UNIVERSITA' DEGLI STUDI DI PADOVA

Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate

DOTTORATO DI RICERCA IN MATEMATICA COMPUTAZIONALE

XII CICLO

Mixed Finite Elements and Finite Volumes for the solution of density dependent flow and transport of radioactive contaminants in porous media

Coordinatore: Ch.mo Prof. Renato Zanovello

Supervisori: Ch.mo Prof. Giorgio Pini *e*
Ch.mo Prof. Donato Trigiante

Dottoranda: Annamaria Mazzia

31 Dicembre 1999

Contents

Ringraziamenti (non formali)

Desidero ringraziare un gran numero di persone il cui aiuto, diretto o indiretto, esplicito o implicito, è stato molto importante per i miei studi di Dottorato e per la preparazione della mia tesi.

Prima di tutto, ringrazio i miei relatori, il Prof. G. Pini (Università di Padova) e il Prof. D. Trigiante (Università di Firenze): mi hanno permesso di allargare gli orizzonti delle mie conoscenze su vari campi della matematica applicata.

Un grazie particolare va anche al Prof. G. Gambolati (Università di Padova): ha sempre mostrato interesse sulla mia *familiarizzazione* con un argomento di ricerca in qualche modo nuovo e inusuale per un matematico. Più di una volta mi ha chiesto se mi trovassi bene a lavorare in un gruppo con più ingegneri (come lui) che matematici!

Ma un grazie speciale e davvero grande va a *Mario* e a *Luca*, noti anche come Dr. Ing. *M. Putti* e Dr. *L. Bergamaschi* (entrambi dell'Università di Padova): hanno seguito tutti i miei studi, giorno dopo giorno, con pazienza (reciproca, per essere sinceri), aiutandomi e incoraggiandomi nei momenti difficili.

Infine, ringrazio tutti gli amici conosciuti durante questi anni di Dottorato, gli amici del DMMMSA (Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate) e gli amici *di fuori* - Felix, d. Giovanni, tutti i ragazzi della Cappella Universitaria di Padova e tutti gli altri amici sparsi un po' dovunque in tutto il mondo: sono stati un aiuto *implicito* ma immenso per i miei studi. Spero che l'amicizia con ciascuno di essi possa continuare con entusiasmo e con gioia!

E per concludere, una dedica: a papà. In questa occasione, sarebbe stato certamente molto felice e orgoglioso.

Annamaria Mazzia

Elementi Finiti Misti e Volumi Finiti per la soluzione del problema di flusso e trasporto di contaminanti radioattivi pesanti in mezzi porosi

riassunto

In questa Tesi viene sviluppato un **metodo numerico** accurato ed efficiente per risolvere problemi accoppiati di **flusso** e **trasporto** di **contaminanti radioattivi** in acque sotterranee. Il lavoro è finalizzato, soprattutto, allo studio del sito più contaminato del mondo, il **lago Karachai**, negli Urali del Sud (Russia). Questo lago fu utilizzato, fin dagli anni cinquanta, per immagazzinarvi residui radioattivi provenienti da esperimenti nucleari e, successivamente, come discarica dei rifiuti liquidi radioattivi della centrale nucleare di Mayak. Per comprendere la gravità del problema, basti pensare che la quantità di contaminanti radioattivi presenti nel lago Karachai è maggiore, in termini di radioattività, a quanto rilasciato dall'incidente di Chernobyl nel 1986. Inoltre, bastano appena poche ore sulla riva del lago per ricevere una dose fatale di radiazioni.

Questa ricerca rientra nell'ambito del progetto RaCoS, un contratto europeo attivo presso il Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate dell'Università di Padova, di cui coordinatore è il prof. G. Gambolati. Il progetto ha come finalità lo sviluppo di un modello predittivo del trasporto di radionuclidi nell'acquifero sottostante il lago Karachai. Tale modello dovrà essere utilizzato per la valutazione dell'efficienza di tecniche di contenimento atte alla prevenzione della contaminazione da sostanze radioattive dei fiumi che scorrono nelle vicinanze del lago.

La centrale di Mayak riversa nel lago sostanze radioattive quali Cs^{137} e Sr^{90} assieme a soluzioni saline ad alta densità. Queste soluzioni pesanti infiltrano nel terreno trasportando con sè i radionuclidi. La simulazione dei processi di trasporto è dunque governata dal modello matematico del *flusso di densità*: un sistema nonlineare formato dall'equazione di flusso del fluido e di trasporto dei contaminanti. Per risolvere numericamente il problema accoppiato di flusso e trasporto, noi consideriamo lo schema iterativo di Picard, risolvendo per prima l'equazione del flusso, calcolando, quindi, il campo delle velocità e risolvendo, infine, l'equazione di trasporto, che fornisce, dunque, la concentrazione del contaminante. Questa procedura è ripetuta fino a che non si ottiene convergenza (pag.93). Per ottenere soluzioni accurate, è fondamentale che sia accurato il campo delle velocità. Inoltre, visto che in problemi reali i domini sono irregolari, è opportuno operare su griglie non strutturate. L'originalità di questo lavoro consiste nel risolvere il problema di flusso e trasporto (pag. 89) combinando un metodo agli **Elementi Finiti Misti Ibridi** (EFMI) standard per la soluzione del problema di flusso, con una tecnica originale di **time-splitting** per la soluzione del problema di trasporto, entrambi sviluppati su griglie non strutturate.

L'approccio degli EFMI garantisce proprietà come quella della conservazione della massa e velocità conservative, assicurandoci un campo di velocità accurato.

La tecnica di time-splitting per il trasporto ci permette di ottenere soluzioni altrettanto accurate per la concentrazione, in quanto i due termini di **avvezione** e **dispersione** nel trasporto vengono risolti separatamente considerando le discretizzazioni *migliori* che garantiscano soluzioni efficienti e, per il caso dell'avvezione, capaci di catturare bene gli shocks e le discontinuità. Perció uno schema ai **Volumi Finiti** (VF) è utilizzato per risol-

vere la parte avvettiva dell'equazione di trasporto, mentre lo schema agli EFMI è usato per la parte dispersiva. La scelta è stata guidata oltre che da ragioni di accuratezza, robustezza ed efficienza dei due metodi, anche dal fatto che sono entrambi basati sulla formulazione debole dell'equazione di partenza e usano simili spazi funzionali.

Il time-splitting è spiegato nel Capitolo 4, dove viene dapprima descritto uno schema del second'ordine nello spazio e del prim'ordine nel tempo, con la possibilità di utilizzare diversi passi temporali per l'avvezione e per la dispersione (pag. 65), e, in seguito, viene considerata un'estensione al second'ordine anche nel tempo (si vedano, a proposito, il teorema e il lemma di pag. 79 e pag. 84, rispettivamente). Gli esempi numerici confermano la validità del metodo, in quanto le soluzioni rivelano diffusione numerica minima senza oscillazioni, conservano la simmetria e sono capaci di catturare bene gli shocks.

Avendo, dunque, due strumenti validi per risolvere il flusso e il trasporto, si può affrontare la risoluzione del problema accoppiato, cui è dedicato tutto il Capitolo 5. La procedura numerica viene sperimentata su un caso test della letteratura, il problema di Elder (pag. 95): i risultati confermano che l'approccio utilizzato è accurato, non presenta oscillazioni numeriche nè introduce eccessiva diffusione numerica come accade con metodi più convenzionali. Inoltre, la presenza di griglie non strutturate non influenza la soluzione ma le simmetrie sono ben conservate. Tutto ciò suggerisce di utilizzare questa tecnica per risolvere situazioni più complesse, come ad esempio, quella del lago Karachai, cui è dedicata l'ultima parte del Capitolo (pag. 96).

I Capitoli 4 e 5 evidenziano, dunque, i risultati ottenuti nel corso della ricerca di Dottorato. La parte iniziale della Tesi è, invece, di carattere più introduttivo. Infatti, mentre il Capitolo 1 ci dà una visione generale del lavoro svolto, i Capitoli 2 e 3 forniscono gli strumenti di base utilizzati per risolvere il problema accoppiato di flusso e trasporto, cioè i VF e gli EFMI. In questo modo si ha un quadro completo ed esauriente delle equazioni di avvezione e di dispersione e dei metodi numerici utilizzati per risolverle. L'appendice finale, inoltre, presenta alcune note riguardanti l'uso delle norme per il calcolo dell'errore, ed alcuni rudimentali concetti di idrologia.

Mixed Finite Elements and Finite Volumes for the solution of density dependent flow and transport of radioactive contaminants in porous media

Abstract

In this Thesis a **numerical method** that is accurate and efficient at the solution of the coupled **flow** and **transport** problem in groundwater is developed. The work is particularly aimed at studying the most contaminated site on earth, **Lake Karachai**, in the South Urals (Russia). In the 1950s, this lake was used as storage reservoir for medium level radioactive wastes from nuclear tests, and as a repository for liquid radioactive wastes from the nearby Mayak nuclear plant. The problem is very serious: the total amount of radioactivity in Lake Karachai is comparable to that released by the 1986 Chernobyl accident. For example, a person standing on the shoreline of the lake may receive a fatal dose of radiation in just few hours.

This research is part of the RaCoS project, a Research contract funded by the European Union, whose coordinator is the Department of Mathematical Methods and Models for Scientific Applications of the University of Padova (Prof G. Gambolati). The project is aimed at devoloping a predictive model of radionuclides transport in the aquifer underlying Lake Karachai. This model must be used to evaluate the effectiveness of containment measures for preventing radioactive contamination of the rivers flowing near the lake.

The Mayak plant dumps in the lake radioactive wastes containing cesium-137 and strontium-90 together with high density saline solutions. These heavy solutions infiltrate through the soil and carry with them the radionuclides. The simulation of the waste transport processes is thus governed by the so-called mathematical model of *density-driven flow*: a nonlinear system of partial differential equations formed by the coupled fluid flow and contaminant transport equations. The usual procedure to numerically solve the coupled flow and transport problem is given by the Picard iterative procedure, by which first the flow equation is solved, then the velocity field is calculated, and, finally, the transport equation is solved for the concentration values. The iteration is repeated until convergence is achieved (pag.93).

To obtain accurate solutions, accurate evaluation of velocity fields is required. In addition, since real world simulations generally involve irregular boundaries, it is important to work on general unstructured grids. The originality of this work consists in solving the flow and transport problem (pag. 89) by combining a **Mixed Hybrid Finite Element** (MHFE) method for flow with a **time-splitting** technique for transport, both developed on unstructured grids.

The MHFE approach preserves properties as conservation of mass and conservative velocities, leading to accurate velocities field. The time-splitting technique for transport allows accurate solution for concentration, since **advection** and **dispersion** in transport are solved separately, by considering the *best* discretizations available to get efficient solutions and, in the case of advection, capable of well capturing shocks and discontinuities.

To this aim, a **Finite Volume** (FV) scheme is used for solving the advection term of the transport equation, while the dispersive flux is discretized using a MHFE technique. The choice of these two schemes is dictated, on one hand by their accuracy, robustness and

efficiency in handling nonuniform meshes and highly variable coefficients. On the other hand, both FV and MHFE are based on the weak formulation of the governing equation and use similar functional spaces for the approximation of the dependent variable, making them ideally suited for combination in a time-splitting approach.

The time-splitting technique is developed in Chapter 4. First a second order accurate in space and first order accurate in time scheme, with different time step sizes for the advection and dispersion terms is presented (p 65). Next its extension to second order accuracy in time is considered (see theorem and lemma in pp 79 and 84, respectively). Numerical examples ascertain the validity of the method and show that the solutions obtained with the proposed scheme display small amount of numerical diffusion, preserve simmetry and well capture discontinuities.

With the tools thus developed the coupled flow and transport equations, to which the entire Chapter 5 is devoted, are tackled. The numerical procedure is tested on a case taken from literature, Elder's problem (p 95): the results confirm that the approach is accurate, does not present numerical oscillations, and does not introduce excessive numerical diffusion, as it may happen with other types of discretization approaches. Furthermore, the presence of unsymmetric meshes does not dramatically influence the solution, but simmetries are well preserved. This suggests that the technique developed to solve the coupled flow and transport problem by MHFE method and the time splitting technique can be used to simulate more complex situations, in particular, the Lake Karachai problem, to which the final part of the Chapter is devoted (p 96).

As an overview, Chapters 4 and 5 underline the main results obtained during the PhD research. Instead, the first part of the Thesis is more introductory. Chapter 1 gives a general idea of the work done, while Chapter 2 and 3 provide the basic tools for solving the coupled flow and transport problem, that is the FV and MHFE methods. In this way, we try to present a complete and exhaustive picture about advection and dispersion equation and numerical methods to solve them.

Finally, in the appendix a few notes about the use of norms to compute error, and some rudimentary concepts of hydraulic engineering are reported.

1 Introduction to the problem and motivation for the study

In 1997 the RaCoS Project (Radionuclide Contamination of Soil and Groundwater at the Lake Karachai Waste Disposal Site (Russia) and the Chernobyl Accident Site (Ukraine)) was financed by the European Union D.G. XII under the 4th Framework RTD premises [48]. It is an INCO/COPERNICUS project (Cooperation with NIS Countries), whose coordinating partner is the Department of Mathematical Methods and Models for Scientific Applications of the University of Padova (Prof G. Gambolati). The other partners are the Center for the Research and Development of High Studies in Sardinia (CRS4), the Department of Water Management of Delft University of Technology, the Institute of Environmental Geology of Russian Academy of Science in St. Petersburg, and the Research and Development Center of Radioecological Studies of the National Academy of Science in Ukraine.

The main objectives of the RaCoS project are to implement appropriate modelling techniques to investigate the characteristics of radioactive contaminants and the processes that control their spreading, to assess the present and future situation with respect to the contaminated area, and to evaluate containment and remediation measures for the mitigation of pollution and the restoration of subsurface resources at the Lake Karachai (Russia) and Chernobyl (Ukraine) sites [49].

For this purpose, the problem at the Lake Karachai site is concerned with the simulation of the movement of radioactive heavy brines (nitrates carrying with them radionuclides) infiltrating from the lake into the Quaternary aquifer and possibly threating contamination of rivers in the area [48]. The modeling study is performed by means of a three-dimensional Finite Element code (CODESA-3D) solving the coupled density dependent flow and transport groundwater equations [24]. During the course of the study, concerns were expressed about the effects of the large amount of numerical diffusion necessarily introduced by CODESA-3D to maintain stability. It was argued in fact that artificial dispersion and/or negative concentrations (even though with small absolute values) could drastically affect the fate of the contaminant by altering the adsorption/desorption pattern of the nitrates onto solid grains, thus changing the speed at wich the contaminant plume moves. For this reason, it was decided to develop a highly accurate two-dimensional model used to verify the results of CODESA-3D.

In this Thesis, we develop a two dimensional code based on the combination of a Mixed Hybrid Finite Element (MHFE) method for the solution of the flow equation, with a time-splitting technique to solve the transport equation. In the time-splitting technique dispersion and advection fluxes are splitted into two separate partial differential equations (PDEs) and discretized by a triangle-based high resolution TVD Finite Volume (FV) scheme and a MHFE technique, respectively. The combination of MHFE and FV in this fashion guarantees a high accuracy of the solution with introduction of minimal numerical diffusion while maintaining stability and without ever producing negative concentration values.

FIGURE 1.1: *Location of the Chelyabinsk province*

1.1 The Lake Karachai: an overview

There are several surface reservoirs located within the South Ural province of Chelyabinsk in Russia (see Figure 1.1) that have been operated for over 45 years for storage of medium and low level radioactive liquid wastes. The greatest threat to the subsurface environment is represented by leakage from the Lake Karachai, filled with medium level wastes, that has caused a subsurface contaminant plume. According to the Worldwatch Institute report of 1991 on nuclear waste, Lake Karachai is the *most polluted site* on earth. Lake Karachai was constructed inside the grounds of the Mayak Chemical Combine, Russia's main nuclear reprocessing plant and formerly a major nuclear weapons production site. Since the 1950s it was used as a storage reservoir for medium level wastes from nuclear tests, and to hold liquid nuclear wastes previously dumped in the River Techa by the Mayak NPP (Nuclear Power Plant). It is estimated that from 1949 to 1951 Mayak discharged about 2.76 million Ci of liquid radioactive wastes into the River Techa.

When the River Techa was dammed in 1956 and 1963, releases of irradiated waters were reduced. However, the series of pools and marshes created by the dams continue to be a source of pollution. The Assanov Marshes, which cover a 30 km^2 area below dam number 11, contains about 6000 Ci of strontium-90 and cesium-137 and is a continuous open source of radioactivity into the River Techa.

In 1957 one of the storage tanks containing radioactive sediments exploded, spreading irradiation over an estimated 23000 km^2 area. After this disaster, about 10500 people along the Techa river were evacuated. Contamination of the area was more than 1 Ci/km^2. About 20 millions of different radionuclides were released to the environment as a result of the accident and led to the formation of the East Ural Radioactive Tracer.

It is estimated that the total accumulated activity of these wastes is up to 120

millions Ci of high-level waste, which is comparable to the total amount released by the 1986 Chernobyl accident [48]. Another point of comparison is offered by the fact that the total amount of radioactive discharges from the Russian Navy fleets adds up to only 0.5 per cent of the radioactivity in Lake Karachai. In 1991, U.S. experts measured a dose rate of 300 to 600 millirems per hour near the shores of the lake, which is three to six times the exposure permitted per year by U.S. regulations. It was estimated that just one minute standing on the shore without full protection would mean certain death. At the present time, a person standing at some points on the shoreline would receive a fatal dose of radiation in a few hours.

Efforts have been made to fill in the lake with large rocks and concrete, and just over a third of the project to *close* Lake Karachai has been completed. The decontamination plan involves covering the area with u-shaped blocks to keep radioactive silts at the bottom of the lake, and filling in the water with gravel. The project has received only meagre funding each year, and does not yet include any measures to prevent downward contamination.

If the underground pollutants beneath Karachai reach the Irtysh River system, the radioactive contamination could eventually reach the Artic Ocean. The plume of polluted water from the lake is drifting at a rate of approximately 80 m/y toward the Irtysh water system.

The plume is made up mainly of heavy brines (nitrates) that carry with them the radioactive ions. Being heavier than pure water, the brine movement is driven downward by gravity effects and spread laterally by the regional flow regime. Because of the high density of these brines ($\rho \sim 110$ Kg/m^3) gravity effects introduce instabilities in the flow regime causing fingering in the concentration plume. The main flow direction is however downward and the pollutants tend to accumulate at the bottom of the aquifer.

The Lake Karachai site represents a unique field laboratory for studying and solving complex problems of radionuclide migration. In fact, monitored concentrations of some radionuclides exceed the safety standard level by up to 6 orders of magnitude. Increasing in volume, the plume moves toward the zones of groundwater discharge - the velocity of the advancing front of the contaminant plume ranges between 0.2 and 1.5 m/d - where observations reveal that radionuclides are already being released into the surface environment at streams and well-fields, thus potentially contaminating springs and source of drinking water. An analysis of monitoring data shows that the contaminant plume is of ellipsoidal shape in plan view and that the preferential flowpaths of brines differ considerably from the direction of the regional natural-gradient flow. Cross sectional groundwater quality stratification results from the differences in density between radioactive brines and ambient water.

The geological cross section at the Lake Karachai site is represented by fractured moderately metamorphosed effusive Lower-Silurian rocks. Under natural conditions groundwater is recharged primarily by precipitation. Moreover, being the Lake Karachai located within the watershed between two river valleys (the Myshelyak and Techa, respectively), these valleys serve as groundwater discharge area. The total aquifer transmissivity has been estimated according to the data of about 200 multi-well pumping tests and slug tests. The test results have demonstrated the complex and strongly heterogeneous nature of the transmissivity in this area.

The brines recharging into the aquifer are radioactive and of high density. There

are about 100 observation and prospecting wells at the Lake Karachai site. It has been established that the solutions contain a wide range of long-living artificial radionuclides (such as strontium-90, cesium-137, ruthenium-106, cobalt-60, cerium-144, and isotopes of uranium and plutonium).

The following physical and chemical processes are considered to be among the most important at the Lake Karachai site: density-dependent advection, which dramatically enhances the displacement of natural subsurface water and results in highly irregular movements of the brines; advection due to natural-gradient flow in strongly heterogeneous fractured rocks; longitudinal and transversal dispersion; diffusion of radionuclides into porous matrix; radioactive decay; physical and chemical interactions, in particular, adsorption onto fracture walls. This last process retards the movements of individual radionuclides with respect to the transport of principal chemical components. Moreover, depending on the process of interaction, the mass transport potential of the contaminant may be either enhanced or deteriorated [48].

One of the actions included in the decontamination plan for the Mayak area consists in eliminating the discharge of radioactive waste in the Lake Karachai. After closing the lake as a liquid reservoir, the major input of radioactive contaminants from the surface will be caused by infiltration of water moving through the lake bottom deposits and the underlying rock. As this water is supposed to be of a low total solid content, it would not follow the dense plume pathways. Therefore, this part of radioactive contamination will be spreading to the river Myshelyak valley in the North-East direction, in accordance with the main flow in the most permeable section of the aquifer. As the radionuclides transport in this case would not be controlled by the brine movement, the appropriate plume is supposed to be characterized by a very wide transition zone, which makes the reliable forecast of radionuclides inflow into the river rather questionable. The appropriate scenarios of the process are currently under investigation.

As for the dense plume remnants, their further fate is two-fold. The brines will partly spread to the lower portion of the aquifer and the underlying aquitard while the lighter part will tend to move upwards towards the river Myshelyak. All this means that the radionuclides will go upward mostly along the river valley. In this connection protective measures against surface water contamination need to be substiantiated. The groundwater technical remediation at this site seems unreal. The most radical measure of this sort could be connected with "pump-and treat" approach. At the same time, it is most questionable that such an approach could be realized on practice. Natural remediation approach, i.e. the underground spaces are sacrified in favor of the surface water protection and used as a natural reactor, seems the only reliable alternative. Such an approach relies mostly upon some rigid limitations for the water usage and upon the natural attenuation effects of the groundwater [31]. Containment strategies need to be investigated.

The mathematical model of the Lake Karachai waste disposal site requires the analysis of brine dynamics. Since the brine component affects fluid density, and the hydrodynamic dispersion is depedendent on the local velocity field and brine concentration gradient, the mathematical model of the Lake Karachai, represented by flow and advective-dispersive transport of the dense brine, is strongly coupled. These coupling cause nonlinearities in the equations and preclude analytical solutions, with significant challenges for numerical simulation. Another major source of nonlinearity enters the equations through velocity dependent hydrodynamic dispersion and the advective transport [41, 44].

For all these reasons, we need accurate evaluation of the velocity field and of the advective part of the trasport equation to obtain an accurate simulation of the brine movement.

1.2 The mathematical model

The mathematical formulation governing the contamination process at the Lake Karachai is given by a coupled system of variably saturated flow and miscible salt transport equations [24]

$$\begin{aligned} \sigma\frac{\partial\psi}{\partial t} &= \vec{\nabla}\cdot[K_s\frac{1+\epsilon c}{1+\epsilon' c}K_r(\vec{\nabla}\psi+(1+\epsilon c)\eta_z)]+ \\ &\quad -\phi S_w\epsilon\frac{\partial c}{\partial t}+\frac{\rho}{\rho_0}q^*+q \qquad (1.1a) \\ \vec{v} &= -K_s\frac{1+\epsilon c}{1+\epsilon' c}K_r(\vec{\nabla}\psi+(1+\epsilon c)\eta_z) \qquad (1.1b) \\ \phi\frac{\partial S_w c}{\partial t} &= \vec{\nabla}\cdot(D\vec{\nabla}c)-\vec{\nabla}\cdot(c\vec{v})+qc^*+f \qquad (1.1c) \end{aligned}$$

The above system will be studied in detail in Chapter 5, with appropriate initial and boundary conditions that complete the mathematical formulation of the flow and transport problem. Here it is important to underline that equation (1.1a) represents the mass balance for the mixture (or *flow* equation); equation (1.1b) is the Darcy's velocity, while equation (1.1c) is the mass balance for the salt (or *transport* equation). For the coupled system of equations describing the movement of density-dependent flow in aquifers, the usual procedure is to decouple the problem by first solving the flow equation for the unknown pressure head ψ and the velocity field $\vec{v}$, next solving the transport equation for the unknown concentration c. Convergence of this iterative procedure is obtained when the absolute norm of the differences between the pressure head and concentration corresponding to two successive iterations fall below a prescribed tolerance.

The numerical solution of the coupled flow and radioactive contaminant transport model in aquifers requires the accurate evaluation of Darcy's velocities fields. In the simulation of real world aquifers general boundaries consistent with different geological characteristics need to be employed. It is thus important to apply the numerical discretization of the governing equation in general unstructured grids. This approach may lead to innacurate calculation of velocities, and thus to innacurate predictions and numerical difficulties. Accurate numerical flow field may be obtained using the Mixed Hybrid Finite Element approach. MHFE preserves important properties, such as conservation of mass, and leads to accurate solutions [4].

Accurate simulation of radionuclide contaminant transport may be achieved by a combined use of the Finite Volume method with high-resolution upwind schemes and MHFE. FV techniques in conjunction with MHFE offer an ideal framework for the solution of the numerical difficulties arising when advection dominates dispersion. These difficulties translate into oscillations in the numerical solution that may deteriorate both the accuracy of the prediction and the convergence of the nonlinear iterations. The combination of FV and MHFE requires the use of a time-splitting technique, whereby advec-

tion is approximated by a Godunov-type procedure, and diffusion by a MHFE method [9, 10, 11, 12, 37, 35, 36].

The results obtained for the time-splitting technique show good approximation to the transport equation without introducing oscillation or numerical diffusion. Therefore this technique is used to solve equation (1.1c) in the simulation of the Lake Karachai site, together with the MHFE method that solves equations (1.1a) and (1.1b).

1.3 Literature review

The key words to proceed with the construction of the numerical method for the simulation of the Lake Karachai site are the following:

- advection equation, Finite Volume scheme;
- dispersion equation, Mixed Hybrid Finite Element method;
- advection-dispersion equation, time-splitting technique;
- density dependent coupled flow and transport of radioactive contaminant in porous media, MHFE for the flow equation, time-splitting for the transport, Picard iteration.

Advection equation, Finite Volume scheme. Advection equation, a particular hyperbolic conservation law, is involved in many practical problems of science and engineering. In general it is not possible to derive exact solutions of this equation, and therefore we have to devise and study numerical methods. Special difficulties are associated with solving advection equations, that must be dealt with carefully in developing numerical methods. In fact, methods based on standard finite difference approximations may behave well for smooth solutions but can give disastrous results when discontinuities are present. Therefore, the theory known about the mathematical structure of these equations and their solutions needs to be exploited to develop appropriate methods that overcome some of the numerical difficulties arising from a more standard approach.

Many methods have been derived for advection equations based on finite difference approximation. A survey of these methods is given in [5, 27, 28, 33]. Methods developed using standard finite difference discretizations are inaccurate near discontinuities, since they are based on truncated Taylor series expansions. Thus, different approaches need to be considered.

A well known class of numerical methods that solve hyperbolic conservation laws are the Godunov-type methods [42]. These methods use, in some way, the exact solution of the Riemann problem and do not introduce numerical oscillations or discontinuities. Unfortunately, these methods are only first order accurate; hence the solutions are smoothed around discontinuities. Based on Godunov methods, other methods have been developed, the high-resolution methods, which are second order accurate in smooth regions and give good results, i.e. no oscillations, around shocks [52, 59]. In particular second order total variation diminishing (TVD) schemes [55] and essentially non oscillatory (ENO) schemes [40] (which surpass the second order accurate barrier associated with TVD methods) eliminate unphysical spurious oscillations. A survey of Godunov-type methods and high-resolution methods is given in [25, 33, 54, 57].

Historically, these numerical schemes have been first derived for one dimensional scalar conservation laws. Extensions to multidimensions are often obtained by means of Finite Volume formulations on structured meshes [42]. These are tipically designed in a dimension by dimension fashion and inherit all the limitations of a structured regular grid. On the other hand, unstructured grids, such as triangular meshes, offer a greater flexibility when dealing with complex geometries and limit grid orientation effects. However, extensions of higher order schemes for hyperbolic conservation laws on unstructured discretizations is not so immediate [46, 16]. A survey of different approaches for the solution of hyperbolic conservation laws on unstructured grids is shown in [25].
In this Thesis we consider the scheme developed by [16, 34]: it is based on a FV type discretization and relies on a very local adaptive interpolation idea, which results in computational efficiency.

Dispersion equation, Mixed Hybrid Finite Element method. The numerical solution of dispersion equation, the groundwater flow equation (1.1a) in conjunction with (1.1b), whose unknown are pressure and velocity, or the dispersion part related to (1.1c), whose unknown is concentration, yields a set of discrete values of pressure and velocity, or concentration and dispersion flux, respectively.

In the first case, it is important to obtain accurate velocities, especially when pressure and velocity fields are used for the solution of a contaminant transport problem. Accuracy of the discrete velocities is a necessary but not sufficient condition for a correct solution of the transport equation. Most importantly, the velocity fields need to conserve mass. The approach used with Finite Element (FE) simulations may lead to violation of the mass conservation principle, and thus to inherently inaccurate contaminant fate predictions [18, 43]. The Mixed Finite Element (MFE) method provides an attractive framework for these type of problems, because it approximates pressure (or concentration) and normal fluxes (velocities or dispersive fluxes) simultaneously, and satisfies the mass conservation principle. At the same time, it maintains the flexibility of FE in handling general boundary conditions and domain shapes. The MFE method has been extensively used for the solution of parabolic equations arising in different application fields, such as groundwater flow or petroleum reservoir simulation. In steady-state problems, i.e. elliptic equations, the indefinite mixed matrix system becomes ill-conditioned. A common solution method is the Mixed Hybrid Finite Element technique. Through the definition of an extra variable representing the pressure (or concentration) head at element edges, MHFE gives rise to a symmetric positive definite system matrix with good conditioning properties. An exhaustive analysis of Mixed and Hybrid Finite Element methods can be found, for example, in [6, 47, 51].

Special consideration has been devoted to lowest order mixed finite elements [51], that display global first order of accuracy for both pressure head and velocities fields. In linear problems, under suitable conditions on the mesh and on the regularity of the solution of the continuous problem, it has been proved that this method achieves superconvergence for the pressure head and the normal fluxes in specific points of the mesh [13]. For general triangular meshes, second order accuracy at midpoint edges in grids formed by triangles having edges parallel to three different lines has been observed [14].

A development of a two-dimensional MHFE model for the solution of the nonlinear

equation of variably saturated flow in groundwater on unstructured triangular meshes is given in [4]. This approach is followed in this Thesis.

Advection-dispersion equation, time-splitting technique. Approximation of the advection-dispersion equation (1.1c) leads to difficulties when advection dominates because sharp concentration front tend to develop and move without changing form. Standard finite difference and finite element methods may not work well for problems with sharp fronts, showing non-physical oscillations and numerical diffusion. Two approaches are generally used to overcome these phenomena. One is based on the definition of a proper control volume where upwind techniques can be used for approximating the advective flux. In this case the stability of the scheme is obtained by adding an amount of numerical diffusion that is dependent on the approach used [42, 28]. The other class of methods originates from the splitting of the dispersion and advection fluxes into two separate partial differential equations containing one the dispersive and the other the advective term, respectively. These two equations are then discretized, each with the technique deemed most appropriate. Splitting allows the combination of explicit time-stepping for advective fluxes with implicit time-stepping for dispersive fluxes. This approach lessens the stability constraint connected with explicit discretization of the dispersion term but maintains the possibility of using efficient explicit schemes for advection. Belonging to this class are the Eulerian-Lagrangian schemes [39, 7] or the fully Eulerian Godunov-Mixed Methods (GMM) [9, 10, 11, 12]. In this latter approach, a time-explicit, spatially second-order accurate Godunov method is used to treat advection, and a time-implicit, spatially second order accurate Mixed Finite Element method is used for modeling dispersion.

The time-splitting technique developed in this Thesis is similar in spirit to the GMM approach, by considering a triangle-based, high resolution FV scheme [46, 16, 34] to discretize the advective term, while the dispersive flux is discretized using a Mixed Hybrid Finite Element (MHFE) technique.

Density dependent flow and transport of radioactive contaminants in porous media, MHFE for the flow equation, time-splitting for the transport, Picard iterations. Coupled flow and transport equations (1.1) represent the mathematical model of density driven contamination in groundwater [1, 22, 23, 24, 30].

Coupling in system (1.1) is due to the concentration terms in the flow equation (1.1a) and the head terms that appear in the transport equation (1.1b) via the Darcy velocities. In the simpler case of non-density dependent flow and transport, the system is coupled only through the head terms in the transport equation. In this case there is physical coupling, but mathematically the system can be reduced and solved sequentially, first the flow and then the transport equation, without iteration. In the density dependent case, the system is irreducible and any sequential solution procedure requires iteration. To this purpose the most commonly used algorithm is given by a Picard iterative procedure, as explained in [24, 41, 44, 45].

1.4 Schematical content of the Thesis

This work is organized in two parts. In the first we present a review of the FV and MHFE method, fundamental tools to develop the second part, where the time-splitting technique to solve transport equation and some applications to the coupled flow and transport problem (and, in particular, to the Lake Karachai problem) are exposed.

1.4.1 Part I. Numerical Tools

Chapter 2 is devoted to the numerical approximation of advection equations by the FV method. To this aim, the resolution of the Riemann problem for hyperbolic conservation laws is given, including the classical notions of integral and differential form, classical and weak solutions, rarefaction and shock waves, entropy conditions.

Numerical approximation of hyperbolic conservation laws by Godunov-type methods are considered next: for the sake of completeness special emphasis is devoted to the study of Godunov and van Leer's methods for one-dimensional spatial case. Numerical examples with both periodic and non periodic boundary conditions validate the theoretical statements.

The two-dimensional spatial case is studied at the end of Chapter 2. After the definition of the Finite Volume method and the presentation of general principles and properties, the attention is focused on a high resolution triangular FV scheme by following the scheme developed in [16] and next modified by [34]. Numerical experiments verify the numerical convergence of the proposed scheme.

Chapter 3 is devoted to the development of the MHFE method. Some general definitions and theorems about the Mixed Finite Element method are given. Then, attention is directed to the presentation of the MHFE formulation by introducing the concept of the Lagrange multipliers and some superconvergence results.

Since our goal is to solve dispersive flux in groundwater, we develop its numerical approximation by using the MHFE method. Some numerical results complete the description.

1.4.2 Part II. New results

Chapter 4 is devoted to the development of a time-splitting technique. This novel approach to solve advection-dispersion equation is dictated by the fact that standard finite difference and finite element methods may not work well for problem with sharp fronts, showing non-physical oscillations. The time-splitting technique overcomes this phenomenon combining numerical stability with minimal artificial diffusion: the dispersion and advective fluxes are splitted into two separate partial differential equations, containing one the dispersive and the other the advective term, respectively. Then, these two equations are discretized, each with the technique deemed most appropriate. A time-explicit, spatially second order accurate Godunov method is used to treat advection, and a time-implicit, spatially second order accurate MFE method is used for modeling dispersion. We combine a triangle-based, high-resolution FV scheme for advection, with a MHFE method for dispersion. Some of the original results obtained with this technique are presented and showed in this Chapter.

We first consider a scheme of second order accuracy in space and first order in time. Numerical tests on an analytical one dimensional example ascertain the convergence prop-

erties of the scheme when advection or dispersion is dominant. Results on a realistic test case of groundwater contaminant transport confirm the validity of the proposed scheme. Then, the study of this technique is extended to obtain second order accuracy also in time. The theoretical presentation is accompanied by some numerical results.

Chapter 5 is devoted to the application of the time-splitting technique in conjunction whith the MHFE method discretization of the flow equation, to solve the coupled flow and transport problem. To this aim system (1.1) is studied in detail, taking into account coupling and nonlinearities. The numerical solution is obtained by means of the Picard iterative scheme.

A first application to validate the proposed method concerns Elder's problem, that represents a fluid flow driven purely by density differences. The results indicate that the proposed approach is accurate and reliable, does not suffer from numerical oscillation and does not introduce large amounts of numerical diffusion, as typically done by more conventional upwind discretizations. Also, the presence of unsymmetric meshes does not dramatically influence the solution, as reported in the literature for other types of discretization approaches. This suggests that the technique developed may be efficiently used to simulate more complex situations. Therefore, we arrive at the object of all our efforts: application to the Lake Karachai problem.

Finally, an appendix concludes the work: we report some observations, which come to light during the study of the numerical experiments, about the proper use of norms, and some notes regarding modeling groundwater that may seem obvious to hydraulic engineers but not too much obvious to mathematicians.

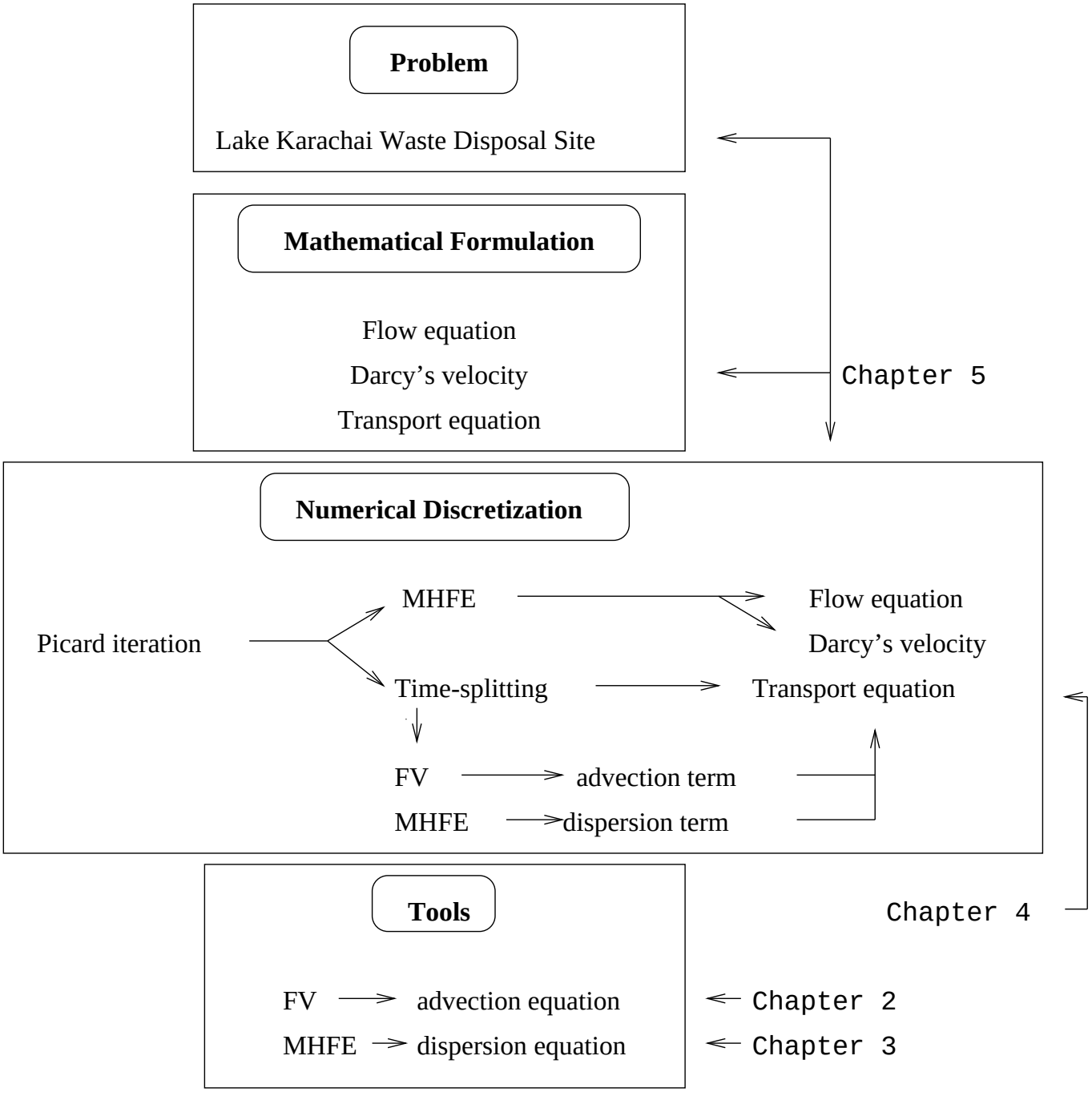

FIGURE 1.2: *Schematical content of the Thesis*

Part I
Numerical tools

2 Finite Volume Methods for the advection equation

In this Chapter we study the numerical solution of the linear advection equation, a particular case of a hyperbolic conservation law. Our goal is to develop an accurate method to solve advection-dispersion equation by a time-splitting technique, taking into account the *best* discretizations available to solve advection and dispersion, respectively. To this purpose, we consider the solution of the advection equation by Godunov-type and high-resolution methods, in one and two dimensions. After a short review of numerical solutions of hyperbolic conservation laws, we describe one dimensional high resolution Finite Volume (FV) methods and report some theoretical and numerical results. Next, we concentrate on triangle based adaptive stencils for hyperbolic conservation laws. Theoretical results are verified by numerical experiments.

2.1 Introduction to Hyperbolic Conservation Laws

The study of numerical solution of hyperbolic conservation laws is an important and interesting field of research in itself because of the special difficulties associated with the presence of shocks and discontinuities in the solution. Numerical methods based on simple finite-difference approximations may behave well for smooth solutions but can give disastrous results when sharp fronts occur [27]. The study of linear conservation laws is important for understanding the behavior of a numerical scheme, but it is also very important to consider that the introduction of nonlinearities changes dramatically the nature of the problem because it induces a loss of the uniqueness of the solution [33, 54]. The solution that is physically relevant has then to be properly characterized and the numerical approximations have to respect this characterization otherwise they would converge to a solution which has no physical meaning.

2.2 Integral and Differential Form

Hyperbolic scalar conservation laws may be expressed by the following equation:

$$\frac{\partial u}{\partial t} + \frac{\partial f}{\partial \vec{x}} = 0 \tag{2.1}$$

where $u : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^n$ is a n-dimensional vector of conserved quantities, or state variables, t is time, $f = f(u)$ is a n-dimensional vector called flux function and $\vec{x} \in \mathbb{R}^d$ is the spatial coordinate system. In the following, for simplicity, we consider the case $n = 1$ and $d = 1$.

Equation (2.1) derives from physical principles. As an example, we consider the equation for conservation of mass in a one-dimensional groundwater contamination problem. Assume that no diffusion nor mechanical dispersion occur and that density of the contaminant and the water and velocity of the fluid mixture (contaminant-water) are constant.

Let $\rho(x, t)$ be the density of the contaminant at point x and time t. This density is defined in such a way the total pollutant mass in two sections 1 and 2 is given by the integral of the density taken between x_1 and x_2. If no source or sinks are present, the mass contained in the volume thus identified can change only because of the mixture flowing across the endsections in x_1 or x_2.

Let $v(x,t)$ be the velocity of the fluid mixture at point x and time t. Then its rate of flow, or flux, past this point, is given by the product $\rho(x,t)v(x,t)$.
The rate of mass change in $[x_1, x_2]$ is given by the difference in fluxes at x_1 and x_2:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{x_1}^{x_2} \rho(x,t)\ \mathrm{d}x = \rho(x_1,t)v(x_1,t) - \rho(x_2,t)v(x_2,t)$$

By integrating this in time from t^k to t^{k+1} ($t^{k+1} > t^k$) we obtain:

$$\int_{x_1}^{x_2} \rho(x,t^{k+1})\ \mathrm{d}x =$$

$$= \int_{x_1}^{x_2} \rho(x,t^k)\ \mathrm{d}x + \int_{t^k}^{t^{k+1}} \rho(x_1,t)v(x_1,t)\ \mathrm{d}t - \int_{t^k}^{t^{k+1}} \rho(x_2,t)v(x_2,t)\ \mathrm{d}t$$

This is the integral form of the conservation law. To obtain the differential form, we assume that the functions ρ and v are differentiable functions. Then, the following equalities hold:

$$\int_{t^k}^{t^{k+1}} \frac{\partial}{\partial t}\rho(x,t)\ \mathrm{d}t = \rho(x,t^{k+1}) - \rho(x,t^k)$$

and

$$\int_{x_1}^{x_2} \frac{\partial}{\partial x}(\rho(x,t)v(x,t))\ \mathrm{d}x = \rho(x_2,t)v(x_2,t) - \rho(x_1,t)v(x_1,t).$$

By substituting these expressions in the previous equation, we obtain:

$$\int_{t^k}^{t^{k+1}} \int_{x_1}^{x_2} \{\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}(\rho(x,t)v(x,t))\}\ \mathrm{d}x\ \mathrm{d}t = 0.$$

Since this must hold for any section $[x_1, x_2]$ and over any time interval $[t^k, t^{k+1}]$, we conclude that the integrand of this equation must be identically zero, i.e.,

$$\rho_t + (\rho v)_x = 0$$

This is the differential form of the conservation law for the conservation of the mass and can be solved in isolation only if the velocity v is known a priori or is known as a function of ρ (in this case we have a scalar conservation law for ρ), otherwise we can solve this equation in conjunction with other equations, typically with equations for the conservation of momentum, i.e. a flow equation, and so we have a system of conservation laws.

2.3 Classical and Weak Solutions

Equation (2.1) must be augmented by some initial conditions and also possible boundary conditions on a bounded spatial domain. The simplest problem is the initial value problem, or *Cauchy problem*, defined for $-\infty < x < \infty$ and $t \geq 0$. We must specify initial conditions only:

$$u(x,0) = u_0(x) \qquad -\infty < x < \infty. \tag{2.2}$$

It is very easy to see that classical solutions of (2.1)-(2.2) are constant along the *characteristics*, which are curves $(x(t), t)$ defined by

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = f'(u(x(t),t)) \quad t \geq 0 \\ x(0) = x_0 \end{cases} \tag{2.3}$$

In fact, differentiating $u(x,t)$ along one of these curves, we find the rate of change of u along the characteristics and we get

$$\begin{aligned} \frac{\mathrm{d}u(x(t),t)}{\mathrm{d}t} = & \frac{\partial u(x(t),t)}{\partial t} + \frac{\partial u(x(t),t)x'(t)}{\partial x} \\ & = u_t + f'(u)u_x \\ & = u_t + (f(u))_x \\ & = 0, \end{aligned}$$

confirming that u is constant along these characteristics. Moreover, this shows that the characteristics travel at constant velocity which is equal to $f'(u_0(x_0))$.

Simple arguments show that if $u_0(x)$ is increasing (decreasing) and $f(u)$ is convex (concave), the classical solution of (2.1)-(2.2) is well defined for all $t > 0$. However, in the general case, classical solutions fail to exist for all $t > 0$ even if u_0 is very smooth [33, 54]. This happens when $\inf_x u_0'(x) f''(u_0(x)) < 0$: then classical solutions exist only for t in $[0, T^*]$ where

$$T^* = -\frac{1}{\inf_x u_0'(x) f''(u_0(x))}.$$

At the time $t = T^*$ the characteristics first cross, the function $u(x,t)$ has an infinite slope – the wave is said to *break* by analogy with waves on a beach – and a shock forms.

We state this result in the following theorem.

THEOREM 2.1. *If we solve (2.1)-(2.2) with smooth initial data $u_0(x)$ for which*

$$u_0'(x) f''(u_0(x) \quad \textit{is somewhere negative,}$$

then the wave will break at time

$$T^* = -\frac{1}{\inf_x u_0'(x) f''(u_0(x))}.$$

Proof. Since along characteristics $u(x(t),t)$ is equal to $u_0(x_0)$, we can write $x(t) = x_0 + t f'(u_0(x_0))$. We can calculate the blow up time (i.e., the first time when two different characteristics arrive at same point (x,t)). In this case there are two points, x_0 and $\overline{x}_0$, such that

$$x = x_0 + t f'(u_0(x_0)) = \overline{x}_0 + t f'(u_0(\overline{x}_0)),$$

that is,

$$t = -\frac{\overline{x}_0 - x_0}{f'(u_0(\overline{x}_0)) - f'(u_0(x_0))} = -\frac{1}{u_0'(\xi) f''(u_0(\xi))},$$

where ξ lies between x_0 and $\overline{x}_0$. Obviously, this expression for t makes sense when $\dfrac{1}{u_0'(\xi) f''(u_0'(\xi))}$ is negative. Thus, the blow up occurs if $u_0'(x) f''(u_0(x))$ is somewhere negative: at $t = T^*$ the solution forms a shock wave. ∎

To allow discontinuities, which arise in a natural way in this situation, we define a weak solution of conservation law.

DEFINITION 2.2. *A function $u(x,t)$, bounded and measurable, is called a weak solution of the conservation law (2.1)-(2.2), if for each $\phi \in \mathcal{C}_0^1(\mathbb{R} \times \mathbb{R}^+)$, the following equality holds:*

$$\int_0^\infty \int_{-\infty}^{+\infty} [\phi_t u + \phi_x f(u)] \; dx \; dt = -\int_{-\infty}^{+\infty} \phi(x,0)u(x,0) \; dx. \tag{2.4}$$

Here $\mathcal{C}_0^1(\mathbb{R} \times \mathbb{R}^+)$ is the space of functions that are continuously differentiable with compact support, that is, $\phi(x,t)$ is identically zero outside of some bounded set and so the support of the function lies in a compact set.

In this way we rewrite the differential equation in a form where less smoothness is required to define the solution. In fact, the basic idea to define a weak solution of conservation law is to take the PDE, multiply by a smooth test function, integrate one or more times over some domain, and then use integration by parts to move derivatives off the function u and onto the smooth test function. The result is an equation involving fewer derivatives on u, and hence requiring less smoothness.

2.4 The Riemann Problem

A Riemann problem is simply the conservation law together with particular initial data consisting of two constant states separated by a single discontinuity,

$$u_0(x) = \begin{cases} u_l & x < 0, \\ u_r & x > 0. \end{cases} \tag{2.5}$$

As an example, consider Burgers' equation, in which $f(u) = \frac{1}{2}u^2$, so that our conservation law becomes:

$$u_t + (\frac{1}{2}u^2)_x = 0. \tag{2.6}$$

This is also called *inviscid* Burgers' equation, since the equation studied by Burgers also includes a viscous term:

$$u_t + (\frac{1}{2}u^2)_x = \epsilon u_{xx}. \tag{2.7}$$

Equation (2.7) is the simplest model that includes the nonlinear and viscous effects of fluid dynamics. The analitic solution is available through a transformation known as the Cole-Hopf transformation, because, around 1950, Hopf, and independently Cole, solved exactly this equation [8, 29]. Thus Burgers' equation provides an important test for many proposed numerical methods dealing with nonlinear PDEs.

Consider the Riemann problem applied to inviscid Burgers' equation (2.6), with piecewise constant initial data (2.5). The form of the solution depends on the relation beetwen u_l and u_r.

First case: $u_l > u_r$. In this case there is a unique weak solution,

$$u(x,t) = \begin{cases} u_l & x < st \\ u_r & x > st. \end{cases}$$

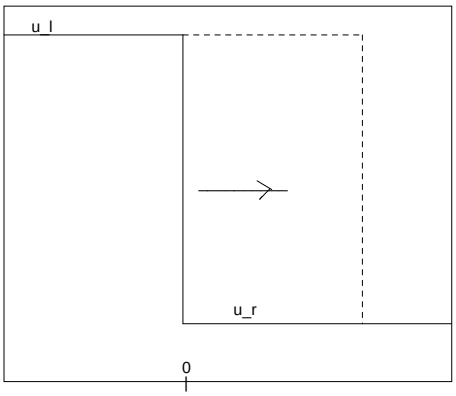

FIGURE 2.1: *Shock wave:* $u_l > u_r$.

where

$$s = \frac{(u_l + u_r)}{2}$$

is the shock speed, the speed at which the discontinuity travels.

Second case: $u_l < u_r$**.** In this case there are infinitely many weak solutions, since between the points $u_l t < x < u_r t$, there is no information available from the characteristics. To determine the correct physical behavior we adopt the vanishing viscosity approach by considering equation (2.7): equation (2.7) is a model of (2.6) valid only for small ϵ and smooth u. If the initial data is smooth and ϵ very small, then before the wave begins to break the ϵu_{xx} term is negligible compared to other terms and the solutions to the two PDEs look nearly identical. As the wave begins to break, the term u_{xx} grows much faster than u_x and at some point the ϵu_{xx} term is comparable to the other terms and begins to play a role. This term keeps the solution smooth for all time, preventing the breakdown of solutions that occurs for the hyperbolic problem. As ϵ goes to zero the solution of the viscous Burgers' equation becomes sharper and sharper and approaches the discontinuous solution of the inviscid Burgers' equation. Therefore, the physically correct weak solution for this Riemann problem is the solution that is stable to perturbations and is obtained as the vanishing viscosity generalized solution. In the $x - t$ plane the solution forms a wave from which the characteristics emanate with continuous slopes between u_l and u_r. This is called a *rarefaction wave* or *expansion fan* and is given by:

$$u(x,t) = \begin{cases} u_l & x < u_l t \\ x/t & u_l t \leq x \leq u_r t \\ u_r & x > u_r t \end{cases}$$

Thus, shock or rarefaction waves are the two possible solutions of the Riemann problem.

More generally, for arbritrary flux function $f(u)$ we have the following relation between the shock speed s and the states u_l and u_r, called *the Rankine-Hugoniot jump condition*:

$$f(u_l) - f(u_r) = s(u_l - u_r). \tag{2.8}$$

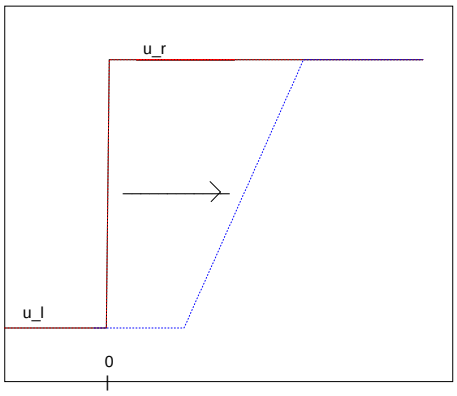

FIGURE 2.2: *Rarefaction wave:* $u_l < u_r$.

For scalar problems this gives simply

$$s = \frac{f(u_l) - f(u_r)}{u_l - u_r} = \frac{[f]}{[u]}$$

where $[\cdot]$ indicates the jump in some quantity across the discontinuity.

As shown above, there are situations in which the weak solution is not unique and an additional condition is required to identify the physically relevant solution. Since the condition that defines this solution as the limiting solution of the viscous equations as ϵ goes to zero is not easy to work with, we look for simpler conditions. To this aim the concept of *entropy condition* is introduced. We consider the following definition [5, 33].

DEFINITION 2.3. *A discontinuity propagating with speed s given by (2.8) satisfies the entropy condition if*

$$f'(u_l) > s > f'(u_r).$$

By considering the previous example, when $u_l < u_r$ the entropy condition is violated: in fact, characteristics come out of the wave as time advances and the propagating discontinuity is unstable to perturbations. Therefore the solution is not a shock wave but a rarefaction wave.

2.5 One dimensional case

2.5.1 Spatial discretization

Let us consider the Cauchy problem for conservation laws (2.1)-(2.2). When we attempt to calculate the solutions numerically, new problems arise. A finite-difference discretization of the conservation law (2.1) is expected to be inappropriate near discontinuities, since it is based on truncated Taylor-series expansions. Indeed, if we compute discontinuous solutions to conservation laws using standard methods, we typically obtain numerical results that are very poor. For example, natural first order accurate numerical methods have a large amount of *numerical viscosity* that smoothes the solution in much the same way physical

viscosity would, while a standard second order method eliminates this numerical viscosity but introduces *dispersive effects* that lead to large oscillations in the numerical solution. Therefore we would like to have numerical methods constructed *ad hoc* to solve hyperbolic conservation laws, which are accurate in smooth regions and give good results around discontinuities or sharp fronts. First we will consider Godunov's method: it uses the exact solution of the local Riemann problem and does not produce oscillations around discontinuities. Unfortunately, it is only first order accurate and the solutions display large numerical viscosity. A generalization of Godunov's method is represented by van Leer's method, an example of high resolution methods, characterized by second order accuracy on smooth solutions and the absence of spurious oscillations.

Both methods are *conservative* according to the following

DEFINITION 2.4. *Given a uniform grid with time step Δt and spatial mesh size Δx, a numerical method is said to be conservative if the corresponding scheme can be written as:*

$$v_j^{n+1} = v_j^n - \lambda(g_{j+\frac{1}{2}}^n - g_{j-\frac{1}{2}}^n), \quad j \in \mathbb{Z} \quad n \geq 0 \tag{2.9}$$

where v_j^n approximates $u(x_j, t^n)$ at the point ($x_j = j\Delta x, t^n = n\Delta t$), $\lambda = \dfrac{\Delta t}{\Delta x}$ and $g : \mathbb{R}^{2k} \longrightarrow \mathbb{R}$ is a continuous function, called the numerical flux *(function), that defines a $(2k+1)$-point scheme.*

$$g_{j+\frac{1}{2}}^n = g(v_{j-k+1}^n, \ldots, v_{j+k}^n).$$

The values v_j^0 are given by initial conditions.

This form of the scheme arises naturally if we view v_j^n as an approximation of the average $\overline{u}_j^n$ of $u(\cdot, t^n)$ on the cell $[x_{j-1/2}, x_{j+1/2}]$ (where $x_{j\pm1/2} = x_j \pm \dfrac{\Delta x}{2}$), defined by

$$\overline{u}_j^n = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, t^n)\ \mathrm{d}x.$$

Since the weak solution $u(x,t)$ satisfies the integral form of the conservation law, we have:

$$\begin{aligned} \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, t^{n+1})\ \mathrm{d}x = & \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, t^n)\ \mathrm{d}x \\ & -[\int_{t^n}^{t^{n+1}} f(u(x_{j+1/2}, t))\ \mathrm{d}t - \int_{t^n}^{t^{n+1}} f(u(x_{j-1/2}, t))\ \mathrm{d}t]. \end{aligned}$$

Dividing by Δx and using the averages $\overline{u}_j^n$ we get

$$\overline{u}_j^{n+1} = \overline{u}_j^n - \frac{1}{\Delta x}[\int_{t^n}^{t^{n+1}} f(u(x_{j+1/2}, t))\ \mathrm{d}t - \int_{t^n}^{t^{n+1}} f(u(x_{j-1/2}, t))\ \mathrm{d}t].$$

Comparing this to (2.9), we see that the numerical flux function can be considered as an average flux through $x_{j+1/2}$ over the time interval $[t^n, t^{n+1}]$,

$$\overline{g}_{j+1/2}^n = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(u(x_{j+1/2}, t))\ \mathrm{d}t.$$

An important property is the *consistency* with the original conservation law, that is the numerical flux function g reduces to the true flux f for the case of constant flow:

$$g(u, u, \ldots, u) = f(u) \quad \forall u \in \mathbb{R}.$$

For consistency it is sufficient that g is a Lipschitz continuous function of each variable, i.e. there is a constant $\mathcal{K}$ such that

$$|g(u_{j-k+1}, \ldots, u_{j+k}) - f(u)| \leq \mathcal{K} \max_{-k+1 \leq i \leq k} |u_{j+i} - u|,$$

for all u_{j+i} sufficiently close to u.

The main advantage of conservative and consistent schemes is that, when they converge, they converge to solutions whose shocks or discontinuity satisfy automatically the jump conditions, that is, the discontinuities always travel at the correct velocity. This important result, which is not true for non conservative or non consistent schemes, is due to Lax and Wendroff (the proof is given in [33]).

THEOREM 2.5 (LAX-WENDROFF). *Assume that the scheme (2.9) is consistent with the conservation law (2.1)-(2.2) and that it generates a sequence that converges to a function u^* as the gridsizes Δx, Δt go to zero. Then, u^* is a weak solution of the conservation law.*

Godunov's method. Godunov's method is an example of a conservative scheme. The solution is considered piecewise constant over each mesh cell at a fixed time and its evolution to the next time step results from the wave interactions originating at the boundaries between adjacent cells. The cell interfaces separate two different states at the left and at the right side, and the resulting interaction can be exactly resolved by solving a local Riemann problem. Complete definition of the interaction between adjacent cells is attained when the time interval over wich the waves are allowed to propagate is limited by the condition that adjacent Riemann problems do not interfere. This leads to a form of Courant-Friedrichs-Lewy (CFL) condition.

Godunov's method can be described as follows [5, 33]:

1. Given data v_j^n at time t^n, construct a piecewise constant function $\hat{v}_j^n(x, t^n)$ (see Figure 2.3) defined by

$$\hat{v}_j^n(x, t^n) = v_j^n \qquad x_{j-1/2} \leq x \leq x_{j+1/2}. \tag{2.10}$$

2. Solve the local Riemann problem at the cell interfaces, that is, on each subinterval $[x_j, x_{j+1}]$ and for $t \geq t^n$, solve

$$\begin{cases} \dfrac{\partial \hat{v}_j^n}{\partial t} + \dfrac{\partial f(\hat{v}_j^n)}{\partial x} = 0 \\ \hat{v}_j^n(x, t^n) = \begin{cases} v_j^n, & x_j < x < x_{j+1/2} \\ v_{j+1}^n, & x_{j+1/2} < x < x_{j+1} \end{cases} \end{cases} \tag{2.11}$$

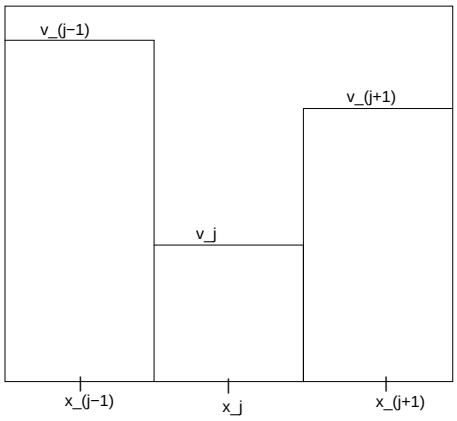

FIGURE 2.3: *First stage of Godunov's scheme at time* t^n.

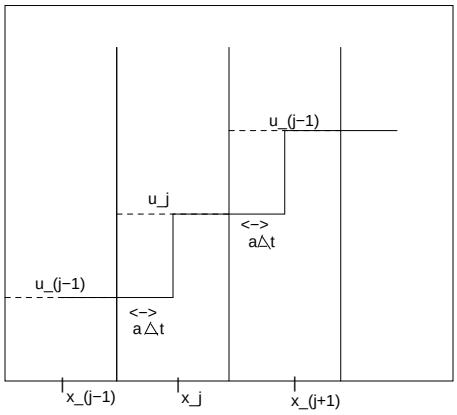

FIGURE 2.4: *Linear convection: translation of discontinuity.*

3. Define the approximation v_j^{n+1} at time t^{n+1} by averaging the Riemann problem solution $\hat{v}_j^n$ at the time t^{n+1}, so that

$$v_j^{n+1} = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} \hat{v}_j^n(x, t^{n+1}) \, \mathrm{d}x. \tag{2.12}$$

These values are then used to define new piecewise constant data $\hat{v}_j^{n+1}(x, t^{n+1})$ and the process repeats.

The first and third stages are of numerical nature and can be considered as *projection steps*, while the second stage, the physical one, is the *evolution step*. The basics of Godunov approach can be exemplified by an application to the simple linear advection equation in one dimension $u_t + au_x = 0$, with $a > 0$. The first step is independent of the equation to be solved. The second step is obtained by translation of the discontinuity at the interface over the distance $a\Delta t$ as shown in Figure 2.4. The new approximation at time level $n + 1$

results from the averaging of this new state, obtaining

$$\begin{aligned} v_j^{n+1} &= \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \hat{v}_j^n(x,t^{n+1})\ \mathrm{d}x \\ &= \frac{1}{\Delta x}[a\Delta t v_{j-1}^n + (\Delta x - a\Delta t)v_j^n] \\ &= v_j^n - a\frac{\Delta t}{\Delta x}(v_j^n - v_{j-1}^n) \end{aligned}$$

In practice Godunov's scheme is equal to the first-order upwind scheme when solving the linear advection equation.

More in general, provided we assume the CFL condition

$$\lambda \max_u |f'(u)| \leq \frac{1}{2},$$

where $\lambda = \dfrac{\Delta t}{\Delta x}$ so that the waves emanating from the points $x_{j-1/2}$ and $x_{j+1/2}$ do not interact, the solution is obtained by solving a juxtaposition of local Riemann problems and

$$\hat{v}_j^n(x,t) = \hat{v}_R(\frac{x - x_{j+1/2}}{t - t^n}; v_j^n, v_{j+1}^n), \qquad x_j \leq x \leq x_{j+1}, \tag{2.13}$$

for all $t > t^n$, where $\hat{v}_R$ is the solution of the local Riemann problem. We recall that the CFL condition is a necessary stability condition stating that the domain of dependence of the method includes the domain of dependence of the PDE.

In order to derive a general form of the scheme, let us integrate equation (2.11) over the rectangle $\left[x_{j-1/2}, x_{j+1/2}\right] \times [t^n, t^{n+1}]$. Since the function is piecewise smooth, we obtain:

$$\int_{x_{j-1/2}}^{x_{j+1/2}} (\hat{v}_j^n(x,t^{n+1}) - \hat{v}_j^n(x,t^n))\ \mathrm{d}x + \int_{t^n}^{t^{n+1}} (f(\hat{v}_j^n(x_{j+1/2}^-,t)) - f(\hat{v}_j^n(x_{j-1/2}^+,t)))\ \mathrm{d}t = 0,$$

where we consider the usual notation $x^+ = \lim_{x\to x^+} x$ and $x^- = \lim_{x\to x^-} x$. Using (2.10) and (2.12), we get

$$\Delta x(v_j^{n+1} - v_j^n) + \int_{t^n}^{t^{n+1}} (f(\hat{v}_j^n(x_{j+1/2}^-,t) - f(\hat{v}_j^n(x_{j-1/2}^+,t)))\ \mathrm{d}t = 0$$

At this point we note that the integral we need to compute in the previous equation is trivial because the integrand is independent of t. This follows by using (2.13) and from the fact that the solution of the Riemann problem at $x_{j+1/2}$ is a similarity solution, constant along each ray $(x - x_{j+1/2})/(t - t^n) = constant$. Therefore we have:

$$v_j^{n+1} = v_j^n - \lambda\{f(\hat{v}_R(0^-; v_j^n, v_{j+1}^n)) - f(\hat{v}_R(0^+; v_{j-1}^n, v_j^n))\}.$$

Since the function $x \longrightarrow f(\hat{v}_R(x; u_l, u_r))$ is continuous at the origin because of the Rankine-Hugoniot conditions, Godunov's method can be written in the conservative form

$$v_j^{n+1} = v_j^n - \lambda\{f(\hat{v}_R(0; v_j^n, v_{j+1}^n)) - f(\hat{v}_R(0; v_{j-1}^n, v_j^n))\}. \tag{2.14}$$

and its numerical flux is given by

$$g(u,v) = f(\hat{v}_R(0;u,v)). \tag{2.15}$$

Van Leer's method. Since the first and third steps of Godunov's methods are of a numerical nature, they can be modified without influencing the physics, for instance by replacing the piecewise constant approximation by a piecewise linear variation inside each cell. This leads to the definition of a spatially second order accurate scheme, known as van Leer's or MUSCL (Monotone Upstream-centered Scheme for Conservation Laws) method [59]. However, the straightforward replacement of the first-order scheme by a second-order accurate interpolation leads to the generation of oscillations around discontinuities. To overcome this limitation and achieve the goal of oscillation-free, spatially second-order accurate schemes, non linear components are introduced. Non linear discretizations imply that the schemes will be non linear even when applied to linear equations. This concept was introduced initially by van Leer under the form of *limiters*, i.e. functions that control the gradient of the computed solution with the aim of preventing the appearance of unphysical overshoots or undershoots.

Consequently, we study van Leer's method as an example of a second-order slope limiter method. The three main steps of van Leer's MUSCL approach are the following [25, 27, 59]:

1. *reconstruction step:* the dependent variable is interpolated using a piecewise linear function $\hat{v}$ starting from the cell averages v_j^n, v_{j-1}^n, v_{j+1}^n. To this end define S_j^n to be the *slope* on the jth cell calculated using v_{j-1}^n and v_{j+1}^n. Then

$$\hat{v}^n(x) = \begin{cases} v_j^n + (x - x_j)\dfrac{S_j^n}{\Delta x} & x_{j-1/2} < x < x_{j+1/2}, \\ \hat{v}^n(x_{j-1/2}) & x \leq x_{j-1/2} \\ \hat{v}^n(x_{j+1/2}) & x \geq x_{j+1/2}. \end{cases} \tag{2.16}$$

   Note that taking $S_j^n = 0$ for all j and n recovers Godunov's method;

2. *evolution step*: the waves are propagated across cell interfaces according to an exact or approximate solution of a local Riemann problem that uses the interpolated values $\hat{v}^n(x)$ as initial conditions. One solves

$$\begin{cases} \dfrac{\partial}{\partial t} w + \dfrac{\partial}{\partial x} f(w) = 0 \quad x \in \mathbb{R}, t^n \leq t \leq t^{n+1} \\ w(x, t^n) = \hat{v}^n(x). \end{cases} \tag{2.17}$$

   This step yields $w(\cdot, t^{n+1})$.

3. *cell-averaging step*: v_j^{n+1} is obtained by projecting the solution $w(x, t^{n+1})$ onto the piecewise constant functions

$$v_j^{n+1} = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} w(x, t^{n+1}) \, \mathrm{d}x. \tag{2.18}$$

The cell average of $\hat{v}^n(x)$ over $[x_{j-1/2}, x_{j+1/2}]$ is equal to v_j^n for any choice of S_j^n and thus step 1 is conservative. Since steps 2 and 3 are also conservative, the overall method is conservative for any choice of S_j^n. These three steps can be visualized graphically as in Figure 2.5 in the case of linear advection. Provided we assume some convenient CFL

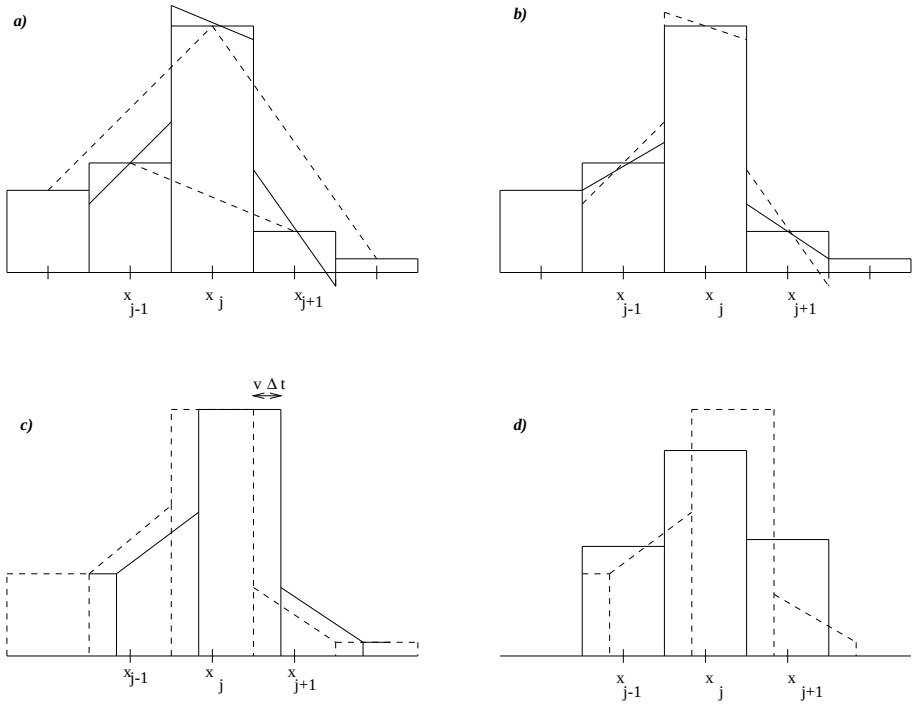

FIGURE 2.5: *Visualization of van Leers's scheme for linear advective flux: a) reconstruction; b) limiting; c) evolution; d) cell-averaging step.*

condition so that the waves issued from the points $x_{j-1/2}$ and $x_{j+1/2}$ do not interact, the solution can be viewed as a juxtaposition of local Riemann problem solution. In order to derive a more explicit form of the scheme, we integrate equation (2.17) over $\left[x_{j-1/2}, x_{j+1/2}\right] \times [t^n, t^{n+1}]$,

$$\int_{t^n}^{t^{n+1}} \int_{x_{j-1/2}}^{x_{j+1/2}} (\frac{\partial w}{\partial t} + \frac{\partial f(w)}{\partial x}) \, \mathrm{d}x \, \mathrm{d}t = 0.$$

We obtain

$$\int_{x_{j-1/2}}^{x_{j+1/2}} (w(x, t^{n+1}) - w(x, t^n)) \, \mathrm{d}x + \int_{t^n}^{t^{n+1}} (f(w(x_{j+1/2}, t)) - f(w(x_{j-1/2}, t))) \, \mathrm{d}t = 0$$

since the flux is continuous, and then by (2.18)

$$\Delta x(v_j^{n+1} - v_j^n) + \int_{t^n}^{t^{n+1}} (f(w(x_{j+1/2}, t)) - f(w(x_{j-1/2}, t))) \, \mathrm{d}t = 0$$

We are left with the evaluation of the numerical flux

$$g_{j+1/2}^n = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(w(x_{j+1/2}, t)) \, \mathrm{d}t.$$

Using the midpoint rule, we can write

$$\frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(w(x_{j+1/2}, t)) \, \mathrm{d}t = f(w(x_{j+1/2}, t^n + \frac{\Delta t}{2})) + \mathcal{O}(\Delta t^2).$$

Following [58], define the updated values $v^{n+1/2}_{j+1/2,\pm}$ at time $t^n + \dfrac{\Delta t}{2}$ by

$$\begin{cases} v^{n+1/2}_{j+1/2,-} = v^n_{j+1/2,-} - \dfrac{\Delta t}{2\Delta x}(f(v^n_{j+1/2,-}) - f(v^n_{j-1/2,+})), \\ \\ v^{n+1/2}_{j+1/2,+} = v^n_{j+1/2,+} - \dfrac{\Delta t}{2\Delta x}(f(v^n_{j+3/2,-}) - f(v^n_{j+1/2,+})), \end{cases}$$

where

$$\begin{cases} v^n_{j+1/2,-} = \hat{v}(x^-_{j+1/2}) = v^n_j + \dfrac{S^n_j}{2}, \\ v^n_{j+1/2,+} = \hat{v}(x^+_{j+1/2}) = v^n_{j+1} - \dfrac{S^n_{j+1}}{2}, \end{cases}$$

Then, solve the Riemann problem at the point $x_{j+1/2}$ with piecewise constant initial data $v^{n+1/2}_{j+1/2,\pm}$

$$\begin{cases} \dfrac{\partial \hat{w}}{\partial t} + \dfrac{\partial f(\hat{w})}{\partial x} = 0 \\ \hat{w}(x,0) = \begin{cases} v^{n+1/2}_{j+1/2,-} & x < x_{j+1/2} \\ v^{n+1/2}_{j+1/2,+}, & x > x_{j+1/2} \end{cases} \end{cases}$$

whose solution is noted as $v_R(\dfrac{x - x_{j+1/2}}{t}; v^{n+1/2}_{j+1/2,-}, v^{n+1/2}_{j+1/2,+})$.

Replacing $w(x_{j+1/2}, t)$ by $v_R(0; v^{n+1/2}_{j+1/2,-}, v^{n+1/2}_{j+1/2,+})$, the numerical flux can be approximated by $g^n_{j+1/2} = f(v_R(0; v^{n+1/2}_{j+1/2,-}, v^{n+1/2}_{j+1/2,+}))$. This time stepping scheme is equivalent to a two-stage second order accurate Runge scheme.

The numerical scheme is completely defined after we have specified S^n_j. As an example, in the linear advection equation (i.e. $f(u) = au$ with $a > 0$), if $S^n_j = v^n_{j+1} - v^n_j$ and the advection equation is solved exactly in step 2, then the method reduces to the Lax-Wendroff method [33]. This illustrates that it is possible to obtain second order accuracy by this approach. The oscillations which arise with Lax-Wendroff method can be interpreted geometrically as being caused by a poor choice of slopes. In fact oscillations are created when the slope in a cell becomes larger than the difference of adjacent mean values. Thus, a scheme without overshoots around discontinuities can be obtained if excessively large gradients are avoided. Therefore we have to control each cell at each time step to keep the gradients within the proper bounds. The generation of oscillations can be prevented by acting on their production mechanism and introducing nonlinear correction factors, so called the limiters, that force the method to be *total variation diminishing* (TVD)(Lax-Wendroff scheme is not TVD).

We have the following:

DEFINITION 2.6. *A numerical method to solve hyperbolic conservation laws is called total variation diminishing if*

$$\sum_{j=-\infty}^{+\infty} |v^{n+1}_{j+1} - v^{n+1}_j| \leq \sum_{j=-\infty}^{+\infty} |v^n_{j+1} - v^n_j|$$

The advantage of a TVD scheme stems from the fact that, whenever stability is assured, it is free of numerical oscillations.

To obtain a slope limiter we follow [57]. Let

$$S_j^n = \hat{S}_j^n \Phi_j^n$$

where $\hat{S}_j^n$ represents the actual slope approximation and $\Phi_j = \Phi(\theta_j^n)$ is a limiter function, defined in such a way that the method is TVD and θ_j^n is the ratio of two consecutive gradients, i.e.:

$$\theta_j^n = \frac{v_j^n - v_{j-1}^n}{v_{j+1}^n - v_j^n}.$$

The slope $\hat{S}_j^n$ can be defined by

$$\hat{S}_j^n = \frac{1}{2}(1+\omega)(v_j^n - v_{j-1}^n) + \frac{1}{2}(1-\omega)(v_{j+1}^n - v_j^n)$$

where ω is a parameter in the real interval $[-1, 1]$. For $\omega = 0$, $\hat{S}_j^n$ represents central difference approximation to the first spatial derivative of the numerical solution at time level n.

A TVD scheme is obtained if (see, for example, [33])

$$\Phi(\theta) = 0 \qquad \text{for } \theta \leq 0$$

and

$$0 \leq \Phi(\theta) \leq 2\theta.$$

Second order accuracy requires additional conditions on Φ. In fact, when $\theta \to 1$, which means that the numerical solution is smooth, the limiting function must evidently approach unity [55].

Various limiter functions have been defined in literature. In the subsequent numerical tests we consider the following three limiter functions:

- van Leer's limiter [59]:
$$\Phi(\theta) = \frac{|\theta| + \theta}{1 + |\theta|},$$
- *minmod* limiter that represents the lowest boundary of the second-order TVD region [33]:
$$\Phi(\theta) = \begin{cases} \min(\theta, 1) & \text{if } \theta > 0 \\ 0 & \text{if } \theta \leq 0. \end{cases}$$
It is a particular case of the minmod function, defined as the function that selects the number with the smallest modulus from a series of numbers when they all have the same sign, and zero otherwise. For two arguments:
$$\text{minmod}(x, y) = \begin{cases} x & \text{if } |x| < |y| \quad \text{and } xy > 0 \\ y & \text{if } |x| > |y| \quad \text{and } xy > 0 \\ 0 & \text{if } xy < 0 \end{cases}$$
- the *superbee* limiter, that represents the upper limit of the second-order TVD region and has been introduced by Roe [52]:
$$\Phi(\theta) = \max[0, \min(2\theta, 1), \min(\theta, 2)].$$

2.5.2 Remark about time integration

As explained in the previous Section, linear schemes with high order of accuracy generate spurious oscillation wherever the solution is not smooth. To prevent spurious oscillation the concept of TVD is imposed, making the scheme nonlinear even if applied to a linear equation.

When explicit Euler is used to approximate the time derivative (first order in time), the linear version of the scheme (no limiter is adopted) is unconditionally unstable. Only with the TVD constraint imposed, i.e. the nonlinear version, the scheme is conditionally stable. In the following we explain why the unlimited linear form of the above scheme is unconditionally unstable.

Rewriting the Euler scheme for the simple model equation in which $f(u) = au$, a positive constant, we get

$$v_j^{n+1} = v_j^n - \lambda(g_{j+\frac{1}{2}}^n - g_{j-\frac{1}{2}}^n), \qquad \lambda = \frac{\Delta t}{\Delta x},$$

where the numerical flux is

$$g_{j+\frac{1}{2}}^n = a(v_j + \frac{1}{2}\delta v_j)$$

and

$$g_{j-\frac{1}{2}}^n = a(v_{j-1} + \frac{1}{2}\delta v_{j-1}).$$

The gradient δv_j should be limited to obtain a conditionally stable scheme. Without TVD constraint, we consider the linear average for the gradient, that is

$$\delta v_j = \frac{v_{j+1} - v_{j-1}}{2}$$

and get the scheme

$$v_j^{n+1} = v_j^n - \frac{a\lambda}{4}(3v_j + v_{j+1} - 5v_{j-1} + v_{j-2}). \tag{2.19}$$

Inserting Taylor expansions into the full discretized equation, the truncation error can be easily found to be $-a^2\lambda\Delta x u_x x/2$ [56].

A necessary condition for stability of our numerical scheme is given by the Godunov-Ryabenkii condition [27]:

LEMMA 2.7. *If equation (2.19), with periodic boundary conditions, has a solution*

$$v_j = e^{st} f_j, \qquad \|f\| \leq \infty$$

for some complex number s with $\Re(s) > 0$ and some stepsize Δx_0, then the appoximation is not stable.

Proof. If such a solution exists, for some $\Delta x = \Delta x_0$, $s = s_0$, $\Re(s) > 0$, then

$$\tilde{s_0} f_j = -\frac{a}{4}(3f_j + f_{j+1} - 5f_{j-1} + f_{j-2}), \qquad \tilde{s_0} = \Delta x_0 s_0,$$

$$\|f\| < \infty.$$

Therefore, for any Δx,

$$\tilde{v_j} = e^{(\tilde{s_0}/\Delta x)t} f_j$$

is also a solution. Thus, as Δx goes to zero, we can construct solutions to the difference equation that grow arbitrarily fast. □

We can formulate this in terms of the eigenvalue problem

$$\tilde{s}\phi_j = -\frac{a}{4}(3\phi_j + \phi_{j+1} - 5\phi_{j-1} + \phi_{j-2}), \qquad \tilde{s} = \Delta x s, \tag{2.20}$$

$$\|\phi\| < \infty.$$

We have the following theorem

THEOREM 2.8. *The approximation (2.19) is not stable because equation (2.20) has eigenvalues $\tilde{s}$ with $\Re(\tilde{s}) > 0$.*

Proof. Equation (2.20) is an ordinary difference equation with constant coefficients, and its solution has the form

$$\phi_j = \sigma_1 k_1^j + \sigma_2 k_2^j + \sigma_3 k_3^j$$

where k_1, k_2 and k_3 are solutions of the characteristic equation

$$k^3 + (\frac{4\tilde{s}}{a} + 3)k^2 - 5k + 1 = 0, \qquad \tilde{s} = \Delta x s \tag{2.21}$$

Assuming that equation (2.21) has a solution $k = e^{\sqrt{-1}\xi}$, ξ real, we get

$$\tilde{s} = \frac{a}{4}[(4\cos\xi - 2\cos^2\xi - 2) + \sqrt{-1}(\sin 2\xi - 6\sin\xi)].$$

Now, it is straightforward to see that $\Re(\tilde{s})$ vanishes only for $\cos\xi = 1$. Thus, there exist solutions $\tilde{s}$ with $\Re(\tilde{s})$ positive. Then, $v_j = e^{st} f_j$ is a solution of (2.19) and, for Lemma 2.7 approximation (2.19) is unconditionally unstable. □

In conclusion, the TVD constraint to the scheme makes it conditionally stable. However also in this case, attention has to be posed when using the Euler explicit scheme in time. Indeed, as reported in [56], one may note that the solution profile displays a stair-like behavior when using a minmod or superbee limiter. To avoid this stair-like solution a much smaller Courant number should be used, but in this way the attractiveness of this method is lost. A better solution is to use high-order discretization not only in space but also in time, as for example Runge-Kutta schemes.

2.5.3 Numerical tests

In this Section the previously analyzed methods (Godunov's and van Leer's schemes) are applied to different model problems. The two sample tests used are the linear advection equation, which serves as a model for contact discontinuities in fluid dynamics, and Burgers' equation used to study how the methods treat shocks.

Periodic boundary conditions. To apply any of the previous methods in a finite domain, we are immediately faced with the problem of how to discretize the equations at the boundary points. The simplest approach that avoids this problem consists in using periodic boundary conditions (p.BCs),

$$u(x_{min}, t) = u(x_{max}, t),$$

so that periodicity is used to provide the extra needed values to implement Godunov and van Leer's schemes at the boundaries.

Example 1 p.BCs. We begin by studying the simple hyperbolic equation

$$\begin{array}{ll} u_t + u_x = 0 & x \in [-1, 1] \quad t \geq 0 \\ u(x, 0) = u_0(x) = \sin(2\pi x) & x \in [-1, 1] \end{array}$$

Its exact solution is the wave $u(x, t) = \sin(2\pi(x - t))$, constant along the characteristics $x - t = constant$, and with speed of propagation equal to $\mathrm{d}x/\mathrm{d}t = 1$. In this case the CFL number is $\lambda = \Delta t/\Delta x$, because $|f'(u)| = 1$.

We calculate the solution at time $t = 2$ s with a CFL number equal to $\lambda = 0.5$, with $\Delta t = 1 \times 10^{-2}$ s and $\Delta x = 2 \times 10^{-2}$ m. Figure 2.6 shows numerical solutions to this problem by using Godunov's and van Leer's method: the results are plotted at time $t = 2$ s along with the exact solution. Godunov's solutions are smeared because being a first order method it introduces large amount of numerical diffusion, while van Leer's method approximates the exact solution much better. The behavior of different limiters can be observed in Figure 2.6 (b), (c) and (d). The minmod limiter reduces locally the accuracy of the solution around the extrema, while the superbee limiter artificially sharpens the fronts. The overcompressive property of superbee limiter is not too adequate for smooth profiles. Instead, the van Leer limiter display an intermediate behavior between the minmod and superbee limiters.

Example 2 p.BCs. To simulate contact discontinuities we again solve the linear advection equation $u_t + u_x = 0$, with initial condition $u_0(x)$ periodic of period 2 defined on the interval $[-1, 1]$ as

$$u_0(x) = \begin{cases} 1 & -1 \leq x \leq -0.75 \\ 0 & -0.75 < x < -0.25 \\ 1 & -0.25 \leq x \leq 0.25 \\ 0 & 0.25 < x < 0.75 \\ 1 & 0.75 \leq x \leq 1 \end{cases}$$

We solve this problem at time $t = 2$ s with $\lambda = 0.5$, by setting $\Delta t = 1 \times 10^{-2}$ s and $\Delta x = 2 \times 10^{-2}$ m. The expected behavior of the considered numerical methods is plotted in Figure 2.7. All the schemes are monotone but the first order Godunov scheme shows again excessive numerical diffusion, while the second order schemes are more accurate with the superbee limiter being the most compressive and the minmod limiter the most diffusive one.

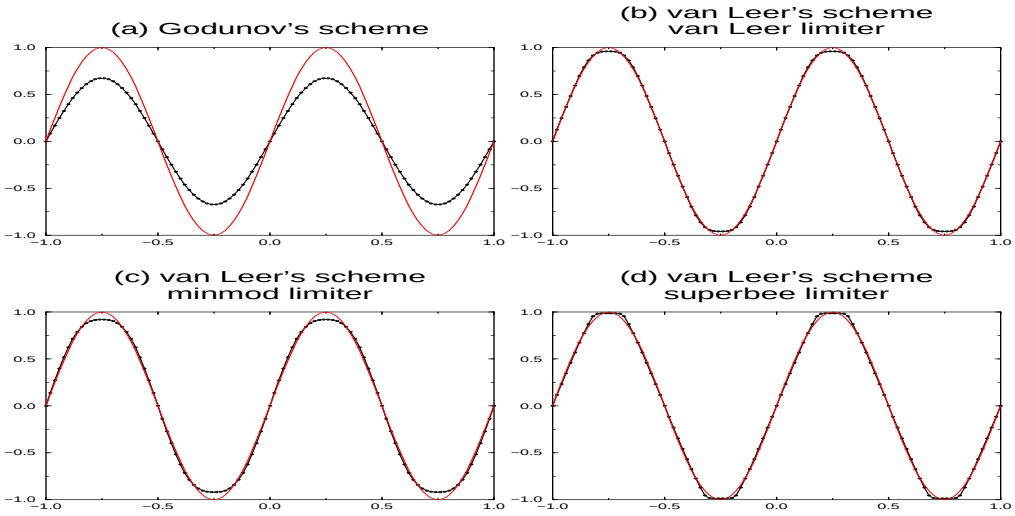

FIGURE 2.6: *Example 1 p.BCs: numerical (— • —) and exact (–) solution with* $\lambda = 0.5$, $\Delta x = 2 \times 10^{-2}$ *m,* $t = 2$ *s.*

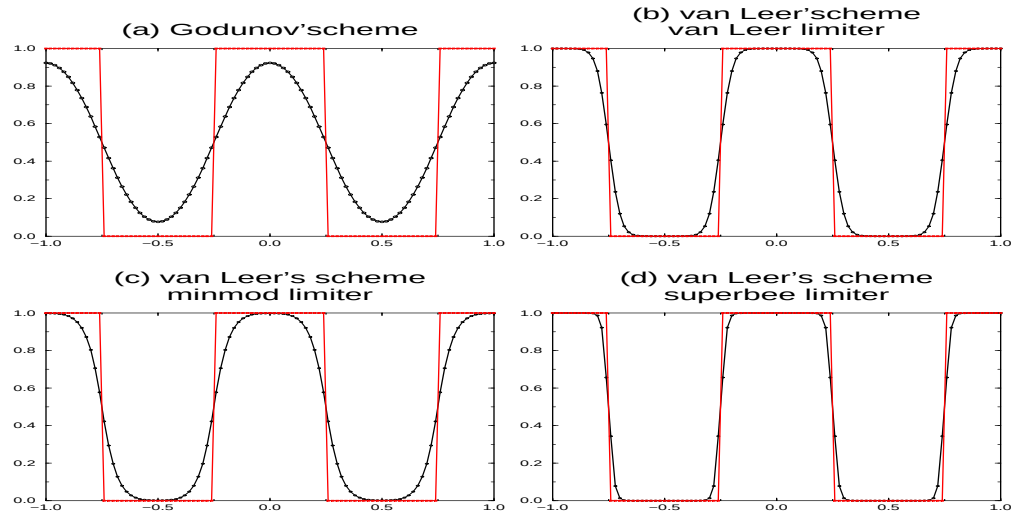

FIGURE 2.7: *Example 2 p.BCs: numerical (— • —) and exact (–) solution with* $\lambda = 0.5$, $\Delta x = 2 \times 10^{-2}$ *m,* $t = 2$ *s.*

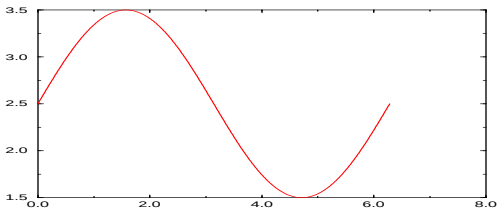

FIGURE 2.8: *Example 3 p.BCs: initial condition.*

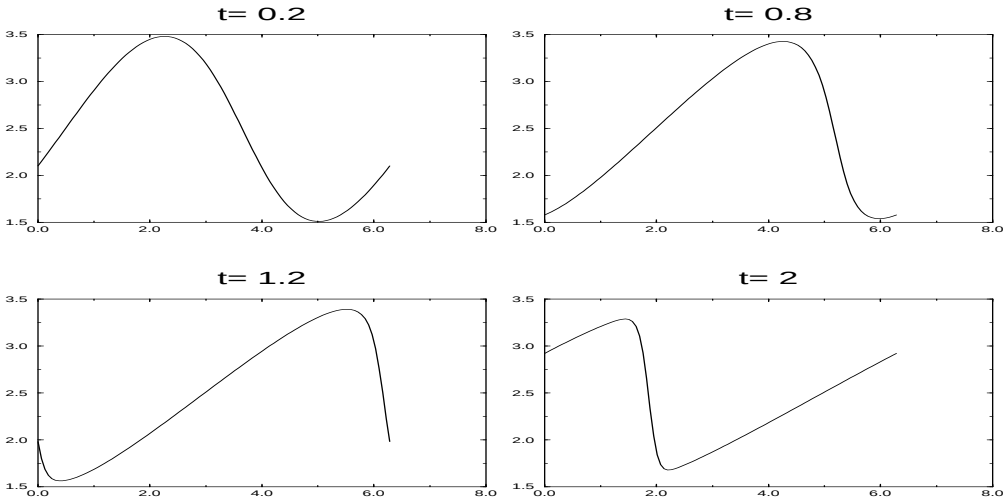

FIGURE 2.9: *Example 3 p.BCs: evolution in time by Godunov's method.*

Example 3 p.BCs. As nonlinear test problem with periodic initial and boundary conditions, we consider the inviscid Burger's equation (2.6) on the interval $[0, 2\pi]$ and $t \geq 0$, with periodic boundary conditions $u(0,t) = u(2\pi,t)$ and initial condition $u_0(x) = 5/2 + \sin x$. As t increases from the origin, $u_0(x)$ is transported with unit speed to the right and simultaneously evolves into a function with an increasingly sharper profile which, after a while, is discontinuous: a shock appears. Figure 2.8 displays the initial condition. The evolution of the numerical solution obtained by Godunov's method at times $t = 0.2,\ 0.8,\ 1.2$ and 2 s is plotted in Figure 2.9, while Figure 2.10 shows the same solutions obtained by van Leer's method with van Leer's limiter. We have chosen the van Leer limiter, because its behavior is intermediate between the minmod and superbee limiters behavior. For both method we have used $\Delta t = 2 \times 10^{-3}$ s and $\Delta x = 2\pi/100$ m.
The smooth initial solution degenerates into a discontinuous one, but Godunov's method tends to smooth the profile while van Leer's method is more efficient to control the shock.

Non periodic boundary conditions. Particular attention is needed to discretize boundary conditions, since stable schemes can be strongly affected by unsuitable boundary approximations, leading to possible instability or to reduction of unconditional to conditional stability of the numerical scheme. With regard to accuracy, it is well known (see, for example, [27]) that, for linear equations, the boundary scheme can be one order of

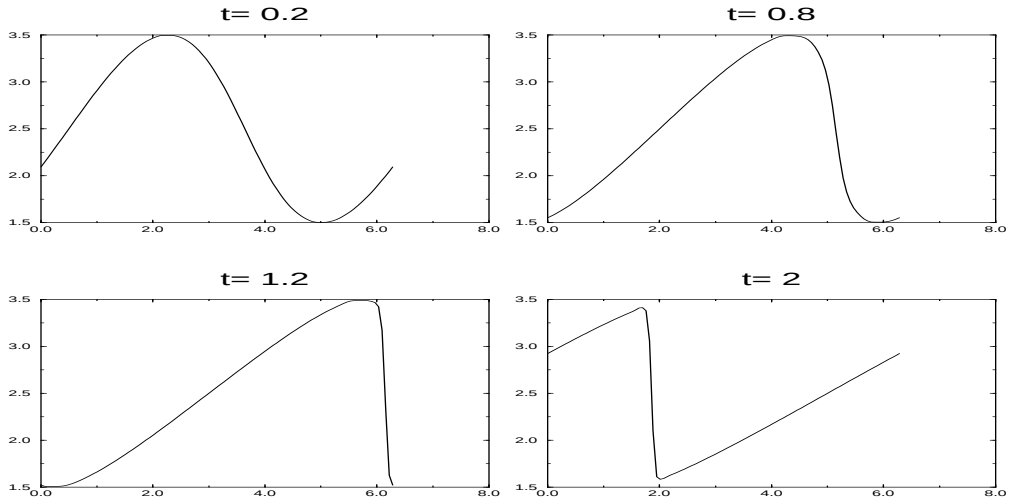

FIGURE 2.10: *Example 3 p.Bcs: evolution in time by van Leer's method with van Leer limiter.*

accuracy lower than the basic discretization without reducing the global order of accuracy of the complete method.
Consider the scalar conservation law

$$\begin{aligned} &u_t + (f(u))_x = 0, \qquad x \in [x_{min}, x_{max}],\ t > 0 \\ &u(x,0) = u_0(x), \end{aligned}$$

with prescribed boundary condition $u(x_{min}, t) = g(t)$ if the characteristic is ingoing, that is if $f' > 0$, or imposing $u(x_{max}, t) = g(t)$ otherwise. For example, for the linear advection equation $u_t + au_x = 0$, if $a > 0$, the characteristics are leaving from the boundary $x = x_{min}$, thus coming into the domain. Therefore, we prescribe the solution at the boundary $x = x_{min},\ u(x_{min}, t) = g(t),\ t > 0$. Instead, if $a < 0$ we have $u(x_{max}, t) = g(t),\ t > 0$. Therefore, when we consider the numerical scheme, at most one of the boundary values is known by the boundary conditions.

Let the space interval $[x_{min}, x_{max}]$ be divided into $N-1$ cells of length Δx, with $j = 1$ at $x = x_{min}$ to $j = N$ at $x = x_{max}$, and consider a 3-point linear conservative scheme (Godunov's scheme) for approximating the linear advection equation with $a > 0$:

$$v_j^{n+1} = v_j^n - \lambda(g_{j+\frac{1}{2}}^n - g_{j-\frac{1}{2}}^n),$$

with $g_{j+\frac{1}{2}}^n = g(v_j^n, v_{j+1}^n)$, and j varying from 2 to $N-1$. Note that v_1^n is given by the boundary condition, i.e. $v_1^n = g(t^n)$. We can not write the previous formula for $j = N$ because it requires values outside of the computational domain. In the same way, if we use a 5-point scheme, (e.g. the van Leer scheme), again $v_1^n = g(t^n)$, but the computation of v_2^n and v_{N-1}^n requires the evaluation of auxiliary values v_0^n, v_N^n and v_{N+1}^n. To this aim we consider schemes based on extrapolations of the internal variables toward the boundary. In particular, we consider the following formulae for a 3-point scheme (we write the conditions for an outlet boundary $j = N$. The transposition to inlet conditions is straightforward, replacing $j = N$ by $j = 1$, $j = N-1$ by $j = 2$ and so on):

- space extrapolation
  - zero-order extrapolation: $v_N^{n+1} = v_{N-1}^{n+1}$
  - first-order extrapolation: $v_N^{n+1} = 2v_{N-1}^{n+1} - v_{N-2}^{n+1}$
- space-time extrapolation
  - zero-order: $v_N^{n+1} = v_{N-1}^{n}$
  - first order in space/zero order in time: $v_N^{n+1} = 2v_{N-1}^{n} - v_{N-2}^{n}$
  - first order in space and time: $v_N^{n+1} = 2v_{N-1}^{n} - v_{N-2}^{n-1}$
- time extrapolation
  - zero order: $v_N^{n+1} = v_N^{n}$
  - first order: $v_N^{n+1} = 2v_N^{n} - v_N^{n-1}$

In the case of a 5-point scheme, one proceeds on the same manner, for example

$$v_N^{n+1} = 2v_{N-1}^{n+1} - v_{N-2}^{n+1}$$

and

$$v_{N+1}^{n+1} = 2v_N^{n+1} - v_{N-1}^{n+1}.$$

In the following numerical experiments, we consider the space-time zero order extrapolation in Godunov's method, while, for van Leer's scheme, we use first-order in space and time formula to obtain v_N^{n+1} and v_{N+1}^{n+1} for $n \geq 1$. For $n = 0$ we consider the first order in space and zero order in time formula, because only the initial values for $t = t^0$ are known. For the auxiliary value v_0^{n+1} we use a zero-order space extrapolation.

Example 1 non p.BCs. As a first example we consider the linear advection equation, $u_t + u_x = 0$, with zero boundary conditions, and the following initial condition:

$$u(x,0) = u_0(x) = \begin{cases} \sin(\pi x) & 0 \leq x \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

This test case allows us to test the diffusion properties of the schemes with numerical boundary conditions. The solutions at time $t = 2$ s obtained by Godunov's method and by application of the various limiters to van Leer's scheme can be seen from a comparison of Figures 2.11 (a) $\rightarrow$ (d). The behavior of the different scheme is similar to the one reported previously. Note that no oscillations are introduced at the boundary, confirming the fact that the overall scheme is stable.

Example 2 non p.BCs. A simple nonlinear test on Burgers' equation is given by considering a Riemann problem, that is, the initial condition is:

$$u(x,0) = \begin{cases} 1 & x < 0 \\ 0 & x > 0. \end{cases}$$

Therefore, we have a shock propagating at speed $s = 1/2$ with unmodified intensity $[u] = 1$. In Figure 2.12 we plot the numerical solution at time $t = 4$ s, with $\Delta t = 2 \times 10^{-2}$ m and $\Delta x = 4 \times 10^{-2}$ s.

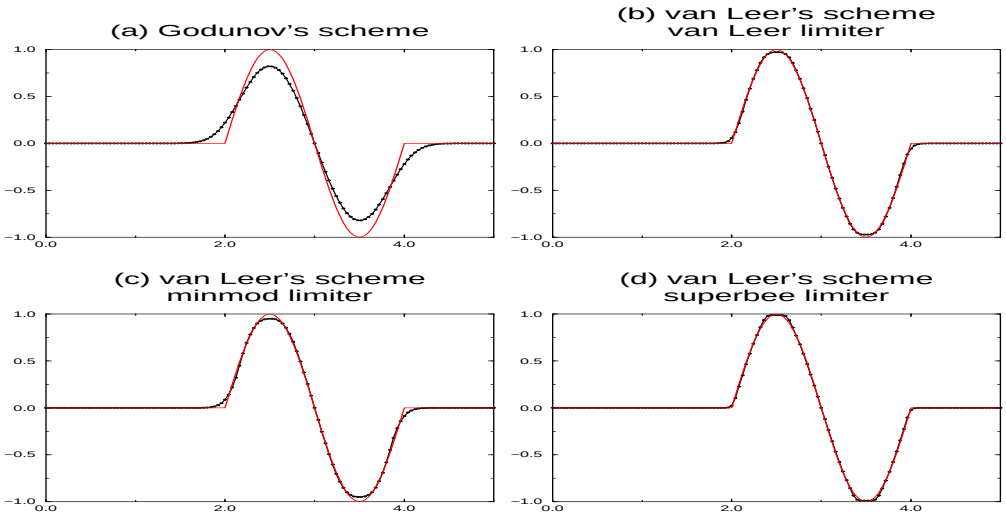

FIGURE 2.11: *Example 1 non p.BCs: numerical* $(- \cdot -)$ *and exact* $(-)$ *solution with* $\lambda = 0.5$, $\Delta x = 4 \times 10^{-2}$ *m*, $t = 2$ *s*.

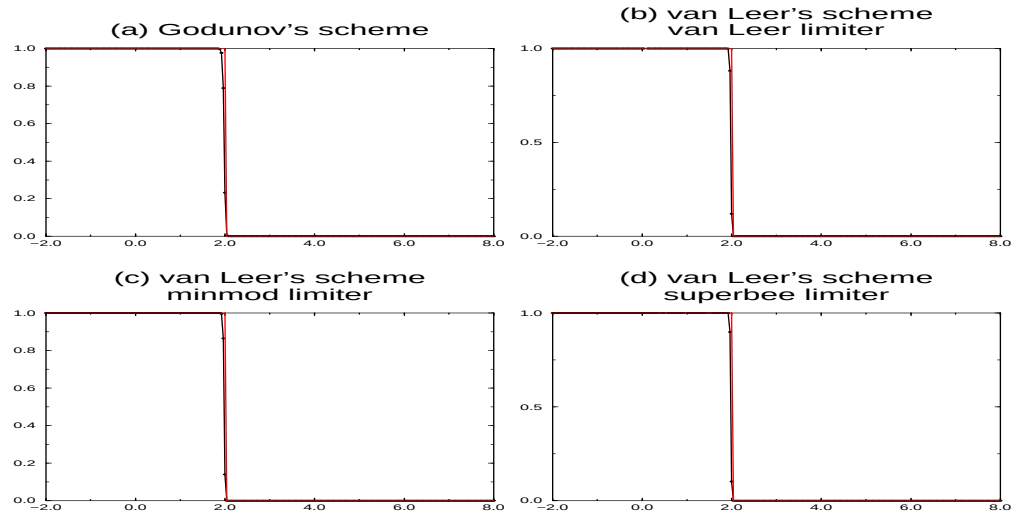

FIGURE 2.12: *Example 2 non p.BCs: numerical* $(- \cdot -)$ *and exact* $(-)$ *solution with* $\lambda = 0.5$, $\Delta x = 4 \times 10^{-2}$ *m*, $t = 4$ *s*.

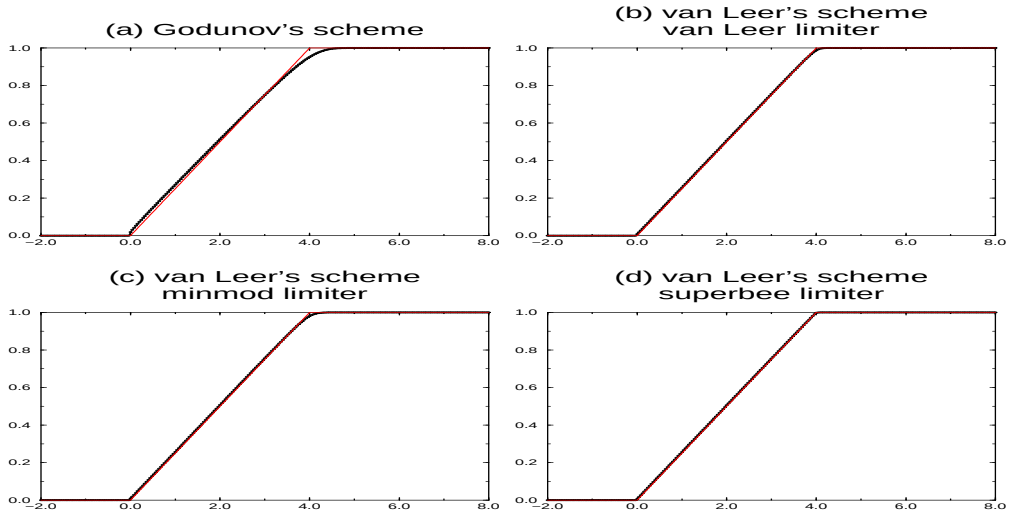

FIGURE 2.13: *Example 3 non p.BCs: numerical (— · —) and exact (—) solution with* $\lambda = 0.5$, $\Delta x = 4 \times 10^{-2}$ *m,* $t = 4$ *s.*

Example 3 non p.BCs. As last example we consider the Riemann problem whose solution is a rarefaction wave:

$$u(x,0) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases}$$

The behavior for the methods considered is displayed in Figure 2.13.

2.6 Two dimensional case

2.6.1 Spatial discretization

A straight forward extension of MUSCL type finite volume scheme to two spatial dimension can be obtained by employing rectangular discretizations. The advective flux can then be split along the x and z directions and the one dimensional techniques previously described can be applied. This approach inherits all the limitations of a structured regular grid, and is affected by grid alignment problems. On the other hand, unstructured grids (triangulations) offer a greater flexibility when dealing with complex geometries and limit grid orientation effects. But extensions of TVD higher order schemes to two dimensional triangulations require the definition of inherently two-dimensional limiters, and maintaining second order accuracy is a challenge. The scheme we considered has been developed by [16] and next modified by [34], is a second order accurate scheme of TVD type in the sense that the fluxes are approximated to the second order accuracy, and is applicable to an unstructured triangular grid. The scheme relies on a very local adaptive interpolation idea, which results in computational efficiency. Before detailed presentation of this scheme, a brief introduction to the Finite Volume (FV) methods for multidimensional conservation laws is presented [25].

Hyperbolic conservation laws in multidimensions space can be written as

$$\begin{aligned} &u_t + \vec{\nabla} \cdot \vec{F}(u) = 0 \\ &u(x,0) = u_0(x), \end{aligned} \tag{2.22}$$

where $x \in \mathbb{R}^d$. Here we consider hyperbolic conservation laws (2.22) in two space dimensions (d =2). As in one spatial dimension, the FV method is based on the local conservation property satisfied by the solutions of a conservations law, and thus can be defined control volumes of general shape, in particular triangulations. The computational domain can be subdivided into cells, or control volumes, T_l, with center x_l. On T_l, $u(\cdot, t)$ is approximated by a constant volume $u_l(t)$, considered as an approximation of the mean value of u over the cell T_l,

$$u_l(t) \cong \frac{1}{|T_l|} \int_{T_l} u(x,t)\ \mathrm{d}x,$$

where $|T_l|$ denotes the area of T_l. The differential equation defining $u_l(t)$ is obtained as follows. First, integrating equation (2.22) over T_l yields

$$\frac{\partial}{\partial t}\left(\int_{T_l} u(x,t)\ \mathrm{d}\Delta\right) + \int_{\partial T_l} \vec{F}(u(\cdot,t)) \cdot n_l\ \mathrm{d}\Gamma = 0, \tag{2.23}$$

where ∂T_l is the boundary of T_l and n_l the outward unit normal vector to T_l. The first term in (2.23) is naturally approximated by

$$\frac{\partial}{\partial t}\left(\int_{T_l} u(x,t)\ \mathrm{d}\Delta\right) \cong |T_l| \frac{\partial u_l(t)}{\partial t}.$$

Since the approximation is not continuous across ∂T_l, we have to discretize the flux across the boundary of the cell, that is $\displaystyle\int_{\partial T_l} \vec{F}(u(\cdot,t)) \cdot n_l\ \mathrm{d}\Gamma$. This can be written as

$$\int_{\partial T_l} \vec{F}(u) \cdot n_l\ \mathrm{d}\Gamma = \sum_{e \subset \partial T_l, e = e_{lj}} \int_{e_{lj}} \vec{F}(u) \cdot n_l\ \mathrm{d}\Gamma,$$

where the sum is taken over all the edges e of the cell, and $\partial T_l = \cup e_{lj}$, where $e_{lj} = T_l \cap T_j$ is the face separating T_l and T_j. The problem is then to define the numerical fluxes approximating $\displaystyle\int_{e_{lj}} \vec{F}(u) \cdot n_l\ \mathrm{d}\Gamma$, using only the values $u_l(t)$. The usual way consists in introducing a function H such that for $e = e_{lj} \subset \partial T_l$

$$\int_{e_{lj}} \vec{F}(u) \cdot n_l\ \mathrm{d}\Gamma \cong |e| H(u_l, u_j, n_e), \tag{2.24}$$

where n_e denotes the unit normal to e pointing in the direction of T_j (thus outward to T_l), and $|e|$ is the length of e. In this way we have assumed that the numerical flux depends only on the values on each side of the edge and on the normal direction to the edge (obviously it depends also on the continuous flux $\vec{F}$).
In the general case, the numerical flux H is assumed to be locally Lipschitz continuous and must be both conservative and consistent.

DEFINITION 2.9. *Conservation property of the numerical flux H defined by equation (2.24) can be written as*

$$H(u_l, u_j, n) = -H(u_j, u_l, -n).$$

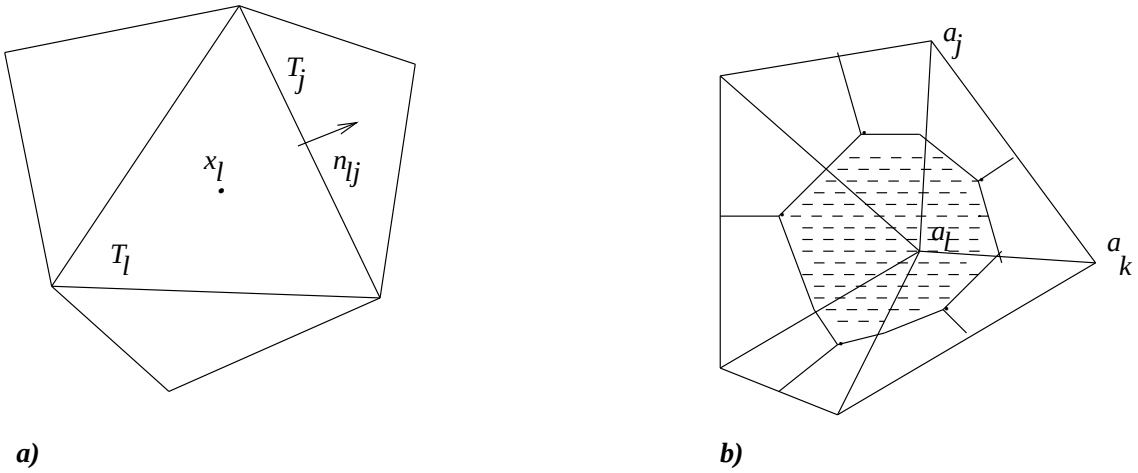

FIGURE 2.14: *a) Cell center; b) Cell vertex.*

This property is directly inherited from the continuous flux (2.24) and means that, in the absence of a source term, the approximate flux at the boundary separating T_l and T_j is the same as the flux at the boundary separating T_j and T_l (since $-n$ is the unit normal to e pointing in the direction of T_l).

DEFINITION 2.10. *Consistency property of the numerical flux H can be written as*

$$H(u, u, n) = \vec{F}(u) \cdot n.$$

Again, this arises naturally from (2.24).
Once this numerical flux has been applied to the approximation of the flux across the boundary of each cell, the following method of lines is obtained

$$|T_l| \frac{\partial u_l(t)}{\partial t} + \sum_{e \subset \partial T_l, e = e_{lj}} |e| H(u_l, u_j, n_e) = 0. \tag{2.25}$$

This ordinary differential equation can be solved by means of the explicit Euler scheme:

$$|T_l|(u_l^{k+1} - u_l^k) + \Delta t \sum_{e \subset \partial T_l, e = e_{lj}} |e| H(u_l, u_j, n_e) = 0,$$

where $u_l^k \cong u_l(t^k)$ and where u_l^0 is given.

There are several examples of FV methods. In particular there are the *cell center* schemes and the *cell vertex* schemes. In the cell center scheme, one defines a *control cell* as a triangle T_l, and the *center* of the cell is the centroid x_l of the triangle T_l (see Figure 2.14 a). Then the solution of (2.22) can be approximated, for example, by a function that is piecewise constant on each triangle. In this way the values u_l are associated to the centroid x_l since

$$u_l \cong \frac{1}{|T_l} \int_{T_l} u(x) \, \mathrm{d}\Delta = u(x_l, t) + \mathcal{O}(\Delta x^2),$$

where Δx is the diameter of the triangulation [25].

In the cell vertex scheme, the quantities are defined at the vertices of the triangulation. Thus, starting from a triangulation of the domain, one defines the *centers* as the

vertices of the triangles (the nodes of the triangulation, a_l), and the *control cell* associated to a_l is the *dual cell* of the node a_l. The polygonal boundary of the dual cell is obtained by joining, for each triangle having the vertex a_l in common, the midpoint of each triangle edge issued from a_l to the triangle barycenter. The boundary is thus composed of medians and a_l is not necessarily the centroid of the control cell [25] (see Figure 2.14 b).

The most simple usual FV schemes are first order accurate, as it can be seen in [25], where stability and convergence results are illustrated and a description of the most popular schemes as Roe's scheme is given.

Our attention is instead devoted to cell centered second order FV schemes. A second order (in space) version of a FV scheme can be obtained via the MUSCL approach. A piecewise linear reconstruction is used instead of piecewise constant function, together with a limiting procedure.

In formula (2.25), the first order flux $H(u_l, u_j, n_{lj})$, where u_l and u_j are the constant values on each side of the edge e_{lj}, is now replaced by $H(u_{lj}, u_{jl}, n_{lj})$, where u_{lj} and u_{jl} are second order approximations of the solution on each side of the edge e_{lj}. These second order approximations are computed through the following steps:

1. prediction of the gradients $\vec{\nabla} u_l$ in each cell;

2. linear extrapolation to define the values u_{lj}, u_{jl} on each side of the edge;

3. limiting procedure to damp over or under shoots.

In this way we choose to work with the MUSCL-type cell-centered approach developed by [16] and then modified by [34].

2.6.2 Triangular Finite Volume scheme

The technique developed by [16, 34] can be described as follows.

Equation (2.25) can be written as

$$\frac{\partial u_l}{\partial t} = -\frac{1}{|T_l|} \sum_{j=1}^{3} H_j(u_{lj}, u_{jl}, n_{lj}) |e_j|,$$

where e_{lj}, $j = 1, 2, 3$ are the edges of triangle T_l and $\vec{n}_{lj}$ is the corresponding outward unit normal. For simplicity, we write u_l instead of $u_l(t)$. The approximation of the three line integrals H_j, $j = 1,\ 2,\ 3$, in the above equation is obtained by a two step procedure. The reconstruction step approximates the values of u_l over each triangle. Second order accurate reconstruction is achieved by linear interpolation in combination with a limiting procedure that explicitly prevents the formation of overshoots and undershoots. The reconstructed values are used in the second step to build a two-point Lipschitz conservative and consistent monotone numerical flux approximating $\vec{F}(u_l) \cdot \vec{n}_{lj}$.

The reconstruction step, following [34], proceeds as follows: for the triangle T_l with centroid x_l, three linear interpolants are built using the values of the nearby triangles, say T_p, T_q, T_r of Figure 2.15. Denoting by x_j the pair of coordinates of the centroid of T_j, $j = l, r, p, q$, we construct L_l^1 as the linear interpolant of the points $\{(x_l, u_l),\ (x_p, u_p),\ (x_q, u_q)\}$, while L_l^2 is the linear interpolant of $\{(x_l, u_l),\ (x_q, u_q),\ (x_r, u_r)\}$

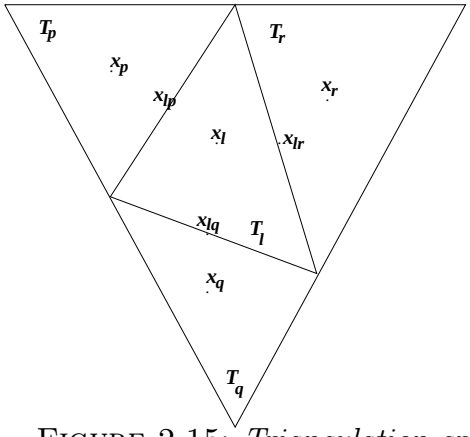

FIGURE 2.15: *Triangulation grid.*

and L_l^3 is the linear interpolant of $\{(x_l, u_l),\ (x_r, u_r),\ (x_p, u_p)\}$. If an edge of T_l is on the boundary, the value of u on the midpoint of that edge is used instead of the centroid value in the linear interpolation.

Once the three interpolants are calculated, the magnitude of the gradient of L_l^j can be expressed as:

$$|\nabla L_l^j| = \sqrt{(\frac{\partial}{\partial x} L_l^j)^2 + (\frac{\partial}{\partial y} L_l^j)^2} \qquad j = 1, 2, 3.$$

Starting from the L_l^j with maximum gradient and going toward the L_l^j with minimum gradient, we choose the first j such that

$$\begin{aligned} &L_l^j(x_{lp}) \text{ is between } u_l \text{ and } u_p \\ &L_l^j(x_{lq}) \text{ is between } u_l \text{ and } u_q \\ &L_l^j(x_{lr}) \text{ is between } u_l \text{ and } u_r, \end{aligned}$$

where x_{lp} is the midpoint of the edge sharing T_l and T_p, and so on. If no L_l^j satisfies these three requirements, we compute the local upper bound UB_l and the local lower bound LB_l of triangle T_l. They are defined, respectively, as the maximum and the minimum of the concentration values at the nearby points of the centroid of T_l.

DEFINITION 2.11. *We call x_j a nearby point of x_l, if it is the centroid of a triangle that has at least a common point with the triangle T_l.*

Therefore, starting again from L_l^j with maximum gradient we choose the first j such that L_l^j satisfies

$$\begin{aligned} UB_l &\geq \max\left(L_l^j(x_{lp}), L_l^j(x_{lq}), L_l^j(x_{lr})\right) \\ LB_l &\leq \min\left(L_l^j(x_{lp}), L_l^j(x_{lq}), L_l^j(x_{lr})\right). \end{aligned}$$

If no interpolant satisfies these inequalities, we choose as interpolant a piecewise constant reconstruction, that is L_l assumes a constant value equal to u_l.

Once the linear interpolation L_l is obtained, the reconstructed values at the midpoints of each edge of T_l from inside and outside the triangle, i.e. $L_l(x_{lj}^{\text{in}})$ and $L_l(x_{lj}^{\text{out}})$, respectively, are the boundary conditions for the local Riemann problem. The line integral

$\int_{e_{lj}} \vec{F}(u_l) \cdot \vec{n}_{lj} \, \mathrm{d}\Gamma$ is approximated using the midpoint formula by $H_j(L_l(x_{lj}^{\text{in}}), L_l(x_{lj}^{\text{out}}))|e_j|$, where H_j is the Godunov flux, and $|e_j|$ is the length of e_{lj}. The second order accurate semidiscrete approximation to (2.22) is then:

$$\frac{\partial u_l}{\partial t} = -\frac{1}{|T_l|} \sum_{j=1}^{3} H_j(L_l(x_{lj}^{\text{in}}), L_l(x_{lj}^{\text{out}}))|e_j| \tag{2.26}$$

We remark that the expression *second order accurate* means that the scheme approximates the flux to second order accuracy. At this point, the right-hand side of (2.26) can be integrated in time with time step Δt over the time interval $[t^k, t^{k+1}]$. In [16, 34] the time integration is accomplished via a second-order TVD Runge-Kutta procedure (Heun formula):

$$\begin{aligned} u_l^1 &= u_l^k - \Delta t E(x, u^k), \\ u_l^{k+1} &= \tfrac{1}{2} u_l^k + \tfrac{1}{2}(u_l^1 - \Delta t E(x, u^1)), \end{aligned}$$

where E is the numerical flux defined as

$$E(x,u) = \frac{1}{|T_l|} \sum_{j=1}^{3} H_j(L_l(x_{lj}^{\text{in}}), L_l(x_{lj}^{\text{out}}))|e_j|.$$

Another way of time integration is to consider the midpoint rule (Runge formula),

$$\begin{aligned} u_l^1 &= u_l^k - \tfrac{1}{2}\Delta t E(x, u^k), \\ u_l^{k+1} &= u_l^k - \Delta t E(x, u^1). \end{aligned}$$

It has been proved in [34] that the scheme with the above linear interpolation and the second order TVD Runge-Kutta scheme satisfies the maximum principle under the CFL condition:

$$\text{CFL} \leq \frac{1}{3},$$

where

$$\text{CFL} = \Delta t_a \sup \frac{\overline{T_l}}{|T_l|} \sup |\frac{d\vec{F}}{dc}|,$$

$\overline{T_l}$ being the perimeter of T_l.

2.6.3 Numerical experiments

Now we present some numerical model problems to verify the numerical convergence rate of the proposed scheme. We must take into account that both schemes, the one introduced by [16] and the one developed by [34], do not reach global second order accuracy but only display superlinear convergence. In [16] the observed order of convergence, in L_1 norm, varies from 1.55 to 1.85 - depending on the test experiments. In [34] the observed order of accuracy is 1.6 in L_1 norm. Therefore we do not expect that a different time integration scheme can dramatically change the results. To this aim we compare the relative errors computed in L_1 and L_2 norms ($\mid e_{\ell,1} \mid$ and $\mid e_{\ell,2} \mid$) and the corresponding convergence rates,

by applying as time integration schemes the classical first order Euler scheme (Eu), the midpoint rule (Mp) and the Runge-Kutta scheme (RK).

For all the following tests we consider the solution at time $t^k = 0.1$ s and $t^k = 1$ s. The domain of integration is the square $[0,1] \times [0,1]$ discretized using uniform triangular elements. The coarsest mesh ($\ell = 1$) is characterized by 200 triangles and 121 edges, while the finest level ($\ell = 4$) is characterized by 12800 triangles and 6561 edges.

We consider the simple linear conservation law

$$u_t + \vec{\nabla} \cdot (\vec{v}u) = 0.$$

Periodic boundary conditions are imposed in both the $x-$ and the $z-$ directions.

Example 1. First of all, we consider the case $\vec{v} = (1, 0)$, with initial condition

$$u_0(x, z) = \sin(2\pi x) \sin(2\pi z).$$

For our simulations, the CFL number is set to 0.27. The time step size ranges from $\Delta t = 4 \times 10^{-3}$ s ($\ell = 1$) to $\Delta t = 0.5 \times 10^{-3}$ s ($\ell = 4$), and is halved at each level. As we can observe from Table 2.1, at time $t^k = 0.1$ s difference between the three temporal schemes (Euler, midpoint or Runge-Kutta) is not relevant. If we consider least squares approximation (l.s.a.), we obtain an order of accuracy of about 1.36 in the L_1 norm and 1.24 in the L_2 norm. These results are in agreement with those achieved by [16], since in that paper least squares linear fits give 1.55 for the same test problem in L_1 norm, but with different CFL values.

The fact that the Euler scheme does not interfer with the overall accuracy can be explained by considering that the final time of observation is very small, so that the spatial order of accuracy is prevailing. In Table 2.2 we have put the errors and rate of convergence for $t^k = 1$ s. We can observe that the Euler scheme achieves an order of accuracy of 1.12 in the L_1 norm and 1.09 in the L_2 norm. By least square approximation we obtain 1.14 and 1.13, respectively. This is in accordance with the fact that we are using a scheme that is first order accurate in time. By using the second order in time schemes we do not observe a decrease in accuracy. In fact we obtain almost 1.24 in the L_1 norm and 1.20 in the L_2 norm. In terms of least squares approximation a value of 1.32 is achieved in L_1 norm, while 1.30 is reached in L_2 norm for both second-order schemes.

Example 2. As a last example we consider a case showed in [34] with $\vec{v} = (-1, 0)$. As initial condition we set

$$u_0(x, z) = \sin(2\pi x) \sin(4\pi z/\sqrt{3}).$$

Again, CFL is set equal to 0.27, with the same Δt as in the previous example (indeed velocity does not change in absolute value). In Table 2.3 we compare the error values obtained with Euler, midpoint and Runge-Kutta in the L_1 and L_2 norms at time $t^k = 0.1$ s. The results are in agreement with those obtained in [34], where an order of convergence of 1.6 is achieved for the same test case, even if on a different spatial grid (an equilateral triangulation). In our example we observe a similar behavior of convergence in L_1 norm, where we obtain 1.60 with the Euler scheme and 1.66 with the midpoint and Runge-Kutta schemes. In terms of least squares approximation we achieve 1.59 with Euler, and 1.61

TABLE 2.1: *Example 1:* $t = 0.1$ *s.*

	Eu		Mp		RK	
ℓ	$\mid e_{\ell,1} \mid$	rate	$\mid e_{\ell,1} \mid$	rate	$\mid e_{\ell,1} \mid$	rate
1	1.12e-1		1.15e-1		1.12e-1	
2	4.09e-2	1.45	4.13e-2	1.48	4.01e-2	1.48
3	1.62e-2	1.34	1.63e-2	1.34	1.60e-2	1.32
4	6.57e-3	1.30	6.63e-3	1.30	6.59e-3	1.28
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.36		1.37		1.36
ℓ	$\mid e_{\ell,2} \mid$	rate	$\mid e_{\ell,2} \mid$	rate	$\mid e_{\ell,2} \mid$	rate
1	1.12e-1		1.13e-1		1.11e-1	
2	4.42e-2	1.34	4.42e-2	1.35	4.28e-2	1.37
3	1.87e-2	1.24	1.91e-2	1.21	1.86e-2	1.20
4	8.18e-3	1.19	8.43e-3	1.18	8.38e-3	1.15
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.26		1.24		1.24

TABLE 2.2: *Example 1:* $t = 1$ *s.*

	Eu		Mp		RK	
ℓ	$\mid e_{\ell,1} \mid$	rate	$\mid e_{\ell,1} \mid$	rate	$\mid e_{\ell,1} \mid$	rate
1	2.24e-1		2.40e-1		2.40e-1	
2	9.82e-2	1.19	8.69e-2	1.46	8.65e-2	1.47
3	4.51e-2	1.12	3.62e-2	1.26	3.57e-2	1.28
4	2.08e-2	1.12	1.53e-2	1.24	1.51e-2	1.24
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.14		1.32		1.32
ℓ	$\mid e_{\ell,2} \mid$	rate	$\mid e_{\ell,2} \mid$	rate	$\mid e_{\ell,2} \mid$	rate
1	2.27e-1		2.45e-1		2.46e-1	
2	9.91e-2	1.19	8.79e-2	1.48	8.78e-2	1.49
3	4.59e-2	1.11	3.71e-2	1.24	3.66e-2	1.26
4	2.16e-2	1.09	1.62e-2	1.19	1.59e-2	1.20
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.13		1.30		1.31

TABLE 2.3: *Example 2:* $t = 0.1$ *s.*

	Eu		Mp		RK	
ℓ	$\mid e_{\ell,1} \mid$	rate	$\mid e_{\ell,1} \mid$	rate	$\mid e_{\ell,1} \mid$	rate
1	1.30e-1		1.29e-1		1.28e-1	
2	4.06e-2	1.68	4.06e-2	1.67	4.13e-2	1.63
3	1.43e-2	1.50	1.41e-2	1.52	1.38e-2	1.58
4	4.71e-3	1.60	4.47e-3	1.66	4.45e-3	1.63
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.59		1.61		1.61
ℓ	$\mid e_{\ell,2} \mid$	rate	$\mid e_{\ell,2} \mid$	rate	$\mid e_{\ell,2} \mid$	rate
1	1.28e-1		1.26e-1		1.25e-1	
2	4.55e-2	1.49	4.50e-2	1.48	4.60e-2	1.44
3	1.72e-2	1.40	1.71e-2	1.39	1.66e-2	1.47
4	6.29e-3	1.45	6.09e-3	1.49	6.05e-3	1.46
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.44		1.45		1.46

with the other schemes, in perfect accordance with [34]. In L_2 norm we achieve about 1.50, while in least squares approximations we reach about 1.45 for the three schemes.

At time $t^k = 1$ s (Table 2.4), like in the previous example, we achieve first order of accuracy with the Euler scheme and about 1.30 for both norms and both second order schemes.

TABLE 2.4: *Example 2:* $t = 1$ *s.*

	Eu		Mp		RK	
ℓ	$\mid e_{\ell,1} \mid$	rate	$\mid e_{\ell,1} \mid$	rate	$\mid e_{\ell,1} \mid$	rate
1	2.69e-1		2.66e-1		2.65e-1	
2	1.58e-1	0.77	1.30e-1	1.03	1.31e-1	1.02
3	6.47e-2	1.29	4.84e-2	1.42	4.91e-2	1.41
4	2.97e-2	1.12	2.10e-2	1.20	2.07e-3	1.25
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.08		1.24		1.24
ℓ	$\mid e_{\ell,2} \mid$	rate	$\mid e_{\ell,2} \mid$	rate	$\mid e_{\ell,2} \mid$	rate
1	2.63e-1		2.58e-1		2.57e-1	
2	1.59e-1	0.73	1.32e-1	0.97	1.33e-1	0.95
3	6.86e-2	1.21	5.23e-2	1.33	5.27e-2	1.33
4	3.21e-2	1.09	2.41e-2	1.12	2.38e-2	1.15
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.03		1.16		1.16

3 Mixed Hybrid Finite Element Method for the dispersion equation

In this Chapter we present the development of the Mixed Hybrid Finite Element (MHFE) method on unstructured triangular meshes for the solution of dispersion equations. This approach will be coupled to the FV scheme described in the previous Chapter to yield the time-splitting technique that will be developed in Chapter 4. It will also be used in the discretization of the flow equation of the coupled equations that govern brine transport at the Lake Karachai. Therefore, we will use the same tool for the solution of the flow equation and of the dispersive part of the transport equation.

3.1 Introduction to Mixed Finite Element methods

An exhaustive analysis of Mixed and Hybrid Finite Element methods is formulated in many books and papers, like [6, 51, 47], as they have been widely developed since the seventies. Here we simply give a brief general abstract setting for the mixed formulation and some general existence and approximation results.

The starting-point of the classical Finite Element method is the solution of a physical problem that minimizes some functional (usually an *energy functional*) in a *well chosen* space of admissible functions W (usually a Hilbert space),

$$\inf_{w \in W} J(w). \tag{3.1}$$

If the functional $J(\cdot)$ is differentiable, the minimum (whenever it exists) will be characterized by a variational equation.

An approximate solution of (3.1) consists in looking for $w_m \in W_m$, where W_m is a finite-dimensional subspace of W, which is solution of the problem

$$\inf_{w_m \in W_m} J(w_m).$$

Therefore, the Finite Element (FE) method is a general technique to build finite-dimensional subspaces of a Hilbert space W, based on a few simple ideas: first of all, the domain $\Omega \subset \mathbb{R}^d, \quad d = 2, 3$, in which the problem is posed, is partitioned into a set of *simple* subdomains, called elements. These elements may be triangles, quadrilaterals, tetrahedra. A space W of functions defined on Ω is then approximated by *simple* functions, defined on each element with suitable matching conditions at interfaces. Simple functions are commonly polynomials or functions obtained from polynomials by a change of variables. In this way, we can state that a FE method can only be considered in relation with a variational principle and a functional space. Changing the variational principle and the space in which it is posed leads to a different FE approximation, even if the solution for the continuous problem can remain the same.

Now, let us consider the Mixed Finite Element (MFE) method, starting from a simple problem. Equation (1.1c) when only dispersion is considered and in the steady state case can be written as:

$$-\vec{\nabla} \cdot (D\vec{\nabla}c) = f \qquad \text{in } \Omega \tag{3.2a}$$

$$c = 0 \qquad \text{on } \partial\Omega \tag{3.2b}$$

where $D = D(x)$ is the dispersion tensor, and c represents the concentration.

Introducing the dispersive flux $\vec{G}$, we can write the Fick law,

$$\vec{G} = -D\vec{\nabla}c. \tag{3.3}$$

Then, it could be desiderable to approximate $\vec{G}$ and c simultaneously using a different Finite Element space for the two variables. With this purpose the problem (3.2) is decomposed into a first order system as follows:

$$\begin{aligned} \vec{G} + D\vec{\nabla}c &= 0 \qquad \text{in } \Omega && \text{(3.4a)} \\ \vec{\nabla}\cdot\vec{G} &= f \qquad \text{in } \Omega && \text{(3.4b)} \\ c &= 0 \qquad \text{on } \partial\Omega && \text{(3.4c)} \end{aligned}$$

Equation (3.4a) can be written as

$$D^{-1}\vec{G} + \vec{\nabla}c = 0 \qquad \text{in } \Omega.$$

Therefore, multiplying by test functions and integrating by parts we obtain the following weak formulation of problem (3.4):

$$\begin{aligned} &\int_\Omega D^{-1}\vec{G}\cdot\vec{w}\ \mathrm{d}\Delta - \int_\Omega c\vec{\nabla}\cdot\vec{w}\ \mathrm{d}\Delta = 0 && \forall\vec{w}\in H(div,\Omega) \\ &\int_\Omega \psi\vec{\nabla}\cdot\vec{G}\ \mathrm{d}\Delta = \int_\Omega f\psi\ \mathrm{d}\Delta && \forall\psi\in L^2(\Omega) \end{aligned} \tag{3.5}$$

where

$$H(div,\Omega) = \{\vec{w}\in L^2(\Omega)^d : \vec{\nabla}\cdot\vec{w}\in L^2(\Omega)\}$$

is the Hilbert space with the norm

$$\|\vec{w}\|_{H(div,\Omega)} = \|\vec{w}\|_{L^2} + \|\vec{\nabla}\cdot\vec{w}\|_{L^2}.$$

We observe that the weak formulation (3.5) involves the divergence of the solution and test functions and not arbitrary first derivatives. This fact allows us to work with the space $H(div,\Omega)$ formed by piecewise polynomial vector functions with continuous normal component.

Problem (3.5) can be written in the space $H(div,\Omega)\times L^2(\Omega)$ with the symmetric bilinear form:

$$b((\vec{G},c),(\vec{w},\psi)) = \int_\Omega D^{-1}\vec{G}\cdot\vec{w}\ \mathrm{d}\Delta - \int_\Omega c\vec{\nabla}\cdot\vec{w}\ \mathrm{d}\Delta - \int_\Omega \psi\vec{\nabla}\cdot\vec{G}\ \mathrm{d}\Delta$$

and the linear form

$$L((f,\psi)) = -\int_\Omega f\psi\ \mathrm{d}\Delta.$$

Indeed, $(\vec{G},c)\in H(div,\Omega)\times L^2(\Omega)$ is the solution of (3.5) if and only if

$$b((\vec{G},c),(\vec{w},\psi)) = L((f,\psi)) \qquad \forall(\vec{w},\psi)\in H(div,\Omega)\times L^2(\Omega).$$

In fact, taking $(\vec{w}, 0)$ and $(\vec{0}, \psi)$, we recover the two equations (3.5).

In order to define finite element approximations to the solution $(\vec{G}, c)$ of (3.5), we need to have finite element subspaces of $H(div, \Omega)$ and $L^2(\Omega)$.
Let $T = \{T_l\}_{l=1}^m$ be a triangulation of Ω, i.e. $\Omega = \cup_{T_l \in T} T_l$ with diameter $\leq h$. The triangulation is admissibile if the intersection of two triangles is either empty, or a vertex, or a complete side.
Thus, we have to construct piecewise polynomials spaces W_h and Ψ_h associated with T_l such that $W_h \subset H(div, \Omega)$ and $\Psi_h \subset L^2(\Omega)$.
The MFE approximation $(\vec{G}_h, c_h) \in W_h \times \Psi_h$ is defined by

$$\begin{aligned} \int_\Omega D^{-1}\vec{G}_h \cdot \vec{w} \, \mathrm{d}\Delta - \int_\Omega c_h \vec{\nabla} \cdot \vec{w} \, \mathrm{d}\Delta = 0 \qquad & \forall \vec{w} \in W_h \\ \int_\Omega \psi \vec{\nabla} \cdot \vec{G}_h \, \mathrm{d}\Delta = \int_\Omega f\psi \, \mathrm{d}\Delta \qquad & \forall \psi \in \Psi_h. \end{aligned} \tag{3.6}$$

In order to have stability and convergence W_h and Ψ_h can not be chosen arbitrarily but they have to be related. First of all we assume that

$$\vec{\nabla} \cdot W_h = \Psi_h. \tag{3.7}$$

Next, let $\pi_2 \vec{G}$ be the L^2- projection of $\vec{G}$ into Ψ_h, such that

$$\int_\Omega \vec{\nabla} \cdot (\vec{G} - \pi_2 \vec{G})\psi \, \mathrm{d}\Delta = 0 \qquad \forall \vec{G} \in H^1(\Omega)^d, \qquad \forall \psi \in \Psi_h \tag{3.8}$$

where $H^1(\Omega)$ is the well-known Sobolev space.

In the following, we recall the definition of the Sobolev space $H^m(\Omega)$.

DEFINITION 3.1. *Given m integer ≥ 0, the Sobolev space $H^m(\Omega)$ is defined as*

$$H^m(\Omega) = \{v \in L^2(\Omega) : D^\alpha v \in L^2(\Omega), \forall |\alpha| \leq m\}$$

where $D^\alpha v$ represents a partial derivative taken in the sense of distributions,

$$D^\alpha v = \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}, \ |\alpha| = \alpha_1 + \dots \alpha_d.$$

On this space, the norm is given by

$$\|v\|_{H^m(\Omega)} = \sum_{k \leq m} |v|_{k,\Omega}^2,$$

where

$$|v|_{k,\Omega}^2 = \sum_{|\alpha|=k} |D^\alpha v|_{L^2(\Omega)}^2.$$

The space $L^2(\Omega)$ is then $H^0(\Omega)$ while $H(div, \Omega)$ is a subset of $H^1(\Omega)$.
In the following we state two theorems without proof, regarding the convergence of the Mixed formulation. In particulat error estimates are specified for $\vec{G}$ and c, respectively.

THEOREM 3.2. *If the spaces W_h and Ψ_h are such that properties (3.7) and (3.8) hold, there exists a constant $\mathcal{C} > 0$ depending only on the bounds of the coefficients of the differential equation such that*

$$\|\vec{G} - \vec{G}_h\|_{H(div,\Omega)} \leq \mathcal{C}\|\vec{G} - \pi_2\vec{G}\|_{H(div,\Omega)}.$$

THEOREM 3.3. *If the spaces W_h and Ψ_h satisfy properties (3.7) and (3.8), and $\pi_2\vec{w}$ satisfies the relation*

$$\|\pi_2\vec{w}\|_{L^2} \leq \mathcal{C}\|\vec{w}\|_{H^1(\Omega)} \tag{3.9}$$

then there exists a constant $\mathcal{C}$ such that

$$\|c - c_h\|_{L^2(\Omega)} \leq C\{\|c - \tilde{\pi}_2 c\|_{L^2(\Omega)} + \|\vec{G} - \pi_2\vec{G}\|_{H(div,\Omega)}\} \tag{3.10}$$

where $\tilde{\pi}_2 c$ is the L^2-projection of c into Ψ_h.

3.2 Theorem of existence and uniqueness

In the previous Section we have stated some general results about error bounds assuming that there exists a solution to our problem. In this Section we state two theorems about existence and uniqueness of the solution of problem (3.5) and (3.6) respectively [51].

THEOREM 3.4. *The problem of finding a pair of functions $(\vec{G}, c) \in H(div,\Omega) \times L^2(\Omega)$ such that (3.5) holds has a unique solution. In addition, c is the solution of the elliptic problem (3.2) and $\vec{G} = -D\vec{\nabla}c$.*

Proof. For the sake of simplicity we prove the theorem for the case in which D is the identity matrix. In this way equations (3.5) become:

$$\int_\Omega \vec{G}\cdot\vec{w}\ \mathrm{d}\Delta - \int_\Omega c\vec{\nabla}\cdot\vec{w}\ \mathrm{d}\Delta = 0 \qquad \forall\vec{w}\in H(div,\Omega) \tag{3.11a}$$

$$\int_\Omega \psi\vec{\nabla}\cdot\vec{G}\ \mathrm{d}\Delta = \int_\Omega f\psi\ \mathrm{d}\Delta \qquad \forall\psi\in L^2(\Omega) \tag{3.11b}$$

Let us first check the uniqueness of the solution. Hence, assume that $f = 0$; from (3.11b) we get $\vec{\nabla}\cdot\vec{G} = 0$. Taking $\vec{w} = \vec{G}$ in (3.11a), we obtain $\vec{G} = 0$. Therefore, we have

$$\int_\Omega c\vec{\nabla}\cdot\vec{w}\ \mathrm{d}\Delta = 0 \qquad \forall\vec{w}\in H(div,\Omega) \tag{3.12}$$

Now, let $\psi \in H^1(\Omega)$ be a function such that

$$\vec{\nabla}\cdot(\vec{\nabla}\psi) = c \qquad \text{in } \Omega.$$

Then, by choosing $\vec{w} = \vec{\nabla}\psi$ in (3.12), we get $c = 0$.
It remains to show that the pair $(\vec{G} = -\vec{\nabla}c, c)$ is a solution of equations (3.11), where c is the solution of problem (3.2) (with $D = I$).
On the one hand, we have

$$-\vec{\nabla}\cdot\vec{G} + f = \vec{\nabla}\cdot\vec{\nabla}c + f = 0$$

and equation (3.11b) is proved.
On the other hand, since $c = 0$ on $\partial\Omega$, we get, by using the Green's formula,

$$\int_\Omega (\vec{G} \cdot \vec{w} - c\vec{\nabla} \cdot \vec{w}) \, \mathrm{d}\Delta = -\int_{\partial\Omega} c\vec{w} \cdot \vec{n} \, \mathrm{d}\Gamma = 0$$

and equation (3.11a) is verified.
We recall the Green's formula

$$\int_\Omega (\vec{\nabla}\psi \cdot \vec{w} + \psi\vec{\nabla} \cdot \vec{w}) \, \mathrm{d}\Delta = \int_{\partial\Omega} \psi\vec{w} \cdot \vec{n} \, \mathrm{d}\Gamma$$

where $\vec{n}$ represent the unit outward normal along the boundary $\partial\Omega$.

By considering $D \neq I$ the proof is slightly modified but in not relevant way. So the proof is concluded. □

When we consider the mixed formulation defined in two finite-dimensional spaces, we have the problem of finding a pair of functions $(\vec{G}_h, c_h) \in W_h \times \Psi_h$ such that equations (3.6) are verified.

About existence and uniqueness of the pair $(\vec{G}_h, c_h)$ we can state the following theorem, whose proof is omitted.

THEOREM 3.5. *Assume that*

$$\begin{cases} \vec{w}_h \in W_h \\ \int_\Omega \psi_h \vec{\nabla} \cdot \vec{w}_h \; d\Delta = 0 \qquad \forall \psi_h \in \Psi_h, \quad \Rightarrow \vec{\nabla} \cdot \vec{w}_h = 0 \end{cases}$$

and that there exists a constant $\mathcal{D}$ such that

$$\sup_{\vec{w}_h \in W_h} \left(\frac{\int_\Omega \psi_h \vec{\nabla} \cdot \vec{w}_h \; d\Delta}{\|\vec{w}_h\|_{H(div,\Omega)}} \right) \geq \mathcal{D} \|\psi_h\|_{L^2(\Omega)}.$$

Then the problem (3.6) has a unique solution $(\vec{G}_h, c_h) \in W_h \times \Psi_h$ and there exists a constant $\mathcal{C}$ which depends only on $\mathcal{D}$ such that

$$\|\vec{G} - \vec{G}_h\|_{H(div,\Omega)} + \|c - c_h\|_{L^2(\Omega)} \leq \mathcal{C}(\inf_{\vec{w}_h \in W_h} \|\vec{G} - \vec{w}_h\|_{H(div,\Omega)} + \inf_{\psi_h \in \Psi_h} \|c - \psi_h\|_{L^2(\Omega)}).$$

3.3 The Lagrange multipliers

To apply the theoretical results about the mixed formulation, we need to define the spaces W_h and Ψ_h that approximate $W \subset H(div, \Omega)$ and $\Psi \subset L^2(\Omega)$ respectively. In the following, we will use the Raviart-Thomas spaces defined on a generic element $T_l \subset \Omega$ as

$$RT_k = (P_k)^d + xP_k$$

where k is an integer ≥ 0 and P_k is the space of polynomials of degree $\leq k$.
It can easily be seen that the dimension of RT_k is given by [6]

$$\dim RT_k = \begin{cases} (k+1)(k+3) & \text{for } d = 2 \\ \frac{1}{2}(k+1)(k+2)(k+4) & \text{for } d = 3. \end{cases}$$

LEMMA 3.6. *For $\vec{w}_h \in RT_k$ the following relations hold:*

$$\begin{aligned} \vec{\nabla} \cdot \vec{w}_h &\in P_k \\ \vec{w}_h \cdot \vec{n}|_{\partial T_l} &\in R_k. \end{aligned}$$

where R_k is the polynomial space defined on the edges e_j of each element T_l:

$$R_k = \{\phi : \phi \in L^2(\partial T_l), \phi|_{e_j} \in P_k, \forall e_j\}.$$

We consider the Raviart-Thomas space of degree zero, whereby the functions $\vec{G}$ and c can be approximated by:

$$\begin{aligned} \vec{G} \simeq \tilde{G} &= \sum_{l=1}^{m} g_l \vec{w}_l \\ c \simeq \tilde{c} &= \sum_{l=1}^{m} c_l \psi_l \end{aligned} \tag{3.13}$$

where $\vec{w}_l$ and ψ_l are vector and scalar basis functions. Since we are considering the RT_0 spaces, $\vec{w}_l$ are first order polynomial of the type:

$$\vec{w}_l = \begin{pmatrix} ax + b \\ az + c \end{pmatrix},$$

while ψ_l are P_0 polynomials equal to one on element T_l and zero elsewhere.

Using Lemma 3.6 one can easily see that the basis functions $\vec{w}_l$ are defined on the edge e_j of T_l and can be chosen as to satisfy

$$\int_{e_j} \vec{w}_l \cdot \vec{n}_j \, \mathrm{d}\Delta = \delta_{lj}$$

where δ_{lj} is the Kronecker function.

The Lagrange multipliers have been introduced to avoid the trouble of solving a system of equations with an indefinite matrix. In fact, when we solve an elliptic problem by a mixed finite element formulation we obtain a system of the form:

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \mathbf{g} \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{f} \end{pmatrix} \tag{3.14}$$

and the matrix is indefinite, i.e. it has both positive and negative eigenvalues. Preconditioned Conjugate Gradient methods can be used to solve the above system, but they may perform poorly [4].

We note that when Dirichlet conditions are imposed on the boundary, i.e. $c = b_D$ on $\partial\Omega$, the vector $(\mathbf{0},\ \mathbf{f})^T$ becomes $(\mathbf{b},\ \mathbf{f})^T$, since, by applying the Green's formula we have to consider the term $-\int_{\partial\Omega} b_D \vec{w} \cdot \vec{n} \, \mathrm{d}\Gamma$.

To overcome the problem of solving an indefinite system, the hybrid formulation relaxes the continuity of the normal velocity across interelement edges. In this way, matrices A and B of (3.14) become both block diagonal and hence easily invertible. The continuity

of normal velocity, which is essential in our discretization, is then back imposed through the introduction of an extra variable, the so-called Lagrange multipliers.

The Lagrange multipliers are constant functions on each edge of T_l, when we use the RT_0 spaces. In this way, we add a new variable in our problem: instead of finding a couple $(\vec{G}, c)$, we find a triple $(\vec{G}, c, \lambda)$. The function λ plays an important role as it represents the value of the concentration head at the boundary of each element T_l.

We can summarize this new situation in the following theorem [6], by considering the model problem (3.2) with Dirichlet and Neumann boundary conditions, that is

$$\begin{array}{ll} -D\vec{\nabla}c = f & \text{in } \Omega \\ c = b_D & \text{on } \Gamma_D \\ \dfrac{\partial c}{\partial n} = 0 & \text{on } \Gamma_N \end{array} \tag{3.15}$$

THEOREM 3.7. *Let $(\vec{G}_h, c_h)$ be the solution of (3.6) with Dirichlet and Neumann boundary conditions as in (3.15), and let λ_h be defined in the space of Lagrange multiplier by $\lambda_h = \sum_{j=1}^{m'} \lambda_{lj}\mu_j$, where m' represents the number of internal edges of triangulation and μ_j are piecewise constant basis functions.*
Then, the triple $(\vec{G}_h, c_h, \lambda_h)$ is the unique solution of the following equations:

$$\int_{T_l} D^{-1}\vec{G}_h \cdot \vec{w}\ d\Delta - \int_{T_l} c_h\vec{\nabla}\cdot\vec{w}\ d\Delta + \int_{\partial T_l} \lambda_h \vec{w}\cdot\vec{n}\ d\Gamma = \int_{\Gamma_D\cap T_l} b_D\vec{w}\cdot\vec{n}\ d\Gamma \tag{3.16a}$$
$$\forall \vec{w} \in W_h,\ \forall l = 1,\dots,m$$

$$\int_\Omega \psi\vec{\nabla}\cdot\vec{G}_h\ d\Delta = \int_\Omega f\psi\ d\Delta \tag{3.16b}$$
$$\forall \psi \in \Psi_h,\ \forall l = 1,\dots,m$$

$$\sum_{T_j}\int_{\partial T_j} \mu_l\vec{G}_h\cdot\vec{n}\ d\Gamma = 0 \tag{3.16c}$$
$$\forall \mu_l,\ \forall l = 1,\dots,m$$

In the mixed formulation there is continuity of the normal components of $\vec{G}$ across the interelement boundaries, that is, on $e_j = T_l \cap T_k$ (T_l and T_k with diameters h_1 and h_2, respectively), we can write

$$\vec{G}_{h_1}\cdot\vec{n}_{lj} + \vec{G}_{h_2}\cdot\vec{n}_{kj} = 0 \qquad j = 1,\dots,m'. \tag{3.17}$$

Now, there is a jump in the normal flux and we loose this continuity. We can also say that global mass conservation is not satisfied. Enforcement of mass conservation in the hybrid formulation is achieved by adding the weak form of equation (3.17), represented by equation (3.16c), that is

$$\int_{e_j} \vec{G}_{h_1}\cdot\vec{n}_l\ \mathrm{d}\Gamma + \int_{e_j}\vec{G}_{h_2}\cdot\vec{n}_k\ \mathrm{d}\Gamma = 0 \qquad j = 1,\dots,m'.$$

As consequence of the introduction of Lagrange multiplier, MHFE formulation gives rise to a symmetric positive definite system matrix with good conditioning properties, as we will see in Section 3.5.

3.4 Superconvergence results

When we consider lowest order Raviart-Thomas spaces, the MFE formulation displays global first order of accuracy for both $\vec{G}$ and c. In linear problems, under suitable conditions on the mesh and on the regularity of the solution of the continuous problem, it has been proved that this method achieves superconvergence for c in specific points of the mesh, called superconvergence points, that is the centroids of the triangles [13].
On the other hand, for $\vec{G}$ no superconvergence results are known for general triangular meshes. In [14] second order accuracy was observed at midpoint edges in a three-lines mesh, that is in a mesh of similar triangles for which every edge is parallel to one of three designated lines. In fact, for the elliptic problem

$$-\vec{\nabla} \cdot (D\vec{\nabla}c(x)) = f(x),$$

where $D = D(x)$ is a positive definite matrix, setting $\vec{G} = -D\vec{\nabla}c$ and imposing boundary conditions of the form $\vec{G} \cdot \vec{n} = b_N$, the following results for a three-lines mesh have been proved in [14].

LEMMA 3.8. *On a three-lines mesh, the vector* $\mathbf{g}$ *computed by the mixed method is super-close to the* π_2*-projection of the true solution* $\vec{G}$ *in the sense that*

$$\|\mathbf{g} - \pi_2\vec{G}\|_{L^2} \leq \mathcal{C}h^2\|\vec{G}\|_{H^2}$$

for $\vec{G} \in (H^2(\Omega))^2$*, being* h *the maximum diameter of the triangulation.*

In this way, given a postprocessing scheme $R : W_h \to L^2(\Omega)$, the following theorem can be stated.

THEOREM 3.9. *Let* $R : W_h \to L^2(\Omega)$ *be a linear operator satisfying*

$$\|\vec{G} - R(\pi_2\vec{G})\|_{L^2} \leq \mathcal{C}h^2\|D^2\vec{G}\|_{L^2} \qquad \forall \vec{G} \in (H^2(\Omega))^2$$

and

$$\|R(\vec{w})\|_{L^2} \leq \mathcal{C}\|\vec{w}\|_{L^2} \qquad \forall \vec{w} \in W_h.$$

Then in any contest for which the vector $\mathbf{g}$ *computed by the mixed method is super-close to the* π_2*-projection of the true solution* $\vec{G}$ *in the sense described in Lemma 3.8, we have that*

$$\|R(\mathbf{g}) - \vec{G}\|_{L^2} \leq \mathcal{C}h^2\|\vec{G}\|_{H^2}.$$

Therefore a local postprocessing technique recovers second order accuracy of $\vec{G}$ on three-lines meshes.

Remark. Do not confuse the partial derivative taken in the sense of distributions, $D^2\vec{G}$, written in the previous theorem, with the square of the matrix D.

3.5 The numerical solution of the dispersive flux

In this Section we consider the solution of parabolic equations tha we obtain adding a temporal partial derivative of concentration in the first side of equation (3.2). This kind of equation arises in different application fields, such as potential flow, groundwater flow,

petroleum reservoir simulation. In particular, the numerical solution of groundwater flow equation yields a system of pressure and velocity values. Often the pressure and velocity fields are subsequently used for the solution of a companion contaminant transport problem, leading to a coupled system of flow and transport equations. For this reason it is important to well approximate pressure and velocity, to assure accuracy and, most importantly, to conserve mass of the velocity fields. The MFE method provides an attractive framework for these type of problems. By simultaneously approximating pressure and normal fluxes, velocities calculated by MFE method automatically satisfy conservation of mass. Since in elliptic equations, that is in steady-state flow problems, the indefinite mixed matrix system becomes ill-conditioned, it is preferable to use the MHFE method, that gives rise to a symmetric positive definite system matrix with good conditioning properties.

In [4] a two-dimensional MHFE model is developed for the solution of the nonlinear equation of variably saturated flow in groundwater on unstructured meshes. In [35] a MHFE scheme for linear dispersion equations on unstructured meshes is presented. Here we develop the same MHFE method, and obtain an accurate and efficient implicit in time scheme (nonlinearities are resolved by iteration, like in [4]), with no restriction on time step. Moreover we define the Lagrange multipliers on all edges of the triangulation.

In Chapter 5 we will consider a coupled flow and transport problem for radioactive contaminant and it will be very important to have accurate solutions for the pressure and velocity fields. On the other hand, to solve the transport equation (see Chapter 4), we will use a time-splitting technique, by solving two different PDEs separately, one containig the advective flux and one containing the dispersive flux. Thus, for both the flow equation and the dispersive part of the transport equation, we will use MHFE as described in the following paragraph.

For simplicity, we continue to use the same notations of the previous Sections and we consider the dispersion equation for the concentration:

$$\begin{aligned} &\frac{\partial \phi c}{\partial t} + \vec{\nabla} \cdot (-D\vec{\nabla} c) = f \qquad \text{on } \Omega \times (0,T], \\ &c = c^0 \qquad \text{on } \partial\Omega \times 0, \\ &c = b_D \qquad \text{on } \Gamma_D \times (0,T], \\ &-D\vec{\nabla} c \cdot \vec{n} = b_N \qquad \text{on } \Gamma_N \times (0,T] \end{aligned} \tag{3.18}$$

where ϕ is the porosity of the medium.
Denoting again by $\vec{G}$ the dispersive flux, $\vec{G} = -D\vec{\nabla} c$, equation (3.18) may be written as:

$$\frac{\partial \phi c}{\partial t} + \vec{\nabla} \cdot \vec{G} = f \qquad \text{on } \Omega \times (0,T] \tag{3.19a}$$

$$\vec{G} = -D\vec{\nabla} c \tag{3.19b}$$

We apply a lowest order Raviart-Thomas MHFE formulation to this problem working on unstructured meshes. Thus Ω is discretized into m triangles, T_l, $l = 1, \ldots, m$.

In this way, concentration c can be approximated by

$$c \simeq \tilde{c} = \sum_{l=1}^{m} c_l \psi_l, \tag{3.20}$$

while dispersive flux $\vec{G}$ can be defined on each triangle by

$$\vec{G}_l \simeq \tilde{G}_l = \sum_{j=1}^{3} \hat{g}_{jl} \vec{w}_{jl} \quad l = 1, \ldots, m \tag{3.21}$$

where $\vec{w}_{jl}$ are the discontinous $RT0$ vector basis functions. We note that the basis functions for the flux are defined as in the mixed case, but their local support is only one triangle. This is due to the fact that we eliminate the continuity of the normal components of the numerical flux $\vec{G}$ across the inter-element boundaries.

Multiplying equation (3.19a) by ψ_l and integrating in space and time, with time-step Δt over the time interval $[t^k, t^{k+1}]$, the following semidiscrete equations are obtained:

$$\phi_l^{k+1} c_l^{k+1} = \phi_l^k c_l^k - \frac{\Delta t}{|T_l|} \int_{T_l} [\vec{\nabla} \cdot \vec{G}(c^{k+\theta}) - f^{k+\theta}] \, \mathrm{d}\Delta \qquad l = 1, \ldots, m \tag{3.22}$$

where c_l^k is the volume average over T_l defined by

$$c_l^k = \frac{\int_{T_l} c(\cdot, t^k) \, \mathrm{d}\Delta}{|T_l|},$$

$|T_l|$ is the area of T_l, ϕ is considered constant within each triangle, and a weighted scheme is used for the time quadrature with weighting parameter $\theta \in \{0.5, 1\}$ and $c^{k+\theta} = \theta c(\cdot, t^{k+1}) + (1-\theta) c(\cdot, t^k) = \theta c^{k+1} + (1-\theta) c^k$. Therefore, for $\theta = 1$ we have the implicit Euler scheme, while for $\theta = 0.5$ we use the Crank-Nicolson integration.

Implementation of the MHFE produces the following system of linear equations:

$$\int_{T_l} D_l^{-1} \vec{G}_l \cdot \vec{w}_{il} \, \mathrm{d}\Delta - \int_{T_l} c \vec{\nabla} \cdot \vec{w}_{il} \, \mathrm{d}\Delta + \int_{\partial T_l} \lambda \vec{w}_{il} \cdot \vec{n}_l \, \mathrm{d}\Gamma = 0 \tag{3.23a}$$

$$\frac{\phi_l^{k+1} |T_l| c_l^{k+1}}{\Delta t} + \int_{T_l} \vec{\nabla} \cdot \vec{G}_l \, \mathrm{d}\Delta = \frac{\phi_l^k |T_l| c_l^k}{\Delta t} + \int_{T_l} f_l \, \mathrm{d}\Delta \tag{3.23b}$$

$$\int_{e_j} \vec{G}_l \cdot \vec{n}_l \, \mathrm{d}\Gamma + \int_{e_j} \vec{G}_r \cdot \vec{n}_r \, \mathrm{d}\Gamma = 0 \qquad \text{if } e_j \in T_l \cap T_r \tag{3.23c}$$

$$\int_{e_j} \vec{G}_l \cdot \vec{n}_l \, \mathrm{d}\Gamma = b_N \qquad \text{if } e_j \in \Gamma_N \cap T_l \tag{3.23d}$$

$$\lambda_j = b_D \qquad \text{if } e_j \in \Gamma_D \tag{3.23e}$$

where $i = 1, 2, 3$, $l = 1, \ldots, m$, $j = 1, \ldots, n$, n being the number of edges, while the quantities with subscript l are defined over element T_l. The unknown Lagrange multiplier λ is expressed as

$$\lambda = \sum_{j=1}^{n} \lambda_j \mu_j,$$

where λ_j represents the trace of the concentration on e_j.

In the above system, equation (3.23a) is the MHFE discretization of the dispersive flux (3.19b); equation (3.23b) represents the discretized version of (3.19a); equation (3.23c) guarantees continuity of the normal flux across interelement edges, while equations (3.23d-3.23e) are the explicitly imposed Neumann and Dirichlet boundary conditions, respectively. The final hybrid formulation can be written in matrix notation as:

$$\begin{pmatrix} A & -B & Q \\ B^T & P' & 0 \\ Q^T & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{g}^{k+\theta} \\ \mathbf{c}^{k+\theta} \\ \boldsymbol{\lambda}^{k+\theta} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{f} + \tilde{P}\mathbf{c}^k \\ \mathbf{b}_N \end{pmatrix} \tag{3.24}$$

where $A = \text{diag}[A_1, \ldots, A_m]$, $B = \text{diag}[B_1, \ldots, B_m]$ and

$$\begin{aligned} A_l = (a_{ik}) &= \int_{T_l} D_l^{-1} \vec{w}_{il} \cdot \vec{w}_{kl} \, \mathrm{d}\Delta \quad B_l = (b_i) = \int_{T_l} \vec{\nabla} \cdot \vec{w}_{il} \, \mathrm{d}\Delta \\ Q = (q_{rj}) &= \int_{\partial T_l} \mu_j \vec{w}_{il} \cdot \vec{n}_l \, \mathrm{d}\Gamma \qquad \mathbf{g} = (g_r) = g_{il} \\ \mathbf{f} = (\hat{f}_l) &= \int_{T_l} f_l \, \mathrm{d}\Delta \end{aligned}$$

where $i, k = 1, 2, 3$, $j = 1, \ldots, n$, $r = 3(l-1) + i$ and $\mathbf{c} = (c_l)$, $\boldsymbol{\lambda} = (\lambda_j)$, and $\mathbf{b}_N = (b_{Nj})$ where b_{Nj} assumes a non vanishing value only if there is a Neumann condition on a boundary edge e_j. Setting $P^k = \text{diag}[p_1^k, \ldots, p_m^k]$ with $p_l^k = \phi_l^k |T_l| / \Delta t$ then $P' = \dfrac{P^{k+1}}{\theta}$ and $\tilde{P} = P' - P^{k+1} + P^k$. If p_l^k is not dependent on time, then $P' = \tilde{P}$.
The system (3.24) is generally not definite and with a number of unknowns equal to $4m + n$, approximately 6 times the num a non vanishing value only if there is a Neumaner of triangles of the discretized domain. Thus, the dimensionality of the problem is reduced by appriopriately transforming the system by means of Schur complements. The solution of the system proceed as follows. First of all, we obtain $\mathbf{g}^{k+\theta}$ from the upper system block:

$$\mathbf{g}^{k+\theta} = A^{-1}(B\mathbf{c}^{k+\theta} - Q\boldsymbol{\lambda}^{k+\theta}). \tag{3.25}$$

Note that matrix A is block-diagonal with 3×3 blocks and is easily invertible. By substituting this expression into the remaining equations, we obtain a reduced system with $\mathbf{c}^{k+\theta}$ and $\boldsymbol{\lambda}^{k+\theta}$ as unknowns:

$$\begin{pmatrix} P' + B^T A^{-1} B & -B^T A^{-1} Q \\ Q^T A^{-1} B & -Q^T A^{-1} Q \end{pmatrix} \begin{pmatrix} \mathbf{c}^{k+\theta} \\ \boldsymbol{\lambda}^{k+\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{f} + \tilde{P}\mathbf{c}^k \\ \mathbf{b}_N \end{pmatrix}. \tag{3.26}$$

Since P' and $B^T A^{-1} B$ are diagonal matrices, we can eliminate the unknown $\mathbf{c}^{k+\theta}$ from the first equation. By setting $H = P' + B^T A^{-1} B$ and $S = A^{-1} B$, we obtain:

$$\mathbf{c}^{k+\theta} = H^{-1}(S^T Q \boldsymbol{\lambda}^{k+\theta} + \mathbf{f} + \tilde{P}\mathbf{c}^k). \tag{3.27}$$

Again, matrix H is block-diagonal and easily invertible. By substituting equation (3.27) in the last equation of system (3.26), we obtain:

$$Q^T S H^{-1} S^T Q \boldsymbol{\lambda}^{k+\theta} - Q^T A^{-1} Q \boldsymbol{\lambda}^{k+\theta} = -Q^T S H^{-1}(\mathbf{f} + \tilde{P}\mathbf{c}^k) + \mathbf{b}_N.$$

Setting $M = A^{-1} - SH^{-1}S^T$, we have:

$$Q^T M Q \boldsymbol{\lambda}^{k+\theta} = Q^T S H^{-1}(\mathbf{f} + \tilde{P}\mathbf{c}^k) - \mathbf{b}_N. \tag{3.28}$$

The matrix $Q^T M Q$ is symmetric and positive definite and we can solve the system to obtain $\boldsymbol{\lambda}^{k+\theta}$.

To obtain $\boldsymbol{\lambda}^{k+1}$, we define $\boldsymbol{\lambda}^k_{diff} = \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k$. By the relationship $\boldsymbol{\lambda}^{k+\theta} = \theta\boldsymbol{\lambda}^k_{diff} + \boldsymbol{\lambda}^k$, we have:

$$Q^T M Q \boldsymbol{\lambda}^k_{diff} = \frac{1}{\theta} Q^T S H^{-1}(\mathbf{f} + \tilde{P}\mathbf{c}^k) + \frac{1}{\theta}\mathbf{b}_N - \frac{1}{\theta} Q^T M Q \boldsymbol{\lambda}^k.$$

Once $\boldsymbol{\lambda}^{k+1}$ is calculated $\mathbf{c}^{k+1}$ can be evaluated using equation (3.27). If $\tilde{P} = P'$ we get:

$$\mathbf{c}^{k+1} = (I - \frac{H^{-1}B^T S}{\theta})\mathbf{c}^k + \frac{H^{-1}}{\theta}(\mathbf{f} + S^T Q \boldsymbol{\lambda}^{k+\theta}),$$

otherwise

$$\mathbf{c}^{k+1} = (I - \frac{H^{-1}(B^T S + P^{k+1} - P^k)}{\theta})\mathbf{c}^k + \frac{H^{-1}}{\theta}(\mathbf{f} + S^T Q \boldsymbol{\lambda}^{k+\theta}).$$

It is now straightforward to obtain $\mathbf{g}$ from (3.25).

3.6 Numerical experiments

In this Section we report the numerical results and convergence rate obtained in the simulation of one and two-dimensional problems for c. The relative error ($\mid e_\ell \mid$) is computed at the centroids of the triangles, by using the L_2 norm. The rate of convergence is computed level after level and globally, by least squares approximation (l.s.a.).

The one-dimensional model problems are solved in a two-dimensional grid system. To this aim, five grid levels are used and defined as follows. At the coarsest level ($\ell = 1$) the rectangular domain is discretized into three layers of rectangular elements that are further subdivided into two triangles. The refined triangulations ($\ell = 2, \ldots, 5$) are obtained by connecting the midpoints of the three edges of each triangle. To reduce the dimensionality of the mesh, the height of the domain is always halved in passing from a coarser to the next finer level, in such a way that the shape of the triangles at the different levels is preserved. The coarsest mesh is defined on the rectangle $[0, 1] \times [0, 0.1]$ and is characterized by 300 triangles and 204 edges, while the finest level ($\ell = 5$) is defined on the rectangle $[0, 1] \times [0, 6.25 \times 10^{-3}]$ and is characterized by 4800 triangles and 3204 edges.

In the first two examples, Δt starts from 2×10^{-3} s ($\ell = 1$) and goes until 1.25×10^4 s ($\ell = 5$): at each next grid level we reduce the time step by half.

Example 1. In this example, we consider the simple problem whose analytical solution is given by:

$$c(x, t) = \mathrm{erfc}\frac{x}{2\sqrt{Dt}} \tag{3.29}$$

where the dispersion coefficient D is chosen so that $c(0, t) = 1$ and $c(1, t) = 0$. Convergence results for $D = 2 \times 10^{-2}$ m^2/s are reported in Table 3.1. Like predicted from the theory, we observe first order of convergence when using Euler scheme in time, and second order when using Crank-Nicolson.

TABLE 3.1: *Example 1:* $t = 1$ *s.*

	Eu		CN	
ℓ	$\mid e_\ell \mid$	rate	$\mid e_\ell \mid$	rate
1	1.13e-3		7.61e-4	
2	3.76e-4	1.59	1.90e-4	2.00
3	1.40e-4	1.42	4.74e-5	2.00
4	5.84e-5	1.26	1.19e-5	1.99
5	2.62e-5	1.16	2.98e-6	2.00
		l.s.a. rate		l.s.a. rate
		1.35		2.00

TABLE 3.2: *Example 2:* $t = 1$ *s.*

	Eu		CN	
ℓ	$\mid e_\ell \mid$	rate	$\mid e_\ell \mid$	rate
1	1.05e-1		3.85e-3	
2	5.06e-2	1.05	9.61e-4	2.00
3	2.48e-2	1.03	2.40e-4	2.00
4	1.23e-2	1.01	6.00e-5	2.00
5	6.12e-3	1.01	1.50e-5	2.00
		l.s.a. rate		l.s.a. rate
		1.02		2.00

Example 2. Again, in this example we prove first and second order convergence rate when time integration is given by Euler and Crank-Nicolson schemes, respectively. We consider the following one-dimensional linear problem:

$$\begin{aligned} &c_t - c_{xx} = 0, \qquad x \in [0,1], \quad t \in [0,1], \\ &c(0,t) = 0, \; c(1,t) = 0 \\ &c(x,0) = 2\sin(\pi x) \end{aligned}$$

whose exact solution is:

$$c(x,t) = 2e^{\pi^2 t}\sin(\pi x).$$

In Table 3.2 we report the relative errors and convergence rates of the numerical solution. The results are in good agreement with the theory.

Example 3. As last one-dimensional test case, we consider a steady-state problem, that is the solution is not dependent on time. Indeed, we consider the problem associated to the exact solution:

$$c(x,t) = \sin(\pi x).$$

Thus we can verify superconvergence at the centroid of the triangles, since error does not depend on time. In fact, we obtain the same results when we use Euler or Crank-Nicolson

TABLE 3.3: *Example 3:* $t = 1$ *s.*

ℓ	$\mid e_\ell \mid$	rate
1	2.93e-3	
2	7.31e-5	2.00
3	1.83e-5	2.00
4	4.57e-6	2.00
		l.s.a. rate
		2.02

TABLE 3.4: *Example 4:* $t = 10$ *s.*

	Eu		CN	
ℓ	$\mid e_\ell \mid$	rate	$\mid e_\ell \mid$	rate
1	7.44e-2		2.00e-3	
2	3.72e-2	1.00	5.20e-4	1.95
3	1.86e-2	1.00	1.33e-4	1.97
4	9.32e-3	1.00	3.37e-5	1.98
		l.s.a. rate		l.s.a. rate
		1.00		1.96

scheme in time. The values reported in Table 3.3 are obtained at time $t = 1$ s starting from $\Delta t = 4 \times 10^{-2}$ s ($\ell = 1$) and halving the time step at each next grid level. Superconvergence is achieved in good agreement with the theory.

Example 4. To test the Mixed Hybrid approach in a two-dimensional problem, we consider the problem whose exact solution is given by

$$c(\vec{x}, t) = \sin\,[t(\frac{x^3}{3} + \frac{z^3}{3} - \frac{x^2}{2} - \frac{z^2}{2})]$$

and calculate the source term f a posteriori from the governing equation. The domain is the unit square $[0, 1] \times [0, 1]$, discretized via right triangular elements. The coarsest mesh ($\ell = 1$) is characterized by 200 triangles and 121 edges, while the finest level ($\ell = 4$) is characterized by 12800 triangles and 6561 edges. The timestep is $\Delta t = .5$ s at $\ell = 1$ and, halving it at each next spatial grid, $\Delta t = 6.25 \times 10^{-2}$ s at $\ell = 4$. The errors are computed at the final time $t = 10$ s and reported in Table 3.4 together with convergence rates. As in the previous examples we note good agreement with the theory.

Part II
New results

4 A time-splitting technique for the advection-dispersion equation

In this Chapter we turn on attention to the solution of the advection-dispersion equation, with special reference to the equation describing solute transport in groundwater. The numerical tools described in the previous Chapters will be combined to form an accurate and efficient algorithm for the solution of this equation.

The main idea that is behind this approach consists in employing a time-splitting technique to allow the use of different spatial discretization schemes, namely MHFE and FV, applied to the dispersive and advective fluxes, respectively. In the following a first order in time technique is presented and studied in details. Next, we propose an extension to second order accuracy in time and space.

4.1 Introduction to the time-splitting technique

The advection-dispersion equation is difficult to approximate when advection dominates because sharp concentration fronts tend to develop and move without changing form. It is well known that standard finite difference and finite element methods may not work well for problems with sharp fronts, showing non-physical oscillations and numerical diffusion. To overcome these phenomena, numerical schemes try to combine numerical stability with minimal artificial diffusion. Two approaches are generally used in these situations. One is based on the definition of a proper control volume where upwind techniques can be used for approximating the advective flux. In this case the stability of the scheme is obtained by adding an amount of numerical diffusion that is dependent on the approach used [42]. The other class of methods originates from the splitting of the dispersion and advection fluxes into two separate partial differential equations (PDEs) containing one the dispersive and the other the advective term, respectively. These two equations are then discretized, each with the technique deemed most appropriate. Splitting allows the combination of explicit time-stepping for advective fluxes with implicit time-stepping for dispersive fluxes. This approach lessens the stability constraint connected with explicit discretization of the dispersion term but maintains the possibility of using efficient explicit schemes for advection. Belonging to this class are the Eulerian-Lagrangian schemes [39, 7] or the fully Eulerian Godunov-Mixed Methods (GMM) [9, 10, 11, 12]. In this latter approach, a time-explicit, spatially second-order accurate Godunov method is used to treat advection, and a time-implicit, spatially second order accurate Mixed Finite Element method is used for modeling dispersion.

Similar in spirit to the GMM approach, using Euler time-stepping, the advective term is discretized by a triangle-based, high resolution Finite Volume (FV) scheme [46, 16, 34], while the dispersive flux is discretized using a Mixed Hybrid Finite Element (MHFE) technique. The choice of these two schemes is dictated, on one hand by their accuracy, robustness and efficiency in handling nonuniform meshes and highly variable coefficients. On the other hand, both FV and MHFE are based on the weak formulation of the governing equation and use similar functional spaces for the approximation of the dependent variable, making them ideally suited for combination in a time-splitting approach. The main difference between the GMM approach and the proposed one lies in the unstruc-

tured character of the spatial discretizations of the latter. This implies, in contrast with the original GMM approach, the use of triangular meshes together with fully multidimensional slope limiters in the FV phase. More precisely, the time-splitting scheme employs explicit and implicit Euler time-stepping for FV and MHFE, respectively, while piecewise constant basis functions are used by both techniques to approximate concentration. Second order accuracy in space is obtained by MHFE at special superconvergence points (the centroids of the triangles) [13]. The FV approach achieves spatial second order accuracy (away from sharp fronts) by employing linear reconstruction plus slope limiting, combined in such a way as to locally satisfy the maximum principle [34]. The resulting numerical scheme is first order accurate in time and second order accurate in space. As we will see, to obtain second order accuracy also in time, it is not sufficient to consider a scheme of second order accuracy in time, but it is necessary to add a correction term in the advection equation.

Both MHFE and FV are locally (at the element level) conservative and monotone. The combination of the two methods in the time splitting approach should maintain these two properties as long as stability requirements are met, as confirmed also by numerical results. In principle there are no difficulties in extending our technique to three dimensions. This is done by merely employing three dimensional versions of MHFE and FV. Proper implementation of FV requires the development of tetrahedra based slope limiters, a field that is still subject of active research.

4.2 The numerical scheme

Subsurface contaminant transport is governed by an advection-dispersion equation of the form

$$\begin{array}{ll} \dfrac{\partial \phi c}{\partial t} + \vec{\nabla} \cdot (\vec{v}c - D\vec{\nabla}c) = f & \text{on } \Omega \times (0,T], \\ c = c^0 & \text{on } \Omega \times 0, \\ c = b_D & \text{on } \Gamma_D \times (0,T], \\ -D\vec{\nabla}c \cdot \vec{n} = b_N & \text{on } \Gamma_N \times (0,T] \\ (\vec{v}c - D\vec{\nabla}c) \cdot \vec{n} = b_C & \text{on } \Gamma_C \times (0,T] \end{array} \tag{4.1}$$

where c is the concentration of the solute, $\phi(t)$ is the porosity of the medium, $\vec{v}(\vec{x},t)$ is Darcy's velocity, $D = D(\vec{v})$ is the tensor accounting for mechanical dispersion and molecular diffusion, and f is a source or sink term ($\Omega \in \mathbb{R}^2$ and $\Gamma = \partial\Omega$).

Denoting by $\vec{F}$ and $\vec{G}$ the advective and dispersive fluxes, respectively, equation (4.1) may be written as:

$$\frac{\partial \phi c}{\partial t} + \vec{\nabla} \cdot (\vec{F} + \vec{G}) = f \qquad \text{on } \Omega \times (0,T] \tag{4.2a}$$

$$\vec{F} = \vec{v}c \tag{4.2b}$$

$$\vec{G} = -D\vec{\nabla}c \tag{4.2c}$$

As the geometry of the physical domain Ω is often complex when dealing with real world applications, we choose to work with unstructured meshes, and thus Ω is discretized into m triangles, T_l, $l = 1, \ldots, m$. Concentration c can be approximated as in (3.20) by:

$$c \simeq \tilde{c} = \sum_{l=1}^{m} c_l \psi_l,$$

where ψ_l are $P_0(T_l)$ scalar basis functions defined in Chapter 3.

Multiplying equation (4.2a) by ψ_l and integrating in space and time, with time-step Δt over the time interval $[t^k, t^{k+1}]$, the following semidiscrete equations are obtained:

$$\phi_l^{k+1} c_l^{k+1} = \phi_l^k c_l^k - \frac{\Delta t}{|T_l|} \int_{T_l} [\vec{\nabla} \cdot (\vec{F}(c^{k+1-\theta}) + \vec{G}(c^{k+\theta})) - f^{k+\theta}] \, \mathrm{d}\Delta \qquad l = 1, \ldots, m$$

where c_l^k is the volume average over T_l defined by

$$c_l^k = \frac{\int_{T_l} c(\cdot, t^k) \, \mathrm{d}\Delta}{|T_l|},$$

$|T_l|$ is the area of T_l, ϕ is considered constant within each triangle, and a weighted scheme is used for the time quadrature with weighting parameter $\theta \in [0.5, 1]$ and $c^{k+\theta} = \theta c^{k+1} + (1-\theta)c^k$.

Denoting by L_d the spatial discretization operator for dispersion, where we also include the source term f, and by L_a the spatial discretization operator for advection, the fully discretized equations become:

$$\phi_l^{k+1} c_l^{k+1} = \phi_l^k c_l^k + \Delta t \left[L_d(c^{k+\theta}) + L_a(c^{k+1-\theta}) \right] \qquad l = 1, \ldots, m \tag{4.3}$$

where, for $\theta = 1$ we have the implicit Euler scheme for L_d and explicit Euler scheme for L_a. For $\theta = 0.5$ the above equation reduces to the midpoint rule for both L_d and L_a. In the first part of this Chapter we consider only Euler schemes, i. e. $\theta = 1$.

The numerical fluxes L_a and L_d are evaluated by means of the discretization methods that are deemed more appropriate to solve, respectively, the advection and diffusion equations. For advection, we consider a high resolution triangular Finite Volume (FV) discretization by following the scheme studied in Chapter 2. This method requires explicit time-stepping and thus stability is guaranteed by a CFL restriction on Δt. Possible nonlinearities can also be resolved without iteration. This can be seen as a disadvantage of the time-splitting method with respect to fully implicit schemes, which are not impaired by stability constraints. However, in many problems of practical importance, CFL numbers less than unity are required to maintain accuracy. Thus stability constraints are automatically satisfied and do not pose limitations. The dispersive flux is discretized by an implicit MHFE method. This approach has been chosen because of its intrinsic compatibility with the FV method. Since it is implicit in time, there is no stability restriction on the time-step connected with MHFE.

4.2.1 The time-splitting technique

The time-splitting technique can be viewed as a predictor-corrector approach and can be described by the following algorithm:

ALGORITHM 4.1.
For each time step do:

- *advection step: for each T_l solve n_a times with the explicit FV scheme, with Δt_a as the time step, determining the predictor concentration $\hat{c}_l^{k+1}$*

1. $\phi_l^{(0)} := \phi_l^k \qquad c_l^{(0)} := c_l^k$

2. DO $i_a = 0, n_a - 1$

$$\phi^{(i_a+1)} c_l^{(i_a+1)} = \phi^{(i_a)} c_l^{(i_a)} + \Delta t_a \left[L_a(c_l^{(i_a)}) \right] \tag{4.4}$$

END DO

3. $\hat{c}_l^{k+1} := c_l^{(n_a)}$

- *dispersion step: for each T_l solve with implicit MHFE method using $\hat{c}_l^{k+1}$ as initial condition*

$$\phi^{k+1} c_l^{k+1} = \phi^k \hat{c}_l^{k+1} + \Delta t_d \left[L_d(c_l^{k+1}) \right] \tag{4.5}$$

with $\Delta t_d = n_a \Delta t_a$, obtaining the final approximation c_l^{k+1}.

Because the stability of the advection step is determined by the CFL constraint, while the dispersive time step is not subject to stability restrictions, we use two different time steps for advection and diffusion, Δt_a and Δt_d respectively. Therefore a finer advection time step together with a coarser diffusive time step can be employed. The convergence rate of the scheme is influenced by the convergence rates of the two spatial discretization methods employed. Namely, first order accuracy in time and second order accuracy in space is expected at the centroids of the triangles, as both FV and MHFE are spatially second order accurate. Accuracy is also influenced by the different time step sizes that can be used in the advection and dispersion discretizations. A heuristic analysis trying to find the optimal n_a in different situations will be reported in a later Section.

Finite Volume discretization. In the advection step, equation (4.4) can be explicitly written as:

$$\phi_l^{k+1} c_l^{k+1} = \phi_l^k c_l^k - \frac{\Delta t}{|T_l|} \int_{T_l} \vec{\nabla} \cdot (\vec{F}(c_l^k)\, \mathrm{d}\Delta \qquad l = 1, \ldots, m \tag{4.6}$$

and is solved using as initial condition the solution calculated at the end of the previous time step.

The spatially second order accurate discrete approximation to (4.6) is then:

$$\phi_l^{k+1} c_l^{k+1} = \phi_l^k c_l^k - \frac{\Delta t}{|T_l|} \sum_{j=1}^{3} H_j(L_l(x_{lj}^{\mathrm{in}}), L_l(x_{lj}^{\mathrm{out}}))|e_j| \tag{4.7}$$

where the terms are described in Section 2.6.2.

MHFE discretization. In the dispersion step, our implementation of the MHFE applied to the discretization of (4.5) produces the following system of linear equations:

$$\int_{T_l} D_l^{-1}\vec{G}_l \cdot \vec{w}_{il}\ \mathrm{d}\Delta - \int_{T_l} c\vec{\nabla}\cdot\vec{w}_{il}\ \mathrm{d}\Delta + \int_{\partial T_l} \lambda\vec{w}_{il}\cdot\vec{n}_l\ \mathrm{d}\Gamma = 0 \tag{4.8a}$$

$$\frac{\phi_l^{k+1}|T_l|c_l^{k+1}}{\Delta t_d} + \int_{T_l}\vec{\nabla}\cdot\vec{G}_l\ \mathrm{d}\Delta = \frac{\phi_l^{k}|T_l|c_l^{k}}{\Delta t_d} + \int_{T_l} f_l\ \mathrm{d}\Delta \tag{4.8b}$$

$$\int_{e_j}\vec{G}_l\cdot\vec{n}_l\ \mathrm{d}\Gamma + \int_{e_j}\vec{G}_r\cdot\vec{n}_r\ \mathrm{d}\Gamma = 0 \qquad \text{if } e_j \in T_l \cap T_r \tag{4.8c}$$

$$\int_{e_j}\vec{G}_l\cdot\vec{n}_l\ \mathrm{d}\Gamma = b_N \qquad \text{if } e_j \in \Gamma_N \cap T_l \tag{4.8d}$$

$$\lambda_j = b_D \qquad \text{if } e_j \in \Gamma_D \tag{4.8e}$$

where $i = 1, 2, 3$, $l = 1, \ldots, m$, $j = 1, \ldots, n$, n being the number of edges. The quantities with subscript l are defined over element T_l, while $|T_l|$ denotes the area of T_l. The dispersive flux $\vec{G}$ is approximated for each T_l by

$$\vec{G}_l = \sum_{j=1}^{3} g_{jl}\vec{w}_{jl} \quad l = 1, \ldots, m$$

where $\vec{w}_{jl} \in W_l$ are the discontinuous RT0 vector basis functions described in the previous Chapter.

In the above system, equation (4.8a) is the MHFE discretization of the dispersive flux (4.2c); equation (4.8b) represents the discretized version of (4.2a) without advective terms, as required by the time-splitting approach; equation (4.8c) guarantees continuity of the normal flux across inter-element edges, while equations (4.8d-4.8e) are the explicitly imposed Neumann and Dirichlet boundary conditions, respectively.

Again the developments that lead to the final system of equations are reported in Section 3.5.

4.2.2 Boundary conditions

The intrinsic nature of the time-splitting approach requires careful implementation of boundary conditions. In this respect, however, we are facilitated by the fact that in groundwater contaminant transport problems only a limited variety of boundary conditions is physically admissible (e.g. in general no boundary layers occur). To better describe how the boundary conditions are implemented in the proposed approach we distinguish between inflow and outflow boundaries. Inflow boundaries are characterized by having flow velocity normal components $\vec{v}\cdot\vec{n}$ directed inside the domain. Dirichlet or Cauchy boundary conditions may be of use in this situation. In the special case of $\vec{v}\cdot\vec{n} = 0$, Neumann type boundary conditions can also be employed. Implementation of these types of boundary conditions in the time splitting algorithm is obtained by specifying Dirichlet-type boundary conditions in the advective step and Neumann-type boundary conditions in the dispersive step. For example, inflow from a distributed source of contaminant can be specified as:

$$\vec{v}c\cdot\vec{n} = \vec{v}c_1\cdot\vec{n} \Rightarrow \text{ i.e. Dirichlet b. c. } c = c_1 \text{ for the advection step}$$
$$-D\vec{\nabla}c\cdot\vec{n} = 0 \Rightarrow \text{ i.e. zero Neumann flux for the dispersion step}$$

Outflow boundaries are characterized by outgoing velocities, and are easily implemented by imposing in the dispersion equation zero Neumann fluxes, as the outgoing advective flux is governed only by the velocity field.

Other type of conditions that may occur concern the presence of internal injection or extraction wells. Also in this case inflow or outflow are governed by the flow field and possible boundary conditions are easily implemented by Dirichlet plus zero Neumann and by zero Neumann conditions for injection and extraction wells, respectively.

4.2.3 Numerical results

Numerical tests on a one dimensional sample problem (movement of a tracer in a semi infinite column) are used to validate the theoretical results. A heuristic analysis on the relative role of the two discretization schemes in the convergence behavior of the proposed approach is aimed at determining the best time-stepping strategy for the explicit and the implicit schemes. At each implicit step, a number of explicit time steps can be performed, according to accuracy and stability requirements. In this way the proposed approach can be viewed also as a sub-stepping technique for the solution of the advective phase. In addition, we present a realistic test case - the Gureghian test - and compare the numerical results obtained by the proposed scheme with those available in the literature [23]. These test cases show that the proposed approach does not suffer from Peclet limitations and always displays small amounts of numerical diffusion, maintaining high order of accuracy across the entire spectrum of Peclet numbers and high computational efficiency.

We recall that the behavior of the proposed numerical scheme can be characterized as a function of two grid related dimensionless numbers, the Courant-Friedrichs-Lewy (CFL) number and the Peclet (Pe) number. The CFL number can be defined for each triangle T_l as [34]

$$\mathrm{CFL} = \Delta t_a \sup \frac{\overline{T_l}}{|T_l|} \sup |\frac{\mathrm{d}\vec{F}}{\mathrm{d}c}| \tag{4.9}$$

where $\overline{T_l}$ and $|T_l|$ denote the perimeter and the area of T_l, respectively. Stability of the FV scheme requires that $\mathrm{CFL} \leq \frac{1}{3}$. The Peclet number represents the ratio between the advective and the dispersive term and can be defined in our case as [46]

$$\mathrm{Pe} = \frac{\mathrm{CFL}}{\gamma} \tag{4.10}$$

where the dispersion number γ is given by:

$$\gamma = |D| \Delta t_a \sup \frac{1}{|T_l|} \tag{4.11}$$

and $|D|$ is the norm of tensor D. Low Peclet numbers indicate that dispersion is predominant over advection, and vice versa.

One dimensional tests. The numerical convergence rate of the time-splitting technique is tested on a one dimensional model problem solved in a two dimensional grid system. We

consider the partial differential equation describing the movement of a tracer in a semi-infinite column and simulate it on a rectangular domain of unit length, with $\vec{v} = (v, 0)$ and $D = \mathrm{diag}(D_1, D_1)$. The boundary conditions $c = 1$ at $x = 0$ and $c = 0$, for $x = \infty$ are imposed. Zero concentration is used as initial condition. This situation is simulated numerically by employing a grid of unitary length and making sure that at the time at which the relative error is evaluated the solution vanishes naturally at the right boundary. The analytical solution to this problem is [1]:

$$c(x,t) = \frac{1}{2}(\mathrm{erfc}\frac{x - vt}{2\sqrt{D_1 t}} + \exp\frac{vx}{D_1} \cdot \mathrm{erfc}\frac{x + vt}{2\sqrt{D_1 t}}). \tag{4.12}$$

The numerical convergence behavior of the scheme is evaluated by calculating errors at different grid levels. For a given level ℓ, we calculate the error $|e_\ell|$ in L_2 norm.

For all the subsequent test runs we consider the solution at $t^k = 0.1$ s.

Five grid levels are used and defined as follows. At the coarsest level ($\ell = 1$) the rectangular domain is discretized into three layers of rectangular elements that are further subdivided into two triangles. The refined triangulations ($\ell = 2, \ldots, 5$) are obtained by connecting the midpoints of the three edges of each triangle. To reduce the dimensionality of the mesh, the height of the domain is always halved in passing from a coarser to the next finer level, in such a way that the shape of the triangles at the different levels is preserved. The coarsest mesh is defined on the rectangle $[0, 1] \times [0, 0.1]$ and is characterized by 300 triangles and 204 edges, while the finest level ($\ell = 5$) is defined on the rectangle $[0, 1] \times [0, 6.25 \times 10^{-3}]$ and is characterized by 4800 triangles and 3204 edges. In the case of constant coefficients, Pe decreases by a factor of 2 in passing from a coarser to a finer level.

The first set of simulations is aimed at numerically verifying the theoretical convergence rate of the time-splitting scheme under different Pe and CFL numbers. Second order convergence rate can be observed on a problem with smooth solution (small Pe number) employing $\Delta t_d = \Delta t_a = h^2$ (h being the diameter of the triangulation). Table 4.1 reports the errors and convergence rates at the different levels for a case with $D_1 = D_2 = 1 \times 10^{-2}$ m^2/s and $v = 1$ m/s. Correspondingly the grid Peclet number varies between 9.22 ($\ell = 1$) to 0.58 ($\ell = 5$), while CFL goes to zero. First order convergence rate is instead achieved when $\Delta t = \mathcal{O}(h)$. Table 4.2 reports the results to the same problem obtained with constant CFL numbers (0.28 and 0.14, respectively) using a constant $\Delta t_d = \Delta t_a$. Convergence is still superlinear but seems to tend asymptotically to first order, as predicted by the theory.

The second set of simulations is aimed at determining the best time stepping strategy, i.e. the number n_a of advective time steps per dispersive time step, for which the error remains reasonably small and CPU time is minimal. It is intuitive to think that the behavior of the time-splitting approach depends on the given Peclet number. For small Pe, i.e. dominant dispersion, one expects convergence to be mainly driven by the MHFE technique discretizing the dispersion terms. The transient behavior of the solution should be well captured even for $n_a = 1$, i.e. $\Delta t_a = \Delta t_d$. On the other hand, for large Pe, the advective terms become important and thus the advective transient has to be accurately captured. We expect for this case the best accuracy when $\Delta t_a < \Delta t_d$, or $n_a > 1$. Verification of this behavior is obtained for a given Peclet number by comparing errors $\mid e_\ell \mid$, computed in L_2 norm, and CPU times for different values of n_a on a fixed mesh level. For this purpose we choose the mesh with 1200 triangles ($\ell = 3$). The dispersion coefficient varies in the range

TABLE 4.1: *One dimensional example: convergence behavior for* $\Delta t_d = \Delta t_a = h^2$.

ℓ	Pe	$\mid e_\ell \mid$	rate
1	9.22	1.24e-2	
2	4.61	3.67e-3	1.76
3	2.31	8.01e-4	1.98
4	1.15	2.07e-4	1.97
5	0.58	7.13e-5	1.86

TABLE 4.2: *One dimensional example: convergence behavior for* $\Delta t_d = \Delta t_a$*; CFL*=0.28 *(left) and CFL*= 0.14 *(right).*

ℓ	Pe	$\mid e_\ell \mid$	rate
1	9.22	1.47e-2	
2	4.61	5.40e-3	1.44
3	2.31	1.84e-3	1.50
4	1.15	7.67e-4	1.42
5	0.58	3.54e-4	1.34

ℓ	Pe	$\mid e_\ell \mid$	rate
1	9.22	1.27e-2	
2	4.61	4.25e-3	1.58
3	2.31	1.14e-3	1.74
4	1.15	4.02e-4	1.66
5	0.58	1.65e-4	1.57

$D = 2 \times 10^{-2} \div 0.5 \times 10^{-4}$ m^2/s, while velocity is kept constant at $v = 0.5$ m/s. These values correspond to Peclet numbers varying from 0.28, a dispersion dominated problem, to 115, a convection dominated case. The results of the different simulations are reported in Tables 4.3 to 4.6. Each column of the tables contains the results ($\mid e_3 \mid$, and CPU times in second) for a fixed Δt value and for the different n_a values tested. Subsequent columns (rows) are characterized by double Δt_d (n_a) values. The CFL number as well as Δt_a are thus constant along the main diagonals of the tables. For example, in Table 4.3 the advective time step is the same ($\Delta t_a = 0.25 \times 10^{-1}$ s) for the three cases $\Delta t_d = 0.25 \times 10^{-3}$ s and $n_a = 1$, $\Delta t_d = 0.5 \times 10^{-3}$ s and $n_a = 2$, and $\Delta t_d = 1 \times 10^{-3}$ s and $n_a = 4$.

When dispersion dominates, i. e. Pe = 0.28, the accuracy of the scheme is mainly influenced by the size of Δt_d, as can be seen in Table 4.3. First order convergence rate can be seen in every row where the values increase linearly with Δt_d. It is worth noting that

TABLE 4.3: *One dimensional example. Relative error norm* $\mid e_3 \mid$ *and CPU times for* $Pe = 0.28$.

Δt_d	$.25 \times 10^{-3}$		$.5 \times 10^{-3}$		1×10^{-3}	
n_a	$\mid e_3 \mid$	CPU	$\mid e_3 \mid$	CPU	$\mid e_3 \mid$	CPU
1	0.69e-3	35.94	1.42e-3	20.20	2.87e-3	12.07
2	0.56e-3	38.20	1.13e-3	21.51	2.30e-3	12.71
4	0.50e-3	44.16	1.00e-3	22.60	2.01e-3	13.95

TABLE 4.4: *One dimensional example. Relative error norm* $|\, e_3 \,|$ *and CPU times for* $Pe = 2.88$.

Δt_d	$.25 \times 10^{-3}$		$.5 \times 10^{-3}$		1×10^{-3}		2×10^{-3}	
n_a	$\mid e_3 \mid$	CPU	$\mid e_3 \mid$	CPU	$\mid e_3 \mid$	CPU	$\mid e_3 \mid$	CPU
1	2.21e-3	29.28	4.06e-3	14.77	7.91e-3	7.72		
2	1.55e-3	32.17	2.46e-3	16.06	4.42e-3	8.43	8.57e-3	4.67
4	1.23e-3	38.09	1.76e-3	19.08	2.86e-3	9.90	5.25e-3	5.44
8	1.10e-3	50.16	1.43e-3	25.10	2.13e-3	12.83	3.52e-3	6.78
16			1.31e-3	36.84	1.80e-3	18.48	2.62e-3	9.65
32					1.62e-3	30.25	2.30e-3	15.18

TABLE 4.5: *One dimensional example. Relative error norm* $|\, e_3 \,|$ *and CPU times for* $Pe = 28.8$.

Δt_d	1×10^{-3}		2×10^{-3}		4×10^{-3}		8×10^{-3}		16×10^{-3}		32×10^{-3}	
n_a	$\mid e_3 \mid$ $\times 10^2$	CPU (s)	$\mid e_3 \mid$ $\times 10^2$	CPU (s)	$\mid e_3 \mid$ $\times 10^2$	CPU (s)	$\mid e_3 \mid$ $\times 10^2$	CPU (s)	$\mid e_3 \mid$ $\times 10^2$	CPU (s)	$\mid e_3 \mid$ $\times 10^2$	CPU (s)
1	4.62	7.19										
2	2.25	7.98	4.70	3.98								
4	1.20	9.54	2.24	4.70	4.75	2.35						
8	0.78	12.55	1.19	6.21	2.26	3.08	4.47	1.66				
16	0.66	18.60	0.79	9.25	1.21	4.59	2.29	2.42	4.11	1.23		
32	0.61	37.80	0.63	15.37	0.79	7.59	1.28	3.94	2.30	2.02	3.76	0.99
64	0.60	55.56	0.59	27.40	0.63	13.52	0.86	7.17	1.36	3.50	2.38	1.73
128			0.59	51.01	0.59	25.50	0.70	13.48	0.99	6.65	1.54	3.08

TABLE 4.6: *One dimensional example. Relative error norm* $\mid e_3 \mid$ *and CPU times for Pe* $= 115$.

Δt_d	1×10^{-3}		2×10^{-3}		4×10^{-3}		8×10^{-3}		16×10^{-3}		32×10^{-3}	
n_a	$\mid e_3 \mid$ $\times 10^2$	CPU (s)	$\mid e_3 \mid$ $\times 10^2$	CPU (s)	$\mid e_3 \mid$ $\times 10^2$	CPU (s)	$\mid e_3 \mid$ $\times 10^2$	CPU (s)	$\mid e_3 \mid$ $\times 10^2$	CPU (s)	$\mid e_3 \mid$ $\times 10^2$	CPU (s)
1	6.06	7.28										
2	2.67	8.03	5.62	4.00								
4	1.78	9.51	2.64	4.71	5.85	2.32						
8	1.90	12.33	1.74	6.16	2.92	3.06	5.48	1.58				
16	2.10	18.39	1.88	9.03	1.60	4.50	2.66	2.33	4.46	1.16		
32			2.10	14.90	1.80	7.45	1.44	3.82	2.39	1.91	3.57	0.97
64					2.02	13.27	1.65	6.75	1.63	3.41	1.88	1.71
128									1.66	6.33	1.56	3.20
256											1.81	6.19

the relative error significantly decreases from the case with $n_a = 1$ to $n_a = 4$ (by about 30 %), whereas the CPU time increases only by 15% on the average. This fact suggests using $n_a > 1$ also for very small Peclet numbers.

At Pe = 2.88, an intermediate value of the Peclet number, (Table 4.4), we still recognize first order convergence rate along the rows. However, the error decrease along the diagonals is now much less pronounced, indicating that in this case small Δt_a are needed to maintain accuracy. These results are exemplified graphically in Figure 4.1 where four plots of $\mid e_3 \mid$ vs CPU time corresponding to four values of Δt_d are reported. For each curve, the data points refer to the n_a values of Table 4.4. Hence, the optimal $(\Delta t_d, n_a)$ combination can be found on the intersection of the envelope of the curves with the horizontal line corresponding to the desired accuracy. Obviously, there is no unique strategy to choose the optimal Δt values, however a few observations can be helpful for this purpose. Efficiency reasons demand that Δt_d be not too small, so as to minimize the number of linear system solutions. On the other hand, accuracy at this level of Peclet numbers already requires small Δt_a, and thus large $n_a (> 4)$. For example, a reasonable choice for our test case could be $\Delta t_d = 10^{-3}$ and $n_a = 16$.

At Pe = 28.8 advection starts to dominate over dispersion. The value of n_a is now important, as can be seen from the significant error decrease in the columns of Table 4.5. We also note that the error remains almost constant along the diagonals, i. e. for constant Δt_a and increasing Δt_d. Note that, for each Δt_d value, the increase in accuracy tends to be smaller as n_a increases, suggesting that after a certain value of n_a, convergence tends to stagnate. From this observation, we may argue that the truncation error of the scheme is proportional to Δt_d and to Δt_a:

$$\mid \epsilon_T \mid \approx O\left(\Delta t_d, h^2\right) + O\left(\Delta t_a, h^2\right) = \mid \epsilon_{T_d} \mid + \mid \epsilon_{T_a} \mid. \tag{4.13}$$

With this model we have that, for constant Δt_d, $\lim_{n_a \to \infty} \mid \epsilon_T \mid = \lim_{\Delta t_a \to 0} \mid \epsilon_T \mid = \mid \epsilon_{T_d} \mid$ explaining the experimented numerical behavior.

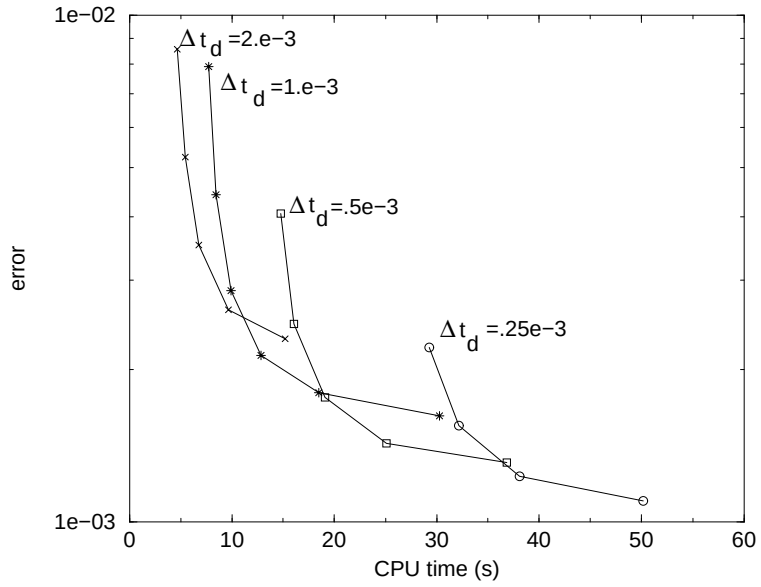

FIGURE 4.1: *Relative accuracy vs CPU time for different Δt_d and fixed Peclet number* $= 2.88$.

At even higher Peclet numbers, (Table 4.6, Pe = 115) the same behavior can be observed with one notable exception. For constant Δt_d the error attains a minimum value for a specific n_a. This behavior suggests that errors due to the operator splitting technique accumulate with the advective time step and thus $\epsilon_{T_a} = \epsilon_{T_a}(n_a)$.

Two dimensional infiltration of chloride ion in a surface aquifer. The applicability of the proposed approach is shown on a realistic two dimensional problem of infiltration of a conservative contaminant into a saturated-unsaturated surface aquifer. The test case considers a ditched-drained aquifer with incident steady rainfall and trickle infiltration of chloride ion [26]. The geometry of the domain and the boundary conditions employed in the solution of the flow and transport problems are described in Figure 4.2. The boundary conditions imposed along face AG correspond to the presence of a seepage face and thus to an outflow condition from which the aquifer is drained. The saturated-unsaturated flow equation is solved in steady state conditions by means of a Richards' equation solver based on the Mixed Hybrid Finite Element method [4]. The physical parameters of the simulation assume the following values: $V_r = 0.1$ cm/d, $V_s = 0.05$ cm/d (Fig. 4.2), $K_s = 1$ cm/d. The moisture retention curves of [30] are used with the following parameter values: $\alpha = 0.015$, $\beta = 2$, $\gamma = 3$, $a = 2$, $b = 3.5$, $\psi_s = -10$ cm, $S_{wr} = 0.01$. The Darcy velocity field $\vec{v} = (v_x, v_y)^T$ and water saturation values S_w as calculated from the solution of the flow problem are used in a 120-day simulation of the transport of the chloride ion. For this latter problem we have used a dispersion tensor $D = \text{diag}(D_1, D_2)$ as given by [1]: $D_1 = \alpha_L|\vec{v}| + nS_wD_0$ and $D_2 = \alpha_T|\vec{v}| + nS_wD_0$ where $|\vec{v}| = \sqrt{v_x^2 + v_y^2}$, $\alpha_L = 0.5$ cm is the longitudinal dispersivity, $\alpha_T = 0.1$ cm is the transverse dispersivity, $\phi = 0.30$ is the porosity of the medium, and $D_0 = 1.e-06$ cm^2/s is the molecular diffusion coefficient.

The mesh employed is made up of 2501 nonuniform triangles and 4800 edges. The

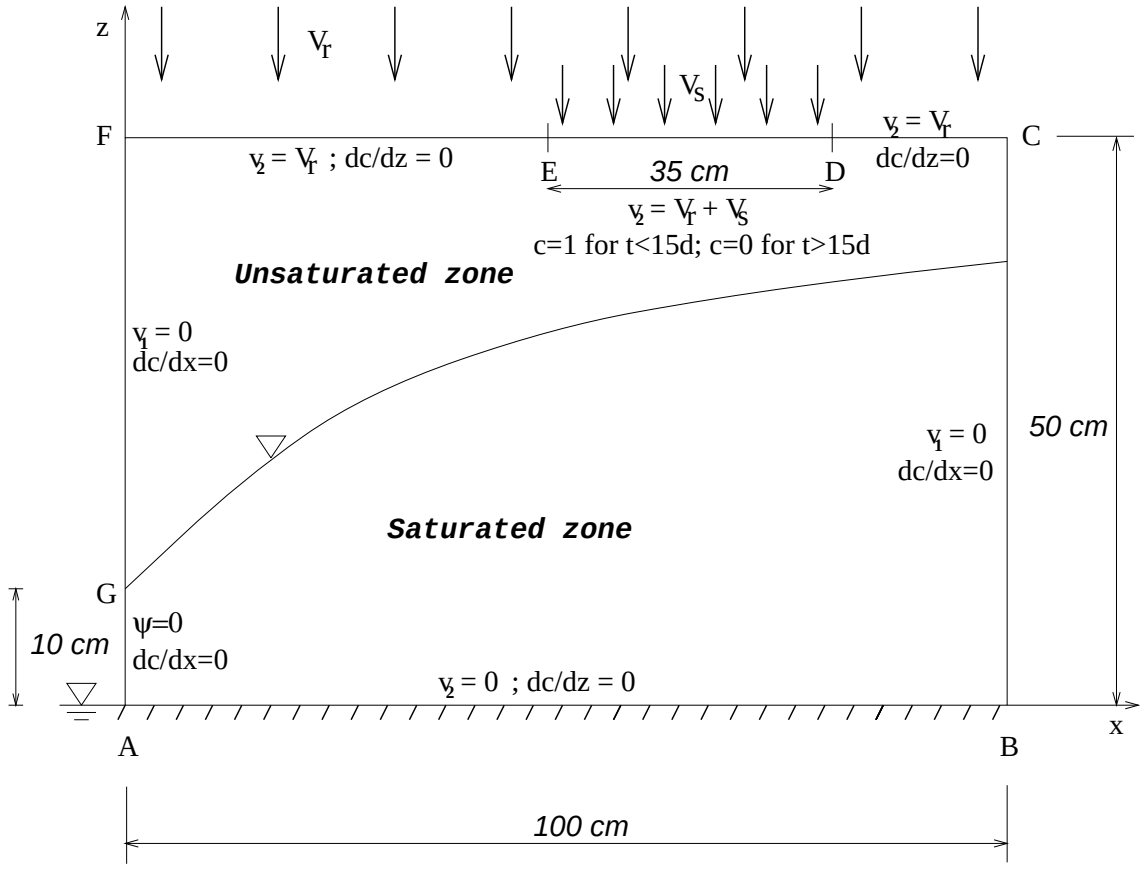

FIGURE 4.2: *Schematic description of the domain and boundary conditions for the two dimensional test case.*

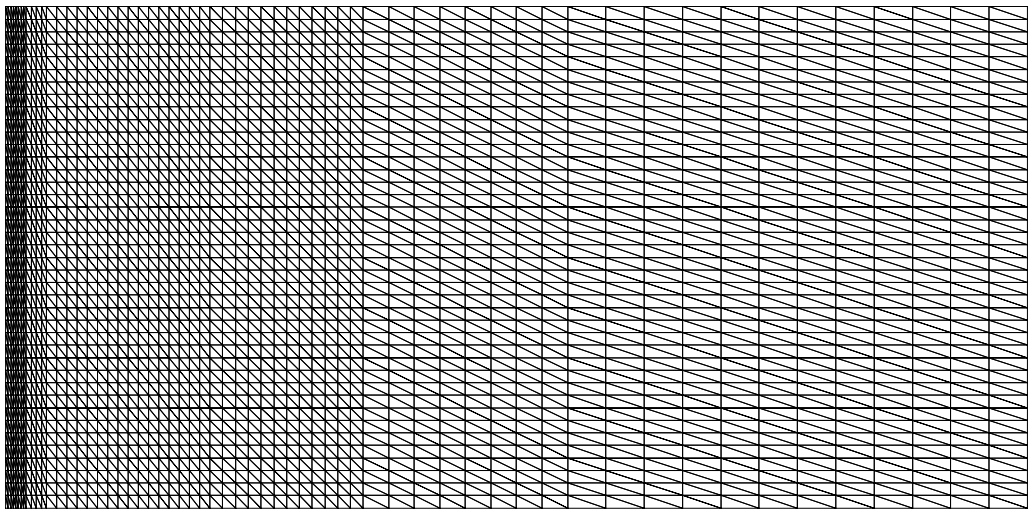

FIGURE 4.3: *Computational mesh used in the two dimensional test case.*

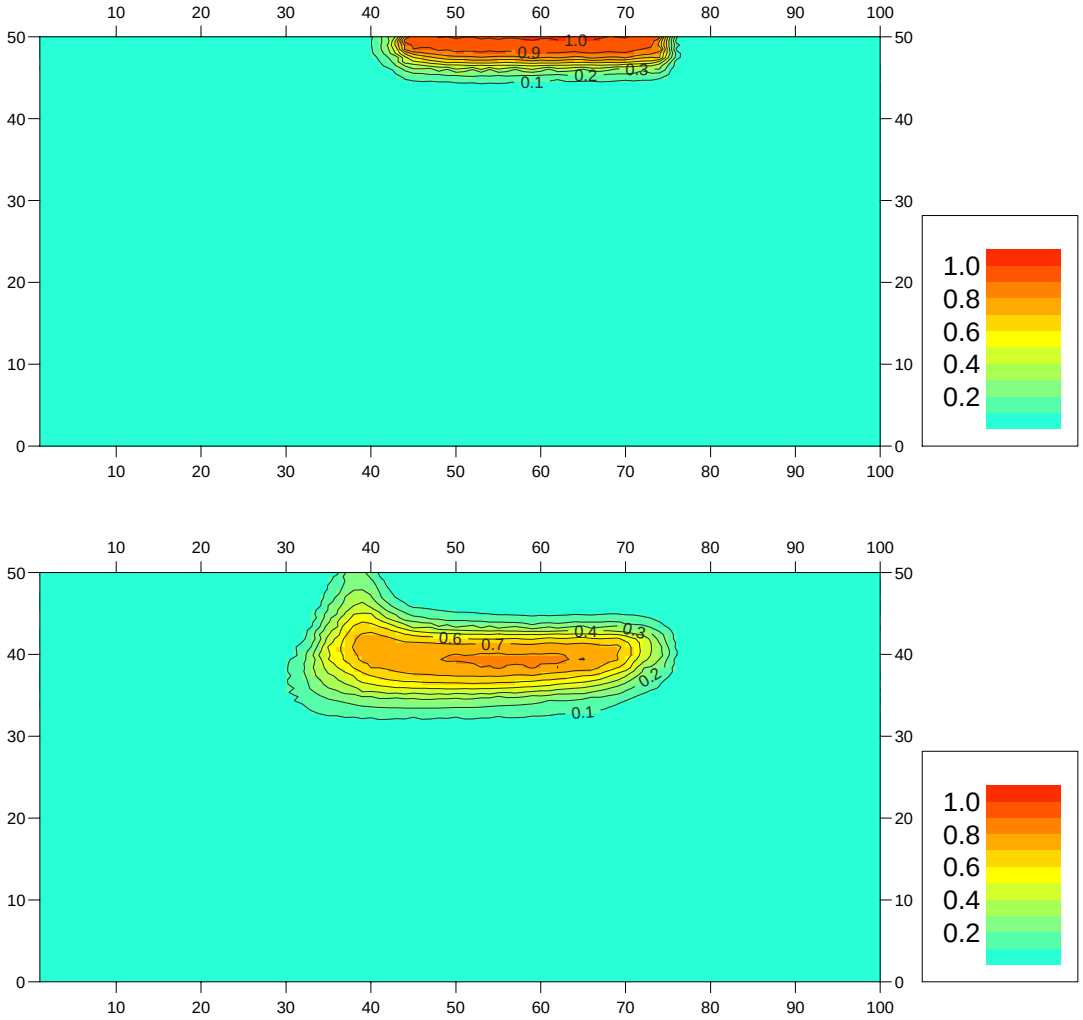

FIGURE 4.4: *Two dimensional test-case: concentration contours at 14.7 days (top) and 45.2 days (bottom).*

time step sizes are $\Delta t_d = 0.5$ days together with $n_a = 50$. The simulations is characterized by CFL = 0.44 and Pe = 5.55. With this choice of parameters the advective phase of the time-splitting algorithm is approximately twice as expensive than the dispersive phase (0.73 and 0.41 seconds per time step, respectively, on a DEC Alpha-600 workstation).

In Figures 4.4 and 4.5 the solute concentration contours at 14.7, 45.2, 90.2, and 120 days of simulation are shown. At the beginning the concentration plume infiltrates downwards with the unsaturated flow. Once it reaches the water table, it finds a more pronounced horizontal velocity in the saturated zone, starts moving towards the seepage face, and exits the domain.

A few observations from the numerical standpoint are worth mentioning. The solution obtained with the proposed approach does not present oscillations in any part of the domain and at any time. This verifies that the property of the FV scheme of being TVD is retained in the time-splitting algorithm, which maintains monotonicity in all our

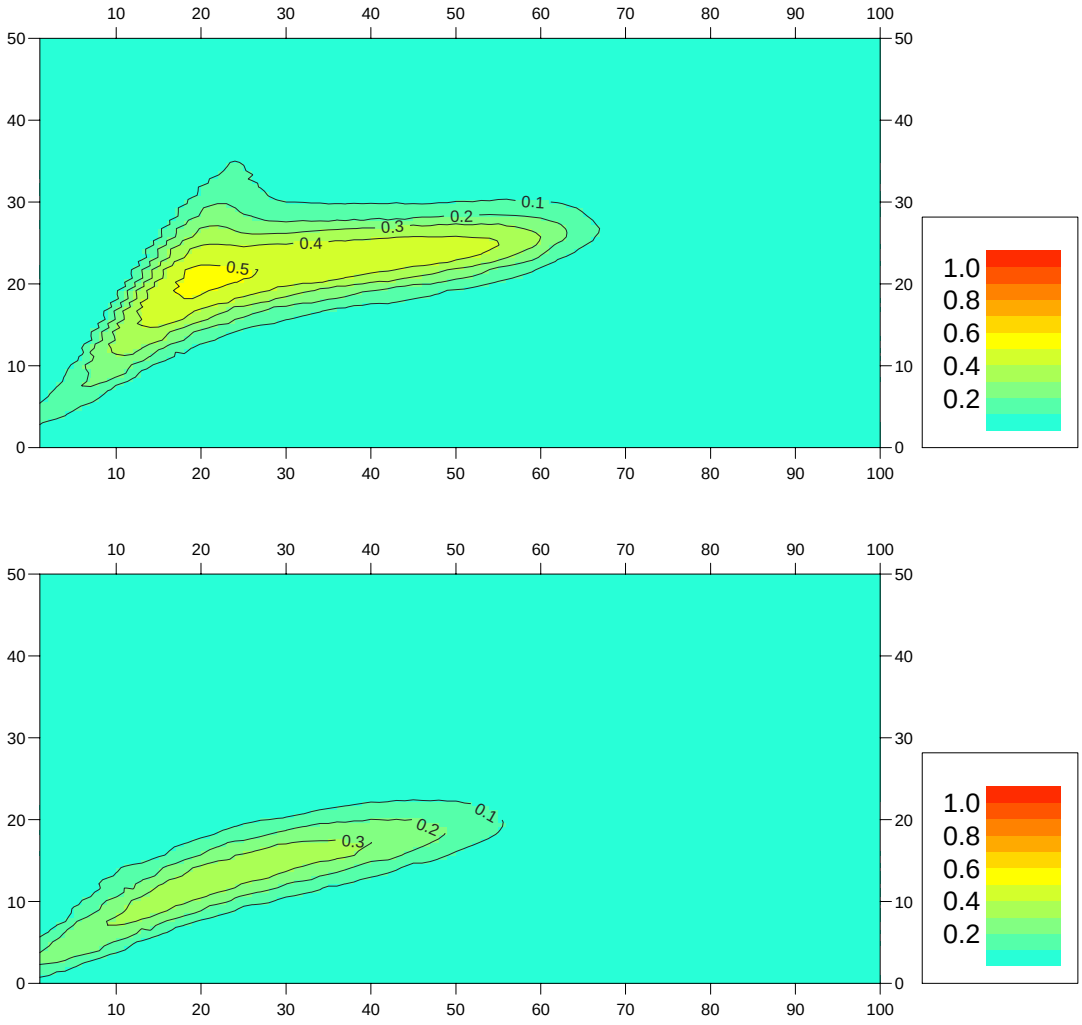

FIGURE 4.5: *Two dimensional test case: concentration contours at 90.2 days (top) and 120 days (bottom).*

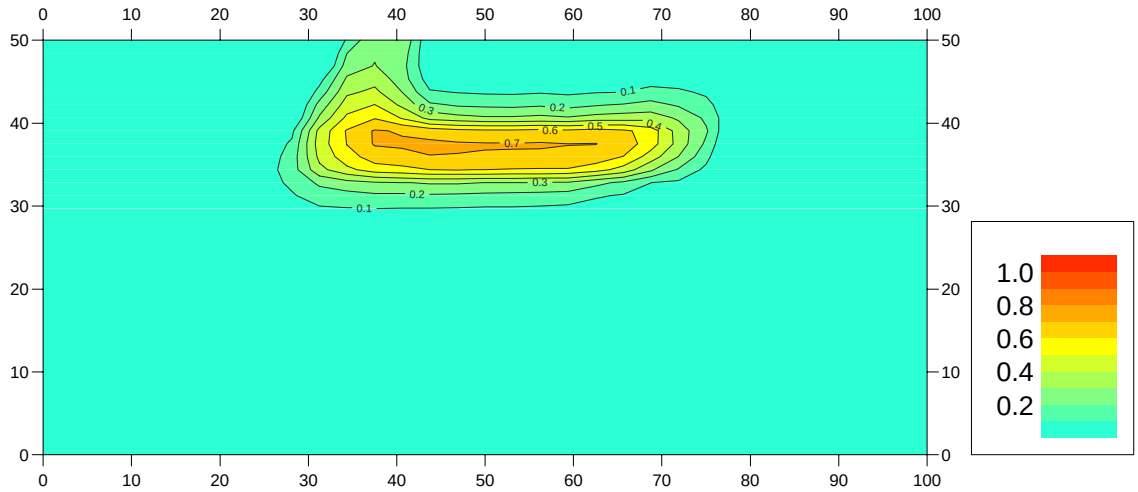

FIGURE 4.6: *Two dimensional test-case: concentration contours at 45.2 days as obtained by standard Galerkin FEM scheme.*

simulations. The plume shape in Figures 4.4 and 4.5 shows a front that is slightly steeper than the corresponding front calculated by standard Galerkin Finite Elements (GFE) with no upwind, as can be seen from Figure 4.6 that shows the solution to the same problem at 45.2 days as obtained by GFE [23]. This confirms the fact that the combination of MHFE and FV introduces less numerical diffusion than standard FE. Finally, the mass balance in all simulations is satisfied within the accuracy of the linear system solution in the dispersive phase (i.e. 10^{-10}), in contrasts with FE applications where mass balance errors of few percents are commonly observed. These characteristic features of the time-splitting algorithm developed here do not change at larger Peclet numbers as long as the stability criteria for the advection step are satisfied.

4.3 Extension of the time-splitting technique to second order accuracy in time

As seen in Section 4.2.3, the time-splitting algorithm introduces an error of the order of Δt in the overall approach. It is obvious then that even if we set $\theta = 0.5$ in equation (4.3), i.e. we use the second order time discretizations (Crank-Nicolson + Midpoint), global second order accuracy in time is not achieved.

A careful analysis of the truncation error shows that the $\mathcal{O}(\Delta t)$ error is proportional to the dispersion and advection discrete fluxes. A correction term can be calculated to recover second order accuracy in time.

4.3.1 The second order algorithm

Let $x = (x_1, x_2)$ denote the coordinates of a point in the domain Ω with convex boundary $\bar{\Omega}$ in $\mathbb{R}^2$. The advection-dispersion equation (4.1) can be rewritten as (for simplicity we

assume $\phi = 1$ and omit the vector notation),

$$\begin{aligned} \frac{\partial c}{\partial t} + \nabla \cdot (F + G) &= f \qquad \text{on } \Omega \times (0, T] && (4.14a) \\ F &= vc && (4.14b) \\ G &= -D\nabla c && (4.14c) \end{aligned}$$

where F and G are the advective and dispersive fluxes, respectively. The domain Ω is again discretized into m triangles T_l, $l = 1, \ldots, m$ with diameters $\leq h$.

Recall that we have called the approximations to c and G by $\tilde{c}$ and $\tilde{G}$, respectively, as defined in equation (3.20) and (3.21).

Given $\tilde{c}^k \in \Psi_h$, application of the operator splitting algorithm is summarized as follow: solve

$$\frac{\partial \hat{c}}{\partial t} + \nabla \cdot F(x, t, \hat{c}) = 0 \qquad \text{on } \Omega \times (t^k, t^{k+1}], \tag{4.15}$$

with initial condition $\tilde{c}^k$. We denote the solution generated at this step by $\hat{c}^{k+1}$. Next solve

$$\frac{\partial c^*}{\partial t} + \nabla \cdot G(x, t, c^*) = 0 \qquad \text{on } \Omega \times [t^k, t^{k+1}], \tag{4.16}$$

with initial condition $\hat{c}^{k+1}$. The solution generated here will approximate c^{k+1}.

For the discretization of the equations we start from equation (4.3) with $\theta = 0.5$ and $\Phi_l = 1$. The advection flux equation is approximated by the Finite Volume scheme on unstructured grids reported in Chapter 2. For this purpose we rewrite equation (4.6) as:

$$\hat{c}_l^{k+1} = \tilde{c}_l^k - \Delta t E^{k+1/2}(x, \tilde{c}^{k+1/2}), \tag{4.17}$$

where $E^{k+1/2}(x, \tilde{c}^{k+1/2}) \in W_h$ is the approximation to F at time $t^{k+1/2} = \dfrac{t^k + t^{k+1}}{2}$, i.e.:

$$\begin{aligned} &E(x, \tilde{c}^{k+1/2}) = E^{k+1/2} = \\ &= \frac{1}{|T_l|} \sum_{j=1}^{3} H_j(L_l(x^{\text{in}}, \tilde{c}^{k+1/2}), L_l(x^{\text{out}}, \tilde{c}^{k+1/2}))|e_j| \approx \frac{1}{\Delta t} \int_{t^k}^{t^{k+1}} \int_{\partial T_l} F(x, t, \hat{c}) \cdot n. \end{aligned}$$

The crucial point for obtaining second order accuracy when this step is joined up with the dispersion step is in the construction of $E^{k+1/2}$, as we will see later.

Next, we apply the discontinuous RT0 Mixed Hybrid Finite Element method together with the Crank-Nicolson scheme in time. Equations (4.8a-4.8b) can be written in this case as:

$$\begin{aligned} ((D^{-1}\tilde{G})^{k+1/2}, w) - (\tilde{c}^{k+1/2}, \nabla \cdot w) &= -(\lambda^{k+1/2}, w \cdot n)_{\partial\Omega}, \quad w \in W_h && (4.18a) \\ (\partial_t \tilde{c}^{k+1}, \psi) + (\nabla \cdot \tilde{G}^{k+1/2}, \psi) &= -(\nabla \cdot E^{k+1/2}, \psi), \quad \psi \in \Psi_h, && (4.18b) \end{aligned}$$

where λ is the Lagrange multiplier function, while $\partial_t \tilde{c}^{k+1}$ represents the backward differencing in time

$$\partial_t \tilde{c}^{k+1} = \frac{\tilde{c}^{k+1} - \tilde{c}^k}{\Delta t}.$$

At this point system (4.18) can be solved, by considering that the unknowns are now $\tilde{c}^{k+1}$, $\tilde{G}^{k+1}$ and λ^{k+1}. Note that the solution pair (c, G) satisfies

$$\begin{aligned}((D^{-1}G)^{k+1/2}, w) - (c^{k+1/2}, \nabla \cdot w) &= -(\lambda^{k+1/2}, w \cdot n)_{\partial\Omega}, \quad w \in W_h \qquad (4.19a)\\ (\frac{\partial c^{k+1}}{\partial t}, \psi) + (\nabla \cdot G^{k+1/2}, \psi) &= -(\nabla \cdot F^{k+1/2}, \psi), \quad \psi \in \Psi_h. \qquad (4.19b)\end{aligned}$$

In order to derive an error estimate for our scheme we will compare our approximate solution with the approximate solution of the elliptic projection of (4.18). Let $\bar{c}(\cdot, t) \in \Psi_h$, $\bar{G}(\cdot, t) \in W_h$ denote the mixed hybrid method solutions to the elliptic problem associated with (4.14); that is, for each $t \in [0, T]$,

$$\begin{aligned}((D^{-1}\bar{G}(\cdot, t), w) - (\bar{c}(\cdot, t), \nabla \cdot w) &= -(\lambda(\cdot, t), w \cdot n)_{\partial\Omega}, \quad w \in W_h \qquad (4.20a)\\ (\nabla \cdot \bar{G}(\cdot, t), \psi) &= (\nabla \cdot G(\cdot, t), \psi)\\ &= -(\frac{\partial c(\cdot, t)}{\partial t}, \psi) - (\nabla \cdot F(\cdot, t), \psi), \qquad (4.20b)\\ &\qquad \psi \in \Psi_h. \qquad (4.20c)\end{aligned}$$

For the following it is useful to introduce $\pi_2 c(\cdot, t)$ defined as the L_2-projection of $c(\cdot, t)$ into Ψ_h. It has been shown that [38, 17]

$$\|\bar{c} - \pi_2 c\| \leq \mathcal{C} h^2,$$

where $\mathcal{C}$ is a generic constant, independent of h and Δt.

Let $\tilde{F} \in W_h$ be such that $\tilde{F} \cdot n$ at the midpoint of a edge of any triangle T_l is equal to the integral average of $F \cdot n$ over that edge; then $\tilde{F}$ satisfies

$$(\nabla \cdot (\tilde{F} - F)(\cdot, t), \psi) = 0, \quad \psi \in \Psi_h. \qquad (4.21)$$

Let

$$\xi = \tilde{c} - \bar{c}, \quad \eta = \tilde{G} - \bar{G}, \quad \beta = \pi_2 c - \bar{c}. \qquad (4.22)$$

For c and the data sufficiently smooth, we assume the numerical flux $E^{k+1/2}(x, \tilde{c}^{k+1/2})$ satisfies the following approximation result:

$$\|E^{k+1/2}(\cdot, \tilde{c}^{k+1/2}) - \tilde{F}^{k+1/2}\| \leq \mathcal{C}(\|\xi^k\| + \|\xi^{k-1}\| + \|\beta^k\| + \|\beta^{k-1}\| + h^2 + \Delta t^2). \qquad (4.23)$$

We will directly verify (4.23) for the Finite Volume scheme described in paragraph 4.3.2.

Then, we can state the following error estimate for the method outlined above, showing second order accuracy in space and time.

THEOREM 4.2. *Let the data and the solution pair (c, G) be sufficiently smooth and assume (4.23) holds. Let ξ, η, β defined as in (4.22). Assume $0 < D^{-1} < d^*$, where d^* is a positive constant. Then*

$$\begin{aligned}&\max_k \|\xi^{k+1}\| + (\sum_{k=0}^{N-1} \|(D^{-1/2}\eta)^{k+1/2}\|^2 \Delta t)^{\frac{1}{2}} \leq\\ &\leq \mathcal{C}(d^*)[h^2 + \Delta t^2 + (\sum_{k=0}^{N} \|\beta^k\|^2 \Delta t)^{\frac{1}{2}} + \qquad (4.24)\\ &+ (\sum_{k=1}^{N-1} \|\beta^{k-1}\|^2 \Delta t)^{\frac{1}{2}} + (\int_0^T \|\frac{\partial \beta(\cdot, t)}{\partial t}\|^2 dt)^{\frac{1}{2}}],\end{aligned}$$

where $\mathcal{C}(d^*)$ is a constant dependent only on d^*.

Proof. Subtracting (4.20a) from (4.18a) we have

$$((D^{-1}\eta)^{k+1/2}, w) - (\xi^{k+1/2}, \nabla \cdot w) = 0, \quad w \in W_h. \tag{4.25}$$

Subtracting (4.20b) from (4.18b), using the L_2- projection of c and applying (4.21), we get

$$\begin{aligned}(\partial_t \xi^{k+1}, \psi) + (\nabla \cdot \eta^{k+1/2}, \psi) = \quad & -(\partial_t \bar{c}^{k+1}, \psi) + (\frac{\partial c^{k+1/2}}{\partial t}, \psi) + \\ & +(\nabla \cdot (F^{k+1/2}, \psi) - (\nabla \cdot E^{k+1/2}, \psi) = \\ & = (\partial_t \beta^{k+1}, \psi) + (\frac{\partial c^{k+1/2}}{\partial t} - \partial_t c^{k+1}, \psi) + \\ & +(\nabla \cdot (\tilde{F}^{k+1/2} - E^{k+1/2}), \psi), \\ & \psi \in \Psi_h,\end{aligned} \tag{4.26}$$

Setting $w = \eta^{k+1/2}$ and $\psi = \xi^{k+1/2}$ and adding (4.25) and (4.26), we obtain

$$\begin{aligned}(\partial_t \xi^{k+1}, \xi^{k+1/2}) + (D^{-1}\eta^{k+1/2}, \eta^{k+1/2}) = \quad & (\partial_t \beta^{k+1}, \xi^{k+1/2}) + \\ & +(\frac{\partial c^{k+1/2}}{\partial t} - \partial_t c^{k+1}, \xi^{k+1/2}) + \\ & +(\nabla \cdot (\tilde{F}^{k+1/2} - E^{k+1/2}), \xi^{k+1/2}).\end{aligned} \tag{4.27}$$

An upper bound to the first two terms of (4.27) is given by

$$\begin{aligned}& (\partial_t \beta^{k+1} + \frac{\partial c^{k+1/2}}{\partial t} - \partial_t c^{k+1}, \xi^{k+1/2}) \leq \\ & \leq \frac{1}{2}\|\partial_t \beta^{k+1} + \frac{\partial c^{k+1/2}}{\partial t} - \partial_t c^{k+1}\|^2 + \frac{\|\xi^{k+1/2})\|^2}{2} \leq \\ & \leq \|\partial_t \beta^{k+1}\|^2 + \|\frac{\partial c^{k+1/2}}{\partial t} - \partial_t c^{k+1}\|^2 + \|\frac{\xi^{k+1} + \xi^k}{2}\|^2 \leq \\ & \leq \mathcal{C}_1(\frac{1}{\Delta t}\int_{t^k}^{t^{k+1}} \|\frac{\partial \beta}{\partial t}\|^2 dt + \Delta t^4) + \mathcal{C}_2(\|\xi^k\|^2 + \|\xi^{k+1}\|^2).\end{aligned} \tag{4.28}$$

In fact

$$\frac{\partial c^{k+1/2}}{\partial t} - \partial_t c^{k+1} = \frac{\partial c^{k+1/2}}{\partial t} - \frac{c^{k+1} - c^k}{\Delta t} = \mathcal{O}(\frac{\Delta t^2}{24}),$$

while $\partial_t \beta^{k+1} = \dfrac{1}{\Delta t}\displaystyle\int_{t^k}^{t^{k+1}} \frac{\partial \beta}{\partial t}$, therefore

$$\begin{aligned}|\partial_t \beta^{k+1}| \leq \quad & \frac{1}{\Delta t}\int_{t^k}^{t^{k+1}} |\frac{\partial \beta}{\partial t}| \leq \\ & \leq \frac{1}{\Delta t}(\int_{t^k}^{t^{k+1}} 1)^{1/2}(\int_{t^k}^{t^{k+1}} |\frac{\partial \beta}{\partial t}|^2)^{1/2} = \\ & = \frac{1}{(\Delta t)^{1/2}}(\int_{t^k}^{t^{k+1}} |\frac{\partial \beta}{\partial t}|^2)^{1/2}\end{aligned}$$

so that

$$\|\partial_t \beta^{k+1}\|^2 = \int_\Omega |\partial_t \beta^{k+1}|^2 \leq \frac{1}{(\Delta t)}(\int_{t^k}^{t^{k+1}} \|\frac{\partial \beta}{\partial t}\|^2),$$

where $\|\frac{\partial\beta}{\partial t}\| \leq \mathcal{C}h$ and for c and the data sufficiently smooth $\|\frac{\partial\beta}{\partial t}\| \leq \mathcal{C}h^2$ [61, 11].
To bound the last term of equation (4.27), we set $w = \tilde{F}^{k+1/2} - E^{k+1/2}$ in (4.25), obtaining

$$
\begin{aligned}
&(\nabla\cdot(\tilde{F}^{k+1/2} - E^{k+1/2}), \xi^{k+1/2}) = ((D^{-1}\eta)^{k+1/2}, \tilde{F}^{k+1/2} - E^{k+1/2}) \leq \\
&\leq \frac{1}{2}\|(D^{-1/2}\eta)^{k+1/2}\|^2 + \frac{d^*}{2}\|\tilde{F}^{k+1/2} - E^{k+1/2}\|^2 \leq \\
&\leq \frac{1}{2}\|(D^{-1/2}\eta)^{k+1/2}\|^2 + \mathcal{C}_3(d^*)(\|\xi^k\|^2 + \|\xi^{k-1}\|^2 + \|\beta^k\|^2 + \|\beta^{k-1}\|^2 + h^4 + \Delta t^4).
\end{aligned}
\tag{4.29}
$$

In the last inequality we have applied (4.23).
Now, substituting (4.29) and (4.28) into (4.27) and using

$$((D^{-1}\eta)^{k+1/2}, \eta^{k+1/2}) = \|(D^{-1/2}\eta)^{k+1/2}\|^2$$

we obtain

$$
\begin{aligned}
&(\partial_t\xi^{k+1}, \xi^{k+1/2}) + \frac{1}{2}\|(D^{-1/2}\eta)^{k+1/2}\|^2 \leq \\
&\leq \mathcal{C}_1\frac{1}{\Delta t}\int_{t^k}^{t^{k+1}}\|\frac{\partial\beta}{\partial t}\|^2 dt + \mathcal{C}_2(\|\xi^k\|^2 + \|\xi^{k+1}\|^2) + \\
&+\mathcal{C}_3(d^*)(\|\xi^k\|^2 + \|\xi^{k-1}\|^2 + \|\beta^k\|^2 + \|\beta^{k-1}\|^2 + h^4) + \\
&+(\mathcal{C}_3(d^*) + \mathcal{C}_1)\Delta t^4.
\end{aligned}
\tag{4.30}
$$

Considering that

$$(\partial_t\xi^{k+1}, \xi^{k+1/2}) = \frac{1}{\Delta t}(\|\xi^{k+1}\|^2 - \|\xi^k\|^2)$$

multiplying (4.30) by $2\Delta t$ and using, for simplicity, a unique constant $\mathcal{C}(d^*)$, we get

$$
\begin{aligned}
&\|\xi^{k+1}\|^2 - \|\xi^k\|^2 + \Delta t\|(D^{-1/2}\eta)^{k+1/2}\|^2 \leq \\
&\leq 2\mathcal{C}(d^*)\int_{t^k}^{t^{k+1}}\|\frac{\partial\beta}{\partial t}\|^2 dt + \mathcal{C}(d^*)(\|\xi^{k-1}\|^2 + \|\xi^k\|^2 + \|\xi^{k+1}\|^2)2\Delta t + \\
&+\mathcal{C}(d^*)(\|\beta^k\|^2 + \|\beta^{k-1}\|^2)2\Delta t + \mathcal{C}(d^*)(h^4 + \Delta t^4)2\Delta t.
\end{aligned}
$$

Summing on k, we obtain:

$$
\begin{aligned}
&\|\xi^N\|^2 - \|\xi^0\|^2 + 2\mathcal{C}(d^*)\sum_{k=0}^{N-1}\|(D^{-1/2}\eta)^{k+1/2}\|^2\Delta t \leq \\
&\leq 2\mathcal{C}(d^*)\int_0^T\|\frac{\partial\beta}{\partial t}\|^2 dt + 2\mathcal{C}(d^*)\sum_{k=0}^{N-1}(\|\beta^k\|^2 + \|\beta^{k-1}\|^2)\Delta t + \mathcal{C}(d^*)(h^4 + \Delta t^4)\Delta t + \\
&+2\mathcal{C}(d^*)\sum_{k=0}^{N-1}(\|\xi^{k-1}\|^2 + \|\xi^k\|^2 + \|\xi^{k+1}\|^2)\Delta t.
\end{aligned}
$$

Isolating on the left hand side the term $\|\xi^N\|^2$, the previus expression can be transformed as

$$\|\xi^N\|^2 \leq \gamma^N + \mathcal{C}(d^*)\Delta t\sum_{k=0}^{N-1}\|\xi^k\|^2,$$

where

$$\gamma^N = -2\mathcal{C}(d^*)\sum_{k=0}^{N-1} \|(D^{-1/2}\eta)^{k+1/2}\|^2\Delta t + 2\mathcal{C}(d^*)\int_0^T \|\frac{\partial\beta}{\partial t}\|^2 dt +$$
$$+2\mathcal{C}(d^*)\sum_{k=0}^{N-1}(\|\beta^k\|^2 + \|\beta^{k-1}\|^2)\Delta t + \mathcal{C}(d^*)(h^4 + \Delta t^4)\Delta t.$$

Therefore we can apply the discrete Gronwall's Lemma, and we obtain (4.24). Thus, the proof of the theorem is completed. □

About Gronwall's Lemma. We recall that the discrete Gronwall's Lemma can be stated in the following way:

LEMMA 4.3 (DISCRETE GRONWALL'S LEMMA). *Let f, g be non negative functions defined on the grid $\{x_j = jh\}_{j=0}^J$, with g being non decreasing ($g(t) \leq g(v)$ if $v \geq t$).*
If

$$f(x_j) \leq g(x_j) + Ch\sum_{k=0}^{j-1} f(x_k) \qquad 0 \leq j \leq J$$

then

$$f(x_j) \leq g(x_j)e^{Cx_j} \qquad 0 \leq j \leq J.$$

4.3.2 Numerical flux approximation

As we have shown in Chapters 2 and 3, piecewise linear reconstruction together with a limitating procedure assures second order accuracy in space in the FV fornulation, while the MHFE scheme achieves superconvergence (quadratic) at the centroids of the triangles. Therefore second order in space is obtained. Theorem 4.2 assures second order accuracy also in time when the midpoint rule for the FV scheme and the Crank-Nicolson procedure for the MHFE method are used as time discretizations, only if condition (4.23) is satisfied by FV. In the following we will prove that this inequality holds only if a correction term is added to the time stepping procedure of the advection equation.
The usual midpoint (or Runge) scheme applied to the FV algorithm can be written as

$$\tilde{c}^1 = \tilde{c}^k - \frac{\Delta t}{2}E(x, \tilde{c}^k)$$
$$\tilde{c}^{k+1} = \tilde{c}^k - \Delta t E(x, \tilde{c}^1)$$

This technique is derived by means of a Taylor series expansion of the type:

$$\tilde{c}^{k+1/2} = \tilde{c}^k + \frac{\Delta t}{2}\tilde{c}_t^k + \mathcal{O}(\Delta t^2). \tag{4.31}$$

This fact has to be taken into account in the FV linear interpolation of the reconstruction phase. The linear interpolant $L_l(x)$ over T_l at time t^k is given by the plane interpolating the values of concentration of the centroids of the three neighboring triangles, e.g., $\tilde{c}_a^k$ at the centroid $x_a = (x_{1a}, x_{2a})$, $\tilde{c}_b^k$ at x_b, $\tilde{c}_c^k$ at x_c. Its expression can be written as

$$L_l(x, \tilde{c}^k) = \tilde{c}_a^k + L_{lx_1}(\tilde{c}^k)(x_1 - x_{1a}) + L_{lx_2}(\tilde{c}^k)(x_2 - x_{2a}).$$

where the coefficients $L_{lx_1}(\tilde{c}^k)$ and $L_{lx_2}(\tilde{c}^k)$ take on the form:

$$L_{lx_1}(\tilde{c}^k) = \frac{(\tilde{c}_b^k - \tilde{c}_a^k)(x_{2c} - x_{2a}) + (\tilde{c}_c^k - \tilde{c}_a^k)(x_{2a} - x_{2b})}{(x_{1b} - x_{1a})(x_{2c} - x_{2a}) + (x_{1c} - x_{1a})(x_{2a} - x_{2b})} \tag{4.32}$$

$$L_{lx_2}(\tilde{c}^k) = -\frac{(\tilde{c}_c^k - \tilde{c}_a^k)(x_{1b} - x_{1a}) + (\tilde{c}_a^k - \tilde{c}_b^k)(x_{1c} - x_{1a})}{(x_{1b} - x_{1a})(x_{2c} - x_{2a}) + (x_{1c} - x_{1a})(x_{2a} - x_{2b})}. \tag{4.33}$$

We can prove the following lemma that will be used in the proof of (4.23):

LEMMA 4.4. *The linear interpolant constructed for the proposed FV scheme satisfies the inequalities:*

$$h|L_{lx_1}(\tilde{c}^k)| \leq \sum_{i \in \mathcal{T}} |\tilde{c}_i^k| \quad \textit{and } h|L_{lx_2}(\tilde{c}^k)| \leq \sum_{i \in \mathcal{T}} |\tilde{c}_i^k|,$$

where $\mathcal{T}$ is the set of indeces corresponding to triangle T_l and its neighbors.

Proof. Since h is the diameter of the triangulation and the distance between two neighboring centroids is greater than h but smaller than $2h$, from equation (4.32) we have:

$$\begin{aligned} &|(\tilde{c}_b^k - \tilde{c}_a^k)(x_{2c} - x_{2a}) + (\tilde{c}_c^k - \tilde{c}_a^k)(x_{2a} - x_{2b})| \leq \\ &\leq |\tilde{c}_b^k - \tilde{c}_a^k|2h + |\tilde{c}_c^k - \tilde{c}_a^k|2h \leq 2h \sum_{i \in \mathcal{T}} |\tilde{c}_i^k| \end{aligned}$$

and

$$|(x_{1b} - x_{1a})(x_{2c} - x_{2a}) + (x_{1c} - x_{1a})(x_{2a} - x_{2b})| \geq 2h^2.$$

Therefore,

$$|L_{lx_1}(\tilde{c}^k)| \leq \frac{2h \sum_{i \in \mathcal{T}} |\tilde{c}_i^k|}{2h^2}$$

and

$$h|L_{lx_1}(\tilde{c}^k)| \leq \sum_{i \in \mathcal{T}} |\tilde{c}_i^k|.$$

The same can be done for $L_{lx_2}(\tilde{c}^k)$ and the lemma is proved. ∎

In the first step of the Runge scheme we need to calculate $L_l(x, \tilde{c}^{k+1/2})$. Since the initial conditions $\tilde{c}^k$ come from the MHFE solution (4.18) at the previous time step, we can use (4.31) to evaluate $\tilde{c}^{k+1/2}$. Thus we need to evaluate $\tilde{c}_t^k$. This can be accomplished in two ways. One way consists in evaluating the spatial discretization of the dispersion and advection terms at t^k, since $c_t^k = [-\nabla \cdot (F + G) + f]^k$. Alternatively, we can simply evaluate $\tilde{c}_t^k$ by backward finite difference, and obtain:

$$\tilde{c}^{k+1/2} = \frac{3}{2}\tilde{c}^k - \frac{1}{2}\tilde{c}^{k-1} + \mathcal{O}(\Delta t^2).$$

Thus, the linear reconstruction at time $t^{k+1/2}$ takes on the following form:

$$L_l(x, \tilde{c}^{k+1/2}) = \frac{3}{2}\tilde{c}_a^k - \frac{1}{2}\tilde{c}_a^{k-1} + L_{lx_1}(\tilde{c}^{k+1/2})(x_1 - x_{1a}) + L_{lx_2}(\tilde{c}^{k+1/2})(x_2 - x_{2_a}) \tag{4.34}$$

where

$$\begin{aligned} &L_{lx_1}(\tilde{c}^{k+1/2}) = \\ &= \frac{\left[\frac{3}{2}\tilde{c}_b^k - \frac{1}{2}\tilde{c}_b^{k-1} - (\frac{3}{2}\tilde{c}_a^k - \frac{1}{2}\tilde{c}_a^{k-1})\right](x_{2c} - x_{2a}) + \left[\frac{3}{2}\tilde{c}_c^k - \frac{1}{2}\tilde{c}_c^{k-1} - (\frac{3}{2}\tilde{c}_a^k - \frac{1}{2}\tilde{c}_a^{k-1})\right](x_{2a} - x_{2b})}{(x_{1b} - x_{1a})(x_{2c} - x_{2a}) + (x_{1c} - x_{1a})(x_{2a} - x_{2b})} \end{aligned}$$

$L_{lx_2}(\tilde{c}^{k+1/2})$ is analogously defined. A similar result as in Lemma 4.4 can be proved for $L_{lx_1}(\tilde{c}^{k+1/2})$ and $L_{lx_2}(\tilde{c}^{k+1/2})$. In fact it is straightforward to prove the following inequalities:

$$h|L_{lx_1}(\tilde{c}^{k+1/2})| \leq \frac{3}{2}\sum_{i\in\mathcal{T}}|\tilde{c}_i^k| + \frac{1}{2}\sum_{i\in\mathcal{T}}|\tilde{c}_i^{k-1}| \leq \mathcal{C}\sum_{i\in\mathcal{T}}(|\tilde{c}_i^k| + |\tilde{c}_i^{k-1}|) \tag{4.35}$$

and

$$h|L_{lx_2}(\tilde{c}^{k+1/2})| \leq \mathcal{C}\sum_{i\in\mathcal{T}}(|\tilde{c}_i^k| + |\tilde{c}_i^{k-1}|) \tag{4.36}$$

Finally, the numerical flux $E^{k+1/2}$ is given by

$$E^{k+1/2} = \frac{1}{|T_l|}\sum_{j=1}^{3} H_j(L_l(x^{\text{in}}, \tilde{c}^{k+1/2}), L_l(x^{\text{out}}, \tilde{c}^{k+1/2})) \tag{4.37}$$

where H_j is the two-point Lipschitz monotone flux defined in Chapter 2.

In order to verify (4.23), we similarly construct $\bar{c}_{x_{lj}}^{\text{in},k+1/2}$ and $\bar{c}_{x_{lj}}^{\text{out},k+1/2}$ from $\bar{c}$ given in (4.20a), and $\pi_2 c_{x_{lj}}^{\text{in},k+1/2}$, $\pi_2 c_{x_{lj}}^{\text{out},k+1/2}$, $c_{x_{lj}}^{\text{in},k+1/2}$ and $c_{x_{lj}}^{\text{out},k+1/2}$ using $\pi_2 c$ and c, respectively. Thus, we can state the following lemma.

LEMMA 4.5. *Relation (4.23), i.e.*

$$\|E^{k+1/2}(\cdot, \tilde{c}^{k+1/2}) - \tilde{F}^{k+1/2}\| \leq \mathcal{C}(\|\xi^k\| + \|\xi^{k-1}\| + \|\beta^k\| + \|\beta^{k-1}\| + h^2 + \Delta t^2)$$

holds when in the FV scheme the linear interpolant is given by (4.34).

Proof. Consider that

$$\begin{aligned}
&H_j^{k+1/2} - \tilde{F}_{lj}^{k+1/2} = \\
&H_j(L_l(x^{\text{in}}, \tilde{c}^{k+1/2}), L_l(x^{\text{out}}, \tilde{c}^{k+1/2})) - \frac{1}{e_j}\int_{e_j} F(c(x, t^{k+1/2}))\cdot n_j = \\
&[H_j(L_l(x^{\text{in}}, \tilde{c}^{k+1/2}), L_l(x^{\text{out}}, \tilde{c}^{k+1/2})) - H_j(L_l(x^{\text{in}}, \bar{c}^{k+1/2}), L_l(x^{\text{out}}, \bar{c}^{k+1/2}))] + \\
&+[H_j(L_l(x^{\text{in}}, \bar{c}^{k+1/2}), L_l(x^{\text{out}}, \bar{c}^{k+1/2})) - H_j(L_l(x^{\text{in}}, \pi_2 c^{k+1/2}), L_l(x^{\text{out}}, \pi_2 c^{k+1/2}))] + \\
&+[H_j(L_l(x^{\text{in}}, \pi_2 c^{k+1/2}), L_l(x^{\text{out}}, \pi_2 c^{k+1/2})) - H_j(L_l(x^{\text{in}}, c^{k+1/2}), L_l(x^{\text{out}}, c^{k+1/2}))] + \\
&+[H_j(L_l(x^{\text{in}}, c^{k+1/2}), L_l(x^{\text{out}}, c^{k+1/2})) - F(c_j^{k+1/2})\cdot n] + \\
&+[F(c_j^{k+1/2})\cdot n - \frac{1}{e_j}\int_{e_j} F(c(x, t^{k+1/2}))\cdot n_j] = \\
&R_1 + \ldots + R_5.
\end{aligned} \tag{4.38}$$

We now examine the terms R_1 through R_5.
By Lipschitz continuity of H_j and (4.35),

$$\begin{aligned}
&|R_1| \leq \mathcal{C}(|\tilde{c}_{x_{lj}}^{\text{in},k+1/2} - \bar{c}_{x_{lj}}^{\text{in},k+1/2}| + |\tilde{c}_{x_{lj}}^{\text{out},k+1/2} - \bar{c}_{x_{lj}}^{\text{out},k+1/2}|) \leq \\
&\mathcal{C}\sum_{i\in\mathcal{T}}(|\tilde{c}_{x_i}^k - \bar{c}_{x_i}^k| + |\tilde{c}_{x_i}^{k-1} - \bar{c}_{x_i}^{k-1}|) = \mathcal{C}\sum_{i\in\mathcal{T}}(|\xi_i^k| + |\xi^{k-1}|)
\end{aligned} \tag{4.39}$$

Applying similar arguments to R_2 and R_3,

$$\begin{aligned}
&|R_2| \leq \mathcal{C}\sum_{i\in\mathcal{T}}(|\bar{c}_{x_i}^k - \pi_2 c_{x_i}^k| + |\bar{c}_{x_i}^{k-1} - \pi_2 c_{x_i}^{k-1}|) = \\
&= \mathcal{C}\sum_{i\in\mathcal{T}}(|\beta_i^k| + |\beta_i^{k-1}|)
\end{aligned} \tag{4.40}$$

and, using the fact that $|\pi_2 c_i - c_i| = \mathcal{O}(h^2)$,

$$|R_3| \leq \mathcal{C} \sum_{i \in \mathcal{T}} (|\pi_2 c_{x_i}^k - c_{x_i}^k| + |\pi_2 c_{x_i}^{k-1} - c_{x_i}^{k-1}|) \leq \mathcal{C} h^2. \tag{4.41}$$

By observing that the linear reconstruction is of second order accuracy in space and time and by (4.35)

$$\begin{aligned} &|R_4| \leq \mathcal{C}(|c_{x_i}^{\text{in},k+1/2} - c_{x_i}^{k+1/2}| + |\pi_2 c_{x_i}^{\text{out},k+1/2} - c_{x_i}^{k+1/2}|) \leq \\ &\mathcal{C}(\Delta t^2 + h^2). \end{aligned} \tag{4.42}$$

Finally, by considering that we have used the midpoint rule to integrate in space and time, it is easy to see that

$$|R_5| \leq \mathcal{C}(\Delta t^2 + h^2) \tag{4.43}$$

Multiplying (4.39)-(4.43) by h^2, summing on all over the edges of the triangulation and using equivalence of norms,

$$\begin{aligned} &\|E^{k+1/2} - \tilde{F}^{k+1/2}\| \leq \mathcal{C}^* \sum_{l=0}^{m} \sum_{j=1}^{3} |H_{lj}^{k+1/2} - \tilde{F}_{lj}^{k+1/2}|^2 h^2 \leq \\ &\mathcal{C}(\mathcal{C}^*)(\|\xi^k\| + \|\xi^{k-1}\| + \|\beta^k\| + \|\beta^{k-1}\| + h^2 + \Delta t^2), \end{aligned}$$

where C^* is an equivalence constant between the continuous and discrete L_2 norms. In this way, we have proved that (4.23) holds. □

4.3.3 Numerical results about second order accuracy

The numerical convergence rate of the fully second order accurate in time scheme is tested on the same one-dimensional model problem studied in Section 4.2.3, solved in a two-dimensional grid system. The numerical convergence of the scheme is evaluated by calculating L_1 and L_2 errors at different grid levels ($|e_{\ell,1}|$, $|e_{\ell,2}|$). Since the L_2 norm is bounded by the L_1 norm, we expect a smaller order of accuracy in the latter case. The rates of convergence are computed level after level and then also by applying a least squares approximation (l.s.a.).

For all the subsequent test runs the CFL number is set constant and equal to 0.28. Passing from one level to the next, the time step Δt is halved, and thus the Pe number decreases by a factor of 2 in passing from a coarser to a finer level.

The simulations are aimed at numerically verifying the theoretical convergence rate of the time-splitting technique scheme under different Pe numbers. We compare the results obtained by using a first order in time scheme, that is the Euler scheme (Eu), with those obtained by using a second order in time scheme. We consider three different time discretization techniques: the midpoint rule taking into account the correction term (MP+disp), the midpoint and Runge-Kutta schemes without considering the correction term (MP+no-disp and RK+no-disp respectively).

Table 4.7 and 4.8 report the errors and convergence rate at the different levels for the case in which dispersion is $D = 4 \times 10^{-3}$ m^2/s and velocity is $v = 0.5$ m/s. These values correspond to Peclet numbers varying from 11.5 ($\ell = 1$) to 1.44 ($\ell = 4$), that is dispersion is dominant. The time step Δt varies in the interval $[2 \times 10^{-3}, 2.5 \times 10^{-4}]$. In Table 4.7,

TABLE 4.7: $t^k = 0.1$ *s. Case of dominant dispersion.*

	Eu		MP+disp		MP+no-disp		RK+no-disp	
ℓ	$\|e_{\ell,1}\|$	rate	$\|e_{\ell,1}\|$	rate	$\|e_{\ell,1}\|$	rate	$\|e_{\ell,1}\|$	rate
1	2.69e-2		2.20e-2		2.15e-2		2.07e-2	
2	1.24e-2	1.12	9.26e-3	1.25	8.91e-3	1.27	8.75e-3	1.24
3	4.29e-3	1.53	2.45e-3	1.92	2.37e-3	1.91	2.31e-3	1.92
4	1.78e-3	1.27	6.34e-4	1.95	5.60e-4	2.08	5.61e-4	2.04
		l.s.a. rate		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.33		1.73		1.77		1.75
ℓ	$\|e_{\ell,2}\|$	rate	$\|e_{\ell,2}\|$	rate	$\|e_{\ell,2}\|$	rate	$\|e_{\ell,2}\|$	rate
1	2.68e-2		2.23e-2		2.20e-2		2.15e-2	
2	1.20e-2	1.16	8.83e-3	1.34	8.90e-3	1.30	8.72e-3	1.30
3	4.09e-3	1.55	2.38e-3	1.89	2.41e-3	1.88	2.32e-3	1.91
4	1.62e-3	1.34	6.44e-4	1.88	6.10e-4	1.98	5.81e-4	2.00
		l.s.a. rate		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.37		1.72		1.74		1.75

the results are relative to a final time $t^k = 0.1$ s, while in Table 4.8 t^k is equal to 1 s. As we can observe from Table 4.7, the errors computed in L_1 norm are slightly smaller than those computed in L_2 norm. The results obtained by using the Euler scheme in time show, as expected, first order accuracy. For all the other time integration schemes global second order convergence rates are achieved. The correction term does not influence the solution at this early time. However, at $t^k = 1$ s, we can observe that the schemes MP+no-disp and RK+no-disp are only first order accurate (but with errors that are smaller than the Eu scheme), while MP+disp displays a superlinear rate of convergence: its rate is equal to 1.61 passing from $\ell = 2$ to $\ell = 3$ against 1.31 for MP+no-disp and 1.32 for RK+no-disp, and it is 1.32 passing from $\ell = 3$ to $\ell = 4$ against 1.08 and 1.11 for the other two schemes, respectively.

Hence, as predicted by the theory developed in Section 4.3, the introduction of the corrective term in the FV scheme is crucial to improve accuracy when dispersion is not negligible. When advection is dominant, the truncation error due to dispersion is always small compared to that one due to the advection flux, and we therefore expect only little differences between the three second order time discretizations. We study this case being $D = 4 \times 10^{-5}$ m^2/s and $v = 0.5$ m/s. Correspondingly, the Peclet number varies between 1150 ($\ell = 1$) and 144 ($\ell = 4$). In Table 4.9 we have reported errors and convergence rates at the time $t^k = 1$ s. As we can see, the Eu scheme does not attains asymptotic convergence for these parameters. The error displays a minimum value at grid level $\ell = 2$ and further one refinement does not improve accuracy. The other schemes display more than first order accuracy. We observe about 1.34 convergence rate in L_1 norm and a little more than first order in L_2 norm. Here, we observe a different behavior between the two norms used, because of the advection domination. As expected, for MP+disp the errors are slightly smaller than for the other schemes, but the behavior is very similar.

TABLE 4.8: $t^k = 1$ *s. Case of dominant dispersion.*

	Eu		MP+disp		MP+no-disp		RK+no-disp	
ℓ	$\|e_{\ell,1}\|$	rate	$\|e_{\ell,1}\|$	rate	$\|e_{\ell,1}\|$	rate	$\|e_{\ell,1}\|$	rate
1	5.56e-3		2.84e-3		2.70e-3		2.63e-3	
2	2.80e-3	0.99	9.39e-4	1.60	9.44e-4	1.52	9.04e-4	1.54
3	1.36e-3	1.04	3.07e-4	1.61	3.80e-4	1.31	3.62e-4	1.32
4	6.85e-4	0.99	1.23e-4	1.32	1.80e-4	1.08	1.68e-4	1.11
		l.s.a. rate		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.01		1.52		1.30		1.32
ℓ	$\|e_{\ell,2}\|$	rate	$\|e_{\ell,2}\|$	rate	$\|e_{\ell,2}\|$	rate	$\|e_{\ell,2}\|$	rate
1	7.39e-3		4.18e-3		4.12e-3		4.07e-3	
2	3.47e-3	1.09	1.34e-3	1.64	1.40e-3	1.56	1.35e-3	1.59
3	1.63e-3	1.09	4.27e-4	1.65	5.15e-4	1.44	4.96e-4	1.44
4	8.20e-4	0.99	1.63e-4	1.39	2.29e-4	1.17	2.18e-4	1.19
		l.s.a. rate		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.06		1.57		1.39		1.41

TABLE 4.9: $t^k = 1$ *s. Case of dominant advection.*

	Eu		MP+disp		MP+no-disp		RK+no-disp	
ℓ	$\|e_{\ell,1}\|$	rate	$\|e_{\ell,1}\|$	rate	$\|e_{\ell,1}\|$	rate	$\|e_{\ell,1}\|$	rate
1	1.94e-2		2.88e-2		2.90e-2		2.88e-2	
2	5.48e-3	1.82	1.18e-2	1.29	1.19e-2	1.28	1.19e-2	1.27
3	4.55e-3	0.27	4.85e-3	1.28	4.83e-3	1.30	4.88e-3	1.29
4	5.28e-3	-	1.75e-3	1.47	1.78e-3	1.44	1.78e-3	1.45
		l.s.a. rate		l.s.a. rate		l.s.a. rate		l.s.a. rate
		0.59		1.34		1.34		1.33
ℓ	$\|e_{\ell,2}\|$	rate	$\|e_{\ell,2}\|$	rate	$\|e_{\ell,2}\|$	rate	$\|e_{\ell,2}\|$	rate
1	6.45e-2		7.88e-2		7.93e-2		7.92e-2	
2	2.01e-2	1.68	3.52e-2	1.16	3.60e-2	1.14	3.58e-2	1.14
3	2.02e-2	-	1.76e-2	1.00	1.77e-2	1.02	1.78e-2	1.01
4	2.24e-2	-	6.70e-3	1.39	6.83e-3	1.37	6.79e-3	1.39
		l.s.a. rate		l.s.a. rate		l.s.a. rate		l.s.a. rate
		0.46		1.17		1.16		1.16

5 Coupled flow and transport problem in groundwater

Many problems in subsurface hydrology can be studied by a mathematical model of density-dependent flow and transport coupled via the fluid density. For example, it is used to describe the direction of dissolved contaminant migration, to define the limits of a capture zone for a contamination recovery well or to delineate a water well protection area for a water supply. In this Chapter we present the mathematical model and study its numerical discretization in a two-dimensional case through Mixed Hybrid Finite Element scheme and a Time Splitting technique. Finally we consider some applications to Elder's problem [53] - a test case from the literature - and to the realistic case of the Lake Karachai.

5.1 The mathematical model

The mathematical model of density-dependent flow and transport in groundwater can be expressed in terms of an equivalent freshwater head h, defined as [1, 19]

$$h = \psi + z$$

where $\psi = p/(\rho_0 g)$ is the equivalent freshwater pressure head, p is the pressure, ρ_0 is the freshwater density, g is the gravitational constant, and z is the vertical coordinate directed upward. The density ρ of the saltwater solution is written in terms of the reference density ρ_0 and the normalized (actual divided by maximum) salt concentration c:

$$\rho = \rho_0(1 + \epsilon c) \tag{5.1}$$

where $\epsilon = (\rho_s - \rho_0)/\rho_0$ is the density ratio, tipically $\ll 1$, and ρ_s is the density of the solution at $c = 1$, and may represent, for instance, the density of seawater, or of the solution nearest a surface salt mound or around an underground saline diapir. The dynamic viscosity μ of the saltwater mixture is also expressed as a function of c and of the reference viscosity μ_0 as

$$\mu = \mu_0(1 + \epsilon' c) \tag{5.2}$$

where $\epsilon' = (\mu_s - \mu_0)/\mu_0$ is the viscosity ratio and μ_s is the viscosity of the solution at $c = 1$. With these definitions, the coupled system of variably saturated flow and miscible salt transport equations is [24]:

$$\begin{aligned}
\sigma \frac{\partial \psi}{\partial t} &= \vec{\nabla} \cdot [K_s \frac{1 + \epsilon c}{1 + \epsilon' c} K_r (\vec{\nabla}\psi + (1 + \epsilon c)\eta_z)] \\
&\quad - \phi S_w \epsilon \frac{\partial c}{\partial t} + \frac{\rho}{\rho_0} q^* + q && \text{(5.3a)} \\
\vec{v} &= -K_s \frac{1 + \epsilon c}{1 + \epsilon' c} K_r (\vec{\nabla}\psi + (1 + \epsilon c)\eta_z) && \text{(5.3b)} \\
\phi \frac{\partial S_w c}{\partial t} &= \vec{\nabla} \cdot (D\vec{\nabla} c) - \vec{\nabla} \cdot (c\vec{v}) + q c^* + f && \text{(5.3c)}
\end{aligned}$$

where K_s is the saturated hydraulic conductivity tensor at the reference density, $K_r(\psi)$ is the relative conductivity, η_z is a vector equal to zero in its x components and 1 in its

z component, $\sigma(\psi, c)$ is the general storage term or overall storage coefficient, t is time, ϕ is the porosity, $S_w(\psi)$ is the water saturation, q^* is the injected and q the extracted volumetric flow rate, $\vec{v}$ is the Darcy velocity vector, D is the dispersion tensor, c^* is the normalized concentration of salt in the injected/extracted fluid, and f is the volumetric rate of injected (positive)/ extracted (negative) solute that does not affect the velocity field.

In the following, we will consider the case with $S_w \equiv 1$ and $K_r \equiv 1$, that is the fully saturated case.

Initial conditions and Dirichlet, Neumann, or Cauchy boundary conditions are added to complete the mathematical formulation of the flow and transport problem (5.3). For the flow equation (5.3a), these take the form

$$\begin{aligned} \psi(\vec{x}, 0) &= \psi_0(\vec{x}) && && (5.4a) \\ \psi(\vec{x}, t) &= \psi_P(\vec{x}, t) && \text{on } \Gamma_1 && (5.4b) \\ \vec{v} \cdot \vec{n} &= -q_N(\vec{x}, t) && \text{on } \Gamma_2 && (5.4c) \end{aligned}$$

where $\vec{x} = (x, z)^T$ is the Cartesian spatial coordinate vector, ψ_0 is the pressure head at time 0, ψ_P is the prescribed pressure head (Dirichlet conditions) on boundary Γ_1, $\vec{n}$ is the outward normal unit vector, and q_N is the prescribed flux (Neumann condition) across boundary Γ_2. We use the sign convention of q_N positive for an inward flux and negative for an outward flux, consistent with the convention used for q and f in system (5.3).

For the transport equation (5.3c), the initial and boundary conditions are

$$\begin{aligned} c(\vec{x}, 0) &= c_0(\vec{x}) && && (5.5a) \\ c(\vec{x}, t) &= c_P(\vec{x}, t) && \text{on } \Gamma_3 && (5.5b) \\ D\vec{\nabla} c \cdot \vec{n} &= q_D(\vec{x}, t) && \text{on } \Gamma_4 && (5.5c) \\ (\vec{v}c - D\vec{\nabla} c) \cdot \vec{n} &= -q_C(\vec{x}, t) && \text{on } \Gamma_5 && (5.5d) \end{aligned}$$

where c_0 is the initial concentration, c_P is the prescribed concentration (Dirichlet condition) on boundary Γ_3, q_D is the prescribed dispersive flux (Neumann condition) across boundary Γ_4, and q_C is the prescribed total flux of solute (Cauchy condition) across boundary Γ_5. The sign convention for q_D and q_C is the same as for q_N, q and f.

5.2 About coupling and nonlinearity

Coupling in system (5.3) is due to the concentration terms in the flow equation (5.3a) and the head terms that appear in the transport equation (5.3c) via the Darcy velocities. In the simpler case of non-density-dependent flow and transport (that is $\epsilon = 0$), the system is coupled only through the head terms in the transport equation. In this case there is physical coupling, but mathematically the system can be reduced (decoupled) and solved sequentially, first the flow and then the transport equation, without iteration. For our density-dependent case, the system is irreducible and any sequential solution procedure requires iteration.

The nonlinearity of the coupled model (5.3) is due to the dependence of solution density on concentration that arises from relationship (5.1). For the flow equation (5.3a), the saturated conductivity $(K_s(1 + \epsilon c)/(1 + \epsilon' c))$, total head, and time derivative terms

are affected. As a consequence of the dependence on concentration in the flow equation, the transport equation (5.3c) is also nonlinear, in its convective and dispersive flux terms. When the density variations become larger than 3-4 %, flow and transport begin to be strongly coupled and the problem becomes increasingly nonlinear [41, 44].

5.3 Numerical Discretization

The solution of the discretized system of equations is usually addressed with an iterative Picard-like scheme by which the problem is decoupled by first solving the flow equation, then calculating the velocity field, and finally solving the transport equation. This three-step sequence is repeated until convergence. A difficulty for the successful application of the above procedure is the requirement that accurate velocity fields be obtained from the solution of the flow equation. Another problem that may influence the convergence of the Picard method is oscillatory behavior of the solution when the transport equation is advection dominated.

Several techniques have been used to solve numerically the coupled flow and transport problem (5.3). In [23, 24] a Finite Element (FE) discretization for the flow equation is joined with an analogous development for the transport equation. In [18] it has been shown that linear Galerkin finite element discretizations of the groundwater flow equation may violate the positive transmissibility property. This property ensures that the discrete flux is in the opposite direction of the head gradient. Violation of this condition means that, locally, nonphysical discrete fluxes can be generated, determining unacceptable approximation errors. This happens also when using the Galerkin approach in three-dimensional triangulations. To avoid this problem, a modification of the three-dimensional Galerkin scheme through a scheme called Orthogonal Subdomain Collocation is proposed in [43]. A more accurate approach to this problem in both two- and three-dimensional triangulations is given by Mixed Hybrid Finite Element (MHFE) method for solving flow, and Finite Volume scheme for solving transport equation [3, 15].

To solve the transport equation we apply the time-splitting technique developed in Chapter 3. The combination of these two approaches ensures high accuracy in both velocity and concentration fields, thus warranting the best possible convergence properties in the Picard iteration used to resolve coupling and nonlinearities.

5.3.1 MHFE method for flow

The MHFE formulation applied to the discretization of the flow equation is similar to the development presented in Chapter 3. Given the triangulation T_l, $l = 1, \ldots, m$, the time step Δt on the time interval $\left[t^k, t^{k+1}\right]$, discretization of equations (5.3a-5.3b) by MHFE

yields:

$$\int_{T_l} K_l^{-1}\vec{v}_l\cdot\vec{w}_{il}\ \mathrm{d}\Delta - \int_{T_l}\psi\vec{\nabla}\cdot\vec{w}_{il}\ \mathrm{d}\Delta + \int_{\partial T_l}\lambda\vec{w}_{il}\cdot\vec{n}_l\ \mathrm{d}\Gamma = -\int_{T_l}(1+\epsilon c)\eta_z\cdot\vec{w}_{il}\ \mathrm{d}\Delta \tag{5.6a}$$

$$\int_{T_l}\sigma_l\ \mathrm{d}\Delta\frac{\psi_l^{k+1}}{\Delta t} + \int_{T_l}\vec{\nabla}\cdot\vec{v}_l\ \mathrm{d}\Delta = \int_{T_l}\sigma_l\ \mathrm{d}\Delta\frac{\psi_l^{k}}{\Delta t} + \int_{T_l} f_l\ \mathrm{d}\Delta \tag{5.6b}$$

$$\int_{e_j}\vec{v}_l\cdot\vec{n}_l\ \mathrm{d}\Gamma + \int_{e_j}\vec{v}_r\cdot\vec{n}_r\ \mathrm{d}\Gamma = 0 \qquad \text{if } e_j\in T_l\cap T_r \tag{5.6c}$$

$$\int_{e_j}\vec{v}_l\cdot\vec{n}_l\ \mathrm{d}\Gamma = -q_N \qquad \text{if } e_j\in\Gamma_2\cap T_l \tag{5.6d}$$

$$\lambda_j = \psi_P \qquad \text{if } e_j\in\Gamma_1 \tag{5.6e}$$

where $i=1,2,3$, $l=1,\ldots,m$, $j=1,\ldots,n$, n being the number of edges, while the quantities with subscript l are defined over element T_l. Matrix K is given by $K=K_s\dfrac{1+\epsilon c}{1+\epsilon' c}K_r$ while $f=-\phi S_w\epsilon\dfrac{\partial c}{\partial t}+\dfrac{\rho}{\rho_0}q^*+q$. The partial derivative $\dfrac{\partial c}{\partial t}$ is approximated by the backward Euler scheme.

In the above system, equation (5.6a) is the MHFE discretization of the mass conservation equation (5.3b); equation (5.6b) represents the discretized version of (5.3a); equation (5.6c) guarantees continuity of the normal flux across interelement edges, while equations (5.6d-5.6e) are the explicitly imposed Neumann and Dirichlet boundary conditions, respectively.
By considering a weighted scheme for the time quadrature with weighting parameter $\theta=0.5$ (Crank-Nicolson) or $\theta=1$ (Euler), ($p^{k+\theta}=\theta p(\cdot,t^{k+1})+(1-\theta)p(\cdot,t^k)=\theta p^{k+1}+(1-\theta)p^{k+1}$), the final hybrid formulation can be written in matrix notation as:

$$\begin{pmatrix} A & -B & C\\ B^T & D' & 0\\ C^T & 0 & 0\end{pmatrix}\begin{pmatrix}\mathbf{v}^{k+\theta}\\ \boldsymbol{\psi}^{k+\theta}\\ \boldsymbol{\lambda}^{k+\theta}\end{pmatrix} = \begin{pmatrix}\mathbf{g}_1\\ \mathbf{g}_2+\tilde{D}\boldsymbol{\psi}^k\\ \mathbf{g}_3\end{pmatrix} \tag{5.7}$$

where $A=\mathrm{diag}[A_1,\ldots,A_m]$, $B=\mathrm{diag}[B_1,\ldots,B_m]$ and

$$A_l=(a_{ik})=\int_{T_l}K_l^{-1}\vec{w}_{il}\cdot\vec{w}_{kl}\ \mathrm{d}\Delta \qquad B_l=(b_i)=\int_{T_l}\vec{\nabla}\cdot\vec{w}_{il}\ \mathrm{d}\Delta$$
$$C=(c_{rj})=\int_{\partial T_l}\mu_j\vec{w}_{il}\cdot\vec{n}_l\ \mathrm{d}\Gamma \qquad \mathbf{v}=(v_r)=v_{il}$$
$$\mathbf{g}_1=(g_{1l})=\int_{T_l}(1+\epsilon c)\eta_z\cdot\vec{w}_{il}\ \mathrm{d}\Delta \quad \mathbf{g}_2=(g_{2l})=\int_{T_l}f_l\ \mathrm{d}\Delta.$$

In the above equations $i,k=1,2,3$, $j=1,\ldots,n$, $r=3(l-1)+i$ and $\boldsymbol{\psi}=(\psi_l)$, $\boldsymbol{\lambda}=(\lambda_j)$, and $\mathbf{g}_3=(-q_{Nj})$ where $-q_{Nj}$ assumes a non vanishing value only if there is a Neumann

condition on a boundary edge e_j. Setting $D^k = \mathrm{diag}[d_1^k, \ldots, d_m^k]$ with $d_l^k = \int_{T_l} \sigma_l^k \, \mathrm{d}\Delta/\Delta t$, then $D' = \dfrac{D^{k+1}}{\theta}$ and $\tilde{D} = D' - D^{k+1} + D^k$. If d_l^k is not dependent on time, then $D' = \tilde{D}$.

The velocity field $\mathbf{v}^{k+\theta}$ is obtained from the upper system block:

$$\mathbf{v}^{k+\theta} = A^{-1}(B\boldsymbol{\psi}^{k+\theta} - C\boldsymbol{\lambda}^{k+\theta} + \mathbf{g}_1). \tag{5.8}$$

Substitution of this expression into the remaining equations yields a reduced system with $\boldsymbol{\psi}^{k+\theta}$ and $\boldsymbol{\lambda}^{k+\theta}$ as unknowns:

$$\begin{pmatrix} D' + B^T A^{-1} B & -B^T A^{-1} C \\ C^T A^{-1} B & -C^T A^{-1} C \end{pmatrix} \begin{pmatrix} \boldsymbol{\psi}^{k+\theta} \\ \boldsymbol{\lambda}^{k+\theta} \end{pmatrix} = \begin{pmatrix} -B^T A^{-1}\mathbf{g}_1 + \mathbf{g}_2 + \tilde{D}\boldsymbol{\psi}^k \\ -C^T A^{-1}\mathbf{g}_1 + \mathbf{g}_3 \end{pmatrix}. \tag{5.9}$$

Setting $H = D' + B^T A^{-1} B$ and $S = A^{-1}B$, the pressure head is given by:

$$\boldsymbol{\psi}^{k+\theta} = H^{-1}(S^T C\boldsymbol{\lambda}^{k+\theta} + S^T\mathbf{g}_1 + \mathbf{g}_2 + \tilde{D}\boldsymbol{\psi}^k), \tag{5.10}$$

Finally, setting $M = A^{-1} - SH^{-1}S^T$, the final system of equations is obtained:

$$C^T M C\boldsymbol{\lambda}^{k+\theta} = C^T M\mathbf{g}_1 + C^T S H^{-1}(\mathbf{g_2} + \tilde{D}\boldsymbol{\psi}^k) - \mathbf{g}_3, \tag{5.11}$$

from which we get $\boldsymbol{\lambda}^{k+\theta}$. Once $\boldsymbol{\lambda}^{k+\theta}$ is calculated, $\boldsymbol{\psi}^{k+\theta}$ and $\mathbf{v}^{k+\theta}$ can be evaluated using equations (5.10) and (5.8).

Note that the matrix $C^T MC$ depends on the unknown values of concentration c^{k+1} and thus system (5.11) is nonlinear. We use a Picard iterative approach to linearization, i.e. in solving (5.11) for $\boldsymbol{\lambda}^{k+1,r+1}$ where r is the iteration counter, we set $\mathbf{c}^{k+1} = \mathbf{c}^{k+1,r}$. The linearized system is thus symmetric and positive definite and can be solved efficiently by the Preconditioned Conjugate Gradient method.

5.3.2 Time-splitting for transport

The time-splitting technique presented in the previous Chapter is applied to solve the transport equation. Equation (5.3c) is splitted into two separate partial differential equations containing one the dispersive flux $-D\vec{\nabla}c$ plus the terms $qc^* + f$, and the other containing the advective flux $c\vec{v}$. We consider the first order accurate in time and second order accurate in space version and use $n_a > 1$. The advective time step is determined by the CFL constraint while the dispersive time step is determined by accuracy considerations.

5.3.3 The Picard method

The Picard iteration for the solution of the nonlinearly coupled system of equations proceeds as follows. To advance in time from t^k to t^{k+1}, at the $(r+1)$-st iteration the flow equation is first solved for $\boldsymbol{\psi}^{k+1,r+1}$ and $\mathbf{v}^{k+1,r+1}$, freezing the values of the concentration, $\mathbf{c}^{k+1,r}$, at the previous iteration. Employing these updated values of velocity and pressure head, the transport equation is then solved for $\mathbf{c}^{k+1,r+1}$. This procedure is repeated until convergence is achieved. The algorithm can be expressed mathematically in the following way:

- flow equation:

$$C^T M C \boldsymbol{\lambda}^{k+\theta,r+1} = C^T M \mathbf{g}_1 + C^T S H^{-1}(\mathbf{g_2} + \tilde{D}\boldsymbol{\psi}^k) - \mathbf{g}_3$$
$$\boldsymbol{\psi}^{k+\theta,r+1} = H^{-1}(S^T C \boldsymbol{\lambda}^{k+\theta,r+1} + S^T \mathbf{g}_1 + \mathbf{g}_2 + \tilde{D}\boldsymbol{\psi}^k)$$
$$\mathbf{v}^{k+\theta,r+1} = A^{-1}(B\boldsymbol{\psi}^{k+\theta,r+1} - C\boldsymbol{\lambda}^{k+\theta,r+1} + \mathbf{g}_1)$$

  where dependence on concentration is implicit;

- transport equation:

$$\hat{\mathbf{c}}^{k+1,r+1} = \mathbf{c}^k + \Delta t_a \left[L_a(\mathbf{c}^k)\right]$$
$$\mathbf{c}^{k+1,r+1} = \hat{\mathbf{c}}^{k+1,r+1} + \Delta t_d \left[L_d(\mathbf{c}^{k+1,r+1})\right].$$

Since the advection equation is solved explicitly and depends only on the values at time t^k, if the dispersion flux is null in the transport equation, we would reach convergence in just one iteration. In the general case, convergence is achieved when:

$$\begin{aligned} \|\boldsymbol{\psi}^{k+1,r+1} - \boldsymbol{\psi}^{k+1,r}\|_2 &\leq \mathrm{tol}_\psi \\ \|\mathbf{c}^{k+1,r+1} - \mathbf{c}^{k+1,r}\|_2 &\leq \mathrm{tol}_c \end{aligned} \tag{5.12}$$

where tol_ψ and tol_c are two fixed tolerances.

In the applications, the time step size is adaptively adjusted according to the behavior of the previous iteration. The simulation begins with a time step size of Δt_0 and an initial guess given by input values or by the values calculated at the previous time step. The size of the time step has a direct effect on the convergence of the iteration because of its influence on the quality of the initial solution estimate. After the convergence of the Picard iteration, the time step size for the next time step is increased by a factor of Δt_{incr} (to a maximum size of Δt_{max}) if convergence was achieved in fewer than $maxit_1$ iterations; it is left unchanged if convergence was reached between $maxit_1$ and $maxit_2$ iterations; and it is decreased by a factor of Δt_{red} (to a minimum of Δt_{min}) if convergence required more than $maxit_2$ iterations. If convergence is not achieved (that is relations (5.12) do not hold within a maximum number of iterations, $maxit$), the time step is repeated (we back-step) using a reduced time step size (by Δt_{red}, unless the time step cannot be reduced any further, in which case we set $\Delta t = \Delta t_{min}$). The values of the various Δt's and $maxit$'s are chosen empirically.

5.4 Notes on the definition of flow and transport coefficients

To be able to solve correctly the coupled flow and transport problem, all the coefficients in equations (5.3) must be specified. Recall that the flow and velocity equations (5.3a- 5.3b) take on the form:

$$\sigma \frac{\partial \psi}{\partial t} = \vec{\nabla} \cdot [K_s \frac{1+\epsilon c}{1+\epsilon' c} K_r (\vec{\nabla}\psi + (1+\epsilon c)\eta_z)] - \phi S_w \epsilon \frac{\partial c}{\partial t} + \frac{\rho}{\rho_0} q^* + q$$
$$\vec{v} = -K_s \frac{1+\epsilon c}{1+\epsilon' c} K_r (\vec{\nabla}\psi + (1+\epsilon c)\eta_z)$$

We have defined only ϵ, ϵ', ρ, ρ_0 and the function η_z. The definition of the remaining coefficients is taken from [4]. The general storage term or overall storage coefficient, σ is given by:

$$\sigma = S_s(1+\epsilon c)$$

where S_s is the elastic storage coefficient. (In the general case of variably saturated flow and transport equations, we have $\sigma = S_w S_s(1+\epsilon c) + \phi\rho_0(1+\epsilon c)\dfrac{\partial S_w}{\partial \rho}$. Now we are considering $S_w = 1$.) The saturated hydraulic conductivity tensor K_s is defined by:

$$K_s = \frac{\rho_0 g k}{\mu}$$

where k is the intrinsic permeability tensor of the porous medium. K_r, the relative conductivity, is constant and equal to 1 in our applications. In general it is a function that depends on the pressure head ψ. The porosity ϕ and the volumetric flow rates q^* and q are constant or functions that do not depend on pressure or concentration.

The transport equation is given by

$$\phi\frac{\partial S_w c}{\partial t} = \vec{\nabla}\cdot(D\vec{\nabla}c) - \vec{\nabla}\cdot(c\vec{v}) + qc^* + f$$

where the dispersion tensor D is given by [1]:

$$D_{ij} = \alpha_T|\vec{v}|\delta_{ij} + (\alpha_L - \alpha_T)\frac{v_i v_j}{|\vec{v}|} + \phi S_w D_0 \tau \delta_{ij} \qquad i,j = x,z$$

with α_L and α_T being the longitudinal and transverse dispersivity coefficients, respectively, $|\vec{v}| = \sqrt{v_x^2 + v_z^2}$, δ_{ij} the Kronecker delta, D_0 the molecular diffusion coefficient, and τ the tortuosity ($\tau = 1$ is usually assumed). Nonzero off-diagonal coefficients in D are avoided by aligning the local coordinate system of each triangle to the principal axis of anisotropy. To this aim, we consider a new coordinate system, aligned with the direction of Darcy's velocity, so that $x' = (v_x x - v_z z)/|\vec{v}|$ and $z' = (v_z x - v_x z)/|\vec{v}|$. Then the elements of the dispersion tensor become

$$\begin{aligned} D_{xx} &= \alpha_L|\vec{v}| + \phi S_w D_0 \tau \\ D_{zz} &= \alpha_T|\vec{v}| + \phi S_w D_0 \tau \\ D_{xz} &= D_{zx} = 0 \end{aligned}$$

5.5 Application to Elder's problem

Our first example is a classical natural convection problem used to validate the coupled flow and transport solvers. It was suggested in the literature [60] as an exercise to test the accuracy of models in representing fluid flow driven purely by density differences. Elder's problem is concerned with the movement of a solute moving in a closed rectangular box by diffusion [53].

A source of solute with constant unit value is applied at the top of a closed rectangular box while concentration at the base is maintained at zero. The pressure is initially hydrostatic, that is, by assuming a uniform surface elevation of the top layer, equal to z_{max}, the inital pressure head is $\psi(\vec{x}, 0) = z_{max} - z$. The two upper corners of the box are held at zero pressure head. The solute enters the pure water initially by diffusion, increases its density and a circulation process begins. The motion develops as a set of eddies forming at the two ends of the source, because of the solute density. Small eddies of reverse circulation are associated with the end eddies, followed by a further set of eddies growing near the ends. This process continues in time until the end eddies merge into single large eddies.

TABLE 5.1: *Parameters for Elder's problem.*

Domain	Rectangular section of 600 m × 150 m
Source location	Centered on the upper boundary at 150 m $\leq x \leq$ 450 m
Permeability:	
K_{sx}	0.410654 m/d
K_{sz}	0.410654 m/d
Elastic storage S_s	9.8e-3
Density ratio, ϵ	0.2
Viscosity ratio, ϵ'	0
Porosity ϕ	0.1
Dispersivity α_L, α_T	0 m
Diffusion coefficient D_0	0.308016 m^2/d
BCs for flow	No flow Zero pressure head at the two upper corners
BCs for mass transport	$c = 0$ along the base and $c = 1$ at the source. Zero concentration gradient elsewhere
IC for pressure head	Hydrostatic
IC for concentration	$c = 0$
Grid characteristics	2000 triangles (100 × 20) and 1070 edges
Horizontal nodal spacing, Δx	12 m
Vertical nodal spacing, Δz	7.5 m
Time increment, Δt	30 d
Convergence criteria, tol_ψ, tol_c	1.e-3

The parameters used in our simulation are given in Table 5.1, where we use the notation BCs to represent Boundary Conditions and IC to represent Initial Condition. The 600 m × 150 m domain shown in Figure 5.1 is discretized with a triangular grid of 2000 elements. Figures 5.2 and 5.3 show the results obtained for $t = 1$, 2, 3, and 4 years. The solute transport solution agrees very well with the results of the literature in [53, 60]. The results indicate that the proposed approach is accurate and reliable, does not suffer from numerical oscillation and does not introduce large amounts of numerical diffusion, as typically done by more conventional upwind discretizations. Also, the presence of unsymmetric meshes does not dramatically influence the solution, as reported in the literature for other types of discretization approaches.

5.6 Application to the Lake Karachai problem

Existing three dimensional numerical models and codes developed for the simulation of groundwater flow and contaminant transport are extended and enhanced by the partners of the RaCoS project (see Chapter 1) to perform simulations at the Lake Karachai. The procedure here developed to solve density dependent flow and transport in groundwater, being a highly accurate two dimensional simulation model, is deemed necessary to verify

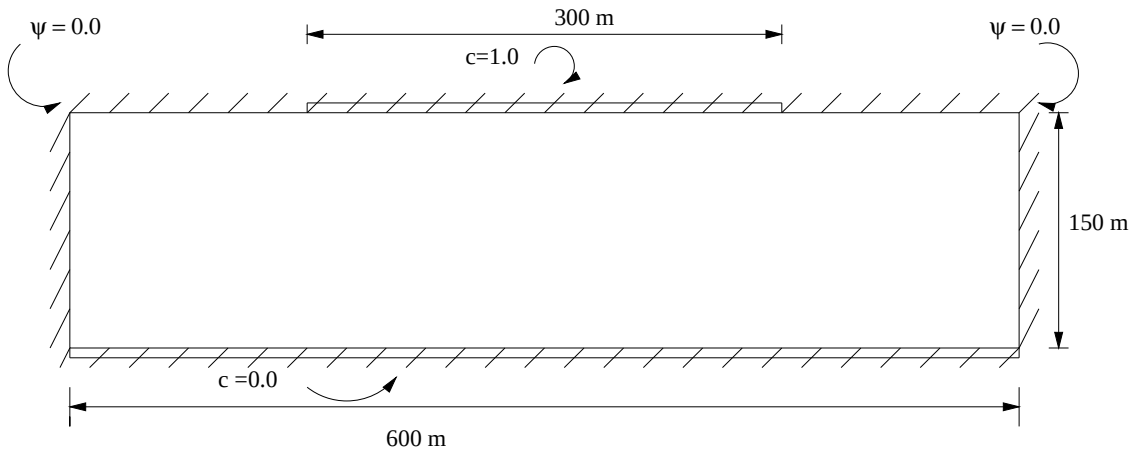

FIGURE 5.1: *Voss and Souza definition of Elder's problem.*

the results obtained by the three dimensional codes. In particular, it is important to assess how the large amount of numerical diffusion necessarily introduced by the three dimensional codes to maintain stability, and possible appearance of negative concentration values affect the prediction of the radionuclide movement. In fact these numerical phenomena may alter the adsorption/desorption pattern of the solute and change the speed at which the contaminant plume moves [50].

The combination of the MHFE method for flow and of the time-splitting technique for transport can be now applied to determine the extent and the character of the contaminant plume at the Lake Karachai site in two dimensional simulations, to predict the behavior of the contaminant plume under different scenarios and to verify the results obtained by three dimensional codes. A vertical section of the complex Karachai site is considered in Figure 5.4. The Lake Karachai is 400 m long and is almost centered within the domain of interest. According to the data acquired by field measurements, three main zones have been idientified with different hydraulic conductivity values. In addition, since precipitation is the primary recharge mechanism for the underlying aquifer, prescribed Neumann conditions have been considered as boundary conditions for pressure. The bottom and the left side of the domain are assumed to be impermeable. The test cases reported next are taken from the RaCoS project yearly reports and are considered to be representative of the real situation [49, 50, 21, 32, 20, 31].

5.6.1 Case 1

In this first case, we consider a schematic domain of the Lake Karachai. A rectangular section of size 500 m $\times$ 100 m is considered in Figure 5.5. The source simulating the Lake is located at the top of the domain at 225 m $\leq x \leq$ 260 m (face BC). The boundary conditions for the flow equation are as follows. Zero flux is imposed on the vertical faces DE and EF. Faces AB and CD are subjected to a Neumann flux equal to q_N= 2.16e-4 m^2/d, while at the source BC q_N = 3.024e-3 m^2/d is prescribed. On side FA hydrostatic conditions with pressure head $\psi = 0$ at the top are assumed. For the transport equation we set at the source BC Dirichlet boundary conditions $c = 1$ and zero dispersive flux elsewhere.

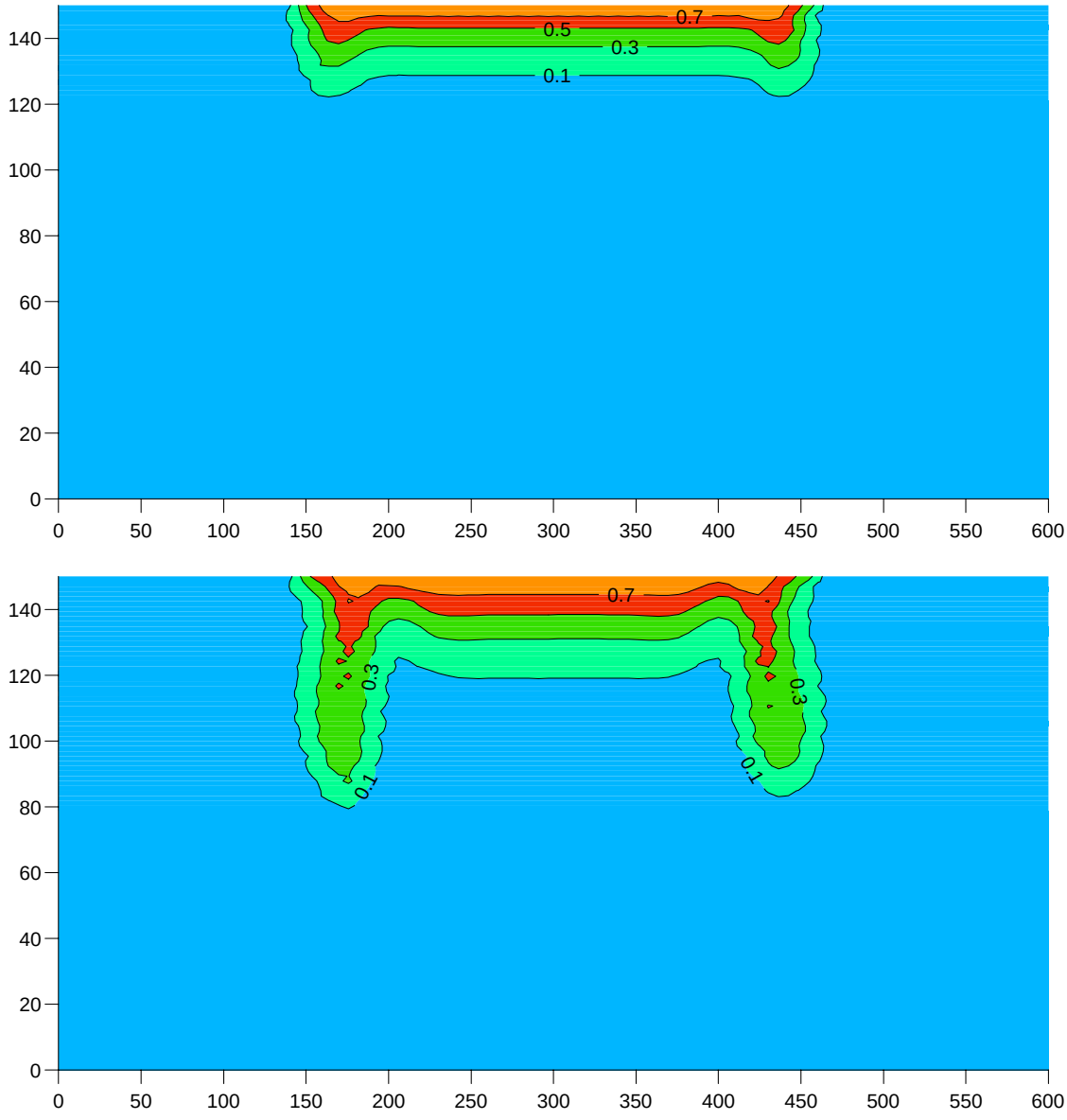

FIGURE 5.2: *Concentration contours for Elder's problem at 360 days (top) and 720 days (bottom).*

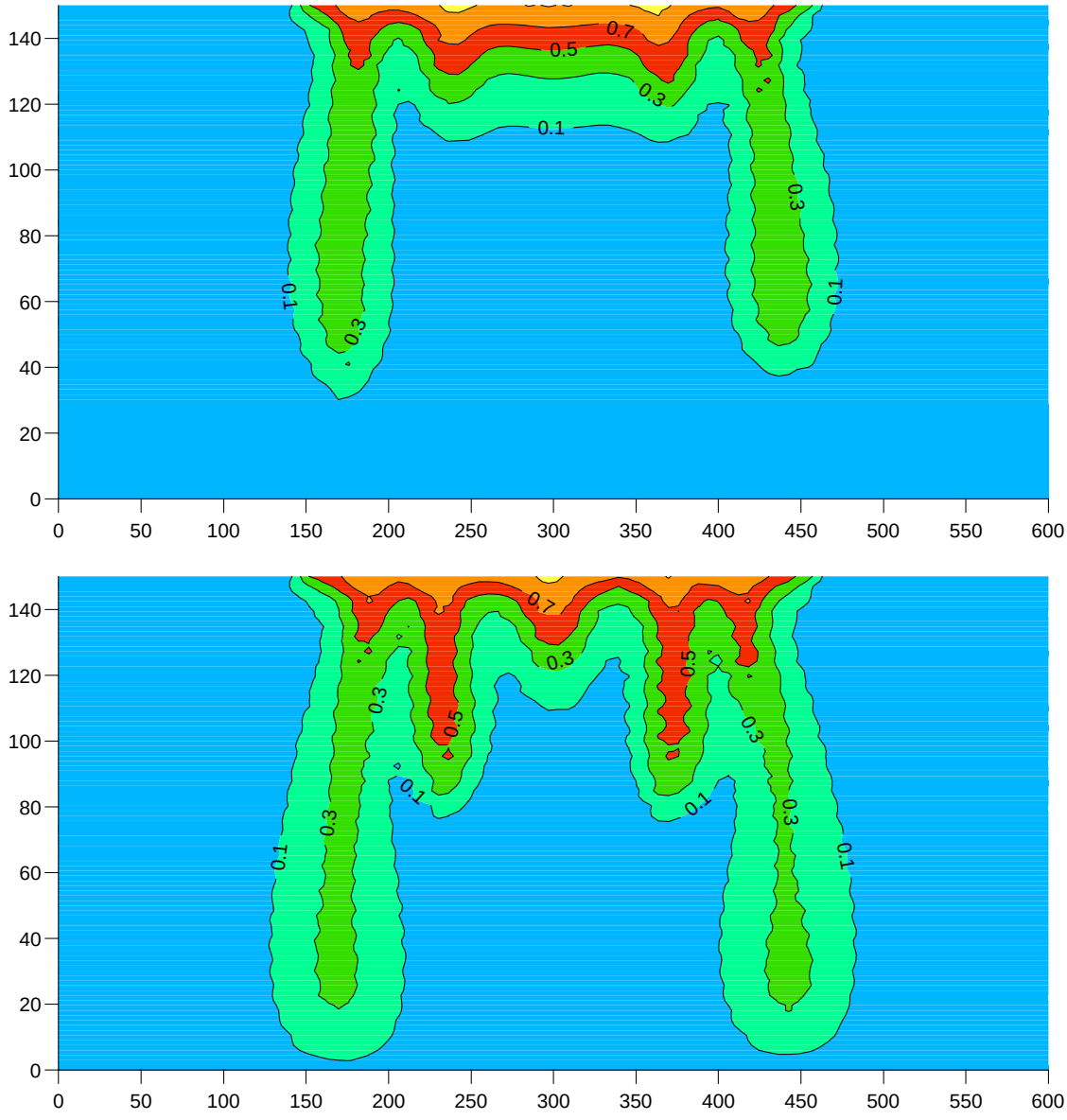

FIGURE 5.3: *Concentration contours for Elder's problem at 3 years (top) and 4 years (bottom).*

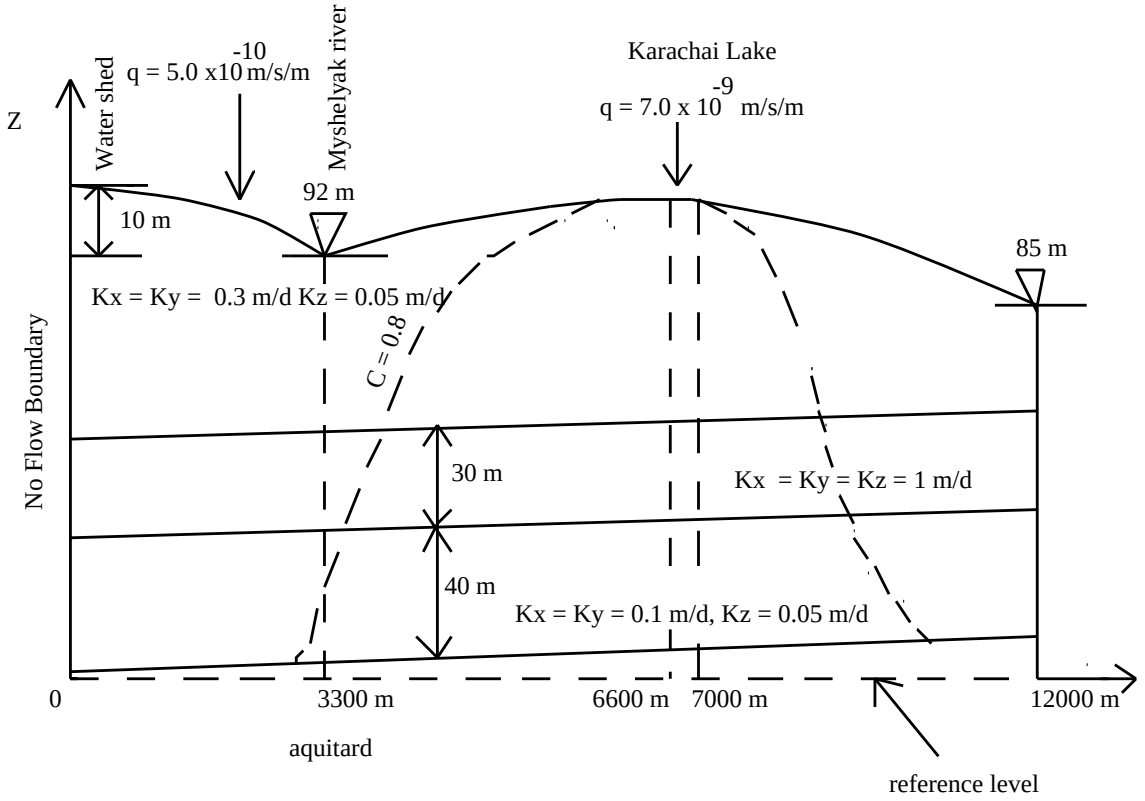

FIGURE 5.4: *Vertical section of the Lake Karachai site.*

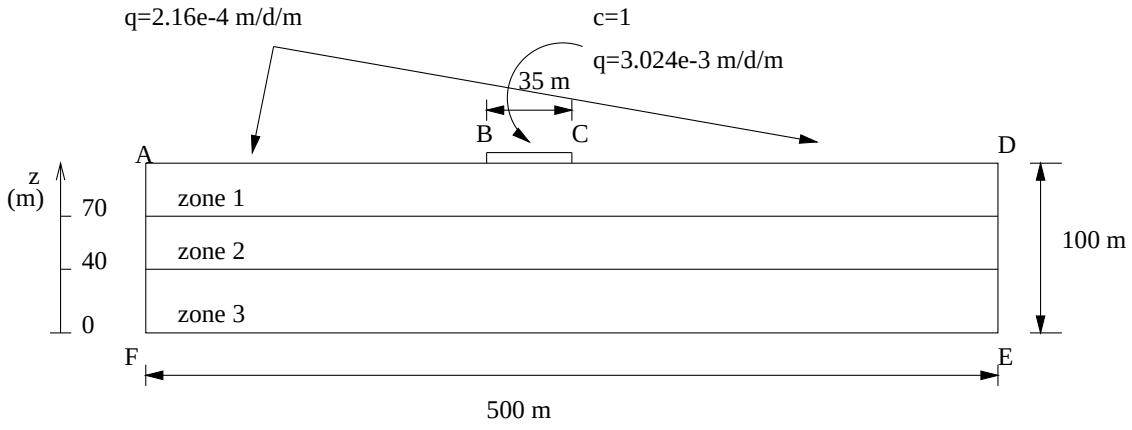

FIGURE 5.5: *Domain for the Case 1.*

According to data acquired by field measurement, we consider three different permeability zones, (Figure 5.5), with a more permeable central layer (zone 2). The parameters used in this simulation are summarized in Table 5.2.

The distributed water infiltration at the surface together with the hydrostatic condition at the left side of the domain create a regional flow from left to right.

Figures 5.6, 5.7 and 5.8 display the contaminant plume as calculated with the proposed approach after 500, 1000, 2000, 3000, 4000, and 5000 days. At the beginning the plume moves symmetrically downward driven by gravity. It widens when it encounters the zone with larger permeability, and tends to deposit at the bottom of the aquifer. At t=2000 days (Figure 5.7), the contaminant tilts towards the right in accordance to the regional flow.

5.6.2 Case 2

This simulation, simpler than the previous one, considers a rectangular domain of 1500 m $\times$ 100 m, discretized with a triangular grid of 600 elements. The test case tries to assess the effects that physical instabilities may have on a long-term simulation (t $\geq$ 2000 days). It is expected that after a period of time the fingering due to density differences will not affect the shape of the contamination plume. The source of contaminant (the lake) with constant unit value is applied in the left upper area of the domain, at 0 m $\leq x \leq$ 300 m (see Figure 5.9). At the source there is an incoming water flux q_N= 5.e-2 m^2/d. On the right upper corner a Dirichlet condition of prescribed pressure head ψ=150 m is fixed to simulate the presence of a river. Initial pressure and concentrations are zero. The domain is assumed to be homogeneous. The parameters used are summarized in Table 5.3.

Figures 5.10 and 5.11 display the concentration contours at 500, 1000, 1500, and 2000 days. The instabilities mentioned above are clearly visible at the early times but are almost completely disappeared already at t=1000 days. The transport of the contaminant from the top to the bottom of the aquifer is simulated accurately and does not reveal overly dispersive concentration fronts, even through we are using a very coarse grid. To obtain similar results with standard FEM a much more refined mesh needs to be employed.

5.6.3 Case 3

The final simulation considers a vertical cross section comprising the Lake Karachai (CD), and the rivers Myshelyak (B) and Techa (E) (Figure 5.12). A domain of 12000 m of length and 85 m of depth is discretized with 19200 triangles. The Lake Karachai is located at 6600 m $\leq x \leq$ 7000 m. Furthermore, at points B and E at the top of the domain (x= 3300 m and x=12000 m), the pressure head is fixed to simulate the presence of the Myshelyak and Techa river, respectively. The domain is assumed to be heterogeneous and the three main layers of test case 1 are used. The parameters for the simulation are reported in Table 5.4.

The initial conditions are given by the steady state flow distribution, to ensure a stable initial flow for the coupled flow and transport problem. Figures 5.13, 5.14 and 5.15 show the concentration contours after 1, 2, 4, 8, 10, and 15 years, respectively. The solution displays a very tight dispersion front, reflecting the ability of the proposed approach to minimize numerical diffusion. The oscillations that are visible in the middle of the plume

TABLE 5.2: *Parameters for the Case 1.*

Domain	Rectangular section of 500m × 100 m
Source location	Centered on the upper boundary at 225 m $\leq x \leq$ 260 m
Permeability	3 zones
zone 1 ($70 \leq z \leq 100$)	
K_{sx}	0.30 m/d
K_{sz}	0.05 m/d
zone 2 ($40 \leq z \leq 70$)	
K_{sx}	1. m/d
K_{sz}	1. m/d
zone 3 ($0 \leq z \leq 40$)	
K_{sx}	0.1 m/d
K_{sz}	0.05 m/d
Elastic storage S_s	1.e-5 m^{-1}
Density ratio, ϵ	0.07
Viscosity ratio, ϵ'	0
Porosity ϕ	0.025
Dispersivity α_L	5 m
Dispersivity α_T	1 m
Diffusion coefficient D_0	0 m^2/d
BCs for flow	q_N= 2.16e-4 m^2/d on faces AB ad CD q_N = 3.024e-3 m^2/d at the source hydrostatic frewshwater head with ψ=0 m at the top on side FA zero flux elsewhere
BCs for mass transport	$c = 1$ at the source no dispersive flux elsewhere
IC for pressure head	Hydrostatic
IC for concentration	$c = 0$
Grid characteristics	4000 triangles (200 × 20) and 2021 edges
Horizontal nodal spacing, Δx	5 m
Vertical nodal spacing, Δz	5 m
Time increment, Δt	30 d
Convergence criteria, tol_ψ, tol_c	1.e-3

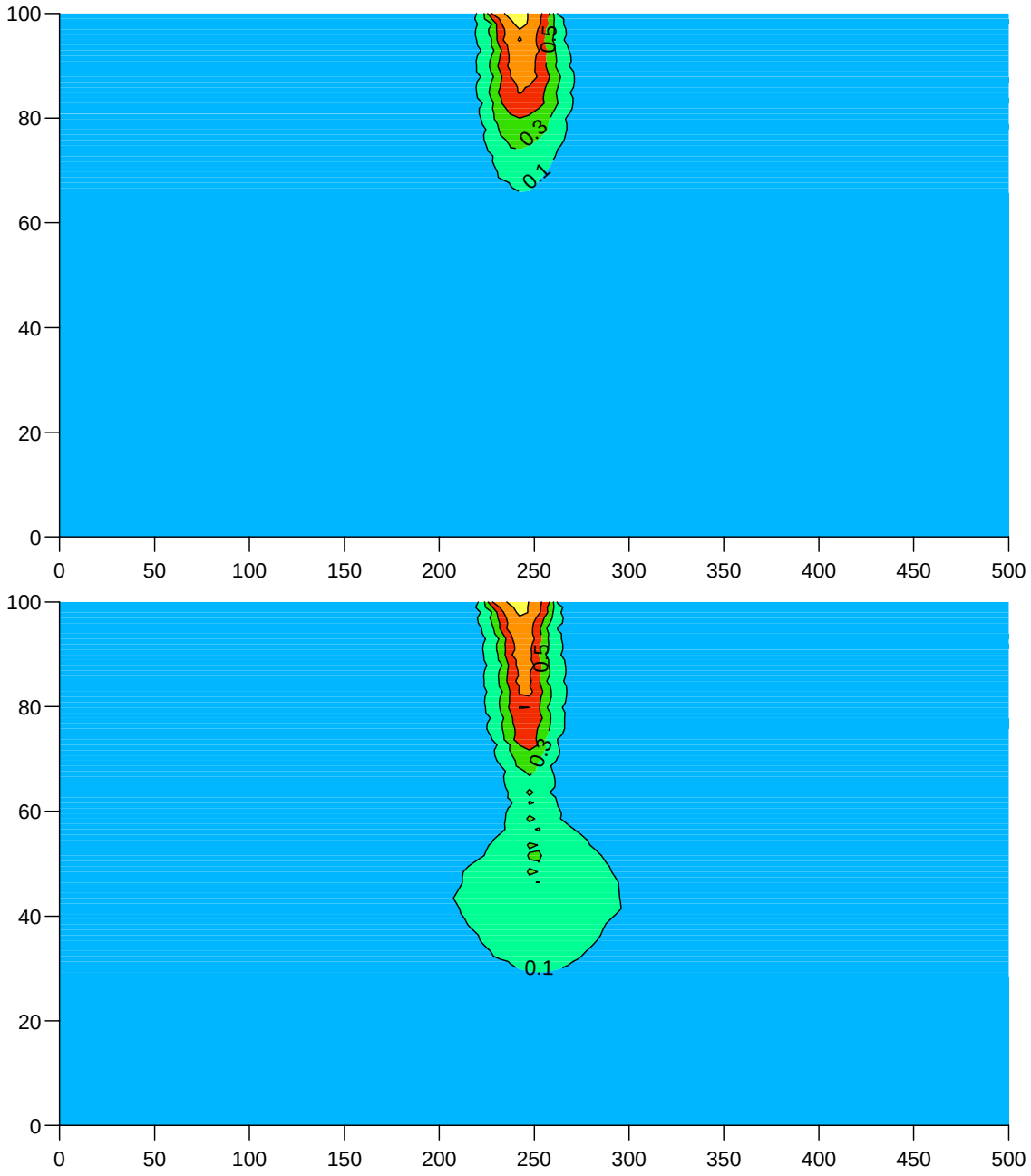

FIGURE 5.6: *Case 1: concentration contours for the Lake Karachai problem at 500 days (top) and 1000 days (bottom).*

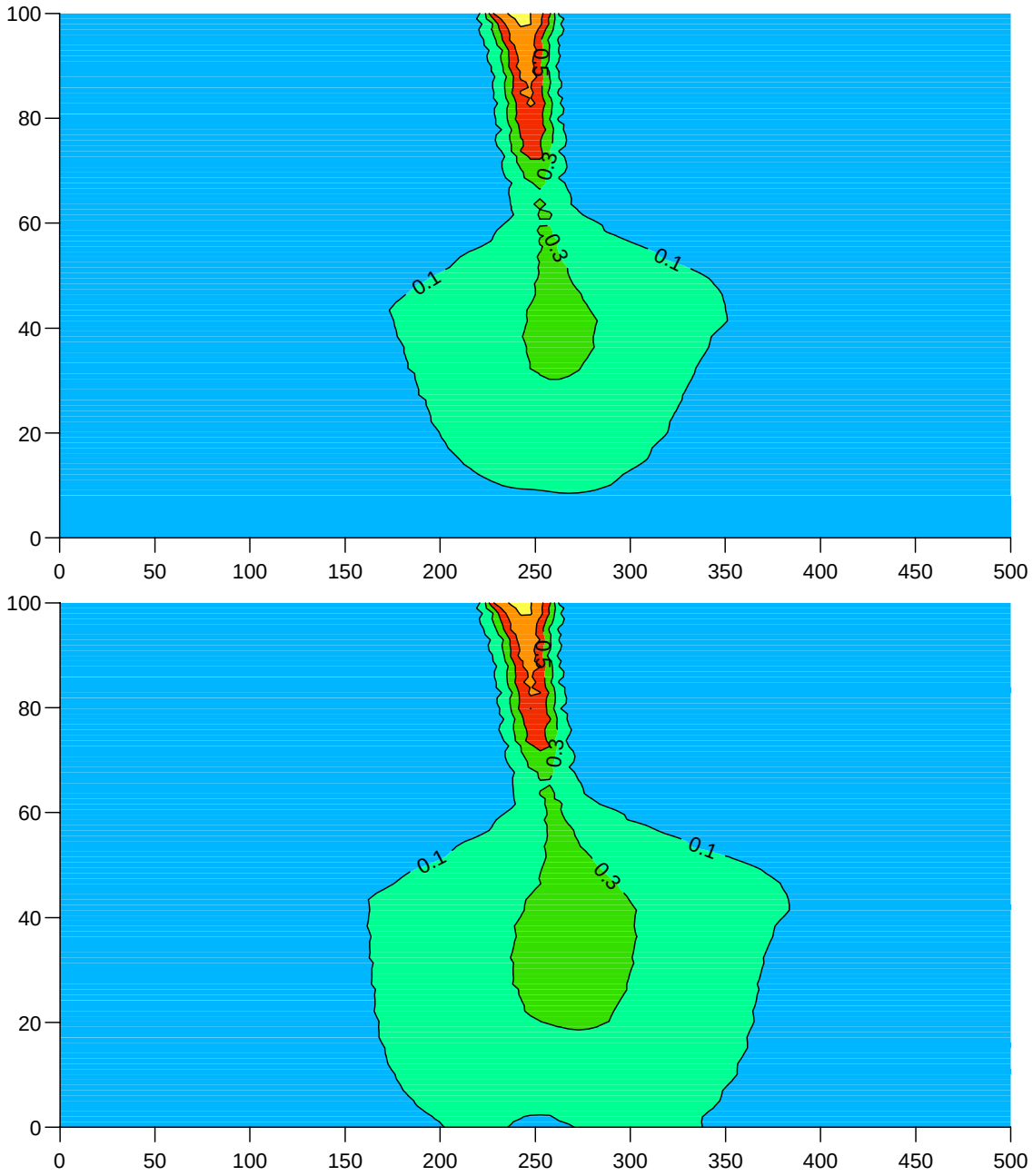

FIGURE 5.7: *Case 1: concentration contours for the Lake Karachai problem at 2000 days (top) and 3000 days (bottom).*

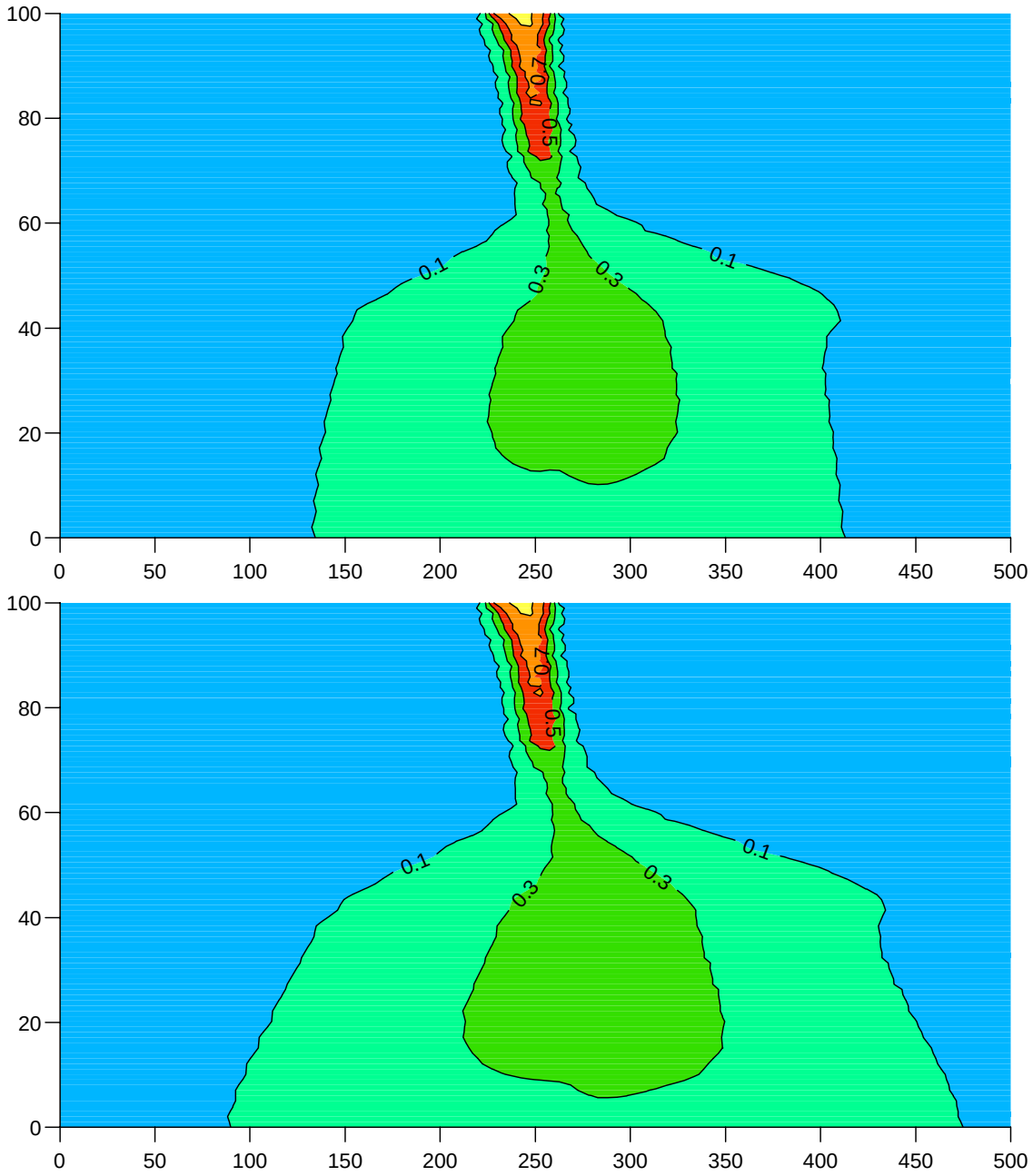

FIGURE 5.8: *Case 1: concentration contours for the Lake Karachai problem at 4000 days (top) and 5000 days (bottom).*

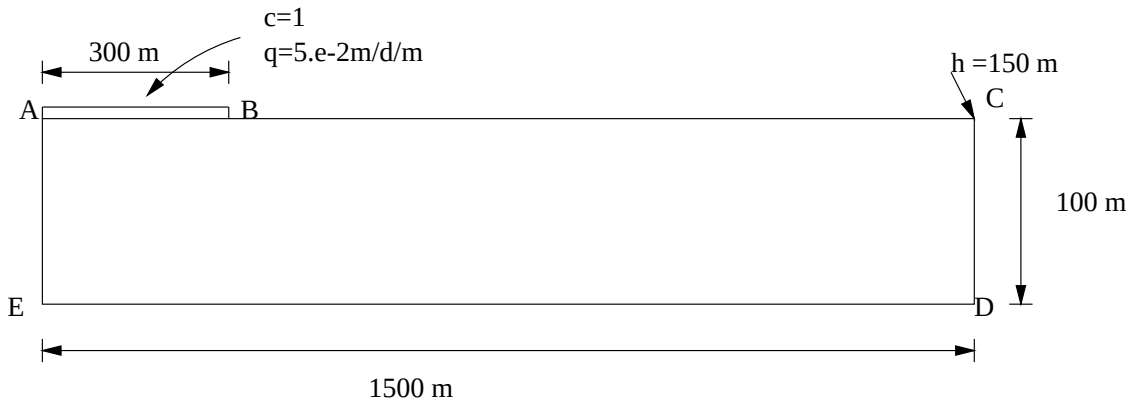

FIGURE 5.9: *Domain for the Case 2.*

TABLE 5.3: *Parameters for the Case 2.*

Domain	Rectangular section of 1500m $\times$ 100 m
Source location	Centered on the upper boundary
	at 0 m $\leq x \leq$ 300 m
Permeability	
K_{sx}	1 m/d
K_{sz}	1 m/d
Elastic storage S_s	1.e-5 m^{-1}
Density ratio, ϵ	0.07
Viscosity ratio, ϵ'	0
Porosity ϕ	0.010
Dispersivity α_L	50 m
Dispersivity α_T	10 m
Diffusion coefficient D_0	0 m^2/d
BCs for flow	q_N= 5.e-2 m^2/d at the source
	frewshwater head ψ=150 m on the right upper corner
	zero flux elsewhere
BCs for mass transport	$c = 1$ at the source
	zero dispersive flux elsewhere
IC for pressure head	$\psi = 0$
IC for concentration	$c = 0$
Grid characteristics	600 triangles (60 $\times$ 10) and 341 edges
Horizontal nodal spacing, Δx	50 m
Vertical nodal spacing, Δz	10 m
Time increment, Δt	10 d
Convergence criteria, tol_ψ, tol_c	1.e-4

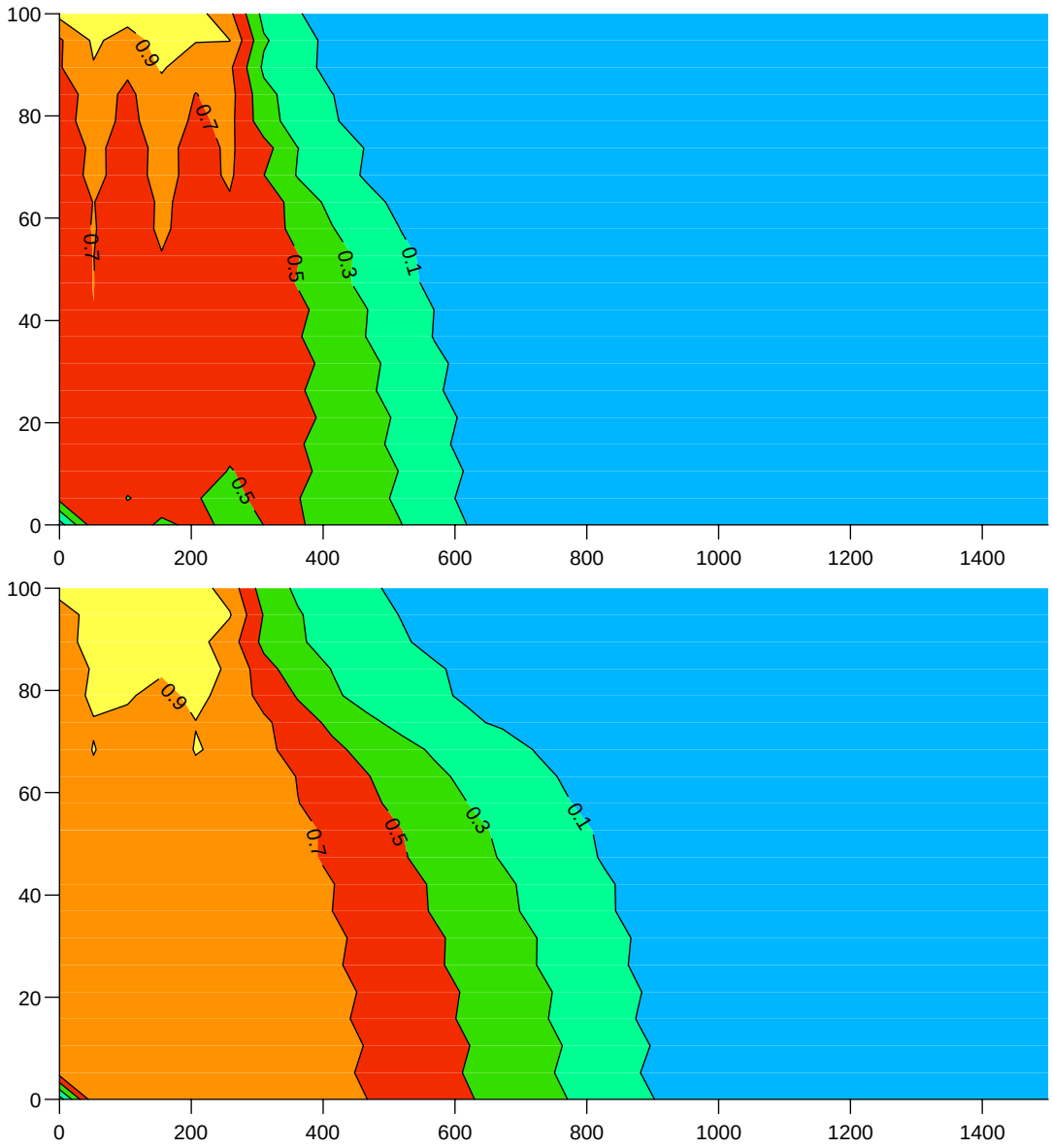

FIGURE 5.10: *Case 2: concentration contours at 500 days (top) and 1000 days (bottom).*

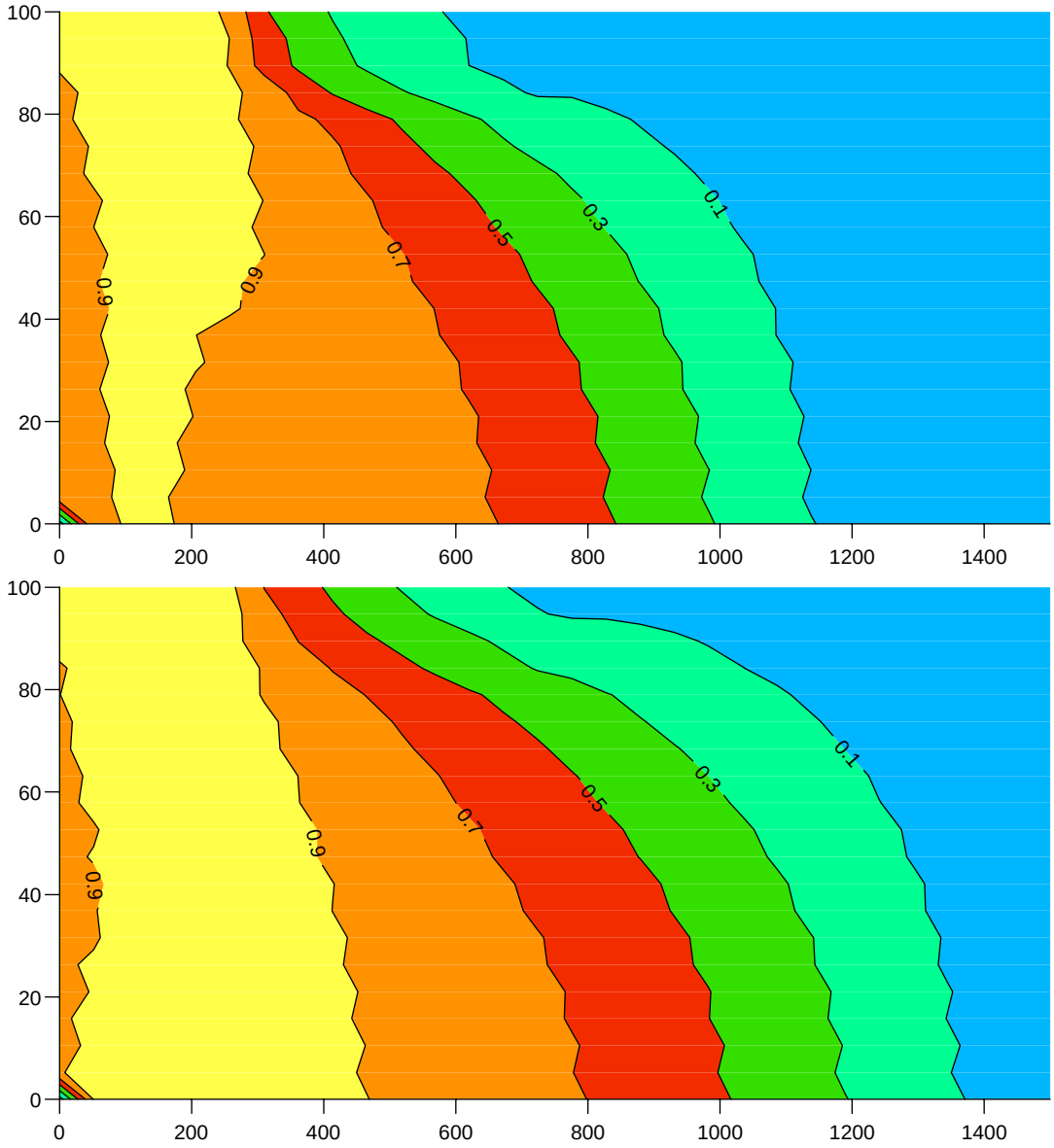

FIGURE 5.11: *Case 2: concentration contours at 1500 days (top) and 2000 days (bottom).*

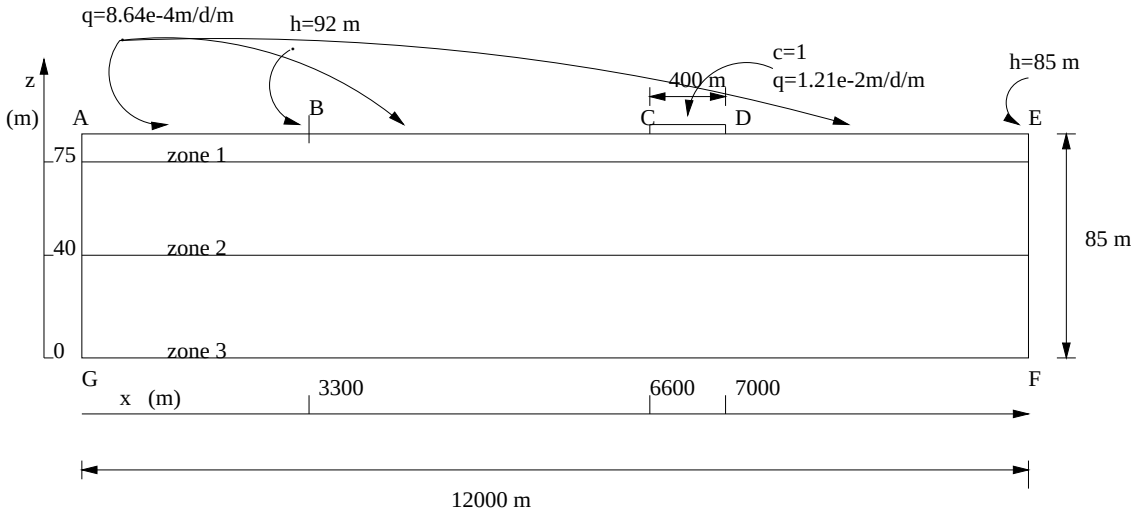

FIGURE 5.12: *Domain for the Case 3.*

are considered to by physical instabilities and disappear at t=15 years. The results are in very good agreement with the experimental observations. Comparison with the results obtained by the FEM model seems to indicate that the overly dispersive front of the latter does not dramatically influence the global qualitative behavior of the plume. However a quantitative assessment of the contaminant flux that may be discharged in the Techa and Myshelyak rivers needs to be assessed more accurately than it is possible with the FE approach.

Future work in this direction will be performed in portions of the domain surrounding the two rivers. Starting with initial and boundary conditions given by the three dimensional FE model, the amount of contaminant discharged into the rivers will be calculated using the approach proposed in this Thesis.

TABLE 5.4: *Parameters for the Case 3.*

Domain	Rectangular section of 12000 m $\times$ 85 m
Source location	Centered on the upper boundary
	at 6600 m $\leq x \leq$ 7000 m
Sink location	at x = 3300 m
Permeability	3 zones
zone 1 ($75 \leq z \leq 85$)	
K_{sx}	0.30 m/d
K_{sz}	0.05 m/d
zone 2 ($40 \leq z \leq 75$)	
K_{sx}	1. m/d
K_{sz}	1. m/d
zone 3 ($0 \leq z \leq 40$)	
K_{sx}	0.1 m/d
K_{sz}	0.05 m/d
Elastic storage S_s	1.e-2 m^{-1}
Density ratio, ϵ	0.035
Viscosity ratio, ϵ'	0
Porosity ϕ	0.025
Dispersivity α_L	10 m
Dispersivity α_T	1 m
Diffusion coefficient D_0	0 m^2/d
BCs for flow	q_N= 8.64e-4 m^2/d on the faces AB, BC, DE
	q_N = 1.21e-2 m^2/d at the source
	frewshwater head ψ=92 m at point B
	frewshwater head ψ=85 m at point E
	zero flux elsewhere
BCs for mass transport	$c = 1$ at the source
	$c = 0$ elsewhere at top of the domain
IC for pressure head	from a steady state flow simulation
IC for concentration	$c = 0$
Grid characteristics	19200 triangles (1200 $\times$ 16) and 10217 edges
Horizontal nodal spacing, Δx	20 m
Vertical nodal spacing, Δz	5.3125 m
Time increment, Δt	2 d
Convergence criteria, tol_ψ, tol_c	1.e-3

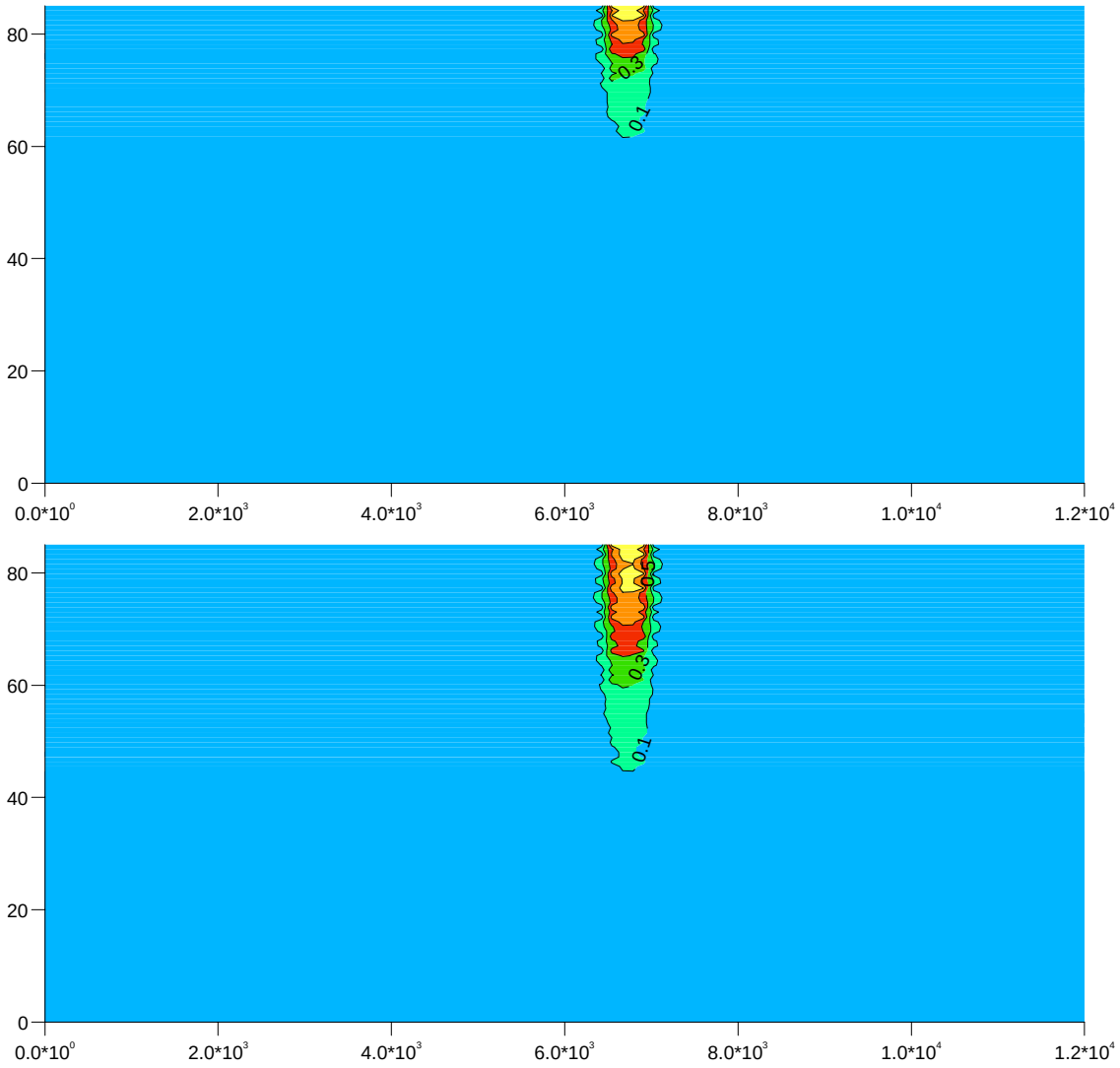

FIGURE 5.13: *Case 3: concentration contours at 1 (top) and 2 years (bottom).*

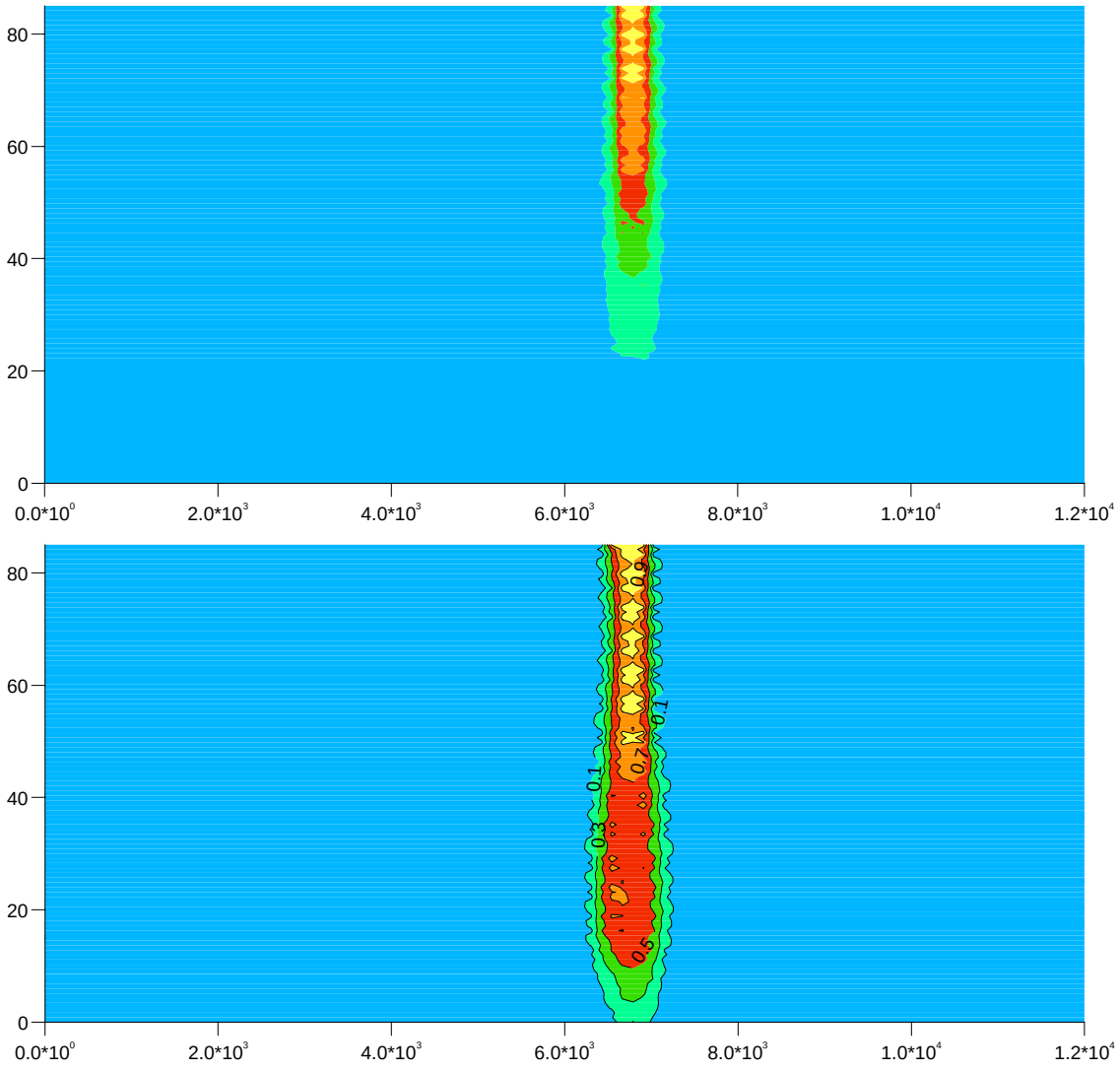

FIGURE 5.14: *Case 3: concentration contours at 4 (top) and 8 years (bottom).*

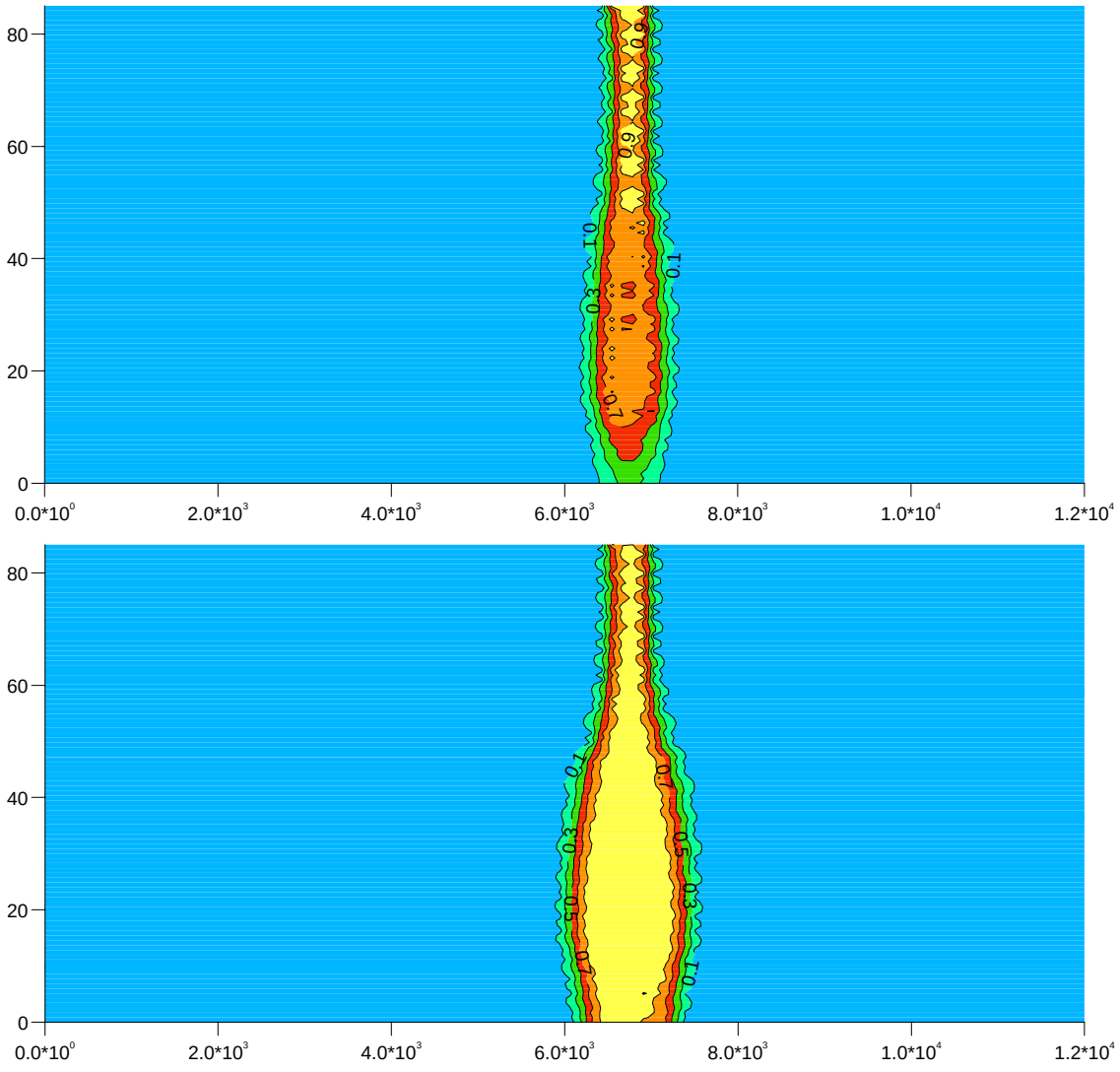

FIGURE 5.15: *Case 3: concentration contours at 10 (top) and 15 years (bottom).*

A About norms and convergence rates

If a numerical scheme has order of convergence p, then the error $|e_\ell|$ (for simplicity we consider the dependence on gridsize h) can be written as $|e_\ell| = \mathcal{C}h_\ell^p$ and $|e_{\ell+1}| = \mathcal{C}h_{\ell+1}^p$, where $h_{\ell+1} = \dfrac{h_\ell}{2}$. Therefore

$$\frac{|e_\ell|}{|e_{\ell+1}|} = \frac{h_\ell^p}{h_{\ell+1}^p} = \frac{h_\ell^p}{\frac{h_\ell^p}{2^p}} = 2^p.$$

and the order of convergence of the scheme is calculated numerically as:

$$\frac{\log \dfrac{|e_\ell|}{|e_{\ell+1}|}}{\log 2} = p \tag{1.1}$$

where $|e_\ell|$ and $|e_{\ell+1}|$ represent the norms of the error at two consecutively refined grid levels ℓ and $\ell + 1$.

In all our simulations we have used L_1 and L_2 relative norms, by comparing the results obtained by these two norms. Thus we have considered the following formulae:

- L_1 relative norm

$$|e_\ell| = \frac{\sum_{l=1}^{m} |c(x_l, t^k) - c_l^k|}{\sum_{l=1}^{m} |c(x_l, t^k)|};$$

- L_2 relative norm

$$|e_\ell| = \frac{\sqrt{\sum_{l=1}^{m} (c(x_l, t^k) - c_l^k)^2}}{\sqrt{\sum_{l=1}^{m} c(x_l, t^k)^2}},$$

where $c(x_l, t^k)$ is the analytical solution on the centroid of triangle T_l at time t^k and c_l^k is the corresponding numerical solution.

The corresponding absolute norms are:

- L_1 absolute norm

$$|e_\ell| = \sum_{l=1}^{m} |(c(x_l, t^k) - c_l^k)||T_l|;$$

- L_2 absolute norm

$$|e_\ell| = \sqrt{\sum_{l=1}^{m} (c(x_l, t^k) - c_l^k)^2 |T_l|},$$

where $|T_l|$ is the area of T_l.

A.1 In presence of fixed domain

Using absolute or relative norms to compute errors and, consequently, the rates of convergence, it is irrelevant when the domain does not change shape. As an example, we consider again the advection equation solved with Finite Volume scheme

$$u_t + \vec{\nabla} \cdot (\vec{v}u) = 0,$$

with $\vec{v} = (-1,\, 0)$, initial condition

$$u^0(x, y) = \sin(2\pi x) \sin(4\pi y/\sqrt{3}),$$

and a square domain $[0, 1] \times [0, 1]$ discretized using right triangular elements. The coarsest mesh ($\ell = 1$) is characterized by 200 triangles and 121 edges, while the finest level ($\ell = 4$) is characterized by 12800 triangles and 6561 edges. Therefore, at each level, the area of the domain does not change. To advance in time we use Euler, Midpoint and Runge-Kutta scheme, respectively. In Table A.1 we show the error norms and convergence rates obtained with both absolute and relative norms at time $t = 0.1$ s. Since the L_1 norm is bounded above by the L_2 norm, there is a difference in the convergence rate (L_1 rate is greater than the L_2 rate), but there is no difference between absolute or relative norms.

A.2 In presence of shrinking domain

If we consider a case of shrinking domain, the situation changes completely.

Let us consider the simple diffusive problem solved by the MHFE method,

$$c_t - Dc_{xx} = 0,$$

with appropriate initial and boundary conditions so that its analytical solution is given by

$$c(x, t) = erfc\frac{x}{2\sqrt{Dt}}.$$

Table A.2 reports the values obtained in L_1 and L_2 absolute and relative norms, by using Euler or Crank-Nicolson rule in time, respectively.

From the theory, we know that the scheme is of second order accuracy in the centroids of the triangles and of first or second order accuracy in time, if we use Euler or Crank-Nicolson schemes, respectively. Thus, halving at each level the gridsizes of the spatial and temporal meshes, we expect first or second global order of accuracy.

From Table A.2, the rates of convergence follow the theory when computed from the relative errors (there is no difference if we consider L_1 or L_2 relative norms), but not the absolute norms. The rates of convergence are greater than the expected once by a factor one in L_1 absolute norm and of 0.5 in L_2 absolute norm. This behavior is due to the fact that we are considering a shrinking domain at each level. In fact, since we are considering a one-dimensional example in a two-dimensional scheme, the height of the domain is always halved in passing from a coarser to the next finer level, in such a way that the shape of the triangles at the different levels is preserved. Therefore, the coarsest mesh is defined on the rectangle $[0, 1] \times [0, 0.1]$ and is characterized by 300 triangles and 204 edges, while the

TABLE A.1: *Convergence rates in the case of fixed domain.*

	Eu		Mp		RK	
ℓ	L_1 abs	rate	L_1 abs	rate	L_1 abs	rate
1	4.66e-2		4.69e-2		4.62e-2	
2	1.44e-2	1.69	1.42e-2	1.72	1.46e-2	1.66
3	5.06e-3	1.51	4.92e-3	1.53	4.99e-3	1.55
4	1.71e-3	1.56	1.62e-3	1.60	1.60e-3	1.64
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.58		1.61		1.61
ℓ	L_1 rel	rate	L_1 rel	rate	L_1 rel	rate
1	1.20e-1		1.20e-1		1.19e-1	
2	3.70e-2	1.70	3.64e-2	1.72	3.64e-2	1.66
3	1.30e-2	1.51	1.26e-2	1.53	1.28e-2	1.55
4	4.40e-3	1.56	4.18e-3	1.59	4.12e-3	1.63
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.58		1.61		1.61
ℓ	L_2 abs	rate	L_2 abs	rate	L_2 abs	rate
1	5.67e-2		5.66e-2		5.61e-2	
2	1.89e-2	1.58	1.84e-2	1.62	1.89e-2	1.57
3	7.21e-3	1.39	6.97e-3	1.40	7.12e-3	1.41
4	2.80e-3	1.36	2.66e-3	1.39	2.62e-3	1.44
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.44		1.46		1.47
ℓ	L_2 rel	rate	L_2 rel	rate	L_2 rel	rate
1	1.17e-1		1.17e-1		1.16e-1	
2	3.90e-2	1.58	3.80e-2	1.62	3.81e-2	1.56
3	1.49e-2	1.39	1.44e-2	1.40	1.47e-2	1.41
4	5.79e-3	1.36	5.50e-3	1.39	5.41e-3	1.44
		l.s.a. rate		l.s.a. rate		l.s.a. rate
		1.44		1.46		1.47

finest level ($\ell = 5$) is defined on the rectangle $[0,1] \times [0, 6.25 \times 10^{-3}]$ and is characterized by 4800 triangles and 3204 edges.

When we compute errors in the L_1 absolute norm, $|e_\ell|/|e_{\ell+1}|$ is given by

$$\frac{|e_\ell|}{|e_{\ell+1}|} = \frac{h^p|\Omega|}{(h/2)^p|\Omega/2|}.$$

where $|\Omega|$ is the area of the domain, halved in the next finer level. Thus,

$$\frac{|e_\ell|}{|e_{\ell+1}|} = 2^p 2$$

so that

$$\log \frac{|e_\ell|}{|e_{\ell+1}|} / \log 2 = p + 1.$$

We do not obtain p directly! In the case of L_2 absolute norm, we have

$$\frac{|e_\ell|}{|e_{\ell+1}|} = \frac{h^p\sqrt{|\Omega|}}{(h/2)^p\sqrt{|\Omega/2|}}.$$

and

$$\frac{|e_\ell|}{|e_{\ell+1}|} = 2^p\sqrt{2}.$$

Therefore

$$\log \frac{|e_\ell|}{|e_{\ell+1}|} / \log 2 = p + \frac{1}{2}.$$

TABLE A.2: *Convergence rates in the case of shrinking domain.*

	Eu		CN	
ℓ	L_1 abs	rate	L_1 abs	rate
1	2.27e-5		1.53e-5	
2	3.77e-6	2.59	1.90e-6	3.01
3	7.05e-7	2.42	2.38e-7	3.00
4	1.47e-7	2.26	3.00e-8	2.99
5	3.30e-8	2.15	3.84e-9	2.90
		l.s.a. rate		l.s.a. rate
		2.35		2.99
ℓ	L_1 rel	rate	L_1 rel	rate
1	1.42e-3		9.57e-4	
2	4.73e-4	1.58	2.39e-4	2.00
3	1.77e-4	1.42	5.97e-5	2.00
4	7.35e-5	1.27	1.50e-5	1.99
5	3.31e-5	1.15	3.85e-6	1.96
		l.s.a. rate		l.s.a. rate
		1.35		1.99
ℓ	L_2 abs	rate	L_2 abs	rate
1	1.09e-4		7.36e-5	
2	2.57e-5	2.08	1.30e-5	2.50
3	6.79e-6	1.92	2.29e-6	2.50
4	1.99e-6	1.77	4.05e-7	2.50
5	6.34e-7	1.65	7.19e-8	2.49
		l.s.a. rate		l.s.a. rate
		1.85		2.50
ℓ	L_2 rel	rate	L_2 rel	rate
1	1.13e-3		7.61e-4	
2	3.76e-4	1.59	1.90e-4	2.00
3	1.40e-4	1.42	4.74e-5	2.00
4	5.84e-5	1.26	1.19e-5	1.99
5	2.62e-5	1.16	2.98e-6	2.00
		l.s.a. rate		l.s.a. rate
		1.35		2.00

B Groundwater modeling

Groundwater modeling may be defined, briefly, as the simulation of the current behavior of an aquifer and the prediction of its future conditions. An exhaustive introduction and analysis can be found, for example, in [1, 2]. In the following we give some rudimentary concepts useful to better understand the coupled flow and transport equations.

- An *aquifer* is a geological formation which contains water and permits significant amounts of water to move through it under ordinary field conditions. The word aquifer comes from the latin words *aqua* and *ferre*, i.e., *to bear water*.

- That portion of the rock formation which is occupied by solid matter is called the *solid matrix*. The remaing part is called the *void space*. The void space is occupied by one (water) or two (water and air) fluid phases.

- Subsurface formations containing water may be divided vertically into two horizontal zones, according to the relative proportion of the void space which is occupied by water:

  - *zone of saturation*: the entire void space is filled with water;
  - *unsaturated zone*: the pores contain both gases and water.

  The unsaturated zone is overlying the zone of saturation.

- A porous medium domain is said to be *homogeneous* with respect to its permeability, if the latter is the same at all its point. Otherwise the domain is said to be *heterogeneous*. If the permeability at a considered point is independent of direction, the porous medium is *isotropic* at that point. Otherwise, the porous medium is *anisotropic*.

B.1 Darcy's law

In 1856, Henry Darcy investigated the flow of water in vertical homogeneous sand filters in connection with the fountains of the city of Dijon. From his experiments, Darcy concluded that the rate of flow (i.e. volume of water per unit time), Q, is:

1. proportional to the cross-sectional area, say A;

2. proportional to the difference in water level elevations in the inflow and exit reservoirs of the filter, say $(h_1 - h_2)$;

3. inversely proportional to the filter's lenght, say L.

Therefore, the first form of Darcy's law is:

$$Q = \frac{KA(h_1 - h_2)}{L},$$

where K is a constant. Darcy's law extended to flow through an inclined homogeneous porous medium column takes the form:

$$Q = \frac{KA(h_1 - h_2)}{L},$$

where now h represents the piezometric head defined by $h = \psi + z$, as seen in Chapter 5, with z the elevation of the point. The piezometric head expresses the sum of the potential energy ad pressure energy, per unit weight of water, while $\frac{h_1 - h_2}{L}$ is the hydraulic gradient. Denoting this gradient by J and defining the specific discharge, v, as the volume of water flowing per unit time through a unit cross-sectional area normal to the direction of flow, we obtain

$$q = KJ,$$

where $q = Q/A$.

Let us consider a point along the column's axis and a segment of the column of length s along the column's axis, on both sides of the point. Darcy's law becomes:

$$q_s = -K\frac{\partial h}{\partial s} \equiv KJ_s \qquad J_s \equiv \frac{\partial h}{\partial s}.$$

This expression gives the component of the specific discharge in the direction s at any point in a porous medium domain, given K and the spatial distribution of the piezometric head. By considering the volumetric porosity ϕ, it can be shown that the average areal porosity is equal to ϕ and, thus, the portion of the area A available to flows is ϕA. Accordingly, the average velocity, v, of the flow through the column is given by

$$v = \frac{Q}{\phi A} = \frac{q}{\phi}.$$

We are now able to generalize Darcy's law when the flow is three-dimensional:

$$\vec{q} = K\vec{J} = -K\vec{\nabla}h \qquad \vec{v} = \vec{q}/\phi,$$

where $\vec{v}$ is the velocity vector, $\vec{q}$ is the specific discharge vector, $\vec{J} = -\vec{\nabla}h$ is the hydraulic gradient.

B.2 Hydraulic conductivity

The coefficient of proportionality K appearing in Darcy's law is the hydraulic conductivity of the porous medium (it is the same K_s parameter that appears in the coupled flow and transport equations). In an isotropic medium it may be defined as the specific discharge per unit hydraulic gradient. It expresses the ease with which a fluid is transported through the tortuous void space. It is therefore a coefficient that depends on both matrix and fluid properties. The relevant properties are the density, ρ, and viscosity, μ. The relevant solid matrix properties are mainly grain- (or pore-) size distribution, shape of grains (or pores), tortuosity, specific surface, and porosity.
The hydraulic conductivity K may be expressed as

$$K = \frac{k\rho g}{\mu},$$

where g is the acceleration of gravity and where k - called the permeability of the porous medium - depends solely on the properties of the solid matrix.

B.3 Moisture content and saturation

In unsaturated flow, the void space is partly filled by air and partly by water. Two state variables may be used to define the relative quantity of water at a certain time at a point in a porous medium domain:

$$\theta_w = \frac{\text{volume of water}}{\text{total volume}} \qquad 0 \leq \theta_w \leq \phi$$

$$S_w = \frac{\text{volume of water}}{\text{volume of voids}} \qquad 0 \leq S_w \leq 1$$

θ_w is called the water (or moisture) content and S_w is the water saturation. In this way, $\theta_w = \phi S_w$.

B.4 Molecular diffusion and mechanical dispersion

Molecular diffusion, caused by the random movement of molecules in a fluid, produces an additional flux of tracer particles (at microscopic level) from regions of higher tracer concentration to that of lower concentration. In addition to the role played at the microscopic level by molecular diffusion in enhancing the transversal component of mechanical dispersion, it produces a macroscopic flux of its own.

We refer to the spreading caused by the velocity at the microscopic level, enhanced by molecular diffusion, as macroscopical or hydrodynamic dispersion.

The dispersion tensor D of transport equation, studied in Chapters 4 and 5, takes into account for mechanical dispersion and molecular diffusion.

References

[1] J. Bear, *Hydraulics of Groundwater*, McGraw-Hill, New York, 1979.

[2] J. Bear and A. Verruijt, *Modeling Groundwater Flow and Pollution*, Reidel Publications, Dordecht, Holland, 1987.

[3] L. Bergamaschi, C. Gallo, G. Manzini, C. Paniconi, and M. Putti, *A mixed finite element/TVD finite volume scheme for saturated flow and transport in groundwater*, in Finite Elements in Fluids, M. Morandi Cecchi, K. Morgan, J. Periaux, B. A. Schrefler, and O. C. Zienkiewicz, eds., University of Padua, Italy, 1995, Dip. Matematica Pura e Applicata, pp. 1223–1232.

[4] L. Bergamaschi and M. Putti, *Mixed finite elements and Newton-type linearization for the solution of Richard's equation*, Int. J. Numer. Methods Engrg., 45 (1999), pp. 1025–1046.

[5] A. C. Berkenbosch, E. Kaasschieter, and J. ten Thije Boonkkamp, *Finite-difference methods for one-dimensional hyperbolic conservation laws*, Num. Meth. PDE, 8 (1992), pp. 267–276.

[6] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, Berlin, 1991.

[7] V. Casulli, *Eulerian-Lagrangian methods for hyperbolic and convection dominated parabolic problems*, in Computational Methods for Nonlinear Problems, Swansea, 1987, Pineridge, pp. 239–269.

[8] J. Cole, *On a quasilinear parabolic equation occurring in aerodynamics*, Q. Appl. Math., 9 (1951), pp. 225–236.

[9] C. N. Dawson, *Godunov-mixed methods for immiscible displacement*, Int. J. Numer. Methods Fluids, 11 (1990), pp. 835–847.

[10] ——, *Godunov-mixed methods for advective flow problems in one space dimension*, SIAM J. Num. Anal., 28 (1991), pp. 1282–1309.

[11] ——, *Godunov-mixed methods for advection-diffusion equations in multidimensions*, SIAM J. Num. Anal., 30 (1993), pp. 1315–1332.

[12] ——, *High resolution upwind-mixed finite element methods for advection diffusion equations with variable time-stepping*, Num. Meth. PDE, 11 (1995), pp. 525–538.

[13] J. Douglas Jr. and J. E. Roberts, *Global estimates for mixed methods for second order elliptic equations*, Math. Comp., 44 (1985), pp. 39–52.

[14] T. F. Dupont and P. T. Keenan, *Superconvergence and postprocessing of fluxes from lowest order mixed methods on triangles and tetrahedra*, SIAM J. Sci. Comput., 19 (1998), pp. 1322–1332.

[15] L. J. Durlofsky, *A triangle based mixed finite element-finite volume technique for modeling two phase flow through porous media*, J. Comp. Phys., 105 (1993), pp. 252–266.

[16] L. J. Durlofsky, B. Engquist, and S. Osher, *Triangle based adaptive stencils for the solution of hyperbolic conservation laws*, J. Comp. Phys., 98 (1992), pp. 64–73.

[17] R. E. Ewing, R. D. Lazarov, and J. Wang, *Superconvergence of the velocity along the gauss lines in mixed finite element methods*, SIAM J. Num. Anal., 28 (1991), pp. 1015–1029.

[18] P. A. Forsyth, *A control volume finite element approach to NAPL groundwater contamination*, SIAM J. Sci. Stat. Comput., 12 (1991), pp. 1029–1057.

[19] E. O. Frind, *Simulation of long-term transient density-dependent transport in groundwater*, Adv. Water Resources, 5 (1982), pp. 73–88.

[20] C. Gallo, *Brine-driven transport of radioactive contaminants: a one-dimensional numerical investigation*, technical report, CRS4 Environment Group, March 1999.

[21] C. Gallo and E. Leonardi, *Upgrade of the CODESA-3D density-dependent simulator to the strongly-coupled (brine) case*, technical report, CRS4 Environment Group, March 1999.

[22] G. Gambolati, C. Paniconi, and M. Putti, *Numerical modeling of contaminant transport in groundwater*, in Migration and Fate of Pollutants in Soils and Subsoils, D. Petruzzelli and F. G. Helfferich, eds., vol. 32 of NATO ASI Series G: Ecological Sciences, Berlin, 1993, Springer-Verlag, pp. 381–410.

[23] G. Gambolati, G. Pini, M. Putti, and C. Paniconi, *Finite element modeling of the transport of reactive contaminants in variably saturated soils with LEA and non-LEA sorption*, in Environmental Modeling, Vol. II: Computer Methods and Software for Simulating Environmental Pollution and its Adverse Effects, P. Zannetti, ed., Computational Mechanics Publications, Southampton, UK, 1994, ch. 7, pp. 173–212.

[24] G. Gambolati, M. Putti, and C. Paniconi, *Three-dimensional model of coupled density-dependent flow and miscible salt transport in groundwater*, in Seawater Intrusion in Coastal Aquifers; Concepts, Methods and Practices, J. Bear, A. H.-D. Cheng, S. Sorek, D. Ouazar, and I. Herrera, eds., Kluwer Academic Publ., Dordrecth, The Netherlands, 1999, ch. 10, pp. 315–362.

[25] E. Godlewski and P. Raviart, *Numerical Approximation of Hyperbolic Systems of Conservation Laws*, Springer-Verlag, New York, 1996.

[26] A. B. Gureghian, *TRIPM, a two-dimensional finite element model for the simultaneous transport of water and reacting solutes through saturated and unsaturated porous media*, Technical Report ONWI-465, Off. of Nuclear Water Isolation, Columbus, Ohio, 1983.

[27] B. Gustaffson, H. Kreiss, and J. Oliger, *Time Dependent Problems and Difference Methods*, John Wiley & Sons, Inc., New York, 1995.

[28] C. Hirsch, *Numerical Computations of Internal and External Flows (Vol. 2)*, John Wiley & Sons, Inc., New York, 1988.

[29] E. Hopf, *The partial differential equation* $u_t + uu_x = \delta u_{xx}$, Commun. Pure Appl. Math., 3 (1950), pp. 201–230.

[30] P. S. Huyakorn, S. D. Thomas, and B. M. Thompson, *Techniques for making finite elements competitive in modeling flow in variably saturated porous media*, Water Resour. Res., 20 (1984), pp. 1099–1115.

[31] IEG.HFMteam, *Partner progess report (RaCoS - second year)*, technical report, Inst. of Environmental Geology of Russian Academy of Science (St. Petersburg Division), March 1999.

[32] E. Leonardi and C. Gallo, *Preliminary study on density-dependent flow and transport in the lake karachai area (UE RaCoS project): adequacy of CODESA-3D to model a plume of contaminant brine*, technical report, CRS4 Environment Group, February 1999.

[33] R. LeVeque, *Numerical Methods for Conservation Laws*, Birkhäuser Verlag, Zurich, 1992.

[34] X.-D. Liu, *A maximum principle satisfying modification of triangle based adaptive stencils for the solution of scalar hyperbolic conservation laws*, SIAM J. Num. Anal., 30 (1993), pp. 701–716.

[35] A. Mazzia, L. Bergamaschi, and M. Putti, *A time-splitting technique for advection-dispersion equation in groundwater*, J. Comput. Phys., 157 (2000), pp. 181–198.

[36] ——, *Triangular finite volume-mixed finite element discretization for the advection-diffusion equation*, in Large Scale Scientific Computations of Engineering and Environmental Sciences, M. Griebel, S. Margenov, and P. Yalamov, eds., Braunschweig (Ger), 2000, Vieweg, pp. 371–378.

[37] ——, *A second order time-splitting technique for advection-dispersion equation on unstructured grids*, in Godunov Methods Theory and Applications, E. F. Toro, ed., vol. 1, New York, 2001, Academic/Plenum Publishers, pp. 603–610.

[38] M. Nakata, A. Weiser, and M. F. Wheeler, *Some superconvergence results for mixed finite element methods for elliptic problems on rectangular domains*, in The Mathematics of Finite Elements and Applications V, J. Whiteman, ed., London, 1985, Academic Press.

[39] S. P. Neuman, *A Eulerian-Lagrangian numerical scheme for the dispersion convection equation using conjugate space time grids*, J. Comp. Phys., 41 (1981), pp. 270–294.

[40] S. Osher and S. Chakravarthy, *High resolution schemes and the entropy condition*, SIAM J. Num. Anal., 21(5) (1984), pp. 955 – 984.

[41] C. Paniconi and M. Putti, *Modeling variably saturated flow problems using Newton-type linearization methods*, in Advanced Methods for Groundwater Pollution Control, G. Gambolati and G. Verri, eds., CISM Courses and Lectures NO. .364, Wien, New York, 1995, Springer-Verlag, pp. 45–64.

[42] R. Peyret and T. D. Taylor, *Computational Methods for Fluid Flow*, Springer-Verlag, New York, 1983.

[43] M. Putti and C. Cordes, *Finite element approximation of the diffusion operator on tetrahedra*, SIAM J. Sci. Comput., 19 (1998), pp. 1154–1168.

[44] M. Putti and C. Paniconi, *Finite element modeling of saltwater intrusion problems with an application to an Italian coastal aquifer*, in Advanced Methods for Groundwater Pollution Control, G. Gambolati and G. Verri, eds., CISM Courses and Lectures NO. .364, Wien, New York, 1995, Springer-Verlag, pp. 65–84.

[45] M. Putti and C. Paniconi, *Picard and Newton linearization for the coupled model of saltwater intrusion in aquifers*, Adv. Water Resources, 18 (1995), pp. 159–170.

[46] M. Putti, W. W.-G. Yeh, and W. A. Mulder, *A triangular finite volume approach with high resolution upwind terms for the solution of groundwater transport equations*, Water Resour. Res., 26 (1990), pp. 2865–2880.

[47] A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*, vol. 23 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1994.

[48] RaCoS, *Radionuclide Contamination of Soils and Groundwater at the Lake Karachai Waste Disposal Site (Russia) and the Chernobyl Accident Site (Ukraine): Field Analysis and Modeling Study*, Summary Description of the Joint Research Proposal, (1996).

[49] ——, *Radionuclide Contamination of Soils and Groundwater at the Lake Karachai Waste Disposal Site (Russia) and the Chernobyl Accident Site (Ukraine): Field Analysis and Modeling Study*, First year progress Report, (1998).

[50] ——, *Radionuclide Contamination of Soils and Groundwater at the Lake Karachai Waste Disposal Site (Russia) and the Chernobyl Accident Site (Ukraine): Field Analysis and Modeling Study*, Second year progress Report, (1999).

[51] P. A. Raviart and J. M. Thomas, *A mixed finite element method for second order elliptic problems*, in Mathematical Aspects of the Finite Elements Method, I. Galligani and E. Magenes, eds., no. 606 in Lecture Notes in Mathematics, New York, 1977, Springer-Verlag.

[52] P. L. Roe, *Some contributions to the modelling of discontinuous flows*, Lectures Appl. Math., 22 (1985), pp. 163–193.

[53] G. Ségol, *Classic Groundwater Simulations*, Prentice-Hall, Inc., Englewood Cliff, NJ 07632, 1994.

[54] G. A. Sod, *Numerical Methods in Fluid Dynamics*, Cambridge University Press, New York, 1985.

[55] P. K. Sweby, *High resolution schemes using flux limiters for hyperbolic conservation laws*, SIAM J. Num. Anal., 21 (1984), pp. 995–1011.

[56] C. Tai, D. Chiang, and Y. Su, *Explicit time marching methods for the time-dependent Euler computations*, J. Comp. Phys., 130 (1997), pp. 191–202.

[57] E. F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Springer-Verlag, Berlin, 1997.

[58] G. D. van Albada, B. van Leer, and W. W. Roberts Jr., *A comparative study of computational methods in cosmic gas dynamics*, Astron. Astroph., 108 (1982), pp. 76–84.

[59] B. van Leer, *Towards the ultimate conservative difference scheme. V. a second-order sequel to Godunov's method*, J. Comp. Phys., 32 (1979), pp. 101–163.

[60] C. I. Voss and W. R. Souza, *Variable density flow and solute transport simulation of regional aquifers containing a narrow freshwater-saltwater transition zone*, Water Resour. Res., 23 (1987), pp. 1851–1866.

[61] A. Weiser and M. F. Wheeler, *On convergence of block-centered finite differences for elliptic problems*, SIAM J. Num. Anal., 25 (1988), pp. 351–375.