifers (K=constant), Eq. 9.17 can be written in the following two forms:

$$K\left[\frac{\partial}{\partial x}\left(h\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(h\frac{\partial h}{\partial y}\right)\right]+N(x,y,t)=S_y\frac{\partial h}{\partial t} \tag{9.21}$$

$$\frac{\partial^2 h^2}{\partial x^2}+\frac{\partial^2 h^2}{\partial y^2}+\frac{2N(x,y,t)}{K}=\frac{S_y}{Kh}\frac{\partial h^2}{\partial t}. \tag{9.22}$$

Two procedures of linearization are commonly used. According to the first procedure, i.e., the Baumann procedure of linearization, if the variation in h is much less than the initial height of the water table h_0, then the coefficient h appearing on the left hand side of Eq. 9.21 can be replaced by h_0. Then Eq. 9.18 can be rewritten as:

$$T\left(\frac{\partial^2 h}{\partial x^2}+\frac{\partial^2 h}{\partial y^2}\right)+N(x,y,t)=S_y\frac{\partial h}{\partial t}, \tag{9.23}$$

where $T=Kh_0$. Now Eq. 9.23 is linear in h. In the second procedure, i.e., the Hantush's procedure of linearization, h appearing in the denominator on the right hand side of Eq. 9.22 is replaced by the weighted mean of the depth of saturation $\bar{h}$; a constant of linearization which is approximated by $0.5[h_0+h(t_e)]$; t_e is the period at the end of which $\bar{h}$ is to be approximated. Then Eq. 9.23 becomes:

$$\frac{\partial^2 h^2}{\partial x^2}+\frac{\partial^2 h^2}{\partial y^2}+\frac{2N(x,y,t)}{K}=\frac{S_y}{K\bar{h}}\frac{\partial h^2}{\partial t}. \tag{9.24}$$

Now Eq. 9.24 becomes linear in h_2. Substitution of a new variable H, defined as $H=h^2-h_0^2$, Eq. 9.24 can be rewritten as:

$$\frac{\partial^2 H^2}{\partial x^2}+\frac{\partial^2 H^2}{\partial y^2}+\frac{2N(x,y,t)}{K}=\frac{S_y}{K\bar{h}}\frac{\partial H}{\partial t}. \tag{9.25}$$

Equation 9.25 is being extensively used for development of groundwater flow models. Here, it should be mentioned that in case of using Eq. 9.25 for development of mathematical models, the initial and boundary conditions should be also described in the form of H to preserve linearity of the problem.

Now, to make use of Eq. 9.24 (or 9.25) for the development of groundwater flow models, one needs to compute the value of $\bar{h}$. Marino [3] suggested the method of successive approximation for computation of $\bar{h}$ value. In this method, the weighted mean of the depth of saturation is taken as a first approximation equal to the initial depth of saturation, h_0. The first approximated height of the water table is then calculated by using a solution of Eq. 9.25. In the second trial, the weighted mean of the

depth of saturation is approximated by the average of the initial depth of saturation h_0 and the first approximation of the height of the water table. This procedure is repeated until the value of the calculated height of the water table converges. The last estimated value of $\bar{h}$ for which the calculated water table height converges at a given time and position is the desired value of $\bar{h}$ for that particular time and position. Thus, for each time and position one has to compute $\bar{h}$ and the conversed value of water table height for this value of $\bar{h}$ is the desired water table height at the given position and time. By comparing results of analytical solutions based on the Hantush linearization procedure with the experimental results obtained from Hele-shaw model, Marino[3] found that for $N \leq 0.2\ K$ and $h - h_0 \leq 0.5 h_0$, the maximum deviation between both the results was 6%. Even for $h - h_0 \geq 20\ h_0$, the maximum deviation was 12.2%. It shows that the results of analytical model agree reasonably well with the experimental results. Rao and Sarma [4] have reported that both the linearization methods yield results which have satisfactory agreement with those of the experiments (within ±5%) for the rise of the water table up to 40% of its initial height. Beyond this limit the Hantush linearization scheme gave a more satisfactory agreement. Thus, Hantush method was found to have wider applicability. However, Hantush method involves computation of weighted mean of the depth of saturation through successive iteration and hence requires more computation time than the Baumann procedure of linearization. However, time is not at all an issue in this era of fast computers. Equations 9.23–9.25 are popularly known as linearized Boussinesq equations.

9.3.6 Groundwater Flow Equations for Sloping Aquifer

The groundwater flow equation in a sloping 2-D unconfined aquifer is described by: [5–6]

$$\frac{\partial^2 s}{\partial x^2} + \frac{\partial^2 s}{\partial y^2} - 2a\frac{\partial s}{\partial t} + \frac{2N(t)}{K} = \frac{1}{\Re}\frac{\partial s}{\partial t}, \tag{9.26}$$

where $s = h_2$, h = variable water table height, $a = q/2D$, q = slope of the base, D = the mean depth of saturation, and $\Re = KD/Sy$.

9.3.7 Groundwater Flow Equations in Cylindrical Coordinates

This type of equation is used to describe groundwater flow induced by recharging/pumping through circular shape recharge basin/well and is given by: [7–8]

$$S\frac{\partial h}{\partial t} = -\frac{1}{r}\frac{\partial}{\partial r}(rQ) + N(r,t), \tag{9.27}$$

where r is the radial distance measured from the center of recharge basin, and q is defined by Darcy law as:

$$Q = -Kh\left(\frac{\partial h}{\partial r}\right). \tag{9.28}$$

Equations 9.23–9.25 describe 2-D groundwater flow. Equation for 1-D flow, for example in the x direction, can be obtained by simply substituting zero for the derivative of y. Groundwater flow equations for a steady-state condition can be obtained by substituting zero for time derivatives.

Groundwater flow equations presented here are in the form of partial differential equations having infinite numbers of solutions. To obtain a unique solution for a particular problem, some more information about the problem under consideration is needed, such as the values of aquifer parameters, geometry of the flow domain, leakage rate, recharge rate, pumping rate, initial conditions, boundary conditions, etc. depending on the physical condition of the problem under consideration. Aquifer parameters can be deduced from field as well as experimental methods [1, 9]. A brief description about the initial and boundary conditions commonly used in groundwater flow problems are discussed below.

9.3.8 Initial Conditions

Initial conditions describe the distribution of h at all points of the flow domain at the beginning of the investigation, i.e., at $t=0$. This is expressed as:

$$h = h_0(x, y, 0), \tag{9.29}$$

where h_0 is a known value of h for all points of the flow domain at $t=0$. Now, to make use of initial condition for the solution of Eq. 9.25, Eq. 9.29 can be written as:

$$H(x, y, 0) = 0, \tag{9.30}$$

in which $H = h^2 - h_0^2$.

9.3.9 Boundary Conditions

These conditions describe the nature of interaction of the aquifer along its boundaries with its surrounding environs such as reservoir, rivers, groundwater divide, etc. Three types of boundary conditions are generally encountered in groundwater flow problems.

Dirichlet boundary condition: In this case, h is prescribed for all points of the boundary for the entire period of investigation. This is expressed as:

$$h = h_0\,(x, y, t), \tag{9.31}$$

where $h_0(x, y, t)$ are known values of h at all points on the boundary.

Neumann boundary condition: This type of boundary condition prescribes the flux across the boundary of the flow system and can be expressed as:

$$q = \psi_1\,(x, y, t), \tag{9.32}$$

where $\psi_1\,(x, y, t)$ are the known values of flux at boundaries. A special case of this boundary condition is the no-flow boundary condition in which flux is zero. This condition occurs at impermeable surfaces or at the groundwater divide, a surface across which no flow takes place.

Cauchy boundary condition: This boundary condition is encountered at the semi-pervious boundary layer between the aquifer and an open water body such as river. Because of the resistance to the flow offered by the semi-pervious boundary that lies between the aquifer and the river, the water level in the river differs from that in the aquifer on the other side of the semi-pervious boundary. In this case, the flux is defined by:

$$q = K' \frac{h - h_0}{b}, \tag{9.33}$$

where h is the head at $x=0$, h_0 is the water level in the river, b and K' are the thickness and hydraulic conductivity, respectively, of the semi-pervious boundary layer.

9.3.10 Estimation of Rate of Recharge and Pumping

Recharging and pumping are the essential components of groundwater development schemes. The purpose of groundwater recharging is to store groundwater in order to reduce, stop, or even reverse the declining trend of water table. On the other hand, pumping is used for water supply. Thus, recharging and pumping have significant effects on the dynamics of water table. Therefore, accurate estimation of recharge and pumping rates are very crucial for prediction of water table fluctuation. Many mathematical models have been developed to predict water table fluctuations in response to recharge from basins of different geometrical shapes [3, 10–15]. Most of the models are based on the assumption of constant rate of recharge applied continuously. However, rate of recharge largely depends on the infiltration rate which is influenced by several factors. The infiltration rate decreases initially mainly due to dispersion and swelling of soil particles at the bottom of the basin. After some time, it increases owing to displacement of the entrapped air from pores. After attaining a maximum value, it again decreases owing to clogging of the soil pores [1, 16–17].

Clogging is caused by silt and clay deposition over and immediately below the base of basin. The rate of recharge follows almost a similar pattern of variation of

infiltration rate with comparatively less intensity and with some time lag due to the time taken by the infiltrated water to reach the water table. When the rate of recharge decreases to a minimum prescribed level, the recharge operation is discontinued for some time and after drying, cleaning, and if necessary, scrapping of the silted bottom of the basin, recharge rate is brought back almost to its initial value and the basin is again put back to use for the next phase of recharge operation. Zomorodi [18] has demonstrated with the help of field examples that the solutions based on the assumption of constant rate of recharge are unable to predict the rise and subsequent decline of the water table which is due to decrease in the rate of recharge. He suggested that the recharge rate should be treated as a variable in time in order to simulate actual field conditions. Several schemes have been proposed to approximate time varying recharge rate in order to develop predictive groundwater flow models. Rai and Singh [19] have used two linear elements to approximate exponentially decaying recharge rates applied from a strip basin to a semi-infinite aquifer. Some workers [20–22] used exponential function to approximate one cycle of time varying recharge applied from a single basin. Yue-zan and others [23] have used a scheme in which the duration of time varying recharge is divided into several time zones according to the actual variation in the recharge rate. In each time zone, the variation range of recharge rate should be considered so small that it can be represented by the constant average value of recharge rate of that particular zone. Yue-zan and colleagues referred this recharge rate approximation as stepped variable scheme.

Manglik and others [24] have proposed a new scheme for approximation of time varying recharge rate. In this scheme, time varying recharge rate is approximated by a series of line elements of different lengths and slopes. The number, lengths, and slopes of the line elements depend on the nature of variation of recharge rate. Advantage of this approximation scheme is that any complex nature of recharge rate variation can be approximated with more accuracy. This scheme was extended for the recharge operation from multiple basins [25] and combination of recharging and pumping from a number of basins and wells [26]. In a real field condition, artificial recharging and pumping operations are carried out intermittently from more than one site according to necessity. In groundwater flow equations, for example in Eq. 9.25, $N(x, y, t)$ is the sum of all recharge and withdrawal rates from distributed sources (recharge basins, ponds, streams, etc.) and from distributed sinks (wells, leakage sides, etc.). According to the scheme considered in [26], $N(x, y, t)$ is represented by:

$$N(x,y,t) = \begin{cases} \sum_{i=1}^{n} N_i(t) & \text{for } x_{i1} \le x \le x_{i2}, y_{i1} \le y \le y_{i2} \\ 0 & \text{elsewhere} \end{cases}, \tag{9.34}$$

where n is the total number of basins and/or well, $N_i(t)$ is the time-varying recharge (or pumping) rate for the *ith* basin (or well, respectively) and $x_{i1}, x_{i2}, y_{i1}, y_{i2}$ are

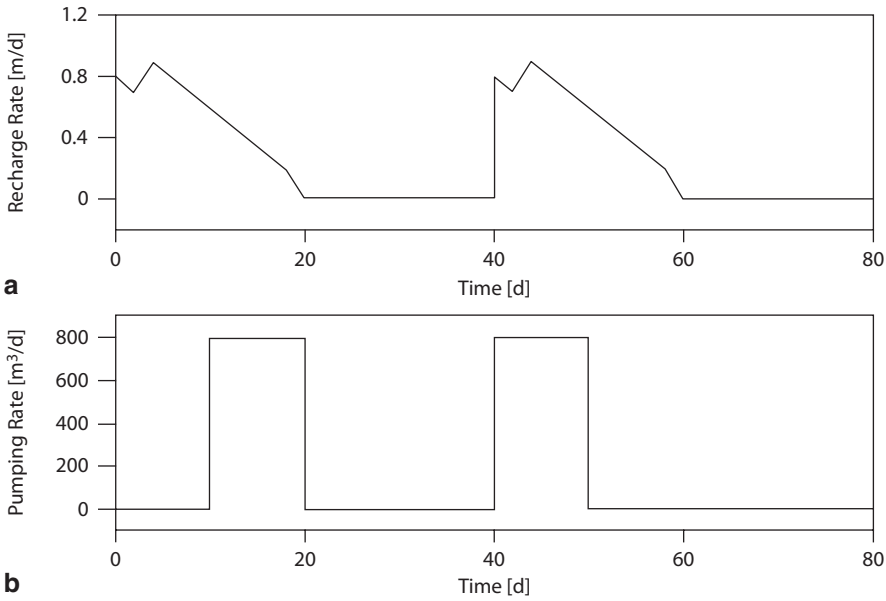

Fig. 9.5 Approximation of two cycles of time varying recharge and pumping. [27]

the coordinates of *ith* basin (or well). $N_i(t)$ is positive for recharge to the aquifer and negative for pumping and leakage out of the aquifer.

For demonstration purpose, Fig. 9.5 illustrates approximation of two cycles of time varying recharge and pumping rates by using this scheme. In this example, two cycles of recharge operations each of 20 days duration with 20 days gap are considered. In each cycle the rate of recharge decreases from 0.8 to 0.7 m/d during the first 2 days and again reaches maximum value of 0.9 m/d during the next 2-day period. Thereafter, it continuously decreases to 0.2 m/d during the next 14 days period. After that recharge operation is discontinued. As a result, the rate of recharge decreases to zero in the next 2 days (Fig. 9.5a). The second cycle of recharge starts after a gap of 20 days. In the second cycle also variation of the rate of recharge is considered in the same way. This kind of time varying recharge rate is approximated by using 11 linear elements of different lengths and slopes. In this example two cycles of pumping each of 10 days duration with a gap of 20 days is considered at a rate of 80 m^3/d. The first cycle of pumping begins after 10th day from the beginning of the first cycle of recharge and continues till the 20th day, i.e., the last day of the first cycle of recharging. After a gap of 20 days, the second cycle of pumping starts on the 40th day and continues till the 50th day (Fig. 9.5b). This time varying pumping rate is approximated by nine linear elements [27–30] used the same scheme to approximate time-dependent recharge, pumping and/or leakage to develop analytical models to predict water table fluctuation. Results of analytical models are verified by comparing with the numerical results obtained by using MODFLOW.

9.4 Analytical Methods of Solution

The purpose of solving a groundwater flow equation is to obtain the values of water table height, $h(x, y, t)$. Generally two types of methods, namely analytical methods and numerical methods are used for this purpose. Most of the field problems are of complex nature because of the inhomogeneous anisotropic nature of flow systems and irregular shape of their boundaries. Such problems are not easily amenable to analytical methods and can be solved by using numerical methods. Development of numerical methods is based on two schemes: finite difference and finite elements. Accordingly, these methods are called finite difference and finite elements. These methods are described in detail in many published works [1, 31–36]. Based on these numerical methods, many computer programs such as SUTRA, MODFLOW, POR-FLOW, etc. are being developed and widely used to solve actual field problems of groundwater flow.

Although the application of analytical solutions is restricted to the relatively homogeneous isotropic flow system having boundaries of simple geometrical shapes, their application is fast and simple compared with that of the numerical methods. Analytical solutions are also useful for other purposes such as sensitivity analysis of the effects of various controlling parameters such as aquifers properties, initial and boundary conditions, intensity and duration of recharge rate, shape, size, and location of the recharge basin, etc. on water table fluctuation. Such information is very essential for making judicious selection of an appropriate development scheme out of many proposed schemes to achieve the preset objectives of groundwater resources management. Besides these applications, analytical solutions are also used for checking validity and calibration of numerical models under development by comparing results obtained from both the approaches. Analytical methods commonly used for the solution of groundwater problems include the Laplace transforms, integral balance methods, method of separation of variables, approximate analytic methods, Fourier transforms, etc. Details about these methods and their applications in the solution of groundwater flow problems or in heat conduction problems can be found in many books [1, 37–42]. A review of analytical solutions has been presented in [43]. Some commonly used analytical methods are discussed in the following subsections.

9.4.1 Laplace Transform

The technique of the Laplace transformation is widely used for solving diffusion type differential equation that contains a first order differential in time. By using the Laplace transform the partial derivative with respect to time variable is removed from the second order partial differential equation (in this case groundwater flow). As a result, the original second order partial differential equation is reduced in second order ordinary differential equation. When the ordinary differential equation is solved and this solution is inverted by using inversion of the associated the Laplace

transform; the desired solution of the unknown variable such as water table height is obtained. The Laplace transform of a function $h(t)$ is defined as:

$$L[h(t)] - h(p) = \int_0^\infty e^{-pt} h(t')dt', \tag{9.35}$$

and its inversion is:

$$h(t) \equiv \frac{1}{2\pi i} \int_{y-i\infty}^{y+i\infty} e^{pt} h(p)\, dp, \tag{9.36}$$

where y is a constant so large that all the singularities lie to the left of the line ($y-i\infty$, $y+i\infty$) on the complex p-plane. Generally, the Laplace transform of a function and its inverse are given in several books [38, 44–45]. Application of the Laplace transform in the solution of groundwater flow problems can be found in [19, 46–49].

9.4.2 The Integral Balance Method

This method is applicable to both linear and nonlinear 1-D transient boundary value problem for certain boundary conditions. The results are approximate. But several solutions obtained by applying this method when compared with the exact solutions have confirmed that the accuracy is generally acceptable. The following steps are followed in the application of this method:

(i) The differential equation describing 1-D groundwater flow is integrated over the length of the aquifer in order to remove the derivative with respect to space coordinate.
(ii) A suitable profile is chosen for the distribution of water table height. A polynomial profile is generally preferred for this purpose. Experience has shown that there is no significant improvement in the accuracy of the solution by choosing a polynomial greater than the fourth degree. Coefficients in the polynomial are determined by applying boundary conditions.
(iii) When the expression of the polynomial profile is introduced into the integrated groundwater flow equation and the indicated operations are performed, a first order ordinary equation is obtained for the average height of the water table with time as the independent variable. The solution of this differential equation subject to the initial condition gives an expression of the initial condition for average height of the water table.
(iv) Now the first order differential equation for the average water table height is solved subject to the initial condition for the average height to get the desired solution of the water table height.

Singh and Rai [50–51] have used this method in the solution of ditch-drainage problems in the presence of time varying recharge rate.

9.4.3 Approximate Analytic Methods

Approximate analytic methods are attempted in solving groundwater flow governing equation in its nonlinear form. One example of such a method is presented by Basak [52] in solving a ditch-drainage problem based on the assumption that the first derivative of the water table height with respect to time is independent of the space coordinate, i.e., $dh/dt \neq f(x)$ and is a function of time only. This approximation is valid when the successive water table profiles are almost parallel. This condition is satisfied almost in the entire region except near the drains. The accuracy of approximate solution is verified by comparing the results with the results of known exact solutions. Basak's solution is found to be in close agreement with an exact solution of the same problem. Singh and Rai, [50, 53] used this method to obtain a model to describe water table fluctuation induced by exponentially decaying recharge rates.

9.4.4 Method of Separation of Variables

In this method groundwater flow equation is separated into ordinary differential equations for each independent variable. The resulting ordinary differential equations are solved and the complete solution is constructed by the linear superposition of all separated solutions. Examples of application of this method can be found in [54] for 1-D sloping aquifer, in [6] for 2-D sloping aquifer, and in [55] for radial flow.

9.4.5 Finite Fourier Transforms

These transforms are useful in solving boundary value problems in which at least two of the boundaries are parallel and separated by a finite distance.

1-D Finite Fourier Sine Transform

This transform is used when value of a variable is specified at the boundaries. For 1-D case, the finite Fourier sine transform $S(m, t)$ with respect to x of a function $H(x, t)$, $0<x<A$ is defined as:

$$F_s[H(x,t)] = S(m,t) = \int_0^A H(x,t) \sin\left(\frac{m\pi x}{A}\right) dx, \tag{9.37}$$

its inversion is given by:

$$H(x,t) = \frac{2}{A}\sum_{m=1}^{\infty} S(m,t)\sin\left(\frac{m\pi x}{A}\right), \tag{9.38}$$

in which m is integer representing the number of Fourier coefficients and A is the length of the aquifer. Application of this transform in the solution of groundwater flow problem can be found in [21, 56, 57].

2-D Finite Fourier Sine Transform

Finite Fourier sine transform, $S(m, n, t)$ with respect to x and y of a function $H(x, y, t)$, $0<x<A$ and $0<y<B$ is given by:

$$F_s[H(x,y,t)] = S(m,n,t) = \int_0^B \int_0^A H(x,y,t)\sin\frac{m\pi x}{A}\sin\frac{n\pi y}{B}dxdy, \tag{9.39}$$

in which m and n are integers representing number of Fourier coefficients and A and B are length and width of the aquifer in the x and y directions, respectively. Its inversion formula is given by:

$$H(x,y,t) = \frac{4}{AB}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} S(m,n,t)\sin\left(\frac{m\pi x}{A}\right)\sin\left(\frac{n\pi y}{B}\right). \tag{9.40}$$

Example of application of this transform in the solution of groundwater flow equation can be found in [27, 58]. Recently, [29] have used this transform to develop a model to describe water table fluctuation in anisotropic aquifer.

1-D Finite Fourier Cosine Transform

This transform is used to solve 1-D flow equation where flux is defined at the boundaries. The finite Fourier cosine transform in the interval $0 < x < A$ is defined as:

$$C(m,t) = \int_0^A H(x,t)\cos\left(\frac{m\pi x}{A}\right)dx, \tag{9.41}$$

and its inverse is given by:

$$H(x,t) = \frac{1}{A}C(0,t) + \frac{2}{A}\sum_{m=1}^{\infty} C(m,t)\cos\left(\frac{m\pi x}{A}\right). \tag{9.42}$$

Examples of application of this transform can be found in [21, 59].

2-D Finite Fourier Cosine Transform

This transform is used for the solution of those problems in which the boundary conditions are characterized by flux across the two parallel boundaries. For 2-D problem, the finite Fourier cosine transform $C(m, n, t)$ with respect to x and y of a function $H(x, y, t)$ in the intervals $0 < x < A$ and $0 < y < B$ is defined as:

$$C(m,n,t) = \int_0^A \int_0^B H(x,y,t) \cos\left(\frac{m\pi x}{A}\right) \cos\left(\frac{n\pi y}{B}\right) dydx, \tag{9.43}$$

and its inverse is given by:

$$H(x,y,t) = \tfrac{1}{AB} C(0,0,t) + \tfrac{2}{AB} \sum_{m=1}^{\infty} C(m,0,t) \cos\left(\tfrac{m\pi x}{B}\right)$$
$$+ \tfrac{2}{AB} \sum_{m=1}^{\infty} C(0,n,t) \cos\left(\tfrac{n\pi y}{A}\right) + \tfrac{4}{AB} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} C(m,n,t) \cos\left(\tfrac{m\pi x}{B}\right) \cos\left(\tfrac{n\pi y}{A}\right). \tag{9.44}$$

Example of application of this transform is in [20, 28].

1-D Extended Finite Fourier Cosine Transform

This transform is used when the flow problem is characterized with mixed boundary conditions, i.e., at one boundary flux is defined and at its parallel boundary head is defined. This transform in the interval $0 < x < A$ is defined as:

$$C_e[m,t] = \int_0^A H(x,t) \cos \frac{(2m+1)\pi x}{2A} dx, \tag{9.45}$$

and its inverse is given by:

$$H(x,t) = \frac{2}{A} \sum_{m=0}^{\infty} C_e(m,t) \cos\left(\frac{(2m+1)\pi x}{2A}\right). \tag{9.46}$$

Application of this transform in the solution of 1-D groundwater flow can be found in [55].

2-D Extended Finite Fourier Cosine Transform

For 2-D flow problem, the transform $C(m, n, t)$ with respect to x and y in the interval $0<x<A$ and $0<y<B$ of a function $H(x, y, t)$ is defined as:

$$C_e(m,n,t) = \int_0^B \int_0^A H(x,y,t) \cos \frac{(2m+1)\pi\, x}{2A} \cos \frac{(2n+1)\pi y}{2B} dxdy, \tag{9.47}$$

its inversion formula is given by:

$$H(x,y,t) = \frac{4}{AB} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_e(m,n,t) \cos\left(\frac{(2m+1)\pi x}{2A}\right) \cos\left(\frac{(2n+1)\pi y}{2B}\right). \tag{9.48}$$

This transform has been used in [25] to develop a groundwater flow model.

9.5 Summary

Recharging and pumping are the essential components of water resources development. Therefore, prediction of water table fluctuations in response to proposed schemes of recharging and pumping are essential to make judicious selection of an appropriate development scheme out of many to achieve the preset objectives of sustainable management. This is accomplished by carrying out sensitivity analysis of the effects of changes in the controlling parameters on the dynamic behavior of the water table. Controlling parameters include shape, size, and location of recharge basins and wells, intensity of recharge and pumping rates, duration and number of cycles of recharge and pumping operations, etc. Solution of the prediction problem lies in the solution of the governing flow equations subject to the initial and boundary conditions associated with the physical problems under consideration. In this chapter groundwater flow equations have been presented to describe 2-D groundwater flows in inhomogeneous anisotropic unconfined aquifer (Eq. 9.16), inhomogeneous, isotropic unconfined aquifer (Eq. 9.17), in leaky unconfined aquifer (Eq. 9.20) in sloping homogeneous isotropic sloping aquifer (Eq. 9.26) in response to intermittently applied time varying recharge and/or pumping from multiple basins of rectangular shapes and wells, respectively, along with the initial and boundary conditions and methods of their solutions. Groundwater flow equations to describe 1-D flow can be obtained by substituting zero for the derivative of one space coordinate.

Governing flow equation in cylindrical coordinate system (Eq. 9.27) is also presented to describe groundwater flow induced by time varying recharge from circular basin. Groundwater flow equations for steady state can be obtained by substituting zero for the time derivative. The above mentioned governing flow equations are used for the development of analytical/numerical models to predict water table fluctuations in the flow system under consideration. Examples of the analytical models have been cited from the papers published in reputed journals which can be easily accessible to the interested readers. Though the application of analytical models are restricted to the flow system having boundaries of simple geometrical shapes, their application is fast and simple compared to numerical methods. Analytical models are also used to check the validity of numerical models under development, because of the assured accuracy of the results of analytical models.

References

1. Bear J (1979) Hydraulics of groundwater. McGraw-Hill, New York, pp 576
2. Rai SN (2002) Groundwater flow modeling. In: Rai SN et al (eds) Dynamics of earth's fluid system. A.A.Balkema, Netherland, pp 291
3. Marino MA (1967) Hele–Shaw model study of the growth and decay of groundwater ridges. J Geophys Res 72:1195–1205
4. Rao NH, Sarma PBS (1980) Growth of groundwater mound in response to recharge. Groundwater 18:587–595
5. Bauman P (1965) Technical development in groundwater recharge. In: Chow VT (ed) Advances in hydroscience, vol 2. Academic Press, New York, pp 209–279
6. Ramana DV, Rai SN, Singh RN (1995) Water table fluctuation due to transient recharge in a 2–D aquifer system with inclined base. Water Resour Manag 9:127–138
7. Mercer JW, Faust CR (1980) Groundwater modeling: mathematical models. Groundwater Nov:212–227
8. Rai SN, Ramana DV, Singh RN (1998) On the prediction of ground water mound formation in response to transient recharge from a circular basin. Water Resour Manag 12:271–284
9. Todd DK (1959) Groundwater hydrology. Wiley, New York, pp 336
10. Hantush MS (1967) Growth and decay of groundwater mounds in response to uniform percolation. Water Resour Res 3:227–234
11. Hunt BW (1971) Vertical recharge of unconfined aquifer. J Hydraul Div ASCE 96(HY7):1017–1030
12. Marino MA (1974) Rise and decline of water table induced by vertical recharge. J Hydrol 23:289–298
13. Rao NH, Sarma PBS (1981) Recharge from rectangular areas to finite aquifers. J Hydrol 53:269–275
14. Rao NH, Sarma PBS (1983) Recharge to finite aquifers from strip basins. J Hydrol 66:245–252
15. Rao NH, Sarma PBS (1984) Recharge to aquifers with mixed boundaries. J Hydrol, 74:43–51
16. Detay M (1995) Rational groundwater reservoir management, the role of artificial recharge. In: Johnson AI, Pyne RDJ (eds) Artificial recharge of groundwater II. ACSE, New York, pp 231–240
17. Dickensen JM, Bachman SB (1995) The optimization of spreading ground operation. In: Johnson AI, Pyne RDJ (eds) Artificial recharge of groundwater II. ACSE, New York, pp 630–639
18. Zomorodi K (1991) Evaluation of the response of a water table to a variable recharge. Hydrol Sci J, 36:67–78
19. Rai SN, Singh RN (1981) A mathematical model of water table fluctuations in a semi-infinite aquifer induced by localized transient recharge. Water Resour Res 17(4):1028–1032
20. Rai SN, Manglik A, Singh RN (1994) Water table fluctuation in response to transient recharge from a rectangular basin. Water Resour Manag 8(1):1–10
21. Rai SN, Singh RN (1995a): An analytical solution for water table fluctuation in a finite aquifer due to transient recharge from a strip basin. Water Resour Manag 9:27–37
22. Rai SN, Singh RN (1996a) On the prediction of ground water mound formation due to transient recharge from a rectangular area. Water Resour Manag 10:189–198
23. Yue-zan T, Mei Y, Bing-feng Z (2007) Solution and its application to transient stream/groundwater model subjected to time dependent vertical seepage. Appl Math Mech 28(9), 1173–1180
24. Manglik A, Rai SN, Singh RN (1997) Response of an unconfined aquifer induced by time varying recharge from a rectangular basin. Water Resour Manag 11:185–196
25. Rai SN, Manglik A (1999) Modeling of water table variation in response to time varying recharge from multiple basins using the linearized Boussinesq equation. J Hydrol 220:141–148

26. Manglik A, Rai SN (2000) Modeling of water table fluctuation in response to time varying recharge and withdrawal. Water Resour Manag 14(5):339–347
27. Rai SN, Manglik A (2012) An analytical solution of Boussinesq equation to predict water table fluctuation due to time varying recharge and withdrawal from multiple basins, well and leakage sites. Water Resour Manag 26:243–252
28. Manglik A, Rai SN, Singh VS (2004) Modeling of aquifer response to time varying recharge and pumping from multiple basins and wells. J Hydrol 292:23–29
29. Manglik A, Rai SN, Singh VS (2013) A generalized predictive model of water table fluctuations in anisotropic aquifer due to intermittently applied time varying recharge from multiple basins. Water Resour Manag 27:25–26
30. Rai SN, Manglik A, Singh VS (2006) Water table fluctuation owing to time-varying recharge, pumping and leakage. J Hydrol 324:350–358
31. Narasimhan TN (1975) A unified numerical model for saturated—unsaturated groundwater flow, Ph.D. dissertation, Department of Civil Engineering, University of California, Berkeley, California.
32. Runchal AK, BudhiSagar (1992) PORFLOW—a model for fluid flow, heat and mass transport in multi-fluid, multiphase, fractured, or porous media. v.2.41, Analytic and computational research Inc
33. Pinder GF, Gray WG (1997) Finite element simulation in surface and subsurface hydrology. Academic press, New York, pp 295
34. Voss CI (1984) SUTRA—a finite element simulation model for saturated—unsaturated fluid density dependent groundwater flow with energy transport or chemically single species contaminant transport. USGS Water Resources, investigate report, 84-4269, pp 409
35. Rushton KR (2002) Groundwater hydrology: conceptual and computational models. Wiley, New York, pp 407
36. Chiang W-H, Kinzelbach W (2003) 3-D groundwater modeling with PMWIN. Springer Verlag, Berlin, pp 346
37. Polubarinova–Kochina PY (1962) Theory of groundwater movement. Princeton University Press, Princeton, pp 613
38. Ozisik MN (1980) Heat conduction. Wiley, New York, pp 687
39. Sneddon IN (1974) The use of integral transforms. Tata McGraw Hill, New Delhi
40. Lee T-C (1999) Applied mathematics in hydrogeology. Lewis, Boca Raton, pp 382
41. Bruggeman GA (1999) Analytical solution of geohydrological problems. Elsevier, Amsterdam, pp 959
42. Rai SN (2004) Analytical methods of solution. In: Rai SN (eds) Role of mathematical modeling. DST, New Delhi, pp 445
43. Rai SN, Singh RN (1996b) Analytical modeling of unconfined flow induced by time varying recharge. Proc Indian Natl Sci Acad 62:253–292
44. Carslaw MS, Jaeger JC. (1959) Conduction of heat in solids. Oxford Univ Press, Oxford, pp 510
45. Abramowidz M, Stegun IA (1970) Hand book of mathematical functions. Dover, New York, pp 1046
46. Rai SN, Singh RN (1979) Variation of water table induced by time varying recharge. Geophys Res Bull 17:97–109
47. Rai SN, Singh RN (1980) Dynamic response of an unconfined aquifer subjected to transient recharge. Geophys Res Bull 18:50–56
48. Rai SN, Singh RN (1985) Water table fluctuations in response to time varying recharge. Scientific basis for water resources management, IAHS Publ. no. 153, 287–294
49. Rai SN, Singh RN (1992) Water table fluctuations in an aquifer system due to time varying surface infiltration and canal recharge. J Hydrol 136:381–387
50. Singh RN, Rai SN (1983) On an approximate analytical solution of the Boussinesq equation for a transient recharge. In: Plate E, Buras N (eds) Scientific procedures applied to planning, design and management of water resources system, IAHS Publ. No. 147, 140–148

51. Singh RN, Rai SN (1989) A solution of the nonlinear Boussinesq equation for phreatic flow using an integral balance approach. J Hydrol 109:313–323
52. Basak P (1979) An analytical solution for the transient ditch drainage problem. J Hydrol 41:377–382
53. Singh RN, Rai SN (1980) On subsurface drainage of transient recharge. J Hydrol 48:303–311
54. Singh RN, Rai SN, Ramana DV (1991) Water table fluctuation in a sloping aquifer with transient recharge. J Hydrol 126:315–326
55. Rai SN, Singh RN (1998) Evolution of the water table in a finite aquifer due to transient recharge from two parallel strip basins. Water Resour Manag 12:199–208
56. Rai SN, Manglik A (2001) Modeling of water table fluctuations due to time-varying recharge from canal seepage. In: Seiler K-P, Wohnlich S (eds) New approaches characterizing groundwater flow. Balkema, The Netherlands, pp 775–778
57. Rai SN, Ramana DV, Thiagarajan S, Manglik A (2001) Modeling of groundwater mound formation due to transient recharge. Hydrol Process 15:1507–1514
58. Rai SN, Singh RN (1995b) Two dimensional modeling of water table fluctuation in response to localized transient recharge. J Hydrol 167:167–174
59. Rai SN, Ramana DV, Manglik A (1997) Modelling of water table fluctuation in finite aquifer system in response to transient recharge. In: Singhal DC et al (eds) Proceeding of international symposition on emerging trends in hydrology 1, 243–250

Chapter 10
Two-Dimensional Solute Transport from a Varying Pulse-Type Point Source

Premlata Singh, Sanjay Kumar Yadav and Alexander V. Perig

10.1 Introduction

There are many sources from which certain mass in the form of solute particles or tracer particles are released and spread through a medium. Those of pollutants category degrade the environment. The sources of such pollutants are from all walks of human life: industries, agricultural lands, municipal drainage systems, garbage disposal sites, mines, hydro, thermal, nuclear power plants, etc. Some of the sources are natural too, like the one due to volcanic eruption. Pollutants from these sources, in one or the other form, get mixed with air, surface water, groundwater, and soil making the environment unsuitable for life to survive on the earth. Instantaneous sources are a few. So, their impact also exists for a short period of time. Most of these sources are of continuous nature, i.e., they are active for a long time and are bound to remain as such in the future. They are of uniform nature (the input pollutant remains uniform throughout its existence) or of varying nature (the input is either increasing or decreasing with time). Once the cause of the pollution is removed, the input concentration either becomes zero or starts decreasing with time, and the region starts getting rehabilitated. Continuously increasing the sources of pollutants have become a matter of great concern. To assess the impact of pollutants in the presence and after the removal of a source on the environment, it is necessary to know the solute-mass-transport behavior and the variation pattern in its

P. Singh (✉)
Department of Applied Mathematics, Birla Institute of Technology,
Mesra, Extension center, Patna 800014, India
e-mail: prema.singh@rediffmail.com

S. K. Yadav
Department of Mathematics, Banaras Hindu University,
Varanasi 221005, India
e-mail: skyvns@gmail.com

A. V. Perig
Department of Technical Mechanics, Donbass State Engineering Academy,
Shkadinova 72, Kramatorsk 84313, Ukraine
e-mail: olexander.perig@gmail.com

S. K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,*
Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_10,

concentration level with position and time. One way is to use mathematical models in continuum approach. In this approach, the solute transport through a convective medium is described by a parabolic-type partial differential equation, derived on the principle of conservation of mass and is known as the advection-diffusion equation (ADE). Its general form in one dimension is:

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D(x,t)\frac{\partial c}{\partial x} - u(x,t)c \right), \tag{10.1}$$

where $D(x,t)$ is known as the solute dispersivity parameter and $u(x,t)$ is the velocity of the flow domain through a medium transporting the solute particles, and c is the solute concentration at a position x at time t. The determination of the longitudinal distribution of waste products in a waterway is often based on one-dimensional mass balance equation which includes advection and dispersion terms. The advective mass transfer is represented by the velocity averaged over a cross section normal to the longitudinal axis of the waterway. The dispersion term accounts for longitudinal mixing which results from the combined effects of turbulent diffusion and the shear-induced velocity distribution in both the transverse and vertical directions.

Mathematical modelers use ADE to describe the concentration levels at different positions and time away from its source through its analytical and numerical solutions. Most of the analytical solutions for advection-diffusion transport problems in ideal conditions with growth and decay terms, subject to various initial and boundary conditions in semi-infinite and finite regions have been compiled by different researchers [1–3]. Predicting the fate of pollutants in natural environments such as rivers and man-made channels is one of the major concerns. The transport of reacting species is affected by the changes produced in the chemical composition of the environment. The number of studies in reactive transport has increased, showing the importance of this issue [4–7]. In a recent work [8], a model was developed for investigating the solute transport into a sub-aqueous sediment bed, under an imposed standing water surface. Following the formulation of the ADE and the theories relating D and u [9–11], the number of solute transport studies has increased considerably. Many solute transport models consider homogeneous media but in reality the ability of solute to permeate though the medium of air, soil, or groundwater varies with position, which is referred to as heterogeneity. Early efforts to describe heterogeneity were achieved by making the use of stratification and defining porosity–distance relationship [12–15]. In the former situation, the number of layers cannot be large and is restricted up to two or three only to get analytical solution for each layer. In the latter situation, the relationships are valid for finite domains and only numerical method is the option to deal with the dispersion problems. Later, scale-dependent dispersion has been attributed to heterogeneity.

According to another theory [16], some large subsurface formations exhibit variable dispersivity as a function of position or time variables while the flow domain remains uniform. This theory was fully supported through later works [17]. Analytical solutions to solute transport problem in a semi-infinite medium were

obtained [18] based on such observations, where the dispersion parameter depends on distance and increases up to a limited value. This problem was extended [19] for periodic input condition on a semi-infinite domain and porous media with fracture zones, and included the adsorption effects. An analytical solution was derived in [20] for linear-asymptotic distance-dependent dispersion problems. The limitations of analytical solution for ADE with coefficients being the function of space variable have also been analyzed [21]. But it is possible to solve this equation analytically only in some particular cases. In a more general situation, numerical techniques are required [22]. The literature presents several analytical methods, to solve the partial differential equations governing transport phenomena. Exact solutions of linear diffusion problems by the classical integral transform techniques were reviewed and classified in [23]. This work identified and unified seven classes of problems and demonstrated many applications in heat and mass diffusion. This work was later generalized and extended [24], thereby creating a new systematic procedure referred to as the generalized integral transform technique (GITT) used in later works [25–27]. Most of the analytical methods are also reviewed in a recent work [28].

In both [9] and [29], it is pointed out that their analyses of the dispersion coefficient are limited to asymptotically large times after the introduction of pollutant into the flow. This means that one-dimensional ADE is not applicable to a dispersing mass cloud immediately after the introduction of the pollutant. So, modeling the dispersion problems in two and three dimensions have been thought to be a better option compared with the one-dimensional model. Several other authors have used simplified one- and two-dimensional models that to some extent incorporate variable coefficients [30–33]. The ADE with spatially variable coefficients in three dimensions with particular functional forms for the coefficients was analytically solved in [34]. A general methodology to develop dispersion models in three-dimensional heterogeneous aquifers under nonstationary conditions was presented in [35]. Analytical solutions of one-, two-, and three-dimensional ADEs were obtained in [36]. The author assumed that dispersivity increases directly with the first power of the flow length in the steady and unsteady flows. Depth-averaged solute transport and lateral diffusive transport were modeled in a two-layer system of contrasting permeabilities [37]. They obtained two-dimensional analytical solutions for the first-order rate model in an infinite medium, using the methods of Fourier and Laplace integral transformations. The analytical solution for two-dimensional chemical transports through an aquifer was presented in [38]. Time-dependent infinite element approach was used to simulate contaminant transport problems in infinite media [39]. They considered one- and two-dimensional ADEs with constant or variable coefficients, addressing the anisotropy of the media as well as leakage effects in porous fissured media. A two-dimensional semi-analytical solution was presented to analyze stream–aquifer interactions in a coastal aquifer, where groundwater is affected by tidal effects [40]. The work in [41] is devoted to mathematical modeling and computer simulation of diffusion and transport of chemicals in rivers, in which one-, two-, and three-dimensional models in terms of time-dependent convection-diffusion-reaction differential equations are presented. Analytical solution for time-varying dispersion coefficients have been presented in one dimension [42, 43], two

dimensions [44], and three dimensions [45, 46]. The Hankel transform technique was used in [47] to describe two-dimensional solute transport along unsteady flow from a time-dependent point source.

In view of such wide applicability of ADE, its analytical solutions in more real dispersion problems are always in demand. The present work is an attempt in that direction using Laplace integral transform. A two-dimensional ADE is solved analytically in a heterogeneous horizontal semi-infinite plane. The solute transport is due to a varying pulse-type stationary point source. Solutions with such type of input condition are very useful in estimating rehabilitation time of a polluted domain, once the source of pollution is eliminated. Particularly in groundwater domain, where one major source of pollution is the disposal of a variety of wastes on the surface, the input concentration increases with time in the presence of the source. Once it is dumped or removed, the input concentration starts decreasing, instead of becoming zero at once (as is the case with uniform point source). Due to heterogeneity of the semi-infinite medium velocity, the dispersivity components are considered as linear function of the respective space variable. Owing to the observations in [17, 21], to accommodate other causes affecting the two parameters of the solute transport over long space and time domains, both are considered temporally dependent too. The two-dimensional ADE with variable coefficients is reduced into an ADE in one dimension with constant coefficients.

10.2 Mathematical Formulation and Analytical Solution

Let the longitudinal and lateral directions at the origin be taken as the x and y axes, respectively. Let c be the contaminants concentration in the aquifer at any time t at position (x, y), $u(x,t)$, and $v(y,t)$ be the velocity components while $D_x(x,t)$ and $D_y(y,t)$ be the dispersions coefficients at the same position along the x and y axes, respectively. The linear advection–diffusion partial differential equation in two-dimensional horizontal isotropic but heterogeneous medium in general form may be written as:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left[D_x(x,t)\frac{\partial C}{\partial x} - u(x,t)C \right] + \frac{\partial}{\partial y}\left[D_y(y,t)\frac{\partial C}{\partial y} - v(y,t)C \right] \quad (10.2)$$

It is assumed that due to heterogeneity, velocity increases linearly with position variable in a direction. In a finite domain $0 \le x \le \ell_1$ along the longitudinal direction, let the velocity at $x = 0$ be $u = u_0$, and at $x = \ell_1$ be $u_0(1+\alpha_1)$, where α_1 is a nondimensional constant greater than zero. So, the velocity at an intermediate position may be interpolated as $u_0(1+ax)$, where $a = (\alpha_1 / \ell_1)$ may be considered as the heterogeneity parameter along the longitudinal direction as inverse of the space variable. Similarly, the lateral velocity component may be interpolated as $v_0(1+by)$ in a domain $0 \le y \le \ell_2$, where $b = (\alpha_2 / \ell_2)$ may be termed as the heterogeneity parameter along the lateral direction. As the flow domain has to satisfy the Darcy

law (in case of porous flow) or laminar conditions (in the case of air or water flow), the values of a and b should be of small order.

In case both have different values, it means heterogeneity along the longitudinal direction is different from that along the lateral direction. The different values represent media of different heterogeneities. The dispersivity due to heterogeneity is considered square of the velocity, in each direction. Additionally, unsteadiness of velocity and dispersivity are considered. These are described by $f_1(mt)$ and $f_2(mt)$, respectively, where the coefficient m is termed as the unsteady parameter with dimension inverse of t. Heterogeneity along both the directions is considered different but unsteadiness is considered the same. Expressions for each coefficient are considered in degenerate forms as in [48–50].

$$u(x,t) = u_0 f_1(mt)\,(1+ax) \quad \text{and} \quad v(y,t) = v_0 f_1(mt)\,(1+by), \tag{10.3}$$

$$D(x,t) = D_{x0} f_2(mt)\,(1+ax)^2 \quad \text{and} \quad D(y,t) = D_{y0} f_2(mt)\,(1+by)^2 \tag{10.4}$$

While choosing expressions for $f_1(mt)$ and $f_2(mt)$, it is ensured that $f_1(mt) = 1 = f_2(mt)$ for $m = 0$ or $t = 0$ The former case represents the steady flow and dispersion while the latter case represents velocity and dispersion at the initial stage. So the coefficients u_0, v_0 and D_{x0}, D_{y0} in the above equations may be referred to as the uniform velocity components of dimension (LT^{-1}) and the initial dispersion coefficient components of dimension (L^2T^{-1}), respectively.

In this chapter, we are considering two subsections based on the relationship:

$$f_2(mt) = f_1^n(mt) \tag{10.5}$$

For any value of n between 1 and 2, the theory [51] holds good. In the first subsection $n = 2$ is considered, and in the second one $n \neq 2$ is considered. For $n = 2$, each dispersivity component, due to both heterogeneity and unsteadiness, becomes proportional to the square of the respective velocity component, it thus satisfies the theories in [9–10].

For $n = 1$, we get the theory in [11]. For $f_1(mt) = 1$ and $f_2(mt) = f(mt)$, the theory in [16] is applicable. To solve ADE (Eq. 10.2), the following initial and boundary conditions are assumed:

Initially, the domain is considered solute free, i.e.,:

$$c(x,y,t) = 0,\ x \geq 0,\ y \geq 0, \quad t = 0 \tag{10.6}$$

The point source is a varying pulse type and is supposed to be at the origin of the horizontal medium, i.e., this condition is the combination of the following two equations:

$$-D_x(x,t)\frac{\partial c}{\partial x} + uc(x,y,t) = \begin{cases} uC_0, & 0 < t \leq t_0 \\ 0, & t > t_0 \end{cases},\ x = 0, y = 0, \tag{10.7a}$$

$$\text{and} -D_y(y,t)\frac{\partial c}{\partial y} + vc(x,y,t) = \begin{cases} vC_0, & 0 < t \leq t_0 \\ 0, & t > t_0 \end{cases},\ x = 0, y = 0 \tag{10.7b}$$

It describes that the concentration at the origin increases with time in the presence of the source. Once the source of pollution is eliminated at $t = t_0$, the input concentration starts decreasing. The second boundary conditions at the far ends along both the directions are of homogeneous flux type:

$$\frac{\partial c}{\partial x} = 0, \frac{\partial c}{\partial y} = 0, x \to \infty, y \to \infty, t \geq 0 \tag{10.8}$$

10.2.1 Dispersivity as a Square of the Velocity

Using the Eq. 10.5 for $n = 2$, in the Eqs. 10.2 and 10.3, ADE (10.2) becomes:

$$\frac{\partial C}{\partial T^*} = \frac{\partial}{\partial x}\left[D_{x0} f^2 (mt)(1+ax)^2 \frac{\partial C}{\partial x} - u_0 f(mt)(1+ax)C \right]$$
$$+\frac{\partial}{\partial y}\left[D_{y0} f(mt)^2 (1+by)^2 \frac{\partial C}{\partial y} - v_0 f(mt)(1+by)C \right], \tag{10.9}$$

where a new independent variable T^* is introduced by using the transformation [52]:

$$T^* = \int_0^t f(mt)dt \quad \text{or} \quad \frac{\partial T^*}{\partial t} = f(mt) \tag{10.10}$$

As m occurs in the denominator, hence T^* may be referred to as a new time variable. Also, an expression of $f(mt)$ is chosen such that for $t = 0$, we get $T^* = 0$; the nature of the initial condition does not change in the new time domain. Let the time of elimination of the source have a corresponding value T_0^*. Further, the following new space variables are introduced through the transformations [48, 49]:

$$X = \int \frac{dx}{(1+ax)} = \frac{1}{a}\log(1+ax), \tag{10.11a}$$

$$\text{and} \quad Y = \int \frac{dx}{(1+by)} = \frac{1}{b}\log(1+by) \tag{10.11b}$$

The longitudinal and lateral directions remain semi-infinite in the new space variables. The partial differential equation given by the Eq. 10.9 together with the initial and boundary conditions (Eqs. 10.6 through 10.8) become:

$$\frac{\partial c}{\partial T^*} = D_{x0} f(mt)\frac{\partial^2 c}{\partial X^2} - \left\{u_0 - aD_{x0} f(mt)\right\}\frac{\partial c}{\partial X} - au_0 c$$
$$+D_{y0} f(mt)\frac{\partial^2 c}{\partial Y^2} - \left\{v_0 - bD_{y0} f(mt)\right\}\frac{\partial c}{\partial Y} - bv_0 c, \tag{10.12}$$

$$c(X, Y, T^*) = 0, X \geq 0, Y \geq 0 \quad T^* = 0, \tag{10.13}$$

$$-D_{x0}f(mt)\frac{\partial c}{\partial X}+u_0c(X,Y,T^*)=\begin{cases}u_0C_0, & 0<T^*\le T_0^*\\ 0, & T^*>T_0^*\end{cases},\quad X=0,\ Y=0, \tag{10.14}$$

$$-D_{y0}f(mt)\frac{\partial c}{\partial Y}+v_0c(X,Y,T^*)=\begin{cases}v_0C_0, & 0<T^*\le T_0^*\\ 0, & T^*>T_0^*\end{cases},\quad X=0,Y=0, \tag{10.15}$$

$$\frac{\partial c}{\partial X}=0,\frac{\partial c}{\partial Y}=0,\quad X\to\infty,Y\to\infty,T^*\ge 0 \tag{10.16}$$

Further, using another space variable introduced through a transformation similar to the one used in an earlier work [53]:

$$Z=aX+bY, \tag{10.17}$$

and assuming:

$$f_1(mt)=1-\lambda f(mt), \tag{10.18}$$

where $\lambda=(D/U)$, $D=a^2D_{x0}+b^2D_{y0}$, and $U=au_0+bv_0$, the present initial and boundary value problem becomes:

$$\frac{\partial c}{\partial T^*}=Df(mt)\frac{\partial^2 c}{\partial Z^2}-Uf_1(mt)\frac{\partial c}{\partial Z}-Uc, \tag{10.19}$$

$$c(Z,T^*)=0,\quad Z\ge 0,T^*=0, \tag{10.20}$$

$$-f(mt)D^*\frac{\partial c}{\partial Z}+U^*c(Z,T^*)=\begin{cases}U^*C_0, & 0<T^*\le T_0^*\\ 0, & T^*>T_0^*\end{cases},\quad Z=0, \tag{10.21}$$

$$\frac{\partial c}{\partial Z}=0,\quad Z\to\infty,T^*\ge 0, \tag{10.22}$$

where $D^*=aD_{x0}+bD_{y0}$, $U^*=u_0+v_0$

Now the first-order decay term in the Eq. 10.19 may be eliminated by using the transformation:

$$c(Z,T^*)=C(Z,T^*)\exp(-UT^*) \tag{10.23}$$

Finally, using the transformations:

$$\eta=Z\frac{f_1(mt)}{f(mt)}, \tag{10.24}$$

$$\text{and } T=\int_0^t f_1^2(mt)dt, \tag{10.25}$$

one by one, the variable coefficients of the ADE are reduced into constant coefficients. Thus, we get the one-dimensional ADE with constant coefficients in the new independent variables (η, T) as:

$$\frac{\partial C}{\partial T} = D\frac{\partial^2 C}{\partial \eta^2} - U\frac{\partial C}{\partial \eta}. \tag{10.26}$$

While choosing an expression for $f(mt)$, it is also ascertained that through the transformation (Eq. 10.25), we get $T = 0$, for $t = 0$, so that the nature of the initial condition does not change in this time domain. In this time domain, let the corresponding value of time of elimination of the source be T_0 Thus, initial condition (Eq. 10.20) becomes:

$$C(\eta, T) = 0,\ T = 0,\ \eta \geq 0. \tag{10.27}$$

Now to write the input condition (Eq. 10.21) in the time domain T, the time variable T^* has to be expressed explicitly in terms of T. For the purpose, expressions of $f_1(mt)$ and $f_2(mt)$ may be chosen. Most of the changes are of exponential nature, hence an expression of exponentially decreasing or increasing nature is considered as:

$$f(mt) = \exp(-mt) \tag{10.28a}$$

$$\text{or} \quad f(mt) = \exp(mt) \tag{10.28b}$$

Small value of m less than 1.0 ensures that the changes are of small order. The two expressions represent decelerating flow and accelerating flow, respectively. Using the decelerating function, from the Eq. 10.10, we have:

$$T^* = \frac{1}{m}\left[1 - \exp(-mt)\right] \quad \text{or} \quad mt = -\log(1 - mT^*)$$

Also from the Eq. 10.25, we have:

$$T = t + \frac{\lambda^2}{2m}\left\{1 - \exp(-2mt)\right\} - \frac{2\lambda}{m}\left\{1 - \exp(-mt)\right\}$$

$$\text{or} \quad T = \frac{1}{m}\left[-\log(1 - mT^*) + \frac{\lambda^2}{2}\left\{1 - (1 - mT^*)^2\right\} - 2\lambda\left\{1 - (1 - mT^*)\right\}\right].$$

In $f(mt)$, m is chosen much smaller than 1, so its second- and higher-degree terms in the logarithmic and binomial expansions in the above equation may be omitted. So, we may get:

$$T^* = \gamma T, \text{ where } \gamma = (1 - \lambda)^{-2} \tag{10.29}$$

It may be verified that the same relationship may be obtained for the expression in the Eq. 10.28b. Using this relationship, the input condition (Eq. 10.21) for decelerating flow, i.e., for the expression in the Eq. 10.28a may be written as:

$$-D^*(1-\lambda+\lambda m\gamma T)\frac{\partial C}{\partial \eta}+U^*C=\begin{cases}U^*C_0\exp(\gamma UT), & 0<T\le T_0\\ 0, & T>T_0\end{cases},\quad \eta=0. \tag{10.30a}$$

For the accelerating expression (Eq. 10.28b), it becomes:

$$-D^*(1-\lambda-\lambda m\gamma T)\frac{\partial C}{\partial \eta}+U^*C=\begin{cases}U^*C_0\exp(\gamma UT), & 0<T\le T_0\\ 0, & T>T\end{cases},\quad \eta=0. \tag{10.30b}$$

The second boundary condition (Eq. 10.22) may be obtained as:

$$\frac{\partial C}{\partial \eta}=0,\ \eta\to\infty,\ T\ge 0 \tag{10.31}$$

The convective term in the advection-diffusion Eq. 10.26 is eliminated by using the transformation:

$$C(\eta,T)=K(\eta,T)\exp\left(\frac{U}{2D}\eta-\frac{U^2}{4D}T\right) \tag{10.32}$$

We apply the above transformation on the Eqs. 10.26, 10.27, 10.30a, and 10.32 and then the Laplace transformation (parameter p); a second-order ordinary boundary value problem is obtained as:

$$D\frac{d^2\bar{K}}{dZ^2}=p\bar{K}, \tag{10.33}$$

$$\begin{aligned}&D^*\left[(\lambda-1)+m\lambda\gamma\frac{d}{dp}\right]\left[\frac{d\bar{K}}{dZ}+\frac{U}{2D}\bar{K}\right]+u\bar{K}\\&=\frac{U^*C_0}{p-\beta^2}\left[1-\exp-(p-\beta^2)T_0\right],\eta=0,\end{aligned} \tag{10.34a}$$

$$\text{and}\quad \frac{d\bar{K}(\eta,p)}{d\eta}+\frac{U}{2D}\bar{K}(\eta,p)=0,\ \eta\to\infty, \tag{10.35}$$

where $\beta^2=\left(U\gamma+\frac{U^2}{4D}\right)$ Its particular solution may be obtained as:

$$\bar{K}(\eta,p)=\frac{2U^*C_0\sqrt{p}\left[1-\exp-(p-\beta^2)T_0\right]}{A_1\left(p-\beta^2\right)\left(p+A_{21}\sqrt{p}+A_{31}\right)}\exp\left(-\eta\sqrt{p/D}\right), \tag{10.36}$$

A_1, A_2, etc. are defined with others after Eq. 10.38. Now $K(\eta,T)$ may be obtained by taking the inverse Laplace transform of the solution in the Eq. 10.36. For that,

the factor $p + A_{21}\sqrt{p} + A_{31}$ is assumed as $(\sqrt{p} + b)^2$. Then, $2b = A_{21}$ and $b^2 = A_{31}$ are substituted in $K(\eta, T)$. Using the necessary transformations back, the desired analytical solution may be obtained as:

$$c(x,y,t) = \begin{cases} F(x,y,t), & 0 < t \le t_0 \\ F(x,y,t) - F(x,y,t-t_0), & t > t_0 \end{cases}, \qquad (10.37a, b)$$

where:

$$\begin{aligned} F(x,y,t) = 2U^* C_0 \big[& B_1 (A + \beta E) + B_2 (A - \beta F) \\ & + B_3 \left\{ (1 + 2A_{31}T) A - (A_{21} + A_{31}Z + A_{21}A_{31}T) F \right\} \\ & + B_4 \left\{ (1 + (A_{21}/2)Z + 2A_{31}T) F - A_{21}TA \right\} \big], \end{aligned} \qquad (10.38)$$

$$A_1 = 2(1-\lambda)D^*\sqrt{D},\ A_2 = 2DU^* - (1-\lambda)D^*U,\ A_3 = -m\lambda\gamma D^*\sqrt{D},\quad \mathrm{A}_{21} = A_2 / A_1,$$

$$\mathrm{A}_{31} = A_3 / A_1,\quad D = a^2 D_{x0} + b^2 D_{y0},\quad U = au_0 + bv_0,\quad D^* = aD_{x0} + bD_{y0},$$

$$U^* = u_0 + v_0,\ B_1 = \frac{1}{2A_1(\beta^2 + \beta A_{21} + A_{31})},\quad B_2 = \frac{1}{2A_1(\beta^2 - \beta A_{21} + A_{31})},$$

$$B_3 = \frac{\beta^2 + A_{31}}{A_1(\beta A_{21})^2 - A_1(\beta^2 + A_{31})^2},\ B_4 = \frac{A_{21}A_{31}}{A_1(\beta A_{21})^2 - A_1(\beta^2 + A_{31})^2},$$

$$\beta^2 = \left(\gamma U + \frac{U^2}{4D} \right), \gamma = (1-\lambda)^{-2},\ \lambda = (D/U),\ f_1(mt) = 1 - \lambda f(mt),$$

$$A = \frac{1}{\sqrt{\pi T}} \exp\left(\frac{u_0}{2D_0}\eta - \frac{Z^2}{4D_0 T} - \beta^2 T \right),\ \eta = Z\frac{f_1(mt)}{f(mt)},\ Z = \log\{(1+ax)(1+by)\},$$

$$E = \exp\left(\frac{u_0}{2D_0} - \frac{\beta}{\sqrt{D_0}} \right) \eta \cdot \mathrm{erfc}\left(\frac{\eta}{2\sqrt{TD_0}} - \beta\sqrt{T} \right),\ T = T^*/\gamma,\ f(mt) = \exp(-mt),$$

$$F = \exp\left(\frac{u_0}{2D_0} + \frac{\beta}{\sqrt{D_0}} \right) \eta \cdot \mathrm{erfc}\left(\frac{\eta}{2\sqrt{TD_0}} + \beta\sqrt{T} \right),\quad \text{and}$$

$$T^* = \int_0^t f(mt)dt = \frac{1}{m}\left[1 - \exp(-mt)\right]$$

The solution given by the Eq. 10.37a describes the solute transport in the time domain $0 \le t \le t_0$, i.e., in the presence of the source, and that given by the Eq. 10.37b describes the same, once the source is eliminated, i.e., in the time domain $t > t_0$. For $f(mt) = \exp(mt)$, proceeding with the input condition (Eq. 10.30b) the same solution as in the Eq. 10.37a and 10.37b will be obtained with only one change: $A_3 = m\lambda\gamma D^* \sqrt{D}$.

10.2.2 Unsteadiness of Dispersion and the Velocity Being Related in a General Way

The value of n in the Eq. 10.5 may be between 1 and 2 or outside of this range. In other words, the relation between unsteadiness of dispersion and that of velocity is considered general in nature. We use the expressions for velocity and dispersivity components from Eqs. 10.3 and 10.4 in Eq. 10.2. We then use the transformations in Eqs. 10.11a and 10.11b to get:

$$\begin{aligned}\frac{\partial c}{\partial T^*} &= \frac{f_2(mt)}{f_1(mt)}\left(D_{x0}\frac{\partial^2 c}{\partial X^2} + D_{y0}\frac{\partial^2 c}{\partial Y^2} + aD_{x0}\frac{\partial c}{\partial X} + bD_{y0}\frac{\partial c}{\partial Y}\right) \\ &-u_0\frac{\partial c}{\partial X} - v_0\frac{\partial c}{\partial Y} - (au_0 + bv_0 c),\end{aligned} \tag{10.39}$$

where:

$$T^* = \int_0^t f_1(mt)dt, \tag{10.40}$$

is another time variable. The input conditions (Eqs. 10.7a and 10.7b) become:

$$-D_{x0}\frac{f_2(mt)}{f_1(mt)}\frac{\partial c}{\partial X} + u_0 c = \begin{cases} u_0 C_0, & 0 < T^* \le T_0^* \\ 0, & T^* > T_0^* \end{cases}, \quad X = 0, Y = 0, \tag{10.41a}$$

$$-D_{y0}\frac{f_2(mt)}{f_1(mt)}\frac{\partial c}{\partial Y} + v_0 c = \begin{cases} v_0 C_0, & 0 < T^* \le T_0^* \\ 0, & T^* > T_0^* \end{cases}, \quad X = 0, Y = 0. \tag{10.41b}$$

Using the transformation (Eqs. 10.17 and 10.23), the ADE and the input condition become

$$\frac{\partial C}{\partial T^*} = D\frac{f_2(mt)}{f_1(mt)}\frac{\partial^2 C}{\partial Z^2} - Uf_3(mt)\frac{\partial C}{\partial Z}, \tag{10.42}$$

$$-D^*\frac{f_2(mt)}{f_1(mt)}\frac{\partial C}{\partial Z} + U^* C = \begin{cases} U^* C_0 \exp\left(UT^*\right), & 0 < T^* \le T_0^* \\ 0, & T^* > T_0^* \end{cases}, \quad Z = 0, \tag{10.43}$$

respectively, where:

$$f_3(mt) = 1 - \lambda \frac{f_2(mt)}{f_1(mt)}, \tag{10.44}$$

$\lambda = (D/U)$, $D = a^2 D_{x0} + b^2 D_{y0}$, $U = au_0 + bv_0$, $D^* = aD_{x0} + bD_{y0}$, and $U^* = u_0 + v_0$.

Further, using the new space and time variables introduced through the respective transformations:

$$\eta = Z \frac{f_1(mt)}{f_2(mt)} f_3(mt), \tag{10.45}$$

$$T = \int_0^t \frac{f_1^2(mt) f_3^2(mt)}{f_2(mt)} dt, \tag{10.46}$$

the ADE (Eq. 10.42) becomes:

$$\frac{\partial C}{\partial T} = D \frac{\partial^2 C}{\partial \eta^2} - U \frac{\partial C}{\partial \eta} \tag{10.47}$$

It is the same as Eq. 10.26, but the new independent variables η and T have different expressions (Eqs. 10.45 and 10.46, respectively). Similarly, the initial and the second boundary conditions may be obtained as

$$C(\eta, T) = 0, \quad T = 0, Z \geq 0, \tag{10.48}$$

$$\text{and } \frac{\partial C}{\partial \eta} = 0, \quad \eta \to \infty, \tag{10.49}$$

respectively. To write the input condition (Eq. 10.43) in the new time variable T, it has to be explicitly related with T^*. For this, we consider:

$$f_1(mt) = \exp(mt), f_2(mt) = \exp(-mt), \tag{10.50a}$$

$$\text{and} \quad f_1(mt) = \exp(-mt), f_2(mt) = \exp(mt) \tag{10.50b}$$

The first combination represents accelerating flow and decelerating dispersivity, and the second one represents decelerating flow and accelerating dispersivity. For the combination in Eq. 10.50a and using the transformation in Eq. 10.40, we have:

$$T^* = \frac{1}{m}\left[\exp(mt) - 1\right] \quad \text{or} \quad \exp(-mt) = (1 + mT^*)^{-1}$$

Also, using Eq. 10.44 and the transformation (Eq. 10.46), we get:

$$T = \frac{1}{3m}(1+mT^*)^3 - \frac{\lambda^2}{m}(1+mT^*)^{-1} - \frac{2\lambda}{m}(1+mT^*) - \frac{1}{3m} + \frac{\lambda^2}{m} + \frac{2\lambda}{m}$$

The unsteady parameter m is considered much smaller than 1. Omitting its second- and higher-degree terms in the binomial expansions, we get the same relationship as in Eq. 10.29, i.e.:

$$T^* = \gamma T, \quad \text{where} \quad \gamma = (1-\lambda)^{-2}$$

It may be verified that this relationship also holds good for the second combination given by the Eq. 10.50b. Thus, using the combination (Eq. 10.50a) and the above relationship, input condition (Eq. 10.43) may be written as:

$$-D^*(1-\lambda+2\lambda m\gamma T)\frac{\partial C}{\partial \eta} + U^*C = \begin{cases} U^*C_0\exp(\gamma UT); & 0 < T \le T_0 \\ 0 & ; T > T_0 \end{cases}, \eta = 0. \quad (10.51a)$$

For the combination (Eq. 10.50.2), the input condition will be:

$$-D^*(1-\lambda-2\lambda m\gamma T)\frac{\partial C}{\partial \eta} + U^*C = \begin{cases} U^*C_0\exp(\gamma UT), & 0 < T \le T_0 \\ 0, & T > T_0 \end{cases}, \quad \eta = 0. \quad (10.51b)$$

Comparing the initial and boundary value problems in the (η, T) domain, in both the sections, it may be concluded that the solution of Eqs. 10.45–10.47 along with the condition (Eq. 10.51a) will also be given by the solution (Eq. 10.37a, b) with one change, $A_3 = -2m\lambda\gamma D^*\sqrt{D}$. Similarly, using the input condition (Eq. 10.51b), the same solution will be obtained with one change: $A_3 = 2m\lambda\gamma D^*\sqrt{D}$.

10.2.3 Particular Cases

1. For $f_1(mt) = f(mt)$ and $f_2(mt) = f_1^2(mt)$, the results of Sect. 2.2 reduce to the respective results of Sect. 2.1. In other words, we get the solution of a dispersion problem subject to the same conditions, satisfying the theory [18–19]
2. For $f_1(mt) = f(mt) = f_2(mt)$, we get the solution according to the theory in [19]
3. For $m = 0$, $f_1(mt) = f_2(mt) = 1$, we get the solution for a dispersion problem in which dispersivity and velocity both are steady. But, we should ensure that the heterogeneity parameters along both the directions are not equal, i.e., $a \neq b$. Otherwise, the denominator of B_2 in the expression of the Eq. 10.38 becomes zero
4. For $f_1(mt) = 1$ and $f_2(mt) = f(mt)$, we get a solution holding the theory [26]
5. For the last combination, $f_1(mt) = f(mt)$ and $f_2(mt) = 1$, we may get a solution

10.3 Illustration and Discussion

The solutions obtained in the Eqs. 10.37a, b, satisfying the conditions of both the Sects. 2.1 and 2.2, are illustrated in Figs. 10.1, 10.2, 10.3, 10.4. The concentration values are evaluated in a finite domain $0 \le x\,(\text{km}) \le 1$ and $0 \le y\,(\text{km}) \le 1$ along the longitudinal and transverse directions, for a set of input data: reference concentration $C_0 = 1.0$, initial velocity, and dispersivity components along both the directions as: $u_0 = 1.05\text{km/year}$, $v_0 = 0.15\text{km/year}$, $D_{x0} = 1.1\text{km}^2/\text{year}$, $D_{y0} = 0.1\text{km}^2/\text{year}$, unsteady parameter $m = 0.1\ \text{year}^{-1}$. Heterogeneity parameters are chosen as $a = b = 0.1\text{km}^{-1}$, except in Fig. 10.4 where both the parameters have different values (see the 3rd particular case in Sect. 2.3). The source of solute is supposed to be eliminated at time $t_0 = 2.0$ year. From the figures, it is evident that the input concentration, i.e., (c/C_0) at the origin $(x = 0, y = 0)$, described by the conditions in the Eqs. 10.7a and b, varies with time. In Figs. 10.1a and 10.2a, the input concentration increases with time in the presence of the source, i.e., in the time domain $0 < t < t_0$, and in Figs. 10.1b and 10.2b the input concentration decreases with time after the elimination of the source, i.e., in the time domain $t > t_0$. Fig. 10.1a and b illustrate the solutions given by the Eqs. 10.37a and 10.37b, respectively, when the dispersivity in each direction is proportional to the square of the respective velocity (Sect. 2.1).

In both the figures $f(mt) = \exp(-mt)$ is considered. Fig. 10.1a is drawn at $t\,\text{year} = 1.0$ and 1.5. The input concentration increases with time, its values at both the times are 0.46 and 0.56, respectively. Fig. 10.1b is drawn at $t\,\text{year} = 2.5$ and 3.5. The input concentration decreases with time, its values at both the times are 0.39 and 0.24, respectively. The concentration values at a position (*x, y*) increases with time, in the presence of the source and decreases with time in the absence of the source. Fig. 10.2a and b are drawn by evaluating concentration values from the solutions of the Eqs. 10.37a and 10.37b, subject to the provisions of subsection 2.2. The combination given by the Eq. 10.50b, i.e., that of decelerating flow and accelerating dispersivity is considered.

The concentration distribution pattern is as expected. Fig 10.3a and b compare the two-dimensional solute transport for both the combinations given by the Eqs. 10.50a and 10.50b in the presence and after the removal of the source, respectively. The former figure is drawn at $t\ (\text{year}) = 1.0$ and the latter figure is drawn at $t\ (\text{year}) = 2.5$. It may be observed that in the presence of the source, the input concentration for the combination in the Eq. 10.50b is lower (0.43) than (0.50) for the combination given by the Eq. 10.50a. But the trend reverses in the absence of the source, though the difference is not so distinct (the two are 0.375 and 0.371, respectively). The solute transport in the presence of the source is slower but the rehabilitation process is faster for the combination in the Eq. 10.50a than that for the combination in the Eq. 10.50b. Though after the removal of the source a peak near the origin is formed in the case of the combination in the Eq. 10.50a, but it attenuates very fast. Fig. 10.4a and b compare the solute transport in the case of steady dispersivity along steady flow with that in the case of decelerating dispersivity along accelerating horizontal flow domain, in the presence of the source, at

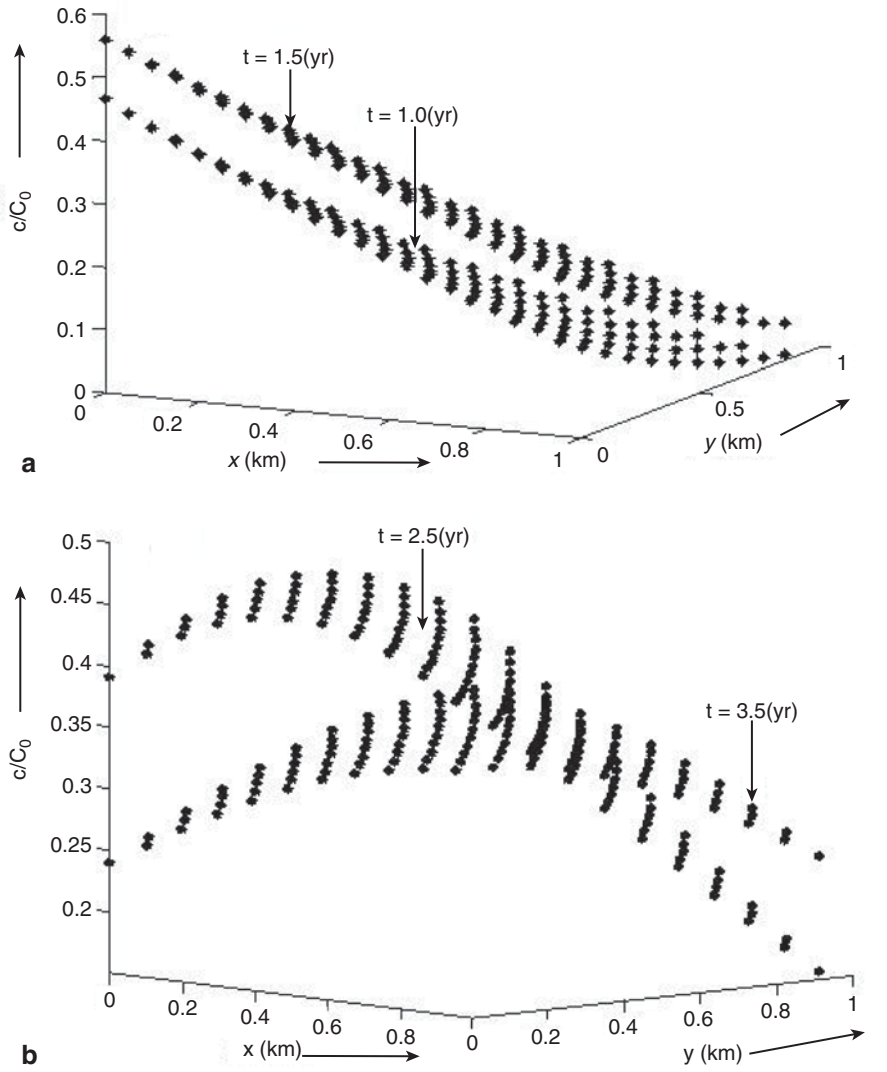

Fig. 10.1 Two-dimensional solute transport (Sect. 2.1) for $f(mt) = \exp(-mt)$, described by **a** Eq. 10.37a and **b** Eq. 10.37b, $t_0 = 2.0(\text{year})$

$t = 1.5$ (year), and after its removal at $t = 3.5$ (year), respectively. But, as stated in the previous paragraph of this section, the heterogeneity parameters are chosen as $a = 0.15$ and $b = 0.12\text{km}^{-1}$. In the presence of the source, the input concentration in the former case is 0.62 while that for the latter case is 0.63. After the source is removed, the two values are 0.157 and 0.143, respectively. It may be observed that the pollution rate in the domain is slower in the latter case but rehabilitation is faster.

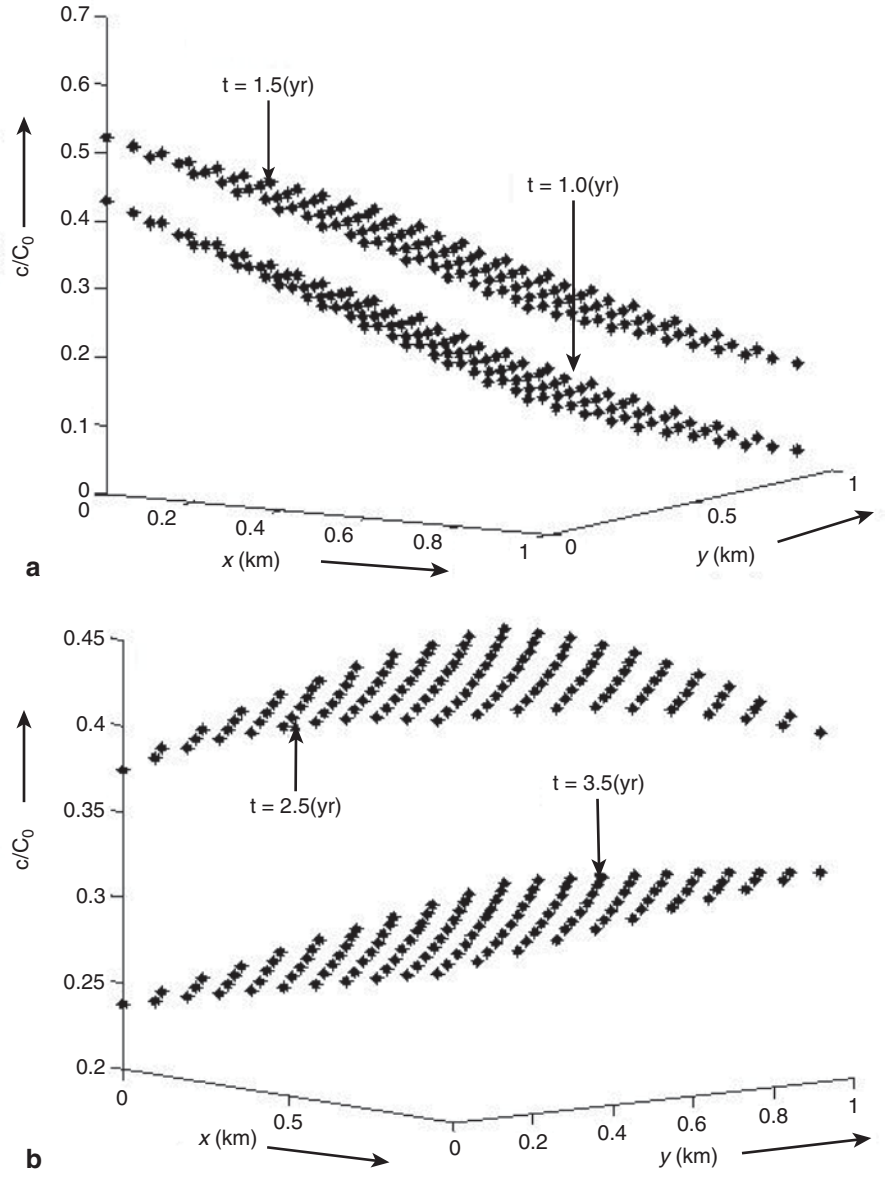

Fig.10.2 Two-dimensional solute transport (Sect. 2.2) for the combination (Eq. 10.50b), described by **a** Eq. 10.37a and **b** Eq. 10.37b, $t_0 = 2.0(\text{year})$

It is evident from these figures that the lateral solute transport is also significant and cannot be ignored in comparison with the longitudinal component, even though the lateral component input value is considered only one tenth of that of the longitudinal component. In other words, neglecting lateral components of velocity

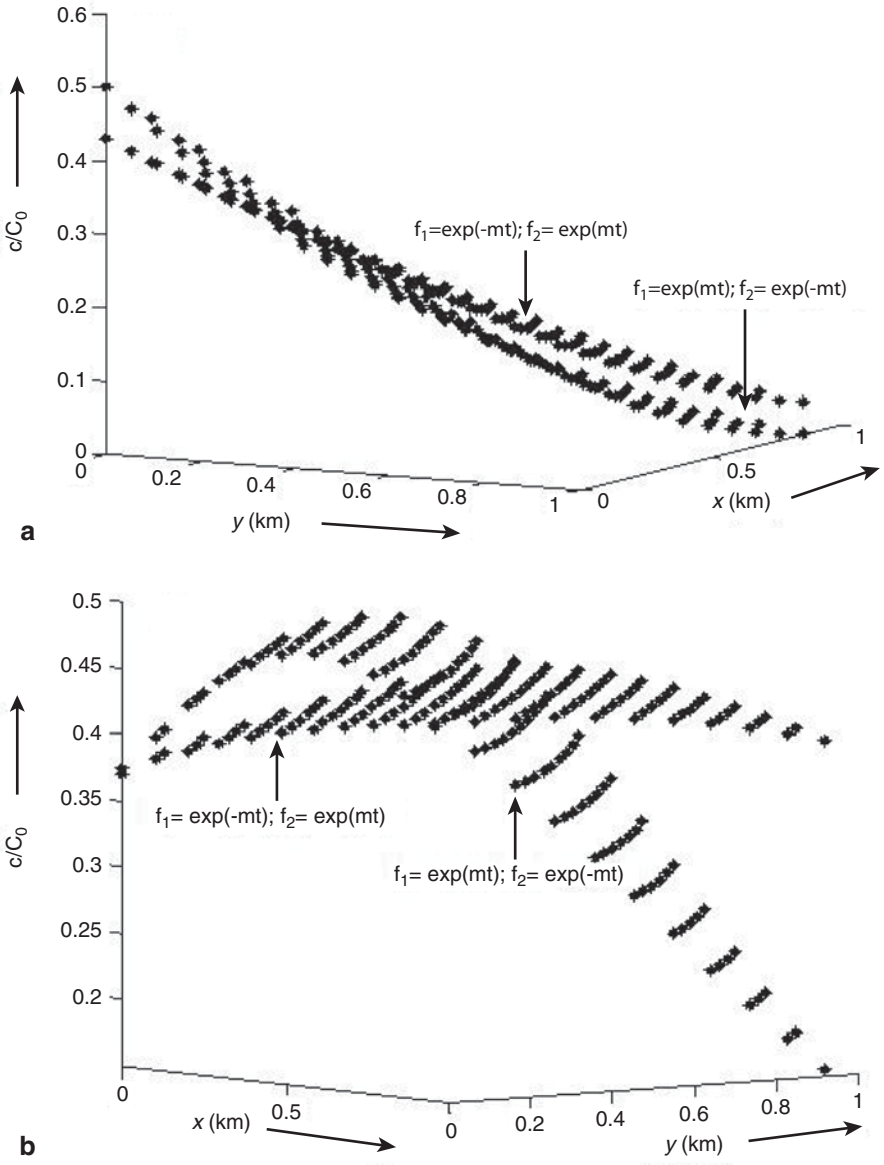

Fig. 10.3 Comparison of solute transports (Sect. 2.2) for the two combinations in Eqs. 10.50a and b, described by **a** Eq. 10.37a at $t = 1.0(\text{year})$, and **b** Eq. 10.37b at $t = 2.5(\text{year})$, $t_0 = 2.0(\text{year})$

and dispersivity, one is estimating virtually higher concentration at a longitudinal position than the actual one. A two-dimensional model is always better than one-dimensional model in estimating the time in which a domain gets polluted and its rehabilitation time, once the source of pollution is removed. One time dependent

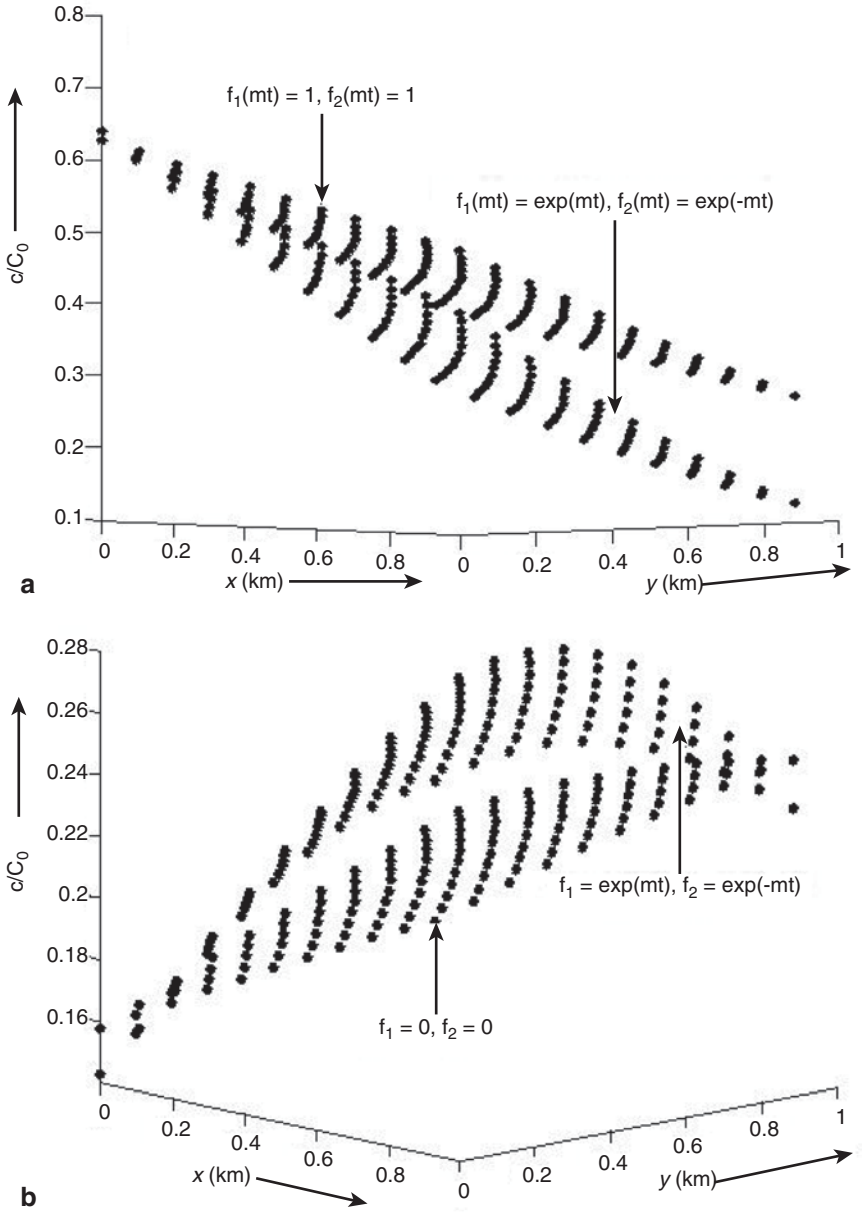

Fig. 10.4 Comparison of solute transports (Sect. 2.2) for decelerating dispersivity along accelerating flow, and steady dispersivity along steady flow, both described by **a** Eq. 10.37a at $t = 1.5$ (year), and **b** Eq. 10.37b at $t = 3.5$ (year), $t_0 = 2.0$ (year)

form $f(mt) = \exp(mt)$ tends to infinity as $t \to \infty$, hence it may be argued that ultimately the solute transport or velocity of the flow field increases enormously. But this situation will never occur because (as evident from the figures), the concentration gradient with respect to position decreases with time. So, after a certain time period an equilibrium stage is reached; the source becomes ineffective as far as its effect of polluting the domain is concerned. Similarly, once the source is removed, the rehabilitation process is almost completed in a finite time and its rate increases with time. In other words, the concentration value at a particular position does not change uniformly with time. For example, the concentration at the origin in the presence of the source at $t = 1.0, 1.5$, and 2.0 (years) are 0.46, 0.56, and 0.64, respectively. The same in the absence of the source at $t = 2.5, 3.5$, and 4.5 (years) are 0.39, 0.24, and 0.04, respectively. This represents a real scenario. This may be achieved through a third type of boundary condition at the origin, as considered in the Eqs. 10.7a and 10.7b. The unsteadiness and heterogeneity in different media varies. In air, the two may be on the higher side, while in soil and in aquifer they are on the lower side. The present study may be useful to account for the dispersion through any medium through selection of the appropriate values of the respective parameters.

10.4 Summary

ADE describes solute transport originating from a source. Dispersivity may not be uniform in a real situation. It depends upon convection through a medium, which is seldom homogeneous. So, heterogeneous nature of a medium affects the velocity as well as dispersivity and hence the solute transport. Apart from it, in a large time domain and long extent of the medium, the factors other than heterogeneity may affect both the parameters. To accommodate these factors, one way is to assume velocity and dispersivity temporally dependent [43, 47]. With all such factors, a two-dimensional solute transport from a varying pulse-type point source has been studied along exponentially time-dependent flow through a heterogeneous, initially solute-free, semi-infinite medium. The heterogeneity is described by spatially dependent advective and dispersivity parameters. The dependence considered increases linearly. By changing the value of the heterogeneity parameter, media of different heterogeneities may be represented. Small value of the parameter ensures that effects on solute transport remains confined to a small range. The unsteadiness of dispersivity is considered as a square of the velocity, and inverse of the velocity, in two separate sections, respectively. Such real assumptions have made the coefficients of the ADE variable. These are reduced to constant coefficients, through formulation of suitable moving coordinate transformations. It helps the use of Laplace integral transformation technique in obtaining the analytical solution. This technique has the least error of approximation, compared to most of the other such analytical methods. To keep the error further low, the two-dimensional ADE is reduced to one-dimensional equation, by using another coordinate transforma-

tion. The effect of lateral solute transport has been found significant in the context of a varying source of the solute mass too. To establish this fact, the values of the lateral components of velocity and dispersivity parameters are considered only one tenth of the respective longitudinal components. The point source considered in this model is stationary. A moving source or moving input boundary condition of varying pulse type may be an interesting extension to assess the pollution level of the air due to heavy vehicular transports.

Acknowledgments The first author, Premlata Singh gratefully acknowledges the financial support in the form of Dr. D. S. Kothari postdoctoral fellowship awarded by the University Grants Commission, Government of India during 2010–2011.

References

1. van Genuchten MTh, Alves W J (1982) Analytical solutions of the one-dimensional convective-dispersive solute transport equation. USDA ARS Technical Bulletin Number 1661, U.S. Salinity Laboratory
2. Javandel I, Doughty C, Tsang CF (1984) Groundwater transport: handbook of mathematical models, American Geophysical Union Water Resources Monograph Series 10. American Geophysical Union, Washington DC
3. Domenico PA, Schwartz FW (1998) Physical and chemical hydrogeology, 2nd edn. Wiley, New York
4. Jørgensen PR, Urup J, Helstrup T, Jensen MB, Eiland F, Vinther FP (2004) Transport and reduction of nitrate in clayey till underneath forest and arable land. J Contam Hydrol 73:207–226
5. Stutter MI, Deeks LK, Billett MF (2005) Transport of conservative and reactive tracers through a naturally structured upland podzol field lysimeter. J Hydrol 300:1–19
6. Wriedt G, Rode M (2006) Modeling nitrate transport and turnover in a lowland catchment system. J Hydrol 328:157–176
7. Ojo TO, Bonner JS, Page CA (2007) Similation of constituent transport using a reduced 3-D constituent transport model (CTM) driven by HF Radar: modeling, application and error analysis. Environ Model Soft 22:488–501
8. Qian Q, Voller V, Stefan HG (2007) Modeling of solute transport into sub-aqueous sediments. Appl Math Model 31:1461–1478
9. Taylor GI (1953) Dispersion of soluble matter in solvent flowing slowly through a tube. Proc R Soc Lond A 219:186–203
10. Scheidegger A (1957) The physics of flow through porous media. University of Toronto Press, and Oxford University Press, London
11. Rumer RR (1962) Longitudinal dispersion in steady and unsteady flow. J Hydraul Div 88:147–173
12. Coats KH, Smith BD (1964) Dead-end pore volume and dispersion in porous media. Soc Petrol Eng J 4:73–84
13. Shamir UY, Harleman DRF (1967) Dispersion in layered porous media. J Hydraul Div 95:237–260
14. Lin SH (1977) Nonlinear adsorption in porous media with variable porosity. J Hydrol 35:235–243
15. Valocchi AJ (1989) Spatial moment analysis of the transport of kinetically adsorbing solute through stratified aquifers. Water Resour Res 25:273–279
16. Matheron G, de Marsily G (1980) Is transport in porous media always diffusive? A counter example. Water Resour Res 16:901–917

17. Dagan G, Neuman SP (1977) Subsurface flow and transport: a stochastic approach. Cambridge University Press, Cambridge
18. Yates SR (1992) An analytical solution for one dimensional transport in porous media with an exponential dispersion function. Water Resour Res 28:2149–2154
19. Logan JD, Zlotnik V (1995) The convection-diffusion equation with periodic boundary conditions. Appl Math Lett 8(3):55–61
20. Huang K, van Genuchten MTh, Zhang R (1996) Exact solutions for one dimensional transport with asymptotic scale dependent dispersion. Appl Math Model 20:297–308
21. Neelz S (2006) Limitations of an analytical solution for advection-diffusion equation with variable coefficients. Commun Numer Methods Eng 22:387–396
22. Morton KW, Mayers DF (1994) Numerical solution of partial differential equations. Cambridge University Press, Cambridge
23. Mikhailov MD, Ozisik MN (1984) Unified analysis and solutions of heat and mass diffusion. Wiley, New York
24. Cotta RM (1993) Integral transforms in computational heat and fluid flow. CRC Press, Boca Raton
25. Moreira DM, Vilhena MT, Buske D, Tirabassi T (2009) The state-of-art of GITT method to simulate pollutant dispersion in the atmosphere. Atmos Res 92:1–17
26. Cassol M, Wortmann S, Rizza U (2009) Analytical modeling of two-dimensional transient atmospheric pollutant dispersion by double GITT and laplace transformation techniques. Environ Mod Soft 24:144–151
27. Guerrero JSP, Skaggs TH (2010) Analytic solutions for one dimensional advection-dispersion transport equation with space-dependent coefficients. J Hydrol 390:57–65
28. Guerrero JSP, Pimentel LCG, Skaggs TH, van Genuchten MTh (2009) Analytical solution of the advection-diffusion transport equation using a change-of-variable and integral transform technique. Int J Heat Mass Transfer 52:3297–3304
29. Aris R (1956) On the dispersion of a solute in fluid flowing through a tube. Proc R Soc Lond A 235:67–77
30. Wang ST, McMillan AF, Chen BH (1978) Dispersion of pollutants in channels with nonuniform velocity distribution. Water Resour Res 12:389–394
31. Yeh GT, Tsai YJ (1976) Dispersion of water pollutants in a turbulent shear flow. Water Resour Res 12(6):1265–1270
32. Basha HA (1997) Analytical model of two-dimensional dispersion in laterally nonuniform axial velocity. J Hydraul Eng 123(10):853–862
33. Chen JS, Liu CW, Liao CM (2003) Two-dimensional laplace-transformed power series solution for solute transport in a radially convergent flow field. Adv Water Resour 26:1113–1124
34. Nokes RI, Wood IR (1988) Vertical and lateral turbulent dispersion: some experimental results. J Fluid Mech 187:373–394
35. Serrano SE (1996) Hydrologic theory of dispersion in heterogeneous aquifers. ASCE J Hydrol Eng 1:144–151
36. Hunt B (1998) Contaminant source solutions with scale dependent dispersivities. ASCE J Hydrol Eng 3:268–275
37. Hantush MM, Marino MA (1996) Inter layer diffusive transfer and transport of contaminants in stratified formations. II: analytical solutions. J Hydrol Eng 3:241–247
38. Lotinopoulos P, Tolikas D, Mylopoulos Y (1998) Analytical solutions for two-dimensional chemical transport in aquifers. J Hydrol 98:11–19
39. Zhao C, Valliappan S (1994a) Transient infinite element for contaminant transport problems. Int J Numer Methods Eng 137:113–115
40. Kue-Yong Kim, Kim T, Kim Y, Nam-Chil W (2007) A semi-analytical solution for groundwater responses to stream-stage variations and tidal fluctuations in coastal aquifers. Hydrol Process 21:665–674
41. Kachiashvili K, Gordeziani D, Lazarov R, Melikdzhanian D (2007) Modeling and simulation of pollutants transport in rivers. Appl Math Model 31:1371–1396
42. Kumar N (1983) Unsteady flow against dispersion in finite porous media. J Hydrol 63:345–358

43. Yadav RR, Vinda RR, Kumar N (1990) One-dimensional dispersion in unsteady flow in an adsorbing porous medium: an analytical solution. Hydrol Process 4:189–196
44. Aral MM, Liao B (1996) Analytical solution for two-dimensional transport equation with time dependent dispersion coefficients. J Hydrol Eng 1(1) 0–32
45. Zou S, Ma J, Koussis AD (1996) Analytical solutions to non-Fickian subsurface dispersion in uniform groundwater flow. J Hydrol 179:237–258
46. Zoppou C, Knight JH (1997) Analytical solution for advection-diffusion equations with spatially variable coefficients. J Hydraul Eng 123:144–148
47. Singh MK, Singh P, Singh VP (2010) Analytical solution for two dimensional solute transport in finite aquifer with time dependent source concentration. J Eng Mech 136(10):1309–1315
48. Kumar A, Jaiswal DK, Kumar N (2010) Analytical solutions to one—dimensional advection—diffusion equation with variable coefficients in semi—infinite media. J Hydrol 380:330–337
49. Jaiswal DK, Kumar A, Kumar N, Singh MK (2011) Solute transport along temporally and spatially dependent flows through horizontal semi-infinite media: dispersion being proportional to square of velocity. J Hydrol Eng 16(3):228–238
50. Su N, Sander GC, Liu F, Anh V, Barry DA (2005) Similarity solutions for solute transport in fractal porous media using a time- and scale dependent dispersivity. Appl Math Model 29:852–870
51. Freeze RA, Cherry JA (1979) Groundwater. Prentice-Hall, New Jersey
52. Crank J (1975) The mathematics of diffusion. Oxford University Press, UK
53. Carnahan CL, Remer JS (1984) Non-equilibrium and equilibrium sorption with a linear sorption isotherm during mass transport through porous medium: some analytical solutions. J Hydrol 73:227–258

Chapter 11
The Problem of Futile Cycles in Metabolic Flux Modeling: Flux Space Characterization and Practical Approaches to Its Solution

Wynand S. Verwoerd and Longfei Mao

11.1 Introduction

A metabolic flux model of an organism can be developed from the genome-scale metabolic network (GEM) for quantitatively understanding and simulating the phenotypes of metabolic systems. For the development of the flux model, the GEM serves as a framework that integrate all of the massive "omics" data derived from systems biology research, comprising (i) gene detection (genomics), (ii) gene expression (transcriptomics), (iii) protein expression and modifications (proteomics), (iv) primary and secondary metabolites production (metabolomics), (v) measurement and estimation reaction rates (fluxes) for a network of reactions that occur in an organism (fluxomics) [1], and (vi) large-scale literature mining (bibliomics) [1–3].

Due to the advances in the high-throughput "omics" technologies and computer capabilities, there has been an exponential increase in GEMs reconstructed for a wide variety of organisms since the first GEM was built in 1999 [4]. As the GEMs intend to include as large part of the cell metabolism as possible and as much biological information as possible [5], they provide a detailed representation of biological reaction networks and their functional states [6], and can be used as analysis platforms for computational systems approaches such as constraint-based modeling [3], to characterize the flux profiles of the microbial phenotypes. This type of analysis can generate new knowledge that facilitates metabolic engineering of interesting biotechnological processes at the whole-cell level and can overcome the difficulties experienced in reductionist investigative strategies. Therefore, using in silico modeling approaches to develop strains has been considered as a promising area in the field of systems biology.

W. S. Verwoerd (✉) · L. Mao
Faculty of Ag & Life Science, Centre for Advanced Computational Solutions (C-fACS), Lincoln University, Lincoln, New Zealand
e-mail: wynand.verwoerd@lincoln.ac.nz

L. Mao
e-mail: longfei.mao@lincoln.ac.nz

S.K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_11,

In this chapter, we illuminate the basic concepts and principles of the constraint-based flux analysis, discuss two issues commonly encountered in the analysis, namely, alternate optimal solutions (AOSs) and futile cycles, and provide some solution methods based on our perspective. We hope this chapter will guide the audience whom we envisaged as researcher relatively new to metabolic flux modeling to understand the essential features of this technique.

11.2 The Mathematical Expression of GEMs

The GEMs can be converted into a mathematical format as stoichiometric matrix that is fundamentally required by all in silico modeling approaches [7]. Stoichiometry is the calculation of quantitative relationships of the reactants and products in chemical reactions. Metabolic networks are represented by a stoichiometric matrix S with each column of the stoichiometric matrix corresponding to a chemical or transport reaction and rows corresponding to the metabolites [8, 9]. The elements in S are the stoichiometric coefficients of the associated reactions. Negative and positive coefficients indicate the directionality of the reaction, i.e., the substrate metabolites have negative coefficients, whereas the product metabolites have positives ones. In real situations, the number of reactions (n) is usually larger than that of compounds (m), which indicates that there are more unknown variables than equations and thus the system of equations has more than one solution [10]. The stoichiometric matrix can be annotated by including other important information linked to either the reactions or the metabolites, such as the genome and gene expression data for use in certain applications [7]. The stoichiometric matrix based on a reconstructed metabolic network serves as a backbone for mathematical modeling approaches to perform predictive, hypothesis-driven in silico experiments [11]. These modeling approaches can be divided into three categories: (i) characterizing the general network structure through null-space analysis such as singular value decomposition (SVD), (ii) analyzing all possible flux distributions in a network through extreme pathway analysis (ExPa) and elementary mode analysis (EM), and (iii) constraint-based flux analysis, specifically flux balance analysis (FBA) that identifies the flux distribution of a particular network state [7]. However, constraint-based modeling is the only approach that is tractable for genome-scale description of metabolic process.

11.3 Flux Characterization of Cellular Phenotypes by Constraint-based Flux Modeling

The constraint-based modeling approach is used to predict possible cellular phenotypes by interrogating capabilities of the GEM through the imposition of physicochemical constraints. The constraints are formulated based on the consideration

of stoichiometry, thermodynamics, flux capacity, and regulatory restraints under which reactions operate in a metabolic network [12, 13]. These constraints are applied to reduce the range of attainable flux distributions or metabolic phenotypes achievable for an organism. As a result, a steady-state flux space is defined that contains all possible functional states of the network [14].

Since far fewer model parameters are needed for steady-state analysis of a metabolic network than for dynamic modeling, constraint-based analysis can be performed on a GEM. Nonetheless, this method generally offers no information about metabolite concentrations or about the temporal dynamics of the system [15–17]. Although the other quantitative approach, i.e., kinetic modeling, can characterize the detailed mechanisms of the metabolic reaction systems, obtaining a large number of kinetic parameters is not an easy task. In addition, many of the kinetic values cannot be trusted because they are acquired from in vitro rather than in vivo measurement [18]. Therefore, compared with kinetic modeling, constraint-based modeling is the more appropriate tool for in silico engineering of the complex biological systems. The cornerstone of the constraint-based modeling approach is FBA [19].

11.3.1 FBA

FBA relies on data-driven constraints and linear optimization theories [20]. The constraints used in FBA can be classified into three groups: physicochemical, topological, and environmental. Physicochemical constraints are physical laws such as the stoichiometry of the reactions and thermodynamics on reaction directions; topological constraints represent spatial restrictions on metabolites within cellular compartments; and environmental constraints include nutrient availability, pH, and temperature [20]. FBA can analyze metabolic networks to relate genotypes to phenotypes because all expressed phenotypes of a given biological system must satisfy basic physicochemical, topological, and environmental constraints that are imposed on the functions of all cells [21, 22].

Once constraints are set, to evaluate the performance of the biological system at various perturbations, FBA requires a physiologically relevant objective function. Objective functions can be in many forms such as physiologically meaningful objectives or design objectives for the interrogation or exploitation of a given system. Examples of common objective functions include maximizing biomass or cell growth, maximizing ATP production or maximizing the rate of synthesis of a particular product [23]. Because the physicochemical constraints are readily defined from the annotated genome sequence that identifies the enzymes present in an organism, and hence the biochemical reactions for which stoichiometries are mostly known, as well as measured enzymatic capacities, FBA needs a minimal amount of experimental data.

FBA was originally specifically designed to quantitatively compute growth phenotypes [8, 24], by maximizing the biomass reaction flux (representing the growth rate) given a set of bounded intake rates for external substrates [25]. In FBA, the

biomass equation is usually set as a linear combination of a range of macromolecular components, including proteins, DNA, RNA, lipids, lipoteichoid acids, peptidoglycan, and polysaccharides, which accounts for all known biomass constituents and their fractional contributions to the overall cellular biomass [26, 27]. This equation is used to reflect the observation that cell growth requires synthesis of a range of metabolites. The individual composition of every precursor metabolite in the biomass equation is maintained at a fixed stoichiometry, which is determined by examining the relevant literature or adapting known biomass compositions of related organisms [25]. The stoichiometric combination makes the equation consume each biomass precursor metabolite in proportion to their ratio in the biomass composition [8] and thus independent of the specific growth rate [26]. The detailed method to formulate the biomass reaction for genome-scale network reconstruction can be found in [27].

While constraints represent fixed physical laws, the choice of an objective function is more ad hoc and generally represents biological reality. Taking biomass maximization as the FBA objective function reflects the assumption that microbes evolve to optimize their metabolisms for maximizing biomass production (growth rate) [24, 28–30]. This assumption has been confirmed by experiments in many cases [29, 31, 32]. It was also found that even when the metabolic network did not initially operate according to the optimal growth principle, under selection pressure *Escherichia coli* growing on glycerol evolved to eventually maximize their growth rate [31]. Besides, flux balance models have been successful for different specific metabolic engineering applications, such as production of lycopene and vanillin [33–36].

A previous study used a central metabolic model of *E. coli* to evaluate the prediction accuracy of FBA with different objectives and the results showed that maximization of adenosine tri-phosphate (ATP) or biomass yield per unit flux could achieve higher accuracy under specific conditions, such as unlimited growth on glucose under aerobic conditions, than maximization of biomass production ("growth rate") [37]. However, this indication is questioned by a later study based on a genome-scale model of *E. coli* [38]. The study found that there was very little difference between maximizing biomass production rate and biomass yield per unit flux for predicting gene expression changes seen after adaptive evolution and both of these two objectives achieved high prediction accuracy. On the other hand, maximizing ATP yield led to poor prediction involving few genes. It was thought that maximization of ATP selects against the usage of biosynthetic pathways since the end products are not included in this objective [38].

The biomass objective function is now built into almost every FBA-based metabolic network simulation of a microorganism. It can be regarded as a short cut, to circumvent the need for including the regulatory and signaling networks. Conversely, if metabolic, regulatory, and signaling networks are all reconstructed, an extrinsic objective will not be required for a growth-phenotype simulation [19].

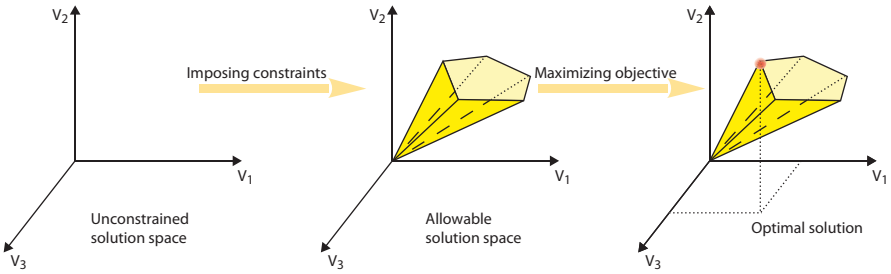

Fig. 11.1 The conceptual basis of FBA. (Visualization inspired by [10, 39])

11.3.2 Mathematical Basis of FBA

FBA is an implementation of constraints-based analysis. The static metabolic state of a cell is represented by the vector of all flux values, each of which is associated with a particular reaction. Without any constraints, the fluxes can take on any real value. After addition of constraints such as stoichiometric, thermodynamic, and enzyme capability rules, the total flux solution space shrinks into a bounded polyhedral cone, termed the allowable solution space. Any point outside of the polyhedral cone violates one or more of the applied constraints and are biologically infeasible. Linear optimization that can maximize or minimize a defined objective such as biomass production can further reduce the steady-state solution space, possibly to a single solution. This circumvents the need for complete knowledge of the biochemical reaction network.

Mathematically, FBA involves computing a basis of the underlying polyhedral cone and identifies a single optimal flux distribution point within the cone using linear optimization. Linear optimization underlying FBA modeling comprises four major stages (Fig. 11.1): (i) Metabolite balancing and stoichiometry under steady-state assumption (connectivity and mass conservation) is mathematically expressed as a subspace of universal flux space. (ii) Irreversibility of biochemical fluxes limits all fluxes to positive values, yielding a semi-infinite convex cone. (iii) Minimum and maximum fluxes (capacity constraints) further reduce this to a confined convex space. (iv) A linear objective represents a hyperplane that intersects the convex cone, and maximizing or minimizing amounts to adjusting this hyperplane to touch the polyhedral convex cone. If it touches the cone at a single vertex, this is the optimal point that represents a state of the flux distribution that achieves the optimal objective function value consistent with all constraints. More generally, it may touch the cone along a line giving multiple optimal points.

A three-dimensional flux space based on a hypothetical metabolic network is depicted in Fig. 11.1. Progressive application of constraints reduces the feasible solution space to a convex cone, and optimizing the objective further may reduce this to a single optimal point.

Linear Programming Formulation of FBA

In FBA, all reactions in which a metabolite participates are considered, which can be represented by a mass balance equation:

$$\frac{dx_i}{dt} = \sum_j s_{ij} v_j . \tag{11.1}$$

Here S_{ij} is the stoichiometric coefficient of metabolite x_i for flux v_j. The equation neglects metabolite dilution because this effect on concentration is negligible compared to the fluxes for each reaction in which the metabolite participates [40].

For FBA, the metabolic reactions of the metabolic model are converted into a stoichiometric matrix. The system is assumed in steady state, i.e., the concentrations x of internal metabolites are constant [41]. The maximum and minimum ranges of the flux for each reaction are defined by the capacity constraints. These constraints provide a feasible space of the flux distribution in the metabolic model. FBA defines these constraints as a system of linear equations, which are solved by application of a linear programming (LP) technique to obtain a flux distribution in the feasible space that maximizes or minimizes an appropriate objective function. This LP problem is represented by the following equation:

$$\begin{aligned} &\text{maximize: } c^T \bullet v \\ &\text{subject to: } S \bullet v = 0 \\ &v_{\min} \le v \le v_{\max} . \end{aligned} \tag{11.2}$$

Here c is a vector containing the flux coefficients of the linear objective function. In general, a published network model always includes an experimentally formulated biomass reaction as the objective function. The flux through this reaction is the growth rate (h^{-1}) predicted by modeling. S is the $m \times n$ stoichiometric matrix of all the reactions in the metabolic network; m is the number of metabolites, n is the number of fluxes (reaction rates), and v denotes a vector of flux of each metabolic reaction.

$v_{\min}$ and $v_{\max}$ represent the minimum and maximum flux of each reaction based on the maximal enzymatic reaction rate, any constant uptake rate from the environment and possibly the irreversibility of the reaction. Also, thermodynamic constraints on reactions may exist to restrict the directional flow of the reaction. These constraints are obtained from the literature where the models are retrieved.

For a cellular metabolism, the number of metabolites defines the number of balance equations in Eq. 11.2 while the number of reactions represents the number of unknowns in Eq. 11.2. If the number of reactions n is larger than the number of metabolites m, the system has a mutual degree of freedom $F \ge n - m$. Hence, if a system has a number of measured fluxes less than F the system becomes underdetermined

by nature. In general, the metabolic network of an organism is an underdetermined system.

Network reactions include those by which the cell exchanges metabolites (nutrients and waste products) with its environment. A given metabolic environment usually dictates upper bounds on such exchange reaction fluxes. The fluxes (or reaction rates) are expressed in mmol per gram dry weight and hour (mmol/(gDW/h)). Under physiological conditions, each reaction will have a limited maximal flux due to thermodynamic constraints. $v_{\min}$ is limited to zero for irreversible reactions. The exact in vivo limits are usually not known, but it is biologically meaningful to restrict the reaction rates to a high value (for example 10,000 mmol/gDW/h) that act as additional constraints in the modeling. These high values are also known as explicitly encoded infinity constraints, i.e., bounds on reactions represented by a large number.

$$0 \le v \le 10000$$
$$-10000 \le v \le 10000.$$

To further constrain an underdetermined system that does not yield a unique flux distribution, it may be possible to use experimentally measured exchange fluxes as fixed values or a proper range of values.

The static nature of FBA makes the model structure for this modeling framework significantly different from others based on dynamical modeling systems of ordinary differential equations (ODE). Additional introductory expositions of the formulation of FBA can be found in the recent three chapters [10, 42, 43].

11.3.3 Comparison of FBA and ^{13}C-Based Metabolic Flux Analysis (^{13}C-MFA)

^{13}C-MFA is another fluxomics tool for experimental determination of reaction fluxes in an organism. During the measurement, a ^{13}C isotopic labelled carbon source is used to culture microbes and then intracellular fluxes can be quantitatively determined by tracing the transition path of the labelled atoms between metabolites in a biochemical network. Both FBA and ^{13}C-MFA require the use of a metabolic network and the assumption of a steady state for internal metabolites disregarding dynamic intracellular behavior. However, FBA characterizes the "optimal" metabolism for the desired functional metabolic output (phenotype) and can be implemented on GEM, whereas ^{13}C-MFA profiles in vivo metabolic flux distribution in a metabolic network and current technique only allows it to work on a small-sized central metabolic network [44]. Nevertheless, as the ^{13}C-MFA approach determines enzymatic rates at a specific growth condition experimentally, its resultant flux values are more precise than the prediction results of FBA. The metabolic fluxes experimentally measured by ^{13}C-MFA can be complementary to the metabolism predicted by FBA for identifying competitive pathways or toxic by-products, and

thus helps reducing gene targets underlying the enzymatic hurdle for improving desired product yield [45].

11.3.4 Variants of FBA

FBA has been considered as a general guideline and a viable first step for metabolic engineering of microorganisms for improving biosynthetic yield [46]. Nonetheless, the in silico prediction results of FBA could be different from in vivo observation. In an attempt to resolve this inconsistency, a number of optimization algorithms and computational strategies have been proposed to constrain the solution space of FBA, i.e., energy balance analysis in conjunction with FBA can take into account the thermodynamic principle to improve theoretical prediction [46, 47]. However, energy balance analysis can only be conducted on small-sized network and requires biochemical thermodynamics parameters of all reactions. On the other hand, to better describe metabolic behaviors divergent from optimal prediction, FBA can be performed under a bi-level optimization framework to estimate the potential trade-off between the maximizations of biomass production rate and the other desired product yield [48]. FBA can also use other objective functions such as minimization of metabolic adjustment (MOMA) and regulatory on/off minimization (ROOM). MOMA has been successfully used to engineer strains with increased production of many products such as lycopene [49, 50], valine [51], threonine [52], and polylactic acid [53]. Besides, transcriptional regulation is incorporated into several other variants of FBAs, such as regulatory FBA (rFBA) [54], steady-state rFBA (SR-FBA) [55], and probabilistic regulation of metabolism (PROM). These variant approaches center on predicting the immediate behavior of knockout strains, which is different from ordinary FBA that is used to predict cellular behaviors after strains have undergone adaptive evolution [30].

Furthermore, FBA can also be combined with metabolic pathway analysis, such as EM and ExPA, which can characterize all metabolic routes in a metabolic network without a prior knowledge of reaction rates. The computation of EMs and ExPAs are restricted to metabolic networks of moderate size and connectivity, because the number of modes and the computation time rise exponentially with increasing network complexity [56]. For example, 71 million elementary flux modes (EFMs) were found in a medium size metabolic network of *Saccharomyces cerevisiae* (230 reactions and 218 metabolites) [57]. Computation of EMs for a central metabolism network of *E. coli* (106 reactions and 89 metabolites) results in about 26 millions [56]. According to our empirical observation, computation of EMs for the metabolic network of *Geobacter sulfurreducens* (709 reactions; performed with CellNetAnalyzer 9.9 on a i7 four cores CPU, 8 GB memory and solid state disk home computer) could not terminate after several days. The large number of EMs is biologically meaningless and could only be interpreted for the fact that the network contains a large quantity of parallel (redundant) pathways.

11.4 Exploration of AOS and Futile Loops

The identified solutions by FBA are generally not unique and an infinite number of different flux vectors may exist, producing an identical optimal objective value. In the context of metabolic models, these flux vectors are called AOS or equivalent phenotypic states [58–61], and constitute an optimal solution space. Previous methods for the determination of AOS include: (a) vertex enumeration [60] such as EMs or ExPAs, and (b) flux variability analysis (FVA) [62]. As mentioned before, the drawback of the computation of EMs and ExPAs is that these two methods are usually not suitable for network models beyond a medium size (up to 100 reactions). To explore the whole optimal solution space of the genome-scale network, it is required to use FVA, which can calculate the feasible range of flux values for each reaction [59]. FVA is an extension of FBA that calculates a full range of possible numerical values for each reaction flux in a network while achieving a particular objective function value. This can help clarify the entire range of achievable cellular functions and examine the redundancy in a metabolic network. However, ordinary FVA results usually contain large futile values due to the presence of loops in the network.

11.4.1 Linear Programming Formulation of FVA

The FVA method tests for the range (min, max) of certain fluxes that fit the optimal solution. FVA can be formulated as:

$$\begin{aligned}
&\text{minimize or maximize } v_i \text{ for all } i \in \{1,\ldots,n\}:\\
&\text{subject to } Sv = 0\\
&Z = c^T \bullet v = Z_{\text{optimal}}\\
&v_{i,\min} \le v_i \le v_{i,\max}.
\end{aligned}$$

Here Z is a known optimal value for the biological objective; S is an $m \times q$ stoichiometric matrix which consists of m rows (i.e., metabolites) and q columns (i.e., reactions with the respective stoichiometric coefficients for the participating metabolites); v is the flux of each reaction in the model. FVA shares the same central assumption with FBA, i.e., the system is in steady state which means the concentrations of internal metabolites are constant. FVA results address the variability of v and thus can capture the effect of perturbations to the metabolic reactions in terms of the changes in flux output range.

Specifically, FVA calculates for each reaction i, the minimal possible flux and the maximal possible flux that reaction i can carry in any flux distribution that is consistent with the constraints, including the optimal Z value as a constraint. In other words, the information obtained from FVA is the physiologically feasible flux

range for the unknown reactions. Moreover, if the computed minimal and maximal rate of a reaction coincide, $v_{i,min} = v_i max$, the reaction rate follows to be uniquely determined.

In contrast to FBA, FVA as described above does not make any assumption about biological objectives [10]. The objective function $Z_{optimal}$ is only used as a tool to define a desired metabolic state (such as a redox perturbed state) for FVA to compute the feasible flux ranges. The relationship between FBA and FVA is illustrated in Fig. 11.2. This figure visualizes a three-dimensional flux space, and shows that the FBA solution space is a projection on the plane represented by a fixed (maximal) objective value. In the illustrated case it is a polygon, and the vertices are extreme points that satisfy the FBA constraints represented by the finite convex cone. Any point inside the polygon, is a flux combination that is both feasible and gives the same objective value, and can be mathematically constructed by a suitable combination of the edge vectors of the polygon. Hence, knowledge of all vertices would be enough to fully specify the FBA optimal solution space. Note that an FBA calculation only gives a single vertex; and since all vertices have the same objective value, it is essentially random which one. Minor changes to constraints, a different linear optimizer or even the ordering of variables can produce a different vertex. Hence, it is virtually impossible to get full specification of the optimality space from FBA especially for large networks.

In a FVA calculation, one flux at a time is allowed to vary while all others are kept fixed. For example, Fig. 11.2b shows the case where the feasible range of flux v_1 is explored while v_2 and v_3 are kept fixed. This corresponds to the determination of two flat parallel planes, perpendicular to the v_1 axis that touches the polygon at opposite vertices.

Repeating this for all fluxes, the result is a cubical bounding box containing the optimality polygon. So FVA yields only an approximation to the optimality space, as the intersection of the optimality plane and the bounding box; the actual polygon is contained within this, but it could also contain points that are outside the polygon and hence infeasible. Nevertheless, this approximate description is a major improvement on simple FBA and may suffice for practical use.

The situation illustrated by Fig. 11.2, where the optimality space is a finite polygon, is simpler than most real network models where the convex cone and hence the optimality subspace is unbounded in some directions. This is more fully discussed in Sect. 4.5 below but a similar comparison between FBA and FVA solutions can be developed in that case.

11.4.2 Futile Loops

In constraint-based modeling, the constraints mainly used are mass balance and metabolite uptake rates. The thermodynamic limitations associated with the reactions are difficult to implement and thus commonly overlooked. Without thermodynamic constraints, nonphysical fluxes can be computed for some metabolic reactions if

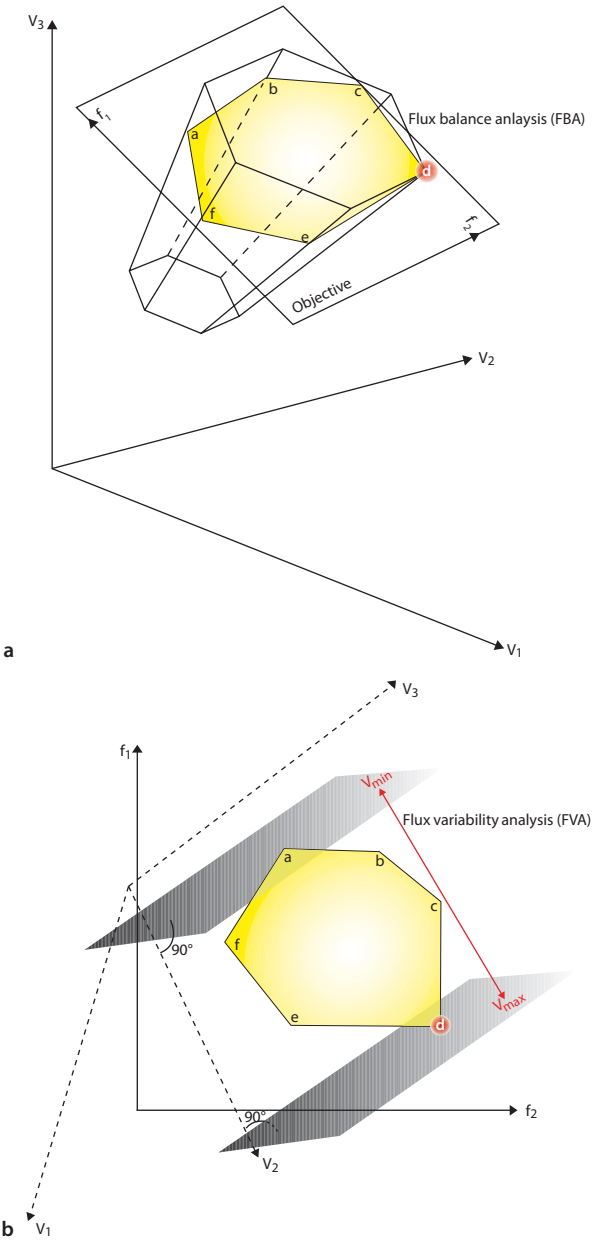

Fig. 11.2 Optimization principles underlying FBA and FVA (feasible space resulting from FBA and the *yellow polygon* representing the optimal solution space that contains points computed by maximizing the objective function (the objective plane), the coordinates are three arbitrary representative fluxes, a simplified version of the multidimensional flux space in (**a**), FVA to identify the possible range of each flux such that the objective value does not degenerate in (**b**))

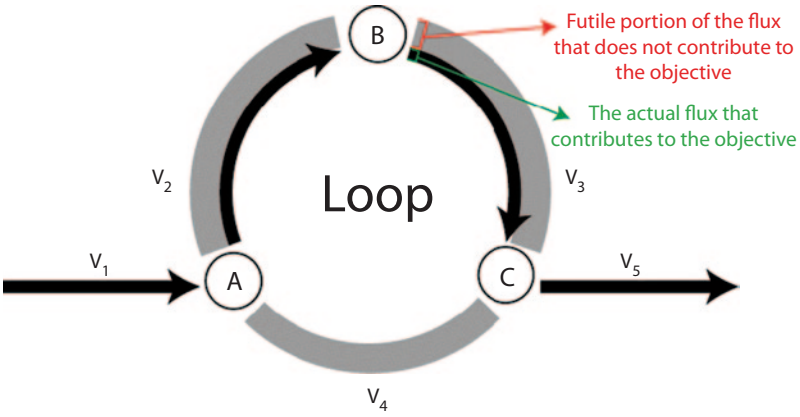

Fig. 11.3 An example set of reactions that form a loop cycling three metabolites internally (*capital letters* represent hypothetical metabolites, FBA and FVA can particularly not resolve reversible, circular, and parallel reactions)

they form an internal cycle (Fig. 11.3). Such a cycle of reactions violate a "loop law" that is analogous to Kirchhoff's second law for electrical circuits, as discussed previously by Beard et al [63].

Futile cycles that consume biochemical energy are not necessarily "futile" [64]. Although the exact metabolic function of futile cycles has not been fully elucidated, their activities have been suggested as mechanisms of biochemical regulation [64–66]. The amount of energy expended by (or the change of the flux through) a futile cycle determines the magnitude of changes in the effective equilibrium constants for biochemical reactions and the sensitivity and robustness of intermediate concentrations [64]. Also, the cellular ATP/ADP ratio could be dependent on the specific growth rate and environmental conditions [67]. For example, futile cycle activity in *Geobacter metallireducens* might be associated with the regulation of the substrate utilization for gluconeogenesis and the oxidative tricarboxylic acid cycle (TCA) cycle [68].

In practical FBA calculations, the presence of a futile loop is often signaled by implausibly high flux values for some reactions. It is a common practice to set an artificial upper limit (for example 10,000) on all fluxes, simply to avoid numerical problems in the optimization algorithm. These large values are sometimes referred to as explicitly encoded infinity constraints. Fluxes of loop reactions are often observed to reach values of this magnitude, but simple strategies such as blocking these reactions by constraining their fluxes to zero only work in the simplest cases as discussed further in Sect. 4.3 below.

One reason for this is that individual loop reactions may simultaneously also participate in non-futile fluxes carried by productive pathways. In this case, flux values along the loop may all be high, but with different values, because the observed fluxes result from superimposition of productive and futile fluxes. This has to be disentangled without affecting productive fluxes. This problem is particularly

common when loops involve a currency metabolite such as nicotinamide adenine dinucleotide (NADH).

Another reason is that as loop fluxes do not contribute to the objective value, it is conceivable that some such fluxes may have another value, even zero, in the particular flux distribution that is selected from the solution space by the optimization algorithm.

Elaborate strategies proposed to remove these thermodynamically infeasible futile cycles include known flux directionality [69], energy-balance equations [63], known [70–75] or predicted [76] thermodynamic parameters, and nonlinear constraints to eliminate flux distributions that utilize reactions which cannot be thermodynamically feasible under physiological conditions [77]. But these thermodynamic constraints are difficult to obtain and nonlinear constraints can make application to large-scale systems computationally challenging [71]. Therefore, the listed methods above cannot be applied on genome scale systems [71].

In addition, another algorithm called loopless-COBRA was recently proposed to eliminate loops [78]. This method imposes the second law of thermodynamics by using a mixed-integer linear programming (MILP) approach to constrain flux solutions so that they obey the loop law and does not require additional data such as metabolite concentrations or thermodynamic parameters. This method was mathematically proved not to over-constrain the problem beyond the elimination of the loops themselves [79]. However, such a method was unable to be conducted with the commonly used solver GNU Linear Programming Kit (GLPK) and necessitates a commercial solver (TOMLAB/CPLEX package (Tomlab Research, Pullman, WA)). In addition, the algorithm is computationally intensive and its running time is significantly increased with the network size.

Another previous study indirectly constrains the futile loops by coupling reaction flux to enzyme synthesis costs, but this approach necessitates comprehensive knowledge about the transcriptional and translational machinery [61].

Alternatively, loops can be removed by minimizing network flux [38, 70, 80, 81]. The application of flux minimization can produce a most likely stationary flux distribution. The notion behind this method is that the flux through any reaction of a metabolic network requires some “effort” and a metabolic network is inclined to fulfill the same biological objective, such as growth with minimum metabolic effort [70]. A drawback of this approach is that the use of the minimum sum of all fluxes in the network as an auxiliary criterion in FBA by previous methods, can forestall variability within optimal flux distribution.

We present a simple method that resolves the accidently activated futile values in the FBA results below. The simple method can remove the futile cycle without impacting simulations of the cellular phenotype such as biomass growth, but in some situations will unintentionally eliminate alternative optimal solutions. On the other hand, for concurrently determining AOS and efficiently eliminating the flux loop without side effects, we also devised another method termed FVA with target flux minimization (FATMIN) that combines the functions of FVA and FBA to determine AOS and eliminates flux loops in large metabolic networks. Importantly, FATMIN requires the least amount of information to quantify the fluxes and analyze the metabolic system.

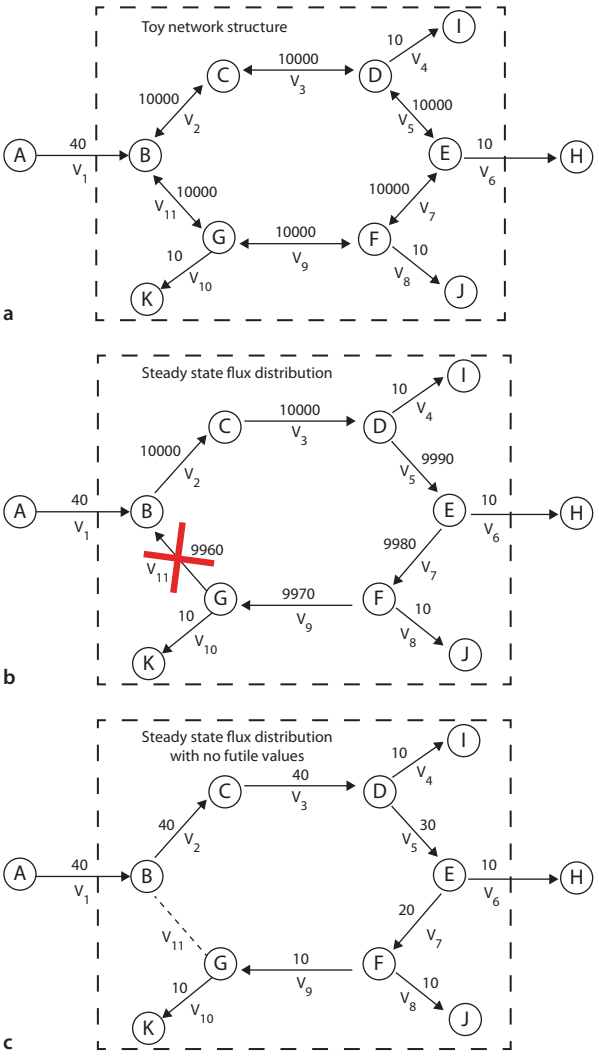

Fig. 11.4 A toy network with loop ((**a**) reactions v_2, v_3, v_5, v_7, v_9, and v_{11} form a loop, reactions v_4, v_8, v_{10} are constraints, v_1 is the exchange reaction providing source for the network, v_6 is the reaction that should be maximized (**b**) FBA returns a solution containing a loop and an objective value of 10, elimination of the less meaningful flux values in this case, v_{11} is constrained to zero as it has the lowest absolute flux value (**c**) eliminating reaction v_{11}, biologically meaningful flux values for v_2, v_3, v_5, v_7, and v_9 obtained without degenerating the objective value of 10)

11.4.3 A Simple Method to Remove Futile Reactions Encountered in FBA Modeling

An example of a network fragment that contains a superposition of productive fluxes and a futile loop is shown in Fig. 11.4.

A simple method to remove those loops without affecting productive fluxes, is to constrain the flux with the smallest absolute value to zero in order to break the loop. This still allows other fluxes in the loop to attain the values they contribute to productive (non-loop) pathways.

In the toy network of Fig. 11.4, fluxes towards the nodes I, J, and K are considered to be constrained while the yield of product H is to be maximized. In the optimal solution, all flux contributions necessary to supply the constraints and yield are taken as productive. However, the solution illustrated in Fig. 11.4b in addition contains a circulating flux. Because it is the sum of circulating and productive fluxes for a reaction that is limited by the explicitly encoded infinite capacity constraints as shown in Fig. 11.4a, it seems plausible to expect the smallest absolute flux value to be in the link along the loop that carries only the circular flux component. Constraining this flux (v_{11}) to zero, indeed produces a feasible steady state flux distribution (v_1,v_2,…v_{11})=(40, 40, 40, 10, 30, 10, 20, 10, 10, 10, 0) without loops as shown in Fig. 11.4c.

However, this is not the only compatible distribution; for example, the flux vectors (40, 30, 30, 10, 20, 10, 10, 10, 0, 10, −10) and (40, 0, 0, 10, −10, 10, −20, 10, −30, 10, −40) satisfy the same conditions as would suitably normalized combinations of these vectors. These alternative distributions only occur because by assumption, all loop reactions in the toy network are reversible. In real networks, that is often not the case and so the simple method is adequate to eliminate futile loops. Even if this does not apply, depending on the purpose of the modeling it may only be required to produce one distribution free of futile loops, regardless of whether others also exist.

The simple method also has the drawback that it may overlook some hidden futile cycles, as sometimes those futile cycle reactions could have flux rates in a particular FBA solution within a reasonable range (for example, the absolute flux value is less than 100 mmol/gDW/h) and appears to be “normal.” For these reasons, we have developed a more elaborate method to deal with futile loops as described in the next subsection.

11.4.4 The FATMIN Algorithm

The basis of FATMIN is FVA, which is used to probe the feasible flux ranges of desired reactions, such as reactions that involve NADH. FATMIN constrains the pointlessly high flux values of the loop reactions to biologically meaningful values. It is intended to deal with such loops involving one or more *target metabolites* that are of specific interest.

All reactions that *produce* (not consume) the target metabolite, and for which the FVA range extends higher than a threshold value (that is chosen to be higher than any realistic flux value) are identified as *target reactions*. Note that, because FVA probes the full range of fluxes compatible with the constraints, the problem of hidden cycles that can occur in a single FBA solution is eliminated.

To identify which portion of the high value is compulsory to achieve the value of the FBA objective function, we add an artificial metabolite to each target reaction, and add another reaction (for example, reaction ID: FluxMin) that only comprises the artificial metabolite. FluxMin acts as a drain reaction for the artificial metabo-

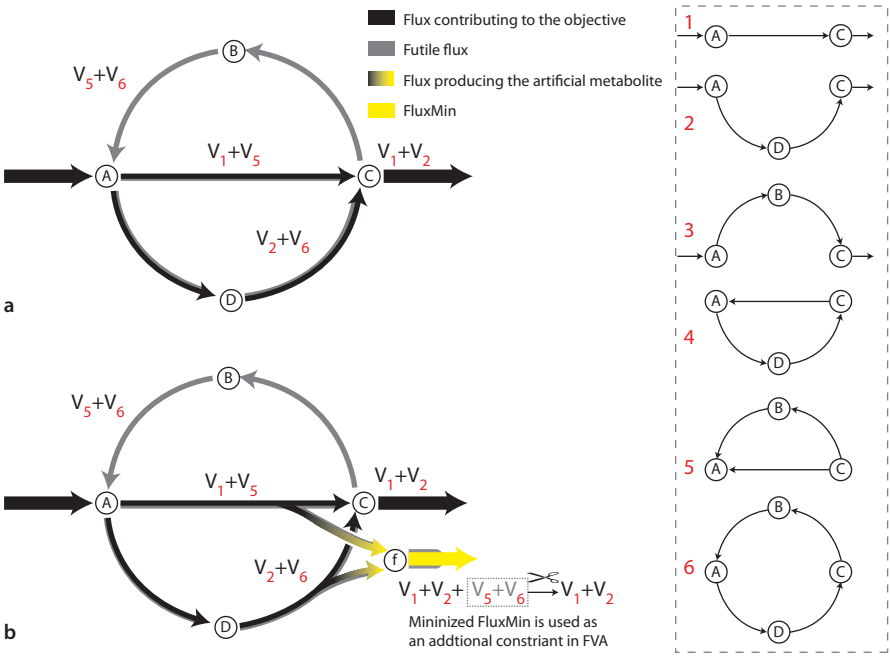

Fig. 11.5 A schematic diagram to illustrate elimination of the futile cycle in an imaginary network by FATMIN ((**a**) flux distribution at steady state without employment of FATMIN, (**b**) minimization in FATMIN to identify another constraint that can remove the futile values associated with cycles 5 and 6, the inset shows alternate path S through the network, with numbers used as flux indices in the main figure, path 3 and 4 carry no flux with reaction directions as shown, but are examples of paths that become active if reactions are made reversible) [82]

lite, and its flux is the sum of all production fluxes. Minimizing this flux while maintaining the objective value, in effect eliminates futile fluxes.

The idea is demonstrated in Fig. 11.5a. Metabolite C is taken as the target metabolite that this simple network produces. The network contains two alternative routes (1 and 2) for producing C and two cycles (5 and 6). All reactions are irreversible, so with paths through the network chosen as indicated, all flux values are nonnegative. A new metabolite labelled "F" is added to both reactions that produce C, as well as the associated drain reaction, to give the augmented network in Fig. 11.5b. Balancing fluxes at all nodes of the augmented network shows that although the net production of C (made up of the contributions v_1 and v_2) is independent of the futile fluxes, the flux of the added reaction FluxMin is $v_1+v_2+v_5+v_6$. Since the value of v_1+v_2 remains fixed for a chosen objective value and negative values for v_5 and v_6 are not allowed, minimization of FluxMin reduces both of the cyclic fluxes to zero. The resulting minimal value is subsequently used as a constraint on FluxMin in a repeated FVA calculation for the augmented network.

Although the explanation above referred to fluxes along individual pathways—in effect, EMs—the strength of the FATMIN method is that these do not in fact

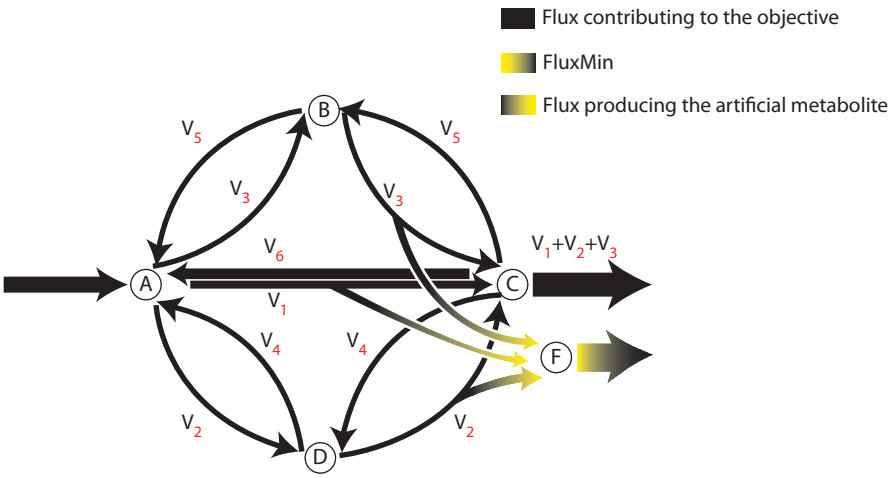

Fig. 11.6 Illustration of all possible flux paths in the imaginary network with reversible loop reactions

have to be identified to apply the method. Simply adding FluxMin to the network and minimizing its flux ensures that the cyclic components implicitly contained in it are eliminated.

This result relies on the fact that negative flux values are excluded. Realistic networks containing reversible reactions are usually handled in FBA by allowing positive and negative flux values. To maintain the exclusion of negative fluxes, in the FATMIN minimization step reversible reactions involving the target metabolite are conceptually split into two counter directional reactions, and the artificial metabolite only entered into the branch that produces the target metabolite. This is illustrated in Fig. 11.6.

In practice, because the other (consuming) branch does not contribute to FluxMin, it can be omitted. So the procedure followed is that where the FVA range of a target reaction exceeds the threshold on both negative and positive fluxes, the negative lower limit constraint is replaced by zero for finding the minimal FluxMin flux value. We stress that for the subsequent FluxMin-constrained FVA calculations the lower limit is restored to its original value.

Another point illustrated by the example is that the FATMIN strategy does not restrict the optimal solution space; for example all v_1 and v_2 value pairs allowed by the original objective value remain viable in the augmented network. Finally, we remark that the strategy described here is easily extended for multiple target metabolites by simply introducing a unique artificial metabolite and drain reaction for each.

Overall, FATMIN can be summarized as a modeling sequence FBA+FVA+FBA+FVA, which is implemented as a computational pipeline consisting of the following steps, where we take NADH as the target metabolite:

1. Perform FBA to calculate the optimal objective $Z_{optimum}$. Then set the obtained optimal $Z_{optimum}$ as the constraint and perform FVA to calculate a feasible flux range for each reaction.
2. Extract NADH reactions and calculate the feasible range of NADH flux values associated with each reaction by multiplying its reaction rate with the NADH stoichiometry coefficient.
3. Any reaction with an absolute value of the NADH flux boundaries higher than a threshold T (here, we chose T = 100 mmol/gDW/h) is identified as a target reaction. If the target reaction is reversible and the upper limit is higher than T, we decompose the target reaction into two reactions in forward and backward direction, respectively.
4. Extend the network by adding an artificial metabolite F as a product to each target reaction, and an F drain reaction FluxMin to the network model.
5. Use the FBA optimum ($Z_{optimum}$) as the constraint and minimize the flux of reaction FluxMin.
6. Perform FVA on the augmented network while using the minimal flux value obtained at step 5 as constraint for the flux of FluxMin.
7. Recalculate the numerical range of NADH flux values using the method in step 2.

This method shares some ideas (further detailed in the next subsection) with [57], in which the solution space is described in terms of three key characteristics: linealities which are the reversible (bidirectional) infinite reactions, rays which are the irreversible infinite reactions, and vertices which are the corner points of the shape formed by interception of the polyhedral cone representing the convex constraint space with the objective plane. However, there could be millions of vertices, from which one cannot identify biological significance. Therefore, FATMIN intends to remove only the rays and linealities, which matched irreversible and reversible cycles. The rays and linealities are independent of the growth medium [57]. Since FATMIN is based on FVA, the method inherits the function of FVA in elucidating the phenomena of equivalent optima in the network. Nonetheless, unlike the results of FVA that could contain infeasible, futile values, the result of FATMIN are the feasible flux ranges for the reactions.

The drawback of FVA is that this tool only produces the outline of a rectangular cuboid or “box” encompassing the polygon of the solutions, but cannot illustrate the real shape of the polygon, which indicates that FVA results cannot reflect the relationships between different reaction fluxes, for example, how the increase in the fluxes of some reactions influence the fluxes of the other reactions.

11.4.5 Mathematical Explanation for the Optimal Solution Space of FBA

There are two ways to describe the set of optimal solutions of a FBA problem. In the first description, corresponding to the LP formulation discussed in Sect. 3.3, optimal

solutions arising from FBA are defined by a finite number of (a) linear inequalities, i.e., the constraints in the LP, and (b) equalities, i.e., the constraint that the objective value of the LP is equal to the optimal value. Therefore, the optimal solution space is a polyhedron (or polytope, when bounded, for example, when none of the fluxes can reach minus or plus infinity). This representation of a polyhedron is called the "outer description" and can be mathematically formulated as follows:

$$P = \left\{ x \in R^n \,\middle|\, Ax = 0 \right\} \qquad A \in R^{m \times n}.$$

The other description of the polyhedron is termed "inner description" or "parametric representation," which involves use of Minkowsky sum to decompose polyhedron into three sets: a linear combination of so-called lines (columns of matrix L), and a positive combination of extremal rays, which form a cone (columns of matrix R), and a convex combination of vertices (columns of matrix V). Thus, the Minkowski characterization of any optimal flux vector can be formulated as follows:

$$P = \{x \mid x = L\lambda + R\mu + V\upsilon,\ \mu,\ \upsilon \geq 0,\ \textstyle\sum \upsilon = 1\}, \tag{11.3}$$

where λ, μ, and υ denote lineality, ray, and vertex.

Linealities and rays correspond to reversible and irreversible cycles in the network, respectively [83]. In linear algebraic terms, for any flux v in this lineality space and any optimal flux v', the flux $v' + \mu v$ is also optimal for every value of μ. And for any flux v in the cone of rays and any optimal flux v, the flux $v' + \upsilon v$ is also optimal for every value of $\upsilon > 0$. The rays and the linealities do not belong to the optimal solution space themselves and thus they do not contribute to optimization of the metabolic objective. In fact, the rays and linealites only carry the information about the directions in which the solution space is unbounded. Conversely, vertices are feasible and optimal flux vectors and they correspond to corner points (a suitably chosen projection) of the polyhedron describing the solution space. Therefore, any optimal flux vector can then be written as a linear combination of the linealities plus a positive linear combination of the rays of the cone plus a convex combination of the vertices of the polytope, as shown in formula (11.3) [83]. This Minkovski description indicates that it is the μv of linealities $v' + \mu v$ and the υv of rays $v' + \upsilon v$ which are removed by the minimization of the flux through the FluxMin reaction in the FATMIN algorithm.

Compared with the elimination of the linealities and rays, identification of all the vertices in a network is much more difficult, because such a task is highly challenging from the point of computation. The vertices are reaction paths through the network that give rise to the maximization of an objective and the total number can in general be much larger than the number of reactions in the system. For example, the FBA of growth of *E. coli* (model version iJR904) on glucose in a mineral medium indicated that only 59 reactions out of 1,066 reactions in the model are variable, but it gave rise to 17,280 vertex solutions [83]. Thus, this suggests the enumeration of vertices in a GEM would not help elucidation of the metabolic behavior of the microorganism from a biological perspective. For characterization

of the phenotypes of microorganisms, an algorithm that aims at removing the lineralities and rays present in the solution space, such as FATMIN, would be far more efficient.

11.5 Summary

The metabolic capabilities and behavior of an organism can be elucidated by development of flux models of genome scale (GEM's) with FBA. FBA characterizes the solution space by identifying one single optimal solution for the flux distribution of a network. On the other hand, the existence of AOS, i.e., multiple equally valid optima that spans the whole optimal solution space, can be explored by an extension of FBA termed as FVA. This technique can help clarify the entire range of achievable cellular functions and examine the redundancy in a GEM.

The primary challenge in the use of FBA and FVA is that many biological networks are underdetermined systems and it is thus difficult to shrink the solution space by the use of stoichiometric constraints alone. The straightforward idea to reduce the number of the futile cycles in FBA results is to fix reaction directions by incorporating energy constraints. However, in most cases, the required kinetic and thermodynamic knowledge of the biochemical reactions are not easy to obtain. And there is a challenge for computation of a genome-scale energy balance model. Therefore, it is desired to develop methods based on available stoichiometry to efficiently eliminate futile loops. Such a method can greatly increase the predictive power of a genome-scale model.

In this chapter, for elimination of futile cycles in the FBA results, we introduce a simple notion to cut off the circulating flux layer while obtaining the same objective value. To comprehensively elucidate the AOS without the interference of futile values, we present FATMIN, a combined pipeline approach based on FBA and FVA.

FATMIN circumvents the need to fully characterize the optimal solution space of genome-scale stoichiometric models (a polyhedron) corresponding to FBA to remove futile cycles and captures the alternate optimal solutions at genome-scale in the form of flux ranges. Because vertices generally appear in large numbers that are quite meaningless to experimental biologists, this approach does not intend to identify all vertices that represent combinations of reaction paths through the network that maximize an objective.

Instead, the outcome of the approach is a list of reactions (enzymes) that produce some target metabolite at high rate. This information would enable experimental biologists to get essential information about candidate pathways for bioengineering, and about biological burdens that restrain high production rates of the target metabolite.

The field of using metabolic flux models to study cellular phenotypes is growing fast, with development of new constraint-based techniques and reconstructions of the GEMs for a wide range of organisms. Therefore, one challenge for the future

is to find different manipulation techniques of these models for engineering new strains of cells and organisms beneficial to human beings.

References

1. De Filippo C, Ramazzotti M, Fontana P, Cavalieri D (2012) Bioinformatic approaches for functional annotation and pathway inference in metagenomics data. Brief Bioinform 13:696–710
2. Zhu H, Snyder M (2002) 'Omic' approaches for unraveling signaling networks. Curr Opin Cell Biol 14:173–179
3. Thiele I, Jamshidi N, Fleming RMT, Palsson B.Ø (2009) Genome-scale reconstruction of *Escherichia coli's* transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. PLoS Comput Biol 5:e1000312
4. Kim TY, Sohn SB, Kim YB, Kim WJ, Lee SY (2012) Recent advances in reconstruction and applications of genome-scale metabolic models. Curr Opin Biotechnol 23:617–623
5. Osterlund T, Nookaew I, Nielsen J (2011) Fifteen years of large scale metabolic modeling of yeast: developments and impacts. Biotechnol Adv 30(5):979–988
6. Price ND, Papin JA, Schilling CH, Palsson BO (2003) Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol 21:162–169
7. Terzer M, Maynard ND, Covert MW, Stelling J (2009) Genome-scale metabolic networks. Wiley Interdiscip Rev Syst Biol Med 1:285–297
8. Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2:886–897
9. Rocha I, Forster J, Nielsen J (2008) Design and application of genome-scale reconstructed metabolic models. Methods Mol Biol 416:409–431
10. Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? Nat Biotechnol 28:245–248
11. Chuang HY, Hofree M, Ideker T (2010) A decade of systems biology. Annu Rev Cell Dev Biol 26:721–744
12. Varma A, Palsson BO (1994) Metabolic flux balancing: basic concepts, scientific and practical use. Nat Biotechnol 12:994–998
13. Price ND, Papin JA, Schilling CH, Palsson BO (2003) Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol 21:162–169
14. Price ND, Schellenberger J, Palsson BO (2004) Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. Biophys J 87:2172–2186
15. Lee JM, Gianchandani EP, Eddy JA, Papin JA (2008) Dynamic analysis of integrated signaling, metabolic, and regulatory networks. PLoS Comput Biol 4:e1000086
16. Covert MW, Xiao N, Chen TJ, Karr JR (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli. Bioinformatics 24:2044–2050
17. Jamshidi N, Palsson Bÿ (2010) Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models. Biophys J 98:175–185
18. Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, Walsh MC, Bakker BM, van Dam K, Westerhoff HV, Snoep JL (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. Eur J Biochem 267:5313–5329
19. Oberhardt MA, Palsson BO, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Mol Systs Biol 5:320
20. Lee JM, Gianchandani EP, Papin JA (2006) Flux balance analysis in the era of metabolomics. Brief Bioinform 7:140–150

21. Covert MW, Famili I, Palsson BO (2003) Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology? Biotechnol Bioeng 84:763–772
22. Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Micro 2:886–897
23. Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. Curr Opin Biotechnol 14:491–496
24. Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol 60:3724–3731
25. Durot M, Bourguignon PY, Schachter V (2009) Genome-scale models of bacterial metabolism: reconstruction and applications. FEMS Microbiol Rev 33:164–190
26. Oliveira A, Nielsen J, Forster J (2005) Modeling *Lactococcus lactis* using a genome-scale flux model. BMC Microbiol 5:39
27. Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat protoc 5:93–121
28. Famili I, Förster J, Nielsen J, Palsson BO (2003) *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci U S A 100:13134–13139
29. Edwards JS, Ibarra RU, Palsson BO (2001) In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol 19:125–130
30. Fong SS, Palsson BO (2004) Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nat Genet 36:1056–1058
31. Ibarra RU, Edwards JS, Palsson BO (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420:186–189
32. Burgard AP, Maranas CD (2003) Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotechnol Bioeng 82:670–677
33. Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, Galagan JE, Church GM (2009) Large-scale identification of genetic design strategies using local search. Mol Syst Biol 5:296
34. Patil K, Rocha I, Forster J, Nielsen J (2005) Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 6:308
35. Xu P, Ranganathan S, Fowler ZL, Maranas CD, Koffas MAG (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab Eng 13:578–587
36. Brochado A, Matos C, Moller B, Hansen J, Mortensen U, Patil K (2010) Improved vanillin production in baker's yeast through in silico design. Microb Cell Fact 9:84
37. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol 3:119
38. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, Konig R, Smith RD, Palsson BO (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Mol Syst Biol 6:390
39. Reed J, Vo T, Schilling C, Palsson B (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol 4:R54
40. Stephanopoulos G, Aristidou AA, Nielsen JH, Nielsen J (1998) Metabolic engineering: principles and methodologies Academic Press, San Diego
41. Palsson BO (2006) Systems biology: properties of reconstructed networks. Cambridge university press, Cambridge
42. Santos F, Boele J, Teusink B (2011) Chapter twenty-four—a practical guide to genome-scale metabolic models and their analysis. In: Daniel Jameson, MV, Hans, VW (eds) Methods enzymol, vol 500. Academic Press, Burlington pp 509–532
43. Feist AM, Palsson BO (2010) The biomass objective function. Curr Opin Microbiol 13:344–349

44. Kim IK, Roldao A, Siewers V, Nielsen J (2012) A systems-level approach for metabolic engineering of yeast cell factories. FEMS Yeast Res 12:228–248
45. Tang YJ, Sapra R, Joyner D, Hazen TC, Myers S, Reichmuth D, Blanch H, Keasling JD (2009) Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain. Biotechnol Bioeng 102:1377–1386
46. Blazeck J, Alper H (2010) Systems metabolic engineering: genome-scale models and beyond. Biotechnol J 5:647–659
47. Boghigian BA, Seth G, Kiss R, Pfeifer BA (2010) Metabolic flux analysis and pharmaceutical production. Metab Eng 12:81–95
48. Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84:647–657
49. Alper H, Jin YS, Moxley JF, Stephanopoulos G (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab Eng 7:155–164
50. Alper H, Miyaoku K, Stephanopoulos G (2005) Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat Biotechnol 23:612–616
51. Park JH, Lee KH, Kim TY, Lee SY (2007) Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. Proc Natl Acad Sci U S A 104:7797–7802
52. Lee KH, Park JH, Kim TY, Kim HU, Lee SY (2007) Systems metabolic engineering of *Escherichia coli* for L-threonine production. Mol Syst Biol 3:149
53. Jung YK, Kim TY, Park SJ, Lee SY (2010) Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. Biotechnol Bioeng 105:161–171
54. Covert MW, Palsson BO (2002) Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. Journal Biol Chem 277:28058–28064
55. Shlomi T, Eisenberg Y, Sharan R, Ruppin E (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Mol Syst Biol 3:101
56. Terzer M, Stelling J (2008) Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24:2229–2235
57. Jol SJ, Kümmel A, Terzer M, Stelling J, Heinemann M (2012) System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes. PLoS Comput Biol 8:e1002415
58. Lee K, Berthiaume F, Stephanopoulos GN, Yarmush DM, Yarmush ML (2000) Metabolic flux analysis of postburn hepatic hypermetabolism. Metab Eng 2:312–327
59. Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng 5:264–276
60. Reed JL, Palsson BØ (2004) Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res 14:1797–1805
61. Thiele I, Fleming RMT, Bordbar A, Schellenberger J, Palsson BØ (2010) Functional characterization of alternate optimal solutions of *Escherichia coli's* transcriptional and translational machinery. Biophys J 98:2072–2081
62. Duarte N, Herrgard M, Palsson B (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 14:1298–1309
63. Beard DA, Liang SD, Qian H (2002) Energy balance for analysis of complex metabolic networks. Biophys J 83:79–86
64. Qian H, Beard DA (2006) Metabolic futile cycles and their functions: a systems analysis of energy and control. Syst Biol, IEE Proc 153:192–200
65. Ibarguren I, Diaz-Enrich MJ, Cao J, Fernandez M, Barcia R, Villamarin JA, Ramos-Martinez JI (2000) Regulation of the futile cycle of fructose phosphate in sea mussel. Comp Biochem Physiol B Biochem Mol Biol 126:495–501

66. Samoilov M, Plyasunov S, Arkin AP (2005) Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. Proc Natl Acad Sci U S A 102:2310–2315
67. Dauner M, Storni T, Sauer U (2001) *Bacillus subtilis* metabolism and energetics in carbon-limited and excess-carbon chemostat culture. J Bacteriol 183:7308–7317
68. Tang YJ, Chakraborty R, Martín HG, Chu J, Hazen TC, Keasling JD (2007) Flux analysis of central metabolic pathways in *Geobacter metallireducens* during reduction of soluble Fe(III)-Nitrilotriacetic acid. Appl Environ Microbiol 73:3859–3864
69. Varma A, Palsson BO (1993) Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. J Theor Biol 165:477–502
70. Holzhütter HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur J Biochem 271:2905–2922
71. Henry CS, Broadbelt LJ, Hatzimanikatis V (2007) Thermodynamics-based metabolic flux analysis. Biophys J 92:1792–1805
72. Fleming RMT, Thiele I, Provan G, Nasheuer HP (2010) Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism. J Theor Biol 264:683–692
73. Fleming RMT, Thiele I (2011) von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models. Bioinformatics 27:142–143
74. Cogne G, Rügen M, Bockmayr A, Titica M, Dussap C-G, Cornet J-F, Legrand J (2011) A model-based method for investigating bioenergetic processes in autotrophically growing eukaryotic microalgae: application to the green algae *Chlamydomonas reinhardtii*. Biotechnol Prog 27:631–640
75. Noor E, Bar-Even A, Flamholz A, Lubling Y, Davidi D, Milo R (2012) An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. Bioinformatics 28:2037–2044
76. De Martino D, Figliuzzi M, De Martino A, Marinari E (2012) A scalable algorithm to explore the Gibbs energy landscape of genome-scale metabolic networks. PLoS Comput Biol 8:e1002562
77. Beard DA, Qian H (2005) Thermodynamic-based computational profiling of cellular regulatory control in hepatocyte metabolism. Am J Physiol Endocrinol Metab 288:E633–644
78. Schellenberger J, Lewis NE, Palsson BO (2011) Elimination of thermodynamically infeasible loops in steady-state metabolic models. Biophys J 100:544–553
79. Noor E, Lewis NE, Milo R (2012) A proof for loop-law constraints in stoichiometric metabolic networks. BMC Syst Biol 6:140
80. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U (2012) Multidimensional optimality of microbial metabolism. Science 336:601–604
81. Klitgord N, Segrè D (2010) The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. Genome Inform 22:41–55
82. Mao L, Verwoerd WS (2013) Model-driven elucidation of the inherent capacity of *Geobacter sulfurreducens* for electricity generation. J Biol Eng 7:14
83. Kelk SM, Olivier BG, Stougie L, Bruggeman FJ (2012) Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. Sci Rep 2:580

Chapter 12
Contaminant Concentration Prediction Along Unsteady Groundwater Flow

Mritunjay Kumar Singh and Priyanka Kumari

12.1 Introduction

A mathematical model in continuum approach comprises of one or more ordinary and/or partial differential equations. The solutions of such differential equations may be obtained through analytical, numerical, and statistical approaches. Each one of these has its own importance. The prime objective in this chapter is to focus on analytical and numerical approaches to solve a solute transport problem in geological formations in general, and confined aquifer in particular. The geological formations have characteristics such as porosity, permeability, hydraulic conductivity. The solute transport involves mechanisms such as advection, dispersion, adsorption, etc. The solute transport problem is dealt with initial and boundary conditions. The aquifers may be clean or may not be so before the source of the pollution is introduced in the domain. The initial condition, normally defined at time $t = 0$, describes this feature. The boundary conditions describe the nature of the source of pollution and its interaction with the flow system and the surrounding. So the concentration is prescribed for all the points of the boundary for the entire period of investigation. There are three types of boundary conditions that occur in the literature: (i) Dirichlet type or the first type in which the solution is prescribed at a position, (ii) Neumann type or the second or flux type in which the derivative of the solution is prescribed, and (iii) the mixed or third type, which is a linear combination of the first- and second-type conditions. The condition may be homogeneous or inhomogeneous.

A geological formation is contaminated by means of various point or surface sources such as domestic, industrial, environmental, and agricultural, etc. On-site septic systems, spills of industrial chemicals, underground injection wells (industrial waste), etc. are the examples of point sources. Examples of nonpoint source contamination are agricultural fields with excessive dosages of fertilizers and pesticides, distribution of waste drugs on the earth surface, etc. [1–4].

M. K. Singh (✉) · P. Kumari
Department of Applied Mathematics, Indian School of Mines, 826004, Dhanbad, India
e-mail: drmks29@rediffmail.com

S.K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_12,

A considerable interest is growing in understanding and evaluating contaminant transport along an aquifer to monitor the quality of the groundwater. The traditional advection–dispersion equation represents a standard model to predict the contaminant movement along the groundwater flow system, based on the Fick's law of diffusion and conservation of mass [5–7]. The literature presents several methods to solve the advection–dispersion equation under a variety of initial and boundary conditions [8–13]. In the last few decades, attempts have been made to predict the contaminant concentration in aquifers for one-, two-, and three-dimensional cases [14–17].

The one-dimensional advection–dispersion equation with adsorption, zero order production, and first-order decay was solved analytically, considering constant dispersion parameters. The solution was obtained using the Laplace transform for the first as well as the third-type boundary conditions [8]. Using the Fourier series and the method of superposition principle, analytical solutions of one-dimensional advection–dispersion equation in semi-infinite and infinite domains were presented for the cases of periodically fluctuating concentration and concentration pulse [18]. Considering scale-dependent dispersion coefficient and periodic type time-dependent boundary condition, the analytical solution of one-dimensional advection–dispersion equation was obtained in the form of hyper-geometric function [19]. The solutions obtained in these studies were developed with a variety of integral transforms. However, it may be obtained more conveniently with the help of the Green's function method (GFM). Keeping these facts in mind, a multidimensional solute transport model for the first, the second, and the third-type boundary conditions using GFM was explored [20].

The time- and space-dependent nature of the dispersion coefficient in solute transport problems has been explored in the literature for both homogeneous and heterogeneous mediums. The heterogeneity arises from the assumption of a scale-dependent dispersion coefficient. The analytical solution of one-dimensional solute transport equation with asymptotic scale-dependent dispersivity was studied [21], in which the dispersion coefficient was assumed to be increasing linearly with distance up to some distance and then reaching an asymptotic value. The analytical solution of one-dimensional advection–dispersion equation was obtained with the help of the power series method, assuming radial distance-dependent velocity and dispersion coefficient for different values of Peclet number [22]. Using the generalized integral transform technique (GITT), a general analytical solution was presented for the linear, one-dimensional advection–dispersion equation with a variety of distance-dependent coefficients in the form of linear, parabolic, exponential, and asymptotic functions [23]. Considering linear equilibrium adsorption and first-order decay, the analytical solution of a one-dimensional solute transport equation was studied. The dispersion coefficient was taken in linear, asymptotic, and exponential forms. The solution was obtained with the method of superposition principle for instantaneous and continuous injection [24]. The contaminant concentration pattern, along uniform groundwater flow for uniform source concentration was discussed with the help of the Laplace integral transform technique (LITT) [25, 26].

Most of the analytical solutions of the advection–dispersion equation available in the literature are based on simple initial and boundary conditions. However, the complexity of boundary condition makes the dispersion problem difficult in getting the analytical solution. The analytical solution of one-dimensional transient storage model was presented using the Laplace transform technique assuming the channel flow velocity and longitudinal dispersion coefficient as constants. The initial condition for both the main channel and the storage zone concentration was considered zero and the upstream boundary conditions for the channel and the storage zone were represented by the Heaviside function involving the time-dependent term [27]. Some of the problems dealt with by previous researchers are enlisted in Tables 12.1 and 12.2 for different set of initial and boundary conditions in confined and unconfined aquifers.

An example of one-dimensional contaminant transport model in a homogeneous, isotropic, semi-infinite aquifer is discussed below. The contaminant on the surface infiltrates, percolates downwards, and mixes with the groundwater; therefore, the initial concentration in the aquifer cannot be ignored. It is assumed here as an exponentially decreasing function of a space-dependent term. The input boundary condition at the origin is linear pulse type and the time-dependent boundary condition is in the form of shifted Heaviside function. The dispersion coefficient is considered directly proportional to the seepage velocity [50, 51]. The study of solute transport modelling along unsteady groundwater flow in an aquifer was explored extensively [52]. The analytical solution is derived with the help of the Laplace transform technique for different types of unsteady groundwater velocity distribution taken from [24] and compared with that of numerical solution obtained by the finite difference method.

12.2 Mathematical Models

A one-dimensional advection–dispersion equation in an isotropic, homogeneous, semi-infinite aquifer may be written as:

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - U\frac{\partial c}{\partial x}, \tag{12.1}$$

where c is the concentration of solute mass, D is the diffusion coefficient, and U is the Darcy velocity.

Initially, groundwater contains solutes. Some initial background concentration is supposed to decrease exponentially with distance. Let it be represented as:

$$c(x,t) = c_i \exp(-\gamma^* x),\ x > 0,\ t = 0, \tag{12.2}$$

where $c_i[ML^{-3}]$ is the initial concentration and $\gamma^*[L^{-1}]$ is the decay parameter. The input point source concentration at the origin, where the contaminants reach the

Table 12.1 Input parameters for one-dimensional problem. [13, 28–49]

Author's name	Dispersion coefficients	Seepage velocities
Savovic and Djordjevich [13]	$D_0 = 0.71\,\text{km}^2$ / year	$u_0 = 0.60\text{km}$ / year
Liu et al. [28]	$D_0 = 0.1\,\text{cm}^2$ / day	$u = 2cm/h - 200cm/h$ $\alpha_L = 0.1$ cm
Serrano, S. E.[29]	$D = 1\,m^2$ / month	$u = 1\text{m}$ / month
Shukla, V. P. [30]	$D = 13.93\text{m}^2$ / s	$u = 25.6\text{km}$ / day $= 0.296\text{m}$ / s
Didierjean et al. [31]	$D_{x1} = 4.63\times10^{-6}\text{m}^2$ / s $D_{x2} = 1.16\times10^{-7}\text{m}^2$ / s	$v_D = 5.2\times10^{-6}\text{m}$ / s
Kumar et al. [32]	$D_0 = 1.0\,\text{hm}^2$ / day	$u_0 = 0.01\text{hm}$ / day
Karahan, H. [33]	$D = 0.005\text{m}^2$ / s	$u = 0.8\text{m}$ / s
Smedt, F. D. [34]	$D = 5\text{m}^2$ / s	$v = 1\text{m}$ / s
Kumar et al. [35]	$D = 5.5\times10^{-2}\text{cm}^2$ / h	$u = 0.9$ cm / h
Yeh and Yeh [36]	$D = 1\text{m}^2$ / day	$u = 1\text{m}$ / day
Srinivasan and Clement [37, 38]	$D_x = 0.18\,\text{cm}^2$ / h for Dirichlet 10 m^2 / year for Cauchy	$v = 1$ cm / h for Dirichlet 100 m / year for Cauchy
Guerrero et al. [39]	$D_x = 1\,\text{cm}^2$ / h	$u = 1\text{cm}$ / h and 10 cm / h
Jaiswal et al. [40]	$D_0 = 0.14\text{km}^2$ / year	$u_0 = 0.25\text{km}$ / year
Yudianto and Yuebo [41]	$D_x = 1.33\text{m}^2$ / s	$u_x = 1.26$ m / s
Kumar et al. [42]	$D_0 = 1.29\text{m}^2$ / day	$u_0 = 1.05$ m / day
Wang et al. [43]	$D_L = 0.4\text{m}^2$ / day	$v = 0.25\text{m}$ / day
Ahsan, M. [44]	$D = 50\text{m}^2$ / s, 100 m^2 / s	$u = 1$ m / s
Guerrero et al. [45]	$D_m = 20,50\text{cm}^2$ / day, two layer $D_m = 7,18\,\text{cm}^2$ / day, five layer	$u_m = 40,25\text{cm}$ / day, two layer $u_m = 10,8\text{cm}$ / day, five layer
Su et al. [46]	$D_1 = 0.0952\,\text{m}^2$ / h 0.001m^2 / h for$m = 0, \lambda = 1$ $D_1 = 0.00481$ $0.046\text{for}m = 2, \lambda = -1$	$v = 1.4$ m / h 0.14 m / h
Liang et al. [47]	$D_{xx} = 1.02\text{m}^2$ / s $D_{yy} = 0.094\text{m}^2$ / s	$U = V = 1/\sqrt{2}\text{m}$ / s
Kong et al. [48]	$K_L = 1\,\text{m}^2$ / s, $K_T = 0.1\,\text{m}^2$ / s	$u = v = 0.5\text{m}$ / s

groundwater level, is considered linear pulse type and the time-dependent boundary condition in the form of the Heaviside unit step function. The concentration gradient at the infinite extent is supposed to be zero. Thus, the two boundary conditions are:

$$c(x,t) = c_0 q(qt+b)\left[u(t-t_1) - u(t-t_2)\right], \quad x = 0, \quad t > 0 \tag{12.3a}$$

Table 12.2 Set of geological formations with initial and boundary conditions. [13, 28–45, 49]

Author's name	Dimension	Nature	Initial conditions	Boundary conditions	Method used
Savovic and Djordjevich [13]	One	Inhomogeneous, semi-infinite	Zero	Uniform	Explicit finitedifference method
Liu et al. [28]	One	Homogeneous, isotropic	Space dependent	Mixed type	Generalized integral transform technique
Serrano, S.E. [29]	One	Homogeneous	Uniform with Dirac delta function	First type	Method of decomposition
Shukla V. P. [30]	One	Homogeneous	Exponential decreasing space dependent	First type, time-dependent sine and cosine form	Fourier transform technique
Didierjean et al. [31]	One	Heterogeneous	Zero	First type	Quadrupole method
Kumar et al. [32]	One	Inhomogeneous, finite and semi-infinite	Space dependent	Uniform pulse type	Explicit finite-difference method
Karahan H. [33]	One	Homogeneous, finite	Space dependent	First type, time dependent	Implicit finite-difference method
Smedt F. D. [34]	One	Homogeneous, infinite	Uniform pulse type	Zero	Laplace transform technique
Kumar et al. [35]	One	Homogeneous, finite	Uniform	Dirichlet and Neumann	Element-free Galerkin method
Yeh and Yeh [36]	One	Homogeneous, isotropic, semi-infinite	Zero	First and third type with fixed concentration	Laplace transform technique
Srinivasan and Clement [37, 38]	One	Homogeneous, semi-infinite	Space dependent exponentially distributed	Dirichlet and Cauchy type	Laplace and linear tansformation
Guerrero et al. [39]	One	Homogeneous	Zero and uniform	First and third type	Generalized integral transform technique
Jaiswal et al. [40]	One	Homogeneous, isotropic, semi-infinite	Space dependent exponentially decreasing	Uniform and varying pulse type	Laplace transform technique
Yudianto and Yuebo [41]	One	Homogeneous, finite	Uniform	Dirichlet and Neumann type	Finite difference, shooting and collocation method
Kumar et al. [42]	One	Homogeneous, semi-infinite	Uniform	Uniform and varying pulse type	Laplace transform technique

Table 12.2 (continued)

Author's name	Dimension	Nature	Initial conditions	Boundary conditions	Method used
Wang et al. [43]	One	Homogeneous, semi-infinite	Non-zero	First and third type	Stepwise superposition approximation approach
Ahsan, M. [44]	One	Homogeneous, finite	Space dependent function	Uniform	Laplace transform finite analytical method
Guerrero et al. [45]	One	Homogeneous, multilayered	Arbitrary function	Uniform mixed type	Classic integral transform technique
Pang and Hunt [49]	One	Homogeneous	Zero	First type	Superposition principle

$$\frac{\partial c}{\partial x} = 0, \quad x \to \infty, \quad t > 0, \tag{12.3b}$$

where $c_0[ML^{-3}]$ is the solute concentration. $q[T^{-1}]$ is decay rate coefficient, b is the constant parameter of linear pulse boundary condition at $x = 0$, the source is present during $t_1 \le t \le t_2$, and $u(t - t_i)$ is the shifted Heaviside function, defined as:

$$u(t - t_i) = \begin{cases} 0, & t < t_i \\ 1, & t \ge t_i \end{cases}. \tag{12.4}$$

$$Let \quad U = U_0 f(mt), \tag{12.5}$$

where $U_0[LT^{-1}]$ is initial seepage velocity. Here, $f(mt)$ is the time varying expression in which $m[T^{-1}]$ is the flow resistance coefficient. For many types of porous media, the dispersion coefficient often varies with the seepage velocity [50]. So we assume:

$$D \propto U, D = aU \Rightarrow D = aU_0 f(mt) \Rightarrow D = D_0 f(mt), \tag{12.6}$$

where $a[L]$ is the dispersivity and $D_0 = aU_0$ is the initial dispersion coefficient. Using Eqs. 12.5 and 12.6 in Eq. 12.1, we get:

$$\frac{1}{f(mt)} \frac{\partial c}{\partial t} = D_0 \frac{\partial^2 c}{\partial x^2} - U_0 \frac{\partial c}{\partial x}. \tag{12.7}$$

We introduce a new time variable T^* [53] as:

$$T^* = \int_0^t f(mt) dt. \tag{12.8}$$

Equation 12.7, with the initial and boundary conditions given in Eqs. 12.2 and 12.3a, b, may be written as:

$$\frac{\partial c}{\partial T^*} = D_0 \frac{\partial^2 c}{\partial x^2} - U_0 \frac{\partial c}{\partial x} \tag{12.9}$$

$$c(x, T^*) = c_i \exp(-\gamma^* x), \quad x > 0, \quad T^* = 0 \tag{12.10}$$

$$c(x, T^*) = c_0 q(qT^* + b)\left[u(T^* - T_1^*) - u(T^* - T_2^*)\right], \quad x = 0, \quad T^* > 0 \tag{12.11a}$$

$$\frac{\partial c}{\partial x} = 0, \quad x \to \infty, \quad T^* > 0. \tag{12.11b}$$

To reduce the number of parameters from the above problem, a set of nondimensional variables are introduced as follows:

$$C = \frac{c}{c_0}, \quad X = \frac{xU_0}{D_0}, \quad T = \frac{U_0^2}{D_0} T^*, \quad \gamma = \frac{\gamma^* D_0}{U_0}, \quad Q = \frac{qD_0}{U_0^2},$$
$$T_1 = \frac{U_0^2}{D_0} T_1^*, \quad T_2 = \frac{U_0^2}{D_0} T_2^*. \tag{12.12}$$

Using Eq. 12.12, Eqs. 12.9 through 12.11a, b reduce to:

$$\frac{\partial C}{\partial T} = \frac{\partial^2 C}{\partial X^2} - \frac{\partial C}{\partial X} \tag{12.13}$$

$$C(X, T) = \frac{c_i}{c_0} \exp(-\gamma X), \quad X > 0, \quad T = 0 \tag{12.14}$$

$$C(X, T) = Q(QT + b)\left[u(T - T_1) - u(T - T_2)\right], \quad X = 0, \quad T > 0 \tag{12.15a}$$

$$\frac{\partial C}{\partial X} = 0, \quad X \to \infty, \quad T > 0. \tag{12.15b}$$

12.2.1 Analytical Solution

To find the analytical solution of the problem, the convective term is reduced from Eq. 12.13, using the following transformation:

$$C(X,T) = K(X,T)\exp\left(\frac{X}{2} - \frac{T}{4}\right). \tag{12.16}$$

Equations 12.13 through 12.15a, b become:

$$\frac{\partial K}{\partial T} = \frac{\partial^2 K}{\partial X^2} \tag{12.17}$$

$$K(X,T) = \frac{c_i}{c_0}\exp\left(-\gamma - \frac{1}{2}\right)X, \quad X > 0, \quad T = 0 \tag{12.18}$$

$$K(X,T) = Q(QT+b)\left[u(T-T_1) - u(T-T_2)\right]\exp\left(\frac{T}{4}\right), \quad X = 0, \quad T > 0 \tag{12.19a}$$

$$\text{and} \quad \frac{\partial K}{\partial X} + \frac{K}{2} = 0 \quad at\ X \to \infty, \quad T > 0. \tag{12.19b}$$

Applying LITT to the Eqs. 12.17 through 12.19a, b, we get the solution $K(X,T)$ as follows:

$$\begin{aligned} K(X,T) = {} & QF_1(X,T) + Q^2F_2(X,T) - QF_3(X,T) - Q^2F_4(X,T) \\ & + \frac{c_i}{c_0}\exp\left(-\gamma - \frac{1}{2}\right)XF_5(X,T) - \frac{c_i}{c_0}F_6(X,T), \end{aligned} \tag{12.20}$$

where

$$F_1(X,T) = (QT_1 + b)\exp\left(\frac{T_1}{4}\right)$$

$$\begin{cases} \dfrac{1}{2}\exp\left(\dfrac{T}{4} - \dfrac{X}{2}\right)\operatorname{erfc}\left(\dfrac{X}{2\sqrt{T}} - \dfrac{\sqrt{T}}{2}\right) + \dfrac{1}{2}\exp\left(\dfrac{T}{4} + \dfrac{X}{2}\right)\operatorname{erfc}\left(\dfrac{X}{2\sqrt{T}} + \dfrac{\sqrt{T}}{2}\right)\Bigg|_{T=T-T_1} & T > T_1, \\ 0 & T < T_1 \end{cases} \tag{12.21a}$$

$$F_2(X,T)=\exp\left(\frac{T_1}{4}\right)\begin{cases}\left.\begin{aligned}&\frac{1}{2}(T-X)\exp\left(\frac{T}{4}-\frac{X}{2}\right)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}}-\frac{\sqrt{T}}{2}\right)\\&+\frac{1}{2}(T+X)\exp\left(\frac{T}{4}+\frac{X}{2}\right)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}}+\frac{\sqrt{T}}{2}\right)\end{aligned}\right|_{T=T-T_1} & T>T_1\\ 0 & T<T_1\end{cases} \tag{12.21b}$$

$$F_3(X,T)=(QT_2+b)\exp\left(\frac{T_2}{4}\right)$$

$$\begin{cases}\left.\frac{1}{2}\exp\left(\frac{T}{4}-\frac{X}{2}\right)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}}-\frac{\sqrt{T}}{2}\right)+\frac{1}{2}\exp\left(\frac{T}{4}+\frac{X}{2}\right)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}}+\frac{\sqrt{T}}{2}\right)\right|_{T=T-T_2} & T>T_2\\ 0 & T<T_2\end{cases} \tag{12.21c}$$

$$F_4(X,T)=\exp\left(\frac{T_2}{4}\right)\begin{cases}\left.\begin{aligned}&\frac{1}{2}(T-X)\exp\left(\frac{T}{4}-\frac{X}{2}\right)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}}-\frac{\sqrt{T}}{2}\right)\\&+\frac{1}{2}(T+X)\exp\left(\frac{T}{4}+\frac{X}{2}\right)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}}+\frac{\sqrt{T}}{2}\right)\end{aligned}\right|_{T=T-T_2} & T>T_2\\ 0 & T<T_2\end{cases} \tag{12.21d}$$

$$F_5(X,T)=\exp\left(\gamma+\frac{1}{2}\right)^2 T \tag{12.21e}$$

$$\begin{aligned}F_6(X,T)=&\frac{1}{2}\exp\left(\gamma^2T+\frac{T}{4}+\gamma T-\gamma X-\frac{X}{2}\right)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}}-\gamma\sqrt{T}-\frac{\sqrt{T}}{2}\right)\\&+\frac{1}{2}\exp\left(\gamma^2T+\frac{T}{4}+\gamma T+\gamma X+\frac{X}{2}\right)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}}+\gamma\sqrt{T}+\frac{\sqrt{T}}{2}\right).\end{aligned} \tag{12.21f}$$

Now using the transformation given in the Eq. 12.16, Eq. 12.20 can be written as:

$$\begin{aligned}C(X,T)=&QG_1(X,T)+Q^2G_2(X,T)-QG_3(X,T)\\&-Q^2G_4(X,T)+\frac{c_i}{c_0}G_5(X,T)-\frac{c_i}{c_0}G_6(X,T),\end{aligned} \tag{12.22}$$

where

$$G_1(X,T) = (QT_1 + b)\exp\left(\frac{T_1}{4}\right)$$

$$\begin{cases} \frac{1}{2}\operatorname{erfc}\left(\frac{X}{2\sqrt{T}} - \frac{\sqrt{T}}{2}\right) + \frac{1}{2}\exp(X)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}} + \frac{\sqrt{T}}{2}\right)\Big|_{T=T-T_1} & T > T_1 \\ 0 & T < T_1 \end{cases}$$

(12.23a)

$$G_2(X,T) = \exp\left(\frac{T_1}{4}\right)$$

$$\begin{cases} \frac{1}{2}(T-X)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}} - \frac{\sqrt{T}}{2}\right) + \frac{1}{2}(T+X)\exp(X)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}} + \frac{\sqrt{T}}{2}\right)\Big|_{T=T-T_1} & T > T_1 \\ 0 & T < T_1 \end{cases}$$

(12.23b)

$$G_3(X,T) = (QT_2 + b)\exp\left(\frac{T_2}{4}\right)$$

$$\begin{cases} \frac{1}{2}\operatorname{erfc}\left(\frac{X}{2\sqrt{T}} - \frac{\sqrt{T}}{2}\right) + \frac{1}{2}\exp(X)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}} + \frac{\sqrt{T}}{2}\right)\Big|_{T=T-T_2} & T > T_2 \\ 0 & T < T_2 \end{cases}$$

(12.23c)

$$G_4(X,T) = \exp\left(\frac{T_2}{4}\right)$$

$$\begin{cases} \frac{1}{2}(T-X)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}} - \frac{\sqrt{T}}{2}\right) + \frac{1}{2}(T+X)\exp(X)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}} + \frac{\sqrt{T}}{2}\right)\Big|_{T=T-T_2} & T > T_2 \\ 0 & T < T_2 \end{cases}$$

(2.23d)

$$G_5(X,T) = \exp(\gamma^2 T + \gamma T - \gamma X) \tag{12.23e}$$

$$G_6(X,T)=\frac{1}{2}\Bigg[\exp\Big(\gamma^2T+\gamma T-\gamma X\Big)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}}-\gamma\sqrt{T}-\frac{\sqrt{T}}{2}\right)+\exp\Big(\gamma^2T+\gamma T+\gamma X+X\Big)\operatorname{erfc}\left(\frac{X}{2\sqrt{T}}+\gamma\sqrt{T}+\frac{\sqrt{T}}{2}\right)\Bigg] \tag{12.23f}$$

For different values of T, the analytical solution can be written as:

$$C(X,T)=\frac{c_i}{c_0}\exp\Big(\gamma^2T+\gamma T-\gamma X\Big)-\frac{c_i}{2c_0}\Bigg[\exp\Big(\gamma^2T+\gamma T-\gamma X\Big)erfc\left(\frac{X}{2\sqrt{T}}-\gamma\sqrt{T}-\frac{\sqrt{T}}{2}\right)+\exp\Big(\gamma^2T+\gamma T+\gamma X+X\Big)erfc\left(\frac{X}{2\sqrt{T}}+\gamma\sqrt{T}+\frac{\sqrt{T}}{2}\right)\Bigg]\quad 0\le T<T_1 \tag{12.24a}$$

$$\begin{aligned}C(X,T)&=Q\Big(QT_1+b\Big)\exp\left(\frac{T_1}{4}\right)\left\{\frac{1}{2}erfc\left(\frac{X}{2\sqrt{T-T_1}}-\frac{\sqrt{T-T_1}}{2}\right)+\frac{1}{2}\exp(X)erfc\left(\frac{X}{2\sqrt{T-T_1}}+\frac{\sqrt{T-T_1}}{2}\right)\right\}\\&+Q^2\exp\left(\frac{T_1}{4}\right)\left\{\frac{1}{2}\Big(T-T_1-X\Big)erfc\left(\frac{X}{2\sqrt{T-T_1}}-\frac{\sqrt{T-T_1}}{2}\right)\right\}\\&+\frac{1}{2}Q^2\exp\left(\frac{T_1}{4}\right)\Big(T-T_1+X\Big)\exp(X)erfc\left(\frac{X}{2\sqrt{T-T_1}}+\frac{\sqrt{T-T_1}}{2}\right)\\&+\frac{c_i}{c_0}\exp\Big(\gamma^2T+\gamma T-\gamma X\Big)\\&-\frac{c_i}{2c_0}\Bigg[\exp\Big(\gamma^2T+\gamma T-\gamma X\Big)erfc\left(\frac{X}{2\sqrt{T}}-\gamma\sqrt{T}-\frac{\sqrt{T}}{2}\right)\\&+\exp\Big(\gamma^2T+\gamma T+\gamma X+X\Big)+erfc\left(\frac{X}{2\sqrt{T}}+\gamma\sqrt{T}+\frac{\sqrt{T}}{2}\right)\Bigg]\quad T_1<T<T_2\end{aligned} \tag{12.24b}$$

$$C(X,T)=Q(QT_1+b)\exp\left(\frac{T_1}{4}\right)\left\{\frac{1}{2}erfc\left(\frac{X}{2\sqrt{T-T_1}}-\frac{\sqrt{T-T_1}}{2}\right)+\frac{1}{2}\exp(X)\operatorname{erfc}\left(\frac{X}{2\sqrt{T-T_1}}+\frac{\sqrt{T-T_1}}{2}\right)\right\}$$
$$+Q^2\exp\left(\frac{T_1}{4}\right)\left\{\frac{1}{2}(T-T_1-X)erfc\left(\frac{X}{2\sqrt{T-T_1}}-\frac{\sqrt{T-T_1}}{2}\right)\right\}$$
$$+Q^2\exp\left(\frac{T_1}{4}\right)\left\{\frac{1}{2}(T-T_1+X)\exp(X)erfc\left(\frac{X}{2\sqrt{T-T_1}}+\frac{\sqrt{T-T_1}}{2}\right)\right\}$$
$$+Q(QT_2+b)\exp\left(\frac{T_2}{4}\right)\left\{\frac{1}{2}erfc\left(\frac{X}{2\sqrt{T-T_2}}-\frac{\sqrt{T-T_2}}{2}\right)\right\}$$
$$+Q(QT_2+b)\exp\left(\frac{T_2}{4}\right)\left\{\frac{1}{2}\exp(X)erfc\left(\frac{X}{2\sqrt{T-T_2}}+\frac{\sqrt{T-T_2}}{2}\right)\right\}$$
$$+Q^2\exp\left(\frac{T_2}{4}\right)\left\{\frac{1}{2}(T-T_2-X)erfc\left(\frac{X}{2\sqrt{T-T_2}}-\frac{\sqrt{T-T_2}}{2}\right)\right\}$$
$$+Q^2\exp\left(\frac{T_2}{4}\right)\left\{\frac{1}{2}(T-T_2+X)\exp(X)erfc\left(\frac{X}{2\sqrt{T-T_2}}+\frac{\sqrt{T-T_2}}{2}\right)\right\}$$
$$+\frac{c_i}{c_0}\exp\left(\gamma^2T+\gamma T-\gamma X\right)-\frac{c_i}{2c_0}\left[\exp\left(\gamma^2T+\gamma T-\gamma X\right)erfc\left(\frac{X}{2\sqrt{T}}-\gamma\sqrt{T}-\frac{\sqrt{T}}{2}\right)\right.$$
$$\left.+\exp\left(\gamma^2T+\gamma T+\gamma X+X\right)erfc\left(\frac{X}{2\sqrt{T}}+\gamma\sqrt{T}+\frac{\sqrt{T}}{2}\right)\right] \qquad T>T_2. \tag{12.24c}$$

Combining the Eqs. 12.24a and 12.24c and using the shifted Heaviside function, the solution for a point source taken at the origin can be written as:

$$C(X,T,T_1,T_2)=F(X,T,T_1,-1)+\exp(X)F(X,T,T_1,1)$$
$$-F(X,T,T_2,-1)-\exp(X)F(X,T,T_2,1)+G(X,T), \tag{12.25}$$

where $F(X,T,T_i,\lambda)$ and $G(X,T)$ are defined as:

$$F(X,T,T_i,\lambda)=u(T-T_i)\frac{1}{2}\exp\left(\frac{T_i}{4}\right) \tag{12.26}$$
$$(Q^2T+Qb+\lambda Q^2X)erfc\left(\frac{X}{2\sqrt{T-T_i}}+\frac{\lambda\sqrt{T-T_i}}{2}\right)$$

and

$$G(X,T)=\frac{c_i}{c_0}\exp(\gamma^2T+\gamma T-\gamma X)-\frac{c_i}{2c_0}\exp(\gamma^2T+\gamma T-\gamma X)erfc\left(\frac{X}{2\sqrt{T}}-\gamma\sqrt{T}-\frac{\sqrt{T}}{2}\right)$$
$$-\frac{c_i}{2c_0}\exp(\gamma^2T+\gamma T+\gamma X+X)erfc\left(\frac{X}{2\sqrt{T}}+\gamma\sqrt{T}+\frac{\sqrt{T}}{2}\right), \tag{12.27}$$

where u is the Heaviside function defined in Eq. 12.4. For the case of $0 \le T < T_1$, the values of $u(T-T_1)$ and $u(T-T_2)$ are zero. Therefore, the terms involving $u(T-T_1)$ and $u(T-T_2)$ vanish and only the last term of the Eq. 12.25 will appear in the solution. This is the case when the input concentration is taken as zero. For the case of $T_1 \le T \le T_2, u(T-T_1)$ is 1 while $u(T-T_2)$ is 0. Hence, in that case the terms involving $u(T-T_2)$, i.e. the third and fourth terms of Eq. 12.25 will be zero.

In the case of $T > T_2$ the values of $u(T-T_1)$ and $u(T-T_2)$ are 1. Therefore, the solution involves all the terms involving $u(T-T_1)$ and $u(T-T_2)$.

12.2.2 Numerical Solution

The equations 12.13 through 12.15b describe the one-dimensional solute transport problem in a semi-infinite domain. In order to get its numerical solution using a finite difference scheme, the semi-infinite domain $X \in (0,\infty)$ is converted into a finite domain $Y \in (0,1)$ using the transformation:

$$Y = 1 - \exp(-X). \tag{12.28}$$

As a result, Eq. 12.13 is transformed into:

$$\frac{\partial C}{\partial T} = (1-Y)^2 \frac{\partial^2 C}{\partial Y^2} - 2(1-Y)\frac{\partial C}{\partial Y}. \tag{12.29}$$

The initial and boundary conditions given by Eqs. 12.14 through 12.15b in the Y domain become:

$$C(Y,T) = \frac{c_i}{c_0}\exp\left(\gamma \log(1-Y)\right), \quad Y > 0, \quad T = 0 \tag{12.30}$$

$$C(X,T) = Q\,(QT+b)\left[u\left(T-T_1\right) - u\left(T-T_2\right)\right], \quad Y = 0, \quad T > 0 \tag{12.31}$$

$$\frac{\partial C}{\partial Y} = 0, \quad Y = 1, \quad T > 0. \tag{12.32}$$

The boundary $Y = 1$ corresponds to $x \to \infty$; but getting concentration values at infinity is not possible. Therefore, the values are evaluated up to some finite length $x = l$, along the longitudinal direction. It corresponds to $Y = 1 - \exp\left(-(u_0 l / D_0)\right) = Y_0$. To find the numerical solution of the problem two-level explicit finite difference schemes are used.

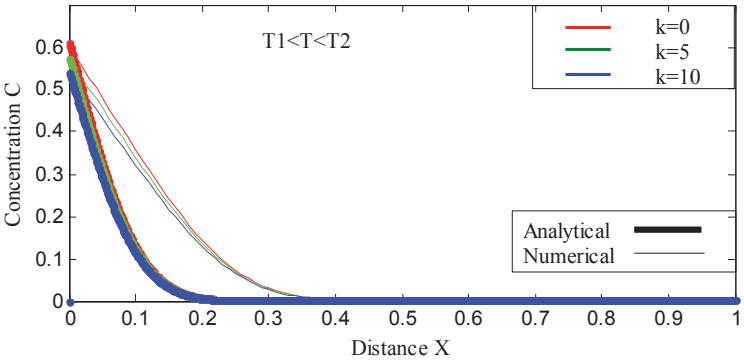

Fig. 12.1 Concentration pattern for exponential velocity during $T_1 < T < T_2$

12.3 Numerical Results and Discussion

Three forms of the time-dependent seepage velocity are considered. They are: (a) asymptotic function, (b) exponential function, and (c) sigmoid function. Mathematical expressions for these functions are:

$$(a) \quad f(mt) = \frac{mt}{(mt+k)} \tag{12.33}$$

$$(b) \quad f(mt) = 1 - k\exp(-mt) \Rightarrow T = \frac{U_0^2}{mD_0}\left[mt + k(\exp(-mt) - 1)\right] \tag{12.34}$$

$$(c) \quad f(mt) = \frac{mt}{\sqrt{k^2 + (mt)^2}} \Rightarrow T = \frac{U_0^2}{mD_0}\left(\sqrt{k^2 + (mt)^2} - k\right). \tag{12.35}$$

The analytical and numerical solutions of the present problem are computed for: $c_i = 0.0001$, $c_0 = 1.0$, $U_0 = 0.001$, $D_0 = 0.01$, $\gamma = 0.3$, $q = 0.0001$, $m = 0.0165$ and $b = 0.5$. The length of the aquifer is assumed to be 1 km. The origin is assumed to be at the source of contaminants; the source is supposed to be active from $t_1 = 969$ days to $t_2 = 1,060$ days (for 3 months approximately). The concentration distribution patterns with position and time are similar in all the three forms of velocity. The effect of unsteady parameter k on the concentration level is shown in Fig. 12.1 for exponential velocity at a time when the source is active. This figure is drawn at $t = 1{,}000$th day for $k = 0$, 5, and 10 ; $k = 0$ represents the uniform seepage velocity and dispersion coefficient. It may be noted that unsteadiness decreases with this

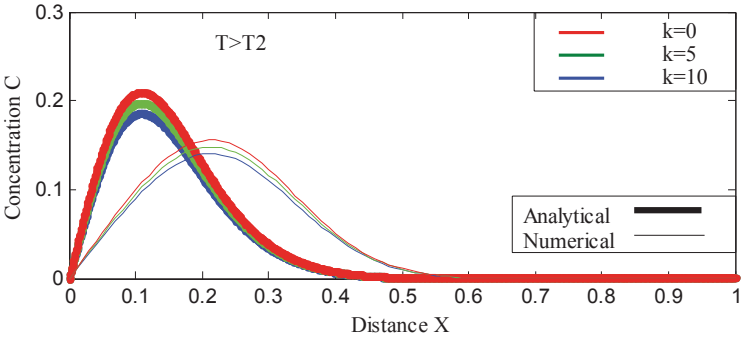

Fig. 12.2 Concentration pattern for exponential velocity during $T > T_2$

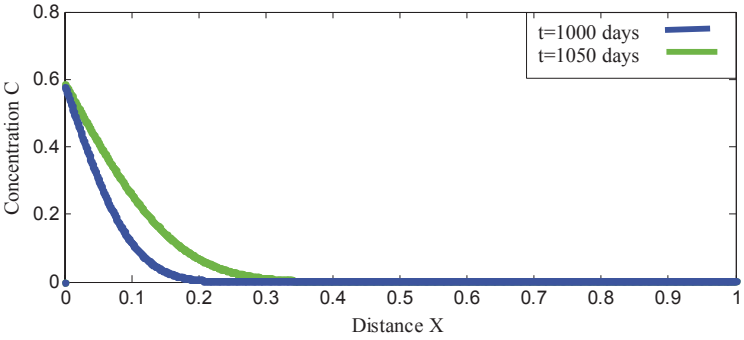

Fig. 12.3 Analytical solution for the sigmoid-type velocity expression during $T_1 < T < T_2$

parameter, as is evident from Eqs. 12.33, 34, and12.35. This figure also compares the analytical and numerical solutions. Both the solutions are similar and this fact mutually validates each other. The input concentration (concentration value at the origin) in both the solutions is the same, and it decreases with k. The same effect is shown in Fig. 12.2 at the $t = 1{,}090$th day, when the source becomes inactive. It may be observed that the input concentration becomes zero and the peak concentration value decreases with k. The peak value describes the level of existing pollution after the removal of its source.

The effect of time on the concentration level is shown in Fig. 12.3 for sigmoid velocity in the presence of source for $k = 5$. This figure is drawn at $t = 1{,}000$ and $1{,}050$ days. The pollution level increases with time at a particular position. After the removal of the source, this effect is shown in Fig. 12.4. This figure is drawn at $t = 1{,}090$ and $1{,}120$ days, and it shows the pattern of rehabilitation of the polluted domain with time.

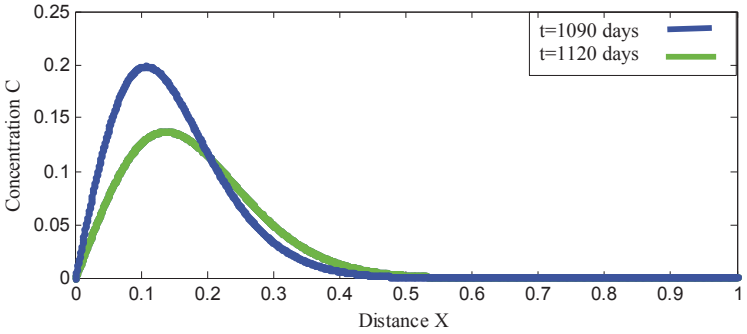

Fig. 12.4 Analytical solution for the sigmoid type velocity expression during $T > T_2$

12.4 Conclusion

The one-dimensional contaminant concentration pattern governed by the advective-diffusive process along a homogeneous semi-infinite aquifer is depicted with a pulse-type input condition. Linear, exponentially decreasing, and sigmoid forms of Darcy velocity are considered. Using suitable transformations variable coefficients (due to unsteadiness of the two dispersive–advective parameters) are reduced into constant coefficients. As a result, it becomes possible to use the Laplace transformation technique to get the analytical solutions. Numerical solutions of the advection diffusive problem are also obtained. To get the numerical solution, the semi-infinite domain is converted into a finite domain. Analytical and numerical solutions are compared and they are found to be in good agreement. The unsteadiness of velocity is defined with the help of two parameters m and k .

12.5 Notations

c	Contaminant concentration in the aquifer at any time ML^{-3}
t	Time variable T
D	Dispersion coefficient along x -axis L^2T^{-1}
U	Groundwater velocity component along x -axis LT^{-1}
x	Space variable along x -axis L
c_{i_*}	Initial background solute concentration ML^{-3}
γ	Decay parameter L^{-1}
c_0	Solute concentration at the source ML^{-3}
q	Decay coefficients T^{-1}
b	Constant parameters

u Shifted Heaviside function
t_1 Beginning time of the activated source T
t_2 Ending time of the activated source T
U_0 Initial seepage velocity LT^{-1}
a Dispersivity L
D_0 Initial dispersion coefficient L^2T^{-1}
f Temporally dependent function
m Flow resistance coefficient T^{-1}
C Non-dimensional solute concentration in the aquifer
X Non-dimensional space variable
T^* New time variable T
T Non-dimensional time variable
γ Non-dimensional decay parameter
Q Non-dimensional decay rate coefficient
T_1 Non-dimensional beginning time of the activated source
T_2 Non-dimensional ending time of the activated source
K Non-dimensional solute concentration in the aquifer
λ Constant parameter
k Arbitrary constant

References

1. Bear J, Verruijt A (1987) Modelling groundwater flow and pollution. Reidel Publishing Co., Dordrecht, p 414
2. Ghosh NC, Sharma KD (2006) Groundwater modelling and management. Capital Publishing Company, New-Delhi
3. Batu V (2006) Applied flow and solute transport modeling in aquifers: fundamental principles and analytical and numerical methods. CRC Press, Taylor and Francis, USA
4. Singh P, Singh MK, Singh VP (2010) Contaminant transport in unsteady groundwater flow: analytical solutions. LAP LAMBERT Academic Publishing, Germany
5. Fried JJ, Combarnous MA (1971) Dispersion in porous media. Adv Hydrosci 7:169–281
6. Bear J (1972) Dynamics of fluids in porous media. Elsevier, New York
7. Chrysikopolous CV, Roberts PV, Kitanidis PK (1990) One-dimensional solute transport in porous media with partial well-to-well recirculation: application to field experiment. Water Resour Res 26(6):1189–1195
8. van Genucheten MT, Alves WJ (1982) Analytical solution of one-dimensional convective-dispersion solute transport equation. Tech Bull 1661:1–55
9. Celia MA, Russell TF, Herrera I (1990) An Eulerian-Lagrangian localized adjoint method for the advection-diffusion equation. Adv Water Resour 13(4):187–206
10. Hunt B (2002) Scale dependent dispersion from a pit. J Hydraul Div 104:75–85
11. Chen JS, Ni CF, Liang CP (2008a) Analytical power series solutions to the two-dimensional advection-dispersion equation with distance-dependent dispersivities. Hydrol Process 22(24):4670–4678
12. Chen JS, Ni CF, Liang CP, Chiang CC (2008b) Analytical power series solution for contaminant transport with hyperbolic asymptotic distance-dependent dispersivity. J Hydrol 362(1–2):142–149

13. Savovic S, Djordjevich A (2012) Finite difference solution of the one-dimensional advection-diffusion equation with variable coefficients in semi-infinite media. Int J Heat and Mass Trans 55:4291–4294
14. Yeh GT (1981) *AT123D*: analytical transient one-, two-, and three-dimensional simulation of waste transport in the aquifer system. Oak Ridge National Laboratory, USA (ORNL-5602)
15. Sim Y, Chrysikopolous CV (1999) Analytical solutions for solute transport in saturated porous media with semi-infinite or finite thickness. Adv Water Res 22(5):507–519
16. Park E, Zhan H (2001) Analytical solutions of contaminant transport from finite one, two, and three-dimensional sources in a finite-thickness aquifer. J Contam Hydrol 53:41–61
17. Costa CP, Vilhena MT, Moreira DM, Tirabassi T (2006) Semi-analytical solution of the steady three-dimensional advection-diffusion equation in the planetary boundary layer. Atmos Environ 40:5659–5669
18. Valocchi AJ, Roberts PV (1983) Attenuation of groundwater contaminant pulses. J Hydraul Eng 109(12):1665–1682
19. Logan JD (1996) Solute transport in porous media with scale dependent dispersion and periodic boundary conditions. J Hydrol 184(3–4):261–276
20. Leij FJ, van Genuchten MTh (2000) Analytical modeling of nonaqueous phase liquid dissolution with Green's functions. Trans Por Med 38:141–166
21. Huang K, van Genuchten MT, Zhang R (1996) Exact solutions for one-dimensional transport with asymptotic scale-dependent dispersion. Appl Math Model 20:298–308
22. Chen JS, Liu CW, Liao CM (2002) A novel analytical power series solution for solute transport in a radially convergent flow field. J Hydrol 266(1–2):120–138
23. Guerrero JSP, Skaggs TH (2010) Analytical solution for one-dimensional advection-dispersion transport equation with distance-dependent coefficients. J Hydrol 390:57–65
24. Basha HA, EI-Habel FS (1993) Analytical solution of the one-dimensional time dependent transport equation. Water Resour Res 29(9):3209–3214
25. Singh MK, Mahato NK, Singh P (2008) Longitudinal dispersion with time dependent source concentration in semi-infinite aquifer. J Earth Syst Sci 117(6):945–949
26. Singh MK, Ahamad S, Singh VP (2012) Analytical solution for one-dimensional solute dispersion with time-dependent source concentration along uniform groundwater flow in a homogeneous porous formation. J Eng Mech. doi:10.1061/(ASCE)EM.1943-7889.0000384
27. Alhan CMK (2008) Analytical solutions for contaminant transport in streams. J Hydrol 348:524–534
28. Liu C, Szecsody JE, Zachara JM, William PB (2000) Use of the generalized integral transform method for solving equations of solute transport in porous media. Adv Water Res 23:483–492
29. Serrano SE (2001) Solute transport under non-linear sorption and decay. Water Res 35(6):1525–1533
30. Shukla VP (2002) Analytical solutions for unsteady transport dispersion of nonconservative pollutant with time-dependent periodic waste discharge concentration. J Hydraul Eng 128(9):866–869
31. Didierjean S, Maillet D, Moyne C (2004) Analytical solutions of one-dimensional macrodispersion in stratified porous media by the quadrupole method: convergence to an equivalent homogeneous porous medium. Adv Water Resour 27:657–667
32. Kumar N, Singh VP, Yadav RR (2005) Distribution of pulse type uniform input in aquifers in tropical regions. Ind J Eng Mat Sci 12:356–362
33. Karahan H (2006) Implicit finite difference techniques for the advection-diffusion equation using spreadsheets. Adv Eng Soft 37:601–608
34. Smedt FD (2006) Analytical solutions for transport of decaying solutes in rivers with transient storage. J Hydrol 330:672–680
35. Kumar RP, Dodagoudar GR, Rao BN (2007) Meshfree modelling of one-dimensional contaminant transport in unsaturated porous media. Geomech Geoeng 2(2):129–136
36. Yeh HD, Yeh GT (2007) Analysis of point-source and boundary-source solutions of one-dimensional groundwater transport equation. J Env Eng 133(11):1032–1041

37. Srininivasan V, Clement TP (2008a) Analytical solutions for sequentially coupled one-dimensional reactive transport problems—part I: mathematical derivations. Adv Water Resour 31:203–218
38. Srininivasan V, Clement TP (2008b) Analytical solutions for sequentially coupled one-dimensional reactive transport problems—part II: special cases, implementation and testing. Adv Water Resour 31:219–232
39. Guerrero JSP, Pimentel LCG, Skaggs TH, vanGenuchten MT (2009) Analytical solution of the advection-diffusion transport equation using a change-of-variable and integral transform technique. Int J Heat and Mass Trans 52:3297–3304
40. Jaiswal DK, Kumar A, Kumar N, Yadav R (2009) R.: Analytical solutions for temporally and spatially dependent solute dispersion of pulse type input concentration in one-dimensional semi-infinite media. J Hydro Environ Res 2:254–263
41. Yudianto D, Yuebo X (2010) A comparison of some numerical methods in solving 1-D steady-state advection dispersion reaction equation. Civil Eng Env Sys 27(2):155–172
42. Kumar A, Jaiswal DK, Yadav RR (2011) One-dimensional Ssolute transport for uniform and varying pulse type input point source with temporally dependent coefficients in longitudinal semi-infinite homogeneous porous domain. Int J Maths Scientific Comput 1(2):56–66
43. Wang H, Han R, Zhao Y, Lu W, Zhang Y (2011) Stepwise superposition approximation approach for analytical solutions with non-zero initial concentration using existing solutions of zero initial concentration in contaminate transport. J Env Sci 23(6) 923–930
44. Ahsan M (2012) Numerical solution of the advection-diffusion equation using Laplace transform finite analytical method. Int J River Basin Manag 10(2):177–188
45. Guerrero JSP, Pontedeiro EM, Skaggs TH, vanGenuchten MT (2013) Analytical solutions of the one-dimensional advection-dispersion solute transport equation subject to time-dependent boundary conditions. Chem Eng J 221:487–491
46. Su N, Sander GC, Fawang L, Anh V, Barry DA (2005) Similarity solutions for solute transport in fractal porous media using a time and scale-dependent dispersivity. Appl Math Model 29:852–870
47. Liang D, Wang X, Falconer RA, Bockelmann-Evans BN (2010) Solving the depth-integrated solute transport equation with a TVD-MacCormack scheme. Environ Model Soft 25:1619–1629
48. Kong J, Xin P, Shen CJ, Song ZY, Li L (2013) A high-resolution method for the depth-integrated solute transport equation based on an unstructured mesh. Environ Model Soft 40:109–127
49. Pang L, Hunt B (2001) Solutions and verification of scale-dependent dispersion model. J Contam Hydrol 53:21–39
50. Ebach EH, White R (1958) Mixing of fluid flowing through beds of packed solids. J Am Inst Chem Eng 4:161–164
51. Kumar N (1983) Unsteady flow against dispersion in finite porous media. J Hydrol 63:345–358
52. Mahato NK (2012) Study of solute transport modeling along unsteady groundwater flow in aquifer. Ph. D. Thesis, Indian school of Mines, Dhanbad
53. Crank J (1975) The mathematics of diffusion. Oxford University Press, Oxford

Chapter 13
Wavelet-Multigrid Method for Solving Modified Reynolds Equation Modeling Synovial Fluid Flow in a Normal Human Knee Joint

Chandrasekhar Salimath

13.1 Introduction

Wavelets are recognized as a powerful mathematical tool in signal and image processing, time-series analysis, geophysics, computer graphics, etc. The last few years have witnessed tremendous amount of interest and activity in the application of the wavelet theory and its associated multi-resolution analysis to different areas of science and technology. Lately, an area in which wavelets are gaining currency is numerical simulations of ordinary and partial differential equations. What makes them particularly appealing to the solution of partial differential equations is their hierarchical structure and localization property, coupled with their time and scale invariance. Wavelet-multigrid method is applied to solve modified Reynolds equation arising in biomedical engineering, an illustration of the advantages of wavelet methods over traditional methods. The unique feature of this method is combining classical multigrid scheme with the modern wavelet theory in such a way that each benefits from the other.

We turn our attention to synovial joints [1] of the human body. Recently, the research on synovial joints has focused on two aspects: To investigate the fundamental lubrication processes occurring in the natural joints of the human body and the development of artificial, replacement joints based on theoretical results. Mathematical models of human joints serve to predict quantities which are difficult to measure experimentally and to simulate changes to the physiological conditions. With particular reference to the human knee joint, many researchers are involved in modeling of the bio-mechanism of the joint lubrication, proposing sophisticated models that involve complex numerical computation in order to obtain the desired solutions.

The main aim of this chapter is to propose an analytical approximate squeeze film lubrication model of the human knee joint for the quick assessment of the synovial fluid pressure and the associated load carrying capacity. The proposed

C. Salimath (✉)
JSS Academy of Technical Education, Noida, India
e-mail: salimathcs@yahoo.com

S.K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_13,

model is based on a modified Reynolds equation; the solution of which gives the fluid pressure and consequently the load-carrying capacity. Normal synovial fluid is viscous, and its functions are nutrition and lubrication of cartilage, load bearing and shock absorption, ensuring efficient lubrication and proper functioning of the synovial joints.

The modified Reynolds equation which incorporates randomized surface roughness structure as well as the elastic nature of the articular cartilage with viscous and non-Newtonian synovial fluid as lubricant is derived. A simplified mathematical model has been developed for understanding the combined effects of surface roughness, poroelasticity, and lubrication aspects of viscous and non-Newtonian fluid of human knee joint.

Retrospectively, the simplest and the most successful linear biphasic model for articular cartilage has been developed by Mow and others [2]. This model includes small deformation of poroelastic material which corresponds to Biot [3] theory for soil consolidation. Using this (Biot's) theory, the governing equations for cartilage deformation and motion of interstitial fluid were formulated. Monsour and others [4] modeled the joint as porous permeable, deformable elastic material (articular cartilage) filled with a single layer of homogenous fluid. The tissue secretes viscous and highly non-Newtonian fluid called synovial fluid which mainly consists of hyaluronic acid, nutrients, glycoprotein, etc. This synovial fluid bathes and supplies nutrients to both surfaces of the cartilage. Hou and others [5] analyzed the squeeze film lubrication of articular cartilage by assuming synovial fluid to be linearly viscous fluid. Detailed analyses about articular cartilage and non-Newtonian characteristics of synovial fluid are given in [6].

Poroelasticty is a continuum theory for the analysis of porous media with elastic matrix consisting of interconnected fluid filled pores. When fluid permeates into a poroelastic material, the drag force between the fluid and the porous medium may cause deformation in the porous matrix. This leads to volumetric changes in the pores. Since the pores are filled with fluid, the presence of the fluid not only acts as a stiffener of the material but also results in the flow of pore fluid between regions of higher and lower pore pressure. A successful model for cartilage with interstitial fluid has been developed by Mow and his coworkers [6]. This simplest linear version of biphasic mixture includes the small deformation of the porous elastic matrix, which corresponds to Biot's model for soil consolidation. The above biphasic model for a homogeneous and isotropic articular cartilage was used in a series of papers to model the fluid flow across the articular surface in geometrically idealized step-loaded synovial joints [5–7]. Various aspects of articular cartilage and non-Newtonian characteristics of the synovial fluid are presented by Torzilli and Mow [6]. Collins [8] considered Biot's theory to model poroelastic matrix for cartilage which is assumed to satisfy generalized form of Darcy's law for unsteady flow in an elastic porous medium. Later, modified and corrected forms of the constitutive equations were given in [9–12]. Most of the studies on synovial joint mechanics and lubrication have used elastic single-phase models of cartilage and a Newtonian single-phase model for synovial fluid. Recently, Mercer and Barry [13] gave a numerical method on finite difference approximations for the calculation of

deformation, pressure, and flow within a finite two-dimensional poroelastic medium by considering slip and no-slip boundaries.

Sayles and others [14] experimentally revealed that the cartilage surface is rough and that roughness-height distribution is Gaussian in nature. This motivates us to investigate the effect of roughness in the cartilage surface. Christensen [15] developed the stochastic theory to investigate the effect of roughness in hydrodynamic lubrication on the assumption that roughness can be represented as a randomly varying quantity. It is assumed that in classical hydrodynamic lubrication theory, the roughness heights are small compared to the film thickness. This theory consists of two types of roughness structures, namely, longitudinal and transverse roughness. The former one has the roughness striations in the form of long ridges and narrows in horizontal direction whereas the latter one in the vertical direction.

The present chapter is organized as follows. In order to make it as self-contained as possible, the anatomy of the human knee joint and its bio-mechanism is given in Sect. 13.2. The simplified modified constitutive relations of poroelastic material are considered in the Sect. 13.3. In Sect. 13.4, the modified Reynolds equation which holds in the film region is derived for two roughness patterns, namely, longitudinal and transverse roughness structures. In Sect. 13.5, we describe briefly wavelet-multigrid method, recently developed by the author, for the solution of elliptic partial differential equations. Section 13.6 is devoted to the discussion of the results obtained in the previous sections for various parameters. Section 13.7 summarizes the major findings of the present study.

13.2 Anatomy and Bio-Mechanism of a Human Knee Joint

The knee joint is one of the most important joints of our body. It plays an essential role in movement related to carrying the body weight in horizontal (running and walking) and vertical (jumps) directions. The knee joint is one of the largest and most complex joints in the body. The knee joins the thighbone (femur) to the shinbone (tibia). The smaller bone that runs alongside the tibia (fibula) and the kneecap (patella) are the other bones that make the knee joint. Tendons connect the knee bones to the leg muscles that move the knee joint. Ligaments join the knee bones and provide stability to the knee in preventive and self-corrective ways. The anterior cruciate ligament prevents the femur from sliding backward on the tibia (or the tibia sliding forward on the femur).The posterior cruciate ligament prevents the femur from sliding forward on the tibia (or the tibia from sliding backward on the femur). The medial and lateral collateral ligaments prevent the femur from sliding side to side.

Two C-shaped pieces of cartilage called the medial and lateral menisci act as shock absorbers between the femur and tibia. Numerous synovial fluid-filled sacs help the knee move smoothly. When a person walks or stands still, the force L acting through pushes the femoral condyles and the tibial plateau together when $L>0$

and to create a squeeze film effect when $L<0$. The synovial fluid is sucked into or sucked out of the cartilage. Cartilage is avascular (free from blood supply) and the flow of synovial fluid into the cartilage is one of the ways it receives nutrients. However, when a person stands still for an extended period of time, it is believed that the synovial fluid will eventually be squeezed out of the gap between tibial plateau and femoral condyles, causing direct contact between cartilage-coated surfaces. This is an undesirable condition, as it results in cartilage degradation and promotes in the long run osteoarthritis.

The physical configuration of the problem is shown in [16] to predict the performance of the knee joint. The bone ends are covered by articular cartilage to prevent natural abrasion, which is in a sac containing fluid for lubricating the two surfaces. A tough fibrous capsule together with the muscles, ligaments, intra-articular structures, etc. encloses the normal joint cavity. The inner lining of this capsule, the synovial membrane, secretes viscous and highly lubricating fluid called synovial fluid. This fluid bathes both articular surfaces and intra-articular structures. Following Walker and Erkman [17], as the load-bearing area of the synovial knee joint is small, two articular surfaces may be considered to be parallel under high loading conditions. For mathematical simplicity, the average of the three layers of the cartilages is modeled as a single poroelastic layer. So, the problem considered is that of squeeze film lubrication between two rectangular surfaces with finite dimensions.

13.3 Mathematical Modeling and Computer Simulation

We present a simple mathematical model for the human knee joint. The governing equations are the principles of mass and momentum balances applied to the simplified geometry under the following hypotheses.

1. The lubricant in the film region is modeled as Newtonian, i.e., linearly viscous and incompressible fluid.
2. For simplicity, both the tibial plateau and the femoral condyles are modeled as rigid impermeable flat surfaces and the effects of menisci are neglected.
3. Classical lubrication theory is used to describe the flow of synovial fluid in the thin gap.
4. The flow is assumed to be steady state, laminar, incompressible, and three-dimensional.
5. Viscosity is kinematical and depends on the hyaluronic acid concentration.

The application of hypotheses from 1 to 5 to the conservative laws leads to a modified Reynolds lubrication equation giving fluid pressure in the joint which supports the load avoiding the direct contact between the solid surfaces.

The upper rigid rough impervious spherical indenter approaches the lower poroelastic matrix normally with a constant velocity dH / dt. The film thickness between two plates is characterized by:

$$H = h(x, y) + h_s(x, y, \xi), \tag{13.1}$$

where $h(x, y) = h_0 + (x^2 + y^2)/2R$ is the nominal smooth part of the geometry, h_0 is the minimum thickness, R is the radius of the indenter in x–y plane, h_s is the height of the surface asperities measured from the nominal level which is a randomly varying quantity of zero mean and ξ is the index parameter determining a definite roughness structure. All the articulations of knee joint under fluid film lubrication involve cartilage–viscous fluid–cartilage interactions. The lubricant in the film region is assumed to be Newtonian fluid, i.e., linearly viscous. So, the problem considered would be that of three-dimensional squeeze film lubrication between the upper rigid spherical indenter and the lower smooth poroelastic matrix. With usual assumptions of lubrication, the governing equations are:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{13.2}$$

$$\frac{\partial^2 u}{\partial z^2} = \frac{1}{\mu}\frac{\partial p}{\partial x}, \tag{13.3}$$

$$\frac{\partial^2 v}{\partial z^2} = \frac{1}{\mu}\frac{\partial p}{\partial y} \tag{13.4}$$

$$0 = \frac{\partial p}{\partial z}, \tag{13.5}$$

where $u, v,$ and w are the velocity components in $x, y,$ and z directions, respectively, p is the pressure, and μ is the viscosity of the fluid.

13.3.1 Poroelastic Region

Following Mow and others [2], the poroelastic material is assumed to be made of a linearly elastic solid phase (deformable cartilage matrix) and an inviscid fluid phase in which these two phases are intrinsically incompressible. For the cartilage matrix deformation and viscous fluid contained in its pores, we write modified coupled equations of motion as in [9, 12, 17]:

$$\text{Matrix}: \ \rho_m \frac{\partial^2 \mathbf{U}}{\partial t^2} = div\,\sigma_m - (\phi^f)^2 \frac{\mu}{k}\left(\frac{\partial \mathbf{U}}{\partial t} - \mathbf{V}\right) \tag{13.6}$$

$$\text{Fluid: } \rho_f \frac{\partial \mathbf{V}}{\partial t} = div\,\sigma_f + (\phi^f)^2 \frac{\mu}{k}\left(\frac{\partial \mathbf{U}}{\partial t} - \mathbf{V}\right) \text{ and} \tag{13.7}$$

$$\nabla \cdot \left(\phi^s \frac{\partial \mathbf{U}}{\partial t} + \phi^f \mathbf{V}\right) = 0, \tag{13.8}$$

where ϕ^f is the porosity and $\phi^s = 1 - \phi^f$ is the solidity of the poroelastic material. ρ_m and ρ_f denote the densities of solid matrix and fluid, respectively, $\mathbf{U}$ is the corresponding displacement vector, k is the permeability of the cartilaginous matrix to the fluid. The left hand terms denote the local forces (mass times acceleration), which are counterbalanced by the right hand terms namely the surface forces, $div\,\sigma$, and the porous medium driving forces (Darcy's law), respectively. These two component equations may be simply viewed as generalized forms of Darcy's law for unsteady flow in a deformable porous medium in terms of relative velocity $(\partial \mathbf{U}/\partial t - \mathbf{V})$ between the moving cartilage and the fluid contained in its pores. Also, Eqs. 6 and 7 denote force balances for the linear elastic solid and viscous fluid components of the cartilage, respectively. The classical stress tensor σ for a continuous homogeneous medium may be expressed for the matrix (cartilage) and fluid (synovial). Thus, the constitutive relations for the solid and fluid phases, respectively are

$$\sigma_m = -\phi^s P\mathbf{I} + 2Ne + A'e\mathbf{I}, \tag{13.9}$$

$$\sigma_f = -\phi^f P\mathbf{I} + E'e\mathbf{I}, \tag{13.10}$$

in terms of the elastic parameters N, E', and A' of the cartilage and the hydrostatic pressure P and $\mathbf{I}$ the identity tensor, e the cartilage dilation. The inertial terms in Eqs. 13.6 and 13.7 are neglected, because in the balance of the momentum equation, the fluid–fluid viscous stress is negligible compared with the drag between the fluid and the solid matrix [10]. After neglecting the inertia terms in Eqs. 13.6 and 13.7 and using this result into Eq. 13.8, we get:

$$\nabla^2 e - \frac{\mu}{k(A + E' + 2N)} \frac{\partial e}{\partial t} = 0, \tag{13.11}$$

where $e = \nabla \cdot \mathbf{U}$. The typical stress-strain curve obtained from confined compression tests an articular cartilage under loading stresses is essentially a linear relationship beyond the initial stress level [18]. Following Collins [8], their results may be characterized by a simple linear equation in terms of the corresponding average bulk modulus K and pressure Pas:

$$e = e_0 + \frac{P}{K}. \tag{13.12}$$

Using this empirical relation (Eq. 13.12) in Eq. 13.11, we get the equation describing pressure in the poroelastic region:

$$\nabla^2 P = 0. \tag{13.13}$$

13.3.2 Boundary Conditions

The relevant boundary conditions for the velocity field ($0<z<H$) are:

$$u(x,y,0) = u(x,y,H) = v(x,y,0) = v(x,y,H) = 0, \tag{13.14}$$

$$w(x,y,0) = -w_n, \quad w(x,y,H) = -\frac{dH}{dt}, \tag{13.15}$$

where w_n represents the normal component of the relative velocity of the fluid at the cartilage surface. Conditions (Eq. 13.14) are no-slip velocity conditions.

13.4 Solution Procedure

Integrating Eq. 13.13 with respect to z over the porous layer thickness ($-\delta$, 0) and using the solid backing boundary condition $\partial P/\partial z = 0$ at $z = -\delta$ we get:

$$\left.\frac{\partial P}{\partial z}\right|_{z=0} = -\int_{-\delta}^{0}\left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2}\right)dz. \tag{13.16}$$

Here δ is the thickness of the poroelastic layer. Using the Morgan–Cameron approximation [19] valid for the poroelastic layer to be small and incorporating the pressure continuity condition ($p = p^*$) at the porous interface $z=0$, we get:

$$\left.\frac{\partial P}{\partial z}\right|_{z=0} = -\delta\left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2}\right). \tag{13.17}$$

After neglecting inertia terms, Eq. 13.7 may be arranged in terms of relative velocity in the form:

$$\left(\mathbf{V} - \frac{\partial \mathbf{U}}{\partial t}\right) = -\frac{k}{\mu(\phi^f)^2}(\phi^f \nabla P - E'\nabla e). \tag{13.18}$$

Elimination of e through Eqs. 13.12 and 13.18 gives:

$$\left(\mathbf{V} - \frac{\partial \mathbf{U}}{\partial t}\right) = -\frac{k}{\mu(\phi^f)^2}\nabla P\left(\phi^f - \frac{E'}{K}\right). \tag{13.19}$$

The normal component of the relative fluid velocity at the cartilage surface is

$$w_n = -\left(\mathbf{V} - \frac{\partial \mathbf{U}}{\partial t}\right)_n = -\frac{k}{\mu(\phi^f)^2}\left(\frac{E'}{K} - \phi^f\right)\frac{\partial P}{\partial z}\bigg|_{z=0}. \tag{13.20}$$

Using Eq. 13.17 in Eq. 13.20, we get:

$$w_n = \frac{k\delta}{\mu(\phi^f)^2}\left(\frac{E'}{K} - \phi^f\right)\left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2}\right). \tag{13.21}$$

Equations 13.3 and 13.4 can be integrated for *u* and *v* with respect to *z* using boundary conditions (Eq. 13.14). Substituting *u* and *v* in Eq. 13.2 and integrating across the film thickness from $z = 0$ to $z = H$ with respect to *z* using boundary conditions (Eq. 13.15), we obtain modified Reynolds equation

$$\frac{\partial}{\partial x}\left\{\left[H^3 - \frac{24\delta k}{\left(\phi^f\right)^2}\left(\frac{E'}{K} - \phi^f\right)\right]\frac{\partial p}{\partial x}\right\} + \frac{\partial}{\partial y}\left\{\left[H^3 - \frac{24\delta k}{\left(\phi^f\right)^2}\left(\frac{E'}{K} - \phi^f\right)\right]\frac{\partial p}{\partial y}\right\} = -12\mu\frac{dH}{dt}, \tag{13.22}$$

for the pressure distribution in the joint cavity. For including roughness features, taking the stochastic average of Eq. 13.22 with respect to h_s, we get:

$$\frac{\partial}{\partial x}\left\{E\left[\left[H^3 - \frac{24\delta k}{\left(\phi^f\right)^2}\left(\frac{E'}{K} - \phi^f\right)\right]\frac{\partial p}{\partial x}\right]\right\} + \frac{\partial}{\partial y}\left\{E\left[\left[H^3 - \frac{24\delta k}{\left(\phi^f\right)^2}\left(\frac{E'}{K} - \phi^f\right)\right]\frac{\partial p}{\partial y}\right]\right\} = -12\mu\frac{dE(H)}{dt}, \tag{13.23}$$

where expectancy operator $E(\bullet)$ is defined by:

$$E(\bullet) = \int_{-\infty}^{\infty} (\bullet) f(h_s) dh_s, \tag{13.24}$$

f is the probability density function of the stochastic film thickness h_s. In most of the problems, surfaces show a roughness height distribution which is Gaussian in nature. Therefore, polynomial form which approximates the Gaussian is chosen in

the analysis. Such a probability density function is [12]:

$$f(h_s) = \begin{cases} \dfrac{35}{32c^7}\left(c^2 - h_s^2\right)^3, & -c < h_s < c \\ 0, & \text{elsewhere} \end{cases}$$

where $c = 3\sigma_1$ and σ_1 is the standard deviation.

13.4.1 Longitudinal Roughness

In the context of rough surfaces, there are two types of roughness patterns, which are of special interest. The one-dimensional longitudinal structure where the roughness has the form of long narrows ridges and valleys running in the *x*-direction and the one-dimensional transverse structure where roughness striations are running in the *y*-direction in the form of narrow and valleys. However, the present study is restricted to one-dimensional longitudinal roughness since the one roughness structure can be obtained from other by just rotation of the coordinate axes. For the one-dimensional longitudinal roughness pattern, the film thickness assumes the form:

$$H = h_0(t) + h_s(x, \xi),$$

then, modified *Stochastic Reynolds equation* (Eq. 13.23) takes the form:

$$\frac{\partial}{\partial x}\left\{\left[E(H^3) - \frac{24\delta k}{\left(\phi^f\right)^2}\left(\frac{E'}{K} - \phi^f\right)\right]\frac{\partial E(p)}{\partial x}\right\} +$$
$$\frac{\partial}{\partial y}\left\{\left[\frac{1}{E\left(1/H^3\right)} - \frac{24\delta k}{\left(\phi^f\right)^2}\left(\frac{E'}{K} - \phi^f\right)\right]\frac{\partial E(p)}{\partial y}\right\} = -12\mu\frac{dE(H)}{dt}. \quad (13.25)$$

13.4.2 Transverse Roughness

In this structure, the roughness striations are running in the *y*-direction in the form of narrow ridges and valleys. In this case, the film thickness takes the form:

$$H = h + h_s(y, \xi)$$

and modified Reynolds equation becomes:

$$\frac{\partial}{\partial x}\left\{\left[\frac{1}{E\left(H^{-3}\right)} - \frac{24k\delta}{(\phi^f)^2}\right]\frac{\partial E(p)}{\partial x}\right\} + \frac{\partial}{\partial y}\left\{\left[E(H^3) - \frac{24k\delta}{(\phi^f)^2}\right]\frac{\partial E(p)}{\partial y}\right\} = -12\mu\frac{dE(H)}{dt} \quad (13.26)$$

For the distribution function given by Eq. 13.24, we have:

$$E(H) = h. \tag{13.27}$$

$$E(H^3) = h^3 + \frac{hc^2}{3} \tag{13.28}$$

$$E(1/H^3) = \frac{35}{32c^7}\left[3\left(5h^2 - c^2\right)\left(c^2 - h^2\right)\log\left(\frac{h+c}{h-c}\right) + 2ch\left(15h^2 - 13c^2\right)\right] \tag{13.29}$$

Therefore, Reynolds equation (Eq. 27) for longitudinal roughness takes the form:

$$\frac{\partial}{\partial x}\left\{\left[E(H^3) - \frac{24k\delta}{(\phi^f)^2}\left(\frac{E'}{K} - \phi^f\right)\right]\frac{\partial E(p)}{\partial x}\right\} + \frac{\partial}{\partial y}\left\{\left[\frac{1}{E(H^{-3})} - \frac{24k\delta}{(\phi^f)^2}\left(\frac{E'}{K} - \phi^f\right)\right]\frac{\partial E(p)}{\partial y}\right\} = -12\mu\frac{dh}{dt}. \tag{13.30}$$

Introducing the following nondimensional parameters and variables, in Eq. 13.30:

$$\bar{x} = \frac{x}{a},\ \bar{y} = \frac{y}{a},\quad \bar{H} = \frac{H}{h_0} = 1 + \frac{\bar{x}^2 + \bar{y}^2}{2R_1h_1} + \bar{h}_s,\ R_1 = \frac{R}{a},\ h_1 = \frac{h_0}{a}, \bar{h}_s = \frac{h_s}{h_0},$$
$$\bar{k} = \frac{k\delta}{h_0^3},\ \bar{p} = \frac{E(p)h_0^3}{\mu_0^2 dh/dt},\ c = \frac{C}{h_0},$$

we get

$$\frac{\partial}{\partial \bar{x}}\left\{\left[E(\bar{H}^3) - \frac{24\bar{k}}{(\phi^f)^2}\left(\frac{E'}{K} - \phi^f\right)\right]\frac{\partial \bar{p}}{\partial \bar{x}}\right\} + \frac{\partial}{\partial \bar{y}}\left\{\left[E(\bar{H}^{-3}) - \frac{24\bar{k}}{(\phi^f)^2}\left(\frac{E'}{K} - \phi^f\right)\right]\frac{\partial \bar{p}}{\partial \bar{y}}\right\} = -12. \tag{13.31}$$

Here:

$$E(\bar{H}^3) = \bar{h}_0^3 + \frac{\bar{h}C^2}{3}$$
$$E(\bar{H}^{-3}) = \frac{35}{32C^7}\left(3(5\bar{h}^2 - C^2)(C^2 - \bar{h}^2)log\left(\frac{\bar{h}+C}{\bar{h}-C}\right) + 2C\bar{h}(15\bar{h}^2 - 13C^2)\right),$$

and for pressure distribution, the boundary conditions are:

$$\begin{aligned} \bar{p} &= 0 \text{ at } \bar{x} = \pm 1, \\ \bar{p} &= 0 \text{ at } \bar{y} = \pm 1. \end{aligned} \tag{13.32}$$

13.5 Wavelet-Multigrid Method

Since Eq. 13.32 is an elliptic equation with variable coefficients, it is difficult to solve analytically; we propose to solve it numerically using recently developed wavelet-multigrid method [20]. Using the first order finite difference scheme for the derivatives in Eq. 13.32, we write:

$$A_0 \bar{p}_{i+1,j} + A_1 \bar{p}_{i-1,j} + A_2 \bar{p}_{i,j+1} + A_3 \bar{p}_{i,j-1} + A_4 \bar{p}_{i,j} = A_5, \tag{13.33}$$

where

$$\begin{aligned} &A_0 = \frac{\alpha_{i+1/2,j}}{(\Delta x)^2},\ \mathrm{A}_1 = \frac{\alpha_{i-1/2,j}}{(\Delta x)^2},\ \mathrm{A}_2 = \frac{\beta_{i,j+1/2}}{(\Delta y)^2},\ \mathrm{A}_3 = \frac{\beta_{i,j-1/2}}{(\Delta y)^2}, \\ &A_4 = -(A_0 + A_1 + A_2 + A_3),\ \mathrm{A}_5 = -12, \\ &\alpha_{i,j} = E(\bar{H}^3_{i,j}) - \frac{24\bar{k}}{(\phi^f)^2}\left(\frac{E'}{K} - \phi^f\right),\ \beta_{i,j} = E(\bar{H}^{-3}_{i,j}) - \frac{24\bar{k}}{(\phi^f)^2}\left(\frac{E'}{K} - \phi^f\right). \end{aligned}$$

Let the matrix formulation of Eq. 13.33 be:

$$A\mathbf{x} = \mathbf{y}. \tag{13.34}$$

By considering D-4 family of wavelets, the corresponding transform matrix (just for illustrative purpose) is:

$$W = \frac{1}{\sqrt{2}} \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & 0 & 0 & 0 & 0 \\ d_0 & d_1 & d_2 & d_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_0 & c_1 & c_2 & c_3 & 0 & 0 \\ 0 & 0 & d_0 & d_1 & d_2 & d_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & c_0 & c_1 & c_2 & c_3 \\ 0 & 0 & 0 & 0 & d_0 & d_1 & d_2 & d_3 \\ c_2 & c_3 & 0 & 0 & 0 & 0 & c_0 & c_1 \\ d_2 & d_3 & 0 & 0 & 0 & 0 & d_0 & d_1 \end{bmatrix}, \tag{13.35}$$

where c_i and d_i, 's are the filter coefficients [21].

Now the fast wavelet transform (FWT) is performed on A and y of Eq. 13.35 recursively till the coarsest level is reached at level $-J$. Then matrix equation $\hat{A}_l \hat{\mathbf{x}}_l = \hat{\mathbf{y}}_l$ is used to obtain $\hat{\mathbf{x}}_l$ at the coarsest level using the Gauss–Jordan method.

Finally, integer wavelet transform (IWT) is performed on $\hat{\mathbf{x}}$ to obtain $\hat{\mathbf{x}}_0$ level -1 which gives the distribution of fluid film pressure $\bar{p}$. In the calculation, for all numerical experiments D-4 wavelets are employed. However, one has the freedom and flexibility to choose any wavelet basis. To test the accuracy, the problem is solved at resolutions 2^4 and 2^8. It is observed that there is marginal increase in the accuracy of the solution; better accuracy can be achieved by increasing the resolution and/or the order of the wavelet family. It is also observed that the amount of computational effort is considerably less than that of classical multigrid method.

Once we have obtained the fluid film pressure by using the wavelet-multigrid method, the load carrying capacity can be evaluated straightforwardly. We can find the nondimensional load carrying capacity $\bar{W}$ per unit area of the joint surface using the following formula.

$$\bar{W} = \int_{-1}^{1}\int_{-1}^{1} \bar{p}(\bar{x},\bar{y})d\bar{x}d\bar{y}. \tag{13.36}$$

13.6 Results and Discussion

A simplified mathematical model has been developed for the study of effect of surface roughness on the poroelastic bearing by modeling articulate cartilage as poroelastic matrix and synovial fluid as a linearly viscous fluid. The problem considered is that of the squeeze film lubrication between the rough spherical indenter and smooth poroelastic material. The modified Reynolds equation incorporates the constitutive equations of poroelastic matrix. This Reynolds equation is solved using wavelet-multigrid method. The values of $E', K,$ and $\bar{k}$ are taken from Torzilli [6], which are associated with healthy human cartilage with normal functioning. The parameters $\bar{k}$ and ϕ^f are chosen in such way that the load-carrying capacity that can be sustained by the joints should be more. The pressure $\bar{p}$ and $\bar{w}$ are functions of nondimensional parameters C, $\bar{k}$, and $\bar{H}$ and constant parameters ϕ^f and E'/K. As the radius of the upper rigid indenter increases, it becomes relatively flat and uniform, which leads to increase in the area of the film region. This wide film area avoids exit of lateral fluid and is responsible for retaining the large amount of fluid which enhances the pressure and the load-carrying capacity. For large radius, the spherical indenter becomes relatively flat and thus the study reduces to the squeeze film lubrication between two parallel plates.

The distribution of pressure $\bar{p}$ with rectangular coordinates $\bar{x}$ and $\bar{y}$ is given in Figs. 13.1a and 13.1b. For $C=0.3$, the roughness effects are more pronounced when compared with $C=0.1$. In Fig. 13.2, it is observed that the effect of roughness is to increase the pressure distribution increasing values of C. The roughness increases the pressure in the film region compared with the smooth case. This is because, the presence of surface asperities reduces the fluid flow and therefore large fluid accumulates in the film region, which enhances the pressure built up.

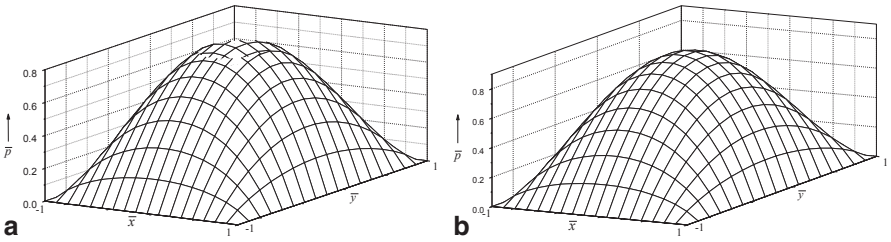

Fig. 13.1 **a** Fluid film pressure distribution for $C=0.3$ with $\overline{k} = 7.65\times10^{-5}$, $\phi^f = 0.8$ and $\mathbf{E'/K = 1.0}$ **b** Fluid film pressure distribution for $C=0.1$ with $\overline{k} = 7.65\times10^{-5}\phi^f = 0.8$ and $\mathbf{E'/K = 1.0}$

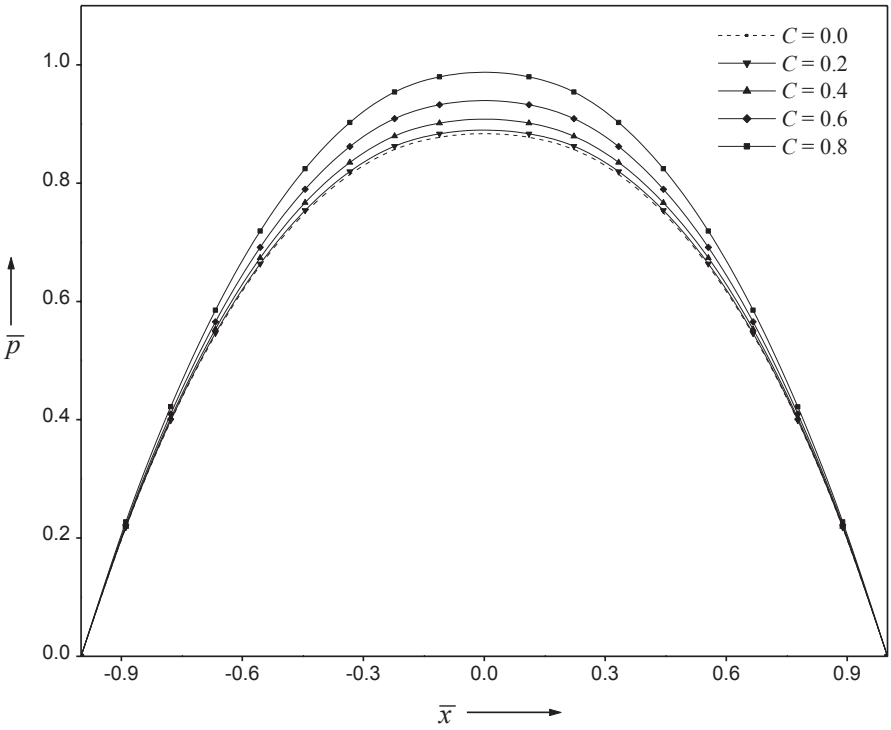

Fig. 13.2 Variation of pressure $\overline{p}$ with $\overline{x}$ for different values of C with $\bar{\mathbf{k}} = \mathbf{7.65\times10^{-5}}, \mathbf{f^f = 0.8}$ and $\mathbf{E'/K = 1.0}$

The variation of load $\overline{w}$ that can be sustained by a joint, with roughness parameter C for different elastic parameter E'/K is shown in Fig. 13.3. The load-carrying capacity increases with increasing E'/K for different roughness parameters C. Presence of hyaluronic acid, molecular, and other molecular weight substances in the synovial fluid results in the formation of a thick dense substance on the cartilage

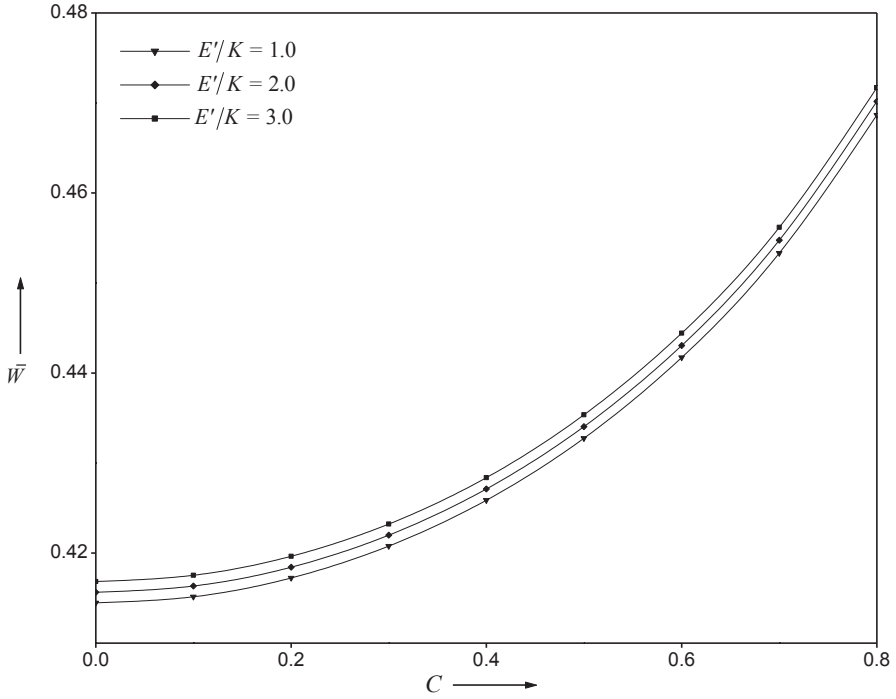

Fig. 13.3 Variation of load $\bar{\mathbf{W}}$ with C for different values of $\mathbf{E}'/\mathbf{K}$ with $\bar{\mathbf{k}} = \mathbf{7.65}\times\mathbf{10}^{-5}, \mathbf{f}^{\mathbf{f}} = \mathbf{0.8}$

surface during squeezing action. This leads to increase in the pressure distribution and in turn increase in the load-carrying capacity. This trend is preserved for all values of E'/K.

13.7 Summary

Wavelet-multigrid method is found to be accurate for the solution of modified Reynolds equation, which incorporates surface roughness and poroelastic nature of articular cartilage. The method is proven to be more efficient than the classical multigrid method. Comparing the rates of convergence of the two methods, it is found that Wavelet-multigrid method converges faster compared with the multigrid scheme. It provides the ability to carry out calculations focusing only in regions where it is needed. It can as well be generalized to nonuniform grids, which is a unique feature of this approach. In classical scheme we find that as the grid size decreases the condition number increases; a serious disadvantage whereas in the present Wavelet-multigrid scheme the conditioning of the matrix is incorporated in the procedure itself without requiring separate analysis for it. It is observed that

the effect of roughness is to increase the pressure built up (and hence the load-carrying capacity) compared with the classical case. The governing equations describing complex structure of cartilage and synovial fluid are complicated because of nonlinearity and joints having a wide range of articulating features. However, the proposed simplified model does predict some of the salient features of bearing characteristics, which would enable the biomedical engineers in selecting suitable design parameters. In addition to being of scientific and medical interest in its own right, the present study has achieved the objective of gaining deeper understanding of the lubrication in the knee joint which is of crucial importance in the total knee-replacement technologies. The results obtained could guide the experimentation with new material for knee replacement with mechanical characteristics analyzed here. Currently, the most promising compliant materials are hydro gels, which have similar properties to those of natural cartilage.

References

1. Torzilli PA (1978) The lubrication of human joints: a review. In: Fleming DG (ed) Handbook of engineering in medicine and biology. CRC Reviews Boca Raton
2. Mow VC, Kuei SC, Lai WM, Armstrong CG (1980) Bi-phasic creep and stress relaxation of articular cartilage in compression: theory and experiments. Trans ASME J Biomech Eng 102:73–84
3. Biot MA (1941) General theory of three-dimensional consolidation. J Appl Phys 12:155
4. Mansour JM, Mow VC, Holmes MH (1973) Fluid flow through aticular cartilage during steady joint motion. Proceedings of the National Engineering and Bioengineering Conference, p. 183
5. Hou JS, Mow VC, Lai WM, Holmes MH (1992) An analysis of the squeeze film of the squeeze-film lubrication mechanism for articular cartilage. J Biomech 25:247
6. Torzilli PA, Mow VC (1976) On the fundamental fluid transport mechanics through normal pathological articular cartilage during function-II. J Biomech 9:587
7. Forster H, Fisher J (1996) The influence of loading time and lubricant on the friction of articular cartilage. Proc Inst Mech Eng 210:109
8. Collins RJ (1982) A model of lubricant gelling in synovial joints. ZAMP 33:93–123
9. Ateshian GA, Wang H, Lai WM (1998) The role of interstitial fluid pressurization and surface porosities on the boundary friction of articular cartilage. J Tribol 120:241
10. Jin ZM, Dowson D, Fisher J (1992) The effect of porosity of articular cartilage on the lubrication of a normal human hip joint. Proceedings of the institution of mechanical engineers, Part H. J Eng Med 206:117
11. Hlavacek M (2000) Squeeze-film lubrication of the human ankle joint with synovial fluid filtrated by articular cartilage with the superficial zone wornout. J Biomech 33:1415
12. Ateshian GA, Lai WM, Zhu WB, Mow VC (1994) An asymptotic solution for the contact of two biphasic cartilage layers. J Biomech 27(11):1347–1360
13. Barry SI, Holmes M (2001) Asymptotic behavior of thin poroelastic layers. IMA J Appl Math 66:175
14. Mercer GN, Barry SI (1999) Flow and deformation in poroelasticity-II. Numerical method. Math Comp Model 30:31.
15. Sayles RS, Thomas TR, Anderson J, Haslock I, Unsworth A (1979) Measurement of surface microgeometry of articular cartilage. J Boimech 12:257
16. Christensen H (1969) Stochastic model for hydrodynamic lubrication of rough surfaces. Proc Inst Mech Eng Pt-I, 184:1013

17. Walker PS, Erkman MJ (1972) Metal-on-metal lubrication in artificial human joints. WEAR 21:377
18. Maroudas A (1975) Biophysical chemistry of cartilaginous tissues with special reference to solute and fluid transport. Biorheology 12:233–248
19. Dowson D, Jin Z. M (1992) Microelastohydrodynamic lubrication of low-elastic-modulus solids on rigid substrates. J Phys D Appl Phys 25 A116–A123
20. Ateshian GA, Huiqun Wang (1995) A theoretical solution for the frictionless rolling contact of cylindrical biphasic articular cartilage layers. J Biomech. 28(11):1341–1355
21. Bujurke NM, Salimath CS, Kudenatti RB, Shiralashetti SC (2007) A fast wavelet-multigrid method to solve elliptic partial differential equations. Appl Math Comp 185:667–680

Chapter 14
A Basic Concept on Modelling Soil Organic Carbon

Nimai Senapati, Subhadip Ghosh, Heiko Daniel and Amitava Rakshit

14.1 Introduction

Soil organic carbon (SOC) is a complex and mixture of diversified materials ranging from rapidly decomposable plant parts to microorganisms and make up a vital part of the soil. Enumeration of SOC dynamics is becoming more important as this can greatly impact soil productivity and sustainability. The amount of SOC is a balance between the build-up from inputs of new plants and animal materials and the constant losses, where the carbon (C) decomposes and the constituents separate to mineral nutrients and gases, or are washed or leached away. Positive build-up of SOC levels is possible when there is abundance of water, nutrients and sunlight, whereas continuous fallow favours SOC loss.

There are five principal pools which regulate the C cycle on earth—oceanic pool (~38,000 Pg, $Pg = 10^{12}$ kg $= 10^{15}$ g $=$ Gt), followed by the geologic (~5,000 Pg), pedologic/soil (~2,500 Pg to one meter depth, organic and inorganic), atmospheric

N. Senapati (✉)
Institut National De La Recherche Agronomique (INRA), Poitou-Charentes, URP3F, 86600 Lusignan, France
e-mail: nimai_senapati@yahoo.com

S. Ghosh
Centre for Urban Greenery and Ecology, National Parks Board, Singapore, 259569 Singapore
e-mail: subhadip_ghosh@nparks.gov.sg

S. Ghosh · H. Daniel
School of Environmental and Rural Sciences, University of New England, 2351 Armidale, NSW, Australia
e-mail: subhadip_ghosh@nparks.gov.sg

H. Daniel
e-mail: hdaniel@une.edu.au

A. Rakshit
Department of Soil Science and Agricultural Chemistry, Institute of Agricultural Science, Banaras Hindu University, Varanasi, 221005 Uttar Pradesh, India
e-mail: amitavar@bhu.ac.in

S. K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_14,

(~760 Pg) and biotic pool (~560 Pg) [1–2]. Soils contain about 1,500 Pg of organic carbon globally in top 1 m, which is roughly twice the amount of C in the atmosphere as carbon dioxide (CO_2) and about three times the amount of C in vegetation [1, 3–4]. Although the SOC pool is small compared with that of the oceans, potentially it is much more labile in the short term [1]. Using respective estimates of 1,500 and 720 Pg of C in soil and the atmosphere, and an atmospheric concentration of 390 ppm for CO_2 a 1 % change in the amount of C stored in soils would equate approximately to an 8 ppm change in atmospheric CO_2 concentration, provided all other components of the C cycle remained constant [5]. Thus, small change in flow of C into or out of soil C pool could have dramatic impact on a global scale [6].

Diffusive process plays an important role in SOC dynamics through mass and gaseous movement and their transportation. The major factors controlling the size of the SOC pool and its movement are land use, land use change (LUC), climate, soil, management practice and technology [2, 6–8]. Agricultural soils are known to be C depleted due to tillage, disturbance, aeration and mineralization [9–10]. However, there is a considerable potential to sequester and store additional C by modifying soil management practices in such soils. Accurate quantification of inputs of C into and outputs of C from soil is essential to assess C sequestration potential and changes in SOC with time. Such measurements will help unravel the mechanisms that control C storage and formulation of robust models of soil C dynamics and turnover. Understanding the response of the soil C reserve and its transportation to change in different factors is of critical importance. The SOC turnover models are able to simulate SOC dynamics under various land uses (forest, grassland, cropland, etc.), management practices (crop rotation, ploughing, fertilizer application, irrigation, stubble management, soil amelioration, etc.), technological improvement (varietal improvement, better machinery, improved agronomic knowledge, etc.) and climatic conditions [8, 11–14]. Thus, they could help in investigation of change in SOC dynamics under different scenarios and may also help in refining our understanding of SOC turnover processes by pointing out our knowledge gap.

14.2 Soil Organic Carbon Models

A model is a simplified representation of reality, designed to meet different objectives [15]. A model may be a physical miniature of real entity, such as a clay sculpture, a hypothetical description of a complex entity or process, or a schematic/mathematical representation of a real-world system, theory, phenomenon, or situation.

SOC models represent turnover or decomposition of SOC as well as its transformation in soil–plant–atmosphere system. One of the earliest SOC models [16] to describe C accumulation or loss from soil was:

$$\frac{dx}{dt} = -kx + A,$$

where dx = change in state variable (e.g. soil C)

k = first order rate constant

A = addition rate (mass t^{-1}) which is independent of loss and the amount present

Since then many SOC models have been developed to meet different specific objectives.

14.3 Application of Soil Organic Carbon Models

The SOC turnover models are widely used in many disciplines—soil science, agronomy and environmental sciences. These models help to improve our understanding of C turnover process in soil as well as underlying C stabilization mechanisms in soil [13–14, 17]. At the same time, the models are now even being extensively used to extrapolate our understanding of SOC dynamics both temporally (future projections) [8, 12] and spatially, i.e. from national [18], regional [19] to global scales [20]. Another important application of SOC models is in agronomy, through its incorporation into decision support systems to improve agronomic efficiency and environmental quality, e.g. APSIM [21], DSSAT-CSM [22], SUNDIAL-FRS [23], etc.

The SOC models can be used to explore and investigate SOC dynamics under different management and environmental scenarios even beyond the realm of experimental work [8, 12]. They can be used for interpolation and extrapolation of experimental data both on time and space, and also making projections of SOC behaviour under current and future environmental and management conditions [14]. The SOC models are now increasingly being used by the policymakers also at the national, regional and global scales, for example, in post-Kyoto debate on capacity of terrestrial ecosystem to store C [24].

14.4 Classification of Soil Organic Carbon Models

SOC turn-over models are broadly classified as:

a. Process-oriented models
b. Organism-oriented/food-web model
c. Cohort model

These models are described below.

14.4.1 Process-Oriented Model

Process-oriented models focus on the processes controlling the movement and transformation of matter or energy. Process-based models can be divided into:

a. Single compartmental model where SOC models consider SOC as one homogeneous compartment. [16]
b. Two compartmental models where SOC is considered as two compartments. [25]
c. Multi-compartmental model considers SOC multi-compartmental, e.g. RothC [26], CENTURY [27], etc.

The first two types of SOC models are mostly static, where the environmental variables remain constant, whereas the models under type (c) are mostly dynamic, where the environmental variables vary with time.

Most of the SOC models are process-based multi-compartmental, describing SOC as a finite number of compartments or SOC pools, each of which is homogeneous, well mixed and characterized by its position in the model's structure and its decay rate [13–14, 28]. Decay rates are usually expressed by first-order kinetics with respect to the concentration (C) of the pool as:

$$\frac{dC}{dt} = -kC,$$

where t is the time, k is the rate constant of first-order kinetics. The pool's half-life ($h = (\ln 2)/k$), or its turnover time ($1/k$) is sometimes used instead of k to characterize a pool's dynamics.

With a lower decay rate constant, the stability of the organic pool is higher with higher half-life and turnover time. The compartments interact by exchanging materials, and by exchanges with the environment [29]. The flow of C within most models represent a sequence of C going from plant and animal debris to the microbial biomass, and then to soil organic pools of increasing stability. The output flow from an organic C pool is usually split into a microbial biomass pool, another C pool, and under aerobic condition into atmospheric CO_2. Two parameters control the split flow, viz., microbial efficiency and stabilization/humification factor, which control the flow of decayed C to the biomass and humus pools, respectively. The sum of the efficiency and humification factors must be lower than one to account for the release of CO_2 [14].

Most of the present day's models are multi-compartmental process-based models. For example, out of the 33 models, currently available within the Global Change and Terrestrial Ecosystems (GCTE), and Soil Organic Matter Network (SOMNET) database [30], 30 are multi-compartmental process-based models.

14.4.2 Organism-Oriented/Food-Web Model

Organism-oriented models focus on flow of energy and matter (C and N) through food webs of soil organisms [31–32]. Such models are dynamic and explicitly account for different trophic levels or functional groups of soil biota. Food-web models require a detailed knowledge of the biology of the system to be simulated and are usually parameterized for application at specific sites.

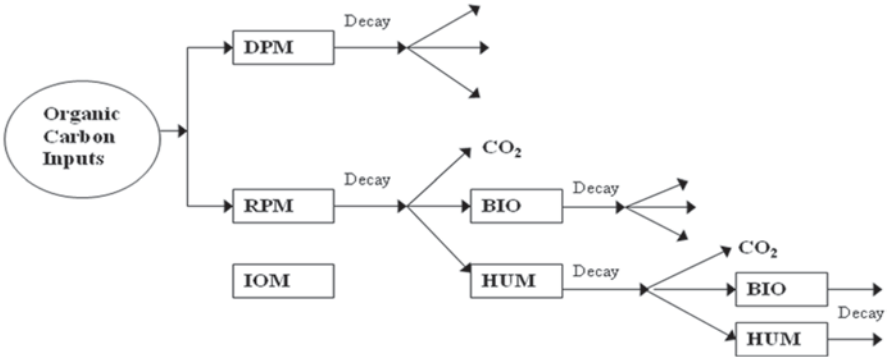

Fig. 14.1 Structure of the Rothamsted carbon model (*RothC*) (after [26]). *DPM* Decomposable plant material, *RPM* Resistant plant material, *IOM* Inert organic matter, *BIO* Microbial biomass, *HUM* Humified organic matter

14.4.3 Cohort Model

This type of model considers each fresh addition of plant debris as a separate cohort, which decays in a continuous way. These models describe decomposition as continuum. Most of such models are dynamic models and account SOC as one pool, which decays with a feedback loop into itself. The SOC pool is divided into an infinite number of components, characterized by its quality with respect to degradability as well as impact on the physiology of the decomposers. One of the examples of this type of model is Q-SOIL [33]. This model is quality dependent and is represented by a single rate equation which represents the dynamics of each SOC component. Exact solutions to the rate equations are obtained analytically [34].

Among the three types of models discussed, process-based multi-compartmental dynamic models are most popular due to (a) ease of use and transferability, (b) their successful coupling with GIS software, (c) their different internal parameters are easier to estimate and calibrate for specific purposes, and (d) they are easily scaled up. Different compartments actually simplify very complex physical, biological and chemical characteristics of SOC, and thus help to understand and depict the underlying turnover process [14]. Figures 14.1 and 14.2 show the structure of the two most popular process-based multi-compartmental models, viz., Rothamsted carbon model (RothC) [26] and CENTURY model [27]. A detailed overview on the SOC turnover models could be found elsewhere [13–14, 35].

14.5 Factors Affecting Turnover of Soil Organic Carbon in Models

Decomposition rate constants of different compartments or SOC pools follow the first-order kinetics and are constant for a given set of biotic and abiotic conditions. For nonoptimum environmental circumstances, the maximum value of rate constant

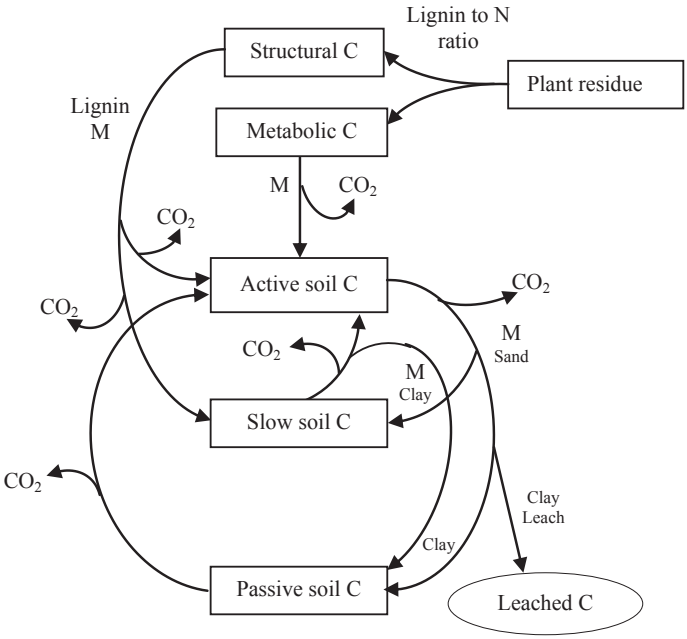

Fig. 14.2 A simple structural diagram of the SOC sub-model of the CENTURY model. (*M* : Multiplier for effect of moisture, temperature, cultivation, *Leach* : H_2O leached below 30 cm)

(*k*) is modified by multiplication with a reduction factor μ —ranging from 0 to 1. Environmental factors considered by SOC models for the modification of *k* include soil temperature, water, oxygen, pH, nitrogen, clay content, cation exchange capacity, salinity, type of vegetation, plant cover and tillage [13–14, 36].

14.6 Initialization of SOC Models

Model initialization, the initial distribution of total SOC between model pools, has been described as the most crucial part of a simulation study as it can influence the model prediction of soil C stock [37–38]. Incorrect or flawed initialization potentially leads to improper assessment of interannual variability, and can also produce fallacious trends in output as the state variables drift back towards the model ideal [38]. As various pools in SOC models are based on qualitative concepts and often do not correspond to measureable fractions [28, 39], the partition of SOC pools may lead to an initialization problem. Thus, the model may reflect not only the effect of the modelling objectives, but also the initialization procedure. The most common approach to solve the initialization problem is to achieve the initial SOC pool distribution by a spin-up run of the model, i.e. a run of the model over several hundreds to thousands of years to find equilibrium/steady state SOC, assuming initial SOC

and its distribution among the model pools are at equilibrium/steady state with the current land use and climatic conditions [8, 40]. The observed soils, however, may not be at equilibrium due to disturbances such as fire, erosion, land use and management changes [41]. The equilibrium/steady state assumption for the ecosystem C cycle has been challenged [42] and its limitation in modelling approaches emphasized [41].

In the second method, where recent land use and management condition is known, the model is started with default or steady-state pool distribution and then run for a few decades to obtain a quasi-steady state distribution under recent conditions [43]. However, there is rarely enough information about the past condition for a large area to initialize the model correctly. A third approach is to run most pools to steady state and then adjust the slowest pool so that the total SOC content matches the measured value [41]. Rapidly turnover pools reflect the current and recent state of the system, while slowly turnover pools reflect historical condition. Additionally, a fourth method is using the statistical model-data fusion technique to match pool distribution with observed C dynamics during the period of model simulation [38, 44]. Another method of initialization of SOC model is initializing model with measured SOC pools.

The use of measured SOC pools has the advantage over other initialization methods, as field measurements are independent of the model or any assumption, and information on previous land use and management history are not required [45]. Any process that is ignored in the model, but which could influence SOC dynamics, is also taken into account in SOC partitioning. Thus, measured fractions reflect better the real site-specific conditions under which SOC is accumulated [45–46]. Once the SOC pools are measured by a reliable fractionation method, model initialization issues can be solved and model performance can also be improved.

14.7 Measured SOC Fractions and Conceptual Modelled Pools

Soil organic matter represents a chemically and spatially heterogeneous mixture of organic materials that exist along a continuum of various stage of decomposition [47]. Thus, SOC cannot be separated easily into different pools, rather it exists as one continuous pool with almost infinite range of turnover times from minutes to millennia [48]. However, to simulate SOC dynamics successfully, most of the process-based SOC models divided SOC into multiple conceptual pools defined by their sizes and turnover times which are governed by first-order rate constants and modified by different climatic and edaphic factors [13–14, 28, 38]. A pool may be defined as a compartment containing material that is chemically indistinguishable and equally accessible to plants or to the soil microbial population [49]. A measured fraction represents a pool when it is unique (i.e. it should not represent only a portion of chemically and biologically indistinguishable material) as well as non-composite [39]. On the other hand, some other researchers [48] have given importance

to the mechanisms and process of SOC stabilization in measured fractions/pools and for modelling SOC dynamics.

The SOC turnover in models has been described using simplified, deterministic, multi-compartmental simulation that is primarily based on empirically derived relationship. A major limitation of process-based multi-compartmental SOC models is that the various model compartments/pools are conceptual and often do not correspond to measureable fractions [39, 45, 50]. The only measured quantity is the microbial biomass and total SOC. Thus, it is often not possible to validate the dynamics of the modelled SOC pools with measured pools. However, measured pools in model have the advantage of thorough model evaluation and detailed initialization [39].

During the last few decades, different chemical methods were proposed to separate SOC into different pools [51–53]. The wet chemical fractionation methods consider the chemical nature of SOC as the only factor to influence the turnover rate of SOC pools, soil protection on SOC is ignored. Protected C is the influence of spatial arrangement of SOC in soil matrix (primary organo-mineral complex) and soil structure or aggregate (secondary organo-mineral complex) on stabilization/decomposition mechanism of SOC [50, 54]. Microorganism needs physical access of soil C in order to use it. In situations where the structure of a system does not control the accessibility of soil C to microbes (for example, forest litter layer), there chemical composition is the sole factor controlling soil C dynamics [50]. Soil microbes generally consume most organic substrate fallen or deposited onto the soil, yet there is accumulation of organic C in the soil. There must be some protection of incoming organic C in the way that prevents it from being completely consumed.

In the last few decades, physical fractionation methods have got the importance over chemical methods in fractionating SOC [55–57]. Although physical SOC fractionation methods elucidate spatial arrangement of primary and secondary organo–mineral complexes on SOC stabilization, they do not consider chemical or biochemical stabilization mechanisms, i.e. they are unable to give importance to chemical or biochemical nature or composition of SOC on its stabilization.

Therefore, to capture all the SOC stabilization mechanisms together, various approaches in SOC fractionation schemes have combined physical and chemical methods together subsequently. The most common approach is density or size fractionation prior to chemical analysis. By separating light fraction (LF) or the sand-sized fraction (particulate organic matter, POC), which contains mainly fresh plant material, through density fractionation or particle size fractionation, rest of the fractions are then further characterized chemically by acid hydrolysis or base extraction [58–61] or oxidation with H_2O_2/NaOCl/$Na_2S_2O_8$ [62]. There are some physicochemical methods which have shown good promise to relate measured SOC fractions with SOC pools. Skjemstad and others [63] have proposed a physicochemical method where they suggested partitioning SOC into POC (particle size fractionation), charcoal carbon (photo-oxidation) and humus (Fig. 14.3). Similarly, Zimmermann and others [45] suggested a physicochemical fractionation method to partition SOC into free LF/POC, dissolved organic carbon (DOC), stable aggregate and sand associated C (S + A), silt and clay associated C (s + c), and inert organic

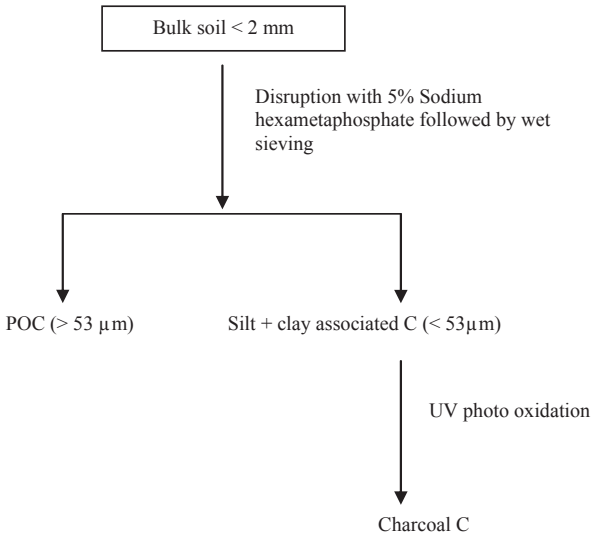

Fig. 14.3 Physicochemical SOC fractionation method POC particulate organic carbon, after [63])

carbon (rSOC) (NaOCl oxidation of silt + clay associated C) (Fig. 14.4). The concepts of relating these SOC fractions with model (RothC) conceptual pools are summarized in Figs. 14.5 and 14.6. Figures 14.3 and 14.5 show the partitioning of SOC into different fractions and relation of SOC fraction to SOC model (RothC) pools after [63]. POC is the measure of the resistant plant material pool (RPM), charcoal C is the measure of inert organic matter pool (IOM) and the quantity of humus pool (HUM) is calculated by HUM = total SOC–(POC + Charcoal carbon).

Similarly, Figs. 14.4 and 14.6 show the partitioning of SOC into different fractions and relation of SOC fraction to model pools after [45]. The measured C in rSOC fraction is directly associated with the IOM. The sum of C in POC and DOC fractions is split into decomposable plant material pool (DPM) and RPM using the ratio of DPM: RPM obtained by RothC model under equilibrium condition. The same procedure is also used to separate the sum of C in S + A fraction, and silt and clay fraction excluding rSOC fraction ($s + c$ –rSOC) into biomass pool (BIO) and HUM.

One of the most important issues in the SOC model is inert organic matter pool (IOM). The issue of the IOM and its representation in models has been discussed [28]. IOM is defined in the model as an inert pool which does not undergo decomposition, means that it is not a dynamic pool. Although, different chemical methods are able to produce an old, chemically recalcitrant and stable SOC fraction, they nevertheless are not completely inert [64]. Further research for harmonizing measured IOM fraction of SOC and the IOM pool may be necessary.

Process-oriented multi-compartmental models could be improved by modifying different pools/compartments so that they are based on measureable SOC fractions which are unique, non-composite and in accordance with different SOC stabilization mechanisms

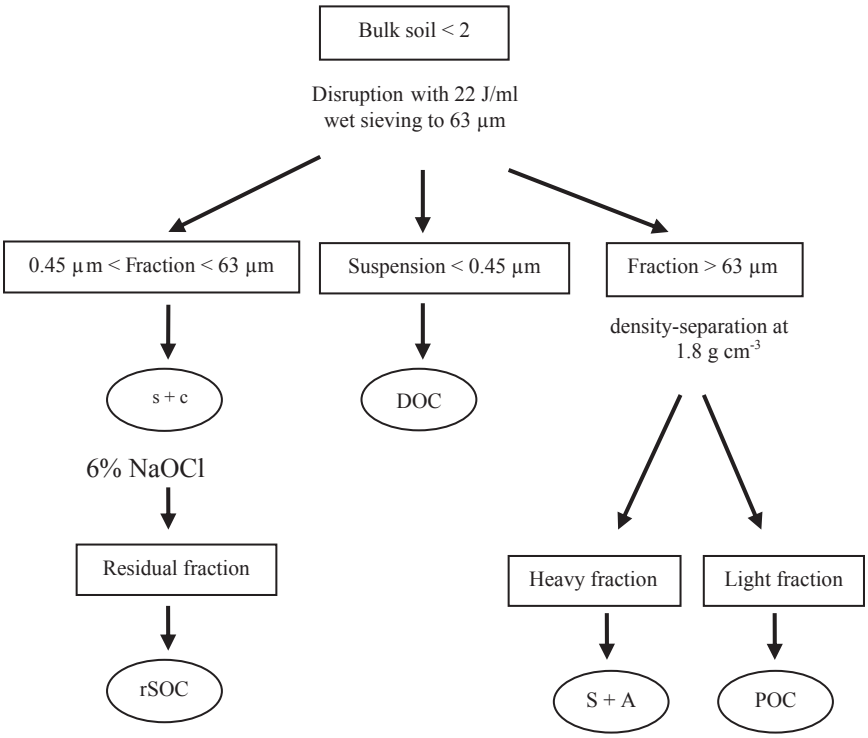

Fig. 14.4 Physicochemical SOC fractionation method. Particulate organic carbon (*POC*), Dissolved soil organic carbon (*DOC*), Soil organic carbon in sand and stable aggregate (*S + A*), Chemically resistant soil organic carbon/inert organic carbon (*rSOC*), Soil organic carbon in silt and clay fraction (*s + c*), after [45]

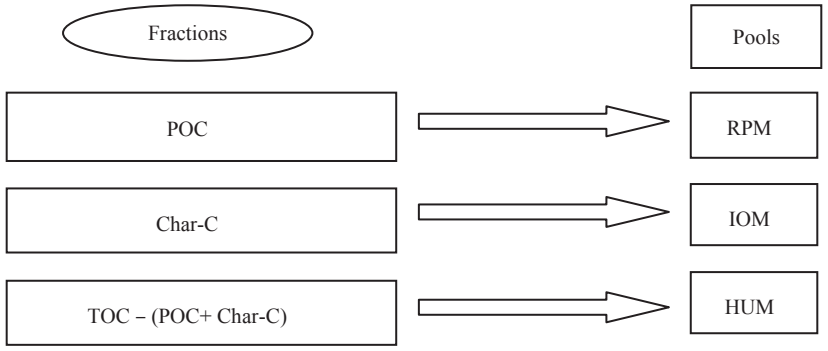

Fig. 14.5 The concept of relating measured soil organic carbon fractions to conceptual pools of RothC model *POC* particulate soil organic carbon, charcoal carbon (*Char-C*), total soil organic carbon (*TOC*), total soil organic carbon excluding POC and Char-C (*TOC (POC + Char-C)*), resistant plant material (*RPM*), inert organic matter (*IOM*) and humified organic matter (*HUM*), after [63]

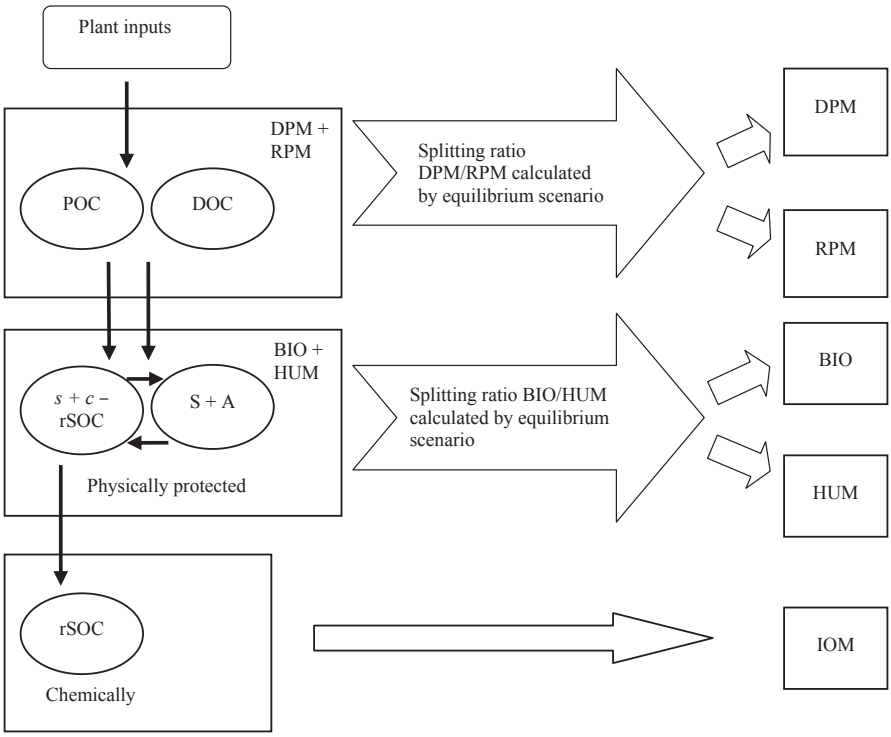

Fig. 14.6 The concept of relating measured soil organic carbon fractions to conceptual pools of RothC model particulate soil organic carbon (*POC*), dissolved soil organic carbon (*DOC*), soil organic carbon in sand and stable aggregate ($S + A$), soil organic carbon in silt and clay fraction excluding inert organic carbon ($s + c$–*rSOC*), chemically resistant soil organic carbon/inert organic carbon (*rSOC*), decomposable plant material (*DPM*), resistant plant material (*RPM*), microbial biomass (*BIO*), humified organic matter (*HUM*) and inert organic matter (*IOM*), *oval* represents fraction, *square* represents pool, after [45])

14.8 Black Carbon and Modelling Soil Organic Carbon

Black carbon (C) is an inorganic carbon compound produced from incomplete combustion of organic materials having a graphitic microstructure [65]. Black C is resistant to oxidation and microbial activity and so is persistent on a geological timescale [66]. Due to this recalcitrant nature, black C is assumed to represent part of the inert organic carbon pool. This pool is expected in higher proportion of total SOC in fire prone regions, for example at the continental scale, black C comprises 0–82 % of the total SOC in Australia [67]. The importance of black C in the simulation of SOC was demonstrated [67] and the result showed a reduction in simulated CO_2 emissions by 18.3 and 24.4 % in two Australian savannah regions in response to a warming of 3 °C over 100 years by including a realistic black C stocks in prediction models. Hence, quantification of the amount of black

C is important in SOC modelling. Historically, black C in soils has been calculated by digesting away organic matter using nitric acid and heat, and then sorting char particles by hand or under a microscope [68]. New analytical methods for estimating char materials in mineral soils include using benzenepolycarboxylic acids (BPDA) during nitric acid digestions as markers for black C [69], chemical digestion [70] and using a combination of high energy ultraviolet photooxidation and nuclear magnetic resonance spectroscopy [71] or a combination of chemical oxidation followed by solid-state 13°C nuclear magnetic resonance spectroscopy [72]. Schmidt and others [73] tested several forms of thermal oxidation, chemical oxidation, photo-oxidation, and a molecular marker method on Australian soils. The resulting black C values for individual sample varied over two orders of magnitude indicating great disparity between individual methods. One of the possible explanations for the wide range of the result is the different operational definitions of black C and clear cut boundaries of the different methods, which were developed for specific research questions.

14.9 Evaluation of SOC Models

Model evaluation is the prerequisite for model application and helps in the assessment of model performance [15]. The benefits of model evaluation are: (a) it shows how well a model can be expected to perform in a given situation, i.e. it determines the accuracy of the simulation, (b) it analyses the behaviour of model, whether the model responds in the expected way to the change in the conditions of simulation, (c) it determines which components of the model are most important in influencing the results, and (d) it can help in improving the understanding of the system as well as the model [13–15].

A graphical analysis gives a quick approximate evaluation, whereas a quantitative analysis determines the statistical accuracy of the simulation. Some statistics measure association (e.g. sample correlation coefficient) and others measure coincidence (root mean square error, relative error, mean difference, modelling efficiency, lack of fit, etc.) between the measured and simulated values. A higher association indicates similar trends between the measured and simulated values, whereas a higher coincidence indicates close proximity of the simulated values with the corresponding measured values. Figure 14.7 shows different model performances with different coincidences and associations between the measured and simulated values. A sensitivity analysis evaluates the behaviour of the model, whereas uncertainty analysis determines how much uncertainty is introduced in model output by each component of the model. All these model evaluation methods/parameters are defined and summarized in the literature [11, 15, 74].

Model performances are generally assessed using a long-term experimental data set. Models can be evaluated at different levels, viz., individual process level, subset process level (e.g. net mineralization), or the model's overall outputs (e.g. changes

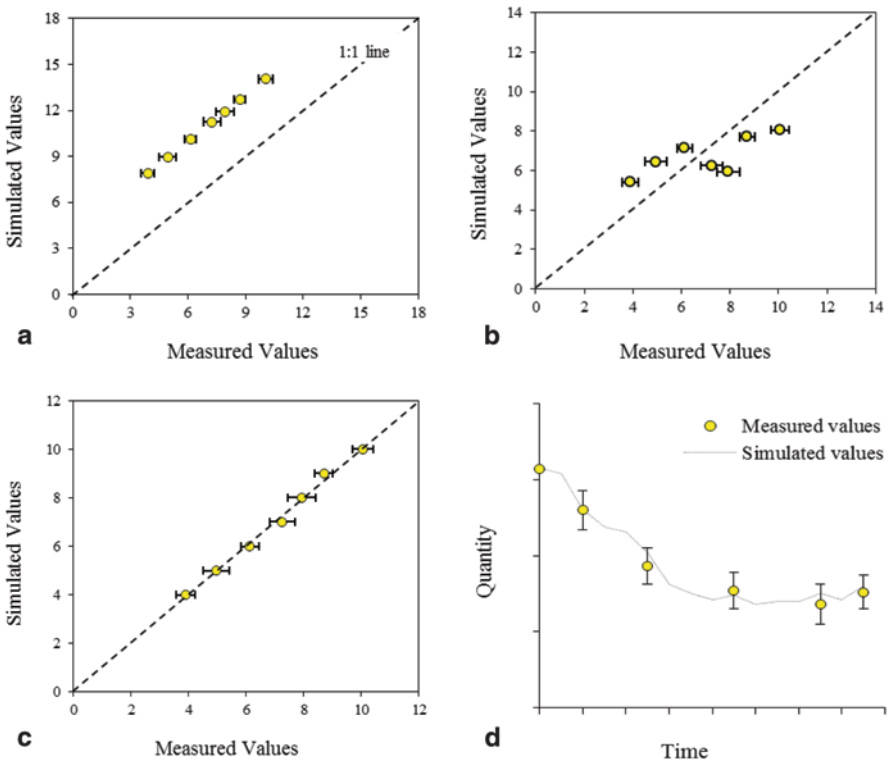

Fig. 14.7 An illustration of different hypothetical model performances with different coincidences and associations between measured and simulated values. **a** Higher association but lower coincidence, bad model performance. **b** Lower association but higher coincidence, bad model performance. **c** Higher association and coincidence, good model performance. **d** Higher association and coincidence in time scale, good model performance

in total soil organic matter (SOM) over time). Most comprehensive model evaluation to date has been done by Smith and others [11] with nine models using 12 data sets from seven long-term experiments covering arable lands, managed-unmanaged grasslands, and woodlands across the world.

14.9.1 Evaluation of SOC Models Under Land Use Change

LUC is a global concern and has adverse effects on the climate through emission of greenhouse gases [75]. Historically, soils have lost between 40–90 Pg C globally through cultivation and disturbance [76], and it is estimated that change in land use alone emitted 1.6 ± 0.8 Pg C yr^{-1} to the atmosphere during the 1990s [4, 77]. Expansion of the cropland is one of the major drivers of LUC and is likely to continue in

the future. The magnitude of changes of SOC following LUC vary widely among sites due to differences in environmental conditions (such as soil type, texture, mineralogy, etc.), climate, ecosystem productivity, plant species, and the management practices [7]. The prediction of changes in SOC as a consequence of rapid changes in land use is an important task for the future, and SOC dynamic models are one of the most promising tools in this respect.

Model performances are generally evaluated using the long-term field experimental data set, collected in chronosequence [11]. Model performance in simulation of changes in SOC under LUC can also be assessed using chronosequence data sets [78]. However, there are extremely few cases where chronosequence data have been collected before and after a LUC. In this respect, paired-sites offer an opportunity to study LUC effects on SOC and can provide a basis for SOM model evaluation. Paired-site consists of two sites, viz., a site where LUC already occurred several years before and an adjacent remnant, native-vegetation site, located on a common soil type, landscape position, and slope angle.

14.10 Projection of Soil Organic Carbon Under Climate Change

A number of studies have examined the potential of soil C sequestration for climate mitigation options [2, 4]; whereas, climate change will be a key driver of change in soil C stock during the twenty-first century [8]. Increases in mean air temperature are likely to accelerate SOC decomposition and loss of SOC in the future, if soil moisture is not a limiting factor [8, 12]. On the other hand, any increase in soil C input from, for example, increased ecosystem productivity through CO_2 fertilization or climate change, or technological and management improvement, could compensate the loss [8]. Increase in extreme climatic events and decrease in rainfall could reduce ecosystem productivity and thus could offset any positive impacts of climate change and CO_2 fertilization on soil C inputs through ecosystem productivity. Ultimately increase in SOC decomposition and changes in soil C inputs will determine the fate of soil C reserve under future climatic changes. Therefore, simulation of the fate of SOC under climatic changes is very important. There are some recent studies on projection of SOC under climatic changes across arable, forests and grasslands [8, 12, 46].

Grassland is one of the most important and widespread terrestrial ecosystems, covering approximately 40% of the global land surface, and containing the largest share (39%) of terrestrial soil C stocks (~580 Gt C) [79]. Any change in the SOC storage in grasslands will have a significant and long-term effect on global C cycles [80]. Hence, estimates of changes in grassland soil C stocks under climatic changes are of critical importance. A loss of 6–10% of the European grassland SOC stocks over 90 years (1990–2080) was projected by Smith and others [8] depending on the emission scenarios, whereas Xu and others [46] reported a loss of 2–6% of grassland SOC stocks in Ireland across different emission scenarios over 40 years

(2021–2060). Similarly, Senapati and others [81] estimated a loss of grassland SOC by 10–11 % due to the direct effect of climatic changes in Australia.

14.11 Limitation/Weakness/Scope of Improvement of the SOC Model

Soil microbial biomass is often represented as an active pool in SOC turnover models and thus remains a black box in terms of functionality [82]. Models often do not include the simulation of feedback of soil microbial biomass resulting from changes of a range of management practices [32]. Although there is no evidence of limiting SOC turnover by soil biota abundance, inclusion of the dynamics of soil biota explicitly and their feedback in the SOC models could improve model performance [83].

Soil protection towards SOC through aggregation is well known [60]. Disruption of soil aggregates is associated with loss of SOC, whereas generation of soil aggregates provides more retention of SOC. However, most of the models do not count the whole process of soil aggregate formation and destruction, and hence ignore soil structure dynamics and its implication on SOC stabilization or decomposition [48]. Inclusion of soil aggregate dynamics in the model might help in model improvement. Although there are some efforts to include soil aggregation factor in model [17], but still feedback between soil structure and microbial activity needs to be improved in the SOC turnover models.

Although the diffusive process plays an important role in movement and transportation of solutes and gaseous substances in SOC dynamics, SOC models often do not simulate the diffusive process explicitly. Further, model development regarding the role of diffusive process in SOC dynamics could be important.

Models often simulate inadequately the dynamics of different macronutrients (N, P, K), and other micronutrients especially their toxicity (e.g. Al toxicity), if any in the soil, along with the SOC dynamics. The SOC models generally simulate faster SOC turnover than reality in acidic soil, as the decomposition rate is up to two-thirds slower under acidic condition [43]. While some models account soil pH in turnover process of SOC, others do not. Adequate simulation of these factors together with soil C in the SOC turnover model could improve model overall performance. SOC models also need better integration with landscape as well as whole ecosystem processes.

14.12 Summary

The SOC turnover models are important tools for investigating the behaviour of SOC under current and future management practices, land use, LUC and climatic condition. The SOC models are now increasingly being used to assist policy

decisions at the national and international levels regarding future guidelines of land use, climate change and management practices. They are the important components of different agronomic decision support systems for improving agronomic efficiency and environmental quality. These models are useful in exploration, extrapolation, interpolation and explanation of experimental data. The SOC models are also important means of improving our understanding of C turnover process as well as underlying C stabilization mechanisms in soil.

Process-based multi-compartmental models are most popular and widely used SOC turnover models compared with other SOC models due to ease of use across different scales in space and time. Different compartments/pools actually simplify very complex physical, biological and chemical characteristics of SOC, and thus help in better understanding the SOC dynamics. However, various pools in multi-compartmental models are conceptual, based on empirically derived relationships and qualitative concepts, and often do not correspond to measureable fractions. Measured SOC fractions need to be unique as well as non-composite to represent a pool in the model. Although a combination of physical fractionation methods with chemical fractionation methods has the potential to relate measured SOC fractions to model pools, its applicability needs to be tested further.

However, the process-oriented multi-compartmental models could be improved by modifying their different pools/compartments so that they are based on measureable SOC fractions which are unique, non-composite, and in accordance with different SOC stabilization mechanisms. Once measured, the different SOC pools can be helpful in initialization of model explicitly. The use of measured SOC pools has the advantage over other initialization methods as the field measurements are independent of the model assumptions. Measured fractions reflect better the real site specific conditions under which SOC is accumulated; information on previous land use and management history are not required.

Model evaluation is a prerequisite for model application and helps in the assessment of model performance. Model evaluation provides confidence and reliability in the modelling. Different SOC models need to be evaluated comprehensively using reliable dataset as widely as possible before their use in a new environment other than their original environment, where they were developed. Model performances are generally evaluated using long-term field experimental data set, collected in chronosequence. In the absence of long-term chronosequence data sets, particularly under the scenario of LUC, paired-sites data sets could offer an opportunity for the evaluation of SOC models.

Climatic change is a key driver of change in soil C dynamics during the twenty-first century. Increases in mean air temperature are likely to accelerate SOC decomposition and loss of SOC in the future, if soil moisture is not a limiting factor. Increases in mean air temperature and CO_2 fertilization could increase ecosystem productivity, whereas increase in extreme climatic events and decrease in rainfall could reduce the same. Hence, projection of the fate of SOC dynamics under climatic change is very important. Reliable measures of black C in the fire-prone regions are very important in projection studies, as it influences the simulation of SOC dynamics.

The SOC models often simulate the dynamics of different macro- and micronutrients along with SOC dynamics inadequately. They also often do not account for soil pH explicitly, and do not simulate the whole process of soil aggregation and the dynamics of soil biota adequately. Inclusion of all these process/factors/parameters in the SOC models could represent the complex real life systems in a better way; it might improve the overall model performance.

References

1. Batjes NH (1996) Total carbon and nitrogen in the soils of the world. Eur J Soil Sci 47:151–163
2. Lal R (2004) Soil carbon sequestration to mitigate climate change. Geoderma 123:1–22
3. IPCC (2007) Climate Change 2007: the physical science basis. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
4. Smith P (2004) Soils as carbon sinks: the global context. Soil Use Manag 20:212–218
5. Baldock JA, Wheeler I, McKenzie N, McBrateny A (2012) Soils and climate change: potential impacts on carbon stocks and greenhouse gas emissions, and future research for Australian agriculture. Crop Pasture Sci 63:269–283
6. Smith P, Fang C, Dawson JJC, Moncrieff JB (2008) Impact of global warming on soil organic carbon. In: Donald LS (ed) Advances in agronomy. Academic Press, San Diego, CA pp 1–43
7. Guo LB, Gifford RM (2002) Soil carbon stocks and land use change: a meta analysis. Glob Chang Biol 8:345–360
8. Smith JO, Smith P, Wattenbach M, Zaehle S, Hiederer R, Jones RJA, Montanarella L, Rounsevell MDA, Reginster I, Ewert F (2005) Projected changes in mineral soil carbon of European croplands and grasslands, 1990–2080. Glob Chang Biol 11:2141–2152
9. Wilson BR, Growns I, Lemon J (2008) Land-use effects on soil carbon and other soil properties on the NW slopes of NSW: implications for soil condition assessment. Aust J Soil Res 46:359–367
10. Ghosh S, Wilson BR, Ghoshal SK, Senapati N, Mandal B (2012) Organic amendments influence soil quality and carbon sequestration in the Indo-Gangetic plains of India. Agric Ecos En 156:134–141
11. Smith P, Smith JU, Powlson DS, McGill WB, Arah JRM, Chertov OG, Coleman K, Franko U, Frolking S, Jenkinson DS, Jensen LS, Kelly RH, Klein-Gunnewiek H, Komarov AS, Li C, Molina JAE, Mueller T, Parton WJ, Thornley JHM, Whitmore AP (1997) A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. Geoderma 81:153–225
12. Smith P, Smith J, Wattenbach M, Meyer J, Lindner M, Zaehle S, Hiederer R, Jones RJA, Montanarella L, Rounsevell M, Reginster I, Kankaanpää S (2006) Projected changes in mineral soil carbon of European forests, 1990–2100. Can J Soil Sci 86:159–169
13. Falloon P, Smith P (2009) Modelling soil carbon dynamics. In: Kutsch WL, Bahn M, Heinemeyer A (eds) Soil carbon dynamics: an integrated methodology. Cambridge University, New York, pp 221–244
14. Smith P (2002) Organic matter modeling. In: Lal R (ed) Encyclopedia of soil science. Marcel Dekker Inc, New York, pp 917–924
15. Smith J, Smith P (2007) Introduction to environmental modelling. Oxford University, New York
16. Jenny H (1941) Factors of soil formation. A system of quantitative pedology. McGraw-Hill, New York

17. Malamoud K, McBratney AB, Minasny B, Field DJ (2009) Modelling how carbon affects soil structure. Geoderma 149:19–26
18. Lee JJ, Phillips DL, Liu R (1993) The effect of trends in tillage practices on erosion and carbon content of soils in the US corn belt. Water Air Soil Pollut 70:389–401
19. Falloon PD, Smith P, Smith JU, Szabó J, Coleman K, Marshall S (1998) Regional estimates of carbon sequestration potential: linking the Rothamsted carbon model to GIS databases. Biol Fertil Soils 27:236–241
20. Post WM, King AW, Wullschleger SD (1996) Soil organic matter models and global estimates of soil organic carbon. In: Powlson DS, Smith P, Smith JU (eds) Evaluation of soil organic matter models using existing long-term datasets. NATO ASI series I: global environmental change, vol 38. Springer, Berlin, pp 201–222
21. McCown RL, Hammer GL, Hargreaves JNG, Holzworth DP, Freebairn DM (1996) APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. Agric Sys 50:255–271
22. Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, Wilkens PW, Singh U, Gijsman AJ, Ritchie JT (2003) The DSSAT cropping system model. Eur J Agron 18:235–265
23. Smith JU, Bradbury NJ, Addiscott TM (1996) SUMDIAL: a PC-based system for simulating nitrogen dynamics in arable land. Agron J 88:38–43
24. IPCC (2000) Land use, land-use change, and forestry. In: Watson RT, Noble IR, Bolin B, Ravindranath NH, Verardo DJ, Dokken DJ (eds) A special report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
25. Jenkinson DS (1977) Studies on the decomposition of plant material in soil. V. The effects of plant cover and soil type on the loss of carbon from ^{14}C labelled ryegrass decomposing under field conditions. J Soil Sci 28:424–434
26. Coleman K, Jenkinson DS (1996) RothC-26.3– a model for the turnover of carbon in soil. In: Powlson DS, Smith P, Smith JU (eds) Evaluation of soil organic matter models using existing long-term datasets. NATO ASI series I: global environmental change, vol 38. Springer, Berlin, pp 237–246
27. Parton WJ (1996) The CENTURY Model. In: Powlson DS, Smith P, Smith JU (eds) Evaluation of soil organic matter models using existing long-term datasets. NATO ASI series I: global environmental change, vol 38. Springer, Heidelberg, pp 283–293
28. Falloon PD, Smith P (2000) Modelling refractory soil organic matter. Biol Fertil Soils 30:388–398
29. Parshotam A (1996) The Rothamsted soil-carbon turnover model—discrete to continuous form. Ecol Model 86:283–289
30. Smith P, Smith JU, Powlson DS (1996) Soil organic matter network (SOMNET): 1996 model and experimental metadata; GCTE Report 7. GCTE Focus 3 Office, Wallingford, p. 225
31. Paustian K (1994) Modelling soil biology and biochemical processes for sustainable agricultural research. In: Pankhurst CE, Doube BM, Gupta VVSR, Grace PR (eds) Soil biota: management in sustainable farming systems. CSIRO, Melbourne, pp 182–193
32. Smith P, AndrÉN O, Brussaard L, Dangerfield M, Ekschmitt K, Lavelle P, Tate K (1998) Soil biota and global change at the ecosystem level: describing soil biota in mathematical models. Glob Chang Biol 4:773–784
33. Bosatta E, ÅGren GI (1995) Theoretical analyses of interactions between inorganic nitrogen and soil organic matter. Eur J Soil Sc 46:109–114
34. Bosatta E, Ågren GI (1994) Theoretical analysis of microbial biomass dynamics in soils. Soil Biol Biochem 26:143–148
35. McGill WB (1996) Review and classification of ten soil organic matter (SOM) models. In: Powlson DS, Smith P, Smith JU (eds) Evaluation of soil organic matter models using existing long-term datasets. NATO ASI series I: global environmental change, vol 38. Springer, Berlin, pp 111–132
36. Setia R, Smith P, Marschner P, Baldock J, Chittleborough D, Smith J (2011) Introducing a decomposition rate modifier in the Rothamsted carbon model to predict soil organic carbon stocks in saline soils. Environ Sci Tech 45:6396–6403

37. Carvalhais N, Reichstein M, Seixas J, Collatz GJ, Pereira JS, Berbigier P, Carrara A, Granier A, Montagnani L, Papale D, Rambal S, Sanz MJ, Valentini R (2008) Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. Glob Biogeochem Cycles 22(2), GB2007
38. Yeluripati JB, van Oijen M, Wattenbach M, Neftel A, Ammann A, Parton WJ, Smith P (2009) Bayesian calibration as a tool for initialising the carbon pools of dynamic soil models. Soil Biol Biochem 41:2579–2583
39. Smith JU, Smith P, Monaghan R, MacDonald AJ (2002) When is a measured soil organic matter fraction equivalent to a model pool? Eur J Soil Sci 53:405–416
40. Coleman K, Jenkinson DS (2008) RothC-26.3. A model for the turnover of carbon in soil. Rothamsted Research, Harpenden
41. Wutzler T, Reichstein M (2007) Soils apart from equilibrium—consequences for soil carbon balance modelling. Biogeosciences 4:125–136
42. Cannell MGR, Thornley JHM (2003) Ecosystem productivity is independent of some soil properties at equilibrium. Plant Soil 257:193–204
43. Kelly RH, Parton WJ, Crocker GJ, Graced PR, Klír J, Körschens M, Poulton PR, Richter DD (1997) Simulating trends in soil organic carbon in long-term experiments using the century model. Geoderma 81:75–90
44. Carvalhais N, Reichstein M, Ciais P, Collatz GJ, Mahecha MD, Montagnani L, Papale D, Rambal S, Seixas J (2010) Identification of vegetation and soil carbon pools out of equilibrium in a process model via eddy covariance and biometric constraints. Glob Chang Biol 16:2813–2829
45. Zimmermann M, Leifeld J, Schmidt MWI, Smith P, Fuhrer J (2007) Measured soil organic matter fractions can be related to pools in the RothC model. Eur J Soil Sci 58:658–667
46. Xu X, Liu W, Kiely G (2011) Modeling the change in soil organic carbon of grassland in response to climate change: effects of measured versus modelled carbon pools for initializing the Rothamsted carbon model. Agric Ecos En 140:372–381
47. Waksman SA (1938) Humus: origin, chemical composition, and importance in nature. The Williams & Wilkins company, Baltimore
48. Krull ES, Baldock JA, Skjemstad JO (2003) Importance of mechanisms and processes of the stabilisation of soil organic matter for modelling carbon turnover. Funct Plant Biol 30:207–222
49. Jenkinson DS, Fox RH, Rayner JH (1985) Interactions between fertilizer nitrogen and soil nitrogen—the so-called 'priming' effect. J Soil Sci 36:425–444
50. Elliot ET, Paustian K, Frey SD (1996) Modelling the measureable or measuring the modelable: a hierarchical approach to issolating meaningful soil organic matter fractions. In: Powlson DS, Smith P, Smith JU (eds) Evaluation of soil organic matter models using existing long-term datasets. NATO ASI series I: global environmental change, vol 38. Springer, Berlin, pp 162–179
51. Blair G, Lefroy R, Lisle L (1995) Soil carbon fractions based on their degree of oxidation, and the development of a carbon management index for agricultural systems. Aust J Agri Res 46:1459–1466
52. Chan KY, Bowman A, Oates A (2001) Oxidizible organic carbon fractions and soil quality changes in an Oxic Paleustalf under different pasture leys. Soil Sci 166:61–67
53. Rovira P, Vallejo VR (2002) Labile and recalcitrant pools of carbon and nitrogen in organic matter decomposing at different depths in soil: an acid hydrolysis approach. Geoderma 107:109–141
54. Christensen BT (1996) Matching measurable soil organic matter fractions with conceptual pools in simulation models of carbon turnover: revision of model structure. In: Powlson DS, Smith P, Smith JU (eds) Evaluation of soil organic matter models using existing long-term datasets NATO ASI series I: global environmental change, vol 38. Springer, Berlin, pp 143–159
55. Cambardella CA, Elliot ET (1992) Particulate soil organic matter changes across a grassland cultivation sequence. Soil Sci Soc Am J 56:777–783

56. Christensen BT (2001) Physical fractionation of soil and structural and functional complexity in organic matter turnover. Eur J Soil Sci 52:345–353
57. Sohi SP, Mahieu N, Arah JRM, Powlson DS, Madari B, Gaunt JL (2001) A procedure for isolating soil organic matter fractions suitable for modeling. Soil Sci Soc Am J 65:1121–1128
58. Martel YA, Paul EA (1974) Effects of cultivation on the organic matter of grassland soils as determined by fractionation and radiocarbon dating. Can J Soil Sci 54:419–426
59. Trumbore SE, Vogel JS, Southon JR (1989) AMS ^{14}C measurements of fractionated soil organic matter: an approach to deciphering the soil carbon cycle. Radiocarbon 31:644–654
60. Six J, Conant RT, Paul EA, Paustian K (2002) Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. Plant Soil 241:155–176
61. Wattel-Koekkoek EJW, Buurman P, Van Der Plicht J, Wattel E, Van Breemen N (2003) Mean residence time of soil organic matter associated with kaolinite and smectite. Eur J Soil Sci 54:269–278
62. Leifeld J, Kögel-Knabner I (2001) Organic carbon and nitrogen in fine soil fractions after treatment with hydrogen peroxide. Soil Biol Biochem 33:2155–2158
63. Skjemstad JO, Spouncer LR, Cowie B, Swift RS (2004) Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools. Aust J Soil Res 42:79–88
64. Bruun S, Thomsen IK, Christensen BT, Jensen LS (2008) In search of stable soil organic carbon fractions: a comparison of methods applied to soils labelled with ^{14}C for 40 days or 40 years. Eur J Soil Sci 59:247–256
65. Braadbaart F, Poole I (2008) Morphological, chemical and physical changes during charcoalification of wood and its relevance to archaeological contexts. J Arch Sci 35:2434–2445
66. Herring JR (1985) Charcoal flux into sediments of the north pacific ocean: the cenozoic record of burning. In: Sundquist ET, Broecker WS (eds) The carbon cycle and atmospheric CO_2: natural variations from archean to present. Geophysical monograph series, vol 32. American Geophysical Union, Washington D.C, pp 419–442
67. Lehmann J, Skjemstad J, Sohi S, Carter J, Barson M, Falloon P (2008) Australian climate-carbon cycle feedback reduced by soil black carbon. Nat Geosci 1:832–835
68. Zackrisson O, Nilsson MC, Wardle DA (1996) Key ecological function of charcoal from wildfire in the boreal forest. Oikos 77:10–19
69. Glaser B, Haumaier L, Guggenberger G, Zech W (1998) Black carbon in soils: the use of benzenecarboxylic acids as specific markers. Org Geochem 29:811–819
70. Kurth VJ, MacKenzie MD, DeLuca TH (2006) Estimating charcoal content in forest mineral soils. Geoderma 137:135–139
71. Skjemstad JO, Janik LJ, Head MJ, McClure SG (1993) High energy ultraviolet photo-oxidation: a novel technique for studying physically protected organic matter in clay and silt-sized aggregates. J Soil Sci 44:485–499
72. Simpson MJ, Hatcher PG (2004) Determination of black carbon in natural organic matter by chemical oxidation and solid-state ^{13}C nuclear magnetic resonance spectroscopy. Org Geochem 35:923–935
73. Schmidt MWI, Skjemstad JO, Czimczik CI, Glaser B, Prentice KM, Gelinas Y, Kuhlbusch TAJ (2001) Comparative analysis of black carbon in soils. Glob Biogeochem Cycles 15:163–167
74. Smith JU, Smith P, Addiscott TM (1996) Quantitative methods to evaluate and compare soil organic matter (SOM) models. In: Powlson DS, Smith P, Smith JU (eds) Evaluation of soil organic matter models using existing long-term datasets NATO ASI series I: global environmental change vol 38. Springer, Berlin, pp 181–199
75. Lal R, Kimble JM (1997) Conservation tillage for carbon sequestration. Nutr Cycl Agroecosys 49:243–253
76. Houghton RA, Hackler JL, Lawrence KT (1999) The U.S. carbon budget: contributions from land-use change. Science 285:574–578
77. IPCC (2001) Climate change 2001: the scientific basis. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, Linden PJ, Dai X, Maskell K, Johnson CA (eds) Contribution of working group i

to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
78. Cerri CEP, Easter M, Paustian K, Killian K, Coleman K, Bernoux M, Falloon P, Powlson DS, Batjes N, Milne E, Cerri CC (2007) Simulating SOC changes in 11 land use change chronosequences from the Brazilian Amazon with RothC and century models. Agric Ecos En 122:46–57
79. White RP, Murray S, Rohweder M (2000) Pilot analysis of global ecosystems: grassland ecosystems. World Resources Institute, Washington, D.C
80. Parton WJ, Scurlock JMO, Ojima DS, Schimel DS, Hall DO, Scopegram Group Members (1995) Impact of climate change on grassland production and soil carbon worldwide. Glob Chang Biol 1:13–22
81. Senapati N, Smith P, Wilson B, Yeluripati JB, Daniel H, Lockwood P, Ghosh S (2013) Projections of changes in grassland soil organic carbon under climate change are relatively insensitive to methods of model initialization. Eur J Soil Sci 64:229–238
82. Chabbi A, Rumpel C (2009) Organic matter dynamics in agro-ecosystems—the knowledge gaps. Eur J Soil Sci 60:153–157
83. Lawrence CR, Neff JC, Schimel JP (2009) Does adding microbial mechanisms of decomposition improve soil organic matter models? A comparison of four models using data from a pulsed rewetting experiment. Soil Biol Biochem 41:1923–1934

Chapter 15
Crop Growth Simulation Modeling

Avnish Kumar Bhatia

15.1 Introduction

Agricultural crops comprise of various annual and perennial plant species grown on farm for food, fodder, fiber, and other raw material for industrial uses. The fields used for growing crops differ in environmental conditions, input availability and management practices resulting in differences in resource use and productivity. It is predicted that there will be around 9.1 billion people on the globe by the year 2050 and food consumption will have to increase by 50–70 % from now [1]. Increase in population requires increase in the agricultural production for food and other requirements. At the same time, there is shrinkage of available resources due to urbanization, developmental activities, and use of land for industries. The demand for food is increasing due to growing population and rising income of majority of the population. There is a possibility of aggravating food crisis because of climate change and diversion of arable lands to other uses [2–3].

Efficient management of available resources is essential to increase productivity of agriculture. In addition, the focus of agricultural production is changing from quantity to desired food quality. Solution to these challenges requires consideration of how crop growth components such as inputs, soils, weather, and management work together to effect plant growth and crop yield. Farmers require increasing amount of information to deal with the transitions in weather, changes in technology, and transition in demand for food in terms of both quantity and quality. The task to collect, collate, store, and analyze the information is demanding in itself. With information technology opening new opportunities, computer programs in the form of crop models that simulate crop growth and crop yield under different input and environments can help farmers and planners make decisions to manage crops better.

Crop modeling is the dynamic simulation of crop growth by numerical integration of constituent processes with the aid of computers [4]. Crop modeling has developed extensively over the past 30 years and a diverse range of crop models have

A. K. Bhatia (✉)
National Bureau of Animal Genetic Resources, 132001 Karnal, Haryana, India
e-mail: avnish@lycos.com

S. K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9_15,

been developed and validated across regions. It combines a scientific approach to enhance understanding with application orientation focusing on output prediction tasks. Crop simulation models have been the desired tools for increasing the efficiency of trials on a crop in diverse regions.

Precision farming aims to apply crop inputs such as seed, fertilizer, pesticide, and irrigation in a timely and efficient manner. It holds great potential to increase crop productivity through taking up agricultural practices according to weather and soil conditions. This involves management of agricultural practices such as sowing, fertilizer, irrigation, and other inputs on the basis of variability in field conditions. Therefore, it depends on crop growth models to accurately predict responses of crop growth under different weather, soil, and other environmental conditions [5].

Changes in climate have resulted in increase in average temperature and CO_2 levels. Environmental degradation due to industrial and agricultural activities has caused dispersion of pollutants in the subsurface. New strategies are needed to assess the impact of these changes. Existing groundwater, soil conservation and agricultural policies and strategies need reconsideration. A system approach, encompassing the development of crop simulation models is useful for policy makers and scientists to get better insights in the complexity and interaction of the processes affecting the nutrients, pollutants, and chemicals in the dynamic soil–water–crop relations [6].

15.2 History of Crop Modeling

Oteng-Darko and others [7] describe the history of crop growth modeling since 1960s. Attempts to model photosynthetic rates of crop canopies and to estimate potential food production for some areas of the world and to provide indications for crop management and breeding resulted in the construction of an elementary crop growth simulator (ELCROS) [8]. The model included a static photosynthesis model. Crop respiration was taken as a daily fixed factor of the biomass. Later, a functional equilibrium between root and shoot growth was added to the model. Addition of micrometeorology and quantification of canopy resistance to gas exchanges improved the simulation of transpiration. It resulted into the basic crop growth simulator (BACROS) [9].

International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) started the development of a model in 1982 for tropics and subtropical environments. It was an attempt for understanding various cropping options through systems analysis and simulation. Major outcome of IBSNAT was the decision support system for agrotechnology transfer (DSSAT) [10], which has been used as a research and teaching tool. It is used to derive recommendations concerning crop management and to investigate environmental and sustainability issues. It enables users to match biological requirements of the crops to physical characteristics of the land to provide them with management options

for improvement in land-use planning. It includes database management for soil, weather, genetic coefficients, management inputs, crop simulation models, and a series of utility, and weather generation programs. It also contains strategy evaluation program to evaluate options such as variety selection, planting date, plant density, row spacing, soil type, irrigation, fertilizer application, initial conditions on yields, water stress in the vegetative or reproductive stages of development, and yield. DSSAT has the potential to reduce substantially the time and cost of field experimentation necessary for adequate evaluation of new cultivars and management systems.

Agricultural production systems simulator (APSIM) has been developed and used for improved risk management in agricultural production [11]. It includes crop models that were mostly species oriented, emphasizing the differences between species.

The species-oriented approach has dominated simulation models of agricultural production systems and resulted in a large number of different crop models [11]. Thus, the science behind the models is often not transparent, improvements cannot be easily transferred between models, and their computer code can hardly be reused. Due to differences in their program structures and modeling approaches and poor modularity, such models are difficult to maintain and component level comparison is not possible while whole model comparison provides little information on how to improve models due to the confounding effects of interactions among different subprocesses.

Alternately, crop models based on the similarities across crops can be thought so that modeling approaches developed and implemented for one crop can be reused for other crops. These similarities are based on the recognition that all crops can be represented using the same basic physiological principles to capture resources, to grow and develop to yield the desired output [12]. Such generic models have been developed recently for wide use of modeling community. InfoCrop [13] is such a generic crop model that is designed to simulate the effects of weather, soils, agronomic management such as planting, nitrogen, residues and irrigation, and major pests on crop yield and associations with environmental impacts.

15.3 Crop Model Development

A crop model is a simple demonstration of growth and yield of a crop. In general, it is used to study crop growth and to calculate growth responses to the environment in which the crop is cultivated. Model simulates or imitates the behavior of a real crop by predicting the growth of its components, such as leaves, roots, stems, and grains. Thus, a crop growth simulation model not only predicts the final state of crop production or harvestable yield but also contains quantitative information about major processes involved in the growth and development of the crop. Reactions and interactions at the level of tissues and organs are combined to form a picture of the crop's growth processes [7].

15.3.1 Crop Growth Processes

Empirical crop growth models, which apply functions fitted to data, are calibrated for a particular species and given site conditions. These models cannot be valid over a wide range of conditions. Understanding how plant growth and development interact with environmental factors such as light, temperature, soil parameters, and water availability is extremely important to understand the yield of crop plants. To model these interactions we need to take into account the physiological processes involved in crop growth. The term "process-based model" was introduced to describe those models that considered the interactions between plant functional processes and abiotic factors [14].

Plant growth is driven by several processes simultaneously and is thus highly complex to understand and to model. Photosynthesis is only one out of several growth drivers. Photosynthesis is simulated based on direct and diffuse light, as in greenhouses. The CO_2 concentration in the atmosphere can have a large impact on the rate of photosynthesis. Other important aspects of plant growth that can also be considered are the way in which dry matter is stored into the different plant organs, the feedbacks involved, and the plant's capacity for additional resource acquisition, as well as the metabolic costs and expenses. Plant investment strategies with regard to the above sink-driving processes are a very important aspect to be considered in analyzing and modeling morphological plasticity. Added to this complexity is the further problem of situating a given species within an ecosystem and understanding the interactions taking place, especially concerning below-ground processes.

The InfoCrop model [13] considers the following processes in the model:

1. Crop growth and development: phenology, photosynthesis, partitioning, leaf area growth, storage organ numbers, source sink balance, transpiration, uptake, allocation, and redistribution of nitrogen
2. Effects of water, nitrogen, temperature, flooding, and frost stresses on crop growth and development
3. Crop–pest interactions: damage mechanisms of insects-pests and diseases
4. Soil water balance: root water uptake, inter-layer movement, drainage, evaporation, runoff, and pond filling
5. Soil nitrogen balance: mineralization, uptake, nitrification, volatilization, inter-layer movement, denitrification, leaching
6. Soil organic carbon dynamics: mineralization and immobilization
7. Emissions of green house gases: CO_2, methane, nitrous oxide

15.3.2 Input and Output

Crop models take a variety of input data on weather, phenology, nutrients, and irrigation, and provide output at various crop growth stages and finally crop yield.

Gertsis and others [15] use the following information for crop simulation model divided into three categories to validate and apply GOSSYM for optimization of cotton production under the cultural, soil, and climatic conditions of Greece and Spain:

1. Crop information: latitude of location, distance between rows, plant population per unit of planted row, date of emergence and cultivar used, time and amount of irrigation, and nitrogen fertilizer in various forms
2. Soil information: initial soil water and fertility levels at 15 cm depth increments, taken at the beginning of each growing season; the soil moisture characteristic of the curve, the saturated hydraulic conductivity and textural analysis, and the bulk density required for the permanent soil hydrology files and for each horizon
3. Weather information: daily maximum and minimum temperature, rainfall, solar radiation, and wind speed

The following information is produced in three formats by the GOSSYM model execution:

1. Analytical tables: general information on crop height, leaf area index, node number, green and open bolls, yield, aborted fruits, fruiting positions, light intercepted, evapotranspiration, water and nitrogen stress, soil water and leaf water potential, weather information, dry weight of the plant parts
2. Maps: plant map, root system map, soil water potential, soil water content, soil temperature, soil nitrogen, vegetative and reproductive branch and plant organs distribution
3. Graphs: cotton lint yield, plant height, bolls, water and nitrogen stress, solar radiation, temperature

InfoCrop [13], a generic crop model, requires the following input to produce certain output as given below:

Input:

1. Weather parameters: radiation, rainfall, temperature, wind speed, humidity, frost
2. Soil parameters: texture, pH, depth, fertility
3. Crop variety: Physiology, phenology, morphology
4. Management/package of practices: Date of planting, fertilizer application, irrigation, residue management
5. Pests: type, population, severity

Output:

1. Economic yield, biomass
2. Crop duration, evapotranspiration, and nitrogen uptake
3. Yield loss due to pests
4. Environmental impact: greenhouse gases, nitrogen losses, percolation
5. Soil carbon, nitrogen, water dynamics

It becomes difficult to have all the required data for simulation. With the available minimum data set, it is possible to execute simulation models for desired purposes.

15.3.3 Programming Languages

The basic crop growth models are developed in Fortran simulation translator/environment (FST/FSE) [16]. FORTRAN simulation environment has been developed to provide a computationally powerful programming. The simulation environment provides functionality for specification of model processes, integration of various rate variables, time update in the model, input of weather data, and reading of model parameters from files. It also provides graphical presentation of results. The crop model software is written in a high-level computer language, usually FORTRAN. Other high-level computer languages such as C ++ and Pascal have also been used.

15.3.4 Other Requirements

The time step of the operating crop simulation model should match the real lifetime intervals during which there is measurable variation in the inputs. One-day time step is taken for crop growth simulation models, because the weather parameters such as rainfall, temperature, wind speed, etc. are recorded on a daily basis.

15.3.5 Model Development Process

Model building can be started after availability of scientific knowledge about the crop growth processes, the growth-controlling factors, and the interactions that are dominant within a particular cropping system. Crop growth models in general contain a set of equations that estimate the production rate of biomass from the captured resources, such as carbon dioxide, solar radiation, and water. Regression analysis describes the effect of one or more variables (independent variables) on a single variable (dependent variable). For the desired output such as crop yield, one equation would be identified. The factors that control daily growth would be identified. Then, the biomass would be partitioned to the economic portion and the other crop matter. These would be specified as mathematical functions with conditional rules and would be looped over the selected time-steps to calculate plant growth and crop yield.

A crop growth simulation model algorithm [17] is given below. It calculates plant growth at a time step as a function of solar radiation, which is a function of crop canopy. The stop condition can be provided as the number of days for which the simulation is desired till the life span of crop.

```
t = 0;  /*Time */
yield(t) = seed sowing;  /* Yield at time t */
DO UNTIL END-condition is satisfied
t = t+1;
Inputsolar-radiation;
solar-radiation intercepted = Defined function of
crop canopy;
  plant-growth(t) = (solar-radiation) * (% radiation
  intercepted)*(radiation-used);
  yield(t) = yield(t-1) + plant growth(t);
END DO
```

Steps in development of a crop growth model are listed below [17]:

1. Flow charts are constructed to depict input, major crop processes, and output. It is a framework depicting the different steps to be followed like reading the data, computing different model components, repetition of calculations, desired reports, etc. in achieving the objective of the model.
2. Phenological development such as vegetative development, canopy expansion, crop maturity is controlled by thermal time. Phenology is the study of periodic biological phenomena. It qualitatively describes the successive stages in the development of plants, from seed germination to flowering to maturity. The crop growth stages, known as phenophases differ from crop to crop. Temperature and photo-period are the two main factors that determine flowering in plants.
3. Photosynthesis is derived as a function of canopy light interception and a radiation-use efficiency factor; and as soon as flowering occurs, daily biomass is allocated to maintenance and grain filling until crop end is encountered. Crop growth results from photosynthesis and is subject to modification by both abiotic and biotic factors. In the early crop growth stages, the rate of dry matter production is proportional to the amount of radiation intercepted, which is a function of leaf area index. Crop growth depends on the quantity of incident light, the proportion of that light intercepted by the photosynthesizing organs of the plant, its efficiency of conversion of light into dry matter and respiratory losses.
4. Computer program for the crop model is developed in a computer language.

15.4 Model Calibration

Calibration is the adjustment of the system parameters so that simulation results reach at a predetermined level of an observation. In many instances, even if a model is based on observed data, simulated values do not exactly comply with the observed data and minor adjustments have to be made for some parameters. This noncompliance may arise from sampling errors as well as from incomplete knowledge of the system. Alternatively, it may arise when the model is used in a situation that

Table 15.1 Examples of model parameters used in literature

Reference	Model and crop	Parameter	Value	Unit
[18]	AquaCrop, maize	Base temperature	8.0	°C
		Cut-off temperature	30.0	°C
		Canopy growth coefficient	19.9	% per day
		Leaf growth threshold (P_{upper})	0.14	% of total available water
		Leaf growth threshold (P_{lower})	0.72	% of total available water
		Leaf growth stress coefficient curve shape	2.9	Unitless (moderately convex curve)
		Time from sowing to emergence	6	days
		Time from sowing to the start of flowering	52	days
		Time from sowing to start senescence	72	days
		Time from sowing to maturity	97	days
		Length of the flowering stage	10	days
[13]	InfoCrop, wheat	Base temperature for sowing to germination	3.6	°C
		Thermal time form sowing to germination	70	°C days
		Base temperature for germination to 50% flowering	5	°C
		Thermal time for germination to 50% flowering	850	°C days
		Base temperature for 50% flowering to physiological maturity	7.5	°C
[19]	Canegro, sugarcane	Maximum fraction of daily dry mass increments partitioned to aerial parts	0.88	
		Partitioning coefficient	0.6	

is markedly different from the one under which it was developed. This is an essential step in model development aimed at deriving parameter values on the basis of experimental data.

Table 15.1 shows an example of parameters in some crop growth model as reported in literature.

Abedinpour and others [18] accomplished calibration of the AquaCrop model for maize crop by using the observed values from the field experiment during the year 2009 as a model input and then simulating the model to predict the output, viz. the yield, biomass, and canopy cover. The predicted output values were compared with the observed yield and biomass of the experimental plot. The difference between the model predicted and the experimental data was minimized by using trial and error approach, in which one specific input variable was chosen as the reference variable

at a time, and adjusting only those parameters that were known to influence the reference variable the most. The procedure is repeated to arrive at the closest match between the simulated and observed values of the experiments for each treatment combination.

15.5 Model Validation

Before a model can be used, it must be validated that involves running the model on historical input data recorded for a real system. Model output is compared with the real system output. Models should be validated with independent data, which has not been used in model development. Model building is an expensive exercise and most researchers are involved in the validation of an existing model to a new situation rather than in model development.

The model validation stage involves the confirmation that the calibrated model closely represents the real situation. Ideally, all models should be validated both at the level of overall system output and at the level of internal components and processes [17]. Validation of internal processes is important because of the occurrence of feedback loops in biological systems; good prediction of the overall output of the model could be attributed to compensating internal errors. However, validation of all the components is not possible due to the lack of detailed datasets and the option of validating only the determinant ones are adopted.

Singh and others [20] perform calibration, validation, and sensitivity analysis of CropSyst model to quantify and verify the interactive effects of different water and nitrogen treatments on the productivity of direct-seeded rice–wheat cropping system using measurements from field experiments. The observed variables were collected from field experiments conducted during 2004–2005 and 2005–2006 at the research farm of Indian Agricultural Research Institute, New Delhi. For direct seeded rice, the model performed well at lower levels of nitrogen (120 kg/ha), whereas at higher levels of nitrogen treatment (150 kg/ha) the predicted values were less than the measured values. The model performed satisfactorily at all levels of nitrogen in the case of wheat. Sensitivity analysis of the model for various crop parameters shows that the model is highly sensitive to the parameters like light to the above biomass conversion, specific leaf area, and phenological degree-days. Thus, more accuracy is required in the determination of these parameters in the model.

Casanova and others [21] test the performance of ORYZA1 for Mediterranean conditions for fully irrigated direct-seeded rice. The model was calibrated and validated with field data of two cultivars, a short-grain (Tebre) and a long-grain cultivar (L-202), grown in various years in the EbroDelta of Spain. Phenological development of rice crop, daily dry matter production, and leaf area development were calibrated. Tebre and L-202 had no significant differences in the total length of the development period. The preheading period, however, was longer and the

postheading period shorter in L-202 than in Tebre. This induced differences in translocation characteristics, spikelet number per unit area, weight of the grains and harvest index. The crop characteristics—extinction coefficient (increased with development stage), dynamics of nitrogen distribution, partitioning of assimilates, relative death rate of leaves, relative growth rate of leaf area during exponential growth, specific leaf area, and a strongly decreasing specific stem green area—were similar between cultivars. The simulated curve fitted the observations. The model simulated rice growth very accurately until flowering. After flowering, however, divergences appeared and increased especially at the yellow ripe stage. From then on, the crop did not grow much more, whereas it continued in the simulation. This reduction of growth rate was usually accompanied by an increase in the relative death rate of leaves and the drying of the grains. The researchers attribute the main source of error to a limited understanding of the ripening and sink limitation processes.

15.6 Available Models

Table 15.2 shows some of the available crop simulation models. The World Food Studies (WOFOST) model has been used to simulate maize crop growth and yield. The crop environment resource synthesis (CERES)—pearl millet and Cropsyst—are being used to study the suitability and yield simulation of pearl millet genotypes. GOSSYM is one of the most common growth models used for cotton crop. APSIM added with several modules is being used in crop rotation, crop sequence, and simulation studies involving perennial crops. DSSAT includes 15 crop models along with databases and other decision tools. Generic models such as InfoCrop have been used for modeling tree crops such as coconut [22].

15.7 Applications of Crop Growth Models

Simulation modeling is increasingly being applied in research, teaching, farm management, policy analysis, and production forecasts [7]. Some applications of crop growth simulation models are listed in the following sub-sections.

15.7.1 Crop Cultivation Practices

Many crop modeling groups have used their models in a research mode to evaluate producer risk by considering yield response to long-term historical weather records for a region, and to optimize planting date, planting density, row spacing, choice of cultivar, and fertilizer application for different soil types [30]. Models can assist in

Table 15.2 Description of some crop growth simulation models

Model	Description	Reference
COTTAM	Determines the morphological development of cotton crop in response to environment	[23]
GOSSYM-COMAX	GOSSYM/COMAX is a model-based decision aid for cotton crop management	[24]
WOFOST	Analyses of the growth and production of field crops under a wide range of weather and soil conditions	[25]
CropSyst	CropSyst is a multi-year, multi-crop, daily time-step cropping systems simulation model	[26]
APSIM	Agricultural Production Systems Simulator (APSIM) contains a suite of modules for modeling framework of crops, soil, climate, and management interactions	[11]
GWM	General weed management (GWM) model consists of a simulation model and databases. The simulation model includes processes of weed population dynamics during a season	[27]
ORYZA1	Simulates rice growth, development, and leaf area index under potential production	[21]
CERES	Crop environment resource synthesis (CERES)—generic modeling environment for cereal crops	[28]
DSSAT	A collection of programs and databases with various crop simulation models and decision tools	[10]
CANEGRO	Simulation of dry matter partitioning in sugarcane crop	[19]
InfoCrop	A generic crop simulation model to provide integrated assessment of the effect of weather, variety, pests, soil, and management practices on crop growth and yield, soil nitrogen, and organic carbon dynamics	[13]
AquaCrop	The Food and Agriculture Organization (FAO) crop model simulates yields of major herbaceous crops as a function of water consumption under rain-fed, supplemental, deficit, and full irrigation conditions	[29]

adjusting package of practices in various crops. Farmers can optimize a package of practices in his/her field with the help of crop growth models.

15.7.2 Cropping System Research Understanding

Model development involves integration of cropping systems' performance across variety of plants, soil, water management, weather, etc. It allows identification of the major factors that drive the system and can highlight areas where knowledge is lacking. Thus, modeling of crop growth contributes towards targeted and efficient research programs on crop cultivation.

Crop models can be used by researchers to examine scientific hypotheses on how a given process works or how the crop responds to a particular factor [30]. The hypothesis tested may be the ability of the model to predict final yield or the ability to predict intermediate processes such as instantaneous or daily canopy assimilation.

15.7.3 Field Experiments Data Management

Crop model development requires experimentation with data from various sites. It also guides crop cultivation under optimal conditions for maximal yield. Simulation model development, testing, and application demand the use of a large amount of technical and observational data and the modeler needs to develop data storage systems that aid in field research. The stored data becomes permanent record of experiments and are available to other researchers.

15.7.4 Climate Change Impacts Studies

The variability of climate is currently one of the major concerns for maintaining crop productivity. The application of crop models to study the potential impact of climate change on crop production has been widely used in various regions. The increased concentration of carbon dioxide and other greenhouse gases are expected to increase the temperature. Elevated temperature and carbon dioxide affects the biological processes like respiration, photosynthesis, plant growth, reproduction, water use, etc. Proper understanding of the effects of climate change on crop yield helps farmers to make crop management decisions such as selection of crops, varieties, sowing dates, and irrigation scheduling to minimize the risks.

15.7.5 Water and Fertilizer Management

Crop models have attempted to optimize water and nitrogen management over the long term. These efforts can suggest the best long-term strategy for water or fertilizer applications. For producers with access to water, this approach can help determine return on investment in irrigation equipment. Considered with prices and complete enterprise production costs, crop models can be used to evaluate economic risk in crop cultivation.

15.7.6 Crop Management Practices

Simulation can be done to determine the best management practices under a certain cropping system. In the past, the main focus of agronomic research has been on crop production. But the quality of the environment has become an important issue in addition to profitable crop production. Agricultural managers require strategies for optimizing the profitability of crop production while maintaining soil quality and minimizing environmental degradation. Solutions to these challenges require consideration of how numerous components interact to effect plant growth. Models

having parameters such as chemical leaching or erosion components can be used to determine the best farming practices over the long term.

15.7.7 Crop Yield Forecasting

Reasonably precise estimates of crop yield over large areas before the actual harvest are of immense value to both the researcher and the farmers in terms of crop cultivation and planning. Crop models are executed using actual weather data during the cropping season for the geographical region of interest to forecast crop yield. Yield forecasts are integrated with other information systems on availability and prices of crop produce to locate areas of surplus yield and possible trade.

15.7.8 Genetic Improvement and Breeding

Development and release of a crop variety is a complex process that may take many years. Crop models integrate different components of agro-ecosystems and can be used to conduct multilocation field experiments to understand genotype by environmental interactions. These studies are useful in reducing the number of sites and seasons required for field evaluation of crop varieties. Crop varieties may be developed in much less time with this approach.

The crop parameter determining the specific interaction could be identified by modeling a range of genotypes in selected environments. Hypothetical values could then be modeled combining the crop parameters conferring the most advantage as an indication of suitable traits and breeding target. Crop simulation models may be used to determine, in the future, which genetic traits would be most profitable for molecular geneticists to manipulate to maximize yields and returns for particular environments [30].

15.8 Limitations of Crop Growth Models

Model use is limited by unavailability of accurate input data. Limitations of model data includes cost of obtaining data, spatial variability in soil characteristics, temporal variability in pest outbreaks, and data quality in terms of measurement and sampling errors [30].

Crop models capture complex physiological interactions in a rough way by using sink requirements with environmental variables, using empirical routines for carbohydrate partitioning, nitrogen dynamics, incomplete description of root development and growth, and the assumption of optimum crop growth conditions [31]. Interactions between production, environment, and cost-cutting measures have many

aspects. Most of the underlying mechanisms are either yet to be quantified or even not known. The lack of corresponding knowledge forces modelers to build idealistic representations of system mechanism.

Crop models are assumed to work in different environments. Their empirical nature often does not allow such extrapolations [32]. The models should be important both for assessing the impacts of different management schemes and for identifying subject area where further basic research is needed.

15.9 Future Prospects in Crop modeling

15.9.1 Inclusion of Plant Architecture

A modeling methodology that takes into account aspects of the spatial, three-dimensional structure of plants, referred to as plant architecture, could answer the questions regarding plant and crop growth and development. Examples of such questions include those related to the effects of manipulation of plant canopies to competition phenomena between plants of the same or of different species such as weeds, or to the plastic response of plant structure to environmental influences.

Functional-structural plant modeling (FSPM) builds upon the classical principles that have been implemented in the widely used process-based crop models, and adds the possibility of explicitly considering plant structure [33]. This approach to plant and crop modeling provides the framework to consider environmental influences on each component of the system as well as mutual influences of structural components. An illustration can be given about cereal crop species such as rice, wheat, and maize. These plants exhibit a regular and coordinated development, making them particularly suitable for FSPM. These crops grow at high population densities and individual plant experiences a high degree of intra-specific competition for light, nutrients, and water. Compared to solitarily grown cereal plants, crop cereals experience various effects of competition, an important one of which is expressed in the tillering that the plant produces. For example, wheat plants grown at a low population density produce a high number of tillers and vice versa. To a large extent, the number of tillers that a wheat plant produces can be traced back to the degree of intra-specific competition for light at the early stages of vegetative growth, when the axillary buds break and tillers are being formed.

15.9.2 Social and Environmental Interactions

With the global population increase, climate change, and decrease in arable areas, identifying new approaches to increase crop productivity is one critical global challenge. Developing the next generation of crop and agro-ecosystem models can help

us identify better farming practices and breeding targets for increased productivity. It can also be helpful in the studies on the mechanistic basis behind observable macroscopic phenotypes, and predict responses of plants or crops to future climate change. The next generation crop and agro-ecosystem models will incorporate recent advances of mechanisms underlying many aspects of crop growth and development, and interactions between plants and their environments [5].

Developing the next generation crop and agro-ecosystem models requires close collaboration between scientists in different disciplines such as agronomy, mathematics, computer science, plant sciences, soil sciences, geology, and social sciences. Such models will play an important role in precision agriculture, where the bottleneck is the lack of robust models of interaction of plants with their environments. These interactions are essential to ensure that technical innovations are socioeconomically viable and will eventually be implemented to the benefit of farmers, consumers, and policy makers.

15.9.3 Genomics and Crop Modeling

In view of the potential added value of robust crop modeling to classical quantitative genetics, model-input parameters are considered to represent genetic coefficients. A number of case studies, in which the effects of quantitative trait loci or genes have been incorporated into existing models to replace model input, have shown promise of using these models in analyzing genotype–phenotype relationships of complex crop traits [34]. Studies of functional genomics will increasingly enable the elucidation of the molecular genetic basis of these model-input traits. Modeling of crop systems biology, which combines modern genomics, traditional physiology, and biochemistry is expected to realize the roles of in-silico modeling in narrowing genotype–phenotype gaps. Modeling tools based on crop systems biology can generate important insights to investigate important societal issues, such as improving food security or zinc supply for human nutrition [35].

15.9.4 Microscale Modeling of Crop Growth Components

Components such as CO_2 diffusion, light propagation in leaves, nutrient dispersion in the soil, ecohydrological processes, etc. affect crop production. Studies on individual crop growth elements at microscale are useful to breed crops for changing climate conditions resulting into severe drought and severe flooding. The microscale studies are also useful for agricultural and environmental policy makers concerned with dispersion of pollutants and fertilizers in the environments. Crop growth models are useful tools for the microscale studies on the effects of the individual elements on crop productivity. One example of such a study is the investigation of transport of CO_2 in leaves by combining a two-dimensional, microscale model with photosynthesis kinetics in wheat leaves [36].

15.10 Summary

Increase in the world population demands increase in the agricultural production for food, fiber, and industrial use. This demands efficient management of resources to increase crop productivity. Precision agriculture requires pertinent technologies, such as crop growth models to facilitate measured amounts of inputs to obtain desired quantity and quality of crop output. Crop growth model-based decision support systems are desirable tools to help agricultural decision makers in policy issues like yield forecasting, crop management, environmental issues, etc.

Crop growth simulation models integrate crop physiology, weather parameters, soil parameters, and management practices to simulate growth and yield of crops. It may not be possible to obtain all the required data for crop simulation. However, with the available minimum data it is possible to run the simulation models for desired applications. Crop simulation models compute growth values on a day to day basis using the relations among values of crop growth and weather parameters. The crop-specific model design results in poor modularity and prevents model sharing. A generic model can be developed using common crop physiological processes.

Validating and fine-tuning of crop model is an important step before using it for actual prediction tasks. A number of crop growth models have been developed since 1980s; many are generic models applicable to various crops. Future crop models should incorporate improved mechanism of interaction with environment and society. Crop models would be more helpful in breeding programs, if model parameters could be linked to genetic information. Microscale studies on individual crop growth components such as nutrient dispersion, CO_2 diffusion, subsurface water movement are possible with crop growth models.

References

1. Jaggard KW, Qi A, Ober ES (2010) Possible changes to arable crop yields by 2050. Philos Trans Royal Soc 365:2835–2851
2. Indian Council of Agricultural Research (2011) Vision ICAR 2030. Krishi Bhawan, New Delhi
3. IAASTD (2008) Agriculture at a crossroads: the synthesis report. Washington, DC: International Assessment of Agricultural Knowledge, Science and Technology for Development. www.agassessment.org/
4. Sinclair TR, Seligman NG (1996) Crop modeling: from infancy to maturity. Agron J 88:698–704
5. Zhu XG, Zhang GL, Tholen D, Wang Y, Xin C, Song Q (2011) The next generation models for crops and agro-ecosystems. Sci China Information Sci 54:589–597
6. Timmerman A, Feyen J (2003) The wave model and its application; simulation of the substances water and agrochemicals in the soil, crop and vadose environment. Revista CORPOICA 4:36–41
7. Oteng-Darko P, Yeboah S, Addy SNT, Amponsah S, Danquah EO (2013) Crop modeling: a tool for agricultural research—a review. J Agric Res Dev 2(1):1–6

8. De Wit CT, Brouwer R, Penning De Vries FWT (1970) The simulation of photosynthetic systems, in prediction and measurement of photosynthetic productivity. In: Setlik I (ed) Proceedings of international biological program/plant production technical meeting. PUDOC, Wageningen
9. De Wit CT, Goudriaan J (1978) Simulation of assimilation, respiration and transpiration of crops. Simulation monograph. PUDOC, Wageningen
10. Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, Wilkens PW, Singh U, Gijsman AJ, Ritchie JT (2003) The DSSAT cropping system model. Europ J Agron 18:235–265
11. McCown RL, Hammer GL, Hargreaves JNG, Holzworth DP, Freebairn DM (1996) APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. Agric Syst 50:255–271
12. Wang E, Robertson MJ, Hammer GL, Carberry PS, Holzworth D, Meinke H, Chapman SC, Hargreaves JNG, Huth NI, Mclean G (2002)Development of a generic crop model template in the cropping system model APSIM. Europ J Agron 18:121–140
13. Aggarwal PK, Kalra N, Chander S, Pathak H (2006) Infocrop: a dynamic simulation model for the assessment of crop yields, losses due to pests, and environmental impact of agroecosystems in tropical environments. I. Model description. Agric Syst 89:1–25
14. Fourcaud T, Zhang X, Stokes A, Lambers H, Korner C (2008) Plant growth modelling and applications: the increasing importance of plant architecture in growth models. Ann Bot (Lond) 101:1053–1063
15. Gertsis AC, Galanopoulou-Sendouca S, Papathanasiou G, Symeonakis A (1997) Use of GOSSYM—a cotton growth simulation model to manage a low input cotton production system in Greece. In: Proceedings for the first European conference for information technology in agriculture, Copenhagen, pp 359–362
16. Van Kraalingen DWG, Rappoldt C, Van Laar HH (2003) The Fortran simulation translator: a simulation language. Europ J Agron 18:359–361
17. Cheeroo-Nayamuth BF: (2000) Crop modelling/simulation: an overview. In: Lalouette JA, Bachraz DY, Sukurdeep N (eds.) Proceedings of the Fourth Annual Meeting of Agricultural Scientists, Réduit, Mauritius, 21–22 October 1999. Réduit, Mauritius
18. Abedinpour M, Sarangi A, Rajput TBS, Singh M, Pathak H, Ahmad T (2012) Performance evaluation of aquacrop model for maize crop in a semi-arid environment. Agric Water Manag 110:55–66
19. Singels A, Bezuidenhout CN (2002) A new method of simulating dry matter partitioning in the Canegro sugarcane model. Field Crops Res 78(2–3):151–164
20. Singh AK, Goyal V, Mishra AK, Parihar SS (2013) Validation of Cropsyst simulation model for direct seeded rice–wheat cropping system. Current Sci 104(10):1324–1331
21. Casanova D, Goudriaan J, Bosch AD (2000)Testing the performance of ORYZA1, an explanatory model for rice growth simulation, for Mediterranean conditions. Europ J Agron 12:175–189
22. Kumar SN, Kasturi Bai KV, Rajagopal V, Aggarwal PK (2008) Simulating coconut growth, development and yield with the InfoCrop-coconut model. Tree Physiol 28:1049–1058
23. Jackson BS, Arkin GF, Hearn AB (1988) The cotton simulation model "COTTAM": fruiting model calibration and testing. Trans ASABE 31(3):846–854
24. Mckinion JM, Baker DN, Whisler FD, Lambert JR (1989) Application of GOSSYM/COMAX system to cotton crop management. Agric Syst 31:55–65
25. Van Diepen CA, Wolf J, van Keulen H (1989) WOFOST: a simulation model of crop production. Soil Use Manag 5(1):16–24
26. Stöckle CO, Martin SA, Campbell GS (1994) Cropsyst, a cropping systems simulation model: water/nitrogen budgets and crop yield. Agric Syst 46(3):335–359
27. Wiles LJ, King RP, Schweizer EE, Lybecker DW, Swinton SM (1996) GWM: general weed management model. Agric Syst 50(4):355–376

28. Greenwald R, Bergin MH, Xu J, Cohan D, Hoogenboom G, Chameides WL (2006) The influence of aerosols on crop production: a study using the CERES crop model. Agric Syst 89:390–413
29. Steduto P, Hsiao TC, Raes D, Fereres E (2009) Aquacrop—the FAO crop model to simulate yield response to water: I. Concepts and underlying principles. Agron J 101(3):426–437
30. Boote KJ, Jones JW, Pickering NB (1996) Potential uses and limitations of crop models. Agron J 88:704–716
31. Amjed A, Sanjani S, Hoogenboom G, Ahmad A, Khaliq T, Wajid SA, Noorka IR, Ahmad S (2012) Application of crop growth models in agriculture of developing countries: a review. New Hor Sci Technol 1(4):95–99
32. Langensiepen MF, Bergamaschi M, Grasle W, Scholberg J (1999) Are the crop models universally applicable? In: Proceedings of International Symposium on Modeling Cropping Systems, June 21–23, 1999, Lleida, Spain
33. Evers JB, Vos J, Buck-Sorlin GH (2008) Considering plant structure in models of plant growth and development. In: Keulen HV, Van Laar HH, Rabbinge R (eds.) 40 years theory and model at Wageningen UR, Wageningen University and Research Centre, The Netherlands, pp 15–19
34. Xinyou Y, Struik PC (2012) Modelling gene-trait-crop relationships: past experiences and future prospects. Acta Hortic 957:181–189
35. Meinke H, Stuik PC, Vos J, Der Werf WV (2008) Modelling that bridges scales and connects disciplines. In: Keulen HV, Van Laar HH, Rabbinge R (eds) 40 Years Theory and model at Wageningen UR, Wageningen University and Research Centre, The Netherlands, pp. 37–45
36. Ho QT, Verboven P, Yin X, Struik PC, Nicolai BM (2012) A microscale model for combined CO_2 diffusion and photosynthesis in leaves. PLoS ONE 7(11):e48376

Authors Index

S. K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9,

Index

S. K. Basu, Naveen Kumar (eds.), *Modelling and Simulation of Diffusive Processes,* Simulation Foundations, Methods and Applications, DOI 10.1007/978-3-319-05657-9,